Chapter 5

Discussion

In this dissertation, I have described three projects that take genetic or phenotypic approaches to understanding developmental disorders using data from next generation sequencing. In Chapter 2, I investigated the phenotype of Wiedemann-Steiner syndrome (WSS) resulting from *KMT2A* mutations, and aimed to define the phenotypic and mutational spectrum. I identified, collected and analysed standardised data from 84 individuals with WSS and *KMT2A* mutations. To my knowledge this is the largest cohort reported in the world to date. I defined the *KMT2A* mutational spectrum and provided the first detailed evaluation of the features associated with *KMT2A*-associated WSS in a cohort more than 15 times larger than the largest previous report (N=5 individuals). I defined the growth pattern and demonstrated that not all individuals with WSS have hypertrichosis. I reported new features of WSS (including genital abnormalities in females) and highlighted that epilepsy is as an important feature of WSS that is associated with poorer developmental outcomes in WSS individuals. I reported the first somatic mosaic individual with a *KMT2A* mutation with a milder clinical phenotype.

My next aim was to investigate how missense mutations affect KMT2A function. I showed that missense mutations cluster within the recognized domains of KMT2A, including the zinc-finger and zinc-binding domains and proposed disease mechanisms for these mutations, including preventing DNA binding. Finally, I aimed to investigate whether WSS caused by *KMT2A* mutations has a recognisable facial appearance. I showed that the facial appearance of individuals with WSS is distinguishable from that of other individuals with developmental disorders by carrying out a facial recognition experiment with experienced Genetics Clinicians. This investigation has significantly advanced the knowledge and understanding of WSS caused by *KMT2A* mutations. The knowledge gained will help clinicians identify and manage individuals with WSS in the future and ultimately improve the medical care of these individuals, who have multiple medical needs.

In Chapter 3, my first aim was to investigate the genetic basis of developmental disorders associated with hypertrichosis using whole exome sequencing. I identified 247 individuals with developmental disorders including hypertrichosis or with WSS or a condition similar to WSS and analysed their exome variant profiles. The *de novo* diagnostic yield from my cohort was 29%, which is higher than the *de novo* diagnostic

yield of the DDD study more generally of 23%. However, a significant proportion of this yield was *KMT2A* mutations and removing those individuals gave a diagnostic yield of 21%. My next aim was to seek evidence of a burden of variants in genes that play a role in maintaining the structure or function of chromatin (chromatin genes). I showed that the DDD hypertrichosis cohort was significantly enriched for variants in chromatin genes. My findings suggest that hypertrichosis is an important signal that an individual could have a chromatin disorder and may be more likely to harbour a diagnostic *de novo* mutation than individuals with developmental disorders more generally. This investigation also highlighted known disease genes implicated in hypertrichosis, which could enable gene panels (for phenotype-focused next generation sequencing) to be curated for individuals with hypertrichosis.

I next sought to identify new genes implicated in developmental disorders associated with hypertrichosis. I identified two individuals with identical missense mutations in *ZMYND11*, suggesting that *ZMYND11,* a well-recognized developmental disorder gene, is specifically associated with hypertrichosis. Although I didn't identify any new genes, I showed in principle this approach can successfully identify disease genes associated with hypertrichosis and will pave the way for further research into the genetic architecture of hypertrichosis using larger cohort sizes in the future.

In Chapter 4, I aimed to investigate the underlying architecture of severe developmental disorders by seeking out evidence of autosomal recessive disease using a population matched control dataset. I generated a novel control dataset of untransmitted diplotypes and analysed 1,080 non-consanguineous trios with developmental disorders in the DDD Study. By the use of the untransmitted diplotypes, I showed that there is a genome wide trend towards over transmission of very rare (MAF < 0.05%) LoF variants to DDD probands, giving evidence that inherited variants contribute to developmental disorders in the DDD study cohort. In addition, by separating out the individuals with a likely dominant cause of their disorder (dominant probands) I showed an enrichment of rare (MAF <5%) biallelic loss of function (LOF) variants in known developmental disorder genes in non-dominant probands compared to dominant probands, providing evidence of recessive disease in the non-dominant probands. To my knowledge, my work using the

untransmitted diplotypes gives the first insight into the contribution of autosomal recessive disease to developmental disorders by studying untransmitted alleles.

My final aim was to contribute to the significant improvement of the diagnosis of children with developmental disorders as a clinician researcher working as a member of the DDD study analysis team. My analyses, clinical knowledge and role in clinical reporting contributed significantly to the DDD study, which has shaped modern day clinical genetics knowledge and practice. Through analysis of 1,133 trios, 31% of probands and their families received a diagnosis for their disorder, and 12 novel genes associated with developmental disorders were discovered. A case-control analysis looking for evidence of mosaicism in 1303 DDD trios, identified 12 structural mosaic abnormalities (0.9%) that were reported back to local clinicians, 10 of which were assessed as highly likely to be pathogenic in causing the individual's developmental disorder. In further analysis of analysis of 4293 trios, the DDD study identified four new genes implicated in recessive diseases and discovered 14 new dominant disease genes. Many of the aspects of the DDD study have been incorporated into modern day clinical genetics practice; for example, the DDG2P is used in Clinical Genetics laboratories throughout the UK. Also, multi-disciplinary meetings to review whole exome sequencing findings, as pioneered by the DDD study, form an important part of the week for a number of Clinical Genetics departments.

The three projects I have presented in this dissertation have four important outcomes that increase knowledge of developmental disorders and will ultimately help individuals with rare diseases and their family members in the future. 1. I have significantly increased knowledge of the phenotypic and mutational spectrum of the rare disorder Wiedemann-Steiner syndrome. This will ultimately improve patient identification, diagnosis and medical care. 2. I have demonstrated that there is a burden of variants in chromatin genes in individuals with hypertrichosis. This suggests that hypertrichosis is an important indicator that an individual could have a chromatin disorder and may be more likely to harbour *de novo* mutation than other individuals with developmental disorders. This knowledge and knowledge of the genes implicated will help in the diagnosis of individuals with hypertrichosis in the future, and also will enable gene panels (for selective next generation sequencing) to be curated for individuals with

hypertrichosis. 3. My work using the untransmitted diplotypes has increased knowledge of the architecture of developmental disorders through studying non-transmitted alleles. I showed that inherited variants are contributing to developmental disorders in the DDD study cohort. Additionally, I found evidence for recessive disease in DDD study individuals by identifying a burden of biallelic loss of function variants in DDD non-dominant probands. 4. I have played a key role as an analyst and clinician researcher in the DDD study which has shaped modern day clinical genetics knowledge and practice.

In conclusion, I have described three investigations that take genetic or phenotypic approaches to understanding developmental disorders using data from next generation sequencing. The themes running throughout this dissertation are dominant versus recessive inheritance, loss of function versus missense variants, the use of next generation sequencing to unravel the underlying causes of developmental disorders and challenges in assigning pathogenicity to variants. Many of these themes are current key issues of Clinical Genetics more widely in the whole exome sequencing era. In the future, further understanding of the architecture, genotypes and phenotypes of developmental disorders will be driven by larger sample sizes, standardisation of phenotype terms, online portals to input and share genotype and phenotype data and large population control databases. Variant interpretation will remain a challenge for many years to come, particularly the understanding of missense variants. Ideally our mutation and control databases would together cover every variant or there would be a valid and reliable functional assay for every disease. However, until this point, a sensible and practical approach to managing patients with variants of uncertain significance is vital and assessment for dysmorphological features will continue to play an important role in variant interpretation. Though understanding more about developmental disorders, we are in a stronger position to manage patients more effectively and develop treatments, ultimately helping individuals with rare diseases and their families.