Chapter 1

Introduction

1.1 Definition, incidence and impact

Developmental disorders are a diverse group of conditions that result in abnormal human development. They demonstrate variability, both within a single disorder and across different types of disorder. Some are life limiting, painful, severely debilitating or degenerative. Developmental disorders may be associated with congenital abnormalities, for example heart or brain malformations or with neurological, cognitive or behavioral phenotypes, for example hypotonia, delayed developmental milestones, intellectual disability or autism. Some developmental disorders have phenotypes known only to affect one organ, for example *NR2F2* (MIM 107773) mutations in congenital heart disease(3). However, many developmental disorders manifest with a multitude of variable phenotypic features affecting a variety of organ systems.

Many individuals with developmental disorders have intellectual disability either as part of a syndrome or as an isolated phenotype. Intellectual disability is defined as substantial impairment of cognitive and adaptive functions that has onset in childhood(4). Severity can range from mild to profound. Developmental disorders can also result from environmental causes, for example *in utero* exposures, trauma or infection. However, many developmental disorders have a genetic cause, in fact the majority of cases of severe intellectual disability are thought to be genetic(5). However, not all rare genetic diseases (including genetic developmental disorders) have been defined and had their underlying cause elucidated. Estimates from the pre-genomics era using human genome mutation rates and the number of essential genes are that there may be around 7750 –15,300 rare-disease-causing genes(6). For those that do receive a genetic diagnosis for their or their child's developmental disorder, the time period leading to diagnosis or 'diagnostic odyssey' may take several years. A study by Rare Diseases Europe (EURODIS) of 6000 families or individuals in 17 countries affected by 8 rare diseases (the majority developmental disorders) showed that for 25% of people the time to diagnosis was 5 to 30 years(7).

Rare diseases are life-threatening or chronically debilitating diseases with low prevalence. Prevalence of rare diseases is defined as less than 1 in 20,000 people in the United States of America and less than 1 in 2000 people in Europe [Commission of the European Communities(8). In Europe it is estimated that five to eight thousand

different rare diseases affect 6-8% of the population(9) Many rare diseases are genetic developmental disorders. Although these disorders are individually rare, collectively they are common and genetic rare diseases affect at least 1 in 50 individuals(10). Some of these genetic developmental disorders can present in the neonatal period, and around a third of these infants will succumb to their rare disease in their first year of life(11-13). In addition to the effects on affected individuals and their families, intellectual disability and other developmental disorders are associated with significant morbidity and mortality and pose enormous socio-economic costs(14-16). These costs include the indirect costs of productivity losses in workplaces or households that occurs when an individual with a developmental disorder is unable to work, or are limited in the amount or type of work they could do or dies prematurely(15).

1.2 A short history of genetic developmental disorders

1.2.1 Copy number change as a cause of developmental disorders

Over time improvements in technology have increased the possible number of genetic diagnoses we are able to make. Copy number change is defined as a gain or loss of chromosomal genetic material compared to the reference human genome. Multiple studies in control populations, have shown that there is tolerance for copy-number change in some regions of the genome(17-19). All humans are estimated to carry copy number variants (CNVs) and they are thought for the most part benign and part of normal variation. However, the effect of a CNV depends on whether it changes the relative location and or sequence of genomic DNA and CNVs are a well-established cause of developmental disorders.

Chromosomal causes of developmental disorders were first identified with the identification of the presence of an extra copy of chromosome 21 in individuals with Down syndrome in 1959(20). This discovery was made on karyotype analysis, a technique honed by Tjio and colleagues who discovered in 1956 that man has 46 chromosomes(21). This was followed, by the discovery of several other chromosome imbalances, including unbalanced translocations, marker chromosomes and large deletions and duplications as the cause of developmental disorders. Karyotyping is able to detect imbalances as small as 5 to 10 Mb and these explain 10-15% of intellectual disability. Karyotyping also has the ability to detect mosaicism (more than one cell

population deriving from a single zygote) a phenomenon apparent in a diverse range of human disorders including developmental disorders(22). Chromosomal mosaicism has been detected from the earliest stages of karyotype use(23).

Developed in the 1980s, Fluorescence in situ Hybridisation (FISH) uses fluorescent labelling to detect chromosome imbalances and provided an accurate method for identifying and confirming small deletions and duplications. The discovery of FISH led to the development of methods to detect subtelomeric chromosomal deletions and duplications ((24) and reviewed by Rudd(25)), these were found to cause 2-5% of previously unexplained intellectual disability(26, 27). More recently, the technique of optical mapping has been reported to successfully detect structural chromosomal variants(28, 29). Optical mapping approaches involve the construction of ordered restriction maps from individual molecules of genomic DNA using single-molecule measurements and computational analysis(30).

Chromosome Microarrays

Chromosome microarrays have increasingly become the 1st line copy number diagnostic test in the developed world(31). Microarray technology can detect smaller gains and losses of DNA sequence than karyotype analysis with a range in length from 1000bp in size. The commonest technologies used diagnostically are array comparative genomic hybridisation (aCGH), which uses fluorescent labelling for comparison to control DNA and single-nucleotide-polymorphism (SNP) arrays which uses fluorescence to label SNPs. Due in the main part to their ability to detect submicroscopic deletions and duplications, chromosome microarrays offer higher diagnosis rates than traditional karyotyping, with 15%–20% of individuals with developmental delay, intellectual disability, autistic spectrum disorder or multiple congenital abnormities receiving a diagnosis(31). However, chromosome microarrays are unable to detect truly balanced translocations, also high-resolution arrays are not in widespread use in a clinical diagnostic setting and small exonic duplications and deletions can go undetected(32).

1.2.2 Single gene causes of developmental disorders

In the 1970s Fred Sanger and colleagues developed a new method of sequencing deoxyribonucleic acid (DNA), the dideoxy chain-termination method(33). This technique

enabled the rapid and accurate sequencing of large stretches of DNA. The introduction of 'Sanger' sequencing together with the introduction and improvement of the polymerase chain reaction (PCR)(34, 35), and the genetic map or catalogue of polymorphisms and linkage methods transformed the diagnostic and research arena for genetic diseases in the 1990s.

Much of early gene discovery in developmental disorders was driven by linkage experiments requiring multiple affected family members. As the presence of female carriers could permit pedigree analysis, this led to the discovery of many X-linked disorders, including fragile X syndrome(36) and Rett syndrome(37). Many autosomal conditions for which gene discovery was possible (neurofibromatosis, myotonic dystrophy, Noonan-spectrum disorders and tuberous sclerosis) are characterised by variable intellectual disability, which increased the possibility of being able to study multiple affected family members, because reproduction was not impaired in all affected individuals by intellectual disability.

1.2.3 Genome wide sequencing approaches in developmental disorders

Sanger sequencing is reliable and robust and was the mainstay of sequencing technology used for over 25 years. However, it is time consuming, and in the diagnostic arena affords little more than targeted sequencing of one or two genes at a time. Second (Next) generational sequencing platforms for genome wide sequencing have become widely available since 2005 and have significantly reduced the cost of DNA sequencing relative to Sanger sequencing(38). These methods carry out massively parallel sequencing of small fragments of DNA from across the entire genome or exome to deliver results rapidly. Initial successes for whole exome sequencing in the clinical arena were a proof of principle analysis in Freeman Sheldon syndrome(39) and a diagnosis of a known condition (congenital chloride diarrhea) in an individual thought to have Bartter syndrome(40). The first developmental disorder of unknown cause unraveled by whole exome sequencing was Miller syndrome which was found to be caused by rare biallelic loss of function variants in *DHODH* (MIM 126064)*(41)*. This seminal paper also illustrated the possibility of incidental or unexpected findings in genome wide sequencing by identifying variants in a ciliary gene in two individuals with

bronchiectasis, recurrent lung infections and chronic obstructive pulmonary disease. Since this time, hundreds more developmental disorder genes have been discovered.

The limitations of genome wide sequencing techniques include the inability or difficulty to detect balanced translocations and variants in repetitive regions of the genome or in regions with highly homologous sequences elsewhere in the genome.

De novo mutations are an increasingly recognized cause of developmental disorders

Exome sequencing confirmed that many undiagnosed developmental disorders result from new germline mutations in autosomal dominant genes arising between generations, known as *de novo* dominant mutations. The first clinically well recognized disorder noted to arise as a *de novo* mutation was Kabuki Make-up syndrome(42). Following this *de novo* mutations were found to underlie a number of distinctive multiple anomaly syndromes including the Say Barber Biesecker type of Ohdo syndrome, Coffin Siris syndrome, and Wiedemann-Steiner syndrome(43-46).

Large projects and consortia

With the increasing widespread use of whole exome sequencing many collaborations have been formed including the nationwide project FORGE in Canada, which aims to discover new genes and identify mutations in known genes(47). In the UK, the Deciphering Developmental Disorders (DDD) Study, is a nationwide study which uses multiple complementary genome wide approaches to decipher the underlying genetic cause of developmental disorders with a trio design(48). I discuss the DDD study in detail in Chapter 2.

1.3 Diagnosing Developmental disorders

1.3.1. Different types of diagnoses for developmental disorders

There are different types or levels of diagnosis for genetic disorders including: clinical, biochemical, genetic (molecular or cytogenetic). A clinical diagnosis means a doctor has examined the individual and has decided their phenotype fits with a certain disorder. This may be specific, i.e. they think that they fit with one particular disorder, e.g. Kabuki-

Make up syndrome, or within a spectrum of disorders, e.g. they have a ciliopathy, i.e. one of a group of disorders that results from impaired ciliary function and have shared features. A clinical diagnosis enables a recurrence risk to be given for future pregnancies. However, there is a possibility that a clinical diagnosis may be incorrect, and thus any given recurrence figures may not be accurate. It also doesn't enable other relatives to be tested for the disorder or for a specific genetic test to be carried out in future pregnancies. Clinical assessment for so called 'dysmorphic features' has played a significant role in the diagnosis and understanding of developmental disorders and making clinical diagnoses. Dysmorphic features are defined as features unusual for a person's age and ethnicity, with Dysmorphology defined as the study of human congenital malformations and syndromes. With the number and rarity of conditions involved, genetics clinicians have traditionally worked together to share knowledge and help diagnose patients. This has led to the occurrence of multiple international meetings for discussing clinical cases and viewing images, on a local, regional and on an international arena for example the Smith Dysmorphology meeting in the USA and the Manchester Dysmorphology Conference. There are also databases set up to help make diagnoses based on phenotypic features, such as the London Medical database www.lmdatabases.com(49) and Possum www.possum.net.au(50).

An international group of clinicians worked together to publish stringent standardized human morphological terms with consensus definitions (and also highlight terms not acceptable for use), illustrating each term with a photograph(51-56). The aim of this work was to increase the accuracy of discussions between dysmorphologists and other specialists such as molecular geneticists and developmental biologists. This standardized list has been accepted as the morphological terms that should be utilized by the American Journal of Human Genetics(57). Ontologies have also been developed to record standardised phenotypic terms, such as the Human Phenotype Ontology (HPO) (58).

Biochemical diagnoses of genetic diseases are most commonly achieved in metabolic disorders where they are able to quantify the enzyme defect or other metabolic disturbance to confirm the diagnosis. Sometimes biochemical testing can be carried out in pregnancy to look for recurrence of metabolic disorders, however genetic testing is the

gold standard for prenatal testing. Also genetic testing can sometimes identify specific subtypes of the metabolic disorders which may direct management as certain subtypes respond better to treatment. For some genetic conditions there are metabolic or chemical tests available that may give evidence as to the diagnosis or carrier state of an individual for example measuring creatine-kinase in duchenne muscular dystrophy or haemoglobin H inclusion bodies in Alpha-Thalassemia X-Linked Intellectual Disability Syndrome **(**ATRX).

Finally, a genetic diagnosis implies having a genetic confirmation of the individual's disorder by identifying the genetic aberration (sequence variant, copy number variant or imprinting defect) that has caused the individual's disorder. This may be a molecular diagnosis, generally meaning a diagnosis achieved through single gene analysis or a cytogenetic diagnosis from microarray or karyotyping. Achieving a genetic diagnosis enables other relatives to be screened to see whether they are carriers for a disorder, it also offers an accurate test to be available in future pregnancies.

1.3.2 Barriers to diagnosing developmental disorders

Barriers that have prevented making a genetic diagnosis in developmental disorders include: the large number of disorders, the diversity of phenotypes associated with developmental disorders, the diversity of genes and mechanisms implicated and the variability of the disorders. Also for some disorders, the small number of families with multiple affected members available and the reproductive disadvantage of the disorder limits the opportunities for studying genes which have been inherited together with the family and which segregate with the disorder to determine the cause (linkage mapping).

1.4 Advantages to making a genetic diagnosis

1.4.1 Benefit to affected individuals, their families and society

For families, being without a genetic diagnosis for their child's or their own developmental disorder, can lead to distress, guilt and anxiety. With a diagnosis may come relief and an end to the diagnostic odyssey and uncertainty. Graungaard *et al* showed that families without a diagnosis find it hard to cope with an uncertain future(59). A study of families with fragile X syndrome showed most viewed having a diagnosis as a

benefit as opposed to a disadvantage(60). Making a genetic diagnosis also enables families to access specific information about their child's condition and join support groups, it may help achieve learning support at school or special educational services.

In pursuit of making a diagnosis, individuals may undergo multiple investigations, many of which are invasive or painful including, blood tests, skin or muscle biopsies, lumbar punctures, brain magnetic resonance imaging (MRI) scanning or an electromyogram (EMG). These investigations are more difficult to carry out in young children especially those with learning or behavioral difficulties and investigations which are straightforward for adults or older children, such as undergoing a brain MRI scan may require a general anesthetic. These investigations are often at a large cost to the healthcare service or provider and require significant time commitments for the individual, family members and caregivers and possibly require an inpatient hospital admission.

Without a confirmed molecular or cytogenetic diagnosis individuals may be given an incorrect clinical diagnosis which may in itself lead to morbidity. In the study by Rare Diseases Europe (EURODIS) for some individuals an incorrect diagnosis led to treatments, including surgery and psychiatric treatment based on an incorrect diagnosis(7).

A genetic diagnosis enables the tailoring of medical care to the individual's specific condition. There are few genetic disorders for which there are specific pharmacological treatments available. Developing treatments has been challenging as not all conditions have not been molecularly defined and with small numbers of affected patients dispersed across the world, securing pharmaceutical funding and setting up of trials is difficult. However, there may be management guidelines or screening recommended for later onset related comorbidities. A genetic diagnosis may also enable individuals take part in clinical trials or be put on a disease registers to be contacted should treatments become available in the future.

Achieving a genetic diagnosis also gives parents accurate information about recurrence risks in future pregnancies. This may enable them also to pursue pre-natal testing or pre-implantation genetic diagnosis in future pregnancies if this is something they would

like. It also enables screening of the wider family to find out whether they are also affected or carriers of the condition. Being a carrier may confer a significant offspring risk for women where the disorder is X-linked or for a consanguineous couple when a recessive disorder has been identified in the family.

Achieving a genetic diagnosis also allows for individuals and their families to take part in phenotypic and natural history studies if they choose. This leads to greater understanding of the natural history of these rare conditions, which will ultimately improve their management. Understanding the phenotypes of these disorders will also enable identification of end-points for clinical trials and help development treatments in the future.

Understanding genetic disorders and their causes, allows screening programs to be developed and implemented, for example cystic fibrosis testing in newborn infants in the UK to alleviate the clinical consequences of the disease by early treatment, and prenatal screening for trisomies in the UK to enable couples to terminate an affected pregnancy if this is something they choose.

Understanding the phenotype and genetic cause of developmental disorders can also give important insights into common and complex disorders. For example Gaucher's disease is a rare lysosomal storage disorder resulting from biallelic variants in the glucocerebrosidase (GBA) gene. Several individuals with this condition and their relatives were noted to have developed Parkinson's disease. Further investigation of this phenomenon showed that Parkinson's disease segregated with the mutant *GBA* alleles in the family(61). This led to several studies, including a multicenter collaborative study of over 5691 individuals with idiopathic Parkinson's disease(62) which confirmed that heterozygous mutant GBA alleles are common and important risk factors for not only Parkinson's disease but also dementia with Lewy bodies (reviewed by Siebert *et al(63))*. More recently mutant GBA alleles have also been shown to be important risk factors for developing multisystem atrophy(64).

Identifying the underlying genetic cause of developmental disorders also advances biology more generally by increasing understanding of molecular pathways and gene and protein function in health and disease in humans and other organisms. For example the discovery through linkage analysis that a heterozygous missense mutation in the

forkhead box P2 gene (FOXP2) resulted in a rare severe speech and language disorder in one three generational family(65-68) uncovered FOXP2 as vital not only for normal language development in humans but also for birdsong in songbirds(69).

1.5 Summary and justification for this investigation

In summary, developmental disorders cause significant mortality, morbidity, distress to families and costs to the health service. For families, diagnosing developmental disorders alleviates stress, possible guilt and anxiety and provides them with coping mechanisms. It also enables the tailoring of medical care to an individual's particular condition and may prevent further invasive diagnostic investigations. Diagnosing developmental disorders can also give us insights into common human diseases as well as advance scientific understanding of other species and the world more generally.

Although advances in sequencing technologies have led to a genetic revolution in knowledge and unraveled the cause of a number of genetic disorders, there remain many challenges at this rapidly moving time. Firstly, not all genetic diseases have been discovered or had their genetic cause elucidated. Secondly there is limited longitudinal phenotypic data available for individuals with molecularly confirmed genetic disorders, meaning that we don't fully understand the wider phenotypes of many of these conditions in particular the phenotypes of many disorders in adulthood. Thirdly, molecularly confirmed developmental disorders are highly variable and little progress has been made to understand this variability or to explain incomplete penetrance of some disorders. Fourthly there are few genetic disorders with available treatments, in order to facilitate the development of treatments in the future, disorders must be molecularly defined and phenotypically characterised over time to allow end points for clinical trials to be identified. Treating genetic disorders is one of the biggest challenges in medical genetics for the next century.

1.6 Outline of this dissertation

This investigation takes traditional and contemporary approaches to understanding developmental disorders:

In Chapter 2, I introduce the developmental disorder Wiedemann-Steiner syndrome (WSS), an autosomal dominant multiple congenital-anomaly syndrome associated with a distinctive facial appearance, developmental delay and hypertrichosis. I led the original gene discovery project for this disorder in 2012(46), by discovering that *de novo* variants in *MLL* (now called *KMT2A*) underlie WSS. Since this time, a number of case reports and small case series have been published, but the full phenotypic spectrum of this disorder is unknown. I investigate the wider phenotypic spectrum associated with WSS by reporting on the phenotype of 84 individuals with this disorder and *KMT2A* mutations.

The theme of Wiedemann-Steiner syndrome continues into Chapter 3, where I present a phenotypic investigation looking for variants in genes underlying developmental disorders associated with hypertrichosis or WSS or related phenotypes. I also present a burden analysis showing that there is a burden of variants in genes encoding chromatin modification enzymes in individuals with hypertrichosis and WSS like phenotypes.

Understanding the architecture of developmental disorders in general is the theme in Chapter 4, where I present the DDD Study and my contributions to this project, including an investigation into the contribution of recessive causation to developmental disorders.

In the Chapter 5 (the final chapter) I will highlight the themes running throughout this dissertation, including 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.

In summary, in this chapter I have introduced developmental disorders and the approaches I will be taking to investigate them. I have detailed how knowledge of genetic disease has improved over time with advancing technology, but that clinical assessment for dysmorphic features has often been vital in gene discovery and continues to be important in the interpretation of variants from broad approaches to sequencing (whole exome and whole genome sequencing). Finally, I have outlined the investigations I will present in this dissertation.