Chapter 2

Wiedemann-Steiner syndrome resulting from mutations in *KMT2A*: A Genotype-phenotype study

2.1 Aims

- **To investigate the phenotype of Wiedemann-Steiner syndrome (WSS) resulting from KMT2A mutations**
- **To investigate the spectrum of mutations in *KMT2A* associated with WSS**
- **To investigate how missense mutations potentially affect *KMT2A* function and cause WSS**
- **To investigate whether experienced clinical geneticists can distinguish the facial appearance of individuals with *KMT2A* mutations from individuals with undiagnosed developmental disorders**

2.2 Introduction

2.2.1 Wiedemann-Steiner syndrome (WSS) and motivation for this investigation

Wiedemann-Steiner syndrome (WSS) is an autosomal dominant multiple congenital-anomaly syndrome associated with a distinctive facial appearance, developmental delay and hypertrichosis(70-72). Since the discovery that *de novo* mutations in *MLL* (now called *KMT2A*) underlie WSS(46) only 15 further individuals with WSS and *KMT2A* mutations have been reported(73-81). These individuals were reported in single case reports or in small case series. Therefore, the full phenotypic and mutational spectrum of WSS remains unknown and as a result, individuals with WSS may not be correctly diagnosed and may have unidentified medical needs.

The ability of clinicians to recognize distinctive facial features has played an important role in the diagnosis of genetic syndromes. As more and more patients with developmental disorders undergo whole exome sequencing to achieve a diagnosis, there will be increasing numbers of individuals identified with missense variants in *KMT2A* in whom a diagnosis of WSS had not been suspected. It is not fully understood how missense variants in *KMT2A* cause WSS and there are missense variants in *KMT2A* in healthy individuals in control databases such as The Exome Aggregation Consortium (ExAC) control database. Therefore, it is vital that missense variants in *KMT2A* are interpreted accurately to correctly diagnose and manage individuals with WSS. Therefore, determining the extent to which clinicians can truly distinguish WSS from other developmental disorders would be useful knowledge when determining the

pathogenicity of *KMT2A* variants identified on whole exome sequencing, in particular missense variants.

2.2.2 Wiedemann-Steiner syndrome (WSS): Definition

In 1989, Wiedemann reported a boy with pre- and post- natal growth deficiency, developmental delay and a distinctive facial appearance(70). His facial features included a round, flat face, hypertelorism, a long philtrum, short palpebral fissures, low set ears and a high arched palate. In addition, he had strabismus and dilatation of the renal calyces(70).

Subsequently, in 2000 Steiner & Marques reported an eight-year-old girl with similar phenotypic features to the individual reported by Wiedemann(2). She had hypotonia, short stature, an unusual facial appearance and intellectual disability. Her facial features included mild synophrys, telecanthus, narrow and down-slanting palpebral fissures, a low nasal bridge, a long and flat philtrum and a thin upper lip. She had a sacral dimple and a high arched palate(71)(Figure 2-1). She had mild hypertrichosis (increased hair) of her arms, legs and back which became accentuated with age(71)(figure 2-2).

Figure 2-1: Facial appearance of the girl reported by Steiner and Marques in 2000(71)
Facial appearance demonstrates telecnathus, mild synophrys, narrow downslanting palpebral fissures, a low nasal bridge, a low nasal bridge and flat and long philtrum. Reproduced from Steiner CE, Marques AP. Growth deficiency, mental retardation and unusual facies. Clin Dysmorphol. 2000;9(2):155-6. Figure 1. Frontal view of the patient's face. Reproduced with permission, copyright Lippincott, Williams & Wilkins.

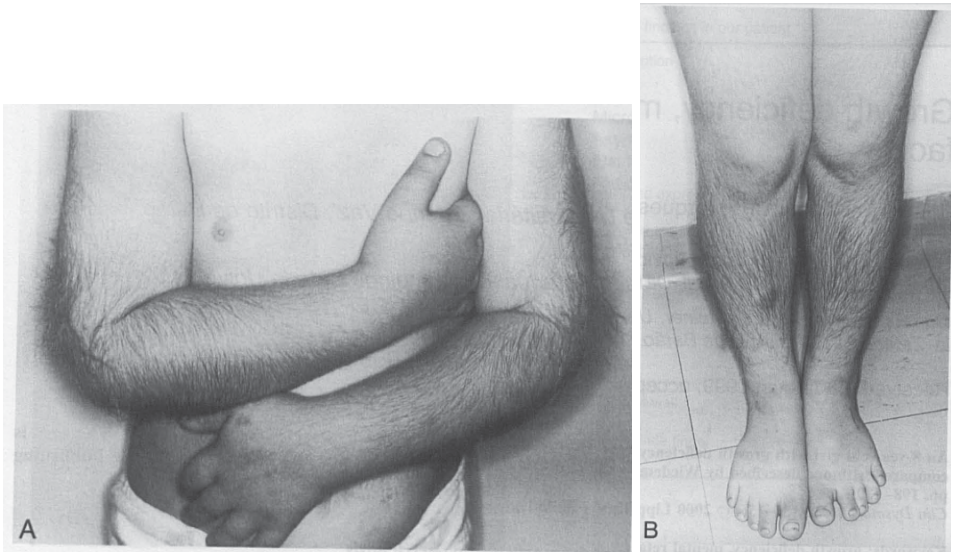
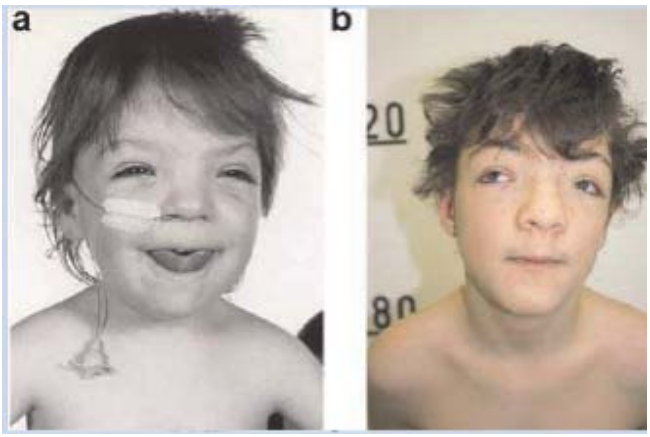


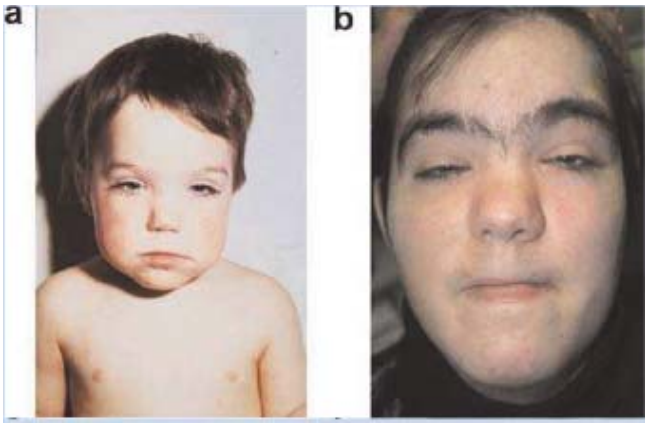
Figure 2-2: The arms and legs of the girl reported by Steiner and Marques in 2000(71)
The arms (A) and legs (B) showing hypertrichosis. Reproduced from Steiner CE, Marques AP. Growth deficiency, mental retardation and unusual facies. Reproduced from Clin Dysmorphol. 2000;9(2):155-6. Figure 2. Hypertrichosis on arms (a) and legs (b). Reproduced with permission, copyright Lippincott, Williams & Wilkins.

Koenig et *al* coined the name 'Wiedemann-Steiner syndrome' in a report on three individuals with a distinctive facial appearance and severe developmental delay(70-72)(Figure 2-3). They felt all three individuals had a similar phenotype to the individuals reported by both Wiedemann and Steiner(70, 71). Facial features common to the individuals reported by Koenig *et al* included: arched or thick eyebrows, hypertelorism, narrow palpebral fissures, and broad nasal bridge and tip(72). All three individuals developed hypertrichosis(72).

I



II



III

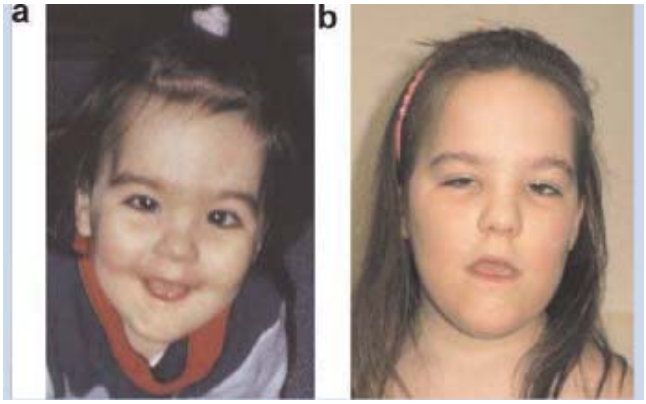


Figure 2-3: I-III: Facial appearance of the individuals reported by Koenig et al in 2010(72). Facial appearance demonstrates arched and or thick eyebrows, hypertelorism, narrow palpebral fissures, broad nasal bridge and tip. Reproduced from Koenig R, Meinecke P, Kuechler A, Schafer D, Muller D. Wiedemann-Steiner syndrome: three further cases. American journal of medical genetics Part A. 2010;152A(9):2372-5. Figure I a, b from Figure 1a, b: Patient 1 at the age of 2 7/12 and 12 4/12 years. Figure II a, b from Figure 3 a, b: Patient 2 at the age of 2 ½ years and 20 years. Figure III a, b from Figure 5: Patient 3 at the age of 19 months and 6 8/12 years. Reproduced with permission, copyright John Wiley and Sons.

2.2.3 *De Novo* mutations in *MLL* cause Wiedemann-Steiner syndrome

In 2012, my colleagues and I demonstrated that de novo mutations in the histone methyltransferase *MLL* (subsequently renamed as *KMT2A*)(82) underlie a distinct phenotype consistent with a diagnosis of WSS(46). This phenotype consisted of hypertrichosis cubiti (excessive hair of the elbow regions), short stature, intellectual disability and a distinctive facial appearance(46). Other associated features observed in the five reported individuals with *KMT2A* mutations were feeding difficulties, behavioral difficulties, skeletal abnormalities and cardiac defects.

2.2.4 15 further individuals with *KMT2A* mutations and WSS reported

Since the discovery that *de novo* mutations in KMT2A cause WSS there have been 15 further individuals with WSS and heterozygous *KMT2A* mutations reported in case reports and case series(73-81). These have increased the phenotypic spectrum of features associated with WSS and *KMT2A* mutations to include epilepsy, renal abnormalities, microphthalmia, congenital immune deficiency and premature eruption of dentition(73, 75, 76, 78). The largest case series reported since the initial gene discovery was by Miyake *et al*, who reported 5 individuals with *KMT2A* variants(74). As well as singletons, Dunkerton *et al* described monozygous twins with WSS and a heterozygous nonsense mutation in *KMT2A*; the twins shared the majority of their observed phenotypic features relating to WSS, differing only in subtle facial dysmorphism(78).

In addition to individuals reported in detailed case reports and case series *KMT2A* variants have been identified in six individuals in large consortia sequencing projects seeking to investigate epilepsy and autism. These include three individuals with epileptic encephalopathy, one individual with unclassified epilepsy (83, 84) and two individuals with autism(85). There is no information available as to whether these individuals have other phenotypic features of WSS.

2.2.5 The mutational spectrum of *KMT2A* mutations in WSS

The individuals reported in my previous work all had frameshift or nonsense mutations in *KMT2A*(46). Using real time PCR (RT-PCR), we showed reduced *MLL (KMT2A)* expression in primary skin fibroblast cells derived from an individual with a *de novo* frameshift mutation in *KMT2A* (c.6913del, p.Ser2305LeufsTer2) compared to unrelated healthy controls. We proposed haploinsufficiency as the disease mechanism(46) reporting their findings were consistent with the notion that transcripts arising from the mutant *MLL (KMT2A)* alleles are subject to nonsense-mediated decay. Subsequently, Strom *et al* expanded the mutational spectrum of *KMT2A* mutations associated with WSS to include a *de novo* variant predicted to affect splicing (c.4086+G>A) and a *de novo* missense mutation (c.4342T>C, p.Cys1448Arg)(76). Mendelsohn *et al* further expanded the observed mutational spectrum to include a multi-exon deletion (of exons 2 to 10) in a girl with WSS(73).

Stellaci *et al* proposed a possible genotype-phenotype correlation for *KMT2A* mutations. They reported a boy with congenital immunodeficiency with low levels of immunoglobulins(75) and severe epilepsy and a *de novo* missense mutation c.3481T>G (p.Cys1161Gly). The c.3481T>G mutation is predicted to disrupt a residue located within the cysteine-rich CXXC DNA binding domain of KMT2A. Stellacci *et al* noted that their individual and two of the other reported individuals who had missense mutations predicted to affect one of two functional domains of KMT2A (the CXXC DNA binding domain and the plant homeo-domain (PHD) zinc finger motif) had a tendency towards infections(74, 76). They proposed that missense mutations affecting these functional domains in a mechanism different from haploinsufficiency might specifically impact the transcriptional control of genes involved in the regulation of haematopoiesis and immune functions(75).

2.2.6 WSS overlaps phenotypically with other developmental disorders

A number of authors have reported phenotypic similarities between WSS and other developmental disorders resulting from mutations in genes that play a role in modifying chromatin structure (chromatin disorders) (46, 74, 77, 79). Jones *et al* and Miyake *et al*

reported similarities of the WSS phenotype to that of Kabuki Make-up syndrome, another congenital multiple anomaly syndrome(46, 74). In fact, three of the six individuals with KMT2A mutations reported by Miyake *et al* had initially been diagnosed with 'atypical' Kabuki Make-up syndrome(74). Kabuki Make-up syndrome results from mutations in the histone methyltransferase KMT2D or the lysine specific demethylase KDM6A(42, 86). KMT2D is a histone methyltransferase which adds trimethylation to histone H3 at lysine 4(87). KMT2D facilitates gene expression through acting as a transcriptional coactivator through interacting with transcriptional machinery at the promoters of target genes to facilitate gene expression(87). KDM6A is a demethylase that removes trimethylation from H3K27 a closed chromatin mark(88). Therefore, KMT2D and KDM6A have complementary functions and loss of function mutations in the genes encoding these enzymes leads to similar phenotypes(89). Yuan *et al* identified a heterozygous *de novo* nonsense mutation in an individual they felt had overlapping features with Cornelia de Lange Syndrome(CdLS)(77). CdLS results from mutations in the *HDAC8* gene (a histone deacetylase) or in genes encoding proteins affecting cohesin structure or function. CdLS is associated with hypertrichosis, short stature, limb defects and a distinctive facial appearance(90-94). Finally, Bramswig *et al* reported a child with a *de novo KMT2A* missense mutation who had overlapping features with Coffin Siris syndrome(79). Coffin Siris syndrome is a multiple congenital anomaly syndrome associated with hypertrichosis, developmental delay, agenesis of the corpus callosum and nail hypoplasia. Coffin Siris syndrome is caused by mutations in genes encoding components of the SWI/SNF chromatin remodelling complex(44, 45).

2.2.7 KMT2A is a histone methyl-transferase with multiple domains

KMT2A encodes the histone methyltransferase enzyme KMT2A, which is expressed in most cell types(82, 95). The KMT2A protein is a large (3,969 aa) multi-domain protein(96) which is one of a family of histone–lysine N-methyltransferase 2 (KMT2) proteins. The KMT2 proteins, including KMT2A, are highly conserved(97) and they play an important role in epigenetic regulation by methylating lysine 4 on the histone H3 tail (H3K4) to modify the structure of chromatin and promote DNA accessibility.

The nomenclature of the KMT2 family of proteins was updated in 2012; they were previously called the MLL family (82). The re-naming sought to alleviate confusion with the previous nomenclature given to these proteins: some enzymes had the same name, and the previous naming system did not reflect the complex function of these enzymes(82).

KMT2A generates mono-, di-, and trimethylated histone H3K4, through its SET domain and interaction with cofactors (Reviewed by Rao *et al(98))*. Mono-, di-, and trimethylated histone H3K4 have been shown to regulate chromatin-mediated transcription, including the transcription of multiple Hox and Wnt genes(99). The protein binding partners of KMT2A are discussed in more detail in Chapter 3: Gene discovery in hypertrichosis. The SET domain, which is responsible for KMT2A's histone-methyltransferase activity(100), is located at the C-terminus of the KMT2A protein (Figure 2-5). This is followed by, a WDR5 interaction (Win) motif, a nuclear receptor motif, a transactivation domain, a bromo domain, and plant homeodomain finger motifs(101). This is followed by a cysteine-rich CXXC domain which binds to unmethylated CpG dinucleotides(102). At the N-terminus there are three AT hooks, which are small DNA-binding protein motifs that also interact with chromatin(100, 102). KMT2A is proteolytically processed into N-terminal and C-terminal fragments at two independent sites(103). The cleaved fragments then form a stable complex that localizes to a sub-nuclear compartment(103).

2.2.8 Facial recognition investigations in dysmorphic syndromes

The delineation of WSS as a developmental disorder and its subsequent genetic characterisation relied on clinicians identifying the distinctive facial appearance of affected individuals(70-72). Clinicians have been recognising features unusual for age and ethnicity (dysmorphic features) and using them to help make genetic diagnoses since at least 1862 when Down syndrome was first described by Down(104). However, only a few investigations have been carried out to determine the accuracy with which clinicians are able to recognise individuals with specific genetic developmental disorders(105-108). No facial recognition experiments have been carried out in individuals with WSS and KMT2A mutations to confirm the facial appearance of

individuals with WSS can be distinguished by clinical geneticists from other developmental disorders.

Facial recognition experiments have given insight into the facial features used by clinicians to make certain diagnoses, highlighted disorders that can successfully be recognized by facial appearance as well as helped guide diagnostic criteria(105-108). These studies have also helped confirm that facial features of developmental disorders can change with age(107). The groups of clinicians used in these facial recognition experiments have included trained Dysmorphologists, and Hepatologists, who in general, would be expected to make few diagnoses based on the facial appearance of their patients. These experiments have involved clinicians being assessed at recognising the facial appearance of Alagille syndrome or Cornelia de Lange Syndrome (CdLS) (for a definition of CdLS, please see above)(105-108). Alagille syndrome is an autosomal dominant developmental disorder associated with cholestasis, cardiac anomalies, skeletal anomalies and other features.

In 2010, Rohatgi *et al* investigated the ability of Dysmorphologists to distinguish individuals with Cornelia de Lange syndrome (CdLS) from individuals with developmental disorders with overlapping features to CdLS, including fetal alcohol syndrome, Floating-Harbor syndrome and Kabuki Make-up syndrome(107). Floating-Harbor syndrome is named after the two hospitals in which the first individuals with the disorder were identified, namely the Boston Floating hospital and the Harbor General Hospital in California(109-111). They asked the clinicians to score each of the individuals as 'Classic', 'mild' or 'non-CdLS' as well as their certainty of the answer. The authors showed, that using facial photographs alone correct diagnoses were made in 90% of classic CdLS and 87% of non-CdLS individuals(107). However, the Dysmorphologists diagnosed only 54% of individuals classified as mild or variant CdLS correctly(107). The Dysmorphologists were asked to document the facial features they used in drawing their conclusions, this enabled the authors to assess the utility of specific features used to make a diagnosis of CdLS. For example, pencilled arched eyebrows, synophrys, long eyelashes, and thin upper lip were features helpful to making a diagnosis of CdLS in individuals with mild CdLS resulting from NIPBL mutations, however heavy straight eyebrows distracted from a diagnosis of CdLS in individuals with

SMC1A mutations(107). As a result of their findings Rohatgi *et al* suggested modifications to the previously used clinical diagnostic criteria for CdLS may be needed(107).

Computational approaches to facial recognition, including the use of machine learning have been developed to recognise individuals with CdLS and other developmental disorders from two dimensional (2D) photographs(112, 113). Ferry *et al* demonstrated that their machine learning driven method of 'Clinical Face Phenotype Space' was able to discriminate between syndromes with a similar accuracy to earlier methods which had utilised three dimensional (3D) image capture(113, 114).

More recently, Basel-Vanagaite *et al* repeated the experiment of Rohatgi *et al* using the same photographs of individuals with CdLS or overlapping disorders. They showed that the average detection rate of the automated Facial Dysmorphology Novel Analysis (FDNA) technology was 87%(112) compared to 77% by the Dysmorphologists in the study of Rohatgi *et al(107)*.

2.2.9 Questions studied in this investigation

There remain a number of unanswered questions about WSS resulting from *KMT2A* mutations, as the largest studies to date have contained only 5 individuals and no one has carried out a large phenotypic study of individuals with WSS. These include: What is the wider phenotypic and mutational spectrum associated with WSS resulting from *KMT2A* mutations? What is the incidence of WSS resulting from *KMT2A* mutations? And in an era where clinicians are interpreting variant findings from next generational sequencing analysis, is WSS caused by *KMT2A* mutations a facially recognisable developmental disorder? It is also not clear whether *KMT2A* mutations do underlie WSS as none of the original individuals reported under this classification have undergone sequencing.

2.3 Methods

2.3.1 Recruitment to the study was genotype or phenotype driven

I identified 84 individuals with *KMT2A* mutations through genotype- or phenotype-driven recruitment from 15 different countries. Genotype-driven recruitment involved the recruitment of individuals with a heterozygous *KMT2A* mutations and a phenotype consistent with WSS. These individuals were recruited either following diagnostic exome sequencing or targeted capillary sequencing of *KMT2A* in their local genetics centre or from a variety of separate research projects including the Deciphering Developmental Disorders (DDD) study, a study investigating individuals with atypical Cornelia de Lange syndrome and a study investigating individuals with Pierpont syndrome. The Deciphering Developmental disorders (DDD) study is collaboration between all the Clinical Genetics units in the United Kingdom and Ireland and the Wellcome Trust Sanger Institute(48), further information about the DDD study is given in Chapter 4: The Deciphering Developmental Disorders Study / Investigations into Autosomal Recessive Developmental Disorders. I considered variants to be pathogenic if they were predicted to result in a loss of protein function (nonsense, frameshift, splice donor and splice acceptor variants or multi-exonic deletions). I considered missense variants to be pathogenic if they were de novo and the phenotype fitted with the published phenotype associated with WSS or if the variant was identical to a variant that had been classified as pathogenic in another individual, or if an individual who had been diagnosed clinically as having WSS was subsequently found to have a *KMT2A* missense variant on sequencing.

For Phenotypic-driven recruitment, I identified 247 individuals with phenotypic features consistent with Wiedemann-Steiner syndrome (as assessed by a Clinical Geneticist) and / or with evidence of increased body hair. Individuals were recruited with one or both parents (duos or trios). Recruitment criteria, including the Human Phenotype Ontology(58) terms used to select patients with increased body hair are listed in given in Chapter 3: Gene discovery in hypertrichosis. 228 of the individuals underwent whole exome sequencing as part of the Deciphering Developmental Disorders study. The 20 remaining individuals were recruited from outside the UK and underwent whole exome sequencing separately at the Wellcome Trust Sanger Institute (WTSI). I identified

individuals with heterozygous pathogenic *KMT2A* mutations (this included variants predicted to result in loss of protein function and missense variants not present in the ExAC database) who were then carried forward for detailed phenotypic study. For DDD study individuals they were only carried forward for further study if the variant had already been clinically reported. For full methods, please see Chapter 3: Gene discovery in hypertrichosis.

2.3.2 Sequencing methods

Each of the different studies from which individuals were recruited or local diagnostic exome or *KMT2A* capillary sequencing had different sequencing methods. For individuals undergoing sequencing as part of the Deciphering Developmental Disorders (DDD) study or the WiSH Study, please see Chapter 3: Gene discovery in hypertrichosis for details of whole exome sequencing methods. Other individuals either underwent whole exome sequencing or targeted capillary sequencing of *KMT2A.*

For individuals not sequenced as part of the DDD study or the WiSH study the KMT2A variant data for each individual, including at least, the DNA sequence change and the predicted protein effect were provided by the patient's clinician. Transcript information and genomic co-ordinates were not available for every variant. I first obtained the correct DNA sequence change for each variant using Human Genome Variation Society (HGVS)(115) nomenclature system for the transcript ENST00000534358.1 and genomic co-ordinates for each variant. To do this I used Genome Reference Consortium human genome build 37 (GRCh37) in the Ensembl Variant Effect Predictor (VEP) programme(116) (release 88, March 2017). As a final check, I used Mutalyzer(117) to confirm the nomenclature of all variants was in line with HGVS nomenclature.

2.3.3 Phenotype analysis of individuals with WSS and *KMT2A* mutations

I devised a phenotype questionnaire which employed a general and targeted approach to capturing phenotype data of individuals with WSS and *KMT2A* mutations (Appendix 1). The questionnaire included general systems based questions, such as: "Urogenital abnormalities?", as well as a more targeted approach to gather negative data for key features of the disorder, for example: "Premature eruption of dentition?" I devised the

final two questions on the questionnaire (do they take any medications? And have they every been admitted to hospital or attended the emergency department?) to gather information on significant illnesses that may not have been elicited by the earlier part of the questionnaire.

To maximise clinical data given clinicians' busy clinical work-load, I calculated that a 2-sided A4 questionnaire was the optimum length. I used short bulleted questions for succinctness. In addition, to capture accurate 'negative data', I gave the options of: 'Yes', 'No', and 'Unknown', to enable clinicians to record that information about that specific feature was unknown. For examination findings, I gave the options of 'Yes', 'No', and 'Not Assessed' to enable clinicians to record that clinical findings had not been specifically looked for on clinical examination and to accurately quantify negative data. Devising the questionnaire was an iterative process and I added further questions, where necessary, as further phenotypic associations were determined.

All individuals entered the study with some growth and phenotypic information either provided by their local clinician via email, or with HPO coded phenotypic information from the Decipher database for the individuals recruited who were part of the DDD study(118). For further information about the Decipher database please see Chapter 4: The Deciphering Developmental Disorders Study / Investigations into Autosomal Recessive Developmental Disorders.

I visited 32/84 of the individuals in their local genetics centre with their Genetics clinician to carry out detailed clinical phenotyping. I asked the families questions based on the phenotype questionnaire and examined the affected individual with their local genetics clinician. 48 further individuals underwent more detailed phenotypic analysis by their local clinician completing the phenotype questionnaire I devised to capture their phenotypic data (appendix 2). Three further individuals did not have a phenotype questionnaire completed by their local clinician, including one individual in the DDD study and one individual who had been previously published in the medical literature for whom there was a significant amount of phenotypic information available.

In order to investigate the facial features of individuals with *KMT2A* mutations and WSS, for consistency, I reviewed facial photographs of 67 of the individuals with WSS and *KMT2A* mutations, including the individuals I had previously examined. I recorded the facial phenotypic features present. If I didn't feel I could confidently assess a phenotypic feature form the photograph, e.g. Synophrys may be difficult to assess in a blond haired individual, I recorded unknown. However, I also took into consideration any facial phenotypic features recorded by the local clinician or those I had recorded when examining each of the individuals in this case, e.g. if synophrys had been recorded on examining them I would record this.

2.3.4 Selection of missense variants for protein modelling.

To investigate how predicted missense mutations may potentially affect KMT2A function and cause WSS, I selected *de novo* missense mutations from my cohort and from the published literature to investigate the potential effect they may have on the 3D structure of KMT2A and how this may impair its biochemical function. It may be that one or more of these variants are pathogenic because they affect, however in this investigation I focused on the possible effect of these variants on the 3D protein structure of *KMT2A*. For the individuals from the medical literature with a phenotype consistent with WSS, I included only individuals reported in case reports or small case series who had *de novo* mutations and published photographs in order to give greater confidence that their KMT2A variant was pathogenic and their phenotype fitted with WSS.

In addition, I included two *de novo* missense variants from large consortia sequencing studies. This included one missense variant reported by de Rubeis *et al* in an individual with autism(85), and one missense variant identified in an individual with epilepsy reported by the Functional genomic variation in the epilepsies research project (EuroEPINOMICS-RES) Consortium *et al* (119). Although there was limited phenotypic information available about these individuals, as the phenotypes autism and epilepsy are observed in individuals with WSS I included them, however it is uncertain whether the phenotype of these individuals is consistent with WSS.

2.3.5 I selected control missense variants from the ExAC database

In order to compare the pathogenic missense mutations I had selected for protein modelling to background normal variation, I generated a control dataset of missense variants from the ExAC database(version 0.3)(120). I selected missense variants in the KMT2A transcript ENST00000534358 with a frequency of >1%. I selected only common variants (frequency >1%) in order to remove variants present in only a single individual and therefore reduce the chance of sequencing error. In addition to remove sequencing errors, I removed Non-PASS variants. In addition, I filtered out variants with multiple reference or alternate alleles.

2.3.6 Clinician recognition of facial features associated with WSS

To determine whether experienced Clinical Geneticists can distinguish the facial appearance of individuals with pathogenic *KMT2A* variants from *KMT2A*-negative individuals with developmental disorders I carried out a facial recognition study. I showed six Clinicians (four consultant Clinical Geneticists and two trainee clinical geneticists) 27 separate facial photographs consecutively in a projected Microsoft PowerPoint presentation. The size of this investigation was limited by the time that could be reasonably be requested of the clinicians for participation. Ideally a larger number of images than 27 would have been used. 19 of these photographs showed individuals with *KMT2A* variants and eight photographs showed randomly selected individuals with developmental disorders from the DDD study. Not all of the individuals in the facial recognition study were part of this cohort of 84 individuals. Also of the 19 individuals with KMT2A variants, 3 of these variants were missense variants identified on whole exome sequencing in individuals in whom a diagnosis of WSS was not certain, of these one was *de novo* and two were inherited. Given Clinical Geneticists are busy individuals, I decided that 27 photographs would be the optimal number of photographs to use to ensure an adequate number of clinicians would be in agreement to take part in the investigation.

I asked each of the clinicians to score each of the facial photographs from 0 (they do not feel the face resembles WSS at all) to 10 (it is a face that they would use as an exemplar of WSS for teaching/training).

2.3.7 Estimation of the incidence of Wiedemann-Steiner syndrome

In order to estimate the incidence of WSS I used the published SNV loss of function *de novo* mutation rate of *KMT2A*(121). I added this to the INDEL loss of function rate to give a per chromosome rate and then doubled this to obtain a per child rate. Finally, I took into account factors that may increase or decrease this rate to give an estimated range for incidence of WSS.

2.3.8 Ethical approval

Each family provided signed consent and ethical approval for this study was obtained from Guy's and St. Thomas' National Health Service (NHS) Foundation Trust local research ethics committee (ref.:08/H0802/84), ''Systematic Characterization of Genes in Inherited Disorders".

2.4 Results

2.4.1 I identified 84 individuals with *KMT2A* mutations

I identified 84 individuals with heterozygous *KMT2A* mutations and a phenotype consistent with WSS, including 73 mutations predicted to result in loss of function of the protein product, 10 missense mutations and one inframe deletion. My cohort consisted of 82 singletons and one set of monozygous twins. One of the individuals, a 51-year-old man (WSSP37), appeared to have a mosaic *KMT2A* frameshift mutation (c.3697delG p.Val1233LeufsTer2). All other 83 individuals from routine testing were understood to have germline *KMT2A* mutations, however deeper interrogation for mosaicism was not carried out. The age of individuals at their last clinical assessment ranged from 1 year 4 months of age to 51 years of age with a mean age of 10.99 years and a median age of 10.08 years. My cohort consisted of 37 males (44%) and 47 females (56%), which is not significantly different from balanced.

2.4.2 98.6% of mutations were *de novo* where inheritance was known and one mutation was inherited from a mosaic father

Where inheritance information was available (70 individuals), 69 mutations were identified as being d*e novo* (98.6%). I conclude from this that WSS is nearly completely penetrant and it is highly likely that the chance of reproduction is reduced in WSS individuals, this is likely due as least in part to learning difficulties and there may be other factors involved.

The one inherited mutation (c.3697delG p.Val1233LeufsTer2) in individual WSSP36 was identified from her 51-year-old father (WSSP37) who was mosaic for this mutation. For details of the sequencing method, please see Chapter 3: Gene discovery in hypertrichosis. Essentially, the mutation was not called in the exome variant profile from either the mother or father using GATK and therefore it appeared *de novo* in the proband. However, when I reviewed the Integrative Genomics Viewer (IGV) Plots for this trio, I detected that her father's variant profile showed 127 reads showing a G at position 118352491127 (73 forward reads, 54 reverse reads) and 11 reads showing a deletion at this position (Figure 2-4). The probands phenotype was consistent with WSS and her father also had some milder phenotypic features of WSS, as discussed further below.

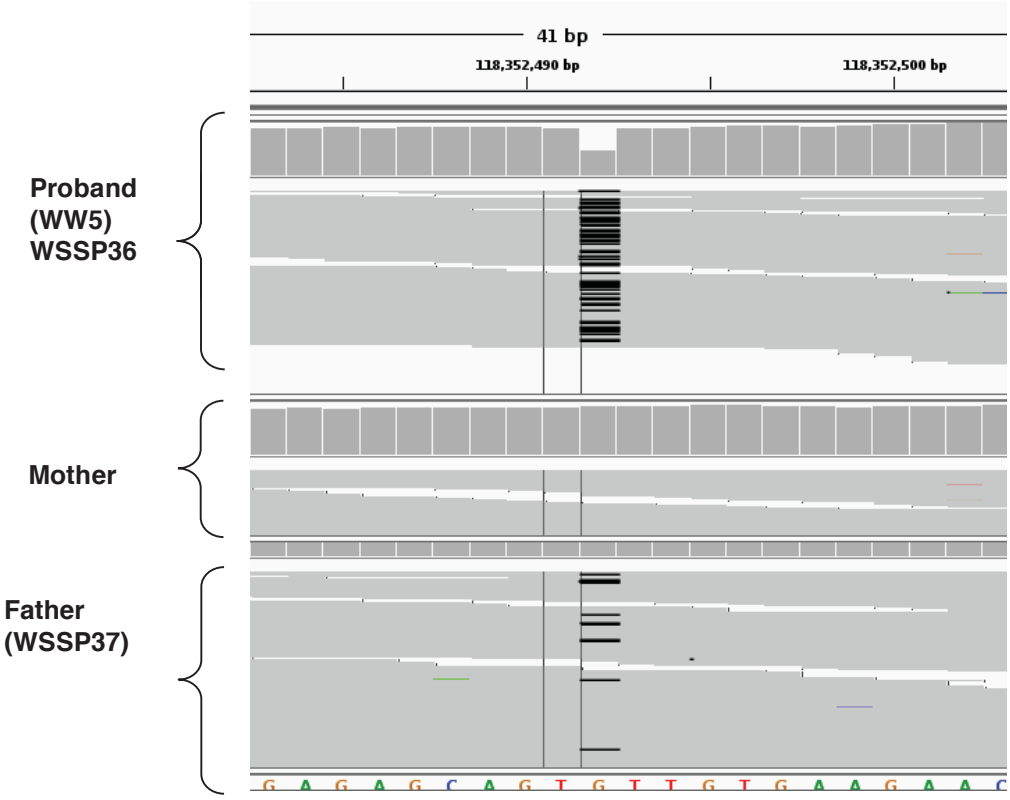


Figure 2-4: Integrative Genomics Viewer Plot showing Mosaic *KMT2A* variant
This Integrative Genomics Viewer (IGV) plot shows reads from proband WW5 (WSSP36) and her mother and father (WSSP37). The proband carries a single nucleotide deletion of a G (guanine) at position Chr11:118352491 (c.3697delG). This is predicted to result in the protein change p.Val1233LeufsTer2. The proband has 56 reads showing a G at position 118352491 (33 forward reads, 23 reverse reads) and 50 reads showing a deletion at this position. Her father has 127 reads showing a G at position 118352491127 (73 forward reads, 54 reverse reads) and 11 reads showing a deletion at this position, consistent with mosaicism.

2.4.3 I confirmed that *de novo* mutations in *KMT2A* cause Wiedemann-Steiner syndrome

None of the original patients reported under the classification of WSS by Wiedemann-Steiner or Koenig had been previously confirmed to have *KMT2A* mutations. Three of the individuals I present in this work (WSSP18, WSSP53 and WSSP58) were the three individuals reported by Koenig *et al*(72) under the classification of Wiedemann-Steiner syndrome (figure 2-3). All three individuals harbored a nonsense mutation in KMT2A (c.3518_3521delGCTT p.Cys1173Ter, c.3790C>T p.Arg1264Ter and c.4012+1G>C). Therefore, I have confirmed that 3 of the 5 individuals originally reported as having WSS have *KMT2A* mutations. Samples from the remaining two individuals originally reported by Wiedemann and Steiner were not available for sequencing.

2.4.4 There were 73 loss of function mutations, 10 missense mutations and one inframe deletion

I identified 36 frameshift mutations, 28 nonsense mutations, 10 missense mutations, 8 mutations predicted to affect splicing, one inframe deletion and one exonic deletion. These mutations are shown in Figure 2-5 with those falling in functional domains displayed also in Figures 2-6(A-C) and given in Appendix 1. There are two familial mutations, one in monozygous twins another in a daughter and her mosaic father. In addition, there are four recurrent mutations: c.2318dupC, p.Ser774ValfsTer12 (identified in WSSP56 and WSSP80), c.3460C>T, p.Arg1154Trp (identified in WSSP33 and WSSP84), c.3790C>T, p.Arg1264Ter(identified in WSSP30, WSSP47,WSSP53, WSSP62), and c.6379C>T p.Arg2127Ter (identified in WSSP59 and WSSP26).

The loss of function mutations are distributed throughout the gene with some clustering in and just after the CXXC zinc finger domain and in the PHD-like zinc-binding domain (zf-HC5HC2H) (figure 2-5). There is clustering of the missense mutations from our cohort and from published reports around the two zinc-finger domains (p=1e-8; this is the lowest value achievable from denovonear which is limited by the maximum number of iterations run, in this case 100 million simulations). The population control variants from the ExAC database are distributed along the length of the gene, with the majority located outside the recognized domain regions. I will discuss the missense variants in detail later in this chapter.

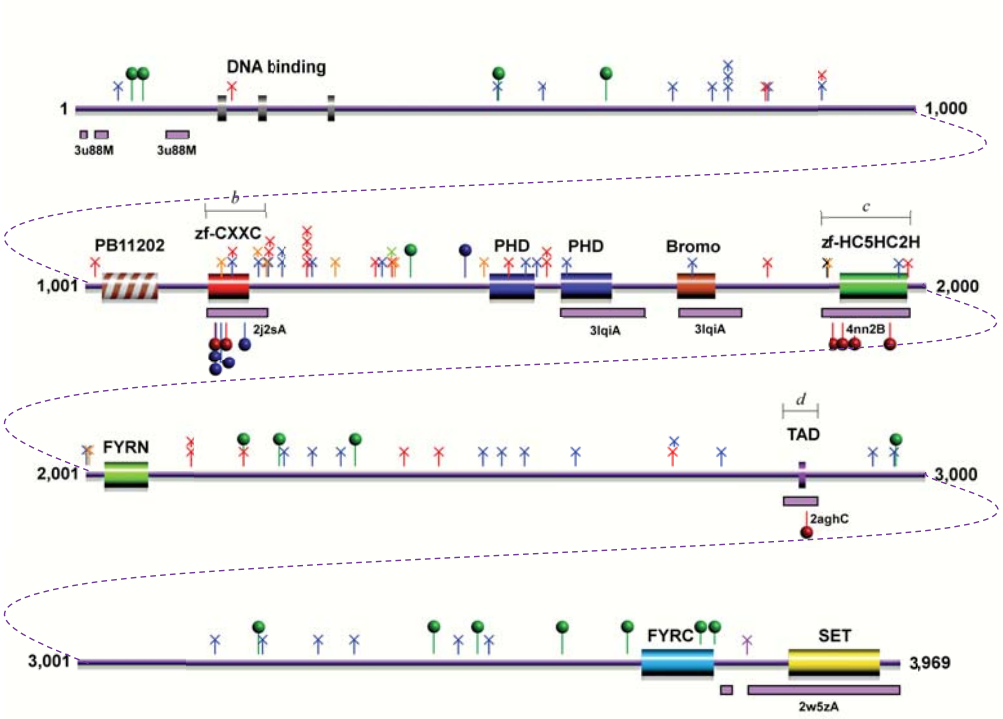


Figure 2-5: Distribution of *KMT2A* variants in individuals with WSS and control missense variants from The ExAC database

A schematic diagram of the KMT2A protein showing its Pfam domains, known 3D structures and KMT2A variants observed in individuals with WSS. The full length of the protein's 3,969 residues is represented by the purple line, cut into four lengths of up to 1,000 residues each. The cylinders along the line represent the Pfam sequence domains: PB11202 (an unclassified Pfam-B domain), zf-CXXC = CXXC zinc finger domain, PHD = PHD finger, zf-HC5HC2H = PHD-like zinc-binding domain, FYRN = F/Y-rich N-terminus, FYRC = F/Y-rich C-terminus, and SET = SET domain. The black-bordered purple bars underneath represent the 3D structural entries in the Protein Databank (PDB) corresponding to that part of the protein. The loss of function (nonsense, frameshift, splice donor and splice acceptor) variants are represented by crosses above the line. Blue crosses represent frameshift variants, red crosses represent nonsense variants, purple crosses represent in-frame deletions, orange crosses represent splice donor/acceptor variants and the green cross denotes the beginning of a multi-exonic deletion. Missense mutations are marked by coloured balls, these are below the line wherever they can be mapped onto a 3D structure. The red balls represent disease-associated missense mutations from this cohort. The blue balls represent *KMT2A* missense mutations reported in the literature. The green balls represent common (>1%) population control missense variants extracted from the ExAC database(120).

2.4.5 The growth profile in WSS resulting from *KMT2A* mutations

Information regarding birth weight was available 35 males and for 43 females (Figure 2-6 and 2-7). The majority of *KMT2A* positive individuals had weights within two standard deviations of the mean (38/43 females, 32/35 males). However, in infancy and childhood the growth charts show evidence for failure to thrive with height and weight in childhood being at the lower end of the normal range or less than two standard deviations (SDs) below the mean in the majority of individuals (Figure 2-8 to 2-11). 48% (16/33) males and 44% (19/43) females have stature >2SDs below the mean (short stature). With increasing age weight becomes more variable between *KMT2A* mutation positive individuals (Figure 2-10 and 2-11). 43% (3/7) of women over 17 years had a weight greater than 2 SDs above the population mean (figure 2-11). Head circumference in *KMT2A* mutation positive individuals was on or below the mean in all individuals, with 39% (12/31) males and 42.5% (17/40) females with a head circumference more than 2SDs below the mean (microcephaly) (figure 2-12 and 2-13).

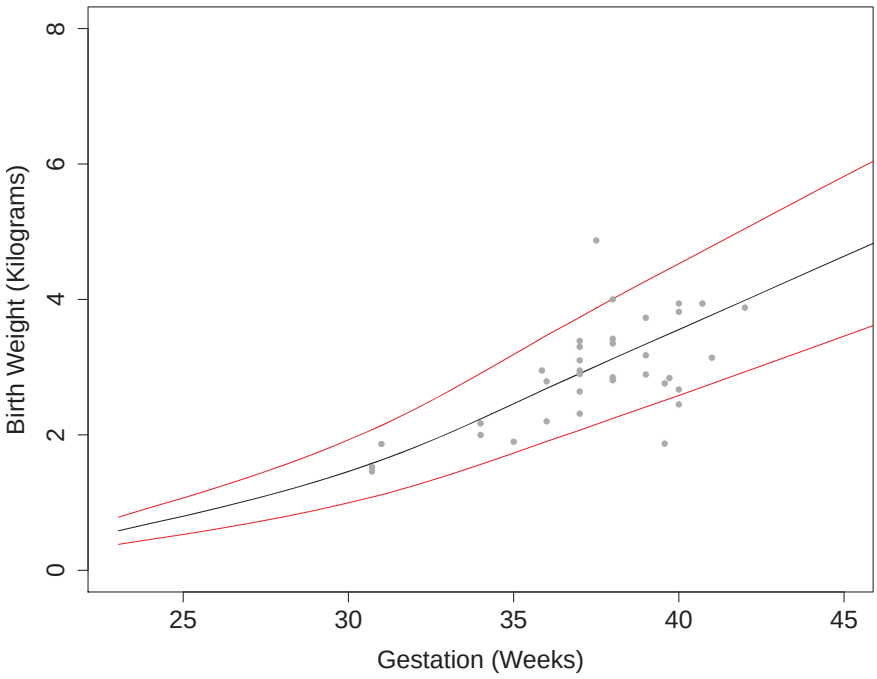


Figure 2-6: Birth weight of males with *KMT2A* mutations (n=35)

Each grey point represents a single individual's weight measurement in Kg at birth. The black line represents the UK population mean, the red lines represent mean +/- 2 standard deviations (The British 1990 Growth Reference(122)).

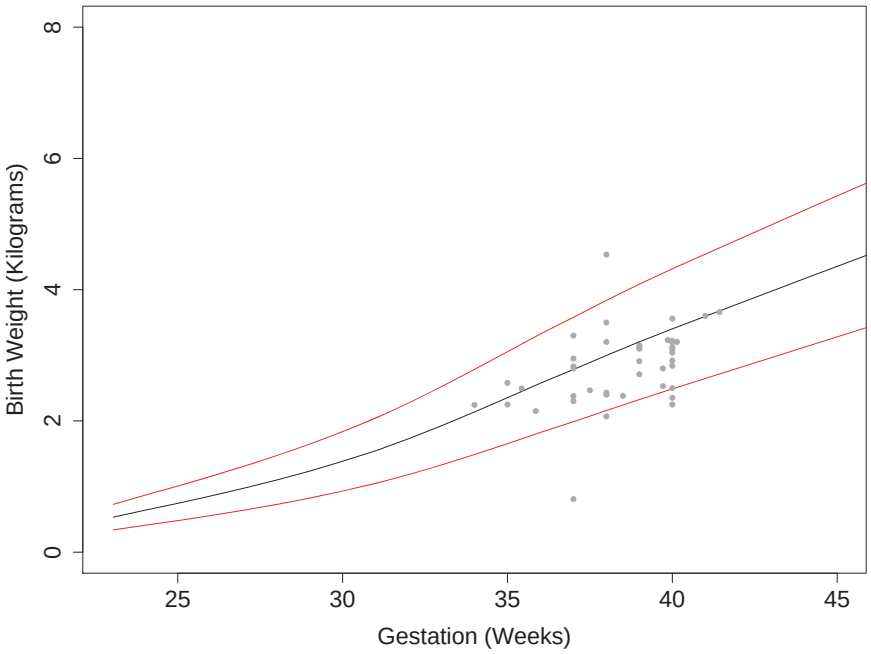


Figure 2-7: Birth weight of females with *KMT2A* mutations (n=43)

Each grey point represents a single individual's weight measurement in Kg at birth. The black line represents the UK population mean, the red lines represent mean +/- 2 standard deviations (The British 1990 Growth Reference(122)).

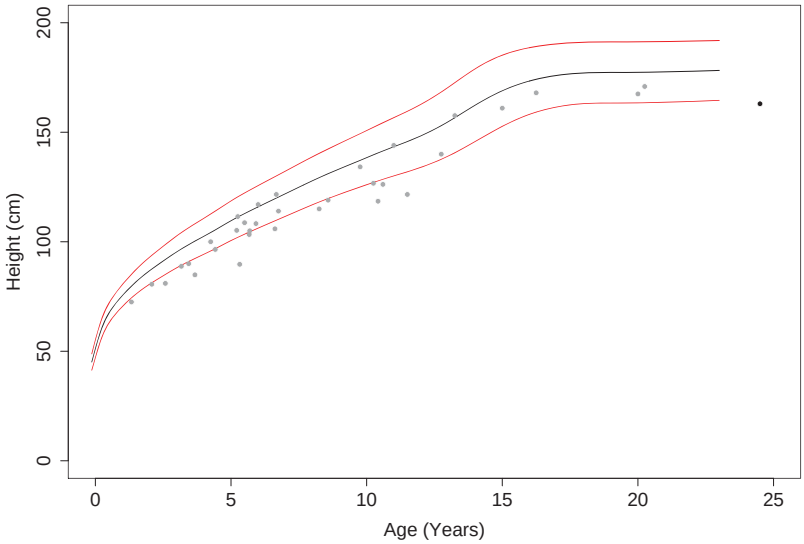


Figure 2-8: Height of males with *KMT2A* mutations (n=33)

Each grey point represents a single individual's height measurement. Where there was more than one data measurement available, data from when there was most recently a full set of OFC, weight and height data available. The black point represents the height of the individual aged 51 years with a mosaic *KMT2A* mutation. The black line represents the UK population mean, the red lines represent mean +/- 2 standard deviations (The British 1990 Growth Reference(122)).

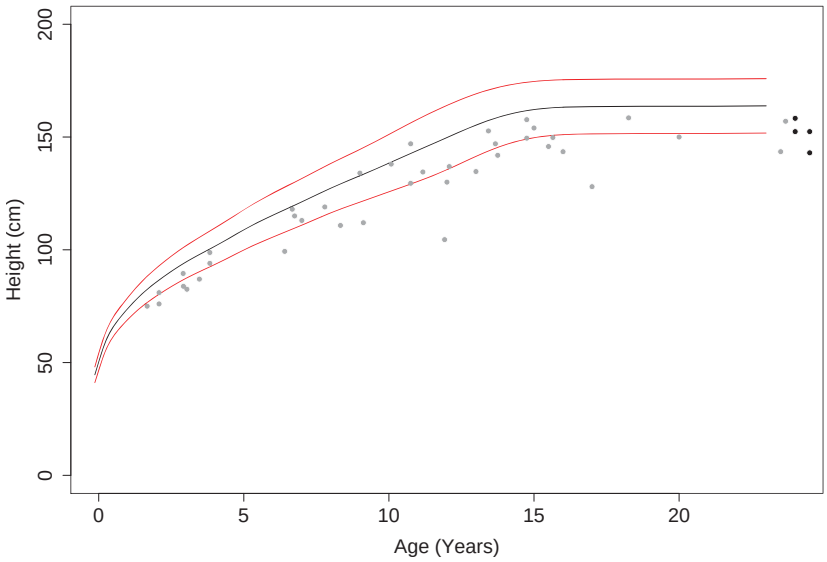


Figure 2-9: Height of females with *KMT2A* mutations (n=43)

Each grey or black point represents a single individual's height measurement. Where there was more than one data measurement available, data from when there was most recently a full set of OFC, weight and height data available. The black dots represent individuals aged greater than 24 years old with jitter. The black line represents the UK population mean, the red lines represent mean +/- 2 standard deviations (The British 1990 Growth Reference(122)).

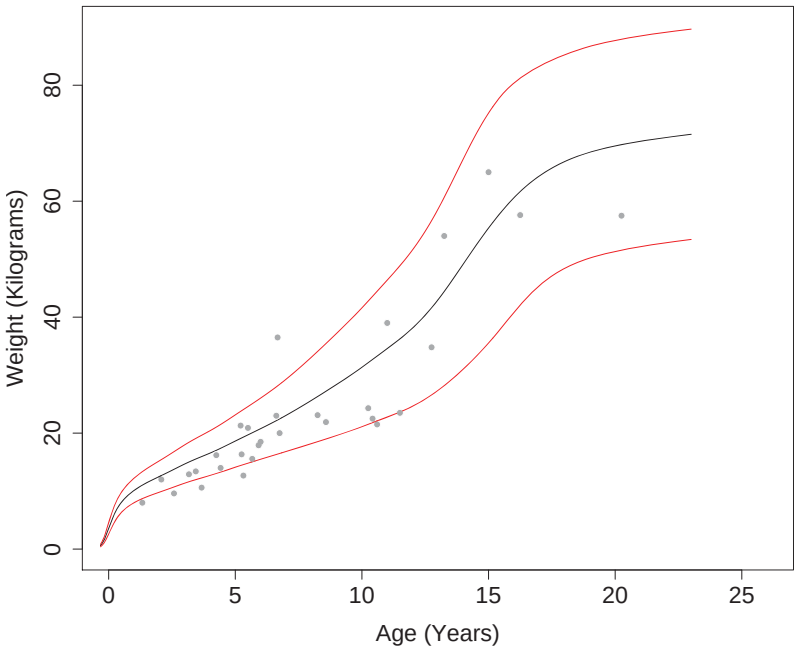


Figure 2-10: Weight of males with *KMT2A* mutations (n=30)

Each grey or black point represents a single individual's height measurement. Where there was more than one data measurement available, data from when there was most recently a full set of OFC, weight and height data available. The black line represents the UK population mean, the red lines represent mean +/- 2 standard deviations (The British 1990 Growth Reference(122)).

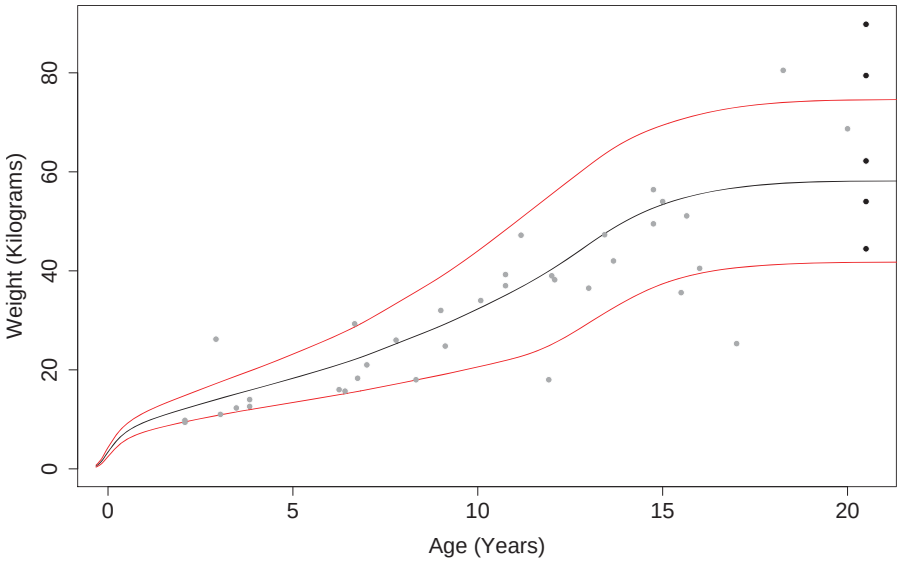


Figure 2-11: Weight of females with *KMT2A* mutations (n=40)

Each grey or black point represents a single individual's height measurement. Where there was more than one data measurement available, data from when there was most recently a full set of OFC, weight and height data available. The black points represent individuals who are greater than 20.5 years old. The black line represents the UK population mean, the red lines represent mean +/- 2 standard deviations (The British 1990 Growth Reference(122)).

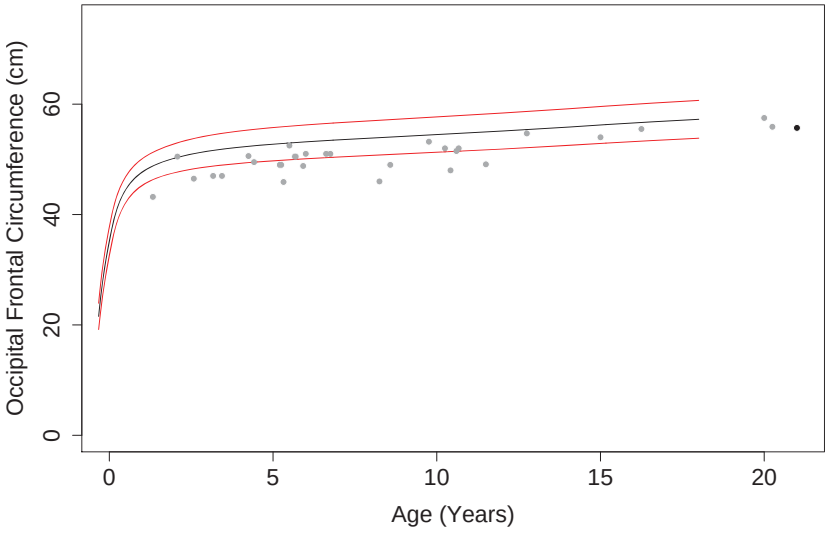


Figure 2-12: Occipital Frontal Circumference of males with *KMT2A* mutations (n=31)

Each grey point represents a single individual's Occipital frontal circumference (OFC) measurement. Where there was more than one data measurement available, data from when there was most recently a full set of OFC, weight and height data available. The black point represents the height of the individual aged 51 years with a mosaic *KMT2A* mutation. The black line represents the UK population mean, the red lines represent mean +/- 2 standard deviations (The British 1990 Growth Reference(122)).

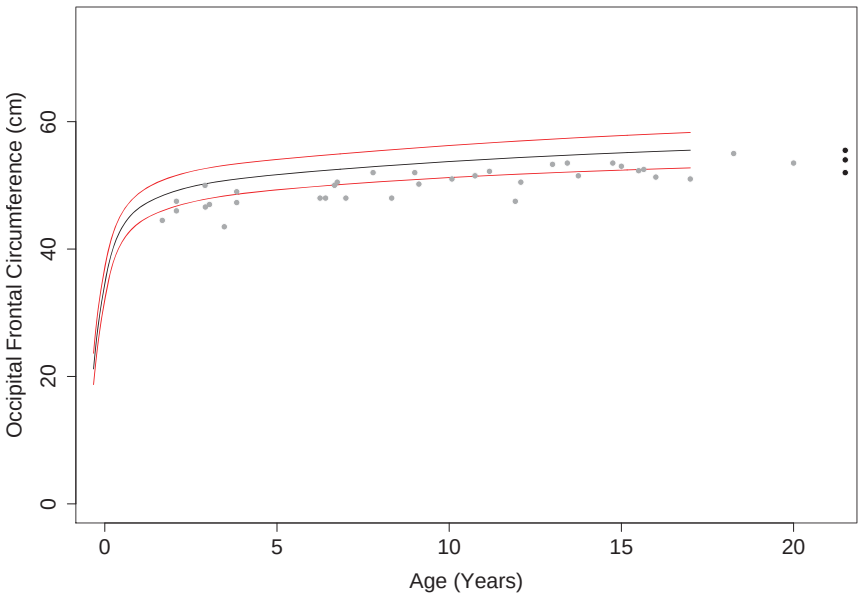


Figure 2-13: Occipital frontal circumference of females with *KMT2A* mutations (n=40)

Each grey or black point represents a single individual's height measurement. Where there was more than one data measurement available, data from when there was most recently a full set of OFC, weight and height data available. The black dots represent individuals aged greater than 24 years old with jitter. The black line represents the UK population mean, the red lines represent mean +/- 2 standard deviations (The British 1990 Growth Reference(122)).

2.4.6 Developmental delay, hypertrichosis, behavioral difficulties, and feeding difficulties were the commonest features

The most common phenotypic features in individuals with *KMT2A* mutations were developmental delay / intellectual disability (all individuals except the mosaic father), hypertrichosis (86%), behavioral difficulties (71%), feeding problems (75%, and requiring PEG or PEJ tube feeding 18%), hypotonia (51%) and constipation (50%), see table 2-1. In terms of facial features, the most commonly observed features were long eyelashes, narrow palpebral fissures, broad eyebrows, flat midface, wide nasal bridge, broad nasal tip and thin upper vermillion border (Table 2-1).

Human Phenotype Ontology Term (Where available)	Feature	Number of Individuals with feature / Number of individuals assessed for that feature	Percentage
Neurological			
HP:0001318	Muscular Hypotonia	43	**51%**
HP:0001250	Seizures	16	**19%**
HP:0011342	Mild global developmental delay	24/72	**33%**
	Mild to moderate global developmental delay	6/72	**8%**
HP:0011343	Moderate global developmental delay	31/72	**43%**
	Moderate to severe developmental delay	4/72	**5%**
HP:0011344	Severe global developmental delay	5/72	**7%**
HP:0012736	Profound global developmental delay	2/72	**3%**
	Mainstream/Regular School(Extra help)	20/61 (14/61)	**33%**
	Special Needs School	40/61	**66%**
HP:0000708	Behavioural abnormality	60	**71%**
	MRI Brain abnormality	14	**17%**
HP:0010832 (HP:0007328)	Abnormality of pain sensation	34	**40%**

HP:0002360	Sleep disturbance	34	**40%**
Gastrointestinal			
HP:0011968	Feeding difficulties	63	**75%**
HP:0011471	Gastrostomy tube feeding in infancy PEJ also	15	**18%**
HP:0011470	Nasogastric tube feeding in infancy (without needing PEG or PEJ feeding)	8	**10%**
HP:0002020	Gastroesophageal reflux	18	**21%**
HP:0002019	Constipation	42	**50%**
Cardiovascular			
HP3000001	Abnormal heart morphology	14	**17%**
Dermatological			
HP:0000998	Hypertrichosis	69	**86%**
HP:0004780	Elbow hypertrichosis	57	**72%**
HP:0011913 HP:0004532 HP:0011914	Lumbar hypertrichosis Sacral hypertrichosis Thoracic hypertrichosis	55	**80%**
	Hypertrichosis of the Lower limbs	54	**78%**
HP:0002219	Facial hypertrichosis	18	**36%**
Urogenital			
HP:0012210	Abnormal renal morphology	9	**10%**
HP:0000811	Abnormal external genitalia	9	**10%**
HP:0000140	Abnormality of the menstrual cycle	9/13	**69%**
Immunological			
HP:0002719	Increased frequency of infection	33	**39%**
Ophthalmological			
	Ophthalmological abnormality	42	**50%**
Ear Nose and Throat			

-	Otitis media with effusion / recurrent otitis media	13	**15%**
HP:0000365	Hearing impairment	11	**13%**
Mouth and palate			
HP:0006292	Abnormality of dental eruption	35	**42%**
Musculoskeletal			
HP:0000960	Sacral dimple	31	**37%**
-	Generalised muscular build	18	**21%**
HP:0007552	Abnormal subcutaneous fat tissue distribution	10	**12%**
	Abnormal planar fat pads	18	**21%**
	Swelling of feet or hands	22	**26%**
HP:0001212	Prominent fingertip pads	23	**27%**
HP:0006191	Deep palmar creases	5	**6%**
Features unusual for age and ethnicity			
HP:0000527	Long eyelashes	52/67	**78%**
	Narrow palpebral fissures	55/67	**82%**
	Eyebrow lateral flare	11/67	**16%**
HP:0011229	Broad eyebrow	56/67	**82%**
	Normal eyebrows	6/67	**9%**
HP:0000508	Ptosis	5/67	**7%**
HP:0000316	Hypertelorism	48/67	**72%**
HP:0040199	Flat midface	53/67	**79%**
HP:0000431	Wide nasal bridge	62/67	**93%**
HP:0000455	Broad nasal tip	64/67	**96%**
HP:0002263	Exaggerated cupids bow upper lip	10/67	**15%**

HP:0000233	Thin upper vermillion border	52/67	**78%**
HP:0012745	Short palpebral fissures	8/67	**12%**
HP:0000637	Long palpebral fissures	26/67	**39%**

Table 2-1: Phenotypic features of 84 individuals with WSS and *KMT2A* mutations

Table showing phenotypic features of 84 individuals with WSS and KMT2A mutations. Where only a proportion of individuals were assessed for a certain phenotypic feature this was stated in column 3, with the denominator reflecting the population assessed for each phenotypic feature. Where Human Phenotype Ontology (HPO)(58) terms are available these are used to code each phenotypic feature.

2.4.7 84% have mild to moderate developmental delay / intellectual disability

Amongst the WSS individuals for whom information about intellect was available 84% (61/72) of individuals were classified as having mild, moderate or mild-to-moderate developmental delay or learning difficulties. With 13% (9/72) individuals classified as having moderate-to-severe or severe developmental delay or learning difficulties and 3% (2/72%) of individuals having profound difficulties. In terms of motor milestones, the mean age of sitting was 11 months, walking was 22 months and mean age of first words was 19 months. Comparing these figures to the normal range adjusted for prematurity (sitting <12 months, walking <21 months and first words <21 months), shows mean age of sitting is within the adjusted normal range however the mean figures for walking and first words are outside the prematurity-adjusted-normal range and delayed.

In terms of education, where information about education is available 66% (40/61) of individuals attend a special needs school and 23% (14/61) of individuals require extra help in mainstream school. There is limited information available about adult outcomes. There are 10 individuals greater than 18 years old in the study. Of the four individuals for whom information is available. One individual works in a mainstream environment, two individuals work in sheltered environment and one individual carries out voluntary work. Three individuals live in sheltered living accommodation with minimal support.

71% (60/84) of individuals have behavioral difficulties. Commonly reported difficulties are abnormal fear / anxiety-related behavior, attention deficit hyperactivity disorder 7% (6/84), autism 26% (22/84), inflexible adherence to routines or rituals 10% (8/84) and anxiety 21% (18/84).

2.4.8 Seizures are associated with poor developmental outcomes

I sought to investigate the individuals with more severe difficulties (severe or profound developmental delay or learning difficulties) and assess whether they had any medical co-morbidities that resulted in their poor developmental outcome. In addition, I had noted in clinic that individuals with seizures often had more severe developmental delay. There is a significant association between seizures and an individual having severe or profound developmental delay / learning difficulties (p=0.0001) (Table 2-2). Other developmental disorders, such as tuberous sclerosis show worse developmental outcomes in individuals with seizures(123). It is possible that the seizures themselves slow developmental progress or affect brain development, however further investigations are needed to determine the cause for the association.

	Mild or moderate delay or learning difficulties	Severe or profound delay or learning difficulties
History of seizures		
Yes	10	6
No	67	1

Table 2-2: Relationship between seizure history and level of learning difficulties N=84

Table to show the association between seizure history and level of learning difficulties in 84 individuals with WSS resulting from KMT2A mutations.

2.4.9 Not all individuals with WSS and *KMT2A* mutations have hypertrichosis

We previously reported all five of our individuals had hypertrichosis(46), in particular referencing the increased hair in the elbow region observed in all individuals with *KMT2A* mutations. In subsequent case reports and case series only two individuals have been

reported not to have hypertrichosis(76, 81). I therefore sought to investigate the proportion of individuals with hypertrichosis. Hypertrichosis was reported in 69/84 (82%) of individuals in my cohort, with 77 clinicians or families asked on direct questioning about the presence of hypertrichosis. Increased hair of the elbow region was observed in 57/84 (68%) individuals, and on the back and legs in 55/84 (65%) and 54/84 (64%) individuals respectively. 18/84 (21%) individuals were noted to have facial hypertrichosis. The incidence of facial hypertrichosis has not previously been investigated in individuals with WSS and *KMT2A* mutations.

2.4.10 Imaging investigations in WSS caused by KMT2A mutations

17% (14/84) of individuals were reported to have congenital heart disease. The true incidence may be higher than this as 46/70 of the remaining individuals have not had an echocardiogram and not all structural cardiac defects produce symptoms in childhood nor can be detected by auscultation using a stethoscope. 11% (9/84) individuals had an abnormality of renal morphology, however only 7/75 of the remaining individuals have had a renal tract ultrasound, therefore abnormalities of renal morphology which can be asymptomatic may have not been detected.

40% of those individuals undergoing brain imaging (34/44) have a structural brain abnormality. However, again the true incidence may be higher than this as only 7 other individuals have had a brain MRI and for 44 individuals there is no information available about MRI brain imaging.

2.4.11 42% of individuals had an abnormality of dental eruption

35 (42%) of individuals were reported to have an abnormality of dental eruption. With the specific abnormalities reported as: advanced eruption of teeth (23), premature eruption of permanent teeth (7), premature loss of primary teeth (6) and persistence of primary teeth (2). Premature eruption of dentition was reported in the individual by Mendelsohn *et al(73)* who had several secondary teeth at four years of age. In this investigation, I expand the phenotype of dental eruption abnormalities to include persistence of primary teeth.

2.4.12 Sleep disturbance is common and may reflect disruption to circadian rhythm

34 individuals (40%) were reported to have sleep disturbance. Disrupted sleep patterns are observed in other developmental disorders such as Smith Magenis syndrome and can cause significant distress to families(124). Increasing evidence shows that chromatin remodelling events play a role in circadian regulation, reviewed by Aguilar-Arnal *et al(125)*. In particular, *KMT2A* has been shown to interact with the core circadian transcription factor complexes CLOCK-BMAL1 and PER-CRY(126). Therefore, the disruption of sleep in WSS may result from an underlying disruption of circadian rhythm. Individuals with WSS also have other phenotypic features that may contribute to sleep disturbance including gastro-oesophageal reflux, tendency towards otitis media with effusion and behavioural difficulties.

2.4.13 Recurrent infections are common in individuals with WSS

39% (33/84) of individuals were reported to have an increased frequency of infections. Increased frequency of infections has been reported previously (46, 74-76). (note the individual reported by Jones *et al* is also included in this cohort). The individual reported by Stellacci *et al* had congenital immune deficiency. The individual reported by Bramswig *et al* died from sepsis at 3 years of age(79). Stellacci *et al* suggested recurrent infections were more common in individuals with missense mutations than loss of function mutations, however I observed no difference in the frequency of increased infections in these groups (p=1.0). Similar disorders, such as Kabuki Make-up syndrome are also associated with an increased susceptibility to infection(127).

I propose the cause of the increased infections in WSS is likely to be multifactorial, with possible contributions resulting from: hypotonia, unsafe swallow, epilepsy, vesico ureteric-reflux and the increased incidence of otitis media. However, it is important to highlight individuals that may have a serious but treatable immune deficiency for further investigation as they may have a congenital immune deficiency. Further work is needed to investigate infections in individuals with WSS and determine the frequency of congenital immune deficiency to guide management and screening.

2.4.14 Other phenotypic features of WSS associated with KMT2A mutations

50% (42/84) of individuals had one or more ophthalmological abnormality, these included strabismus (18), myopia (10) hypermetropia (7) and myopia (10).

26% (22/84) of individuals were reported as having have swollen hands and or feet leading to the suspicion in some that the individual had lymphoedema, however, this has not been confirmed in any individuals in this cohort. In one individual (WSSP10), ultrasound duplex of the extremities was carried out at the age of 6 years to investigate his leg swelling. This was reported as showing an appearance that was felt to be more consistent with excessive subcutaneous fat (or homogeneous swollen subcutaneous fat), than edema. Other features not reported previously, include genital abnormalities in females.

In terms of previously unreported phenotypic features, 40% (34/84) of individuals report an abnormality of pain sensation with 23% (19/84) individuals defining this more specifically as reduced pain sensation. 18% (9/84) of individuals have genital abnormalities, although hypospadias has previously been reported our report expands the spectrum to include genital abnormalities in girls including cliteromegaly and labia minora agenesis. In addition, 2% (2/84) individuals have Raynauld's phenomenon.

2.4.15 A recurrent mutation (p.Arg1154Trp) is associated with seizures

There was no significant difference in the frequency of the most common phenotypic features of WSS (epilepsy, hypertrichosis, behavioral difficulties, feeding difficulties, hypotonia and constipation) between individuals with loss of function mutations versus those individuals with missense mutations. In addition, there was no significant difference between the incidence in the three groups of structural congenital abnormalities (cardiac, renal or brain lesions) sleep disturbance or infections between individuals with loss of function or missense mutations.

However, there is one recurrent missense mutation observed in three individuals and all of these individuals have seizures: c.3460C>T, p.Arg1154Trp. This is seen in two

individuals in this cohort WSSP33 and WSSP84 and in an individual reported by the Euro EPINOMICS consortium, a cohort study of individuals with classic encephalopathies including infantile spasms and Lennox Gastaut syndrome. All three individuals have a history of seizures. For WSSP33 the seizures started before the age of nine years, however no further information is available. WSSP84 had two generalised seizures in the first year of life associated with fever which were associated with a pathological EEG and developmental regression. WSSP33 and in an individual reported by the Euro EPINOMICS consortium, a cohort study of individuals with classic encephalopathies including infantile spasms and Lennox Gastaut syndrome(84). Further data is required to ascertain whether this is a chance association and whether the c.3460C>T, p.Arg1154Trp variant results in an increased frequency of seizures.

2.4.16 Father with mosaic *KMT2A* variant has a milder phenotype

The gentleman with the mosaic KMT2A variant (WSSP37), is of normal intelligence and has short stature. He was reported to have significant hypertrichosis with long hypertrichosis of his back as a child. Given developmental delay and learning difficulties are seen in all individuals with WSS and apparently, germline KMT2A mutations, it is likely because he carries the mutation in the mosaic form that he does not have more significant learning difficulties as a result of this.

2.4.17 The largest cluster of missense mutations lies in the CXXC zinc finger domain

The largest missense mutations cluster is located in the CXXC zinc finger domain (Figure 2-14). Zinc finger domains such as this, contain two zinc ions and selectively bind nonmethyl-CpG-DNA. Each zinc ion is tetrahedrally co-ordinated by four conserved cysteine residues. Some of the disease-associated *KMT2A* missense variants appear to disrupt the binding of one of the two zinc ions by affecting one of the eight co-ordinating cysteines. This would be expected to disrupt the domain's fold affecting DNA binding and recognition (Figure 2-15). Mutations resulting in a protein unfolding and being a target for proteolysis is seen in other disorders, such as ATRX syndrome caused by mutations in the *ATRX* gene(128). The effect of the Arg1154Trp mutation is less clear, however mutations of Arg1154 have been shown to abolish or significantly decrease

DNA binding but have no effect on global protein folding in model organisms(129). It is proposed that the detrimental effect on DNA binding may solely result from the removal of a functionally important side chain(129).

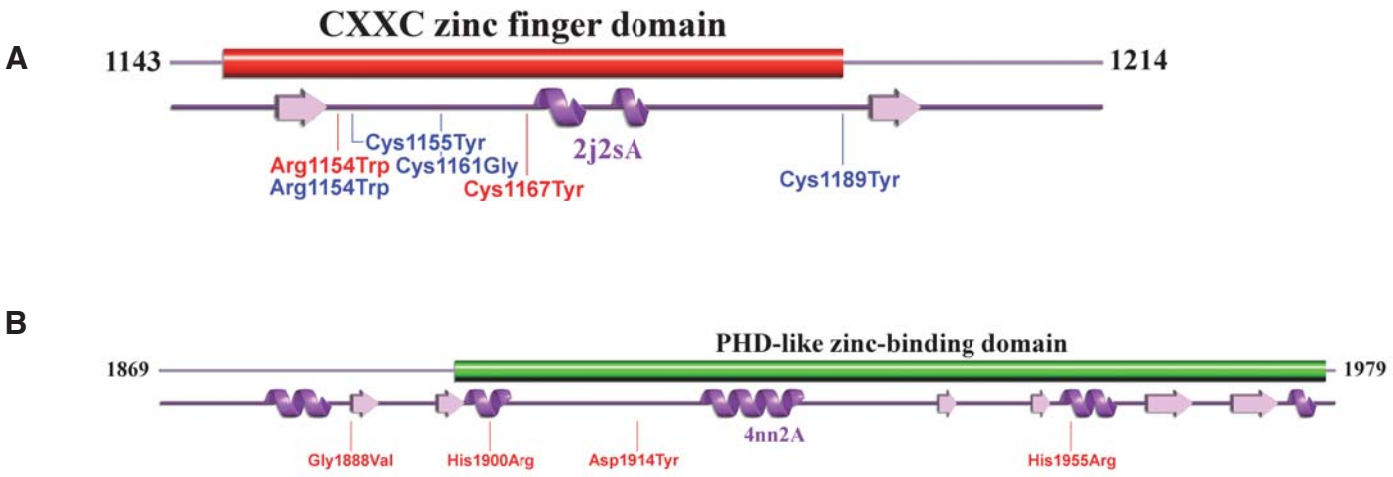


Figure 2-14: Distribution of *KMT2A* missense mutations

A: A zoomed in view of the region 1,143-1,214 corresponding to the 2j2s structure of the CXXC zing finger domain, and the most concentrated cluster of disease-associated KMT2A variants. The secondary structure of the protein is shown as arrows for beta strands and corkscrews for alpha-helices. Missense variants from this cohort are denoted in red. *KMT2A* missense variants reported in the literature are denoted in blue. The Arg1154Trp mutation is present three times, twice in this cohort and previously reported by the EuroEPINOMICS-RES Consortium (130). B.Residues 1,869-1.979. The KMT2A variants were plotted using the Canonical Uniprot protein sequence, this differs by three residues to the nomenclature of the KMT2A protein variants provided in Appendix 2.

The second largest cluster of disease-associated missense mutations is in and around the PHD-like zinc binding domain (Fig 2-16B). Although, there isn't a 3D structure of this domain in the PDB, there is one for a related protein, human Borjeson-Forssman-Lehmann associated protein, (PHF6) a transcriptional regulator that associates with ribosomal RNA promoters(131). Two of the de novo missense mutations, His1900Arg and His1955Arg directly affect zinc-binding residues and therefore are likely to disrupt the domain's fold and disable its DNA-binding ability.

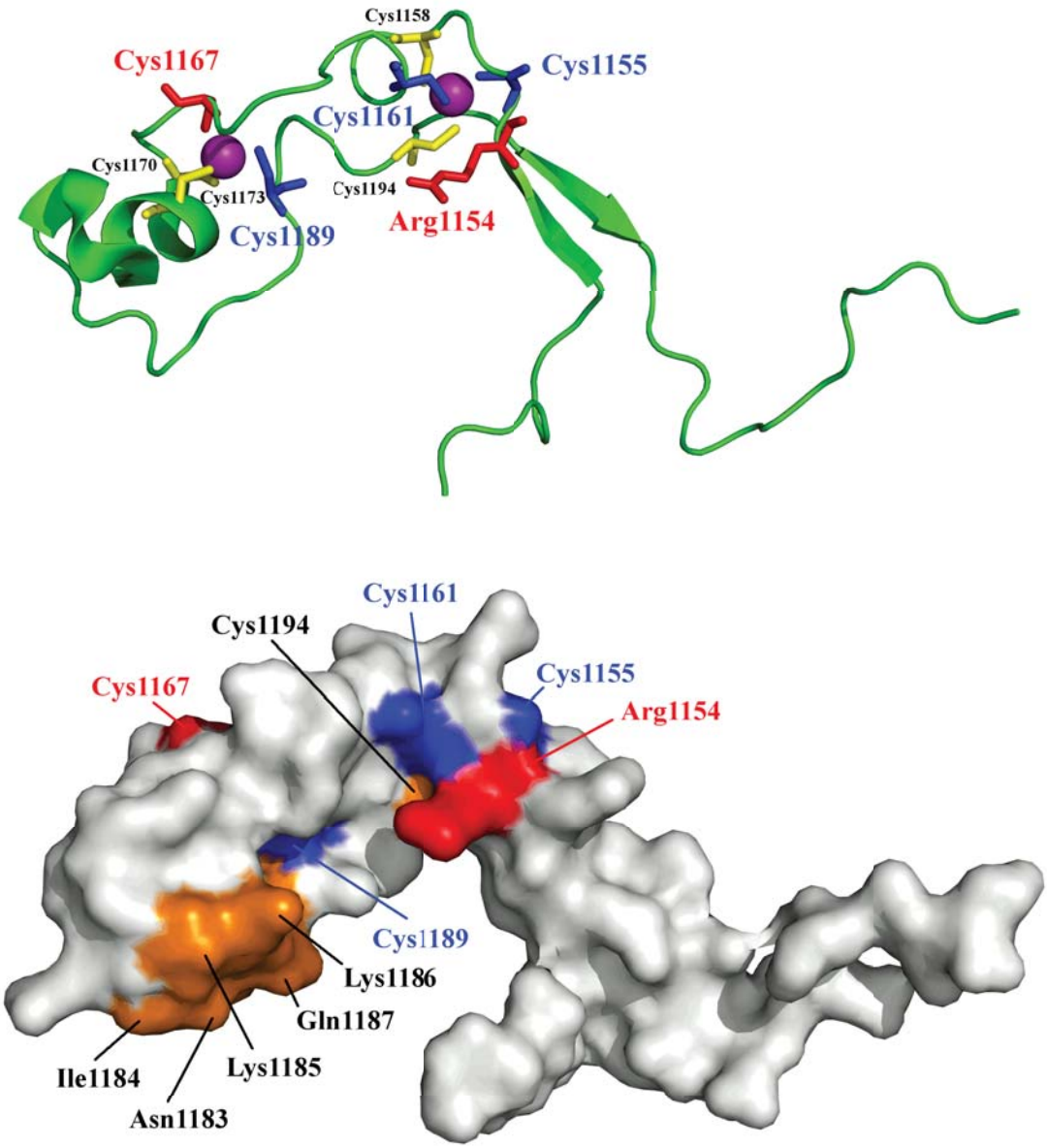


Figure 2-15: The 3D structure of the CXXC domain

A cartoon depiction of the CXXC domain, showing the two zinc ions (purple spheres), each coordinated by four cysteine residues (shown as yellow sticks, unless affected by a mutation when they are coloured red for Cys1167 from this cohort, and blue for Cys1155 and Cys1189 from the published literature). Also shown is Arg1154 (in red), mutations in which are identified as disease-causing in two individuals from this cohort and one individual reported by the EuroEPINOMICS-RES consortium. **b.** A representation of the domain's surface, with the residues thought to interact with DNA shown by the orange and red colouring. The red residue is Arg1154. Its mutation to the larger tryptophan is likely to interfere with DNA binding.

2.4.18 Clinicians can recognize patients with loss of function variants

After modelling the missense mutations, I then assessed whether clinicians can distinguish individuals with *KMT2A* mutations from control individuals with undiagnosed developmental disorders. Overall, the 6 clinicians scored the group of individuals with pathogenic loss of function *KMT2A* variants as more likely to have WSS than they scored the *KMT2A*-negative individuals selected at random from the DDD study (figure 2.16). The difference between the distributions of mean scores for the loss of function group versus the *KMT2A* negative group was significant ($p = 0.001664$, Two-sample Kolmogorov-Smirnov test). There was no significant difference in the distribution of scores between the trainee Clinical Geneticists (each with 1-5 years Dysmorphology experience) and the Consultant Clinical Geneticists (each with 5+ years Dysmorphology experience). Suggesting that years of experience are not necessary to be able to distinguish the facial features of individuals with WSS from those of individuals with developmental disorders.

2.4.19 There was a bimodal distribution for missense variants

The distribution of mean scores for the faces of individuals with missense variants was bimodal. When these scores were interrogated further it was obvious that the three individuals with the lowest scores were individuals in whom the diagnosis of WSS was being questioned (Figure 2-16). Two individuals had inherited missense variants from an affected parent. One individual had a de novo missense variant that was not felt to be causal as his phenotype did not fit with WSS.

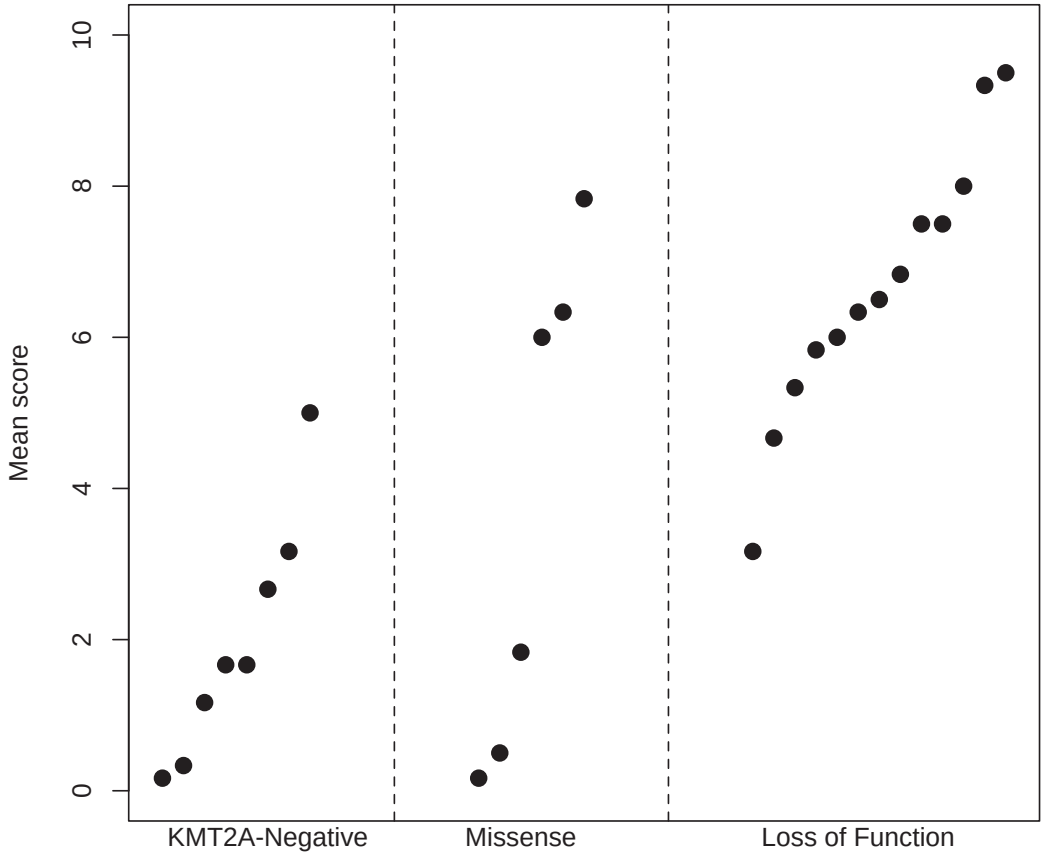


Figure 2-16: Scatterplot showing the mean score of all 6 clinicians for each 'face' by *KMT2A* variant status.

2.4.20 WSS has an estimated prevalence of 1 in 25,000 to 1 in 40,000

The SNV loss of function rate of KMT2A is 8.91935E-06(121). Adding this to the INDEL loss of function rate (gives a figure of 1.48E-05), this is a per transmission rate and therefore needs to be doubled to obtain a per child rate which is 2.96E-05. This equates to a prevalence of 1/34,000. Taking into account factors that may increase or decrease this rate (Table 2-3) I estimated that a prevalence of 1 in 25,000 to 1 in 40,000 would be appropriate for WSS.

Factors that might increase the birth prevalence of WSS from per child CNV and INDEL LoF rate	Factors that might decrease the birth prevalence of WSS from per child CNV and INDEL LoF rate
Include intragenic CNVs or INDELS too big to detect on exome sequencing are not included	Preferential spontaneous miscarriage
Other minor classes of LoF mutations, e.g. structural variants, intronic splicing mutations are not included	
Missense that would act as loss of function mutations are not included	

Table 2-3: Factors that might increase or decrease the birth prevalence of WSS from the per child CNV and INDEL LoF rate.

2.5 Discussion

2.5.1 Summary of findings

In summary, I have identified 84 individuals with *KMT2A* mutations from 82 unrelated families. Using data from these individuals 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 have successfully studied a monogenic disorder by collaborating with multiple large sequencing studies including the DDD study across 15 countries. I have demonstrated the key features of WSS, and broadened the phenotypic spectrum known to be associated with WSS as well as given insight into how missense mutations may cause WSS. In addition, I have confirmed the earlier findings of Jones *et al* that *de novo* mutations in *KMT2A* cause WSS(46, 72) as I report *KMT2A* mutations in all three of the original unscreened individuals reported under the classification of WSS by Koenig *et al*.

In this investigation, I have highlighted the key features of WSS caused by *KMT2A* mutations, namely developmental delay, hypertrichosis, behavioral difficulties, feeding problems, hypotonia and constipation. I have confirmed that sleep difficulties and frequent infections are also common in WSS individuals with *KMT2A* mutations. I have demonstrated the growth pattern of WSS, namely a normal birthweight followed to failure to thrive associated with short stature or height in the lower half of the normal range and a head circumference in the lower half of the normal range or microcephaly. I have

given the first evidence that adult women with WSS have a tendency towards obesity in adulthood. I have also demonstrated that not all individuals with WSS have hypertrichosis.

I have reported new features of WSS including female genital abnormities, abnormalities of pain sensation and Raynaulds phenomenon. I have heighted epilepsy as an important feature of WSS and that epilepsy is associated with poorer outcomes in WSS individuals. I have reported the first somatic mosaic individual with a *KMT2A* mutation and shown he has a milder than other individuals with germline *KMT2A* mutations and has normal intelligence. I have used standardised phenotypic terms (HPO terms) wherever possible throughout this phenotypic analysis to enable other researchers to readily access and understand my findings and perform meta-analyses in the future. I demonstrated through my facial recognition study that the facial appearance of individuals with WSS is distinguishable from that of other individuals with developmental disorders by experienced Clinical Geneticists and trainees in clinical genetics.

2.5.2 Limitations to this investigation

This study is limited to the populations of people living in the vicinity of Clinical Genetics services with the means to be reviewed by a Genetics doctor. As a result, some populations are under-represented in this study. In fact, 83/84 individuals are white. This study is therefore underpowered to investigate WSS in black and Asian or indigenous populations where there may be variability in facial and or other features. Other limitations to this study is that the negative data is not complete as observed from the phenotype Table 2-1. Clinicians are busy people and filling questionnaires is often performed quickly which is likely one contributing factor to the lack of negative data.
Finally, adults are under-represented in this cohort, in order to understand the difficulties faced by adults and know how to look after them from a medical point of view it is important to understand the adult phenotype of individuals with WSS. A further limitation is that in this study I didn't investigate the effect of the missense variants on splicing. Some of the missense variants in my investigation may be pathogenic due to an effect on splicing.

2.5.3 Interpretation of missense variants

In conditions where the predominant mutation mechanism is loss of function alleles, the interpretation of missense findings will remain challenging for many years to come. Without a functional or biochemical assay, current practice is to rely on the patient's clinical phenotype and the predicted effect of the mutation on the protein product. With time, functional work and recurrent mutations missense mutations will be further understood but in the meantime an element of caution needs to be exercised. Clinical phenotyping remains vital to interpret variants and investigations such as this are vital to help clinicians interpret variants found on next generational sequencing platforms. Facial recognition software may play an important role in this process in the future.

KMT2A is a large gene and it may be that phenotypes other than WSS are associated with missense variants in *KMT2A* as is observed with *CREBBP*. Mutations in CREBBP have long been established to be associated with the chromatin disorder Rubinstein Taybi syndrome which shares a number of similarities to WSS, but more recently other phenotypes have been associated with missense variants in *CREBBP(132)*. With the passage of time and with further individuals with developmental disorders undergoing exome sequencing, it will become apparent whether there are other phenotypes resulting from missense mutations in *KMT2A*.

2.5.4 Challenges for phenotypic investigations in the era of genomics

It's well recognised that with next generational sequencing approaches large amounts of sequencing data is generated. However, but given the increased rate of diagnosis with these platforms there will be large accompanying amounts of phenotype data for clinicians to manage in the study of novel disease genes or already recognised genes with broadening phenotypic spectra.

Phenotype study methods long employed by Clinical Geneticists largely consisting of transfer of information by secure email or letter will be put under strain by large sample sizes and the difficulties of managing more data points than ever before. The data management problem has been addressed by research studies such as the DDD and

100,000 genomes project which have incorporated online phenotyping, enabling clinicians to enter their own phenotype data coded by HPO terms into an online portal. This data is then standardised, less prone to error and much more manageable. More ideal yet would be for all clinical data in hospitals to all be recorded by standardised terms at the point of care (clinical review by a doctor or nurse). Then with the patient's consent if they were to join a research study this data set could simply be transferred to the research study. Maintaining genomic information is also important, and ensuring that mutations are recorded correctly in a changing world of genomic builds. This has been recognised by databases such as Decipher who store mutations in a manner which is easily updatable.

2.5.5 Future Directions

Unanswered medical questions

There a number of unanswered questions about WSS that need to be addressed in the future for the accurate management of the affected individuals. This includes (this list is not exhaustive): Further study of infections and immune function; behavioural analysis to help families manage challenging behaviour, assessment of specific learning difficulties to guide teachers with support in the classroom; investigations to difficulties experienced in adulthood with screening and medical management employed as appropriate; investigation of the aetiology of hand and foot swelling in early life. As with other complex developmental disorders, a multidisciplinary approach is most appropriate for the care of these individuals who have multiple medical needs.

To accurately counsel families about recurrence risks further study is needed to determine the true incidence of germline mosaicism. In addition, study of individuals with seemingly germline mutations for evidence of mosaicism may give insight to the individuals at the milder end of the phenotypic spectrum.

How do mutations in KMT2A cause WSS

Studying the effect of missense variants on KMT2A protein levels would help further understand which variants are disease causing and help understand how variants cause

disease. There are commercially available antibodies to KMT2A(133) and these could be used with Western blotting techniques to analyze KMT2A levels. Functional experiments in cells with *KMT2A* mutations in early development may play an important role to determine how mutant *KMT2A* alleles cause disease and develop targets for treatment. But also, to understand how the same mutations in *KMT2A* can be seen in solid tumors and in the germline with very different effects. Greater knowledge of KMT2A in disease will increase understanding of epigenetic regulation in general, and potentially advance knowledge of other chromatin disorders whose aetiology may have shared mechanisms.

Increased knowledge about the sites of KMT2A binding within the genome would also assist understanding of how mutations in *KMT2A* cause WSS. Current knowledge suggests that KMT2A may bind to both gene promoters and gene enhancers(134, 135), however further investigation into the location of KMT2A binding within the genome is required.

Understanding the variability of WSS

Further investigations are needed to fully understand how the difficulties associated with WSS can range from mild to profound. I have shown that there is an association with epilepsy and severe developmental outcomes. However, there may in addition be mutations in other genes involved with the second mutation acting as a modifier or resulting in the individual having two disorders. Interrogation of the exome variant profiles of individuals with WSS stratified for the level of their learning difficulties may help elucidate whether mutations in other genes are involved. Exome sequencing in individuals with extreme phenotypes recently proved successful in identifying *DCTN4* as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis(136) and a similar approach might be possible for WSS and individuals with profound learning difficulties or developmental delay. Emond *et al* in their investigation, which successfully identified DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis, studied only 91 individuals(1). However, the effect size for *DCTN4* was estimated to be relatively large, the collective minor allele frequency of the implicated variants was reasonably high, and the individuals were phenotypically well matched with the exception of the trait of interest. Identifying modifier genes in WSS is likely to be more complex than in cystic fibrosis. Firstly, because the phenotype of WSS is more

complex, making phenotype matching more difficult. Individuals with WSS have multiple problems affecting many more multiple organ systems than are affected in cystic fibrosis including the brain. The number of individuals needed to identify modifiers would depend on the effect size and minor allele frequency of modifier variants, but it would also require accurate phenotype matching with the exception of the phenotype of interest. Therefore, although Emond *et al* showed a sample size of around 100 is possible to identify a modifier in cystic fibrosis(1), a sample size in excess of this is likely to be necessary to identify genetic modifiers in WSS.

Given the association with epilepsy and worse developmental outcomes, to understand this further, mice with heterozygous *KMT2A* mutations could be stimulated to have seizures to see whether this impacted negatively on their development.

2.5.6 Summary of discussion

In summary, I have through molecular and clinical analysis identified 84 individuals with *KMT2A* mutations from 82 unrelated families. Using data from these families 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 have highlighted areas for further investigation in the future, including further immune and behavioural phenotyping, investigations into the variability of WSS and functional experiments to determine how mutations in *KMT2A* cause WSS. A multidisciplinary approach to the medical care of these individuals is vital for the care of these individuals who have multiple medical needs.