De Novo Loss-of-Function Mutations in *SETD5*, Encoding a Methyltransferase in a 3p25 Microdeletion Syndrome Critical Region, Cause Intellectual Disability

Detelina Grozeva,[1] Keren Carss,[2] Olivera Spasic-Boskovic,[1] Michael J. Parker,[3] Hayley Archer,[4] Helen V. Firth,[5] Soo-Mi Park,[5] Natalie Canham,[6] Susan E. Holder,[6] Meredith Wilson,[7] Anna Hackett,[8] Michael Field,[9] James A.B. Floyd,[2,10] UK10K Consortium,[2] Matthew Hurles,[2] and F. Lucy Raymond[1,*]

To identify further Mendelian causes of intellectual disability (ID), we screened a cohort of 996 individuals with ID for variants in 565 known or candidate genes by using a targeted next-generation sequencing approach. Seven loss-of-function (LoF) mutations—four nonsense (c.1195A>T [p.Lys399*], c.1333C>T [p.Arg445*], c.1866C>G [p.Tyr622*], and c.3001C>T [p.Arg1001*]) and three frameshift (c.2177_2178del [p.Thr726Asnfs*39], c.3771dup [p.Ser1258Glufs*65], and c.3856del [p.Ser1286Leufs*84])—were identified in *SETD5*, a gene predicted to encode a methyltransferase. All mutations were compatible with de novo dominant inheritance. The affected individuals had moderate to severe ID with additional variable features of brachycephaly; a prominent high forehead with synophrys or striking full and broad eyebrows; a long, thin, and tubular nose; long, narrow upslanting palpebral fissures; and large, fleshy low-set ears. Skeletal anomalies, including significant leg-length discrepancy, were a frequent finding in two individuals. Congenital heart defects, inguinal hernia, or hypospadias were also reported. Behavioral problems, including obsessive-compulsive disorder, hand flapping with ritualized behavior, and autism, were prominent features. *SETD5* lies within the critical interval for 3p25 microdeletion syndrome. The individuals with *SETD5* mutations showed phenotypic similarity to those previously reported with a deletion in 3p25, and thus loss of *SETD5* might be sufficient to account for many of the clinical features observed in this condition. Our findings add to the growing evidence that mutations in genes encoding methyltransferases regulating histone modification are important causes of ID. This analysis provides sufficient evidence that rare de novo LoF mutations in *SETD5* are a relatively frequent (0.7%) cause of ID.

The identification of over 100 rare but highly penetrant X chromosome genes in which mutations cause intellectual disability (ID) supports the hypothesis that the human genome contains more than 2,000 genes critical to normal intellectual development.[1] When the causative variants are rare and when candidate genes are numerous, the interpretation of a single novel variant in a gene not previously associated with disease is challenging. The recent analysis by Piton et al. looked at evidence of pathogenicity for many of the X chromosome genes in which mutations are reported to cause ID and elegantly demonstrated how previously published evidence of disease causality needs careful review in the light of sequence data of large population sets.[2] As our knowledge of rare variants in the normal population increases, there is a need to establish increasingly stringent criteria to evaluate whether a disease-causing variant has been identified to ensure the accurate translation of new knowledge into safe clinical practice.[3]

In order to identify further Mendelian causes of ID, we screened 996 ID-affected individuals for variants in previously associated genes and candidate genes for ID on the basis of current literature, in-house data, and sequence homology to genes previously implicated in ID. The appropriate ethical approval was obtained (research ethics committee reference 03/0/014), and parents or guardians provided written informed consent. We performed DNA sequence analysis by using next-generation sequencing methods to investigate the coding sequence of 565 genes (Table S1, available online) from 996 individuals with moderate to severe ID (all samples met DNA quality metrics). This was a subset of a large replication study of seven rare diseases and comprised a total of 2,812 individuals who were investigated within the UK10K study. The phenotypes studied were congenital heart disease, ciliopathy, coloboma, ID, neuromuscular disease, severe insulin resistance, and congenital thyroid disease; internal technical control samples were also included for comparison.

The GenomiPhi V2 DNA Amplification Kit (GE Healthcare) was used for whole-genome amplification of the DNA used for sequence analysis with the use of 1 μl of 10 ng/μl template DNA prior to pull-down. A custom-based targeted Agilent SureSelect pull-down array was designed with the SureDesign program (Agilent Technologies). This target was 3.4 Mb of sequence from the coding exons (GRCh37/hg19 human reference sequence, UCSC

[1]Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, UK; [2]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; [3]Sheffield Clinical Genetics Service, Sheffield Children's Hospital, Western Bank, Sheffield S10 2TH, UK; [4]Institute of Medical Genetics, University Hospital of Wales, Heath Park, Cardiff CF14 4XW, UK; [5]Clinical Genetics, Addenbrooke's Treatment Centre, Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge CB2 0QQ, UK; [6]North West Thames Regional Genetics Service (Kennedy Galton Centre), North West London Hospitals NHS Trust, Harrow, Middlesex HA1 3UJ, UK; [7]Department of Clinical Genetics, Children's Hospital at Westmead, Westmead, NSW 2145, Australia; [8]Genetics of Learning Disability Service, Hunter Genetics, Waratah, NSW 2298, Australia; [9]Department of Medical Genetics, Royal North Shore Hospital, St. Leonards, NSW 2298, Australia; [10]The Genome Centre, John Vane Science Centre, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK

*Correspondence: flr24@cam.ac.uk

http://dx.doi.org/10.1016/j.ajhg.2014.03.006.

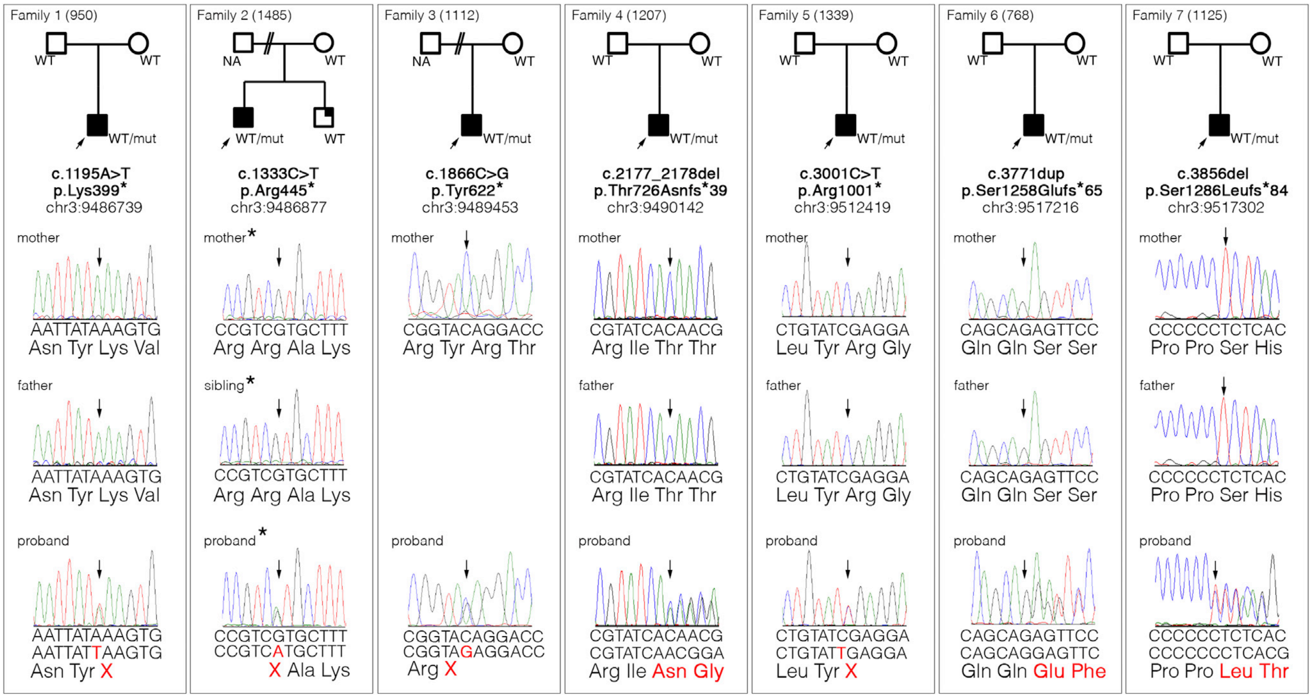

Figure 1. Families Affected by *SETD5* Mutations
The pedigree for each family is shown at the top. Sanger sequencing electropherograms of the mutations are shown below the pedigrees. An arrow indicates the position of the mutation. The genomic coordinates are according to the GRCh37/hg19 human reference sequence. Abbreviations are as follows: NA, not available; WT, wild-type; and mut, mutation. Asterisks indicate that the Sanger sequencing illustration is of the reverse strand.

Genome Browser) of 1,189 genes, of which 565 were ID-related candidate or known genes. Target enrichment and amplification were performed with the HaloPlex Target Enrichment Kit (Agilent Technologies) according to the manufacturer's instructions. The Illumina HiSeq 2000 platform was used to sequence the exons from the targeted regions. Reads were aligned to the reference genome (GRCh37/hg19) with the Burrows-Wheeler Aligner, and single-nucleotide variants (SNVs) and small indels were identified with SAMtools.[4,5] For each sample, variants sites (SNVs and indels) were called with the Genome Analysis Toolkit Unified Genotyper.[6] The calls were then annotated with vcf-annotate (VCFtools).[7] Functional annotations were added with the Ensembl Variant Effect Predictor v.2.8 against Ensembl 70.[8] Standard sequence quality-control criteria were applied to the called variants: variants with a Phred-scaled quality score > 40 and a mapping quality score > 50 were investigated further. The Integrative Genomics Viewer was used for visually inspecting the underlying sequencing data.[9] In addition, only rare variants with a minor allele frequency $< 1\%$ in all of the following data sets were considered for downstream analyses: 1000 Genomes, the UK10K twins cohort, the NHLBI Exome Sequencing Project (ESP), a cohort of 2,172 individuals from whom whole exomes were sequenced at the same laboratory (UK10K), and the UK10K rare replication cohort itself (including all phenotypes). Furthermore, only putative loss-of-function (LoF) variants (nonsense, frameshift, and essential splice-site variants) were analyzed.

We selected the top ten genes with the highest number of rare LoF variants present at frequencies $< 1\%$ (Table S2). We then prioritized the genes for more detailed follow-up on the basis of the following information: (1) the number of different independent LoF variants identified in this cohort, (2) the presence of these variants in a candidate gene, and (3) the paucity of rare LoF variants in the candidate genes in controls according to frequencies in the NHLBI ESP. On the basis of the above, SET-containing-domain 5 (*SETD5*) was selected for further investigation.

Seven independent LoF variants were identified within the coding sequence of *SETD5* in the ID cohort (Figure 1). All seven variants were observed only once within the whole UK10K replication cohort of 2,812 individuals with rare disease. The total number of *SETD5* haplotypes sequenced was 5,624, which represents 1,992 alleles from individuals with ID and 3,632 alleles from individuals with other rare diseases not usually associated with ID. None of these seven LoF variants have been reported in PubMed, ClinVar, HGMD, dbSNP, or 1000 Genomes, and none were identified in the NHLBI ESP, where coverage of the respective exons was available for >4,000 European American individuals. Because none of the individuals within the cohorts contributing to the current data deposited in the NHLBI ESP were described as having intellectual impairment, this was used as an additional independent control set. The seven LoF variants within *SETD5* (CCDS46741.1, RefSeq accession number NM_001080517.1) and their corresponding protein truncations (RefSeq NP_001073986.1) are the following: c.1195A>T (p.Lys399*), c.1333C>T (p.Arg445*), c.1866C>G (p.Tyr622*), c.2177_2178del (p.Thr726Asnfs*39), c.3001C>T (p.Arg1001*), c.3771dup (p.Ser1258Glufs*65), and c.3856del (p.Ser1286Leufs*84) (Figure 1).

We confirmed all of the variants by Sanger sequence analysis by using stored nonamplified genomic DNA from the probands and performed familial segregation analysis to test a de novo hypothesis of disease (Figure 1). For 5/7, we were able to establish molecular evidence of a de novo variant in the proband. For the two families for which a paternal sample was unavailable, the variant was absent from the maternal sample and both parents were reported clinically to have normal intellect, suggesting that de novo inheritance was the most likely cause of disease in these families. In one of the families (family 2) with no paternal sample available, a sibling with a mild intellectual impairment was found not to carry the mutation. The number of de novo LoF variants expected to occur by chance in *SETD5* in a cohort of this size was calculated with the use of the known exome mutation rate,[10] the proportion of variants expected to be LoF,[11] and the length of the coding sequence of *SETD5* (4,329 bp). We compared this number to the observed number (five) of molecularly confirmed de novo LoF variants in *SETD5* in our cohort and found that the probability that they occurred independently by chance was extremely low ($p = 5.25 \times 10^{-9}$, corrected for multiple testing). These results indicate that rare LoF mutations in *SETD5* are a likely cause of ID.

The CCDS46741.1 transcript of the coding sequence of *SETD5* is 4,329 bp long and encodes a protein of 1,442 amino acids. In this transcript, we found only one LoF variant listed in public databases: a 4 bp polymorphic indel (c.4277_4280del [p.Arg1426Profs*82]) with genomic position chr3: 9,517,722. This LoF variant is located within the terminal 16 amino acids of the protein and is reported in 38/3,904 European American adults (from the NHLBI ESP). The seven *SETD5* LoF variants present in the ID cohort are all within the CCDS46741.1 sequence and contribute to the major consensus transcript. The mutations are all located upstream of the single polymorphic LoF variant reported in the NHLBI ESP at the 3′ end of the gene. We then performed further analysis of the DNA sequence from the affected individuals to identify whether there were other more plausible variants that could account for disease. In 6/7 individuals, we did not identify further rare LoF or missense variants in any of the interrogated known genes in which mutations cause syndromic or nonsyndromic ID, nor were there rare LoF variants in candidate genes (565 genes in total). In the family 2 proband, we did identify an essential splice-site variant (c.2914+1G>A, genomic position chr3: 433,481, RefSeq NM_006614.3) in *CHL1* (MIM 607416), a candidate gene not previously associated with ID. This variant was only present in the proband and was absent in the mother and mildly affected brother. In addition, in the family 2 proband, we observed a single missense variant (c.179A>G, RefSeq NM_004595.4) in *SMS* (MIM 300105), an X-linked gene in which mutations are reported to cause ID (Snyder-Robinson syndrome [MIM 309583]) in males. The variant was present in the unaffected mother and mildly impaired brother of the proband. The residue is not well conserved throughout evolution and is not located within a conserved domain of the protein. We concluded that this was not likely to be the primary cause of disease in the family. Furthermore, the phenotype of the mildly affected brother was not in keeping with Snyder-Robinson syndrome. Thus, it is uncertain whether this variant makes any additional contribution to the phenotype. The presence of the *SETD5* variant in the proband and the absence of the variant in the younger brother are compatible with the more severe clinical features in the older brother. The additional contribution of the *CHL1* variant in the proband remains uncertain.

On the basis of the genotypic similarity of the seven affected individuals, the clinical phenotype of each individual was collated from the recruiting physicians, who were blinded to the genotype for minimizing clinical bias in reporting (Figure 2). One family declined permission to publish photographs but permitted review by S.E.H. and F.L.R. Although the clinical features were variable, a number of common features other than ID included a similar facial morphology comprising brachycephaly and a prominent high forehead with striking eyebrows described as full, broad, straight, or with synophrys. The nose morphology was long, thin, and tubular. The morphology around the eyes was similar with long, narrow, and upslanting palpebral fissures; in addition, mild ptosis, unilateral amblyopia, nystagmus, and strabismus were described in single individuals. Ears tended to be large with fleshy lobes, long, and low set; one individual had a preauricular pit. The facial features of individual 2 (Figure 2) were slightly coarser than those of the other six individuals, which might reflect the additional sequence variants present. Feeding problems, particularly difficulties with swallowing and chewing, were noted by several families and physicians. Two children had congenital heart defects; one had a mitral valve prolapse, and the other had a ventricular septal defect with a patent ductus arteriosus (Table 1 and Table S3). Four of the seven children had either an inguinal hernia or hypospadias repaired at a young age. Also, 4/7 children had skeletal abnormalities that required varying degrees of intervention. Thoracic scoliosis, kyphosis, and lordosis were reported, and two children had a significant leg-length discrepancy (one of them also had talipes and hypoplasia of the left calf and required surgery). All children had intellectual impairment, although all were able to talk and communicate their needs. Speech, language, and motor developmental

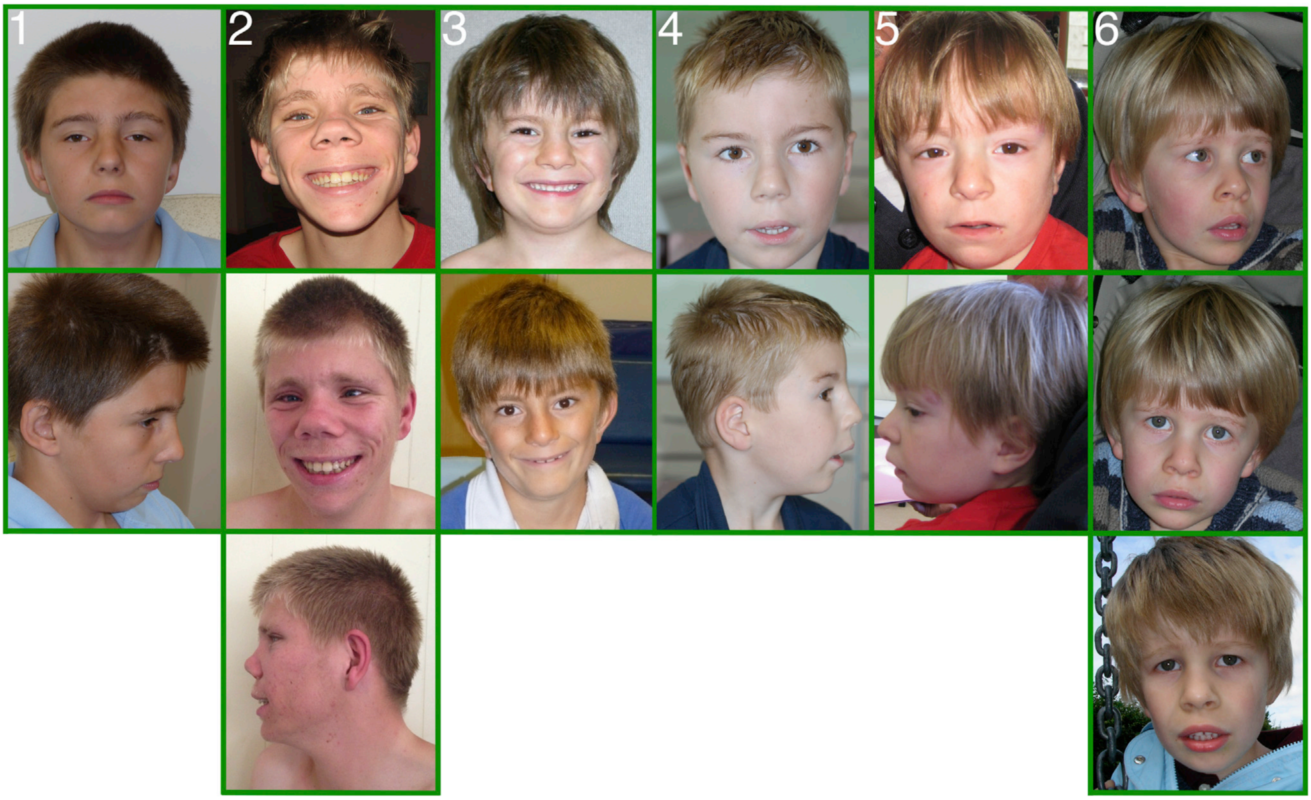

Figure 2. Facial Appearance of the Individuals with *SETD5* Mutations
Columns numbered 1–6 correspond to families 1–6, respectively, in Figure 1. Photographs from family 7 were unavailable for publication.

delay were noted in all individuals. Behavioral problems were a prominent feature of several of the children (5/7) and ranged from obsessive-compulsive disorder to hand flapping with ritualized behavior to features of autism. Involuntary movements and an exaggerated startle response were noted in several individuals, although in none were these a sustained feature over time. Older children had required special schooling because of their ID and behavioral problems, although some attended mainstream school at a young age but required educational statements and extra support. Growth parameters were within the normal range in all children, none had microcephaly or seizures, and all were born without antenatal or postnatal difficulties. Brain MRI was normal in one individual and was not performed in the remaining six individuals.

Although multiple LoF variants in *SETD5* have not been described previously in PubMed, this gene is one of three genes—*THUMPD3*, *SETD5*, and *THUMPD3-AS1* (a non-protein-coding gene)—within the critical region for 3p25 microdeletion syndrome.[12] Distal haploinsufficiency of chromosomal region 3p25 has long been associated with a clinical syndrome characterized by ID, low birth weight, microcephaly, telecanthus, ptosis, micrognathia, cleft palate, and congenital heart disease. Initially, the critical interval that defined the microdeletion syndrome was a 4.3 Mb region that was both large and gene dense.[13] It was not clear whether deletion of a single dosage-sensitive gene within this region was sufficient to cause the syndrome or whether the phenotype was a composite of multiple gene losses. Defining the minimum common overlap of deletions in multiple individuals has reduced the critical region for 3p25 microdeletion syndrome to three genes within a 124 kb interval.[12] The individuals with the smaller deletion within 3p25 have been reported to have a common phenotype of ID, hypotonia, a depressed nasal bridge, and a long philtrum. The presence of congenital heart disease and cleft palate is a more variable feature. Additional features also seen in the individuals reported with the smallest 3p25 microdeletion include synophrys, microcephaly, ptosis, abnormal palpebral fissures, postaxial polydactyly, scoliosis, cleft palate, gastrointestinal anomalies, and seizures.[12,14–16] The similarity between these individuals with haploinsufficiency of 3p25 and the individuals reported here to have *SETD5* truncating mutations is of note (Table 1). We suggest that similar to *EHMT1* mutations in 9q34 for Kleefstra syndrome (MIM 610253) and *KANSL1* mutations in 17q21 for Koolen-de Vries syndrome (MIM 610443), LoF mutations in *SETD5* might be sufficient to cause many of the features of 3p25 microdeletion syndrome.

Table 1. Clinical Features of the Individuals with *SETD5* Mutations and Comparison to Individuals with 3p25 Microdeletion Syndrome

Clinical Features	Individuals with *SETD5* LoF Mutations (n = 7)	Individuals with a 3p25 Deletion (n = 4)
Intellectual disability	7	4
Language delay and/or stammer	6	NA
Ritualized behavior and/or autism	5	NA
Seizures	0	2
Low birth weight and/or growth retardation	0	2
Microcephaly	0	2
Brachycephaly	3	NA
Low-set and/or malformed ears	5	3
Synophrys and/or abnormal eyebrows	5	1
Hypertelorism	0	1
Ptosis	1	2
Upslanting or downslanting palpebral fissures	6	1
Depressed nasal bridge	3	3
Abnormal nasal shape	7	NA
Long, smooth, and/or prominent philtrum	5	3
Thin upper lip	5	NA
Micrognathia	3	NA
Cleft palate	0	1
Postaxial polydactyly	1	1
Scoliosis or kyphosis	4	1
Leg-length discrepancy	2	NA
Feeding difficulties	5	NA
Congenital heart defects	2	2
Gastrointestinal and/or abdominal-wall anomalies	5	1

Data for the 3p25 deletion were adapted from Kellogg et al.[12] The following abbreviations are used: LoF, loss of function; and NA, not available.

Further evidence of the potential pathogenicity of mutations in *SETD5* is the observation of a single de novo LoF variant in an ID cohort and of de novo missense variants in two autism cohorts, although additional detailed phenotypic data have not been reported.[10,17,18]

SETD5 is a methyltransferase on the basis of sequence homology to other SET domain proteins.[19] It is highly conserved throughout mammalian species, suggesting that it is functionally important, although little is known yet of its specific role. *SETD5* is ubiquitously expressed, and especially high levels of *SETD5* expression have been noted in the brain.[20] On the basis of the other family members of this gene group, *SETD5* is likely to be important in the control of histone modification of DNA and to act as a regulator of transcription. Genes encoding methyltransferases specifically and genes encoding histone modifiers in general are increasingly recognized to have a major contribution to the phenotype of ID.[21] Genes encoding histone modifiers include *MECP2* (MIM 300005), *EHMT1* (MIM 607001), *NSD1* (MIM 606681), *KMT2D* (MIM 602113), *KDM6A* (MIM 300128), and *KDM5C* (MIM 314690).[22–28] These genes are all dosage sensitive, and haploinsufficiency alone is recognized to be sufficient to cause disease.[21]

Here, we present an analysis of children and young adults who were recruited to the Genetics of Learning Disability study with moderate to severe ID as the predominant clinical phenotype. This analysis provides sufficient evidence that loss of function of *SETD5* is a relatively frequent cause of ID and occurs as a rare de novo mutational event. The high number of LoF mutations in this cohort (7/996 [0.7%]) suggests that *SETD5* mutations, with a prevalence comparable to that of mutations in *ARID1B*,[29] might be one of the more common causes of ID. The affected individuals showed phenotypic similarity to those previously reported with a deletion in the critical region of 3p25. Prior to mutation analysis, the clinical features alone were not sufficient or consistent for clinicians to delineate this syndrome. In none of the individuals we report was a 3p25 microdeletion syndrome clinically suspected. Genotype-driven syndrome recognition is likely to be increasingly used in the future as more subtle phenotypes emerge. This, however, poses concerns of overinterpreting the phenotypic features and the need for large data sets for distinguishing rare pertinent phenotypic associations from rare incidental findings.

Supplemental Data

Supplemental Data include three tables and can be found with this article online at http://www.cell.com/ajhg.

Acknowledgments

We are indebted to all individuals who participated in the study. This study made use of data generated by the UK10K Project. Funding for the UK10K Project was provided by the Wellcome Trust under award WT091310. A full list of consortium members can be found at the UK10K Project website. This study was supported by grants from Action Medical Research, the Birth Defect Foundation, and the Cambridge National Institute for Health Research Biomedical Research Centre.

Received: January 7, 2014
Accepted: March 11, 2014
Published: March 27, 2014

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://www.1000genomes.org
Genetics Home Reference, Chromatin-modifying enzymes gene family, http://ghr.nlm.nih.gov/geneFamily/chromatinmodifyingenzymes
DECIPHER, http://decipher.sanger.ac.uk/
Ensembl Genome Browser, http://www.ensembl.org/index.html
Mutalyzer, https://mutalyzer.nl/index
NCBI, http://www.ncbi.nlm.nih.gov
NCBI HomoloGene, http://www.ncbi.nlm.nih.gov/homologene
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
Online Mendelian Inheritance in Man (OMIM), http://omim.org
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq
UCSC Genome Browser, http://genome-euro.ucsc.edu
UK10K Project, http://www.uk10k.org/

References

1. van Bokhoven, H. (2011). Genetic and epigenetic networks in intellectual disabilities. Annu. Rev. Genet. *45*, 81–104.
2. Piton, A., Redin, C., and Mandel, J.L. (2013). XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. Am. J. Hum. Genet. *93*, 368–383.
3. Xue, Y., Chen, Y., Ayub, Q., Huang, N., Ball, E.V., Mort, M., Phillips, A.D., Shaw, K., Stenson, P.D., Cooper, D.N., and Tyler-Smith, C.; 1000 Genomes Project Consortium (2012). Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. Am. J. Hum. Genet. *91*, 1022–1032.
4. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
5. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
6. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
7. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.
8. McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., and Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics *26*, 2069–2070.
9. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.
10. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet *380*, 1674–1682.
11. Kryukov, G.V., Pennacchio, L.A., and Sunyaev, S.R. (2007). Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am. J. Hum. Genet. *80*, 727–739.
12. Kellogg, G., Sum, J., and Wallerstein, R. (2013). Deletion of 3p25.3 in a patient with intellectual disability and dysmorphic features with further definition of a critical region. Am. J. Med. Genet. A. *161A*, 1405–1408.
13. Shuib, S., McMullan, D., Rattenberry, E., Barber, R.M., Rahman, F., Zatyka, M., Chapman, C., Macdonald, F., Latif, F., Davison, V., and Maher, E.R. (2009). Microarray based analysis of 3p25-p26 deletions (3p- syndrome). Am. J. Med. Genet. A. *149A*, 2099–2105.
14. Gunnarsson, C., and Foyn Bruun, C. (2010). Molecular characterization and clinical features of a patient with an interstitial deletion of 3p25.3-p26.1. Am. J. Med. Genet. A. *152A*, 3110–3114.
15. Peltekova, I.T., Macdonald, A., and Armour, C.M. (2012). Microdeletion on 3p25 in a patient with features of 3p deletion syndrome. Am. J. Med. Genet. A. *158A*, 2583–2586.
16. Riess, A., Grasshoff, U., Schäferhoff, K., Bonin, M., Riess, O., Horber, V., and Tzschach, A. (2012). Interstitial 3p25.3-p26.1 deletion in a patient with intellectual disability. Am. J. Med. Genet. A. *158A*, 2587–2590.
17. Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., et al. (2012). De novo gene disruptions in children on the autistic spectrum. Neuron *74*, 285–299.
18. Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature *485*, 242–245.
19. UniProt Consortium (2013). Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res. *41* (Database issue), D43–D47.
20. Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K., and Ohara, O. (2000). Prediction of the coding sequences of unidentified human genes. XIX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res. *7*, 347–355.
21. Berdasco, M., and Esteller, M. (2013). Genetic syndromes caused by mutations in epigenetic genes. Hum. Genet. *132*, 359–383.
22. Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat. Genet. *23*, 185–188.
23. Kleefstra, T., Brunner, H.G., Amiel, J., Oudakker, A.R., Nillesen, W.M., Magee, A., Geneviève, D., Cormier-Daire, V., van Esch, H., Fryns, J.P., et al. (2006). Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. Am. J. Hum. Genet. *79*, 370–377.
24. Kurotaki, N., Imaizumi, K., Harada, N., Masuno, M., Kondoh, T., Nagai, T., Ohashi, H., Naritomi, K., Tsukahara, M., Makita, Y., et al. (2002). Haploinsufficiency of NSD1 causes Sotos syndrome. Nat. Genet. *30*, 365–366.
25. Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C., et al. (2010). Exome sequencing

identifies MLL2 mutations as a cause of Kabuki syndrome. Nat. Genet. *42*, 790–793.

26. Hannibal, M.C., Buckingham, K.J., Ng, S.B., Ming, J.E., Beck, A.E., McMillin, M.J., Gildersleeve, H.I., Bigham, A.W., Tabor, H.K., Mefford, H.C., et al. (2011). Spectrum of MLL2 (ALR) mutations in 110 cases of Kabuki syndrome. Am. J. Med. Genet. A. *155A*, 1511–1516.
27. Lederer, D., Grisart, B., Digilio, M.C., Benoit, V., Crespin, M., Ghariani, S.C., Maystadt, I., Dallapiccola, B., and Verellen-Dumoulin, C. (2012). Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. Am. J. Hum. Genet. *90*, 119–124.
28. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian, G., Chelly, J., Fryns, J.P., et al. (2005). Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. Am. J. Hum. Genet. *76*, 227–236.
29. Hoyer, J., Ekici, A.B., Endele, S., Popp, B., Zweier, C., Wiesener, A., Wohlleber, E., Dufke, A., Rossier, E., Petsch, C., et al. (2012). Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. Am. J. Hum. Genet. *90*, 565–572.

The American Journal of Human Genetics, Volume *94*

Supplemental Data

De Novo Loss-of-Function Mutations in *SETD5*, Encoding a Methyltransferase in a 3p25 Microdeletion Syndrome Critical Region, Cause Intellectual Disability

Detelina Grozeva, Keren Carss, Olivera Spasic-Boskovic, Michael J. Parker, Hayley Archer, Helen V. Firth, Soo-Mi Park, Natalie Canham, Susan E. Holder, Meredith Wilson, Anna Hackett, Michael Field, James A.B. Floyd, UK10K Consortium, Matthew Hurles, and F. Lucy Raymond

Table S1: List of the 565 Genes that were Selected for Targeted Next Generation Sequencing

Gene IDs
ABCD1
ACBD6
ACE2
ACIN1
ACOT9
ACSL4
ACTL6A
ACTL6B
ACY1
ADCK3
ADK
ADRA2B
ADSL
AFF2
AGA
AGTR2
AIMP1
AKAP17A
AKAP4
ALDH18A1
ALDH4A1
ALDH5A1
ALG1
ALG12
ALG13
ALG3
ALG6
ALG8
ANK3
AP1S2
AP4B1
AP4E1
AP4M1
AP4S1
ARFGEF2
ARG1
ARHGAP36
ARHGAP6
ARHGEF4
ARHGEF6
ARHGEF9
ARID1A

ARID1B
ARID2
ARIH1
ARL14EP
ARSF
ARX
ASB12
ASCC3
ASCL1
ASH1L
ASMT
ASMTL
ASXL1
ATM
ATP2B3
ATP7A
ATRX
ATXN3L
AUH
AVPR2
AWAT2
BCOR
BCORL1
BDP1
BMP15
BRAF
BRWD3
BTK
C12orf57
CA8
CACNA1F
CACNA1G
CAMK2A
CAMK2G
CAP1
CAPN10
CASK
CASP2
CC2D1A
CC2D2A
CCDC22
CCDC23
CCNA2
CCNB3
CD99
CDH15

CDK16
CDK8
CDKL5
CEP41
CFP
CHD2
CHD7
CHL1
CLCN4
CLCN5
CLIC2
CMC4
CNKSR1
CNKSR2
CNTNAP2
COL4A3BP
COL4A6
COQ5
COX10
CPXCR1
CREBBP
CRLF2
CSF2RA
CSTF2
CTNNB1
CTPS2
CTSD
CTTNBP2
CUL4B
CUX2
CXORF22
CXORF58
CYP7B1
DCHS2
DCX
DDOST
DDX26B
DDX3X
DDX53
DEAF1
DGKH
DHCR7
DHRSX
DHX30
DIAPH2
DKC1

DLG1
DLG2
DLG3
DLG4
DMD
DNMT3B
DOCK11
DPF1
DPF2
DPF3
DYNC1H1
DYRK1A
EEF1A2
EEF1B2
EHMT1
EIF2C1
EIF2S3
ELK1
ELP2
ENOX2
ENTHD2
ENTPD1
EP300
EPPK1
ERCC6
ERLIN2
ESX1
EXOSC3
FAAH2
FAM120C
FAM47B
FAM58A
FASN
FGD1
FKBPL
FKRP
FKTN
FLNA
FMR1
FOXG1
FOXP1
FRMPD4
FRY
FTL
FTSJ1
GAB3

GABRQ
GAD1
GATAD2B
GCDH
GCH1
GDI1
GJC2
GK
GLB1
GLRA2
GM2A
GON4L
GPC3
GPR112
GPR56
GPRASP1
GRB14
GRIA1
GRIA2
GRIA3
GRIK2
GRIN2A
GRIN2B
GSPT2
GTPBP8
HAUS7
HCCS
HCFC1
HDAC4
HDAC8
HDHD1
HEXA
HEXB
HGSNAT
HIST1H4B
HIST3H3
HIVEP2
HPRT1
HRAS
HS6ST2
HSD17B10
HSPD1
HUWE1
IDS
IDUA
IFNAR2

IGSF1
IKBKG
IL1RAPL1
IL3RA
INPP4A
INPP5E
IQSEC2
ITGA4
ITIH6
KANK1
KANSL1
KAT6B
KCNC3
KCND1
KCNH1
KCNK12
KCNQ3
KDM1A
KDM5A
KDM5C
KDM6B
KIAA2022
KIF1A
KIF26B
KIF4A
KIF5C
KIF7
KIRREL3
KLHL15
KLHL21
KLHL34
KLHL4
KMT2D
KRAS
L1CAM
LAMA1
LAMP2
LARP7
LAS1L
LHFPL3
LIMK1
LINS
LRP1
LRP2
LRRK1
MAGEA11

MAGEB1
MAGEB10
MAGEB2
MAGEC1
MAGEC3
MAGED1
MAGEE2
MAGIX
MAGT1
MAN1B1
MAOA
MAOB
MAP2K1
MAP2K2
MAP3K15
MAP7D3
MBD5
MBNL3
MECP2
MED12
MED17
MED23
MEF2C
MGAT5B
MIB1
MID1
MLC1
MLH1
MLL3
MLYCD
MMAA
MMAB
MMADHC
MORC4
MSL3
MTF1
MTMR1
MTMR8
MXRA5
MYO1D
MYO1G
MYT1L
NA
NAA10
NDE1
NDP

NDST1
NDUFA1
NECAB2
NEU1
NF1
NFIX
NHS
NKAP
NLGN3
NLGN4X
NR1I3
NRK
NRXN1
NRXN2
NSD1
NSDHL
NSUN2
NTM
NXF4
NXF5
OCRL
ODF2L
OFD1
OGT
OPHN1
OR5M1
OTC
OXCT1
P2RY4
P2RY8
PABPC5
PAFAH1B1
PAH
PAK3
PARP1
PASD1
PAX6
PBRM1
PC
PCDH10
PCDH19
PCNT
PDHA1
PECR
PEPD
PGK1

PGRMC1
PHACTR1
PHF10
PHF6
PHF8
PHIP
PHKA1
PIGN
PIK3C3
PIN4
PJA1
PLA2G6
PLCXD1
PLP1
PLXNB3
PNKP
POLA1
POLR3A
POLR3B
PORCN
PPP2R5D
PPT1
PQBP1
PRDX4
PRICKLE3
PRMT10
PROX2
PRPS1
PRRG1
PRRG3
PRRT2
PRSS12
PSMA7
PSMD10
PTCHD1
PTEN
PTPN11
PTPN21
RAB39B
RAB3GAP1
RAB40AL
RABL6
RAF1
RAI1
RALGDS
RAPGEF1

RBM10
RENBP
RGAG1
RGN
RGS7
RLIM
RNASET2
RPGR
RPS6KA3
SATB2
SCAPER
SCN2A
SCN8A
SETBP1
SETD5
SETDB2
SGSH
SHANK1
SHANK2
SHANK3
SHOC2
SHOX
SHROOM2
SHROOM4
SLC12A6
SLC16A2
SLC25A22
SLC25A53
SLC25A6
SLC26A9
SLC2A1
SLC31A1
SLC6A1
SLC6A17
SLC6A8
SLC9A6
SMARCA2
SMARCA4
SMARCB1
SMARCC1
SMARCC2
SMARCD1
SMARCD2
SMARCD3
SMARCE1
SMC1A

SMS
SNTG1
SOS1
SOX3
SOX5
SPG11
SPRED1
SPRY3
SPTAN1
SPTLC2
SREBF2
SRGAP3
SRPX2
ST3GAL3
STAB2
STAG1
STARD8
STXBP1
SYN1
SYNCRIP
SYNE1
SYNGAP1
SYP
SYT1
SYTL4
SYTL5
TAF1
TAF2
TAF7L
TANC2
TAT
TBC1D24
TBC1D8B
TCEAL3
TCF4
TCP10L2
TENM1
THAP1
THOC2
ThumpD1
TIMM8A
TKTL1
TLR8
TM4SF2
TMEM132E
TMEM135

TMLHE
TNKS2
TNPO2
TRAPPC9
TREX2
TRIO
TRMT1
TSC1
TSC2
TSC22D3
TSEN2
TSEN34
TSEN54
TSPAN7
TTI2
TUBA1A
TUBA8
TUBAL3
TUBB2B
TUSC3
UBE2A
UBE3A
UBR1
UBR7
UBTF
UPF3B
USP27X
USP9X
UTP14A
VAMP7
VLDLR
VPS13B
VRK1
WAC
WDR11
WDR13
WDR45L
WDR62
WNK3
WWC3
XIAP
XKRX
YY1
ZBTB40
ZC3H14
ZCCHC12

ZCCHC8
ZDHHC15
ZDHHC9
ZEB2
ZFHX4
ZFX
ZFYVE26
ZMYM3
ZMYM6
ZMYND12
ZNF238
ZNF41
ZNF425
ZNF526
ZNF674
ZNF711
ZNF81

Table S2: Rationale for Prioritisation of Genes for Further Investigation

Top ranked genes according to number of rare variants in the ID cohort
(frequency <1% 1000 genomes, UK10K twins cohort, NHLBI GO Exome Sequencing Project Exome Variant Server (NHLBI EVS),
internal cohort 2172 individuals where whole exomes were sequenced at the same laboratory (UK10K) and the UK10K rare replication cohort itself (including all phenotypes))

LoF=Loss of function

ID gene	Number observed LoF variants	Frequency in ID cohort (%)
DCHS2	22	2.2
SYNE1	13	1.3
VPS13B	10	1.0
MIB1	9	0.9
NF1	9	0.9
ATM	8	0.8
PAH	7	0.7
PCDH10	7	0.7
SETD5	7	0.7
ASCC3	6	0.6
ATRX	6	0.6
UTP14A	6	0.6
HEXA	6	0.6
CC2D2A	6	0.6
STAB2	6	0.6

There are 15 genes in the ranking 1-10 as six genes were =10th with six observed LoF variants

The table from above was annotated with information as to how many of the variants are independent or if the same variants have been seen in multiple individuals
The criterion for independence was selected as it is unlikely that recurrent LoF variants in a gene will cause ID within the cohort

ID gene	Number observed LoF variants	Frequency in ID cohort (%)	How many are independent LoFs?	Number Independent LoFs
DCHS2	22	2.2	6 variants seen once; 1 variant seen 2 times; 1 seen 6 times; 1 seen 8 times	9
SYNE1	13	1.3	3 variants seen once; 1 variant seen 2 times; 1 variant seen 8 times	5
VPS13B	10	1.0	5 variants seen once; 1 variant seen 2 times; 1 variant seen 3 times	7
MIB1	9	0.9	5 variants seen once; 2 variants seen 2 times	7
NF1	9	0.9	1 variant seen once; 1 variant seen 2 times; 1 variant seen 6 times	3
ATM	8	0.8	6 variants seen once; 1 variant seen 2 times	7
PAH	7	0.7	4 variants seen once, 1 variant seen 3 times	5
PCDH10	7	0.7	1 variant seen 7 times	1
SETD5	7	0.7	7 variants seen once	7
ASCC3	6	0.6	6 variants seen once	6
ATRX	6	0.6	3 variants seen once; 1 variant seen 3 times	4
UTP14A	6	0.6	1 variant seen 6 times	1
HEXA	6	0.6	2 variants seen once; 2 variants seen 2 times	4
CC2D2A	6	0.6	6 variants seen once	6
STAB2	6	0.6	6 variants seen once	6

Sorted according to "Number Independent LoFs" column; Further information was added about the mode of inheritance, if it is a known or candidate gene, frequency of LoF variants in the NHLBI Exome sequencing Project
and the reason why the corresponding gene is excluded from further investigation

ID gene	Number Independent LoFs	Mode of inheritance	Known or candidate gene	Frequency NHLBI EVS LoF	Reason exclusion further analysis
DCHS2	9	Unknown	Candidate	13	Difficult to make judgement about possible pathogenicity of LoF variants in this gene as many were observed a few times; in addition 13 LoFs in NHLBI EVS
SETD5	7	Unknown	Candidate	1	
ATM	7	Recessive	Known	8	Known gene; recessive inheritance
VPS13B	7	Recessive	Known	17	Known gene
MIB1	7	Unknown	Candidate	13	Difficult to make judgement about possible pathogenicity of LoF variants in this gene as there are 12 LoF variants observed in NHLBI EVS
ASCC3	6	Recessive	Known	7	Known gene; recessive inheritance
CC2D2A	6	Recessive or Autosomal Dominant	Known	8	Known gene
STAB2	6	Unknown	Candidate	12	Difficult to make judgement about possible pathogenicity of LoF variants in this gene as there are 12 LoF variants observed in NHLBI EVS
PAH	5	Recessive	Known	4	Known gene
SYNE1	5	Recessive	Known	17	Known gene
ATRX	4	Hemizygous	Known	0	Known gene
HEXA	4	Recessive	Known	4	Known gene
NF1	3	Autosomal Dominant	Known	3	Known gene
PCDH10	1	Unknown	Candidate	1	No independent variants in this gene
UTP14A	1	Unknown	Candidate	0	No independent variants in this gene

SETD5 was selected for further investigation based on:
One of the genes with high number of independent LoFs
Candidate gene, not previously implicated in ID
Only one LoF observed in the publicly available data from NHLBI EVS

Table S3: Clinical Features of the Individuals with *SETD5* Mutations

FAMILY	**1**	**2**	**3**	**4**	**5**	**6**	**7**
birth weight (kg)	2.47	2.69	2.99	3.66	2.41	2.95	small
gestation (weeks)	34	38	term	term	35+5	term	term
Recent height (percentile)		50-75th	2nd		25-50th		
Recent weight (percentile)			9th		25-50th		
Recent Head Circumference (percentile)	25th	75-98th	50-75th	10-25th	75-91st	75th	10th

SPINE and SKELETON							
leg length discrepancy	y	y					
shortened 4th and 5th metacarpal		y					
hypoplasia of left calf		y					
scoliosis or kyphosis	y	y					
lordosis			y			y	
sacral dimple		y			y		
stiff legged gait					y	y	
bilateral 5th finger clinodactyly			y				
brachdactyly			y				
post axial polydactyly; 2 hands, 1 foot				y			

EARS							
large ears	y		y				
fleshy ear lobes		y			y		
long, narrow, low set ears			y				y
preauricular pit	y						

EYEBROWS							
full eyebrows		y					
synophrys	y		y				y
straight eyebrows	y						
broad eyebrows							y
cysts in eyebrows					y		

HEAD SHAPE							
brachycephaly	y	y					y
prominent high forehead					y		

NOSE							
broad, thickened upturned nasal tip		y			y		
depressed nasal bridge		y			y	y	
anteverted nares		y			y		
prominent high nasal root			y	y	y		y
tubular nose	y			y	y	y	
prominent nares			y				

EYES							
left eye amblyopia		y					
long narrow fissures			y		y		
mild ptosis					y		
nystagmus and strabismus				y			
down slanting palpebral fissures							y
upslanting palpebral fissures	y		y	y	y	y	

MOUTH and LOWER FACE							
long, smooth philtrum	y	y	y	y	y		
small mouth				y			y
short philtrum							y
micrognathia	y			y			y
thin upper lip	y	y	y		y		y
high palate	y						y

FEEDING AND SWALLOWING							
feeding difficuties	y	y					
crowded teeth		y				y	y
dribbling					y	y	y
difficulty chewing, oromotor dyspraxia					y	y	
swallowing difficulties		y				y	

BEHAVIOUR and DEVELOPMENT							
developmental delay	y	y	y	y	y	y	y
walking (age)	y (2yrs)	y (3yrs)	y (18mths)		y (2yrs)	y (3yrs 2mths)	y (20mths)
speech (age first words)	y (4yrs)	y (4 yrs)	y (12mths)	y (late)	y (18 mths)	y (2 years)	
expressive language delay					y	y	y
stammer	y		y	y			
exaggerated startle response						y	y
involutary movements			y			y	y (until 10yrs)
hand flapping and ritualised behaviour	y	y			y		
autistic	y					y	
obsessive compulsive disorder	y					y	y

CONGENITAL HEART DISEASE							
mitral valve prolapse		y					
VSD, PDA	y						

ABDOMINAL ORGAN DEVELOPMENT							
paraumbilical hernia					y		
inguinal hernia		y					y
undescended testes		y			y		
hypospadias	y		y				
nocturnal enuresis							y

OTHER							
fetal finger pads		y					
spiky hair		y					
saggy skin					y		
low hairline			y			y	y
severe constipation			y			y	