Histone lysine methylases and demethylases in the landscape of human developmental disorders.

Víctor Faundes,[1,2] William G. Newman,[1,3] Laura Bernardini,[4] Natalie Canham,[5] Jill Clayton-Smith,[1,3] Bruno Dallapiccola,[6] Sally J. Davies,[7] Michelle K. Demos,[8] Amy Goldman,[3] Harinder Gill,[9] Rachel Horton,[10] Bronwyn Kerr,[3] Dhavendra Kumar,[7] Anna Lehman,[9] Shane McKee,[11] Jenny Morton,[12] Michael Parker,[13] Julia Rankin,[14] Lisa Robertson,[15] Karen Temple,[10] Clinical Assessment of the Utility of Sequencing and Evaluation as a Service (CAUSES) study,[9] The Deciphering Developmental Disorders (DDD) study,[16] Siddharth Banka,[1,3,*]

1. Manchester Centre for Genomic Medicine, Division of Evolution & Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9WL, UK.

2. Laboratorio de Genética y Enfermedades Metabólicas, Instituto de Nutrición y Tecnología de los Alimentos (INTA), Universidad de Chile, Santiago 7830490, Chile.

3. Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation NHS Trust, Health Innovation Manchester, Manchester M13 9WL, UK.

4. Cytogenetics Unit, Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo FG 71013, Italy.

5. North West Thames Regional Genetics Service, London North West Healthcare NHS Trust, Northwick Park Hospital, Watford Road, Harrow HA1 3UJ, UK.

6. Scientific Directorate, Bambino Gesù Children Hospital, IRCCS, Rome 00146, Italy.

7. Institute of Medical Genetics, University Hospital of Wales, Heath Park, Cardiff CF14 4XW, UK.

8. Division of Pediatric Neurology, Department of Pediatrics, University of British Columbia, Vancouver, BC V6H 3V4, Canada.

9. Department of Medical Genetics, Children's & Women's Health Centre of British Columbia, Vancouver, BC V6H 3N1, Canada.

10. Wessex Clinical Genetics Service and Division of Human Genetics, Princess Anne Hospital, Southampton SO16 5YA, UK.

11. Northern Ireland Regional Genetics Service, Belfast City Hospital, Belfast HSC Trust, Belfast BT9 7AB, UK.

12. West Midlands Regional Clinical Genetics Service and Birmingham Health Partners, Birmingham Women's and Children's Hospital NHS Foundation Trust, Birmingham B15 2TG, UK.

13. Sheffield Clinical Genetics Service, Sheffield Children's Hospital NHS Foundation Trust, Western Bank, Sheffield S10 2TH, UK.

14. Peninsula Clinical Genetics Service, Exeter EX1 2ED, UK.

15. North of Scotland Regional Genetics Service, NHS Grampian, Department of Medical Genetics Medical School, Foresterhill, Aberdeen, AB25 2ZD, UK.

16. Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK.

*** Correspondence**

Dr Siddharth Banka

Manchester Centre for Genomic Medicine

St Mary's Hospital

Manchester M13 9WL. U.K.

Tel: +44 (0) 161 70 10980; Fax: +44 (0) 161 27 66145

siddharth.banka@manchester.ac.uk.

ABSTRACT

Histone lysine methyltransferases (KMT) and demethylases (KDM) underpin gene-regulation. Here we demonstrate that variants causing haploinsufficiency of KMTs and KDMs are frequently encountered in individuals with developmental disorders. Using a combination of human variation databases and existing animal models we determine 22 KMTs and KDMs as additional candidates for dominantly inherited developmental disorders. We show that KMTs and KDMs that are associated with, or are candidates for, dominant developmental disorders tend to have higher level of transcription, longer canonical transcripts, more interactors and a higher number and types of post-translational modifications than other KMT/KDMs. We provide evidence to firmly associate *KMT2C*, *ASH1L* and *KMT5B* haploinsufficiency with dominant developmental disorders. While *KMT2C* or *ASH1L* haploinsufficiency results in predominantly neurodevelopmental phenotype with occasional physical anomalies, *KMT5B* mutations cause an overgrowth syndrome with intellectual disability. We further expand the phenotypic spectrum of *KMT2B* related disorders and show that some individuals may have severe developmental delay without dystonia at least till mid-childhood. Additionally, we describe a recessive histone lysine methylation defect caused by homozygous or compound heterozygous *KDM5B* variants resulting in a recognizable syndrome with developmental delay, facial dysmorphism and camptodactyly. Collectively, these results emphasize the significance of histone lysine methylation in normal human development and the importance of this process in human developmental disorders. Our results demonstrate that systematic clinically-oriented pathway-based analysis of genomic data can accelerate the discovery of rare genetic disorders.

MAIN TEXT

Post-translational methylation and demethylation of lysine residues on histone tails is a key dynamic chromatin modification that is mediated by specific methyltransferases (KMTs) and demethylases (KDMs) and underpins gene regulation and several cellular processes[1; 2]. Twenty-seven KMT and 24 KDM encoding genes, classified into eight groups each, are known (Table S1) [3]. Of these, heterozygous variants in seven KMT and four KDM genes are associated with autosomal and X-linked dominant inherited human developmental disorders (DDs) in the Online Mendelian Inheritance in Man database (OMIM) (Table S1)[4-18].

We reviewed published disease-causing variants in KMTs and KDMs in the Human Gene Mutation Database[19] and deduced that ~75% of these were predicted to be heterozygous protein truncating variants (PTVs), suggesting that haploinsufficiency is the predominant mechanism for the associated diseases (Figure 1A) (Table S2). This is consistent with previous studies that have shown a high prevalence of *de novo* (DN) PTVs in dominant DDs[20-22]. We reviewed phenotypes of the available mouse models for KMT and KDM orthologs (Table S1)[23] and found that heterozygous mouse models for six of the 11 known dominant DD-associated KMTs/KDMs and 12 of the 40 of remaining KMTs/KDMs demonstrate anomalies. We reviewed phenotypes of the available zebrafish knockdown (KD) models for KMT and KDM orthologs (Table S1)[24]. Anomalies were observed in KD of seven of the 11 known dominant DD-associated KMTs/KDMs and 18 of the 40 of remaining KMTs/KDMs. The human mutational landscape of KMTs/KDMs and the information from animal models led us to hypothesize that germline heterozygous PTVs in additional KMTs/KDMs may underlie as yet unknown DDs.

For each of the 51 KMTs/KDMs, we compiled selected indices of predicted intolerance to loss of function (LoF) pathogenic variants (Table S1). The pLI (probability of being LoF Intolerant) scores obtained from ExAC Browser[25] were found to be within a narrow range of 0.99-1.0 for KMTs and KDMs already linked with dominant human DDs suggesting a high reliability. The ranges of Residual Variation Intolerance score (0.06-51.92) and Haploinsufficiency Index (3.06-62.96) scores for these genes were broad[26; 27]. We used a pLI score[25] cut-off of >0.9 to determine additional 11 KMTs and 11 KDMs as candidates for as yet unknown dominant human DDs (Figure 1B).

We examined the data from 4,293 trios who underwent exome sequencing as part of the Deciphering Developmental Disorders (DDD) study[22]. All these procedures were in accordance with the ethical standards (Multi-Centre Research Ethics Committee approval 10/H0305/83 and GEN/284/12) and informed consent was obtained from all the participants. The previously described pipeline was used to identify rare high-quality and possibly deleterious variants in our list of 51 KMTs/KDMs. Rare variants were defined as those with minor allele frequencies of <0.001 (for *de novo*, X-linked and dominant heterozygous inheritances) or <0.01 (for compound heterozygous, recessive homozygous) in the Exome Aggregation Consortium [25] (ExAC, Version 0.3.1), the 1000 Genomes Project (1K-G) [28], Ensembl version 80-GRCh37, NHLBI-GO Exome Sequencing Project (ESP) [29], and UK10K [30]. High quality variants were defined with read depth of >20 and a genotype quality score of >20. Truncating or missense variants in canonical transcripts were defined to be possibly deleterious. In total, we identified 218 probands with high-quality rare variants in the 51 KMTs/KDMs (Figure S1) (Tables S3, S4 and S5). Of these, 65 (~1.5% of all the probands) affected individuals had likely causal monoallelic LoF variants (Figure 1C) (Table S3) in the 11 KMTs/KDMs already associated with dominant DDs. Of note, the combined coding size

of the canonical transcripts of these 11 genes is ~0.3% out of the total human exome size (0.092/30 Mb)[31](Figure 1C). Hence, this is an important group of genes in rare undiagnosed developmental disorders. Fifty-two out of these 218 affected individuals had likely benign variants or variants of uncertain significance in these 11 KMTs/KDMs (Figure 1C) (Table S4). [22; 31-33]

One hundred and two of 218 probands had 120 rare high-quality call genetic variants in KMTs/KDMs not yet firmly associated with DDs (Table S5). Of these, 83 variants were in our 22 candidates for dominant DDs, including 9 PTVs and 16 DN protein-altering variants (PAV) (Table S5). Chi-square test revealed a 1.87-fold enrichment (95% Confidence Interval [CI]=0.93-3.76; p=0.072) in the frequency of PTVs in these 22 genes in our cohort against the data from ExAC[25] (Table S6). Similarly, a 4.85-fold enrichment of DN PAVs (95%CI=1.78-13.26; p =0.00065) was observed in these 22 genes in our cohort against the entries marked as 'controls' in "denovo-db"[34] (Table S6).This observation supported our hypothesis that germline heterozygous PTVs in additional KMTs/KDMs may underlie as yet unknown dominant DDs.

We then focused on DN PTVs in our curated list of candidates KMT for dominant DD because these variants are highly likely to be causal (equivalent to category 1 in the American College of Medical Genetics and Genomics guidelines[35]). We interrogated the vcf files of each trio through VarSeq® version 1.3.4 (Golden Helix, Inc., Bozeman, MT) to ensure that the probands did not carry additional causal pathogenic variants in other genes. Collectively, we identified seven variants that fulfilled these criteria. (Table 1) (Table S5). Specifically, these included two DN PTVs each in *ASH1L* (OMIM #607999), *KMT2C* (OMIM # 606833), *KMT5B* (formerly known as *SUV420H1*) (OMIM # 610881) and one in *KMT2B* (OMIM #

606834) (Figure 2). Chi-square test revealed a 5.51-fold enrichment (95% Confidence Interval [CI]=2.3-13.2; p=0.0000165) in the frequency of PTVs in these four genes in our cohort against the data from ExAC[25]. Fisher's exact test revealed a 34.87-fold enrichment of DN PAVs (95%CI=2.0545 to 591.9943; p=0.000039) in these four genes in our cohort against the entries marked as 'controls' in "denovo-db"[34], further supporting a high likelihood of causality. Where possible, variants were confirmed by Sanger sequencing (Table S8) (Figure S2). Importantly, rare variants in these genes have been previously reported in several cases-controls cohorts of individuals with autism, ID, bipolar disorder and congenital heart anomalies but their causality has not been confirmed and the associated phenotypes have not been fully described[20; 36-43]. Detailed phenotype information of the affected individuals was, therefore, collected (Table 1) (Figure 3) (Supplemental Note: case reports).

Of note, we also detected (a) non-truncating DN PAVs in other candidate KMTs and KDMs for dominant DDs (*DOT1L*, *KDM3A*, *PRDM2, SETDB1*). There is insufficient evidence for causality of PAVs in these genes at present; (b) DN PTVs in non-candidate KMTs and KDMs for dominant DDs (*KDM5B* and *SETD1B*). PTVs in these genes could be coincidental or they could be phenotype modifiers in some affected individuals or they could be non-penetrant in some unaffected individuals in the general population; and (c) PTVs in other candidate KMTs and KDMs for dominant DDs (*KDM3A* and *PRDM2*) inherited from a parent who did not share the proband's phenotype. This observation suggests that these PTVs may have incomplete penetrance or that these genes are not haploinsufficiency intolerant, unlike as predicted by their pLI scores (Figure 2). Overall, further studies are needed to determine the pathogenicity of heterozygous PAVs and PTVs in *DOT1L*, *KDM3A*, *KDM5B, PRDM2, SETDB1* and *SETD1B*.

Next, we interrogated the data from >200 individuals from the CAUSES study of children with developmental disorders[44] for potentially pathogenic variants in *KMT2B*, *KMT2C* and *KMT5B*, and identified one additional individual with a DN PTV in *KMT2C* (Table 1; Figure 2).

Copy number variants (CNVs) can be informative in dissecting the molecular basis of genetic disorders[45-47]. We, therefore, examined the DECIPHER database[48] with >41,800 individuals with CNVs, and identified 71 deletions encompassing one of the four genes - *ASH1L, KMT2B*, *KMT2C* or *KMT5B* (Table S7). Where possible, additional detailed phenotype information of the affected individuals was collected (Table 1) (Figure 3) (Supplemental Note: case reports). Of note, only individuals whose deletions did not include other possibly causal DD-related gene(s) were considered for further analysis.

Collectively, we identified three individuals with DN *KMT2C* PTV and 41 deletions encompassing this gene (Tables 1, S5 and S7) (Figure 2 and 3). All affected individuals, for whom detailed clinical information was available, had severe developmental delay and ID (Table 1). KMT2C is a H3K4 methyltransferase[49] that is highly expressed in the developing and adult human brain, specially in the cerebellum[50; 51]. It is interesting to note that individual 3 has hypoplasia of the cerebellar vermis. In mice, a homozygous *Kmt2c* inframe deletion of exons 25 and 26 has been shown to result in partial embryonic lethality and prenatal and postnatal growth retardation[52].

We identified two individuals with *ASH1L* PTVs and five deletions encompassing this gene (Table 1, S5 and S7) (Figure 2). All affected individuals, for whom detailed clinical information was available, displayed variable degrees of global developmental delay or ID,

seizures, hypotonia and aberrant behaviour (Table 1). ASH1L is a methyltransferase that catalyzes mono and di-methylation of H3K36[53]. *ASH1L* is highly expressed in both embryonic and adult human brains[50; 51]. Injection of *ash1a* morpholinos in zebrafish led to a reduction in the number of neurons produced in the epiphysis[54]. Heterozygous and homozygous knock-in mice expressing mutant *Ash1l* containing a short in-frame deletion within the catalytic SET domain display a range of skeletal anomalies[55]. Hypomorphic mice, with an exon 1 *Ash1l* gene trap outside the catalytic SET domain, have reduced levels of normal protein and display impaired fertility[56]. The heterozygous mice for a reporter-tagged deletion allele show impaired pupillary reflex and abnormal coat appearance[23].

We identified two individuals with DN *KMT5B* PTVs and seven deletions encompassing this gene (Tables 1, S5 and S7) (Figure 2 and 3). All affected individuals, for whom detailed clinical information was available, had mild to moderate global developmental delay and ID, macrocephaly, tall stature and similar facial dysmorphism (Table 1). KMT5B is a H4K20 di- and tri- methyltransferase that promotes transcriptional repression[57]. *KMT5B* is highly expressed in both embryonic and adult human brains[50; 51]. The *Kmt5b*-null mice die at embryonic stages, have decreased body length and weight[58], whereas the heterozygous mice have decreased body weight and fat, and vertebral anomalies[23].

We identified one DN heterozygous frameshift and three missense *KMT2B* variants and 18 deletions encompassing this gene (Tables 1, S5 and S7) (Figure 2). All but one of these individuals were recently reported in a study demonstrating PTVs and deletions in this gene associated with childhood-onset dystonia 28 (OMIM # 617284)[59; 60]. The only previously unpublished individual in this cohort is a girl (Table 1) with a *de novo* p.Leu604Profs*72 frameshift variant and severe global developmental delay and additional features (Figure 3).

Importantly, in contrast with the previously described individuals, this girl did not show any evidence of dystonia by the age of 11 years, even with careful reverse phenotyping[61]. Interestingly, some of the previously reported individuals had normal development[59; 60]. Our findings broaden the phenotype of *KMT2B* variants and show that any combination of developmental delay and dystonia can result from heterozygous PTVs in this gene. *KMT2B* is highly expressed in both embryonic and adult human brains [50; 51]. In adults, it is specifically high-expressed in pituitary, cerebellum and bladder[50]. Of note, the affected individual that we describe has growth hormone deficiency, abnormal gait, nystagmus and urinary incontinence. The *Kmt2b* KO mice die before stage E11.5 and display growth retardation, neural tube defects, pericardial effusion, abnormal heart looping, and head abnormalities, whereas the heterozygous mice exhibit fasting hyperinsulinemia, glucose intolerance and fatty liver disease[62; 63].

Next, we systematically explored the differences between the gene/protein-attributes and expression patterns between 33 KMTs/KDMs that are known/candidates for dominant DD and the other 18 KMTs/KDMs using the UniProtKB, GTEx and BrainSpan databases[50; 51; 64]. Mann-Whitney tests were performed with an exact p-value <0.05 considered as significant. Candidate/known dominant DD KMTs/KDMs had longer canonical transcripts, greater number of interactors and a significantly higher number and types of post-translational modifications (adjusted for protein length) (Figure 4) (Tables S9 and S10)[64]. These distinctions are maintained independently for both KMTs and KDMs. This observation is consistent with general properties of genes that are considered to be haploinsufficient (HI)[27] and suggests that candidate/known dominant DD KMTs/KDMs are likely to be key players performing multiple roles in embryogenesis. Similarly, the expression of candidate/known dominant DD KDMs was found to be significantly higher in almost all fetal brain structures

and adult human tissues when compared to other KDMs (Tables S12 and S14) (Figure 4F)[50]which agrees with previous observations regarding HI genes[27]. However, surprisingly we did not find a significant difference between the expression of the candidate/known dominant DD KMTs versus other KMTs in most human tissues. Exceptions were certain brain-areas where the candidate/known dominant DD KMTs are significantly highly expressed before the 10th post-conceptional week (Tables S11 and S13) (Figure 4F)[51]. Further studies will be needed to confirm these unexpected findings. One possibility is that the KMTs that were classified in this study as not being candidates for dominant DDs may be candidates for adult-onset phenotypes. Alternatively, these results may reflect technical limitations of large-scale gene expression experiments such as lack of cell-type level resolution.

Lastly, we turned our focus to test the hypothesis that recessive disorders associated with biallelic variants in some KMTs/KDMs may exist. This hypothesis was based on our observation that five KMT and two KDM homozygous knockout mice are viable, but show multiple anomalies (Table S1). In the cohort of 4,293 subjects from the DDD study, we identified 27/102 probands with bi-allelic variants in KMTs/KDMs. On subsequent analyses, most of these were considered likely non-deleterious. However, one individual had bi-allelic homozygous *KDM5B* (OMIM #605393) PTVs (Table S5) (Figures 2 and 3) and severe global developmental delay (Table 1, Supplemental Note: case reports). Fisher's exact test revealed a 96.89-fold enrichment of homozygous PTVs (95%CI= 3.95 to 2378.87; p=0.03) in *KDM5B* in our cohort against the data from gnomAD[25]. Additionally, no homozygous *KDM5B* knockout genotype was seen in 3,222 adults with high parental relatedness,[65] and the knockout *Kdm5b* mice die prematurely due to respiratory failure, and display disorganized cranial nerves, defects in eye development, increased incidences of exencephaly, and skeletal anomalies [66]. Next, we examined exome data from 5,332 additional individuals from the

DDD study and identified two further individuals with bi-allelic *KDM5B* PTVs and striking overlapping phenotype of severe global developmental delay, camptodactyly and overlapping facial dysmorphism (Table 1, Figures 2 and 3). Hence, bi-allelic *KDM5B* LoF variants cause a recessive DD. KDM5B is a H3K4 demethylase, which modulates RNA polymerase II initiation and elongation rates, and alternative splicing in embryonic stem cells[67].

Overall our results demonstrate the importance of defects in histone lysine methylation in human DDs. In particular, variants in six of eight KMT2 methyltransferases can now be considered to result in dominant DDs [5; 8; 14; 36; 43; 59; 60; 68-71]. KMT2 genes encode enzymes that mono-, di- and/or trimethylate the H3K4[1; 72], and mark active promoters and enhancers[73]. Our observation emphasizes the significance of the correct dosage of KMT2 genes in normal development, despite their apparently redundant enzymatic function. Distinct phenotypes associated with variants in each of the KMT2 genes support their unique biological roles. Furthermore, the possibility of treating some of these conditions makes them highly relevant for future research[74-77]. Our findings enable the grouping of phenotypes based on broad transcriptional consequences of defects in histone lysine methylation. For example, variants in genes promoting transcriptional activity (e.g. H3K4 methyltransferases) appear to cause growth retardation, whereas variants in transcriptional suppressors predominantly result in overgrowth (e.g. *NSD1* [OMIM #606681], *EZH2* [OMIM #601573] and now *KMT5B*). Finally, these results demonstrate a systematic clinically oriented pathway-based approach (e.g. histone lysine methylation in this study) for analysis of large-scale exome or genome sequencing studies can help to reduce the statistical noise and further accelerate the discovery of rare genetic disorders.

SUPPLEMENTAL DATA

Supplemental data include a supplemental note, 2 figures and 14 tables and can be found online with this article.

CONSORTIA

DDD study

Bevan AP, Dixit A, Pridham A, Tivey AR, Sarkar A, Donaldson A, Fryer A, Sifrim A, Henderson A, Magee A, Duncan A, Kraus A, Male A, Ross A, Collins A, Saggar A, Coates A, Nemeth A, Fry A, Green A, Jackson A, Norman A, Barnicoat A, Brady A, Douglas A, Clarke A, Dobbie A, Selby A, Middleton A, Lampe A, Seller A, Procter A, Evans K, Vandersteen A, Weber A, Smith A, Torokwa A, Kaemba B, Treacy B, Fu B, Hutton B, Bernhard B, Kerr B, Castle B, Donnelly C, Gardiner C, Scott C, Brewer C, Wright CF, Langman C, Ogilvie C, Pottinger C, Tysoe C, Taylor C, McWilliam C, Shaw-Smith C, Deshpande C, Longman C, Sequeira C, Patel C, Bennett C, Nellåker C, Wragg C, Kirk C, Turner C, King D, Barrett DM, Perrett D, Pilz DT, Walker D, Baty D, Bohanna D, Bourn D, Goudie D, Bunyan DJ, Jones D, Moore D, FitzPatrick DR, FitzPatrick DR, Rice D, Shears D, Cilliers D, Donnelly D, Williams D, Lim D, Kumar D, McCann E, Donnai D, Baralle D, Johnson D, Rajan D, Wellesley D, McMullan DJ, Simpkin D, Josifova D, de Vries D, Sheridan E, Maher E, Blair E, Roberts E, Chatzimichali E, Prigmore E, Rosser E, Jones E, Sweeney E, Wilkinson E, Gray E, Staff P, Hobson E, Kivuva E, Miles E, Shearing E, Wakeling E, Kinning E, Bragin E, Coomber EL, Yang F, Connell F, Stewart F, Elmslie F, Flinter F, Swaminathan GJ, Kirby G, Cross G, Devlin G, Woods G, Hollingsworth G, Roberts G, Lowther G, Gill H, Archer H, Cox H, Firth H, Kingston H, Murphy H, Stewart H, Firth HV, Purnell H, Mugalaasi H, Ellis I, Scurr I, Simonic I, Colgiu I, Eason J, Awada J, Hurst J, Barrett JC, Morton J, Thomson J, McRae J, Clayton-Smith J, Paterson J, Jarvis J, Kaplanis J, Poulton J, Burn J, Burton J, Dean J, Tolmie J, Berg J, Roberts J, Waters J,

Randall J, Goodship J, Rankin J, Sampson J, Phipps J, Vogt J, Ong KR, Marks K, Brunstrom K, Chandler K, Tatton-Brown K, Smith K, Morley KI, Lachlan K, Temple IK, Martin K, Prescott K, Metcalfe K, McKay K, Ambridge K, Cresswell L, Mason LE, Yates L, Robert L, Islam L, Sneddon L, Bradley L, He L, Gaunt L, Bourdon L, Brueton L, Nevitt L, Wilson L, Harrison L, Hilyard L, Jenkins L, Raymond L, Greenhalgh L, Tischkowitz M, Whiteford M, van Kogelenberg M, Bitner-Glindzicz M, D'Alessandro M, Pollard M, Hurles ME, Balasubramanian M, Lees M, Irving M, McEntagart M, Humphreys M, Parker M, Parker M, Parker M, Wright M, Yau M, Splitt M, Squires M, Suri M, Alvi M, Blyth M, Collinson MN, Ahmed M, Holder M, Akawi N, Canham N, Ghali N, Krishnappa N, Cooper N, Foulds N, Ragge N, Williams N, Carter NP, Shannon N, Pratt N, Quarrell O, Batstone P, Roberts P, Ellis P, Turnpenny P, Greene P, Jones P, Vasudevan P, Harrison R, Rahbari R, Miller R, Fisher R, Gibbons R, Sandford R, Scott R, Andrews R, Taylor R, Singzon R, Hawkins R, Davidson R, Kelsell R, O'Shea R, Banerjee R, Armstrong R, McGowan R, Newbury-Ecob R, Sharif S, Tein M, Al-Turki S, Mansour S, Lynch SA, Davies S, Kazembe S, Gomes Pereira SL, Widaa S, Edkins S, Everest S, Hewitt S, Smithson S, Wallwark S, Wilcox S, Mehta S, Gerety SS, Samant S, McKee S, Mohammed S, Joss S, Ellard S, Morgan S, Banka S, Brent S, Holden S, Douzgou S, Park SM, Clayton S, Hellens S, Payne S, Aitken S, Ingram S, Price S, Clasper S, Gribble S, Holder S, McNerlan S, Tomkins S, Schweiger S, Bumpstead SJ, Naik S, Dabir T, Bayzetinova T, Montgomery T, Singh T, Homfray T, Fendick T, Fitzgerald TW, Fitzgerald TW, Cole T, Maye U, Anjum U, Kini U, Kumar VKA, Harrison V, Murday V, Parthiban V, Varghese V, Clowes V, McConnell V, Sutton V, Straub V, Lam W, Jones WD, Crow Y, Skitt Z, Miedzybrodzka Z.

CAUSES study

Shelin Adam, Christèle du Souich, Alison M. Elliott, Anna Lehman, Jill Mwenifumbo, Tanya N. Nelson, Clara van Karnebeek, and Jan M. Friedman

ACKNOWLEDGEMENTS

We are thankful to all the individuals and their families for taking part in the study. We are thankful to Matthew Hurles for his critical review of the manuscript. Victor Faundes acknowledges to CONICYT, Chile's National Commission for Scientific and Technological Research, for its scholarship support.

We are thankful to the Deciphering Developmental Disorders (DDD) study for the invaluable collaboration. The DDD study (Cambridge South REC approval 10/H0305/83 and the Republic of Ireland REC GEN/284/12) presents independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003], a parallel funding partnership between the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger Institute [grant number WT098051]. The views expressed in this publication are those of the author(s) and not necessarily those of the Wellcome Trust or the Department of Health. The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network. The CAUSES study (www.causes.clinic) (University of British Columbia protocol H-15-00092) is funded by the Mining for Miracles (British Columbia Children's Hospital Foundation) and Genome British Columbia, with support from the British Columbia Provincial Health Services Authority and British Columbia Women's Hospital.

The authors state that there are no conflicts of interests.

WEB RESOURCES

BrainSpan, http://www.brainspan.org/

DECIPHER, https://decipher.sanger.ac.uk

Denovo-db, http://denovo-db.gs.washington.edu/denovo-db/

Ensembl GRCh37, http://grch37.ensembl.org

ExAC Browser, http://exac.broadinstitute.org/

Exome Variant Server, http://evs.gs.washington.edu/EVS/

gnomAD, http://gnomad.broadinstitute.org/

GTEx Portal, https://www.gtexportal.org/home

HGMD® Professional Version, https://www.qiagenbioinformatics.com/products/human-gene-mutation-database/

HUGO Gene Nomenclature Committee, http://www.genenames.org/

IMPC, http://www.mousephenotype.org/

MutationMapper, http://www.cbioportal.org/mutation_mapper.jsp

MutationTaster2, http://www.mutationtaster.org/

OMIM, https://www.omim.org/

RVIS, http://genic-intolerance.org/Search?query=kncn

The 1000 Genomes Project, http://phase3browser.1000genomes.org/index.html

UK10K Project, https://www.uk10k.org/

UniProtKB, http://www.uniprot.org/

ZFIN, http://zfin.org/

REFERENCES

1. Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R., et al. (2007). New nomenclature for chromatin-modifying enzymes. Cell 131, 633-636.

2. Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet 13, 343-357.

3. HUGO Gene Nomenclature Committee. (2016). Gene Family: Chromatin modifying enzymes. In. (

4. Kleefstra, T., Brunner, H.G., Amiel, J., Oudakker, A.R., Nillesen, W.M., Magee, A., Genevieve, 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. American journal of human genetics 79, 370-377.

5. 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. Nature genetics 42, 790-793.

6. Gibson, W.T., Hood, R.L., Zhan, S.H., Bulman, D.E., Fejes, A.P., Moore, R., Mungall, A.J., Eydoux, P., Babul-Hirji, R., An, J., et al. (2012). Mutations in EZH2 cause Weaver syndrome. American journal of human genetics 90, 110-118.

7. 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. American journal of human genetics 90, 119-124.

8. Jones, W.D., Dafou, D., McEntagart, M., Woollard, W.J., Elmslie, F.V., Holder-Espinasse, M., Irving, M., Saggar, A.K., Smithson, S., Trembath, R.C., et al. (2012). De novo mutations in MLL cause Wiedemann-Steiner syndrome. American journal of human genetics 91, 358-364.

9. Luscan, A., Laurendeau, I., Malan, V., Francannet, C., Odent, S., Giuliano, F., Lacombe, D., Touraine, R., Vidaud, M., Pasmant, E., et al. (2014). Mutations in SETD2 cause a novel overgrowth condition. J Med Genet 51, 512-517.

10. 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. Nature genetics 30, 365-366.

11. Chong, J.X., Yu, J.H., Lorentzen, P., Park, K.M., Jamal, S.M., Tabor, H.K., Rauch, A., Saenz, M.S., Boltshauser, E., Patterson, K.E., et al. (2016). Gene discovery for Mendelian conditions via social networking: de novo variants in KDM1A cause developmental delay and distinctive facial features. Genetics in medicine : official journal of the American College of Medical Genetics 18, 788-795.

12. 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. American journal of human genetics 76, 227-236.

13. Laumonnier, F., Holbert, S., Ronce, N., Faravelli, F., Lenzner, S., Schwartz, C.E., Lespinasse, J., Van Esch, H., Lacombe, D., Goizet, C., et al. (2005). Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. J Med Genet 42, 780-786.

14. Singh, T., Kurki, M.I., Curtis, D., Purcell, S.M., Crooks, L., McRae, J., Suvisaari, J., Chheda, H., Blackwood, D., Breen, G., et al. (2016). Rare loss-of-function variants in

SETD1A are associated with schizophrenia and developmental disorders. Nature neuroscience 19, 571-577.

15. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic acids research 43, D789-798.

16. Banka, S., Lederer, D., Benoit, V., Jenkins, E., Howard, E., Bunstone, S., Kerr, B., McKee, S., Lloyd, I.C., Shears, D., et al. (2015). Novel KDM6A (UTX) mutations and a clinical and molecular review of the X-linked Kabuki syndrome (KS2). Clinical genetics 87, 252-258.

17. Banka, S., Howard, E., Bunstone, S., Chandler, K.E., Kerr, B., Lachlan, K., McKee, S., Mehta, S.G., Tavares, A.L., Tolmie, J., et al. (2013). MLL2 mosaic mutations and intragenic deletion-duplications in patients with Kabuki syndrome. Clinical genetics 83, 467-471.

18. Banka, S., Veeramachaneni, R., Reardon, W., Howard, E., Bunstone, S., Ragge, N., Parker, M.J., Crow, Y.J., Kerr, B., Kingston, H., et al. (2012). How genetically heterogeneous is Kabuki syndrome?: MLL2 testing in 116 patients, review and analyses of mutation and phenotypic spectrum. European journal of human genetics : EJHG 20, 381-388.

19. Stenson, P.D., Mort, M., Ball, E.V., Shaw, K., Phillips, A., and Cooper, D.N. (2014). The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Human genetics 133, 1-9.

20. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature 515, 216-221.

21. Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J.D., Romano-Adesman, A., Bjornson, R.D., Breitbart, R.E., Brown, K.K., et al. (2013). De novo mutations in histone-modifying genes in congenital heart disease. Nature 498, 220-223.

22. Deciphering Developmental Disorders, S. (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature 542, 433-438.

23. Koscielny, G., Yaikhom, G., Iyer, V., Meehan, T.F., Morgan, H., Atienza-Herrero, J., Blake, A., Chen, C.K., Easty, R., Di Fenza, A., et al. (2014). The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. Nucleic acids research 42, D802-809.

24. Howe, D.G., Bradford, Y.M., Conlin, T., Eagle, A.E., Fashena, D., Frazer, K., Knight, J., Mani, P., Martin, R., Moxon, S.A., et al. (2013). ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. Nucleic acids research 41, D854-860.

25. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291.

26. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., and Goldstein, D.B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. PLoS genetics 9, e1003709.

27. Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS genetics 6, e1001154.

28. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature 526, 68-74.

29. Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., et al. (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 493, 216-220.

30. UK10K Consortium, Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R., Xu, C., Futema, M., et al. (2015). The UK10K project identifies rare variants in health and disease. Nature 526, 82-90.

31. Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461, 272-276.

32. Office for National Statistics. (2017). Overview of the UK population: July 2017. In. (

33. Central Statistics Office. (2016). Census 2016 Preliminary Report. In. (

34. Turner, T.N., Yi, Q., Krumm, N., Huddleston, J., Hoekzema, K., HA, F.S., Doebley, A.L., Bernier, R.A., Nickerson, D.A., and Eichler, E.E. (2017). denovo-db: a compendium of human de novo variants. Nucleic acids research 45, D804-D811.

35. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine : official journal of the American College of Medical Genetics 17, 405-424.

36. Kleefstra, T., Kramer, J.M., Neveling, K., Willemsen, M.H., Koemans, T.S., Vissers, L.E., Wissink-Lindhout, W., Fenckova, M., van den Akker, W.M., Kasri, N.N., et al. (2012). Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. American journal of human genetics 91, 73-82.

37. De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 515, 209-215.

38. Homsy, J., Zaidi, S., Shen, Y., Ware, J.S., Samocha, K.E., Karczewski, K.J., DePalma, S.R., McKean, D., Wakimoto, H., Gorham, J., et al. (2015). De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science 350, 1262-1266.

39. Kataoka, M., Matoba, N., Sawada, T., Kazuno, A.A., Ishiwata, M., Fujii, K., Matsuo, K., Takata, A., and Kato, T. (2016). Exome sequencing for bipolar disorder points to roles of de novo loss-of-function and protein-altering mutations. Molecular psychiatry 21, 885-893.

40. Deciphering Developmental Disorders Study. (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature.

41. Lelieveld, S.H., Reijnders, M.R., Pfundt, R., Yntema, H.G., Kamsteeg, E.J., de Vries, P., de Vries, B.B., Willemsen, M.H., Kleefstra, T., Lohner, K., et al. (2016). Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. Nature neuroscience 19, 1194-1196.

42. RK, C.Y., Merico, D., Bookman, M., J, L.H., Thiruvahindrapuram, B., Patel, R.V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z., et al. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nature neuroscience 20, 602-611.

43. Stessman, H.A., Xiong, B., Coe, B.P., Wang, T., Hoekzema, K., Fenckova, M., Kvarnung, M., Gerdts, J., Trinh, S., Cosemans, N., et al. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nature genetics.

44. The CAUSES Study Group. (2015). CAUSES research clinic. In. (

45. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. Nature genetics 43, 838-846.

46. Kasher, P.R., Schertz, K.E., Thomas, M., Jackson, A., Annunziata, S., Ballesta-Martinez, M.J., Campeau, P.M., Clayton, P.E., Eaton, J.L., Granata, T., et al. (2016). Small 6q16.1 Deletions Encompassing POU3F2 Cause Susceptibility to Obesity and Variable Developmental Delay with Intellectual Disability. American journal of human genetics 98, 363-372.

47. Banka, S., Cain, S.A., Carim, S., Daly, S.B., Urquhart, J.E., Erdem, G., Harris, J., Bottomley, M., Donnai, D., Kerr, B., et al. (2015). Leri's pleonosteosis, a congenital rheumatic disease, results from microduplication at 8q22.1 encompassing GDF6 and SDC2 and provides insight into systemic sclerosis pathogenesis. Annals of the rheumatic diseases 74, 1249-1256.

48. Bragin, E., Chatzimichali, E.A., Wright, C.F., Hurles, M.E., Firth, H.V., Bevan, A.P., and Swaminathan, G.J. (2014). DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic acids research 42, D993-D1000.

49. Cho, Y.W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G.R., Copeland, T.D., Kalkum, M., et al. (2007). PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. J Biol Chem 282, 20395-20406.

50. GTEx Consortium. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648-660.

51. Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature 508, 199-206.

52. Lee, S., Lee, D.K., Dou, Y., Lee, J., Lee, B., Kwak, E., Kong, Y.Y., Lee, S.K., Roeder, R.G., and Lee, J.W. (2006). Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. Proceedings of the National Academy of Sciences of the United States of America 103, 15392-15397.

53. Eram, M.S., Kuznetsova, E., Li, F., Lima-Fernandes, E., Kennedy, S., Chau, I., Arrowsmith, C.H., Schapira, M., and Vedadi, M. (2015). Kinetic characterization of human histone H3 lysine 36 methyltransferases, ASH1L and SETD2. Biochim Biophys Acta 1850, 1842-1848.

54. Cau, E., and Wilson, S.W. (2003). Ash1a and Neurogenin1 function downstream of Floating head to regulate epiphysial neurogenesis. Development 130, 2455-2466.

55. Miyazaki, H., Higashimoto, K., Yada, Y., Endo, T.A., Sharif, J., Komori, T., Matsuda, M., Koseki, Y., Nakayama, M., Soejima, H., et al. (2013). Ash1l methylates Lys36 of histone H3 independently of transcriptional elongation to counteract polycomb silencing. PLoS genetics 9, e1003897.

56. Brinkmeier, M.L., Geister, K.A., Jones, M., Waqas, M., Maillard, I., and Camper, S.A. (2015). The Histone Methyltransferase Gene Absent, Small, or Homeotic Discs-1 Like Is Required for Normal Hox Gene Expression and Fertility in Mice. Biology of reproduction 93, 121.

57. Yang, H., Pesavento, J.J., Starnes, T.W., Cryderman, D.E., Wallrath, L.L., Kelleher, N.L., and Mizzen, C.A. (2008). Preferential dimethylation of histone H4 lysine 20 by Suv4-20. The Journal of biological chemistry 283, 12085-12092.

58. Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callen, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I.A., et al. (2008). A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. Genes Dev 22, 2048-2061.

59. Zech, M., Boesch, S., Maier, E.M., Borggraefe, I., Vill, K., Laccone, F., Pilshofer, V., Ceballos-Baumann, A., Alhaddad, B., Berutti, R., et al. (2016). Haploinsufficiency of KMT2B, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia. American journal of human genetics 99, 1377-1387.

60. Meyer, E., Carss, K.J., Rankin, J., Nichols, J.M., Grozeva, D., Joseph, A.P., Mencacci, N.E., Papandreou, A., Ng, J., Barral, S., et al. (2016). Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia. Nature genetics.

61. de Goede, C., Yue, W.W., Yan, G., Ariyaratnam, S., Chandler, K.E., Downes, L., Khan, N., Mohan, M., Lowe, M., and Banka, S. (2016). Role of reverse phenotyping in interpretation of next generation sequencing data and a review of INPP5E related disorders. European journal of paediatric neurology : EJPN : official journal of the European Paediatric Neurology Society 20, 286-295.

62. Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K.R., Aasland, R., Anastassiadis, K., Ang, S.L., and Stewart, A.F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development 133, 1423-1432.

63. Goldsworthy, M., Absalom, N.L., Schroter, D., Matthews, H.C., Bogani, D., Moir, L., Long, A., Church, C., Hugill, A., Anstee, Q.M., et al. (2013). Mutations in Mll2, an H3K4 methyltransferase, result in insulin resistance and impaired glucose tolerance in mice. PloS one 8, e61870.

64. Breuza, L., Poux, S., Estreicher, A., Famiglietti, M.L., Magrane, M., Tognolli, M., Bridge, A., Baratin, D., Redaschi, N., and UniProt, C. (2016). The UniProtKB guide to the human proteome. Database : the journal of biological databases and curation 2016.

65. Narasimhan, V.M., Hunt, K.A., Mason, D., Baker, C.L., Karczewski, K.J., Barnes, M.R., Barnett, A.H., Bates, C., Bellary, S., Bockett, N.A., et al. (2016). Health and population effects of rare gene knockouts in adult humans with related parents. Science 352, 474-477.

66. Albert, M., Schmitz, S.U., Kooistra, S.M., Malatesta, M., Morales Torres, C., Rekling, J.C., Johansen, J.V., Abarrategui, I., and Helin, K. (2013). The histone demethylase Jarid1b ensures faithful mouse development by protecting developmental genes from aberrant H3K4me3. PLoS genetics 9, e1003461.

67. He, R., and Kidder, B.L. (2017). H3K4 demethylase KDM5B regulates global dynamics of transcription elongation and alternative splicing in embryonic stem cells. Nucleic acids research 45, 6427-6441.

68. 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.

69. Wang, Q., Li, M., Yang, Z., Hu, X., Wu, H.M., Ni, P., Ren, H., Deng, W., Li, M., Ma, X., et al. (2015). Increased co-expression of genes harboring the damaging de novo mutations in Chinese schizophrenic patients during prenatal development. Scientific reports 5, 18209.

70. Labonne, J.D., Lee, K.H., Iwase, S., Kong, I.K., Diamond, M.P., Layman, L.C., Kim, C.H., and Kim, H.G. (2016). An atypical 12q24.31 microdeletion implicates six genes

including a histone demethylase KDM2B and a histone methyltransferase SETD1B in syndromic intellectual disability. Human genetics 135, 757-771.

71. Dong, S., Walker, M.F., Carriero, N.J., DiCola, M., Willsey, A.J., Ye, A.Y., Waqar, Z., Gonzalez, L.E., Overton, J.D., Frahm, S., et al. (2014). De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder. Cell reports 9, 16-23.

72. Binda, O. (2013). On your histone mark, SET, methylate! Epigenetics-Us 8, 457-463.

73. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature genetics 39, 311-318.

74. Benjamin, J.S., Pilarowski, G.O., Carosso, G.A., Zhang, L., Huso, D.L., Goff, L.A., Vernon, H.J., Hansen, K.D., and Bjornsson, H.T. (2017). A ketogenic diet rescues hippocampal memory defects in a mouse model of Kabuki syndrome. Proceedings of the National Academy of Sciences of the United States of America 114, 125-130.

75. Bjornsson, H.T., Benjamin, J.S., Zhang, L., Weissman, J., Gerber, E.E., Chen, Y.C., Vaurio, R.G., Potter, M.C., Hansen, K.D., and Dietz, H.C. (2014). Histone deacetylase inhibition rescues structural and functional brain deficits in a mouse model of Kabuki syndrome. Science translational medicine 6, 256ra135.

76. Micale, L., Augello, B., Maffeo, C., Selicorni, A., Zucchetti, F., Fusco, C., De Nittis, P., Pellico, M.T., Mandriani, B., Fischetto, R., et al. (2014). Molecular analysis, pathogenic mechanisms, and readthrough therapy on a large cohort of Kabuki syndrome patients. Human mutation 35, 841-850.

77. Bogershausen, N., Tsai, I.C., Pohl, E., Kiper, P.O., Beleggia, F., Percin, E.F., Keupp, K., Matchan, A., Milz, E., Alanay, Y., et al. (2015). RAP1-mediated MEK/ERK pathway defects in Kabuki syndrome. The Journal of clinical investigation 125, 3585-3599.

FIGURE LEGENDS

Figure 1

Variants in histone lysine methyltransferases and demethylases are frequent in developmental disorders and haploinsufficiency is their predominant mechanism

A) The bar graph shows the proportions of postulated disease causing published heterozygous protein-truncating variants (PTV) (in red) and protein altering variants (PAV) (in blue) in known dominant developmental disorders (DD)-associated KMTs and KDMs.

B) A plot of probability of being LoF Intolerant (pLI) scores for all KMTs and KDMs. Red dots represent the pLI scores for known dominant DD-associated KMTs and KDMs, orange dots depict these scores from candidate for dominant DD KMTs/KDMs, and green dots display the pLI scores for non-candidate KMT/KDM genes. The dotted line depicts the cut-off for defining the candidate genes (pLI>0.9)

C) Proportion of canonical transcripts of known DD KMTs and KDMs from the total human exome (left donut graph), proportion of individuals with pathogenic variants in known KMT/KDM genes from the Deciphering Developmental Disorders (DDD) study cohort (central donut graph), and proportion of pathogenic, benign or variants of uncertain significance (VUS) in known KMT/KDM genes, and the percentage of variants in other KMTs/KDMs from the total number of KMT/KDM variants seen in the DDD cohort (right donut graph). The Venn diagram shows the distribution of rare high quality 120 variants, detected in the DDD cohort, in KMTs/KDMs not yet firmly associated with DDs.

The green circle and the ellipse represent the number of variants according to their inheritance, the blue circle and the ellipse represent the number of variants according to their predicted protein effect, and the red circle and the ellipse represent the number of variants detected in candidates for dominant DDs and the other genes.

Figure 2

Variants of interest identified in this study.

Locations of selected plausible candidate variants identified in this study are shown. Candidates genes for dominant DDs with DN PTVs are indicated in red font and the other genes are in black font. The *de novo* (DN) protein-truncating (PTV) in candidate KMTs and KDMs for dominant DD genes (*KMT2B*, *KMT2C*, *ASH1L* and *KMT5B*) (n=8) are highly likely to be causal. We have also shown DN protein-altering variants (PAV) in candidate KMTs and KDMs for dominant DD genes (*KMT2B*, *KMT2C*, *DOT1L*, *KDM3A*, *PRDM2*, *SETDB1*) (n=9) with limited evidence for causality at present (apart from those in *KMT2B* which have been shown to cause early onset dystonia). Inherited PTVs in candidate KMTs and KDMs (*KDM3A* and *PRDM2*) (n=2) are shown. PTVs in these genes may cause non-penetrant phenotypes or this may indicate that these genes tolerate haploinsufficiency unlike as suggested by their pLI scores. DN PTVs in non-candidate KMTs and KDMs for dominant DDs (*KDM5B* and *SETD1B*) (n=4) are also shown. These PTVs could be coincidental or may be acting as phenotype modifiers or could be non-penetrant in some individuals in the general population. Homozygous and compound heterozygous PTVs in *KDM5B* (n=5) show that recessive histone tail lysine methylation disorders also exist.

Figure 3

Photographs from individuals with truncating variants or deletions of *KMT2B*, *KMT2C*, *KMT5B* and *KDM5B*.

The numbers on each picture denote the corresponding individual in Table 1. Individual 1 with *KMT2B* DN PTV has spare scalp hair, large mouth and absent ear lobes; individual 2 with *KMT2C* DN PTV has marked infra-orbital creases, down-slanting palpebral fissures, and a duplicated right thumb. Individual 3 with *KMT2C* DN PTV has marked plagiocephaly and bilateral marked bulging just below the temporal region. Individual 8 with *KMT5B* DN PTV has a broad and large forehead that has persisted in time. Individual 9 with *KMT5B* DN PTV has a prominent forehead, thick ear lobes, broad philtrum, an open mouth appearance and synophrys which is more noticeable in the more recent photograph. Individual 11 with DN *KMT5B* deletion has a long and oval face, ptosis, prominent eyes, protruded ears, open mouth, thick lips and overlapping of 3rd to 2nd toes. Individual 12 with homozygous *KDM5B* PTV has down-slanting palpebral fissures, slightly bulbous nasal tip, low-hanging columella, smooth philtrum and thin upper and lower lips. He has bilateral camptodactyly of 4^{th} and 5^{th} fingers. Individual 14 with compound heterozygous homozygous *KDM5B* PTV has a prominent metopic region, a high nasal bridge, bulbous nasal tip, smooth philtrum, thin lips and a triangular ear with an absent superior crux of helix. He has also mild camptodactyly of the 4^{th} and 5^{th} fingers.

Figure 4

Comparison of gene and protein properties, between known/candidate dominant for DD and other KMTs and KDMs.

The comparisons were made using the data from UniProtKB and Mann-Whitney test was performed with an exact p-value <0.05 considered as significant. The results are represented in dot plots (A) number of post-translational modifications (PTM); (B) number of types of PTM; (C) number of interactors; (D) length of canonical transcripts; (E) number of PTM per 100 amino acids of canonical transcripts in KMT/KDMs; and (F) The median Reads per Kilobase per Million (RPKM) for candidate and non-candidate KMT/KDMs in brain

structures with significant differences across several stages. Black/coloured dots denote known/candidate for dominant DD KMT/KDM genes, white dots/coloured triangles show the non-candidate for dominant KMT/KDM genes, and coloured boxes depict the brain structures. The longer horizontal lines in all the graphs represent the respective medians, the shorter horizontal lines indicate the inter-quartile range and the p values are given at the top of each graph, where relevant.

Table 1. Clinical and genetic characteristics from affected individuals with candidate variants in lysine methyltransferases (KMT) and demethylases (KDM).

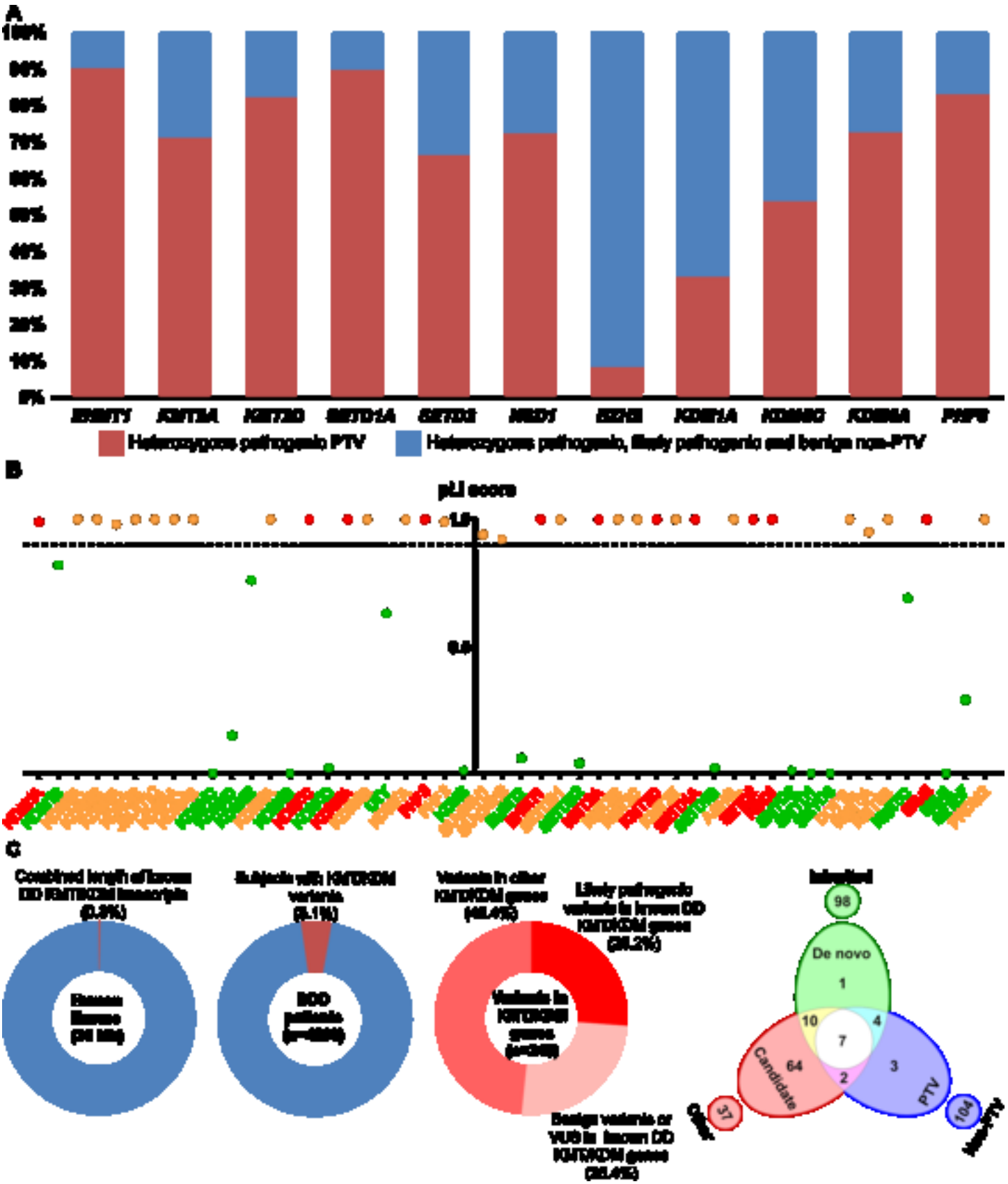
Gene	Sex (age at study)	Individual number	Genomic position (hg19)	cDNA[a] (protein consequence)/ Deletion size	Inh/ zyg	Perinatal history	DD/ ID	Neuropsychiatric disorders/ CNS anomalies	Malformations and anomalies	Height (SD)/ weight (SD)/ OFC (SD)	CD	Other medical issues
KMT2B	F (11y)	1	19:36212057	c.1808dupC p.(Leu604Profs*72)	DN Het	IUGR and feeding difficulties	Severe	Abnormal gait and behavioural problems.	PDA, long & narrow hands, broad halluces	SS (-2.7) LW (-2.9) Mi (-3.34)	Yes	Nystagmus, gastrostomy, urinary incontinence, constipation and growth hormone deficiency
KMT2C	F (17y)	2	7:151884849	c.4744G>T p.(Gly1582*)	DN Het	No	Severe	Elective mutism	Duplicated right thumb and left preauricular tag	SS (-2.1) LW (-2.74) Mi (-2.42)	Yes	Hearing loss and delayed puberty
	F (4y)	3	7:151873688-151873689	c.8849_8850delAT p.(His2950Argfs*17)	DN Het	Hydrocephalus and Dandy-Walker anomaly	Severe	Hydrocephalus and hypoplasia of cerebellar vermis	No	SS (-2) LW (-2) Mi (-1.97)	Yes	No
	F (5y)	4	7:151836279	c.14526dupG p.(Pro4843Alafs*12)	DN Het	No	Severe (motor delay was mild)	Autistic traits, developmental regression, insensitivity to pain and abnormal gait	No	N (0.4) N (0.18) N (-1)	Yes	Constipation
ASH1L	F (13y)	5	1:155449628	c.3033delA p.(Val1014Cysfs*24)	DN Het	Feeding difficulties	Mild	Behavioural problems	Bicuspid AV, VSD and PFO	N (0.2) N (0.82) N (1.16)	Yes	Hypermetropia, precocious puberty and hypermobility
	M (9y)	6	1:155322602	c.7276C>T p.(Arg2426*)	DN Het	Feeding difficulties and hydronephrosis	Severe	Seizures, autistic traits and hypotonia.	Cryptorchidism and inguinal hernia	N (1.6) O (2.33) N (1.36)	Yes	Hypermetropia, hyperacusis and hypermobility
	M (7y)	7	1:155271366-155804269	532.9 Kb	DN	No	Severe	Behavioural problems	Cryptorchidism and blocked nasolacrimal duct	N (-0.59 SD) N (-0.44 SD) Mi (1.72SD)	Yes	Constipation
KMT5B	F (13y)	8	11:67953337	c.219delC p.(Ala74Profs*10)	DN Het	No	Moderate	Autistic traits	No	TS (2.91) N (0.9) Ma (4.43)	Yes	Hypermobility
	M (19y)	9	11:67941365	c.559C>T p.(Arg187*)	DN Het	No	Severe	Seizures,hypotonia and autistic traits	No	N (0.74) N (0.9) N (1.93)	Yes	No
	M (14y)	10	11:67888021-68287033	399.01 Kb	DN	No	Mild	Seizures, enlarged right ventricle and white matter signal alterations	No	N (0.63) Ma (2)	Yes	Strabismus and scoliosis
	M (16y)	11	11:67550395-68389391	839 Kb	DN	No	Mild to moderate	No	Cryptorchidism, pectus excavatum, and overlapping 2-3 toes	N (1.68) N (0.24) N (1.87)	Yes	Strabismus, diabetes mellitus and hypermobility
KDM5B	M (18y)	12	1:202700104	c.4109T>G p.(Leu1370*)	Mat & Pat Hom	Feeding difficulties	Severe	Abnormal gait and agenesis of corpus callosum	Inguinal hernia and camptodactyly of 4th and 5th fingers	N (-0.23) LW (-1.52) N (-1.66)	Yes	Myopia and astigmatism
	M (10y)	13	1:202711635 1:202731850	c.2475-2A>G; c.895C>T (p.Arg299Ter)	Mat & Pat CoHet	No	Moderate	No	Dolichocephaly and supernumerary nipple	N (0.98) N (0.51) N (0.26)	No	

	M (11y)	14	1:202702532 1:202736143	c.3906delC, (p.Asn1302Lysfs*45) c.622dupT (p.Tyr208Leufs*5)	Mat & Pat Het	Feeding difficulties,	Moderate	No	Atrial septal defect, cryptorchidism, hypospadias and camptodactyly of 4th and 5th fingers	N (-0.09 SD) N (-1.05)	Yes	Myopia and strabismus

[a] The transcript IDs are *KMT2B* NM_014727.2; *KMT2C* NM_170606.2; *ASH1L* ENST00000368346.7; *KMT5B* NM_017635.4; *KDM5B* NM_001314042.1.
Abbreviations: AV=Aortic valve; CNS=Central Nervous System; CD=Craniofacial dysmorphisms; CoHet=Compound heterozygous; DD=developmental delay; DN=de novo; F=female; Het=Heterozygous; Hom=Homozygous; ID=intellectual disability; Inh=inheritance; IUGR=intra-uterine growth retardation; LW=Low weight; M=male; Ma=Macrocephaly; mat=maternal; Mi=Microcephaly; N=Normal/Not present; O=Overweight/Obesity; PDA=patent ductus arteriosus; PFO=Patent foramen ovale; SD=Standard deviation; SS=short stature; TS=Tall stature; VSD=Ventricular septal defect; y=years; zyg=zygosity.

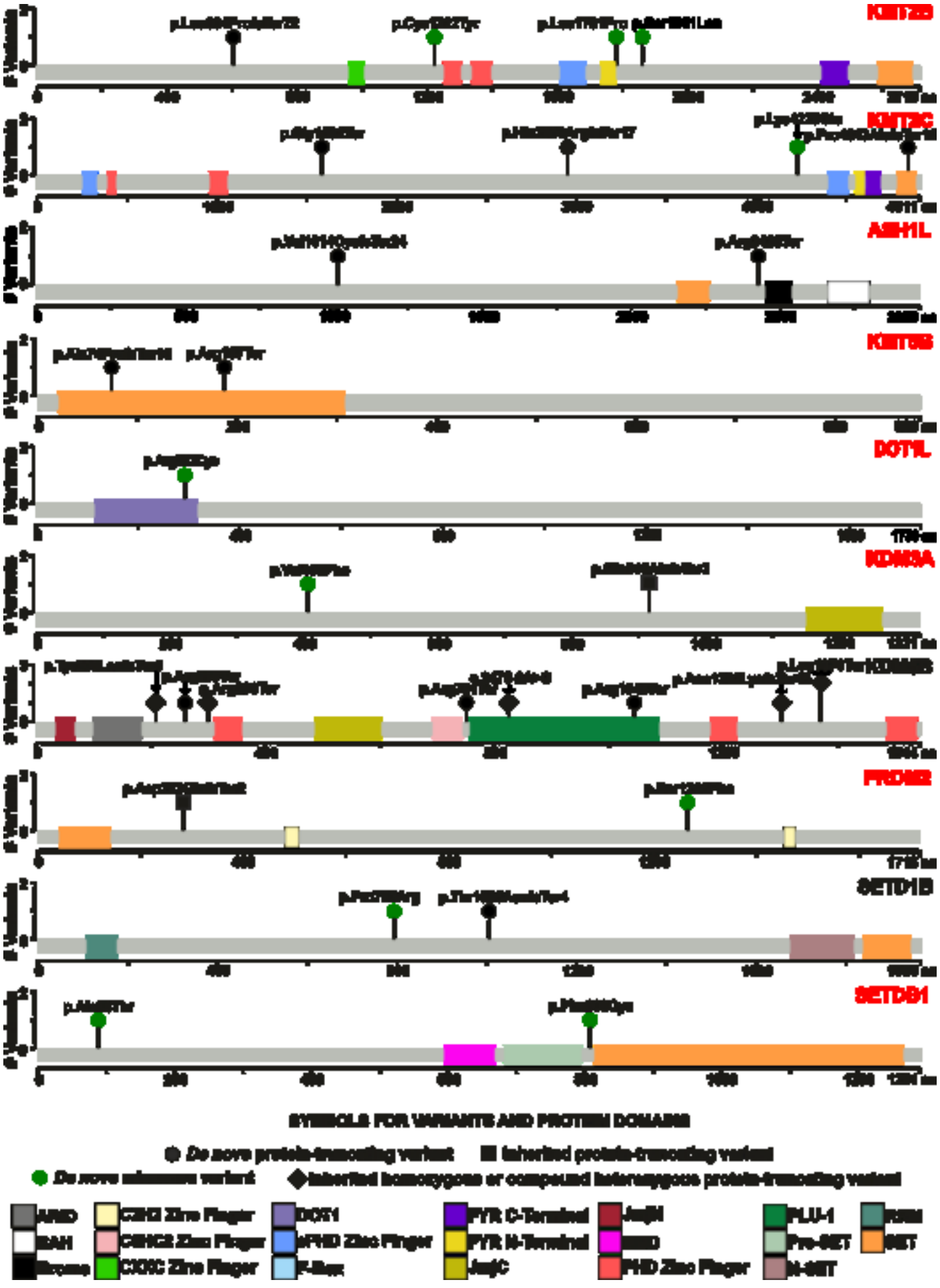
Figure

Click here to download Figure Figure 1.tif



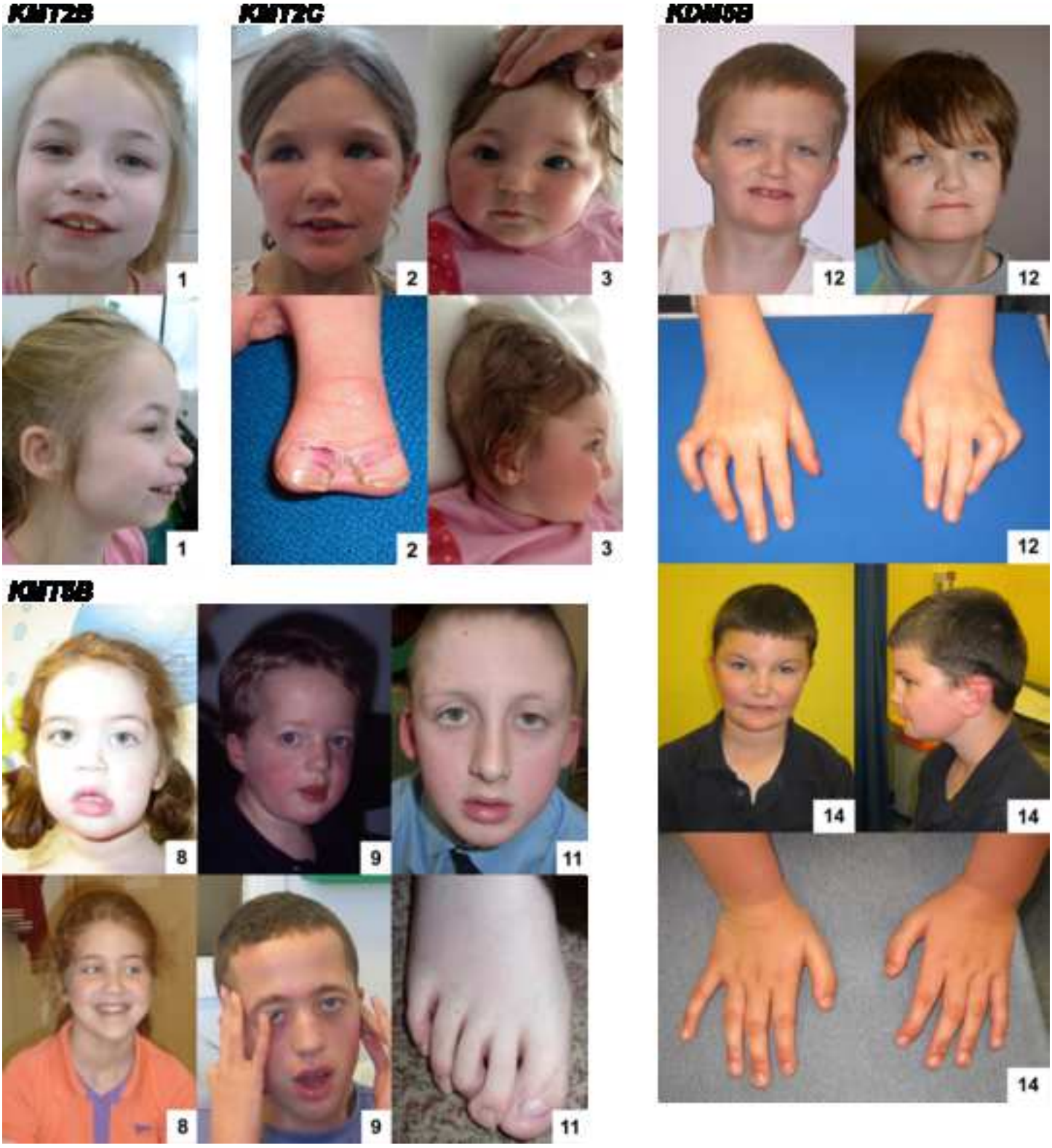
Figure

Click here to download Figure Figure 2.tif



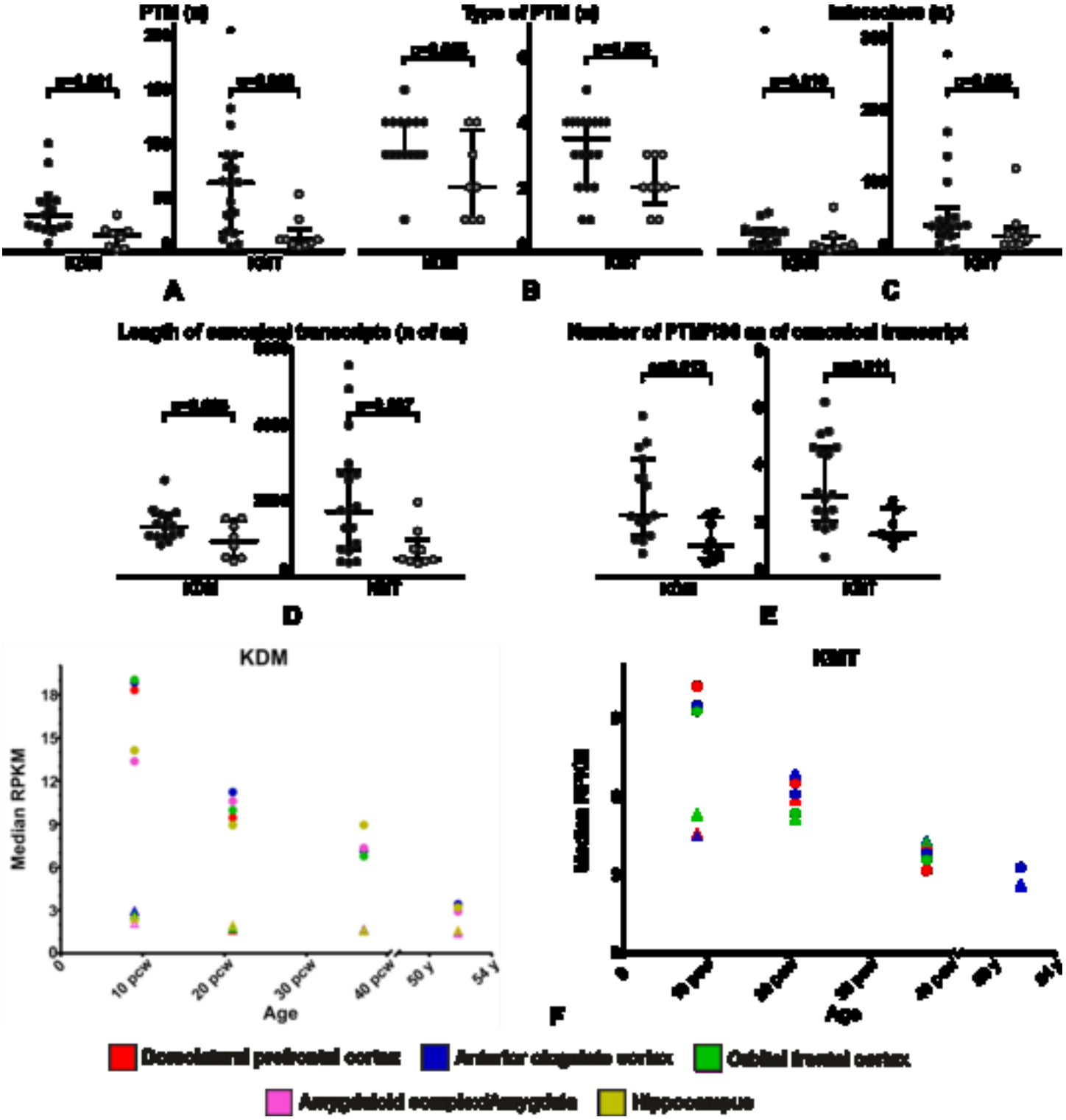
Figure

Click here to download Figure Figure 3.tif



Figure

Click here to download Figure Figure 4.tif



Click here to access/download

Supplemental Movies and Spreadsheets

Tables S7, S13 and S14.xlsx