Research Article

For reprint orders, please contact reprints@futuremedicine.com

Epigenomics

Comparison of methylation episignatures in *KMT2B*- and *KMT2D*-related human disorders

Sunwoo Lee[1], Eguzkine Ochoa[1], Katy Barwick[2], Laura Cif[3], Fay Rodger[1,4], France Docquier[1,4], Belén Pérez-Dueñas[2], Graeme Clark[1,4], Ezequiel Martin[1,4], Siddharth Banka[5], Manju A Kurian[2] & Eamonn R Maher*,[1]

[1]Department of Medical Genetics, University of Cambridge, Cambridge, CB2 0QQ, UK

[2]Developmental Neurosciences, UCL Great Ormond Street Institute of Child Health, Zayed Centre for Research into Rare Disease in Children, London, WC1N 1DZ, UK

[3]Departement de Neurochirurgie, Unite des Pathologies Cerebrales Resistantes, Unite de Recherche sur les Comportements et Mouvements Anormaux, Hopital Gui de Chauliac, Centre Hospitalier Régional Montpellier, Montpellier, France, & Faculte de Medecine, Universite de Montpellier, France

[4]Stratified Medicine Core Laboratory NGS Hub, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK

[5]Division of Evolution, Infection & Genomics, School of Biological Sciences, Faculty of Biology, Medicine & Health, The University of Manchester, Manchester, UK, & Manchester Centre for Genomic Medicine, St. Mary's Hospital, Manchester University Foundation NHS Trust, Health Innovation Manchester, Manchester, M13 9WL, UK

*Author for correspondence: Tel.: +44 0122 374 6714; erm1000@medschl.cam.ac.uk

Aim & methods: To investigate peripheral blood methylation episignatures in *KMT2B*-related dystonia (DYT-*KMT2B*), the authors undertook genome-wide methylation profiling of ~2 M CpGs using a next-generation sequencing-based assay and compared the findings with those in controls and patients with *KMT2D*-related Kabuki syndrome type 1 (KS1). **Results:** A total of 1812 significantly differentially methylated CpG positions (false discovery rate < 0.05) were detected in DYT-*KMT2B* samples compared with controls. Multi-dimensional scaling analysis showed that the 10 DYT-*KMT2B* samples clustered together and separately from 29 controls and 10 with pathogenic variants in *KMT2D*. The authors found that most differentially methylated CpG positions were specific to one disorder and that all (DYT-*KMT2B*) and most (Kabuki syndrome type 1) methylation alterations in CpG islands were gain of methylation events. **Conclusion:** Using sensitive methylation profiling methodology, the authors replicated recent reports of a methylation episignature for DYT-*KMT2B*. These findings will facilitate the development of episignature-based assays to improve diagnostic accuracy.

Plain language summary: The authors compared the DNA methylation patterns in blood from individuals with two rare neurodevelopmental disorders (childhood-onset dystonia [DYT-*KMT2B*] and Kabuki syndrome type 1) and healthy control samples. These two disorders are associated with pathogenic variants in *KMT2B* and *KMT2D*, which encode proteins with related functions but cause distinct inherited disorders. Comparison of the methylation patterns in the two disorders showed that most DNA regions with altered methylation patterns differed between the two disorders and controls. These findings suggest that analyzing DNA methylation patterns could improve diagnostic testing for these disorders and might provide insights into how the clinical features of these disorders are caused.

First draft submitted: 17 December 2021; Accepted for publication: 4 April 2022; Published online: 4 May 2022

Keywords: chromatin disorders • early-onset dystonia • histone lysine methyltransferases (KMTs) • Kabuki syndrome • methylation • neurodevelopmental disorder

Mammalian histone H3 lysine 4 (KMT2) methyltransferases (KMTs) are histone-modifying enzymes that specifically catalyze the transfer of methyl groups (mono-, di- or tri-) to the fourth residue (a lysine) of the Histone 3 protein (H3K4) [1,2]. KMTs are found in active gene promoters (di- or tri-methyltransferases) or gene enhancers

Future Medicine

(mono-methyltransferases) and regulate transcriptional activity [1,3]. In humans, pathogenic variants or dysregulation of KMT2s can cause a variety of rare neurodevelopmental disorders, including Wiedemann–Steiner syndrome (MIM 605130; *KMT2A*), childhood dystonia 28 (MIM 617284 or DYT-*KMT2B*; *KMT2B*), Kleefstra syndrome type 2 (MIM 617768; *KMT2C*), Kabuki syndrome type 1 (MIM 147920; *KMT2D*) and O'Donnel–Luria–Rodan syndrome (MIM 618512, *KMT2E*) [4–10]. These disorders result from monoallelic, mostly *de novo*, loss of function (LOF) variants in the relevant gene and, though they are clinically distinct, a number of clinical features (e.g., global developmental delay, facial dysmorphologies and extra-central nervous system developmental defects) are shared between multiple KMT2-related disorders and with other syndromes caused by pathogenic variants that regulate chromatin structure and function [4,6,11,12]. For example, *KMT2D*-related Kabuki syndrome type 1 (KS1) is associated with a distinctive facial appearance (wide, arched eyebrows; abnormally spaced eyes; flattened nasal tip; and large ears), skeletal abnormalities, short stature, growth delay, mild to severe intellectual disabilities and cardiac defects [13,14]. Similar clinical features are observed in Kabuki syndrome type 2, which is caused by pathogenic variants in the lysine-specific demethylase 6A [15]. However, the clinical phenotype of DYT-*KMT2B* is more distinctive in that most patients present during childhood with dystonia (i.e., twisting, repetitive movement and/or abnormal posture due to involuntarily sustained or intermittent muscle contractions) [6,16–18]. In patients with dystonia, MRI imaging shows bilateral symmetrical abnormalities of the globus pallidus in most cases [19]. In contrast to other KMT2-related disorders, though communication difficulties, intellectual disability and developmental delay can occur in the absence of dystonia, they are not usually the main presenting features [10,19]. Fewer than 150 individuals have been reported with DYT-*KMT2B* and relatively little is known regarding the pathogenesis of the disease, though short-term benefits for refractory dystonia have been reported from deep-brain stimulation to the globus pallidus pars interna [19–21]. A genotype-phenotype correlation has been reported, with the presence of microcephaly, developmental delay/intellectual disability and short stature being more common in individuals with chromosomal microdeletions and protein-truncating variants than in individuals with missense variants [12,19,20,22–25]. Nevertheless, while these features commonly occur in other chromatin gene-related disorders, *KMT2B* is the only gene in its class that typically presents as a genetic dystonia [26].

In recent years, it has become clear that many neurodevelopmental disorders caused by mutations in genes that regulate chromatin function and/or structure exhibit abnormal methylation patterns in blood [27]. Furthermore, these abnormal DNA methylation patterns (episignatures) are distinct for individual neurodevelopmental disorders [27,28]. Clinically this observation provides a pathway to diagnostic assays that can be used to identify candidate genes for patients with an undiagnosed neurodevelopmental disorder and assays that can distinguish between pathogenic and benign variants in chromatin-related disease genes [12,20,22–25,27,29]. Assuming that altered methylation at promoter and regulatory regions is associated with corresponding altered gene expression, episignature analysis might also enable candidate genes germane to the pathogenesis of the relevant disease to be identified (with the caveat that methylation alterations in blood might not reflect those in the central nervous system). Episignatures are most commonly studied using methylation arrays that analyze up to 850,000 CpGs (Illumnia EPIC Beadchip array) in the genome. However, the authors adopted a different strategy by utilizing the targeted bisulphite sequencing approach (Illumina TruSeq Methyl Capture EPIC kit [herein EPIC-NGS]). Though there is considerable overlap in the regions targeted by the two platforms, EPIC-NGS covers regions with a higher density of >3 million CpGs, including 26,981 CpG islands across the genome (97% of human CpG islands) [30]. Therefore, EPIC-NGS has the potential to leverage the power of targeted bisulfite sequencing for a deeper understanding of disorders of the methylome. Therefore, the authors proceeded to investigate the potential basis for the distinct neurological phenotype of DYT-*KMT2B*, by undertaking higher-definition DNA methylation profiling in patients with DYT-*KMT2B* using EPIC-NGS and comparing the findings in DYT-*KMT2B* with those in patients with KS1.

Materials & methods

Study cohort

Genomic DNA from a total of 20 samples from patients with a germline *KMT2B* or *KMT2D* variant was extracted from peripheral blood using Gentra Puregene Blood Kit (Qiagen) or other standard methods and quantified by Qubit™ dsDNA BR Assay Kit (Invitrogen, ThermoFisher). The study cohort included 10 DYT-*KMT2B* patients and 10 KS1 patients diagnosed in expert centers. Details of patient age, gender and *KMT2B* and *KMT2D* variants are reported in Table 1 and the location of the variants are shown in Supplementary Figure 1. The DYT-*KMT2B* cohort had an average age at DNA sampling of 24.1 years and included 3 males and 7 females. The 10 KS1

Table 1. *KMT2B*-related dystonia and Kabuki syndrome type 1 patient variant details.

Patient ID	Disorder	Age†	Gender	Mutation	Protein
*KMT2B*_1 (DY107P)	Early-onset dystonia (DYT28)	41	F	c.3642+ 5G >A	Intron 11, donor splice site lost
*KMT2B*_2 (DY130P)		31	M	c.3147_3160delGGGAGTGGGGGGC	p.Gly1050Profs*33
*KMT2B*_3 (DY52P)		4	F	c.4688del	p.Ala1563Aspfs*83
*KMT2B*_4 (DY69P)		20	F	c.3602delC	p.Pro1201Argfs*154
*KMT2B*_5 (DY78P)		29	F	c.2137dupA	p.Thr713Asnfs*4
*KMT2B*_6 (DY79P)		42	F	c.6439C >T	p.Gln2147*
*KMT2B*_7 (DY80P)		24	F	c.1656dupC	p.Lys553Glnfs*46
*KMT2B*_8 (DY81P)		12	F	c.5658delC	p.Ser1887Profs*8
*KMT2B*_9 (DY98P)		23	M	c.1107dupA	p.Glu370Argfs*19
*KMT2B*_10 (DY158P)		15	M	c.3143_3149del	p.Gly1048Glufs*132
*KMT2D*_1 (10002815)	Kabuki syndrome type 1	11	M	c.14424delC	p.Ser4808Argfs*2
*KMT2D*_2 (10004000)		35	M	c.14946G >A	p.Trp4982*
*KMT2D*_3 (10004089)		33	M	c.8488C >T	p.Arg2830*
*KMT2D*_4 (10007)		27	F	c.14710C >T	p.Arg4904*
*KMT2D*_5 (10012157)		15	M	c.12179_12182delCTGA	p.Thr4060Asnfs*4
*KMT2D*_6 (31210)		34	M	c.16412G >C	p.Arg5471Thr
*KMT2D*_7 (77428)		12	F	c.8727_8730delAAGT	p.Ser2910Argfs*32
*KMT2D*_8 (85546)		25	M	c.13895delC	p.Pro4632Hisfs*8
*KMT2D*_9 (9005905)		11	F	c.5527dupA	p.Thr1843Asnfs*5
*KMT2D*_10 (9901174)		32	M	c.14485dupG	p.Glu4829Glyfs*8

Variant details from 10 KMT2B-related dystonia and 10 Kabuki syndrome type 1 patients. All annotations were aligned by GRCh37/hg19.
†Age of sample was acquired.
F: Female; M: Male.

patients with *KMT2D* variants had an average age of 23.8 years and included seven males and three females. Most individuals with *KMT2D* variants have been described [31]. Methylation profiling results from the DYT-*KMT2B* and KS1 samples were compared with those from 29 healthy control subjects (mean age 19.4 years, 14 males and 15 females). The study was performed in accordance with the Helsinki Declaration, written informed consent was obtained from parents/guardians of the patients and blood samples were collected under local study ethical approval. The study was approved by South Birmingham and Central Manchester Research Ethics Committees.

Targeted DNA methylation sequencing analysis by TrueSeq-EPIC

All DNA samples were sequenced on an Illumina HiSeq 4000 analyzer (Illumina, CA, USA) using TruSeq® Methyl Capture EPIC kit. Bisulfite conversion, library preparation, target enrichment and sequencing were performed at the Cambridge University Department of Medical Genetics Stratified Medicine Core Laboratory (SMCL). Raw sequencing data were extracted as a FASTQ format. After filtering by quality (Phred score ≤30) using Trim-Galore (www.bioinformatics.babraham.ac.uk/projects/trim_galore/), sequenced reads were then aligned to the Genome Reference Consortium Human Build 37 (GRCh37) using Bismark software v0.17.0 (www.bioinformatics.babraha

m.ac.uk/projects/bismark/). After alignment, PCR duplicates were removed with the 'deduplicate bismark' option. Finally, methylation was extracted using 'bismark methylation extractor'.

Methylation analysis

Data normalization, batch effect correction and extraction of differentially methylated CpG sites were conducted by the Bioconductor RnBeads package (https://rnbeads.org/). The RnBeads package was implemented in R software (available on R version 4.0.3). BismarkCov files were directly exported to the RnBeads package as a BED file format. Sequencing coverage threshold was set to 10X and the number of sites above 10X coverage for all 49 samples was approximately 2 million CpGs (1,852,929 CpGs; sequencing reads and coverages per sample are summarized in Supplementary Table 1). CpGs located on sex chromosomes, sites with exceptionally high coverage and NA values were also removed. Differential methylation analysis was then performed by the 'limma' method [32,33]. When computing differentially methylated CpG sites, principal component analysis (PCA) was conducted to interrogate any batch effect (age, gender and batch) and identify significant outliers. If a significant batch effect was detected, target variables were adjusted by surrogate variable analysis (SVA) using the sva package [34]. The identification of significant CpG sites and differentially methylated regions (DMRs) was obtained by Fisher's method from the RnBeads package [32,35]. Only CpG sites (including Open Sea, Shelf, Shore and CpG islands) with combined p-value adjusted less than 0.05 (false discovery rate [FDR] < 0.05 and methylation difference higher than 20% for Open Sea, Shelf, Shore) were considered significant and selected for genome-wide CpG site methylation analysis. For CpG island analysis, all CpG sites within each CpG island were grouped together when calculating p-value. Genes within or close to the significant CpG islands or DMR were annotated. Finally, the identification of significant gain or loss of methylation (GOM/LOM) was determined by median ± 3 standard deviation (3SD) confidence interval compared with the median methylation value of controls (Supplementary Figure 2). In detail, if median methylation values (median of the methylation levels across CpGs) were below the healthy controls' 3SD confidence interval, the authors considered them as LOM and those above 3SDs confidence intervals were designated as GOM. Among four CpG sites (Open Sea, Shelf, Shore and CpG islands), only CpG islands were selected for further genotype-phenotype correlation analysis.

Detection of episignatures

In order to investigate the methylation signature of DYT-*KMT2B* and KS1, and the relationship between the two disorders, hierarchical clustering was performed by two different approaches: multiple dimensional scaling (MDS) and heatmap visualization. MDS plots were plotted using the function 'mdsPlot' in the *minfi* package (version 1.32.0). MDS plots were constructed by calculating the Euclidean distance between samples, and in each case, the most variable CpG positions were specified. Likewise, heatmap clustering was performed by the ComplexHeatmap package (version 2.2.0). In particular, hierarchical clustering was performed using a complete-linkage method by the 'hclust' option. This unsupervised clustering result was then visualized as a dendrogram above the top annotation bar in the heatmap.

Results

DNA methylation profiles for DYT-*KMT2B* & KS1 samples

Genome-wide methylation profiling with EPIC-NGS revealed 1812 (DYT-*KMT2B*) and 89 (KS1) significant differentially methylated CpG positions (DMPs) (FDR < 0.05). MDS analysis using these 1812 CpGs showed that the 10 DYT-*KMT2B* samples clustered together and were separated from 29 control individuals using Euclidean distance (Figure 1A). In addition, the 10 KS1 samples clustered together and were clearly separated from both controls and DYT-*KMT2B* samples (Figure 1C & D & Figure 2). Hierarchical clustering showed similar results and revealed a hypermethylation pattern at almost all 1812 DMPs in all of the DYT-*KMT2B* DNA samples compared with healthy controls; 88.4% (1603/1812) of DMPs showed a significant GOM (above the median of controls + 3SD confidence interval) in at least one sample (Figure 1B & Supplementary Figure 2A). In the KS1 cohort, 36.0% of DMPs (32/89) showed GOM patterns, whereas 23.6% of DMPs (21/89) showed LOM (below the median of controls – 3SD confidence interval) patterns in at least one sample (Figure 1D & Supplementary Figure 2B). Comparison of the DMPs between the 9 (after excluding the KMT2D_5 sample that did not clearly separate from controls) and 10 DYT-*KMT2B* samples identified 677 (including 505 DMPs in 21 CpG islands with FDR <0.05) significant DMPs (Figure 2). Applying a threshold of 3SD confidence interval, on average 73.3% (496/677) of DMPs in 10 DYT-*KMT2B* samples and 47.1% (319/677) DMPs of 9 KS1 samples showed

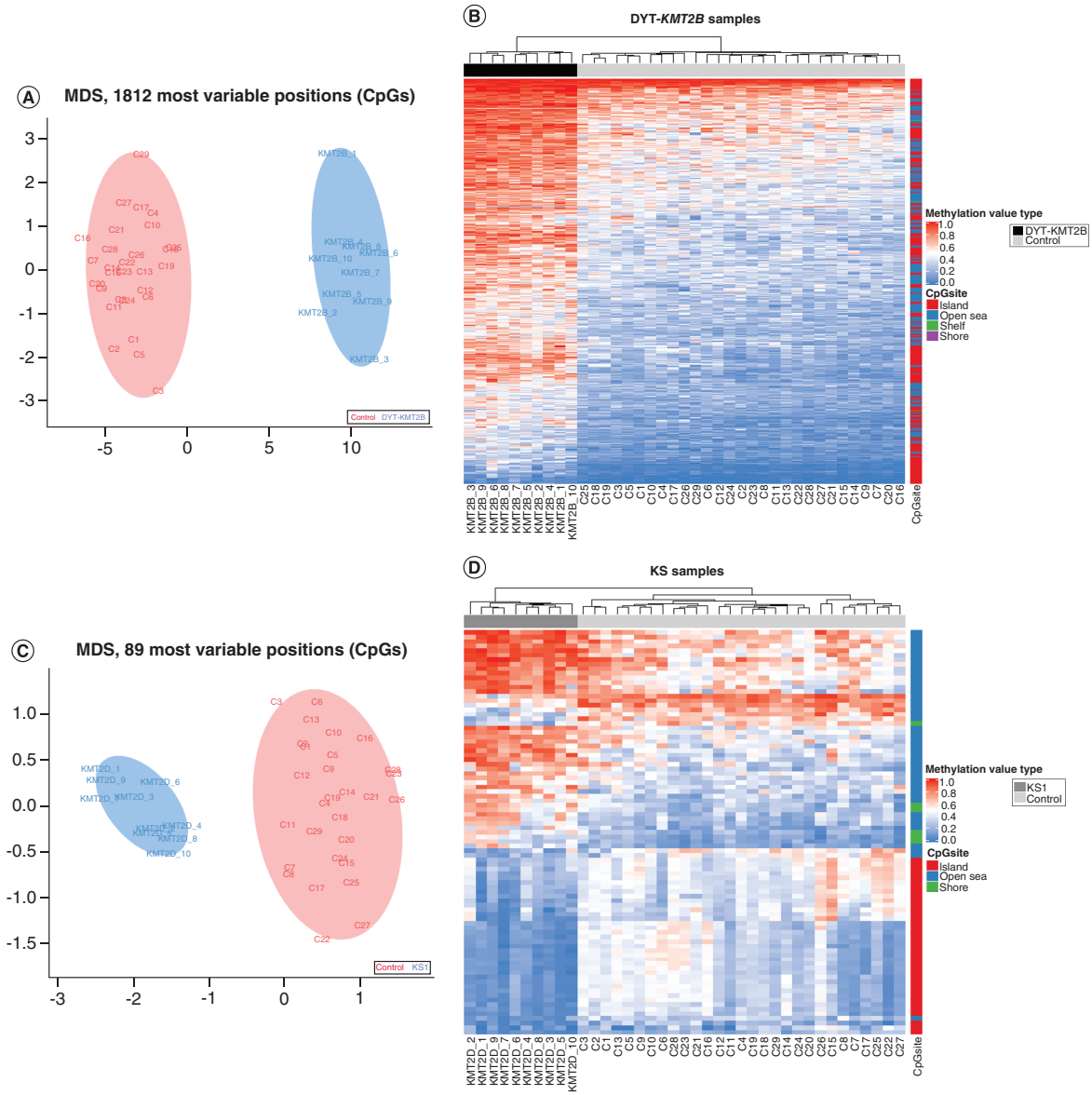

Figure 1. Genome-wide DNA methylation signature of Kabuki syndrome type 1 (KS1) and *KMT2B*-related dystonia (DYT-*KMT2B*) samples (nDYT-*KMT2B* = 10, nKS1 = 10, nControl = 29). (A & C) Multi-dimensional scaling plot is created based on **(A)** 1812 (DYT-*KMT2B*) and **(C)** 89 (KS1) variable positions of significant differentially methylated CpG positions (false discovery rate <0.05) between the disease group and control samples. All controls and samples are annotated with their identification numbers. Red area represents control samples and blue area indicates KMT samples. **(B & D)** Unsupervised clustering results of **(B)** 10 DYT-*KMT2B* and **(D)** 10 KS1 patients are shown in the heatmaps. Hierarchical clustering dendrogram is shown above the annotation bar. Methylation value ranges from 0 to 1 (0 is unmethylated and 1 is methylated differentially methylated CpG positions). Heatmap row indicates each CpG (hierarchical clustering by a complete linkage method, not sorted by location).
DYT-*KMT2B*: *KMT2B*-related dystonia; KS1: Kabuki syndrome type 1; MDS: Multiple dimensional scaling.

significant GOM patterns. In both cohorts, less than 3% of DMPs showed significant LOM patterns at a threshold of 3SD (2.80% in KS1 and 0.74% in DYT-*KMT2B*). Information of CpG sites for DYT-*KMT2B* and KS1 samples is available in Supplementary Table 2.

Methylation episignature comparisons for CpG islands in DYT-*KMT2B* & KS1 samples

A comparison was undertaken of the methylation profiling results of *KMT2D* and *KMT2B* variant samples and controls at DMPs involving CpG islands only. Except for one *KMT2D* variant sample (KMT2D_5), which did

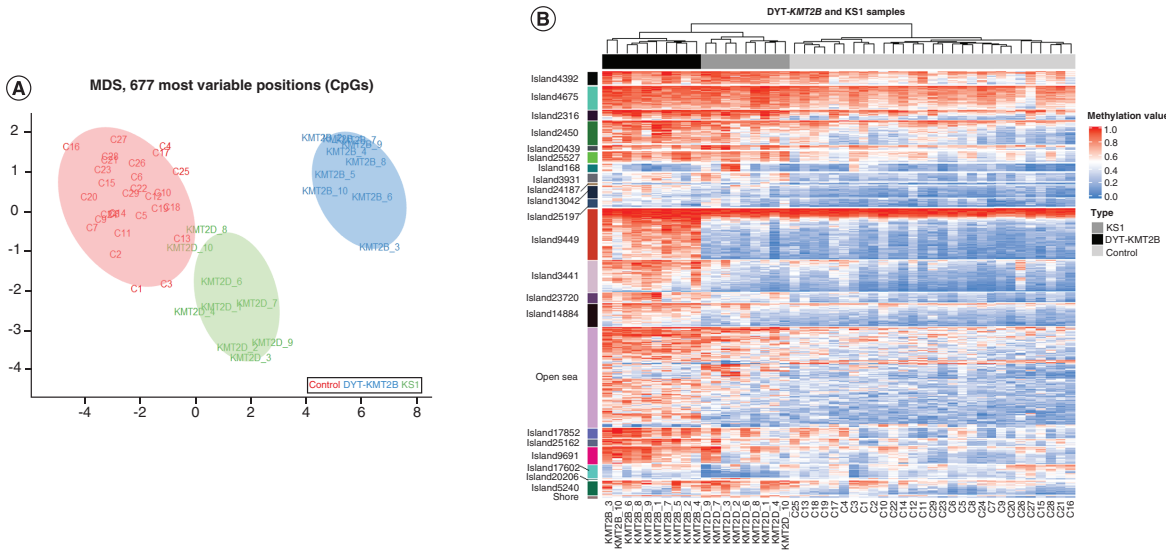

Figure 2. Methylation profiling of *KMT2B*-related dystonia (DYT-*KMT2B*) and Kabuki syndrome type 1 (KS1) samples for the comparative analysis (nDYT-*KMT2B* = 10, nKS1 = 9, nControl = 29). (A) Multiple dimensional scaling plot is created based on 677 variable positions of significant differentially methylated CpG positions (DMPs) (false discovery rate <0.05, including 21 CpG islands with 505 DMPs) among two disease groups (KS1 and DYT-*KMT2B*) compared with controls. All samples are annotated with their identification numbers. **(B)** DNA methylation signature of two disease groups is shown in the heatmap. CpG islands are organized by 'hclust' option in ComplexHeatmap package. Hierarchical clustering dendrogram is shown above the annotation bar. Methylation value ranges from 0 to 1 (0 is unmethylated and 1 is methylated DMPs).
DYT-*KMT2B*: *KMT2B*-related dystonia; KS1: Kabuki syndrome type 1; MDS: Multiple dimensional scaling.

not clearly separate from controls, the results with the 677 DMPs located in 21 CpG islands of 10 DYT-*KMT2B* and 9 KS1 samples clearly separated from each other and from controls (Figure 2). The methylation episignature of 9 KS1 (without the KMT2D_5 sample) is shown in Supplementary Figure 3 (alternative figure for Figure 1C & D). When comparing only CpG islands in 9 KS1 and 10 DYT-*KMT2B* samples (CpG islands specified in Figure 2B; additional heatmap showing only CpG islands' GOM and LOM pattern is in Supplementary Figure 4), interestingly, no CpG island region had LOM patterns in a DYT-*KMT2B* cohort, but 2.77% (14/505 DMPs in at least 1 KS1 sample) of the KS1 cohort showed significant LOM patterns. The identities of CpG islands that were significantly differentially methylated from controls in the DYT-*KMT2B* and/or KS1 cohorts are summarized in Supplementary Table 3.

As shown in Figure 2B, the DYT-*KMT2B* group was clearly separated from KS1 and control groups. Among 21 CpG islands, 8 CpG islands (CpG island 3441, 9449, 9691, 13042, 14884, 23720, 25162, 25527) were significantly differentially methylated in the DYT-*KMT2B* group, but not in KS1 samples (Table 2B & Supplementary Table 3). However, CpG island 3931 was significantly altered (FDR <0.05, see Table 2C & Supplementary Table 3) in the KS1 cohort but not in DYT-*KMT2B.* A further 12 CpG islands in Table 2A were significantly different from controls in both the KS1 and DYT-*KMT2B* groups (see Table 2 & Supplementary Table 3).

To investigate the potential molecular basis of genotype-phenotype correlations, the authors identified the CpG islands that were differentially methylated in DYT-*KMT2B* and/or KS1 and annotated them for their associated genes, tissue expression patterns and previous reports related to neurodevelopmental or neurodegenerative disorders (Table 2). Loci significantly altered in KMT2 groups (DYT-*KMT2B* + KS1) are included in those associated with 15 protein-coding genes (Supplementary Table 3). Two of these genes (*GABRB3*, *GABRA5*) in CpG island 17852 and *RAI1* in CpG island 20439 have been associated with neurodevelopmental disorders. Thus, monoallelic pathogenic variants in the gamma-aminobutyric acid receptors, *GABRB3* and *GABRA5*, are associated with developmental and epileptic encephalopathies 43 (MIM 617113) and 79 (MIM 618559), respectively. *ARHGAP35* has been linked to intellectual disability [36] and heterozygous microdeletions and point mutations in the *RAI1* gene cause Smith–Magenis syndrome (MIM 182290), which is associated with intellectual disability, sleep and behavioral disturbances and congenital anomalies [37–39].

Table 2. Detailed description of genes associated with selected CpG islands in KMT2B-related dystonia and Kabuki syndrome type 1 cohorts.

CpG island ID (location)	Gene (OMIM ID††)	Highest gene expression (GTEx)	Associated inherited disorder?
A. Significantly altered (differentially methylated) in DYT-*KMT2B* and KS1 samples compared with controls (n = 12†)			
17852 (chr15:27138902–27139653)	*GABRA5* (MIM 137142)	Brain	Developmental and epileptic encephalopathy 79 (AR, MIM 618559††)
	GABRB3 (MIM 137192)	Brain	Developmental and epileptic encephalopathy 43 (AR, MIM 617113) Epilepsy, childhood absence, susceptibility to 5 (unknown, MIM 612269)
20439 (chr17:17603584–17604147)	*RAI1* (MIM 607642)	Cervix and others	Smith–Magenis syndrome (AR, MIM 182290) Potocki–Lupski syndrome (AD, MIM 610883)
24187 (chr19:47507307–47507692)	*ARHGAP35* (MIM 605277)	Testis, brain and others	ARHGAP35-related developmental disorder (monoallelic, DD¶)
B. Significantly altered (differentially methylated) in DYT-*KMT2B* but not in KS1 samples (n = 40‡)			
CpG island ID (location)	**Gene (OMIM ID)**	**Highest gene expression (GTEx)**	**Associated inherited disorder?**
3067 (chr2: 87036601–87037001)	*CD8A* (MIM 186910) Regulated#	Spleen, blood	CD8 deficiency, familial (AR, MIM 608957)
4760 (chr3: 113160300–113160641)	*CFAP44* (MIM 617559)	Pituitary	Spermatogenic failure 20 (AR, MIM 617593)
15210 (chr12: 2339163-2339615)	*CACNA1C* (MIM 114205)	Colon	Timothy syndrome (AD, MIM 601005) Brugada syndrome 3 (unknown, MIM 611875) Long QT syndrome 8 (unknown, MIM 618447)
18493 (chr15: 91473315–91473592)	*UNC45A* (MIM 611219)	Artery and others	Osteootohepatoenteric syndrome (AR, MIM 619377)
23720 (chr19: 36484905–36485476)	*SDHAF1* (MIM 612848) Regulated#	Brain and others	Mitochondrial complex II deficiency, nuclear type 2 and Leigh syndrome (AR, MIM 619166)
C. Significantly altered (differentially methylated) in KS1 but not in DYT-*KMT2B* samples (n = 4§)			
CpG Island ID (location)	**Gene (OMIM ID)**	**Highest gene expression (GTEx)**	**Associated inherited disorder?**
21752 (chr18:7011268–7011488)	*LAMA1* (MIM 150320)	Testis	Poretti–Boltshauser syndrome (AR, MIM 615960)

Genes associated with each CpG island are listed above. CpG islands are organized by their unique Illumina TruSeq Methyl Capture EPIC kit IDs and listed in order of the location. The location and gene expression information were extracted from the University of California Santa Cruz (https://genome.ucsc.edu/) and the *Ensembl* genome browser (http://grch37.ensembl.org/index.html). Variants and phenotype information are from DECIPHER (www.deciphergenomics.org/). Information for other CpG islands not listed in this table is summarized in Supplementary Table 3.

†CpG islands 168, 2316, 2450, 4392, 4675, 5240, 17602, 20206, 25197 have no associated inherited disorder.

‡CpG islands 1372, 1545, 1595, 2430, 3388, 3441, 6625, 7814, 8144, 8448, 8781, 9255, 9449, 9458, 9691, 11483, 13042, 13355, 14877, 14878, 14884, 14907, 15752, 17000, 17826, 18566, 19054, 19453, 20506, 20799, 22067, 23279, 25162, 25527, 25653 have no associated inherited disorder.

§CpG islands 3931, 21001, 24842 have no associated inherited disorder.

¶DD: diseases confirmed by DDG2P (www.ebi.ac.uk/gene2phenotype).

#Regulated: genes regulated by the promoter in designated CpG island (not located within CpG island).

††MIM: OMIM ID (www.omim.org).

AD: Autosomal dominant inheritance; AR: Autosomal recessive inheritance; DYT-*KMT2B*: *KMT2B*-related dystonia; KS1: Kabuki syndrome type 1.

Four CpG islands (details in Table 2B) that were significantly different in the KS1 group only were associated with 4 protein-coding genes, 1 of which, *LAMA1*, is associated with an autosomal recessively inherited neurodevelopmental disorder (Poretti–Boltshauser syndrome; MIM 615960) [40].

The prominent occurrence of dystonia distinguishes DYT-*KMT2B* from other KMT2-related disorders, and the authors searched among the significantly altered CpG island-associated genes (compared with KS1 and controls) to see if there were any that might be linked to dystonia (Table 2C). From the developmental disorder-associated genes, none had previously been linked to dystonia and though dystonia may occur in Leigh encephalomyopathy, this has been primarily associated with mutations in *SUCLA2* leading to succinyl-CoA deficiency (MIM 612073) rather than *SDHAF1*, which showed significantly altered methylation (GOM) in DYT-*KMT2B* samples. Biallelic mutations in *SDHAF1* are associated with mitochondrial complex II deficiency (MIM 619166), Leigh syndrome and infantile leukodystrophies [41,42]. Using a training set of 15 dystonia-associated genes (*ANO3, ATP1A3, GCH1, KMT2B, MR1, PRKRA, PRRT2, SGCE, SLC2A1, SUCLA2, TAF1, TH, THAP1, TOR1A, TUBB4A*), the protein-coding genes that were associated with significantly altered CpG islands in *KMT2B* were ranked by a gene

prioritization tool (https://toppgene.cchmc.org/prioritization.jsp). The top-ranking genes were *CACNA1C* (overall p 1.44E-04), *TTBK1* (p = 0.0075), *SLC5A3* (p = 0.073) and *SDHAF1* (p = 0.074) (see Supplementary Table 4).

Discussion

The authors analyzed DNA methylation profiles in two patient cohorts with KMT2-related disorders, DYT-*KMT2B* and KS1. Previously methylation profiling using Illumina methylation arrays has been reported in more than 90 patients with KS1 and, while this study was in progress, 21 patients with DYT-*KMT2B* were reported in two studies [12,20,22,24,25,28,29,43]. The current work utilized EPIC-NGS to analyze more than twice the number of CpGs than array-based methods and included 97% of CpG islands in the human genome (26,918/27,718 CpG islands) [30].

The authors detected an episignature of 1812 significant DMPs and 40 CpG islands in a cohort of 10 DYT-*KMT2B* patients with LOF variants (8 frameshifts, 1 donor splice site lost, 1 stop gain). The authors confirmed the results of Ciolfi *et al.* [20], who reported a methylation episignature in DYT-*KMT2B* patients using 8 pathogenic variants and 10 variants of uncertain significance (VUSs), and Mirza-Schreiber *et al.,* who identified a DYT-*KMT2B* methylation episignature with 13 samples with pathogenic/likely pathogenic variants in *KMT2B* and 4 samples with a VUS [43]. Irrespective of methylation profiling platform (array or EPIC-NGS), all three studies have reported that DYT-*KMT2B* is associated with genome hypermethylation. It has previously been suggested that in mouse embryonic stem cells, *KMT2B* protects developmental genes from transcriptional repression by inhibiting polycomb repressive complex 2 and DNA methylation machineries [44]. Potential therapeutic interventions might therefore include demethylating agents [44].

Comparing the pattern of methylation alterations in DYT-*KMT2B* detected by EPIC-NGS (n = 1812 at an FDR < 0.05) and those reported using methylation arrays (EPIC-array) by Ciolfi *et al.* [20] revealed that many (1649/1812) of the significant EPIC-NGS DMPs were not covered by EPIC array. Previously, 196 significant DMPs (FDR < 0.01 and methylation difference >10%) were detected by EPIC-array but only 4 CpGs in 2 CpG islands (CpG island 1372 and CpG island 18566) overlapped with the EPIC-NGS DMPs.

Comparing the methylation episignatures of the DYT-*KMT2B* and KS1 groups, the authors found that there were differences in the specific patterns of methylation alterations such that although all significant alterations (CpG islands) in the DYT-*KMT2B* cases were GOM events for the KS1 samples, a subset of loci showed hypomethylation alterations. The authors were interested to see whether the hypermethylated loci associated with DYT-*KMT2B* alone might indicate any potential candidate genes for the distinctive phenotypic aspect of this disorder (dystonia). However, none of the altered CpG island-associated genes would appear to provide a compelling explanation for the cause of dystonia. Gene prioritization analysis identified *CACNA1C* as the top-rated candidate gene. In humans, germline gain of function variants in *CACNA1C* are associated with Timothy syndrome, a multi-system disorder associated with long QT syndrome, congenital heart disease, immune abnormalities, syndactyly and short stature [45–47]. Individuals with Timothy syndrome may also display developmental delay, autism and epilepsy [45,46]. However, in a mouse model with homozygous inactivation of *CACNA1A*, null mice developed dystonia and then progressive cerebellar ataxia and died at a few weeks of age [48,49]. It will therefore be of interest to know whether the *CACNA1C*-associated methylation alterations observed in blood DNA could also be detected in affected brain tissue from patients with DYT-*KMT2B* and determine the effect of methylation on transcription.

Conclusion

In conclusion, this work confirmed the presence of a hypermethylation episignature associated with pathogenic variants in *KMT2B.* Direct comparison of EPIC-NGS and EPIC-array on the same samples would facilitate the assessment of the optimum methodology for clinical application of methylation episignature analysis. Identification of episignatures associated with rare inherited disorders can improve the diagnostic and prognostic assessment of these disorders and could contribute to a better understanding of the pathogenesis of these disorders.

Future perspective

There is increasing interest in the definition of specific methylation episignatures in developmental disorders associated with pathogenic variants in genes that regulate chromatin structure and/or function. Methylation episignature analysis is being introduced into clinical practice to inform variant pathogenicity in relevant genes and this practice should grow further. Similarly, its role in the investigation of undiagnosed developmental disorders should increase and could lead to novel disease gene discoveries. Further work is required to define the most

sensitive and cost-effective methodology for defining methylation episignatures in clinical diagnostics and how machine learning-based approaches can improve diagnostic accuracy. If methylation patterns in blood correlate with those in the brain and other affected organs, then methylation profiling might reveal epigenotype-phenotype correlations and a potential role in predicting prognosis.

Summary points

- KMT2 methyltransferases are histone-modifying enzymes that catalyze the methylation of lysine 4 of the Histone 3 protein and regulate transcriptional activity.
- In humans, pathogenic variants in the *KMT2* gene family can cause a variety of rare neurodevelopmental disorders such as early-onset dystonia (*KMT2B*-related dystonia [DYT-*KMT2B*]) and Kabuki syndrome type 1 (KS1).
- Many neurodevelopmental disorders caused by variants in genes regulating chromatin function and/or structure display abnormal DNA methylation patterns (episignatures) in peripheral blood.
- DYT-*KMT2B* is unique among 'chromatin neurodevelopmental disorders' as the most prominent clinical feature is childhood-onset dystonia rather than developmental delay or congenital anomalies.
- Comparison of methylation episignatures in DYT-*KMT2B* and KS1 samples showed that most differentially methylated CpG positions were specific to one disorder and that all (DYT-*KMT2B*) and most (KS1) methylation alterations in CpG islands were gain of methylation events.
- Analysis of genes associated with CpG islands with significant differentially methylated regions suggested potential candidate genes for the molecular pathogenesis of DYT-*KMT2B*.
- Using different methodology for methylation profiling, the authors replicated recent reports of a peripheral blood methylation signature for DYT-*KMT2B*.
- Methylation episignatures could be used to aid pathogenicity interpretation of *KMT2B* variants and can complement mechanistic investigations of the pathogenesis of DYT-*KMT2B*.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/epi-2021-0521

Author contributions

S Lee, E Ochoa, and ER Maher conceived and planned the experiments and the analysis. S Lee and E Ochoa designed the pipeline. Investigations were performed by S Lee, E Ochoa, F Rodger, F Docquier, G Clark, E Martin and ER Maher. Patient data were provided by K Barwick, L Cif, B Perez, S Banka and MA Kurian. E Ochoa and ER Maher supervised the project. S Lee, E Ochoa and ER Maher wrote the original draft, and all authors critically reviewed and edited the manuscript.

Acknowledgments

The authors are grateful to JP Lin, M King, T Lynch, L Carr, B Garavaglia and I Valenzuela for providing samples for this research. The authors acknowledge support from the NIHR UK Rare Genetic Disease Research Consortium.

Financial & competing interests disclosure

This research was co-funded by the NIHR Cambridge Biomedical Research Centre and Rosetrees Trust (to E Ochoa, S Lee and ER Maher). The University of Cambridge has received salary support (ER Maher) from the NHS in the East of England through the Clinical Academic Reserve. The views expressed are those of the authors and not necessarily those of the NHS or Department of Health. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval from South Birmingham and Central Manchester Research Ethics Committees. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Data availability

The processed datasets generated and analyzed during this study are available in Supplementary Tables 5–8. In addition, raw FASTQ files and processed datasets (bismark CpG coverage files and significant differentially methylated regions selected for this study) of 29 healthy controls and 20 samples are available in GEO (reference GSE199836).

References

1. Black JC, Van Rechem C, Whetstine JR. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cel.* 48, 491–507 (2012).
2. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395 (2011).
3. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* doi:10.1038/emm.2017.11 (2017).
4. Husmann D, Gozani O, Struct N, Biol M. Histone lysine methyltransferases in biology and disease HHS Public Access Author manuscript. *Nat. Struct. Mol. Bio.* 26, 880–889 (2019).
5. Jones WD, Dafou D, McEntagart M *et al. De novo* mutations in MLL cause Wiedemann–Steiner syndrome. *Am. J. Hum. Genet.* 91, 358–364 (2012).
6. Zech M, Boesch S, Maier EM *et al.* Haploinsufficiency of KMT2B, encoding the lysine-specific histone methyltransferase 2B, results in early-onset generalized dystonia. *Am. J. Hum. Genet.* 99, 1377–1387 (2016).
7. Kleefstra T, Kramer JM, Neveling K *et al.* Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. *Am. J. Hum. Genet.* 91, 73–82 (2012).
8. Ng SB, Bigham AW, Buckingham KJ *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* 42, 790–793 (2010).
9. O'Donnell-Luria AH, Pais LS, Faundes V *et al.* Heterozygous variants in KMT2E cause a spectrum of neurodevelopmental disorders and epilepsy. *Am. J. Hum. Genet.* 104, 1210–1222 (2019).
10. Faundes V, Newman WG, Bernardini L *et al.* Histone lysine methylases and demethylases in the landscape of human developmental disorders. *Am. J. Hum. Genet.* 102, 175–187 (2018).
11. Gabriele M, Vitriolo A, Cuvertino S *et al. bioRxiv* KMT2D haploinsufficiency in Kabuki syndrome disrupts neuronal function through transcriptional and chromatin rewiring 2 independent of H3K4-monomethylation. doi:10.1101/2021.04.22.440945 (2021).
12. Sobreira N, Brucato M, Zhang L *et al.* Patients with a Kabuki syndrome phenotype demonstrate DNA methylation abnormalities. *Eur. J. Hum. Genet.* 25, 1335–1344 (2017).
13. Adam MP, Banka S, Bjornsson H *et al.* Kabuki syndrome: international consensus diagnostic criteria. *J. Med. Genet.* 56, 89–95 (2019).
14. Adam MP, Hudgins L, Hannibal M. Kabuki syndrome. *GeneReviews*® (2021).
15. Faundes V, Goh S, Akilapa R *et al.* Clinical delineation, sex differences, and genotype–phenotype correlation in pathogenic KDM6A variants causing X-linked Kabuki syndrome type 2. *Genet. Med.* 23, 1202–1210 (2021).
16. Meyer E, Carss KJ, Rankin J *et al.* Mutations in the histone methyltransferase gene *KMT2B* cause complex early-onset dystonia. *Nat. Genet.* 49, 223–237 (2017).
17. Carecchio M, Invernizzi F, Gonzàlez-Latapi P *et al.* Frequency and phenotypic spectrum of *KMT2B* dystonia in childhood: a single-center cohort study. *Mov. Disord.* 34, 1516–1527 (2019).
18. Gorman KM, Meyer E, Kurian MA. Review of the phenotype of early-onset generalised progressive dystonia due to mutations in *KMT2B*. *Eur. J. Paediatr. Neurol.* 22, 245–256 (2018).
19. Cif L, Demailly D, Lin JP *et al. KMT2B*-related disorders: expansion of the phenotypic spectrum and long-term efficacy of deep brain stimulation. *Brain* 143, 3242–3261 (2020).
20. Ciolfi A, Foroutan A, Capuano A *et al.* Childhood-onset dystonia-causing *KMT2B* variants result in a distinctive genomic hypermethylation profile. *Clin. Epigenetics* 13, 1–11 (2021).
21. Zech M, Lam DD, Winkelmann J. Update on *KMT2B*-related dystonia. *Curr. Neurol. Neurosci. Reports* 19, 1–11 (2019).
22. Aref-Eshghi E, Bourque DK, Kerkhof J *et al.* Genome-wide DNA methylation and RNA analyses enable reclassification of two variants of uncertain significance in a patient with clinical Kabuki syndrome. *Hum. Mutat.* 40, 1684–1689 (2019).
23. Aref-Eshghi E, Rodenhiser DI, Schenkel LC *et al.* Genomic DNA methylation signatures enable concurrent diagnosis and clinical genetic variant classification in neurodevelopmental Syndromes. *Am. J. Hum. Genet.* 102, 156–174 (2018).

24. Butcher DT, Cytrynbaum C, Turinsky AL *et al.* CHARGE and Kabuki syndromes: gene-specific DNA methylation signatures identify epigenetic mechanisms linking these clinically overlapping conditions. *Am. J. Hum. Genet.* 100, 773–788 (2017).
25. Aref-Eshghi E, Schenkel LC, Lin H *et al.* The defining DNA methylation signature of Kabuki syndrome enables functional assessment of genetic variants of unknown clinical significance. *Epigenetics* doi:10.1080/15592294.2017.1381807 (2017).
26. Sarmiento IJK, Mencacci NE. Genetic dystonias: update on classification and new genetic discoveries. *Curr. Neurol. Neurosci. Rep.* 21, 8 (2021).
27. Sadikovic B, Levy MA, Kerkhof J *et al.* Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. *Genet. Med.* 23, 1065 (2021).
28. Jeffries AR, Maroofian R, Salter CG *et al.* Growth disrupting mutations in epigenetic regulatory molecules are associated with abnormalities of epigenetic aging. *Genome Res.* 29, 1057–1066 (2019).
29. Aref-Eshghi E, Kerkhof J, Pedro VP *et al.* Evaluation of DNA methylation episignatures for diagnosis and phenotype correlations in 42 Mendelian neurodevelopmental disorders. *Am. J. Hum. Genet.* 106, 356–370 (2020).
30. NGS Methyl-Seq libraries. (Accessed 30 September 2021). https://sapac.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-methyl-capture-epic.html
31. Banka S, Veeramachaneni R, Reardon W *et al.* How genetically heterogeneous is Kabuki syndrome? MLL2 testing in 116 patients, review and analyses of mutation and phenotypic spectrum. *Eur. J. Hum. Genet.* 20, 381–388 (2012).
32. Assenov Y, Müller F, Lutsik P *et al.* Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods* doi:10.1038/nmeth.3115 (2014).
33. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, doi:10.2202/1544-6115.1027 (2004).
34. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. sva: Surrogate Variable Analysis R package version 3.10.0. www.Bioconductor.Org/Packages/Release/Bioc/Html/Sva.Html (2014).
35. Makambi K. Weighted inverse chi-square method for correlated significance tests. *J. Applied Stat.* 30, 225–234 (2010).
36. Kaplanis J, Samocha KE, Wiel L *et al.* Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature 2020* 586, 757–762 (2020).
37. Smith ACM, Dykens E, Greenberg F. Sleep disturbance in Smith–Magenis syndrome (del 17 p11.2). *Am. J. Med. Genet. (Neuropsychiatr. Genet.)* 81, 186–191 (1998).
38. Greenberg F, Lewis RA, Potocki L *et al.* Multi-disciplinary clinical study of Smith–Magenis syndrome (deletion 17p11.2). *Am. J. Med. Genet.* 62, 247–254 (1996).
39. Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH. Mutations in RAI1 associated with Smith–Magenis syndrome. *Nat. Genet.* 33, 466–468 (2003).
40. Aldinger KA, Mosca SJ, Tétreault M *et al.* Mutations in *LAMA1* cause cerebellar dysplasia and cysts with and without retinal dystrophy. *Am. J. Hum. Genet.* 95, 227–234 (2014).
41. Fullerton M, McFarland R, Taylor RW, Alston CL. The genetic basis of isolated mitochondrial complex II deficiency. *Mol. Genet. Metab.* 131, 53–65 (2020).
42. Hoekstra AS, Bayley JP. The role of complex II in disease. *Biochim. Biophys. Acta – Bioenerg.* 1827, 543–551 (2013).
43. Mirza-Schreiber N, Zech M, Wilson R *et al. Brain*. Blood DNA methylation provides an accurate biomarker of *KMT2B*-related dystonia and predicts onset. doi:10.1093/brain/awab360/6378248 (2021). https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awab360/6378248?login=true
44. Douillet D, Sze CC, Ryan C *et al.* Uncoupling histone H3K4 trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation. *Nat. Genet.* 52, 615–625 (2020).
45. Splawski I, Timothy KW, Sharpe LM *et al.* Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. *Cell* 119, 19–31 (2004).
46. Splawski I, Timothy KW, Decher N *et al.* Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations. *Proc. Natl Acad. Sci. U.S.A.* 102, 8089–8096 (2005).
47. Napolitano C, Timothy KW, Bloise R, Priori SG *et al. CACNA1C*-related disorders. *Card. Electrophysiol. From Cell to Bedside Seventh Ed.* 910–916 doi:10.1016/B978-0-323-44733-1.00095-X (2021).
48. Jun K, Piedras-Rentería ES, Smith SM *et al.* Ablation of P/Q-type Ca(2+) channel currents, altered synaptic transmission, and progressive ataxia in mice lacking the alpha(1A)-subunit. *Proc. Natl Acad. Sci. USA* 96, 15245–15250 (1999).
49. Fletcher CF, Tottene A, Lennon VA *et al.* Dystonia and cerebellar atrophy in *CACNA1A* null mice lacking P/Q calcium channel activity. *FASEB J.* 15, 1288–1290 (2001).