


RESEARCH

Open Access



# Childhood-onset dystonia-causing *KMT2B* variants result in a distinctive genomic hypermethylation profile

Andrea Cioffi<sup>1†</sup>, Aidin Foroutan<sup>2,3†</sup>, Alessandro Capuano<sup>4</sup>, Lucia Pedace<sup>5</sup>, Lorena Travaglini<sup>1</sup>, Simone Pizzi<sup>1</sup>, Marco Andreani<sup>5</sup>, Evelina Miele<sup>5</sup>, Federica Invernizzi<sup>6</sup>, Chiara Reale<sup>6</sup>, Celeste Panteghini<sup>6</sup>, Maria Iascone<sup>7</sup>, Marcello Niceta<sup>1</sup>, Ralitzia H. Gavrilova<sup>8</sup>, Laura Schultz-Rogers<sup>8</sup>, Emanuele Agolini<sup>9</sup>, Maria Francesca Bedeschi<sup>10</sup>, Paolo Prontera<sup>11</sup>, Matteo Garibaldi<sup>12</sup>, Serena Galosi<sup>13</sup>, Vincenzo Leuzzi<sup>13</sup>, Paola Soliveri<sup>14</sup>, Rory J. Olson<sup>8</sup>, Giovanna S. Zorzi<sup>15</sup>, Barbara M. Garavaglia<sup>6</sup>, Marco Tartaglia<sup>1\*</sup>  and Bekim Sadikovic<sup>2,3,16\*</sup>

## Abstract

**Background:** Dystonia is a clinically and genetically heterogeneous movement disorder characterized by sustained or intermittent muscle contractions causing abnormal, often repetitive, movements and/or postures. Heterozygous variants in lysine methyltransferase 2B (*KMT2B*), encoding a histone H3 methyltransferase, have been associated with a childhood-onset, progressive and complex form of dystonia (dystonia 28, DYT28). Since 2016, more than one hundred rare *KMT2B* variants have been reported, including frameshift, nonsense, splice site, missense and other in-frame changes, many having an uncertain clinical impact.

**Results:** We characterize the genome-wide peripheral blood DNA methylation profiles of a cohort of 18 patients with pathogenic and unclassified *KMT2B* variants. We resolve the “episignature” associated with *KMT2B* haploinsufficiency, proving that this approach is robust in diagnosing clinically unsolved cases, properly classifying them with respect to other partially overlapping dystonic phenotypes, other rare neurodevelopmental disorders and healthy controls. Notably, defective *KMT2B* function in DYT28 causes a non-random DNA hypermethylation across the genome, selectively involving promoters and other regulatory regions positively controlling gene expression.

**Conclusions:** We demonstrate a distinctive DNA hypermethylation pattern associated with DYT28, provide an epigenetic signature for this disorder enabling accurate diagnosis and reclassification of ambiguous genetic findings and suggest potential therapeutic approaches.

**Keywords:** DNA methylation, Episignature, *KMT2B*, Dystonia 28

## Background

Dystonia is a neurological hyperkinetic movement condition characterized by sustained or intermittent muscle contractions causing abnormal movements and/or postures. Dystonia’s symptoms can be characterized based on body location into focal, segmental or generalized forms, as well as into isolated or combined forms, whether other movement disorders or neurological signs occur or not [1]. Childhood-onset dystonias, which are

\*Correspondence: marco.tartaglia@opbg.net; bekim.sadikovic@lhsc.on.ca

†Andrea Cioffi and Aidin Foroutan have contributed equally to this work and Marco Tartaglia and Bekim Sadikovic have jointly coordinated this work

<sup>1</sup> Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, IRCCS, 00146 Rome, Italy

<sup>2</sup> Department of Pathology and Laboratory Medicine, Western University, London, ON N6A 3K7, Canada

Full list of author information is available at the end of the article



usually genetically determined and can be characterized by other additional neuropsychiatric and systemic features, pose a significant diagnostic challenge for clinicians [1].

The advent of high-throughput sequencing has revolutionized the landscape of dystonia's genetics, enabling identification of several novel disease genes in the last decade, most of them causing complex forms [2]. Recently, heterozygous variants in lysine methyltransferase 2B (*KMT2B*; MIM \*606834), encoding a histone H3 methyltransferase, have been associated with a childhood-onset, progressive and complex form of dystonia named dystonia 28 (DYT28; MIM #617284) [3–5]. Pathogenic *KMT2B* variants generally are de novo events and result in loss-of-function of the methyltransferase, indicating haploinsufficiency as the mechanism of disease [3–6]. Although the pathomechanistic consequences of *KMT2B* variants require further study, *KMT2B* haploinsufficiency is postulated to affect CNS development and function by perturbing the expression of key genes involved in neurodevelopment and motor control [7, 8]. Of note, while *KMT2B* was identified as a dystonia-associated gene only recently, more than one hundred rare or private heterozygous single nucleotide variants (SNVs) as well as insertions/deletions (indels) have been identified in this gene, the majority predicted to cause haploinsufficiency [4–6, 9]. A small proportion of *bona fide* pathogenic *KMT2B* variants are missense and have been reported to affect conserved functional domains of the protein (*i.e.*, the PHD-like and FYR-N domains). However, a number of clinically unclassified missense variants have been described and, in a significant proportion of cases, these variants are inherited from a healthy parent [3, 6]. Currently, no functional assay is available to classify these variants.

*KMT2B* belongs to the family of Set1-Trithorax-type methyltransferases, which are enzymes that specifically methylate histone 3 at lysine 4 (H3K4) and play a key role in chromatin remodeling and gene expression [10]. Mammals possess six SET-related H3K4 methyltransferases (*i.e.*, *KMT2A* (MLL1), *KMT2B* (MLL2), *KMT2C* (MLL3), *KMT2D* (MLL4), *SETD1A* and *SETD1B*), which are dynamically regulated during cell-type specification, in a spatially and temporally non-redundant way [7, 11]. H3K4 can be mono- (H3K4me1), di- (H3K4me2) and tri-methylated (H3K4me3), with each modification differentially distributed in promoters, enhancers and other regulatory regions of actively transcribed genes [12]. While *SETD1A* and *SETD1B* are major contributors of global histone H3K4 tri-methylation (H3K4me3), *KMT2B* and *KMT2A* H3K4me3 marks occur at the transcription start site (TSS) regions of a more restricted set of genes. Different from the other members of the *KMT2*

family, *KMT2C* and *KMT2D* catalyze H3K4me1/2, which are particularly enriched at enhancers [12].

There is a growing evidence that variants in genes encoding proteins involved in the maintenance of chromatin remodeling show unique DNA methylation (DNAm) patterns (known as “episignatures” or Epi-Signs), and that these DNAm signatures can be used as highly specific and robust biomarkers for an increasing number of disorders caused by mutations in these genes [13–16]. These genome-wide DNAm signatures currently include over 40 rare neurodevelopmental disorders associated with more than 60 genes [13, 17]. Of note, these disease-specific episignatures are detectable in peripheral blood despite the variable nature and complexity of diseases and the high variance characterizing the DNAm of genomes from different cells and tissues [13, 17]. As such, DNA methylation testing has recently been implemented in clinical diagnosis of patients with rare disorders [17].

By characterizing the genome-wide DNA methylation profiles of a cohort of clinically and genetically confirmed DYT28 patients, here we define a disease-specific episignature for this disorder and show that the identified methylation pattern could be successfully used to classify *KMT2B* variants of uncertain significance (VUS) and help diagnose clinically unsolved cases. We also provide evidence that dysfunctional *KMT2B* in DYT28 causes DNA hypermethylation of promoters and other regulatory regions positively controlling gene expression, which collectively points out toward a general repression of transcriptional activity as pathogenic mechanism in DYT28.

## Methods

### Study cohort

The study included eight case individuals (five females and three males) with *bona fide* pathogenic *KMT2B* variants and clinically confirmed DYT28 (labelled as DYT28\_Pathogenic), and 10 subjects (seven females and three males) showing variable clinical phenotypes and *KMT2B* VUS (labelled as DYT28\_VUS), according to the American College of Medical Genetics and Genomics/Association for Molecular Pathology (ACMG/AMP) guidelines for interpretation of genomic sequence variants [18] (Additional file 9: Table S1). Additionally, nine control samples were also used (labelled as Control\_Testing). The study was approved by the Ospedale Pediatrico Bambino Gesù Ethical Committee (1702\_OPBG\_2018) and the Western University Research Ethics Board (REB 106302). All DNA samples and clinical records were pseudonymized. DNA specimens were collected following procedures in accordance with the ethical standards of the declaration of Helsinki protocols, with signed informed consents from the participating subjects/

families. Peripheral blood DNA was extracted using the salting out procedure [19] (patients 1, 2, 5–11, 14 and 18) or the Qiagen DNA extraction kit (patients 3, 4, 12, 13, 15–17). All variants had been confirmed by Sanger sequencing using the Big-Dye terminator reaction Kit v.1.1 on a 3100XL Genetic Analyzer Automated Sequencer (Applied BioSystems) (patients 1, 2, 5–11, 14 and 18) or using the Big-Dye terminator reaction Kit v.3.1 on a SeqStudio Genetic Analyzer (Applied Biosystems) (patients 3, 4, 12, 13, 15–17) (Additional file 1: Figure S1).

### DNA methylation profiling and data analysis

Following bisulfite conversion, samples were analyzed using Illumina Infinium MethylationEPIC BeadChips, according to the manufacturer's protocol. Data analysis was carried out as previously reported [13, 14]. Briefly, IDAT files containing methylated and unmethylated signal intensity were imported into R v.4.0.2 for analysis following normalization with background correction using the minfi package [20]. Probes located on X/Y chromosomes or known to cross-react with chromosomal locations other than their target regions contain SNPs at or near the CpG sites, and suggested by Illumina to be cross-reactive were excluded, resulting in 776,314 probes remaining for the analysis. Arrays having more than 5% probe failure rate and those that were previously identified in our database to impose batch effect were excluded from the analysis. Sex of the two unknown samples was predicted using minfi package [20], whereas for the age estimation the wateRmelon package was used [21]. The eight samples with DYT28-causing *KMT2B* variants (DYT28\_Pathogenic) were used to identify the epismature, while the 10 additional samples (DYT28\_VUS) were used for validation and classification by blind testing (Additional file 9: Table S1).

Principal component analysis (PCA) was performed to inspect any batch effect and identify outlier samples. MatchIt package was used to select best-matching controls from EPIC arrays in the EpiSign Knowledge Database (EKD) at the London Health Sciences Center (LHSC) considering age, and sex, as matching variables, providing a control sample size seven times larger than that of tested cases (56 controls labelled as Control\_Training) [13, 17]. The control cohort characterization is provided in Additional file 10: Table S2.

Methylation levels (beta values) were converted to M-values, which were used for linear regression modeling by means of empirical Bayes moderated t-statistic corrected for false discovery rate (FDR) using the Benjamini-Hochberg (BH) method (limma package [22]) to identify differentially methylated probes (DMPs). Estimated blood cell proportions for each sample were added to the model matrix to reduce the bias associated

with those confounding variables [23]. The most informative 1,000 probes were identified considering the interaction between the effect size (absolute mean methylation difference between DYT28 samples and batch controls) and *p*-value [13, 14]. Receiver's operating curve characteristic analysis was performed to identify the top 500 of these 1000 probes, then probes with a Pearson's pairwise correlation >0.9 were removed, resulting in identification of 196 independent probes.

Hierarchical clustering was performed using the gplots package. Multidimensional scaling (MDS) was performed by scaling of the pairwise Euclidean distances between samples. The e1071 R package was used to train a support vector machine (SVM) and for construction of a prediction model to calculate what we refer to as "methylation variant pathogenicity" (MVP) scores [13]. The eight samples with DYT28, and samples from the EKD databases split into two cohorts, 75% of control subjects (>1000 individuals) and 75% of patients from 38 other neurodevelopmental disorders/rare diseases (NDDs/RDs) in the EKD (>1000 individuals) were used as training set. The remaining 25% control subjects and patients with *KMT2B* VUS were considered as the testing set (Additional file 9: Table S1), in order to improve the specificity of the classifier. An MVP plot was generated to assess specificity of the classification model.

### Functional analysis of differentially methylated regions

To detect the differentially methylated regions (DMRs), the DMRcate package was used [24], and regions containing at least five different CpGs within 1 kb with a minimum methylation difference of 10% and a Fisher's multiple comparison *P*<0.01 were selected. Functional analysis of differentially methylated regions was performed by means of missMethyl R package [25] and Web-Gestalt [26]. Genomic region enrichment with respect to the EPIC annotation manifest for the epismature probes was calculated by means of Fisher's exact test; for the region-level analysis of DMRs, we took advantage of bedTools (v.2.30) [27] Fisher command, considering DMRs showing overlapping fraction >50% with the genomic regions annotated as 5'UTR, TSS1500, TSS200, 1stExon, Body, 3'UTR regions. DMRs comparison to the 127 reference epigenomes from the NIH Roadmap Epigenomics Consortium was carried out by means of GIGGLE, a fast and highly scalable genomic interval searching strategy, and evaluated using the GIGGLE score, that combines the estimation of the enrichment for observed versus expected (odds ratio), and Fisher's two tailed tests *p*-value [28].

## Results

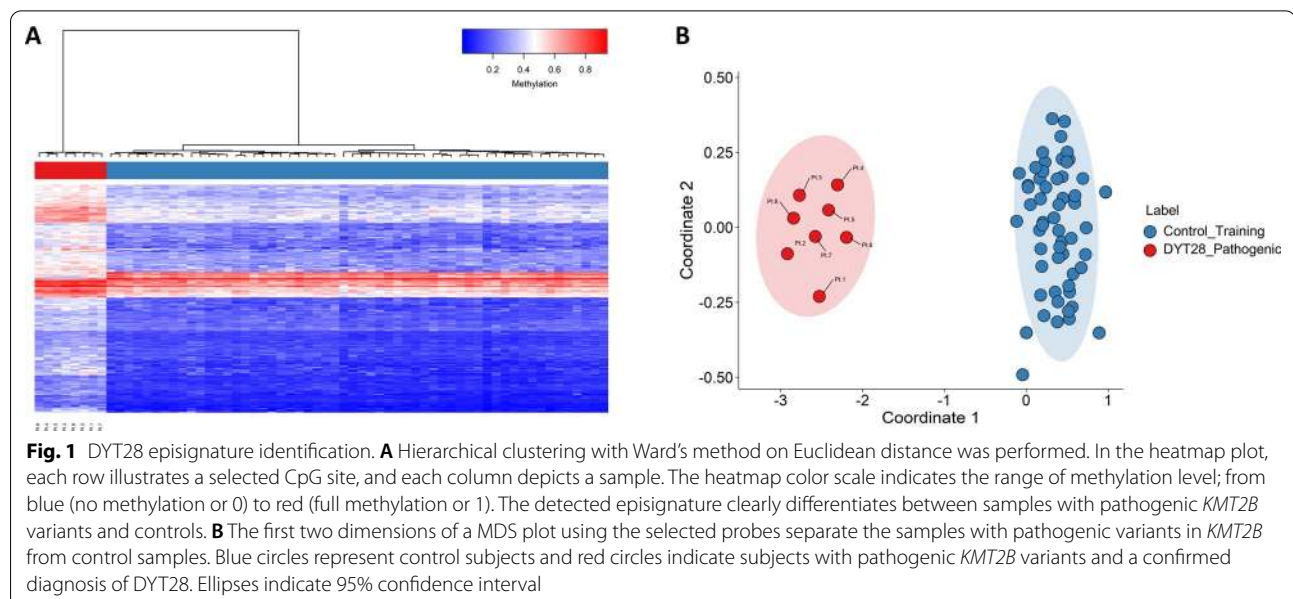
### DYT28 is associated with hypermethylated DNA episignature in blood

Eight individuals carrying *bona fide* pathogenic variants in *KMT2B* (NM\_014727.2), as per the ACMG/AMP guidelines [18], were included in the study. These variants can be considered as representative of the DYT28-causing *KMT2B* variants, as they included frameshift changes (Pt. 1–4 and 7), in-frame deletions (Pt. 6), and missense substitutions (Pt. 5 and 8) (Additional file 9: Table S1). Variants in Pt. 1, 2, 5–8 had been identified by WES in a cohort of 65 patients with genetically unclassified childhood-onset dystonia [3, 5, 6, 9]. The two frameshift variants in Pt. 3 and 4, both resulting in premature termination, were identified in the frame of diagnostic genetic testing. The clinical and molecular characterization of the DYT28 cohort is reported in Additional file 9: Table S1. Comparison of DNA methylation patterns between the peripheral blood DNA specimens of these 8 samples and 56 controls selected from our database based on matching for age and sex resulted in identification of 196 DMPs (methylation difference >10%, FDR <0.01, adjusted for blood cell-type compositions). Notably, more than 96% of probes (189 out of 196) exhibited relative hypermethylation (Additional file 11: Table S3). Of note, while the most robust and significant methylation change described in this episignature classifier involved hypermethylated regions in DYT28 (Additional file 2: Figure S2), the majority of the probes in these samples were found as slightly hypomethylated compared to the control group (Additional file 3: Figure S3). In order to assess the robustness of the episignature in differentiating

between case and control samples, hierarchical clustering (Fig. 1A) and MDS analysis (Fig. 1B) were performed, resulting in clear separation between groups. Eight rounds of cross-validation on MDS plot were performed using different combinations of samples with pathogenic *KMT2B* variants ( $n=7$ ) as training set and single samples with pathogenic variants as testing set. In all steps, the testing samples were correctly clustered with the training samples further providing evidence of a robust common DNA methylation signature (Additional file 4: Figure S4). While the two DYT28-associated missense variants identified in Pt. 5 and Pt.8 satisfied the ACMG/AMP criteria as *bona fide* pathogenic variants [18], the same analysis was also performed excluding those patients (Additional file 5: Figure S5). Comparison of the DNA methylation patterns of these six samples and the same 56 Control-Training samples confirmed the previous results.

### The DYT28-specific episignature allows functional classification of *KMT2B* variants

The generated DYT28-specific episignature was used to test 10 samples with unclassified *KMT2B* variants that had been identified in diagnostic genetic testing (Additional file 9: Table S1). Among these variants, two missense substitutions, for which the inheritance pattern could not be established, had been found in two subjects with childhood-onset dystonia fitting DYT28 (Pt. 9 and 10). Two missense variants, both inherited from an apparently unaffected parent, had previously been reported in subjects with DOPA-responsive cervical and mandibular dystonia (Pt. 14) and slowly progressive dystonia and dyskinesias with caudo-rostral progression (Pt.

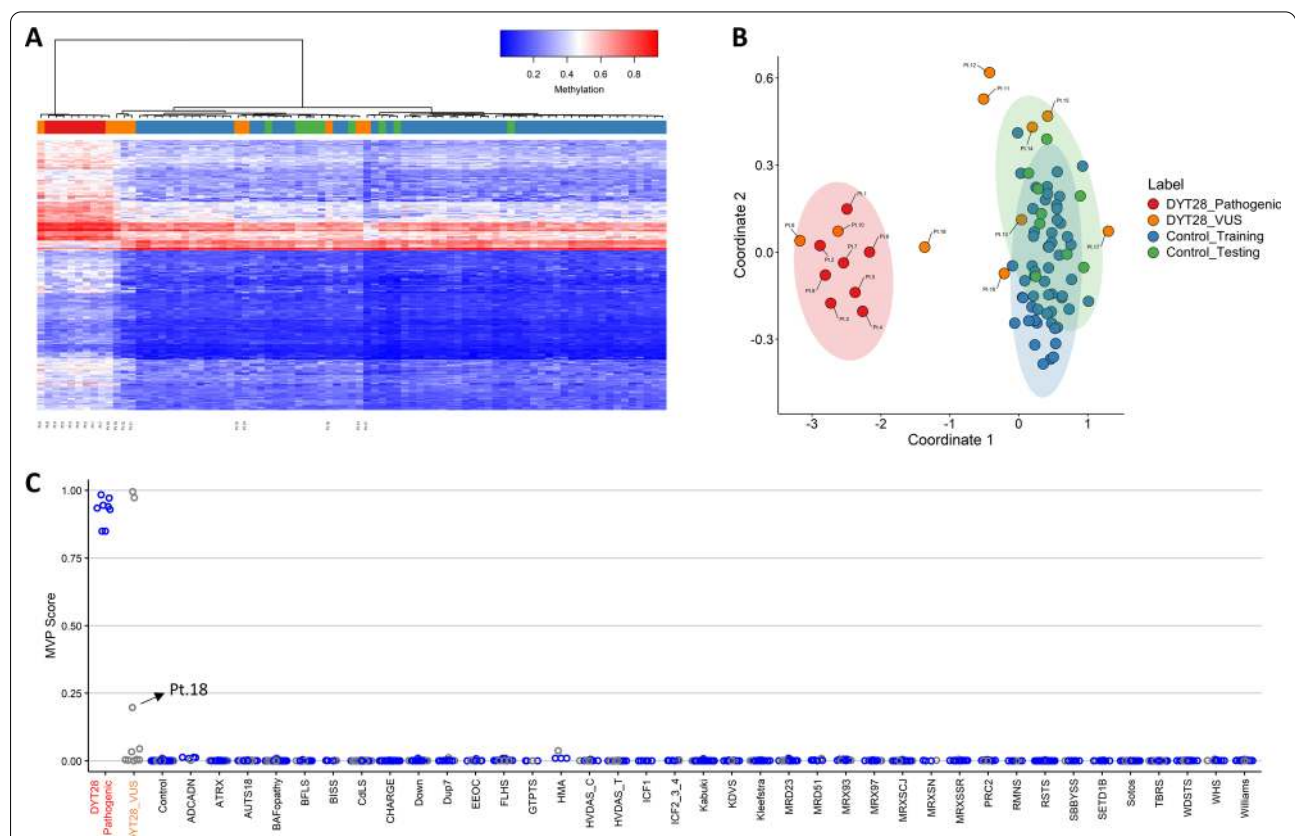


18) without decisive evidence for clinical relevance [3, 9]. Patient 13, who was characterized by progressive generalized dystonia with caudo-cranial progression, presented with an in-frame duplication previously reported as VUS in ClinVar (VCV000808518.4), and annotated in gnomAD v.2.1 ( $MAF=1.18 \times 10^{-3}$ ). The only other variant that had previously been reported in the general population was p.Ser2390Leu identified in patient 15 ( $MAF=4.02 \times 10^{-5}$ , gnomAD), who showed gait instability and dystonia of the upper and lower limbs during paroxysmal attacks. Finally, we selected four additional private missense variants predicted as pathogenic by CADD [29] and MetaDome [30] algorithms (Additional file 9: Table S1) in patients showing late-onset dystonia (Pt. 11 and 12) and congenital movement disorders (Pt. 16 and 17). The de novo occurrence of these variants could be ascertained only in two cases (Pt. 16 and 17).

The genome-wide methylation data obtained from blood DNA specimens of the 10 patients with heterozygous *KMT2B* VUS (Pt. 9 to 18) were analyzed by hierarchical clustering and MDS analysis using the 196

informative probes defining the DYT28-specific epigenature (Fig. 2A, B). Two samples of the testing set (Pt. 9 and 10) clustered with the DYT28 cohort, seven grouped with controls and one sample (Pt. 18) showed an intermediate position.

To test the use of the epigenature in a clinical setting, the recently developed SVM-based classifier was used [13, 17], trained by comparing the eight samples with *bona fide* pathogenic variants (Pt. 1–8) against the 10 VUS samples (Pt. 9–18), controls and a large set of individuals affected by various forms of NDDs or other RDs included in the EKD. All patients from other NDDs/RDs and controls were classified with low probability scores, indicating the high level of specificity of the DYT28 classifier (Fig. 2C). Consistent with the MDS and hierarchical clustering analyses, two missense *KMT2B* VUS, p.Ser1615Leu and p.Arg1777Pro (Pt. 9 and 10), were classified as disease-causing, while all other variants were scored with a significantly lower score, ruling out their clinical relevance in DYT28. Interestingly, the DYT28-related missense variants did not show a preferred



**Fig. 2** DNA methylation pattern analysis yields full sensitivity and specificity in classifying individuals affected by DYT28. Hierarchical clustering analysis (A) and MDS (B) plots are used to classify VUS (used as the testing set) with respect to pathogenic *KMT2B* variants and control samples (used as the training set). C A support vector machine (SVM) was used to classify samples and calculate probability scores reaches full sensitivity and specificity for *KMT2B* variants. The classifier was trained using pathogenic variants, controls and other NDDs/RDs. 75% of controls and NDD/RD samples used for training (blue), 25% for testing (grey). Ellipses indicate 95% confidence interval

localization in putative constrained regions and functional domains along the protein sequence. Nevertheless, we noted the existence of an apparent mutational hotspot for the four pathogenic missense variants between codons 1,615 and 1,777 (Fig. 3), similarly to what previously reported in the literature [3]. The clinical re-evaluation of patients 11 to 18 documented features or a natural history of disease that did not fit the classical presentation of DYT28, clinically validating the conclusions based on the episignature analysis (Additional file 9: Table S1).

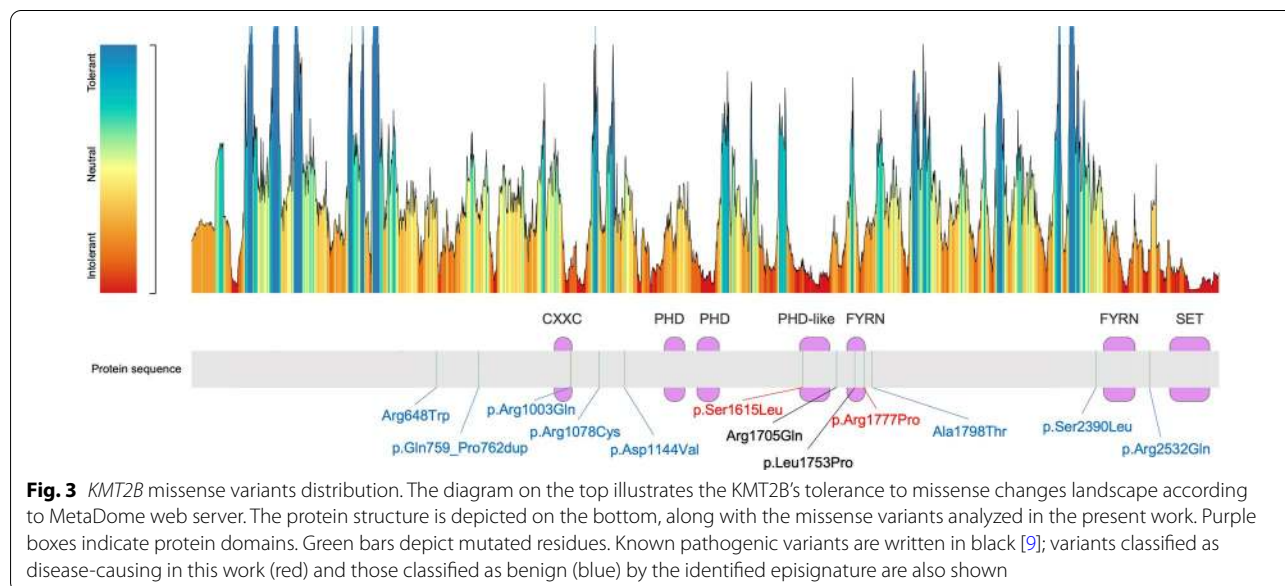
Of note, the same validation analysis was performed considering the episignature generated excluding Pt.5 and Pt.8 (Additional file 6: Figure S6). As shown, hierarchical clustering and MVP classifier properly classified Pt.5, Pt.8, Pt.9 and Pt.10 within the DYT28 group, further documenting the robustness of the approach. An apparently lower specificity, however, was provided by MDS analysis, confirming the dependency of the signature classification efficacy/specificity on sample size.

### DYT28-causing *KMT2B* variants are associated with a non-random distribution of hypermethylation in the genome

H3K4 trimethylation is highly enriched at active promoters near TSS and is positively correlated with transcription [10, 11, 31, 32]. Since this non-random distribution of H3K4me3 throughout the genome, we assessed the genome-wide distribution of DMRs (defined as stretches harboring  $\geq 5$  consecutive CpGs) in DYT28. The analysis allowed us to identify significant methylation changes in 146 genomic regions (hg19 genome assembly), the vast majority represented by relative hypermethylation in Pt.1

to Pt.8 (144 out of 146; Additional file 12: Table S4, Additional file 7: Figure S7). Considering all probes contained in these 146 regions, we first assessed the overall fit of the methylation patterns characterizing the 10 samples with *KMT2B* VUS with those obtained for the DYT28 and control groups (Additional file 8: Figure S8). As shown, the methylation levels for Pt.9 and Pt.10 were more similar to those characterizing patients carrying pathogenic *KMT2B* variants (93 and 89% of probes, respectively), while Pt.11–18 clearly diverge from the DYT28-specific pattern for a significant proportion of probes (from 55 to 12% of probes with methylation levels similar to DYT28). These data further validate the ability of the identified episignature in functionally classifying the tested *KMT2B* VUS.

Aiming to functionally characterize the genomic methylation differences in DYT28, we performed gene set enrichment analyses considering the genes mapping within DMRs by means of missMethyl [25] and Web-Gestalt [26] tools, failing to identify any enrichment for specific biological pathways (MSigDB's hallmark, Gene Ontology, KEGG, Reactome, Panther pathways), including those functionally linked to neurodevelopment and neuronal physiology (data not shown). To further gain insights on the functional implications of the observed hypermethylated status associated with defective *KMT2B* function, we then explored the hypothesis of a non-random distribution of DNA hypermethylation in genomic regions typically enriched in H3K4me3, such as promoters. First, we noted that DMPs constituting the episignature were enriched for genomic regions that are generally poorly methylated in actively transcribed genes (*e.g.*, gene



promoters [TSS1500] and first exon regions [1stExon]) (Fig. 4A). Consistently, gene body regions, which are known to be positively correlated with gene expression when methylated, were significantly underrepresented (Fig. 4A). To further delve into these results, all probes located within the annotated gene regions encompassing the identified DMRs were tested for enrichment analysis (Fig. 4B, Additional file 12: Table S4). Consistently, we observed a qualitatively similar pattern of enrichment involving promoter regions coupled with a depletion of gene body regions (Fig. 4B).

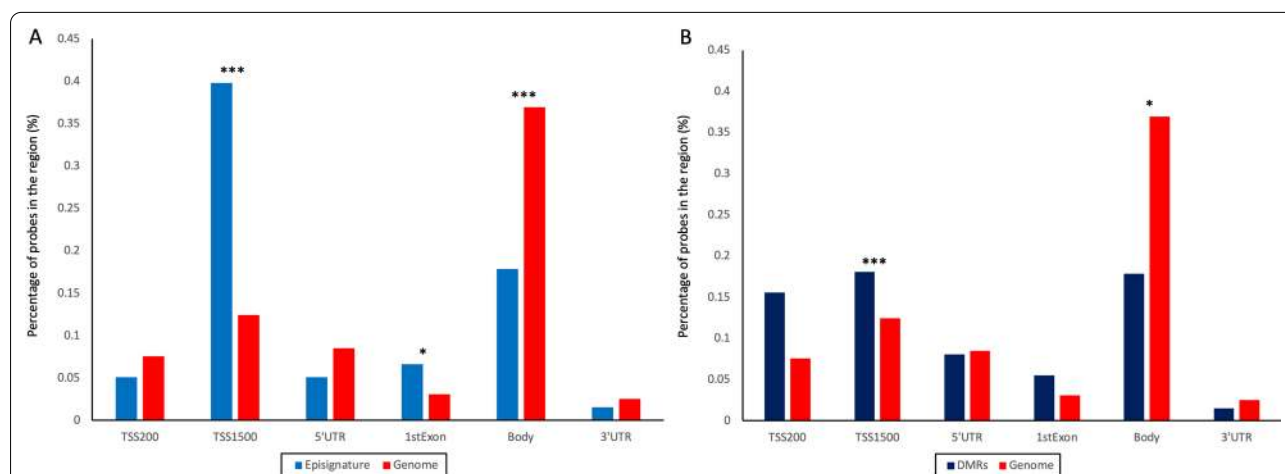
To further support the correlation between DNA methylation levels in DYT28 and the H3K4me3 epigenetic modification, we assessed the overlap of DMRs against the 127 reference epigenomes representing all major cell lineages in human body generated by the NIH Roadmap Epigenomics Consortium [12]. This analysis allowed us to inspect the enrichment of DMRs for regions specifically marked by H3K4me3 (*i.e.*, active TSSs) and the other 15 associated chromatin states as defined by the consortium (Fig. 5). The analysis documented a clear enrichment for the chromatin states associated to H3K4me3 marks (*i.e.*, active TSS and flanking active TSS) (Pairwise Wilcoxon Test's adjusted  $p$ -value  $< 2.1 \cdot 10^{-7}$ ), while regions in quiescent state were consistently underrepresented (Fig. 5).

## Discussion

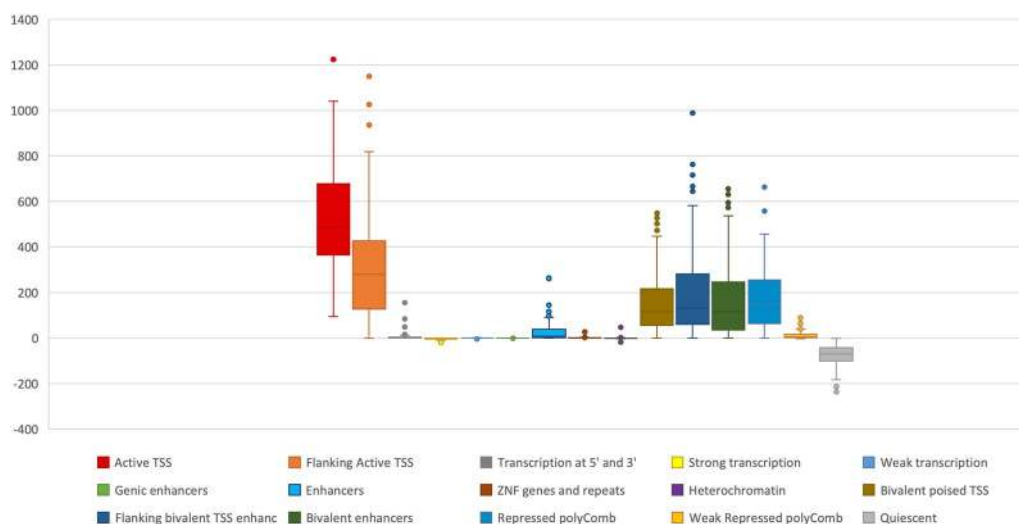
Here we report that DYT28 is associated with a genome-wide hypermethylated profile, showing that the identified methylation episignature can be successfully used to diagnose this disorder. DYT28 is a recently identified

form of dystonia caused by heterozygous inactivating *KMT2B* variants. The disorder is generally characterized by initial lower limb involvement during childhood followed by a more general involvement, usually including the bulbar and cranio-cervical muscles leading to dysarthria and dysphonia [3, 6]. Of note, the clinical phenotype associated with *KMT2B* variants is emerging as variable and has been reported to include conditions not manifesting dystonia at all [6]. Given the variable phenotypic presentation, lack of validated biomarkers and scarcity of experimental assays to promptly and straightforwardly assess *KMT2B* function, current interpreting of the clinical relevance of *KMT2B* variants can be challenging. In line with these considerations, bioinformatic strategies recently proposed showed lack of specificity [6, 33]. Although protein truncating variants and whole-gene deletions classification under a haploinsufficiency paradigm could be straightforward, caution is needed for missense changes and in-frame indels in a clinical setting. Indeed, even using combination of state-of-the-art in silico prediction algorithms (*e.g.*, CADD [29]), regional constraint analyses (MetaDome [30]), as well as segregation and population frequency data, did not help to identify *bona fide* pathogenic variants among the subset of *KMT2B* missense variants included in this study.

The identified episignature in peripheral blood DNA of DYT28 patients is defined by a relatively small number of hypermethylated CpG sites ( $< 200$ ) and proved to be robust and effective in classifying DYT28 patients, with respect to other partially overlapping dystonic phenotypes, other NDDs/RDs, as well as healthy controls. By



**Fig. 4** DYT28 is characterized by a hypermethylation pattern on specific gene regions. Histograms illustrate the non-random gene region distribution for episignature's probes (**A**), and DMRs (**B**) in DYT28 patients. Fisher's exact test was used to report the statistical significance of the enrichment/depletion with respect to genomic background (EPIC array). TSS200, transcription start sites 1–200; TSS, transcription start sites 201–1500; IGR, intergenic regions. **A** Percentage and statistical significance of feature enrichment for DYT28 episignature: \* $P < 0.01$ ; \*\* $P < 10^{-3}$  \*\*\* $P < 10^{-6}$ ; **B** Percentage and statistical significance of feature enrichment for DMRs in DYT28 patients: \* $P < 0.05$ ; \*\* $P < 0.01$  \*\*\* $P < 0.001$



**Fig. 5** DMRs enrichment for 15 chromatin states in 127 reference epigenomes in DYT28. Boxplots display enrichment scores (GIGGLE combo scores) distribution for 15 chromatin states as defined by the NIH Roadmap Epigenomics project. The active states (associated with expressed genes) consist of active transcription start site (TSS) proximal promoter states (Active TSS and Flanking Active TSS), a transcribed state at the 5' and 3' end of genes showing both promoter and enhancer signatures (Transcription at 5' and 3'), actively transcribed states (Strong Transcription and Weak Transcription), enhancer states (Enhancers and Genic Enhancers) and a state associated with zinc finger protein genes (ZNF genes and repeats). The inactive states consist of constitutive heterochromatin, bivalent regulatory states (Bivalent poised TSS, Flanking bivalent TSS enhancers and Bivalent enhancers), repressed PolyComb states (Repressed PolyComb and Weak Repressed PolyComb) and a quiescent state

using this epismature, we correctly classified missense changes not residing in constrained regions (Pt. 8) [30, 34], and excluded pathogenicity in cases with partially overlapping phenotype, or heterozygous for low-frequency alleles predicted as deleterious by in silico algorithms with unknown segregation (Pt. 15) [29, 30, 35], or documented to occur as de novo events (Pt. 16 and 17). Of note, our classification algorithm rejected a diagnosis of DYT28 in patient 18, who carried the p.Arg1003Gln variant in *KMT2B*, previously reported as VUS and predicted to have a destabilizing structural effect by homology modeling [9]. We would note, however, that while a confirmation of a related epismature in a patient with a genetic VUS is considered strong functional evidence for pathogenicity under the PS3/BS3 criterion [13, 17], using the ACMG/AMP sequence variant interpretation framework [18, 36], a negative result in a patient with or without a known genetic variant, while indicative, does not rule out pathogenicity. Although majority of the genes with epismatures currently have a single common epismature mapped, genes with multiple epismatures have been described, and assessment of pathogenicity of variants outside the established reference range warrants caution [17].

The present epigenomic analysis also provides relevant insights on the molecular effects of DYT28-causing *KMT2B* variants. A significant genome-wide shift in DNA methylation was documented: this hypermethylated

status was not associated with a specific enrichment of biological pathways but was characterized by a significant overrepresentation of regulatory regions known to be inversely correlated with gene expression (*i.e.*, gene promoters, and first exons) [11, 37, 38], and a depletion of gene-body regions, where a positive correlation between DNA methylation and active transcription has been described [11, 39]. This non-random hypermethylation pattern is opposite of the relative enrichment of nucleosomes with H3K4me3, which is prominent in TSS of actively transcribed loci [10–12, 40]. Based on the key role of *KMT2B* in H3K4me3, our findings provide new evidence that dysfunctional *KMT2B* in DYT28 causes specific DNA hypermethylation of promoters and other regulatory regions positively controlling gene expression, which collectively points out toward a general repression of transcriptional activity in DYT28. This finding has potentially relevant implications in terms of therapy, as it points to the use of nonspecific DNA methyltransferase inhibitors (*e.g.*, 5-aza-2'-deoxycytidine) or small molecules specifically targeting KDM5 demethylases (*e.g.*, CPI-455), as a potential approach to restore proper DNA methylation levels in DYT28 [41].

## Conclusions

In summary, we demonstrate evidence of a distinct DNA methylation epismature associated with *KMT2B*-related DYT28, enabling accurate diagnosis and reclassification



of ambiguous genetic findings. We also provide insights into the molecular pathophysiology of this disorder, documenting that *KMT2B* haploinsufficiency causes specific DNA hypermethylation of promoters and other regulatory regions positively controlling gene expression, pointing to the use of DNA methyltransferase inhibitors or molecules targeting KDM5 demethylases as potential therapeutic approaches in DYT28.

### Abbreviations

ACMG/AMP: American College of Medical Genetics and Genomics/Association for Molecular Pathology; CNS: Central nervous system; DOPA: 3,4-Dihydroxyphenylalanine; DMP: Differentially methylated probe; DMR: Differentially methylated region; DNAm: DNA methylation; DYT28: Dystonia 28; EKD: EpiSign knowledge database; FDR: False discovery rate; MDS: Multidimensional scaling; NDD: Neurodevelopmental disorders; PCA: Principal component analysis; RD: Rare diseases; VUS: Variant(s) of uncertain significance.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-021-01145-y>.

**Additional file 1: Figure S1.** Chromatograms showing the *KMT2B* variants identified in the 18 patients included in the study.

**Additional file 2: Figure S2.** Volcano plot of differences in the methylation status of individual probes between patients carrying pathogenic *KMT2B* variants and controls versus statistical significance ( $-\log p$ -value) of individual probes. Red dots represent selected, significant differentially methylated probes (DMPs) in Pt. 1-8. Positive and negative mean methylation difference show hypermethylation and hypomethylation, respectively.

**Additional file 3: Figure S3.** Mean methylation difference between patients carrying pathogenic *KMT2B* variants and control samples versus individual probes.

**Additional file 4: Figure S4.** Leave-1-out cross validation carried out by means of MDS plots based on the epismature analysis. For each round of validation, seven of the eight samples with bona fide pathogenic *KMT2B* variants were used for probe selection along with control samples and the one remaining was saved for testing. MDS was used to cluster the samples. Each time, the testing sample clustered with the other *KMT2B* mutated samples.

**Additional file 5: Figure S5.** DYT28 epismature identification excluding samples with missense *KMT2B* variants. (A) Hierarchical clustering with Ward's method on Euclidean distance was performed. In the heatmap plot, each row illustrates a selected CpG site, and each column depicts a sample. The heatmap color scale indicates the range of methylation level; from blue (no methylation or 0) to red (full methylation or 1). The detected epismature clearly differentiates between samples with pathogenic *KMT2B* variants and controls. (B) The first two dimensions of a MDS plot using the selected probes separate the samples with pathogenic variants in *KMT2B* from control samples. Blue circles represent control subjects and red circles indicate subjects with pathogenic *KMT2B* variants and a confirmed diagnosis of DYT28. Ellipses indicate 95% confidence interval.

**Additional file 6: Figure S6.** DNA methylation pattern analysis excluding samples with missense *KMT2B* variants yields full sensitivity and specificity in classifying individuals affected by DYT28. Hierarchical clustering analysis (A) and MDS (B) plots are used to classify VUS/and likely pathogenic missense variants (used as the testing set) with respect to pathogenic *KMT2B* variants in Pt.1-4,6,7 and control samples (used as the training set). (C) A support vector machine (SVM) was used to classify samples and calculate probability scores reaching full sensitivity and specificity for identifying pathogenic *KMT2B* variants. The classifier was trained using bona fide

pathogenic *KMT2B* variants, controls and other NDDs/RDs. 75% of controls and NDD/RD samples used for training (blue), 25% for testing (grey). Ellipses indicate 95% confidence intervals.

**Additional file 7: Figure S7.** Differentially methylated regions (DMRs) in DYT28. For each significant differentially methylated genomic region, the plot displays methylation levels calculated for 8 pathogenic variants (Pt. 1-8), used to define the epismature, versus 56 control samples. Mean difference, along with statistical significance according to Fisher's and Stouffer's methods, are reported for each region.

**Additional file 8: Figure S8.** DMRs methylation levels distribution throughout different genomic regions. Histograms show the DNA methylation levels (as beta values) for different genomic regions (TSS1500, TSS200, 5'UTR, 1stExon, Body, 3'UTR, IGR) in all the probes contained in DYT28's DMRs. Blue track displays median beta values in DYT28 (Pt.1-8); red track, median values for controls; black tracks, subjects with pathogenic *KMT2B* variants; grey tracks, individuals with *KMT2B* VUS not related to DYT28. The numbers on the right show the percentage of CpG probes having a beta value more similar to the median in DYT28 patients compared to controls, calculated for each patient  $i$  as  $ABS(B_i - control\ median) - ABS(B_i - DYT28\ median)$ , where  $ABS$  is the absolute value of the beta difference for the inspected probe. "TSS" indicates transcription start sites, "IGR" indicates intergenic regions, "Body" indicates genomic regions encompassing gene bodies.

**Additional file 9: Table S1.** Clinical characterization of the study cohort.

**Additional file 10: Table S2.** Sex and age of the patient and control groups used in the epismature discovery analysis.

**Additional file 11: Table S3.** Probes defining the methylation epismature associated with DYT28-causing *KMT2B* variants.

**Additional file 12: Table S4.** Regions showing differential methylation in DYT28.

### Acknowledgements

The authors wish to thank Claudia Nardini (Ospedale Pediatrico Bambino Gesù, Rome) for her skillful technical assistance, and acknowledge the "Cell line and DNA bank of genetic movement disorders and mitochondrial diseases" of the Telethon Biobanks Network.

### Authors' contributions

A.C. and A.F. contributed to the conception of the project, performed the bioinformatics and functional analyses and contributed to writing the manuscript. A.Ca. contributed to sample collection and coordinated the clinical data collection and interpretation. L.P., S.P. and E.M. performed the DNA methylation microarray experiments and contribute to data analyses. L.T., M.A., F.I., C.R., C.P., M.I., M.N., L.S.-R., E.A. and B.G. contributed to sample collection and processing, performed the genetic analyses and assisted with data interpretation. R.H.G., M.F.B., P.P., M.G., S.G., V.L., P.S., R.J.O. and G.S.Z. recruited the patients, contributed to the clinical data collection and interpretation. M.T. and B.S. conceived the project, supervised all aspects of the work and wrote the manuscript. All authors read and approved the final manuscript.

### Funding

This work was supported, in part, by Fondazione Bambino Gesù (Vite Coraggiose), Italian Ministry of Health (CCR-2017-23669081, RCR-2020-23670068\_001 and Ricerca Corrente) and Italian Ministry of Research (FOE 2019, 2020 - Sviluppo di protocolli innovativi e applicazione di nuovi strumenti -omici nei pazienti orfani di diagnosi) to M.T., Italian Ministry of Health (Ricerca Corrente) to A.C., and London Health Sciences Molecular Diagnostics Development Fund and Genome Canada Genomic Applications Partnership Program Grant (Beyond Genomics: Assessing the Improvement in Diagnosis of Rare Diseases using Clinical Epigenomics in Canada, EpiSign-CAN) awarded to B.S.

### Availability of data and materials

Clinical data of the study cohort, probes defining the methylation epismature associated with DYT28-causing *KMT2B* variants and list of regions differentially methylated in DYT28 are reported in additional files 8-11. Additional data are available from the corresponding authors upon request.

## Declarations

### Ethics approval and consent to participate

This study was approved by the Ethical Committee of the Ospedale Pediatrico Bambino Gesù (1702 OPBG 2018), and by the Western University Research Ethics Board (REB 106302). DNA specimens from the subjects included in this study were collected following procedures in accordance with the ethical standards of the declaration of Helsinki protocols and approved by the Review Boards of all involved institutions, with signed informed consents from the participating subjects/families.

### Consent for publication

Not Applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, IRCCS, 00146 Rome, Italy. <sup>2</sup>Department of Pathology and Laboratory Medicine, Western University, London, ON N6A 3K7, Canada. <sup>3</sup>Verspeeten Clinical Genome Centre, London Health Sciences Centre, London, Canada. <sup>4</sup>Department of Neuroscience, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy. <sup>5</sup>Department of Pediatric Onco-Hematology and Cell and Gene Therapy, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy. <sup>6</sup>Medical Genetics and Neurogenetics Unit, Fondazione IRCCS Istituto Neurologico C. Besta, Milano, Italy. <sup>7</sup>Medical Genetics Laboratory, ASST Papa Giovanni XXIII, Bergamo, Italy. <sup>8</sup>Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA. <sup>9</sup>Translational Cytogenomics Research Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy. <sup>10</sup>Medical Genetic Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. <sup>11</sup>Maternal-Infantile Department, University Hospital of Perugia, Perugia, Italy. <sup>12</sup>Department of Neuroscience, NESMOS, Sapienza University, Sant'Andrea Hospital, Rome, Italy. <sup>13</sup>Department of Human Neuroscience, Child Neurology and Psychiatry, Sapienza University, Rome, Italy. <sup>14</sup>Department of Neurology, Fondazione IRCCS Istituto Neurologico C. Besta, Milano, Italy. <sup>15</sup>Department of Child Neurology, Fondazione IRCCS Istituto Neurologico C. Besta, Milano, Italy. <sup>16</sup>Molecular Diagnostics Division, London Health Sciences Centre, London, Canada.

Received: 3 June 2021 Accepted: 31 July 2021

Published online: 11 August 2021

## References

- Balint B, Mencacci NE, Valente EM, et al. Dystonia. *Nat Rev Dis Primer*. 2018;4(1):25.
- Lohmann K, Klein C. Update on the genetics of dystonia. *Curr Neurol Neurosci Rep*. 2017;17(3):26.
- Zech M, Lam DD, Winkelmann J. Update on KMT2B-related dystonia. *Curr Neurol Neurosci Rep*. 2019;19(11):92.
- 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*. 2016;99(6):1377–87.
- Meyer E, Carss KJ, Rankin J, et al. Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia. *Nat Genet*. 2017;49(2):223–37.
- Cif L, Demailly D, Lin J-P, et al. KMT2B-related disorders: expansion of the phenotypic spectrum and long-term efficacy of deep brain stimulation. *Brain*. 2020;143(11):3242–61.
- Barbagiovanni G, Germain P-L, Zech M, et al. KMT2B is selectively required for neuronal transdifferentiation, and its loss exposes dystonia candidate genes. *Cell Rep*. 2018;25(4):988–1001.
- Vallianatos CN, Iwase S. Disrupted intricacy of histone H3K4 methylation in neurodevelopmental disorders. *Epigenomics*. 2015;7(3):503–19.
- Carecchio M, Invernizzi F, González-Latapi P, et al. Frequency and phenotypic spectrum of KMT2B dystonia in childhood: a single-center cohort study. *Mov Disord*. 2019;34(10):1516–27.
- Denisov S, Hofmeister H, Marks H, et al. MLL2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas MLL1 is redundant. *Development*. 2014;141(3):526–37.
- Black JC, Van Rechem C, Whetstone JR. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell*. 2012;48(4):491–507.
- Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317–330.
- 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*. 2020;106(3):356–70.
- Radio FC, Pang K, Ciolfi A, et al. SPEN haploinsufficiency causes a neurodevelopmental disorder overlapping proximal 1p36 deletion syndrome with an epismutation of X chromosomes in females. *Am J Hum Genet*. 2021;108(3):502–5166.
- Ciolfi A, Aref-Eshghi E, Pizzi S, et al. Frameshift mutations at the C-terminus of HIST1H1E result in a specific DNA hypomethylation signature. *Clin Epigenet*. 2020;12(1):7.
- Aref-Eshghi E, Bend EG, Colaiacovo S, et al. Diagnostic utility of genome-wide DNA methylation testing in genetically unsolved individuals with suspected hereditary conditions. *Am J Hum Genet*. 2019;104(4):685–700.
- Sadikovic B, Levi MA, Kerkhof J, et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. *Genet Med*. 2021; in press, PMID: 33547396.
- Richards S, Aziz N, Bale S, et al. 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. *Genet Med*. 2015;17(5):405–23.
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1988;16(3):1215.
- Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363–9.
- Pidsley R, Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genom*. 2013;14:293.
- Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47–e47.
- Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinf*. 2012;13:86.
- Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. *Epigenet Chromatin*. 2015;8(1):6.
- Maksimovic J, Oshlack A, Phipson B. Gene set enrichment analysis for genome-wide DNA methylation data. *Bioinformatics*. 2020. <https://doi.org/10.1101/2020.08.24.265702>.
- Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res*. 2019;47(W1):W199–205.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma Oxf Engl*. 2010;26(6):841–2.
- Layer RM, Pedersen BS, DiSera T, Marth GT, Gertz J, Quinlan AR. GIGGLE: a search engine for large-scale integrated genome analysis. *Nat Methods*. 2018;15(2):123–6.
- Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310–5.
- Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. MetaDome: pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum Mutat*. 2019;40(8):1030–8.
- Demers C, Chaturvedi C-P, Ranish JA, et al. Activator-mediated recruitment of the MLL2 methyltransferase complex to the  $\beta$ -globin locus. *Mol Cell*. 2007;27(4):573–84.
- Heintzman ND, Stuart RK, Hon G, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 2007;39(3):311–8.

33. Zech M, Jech R, Havránková P, et al. *KMT2B* rare missense variants in generalized dystonia: *KMT2B* Missense Variants in Dystonia. *Mov Disord*. 2017;32(7):1087–91.
34. Havrilla JM, Pedersen BS, Layer RM, Quinlan AR. A map of constrained coding regions in the human genome. *Nat Genet*. 2019;51(1):88–95.
35. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–43.
36. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence P53/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med*. 2019;12(1):3.
37. Straussman R, Nejman D, Roberts D, et al. Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol*. 2009;16(5):564–71.
38. Brenet F, Moh M, Funk P, et al. DNA Methylation of the First Exon Is Tightly Linked to Transcriptional Silencing. Papavasiliou N, ed. *PLoS ONE*. 2011;6(1):e14524.
39. Ball MP, Li JB, Gao Y, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol*. 2009;27(4):361–8.
40. Shen E, Shulha H, Weng Z, Akbarian S. Regulation of histone H3K4 methylation in brain development and disease. *Philos Trans R Soc B Biol Sci*. 2014;369(1652):20130514.
41. Leadem BR, Kagiampakis I, Wilson C, et al. A kdm5 inhibitor increases global H3K4 trimethylation occupancy and enhances the biological efficacy of 5-Aza-2'-deoxycytidine. *Cancer Res*. 2018;78(5):1127–39.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

