

EHMT2 LOSS-OF-FUNCTION ALTERATIONS CAUSE A KLEEFSTRA-LIKE SYNDROME

Beatriz Martinez-Delgado^{1,2,3}, Estrella Lopez-Martin^{1,2}, Jennifer Kerkhof⁴, Beatriz Baladron^{1,2}, Lidia M. Mielu^{1,2}, Diana Sanchez-Ponce¹, Ariadna Bada-Navarro¹, Marina Herrero-Matesanz¹, Lidia Lopez-Jimenez¹, Jessica Rzasz⁴, Dmitrijs Rots⁷, Marta Fernandez¹, Esther Hernandez-San Miguel^{1,2}, Gema Gomez-Mariano¹, Purificacion Marin-Reina⁵, Rosario Cazorla-Calleja^{6,2}, Javier Alonso^{1,2,3}, Tjitske Kleefstra⁷, Manuel Posada^{1,2,3}, Eva Bermejo-Sanchez^{1,2}, Bekim Sadikovic^{4,8} and Maria J. Barrero^{1,2,*}

1. Institute of Rare Diseases Research (IIER), Spanish National Institute of Health Carlos III (ISCIII), 28220 Madrid, Spain
2. Undiagnosed diseases program SpainUDP, Madrid, Spain
3. Centro de Investigación Biomedica en Red de Enfermedades Raras, (CIBERER)
4. Verspeeten Clinical Genome Centre; London Health Sciences Centre, London, ON N6A 5W9, Canada
5. University and Polytechnic Hospital La Fe, Valencia, Spain
6. Hospital Universitario Puerta de Hierro, Majadahonda, Madrid, Spain
7. Department of Human Genetics, Radboud University Medical Center, 6500 HB, Nijmegen, The Netherlands
8. Department of Pathology and Laboratory Medicine, Western University, London, ON N6A 3K7, Canada

*Corresponding author: mj.barrero@isciii.es

ABSTRACT

Dysregulation of the epigenetic machinery is associated with neurodevelopmental defects in humans. Kleefstra syndrome (KS) is a neurodevelopmental syndrome caused by heterozygous alterations in the gene *EHMT1* that cause loss-of-function. *EHMT1* and *EHMT2* are highly similar histone methyltransferases that play relevant roles in development. Despite their similarity, individuals with alterations in *EHMT2* have never been described. Here, we describe a pediatric patient with a KS-overlapping phenotype and a single base de novo substitution in *EHMT2* that causes the amino acid change p.Ala1077Ser in the catalytic SET domain. This change causes a reduction in the affinity of the catalytic domain for the H3 tail and in the activity of the enzyme by three- to five-fold. DNA methylation, histone methylation and gene expression profiles suggest a significant overlap between the *EHMT2* p.Ala1077Ser variant and KS. Based on this evidence we suggest that *EHMT2* haploinsufficiency causes a Kleefstra-like syndrome. Although we cannot rule out dominant negative effects caused by the *EHMT2* p.Ala1077Ser variant, our data and previously published data suggest that loss of *EHMT2* function is probably more detrimental to cells than loss of *EHMT1*, explaining why individuals with alterations in *EHMT2* are very rare.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

INTRODUCTION

Dysregulation of the epigenetic machinery is associated with neurodevelopmental defects in humans. Kleefstra syndrome (KS) is a neurodevelopmental syndrome caused by heterozygous deletions at chromosome 9q34.3 that include the *EHMT1* gene (~50%) or heterozygous intragenic *EHMT1* pathogenic variants (~50%)^{1,2}. In addition, loss-of-function mutations in the *KMT2C* (MLL3) gene cause a Kleefstra-like syndrome called Kleefstra syndrome-2 (KLEFS2)³⁻⁵. Moreover, although very rare, de novo mutations in epigenetic regulators genes *MBD5*, *MLL3*, *SMARCB1*, and *NR1I3* have been associated with intellectual disability disorders that fall into the Kleefstra syndrome's phenotypic spectrum³.

EHMT1 (also known as GLP) and EHMT2 (also known as G9a) are highly homologous proteins that contain a SET domain that confers them the ability to mono- and dimethylate lysine 9 at histone H3 (H3K9)^{6,7}. H3K9 methylation is finely regulated by several methyltransferases and demethylases and plays important roles in development, differentiation, and cancer⁸. EHMT1 and EHMT2 can form homo- or heterodimers through SET domain interactions, being heterodimers the most active configuration and the preferred form in cells⁹. Both *Ehmt1*^{-/-} and *Ehmt2*^{-/-} mice show severe developmental delays and early embryonic lethality that correlates with lower levels of H3k9me1 and me2^{9,10}.

Despite the overlap in domains and functions of EHMT1 and EHMT2 and their relevance for development, patients with alterations in EHMT2 have never been reported. Here we report a Kleefstra-like syndrome case with a missense variant in the catalytic SET domain of EHMT2 that causes alanine 1077 change to serine reducing its affinity for the H3 tail and the activity of the enzyme by three- to five-fold. Results from our study allow defining this variant as pathogenic following the ACMG criteria¹¹, since it appears de novo in a patient with disease and no family history (PS2), in vitro functional studies are supportive of a damaging effect (PS3), it is absent from controls in gnomAD (PM2) and the variant is located in a critical functional domain (PM1).

MATERIALS AND METHODS

Participants

Patients ND095 and ND120 were recruited by the undiagnosed rare diseases program SpainUDP¹² at the Institute of Rare Diseases Research (IIER), Spanish National Institute of Health Carlos III (ISCIII). Peripheral blood samples and skin biopsies were collected from patients and their parents to perform trio-based whole-exome sequencing and establishment of fibroblast cultures respectively. Informed consents were signed by patient's legal representatives. This research project was approved by the ISCIII Research Ethics Committee. Patient labels ND095 and ND120 correspond to an internal code of the SpainUDP program that is only available to the research team.

Whole Exome sequencing

Whole exome sequencing and data analysis were performed in the probands and their unaffected parents. Genomic DNA was extracted from peripheral blood using the Qiagen QIAamp DNA kit. Whole Exome Sequencing (WES) libraries were prepared using the Nimblegen MedExome + ChrMit as enrichment kit and HS2000 v4, 2x100bp sequencing platform in ND095 family and Nextera Flex DNA Library Prep and Illumina NextSeq500 in ND120 family. Data analysis was performed using two different standardized protocols as previously described¹². These included an in-house analysis and a parallel analysis using the Genomic Analysis module of the RD-Connect Genome-Phenome Analysis Platform (GPAP)¹³. All commonly identified rare variants were further analysed, checking all available scientific evidence through detailed searches in public databases (including Gene-Card, NCBI, UniProt, OMIM, Pubmed and ExAc). Sanger sequencing was performed to validate candidate variants.

Fibroblast culture

Human dermal fibroblasts cultures were established from skin tissue samples. Briefly, fibroblasts were mechanically isolated by dissecting the dermal layer of the skin and the resulting fragments were incubated at 37°C in Dulbeccos' modified Eagle's medium (DMEM) containing 2% of fetal calf serum. Cells were expanded by incubation in 75 cm² culture flasks at 37°C with 5% CO₂ and 95% humidity in DMEM containing 10% fetal bovine serum.

RNA-seq analysis

Total RNA was extracted from fibroblasts using the RNeasy mini kit from Qiagen. RNA-seq was performed at BGI Tech Solutions with two or three biological replicates per condition. Briefly, ribosomal RNAs were removed using a RNase H-based method and first-strand cDNA was generated using random hexamer-primed reverse transcription, followed by a second-strand cDNA synthesis with dUTP instead of dTTP, end repair, A addition and adaptor ligation. The U-labeled second-strand template was digested with Uracil-DNA-Glycosylase (UDG) and amplified by PCR. The resulting library was validated by quality control. The PCR products were then heat denatured and circularized by the splint oligo sequence to generate a single strand circle DNA followed by rolling circle replication to create DNA nanoballs (DNB) for sequencing on the MGI DNBSEQ platform. Raw sequencing data with adapter sequences or low-quality sequences was trimmed or filtered and examined by FastQC for basic quality controls. The sequencing analysis was carried out using Galaxy (<https://usegalaxy.eu/>). Paired reads were aligned to the human hg19 genome build using STAR¹⁴. Gene counts were calculated using HTseq-count¹⁵ and differential expression between patients and healthy fibroblast was interrogated using DESeq¹⁶. Differentially expressed genes were considered at an adjusted p-value <0.05.

For enrichment analysis in differentially expressed genes, we used GSEA preranked¹⁷ with genes ranked according to p-value corrected log₂ of fold change. Interrogated gene sets were Gene Ontology Biological Process (GOBP) and transcription factor

targets (TFT). In addition, we performed enrichment in GOBP terms in significantly upregulated or downregulated genes using Panther (<https://pantherdb.org/>)¹⁸. Scatterplots showing correlations were generated with ggplot2 in Galaxy. Pearson's correlation and bubble plots were calculated using SRPlot¹⁹. Network representation of enriched GOBP terms was generated using Cytoscape Enrichment Map and AutoAnotate Apps²⁰.

Production of recombinant proteins

pET28a-LIC containing the catalytic domain of human EHMT2 (aa913-1193) was a gift from Cheryl Arrowsmith (Addgene plasmid # 25503; <http://n2t.net/addgene:25503>; RRID:Addgene_25503). Mutation p.Ala1077Ser was introduced using the QuikChange II XL Site-Directed Mutagenesis Kit from Agilent. Expression and purification of His-tagged EHMT2 catalytic domain WT and p.Ala1077Ser in bacteria has been previously described²¹. Recombinant H3 was expressed in *E.coli* and purified as previously described²²

Histone methyltransferase assays

Histone methyltransferase assays were performed using the fluorescence-based assay SAM Methyltransferase Assay SAMfluoro™ from G-Biosciences that detects the formation of resorufin over time as a result of the methyltransferase reaction. Assays were performed using 250, 500 or 1000 ng of His-tagged WT or p.Ala1077Ser EHMT2 catalytic domains and 2.5 µg of recombinant histone H3 following the manufacturer instructions. Cumulative reads were measured every minute for 45 minutes. Specific activity was calculated using a resorufin standard.

Histone peptide binding assays

10 µg of biotinylated histone 3 peptides aa 1 to 21 (Active Motif) unmodified, monomethylated at K9 or dimethylated at K9 were incubated overnight at 4°C with 5 µg of recombinant EHMT2 wild type or p.Ala1077Ser mutant catalytic domain in 300 µl of binding buffer 50 mM Tris pH 7.5, 150 mM NaCl, 0.05% NP-40 and 1 mM PMSF. After 1 hour incubation with Pierce streptavidin magnetic beads and extensive washing with binding buffer bound proteins were resolved by SDS-PAGE and stained with Coomassie blue. Gels were scanned using the Bio-Rad GelDoc Go Imaging System and bands quantified using ImageJ²³.

Protein modeling

PDB Protein Data Bank entry 5JJ0 (<https://doi.org/10.2210/pdb5JJ0/pdb>) corresponding to the catalytic domain of EHMT2 complexed with the histone peptide H3K9M and SAM was explored in 3D using PDB resources²⁴.

EpiSign assay

Methylation analysis was conducted using the clinically validated EpiSign assay, following previously established methods^{25–28}. Methylated and unmethylated signal intensities generated from the EPIC array were imported into R 3.5.1 for normalization,

background correction, and filtering. Beta values were then calculated as a measure of methylation level, ranging from 0 (no methylation) to 1 (complete methylation), and processed through the established support vector machine (SVM) classification algorithm for EpiSign disorders. The classifier utilized the EpiSign Knowledge Database, which consists of over 10,000 methylation profiles from reference disorder-specific and unaffected control cohorts, to generate disorder-specific methylation variant pathogenicity (MVP) scores. These MVP scores are a measure of prediction confidence for each disorder and range from 0 (discordant) to 1 (highly concordant). A positive classification typically generates MVP scores greater than 0.5. The final matched EpiSign result is generated using these scores, along with the assessment of hierarchical clustering and multidimensional scaling.

Detection of histone modifications by mass spectrometry

Mass spectrometry was performed by Active Motif. Histones were acid extracted from a cell pellet containing 2.5×10^6 cells, derivatized via propionylation, digested with trypsin, newly formed N-termini were propionylated as previously described²⁹. Histones were extracted by incubating samples at room temperature for 1 hour in 0.2M sulfuric acid with intermittent vortexing. Histones were then precipitated by the addition of trichloroacetic acid (TCA) on ice, and recovered by centrifugation at 10,000 x g for 5 minutes at 4°C. The pellet was then washed once with 1mL cold acetone/0.1% HCl and twice with 100% acetone, and then air dried in a clean hood. The histones were propionylated by adding 1:3 v/v propionic anhydride/2-propanol and incrementally adding ammonium hydroxide to keep the pH around 8, and subsequently dried in a SpeedVac concentrator. The pellet was then resuspended in 100 mM ammonium bicarbonate and adjusted to pH 7-8 with ammonium hydroxide. The histones were then digested with trypsin, resuspended in 100mM ammonium bicarbonate overnight at 37°C, and dried in a SpeedVac concentrator. The pellet was resuspended in 100mM ammonium bicarbonate and propionylated a second time by adding 1:3 v/v propionic anhydride/2-propanol and incrementally adding ammonium hydroxide to keep the pH around 8, and subsequently dried in a SpeedVac concentrator. Histone peptides were resuspended in 50 µL of 0.1% TFA and 3 µL were injected with 3 technical replicates in a Thermo Scientific TSQ Quantum Ultra mass spectrometer coupled with an UltiMate 3000 Dionex nano-liquid chromatography system. The data was quantified using Skyline³⁰, and represents the percent of each modification within the total pool of that amino acid residue.

RESULTS

Clinical characteristics of the proband overlap with Kleefstra syndrome

Proband ND095 was remitted to the undiagnosed rare diseases program SpainUDP¹² after testing negative for candidate gene panels. The proband suffered from global developmental delay characterized by slow developing of gross motor milestones and expressive language delay, hypotonia without weakness, coarctation of the aorta, nephrocalcinosis and renal medullar cystic disease. At the craniofacial level, the

proband had brachicephaly, plagiocephaly with very flat occiput, broad face and midface hypoplasia, synophrys, sparse medial eyebrows, epicanthus, lateral deviation of upslanted palpebral fissures, mildly everted lower eyelids, palpebral ptosis, anteverted nares, smooth philtrum and carp-like mouth. In addition, the proband had dental diastema, microdontia, persistent fetal fingertip pads and scoliosis (a complete clinical history can be provided upon request). From the clinical point of view, the patient's phenotype overlaps significantly with the clinical characteristics of KS (Table 1).

Whole-exome sequencing reveals a VUS in the *EHMT2* gene

Whole-exome sequencing of patient ND095 and his parents carried out at SpainUDP revealed no detectable deleterious variants in the Kleefstra syndrome 1 (KS) causing gene *EHMT1*. In addition, no alterations were detected in *KMT2C*, which causes Kleefstra syndrome 2⁵. Similarly, no pathogenic variants were detected in chromatin-related genes *MBD5*, *MLL3*, *SMARCB1*, and *NR1H3* that have been previously associated with Kleefstra-like syndromes³. Surprisingly, the sequencing revealed a heterozygous de novo variant in the *EHMT2* gene (chr6:31848838C>A; NM_006709.5:c.3229G>T) that resulted in the amino acid change alanine 1077 to serine (p.Ala1077Ser). This variant was not present in gnomAD V4 genomes not exomes, neither reported in ClinVar. Several in silico predictors classified this variant as pathogenic (EIGEN, FATHMM, LRT, MutPred MVP, REVEL) while it was classified as variant of uncertain significance (VUS) by others (Mutation Taster, Mutation Assessor, SIFT, PROVEAN).

The p.Ala1077Ser variant decreases the activity of *EHMT2*

Ala1077 is located in the catalytic SET domain of *EHMT2*, in a highly conserved region among human histone methyltransferases (Figure 1A). Ala1077 is, however, far from the AdoMet cofactor interacting amino acids (Figure 1A) and the presumed dimerization region (Figure 1B). The published *EHMT2* catalytic domain structure in complex with the H3 tail³¹ shows that Ala1077 interacts with the H3 N-terminal tail residue threonine 6 suggesting that it might affect the catalytic activity of *EHMT2* through changes in this interaction (Figure 1C). To test this possibility, we first evaluated the effects of the p.Ala1077Ser variant in the activity of the enzyme. We used an in vitro fluorescence-based assay that allows the detection of methyltransferase activity over time to test the activity of wild type and Ala1077Ser *EHMT2* catalytic domains expressed and purified in *E.coli*, on recombinant histone H3. Figure 2A shows that the p.Ala1077Ser variant drastically reduces the methyltransferase activity of *EHMT2* on histone H3 and the specific activity of the enzyme by five-fold (Figure 2B). Replication of the histone methyltransferase assay using different *EHMT2* concentrations shows that the Ala1077Ser variant reduces the specific activity of the enzyme by three- to five-fold (Figure 2B, Supplementary Figure 1). Next, we evaluated if the p.Ala1077Ser variant affects the interaction of *EHMT2* with histone H3 tail by testing the interaction of wild type *EHMT2* and p.Ala1077Ser mutant with synthetic H3 peptides (aa 1-21) either unmethylated, mono or dimethylated at lysine 9, in an in vitro pull-down assay. Figure 2C and 2D show that the

p.Ala1077Ser variant reduces de affinity of the catalytic domain for histone H3 tail. In addition, we detected a reduced affinity of the WT catalytic domain for the dimethylated peptide compared to unmethylated or monomethylated peptides. These results are in line with previous data that suggests loss of affinity of methyltransferases for its substrate once it is fully methylated^{21,32}. This behavior was not observed in the mutant catalytic domain (Figure 2C and D).

The Episign test classifies patient ND095 as Kleefstra syndrome.

In order to investigate the overlap of patient ND095 with KS-specific DNA methylation signature, we compared the blood DNA methylation profile of patient ND095 with Kleefstra patients using version three of the test EpiSign²⁶. This test, based in blood DNA methylation profiles, allows the diagnosis of over fifty different genetic diseases. Unsupervised hierarchical clustering of DNA methylation profiles shows that ND095 clusters with Kleefstra syndrome patients and clearly separates from healthy controls (Figure 3A). The computational model classified patient ND095 as KS with high confidence, generating a low score for other 56 syndromes (Figure 3C). However, the principal component analysis of DNA methylation showed that ND095 had a particular methylation profile that differs slightly from KS patients (Figure 3B). This data suggests the involvement of an *EHMT1*-related but different causal gene in the phenotype of patient ND095.

Overlap of molecular signatures between patient ND095 and Kleefstra syndrome.

Next, we established fibroblasts cultures from a healthy donor, the ND095 patient carrying the *EHMT2* p.Ala1077Ser variant and a patient (ND120) recently diagnosed with KS through our program, who carries a frameshift variant in *EHMT1* (NM_024757.5:c.1881delT p.His620ThrfsTer12) and quantified histone modifications by mass spectrometry. Figure 4A shows that ND095 had decreased levels of mono-, di- and trimethylated H3K9 while levels of H3K27me3 remained unchanged. In correlation, we found an increase in unmodified and acetylated H3K9. Similar, but less dramatic effects, were found in KS patient ND120 (Figure 4A). These results are in agreement with a previous characterization of changes in histone modifications after knockdown for both *EHMT2* and *EHMT1*³³ and are compatible with the loss of *EHMT2* activity in patient ND095.

Afterwards, we carried out RNA-seq in the fibroblasts cultures and identified differentially expressed genes (DEGs) in ND095 or ND120 fibroblasts compared to healthy fibroblasts. Among DEGs we found both upregulated and downregulated genes with a higher number of DEGs in ND095 compared to ND120 (Figure 4B). This correlates with the more severe effects in H3K9 methylation in ND095 found by mass spectrometry. Despite this difference there was a significant overlap in down and upregulated genes between ND095 and ND120 (Figure 4B) and a significant correlation comparing the fold change of expression of all genes (Figure 4C).

We next investigated the enrichment of GOBP terms in the DEGs using gene set enrichment analysis (GSEA). Results show a significant enrichment of genes involved in

the morphogenesis of diverse tissues and organs derived from the three embryonic layers, and genes involved in the extracellular matrix in genes upregulated in ND095 compared to control fibroblasts (Figure 4D). Significant enrichment in morphogenesis genes in upregulated genes was also identified in patient ND120 (Supplementary Figure 2A) and in genes commonly upregulated in both patients (Supplementary Figure 2B). Genes downregulated in ND095 fibroblasts compared to healthy fibroblasts were mainly involved in cell cycle (Figure 4D). Additionally, a significant enrichment of E2F transcription factors binding sites was found around the transcriptional start site of genes downregulated in ND095 fibroblasts (Supplementary Figure 3A). Moreover, the expression of several E2F factors was significantly downregulated in ND095 fibroblasts (Supplementary Figure 3B). However, in ND120 we could not find enrichment of cell cycle related genes, nor genes bound by E2F transcription factors, in downregulated genes or downregulated E2F factors.

Changes in *EHTM1* and *EHMT2* expression in patients' fibroblasts

Interestingly, the expression of *EHMT2* was significantly downregulated in ND095 compared to healthy control fibroblasts suggesting that p.Ala1077Ser has consequences for *EHMT2* mRNA expression (Supplementary Figure 4A). Despite this, the reads covering the NM_006709.5:c.3229G>T region showed a balance of 50% WT and 50% mutant transcripts (Supplementary Figure 4B). In patient ND120, we detected lower mRNA expression of both *EHMT1* and *EHMT2* (Supplementary Figure 4A). Reads in the *EHMT1* NM_024757.5:c.1881delT region were predominantly WT suggesting that the T deletion affects the levels of RNA likely by nonsense-mediated decay (Supplementary Figure 4B).

DISCUSSION

We describe here a patient with a de novo pathogenic missense variant in the *EHMT2* gene that causes a Kleefstra-like syndrome phenotype. This missense variant causes the change of alanine 1077 to serine in the catalytic SET domain of *EHMT2* reducing its affinity for histone H3 tail and catalytic activity by three- to five-fold. Interestingly, the presence of p.Ala1077Ser change not only reduces the affinity of the catalytic domain for histone H3 but also abrogates the specificity of H3 recognition dependent on its modification. Loss of affinity of *EHMT2* catalytic domain for the H3 tail once is dimethylated might favor the recruitment of other methyltransferases able to trimethylate histone H3. This recruitment might be blocked by the low but persistent binding of the p.Ala1077Ser mutant domain to H3K9me2.

Although the changes in gene expression and H3K9 methylation detected in patients' fibroblasts are more dramatic in ND095 than ND120 there is a significant correlation in changes of gene expression between both patients. In both cases, there is an upregulation of genes involved in the morphogenesis of diverse organs and tissues, suggesting that *EHMT1/2* play a role in preventing the expression of genes belonging to alternative lineages. In accordance, *EHMT1/2* have been previously involved in the silencing of alternative lineage genes during hematopoietic differentiation^{34,35}.

Regarding downregulated genes, a significant enrichment in cell cycle genes and E2F targets was only observed in patient ND095. Similarly, the depletion of EHMT2 in myoblasts has been previously shown to affect cell cycle through downregulation of E2F target genes³⁶. This function has been suggested to be independent of EHMT2 methyltransferase activity and through its association with the E2F1/PCAF complex. Also, a downregulation of cell cycle genes and E2F-regulated factors has been described in KS induced pluripotent stem cells-derived neurons when compared to their healthy counterparts³⁷.

It is surprising that while patients with alterations in *EHMT1* are known since 2004, reported as a syndrome caused by subtelomeric deletions of chromosome 9q³⁸, individuals with genetic alterations in *EHMT2* have not yet been described. In contrast, EHMT2 was reported to be a histone methyltransferase before EHMT1⁶, and has been more extensively studied. Both proteins are highly similar, widely expressed in human tissues and together with the H3K9 methylation mark have been described to play critical roles in differentiation and development. Despite this, several studies suggest that EHMT2 and EHMT1 might have different functions being EHMT2 more relevant for maintaining the cellular levels of H3K9 methylation^{9,39,40}. Although both EHMT1 and EHMT2 can methylate histones in vitro on their own, heterodimers between both proteins are needed to display maximum methylation activity in vivo⁹. Dimer formation-competent but enzymatically inactive mutant Ehmt1 but not Ehmt2 can rescue specific knock out (KO) mouse embryonic stem cells (mESCs) from defective H3K9 methylation, suggesting that the catalytic activity of EHMT1 but not EHMT2 is dispensable for the complex H3K9 methyltransferase activity in vivo³⁹. In addition, it has been suggested that some of the defects observed after EHMT1 depletion could be due to altered EHMT2 activity. The Ehmt1 KO in mESCs results in reduced Ehmt2 expression likely due to the impossibility to form heterodimers, while the Ehmt2 KO had no effects in Ehmt1 expression^{9,40}. In addition, KS causing EHMT1 variants p.Cys1073Tyr and p.Arg1197Trp, which have been reported to have both defective in vitro activity and interaction with EHMT2, did not rescue the levels of H3K9 methylation nor restored EHMT2 levels when overexpressed in Ehmt1 KO mESCs⁴⁰. Therefore, it is likely that the EHMT1 truncations observed in Kleefstra syndrome patients affect the catalytic capacity of EHMT2 through alterations in heterodimer formation. These observations agree with the reduced levels of EHMT2 expression found in the KS patient ND120.

Our H3K9 methylation and gene expression data suggests that EHMT2 loss-of-function has more dramatic consequences regarding loss of H3K9 methylation and altered gene expression than loss of EHMT1 function, despite the fact that the p.Ala1077Ser change still retains some catalytic activity. This data agrees with the previously discussed more relevant role for EHMT2 in controlling the levels of H3K9 methylation than EHMT1. In this context, it is possible that loss of EHMT2 function might be more detrimental to cells than loss of EHMT1 function explaining why inactivating mutations in EHMT2 compatible with life are very rare. In agreement, mouse models have shown that the phenotypes of Ehmt1-/-embryos were mostly identical to those of Ehmt2-/-

embryos, both leading to embryonic lethality^{9,10}. Therefore, it is expected that complete loss of any of these proteins in humans results in severe developmental defects and are not compatible with life. In accordance with the human phenotype, both *Ehmt1*^{+/-} and *Ehmt2*^{+/-} mice are viable and recapitulate certain neurological traits observed in KS patients⁴¹.

However, the interpretation of the effects of missense variants versus truncating variants is not straightforward. Indeed, according to GnomAD⁴² V4 *EHMT1* is much more intolerant to loss-of-function variants than *EHMT2* (pLI=1 and LEUOF = 0.1 90%(0.07 - 0.16) vs. pLI = 0.86 and LEUOF = 0.39 90%(0.31 - 0.49), respectively), while *EHMT2* is more intolerant to missense variants ($Z = 1.98$ and $o/e = 0.87$ (0.84 – 0.91) for *EHMT1* vs. $Z = 4.63$ and $o/e = 0.68$ 90% (0.64 - 0.71) for *EHMT2*). While tools like DOMINO⁴³, which predicts pathogenicity of genes based on their properties rather than variants suggests a very likely dominant inheritance for *EHMT2* in Mendelian disorders (probability: 0.8349), GnomAD probabilities suggest that monoallelic inactivating mutations could be tolerated. If this were the case, the pathogenicity of the monoallelic alteration in patient ND095 could be explained by potential dominant negative effects caused by the *EHMT2* p.Ala1077Ser variant, such as sequestering *EHMT1* into less active heterodimeric forms, that could lead to more severe effects than just the loss of one allele. Consequently, only a limited number of *EHMT2* non-truncating variants could act via this mechanism which could also explain the rarity of cases. Importantly, a clear correlation between *EHMT1* pathogenic variants and phenotype severity has not been yet established for KS⁴⁴, reflecting the complexity of variant interpretation. In addition to deletions in chromosome 9q34.3 that eliminate one *EHMT1* copy, most common forms of *EHMT1* loss-of-function reported in KS patients is single nucleotide changes that cause frameshifts. Missense mutations are rarer and only a couple have been properly validated⁴⁰. Larger efforts would be needed to evaluate the effects of *EHMT1* and *EHMT2* missense variants and their consequences for complex function and disease severity.

Regarding diagnostic perspectives for patients with *EHMT2* alterations, we found a DNA methylation signature for patient ND095 that resembles KS patients, likely due to the formation of one complex between *EHMT1* and *EHMT2*. This shared, but slightly different, DNA methylation profile has been observed also for other disorders of the same protein complex such as BAFopathies⁴⁵. However, the increased number of patients with rare diseases tested for DNA methylation profiles has improved the resolution and specificity of the epigenatures ranging from altered protein complexes to genes, protein domains, and even single nucleotides²⁸. Therefore, it is possible that a specific DNA methylation epigenature for *EHMT2* loss-of-function different from patients with KS can be established in the future after the identification and profiling of more patients with pathogenic *EHMT2* alterations.

ACKNOWLEDGEMENTS

This work was funded with project PID2021-128087OB-I00 by MCIN /AEI /10.13039/501100011033 / FEDER, UE to M.J.B, and AESI PT20CIII/00009 (ISCIII

Platform of Biobanks and Biomodels PT-20). Funding was also partially provided by the Genome Canada and the Ontario Genomics Institute Genomics Applications Partnership Program Grant (OGI-188). This study used tools provided through the RD-Connect GPAP, which received funding originally from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 305444. We also thank the Bioinformatics Unit at ISCIII (I. Cuesta and S. Monzon) for their help in exome analysis, and A. Krepischi and L. Machado for discussions and S. Hortelano for advice on graphs. We thank the patients' association Kleefstra España for their support, the National Biobank of Rare Diseases (BioNER) for helping with patient samples and the entire SpainUDP consortium for their collaboration in this program. The authors also would like to express their gratitude to the patients, their families and their physicians for their active participation in SpainUDP. The authors appreciate the support of the UDNI for international data sharing.

TABLE 1. Anomalies found in patient ND095 compared to Kleefstra syndrome^{1,2,44,46}.

Anomalies	ND095	Kleefstra
High birth weight	-	8-50%
Microcephaly	-	30-80%
Brachycephaly	+	33%
Flat face	+	16%
Hypertelorism	+	55-70%
Arched eyebrows	+	27%
Carp mouth	+	77%
Brachydactyly	+	16%
Synophrys	+	55-80%
Epicanthal folds	+	30-70%
Upslanting palpebral fissures	+	30-70%
Midface hypoplasia	+	55-100%
Anteverted nares	+	40-80%
Protruding tongue	-	40-60%
Ear anomalies	-	45-80%
Dental anomalies	+	10-15%
Hypotonia	+	60-80%
Psychomotor delay/intellectual disability	+	100%
Autism ASD	-	30-75%
Short stature	-	10-39%
Behavioral problems	stereotypy	65-70%
Sleep disorder	-	20-50%
Epilepsy	one episode	20-50%
Brain imaging anomalies	partial empty sella turcica	50-60%
Cardiovascular anomalies	+	40-45%
Renal issues	+	15-30%
Genital anomalies	+	45-50% of males
Umbilical/inguinal hernia	+	15-20%
Skeletal anomalies	+	30-50%
Extremities	+	70%
Hypoacusia	-	20-30%
Ocular anomalies	-	45%
Obesity	-	30-40%
Respiratory complications	+	5-14%
Gastro-esophageal reflux	-	14-19%
Tracheomalacia	-	5-11%
Recurrent infections	-	26-64%
Speech anomalies	non-verbal	~100%

FIGURE LEGENDS

Figure 1. p.Ala1077 is located in the catalytic domain of EHMT2. A. Protein alignment of several human histone methyltransferases. p.Ala1077 and other amino acids described to confer gain of function are highlighted. Adapted from Kato S. et al³². B. Structure of EHMT2 SET-domain dimers with histone H3 (purple) and SAM (blue). C. Magnification shows that residue p.Ala1077 in EHMT2 establishes hydrogen bonds (highlighted in orange) with histone tail residue T6 and EHMT2 residue p.Arg1088.

Figure 2. Molecular alterations in EHMT2 catalytic domain caused by the p.Ala1077Ser change. A. Measure of cumulative methyltransferase activity over time (seconds) using 250 ng of the catalytic domain of EHMT2 wild type (WT) or p.Ala1077Ser mutant and 2.5 µg of recombinant histone H3 as a substrate. Graph shows the mean and standard deviation of two replicates B. Specific activity in nmol/min/mg of WT and p.Ala1077Ser catalytic domains. Mean and standard deviation of two replicates is shown. C. Coomassie blue stained gel depicting the interaction of WT and Ala1077 catalytic domains with histone H3 peptides amino acids 1 to 21 unmodified (H3), monomethylated at lysine 9 (H3k9me1) and dimethylated at lysine 9 (H3k9me3) D. Quantification of WT and p.Ala1077Ser catalytic domains binding to peptides relative to input. Mean and standard deviation of three quantifications is shown. P-values for comparisons between WT and p.Ala1077Ser correspond to * <0.05 , ** <0.005 and *** <0.0005

Figure 3. Results of the Episign test in patient ND095 A. Hierarchical clustering and B. multidimensional scaling plots indicate the case (red) has a DNA methylation profile similar to subjects with a confirmed Kleefstra episignature (blue) and distinct from controls (green). C. MVP score, a multi-class supervised classification system capable of discerning between multiple episignatures by generating a probability score for each episignature. The elevated score for Kleefstra syndrome indicates an episignature similar to the Kleefstra syndrome reference.

Figure 4. Transcriptomic and epigenetic analysis of fibroblast derived from a healthy control, a Kleefstra Syndrome patient (ND120) and the patient carrying the EHMT2 p.Ala1077Ser variant (ND095). A. Relative abundance of histone H3 modifications quantified by mass spectrometry in patient's fibroblasts and a healthy donor. Bars show the mean and standard deviation of triplicates. Significant differences compared to control healthy fibroblast are indicated with asterisks according to P-values * <0.05 , ** <0.005 and *** <0.0005 . B. Overlap of upregulated and downregulated genes found in ND095 or ND120 compared to the healthy control at adjusted p-value <0.05 . C. Scatter plot showing correlation between expression fold change of all genes in ND120 or ND095 compared to control fibroblasts. Pearson's correlation coefficient is indicated. Genes changing expression at an adjusted p-value <0.05 only in ND120 (pink), only in ND095 (purple), in both (orange) and not significantly changing expression (green) are shown. D. Network representation and clustering of GSEA results on GOBP terms at FDR <0.06 . Red nodes correspond to terms enriched in genes

upregulated in ND095 compared to control fibroblasts and blue nodes in
downregulated genes.

REFERENCES

1. Kleefstra, T., Brunner, H.G., Amiel, J., Oudakker, A.R., Nillesen, W.M., Magee, A., Geneviè, D., Cormier-Daire, V., Van Esch, H., Fryns, J.-P., et al. (2006). Loss-of-Function Mutations in Euchromatin Histone Methyl Transferase 1 (EHMT1) Cause the 9q34 Subtelomeric Deletion Syndrome.
2. Willemsen, M.H., Vulto-Van Silfhout, A.T., Nillesen, W.M., Wissink-Lindhout, W.M., Van Bokhoven, H., Philip, N., Berry-Kravis, E.M., Kini, U., Van Ravenswaaij-Arts, C.M.A., Delle Chiaie, B., et al. (2012). Update on Kleefstra syndrome. *Mol Syndromol* 2, 202–212. 10.1159/000335648.
3. Kleefstra, T., Kramer, J.M., Neveling, K., Willemsen, M.H., Koemans, T.S., Vissers, L.E.L.M., Wissink-Lindhout, W., Fenckova, M., Van Den Akker, W.M.R., Kasri, N.N., et al. (2012). Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. *Am J Hum Genet* 91, 73–82. 10.1016/j.ajhg.2012.05.003.
4. Faundes, V., Newman, W.G., Bernardini, L., Canham, N., Clayton-Smith, J., Dallapiccola, B., Davies, S.J., Demos, M.K., Goldman, A., Gill, H., et al. (2018). Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. *Am J Hum Genet* 102, 175–187. 10.1016/j.ajhg.2017.11.013.
5. Koemans, T.S., Kleefstra, T., Chubak, M.C., Stone, M.H., Reijnders, M.R.F., de Munnik, S., Willemsen, M.H., Fenckova, M., Stumpel, C.T.R.M., Bok, L.A., et al. (2017). Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. *PLoS Genet* 13, e1006864. 10.1371/journal.pgen.1006864.
6. Tachibana, M., Sugimoto, K., Fukushima, T., and Shinkai, Y. (2001). SET Domain-containing Protein, G9a, Is a Novel Lysine-preferring Mammalian Histone Methyltransferase with Hyperactivity and Specific Selectivity to Lysines 9 and 27 of Histone H3. *Journal of Biological Chemistry* 276, 25309–25317. 10.1074/jbc.M101914200.
7. Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D.M., and Nakatani, Y. (2002). A Complex with Chromatin Modifiers That Occupies E2F- and Myc-Responsive Genes in G0 Cells. *Science* (1979) 296, 1132–1136. 10.1126/science.1069861.
8. Padeken, J., Methot, S.P., and Gasser, S.M. (2022). Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nat Rev Mol Cell Biol* 23, 623–640. 10.1038/s41580-022-00483-w.
9. Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev* 19, 815–826. 10.1101/GAD.1284005.
10. Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., et al. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* 16, 1779–1791. 10.1101/GAD.989402.

11. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* *17*, 405–424. [10.1038/gim.2015.30](https://doi.org/10.1038/gim.2015.30).
12. López-Martín, E., Martínez-Delgado, B., Bermejo-Sánchez, E., and Alonso, J. (2018). SpainUDP: The Spanish undiagnosed rare diseases program. *Int J Environ Res Public Health* *15*, 1–16. [10.3390/ijerph15081746](https://doi.org/10.3390/ijerph15081746).
13. Laurie, S., Piscia, D., Matalonga, L., Corvo, A., Garcia, C., Fernandez-Callejo, M., Hernandez, C., Luengo, C., Ntalis, A.P., Protassio, J., et al. (2022). The RD-Connect Genome-Phenome Analysis Platform: Accelerating diagnosis, research, and gene discovery for rare diseases. *Hum Mutat.* [10.1002/humu.24353](https://doi.org/10.1002/humu.24353).
14. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21. [10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635).
15. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166–169. [10.1093/bioinformatics/btu638](https://doi.org/10.1093/bioinformatics/btu638).
16. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* *11*, R106. [10.1186/gb-2010-11-10-r106](https://doi.org/10.1186/gb-2010-11-10-r106).
17. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., and Ebert, B.L. (2005). Gene set enrichment analysis : A knowledge-based approach for interpreting genome-wide. *Proc Natl Acad Sci U S A* *102*, 15545–15550.
18. Thomas, P.D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L.P., and Mi, H. (2022). PANTHER: Making genome-scale phylogenetics accessible to all. Preprint at John Wiley and Sons Inc, [10.1002/pro.4218](https://doi.org/10.1002/pro.4218) [10.1002/pro.4218](https://doi.org/10.1002/pro.4218).
19. Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., Wu, S., and Wang, Y. (2023). SRplot: A free online platform for data visualization and graphing. *PLoS One* *18*, e0294236. [10.1371/journal.pone.0294236](https://doi.org/10.1371/journal.pone.0294236).
20. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* *13*, 2498–2504. [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303).
21. Barrero, M.J., and Malik, S. (2006). Two Functional Modes of a Nuclear Receptor-Recruited Arginine Methyltransferase in Transcriptional Activation. *Mol Cell* *24*, 233–243. [10.1016/j.molcel.2006.09.020](https://doi.org/10.1016/j.molcel.2006.09.020).
22. Castaño, J., Morera, C., Sesé, B., Boue, S., Costa, C.B., Martí, M., Roque, A., Jordan, A., and Barrero, M.J. (2016). SETD7 regulates the differentiation of human embryonic stem cells. *PLoS One* *11*, 1–21. [10.1371/journal.pone.0149502](https://doi.org/10.1371/journal.pone.0149502).
23. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* *9*, 671–675. [10.1038/nmeth.2089](https://doi.org/10.1038/nmeth.2089).

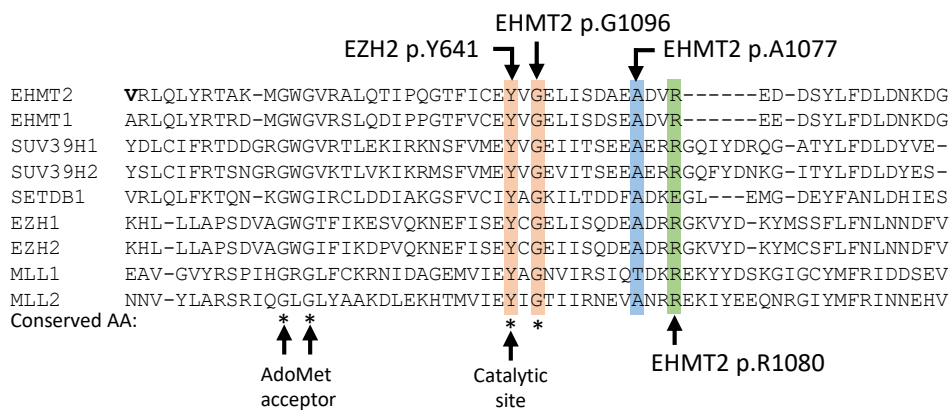
24. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank.
25. Sadikovic, B., Levy, M.A., Kerkhof, J., Aref-Eshghi, E., Schenkel, L., Stuart, A., McConkey, H., Henneman, P., Venema, A., Schwartz, C.E., et al. (2021). Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. *Genetics in Medicine* 23, 1065–1074. 10.1038/s41436-020-01096-4.
26. Aref-Eshghi, E., Bend, E.G., Colaiacovo, S., Caudle, M., Chakrabarti, R., Napier, M., Brick, L., Brady, L., Carere, D.A., Levy, M.A., et al. (2019). Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. *The American Journal of Human Genetics* 104, 685–700. 10.1016/j.ajhg.2019.03.008.
27. Aref-Eshghi, E., Kerkhof, J., Pedro, V.P., Barat-Houari, M., Ruiz-Pallares, N., Andrau, J.-C., Lacombe, D., Van-Gils, J., Fergelot, P., Dubourg, C., et al. (2020). Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *The American Journal of Human Genetics* 106, 356–370. 10.1016/j.ajhg.2020.01.019.
28. Levy, M.A., McConkey, H., Kerkhof, J., Barat-Houari, M., Bargiacchi, S., Biamino, E., Bralo, M.P., Cappuccio, G., Ciolfi, A., Clarke, A., et al. (2022). Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. *Human Genetics and Genomics Advances* 3, 100075. 10.1016/J.XHGG.2021.100075.
29. Garcia, B.A., Mollah, S., Ueberheide, B.M., Busby, S.A., Muratore, T.L., Shabanowitz, J., and Hunt, D.F. (2007). Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat Protoc* 2, 933–938. 10.1038/nprot.2007.106.
30. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–968. 10.1093/bioinformatics/btq054.
31. Jayaram, H., Hoelper, D., Jain, S.U., Cantone, N., Lundgren, S.M., Poy, F., Allis, C.D., Cummings, R., Bellon, S., and Lewis, P.W. (2016). S-adenosyl methionine is necessary for inhibition of the methyltransferase G9a by the lysine 9 to methionine mutation on histone H3. *Proc Natl Acad Sci U S A* 113, 6282–6287. 10.1073/pnas.1605523113.
32. Kato, S., Weng, Q.Y., Insko, M.L., Chen, K.Y., Muralidhar, S., Pozniak, J., Diaz, J.M.S., Drier, Y., Nguyen, N., Lo, J.A., et al. (2020). Gain-of-Function Genetic Alterations of G9a Drive Oncogenesis. *Cancer Discov* 10, 980–997. 10.1158/2159-8290.CD-19-0532.
33. Plazas-Mayorca, M.D., Bloom, J.S., Zeissler, U., Leroy, G., Young, N.L., Dimaggio, P.A., Krugylak, L., Schneider, R., and Garcia, B.A. (2010). Quantitative proteomics reveals direct and indirect alterations in the histone code following methyltransferase knockdown. *Mol Biosyst* 6, 1719–1729. 10.1039/c003307c.
34. Chen, X., Skutt-Kakaria, K., Davison, J., Ou, Y.-L., Choi, E., Malik, P., Loeb, K., Wood, B., Georges, G., Torok-Storb, B., et al. (2012). G9a/GLP-dependent histone H3K9me2

- patterning during human hematopoietic stem cell lineage commitment. *Genes Dev* 26, 2499–2511. 10.1101/gad.200329.112.
35. Verbaro, D.J., Sakurai, N., Kim, B., Shinkai, Y., and Egawa, T. (2018). Cutting Edge: The Histone Methyltransferase G9a Is Required for Silencing of Helper T Lineage–Associated Genes in Proliferating CD8 T Cells. *The Journal of Immunology* 200, 3891–3896. 10.4049/jimmunol.1701700.
 36. Rao, V.K., Ow, J.R., Shankar, S.R., Bharathy, N., Manikandan, J., Wang, Y., and Taneja, R. (2016). G9a promotes proliferation and inhibits cell cycle exit during myogenic differentiation. *Nucleic Acids Res* 44, 8129–8143. 10.1093/nar/gkw483.
 37. Fear, V.S., Forbes, C.A., Anderson, D., Rauschert, S., Syn, G., Shaw, N., Jamieson, S., Ward, M., Baynam, G., and Lassmann, T. (2022). CRISPR single base editing, neuronal disease modelling and functional genomics for genetic variant analysis: pipeline validation using Kleefstra syndrome EHMT1 haploinsufficiency. *Stem Cell Res Ther* 13. 10.1186/s13287-022-02740-3.
 38. Harada, N., Visser, R., Dawson, A., Fukamachi, M., Iwakoshi, M., Okamoto, N., Kishino, T., Niikawa, N., and Matsumoto, N. (2004). A 1-Mb critical region in six patients with 9q34.3 terminal deletion syndrome. *J Hum Genet* 49, 440–444. 10.1007/s10038-004-0166-z.
 39. Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H., and Shinkai, Y. (2008). G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. *EMBO J* 27, 2681–2690. 10.1038/emboj.2008.192.
 40. Yamada, A., Shimura, C., and Shinkai, Y. (2018). Biochemical validation of EHMT1 missense mutations in Kleefstra syndrome. *J Hum Genet* 63, 555–562. 10.1038/s10038-018-0413-3.
 41. Balemans, M.C.M., Huibers, M.M.H., Eikelenboom, N.W.D., Kuipers, A.J., van Summeren, R.C.J., Pijpers, M.M.C.A., Tachibana, M., Shinkai, Y., van Bokhoven, H., and Van der Zee, C.E.E.M. (2010). Reduced exploration, increased anxiety, and altered social behavior: Autistic-like features of euchromatin histone methyltransferase 1 heterozygous knockout mice. *Behavioural Brain Research* 208, 47–55. 10.1016/j.bbr.2009.11.008.
 42. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443. 10.1038/s41586-020-2308-7.
 43. Quinodoz, M., Royer-Bertrand, B., Cisarova, K., Di Gioia, S.A., Superti-Furga, A., and Rivolta, C. (2017). DOMINO: Using Machine Learning to Predict Genes Associated with Dominant Disorders. *The American Journal of Human Genetics* 101, 623–629. 10.1016/j.ajhg.2017.09.001.
 44. Ciaccio, C., Scuvera, G., Tucci, A., Gentilin, B., Baccarin, M., Marchisio, P., Avignone, S., and Milani, D. (2018). New Insights into Kleefstra Syndrome: Report of Two Novel Cases with Previously Unreported Features and Literature Review. *Cytogenet Genome Res* 156, 127–133. 10.1159/000494532.

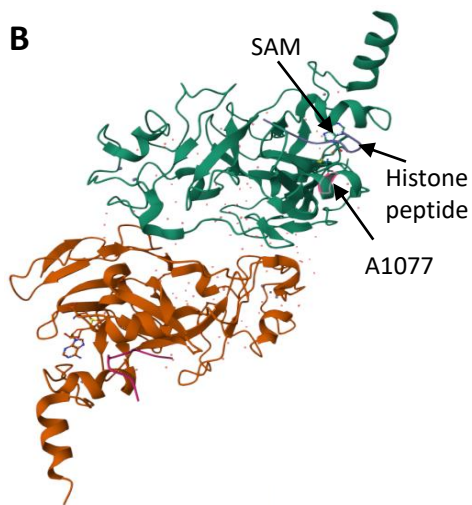
45. Aref-Eshghi, E., Bend, E.G., Hood, R.L., Schenkel, L.C., Carere, D.A., Chakrabarti, R., Nagamani, S.C.S., Cheung, S.W., Campeau, P.M., Prasad, C., et al. (2018). BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin–Siris and Nicolaides–Baraitser syndromes. *Nat Commun* 9, 4885. [10.1038/s41467-018-07193-y](https://doi.org/10.1038/s41467-018-07193-y).
46. Kleefstra, T., van Zelst-Stams, W.A., Nillesen, W.M., Cormier-Daire, V., Houge, G., Foulds, N., van Dooren, M., Willemsen, M.H., Pfundt, R., Turner, A., et al. (2009). Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. *J Med Genet* 46, 598–606. [10.1136/jmg.2008.062950](https://doi.org/10.1136/jmg.2008.062950).

Figure 1

A



B



C

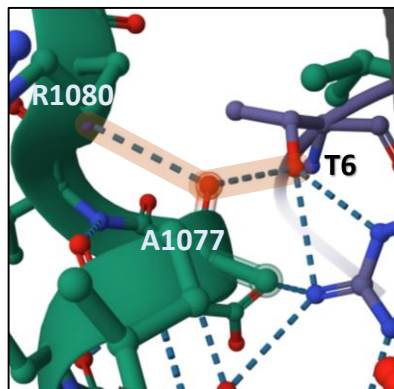


Figure 2

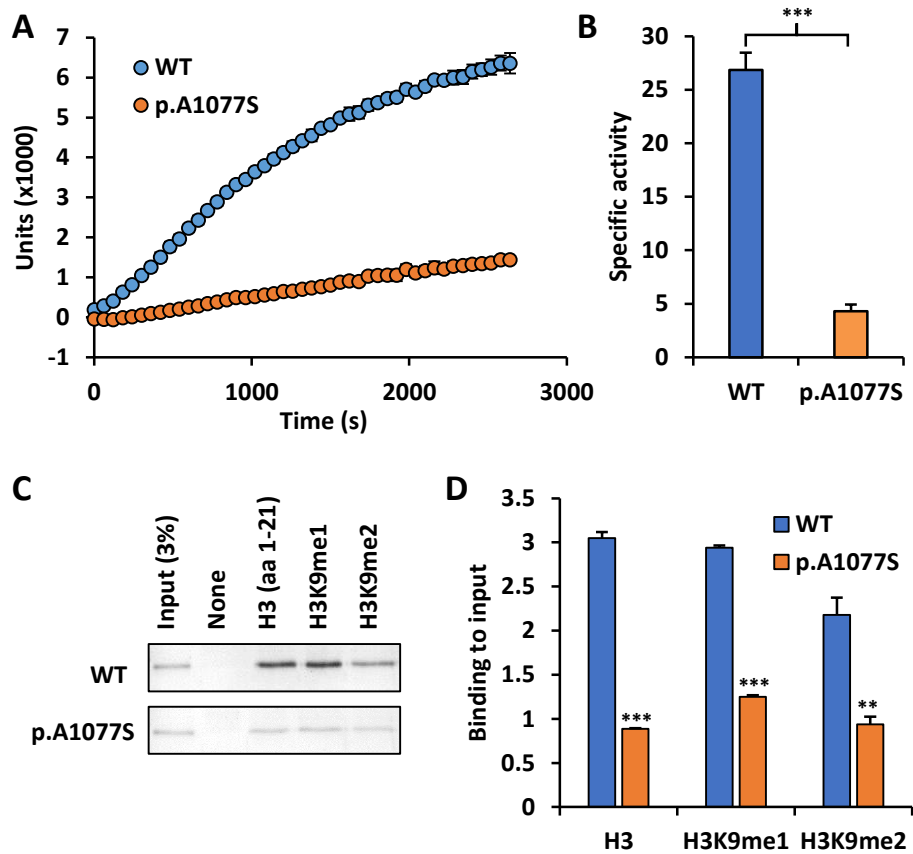


Figure 3

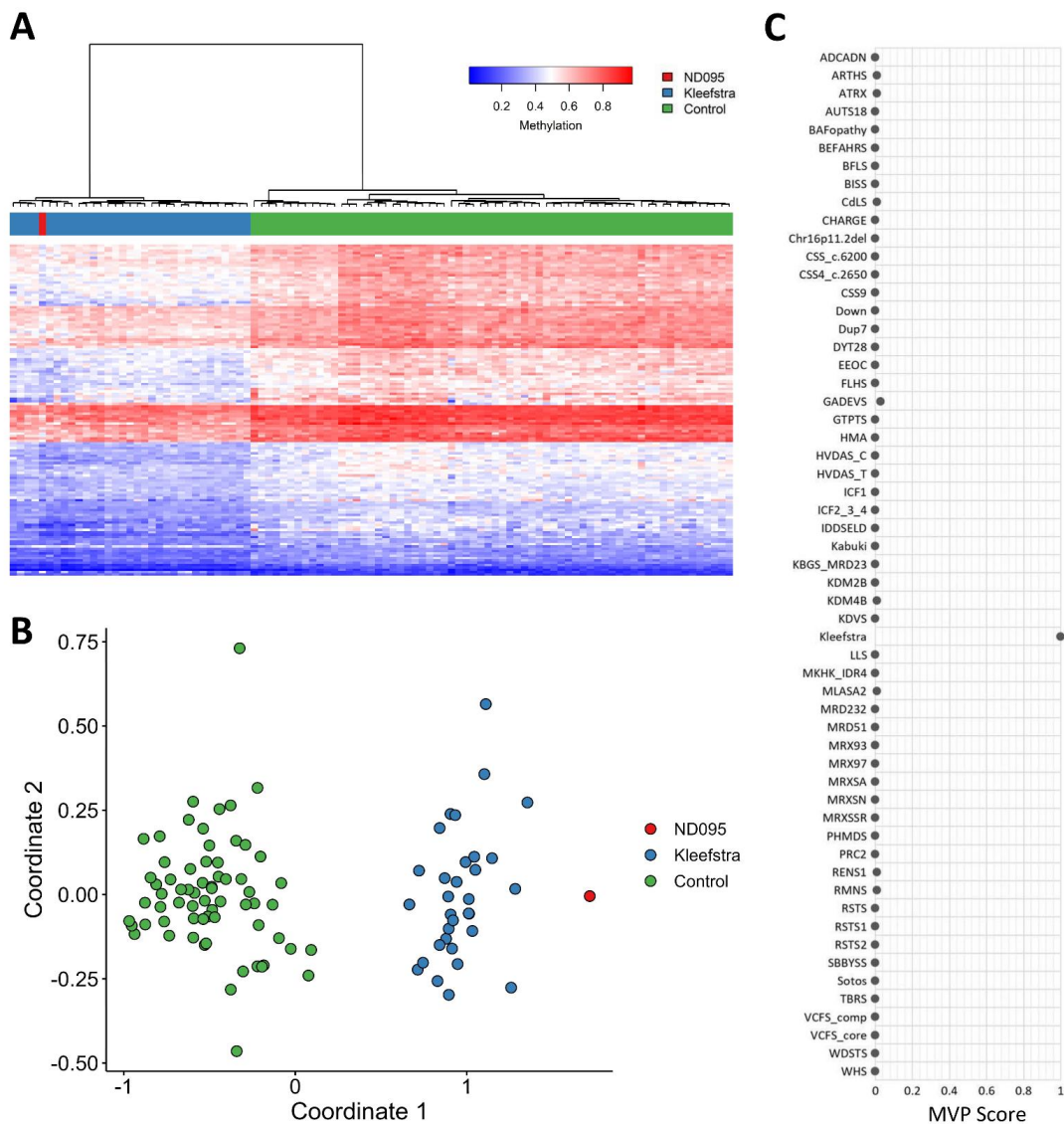


Figure 4

