De novo heterozygous variants in EHMT2 genocopy Kleefstra syndrome via loss of G9a methyltransferase activity.

Aleš Hnízda[1†], Beatriz Martinez-Delgado[2,3,4†], Diana Sanchez-Ponce[2], Javier Alonso[2,3,4], Jeanne Amiel[5] , Tania Attie-Bitach[5], Ariadna Bada-Navarro[2], Beatriz Baladron[2,3], Eva Bermejo-Sanchez[2,3], Vítězslav Brinsa[1,6], Ivana Buková[7], Rosario Cazorla-Calleja[3,8], Sylvie Červenková[7], Shanshan Chow[9], Petr Dušek[10], Marta Fernandez[2], Sourav Ghosh[11], Gema Gomez-Mariano[2], Andrea Gřegořová[12], Mark James Hamilton[13], Hana Hartmannová[1], Esther Hernandez-San Miguel[2,3], Marina Herrero-Matesanz[2], Kateřina Hodaňová[1], Alan Kádek[14], Jennifer Kerkhof[11], Tjitske Kleefstra[15,16], Didier Lacombe[17], Michael A. Levy[11], Estrella Lopez-Martin[2,3], Ruaud Lyse[18], Petr Man[14], Purificacion Marin-Reina[19], Ellen F. Macnamara[9], Haley McConkey[11,20], Petra Melenovská[1], Lidia M. Mielu[2,3] , David Moore[21], Lenka Mrázová[1], Karolína Musilová[1], Petr Nickl[7], Martina Pavlíková[1], Lea Pavlovičová[1], Manuel Posada[2,3], Jan Procházka[7], Kateryna Pysanenko[7], Sheila Ramos del Saz[2], Dmitrijs Rots[15], Jessica Rzasa[11], Radislav Sedláček[7], Viktor Stránecký[1], František Špoutil[7], Matthew L. Tedder[22], Louise Thompson[21], Cynthia J. Tifft[23], Frederic Tran Mau-Them[24], Helena Trešlová[1], Antonio Vitobello[24], Bekim Sadikovic[11,20], Jakub Sikora[1,25], Stanislav Kmoch[1], Maria J. Barrero[2,3*], Lenka Nosková[1*]

*** Corresponding authors: lnosk@lf1.cuni.cz; mj.barrero@isciii.es**

[†]These authors contributed equally to this work.

[1] Research Unit for Rare Diseases, Department of Pediatrics and Inherited Metabolic Disorders, First Faculty of Medicine and General University Hospital, Charles University in Prague, Czech Republic

[2] Institute of Rare Diseases Research (IIER), Spanish National Institute of Health Carlos III (ISCIII), 28220 Madrid, Spain

[3] Undiagnosed diseases program SpainUDP, Madrid, Spain

[4] Centro de Investigación Biomedica en Red de Enfermedades Raras, (CIBERER), Madrid, Spain

[5] Service de Médecine Génomique des Maladies Rares, Hôpital Necker-Enfants Malades, AP-HP, Paris, France

[6] Department of Biochemistry, Faculty of Science, Charles University in Prague, Czech Republic

[7] Czech Centre for Phenogenomics, BIOCEV, Institute of Molecular Genetics, Prague, Czech Republic

[8] Pediatric Neurology Department, Hospital Universitario Puerta de Hierro, Majadahonda, Madrid, Spain

[9] Undiagnosed Diseases Program, National Institutes of Health, Bethesda, MD, USA

[10] Department of Neurology and Center of Clinical Neuroscience, First Faculty of Medicine, Charles University and General University Hospital in Prague, Czech Republic

[11] Verspeeten Clinical Genome Centre, London Health Sciences Centre, London, ON, Canada

[12] Institute of Molecular and Clinical Pathology and Medical Genetics, University Hospital Ostrava, Czech Republic

[13] West of Scotland Clinical Genetics Service, Queen Elizabeth University Hospital, Glasgow, UK

[14] Institute of Microbiology, Czech Academy of Sciences, Prague, Czech Republic

[15] Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands

[16] Department of Human Genetics, Radboud University Medical Center, 6500 HB, Nijmegen, The Netherlands

[17] Génétique Médicale, CHU Bordeaux, INSERM U1211 (MRGM), Université de Bordeaux, France

[18] Département de génétique, APHP-Nord, Paris, France, Service de cytogénétique et génétique médicale, Hôpital de la Mère et de l'Enfant, CHU Limoges, Limoges France

[19] University and Polytechnic Hospital La Fe, Valencia, Spain

[20] Department of Pathology and Laboratory Medicine, Western University, London Ontario Canada

[21] South East Scotland Genetic Service, Western General Hospital, Edinburgh, UK

[22] Greenwood Genetic Center, Greenwood, SC, USA

[23] Office of the Clinical Director and Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, 10 Center Drive, Bethesda, MD, USA

[24] Université Bourgogne Europe, CHU Dijon Bourgogne, Laboratoire de Génomique Médicale, FHU-TRANSLAD, Centre de recherche Translationnelle en Médecine moléculaire – Inserm UMR1231 équipe GAD, Dijon, France

[25] Institute of Pathology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Czech Republic

Abstract

EHMT1 and *EHMT2* genes encode human euchromatin histone lysine methyltransferase 1 and 2 (EHMT1 alias GLP; EHMT2 alias G9a) that form heteromeric GLP/G9a complexes with essential roles in epigenetic regulation of gene expression. While *EHMT1* haploinsufficiency was established as the cause of Kleefstra syndrome twenty years ago, the pathogenesis of G9a

dysfunction in human disease remains largely unknown. Here, we report clinical and molecular correlates of six *de novo EHMT2* variants in patients with clinical presentation, episignatures, histone modifications and transcriptomic profiles similar to those of Kleefstra syndrome. *In vitro* studies revealed that these variants encode for structurally stable G9a proteins that are catalytically incompetent due to aberrant interactions either with histone H3 tail or with S-adenosylmethionine. Heterozygous mice carrying a patient-derived variant (*Ehmt2 c.3385_3396del*) exhibited growth retardation, facial/skull dysmorphia and aberrant behavior. *EHMT2* variants described here likely exert dominant-negative effect on GLP/G9a complexes and thus genocopy the *EHMT1* haploinsufficiency causing Kleefstra syndrome via a distinct molecular mechanism.

Introduction

Histone methylation regulates chromatin compaction and dynamically orchestrates cellular transcription programs. Histone methylation status is maintained by the activity of histone methyltransferases and demethylases[1]. Their dysregulation has been implicated in many neurodevelopmental disorders including monogenic and multifactorial diseases (e.g. Alzheimer or Parkinson disease)[2]. So far, at least eleven distinct genetic syndromes caused by malfunction of histone methyltransferases have been described[3].

EHMT1 and *EHMT2* genes encode human euchromatin histone lysine methyltransferase 1 and 2 (EHMT1 alias GLP; EHMT2 alias G9a), respectively. GLP and G9a have identical domain architecture with high sequence similarity. Their main characteristic is the presence of a highly conserved SET domain that catalyzes the transfer of methyl group from S-adenosylmethionine to lysine 9 of histone H3 (H3K9)[4]. The SET domains assemble into functional dimers[5] and govern the formation of heteromeric GLP/G9a complexes *in vivo*[6]. Despite the ability of these proteins to

form functional homodimers, heterodimers are the preferred form *in vivo*, displaying the maximum enzymatic activity[6,7]. The proper formation and methylation activity of the GLP/G9a complexes is critical for epigenetic regulation of gene expression in many biological processes including oogenesis[8], metabolic regulation[9] or neuronal maturation during brain development[10].

Haploinsufficiency of *EHMT1* was originally described in the *9q* subtelomeric deletion syndrome, a clinically recognizable neurodevelopmental disorder with congenital anomalies, specific dysmorphic features and other characteristic symptoms known as Kleefstra syndrome (KLEFS1, OMIM# 610253)[11,12]. Over time, additional Kleefstra syndrome patients have been identified with pathogenic frameshift, nonsense or missense *EHMT1* variants, which mostly lead to haploinsufficiency through the instability and decreased amount of functional GLP proteins[13,14]. Recently, a broader phenotypic spectrum of Kleefstra syndrome associated with diverse biochemical properties of differently mutated GLP has been described[15].

Homozygous loss of either GLP (*Ehmt1*) or G9a (*Ehmt2*) in mice results in embryonic lethality, underscoring the essential role of both proteins in early development[6]. Moreover, heterozygous *Ehmt1*$^{+/-}$ mice exhibit a spectrum of neurodevelopmental abnormalities, including facial/skull dysmorphia, behavioral deficits[16], synaptic impairments and disrupted hippocampal neurogenesis[17] recapitulating clinical features of Kleefstra syndrome[18]. Importantly, in contrast to severe effect observed for *Ehmt1*$^{+/-}$, *Ehmt2*$^{+/-}$ mice present with only minimal abnormalities[19].

While the molecular mechanisms of epigenetic dysregulation and pathogenesis involved in the clinical symptoms of Kleefstra syndrome are still under investigation, the epigenetic hallmarks that characterize *EHMT1* haploinsufficiency are well described. Recent research revealed that patients with Kleefstra syndrome show a specific DNA methylation signature (episignature) in peripheral blood[20]. This episignature can aid the diagnosis of patients with clinical symptoms

suggestive of Kleefstra syndrome, but with no pathogenic *EHMT1* variants revealed by molecular genetic testing and/or with variants of uncertain clinical significance[15,20,21]. Additionally, abnormal epigenetic states have been described in Kleefstra syndrome neurons derived from induced pluripotent stem cells (iPSCs) and in *Ehmt1*$^{+/-}$ mice, manifested by dysregulated H3K9me2 global levels[22,19].

Despite the recognized functional cooperation and interdependency between GLP and G9a[6,23-25], clinical correlates of G9a dysfunction in humans have not yet been thoroughly described, although the potential involvement of *EHMT2* biallelic loss in a patient with neurodevelopmental delay has been recently suggested[21].

Here we report clinical, structural, biochemical and molecular correlates of six *de novo* heterozygous *EHMT2* variants that we identified through exome/genome sequencing. We show that all these *EHMT2* variants encode for structurally stable but catalytically inactive G9a proteins that seem to possess dominant negative effects on the function of GLP/G9a complexes. Patients harboring these variants showed a clinical phenotype consistent with Kleefstra syndrome. Patient-derived cell lines displayed alterations in episignatures, histone modifications and gene expression profiles closely resembling Kleefstra syndrome. Notably, the *EHMT2* patients showed a specific episignature in blood, enabling the diagnosis of an additional patient carrying a pathogenic *EHMT2* variant. Moreover, a knock-in mouse model harboring a patient-derived heterozygous deletion of four amino acids in the SET domain of G9a recapitulates multiple key phenotypic features observed in patients with *EHMT2* variants reported here.

Results

Multiple de novo variants in EHMT2 are associated with a neurodevelopmental phenotype and facial dysmorphism.

Six patients (P1-P6) carrying six different *de novo* heterozygous variants in *EHMT2* gene have been identified through trio-exome/genome sequencing in several national research or diagnostic programs for rare diseases. The investigators and health care professionals were gathered via GeneMatcher[26], RD-Connect Genome-Phenome Analysis Platform (GPAP)[27] and personal communication. Four of the *EHMT2* variants encode for missense G9a mutations: NM_006709.5: c.3229G>T p.(Ala1077Ser), c.3335A>G p.(Asn1112Ser), c.3472T>C p.(Phe1158Leu) and c.3485A>G p.(Lys1162Arg) and two variants encode for *in-frame* G9a deletions: NM_006709.5: c.3225_3236del p.Glu1076_Val1079del and c.3487_3492del p.(Ser1163_Lys1164del). None of these variants has been previously reported in the Genome Aggregation Database (gnomAD v4.1.0) and ClinVar database. The variants encode for evolutionary conserved residues and are predicted as pathogenic by *in silico* prediction tools (**Table 1**).

The patients` cohort was composed of four females and two males aged 2 – 16 years. All patients presented with facial dysmorphism including midface hypoplasia, synophrys, hypertelorism, dysplastic ears, short philtrum and deeply set eyes, and cardiovascular anomalies including septal defects (ventricular or atrial), pulmonal valve stenosis and coarctation of the aorta. All but one patient presented with hypotonia, psychomotor and developmental delay, absence of speech and behavioral abnormalities. A single patient (P3) with the c.3335A>G p.(Asn1112Ser) variant had normal psychomotor, cognitive and speech abilities but exhibited paroxysmal hyperkinetic (choreoathetoid) movements. Individually, patients presented with various other abnormalities including pes planovalgus or equinovalgus, postaxial polydactyly or big halluxes, kidney cysts and other urogenital abnormalities. Detailed clinical description of patients is provided in **Table 1** and in **Supplementary Material**. Overall, the clinical presentation of the patients was very similar to those of Kleefstra syndrome caused by *EHMT1* haploinsufficiency.

Episignatures of patients with EHMT2 variants overlap with those of Kleefstra syndrome yet reveal a specific hypomethylated episignature.

Analyses of patients' blood DNA methylation profiles using the Episign assay[20,28-30] revealed an episignature profile that matched the Kleefstra syndrome episignature (Supplementary Figure 1). Unsupervised hierarchical clustering and multidimensional scaling (MDS) plots demonstrated that the patients with *EHMT2* variants clustered closely with the Kleefstra syndrome cohort, with one sample (P3) showing complete overlap (Supplementary Figure 1A, B). Furthermore, a multi-class supervised classification model predicted the *EHMT2* patients as Kleefstra syndrome with high confidence, while effectively distinguishing them from other established episignatures associated with genetic disorders in the EpiSign™ Knowledge Database (EKD) (Supplementary Figure 1C). Next, we conducted a discovery analysis to assess the presence of an *EHMT2* specific episignature. For that, we compared the episignatures of the six patients carrying the EHMT2 variants and 54 matched controls, and identified 244 top differentially methylated probes. The mean methylation difference between patients and controls ranged from +45.52% to -35.77% (median = -18.60%), with 88.52% of the signature probes being hypomethylated (Supplementary Table S1). Interestingly, these probes were mapped to 155 genes that were enriched in the Gene Ontology term multicellular organism development at a p-value: 6.1E-7. Cluster analysis validated the feature selection, showing a clear separation between *EHMT2* patients and controls. The analysis also showed a clear separation from the testing Kleefstra syndrome samples in both heatmap and multidimensional scaling (MDS) plots (**Figure 1A, B**). Furthermore, an unsupervised leave-one-out cross-validation on the discovery cases demonstrated that each hold-out test case consistently clustered with the remaining discovery cases in heatmaps and aligned closely with patient group centroids in MDS plots (Supplementary Figure 2). This supports the reproducibility and robustness

of the identified episignature. We then developed a Support Vector Machine (SVM)-based model to evaluate its diagnostic utility for *EHMT2*-related disorders. The model demonstrated high specificity, as nearly all EKD test samples yielded MVP scores close to 0 (**Figure 2C**). It also showed sensitivity to an additional patient with a *de novo EHMT2* variant (EHMT2(NM_006709.5):c.3229G>A p.(Ala1077Thr) that was identified in the EKD database (P7), which had an MVP score greater than 0.75, clustered with discovery cases in the MDS plot, and displayed similar profiles in the heatmap (**Figure 1A, B and C**). The clinical characteristics of this patient were consistent with those of the *EHMT2* patient cohort **(Table 1)**. Additionally, all testing Kleefstra samples had very low MVP scores (< 0.1), did not match the signature, and exhibited profiles more similar to controls. Overall, the episignature analysis suggest that *EHMT2*-related disorders are related to Kleefstra syndrome but represent a distinct molecular entity with its own specific episignature.

EHMT2 patients show histone modification and gene expression signatures that match those of Kleefstra syndrome.

Next, we investigated whether the identified patients had a histone modification profile similar to the one reported for Kleefstra syndrome and in accordance with the loss of G9a activity. We established fibroblast cultures from skin biopsy from patient P2, together with a healthy skin donor and a Kleefstra syndrome patient (KSP) carrying a frameshift variant in *EHMT1* (NM_024757.5: c.1881delT p.(His629ThrfsTer12)). We further generated iPSCs from the established fibroblasts cultures and from peripheral blood mononuclear cells available from patient P1 that were characterized according to the established standards (Supplementary Figure 3 and 4). Quantification of histone modifications in fibroblasts and iPSCs lines by mass spectrometry revealed significantly (p-value<0.05) decreased levels of H3K9me1 or H3K9me2 in all patients'

cell lines compared to control cell lines (**Figure 2A**, Supplementary Table S2). While changes in histone modifications in iPSCs were rather subtle, more pronounced changes were observed in fibroblast cultures in which we detected significant downregulation of H3K9me1/2/3 and upregulation of unmodified and acetylated H3K9 in patients. The changes in histone modifications were consistent between Kleefstra syndrome and P2 patient's fibroblasts, showing coordinated upregulation of several histone marks involved in gene activation such as H3K56ac, H3K64ac and H3K79me3 (**Figure 2B**, Supplementary Figure 5A).

In agreement with the altered histone modification patterns in the fibroblasts and iPSCs patient-derived cell lines, we observed significant changes in gene expression in patients compared to controls that correlated among patients (**Figure 2C**, Supplementary Figures 6A-C). Accordingly, there was a significant overlap between the patients' cell lines in up- and downregulated genes (Supplementary Figure 6D-G). Upregulated genes, most likely G9a direct target genes, were enriched in alternative lineages such as skeletal, muscle and neuronal systems development, and in the cell cycle and Rho GTPase pathways (**Figure 2D**, Supplementary Figures 7 and 8). Overall, the alterations in histone modifications and gene expression observed in patient-derived cells carrying *EHMT2* variants show substantial overlap with those seen in Kleefstra syndrome and are consistent with impaired H3K9 methyltransferase activity.

EHMT2 variants have no effect on G9a transcript and protein levels

To explore the effects of the identified variants, we examined *EHMT2* transcript and G9a protein amounts in the available patient-derived samples. RNA sequencing revealed very similar amounts of wild-type and mutant *EHMT2* transcripts in peripheral leukocytes obtained from P1, P2 and P3, iPSCs from P1 and P2, and fibroblasts from P2 (Supplementary Figure 9). The amounts of G9a protein were also very similar in patient-derived and control iPSCs as determined by western blot

(Supplementary Figure 9). Thus, the identified *EHMT2* variants result in the synthesis of sufficient amounts of structurally stable but likely malfunctional G9a protein.

EHMT2 variants cluster at the G9a enzyme-substrate interface

Sequence alignment and structural modelling (**Figure 3;** Supplementary Figure 10 and Supplementary Table S3) revealed that five of the predicted amino acid changes (Del 1076-9, A1077S, F1158L, K1162R and Del 1163-4) affect the G9a binding site for histone H3 tail, whereas the N1112S substitution affects the G9a binding site for the methyl group donor (S-adenosylmethionine). Thus, all identified variants cluster at the G9a enzyme-substrate interface with potential effects on the histone H3 tail interaction and catalytic activity.

All EHMT2 variants compromise the catalytic activity of G9a through defective enzyme-substrate interactions

Since structural modelling showed that all mutations cluster at the G9a enzyme-substrate interface, we cloned, expressed and purified recombinant methyltransferase domains and examined their catalytic activities. Specifically, we determined the extent of methylation of K9 in an N-terminal histone H3 peptide (amino acid residues 1-20) and nucleosome. Initially, using semi-quantitative analysis by MALDI-TOF mass spectrometry we observed that mono- and di-methylation of histone H3 peptide was decreased with A1077S, K1162R and Del 1163-4 mutations and absent in Del 1076-9, N1112S and F1158L mutations (Supplementary Figure 11). Subsequently, we used a bioluminescence-based assay analysis to quantify the initial velocity rates of lysine methylation on histone H3 peptide or nucleosome. We observed that five of the mutations, Del 1076-9, N1112S, F1158L and K1162R, Del 1163-4, decreased the methylation activity towards histone H3 peptide to less than 10% of the wild-type enzyme (**Figure 4A**). A higher (~ 30%) residual activity was observed for the A1077S mutation. Notably, the methylation of mononucleosome was

reduced to less than 5% of the wild-type enzyme activity with all studied mutations, including the A1077S. These data demonstrate that all studied mutations reduce the catalytic competence of G9a towards the histone H3 tail.

To identify the molecular mechanisms of this reduced catalytic competence, we assessed the structural stability and binding affinity of the recombinant G9a mutants towards their substrates. First, we used native mass spectrometry (nMS), which measures the mass of intact non-covalent protein complexes[31], to quantify the relative amounts of properly folded dimeric forms of the corresponding mutated proteins. Native MS revealed that all G9a mutants predominantly assemble into compactly-folded dimers (**Figure 4B** and Supplementary Figure 12). Using differential scanning fluorimetry (DSF) performed under temperature gradient, we determined that the melting temperatures of free G9a proteins were slightly decreased by N1112S, F1158L, and Del1163-4 mutations and not affected by Del 1076-9, A1077S and K1162R mutations (**Figure 4C**). These data showed that while mutations led to local changes in G9a, none of them has major destabilizing effects on the corresponding overall G9a structures.

Next, we determined the binding affinities of individual G9a mutations for histone H3 peptide and SAM using nMS. Binding affinities were estimated from observed substrate binding visible as 398 Da (for SAM) or 2100 Da (for histone H3 peptide) mass increases of corresponding non-covalent protein-ligand complexes. This analysis revealed that mutants Del 1076-9, A1077S, F1158L, K1162R and Del 1163-4 had reduced binding to histone H3 peptide but unaffected binding to SAM (**Figure 4D**). Opposite results were observed for the N1112S mutation that exhibited normal binding to histone H3 peptide but failed to interact with SAM. Substrate binding behaviour of individual G9a mutations were further corroborated by binding affinity estimated from substrate concentration-dependent increased thermostability of corresponding protein-ligand complexes

using DSF. This analysis further confirmed the decreased binding capacity of Del 1076-9, A1077S, F1158L, K1162R and Del 1163-4 mutations to histone H3 peptide and revealed that the Del 1076-9 and A1077S mutations had partially reduced binding capacities to SAM (**Figure 4C**). Consistently with our observations above, opposite results were recorded for the N1112S mutation that exhibited normal binding to histone H3 peptide but failed to interact with SAM.

Together, nMS and DSF revealed that pathogenic G9a mutations disturb the catalytic competence of the enzyme through defective interaction with one or both of enzyme substrates, while the structural integrity and stability of the protein is not affected.

Using the established protocol, we also examined the heterozygous G9a variant p.T961I previously reported as a variant of unknown significance in a patient with an autistic disorder[32]. This variant is located outside the methyltransferase site and exhibited normal methyltransferase activity and protein stability (Supplementary Figure 13). Given these results, we classify this variant as likely benign.

Pathogenic EHMT1 SET domain variants affect protein stability and dimer formation

To compare the *EHMT2* variants with previously identified pathogenic missense *EHMT1* variants, we also similarly assessed the effects of two previously described SET-domain GLP pathogenic mutants: C1073Y and R1197W[13].

Our biochemical examinations revealed that both mutations render the GLP SET-domain considerably unstable when expressed in *E.coli*, with most of the mutant proteins accumulating in insoluble fractions, which either entirely abolished (R1197W) or decreased (C1073Y) purification yields (Supplementary Figure 14). With reduced but sufficient amounts of purified C1073Y mutant, we were able to determine its decreased thermal stability (Supplementary Table S4) and impaired dimerization (**Figure 4B, D**). Despite this, the C1073Y mutant showed binding affinities

to histone H3 peptide and SAM close to those of wild-type GLP (**Figure 4D**) and had a relatively high (~ 60%) methylation activity with histone H3 peptide but almost no (< 1%) activity with mononucleosome compared to the wild-type protein (Supplementary Figure 14C).

To further examine the aberrant catalysis of mutated methyltransferases, we compared the kinetic parameters of G9a (A1077S) and GLP (C1073Y) variants retaining relatively high residual activity towards histone H3 peptide. This analysis revealed that the C1073Y GLP mutation did not significantly affect enzyme turnover number (k_{cat}), while the A1077S G9a mutation significantly decreased it (Supplementary Table S5). Specifically, k_{cat} values for A1077S G9a were 2.2-fold and 4.5-fold lower towards histone H3 peptide and SAM, respectively, compared to the wild-type enzyme. Therefore, our biochemical and enzymatic analyses confirm that pathogenic G9a variants are catalytically incompetent *per se,* while the tested pathogenic missense GLP mutations lose their activities due to structural instability.

***Ehmt2$^{+/del_1076\text{-}1079}$* mice present with postnatal developmental abnormalities and craniofacial dysmorphia**

To further validate the functional relevance of the *EHMT2* variants at the organismal level, we generated a mouse *knock-in* model (*Ehmt2$^{+/del_1076\text{-}1079}$*) lacking four amino acids in the SET domain of G9a that corresponds to the deletion identified in patient P1 (Supplementary Figure 15). *Ehmt2$^{+/del_1076\text{-}107}$* and littermate *Ehmt2$^{+/+}$* control mice used in the studies were generated by *in-vitro* fertilization. None (out of 13) *Ehmt2$^{+/del_1076\text{-}1079}$* females gave any progeny when mated with fertile *Ehmt2$^{+/+}$* males. Two (out of 18) *Ehmt2$^{+/del_1076\text{-}1079}$* males reproduced once with fertile *Ehmt2$^{+/+}$* females.

Approximately half of the *Ehmt2$^{+/del_1076\text{-}1076}$* animals of both sexes did not survive over postnatal day (p) 30-40 (**Figure 5A**). All surviving mice were terminated no later than p168. In comparison

to sex-matched *Ehmt2*$^{+/+}$ mice, *Ehmt2*$^{+/del_1076\text{-}1076}$ presented with significant (and long-term lasting) weight-gain deficits detectable as early as p1-2 (**Figure 5B,** Supplementary Figure 16). Mice were subjected to facial/skull morphology assessment by microCT, SHIRPA scoring and a battery of behavioral tests (open field, rotarod, grip strength) at the age of 3-4 months. MicroCT and craniometric analyses identified a systematically occurring dysmorphia comprised of saddle and "bent" nose with shorter nasalia, nasal cavity disorganization, shorter and smaller palate and smaller (shorter) neurocranium in *Ehmt2*$^{+/del_1076\text{-}1076}$ mice (**Figure 5C** and **5D**, Supplementary Figures 17-19). Importantly, sphenoid bone formation did not seem to be impacted in *Ehmt2*$^{+/del_1076\text{-}1076}$ in comparison to *Ehmt2*$^{+/+}$ animals (Supplementary Figure 20).

The overall SHIRPA scores were significantly lower in *Ehmt2*$^{+/del_1076\text{-}1076}$ animals mainly due to their lower locomotor activity (Supplementary Figure 21). This observation was consistent with results of the open-field test, in which *Ehmt2*$^{+/del_1076\text{-}1076}$ mice travelled significantly shorter distances at lower speeds and rested more than their *Ehmt2*$^{+/+}$ littermates. Interestingly, no difference in center zone preference/avoidance suggestive of anxious behavior was observed between the two alternative genotypes/sexes (Supplementary Figure 22). Similarly to the open field test, *Ehmt2*$^{+/del_1076\text{-}1076}$ mice had significantly shorter overall times spent, and distances travelled on the rotating rod (rotarod test). While not significantly different, *Ehmt2*$^{+/+}$ mice had higher average difference in the times-to-fall between the first and last (6th) trial on the beam (Supplementary Figure 23) suggesting their better ability to learn in this motor coordination task. Last, grip strength values were comparable between the two genotypes (Supplementary Figure 24). Together, the mouse model verified *in vivo* the pathogenicity of the identified heterozygous *EHMT2* variant.

Discussion

Here, we report for the first time clinical and molecular correlates of *de novo* pathogenic heterozygous loss-of-function variants in the *EHMT2* gene in patients with clinical characteristics overlapping Kleefstra syndrome. Specifically, the probands reported here had recognizable facial dysmorphism (hypertelorism, synophrys, flat face, everted lips, protruding jaw) and cardiac anomalies (septal defects, pulmonal valve stenosis, aortal coarctation). Moreover, the probands exhibited additional phenotypic symptoms typical for Kleefstra syndrome including psychomotor developmental delay, intellectual disability, absent speech, structural brain anomalies, hypotonia, dental anomalies, behavioral and sleep difficulties. These symptoms were observed in all patients except P3, whose clinically milder phenotype was reduced to dysmorphic features and paroxysmal hyperkinetic movement disorder. The clinical similarity of these patients to those with Kleefstra syndrome caused *by EHMT1* haploinsufficiency was accompanied by similar episignature profiles in blood. Similar clustering of episignatures has been reported for epigenetic disorders with malfunction of the BAF complex due to various mutations in its individual subunits[33]. Despite the overlap with the Kleefstra syndrome episignatures, we identified a specific *EHMT2* episignature that can be used to reliably identify additional patients with candidate *EHMT2* variants as demonstrated for patient P7. Most probes in this episignature were hypomethylated compared to healthy controls and mapped to genes involved in development. These findings agree with the reported role of G9a in maintaining DNA methylation at germline genes during mouse embryogenesis[34]. Overall, our observations indicate that mutations in *EHMT1* and *EHMT2* affect the same epigenetic module, albeit with specific functional differences.

Consistent with the aberrant episignature profiles, we observed alteration in histone modifications and gene expression profiles in patient-derived cells. In line with our findings in fibroblasts,

coordinated downregulation of H3K9me2 and upregulation of H3K9ac levels have been described in mouse embryonic stem cells knock-out for G9a[24]. Moreover, similar global changes in histone modifications have been observed by mass spectrometry in HEK293 cells with *EHMT1* and *EHMT2* knock-down, including decreased H3K9 methylation and increased levels of methylated H3K79[35]. Patient-derived iPSCs exhibited only minor changes in histone modifications, limited to a subtle yet significant decrease in H3K9 methylation, which could be explained by low levels of this mark in iPSCs compared to fibroblasts (Supplementary Figure 5B). These differences are consistent with previous reports describing a significant reduction of the repressive marks H3K9me2/3 and H3K27me2/3 during the reprogramming of MEFs to pluripotency[36] and their re-establishment during differentiation to silence lineage-inappropriate gene expression[37]. Accordingly, changes in histone modifications were accompanied by upregulation of genes involved in alternative lineages such as nervous, muscle and skeletal system development, as well as genes involved in cell cycle and Rho signaling pathways. This upregulation is concordant with previously reported gene expression changes resulting from G9a dysfunction[38-42].

Kleefstra syndrome phenotype was recapitulated in the mouse model (*Ehmt2*$^{+/del_1076\text{-}1079}$) harboring a heterozygous *Ehmt2* variant equivalent to Del 1076-9 (identified in patient 1). *Ehmt2*$^{+/del_1076\text{-}1079}$ mice exhibited significantly impaired postnatal survival, growth and weight-gain retardation, decreased locomotor activity, and were mostly infertile. Cranial/facial abnormalities identified in these mice resembled the dysmorphic features observed in the patients and in specific details (e.g. “bent” nose) overlapped with pathologies earlier detected in the *Ehmt1*$^{+/-}$ mouse model[18].

All the identified *EHMT2* variants disturbed the catalytic site of the SET domain and resulted in a loss of methyltransferase activity. Interestingly, all patients presenting with classical symptoms of

Kleefstra syndrome carried *EHMT2* mutations that affected the histone H3 binding site. Conversely, patient P3 with the milder phenotype was heterozygous for the variant N1112S that affects the SAM binding pocket. In addition, the episignature of this patient clustered more closely with Kleefstra syndrome than with the other *EHMT2* patients. These observations suggest that the severity of the phenotype depends on whether the defective enzyme-substrate interactions affects the binding affinity towards histone H3 tail or SAM.

Given the functional interdependency between GLP and G9a, it is remarkable that, twenty years after the initial description of Kleefstra syndrome, no individuals with *EHMT2* alterations have been reported. Kleefstra syndrome is commonly caused by microdeletions or point mutations in *EHMT1* leading to decreased GLP protein levels[22,43]. In agreement with previous reports, we describe that GLP pathogenic mutants C1073Y and R1197W are structurally unstable and/or cause impaired dimerization[13,15], which is compatible with the mechanism of haploinsufficiency. In contrast, the G9a mutants reported here formed structurally stable dimers that were catalytically inactive. Moreover, *EHMT2* variants did not affect the levels of transcript and protein in blood and patient-derived cell-lines indicating that G9a mutants are stable in the patients' cells. These findings suggest that the pathogenic mechanism of *EHMT2* variants differs from that of *EHMT1*, which is consistent with previously published data showing complementarity of GLP and G9a enzymes but non-overlapping functions[44].

Population genomics also point out to a different mechanism of pathogenesis for both genes. The Genome Aggregation Database (gnomAD 4.1.0) shows marked differences in the probability of loss-of-function intolerance scores for *EHMT1* (LOEUF = 0,16) and *EHMT2* (LOEUF = 0,49) suggesting a high level of intolerance to loss of *EHMT1* function, in contrast to *EHMT2*. Similarly, the Database of Genomic Variants (DGV) shows marked differences in occurrence of genic

deletions in the healthy population that are sporadic for *EHMT1*, but frequent for *EHMT2* (Supplementary Figure 25). Consistently, the Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER), a repository for variants identified in patients with rare diseases, shows a high frequency of *EHMT1* deletions, but not that of *EHMT2* (Supplementary Figure 23). Additionally, gnomAD v4.1.0 documented numerous patients of variants (6 variants with 27 allele counts) positioned in methyltransferase active site that were specifically identified in *EHMT1* and not in *EHMT2*. Of particular note is the *EHMT1* variant N1200S, an equivalent to N1112S in EHMT2 patient described here, which was reported in 14 individuals in gnomAD v4.1.0. In summary, the population genomic data support a notion that *EHMT2* haploinsufficiency is not sufficient to cause disease, in contrast to *EHMT1*.

Additional evidence supporting a differential mechanism of pathogenesis for *EHMT1* and *EHMT2* is provided by the mouse model. *Ehmt1*$^{+/-}$ mice have been reported to exhibit significant changes in histone methylation and transcriptome profiles, leading to severe cognitive impairments[17]. In contrast, *Ehmt2*$^{+/-}$ mice display only mild alterations in epigenetic and transcriptomic profiles, resulting in minor phenotypic effects, according to two independent studies[19,24]. However, our mouse model (*Ehmt2*$^{+/del_1076\text{-}1079}$) displays severe phenotype abnormalities. This suggests that the reported SET-domain variants are more pathogenic than the heterozygous loss of *EHMT2*. Overall, our experimental results, population genetics data and phenotypes of the mouse model suggest that the described *EHMT2* pathogenic variants probably act via dominant-negative effects (**Figure 6**).

Mechanistically, differences in the pathogenic mechanisms of *EHMT1* and *EHMT2* can be explained by the different biochemical properties and cellular functions attributed to both enzymes. While mouse embryonic stem cells deficient in GLP show significantly reduced levels of G9a protein, G9a-deficient cells exhibit normal GLP levels[6]. These findings indicate that G9a protein

stability is highly dependent on the presence of GLP, while GLP stability is independent of G9a. Therefore, G9a's biological function seems to be exclusively exercised through the G9a–GLP heterodimer, whereas GLP might be capable of forming both homo- and heterodimeric complexes. In addition, each subunit of the GLP/G9a complex plays a distinctive role in governing the physiological levels of H3K9 methylation. Specifically, GLP contributes to the propagation of H3K9 methylation via recognition of H3K9me1/2 modified nucleosomes through its ankyrin-repeats domain (denoted as reader function)[44], whereas G9a executes the methylation reaction *per se* by its enzymatically active SET domain[25]. Consistently, a recent report evaluating 209 individuals with rare variants in *EHMT1* found that missense variants in the GLP ankyrin repeat domain resulting in structurally stable GLP protein yet impairing its reader function were associated with Kleefstra syndrome[15]. In addition, variants that disrupted the GLP methyltransferase activity without affecting its structural stability could not be correlated with Kleefstra episignatures and phenotype[15]. These data indicate that both normal structural stability and reader function of GLP together with G9a writer activity are necessary conditions for the proper function of the GLP/G9a complexes. It is tempting to speculate that pathogenic variants in *EHMT1* - but not in *EHMT2* - lead to haploinsufficiency caused by the reduced number of functional dimeric units of GLP/G9a. This reduction might be compensated by the formation of GLP homodimers in the case of limited G9a availability, which seems unlikely to apply in the reverse scenario.

In summary, we demonstrate that *EHMT2* missense variants located in the SET domain, which impair its catalytic function lead to clinical symptoms that genocopy Kleefstra syndrome. This is supported by Kleefstra syndrome-like epigenetic changes, aberrant biochemical behavior of the mutant proteins and recapitulation of Kleefstra syndrome phenotype in the mouse model. We also

provide additional evidence that *EHMT2* and *EHMT1* variants act through different pathogenetic mechanisms in the context of the GLP/G9a complex. We believe that our findings will further facilitate the identification of additional individuals affected by pathogenic *EHMT2* variants and promote further research of the physiological and pathogenetic aspects of GLP/G9a biology.

Methods

Identification of *EHMT2* variants

Variants in *EHMT2* were discovered by WES or WGS analyses. Detailed data analysis workflow for individual patients is summarized in Supplementary Material.

DNA Methylation Data Analysis

Data generation and processing

Methylation analysis was conducted using the clinically validated EpiSign assay, following previously established methods[28-30,45]. All computational analyses were performed using R statistical software[46] (version 4.4.1) along with relevant libraries. Raw intensity data files containing methylated and unmethylated signal values underwent quality control using the SeSAMe package[47]. Standard preprocessing of Illumina microarray data involves dye bias correction, background subtraction, and masking of low-quality probes[48]. Detection p-value-based probe masking is further applied using Infinium I out-of-band signal calibration, with a significance threshold set at $p = 0.05$.

Episignature discovery

A signature discovery analysis was conducted using six case samples (P1-P6) carrying variants in the SET domain of *EHMT2*. Prior to analysis, probes were excluded if they met any of the following criteria: located on the X or Y chromosomes, targeting CpG sites overlapping known

single nucleotide polymorphisms (SNPs), identified as cross-reactive, exhibiting high variability in previous Illumina DNA methylation array versions following manufacturing changes, or listed among EPICv1 probes that were omitted in the updated EPICv2 array. Matched control samples for the discovery analysis were selected using the 'matchIt' package[49], with matching criteria based on age, sex, and the array type used for methylation profiling. Samples from the EpiSign™ Knowledge Database (EKD) previously identified as contributing to batch effects, as well as those with more than 5% failed probes, were excluded during the matching process. A total of 54 matched controls were included in the analysis, yielding a match ratio of 1:9. Principal component analysis (PCA) was subsequently performed to assess data structure and identify potential outliers. Differential analysis was conducted on the discovery cohort using the 'limma' package[50] Methylation beta values were first transformed into M-values via a logit transformation. The data was then modeled using multivariate linear regression, where methylation levels served as predictors, case/control status as the response variable, and estimated blood cell compositions as covariates. To enhance statistical reliability, empirical Bayes moderation was applied, and adjusted p-values were calculated using the Benjamini-Hochberg method to control the false discovery rate. To identify probes that define the episignature, a multi-step ranking process was employed. Initially, probes were ranked based on a composite score derived from the absolute mean methylation difference and the negative logarithm of their adjusted p-value. The top subset of these probes was then selected using ROC curve analysis implemented via the 'caret' package[51]. Highly correlated probes were subsequently removed to reduce redundancy. The resulting probe lists were evaluated using unsupervised clustering, including Euclidean distance-based clustering with Ward's method and multidimensional scaling. Visualization was performed using the 'ggplot2'[52]

and 'gplots'[53] packages in R. The probe set that yielded the most distinct clustering was chosen to define the final signature.

The selected probes were utilized to construct a support vector machine (SVM) classification model using the e1071 package[54] as previously described[20,28,45,55]. For each sample, the classifier generates a Methylation Variant Pathogenicity (MVP) score ranging from 0 to 1. An MVP score approaching 1 indicates a methylation pattern closely resembling the EHMT2 episignature, whereas scores near 0 reflect a pattern similar to that of control samples. The SVM classifier was trained using the discovery case samples, matched controls, 75% of the remaining control samples, and 75% of the neurodevelopmental disorder samples from the EKD. The remaining 25% of control and neurodevelopmental disorder samples were reserved for model testing. SVM hyperparameters, including class weights, regularization parameters, and kernel functions were optimized via grid search.

To evaluate the reproducibility and robustness of the identified methylation profile, six rounds of leave-one-out cross-validation were performed on the discovery case samples. In each round, one of the six EHMT2 samples was excluded from probe selection, and the remaining five were used to train the model. The excluded sample was then used for testing and visualized using heatmaps and multidimensional scaling (MDS) plots. To provide an additional independent assessment, the seventh sample in the cohort was subsequently used as a test validation case.

Cell cultures

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 in 75 cm2 culture flasks at 37°C with 5% CO2 and 95% humidity in DMEM containing 10% fetal bovine serum.

Reprogramming to pluripotency

Cell lines were generated from cultured fibroblasts or PBMCs using the CytoTune iPS 2.0 Sendai Reprogramming Kit (Invitrogen). Growing colonies were manually selected according to morphology and passed for at least seven passages before assessing the presence of viral mRNAs.

Mainteinance of hiPSCs

Established cell lines were grown in Essential E8 Media (Gibco) in Vitronectin (Gibco) coated plates. Cells were split twice a week using ReLeSR (Stem Cell Technologies) at a ratio of 1:4-1:10 in 5 uM Rock inhibitor Y-27632 (Tocris) for 1-3 hours before changing to fresh media.

RNA preparation and real-time PCR

RNA was extracted using RNeasy kit (Qiagen) and cDNA prepared using the Maxima First Strand cDNA Synthesis Kit. Oligonucleotides for detecting the expression of transgenes were ones recommended in the reprogramming kit. Clones free from viral expression were selected for further characterization. Oligonucleotides used to detect the expression of pluripotency genes and reprograming factors for have been previously described[56].

Immunofluorescence staining

Cells were fixed using 4% PFA for 20 min at room temperature and washed with phosphate buffered saline (PBS) three times, followed by permeabilization and blocking with 0.25 % Triton X-100 and 5% normal horse serum at room temperature for one hour. Subsequently, cells are incubated with primary antibodies (anti-Sox2 AB5603 Sigma and PA1-094 Fisher Scientific (1:400), anti-NANOG AF1997-SP R&D (1:50), anti-Oct3/4 (C-10) sc-5279 Santa Cruz Biotechnology, anti-TRA1-60 sc-21705 Santa Cruz Biotechnology, anti-Alpha-1-fetoprotein

A000829-2 Agilent Dako (1:500), Alpha-Smooth Muscle A5228 Sigma (1:500), TUBB3 Biolegend 801202(1:500)) overnight at 4°C. The next day, cells were washed with PBS and incubated with secondary antibodies for 1 h at room temperature in the dark and stained with DAPI. Images were captured using a laser confocal microscope (Stellaris 8, Leica Microsystems).

Embryoid body (EB) differentiation

For in vitro differentiation, cells were trypsinized using ReLeSR and 50,000 cells per well were seeded into low attachment 96-well v-bottom plates in Essential 8 media and 5 uM Rock inhibitor Y-27632 (Tocris). After two days, EBs were moved from the plate to a 10 cm low attachment plate with fresh Essential 8 media. The next day they were transferred to gelatin-coated dishes and cultured in differentiation media consisting on DMEM media supplemented with 20% fetal bovine serum up to twenty days.

Karyotype and genotype

The generated iPSCs lines were free from genetic abnormalities according to the hPSC Genetic Analysis Kit from Stem Cell Technology and karyotype from 20 metaphases. To confirm the presence of the mutations, genomic DNA was extracted with the DNeasy Blood and Tissue Kit from Qiagen and mutations verified by amplification and sanger sequencing. Oligonucleotides used for amplifying and sequencing the *EHMT1* gene were primer forward: 5'-CTTCTTCTCTGTGGGGCGAG-3' primer reverse: 5'-CACCATAAGCATCAGCATCAGC-3' and the *EHMT2* gene primer forward 5'TCCATATCGCCCATTCCTGC and primer reverse 5' ATGAAGCGGCTGATGTTGCC.

RNA-seq analysis

Total RNA was extracted from fibroblasts and iPSCs 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 STAR[57]. Gene counts were calculated using HTseq-count[58] and differential expression between patients and healthy fibroblast was interrogated using DEseq[59]. Differentially expressed genes were considered at an adjusted p-value <0.05. PCA plots were generated using the DESeq2 package in R. Enrichment analysis in differentially expressed genes was carried out using DAVID[60]. Scatterplots showing correlations were generated with ggplot2[52] in Galaxy. Pearson's correlation and bubble plots were calculated using SRPlot[61]. Heatmaps were generated using Morpheus (https://software.broadinstitute.org/morpheus/).

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 described[62]. 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 Skyline[63] and represents the percent of each modification within the total pool of that amino acid residue. A total of 92 modification states were quantified including unmodified ones.

Western blot analysis of fibroblasts and iPSCs

Cell extracts were prepared in lysis buffer (0.9x CelLytic MT reagent supplemented with 200 mM NaCl, 1.5 mM magnesium chloride, 20 mM EDTA and protease inhibitor cocktail) with gentle shaking for 30 minutes at 4 °C. Crude cell extracts were centrifuged for 16,000 G for 10 minutes at 4 °C. Supernatants were collected and protein concentration was determined using Bradford assay with bovine serum albumin as standard. Subsequently, extracts were analyzed using SDS-PAGE followed by Western blotting (with 20 µg and 10 µg of total protein loaded in case of

fibroblasts and iPSCs, respectively) with immunodectection against G9a (rabbit anti-G9a, Sigma, G 6919) and GAPDH (mouse Anti-GAPDH, Sigma G8795) as a reference.

Functional studies

In-silico structural mapping

Mutations were mapped into crystal structure of G9a SET-domain bound to the N-terminal peptide of histone H3 and S-adenosylhomocysteine (PDB ID 5JJ0)[64]. Structural models were visualized and assessed by Pymol Viewer. Representative images were prepared using UCSF ChimeraX.

Preparation of recombinant proteins

The methyltransferase domains of G9a (aa residues 913-1193) and GLP (aa 1001-1266) with N-terminal hexa-histidine tag followed by recognition site for TEV protease was expressed in *E.coli* (BL21-CodonPlus (DE3)-RIPL strain) using pET28a plasmid. Bacterial culture was cultivated in Terrific Broth medium supplemented by 0.1 mM zinc chloride. When cells reached OD600 of 1.0, protein expression was induced by an addition of 0.5 mM IPTG at 15 °C for 24 hours.

After cell lysis by sonication in lysis buffer (G9a: 20 mM potassium phosphate (pH 7.5) containing 500 mM NaCl, 20 mM imidazole, 10 mM beta-mercaptoethanol, 5% glycerol and SigmaFast inhibitor cocktail; GLP: 50 mM Tris-HCl (pH 7.5) containing 500 mM NaCl, 20 mM imidazole, 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP), 10% glycerol and SigmaFast inhibitor cocktail), proteins were purified by affinity chromatography using an imidazole gradient from 20 to 500 mM over 20 column volumes with Profinity IMAC column (Biorad) followed by cleavage of histidine tag by TEV-protease overnight at 4 °C (protein:protease w/w ratio of 50:1). In the next step, proteins were separated using size-exclusion chromatography (Enrich SEC 650 column, BioRad) in SEC buffer (G9a: 50 mM potassium phosphate (pH 7.5) containing 100 mM NaCl, 2 mM dithiotreitol, 20 mM imidazole and 5 % glycerol; GLP: 50 mM Tris-HCl (pH 7.5) containing 100

mM NaCl, 0.5 mM TCEP and 10% glycerol) and a second round of affinity chromatography to remove TEV protease containing uncleavable histidine tag. Finally, proteins were dialyzed against storage buffer (G9a: 50 mM potassium phosphate (pH 7.0) containing 100 mM NaCl, 1 mM TCEP and 5 % glycerol; GLP: 25 mM Tris-HCl (pH 8.0) containing 50 mM NaCl, 0.5 mM TCEP and 10 % glycerol), and stored in -80C. Aminoacid sequence of the obtained proteins was verified using on-line pepsin digestion followed by mass spectrometric analysis (details described in[65]).

Solubility test of GLP mutants expressed in E.coli

Bacterial pellets were resuspended in lysate buffer (as described above) and cells were desintegrated by sonication or by an addition of BugBuster reagent, lysozyme (0.5 mg/ml) and benzonase (25 mU/µl) followed by a gentle shaking at room temperature for 20 minutes. Soluble and insoluble fractions were separated by centrifugation at 4 °C at 16,000 G for 20 minutes. Amount of soluble GLP proteins in supernatant was determined using western blotting (with 10 µg of total protein loaded) followed by immunodetection using Anti-6X His tag primary antibody (ab9108 from Abcam) and Rabbit IgG (H+L) Secondary Antibody (31460 from ThermoFischer Scientific). Insoluble fraction was washed for by BugBuster solution followed by centrifugation 4 °C at 5,000 G for 15 minutes and analysed by SDS-PAGE with Coomassie-Blue staining.

Biolumiscence-based methyltransferase assay

Enzyme activity of G9a and GLP variants was assessed using MTase-Glo Methyltransferase Assay (Promega) according to manufacturer instructions. The activity assay was performed with a peptide derived from the N-terminal part of histone H3 (aa residues 1-20; manufactured by Merck Life Science) or mononucleosome (prepared according to[66]). Concentration of the histone H3 peptide and S-adenosylmethionine was kept at 10 µM. Methyltransferase reactions were performed at room temperature for 30 minutes with 5 nM G9a or GLP enzymes. In assay with

mononucleosome, concentration was kept at 0.5 μM and 5 μM for mononucleosome and SAM, respectively. Reaction was executed at 37 °C for 1 hour with 0.5 μM enzyme. Obtained signals were recorded using Spark plate reader (TECAN). Quantification was performed using external calibration standard of S-adenosyl homocysteine solution. Kinetic data were fitted according to Michaelis-Menten equation using OriginPro.

MALDI-TOF MS-based methyltransferase assay

Reactions were performed in 50 mM HEPES (pH 7.9) containing 0.5 mM DTT, 0.25 mM PMSF and 2 mM magnesium chloride with 25 μM histone H3 peptide, 50 μM S-adenosylmethionine and G9a or GLP enzyme (25 or 100 nM) for 30 min at 30 °C. The reaction was quenched by an addition 0.1 % TFA. The resulted mixture was 10-fold diluted in 0.1 % TFA, mixed with saturated solution of α-cyano-4-hydroxycinnamic acid in a ratio of 1:1 (v/v) and deposited to MALDI target plate. Samples were measured using Autoflex Speed (Bruker Daltonics) in a reflecton positive mode with parameters as follows: mass range 700-3500 Da, frequency 1000 Hz and ion delay extraction 130 ns. Instrument was externally calibrated using Peptide calibration standard mix II. Spectra were processed in FlexAnalysis 3.4.

Differential scanning fluorimetry

Proteins (3 μM) dissolved in assay buffer (G9a: 50 mM potassium phosphate (pH 7.0) containing 100 mM NaCl, 1 mM TCEP; GLP: 25mM Tris-HCl (pH 8.0) containing 50 mM NaCl and 0.5 mM TCEP) were analyzed in a free state or bound to histone H3 peptide (50 or 500 μM) or S-adenosylmethinine (50 or 500 μM). After preincubation of the G9a proteins with ligands at room temperature for 10 minutes, Sypro-Orange dye was added to the samples which were subsequently subjected to melting analysis using CFX96 Real-Time System (BioRad). The proteins were heated from 20 °C to 90 °C in increments of 0.5 °C and with 1-minute hold intervals. The signal was

monitored using fluorescence detection (excitation wavelength of 470 nm and emission wavelength of 570 nm). The melting temperatures (Tm) of the proteins were determined as minima from first derivative curves.

Native mass spectrometry

In order to examine the protein dimerization and substrate binding, we performed native ESI-MS analysis on Waters Synapt G2Si instrument. Analysed G9a variants were first transferred into 150 mM ammonium acetate pH 7.5 containing 0.5 mM DTT by two cycles of spin gel filtration (Micro Bio-Spin P-6 Gel columns, 6-kDa cut off; BioRad). The samples (5 µM) were mixed with SAM or histone H3 peptide in equimolar ratio and incubated for 10 min at 4 °C. The protein complexes were electro-sprayed from in-house prepared quartz spraying tips[67] with internal platinum wire connected to a high voltage power source kept at 0.6 kV. Data acquisition was performed in positive ionization sensitivity mode. The trap collision energy was optimized at 40 V with 6 ml/min argon flow, while maintaining 80 V sampling cone voltage, 0 V source offset, and 80 °C source temperature. Quadrupole was operated in broad transmission mode up to 8.000 m/z. Acquired data were summed over 100 scans, smoothed and externally mass recalibrated on caesium iodide clusters in MassLynx 4.1 (Waters Inc.), followed by mass deconvolution in UniDec v6.0.3[68].

Mouse model

Generation of the knock-in mouse model

To model the effects of the variant (c.3225_3236del, p.Glu1076_Val1079del) identified in patient 1, the *Ehmt2* c.3385_G3396del (p.E1129_V1132del, [C57BL/6NCrl-Ehmt2^{em1Ccpcz}/Ph) mouse model (further referenced in the manuscript as *Ehmt2*$^{+/del_1076\text{-}1079}$) was generated on a C57BL/6NCrl background (Charles River Laboratories) by targeting exon 25

(ENSMUSE00001282962) (Supplementary Figure 14A). *Ehmt2* codons E1129, A1130, D1131 and V1132, that are complementary to human *EHMT2* codons 1076-1079, were deleted using CRISPR/Cas9 technology in combination with single-stranded oligonucleotides (ssODN) template: 5'-GGTGCTGCATCCCACCTTGTTATCTAAATCGAAGAGGTAAGAATCATCCTCTCTCGC ATCAGAGATCAGCTCTCCTACGTACCTGTTGGACGGAAACTTCAGTTCTGCATGAAC AGTGAGC -3' as described by[69]. The guide RNAs (gRNAs) of highest score and specificity were designed using http://crispor.tefor.net/.The following guide was selected: gRNA-AGAGCTGATCTCTGATGCCG. The gRNAs (100ng/µl, Integrated DNA Technologies) were assembled into a ribonucleoprotein (RNP) complex with Cas9 protein (500ng/µl; 1081058, 1072532, Integrated DNA Technologies), electroporated into 1-cell zygotes, and transferred into ICR (CD1) pseudopregnant foster females. Putative founders were analyzed by PCR and sequencing. A founder carrying a E1129-V1132del mutation also harbors SnaBI (Eco105I) restriction enzyme site enabling direct detection of the mutation by the restriction enzyme digest. Genotyping was performed by PCR with forward (F) 5′- TTCCCTCAACTTCTCAGACACC -3′ and reverse (R) 5'- ACCCGGACAGGGATGATGTT -3' primers. The F and R product is 775 bp, *SnaBI* digestion of the E1129-D1132del homozygote mice PCR product results in 439bp and 324bp fragments (Supplementary Figure 14B, C). Selected founder was confirmed by sequencing and bred with C57BL/6NCrl wild-type to confirm germ-line transmission of the target deletion. The only male founder carrying E1129-V1132del gave rise to infertile females and sub-viable males, often dying in 10 days after birth. Due to this limitation, the founder was archived in form of frozen sperm. To propagate this line, in vitro fertilization was performed as descrived[70] and viable embryos we implanted into ICR (CD1) surrogate females (Charles River Laboratories).

Animals born after IVF were genotyped and sequenced (Sanger sequencing) within 5 days after birth to detect and monitor positive mice. Bulk sequencing data were further analyzed using DECODR.org online software. The sperm form founder male was sequenced using NGS sequencing to detect all possible *Ehmt2* gene variants in germ line.

SHIRPA scoring

Standardized SHIRPA scoring scale[71] was used to generally asses the phenotype of the *Ehmt2*$^{+/del_1076\text{-}1079}$ and *Ehmt2*$^{+/+}$ mice.

In-vivo microCT mouse scanning

All mice were anesthetized via intramuscular administration of 20% Zoletil-Xylazine (Virbac). In-vivo scans of mice were performed using the SkyScan 1278 (Bruker, Belgium) as Continuous 7 connected scans. Scanning parameters included a voxel size of 50 µm, a 0.5 mm aluminum filter, 180° rotation and no averaging. Source voltage was set to 54 kV and source current to 909 µA. Data reconstruction was carried out using NRecon software, version 2.2.0.6 (Bruker, Belgium) with the following parameters: ring artifact correction at 3, beam hardening at 28 %, threshold for defect pixel mask at 5 %, no smoothing and intensity range set from 0.0025 to 0.10 AU[72]. CTvox software, version 3.3.0.0 (Bruker, Belgium) was used for data visualization. Craniometric measurements (Supplementary Figure 16) were performed in DataViewer, version 1.5.2 (Bruker, Belgium).

Open Field test

The activity of the animals in a novel environment and the level of anxiety displayed were evaluated in the open field test[73]. The area of the open field was a square of 42 × 42 cm uniformly illuminated with a light intensity of 200 lux in the center of the field. The testing arena was virtually divided into periphery and center zones, where the center zone constituted 38% of

the whole arena. Each mouse was placed in the corner of the arena for a 20 min period of free maze exploration. The time spent in each zone, the distance travelled, and other indices were automatically computed based on video recordings (Viewer software 3.0.1.452, Biobserve GmbH, Germany).

Grip strength

A grip strength meter from Bioseb was used to determine the maximal peak force (g) developed by the mouse forelimbs as well as all four limbs when the examiner tried to pull it out of a specially designed grid for mice (Bioseb, BIO-GS3 & BIO-GRIPGS). The maximal effort (g) was used as absolute force (g) and was corrected for body mass (g/g). The examiner was blinded, and a total of 3 measurements were applied to each mouse with a 10-15 min interval between measurements. The average of the 3 values was used for subsequent calculations.

Rotarod

The rotarod system was used to assess the sense of balance, motor learning, and motor coordination in the animals. Three trials per day for each mouse were recorded on a rod with an accelerating speed of rotation (4–40 rpm/5 min, RotaRod, TSE Systems GmhB, Berlin, Germany) during two consecutive days. The average latency to fall was determined from three trials with 15 min intertrial intervals (ITI).

Data availability

Because of the rarity of the disorder, individual participant data beyond that reported in this article cannot be shared, to safeguard patient privacy. Other experimental data are available from the corresponding authors upon reasonable request.

Acknowledgments and Funding

This work was supported from the project MULTIOMICS_CZ (Programme Johannes Amos Comenius, Ministry of Education, Youth and Sports of the Czech Republic,//ID Project CZ.02.01.01/00/23_020/0008540) – Co-funded by the European Union. L.N., A.H., and J.S. were supported by the project National Institute for Neurological Research (Programme EXCELES, ID Project No. LX22NPO5107) - Funded by the European Union – Next Generation EU and by institutional program UNCE/MED/007 of Charles University in Prague and grants from the Spanish Ministry of Science and Innovation PID2021-128087OB-I00 by MCIN /AEI /10.13039/501100011033 / FEDER, UE to M.J.B, Fundación Inocente grant number FII2024-125 and grant AESI PT20CIII/00009 (ISCIII Platform of Biobanks and Biomodels PT-20/23) from the Spanish National Institute of Health Carlos III. Funding was also partially provided by the Genome Canada and the Ontario Genomics Institute Genomics Applications Partnership Program Grant (OGI-188). A.H. was funded from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101003406 (HIPPOSTRUCT). A.K. acknowledges funding by the European Union's ERA fellowship (2D-TOPMASS, grant 101090276). L.N., J.S. and S.K. were supported by grants NU23-07-00281 and NW24-04-00067 from the Ministry of Health of the Czech Republic. L.N., V.S. and S.K thank the National Center for Medical Genomics (LM2023067) for WES analyses. We further acknowledge Solve-RD managed by ERN ITHACA as well as CMS-Biocev ("Biophysical techniques, Crystallization, Diffraction, Structural mass spectrometry") of CIISB, Instruct-CZ Centre, supported by MEYS CR (LM2023042) and European Regional Development Fund-Project „UP CIISB" (No. CZ.02.1.01/0.0/0.0/18_046/0015974). 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 and the Advanced Optical Microscopy Units at ISCIII, the patients' association Kleefstra España, the BioNER and the SpainUDP consortium. We acknowledge Lidia Lopez-Jimenez for her help in generating the iPSC lines. We would like to thank the patients' families for their willingness to participate in this study.

Competing interests

B.S. is a shareholder in Episign, Inc. The other authors declare no conflicts of interest

Ethics Declaration

Informed consent for genetic analyses was obtained for all individuals, and genetic studies were performed as approved by the review boards of corresponding institutes. The patient's parents provided written informed consent for the participation in the study, clinical data and specimen collection, genetic analysis and publication of relevant findings. All data were de-identified. The study adheres to the principles set out in the Declaration of Helsinki. The generation of human iPSCs was approved by the ethical committee of Instituto de Salud Carlos III CEI PI 03_2022. Informed consents for sample donation and iPSCs generation were signed by the legal guardians.

Figure legends

Figure 1. EpiSign (DNA methylation) analysis of peripheral blood from patients with *EHMT2* variants.

Unsupervised clustering of the discovery cohort and assessment of *EHMT2* variants using the identified methylation signature, visualized via (A) heatmap and (B) multidimensional scaling

(MDS) plot. Discovery samples from patients with *EHMT2* variants are shown in red, matched controls in blue, *EHMT2* test sample in orange, and Kleefstra (*EHMT1*) test samples in purple. (C) Methylation Variant Pathogenicity (MVP) scores from the SVM classifier for EpiSign™ Knowledge Database, *EHMT2*, and Kleefstra test samples, averaged over four-fold cross-validation (gray). Higher MVP scores indicate greater similarity to the discovery case methylation profile (blue).

Figure 2. Epigenetic and transcriptomic analyses of cell lines derived from a healthy control, a Kleefstra Syndrome patient (KSP) and patients P1 and P2 carrying *EHMT2* variants.

(A) Relative abundance of the indicated histone H3K9 modifications quantified by mass spectrometry in fibroblasts cultures (Fib) and iPSCs. (B) Correlation of changes in 92 histone modification states detected by mass spectrometry between KSP or P2 and control fibroblast. (C). Pearsons correlation coefficient of gene expression fold change for each patient compared to control fibroblast (Fib) or iPSCs. (D) DAVID enrichments of GOBP, GOCC and Reactome (RE) terms in genes commonly upregulated in KSP and P2 fibroblasts (Fib) and genes commonly upregulated in KSP, P1 and P2 iPSCs.

Figure 3. Sequence and structural positioning of mutations in the G9a protein

(A) Aminoacid sequence alignment shows highly conserved residues in the catalytic pocket of the methyltransferase domain. Positions of the mutations are highlighted by red squares.

(B) Structural mapping demonstrates localization of the mutations in binding pockets for histone H3 tail and SAM in the G9a catalytic site. Mapping was performed using crystal structure of SET-domain of G9a in complex with SAM and histone H3 peptide (PDB ID 5JIN). The variant aminoacid residues are highlighted as red sticks; G9a substrates - histone H3 and SAM - are depicted in green color.

Figure 4. Functional characteristics of G9a variants assessed with recombinant proteins (A) Methylation activity of the G9a mutants with histone H3 peptide (blue) and mononucleosome (red) as substrates studied by quantitative bioluminescence-based assay (Mtase-Glo by Promega). The values represent the means of relative activity to the wild-type enzyme with standard deviations. The symbol "#" indicates "activity not detected". (B) Quantitative assessment for dimer formation of G9a and GLP mutants using nESI-MS. Points represent relative amounts of dimers related to the wild-type proteins. Dimerization is severely affected in GLP but not in G9a mutants. (C) Thermostability of the G9a mutants in their free and substrate-bound states studied by differential scanning fluorimetry in at least two independent measurements. Melting temperatures of free G9a proteins are represented as mean values with standards deviations. Dashed line indicates melting temperature of the wild-type protein. The p-values indicate variants with significantly different changes from the wild-type. whereas "n.s" means changes were not significant, as determined by a two-tailed Student's t-test ($p<0.01$).

The stabilization of G9a proteins upon the addition of individual substrates is demonstrated by a difference in melting temperature in comparison to the free protein. The values represent the means with standard deviations. (D) Binding affinity of the G9a variants towards individual substrates assessed by native mass spectrometry. The protein and substrate concentrations were maintained at 5 µM. Representative deconvoluted nMS spectra revealed specific mass shifts that corresponded G9a complexes with SAM (+398 Da) or histone H3 peptide (+2100 Da) bound as shown by highlighted areas with relative abundance.

Figure 5. Key phenotypic features identified in *Ehmt2*$^{+/del_1076-1079}$ mice

(A) Survival data (5 months) comparing *Ehmt2*$^{+/del_1076-1079}$ *and Ehmt2*$^{+/+}$ mice are shown for each sex. p0/n corresponds to the starting number of animals included into the study. Black dotted line

highlights 50 % survival rate. Statistical significance was calculated using Log-rank (Mantel-Cox) test. (B) Weight curves (5 months) for all animals of both genotypes/sexes included in the survival study. Weights were significantly different ($p<0.05$ by a two-tailed Student's t-test) between the *Ehmt2*$^{+/del_1076\text{-}1079}$ and *Ehmt2*$^{+/+}$mice of each sex at all postnatal time points. Refer to Supplementary Figure 16 for detailed weight comparisons at p1-2 and p21 and weight curves of animals surviving the entire 5 months period. (C) MicroCT analyses in *Ehmt2*$^{+/del_1076\text{-}1079}$ *and Ehmt2*$^{+/+}$ mice. Lateral views of the skulls with (in brown) and without soft facial tissues, horizontal views of the nasal cavity and neurocranium, and 3D reconstruction of the palate (bottom view). Nasalia (grey arrows) are shorter in *Ehmt2*$^{+/del_1076\text{-}1079}$ mice. White arrows highlight the "bent" nose phenotype in *Ehmt2*$^{+/del_1076\text{-}1079}$ animals. For the proportion and range of the shorter nasalia and the "bent nose" abnormalities see Supplementary Figures 18-19. (D) Quantitative microCT craniometric data comparisons in *Ehmt2*$^{+/del_1076\text{-}1079}$ *and Ehmt2*$^{+/+}$ mice. *Ehmt2*$^{+/del_1076\text{-}1079}$ have significantly shorter nasalia and disproportionately shorter neurocranium, which results in a significantly lower ratio of these two values in comparison to *Ehmt2*$^{+/+}$ mice. Palate length is shorter in *Ehmt2*$^{+/del_1076\text{-}1079}$ while the palate width (molar1: molar1 – M1:M1 distance) remains comparable ($p=0{,}083$) to *Ehmt2*$^{+/+}$ mice. As a result, the palate (calculated as a square size of the latter two values) is significantly smaller in *Ehmt2*$^{+/del_1076\text{-}1079}$ animals. Statistical significance was calculated using two-tailed t-test. Values are color/shape-coded based on sex. See Supplementary Figure 17 for technical and acquisition details of the presented craniometric measurements.B and D show average values ± SEM.

Figure 6.

Scheme showing various scenarios for partially dysfunctional GLP/G9a heterodimer, which leads to physiological (upper part) or aberrant (lower part) H3K9 methylation.

References:

1. Greer, E.L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet 13, 343-57 (2012).
2. Younesian, S., Yousefi, A.M., Momeny, M., Ghaffari, S.H. & Bashash, D. The DNA Methylation in Neurological Diseases. Cells 11(2022).
3. Al Ojaimi, M. et al. Disorders of histone methylation: Molecular basis and clinical syndromes. Clin Genet 102, 169-181 (2022).
4. Jenuwein, T., Laible, G., Dorn, R. & Reuter, G. SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cell Mol Life Sci 54, 80-93 (1998).
5. Wei, H. & Zhou, M.M. Dimerization of a viral SET protein endows its function. Proc Natl Acad Sci U S A 107, 18433-8 (2010).
6. Tachibana, M. et al. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes Dev 19, 815-26 (2005).
7. Sanchez, N.A., Kallweit, L.M., Trnka, M.J., Clemmer, C.L. & Al-Sady, B. Heterodimerization of H3K9 histone methyltransferases G9a and GLP activates methyl reading and writing capabilities. J Biol Chem 297, 101276 (2021).
8. Demond, H. et al. Multi-omics analyses demonstrate a critical role for EHMT1 methyltransferase in transcriptional repression during oogenesis. Genome Res 33, 18-31 (2023).
9. Ohno, H., Shinoda, K., Ohyama, K., Sharp, L.Z. & Kajimura, S. EHMT1 controls brown adipose cell fate and thermogenesis through the PRDM16 complex. Nature 504, 163-7 (2013).
10. Ciceri, G. et al. An epigenetic barrier sets the timing of human neuronal maturation. Nature 626, 881-890 (2024).
11. Kleefstra, T. et al. Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. Am J Hum Genet 79, 370-7 (2006).
12. Kleefstra, T. et al. 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 (2009).
13. Yamada, A., Shimura, C. & Shinkai, Y. Biochemical validation of EHMT1 missense mutations in Kleefstra syndrome. J Hum Genet 63, 555-562 (2018).
14. Blackburn, P.R. et al. A Novel Kleefstra Syndrome-associated Variant That Affects the Conserved TPL. J Biol Chem 292, 3866-3876 (2017).
15. Rots, D. et al. Comprehensive EHMT1 variants analysis broadens genotype-phenotype associations and molecular mechanisms in Kleefstra syndrome. Am J Hum Genet 111, 1605-1625 (2024).
16. Balemans, M.C. et al. Reduced exploration, increased anxiety, and altered social behavior: Autistic-like features of euchromatin histone methyltransferase 1 heterozygous knockout mice. Behav Brain Res 208, 47-55 (2010).

17. Balemans, M.C. et al. Hippocampal dysfunction in the Euchromatin histone methyltransferase 1 heterozygous knockout mouse model for Kleefstra syndrome. Hum Mol Genet 22, 852-66 (2013).
18. Balemans, M.C. et al. Reduced Euchromatin histone methyltransferase 1 causes developmental delay, hypotonia, and cranial abnormalities associated with increased bone gene expression in Kleefstra syndrome mice. Dev Biol 386, 395-407 (2014).
19. Iacono, G. et al. Increased H3K9 methylation and impaired expression of Protocadherins are associated with the cognitive dysfunctions of the Kleefstra syndrome. Nucleic Acids Res 46, 4950-4965 (2018).
20. Aref-Eshghi, E. et al. Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. Am J Hum Genet 106, 356-370 (2020).
21. Carvalho, L.M.L. et al. EHMT2 as a Candidate Gene for an Autosomal Recessive Neurodevelopmental Syndrome. Mol Neurobiol (2024).
22. Frega, M. et al. Neuronal network dysfunction in a model for Kleefstra syndrome mediated by enhanced NMDAR signaling. Nat Commun 10, 4928 (2019).
23. Tachibana, M., Sugimoto, K., Fukushima, T. & Shinkai, Y. 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. J Biol Chem 276, 25309-17 (2001).
24. Tachibana, M. et al. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev 16, 1779-91 (2002).
25. Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H. & Shinkai, Y. G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. Embo j 27, 2681-90 (2008).
26. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 36, 928-30 (2015).
27. Laurie, S. et al. The RD-Connect Genome-Phenome Analysis Platform: Accelerating diagnosis, research, and gene discovery for rare diseases. Hum Mutat 43, 717-733 (2022).
28. Levy, M.A. et al. Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. HGG Adv 3, 100075 (2022).
29. Sadikovic, B. et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. Genet Med 23, 1065-1074 (2021).
30. Kerkhof, J. et al. Diagnostic utility and reporting recommendations for clinical DNA methylation episignature testing in genetically undiagnosed rare diseases. Genet Med 26, 101075 (2024).
31. Britt, H.M. & Robinson, C.V. Traversing the drug discovery landscape using native mass spectrometry. Curr Opin Struct Biol 91, 102993 (2025).
32. Balan, S. et al. Exon resequencing of H3K9 methyltransferase complex genes, EHMT1, EHTM2 and WIZ, in Japanese autism subjects. Mol Autism 5, 49 (2014).

33. Aref-Eshghi, E. et al. BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin-Siris and Nicolaides-Baraitser syndromes. Nat Commun 9, 4885 (2018).
34. Auclair, G. et al. EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. Genome Res 26, 192-202 (2016).
35. Plazas-Mayorca, M.D. et al. Quantitative proteomics reveals direct and indirect alterations in the histone code following methyltransferase knockdown. Mol Biosyst 6, 1719-29 (2010).
36. Sridharan, R. et al. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nat Cell Biol 15, 872-82 (2013).
37. Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A. & Feinberg, A.P. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat Genet 41, 246-50 (2009).
38. Roopra, A., Qazi, R., Schoenike, B., Daley, T.J. & Morrison, J.F. Localized domains of G9a-mediated histone methylation are required for silencing of neuronal genes. Mol Cell 14, 727-38 (2004).
39. Maze, I. et al. G9a influences neuronal subtype specification in striatum. Nat Neurosci 17, 533-9 (2014).
40. Schaefer, A. et al. Control of cognition and adaptive behavior by the GLP/G9a epigenetic suppressor complex. Neuron 64, 678-91 (2009).
41. Katoh, K., Yamazaki, R., Onishi, A., Sanuki, R. & Furukawa, T. G9a histone methyltransferase activity in retinal progenitors is essential for proper differentiation and survival of mouse retinal cells. J Neurosci 32, 17658-70 (2012).
42. Wilson, C. et al. The Histone Methyltransferase G9a Controls Axon Growth by Targeting the RhoA Signaling Pathway. Cell Rep 31, 107639 (2020).
43. Yamada, A. et al. Derepression of inflammation-related genes link to microglia activation and neural maturation defect in a mouse model of Kleefstra syndrome. iScience 24, 102741 (2021).
44. Liu, N. et al. Recognition of H3K9 methylation by GLP is required for efficient establishment of H3K9 methylation, rapid target gene repression, and mouse viability. Genes Dev 29, 379-93 (2015).
45. Aref-Eshghi, E. et al. Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. Am J Hum Genet 104, 685-700 (2019).
46. R Core Team. R: A Language and Environment for Statistical Computing. Published online 2022. https://www.r-project.org/.
47. Zhou, W., Triche, T.J., Laird, P.W. & Shen, H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. Nucleic Acids Res 46, e123 (2018).
48. Zhou, W., Laird, P.W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res 45, e22 (2017).

49. Ho, D. E., Imai, K., King, G. & Stuart, E. A. MatchIt : Nonparametric Preprocessing for Parametric Causal Inference. J Stat Softw 42, (2011).
50. Ritchie, M.E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47 (2015).
51. Kuhn, M. Building Predictive Models in R Using the caret Package. J Stat Softw 28, (2008).
52. Wickham, Hadley. Ggplot2 : Elegant Graphics for Data Analysis. (Springer, 2016).
53. Gregory R. Warnes, B. B. L. B. R. G. W. H. A. L. T. L. M. M. A. M. S. M. M. S. B. V. T. G. Gplots: Various R programming tools for plotting data. R package version 3.2.0 (2015).
54. D. MEYER, E. D. K. H. A. W. F. L. E. D. A. W. e1071: misc functions of the department of statistics, probability theory group (Formerly: E1071), TU Wien. R package version 1.7-16 (2018).
55. Aref-Eshghi, E. et al. Genomic DNA Methylation Signatures Enable Concurrent Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental Syndromes. The American Journal of Human Genetics 102, 156–174 (2018).
56. Aasen, T. et al. Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. Nat Biotechnol 26, 1276-84 (2008).
57. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21 (2013).
58. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166-9 (2015).
59. Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biol 11, R106 (2010).
60. Sherman, B.T. et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res (2022).
61. Tang, D. et al. SRplot: A free online platform for data visualization and graphing. PLoS One 18, e0294236 (2023).
62. Garcia, B.A. et al. Chemical derivatization of histones for facilitated analysis by mass spectrometry. Nat Protoc 2, 933-8 (2007).
63. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966-8 (2010).
64. Jayaram, H. et al. 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, 6182-7 (2016).
65. Pacheco-Garcia, J.L. et al. Structural basis of the pleiotropic and specific phenotypic consequences of missense mutations in the multifunctional NAD(P)H:quinone oxidoreductase 1 and their pharmacological rescue. Redox Biol 46, 102112 (2021).
66. Luger, K., Rechsteiner, T.J. & Richmond, T.J. Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol Biol 119, 1-16 (1999).

67. Frecentese, F. et al. Synthesis and in Vitro Screening of New Series of 2,6-Dipeptidyl-anthraquinones: Influence of Side Chain Length on HIV-1 Nucleocapsid Inhibitors. J Med Chem 59, 1914-24 (2016).
68. Marty, M.T. et al. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Anal Chem 87, 4370-6 (2015).
69. Inui, M. et al. Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. Sci Rep 4, 5396 (2014).
70. Takeo, T. & Nakagata, N. Combination medium of cryoprotective agents containing L-glutamine and methyl-{beta}-cyclodextrin in a preincubation medium yields a high fertilization rate for cryopreserved C57BL/6J mouse sperm. Lab Anim 44, 132-7 (2010).
71. Rogers, D.C. et al. Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. Mamm Genome 8, 711-3 (1997).
72. Spoutil, F. et al. Semi-Automated MicroCT Analysis of Bone Anatomy and Mineralization in Mouse Models. Curr Protoc 4, e980 (2024).
73. Syding, L.A. et al. Generation and Characterization of a Novel Angelman Syndrome Mouse Model with a Full Deletion of the. Cells 11(2022).

Figure1

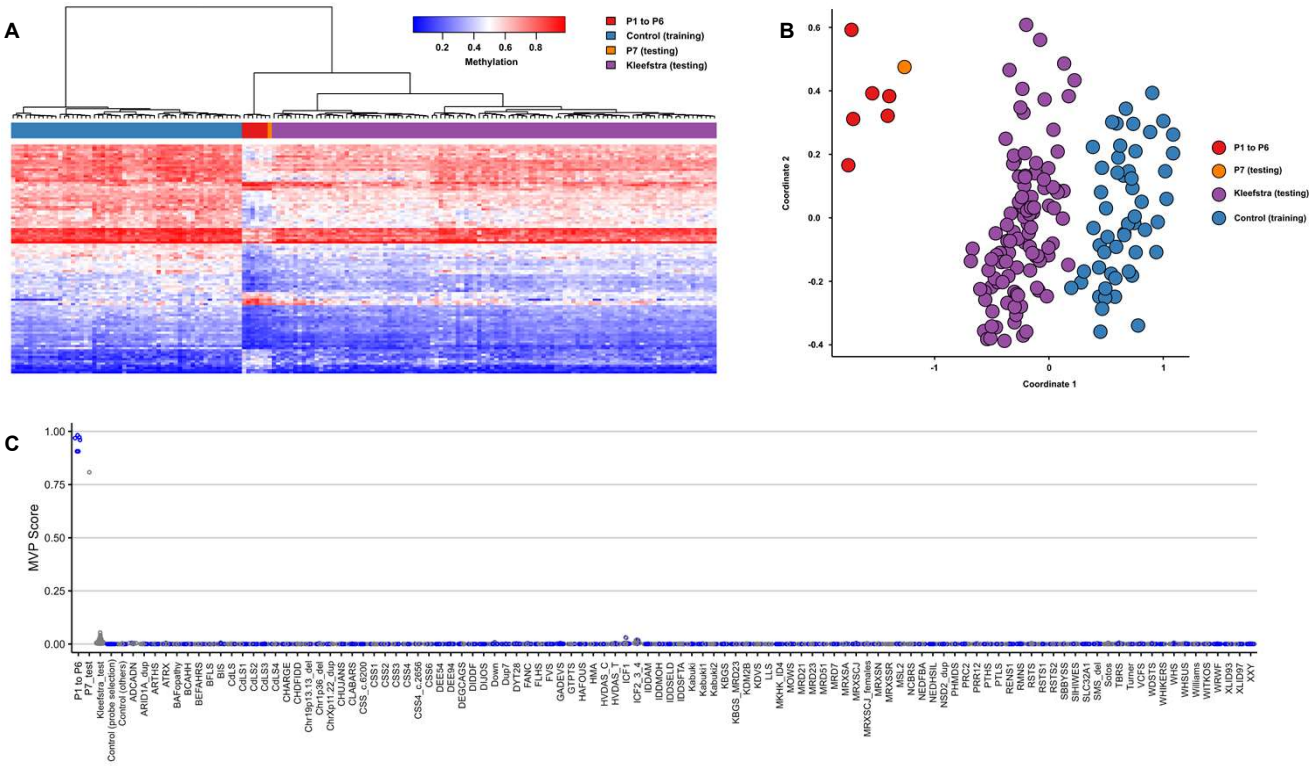

A
0.2
0.4
0.6
0.8
Methylation
P1 to P6
Control (training)
P7 (testing)
Kleefstra (testing)
B
Coordinate 2
Coordinate 1
P1 to P6
P7 (testing)
Kleefstra (testing)
Control (training)
C
MVP Score
1.00
0.75
0.50
0.25
0.00

Figure 2

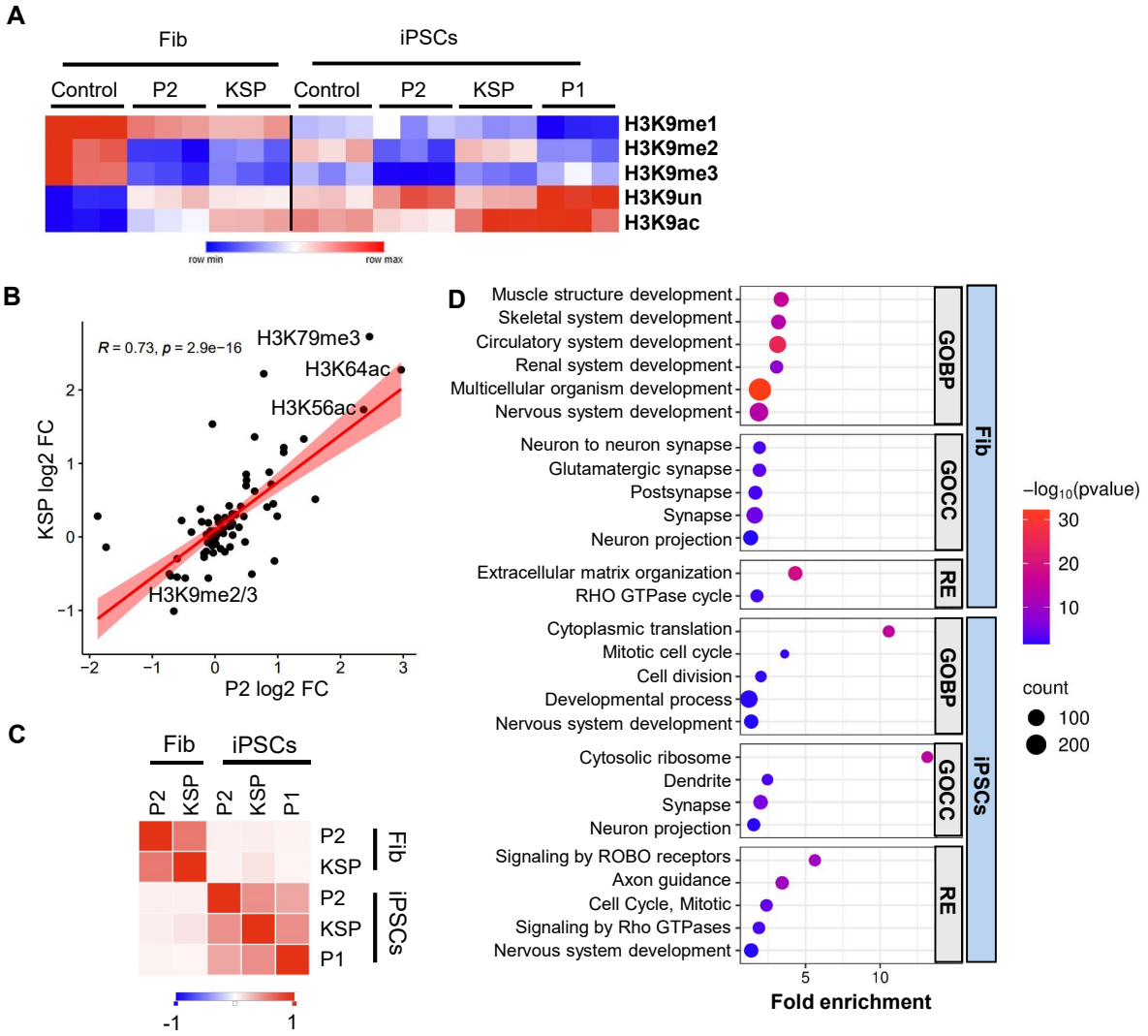
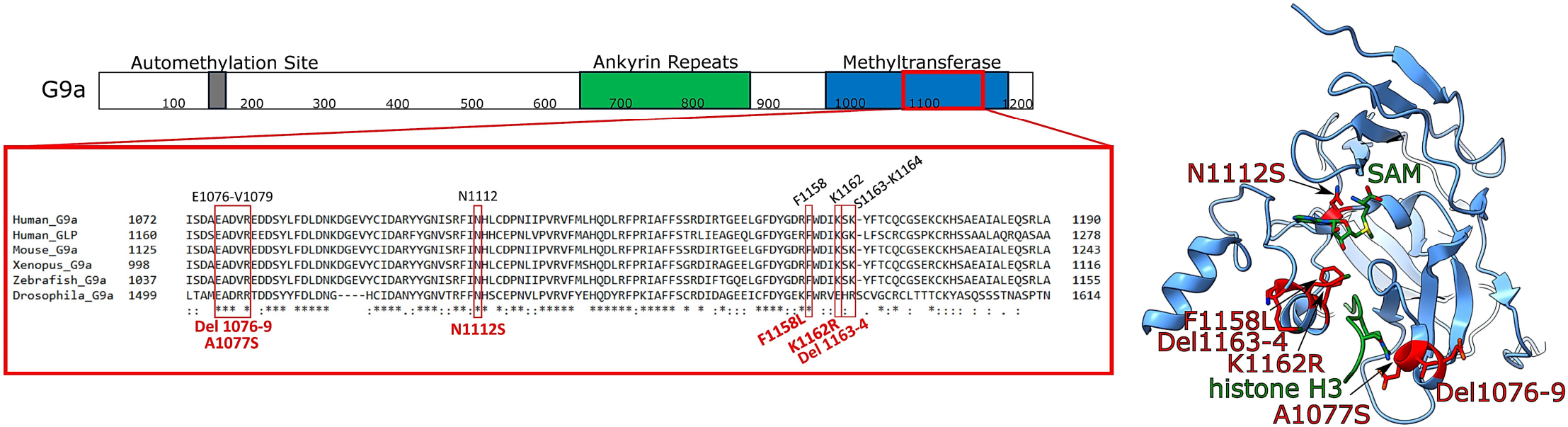
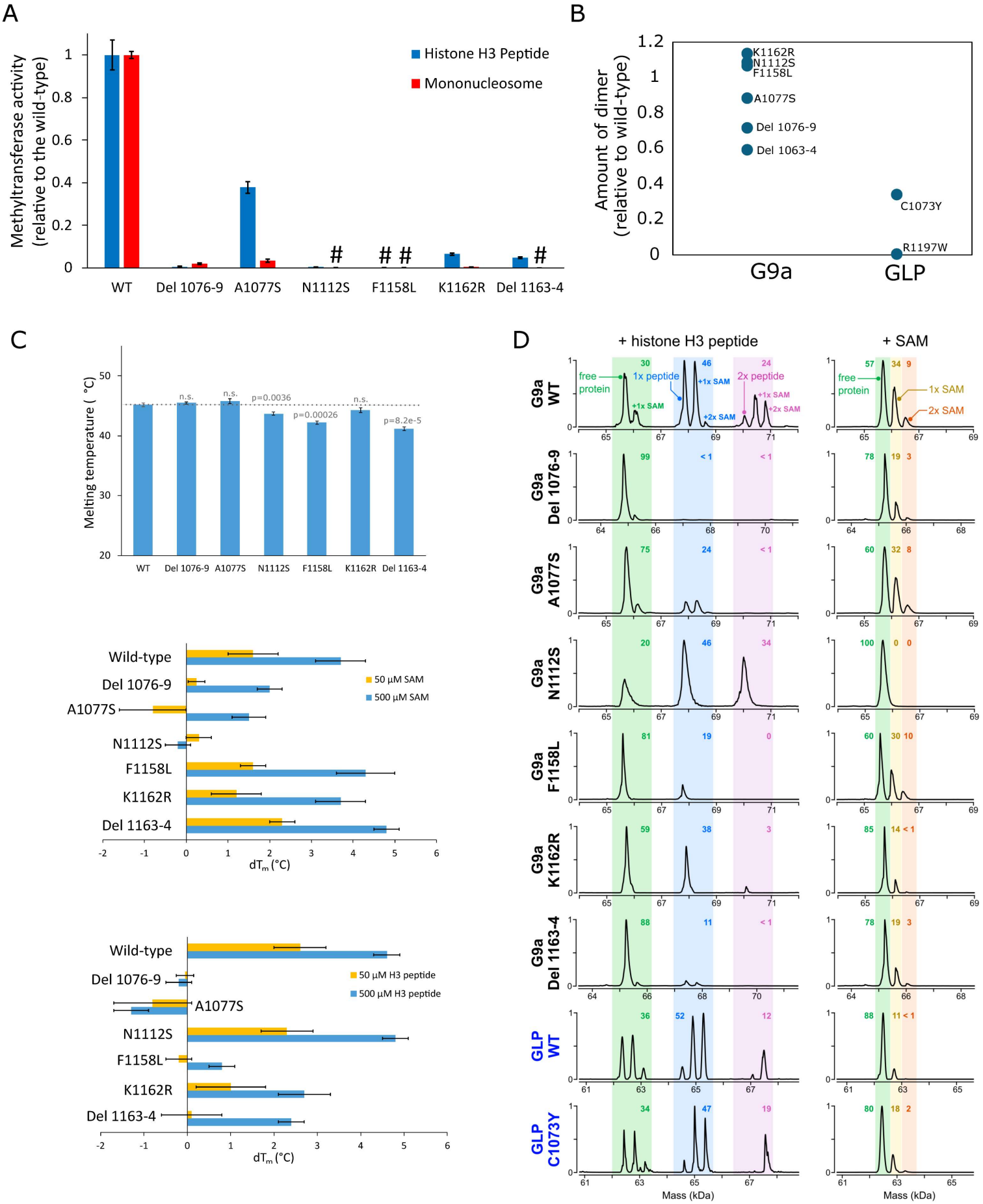

A
Fib
iPSCs
Control
P2
KSP
Control
P2
KSP
P1
H3K9me1
H3K9me2
H3K9me3
H3K9un
H3K9ac
row min
row max
B
R = 0.73, p = 2.9e−16
H3K79me3
H3K64ac
H3K56ac
H3K9me2/3
KSP log2 FC
P2 log2 FC
C
Fib
iPSCs
P2
KSP
P2
KSP
P1
Fib
iPSCs
-1
1
D
Muscle structure development
Skeletal system development
Circulatory system development
Renal system development
Multicellular organism development
Nervous system development
Neuron to neuron synapse
Glutamatergic synapse
Postsynapse
Synapse
Neuron projection
Extracellular matrix organization
RHO GTPase cycle
Cytoplasmic translation
Mitotic cell cycle
Cell division
Developmental process
Nervous system development
Cytosolic ribosome
Dendrite
Synapse
Neuron projection
Signaling by ROBO receptors
Axon guidance
Cell Cycle, Mitotic
Signaling by Rho GTPases
Nervous system development
GOBP
GOCC
RE
GOBP
GOCC
RE
Fib
iPSCs
Fold enrichment
−log10(pvalue)
30
20
10
count
100
200

Figure 3


G9a
Automethylation Site
Ankyrin Repeats
Methyltransferase
100
200
300
400
500
600
700
800
900
1000
1100
1200
E1076-V1079
N1112
F1158
K1162
S1163-K1164
Human_G9a 1072 ISDAEADVREDDSYLFDLDNKDGEVYCIDARYYGNISRFINHLCDPNIIPVRVFMLHQDLRFPRIAFFSSRDIRTGEELGFDYGDRFWDIKSK-YFTCQCGSEKCKHSAEAIALEQSRLA 1190
Human_GLP 1160 ISDSEADVREEDSYLFDLDNKDGEVYCIDARFYGNVSRFINHHCEPNLVPVRVFMAHQDLRFPRIAFFSTRLIEAGEQLGFDYGERFWDIKGK-LFSCRCGSPKCRHSSAALAQRQASAA 1278
Mouse_G9a 1125 ISDAEADVREDDSYLFDLDNKDGEVYCIDARYYGNISRFINHLCDPNIIPVRVFMLHQDLRFPRIAFFSSRDIRTGEELGFDYGDRFWDIKSK-YFTCQCGSEKCKHSAEAIALEQSRLA 1243
Xenopus_G9a 998 ISDAEADVREDDSYLFDLDNKDGEVYCIDARYYGNVSRFINHLCEPNLIPVRVFMSHQDLRFPRIAFFSGRDIRAGEELGFDYGDRFWDIKSK-YFTCQCGSERCKHSAEAIALEQSRLA 1116
Zebrafish_G9a 1037 ISDAEADVREDDSYLFDLDNKDGEVYCIDARYYGNISRFINHLCDPNIIPVRVFMLHQDLRFPRIAFFSSRDIFTGQELGFDYGDRFWDIKSK-YFTCQCGSEKCKHSAEAIALEQSRLA 1155
Drosophila_G9a 1499 LTAMEADRRTDDSYYFDLDNG----HCIDANYYGNVTRFFNHSCEPNVLPVRVFYEHQDYRFPKIAFFSCRDIDAGEEICFDYGEKFWRVEHRSCVGCRCLTTTCKYASQSSSTNASPTN 1614
Del 1076-9
A1077S
N1112S
F1158L
K1162R
Del 1163-4

N1112S
SAM
F1158L
Del1163-4
K1162R
histone H3
A1077S
Del1076-9

Figure 4


A
Methyltransferase activity (relative to the wild-type)
Histone H3 Peptide
Mononucleosome
WT
Del 1076-9
A1077S
N1112S
F1158L
K1162R
Del 1163-4
B
Amount of dimer (relative to wild-type)
K1162R
N1112S
F1158L
A1077S
Del 1076-9
Del 1063-4
C1073Y
R1197W
G9a
GLP
C
Melting temperature (°C)
n.s.
p=0.0036
p=0.00026
p=8.2e-5
Wild-type
Del 1076-9
A1077S
N1112S
F1158L
K1162R
Del 1163-4
50 µM SAM
500 µM SAM
dTm (°C)
50 µM H3 peptide
500 µM H3 peptide
D
+ histone H3 peptide
+ SAM
G9a WT
G9a Del 1076-9
G9a A1077S
G9a N1112S
G9a F1158L
G9a K1162R
G9a Del 1163-4
GLP WT
GLP C1073Y
free protein
1x peptide
2x peptide
+1x SAM
+2x SAM
Mass (kDa)

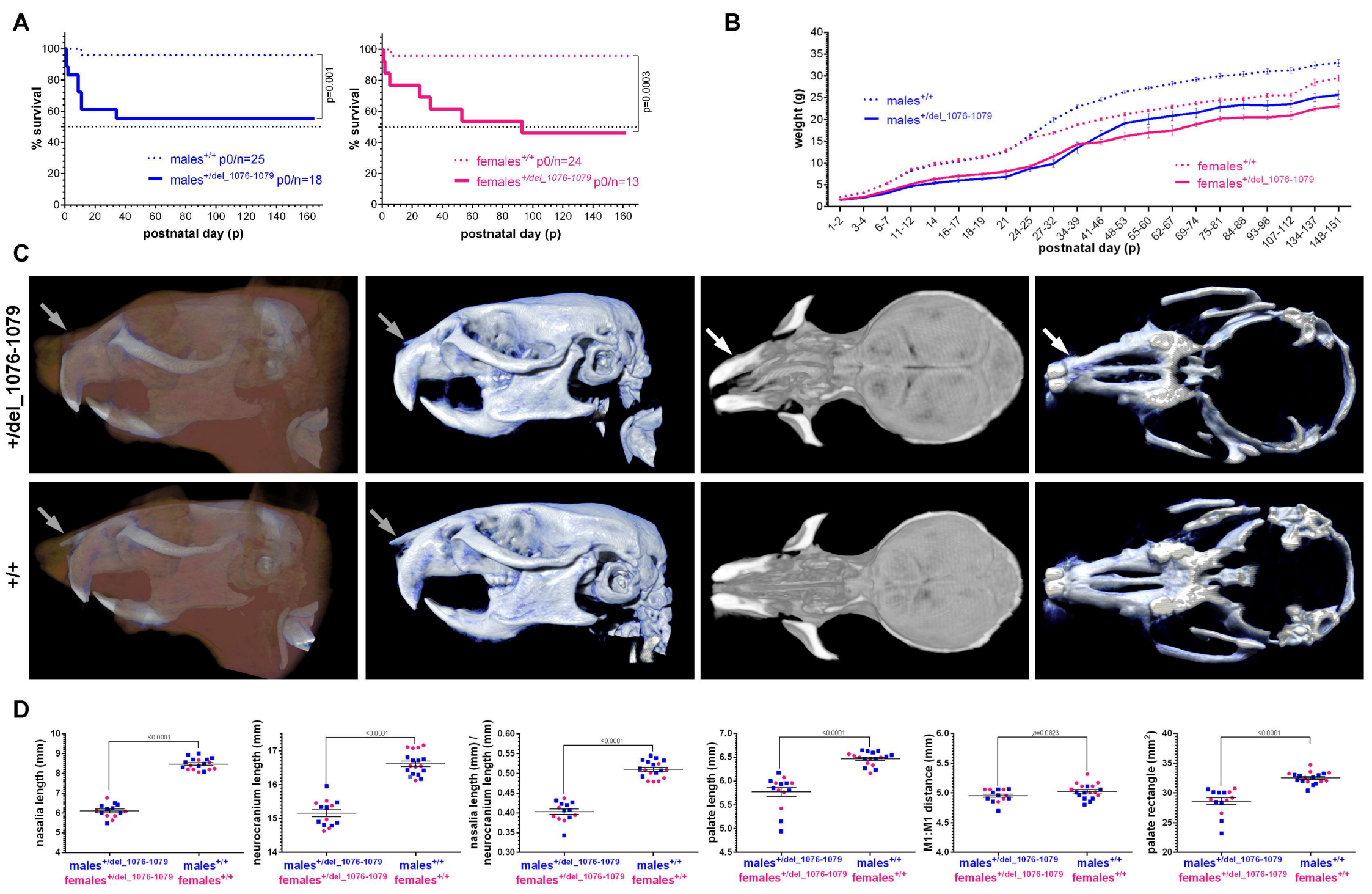

A
% survival
postnatal day (p)
males+/+ p0/n=25
males+/del_1076-1079 p0/n=18
p=0.001
females+/+ p0/n=24
females+/del_1076-1079 p0/n=13
p=0.0003
B
weight (g)
postnatal day (p)
males+/+
males+/del_1076-1079
females+/+
females+/del_1076-1079
C
+/del_1076-1079
+/+
D
nasalia length (mm)
neurocranium length (mm)
nasalia length (mm) / neurocranium length (mm)
palate length (mm)
M1:M1 distance (mm)
palate rectangle (mm²)
<0.0001
p=0.0823
males+/del_1076-1079
females+/del_1076-1079
males+/+
females+/+

Figure 5

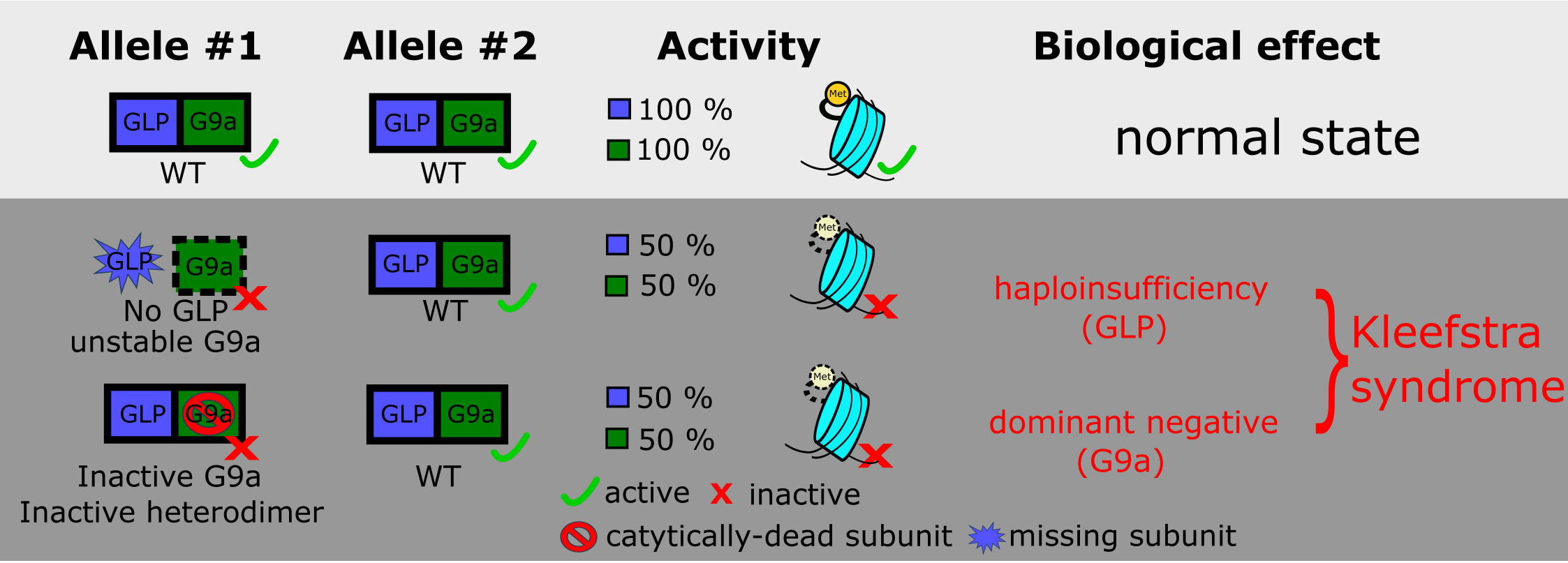

Allele #1
Allele #2
Activity
Biological effect
GLP
G9a
WT
100 %
100 %
Met
normal state
No GLP
unstable G9a
50 %
50 %
haploinsufficiency
(GLP)
Kleefstra
syndrome
Inactive G9a
Inactive heterodimer
dominant negative
(G9a)
active
inactive
catytically-dead subunit
missing subunit

Figure 6

EHMT2 clinical data							
Study Patient ID	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Inheritance	de novo	de novo	de novo	de novo	de novo	de novo	de novo
Zygosity	heterozygous	heterozygous	heterozygous	heterozygous	heterozygous	heterozygous	heterozygous
Mutation (hg38)	6-31881053-CACATCAGCCTCA-C (12bp)	6-31881061-C-A	6-31880790-T-C	6-31880245-A-G	6-31880224-ATTTGCT-A	6-31880232-T-C	6-31881061-C-T
Mutation (NM_006709.5)	c.3225_3236del	c.3229G>T	c.3335A>G	c.3472T>C	c.3487_3492del	c.3485A>G	c.3229G>A
Mutation (NP_006700.3)	p.Glu1076_Val1079del	p.Ala1077Ser	p.Asn1112Ser	p.(Phe1158Leu)	p.(Ser1163_Lys1164del)	p.(Lys1162Arg)	p.Ala1077Thr
gnomAD V4.1 occurence	0x	0x	0x	0x	0x	0x	0x
AlphaMissense	NA	0.98 (pathogenic)	0.94 (pathogenic)	0.95 (pathogenic)	NA	0.97 (pathogenic)	0.99 (pathogenic)
REVEL	NA	0.77 (Pathogenic)	0.8 (Pathogenic)	0.81 (Pathogenic)	NA	0.76 (Pathogenic)	0.75 (pathogenic)
SpliceAI	0 (no predicted effect on splicing)	0 (no predicted effect on splicing)	0.1 (no predicted effect on splicing)	0 (no predicted effect on splicing)	0 (no predicted effect on splicing)	0 (no predicted effect on splicing)	0 (no predicted effect on splicing)
GERP score	4.59	4.59	4.35	4.75	4.75	4.75	4.59
Gender	female	male	female	female	male	female	male
Last follow-up age	15 years	10 years	7 years, 6 months	16 years	2 years	4 years 2months	3 years 10 months
Occipital frontal interference at birth	34,5 cm	34,5 cm	34 cm	32 cm	35.5 at 41 WG + 6 days	33.1 cm	33 cm
Length at birth (cm)	51 cm	50,5 cm	49 cm	48 cm	55 cm	not recorded	47 cm
Weight at birth (g)	3150 g	2440 g	2824 g	2630 g	3100 g	2840 g	2390 g
Occipital frontal interference at last evaluation	51 cm		48.5 cm	51 cm	48.5 cm	49.2cm at 3yr 11 mo	ND
Height (cm)	145 cm	140 cm	100.8 cm	163.5 cm	90 cm	92.4cm at 3yr 3 mo	102.2 cm
Weight (kg)	30 kg	29 kg	15. 3 kg	62 kg	13 kg	14.05 kg at 3yr 3mo	14.7 kg
Neurological findings:							
hypotonia	yes	yes	no	yes	yes	yes	yes
EEG	Nonspecic bilateral focal activity located temporo - occipito-parietally.	Pathological profile at 15 days, normal in later recordings.	Correctly structured wake and sleep pattern apart from a sometimes slower left hemispherical aspect and acute but subclinical flushes? No focus of electric shock abnormalities.	No epilepsy.	ND	No focal or epileptiform anormalities.	Negative for seizure activity.
MRI	Periventricular atrophy, cortical atrophy, megacisterna magna with arachnoid cyst along the left cerebellar hemisphere.	Partially empty sella turcica.	Punctiform FLAIR hypersignal in the left frontal white SB with diffusion restriction, compatible with an ischemic lesion. FLAIR hypersignal from the knee of the corpus callosum on the right, without diffusion restriction, probably a vessel and moderate dilatation of the lateral ventricles.	ND	ND	Mild asymmetry of ventricles and prominence of extra-axial spaces.	Abnormal corpus callosum; mild hypoplasia; mildy enlarged cisterns without Dandy Walker, abnormal subarachnoid space with mildly enlarged bifrontal space, moderate ventriculomegaly without hydrocephalus, and multiple diffuse (mostly subcortical) white matter gliosis.
Seizures	No	one episode	No	No	Febrile seizure at 20 months.	No	Febrile seizures at 1y 10mo of age. Seizure-like activity vs sensory seeking behaviors (had twitching, arches head backwards and forewards, extreme belly laughts without provocation - all followed by period of prolonged sleep); unremarkable EEGs.
Psycho-developmental and behavioural findings:							
Development	severe developmental delay	psychomotor delay	normal	moderate developmental delay	motor delay	global developmental delay	psychomotor delay
Autism	some autistic signs	normal social behavior	no	no	no	no, sociable and enjoys company	ND

Behavior	hyperkinetic movements of hands in emotions, bruxism, self-mutilation in anger	stereotypic hand movements	sudden onset of hyperkinetic movement and some choreoathetosic movements, improvement with L-dopa	behavioral disorder		stereotypic hand and limb movements, tongue-clicking	atypical hand movements
ID	severe mental retardation	severe global developmental delay/intellectual disability	no	moderate developmental delay		moderate-severe global delay (Non-verbal and non-ambulant at age 4 years).	significant global developmental delay/ skills 4-8 months
Motor abilities	walks independently - short distances, longer distances on wheelchair, suboptimal fine motor skills	unsteady walk with frequent falls	normal	walks independently	walked unaided at 24m	at age 4 years, sits independently and stands with support. Not independently ambulant.	significant motor and adaptive functioning delays
Cognitive abilities	understands the instructions	understands the speech	normal, no learning difficulties	understands the instructions		demonstrates receptive understanding, expresses happiness and enjoyment appropriately.	significantly delayed
Verbal abilities	absent speech	no speech	complexe sentences	limited language with simple sentences	a few words	absent speech. Some vocalisations are meaningful in terms of tone.	no speech
Cranial, facial and somatic dysmorphisms:							
Truncal obesity	–	–	–	+	–	–	–
Short neck	+	–	–	–	–	–	–
Brachycephaly	+	+	–	+	–	+	+
Plagiocephaly with flat occipit	+	+	–	+	–	+	+
Flat face	+	–	–	–	–	+	–
Broad face	+	+	+	+	–	+	+
High forehead	+	–	–	–	–	+	–
Midface hypoplasia	+	+	+	+	–	+	+
Prognathia/midface retraction	+	+	+	+	–	+	+
Widow's peak/low anterior hairline	+	+	+	–	–	+	–
Hypertelorism	+	–	+	–	–	–	+
Synophrys	–	+	+	–	+	+	–
Sparse medial eyebrows	–	+	–	–	–	–	–
Deeply set eyes	+	+	+	+	–	–	–
Epicanthus	+	+	–	–	+	–	+
Long eyelashes	+	–	+	–	–	+	+
Flat/low nasal bridge	–	–	–	+	–	+	+
Short nose	+	–	+	-	+	+	+
Beaked nose	–	–	+	+	–	–	–
Anteverted nares/Hypoplastic	–	+	+	+	–	+	–
Protruding columella	+	+	–	+	+	–	–
Short philtrum	+	+	+	+	–	+	+
Microstomia	–	–	–	–	+	+	–
Carp-like mouth	–	+	–	–	–	+	–
Thin lips /thin upper lip	**+ (thin lower lip)**	+	–	+	–	+	+
Widely spaced teeth	+	+	+	–	**+ (primary teeth)**	+	+
Prominent superior incisors	+	–	–	+	–	–	–
Dysplastic/orverfolded helices/ears	+	–	–	+	+	+	+
Crumpled ears	+	–	–	+	+	–	–
Short nails	+	+	–	–	+	+	–
Tapered fingers	+	+	–	+	–	–	+
	+	+	–	–	–		–
Pes planovalgus/neurogenic	+	+	–	–	–	–	–
1st toe longer than others/ Sandal gap	+	+	+	+		–	+
Unilateral postaxial polydactyly of	–	–	–	+	–	–	–

Cardiologic findings:	stenosis of arteria pulmonalis	membranous septal defect, patent foramen ovale, coarctation of the aorta (surgery required)	cardiac anomalies (intra ventricular and auricular communication), surgery required	atrial septal defect	ventricular septal defect (surgically repaired)	coarctation of aortic isthmus; dysplastic pulmonary valve	patent foramen ovale in infancy, resolved
Hearing impairment:	no	no	no	moderate deafness (35 dB)	no	no	no
Visual impairment:	hypermetropia, glasses	no	no	no	?	incipient myopia at last review	alternating extropia; hyperopia with astigmatism, bilateral; maculopathy, bilateral; premature macula
Skeletal anomalies	no	no	no	no	?	no	cutanesous syndactyl 2nd/3rd toes
Gastrointestinal anomalies:	pylorostenosis, surgery	no	no	no	chronic constipation	antenatally detected large stomach and duodenum. Recurrent abdominal distension. Hirschbrung disease excluded on biopsy	GERD, constipation, umbilical hernia
Other:	growth and sexual maturation disorder	kidney cysts in the renal medulla, undescended testicles, inguinal hernia	kidney anomalies (cyst on the left one)	no	stenosis of one lacrymal duct, pyelonephritis at 20 m and posterior urethral valves (surgically repaired)	livedo reticularis; follicular hyperkeratosis of upper limbs; delayed closure of anterior fontanelle; delayed eruption of primary dentition; obstructive sleep apnea on BiPAP therapy	cryptorchidism at birth, resolved without surgery; Seizure-like activity vs sensory seeking behaviors (had twitching, arches head backwards and forewards, extreme belly laughts without provocation - all followed by period of prolonged sleep); unremarkable EEGs
Genetic analyses:							
Microarray	15q26.1 microduplication, inherited from the mother	ND	ND	normal	400 kb	normal SNP microarray	normal male
Karyotype and results of other genetic testing	46, XX, FISH22 negative, William-Beuren syndrome, neuromuscular disorders, SMA, mtDNA deletions, Prader-Willi syndrome, Rett syndrome - all negative	Prader-Willi syndrome, aCGH, clinical exome sequencing - all negative	ND	COH1 gene negative	ND	karyotype 46, XX	negative trio exome; negative mitochondrial genome seq & del/dup
Biochemical testing	ND	ND	ND	ND	ND	Extensive metabolic screens negative.	abnormall LSD at birth, normal at our evaluation
Family History	negative	negative	negative	negative		negative	maternal uncle deceased; brain teratoma
Other notes	NA	polyhydramnion and intrauterine growth restriction, maternal infection by parvovirus during the last trimester of pregnancy	NA	NA	IVF	Antenatal polyhydramnios, large stomach and duodenum.	IUGR; born at 37 weeks gestation, admitted to NICU for respiratory distress for 25 days