Pathogenic variants in KMT2C result in a neurodevelopmental disorder distinct from Kleefstra and Kabuki syndromes

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Rots, D., Choufani, S., Faundes, V., Dingemans, A.J.M., Joss, S., Foulds, N., et al. (2024). Pathogenic variants in KMT2C result in a neurodevelopmental disorder distinct from Kleefstra and Kabuki syndromes. AMERICAN JOURNAL OF HUMAN GENETICS, 111(8), 1626-1642 [10.1016/j.ajhg.2024.06.009].

Availability:
This version is available at: https://hdl.handle.net/11585/981977 since: 2024-09-09

Published:
DOI: http://doi.org/10.1016/j.ajhg.2024.06.009

Pathogenic variants in *KMT2C* result in a neurodevelopmental disorder distinct from Kleefstra and Kabuki syndromes

AUTHORS & AFFILIATIONS

Dmitrijs Rots,[1,2,3,*] Sanaa Choufani,[4,*] Victor Faundes,[5,86,*] Alexander J.M. Dingemans,[1] Shelagh Joss,[6] Nicola Foulds,[7] Elizabeth A. Jones,[8,9] Sarah Stewart[8], Pradeep Vasudevan,[10] Tabib Dabir,[11] Soo-Mi Park,[12] Rosalyn Jewell,[13] Natasha Brown,[14,15] Lynn Pais,[16,17] Sébastien Jacquemont,[18] Khadijé Jizi,[19] Conny M.A. van Ravenswaaij-Arts,[20] Hester Y. Kroes,[21] Constance T. R. M. Stumpel,[22,23] Charlotte W. Ockeloen,[1] Illja J. Diets,[1] Mathilde Nizon,[24] Marie Vincent,[24] Benjamin Cogné,[24] Thomas Besnard,[24] Marios Kambouris,[25] Emily Anderson,[26] Elaine H Zackai,[27] Carey McDougall,[28] Sarah Donoghue,[28] Anne O'Donnell-Luria,[16,17] Zaheer Valivullah,[16] Melanie O'Leary,[16,29] Siddharth Srivastava,[30] Heather Byers,[31] Nancy Leslie,[33] Sarah Mazzola,[34] George E. Tiller,[35] Moin Vera,[35] Joseph J. Shen,[36,37] Richard Boles,[38] Vani Jain,[39] Elise Brischoux-Boucher,[40] Esther Kinning,[41] Brittany N. Simpson,[42] Jacques C. Giltay,[21] Jacqueline Harris,[43,44] Boris Keren,[45] Anne Guimier,[46] Pierre Marijon,[47] Bert B.A. de Vries,[1] Constance S. Motter,[48] Bryce A. Mendelsohn,[49] Samantha Coffino,[50] Erica H. Gerkes,[51] Alexandra Afenjar,[52] Paola Visconti,[53] Elena Bacchelli,[54] Elena Maestrini,[54] Andree Delahaye-Duriez,[55] Catherine Gooch,[56] Yvonne Hendriks,[57] Hieab Adams,[1,2] Christel Thauvin-Robinet,[58] Sarah Josephi-Taylor,[59,60] Marta Bertoli,[61] Michael J. Parker,[62] Julie W. Rutten,[57] Oana Caluseriu,[63] Hilary J. Vernon,[64] Jonah Kaziyev,[17] Jia Zhu,[17] Jessica Kremen,[65] Zoe Frazier,[66] Hailey Osika,[66] David Breault,[17] Sreelata Nair,[67] Suzanne M. E. Lewis,[68] Fabiola Ceroni,[54,69] Marta Viggiano,[54] Annio Posar,[53,70] Helen Brittain,[71] Traficante Giovanna,[72] Gori Giulia,[73] Lina Quteineh,[74] Russia Ha-Vinh Leuchter,[75] Evelien Zonneveld-Huijssoon,[76] Cecilia Mellado,[77] Isabelle Marey,[78] Alicia Coudert,[78] Mariana Inés Aracena Alvarez,[79] Milou G. P. Kennis,[1] Arianne Bouman,[1] Maian Roifman,[80] María Inmaculada Amorós Rodríguez,[81] Juan Dario Ortigoza-Escobar,[82] Vivian Vernimmen,[22,23] Margje Sinnema,[22] Rolph Pfundt,[1] Han G. Brunner,[1,22] Lisenka E.L.M. Vissers,[1,83] Tjitske Kleefstra,[1,2,84,**] Rosanna Weksberg,[4,85,**] Siddharth Banka,[8,86,**]

1 Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands
2 Department of Clinical Genetics, Erasmus MC, Rotterdam, the Netherlands
3 Genetics laboratory, Children's Clinical University Hospital, Riga, Latvia
4 Genetics and Genome Biology Program, Research Institute, the Hospital for Sick Children, Toronto, ON M5G 1X8, Canada

5 Laboratorio de Genética y Enfermedades Metabólicas, Instituto de Nutrición y Tecnología de los Alimentos (INTA), Universidad de Chile, Santiago, Chile
6 West of Scotland Centre for Genomic Medicine, Queen Elizabeth University Hospital, Glasgow, UK
7 Wessex Clinical Genetics Services, University Hospital Southampton NHS Foundation Trust, Southampton SO16 5YA, UK
8 Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester, UK
9 Division of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
10 Department of Clinical Genetics, University Hospitals of Leicester, Leicester Royal Infirmary, Leicester LE1 7RH, UK
11 Northern Ireland Regional Genetics Centre, Belfast City Hospital, Belfast, UK
12 Department of Clinical Genetics, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
13 Yorkshire Regional Genetics Service, Chapel Allerton Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK
14 Victorian Clinical Genetics Service, Murdoch Children's Research Institute, Melbourne, Victoria, Australia
15 Department of Paediatrics, Royal Children's Hospital, The University of Melbourne, Melbourne, Victoria, Australia
16 Center for Mendelian Genomics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
17 Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, USA
18 Department of Pediatrics, University of Montreal, Montreal, Québec, Canada
19 Service de génétique médicale, CHU Ste-Justine, Montréal, Québec, Canada
20 University of Groningen, University Medical Centre Groningen, Dept Genetics, Groningen, the Netherlands
21 Division Laboratories, Pharmacy and Biomedical Genetics, University Medical Center Utrecht, Utrecht, the Netherlands
22 Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, the Netherlands
23 GROW-School for Oncology and Reproduction, Maastricht, the Netherlands
24 Service de Génétique Médicale, Centre Hospitalier Universitaire de Nantes, Nantes, France
25 Division of Genetics, Department of Pathology and Laboratory Medicine Department, Sidra Medicine, Doha, Qatar
26 Liverpool Centre for Genomic Medicine, Liverpool Women's NHS Foundation Trust, Liverpool, UK
27 Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA
28 Division of Human Genetics, Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA
29 Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, USA
30 Department of Neurology, Boston Children's Hospital, Boston, Massachusetts, USA
31 Department of Pediatrics, Stanford University, Stanford, California, USA
32 Department of Pediatrics, Stanford School of Medicine, Stanford, California, USA

33 Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA
34 Center for Personalized Genetic Healthcare, Cleveland Clinic, Cleveland, Ohio, USA
35 Department of Genetics, Kaiser Permanente, Los Angeles, California, USA
36 Division of Genetics, Department of Pediatrics, UCSF Fresno, Fresno, California, USA
37 Division of Genomic Medicine, Department of Pediatrics, University of California Davis, Sacramento, California, USA
38 NeuraBilities healthcare, Pennsylvania, USA
39 All Wales Medical Genomics Service, Wales Genomic Health Centre, Cardiff Edge Business Park, Longwood Drive, Whitchurch, Cardiff, CF14 7YU
40 Centre de Génétique Humaine, CHU de Besançon, Université de Franche-Comté, Besançon, France
41 Clinical Genetics, Birmingham Women's and Children's, Birmingham, UK
42 Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati, Cincinnati, Ohio, USA
43 Kennedy Krieger Institute, Baltimore, Maryland, USA
44 Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
45 Department of Genetics, APHP Sorbonne University, Paris, France
46 Service de Médecine Genomique des Maladies Rares, CRMR Anomalies du développement, Hôpital Necker-Enfants Malades, Assistance Publique des Hôpitaux de Paris, Paris, France
47 Laboratoire de Biologie Médicale Multisites Seqoia FMG2025, 75014 Paris, France
48 Genetic Center, Akron Children's Hospital, Akron, Ohio, USA
49 Department of Medical Genetics, Kaiser Permanente, Oakland, USA
50 Department of Pediatric Neurology, Kaiser Permanente, Oakland, USA
51 Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
52 APHP Sorbonne Université, Centre de Référence Malformations et maladies congénitales du cervelet et déficiences intellectuelles de causes rares, département de génétique et embryologie médicale, Hôpital Trousseau, Paris, France
53 IRCCS Istituto delle Scienze Neurologiche di Bologna, UOSI Disturbi dello Spettro Autistico, Bologna, Italy
54 Pharmacy and Biotechnology Department, University of Bologna, Bologna, Italy
55 Medical Genomics and Clinical Genetics Unit, AP-HP, Hôpital Jean Verdier, Bondy, France
56 Division of Genetics and Genomic Medicine, Department of Pediatrics, Washington University School of Medicine, St Louis, Missouri, USA
57 Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands
58 Unité Fonctionnelle Innovation en Diagnostic Génomique des Maladies Rares, Dijon, France; Inserm, UMR1231, Equipe GAD, Bâtiment B3, Université de Bourgogne Franche Comté, Dijon Cedex, France; Centre de Référence Déficiences Intellectuelles de Causes Rares, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France
59 Department of Clinical Genetics, The Children's Hospital at Westmead, Sydney, NSW, Australia
60 Discipline of Genomic Medicine, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

61 Northern Genetics Service, Newcastle upon Tyne NHS Foundation Trust, Newcastle upon Tyne, UK
62 Department of Clinical Genetics, Sheffield Children's Hospital, Sheffield, UK
63 Department of Medical Genetics, University of Alberta, Edmonton, Canada
64 Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
65 Division of Endocrinology, Boston Children's Hospital, Harvard Medical School, Boston, USA
66 Rosamund Stone Zander Translational Neuroscience Center, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA
67 Department of Fetal Medicine, Lifeline Super Specialty Hospital, Kerala, India
68 Department of Medical Genetics, BC Children's Hospital Research Institute, The University of British Columbia, Vancouver, British Columbia, Canada
69 Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, United Kingdom
70 Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy
71 Department of Clinical Genetics, Birmingham Women's & Children's NHS Trust, Birmingham, UK
72 Medical Genetics Unit, Meyer Children's Hospital IRCCS Florence, Italy
73 Medical Genetics Unit,Meyer Children's Hospital IRCCS, Florence, Italy
74 Division of Genetic Medicine, Geneva University Hospitals, 1205 Geneva, Switzerland
75 Division of Development and Growth, Department of Pediatrics, University of Geneva, Switzerland
76 Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
77 Sección de Genética y Errores Congénitos del Metabolismo, División de Pediatría, Pontificia Universidad Católica de Chile, Santiago, Chile
78 CHU Grenoble Alpes, Grenoble, France
79 Unit of Genetics and Metabolic diseases, Division of Pediatrics, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile
80 The Prenatal Diagnosis and Medical Genetics Program, Division of Maternal Fetal Medicine, Department of Obstetrics and Gynaecology, University of Toronto, Toronto, Canada
81 Deparment of Pediatrics, Hospital Punta Europa Algeciras, Cadiz, Spain
82 Movement Disorders Unit, Institut de Recerca Sant Joan de Déu, CIBERER-ISCIII and European Reference Network for Rare Neurological Diseases (ERN-RND), Barcelona, Spain
83 Research Institute for Medical Innovation, Radboud University Medical Center, Nijmegen, the Netherlands
84 Center of Excellence for Neuropsychiatry, Vincent van Gogh Institute for Psychiatry, Venray, The Netherlands
85 Division of Clinical and Metabolic Genetics, Department of Pediatrics, the Hospital for Sick Children, University of Toronto, Toronto, ON M5G 1X8, Canada
86 Manchester Centre for Genomic Medicine, Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
* contributed equally
** contributed equally; corresponding authors: t.kleefstra@erasmusmc.nl; rweksb@sickkids.ca

ABSTRACT

Trithorax-related H3K4 methyltransferases, KMT2C and KMT2D, are critical epigenetic modifiers. Haploinsufficiency of *KMT2C* was only recently recognised as a cause of neurodevelopmental disorder (NDD), so the clinical and molecular spectrum of the *KMT2C*-related (NDD) (now designated as Kleefstra syndrome 2) are largely unknown. We ascertained 98 individuals with rare *KMT2C* variants, including 75 with protein truncating variants (PTVs). Notably, ~15% *KMT2C* PTVs were inherited. Although the most highly expressed *KMT2C* transcript consists of only the last four exons, pathogenic PTVs were found in almost all the exons of this large gene. We found that *KMT2C* variant interpretation can be challenging due to segmental duplications and clonal hematopoesis induced artefacts. Using samples from 27 affected individuals, divided into discovery and validation cohorts, we generated a moderate strength disorder-specific *KMT2C* DNAm signature and demonstrate its utility in classifying non-truncating variants. Based on 81 individuals with pathogenic/likely pathogenic variants, we demonstrate that the *KMT2C*-related NDDis characterized by developmental delay, intellectual disability, behavioral and psychiatric problems, hypotonia, seizures, short stature, and other comorbidities. Using facial module of PhenoScore on photographs of 34 affected individuals, we show that the *KMT2C*-related facial gestalt is significantly different from general neurodevelopmental disorder population. Finally, using PhenoScore and DNAm signatures, we demonstrate that the *KMT2C*-related NDDis clinically and epigenetically distinct from Kleefstra and Kabuki syndromes. Overall, we define the clinical features, the molecular spectrum, and the DNAm signature of the *KMT2C*-related NDD and prove them as distinct from Kleefstra and Kabuki syndromes highlighting the need to rename this condition.

INTRODUCTION

Mendelian disorders of epigenetic machinery are amongst the most common forms of neurodevelopmental disorders (NDDs)[1]. Humans possess six histone-3 lysine-4 (H3K4) methyltransferases that are divided into three sub-groups, including the trithorax-related subgroup which consists of KMT2C and KMT2D. These proteins are important components of the epigenetic machinery and are involved in spatiotemporal gene expression regulation[2,3]. KMT2C or KMT2D, together with WDR5, RBBP5, ASH2L, DPY30 (i.e., WRAD subunit), KDM6A, and other proteins, form the COMPASS complex[2] that mono- (H3K4me1)[4] and trimethylates H3K4 (H3K4me3)[5] at active chromatin sites of gene enhancers and promoters, respectively.

Heterozygous loss-of-function *KMT2D* (MIM #602113) variants were identified to cause Kabuki syndrome type 1 (MIM #147920) in 2010[6]. However, the consequences of germline *KMT2C* (MIM #606833) variants in humans have been identified more recently. In a phenotype-led study we identified *de novo* loss-of-function *KMT2C* variants in individuals with clinical characteristics overlapping Kleefstra syndrome[3,7]. Kleefstra syndrome (MIM #610253) is caused by haploinsufficiency of euchromatin histone methyltransferase 1 (*EHMT1*; MIM #610253) and is characterized by intellectual disability (ID), autism spectrum disorder (ASD), characteristic facial dysmorphisms, and other variable clinical features[8]. In an independent genotype-led study, we studied variants in histone lysine methyltransferases (KMTs) and demethylases (KDMs) in the Deciphering Developmental Disorders (DDD) cohort and discovered that *KMT2C* loss-of-function variants result in a neurodevelopmental phenotype with occasional physical anomalies[9]. However, the clinical and molecular spectrum of the *KMT2C*-related disorder is largely unknown, as only a few such individuals have been reported. Based on our experience, we hypothesized that the *KMT2C*-related NDD is a unique entity that is clinically and molecularly different from Kleefstra and Kabuki syndromes.

In this study, we provide comprehensive clinical, molecular and DNA methylation (DNAm) data for the *KMT2C*-related NDD based on a large previously unreported cohort of individuals, as well as demonstrate that this disorder is different from the molecularly-related Kleefstra and Kabuki type 1 syndromes.

MATERIALS AND METHODS

Cohort recruitment

This study was approved by the institutional review boards of the South Manchester NHS REC, Radboudumc and the Hospital for Sick Children (Research Ethics Committee Approvals 11/H1003/3/AM02, 2011/188, and #1000038847, respectively). Individuals with rare heterozygous pathogenic (P) or likely pathogenic (LP) variants or variants of uncertain significance (VUSs) in *KMT2C* were identified in clinical diagnostic settings (using standard chromosomal microarray, exome, or genome sequencing[10]) or from large NDD research cohorts (the Deciphering Developmental Disorders (DDD) study[11], 100,000 Genomes project[12], Simons Simplex collection (SSC)[13], and MSSNG[14]). Individuals with rare reported single nucleotide variants (SNVs), indels, and copy number variants (CNVs) in the *KMT2C* gene were included. However, CNVs were limited only to those with deletions <1Mb which did not affect other predicted haploinsufficient genes. Individuals with additional pathogenic variants in other genes were also included in the study but were excluded from the clinical feature frequency analysis. All included individuals or their caregivers/legal representatives consented to participate in this research. Genetic and clinical data from individuals were collected via a customized proforma.

Variant analysis

The variants were annotated using the *KMT2C* MANE select[15] transcript (GenBank:NM_170606.3, GRCh37). All identified variants were re-classified according to the ACMG guidelines 2015[16] based on the clinical and molecular evidence obtained during this study. The variants and their interpretations were submitted to the ClinVar database (ClinVar accession numbers: SCV005044911-SCV005044983 and SCV005044991-SCV005045000). Only individuals with LP/P *KMT2C* variants without known pathogenic variants in other genes were included for further clinical feature analyses.

DNA methylation analysis

Sample processing, DNA methylation (DNAm) signature derivation and analysis were performed similarly as described before[17-19]. Briefly, DNA samples underwent bisulfite conversion using the EpiTect Bisulfite Kit (EpiTect PLUS Bisulfite Kit, QIAGEN) following the manufacturer's protocol. The converted DNAs were then analyzed on the Illumina Infinium Human Methylation EPIC V1 BeadChip (with ~850,000 CpG sites) at The Center for Applied Genomics (TCAG), Hospital for Sick Children Research Institute, Toronto, Ontario, Canada. The affected individuals' and controls' samples were randomly positioned on the arrays.

DNAm analysis was performed using the minfi R package[20]. Briefly, the minfi package was used for data preprocessing, quality control, normalization, transformation to β values and blood cell composition estimation by Housman's method. Probes with a detection p-value >0.05 in >25% of the samples (911 probes), probes located near common polymorphic variants with minor allele frequencies >1% (166,596 probes), non-specific probes (35,613 probes), probes with raw β value equal to 0 or 1 in >25% of samples (240 probes), non-CpG probes (2,377 probes), and chrX and chrY probes (17,485 probes) were removed from the analysis. Additionally, 705 probes behaving like single nucleotide variants were removed using the MethylToSNP package[21]. In total, 645,199 CpG sites remained for the differential methylation analysis. All samples passed the quality control and were suitable for the analysis.

For the signature derivation, the remaining 645,199 CpG sites' β values were transformed to M values, and differentially methylated CpGs were identified by using a linear regression with monocyte count, batch and second principal components as covariates using the limma R package[22]. Only differentially methylated CpGs, with a 10% methylation difference ($|\Delta\beta|>0.10$) and false discovery rate (FDR)-adjusted p-values <0.05 were selected for the analysis. To remove false positive sites, we further excluded 6 CpGs with methylation β values in cases and controls that followed a SNP-like pattern. This resulted in a DNAm signature of 51 CpGs with $|\Delta\beta|>0.10$. The results were visualized through principal component analysis (PCA) and heatmap plots using the Qlucore Omics software.

Machine learning models and classifications

For the sample classification, we have developed a machine learning model – support vector machine (SVM) with linear kernel using *caret* R package[23] as described before[17]. The model was trained on the signature's CpG sites for the discovery cohorts (16 *KMT2C*-related NDD and 50 control samples). Because infants have noticeably different blood DNAm, 3 infant control samples were added to the controls during the model training, to increase model's specificity across different ages. Receiver operating characteristic analysis was used to select the optimal model using the largest value. The SVM model was set to the "probability" mode, so the model generated scores ranging between 0 and 1, where scores <0.25 were interpreted as "negative", >0.5 as "positive", and 0.25-0.5 as "intermediate".

The model's sensitivity and specificity were evaluated using the *KMT2C*-related NDD and control validation samples, as well as Kabuki type 1 and Kleefstra syndromes samples. Finally, the model was used to classify the 22 testing samples with *KMT2C* VUSs.

To evaluate the overlap between *KMT2C*-related NDD DNAm changes with DNAm changes for other disorders of the epigenetic machinery, we analyzed all *KMT2C* samples on 17 other available DNAm signatures deployed on EpigenCentral portal[24], as described before[25]. To evaluate specificity of the *KMT2C*-related NDD DNAm signature, we have utilized six molecularly-confirmed Kabuki syndrome type 1 and six Kleefstra syndrome individuals, as well as 165 healthy controls.

Protein 3D structure analysis

KMT2C is a large (4911 amino acids) multidomain protein, so there is no solved protein structure of the whole KMT2C protein available. Therefore, for the analysis of possible missense variant effects, we have used: 1) solved structures of HMG box (PDB:2YUK), extended PHD6 (PDB:6MLC[26]), and SET (PDB:6KIW[27], 5F6K, 5F59[28]) domains; 2) homology models for FYR (based on PDB:2WZO[29]), and extended PHD2 (based on PDB:4NN2[30], 6U04[31], 7MJU[32], 7D8A[33]) domains; 3) *ab initio* protein models for all missense positions. The homology modeling, analysis, and visualization were performed using YASARA Structure software[34]. The *ab initio* protein models were generated by AlphaFold2[35] (multiple modeled overlapping protein fragments of maximal size 1500 amino acids) and downloaded from the AlphaFold Protein Structure Database[36]. The predicted effect was assessed by at least two different protein structures/models, where available (e.g., solved structure and *ab initio*; two different *ab initio* structures).

Clinical features analyses

Consent acquisition, as well as clinical and molecular description of the recruited individuals were provided by their physicians using a standardized proforma (**Table S1**). However, for statistical analyses, only individuals with LP/P *KMT2C* variants and without confirmed additional pathogenic variants in other genes were considered. For any feature, we excluded individuals with "UNKNOWN" coding as previously described[37]. WHO growth standards per age and sex groups were used to evaluate the growth. Absolute and relative frequencies (expressed as n[%]) were used for describing categorical variables, whereas medians (m) and interquartile ranges (IQR) were used for describing continuous variables. Chi-square/Fisher's exact and Mann–Whitney tests were applied to study categorical and continuous variables, respectively, and their associations with sex (females vs. males), variant group (protein-altering [PAV] vs. protein-truncating variants [PTV]), and inheritance (inherited vs. *de novo* variants), respectively. Two-tailed, Bonferroni-adjusted p-value <0.05 was considered significant for all statistical analyses, which were carried out using the IBM SPSS Version 29 software.

Facial photo analysis

To identify whether *KMT2C*-related NDD has a facial gestalt, facial 2D photos from 29 out of the 34 individuals with a LP/P *KMT2C* variant were compared against 29 sex-, age- and ethnicity- matched individuals with NDDs as controls (sampling different controls 5 times) by using PhenoScore as described before[38]. Not all 34 individuals included in this study could be included in the PhenoScore analyses, because a matched control was not found for five individuals.

Briefly, the facial module of PhenoScore utilizes a state-of-the-art convolutional neural network used in facial recognition (QMagFace[39]) that recognizes facial features and allows for objectively evaluating and statistically comparing different NDD facial gestalts.

Similarly, by using PhenoScore, the photos from these *KMT2C*-related NDD individuals were compared against the sex, age, and ethnicity matched individuals with either *EHMT1* or *KMT2D* pathogenic variants to objectively evaluate whether the *KMT2C*-related NDD facial gestalt significantly differs from those with Kleefstra and Kabuki type 1 syndromes, respectively.

RESULTS

KMT2C-related NDD cohort curation and identification of variant spectrum

To systematically analyze the genetic and clinical spectrum of the *KMT2C*-related NDD, we ascertained 98 individuals with rare reported *KMT2C* variants, irrespective of their phenotypes through research databases, international collaborations, and previously published individuals[3,7,9].

75 out of 98 individuals from 67 families had 61 distinct heterozygous *KMT2C* variants predicted to be protein truncating (PTVs) (20 nonsense, 28 frameshift, 5 splice site and 8 large deletions). These PTVs were classified as pathogenic (P) or likely pathogenic (LP) (**Figure 1**), because loss of function is the currently accepted disease mechanism[3,7,9] and the gene is intolerant to loss-of-function variants in population (pLI=1; o/e = 0.08 (0.06 - 0.12))[40]. Of note, 52 of these P/LP variants were *de novo*, 11 were inherited (6 maternal and 5 paternal), and in 12 individuals the inheritance was unknown. Importantly, *KMT2C* is a large gene consisting of 59 exons (NM_170606.3). Ensembl documents 70 transcripts for *KMT2C*, 31 of which are annotated as protein coding: ranging in amino acid length from 4968 (ENST00000682283.1) to 83 (ENST00000684278.1). The MANE select transcript (canonical isoform NM_170606.3/ENST00000262189.11) encodes a protein of 4911 amino acids with mean expression across all adult tissues in the GTEx database of only 0.83 transcripts per million (TPM), while the most highly expressed transcript (5.08 TPM) is ENST00000485655.2 consisting of only the last four exons of the gene[41]. This, however, might be related to the 3' bias of the short-read GTEx data[42]. In our cohort, we found the PTVs are distributed throughout the gene, with the most 5' and the most distal PTVs being located in the 3rd and 57th *KMT2C* exons, suggesting that in spite of apparently relatively low expression of the canonical (MANE) transcript in adult tissues, PTVs across the gene are likely to be pathogenic. We also did not observe specific clustering of PTVs in the gnomAD database (**Figure S1**), but, based on the skewed variant allele frequency, we noted that some of the observed PTVs are either artifacts due to segmental duplications (e.g., c.2710C>T p.(Arg904Ter) and c.1173C>A p.(Cys391Ter)), or somatic variants (e.g., c.6415C>T p.(Arg2139Ter)), as *KMT2C* is known clonal hematopoiesis driver gene[43].

We observed a relatively high proportion of individuals with pathogenic CNVs in our cohort (n=11, 14%; 5/11 are *de novo*). We, therefore, examined the genomic architecture and found that the *KMT2C* contains several segmental duplications with high homology (>98%) to sequences elsewhere in the genome (**Figure 1B**) making this region a "hot-spot" for structural rearrangements[44]. We examined the Database for Genomic Variants (DGV)[45] and gnomAD[40] for structural variants affecting this coding part of the *KMT2C* in the control population in the UCSC genome browser[46] and identified in total 26 partial or intragenic gains and only a single deletion. This suggests that the partial/intragenic gains are unlikely to be pathogenic (if not leading to frameshift). We, however, cannot exclude frequent occurence of artifacts due to the limitations of current technologies in the regions affected by segmental duplications.

In contrast, non-PTVs have not been classified as pathogenic, so 23 out of 98 individuals from 22 families had 22 VUSs (18 missense and 2 splice variants and 2 large duplications). Of those, 10 were *de novo*, 6 were inherited, and 6 were of unknown inheritance.

Collectively, these data show that P/LP truncating *KMT2C* variants are spread across the gene, affecting only MANE or multiple transcripts, and that they can be inherited in some cases. Additionally, clinical classification of *KMT2C* non-truncating variants is currently challenging.

Protein truncating *KMT2C* variants result in a DNAm signature in peripheral blood

We and others have previously shown that pathogenic variants in genes encoding components of epigenetic regulators are associated with genome-wide DNAm changes[47] from which disease-associated DNAm signatures can be derived. Therefore, we performed genome-wide methylation screening on peripheral blood-derived DNA from 16 *KMT2C*-related NDD individuals (discovery cohort) with pathogenic *KMT2C* PTVs and 50 controls using Illumina methylation Epic arrays. We identified 51 significant differentially methylated CpG sites at $|\Delta\beta|>0.10$, and FDR-corrected $p<0.05$ representing the DNAm signature of the *KMT2C*-related NDD (for simplicity, further called *KMT2C* DNAm signature) (**Table S3**). Most of the signature's CpGs were hypomethylated (42/51, 82%). Interestingly, 4/51 (8%) of the signature's CpGs sites mapped to two CpG islands of the WT1 gene (MIM #607102), which were mostly hypomethylated (**Figure 2C**).

Based on the derived *KMT2C* DNAm signature, we were able to discriminate the discovery *KMT2C* cases from healthy controls based on PCA and heatmap plots (**Figures 2A and 2B**, respectively). Next, we trained an SVM model based on the discovery cohort and tested the sensitivity and specificity of the *KMT2C* DNAm signature using 165 controls without known developmental disorders, and 11 additional validation cases with P/LP PTV *KMT2C* variants. On this SVM model, all controls were classified as negative (SVM values <0.25,100% specificity), and 9/11 validation cases were classified as positive (SVM values >0.5, 82% sensitivity) (**Figure 2D**). Two validation cases were classified negatively: individual #48 presented with typical clinical features and has a *de novo* pathogenic frameshift variant c.13107_13108dup p.(Thr4370Argfs*11); individual #52 is mildly affected with multiple affected children and has a small likely pathogenic *KMT2C* exon 36 and 37 deletion resulting in frameshift. However, we cannot exclude that these cases are high-level mosaics in blood.

Next, to test whether *KMT2C*-related NDD affected individuals share the DNAm signature with other conditions, we analyzed all available samples with a LP/P *KMT2C* variant on 17 available DNAm signatures deployed at EpigenCentral[24]. All *KMT2C*-related NDD samples were classified negatively on all signatures **(Figure 2E)**.

These results show that significant changes in methylation patterns of DNA derived from the peripheral blood of individuals with P/LP *KMT2C* variants exist, and that these changes can be used to generate a moderate-effect DNAm signature that does not overlap with the other known disorder-specific DNAm signatures.

A DNAm signature can be used to re-interpret *KMT2C* VUS

Next, we explored if DNAm signature could enable classification of the *KMT2C* VUSs by testing DNA samples from 22/23 individuals (20 unrelated and one proband-father pair) with 21 distinct *KMT2C* VUSs on the *KMT2C* DNAm signature-derived SVM model (one individual with VUS c.14501T>C p.(Val4834Ala) was not available for testing). The SVM model resulted in classification of 5/21 VUS as positive for the *KMT2C* DNAm signature. One out of 21 VUS was classified as intermediate (SVM score 0.25-0.5), and the rest were classified as negative on the *KMT2C* DNAm signature. These results supported re-classification of 6/21 variants as LP and 13/21 variants as benign (described below), while 2/21 variants remained in the VUS category (one additional VUS was unavailable for testing).

Next, we examined the basis of pathogenicity of the PAVs in this cohort. The details for each variant classification and the criteria applied utilizing various evidence are described in **Table S4.** All six variants reclassified as LP/P were classified as positive (or intermediate) on the DNAm signature. Out of the six VUSs reclassified as LP/P, two (*de novo*) variants were (re)interpreted as PTVs: duplication of exons 3-38 which is predicted to result in frameshift, if present in tandem, so based on the DNAm signature classification we concluded that the duplication most likely is indeed in tandem; one variant was initially reported from trio exome sequencing as *de novo* missense c.3499G>A p.(Asp1167Asn), located outside of known domains, but the signature analysis aided the reanalysis of the variant which was later confirmed by Sanger sequencing as c.3499+1G>T p.?. The remaining four LP PAVs (c.3233G>A p.(Cys1078Tyr), c.13229A>G p.(Asp4410Gly), c.13783C>T p.(Arg4595Cys) and c.14335C>G p.(Arg4779Gly)) are absent from gnomAD, predicted as pathogenic by *in silico* tools (REVEL[48] > 0.65 and AlphaMissense[49] > 0.56), and all are located in well-known functional domains of the protein (PHD6, ePHD2, FYRN and SET domains, respectively). Moreover, these variants are predicted to disrupt the 3D structure of the respective domains (**Figure S3**), except for the c.13229A>G p.(Asp4410Gly) for which no reliable 3D structure or model was available.

In contrast, all 11/13 variants classified as benign are missense and are negative on the DNAm signature, the majority are not classified as pathogenic by *in silico* tools and are not predicted to affect protein 3D structure and/or located outside of the known functional domain's (largely in disordered regions) (**Table S4**). Only three benign PAVs occurred *de novo.* Additionally, one splicing variant c.1735+2dup was inherited from father, predicted to result in inframe skipping of exon 12 that encodes a disordered region of the protein and was classified as negative by the DNAm signature, so the variant was classified as benign. Similarly, one benign partial *KMT2C* duplication (1-55 exons) with unknown inheritance was absent from population databases, but similar *KMT2C* 5' partial duplications are common in the general population (**Figure 1B**), and it is also classified as negative on the DNAm signature.

Finally, three missense VUSs remained classified as VUS due conflicting or insufficient evidence. One of the missense variants (c.14501T>C p.(Val4834Ala)) is located in the SET domain, is predicted as pathogenic by *in silico* tools (REVEL = 0.8 and AlphaMissense = 0.99) and was predicted to disrupt the domain's 3D structure (**Table S2)**, but the sample was not available for DNAm testing. Two missense variants (c.9773A>C p.(His3258Pro) and c.13298C>T p.(Ala4433Val)) were classified negatively on the DNAm signature but had additional significant evidence for pathogenicity: both variants are *de novo*, absent in gnomAD and predicted to be possibly disruptive by *in silico* tools and/or 3D protein analysis (**Table S2, S4**).

These results show that the *KMT2C* DNAm signature can be used to reclassify most, but not all VUSs in the gene and that PAVs disrupting functional domains are more likely to be P/LP.

KMT2C loss-of-function variants result in a phenotypically heterogenous NDD syndrome

In total, we identified 81 individuals from 73 families with P/LP *KMT2C* variants. We gathered detailed clinical information on these individuals to understand the clinical spectrum of *KMT2C*-related NDD. Out of 81 individuals, 6 had a second molecular diagnosis due to a pathogenic variant in another NDD-associated gene. To understand the clinical consequences of P/LP *KMT2C* variants, we analyzed the clinical features of 75 individuals with P/LP *KMT2C* variants without any additional known P/LP variants in other NDD genes **(Table 1, S1)**.

In this cohort, 45.3% (n =34) were females, and the median age at last examination was 11 years (with IQR 5.17;20). The median weight at the last reported time was -1.01 SD (-1.94;-0.21), the median height was -1.56 SD (-2.32;-1.19), and the median head circumference (HC) was -0.44 SD (-1.65;0.51). Importantly, short stature (≤-2SD) was reported in ~54% of individuals, and 5 of them have received recombinant human growth hormone (rhGH) with an apparently good response. Also, 9/53 (~17%) of individuals have displayed microcephaly (≤-2SD) at least on one available measurement. These findings suggest *KMT2C*-related NDD is mainly an undergrowth condition.

Neurodevelopmental problems are the most common features in this cohort. Although gross motor delay was described in 87% of individuals, all individuals older than 3.5 years of age had achieved independent walking. Speech delay was reported in 80% of individuals, and ~15% developed mutism at some point in their lifetime. Developmental regression was reported in 6 individuals. ID was reported in ~86% of individuals with half being classified as having moderate or severe ID. Importantly, a proportion of individuals had normal cognitive abilities (~14%; all with *de novo* variants) and some of them were tested due to non-NDD phenotypes (e.g., short stature, seizures). However, we cannot exclude mosaicism in these cases. Furthermore, behavioral, and psychiatric problems were reported in most individuals: features or diagnosis of ASD or ADHD were reported in ~79% and 61% of individuals, respectively; aggressive (both hetero- and self-) behavior was reported in 18% and obsessive/compulsive behavior – in ~15%.

Abnormalities affecting the central nervous system (~37%; e.g., ventriculomegaly, white-matter anomalies, syringomyelia) and the cardiovascular system (~24%; e.g., septal defects, valvular anomalies) were frequent. Refractive errors (33%) and hearing loss (~29%) were the most frequent sensory system problems. Recurrent infections (~26%), central/obstructive sleep apnea (~17%), seizures (~15%), kyphosis and/or scoliosis (~16%), and constipation (~16%) were some of the other most significant and frequently encountered medical issues (**Table 1**). Feeding difficulty was the most frequent neonatal finding (49%), followed by neonatal hypotonia (~29%).

Next, we compared the frequencies of clinical features between male and female individuals, type of variants (PAVs vs. PTVs), and their inheritance (inherited vs. *de novo*). None of the features was significant after correction for multiple testing ($p_{adj}>0.05$).

Craniofacial dysmorphisms were described in 90% of individuals with P/LP *KMT2C* variants in our cohort. Photographs of 34 individuals were available in this cohort, with 13 consenting for publishing **(Figure 3)**. We used these photographs to analyze the facial features of individuals with *KMT2C*-related NDD. The most frequent findings were frontal bossing/prominent-broad forehead, thick/prominent eyebrows, synophrys, down-slanted palpebral fissures, deep-set eyes, epicanthus, hypertelorism, midface retrusion, anteverted nares, thin vermillion of the upper lip, micrognathia, low-set ears, thick ear helices and posteriorly rotated ears. Although all individuals had craniofacial dysmorphism, overall, we did not identify a clearly recognizable *KMT2C* facial gestalt.

Next, to objectively evaluate whether the *KMT2C*-related NDD has a specific facial gestalt, we used PhenoScore and compared the facial photographs of individuals with the *KMT2C*-related NDD with facial photographs of individuals in a general ID cohort. PhenoScore identified a significantly different *KMT*2C-related facial gestalt (AUC = 0.91, p<0.001) (**Figure S2**), indicating that the facial features the 29 individuals with a LP/P *KMT2C* variant for whom an age-, sex- and ethnicity matched NDD control was available are distinguishable from the general NDD population. The PhenoScore identified periorbital and nasal regions as the most different (**Figure S2**), which fit the most common dysmorphic features described above.

Collectively, these observations show that *KMT2C*-related NDD is clinically a highly heterogeneous disorder characterized by intellectual disability, short stature, congenital anomalies, recurrent infections and craniofacial dysmorphism.

***KMT2C*-related NDD is distinct from Kleefstra and Kabuki syndromes**

OMIM designates the *KMT2C*-related NDD as 'Kleefstra syndrome 2' (MIM #617768) which may be interpreted that the *EHMT1* and *KMT2C* related disorders have significant clinical overlap. However, this overlap has not been examined previously. Also, *KMT2C* and *KMT2D* are part of the same gene family, but their overlap has not been examined previously. For this, firstly, we compared the facial features of individuals with *KMT2C*-related NDD to the photographs of 24 and 10 matched individuals with Kleefstra, and Kabuki type 1 syndromes, respectively. PhenoScore revealed significant differences between the *KMT2C*-related NDD facial gestalt and the facial gestalt of Kleefstra syndrome (AUC=1, $p<0.001$) and Kabuki syndrome type 1 (AUC=0.8, $p=0.04$).

DNAm signatures can also be used to delineate epigenetically distinct disorders[50,51]. Pathogenic loss-of-function variants in the *EHMT1*[52] or *KMT2D*[6] result in unique DNAm signatures[51,53]. We analyzed DNA samples from 6 individuals with Kleefstra and 6 individuals with Kabuki 1 syndromes on the *KMT2C* DNAm signature. Samples from Kleefstra and Kabuki type 1 syndromes were classified negatively by the SVM (**Figure 2D; Table S5**) on the *KMT2C* DNAm signature. Similarly, all tested *KMT2C*-related NDD individuals were classified negatively on the Kleefstra and Kabuki syndrome signatures **(Figure 2E)**, highlighting the DNAm differences among the three conditions.

These findings confirm that the *KMT2C*-related NDD is clinically and epigenetically distinct from both Kleefstra and Kabuki syndromes.

DISCUSSION

In this study, we have demonstrated that heterozygous loss-of-function *KMT2C* variants result in a disorder that is characterized by variable combinations of developmental delay, intellectual disability, short stature, congenital anomalies, craniofacial dysmorphism, recurrent infections, and a moderate strength DNAm signature. We also demonstrate that the *KMT2C*-related NDD is clinically and epigenetically distinct from (*EHMT1*-related) Kleefstra syndrome type 1 and (*KMT2D*-related) Kabuki type 1 syndromes.

Our data show that despite the complexities of the transcripts encoded by *KMT2C,* the PTVs in the gene have similar clinical or epigenetic consequences irrespective of their location in the gene. This observation suggests four possible explanations: 1) the canonical isoform is biologically and disease relevant in spite of its apparent low level of expression in tissues, 2) the developmental expression profile of *KMT2C* may be different from its known adult expression profile, 3) the *KMT2C* transcript expression pattern may not yet be fully known, or 4) the regulatory consequences of loss-of-function variants in the longer transcripts may have complex consequences on the transcriptional network. Additionally, we cannot exclude that the low expression of the canonical transcript in GTEx is a technical artifact of the short-read based RNA sequencing, displaying 3' bias and short-isoform preferences[42]. Our analysis also shows that *KMT2C* is located in a hot-spot for both benign and pathogenic structural rearrangements, and, therefore, CNVs in this gene should be interpreted with caution. In our study, we also showed that rare missense variants that affect conserved residues and disrupt the 3D structure of the PHD Zn-fingers, the FYR domain, or the catalytic SET also cause the same disorder. Interestingly, pathogenic Kabuki syndrome Type 1 causing missense variants in *KMT2D,* a paralogue of *KMT2C,* cluster significantly in several similar functional domains [54]. These findings collectively indicate haploinsufficiency and functional haploinsufficiency as likely mechanisms that underlie *KMT2C*-related NDD.

We also identify a moderate-effect DNAm signature for *KMT2C*-NDD with current sensitivity of ~82%. To identify a reliable signature, we had to utilize a relatively large sample size cohort (of different ages and sex) as initial signatures derived from smaller cohort sizes lacked sensitivity and specificity. It is possible that variable expressivity is typical not only for the clinical features, but also for the DNAm signature. It has been shown previously that the level of DNAm changes correlate with the age of onset of dystonia in the *KMT2B*-related dystonia[55] (MIM #617284), and the Weaver-syndrome (MIM #277590) signature was found to be weaker in mildly affected individuals with a pathogenic *EZH2* (MIM #601573) variant even within a single family[18]. Therefore, it is possible that mildly affected individuals can be classified as intermediate or negative on the DNAm signature despite the presence of pathogenic variants. For example, one mildly affected individual with a pathogenic *KMT2C* variant (exons 36-37 deletions) classified as negative by the DNAm signature is the mother of three affected children. This individual is mildly affected which might explain the DNAm classification results. Unfortunately, DNA from the children was not available for testing. However, the second individual with pathogenic variant c.13107_13108dup p.(Thr4370Argfs*11) was also classified negatively despite having typical *KMT2C*-related NDD clinical features. However, we were not able to be exclude mosaicism in either case, which may explain the negative classification results. In the future, larger studies will be required to further elucidate the biological basis of such correlations, which may have prognostic and management implications.

A large proportion of the identified *KMT2C* signature's CpG sites (8%) mapped to the *WT1* gene's CpG islands, which are hypomethylated in *KMT2C*-related NDD individuals. Hypomethylation of CpG islands, especially those located in the promoter of a gene, is usually associated with upregulation of gene expression[56], which was previously shown for *WT1 in vitro*[57]. WT1 is not a known interactor of *KMT2C*, so it is unclear whether the hypomethylation at the *WT1* CpG islands is a primary effect of KMT2C disruption or secondary to disrupted interactions of multiple genes. Loss-of-function or dominant-negative *WT1* variants are associated with Wilms tumor (MIM #194070) or Denys-Drash syndrome (MIM #194080), respectively, but to date there is no human phenotype associated with *WT1* overexpression. Until now, individuals with *KMT2C*-related NDD have not displayed any tumors or anomalies of sex development that are reported for the *WT1*-related disorders. Recently, a role for WT1 in brain and neuron development was also described[58-60]. Mariotinni et al.[61], proved that Wt1 is important for synaptic plasticity and learning in mice, and that it functions as a long-term memory suppressor. Additionally, they showed that Wt1 overexpression causes a reduction of memory retention[61]. In another study, Ji et al., showed that Wt1 brain-specific loss in mice resulted in depressive-like behavior, but the effects of overexpression were not investigated[60]. At this time, the relevance of *WT1* methylation changes to the phenotype of *KMT2C*-related NDD is not known.

Several neurodevelopmental phenotypes were present in the majority of the individuals in the study cohort, including developmental delay, intellectual disability (mostly mild, but ranging from borderline to severe), ASD, and ADHD. Additionally, the individuals often have other psychiatric issues, including hetero- or self-aggressive behavior, obsessive-compulsive behavior, and selective mutism. In fact, pathogenic or *de novo KMT2C* variants have also been found to be enriched in individuals with schizophrenia[62], bipolar disorder[63] and other psychiatric disorders[64]. This suggests that *KMT2C*-related NDD is associated with multiple, overlapping neurodevelopmental and psychiatric presentations.

Individuals with *KMT2C*-related NDD present with a broad spectrum of clinical features, but the expressivity of the features is highly variable among individuals. For example, ~14% of the individuals in the entire cohort did not have intellectual or learning disabilities. The variable expressivity likely explains the relatively large proportion (11/75, ~15%) of inherited cases in this cohort, as most of the parents with pathogenic variants were mildly affected.

Individuals with the *KMT2C*-related NDD had common dysmorphic features, including frontal bossing or prominent forehead, downslanted palpebral fissures, deep-set eyes, and hypertelorism, among others. These features are mild and non-specific, and therefore, the facial gestalt was unrecognizable to clinicians, unlike Kleefstra and Kabuki 1 syndromes. However, using PhenoScore[38], we were able to show that the *KMT2C*-related NDD had a typical facial gestalt that is significantly different from Kleefstra and Kabuki 1 syndromes.

KMT2C-related NDD, Kleefstra, and Kabuki 1 syndromes are clearly different conditions despite some clinical and molecular overlap, but it is currently impossible to objectively compare the frequencies of specific features among these conditions head-to-head. Kleefstra and Kabuki syndromes are clinically well-known and well-recognized, and most of the published individuals are diagnosed through a phenotype-first approach[8,65]. This results in biased results with only the most typical affected individuals being initially described. However, with the recent widespread application of exome or genome sequencing in clinical diagnostics, it is possible to also identify non-specific and/or mildly affected individuals[66,67], as well as novel morbid entities within the same gene[50], including *KMT2D*[68]. *KMT2C*-related NDD is not clinically recognizable, so all individuals were diagnosed through a genotype-first approach and this study represents a clinical expansion and delineation of this condition. Therefore, studies on the phenotype of Kleefstra and Kabuki 1 syndromes diagnosed through a genotype-first approach would be necessary to better understand their clinical and molecular spectrum, as well as the differences among the three conditions.

Although *KMT2C*-related NDD, Kleefstra and Kabuki 1 syndrome are caused by haploinsufficiency of methyltransferases, where PTVs are the most frequent type of variant, the proportion of inherited variants is considerably different among them. While Kleefstra and Kabuki 1 syndrome mostly occur by *de novo* loss-of-function variants, with only a handful of inherited cases described so far[69,unpublished data], ~15% of *KMT2C*-related NDD individuals inherited their variants from similarly affected parents. Albeit this may be explained by a biased approach to diagnosis, our study also showed that the expressivity of the *KMT2C*-related NDD is highly variable, and individuals with few manifestations can be expected.

Several individuals in our study were primarily investigated due to short stature - a feature present in more than a half of the *KMT2C*-related NDD individuals reported in this study. Short stature is also common in Kleefstra and Kabuki 1 syndromes[8,65]. The causes of short stature in these syndromes are often unknown but likely complex[65]. However, short stature in individuals with monogenic NDDs is rarely investigated in-depth in routine clinical practice as it is usually considered to be part of the syndrome. While it is known that growth hormone deficiency is present in approximately a third of individuals with Kabuki 1 syndrome[70], this can partially explain the presence of short stature. However, treatment with rhGH has proven to be highly effective and safe[71], and therefore, these individuals should be screened for this deficiency. In this cohort, 40/75 indidivuals with endocrine anomalies had proven short stature, and 6/40 had received rhGH with apparently good results. This should prompt further studies on the prevalence of GH deficiency and the utility of rhGH therapy also for individuals with *KMT2C*-related NDD and short stature.

While various psychiatric disorders are seen in *KMT2C*-related NDD, this is not frequent in Kabuki 1 syndrome[72]. Also, although psychotic disorders such as schizophrenia[73] and developmental regression[74-76] have been frequently reported in Kleefstra syndrome, nearly all individuals with Kleefstra syndrome manifest intellectual disability that ranges from moderate to severe[8,76], which contrasts with our finding that ~15% of individuals with *KMT2C*-related NDD demonstrate no cognitive impairment. Again, albeit this may be explained by a bias in our diagnostic approach, our study shows that the penetrance and severity of this finding is consistently reduced in the *KMT2C*-related NDD.

In summary, pathogenic variants in the *KMT2C* gene result in a distinct syndromic neurodevelopmental disorder, which is different from Kleefstra and Kabuki 1 syndromes. Therefore, the OMIM designation of *KMT2C*-related NDD as Kleefstra syndrome 2 should be reconsidered, as it may be misleading to clinicians and patients' families.

DECLARATION OF INTERESTS

RW is a consultant (equity) for Alamya Health. The rest of the authors declare no competing interests.

ACKNOWLEDGEMENTS

Described in the Supplemental information.

WEB RESOURCES

https://epigen.ccm.sickkids.ca

DATA AND CODE AVAILABILITY

Recruited individuals' with pathogenic *KMT2C* variants clinical details are provided in the **Table S1**. Recruited individuals' with benign or VUS *KMT2C* variants details are provided in the **Table S2**. The variants and their interpretations were submitted to the ClinVar database (ClinVar accession numbers: SCV005044911-SCV005044983 and SCV005044991-SCV005045000).

The *KMT2C* DNAm signature is available in supplementary material (**Table S3**). The raw datasets supporting the current study have not been deposited in a public repository due to institutional ethics restrictions. All software and R packages used in the study are publicly available as described in the methods section. The *KMT2C* methylation classifier will be made available through EpigenCentral at https://epigen.ccm.sickkids.ca.

REFERENCES

1. Kleefstra, T., Schenck, A., Kramer, J.M., and van Bokhoven, H. (2014). The genetics of cognitive epigenetics. Neuropharmacology *80*, 83-94. 10.1016/j.neuropharm.2013.12.025.
2. Cenik, B.K., and Shilatifard, A. (2021). COMPASS and SWI/SNF complexes in development and disease. Nat Rev Genet *22*, 38-58. 10.1038/s41576-020-0278-0.
3. Koemans, T.S., Kleefstra, T., Chubak, M.C., Stone, M.H., Reijnders, M.R.F., de Munnik, S., Willemsen, M.H., Fenckova, M., Stumpel, C., Bok, L.A., et al. (2017). Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. PLoS Genet *13*, e1006864. 10.1371/journal.pgen.1006864.
4. Hu, D., Gao, X., Morgan, M.A., Herz, H.M., Smith, E.R., and Shilatifard, A. (2013). The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol Cell Biol *33*, 4745-4754. 10.1128/mcb.01181-13.
5. Rampias, T., Karagiannis, D., Avgeris, M., Polyzos, A., Kokkalis, A., Kanaki, Z., Kousidou, E., Tzetis, M., Kanavakis, E., Stravodimos, K., et al. (2019). The lysine-specific methyltransferase KMT2C/MLL3 regulates DNA repair components in cancer. EMBO Rep *20*. 10.15252/embr.201846821.
6. Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C., et al. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet *42*, 790-793. 10.1038/ng.646.
7. Kleefstra, T., Kramer, J.M., Neveling, K., Willemsen, M.H., Koemans, T.S., Vissers, L.E., Wissink-Lindhout, W., Fenckova, M., van den Akker, W.M., Kasri, N.N., et al. (2012). Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. Am J Hum Genet *91*, 73-82. 10.1016/j.ajhg.2012.05.003.
8. Willemsen, M.H., Vulto-van Silfhout, A.T., Nillesen, W.M., Wissink-Lindhout, W.M., van Bokhoven, H., Philip, N., Berry-Kravis, E.M., Kini, U., van Ravenswaaij-Arts, C.M., Delle Chiaie, B., et al. (2012). Update on Kleefstra Syndrome. Mol Syndromol *2*, 202-212. 10.1159/000335648.
9. Faundes, V., Newman, W.G., Bernardini, L., Canham, N., Clayton-Smith, J., Dallapiccola, B., Davies, S.J., Demos, M.K., Goldman, A., Gill, H., et al. (2018). Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. Am J Hum Genet *102*, 175-187. 10.1016/j.ajhg.2017.11.013.
10. Schobers, G., Schieving, J.H., Yntema, H.G., Pennings, M., Pfundt, R., Derks, R., Hofste, T., de Wijs, I., Wieskamp, N., van den Heuvel, S., et al. (2022). Reanalysis of exome negative patients with rare disease: a pragmatic workflow for diagnostic applications. Genome Med *14*, 66. 10.1186/s13073-022-01069-z.
11. Prevalence and architecture of de novo mutations in developmental disorders. (2017). Nature *542*, 433-438. 10.1038/nature21062.
12. Smedley, D., Smith, K.R., Martin, A., Thomas, E.A., McDonagh, E.M., Cipriani, V., Ellingford, J.M., Arno, G., Tucci, A., Vandrovcova, J., et al. (2021). 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. N Engl J Med *385*, 1868-1880. 10.1056/NEJMoa2035790.
13. Fischbach, G.D., and Lord, C. (2010). The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neuron *68*, 192-195. 10.1016/j.neuron.2010.10.006.
14. Trost, B., Thiruvahindrapuram, B., Chan, A.J.S., Engchuan, W., Higginbotham, E.J., Howe, J.L., Loureiro, L.O., Reuter, M.S., Roshandel, D., Whitney, J., et al. (2022). Genomic architecture of autism from comprehensive whole-genome sequence annotation. Cell *185*, 4409-4427.e4418. 10.1016/j.cell.2022.10.009.
15. Morales, J., Pujar, S., Loveland, J.E., Astashyn, A., Bennett, R., Berry, A., Cox, E., Davidson, C., Ermolaeva, O., Farrell, C.M., et al. (2022). A joint NCBI and EMBL-EBI

transcript set for clinical genomics and research. Nature *604*, 310-315. 10.1038/s41586-022-04558-8.

16. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med *17*, 405-424. 10.1038/gim.2015.30.
17. Choufani, S., McNiven, V., Cytrynbaum, C., Jangjoo, M., Adam, M.P., Bjornsson, H.T., Harris, J., Dyment, D.A., Graham, G.E., Nezarati, M.M., et al. (2022). An HNRNPK-specific DNA methylation signature makes sense of missense variants and expands the phenotypic spectrum of Au-Kline syndrome. Am J Hum Genet *109*, 1867-1884. 10.1016/j.ajhg.2022.08.014.
18. Choufani, S., Gibson, W.T., Turinsky, A.L., Chung, B.H.Y., Wang, T., Garg, K., Vitriolo, A., Cohen, A.S.A., Cyrus, S., Goodman, S., et al. (2020). DNA Methylation Signature for EZH2 Functionally Classifies Sequence Variants in Three PRC2 Complex Genes. Am J Hum Genet *106*, 596-610. 10.1016/j.ajhg.2020.03.008.
19. Awamleh, Z., Choufani, S., Cytrynbaum, C., Alkuraya, F.S., Scherer, S., Fernandes, S., Rosas, C., Louro, P., Dias, P., Neves, M.T., et al. (2023). ANKRD11 pathogenic variants and 16q24.3 microdeletions share an altered DNA methylation signature in patients with KBG syndrome. Hum Mol Genet *32*, 1429-1438. 10.1093/hmg/ddac289.
20. Fortin, J.P., Triche, T.J., Jr., and Hansen, K.D. (2017). Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics *33*, 558-560. 10.1093/bioinformatics/btw691.
21. LaBarre, B.A., Goncearenco, A., Petrykowska, H.M., Jaratlerdsiri, W., Bornman, M.S.R., Hayes, V.M., and Elnitski, L. (2019). MethylToSNP: identifying SNPs in Illumina DNA methylation array data. Epigenetics Chromatin *12*, 79. 10.1186/s13072-019-0321-6.
22. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47. 10.1093/nar/gkv007.
23. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software *28*, 1 - 26. 10.18637/jss.v028.i05.
24. Turinsky, A.L., Choufani, S., Lu, K., Liu, D., Mashouri, P., Min, D., Weksberg, R., and Brudno, M. (2020). EpigenCentral: Portal for DNA methylation data analysis and classification in rare diseases. Hum Mutat *41*, 1722-1733. 10.1002/humu.24076.
25. Awamleh, Z., Goodman, S., Kallurkar, P., Wu, W., Lu, K., Choufani, S., Turinsky, A.L., and Weksberg, R. (2022). Generation of DNA Methylation Signatures and Classification of Variants in Rare Neurodevelopmental Disorders Using EpigenCentral. Curr Protoc *2*, e597. 10.1002/cpz1.597.
26. Liu, Y., Qin, S., Chen, T.Y., Lei, M., Dhar, S.S., Ho, J.C., Dong, A., Loppnau, P., Li, Y., Lee, M.G., and Min, J. (2019). Structural insights into trans-histone regulation of H3K4 methylation by unique histone H4 binding of MLL3/4. Nat Commun *10*, 36. 10.1038/s41467-018-07906-3.
27. Xue, H., Yao, T., Cao, M., Zhu, G., Li, Y., Yuan, G., Chen, Y., Lei, M., and Huang, J. (2019). Structural basis of nucleosome recognition and modification by MLL methyltransferases. Nature *573*, 445-449. 10.1038/s41586-019-1528-1.
28. Li, Y., Han, J., Zhang, Y., Cao, F., Liu, Z., Li, S., Wu, J., Hu, C., Wang, Y., Shuai, J., et al. (2016). Structural basis for activity regulation of MLL family methyltransferases. Nature *530*, 447-452. 10.1038/nature16952.
29. García-Alai, M.M., Allen, M.D., Joerger, A.C., and Bycroft, M. (2010). The structure of the FYR domain of transforming growth factor beta regulator 1. Protein Sci *19*, 1432-1438. 10.1002/pro.404.
30. Liu, Z., Li, F., Ruan, K., Zhang, J., Mei, Y., Wu, J., and Shi, Y. (2014). Structural and functional insights into the human Börjeson-Forssman-Lehmann syndrome-associated protein PHF6. J Biol Chem *289*, 10069-10083. 10.1074/jbc.M113.535351.

31. Klein, B.J., Cox, K.L., Jang, S.M., Côté, J., Poirier, M.G., and Kutateladze, T.G. (2020). Molecular Basis for the PZP Domain of BRPF1 Association with Chromatin. Structure *28*, 105-110.e103. 10.1016/j.str.2019.10.014.

32. Klein, B.J., Deshpande, A., Cox, K.L., Xuan, F., Zandian, M., Barbosa, K., Khanal, S., Tong, Q., Zhang, Y., Zhang, P., et al. (2021). The role of the PZP domain of AF10 in acute leukemia driven by AF10 translocations. Nat Commun *12*, 4130. 10.1038/s41467-021-24418-9.

33. Zheng, S., Bi, Y., Chen, H., Gong, B., Jia, S., and Li, H. (2021). Molecular basis for bipartite recognition of histone H3 by the PZP domain of PHF14. Nucleic Acids Res *49*, 8961-8973. 10.1093/nar/gkab670.

34. Krieger, E., and Vriend, G. (2014). YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics *30*, 2981-2982. 10.1093/bioinformatics/btu426.

35. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589. 10.1038/s41586-021-03819-2.

36. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res *50*, D439-d444. 10.1093/nar/gkab1061.

37. Faundes, V., Goh, S., Akilapa, R., Bezuidenhout, H., Bjornsson, H.T., Bradley, L., Brady, A.F., Brischoux-Boucher, E., Brunner, H., Bulk, S., et al. (2021). Clinical delineation, sex differences, and genotype-phenotype correlation in pathogenic KDM6A variants causing X-linked Kabuki syndrome type 2. Genet Med *23*, 1202-1210. 10.1038/s41436-021-01119-8.

38. Dingemans, A.J.M., Hinne, M., Truijen, K.M.G., Goltstein, L., van Reeuwijk, J., de Leeuw, N., Schuurs-Hoeijmakers, J., Pfundt, R., Diets, I.J., den Hoed, J., et al. (2023). PhenoScore quantifies phenotypic variation for rare genetic diseases by combining facial analysis with other clinical features using a machine-learning framework. Nat Genet *55*, 1598-1607. 10.1038/s41588-023-01469-w.

39. Terhorst, P., Ihlefeld, M., Huber, M., Damer, N., Kirchbuchner, F., Raja, K.B., and Kuijper, A. (2021). QMagFace: Simple and Accurate Quality-Aware Face Recognition. 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 3473-3483.

40. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434-443. 10.1038/s41586-020-2308-7.

41. The Genotype-Tissue Expression (GTEx) project. (2013). Nat Genet *45*, 580-585. 10.1038/ng.2653.

42. Cummings, B.B., Karczewski, K.J., Kosmicki, J.A., Seaby, E.G., Watts, N.A., Singer-Berk, M., Mudge, J.M., Karjalainen, J., Satterstrom, F.K., O'Donnell-Luria, A.H., et al. (2020). Transcript expression-aware annotation improves rare variant interpretation. Nature *581*, 452-458. 10.1038/s41586-020-2329-2.

43. Pich, O., Reyes-Salazar, I., Gonzalez-Perez, A., and Lopez-Bigas, N. (2022). Discovering the drivers of clonal hematopoiesis. Nat Commun *13*, 4267. 10.1038/s41467-022-31878-0.

44. Sahoo, T., Theisen, A., Rosenfeld, J.A., Lamb, A.N., Ravnan, J.B., Schultz, R.A., Torchia, B.S., Neill, N., Casci, I., Bejjani, B.A., and Shaffer, L.G. (2011). Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. Genet Med *13*, 868-880. 10.1097/GIM.0b013e3182217a06.

45. MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res *42*, D986-992. 10.1093/nar/gkt958.

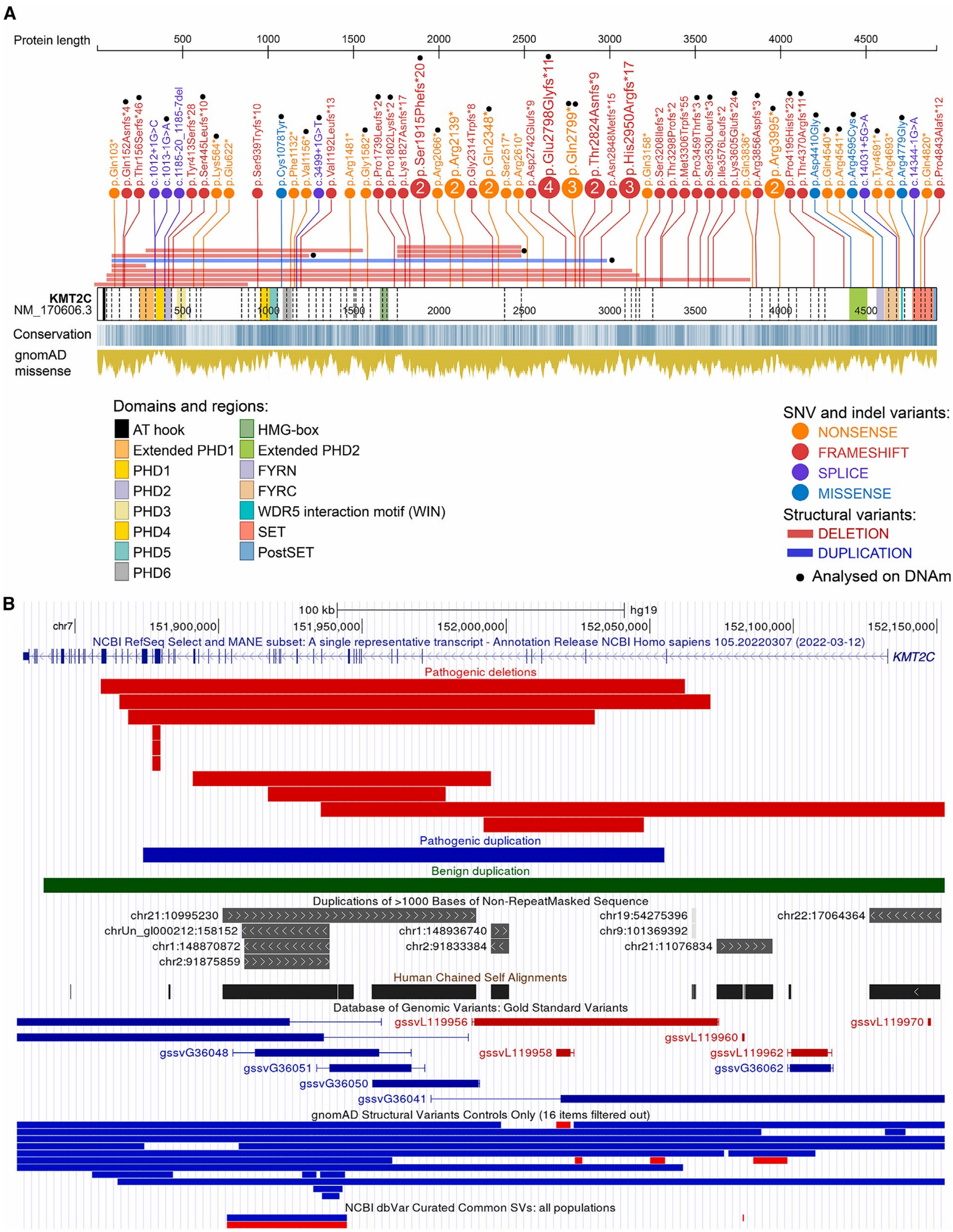
46. Nassar, L.R., Barber, G.P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., et al. (2023). The UCSC Genome Browser database: 2023 update. Nucleic Acids Res *51*, D1188-d1195. 10.1093/nar/gkac1072.
47. Chater-Diehl, E., Goodman, S.J., Cytrynbaum, C., Turinsky, A.L., Choufani, S., and Weksberg, R. (2021). Anatomy of DNA methylation signatures: Emerging insights and applications. Am J Hum Genet *108*, 1359-1366. 10.1016/j.ajhg.2021.06.015.
48. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet *99*, 877-885. 10.1016/j.ajhg.2016.08.016.
49. Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science *381*, eadg7492. 10.1126/science.adg7492.
50. Rots, D., Chater-Diehl, E., Dingemans, A.J.M., Goodman, S.J., Siu, M.T., Cytrynbaum, C., Choufani, S., Hoang, N., Walker, S., Awamleh, Z., et al. (2021). Truncating SRCAP variants outside the Floating-Harbor syndrome locus cause a distinct neurodevelopmental disorder with a specific DNA methylation signature. Am J Hum Genet *108*, 1053-1068. 10.1016/j.ajhg.2021.04.008.
51. Butcher, D.T., Cytrynbaum, C., Turinsky, A.L., Siu, M.T., Inbar-Feigenberg, M., Mendoza-Londono, R., Chitayat, D., Walker, S., Machado, J., Caluseriu, O., et al. (2017). CHARGE and Kabuki Syndromes: Gene-Specific DNA Methylation Signatures Identify Epigenetic Mechanisms Linking These Clinically Overlapping Conditions. Am J Hum Genet *100*, 773-788. 10.1016/j.ajhg.2017.04.004.
52. Kleefstra, T., Brunner, H.G., Amiel, J., Oudakker, A.R., Nillesen, W.M., Magee, A., Geneviève, D., Cormier-Daire, V., van Esch, H., Fryns, J.P., et al. (2006). Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. Am J Hum Genet *79*, 370-377. 10.1086/505693.
53. Goodman, S., Cytrynbaum, C., Chung, B., Chater-Diehl, E., Aziz, C., Turinsky, A., Kellam, B., Keller, M., Ko, J.M., Caluseriu, O., et al. (2020). EHMT1 pathogenic variants and 9q34.3 microdeletions share altered DNA methylation patterns in patients with Kleefstra syndrome. Journal of Translational Genetics and Genomics *4*, 144-158. 10.20517/jtgg.2020.23.
54. Faundes, V., Malone, G., Newman, W.G., and Banka, S. (2019). A comparative analysis of KMT2D missense variants in Kabuki syndrome, cancers and the general population. J Hum Genet *64*, 161-170. 10.1038/s10038-018-0536-6.
55. Mirza-Schreiber, N., Zech, M., Wilson, R., Brunet, T., Wagner, M., Jech, R., Boesch, S., Škorvánek, M., Necpál, J., Weise, D., et al. (2022). Blood DNA methylation provides an accurate biomarker of KMT2B-related dystonia and predicts onset. Brain *145*, 644-654. 10.1093/brain/awab360.
56. Garg, P., Jadhav, B., Rodriguez, O.L., Patel, N., Martin-Trujillo, A., Jain, M., Metsu, S., Olsen, H., Paten, B., Ritz, B., et al. (2020). A Survey of Rare Epigenetic Variation in 23,116 Human Genomes Identifies Disease-Relevant Epivariations and CGG Expansions. Am J Hum Genet *107*, 654-669. 10.1016/j.ajhg.2020.08.019.
57. Hamatani, H., Sakairi, T., Ikeuchi, H., Kaneko, Y., Maeshima, A., Nojima, Y., and Hiromura, K. (2019). TGF-β1 alters DNA methylation levels in promoter and enhancer regions of the WT1 gene in human podocytes. Nephrology (Carlton) *24*, 575-584. 10.1111/nep.13411.
58. Schnerwitzki, D., Perry, S., Ivanova, A., Caixeta, F.V., Cramer, P., Günther, S., Weber, K., Tafreshiha, A., Becker, L., Vargas Panesso, I.L., et al. (2018). Neuron-specific inactivation of Wt1 alters locomotion in mice and changes interneuron composition in the spinal cord. Life Sci Alliance *1*, e201800106. 10.26508/lsa.201800106.

59. Schnerwitzki, D., Hayn, C., Perner, B., and Englert, C. (2020). Wt1 Positive dB4 Neurons in the Hindbrain Are Crucial for Respiration. Front Neurosci *14*, 529487. 10.3389/fnins.2020.529487.
60. Ji, F., Wang, W., Feng, C., Gao, F., and Jiao, J. (2021). Brain-specific Wt1 deletion leads to depressive-like behaviors in mice via the recruitment of Tet2 to modulate Epo expression. Mol Psychiatry *26*, 4221-4233. 10.1038/s41380-020-0759-8.
61. Mariottini, C., Munari, L., Gunzel, E., Seco, J.M., Tzavaras, N., Hansen, J., Stern, S.A., Gao, V., Aleyasin, H., Sharma, A., et al. (2019). Wilm's tumor 1 promotes memory flexibility. Nat Commun *10*, 3756. 10.1038/s41467-019-11781-x.
62. Rees, E., Han, J., Morgan, J., Carrera, N., Escott-Price, V., Pocklington, A.J., Duffield, M., Hall, L.S., Legge, S.E., Pardiñas, A.F., et al. (2020). De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. Nat Neurosci *23*, 179-184. 10.1038/s41593-019-0565-2.
63. Nishioka, M., Kazuno, A.A., Nakamura, T., Sakai, N., Hayama, T., Fujii, K., Matsuo, K., Komori, A., Ishiwata, M., Watanabe, Y., et al. (2021). Systematic analysis of exonic germline and postzygotic de novo mutations in bipolar disorder. Nat Commun *12*, 3750. 10.1038/s41467-021-23453-w.
64. Li, K., Ling, Z., Luo, T., Zhao, G., Zhou, Q., Wang, X., Xia, K., Li, J., and Li, B. (2021). Cross-Disorder Analysis of De Novo Variants Increases the Power of Prioritising Candidate Genes. Life (Basel) *11*. 10.3390/life11030233.
65. Schott, D.A., Blok, M.J., Gerver, W.J., Devriendt, K., Zimmermann, L.J., and Stumpel, C.T. (2016). Growth pattern in Kabuki syndrome with a KMT2D mutation. Am J Med Genet A *170*, 3172-3179. 10.1002/ajmg.a.37930.
66. Rosina, E., Pezzani, L., Pezzoli, L., Marchetti, D., Bellini, M., Pilotta, A., Calabrese, O., Nicastro, E., Cirillo, F., Cereda, A., et al. (2022). Atypical, Composite, or Blended Phenotypes: How Different Molecular Mechanisms Could Associate in Double-Diagnosed Patients. Genes (Basel) *13*. 10.3390/genes13071275.
67. Dyment, D.A., Tétreault, M., Beaulieu, C.L., Hartley, T., Ferreira, P., Chardon, J.W., Marcadier, J., Sawyer, S.L., Mosca, S.J., Innes, A.M., et al. (2015). Whole-exome sequencing broadens the phenotypic spectrum of rare pediatric epilepsy: a retrospective study. Clin Genet *88*, 34-40. 10.1111/cge.12464.
68. Cuvertino, S., Hartill, V., Colyer, A., Garner, T., Nair, N., Al-Gazali, L., Canham, N., Faundes, V., Flinter, F., Hertecant, J., et al. (2020). A restricted spectrum of missense KMT2D variants cause a multiple malformations disorder distinct from Kabuki syndrome. Genet Med *22*, 867-877. 10.1038/s41436-019-0743-3.
69. Bögershausen, N., Gatinois, V., Riehmer, V., Kayserili, H., Becker, J., Thoenes, M., Simsek-Kiper, P., Barat-Houari, M., Elcioglu, N.H., Wieczorek, D., et al. (2016). Mutation Update for Kabuki Syndrome Genes KMT2D and KDM6A and Further Delineation of X-Linked Kabuki Syndrome Subtype 2. Hum Mutat *37*, 847-864. 10.1002/humu.23026.
70. Schott, D.A., Gerver, W.J., and Stumpel, C.T. (2016). Growth Hormone Stimulation Tests in Children with Kabuki Syndrome. Horm Res Paediatr *86*, 319-324. 10.1159/000449221.
71. van Montfort, L., Gerver, W.J.M., Kooger, B.L.S., Plat, J., Bierau, J., Stumpel, C., and Schott, D.A. (2021). Follow-Up Study of Growth Hormone Therapy in Children with Kabuki Syndrome: Two-Year Treatment Results. Horm Res Paediatr *94*, 285-296. 10.1159/000519963.
72. Barry, K.K., Tsaparlis, M., Hoffman, D., Hartman, D., Adam, M.P., Hung, C., and Bodamer, O.A. (2022). From Genotype to Phenotype-A Review of Kabuki Syndrome. Genes (Basel) *13*. 10.3390/genes13101761.
73. Vermeulen, K., Staal, W.G., Janzing, J.G., van Bokhoven, H., Egger, J.I.M., and Kleefstra, T. (2017). Sleep Disturbance as a Precursor of Severe Regression in Kleefstra Syndrome Suggests a Need for Firm and Rapid Pharmacological Treatment. Clin Neuropharmacol *40*, 185-188. 10.1097/wnf.0000000000000226.
74. Vermeulen, K., de Boer, A., Janzing, J.G.E., Koolen, D.A., Ockeloen, C.W., Willemsen, M.H., Verhoef, F.M., van Deurzen, P.A.M., van Dongen, L., van

Bokhoven, H., et al. (2017). Adaptive and maladaptive functioning in Kleefstra syndrome compared to other rare genetic disorders with intellectual disabilities. Am J Med Genet A *173*, 1821-1830. 10.1002/ajmg.a.38280.
75. Verhoeven, W.M., Egger, J.I., Vermeulen, K., van de Warrenburg, B.P., and Kleefstra, T. (2011). Kleefstra syndrome in three adult patients: further delineation of the behavioral and neurological phenotype shows aspects of a neurodegenerative course. Am J Med Genet A *155a*, 2409-2415. 10.1002/ajmg.a.34186.
76. Morison, L.D., Kennis, M.G.P., Rots, D., Bouman, A., Kummeling, J., Palmer, E., Vogel, A.P., Liegeois, F., Brignell, A., Srivastava, S., et al. (2024). Expanding the phenotype of Kleefstra syndrome: speech, language and cognition in 103 individuals. Journal of Medical Genetics, jmg-2023-109702. 10.1136/jmg-2023-109702.

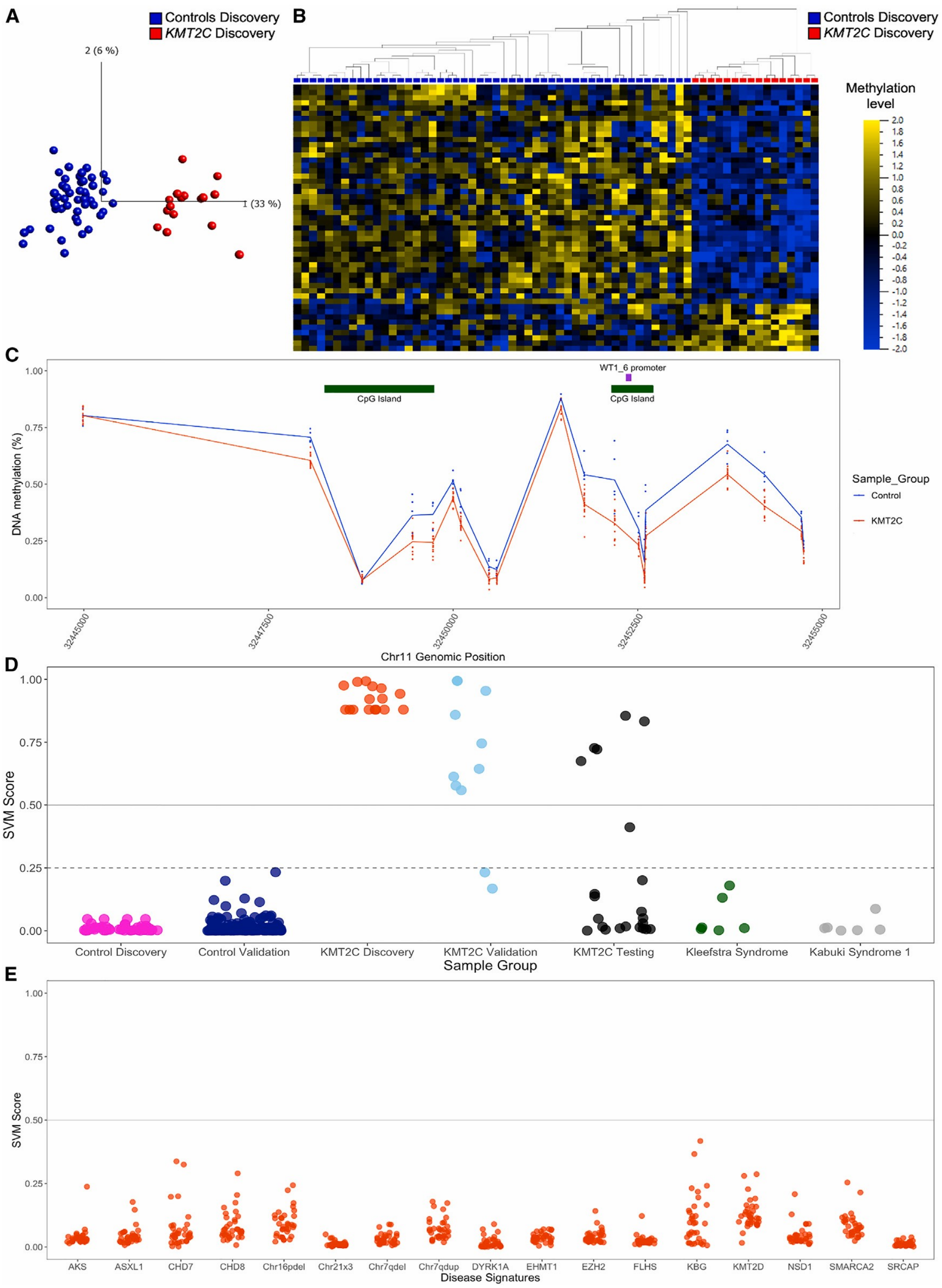
FIGURE AND TABLE LEGENDS:

Figure 1. *KMT2C* variant spectrum among the recruited individuals and population.



***A)** Likely pathogenic and pathogenic single nucleotide, indel and copy number variants identified in this study are shown on the linear structure of the KMT2C protein. **B)** Likely pathogenic and pathogenic, as well as benign copy number variants identified in this study in* *comparison shown on the KMT2C genomic region with segmental duplications (grey), as well as populational deletions and duplications from DGV and gnomAD databases.*

Figure 2. DNAm signature classification results.



***A)** PCA plot for the discovery cohort of 16 KMT2C-related NDD cases (red) and 50 controls (blue). **B)** Heatmap plot with hierarchical clustering for the discovery cohort of 16 KMT2C-related NDD cases (red) and 50 controls (blue), with hypermethylated sites shown in yellow* *and hypomethylated ones in blue. **C)** Differentially methylated region which maps to several WT1 CpG islands (green) and one of the promoters (purple) (data obtained from the UCSC genome browser) where each dot shows methylation level at a CpG site in the region for each discovery cohort sample, and the line depicts the mean methylation for controls (blue) and cases (red). **D)** SVM model classification results for different groups: 16 discovery KMT2C-related NDD cases (red), 50 controls (magenta); validation KMT2C-related NDD cases (light blue) and their matched controls (dark blue); KMT2C VUSs (black); Kleefstra syndrome individuals (green) and Kabuki syndrome individuals (grey). **E)** KMT2C-related NDD classification results against 17 other DNAm signatures.*

Figure 3. Photos of individuals with the *KMT2C*-related NDD.

y.o. = years old – the age of the individuals.

Table 1. Frequencies or distributions of clinical findings of *KMT2C*-related NDD.

Table 1. Continued

Clinical finding[a]	Frequency or distribution
Respiratory anomalies (n [%]) (n¼42)	14 (33.3)
Central/obstructive sleep apnea (n [%])	7 (16.7)
Asthma (n [%])	6 (14.3)
Other (n [%])	2 (4.8)
Palate anomalies (n [%]) (n¼37)	12 (32.4)
High/narrow palate (n [%])	11 (29.7)
Cleft lip/palate (n [%])	1 (2.7)
Cardiovascular anomalies (n [%]) (n¼54)	13 (24.1)
Septal defects (n[%])	6 (11.1)
Conduction disorder (n[%])	6 (11.1)
Valvular anomalies (n [%])	4 (7.4)
Other (n [%])	3 (5.6)
Genitourinary anomalies (n [%]) (n¼39)	8 (20.5)
Uni/bilateral inguinal hernia (n [%])	2 (5.1)
Other (n [%])	7 (17.9)
Dental anomalies (n [%]) (n¼70)	13 (18.8)
Dental crowding (n [%])	6 (8.6)
Prominent upper incisors (n [%])	6 (8.6)
Other (n [%])	1 (1.4)

[a] The number of responders are detailed for every feature, and their frequencies/distributions were calculated according to that number. Individuals with another NDD were excluded. at any momemt documented <-2SD.*

ADHD=Attention Deficit/Hyperactivity Disorder; ASD=Autism Spectrum Disorder; CNS=Central nervous system; IQR=Interquartile range m=median; SD=Standard deviation

SUPPLEMENTAL INFORMATION

Table S1. Recruited individual with *KMT2C*-related NDD detailed clinical description with variants.

Table S2. Recruited individual with benign/VUS *KMT2C* variants detailed clinical description with variants and their classification.

Table S3. *KMT2C* DNAm signature CpG sites.

Table S4. *KMT2C* VUS classification.

Blue = evidence for pathogenicity; Green = benign evidence; Grey – unknown or uncertain significance

Table S5. Description of Kleefstra and Kabuki syndrome 1 samples used for the testing of the *KMT2C* signature.

Figure S1. Distribution of the PTVs across *KMT2C* transcripts with their relative expression in the gnomAD database.

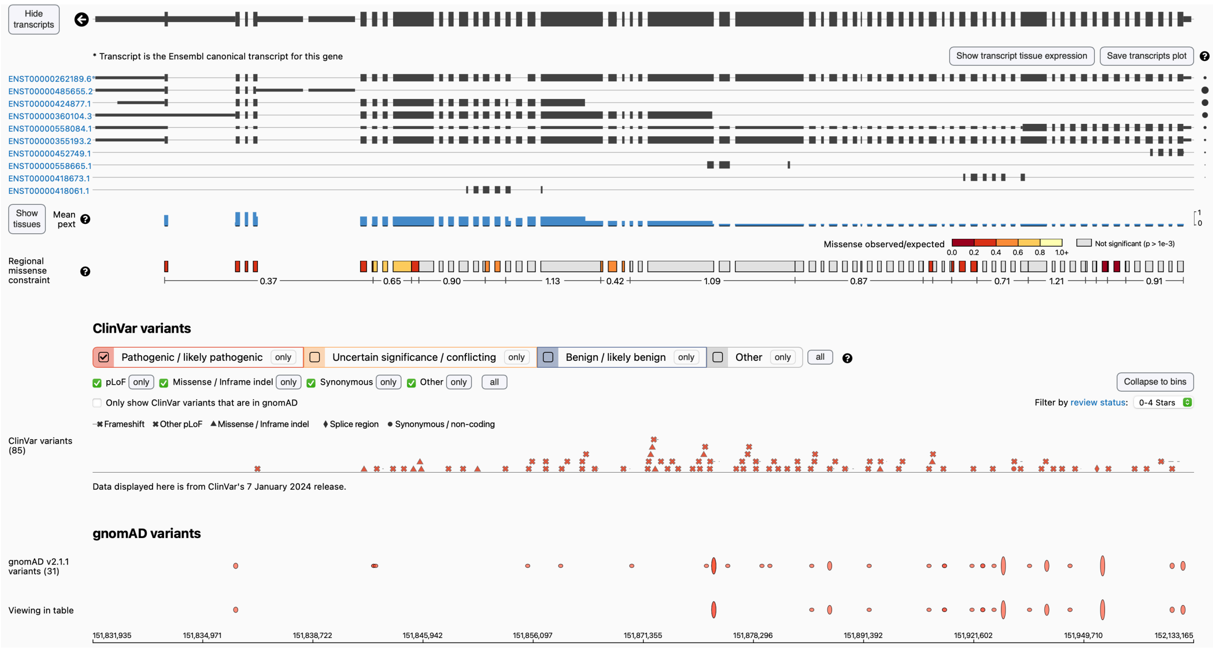
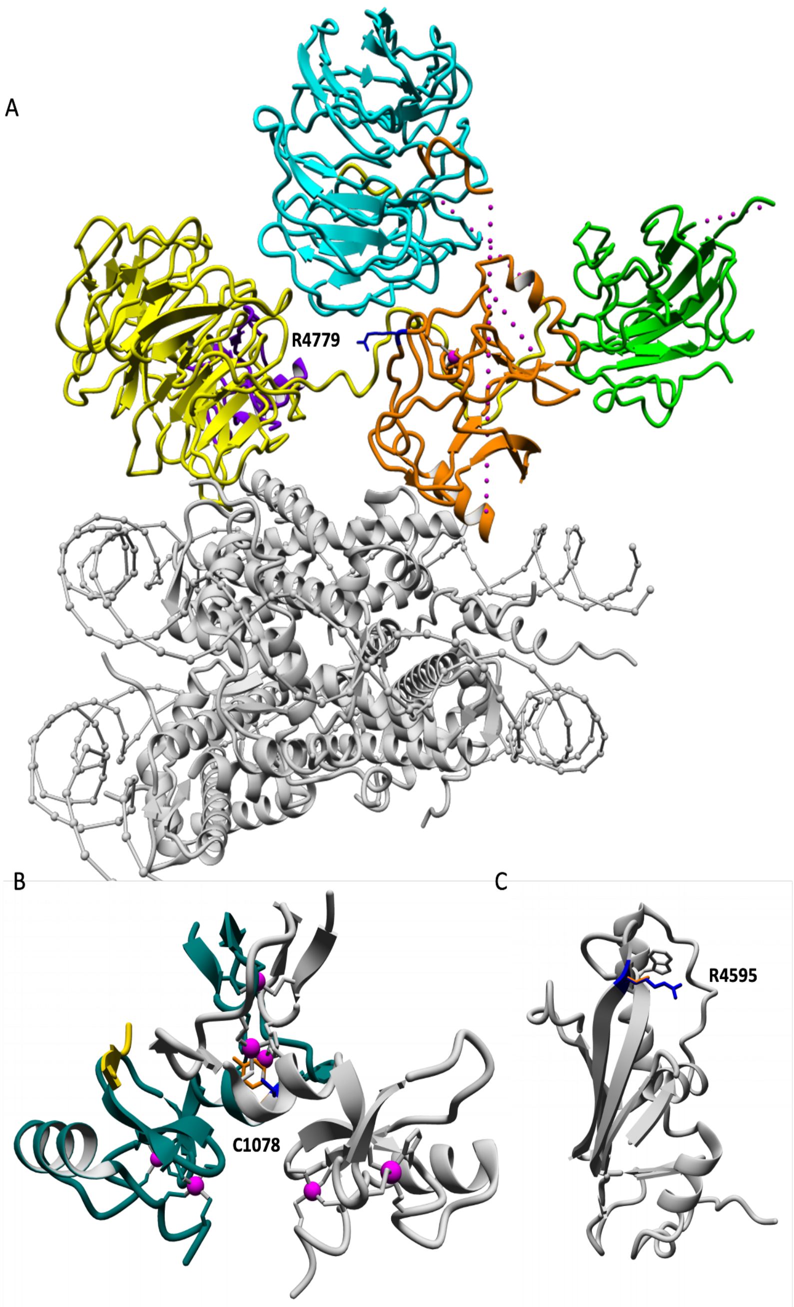


Figure S2. *KMT2C*-related NDD facial gestalt heatmap showing significant facial areas most different from controls.

Figure S3. Protein 3D structure effect predictions of the pathogenic PAVs.



A) *Arg4779 is located on a surface of the SET domain that interacts with WDR5 (PDB:5F6K). Arg change to Gly is a radical change and will likely affect the interaction that is required for the KMT2C activity. Previous study (25561738) showed that Arg4779Pro affected the methyltransferase activity of KMT2C.* ***B)*** *Cys1078 is binding to Zn in ePHD6 to ensure the correct folding and structure of the domain (PDB:6MLC); a change to Tyr would disrupt the Zn binding and, therefore, domain structure****. C)*** *Arg4595 is located on a surface of the FYR domain (AlphaFold2 model) and interacts with Trp4987, so change to smaller Cys would affect local aminoacid packing and interactions, and as result might destabilize the domain or affect binding to other proteins or binding between FYRN and FYRC domain parts.*