Functional correlation of genome-wide DNA methylation profiles in genetic neurodevelopmental disorders

Levy, M.A.; Relator, R.; McConkey, H.; Pranckeviciene, E.; Kerkhof, J.; Barat-Houari, M.; ... ; Sadikovic, B.

Received: 6 December 2021 | Revised: 30 June 2022 | Accepted: 27 July 2022

DOI: 10.1002/humu.24446

RESEARCH ARTICLE

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

Functional correlation of genome-wide DNA methylation profiles in genetic neurodevelopmental disorders

Michael A. Levy[1] | Raissa Relator[1] | Haley McConkey[1] | Erinija Pranckeviciene[1] | Jennifer Kerkhof[1] | Mouna Barat-Houari[2] | Sara Bargiacchi[3] | Elisa Biamino[4] | María Palomares Bralo[5] | Gerarda Cappuccio[6,7] | Andrea Ciolfi[8] | Angus Clarke[9] | Barbara R. DuPont[10] | Mariet W. Elting[11] | Laurence Faivre[12,13] | Timothy Fee[10] | Marco Ferilli[8] | Robin S. Fletcher[10] | Florian Cherick[14,15] | Aidin Foroutan[16] | Michael J. Friez[10] | Cristina Gervasini[17] | Sadegheh Haghshenas[16] | Benjamin A. Hilton[10] | Zandra Jenkins[18] | Simranpreet Kaur[19] | Suzanne Lewis[20] | Raymond J. Louie[10] | Silvia Maitz[21] | Donatella Milani[22] | Angela T. Morgan[23] | Renske Oegema[24] | Elsebet Østergaard[25,26] | Nathalie R. Pallares[2] | Maria Piccione[27] | Astrid S. Plomp[11] | Cathryn Poulton[28] | Jack Reilly[16] | Rocio Rius[29,30] | Stephen Robertson[18] | Kathleen Rooney[1,16] | Justine Rousseau[31] | Gijs W. E. Santen[32] | Fernando Santos-Simarro[5] | Josephine Schijns[33] | Gabriella M. Squeo[34] | Miya St John[23] | Christel Thauvin-Robinet[12,13,35,36] | Giovanna Traficante[3] | Pleuntje J. van der Sluijs[32] | Samantha A. Vergano[37,38] | Niels Vos[11] | Kellie K. Walden[10] | Dimitar Azmanov[39] | Tugce B. Balci[40,41] | Siddharth Banka[42,43] | Jozef Gecz[44,45] | Peter Henneman[11] | Jennifer A. Lee[10] | Marcel M. A. M. Mannens[11] | Tony Roscioli[46,47,48,49] | Victoria Siu[40,41] | David J. Amor[23] | Gareth Baynam[28,50,51] | Eric G. Bend[52] | Kym Boycott[53,54] | Nicola Brunetti-Pierri[6,7] | Philippe M. Campeau[31] | Dominique Campion[55] | John Christodoulou[19] | David Dyment[53,54] | Natacha Esber[56] | Jill A. Fahrner[57] | Mark D. Fleming[58] | David Genevieve[15] | Delphine Heron[59] | Thomas Husson[60] | Kristin D. Kernohan[53,61] | Alisdair McNeill[62] | Leonie A. Menke[33] | Giuseppe Merla[34,63] | Paolo Prontera[64] | Cheryl Rockman-Greenberg[65] | Charles Schwartz[10] | Steven A. Skinner[10] | Roger E. Stevenson[10] | Marie Vincent[66,67] | Antonio Vitobello[12,35] | Marco Tartaglia[8] | Marielle Alders[11] | Matthew L. Tedder[10] | Bekim Sadikovic[1,16]

[1]Verspeeten Clinical Genome Centre, London Health Sciences Centre, London, Ontario, Canada

[2]Autoinflammatory and Rare Diseases Unit, Medical Genetic Department for Rare Diseases and Personalized Medicine, CHU Montpellier, Montpellier, France

[3]Medical Genetics Unit, "A. Meyer" Children Hospital of Florence, Florence, Italy

[4]Department of Pediatrics, University of Turin, Turin, Italy

[5]Institute of Medical and Molecular Genetics (INGEMM), Hospital Universitario La Paz, IdiPAZ, CIBERER, ISCIII, Madrid, Spain

[6]Department of Translational Medicine, Federico II University of Naples, Naples, Italy

[7]Telethon Institute of Genetics and Medicine, Pozzuoli, Italy

[8]Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy

[9]Cardiff University School of Medicine, Cardiff, UK

[10]Greenwood Genetic Center, Greenwood, South Carolina, USA

[11]Amsterdam UMC, University of Amsterdam, Department of Human Genetics, Amsterdam Reproduction and Development Research Institute, Amsterdam, The Netherlands

[12]INSERM-Université de Bourgogne UMR1231 GAD, Génétique Des Anomalies du Développement, FHU-TRANSLAD, UFR Des Sciences de Santé, Dijon, France

[13]Centre de Référence Maladies Rares, Anomalies du Développement et Syndromes Malformatifs, Centre de Génétique, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France

[14]Genetic Medical Center, CHU Clermont Ferrand, France

[15]Montpellier University, Reference Center for Rare Disease, Medical Genetic Department for Rare Disease and Personalize Medicine, CHU Montpellier, Montpellier, France

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

[17]Division of Medical Genetics, Department of Health Sciences, Università degli Studi di Milano, Milan, Italy

[18]Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

[19]Brain and Mitochondrial Research Group, Murdoch Children's Research Institute and Department of Paediatrics, University of Melbourne, Melbourne, Australia

[20]BC Children's and Women's Hospital and Department of Medical Genetics, Faculty of Medicine, University of British Columbia,

[21]Clinical Pediatric Genetics Unit, Pediatrics Clinics, MBBM Foundation, Hospital San Gerardo, Monza, Italy

[22]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

[23]Murdoch Children's Research Institute and Department of Paediatrics, University of Melbourne, Melbourne, Australia

[24]Department of Genetics, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[25]Department of Clinical Genetics, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark

[26]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

[27]Medical Genetics Unit Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties, University of Palermo, Palermo, Italy

[28]Undiagnosed Diseases Program, Genetic Services of Western Australia, King Edward Memorial Hospital, Perth, Australia

[29]Brain and Mitochondrial Research Group, Murdoch Children's Research Institute, Melbourne, Australia

[30]Department of Paediatrics, University of Melbourne, Melbourne, Australia

[31]CHU Sainte-Justine Research Center, University of Montreal, Montreal, Quebec, Canada

[32]Department of Clinical Genetics, LUMC, Leiden, The Netherlands

[33]Department of Pediatrics, Emma Children's Hospital, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

[34]Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Naples, Italy

[35]Unité Fonctionnelle d'Innovation Diagnostique des Maladies Rares, FHU-TRANSLAD, France Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement (TRANSLAD), Centre Hospitalier Universitaire Dijon Bourgogne, CHU Dijon Bourgogne, Dijon, France

[36]Centre de Référence Déficiences Intellectuelles de Causes Rares, Hôpital D'Enfants, CHU Dijon Bourgogne, Dijon, France

[37]Division of Medical Genetics and Metabolism, Children's Hospital of The King's Daughters, Norfolk, Virginia, USA

[38]Department of Pediatrics, Eastern Virginia Medical School, Norfolk, Virginia, USA

[39]Department of Diagnostic Genomics, PathWest Laboratory Medicine, QEII Medical Centre, Perth, Australia

[40]Department of Pediatrics, Division of Medical Genetics, Western University, London, Ontario, Canada

[41]Medical Genetics Program of Southwestern Ontario, London Health Sciences Centre and Children's Health Research Institute, London, Ontario, Canada

[42]Division of Evolution, Infection & Genomics, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

[43]Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester, United Kingdom

[44]School of Medicine, Robinson Research Institute, University of Adelaide, Adelaide, Australia

[45]South Australian Health and Medical Research Institute, Adelaide, Australia

[46]Neuroscience Research Australia (NeuRA), Sydney, Australia

[47]Prince of Wales Clinical School, Faculty of Medicine, University of New South Wales, Sydney, Australia

[48]New South Wales Health Pathology Randwick Genomics, Prince of Wales Hospital, Sydney, Australia

[49]Centre for Clinical Genetics, Sydney Children's Hospital, Sydney, Australia

[50]Undiagnosed Diseases Program, Genetic Services of Western Australia, King Edward Memorial Hospital, Perth, Australia

[51]Division of Paediatrics and Telethon Kids Institute, Faculty of Health and Medical Sciences, Perth, Australia

[52]PreventionGenetics, Marshfield, Wisconsin, USA

[53]Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada

[54]Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada

[55]INSERM U1245, Faculté de Médecine, Rouen, France

[56]KAT6A Foundation, New York, New York, USA

[57]Departments of Genetic Medicine and Pediatrics, Johns Hopkins University, Baltimore, Maryland, USA

[58]Department of Pathology, Boston Children's Hospital, Boston, Massachusetts, USA

[59]AP-HP, Département de Génétique Médicale, Groupe Hospitalier Pitié Salpétrière, Paris, France

[60]Department of Genetics and Reference Center for Developmental Disorders, Normandie Université, UNIROUEN, Inserm U1245 and Rouen University Hospital, Rouen, France

[61]Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, Canada

[62]Department of Neuroscience, University of Sheffield, UK, and Sheffield Children's Hospital NHS Foundation Trust, Sheffield, UK

[63]Laboratory of Regulatory and Functional Genomics, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy

[64]Medical Genetics Unit, University of Perugia Hospital SM della Misericordia, Perugia, Italy

[65]Department of Pediatrics and Child Health, Rady Faculty of Health Sciences, University of Manitoba and Program in Genetics and Metabolism, Shared Health MB, Winnipeg, Manitoba, Canada

[66]Service de génétique Médicale, CHU Nantes, Nantes, France

[67]Institut du thorax, INSERM, CNRS, UNIV Nantes, Nantes, France

Correspondence
Bekim Sadikovic, Verspeeten Clinical Genome Centre, London Health Sciences Centre, London, ON N6A 5W9, Canada.
Email: bekim.sadikovic@lhsc.on.ca

Funding information
Government of Canada, Grant/Award Number: Genome Canada and the Ontario Genomics Institute; Victorian Government, Grant/Award Number: Operational Infrastructure Support Program; South Carolina Department of Disabilities and Special Needs; Royal Children's Hospital Foundation, Grant/Award Number: Chair in Genomic Medicine awarded to JC; London Health Sciences Centre, Grant/Award Number: Molecular Diagnostics Development Fund; Italian Ministry of Research, Grant/Award Number: FOE 2019; Italian Ministry of Health, Grant/Award Numbers: 5x1000, CCR-2017-23669081, RCR-2020-23670068_001

Abstract

An expanding range of genetic syndromes are characterized by genome-wide disruptions in DNA methylation profiles referred to as episignatures. Episignatures are distinct, highly sensitive, and specific biomarkers that have recently been applied in clinical diagnosis of genetic syndromes. Episignatures are contained within the broader disorder-specific genome-wide DNA methylation changes, which can share significant overlap among different conditions. In this study, we performed functional genomic assessment and comparison of disorder-specific and overlapping genome-wide DNA methylation changes related to 65 genetic syndromes with previously described episignatures. We demonstrate evidence of disorder-specific and recurring genome-wide differentially methylated probes (DMPs) and regions (DMRs). The overall distribution of DMPs and DMRs across the majority of the neurodevelopmental genetic syndromes analyzed showed substantial enrichment in gene promoters and CpG islands, and under-representation of the more variable intergenic regions. Analysis showed significant enrichment of the DMPs and DMRs in gene pathways and processes related to neurodevelopment, including neurogenesis, synaptic signaling and synaptic transmission. This study expands beyond the diagnostic utility of DNA methylation episignatures by demonstrating correlation between the function of the mutated genes and the consequent genomic DNA methylation profiles as a key functional element in the molecular etiology of genetic neurodevelopmental disorders.

KEYWORDS
clinical diagnostics, DNA methylation, episignatures, neurodevelopmental syndromes

1 | INTRODUCTION

DNA methylation is a fundamental aspect of mammalian development, and changes in DNA methylation are closely related to variation in the underlying genome (Greenberg & Bourc'his, 2019; Villicana & Bell, 2021). A growing number of genes causing neurodevelopmental syndromes have been shown to be associated with distinct changes in DNA methylation in patients affected with Mendelian disorders (Velasco & Francastel, 2019). These methylation changes may be a direct consequence of the disruption of the gene function as in the chromatin remodeling, DNA methylation, and histone modification genes (Janssen & Lorincz, 2021; Sadikovic et al., 2019). More recent work has shown that changes in DNA methylation are also found in patients with pathogenic variants in genes that have no known direct role in DNA methylation or chromatin remodeling including *FAM50A*, *UBE2A*, and *ZNF711* (Haghshenas et al., 2021; Levy et al., 2022). These indirect changes may be caused by perturbations in the interconnected molecular pathways, including transcriptional regulation and protein signaling, as opposed to directly by DNA methylases or demethylases. Genetic variants that are inherited or that occur at the earliest stages of the embryo can, therefore, have a wide-reaching impact on DNA methylation throughout development. These early changes can be propagated through cell differentiation and tissue development. Hence, an easily accessible tissue such as peripheral blood can be used to demonstrate changes in DNA methylation and develop biomarkers of specific syndromes that occur at early stages of development (Aref-Eshghi et al., 2020; Levy et al., 2022). More than 60 genetic neurodevelopmental syndromes have now been identified that exhibit such alterations in DNA methylation, and the patterns of DNA methylation changes, referred to as episignatures, are now being used in diagnostic clinical testing (Sadikovic et al., 2021).

We have previously described the development of 56 distinct diagnostic episignatures encompassing 65 neurodevelopmental syndromes caused by pathogenic variants in 61 genes, and their use as highly sensitive and specific diagnostic biomarkers (Aref-Eshghi, Bend, et al., 2018; Aref-Eshghi, E., Rodenhiser, 2018; Aref-Eshghi et al., 2017, 2020; Bend et al., 2019; Ciolfi et al., 2020, 2021; Haghshenas et al., 2021; Hood et al., 2016; Kerkhof et al., 2021; Krzyzewska et al., 2019; Levy et al., 2021, 2022; Radio et al., 2021; Rooney et al., 2021; Sadikovic et al., 2021; Schenkel et al., 2017, 2018, 2021). In these studies, we demonstrated that a monogenic syndrome may have more than one episignature depending on the location and/or functional consequence of the underlying genetic variant within the gene. Conversely, similar syndromes, such as those caused by pathogenic variants in genes from the same gene family or same molecular pathway, may share a common episignature. As molecular biomarkers, episignatures are optimized for clinical diagnostics and commonly represent only a fraction of the totality of the DNA methylation change in any given disorders. In this study, we expanded on this work by investigating the broader context of changes in DNA methylation by performing functional genomic assessment and comparison of disorder-specific and overlapping genome-wide DNA methylation changes in these syndromes. We describe disorder-specific and recurring genome-wide differentially methylated probes (DMPs) and differentially methylated regions (DMRs) and correlate them to functional genomic elements including gene promoters and CpG islands. We explore the functional impact of these changes in relation to the corresponding gene pathways and transcriptional networks. By using various correlation analyses, we assess relatedness of the genetic etiology and the consequent DNA methylation profiles in the molecular pathogenesis of genetic neurodevelopmental disorders.

2 | MATERIALS AND METHODS

2.1 | Patient cohorts

The case cohorts consisted of 1381 peripheral blood DNA samples from patients who were diagnosed with one of the 65 neurodevelopmental conditions and who had a positive EpiSign result for one of the corresponding 56 episignatures, that are part of the EpiSign Knowledge Database (EKD; https://episign.lhsc.on.ca/index.html). Mean, median, minimum, and maximum case–cohort sizes were 25, 14, 3, and 191, respectively (Table 1, Supporting Information: Table S1). The control cohort consisted of 4231 samples: 2701 unaffected controls (individuals with no specific neurodevelopmental phenotype and no known pathogenic or suspected pathogenic variant in any of the episignature-related genes) and 1530 unresolved samples (individuals with suspected genetic disorders but with no distinct genetic or EpiSign diagnosis) (Levy et al., 2022). Case samples came from a total of 441 batches: four smaller cohorts were each from a single batch, while all others had at least two batches, the median number of batches per cohort was five. Control samples came from a total of 1957 batches: the minimum number of control batches used per cohort analysis was 13, the median was 36.

2.2 | Sample processing

Peripheral blood DNA was extracted using standard techniques. Bisulfite conversion was performed with 500 ng of genomic DNA using the Zymo EZ-96 DNA Methylation Kit (D5004), and bisulfite-converted DNA was used as input to the Illumina Infinium HumanMethylation450 (450K array) or MethylationEPIC BeadChip array (EPIC array). Array data were generated according to the manufacturer's protocol. Details of sample processing are described in the previous studies (Aref-Eshghi et al., 2020; Aref-Eshghi, Rodenhiser, et al., 2018; Levy et al., 2022).

2.3 | Methylation probe processing and selection

The data analysis pipeline was adapted from previously described methods (Aref-Eshghi et al., 2020; Levy et al., 2022). IDAT files

TABLE 1 List of cohorts

Syndrome	Signature abbreviation	Underlying gene or region	OMIM (2022)	Samples	Probes	Category
X-linked alpha-thalassemia/mental retardation syndrome (ATRX)	ATRX	*ATRX*	301040	30	8666	SWI/SNF chromatin remodeling
Arboleda–Tham syndrome (ARTHS)	ARTHS	*KAT6A*	616268	18	4487	Histone acetyltransferase
Autism, susceptibility to, 18 (AUTS18)	AUTS18	*CHD8*	615032	28	2319	Transcription factor
Beck–Fahrner syndrome (BEFAHRS)	BEFAHRS	*TET3*	618798	16	30,391	DNA demethylase
Blepharophimosis Intellectual disability SMARCA2 Syndrome	BISS	*SMARCA2*	619293	12	10,186	SWI/SNF chromatin remodeling
Börjeson–Forssman–Lehmann syndrome (BFLS)	BFLS	*PHF6*	301900	14	12,321	Transcription factor
Cerebellar ataxia, deafness, and narcolepsy, autosomal dominant (ADCADN)	ADCADN	*DNMT1*	604121	5	151,848	DNA methyltransferase
CHARGE syndrome	CHARGE	*CHD7*	214800	74	840	Transcription factor
Chr16p11.2 deletion syndrome, 593-KB	Chr16p11.2del	Chr16p11.2del	611913	18	10,105	CNV
Coffin–Siris syndrome-1,2 (CSS1,2)	CSS_c.6200[a]	*ARID1B*	135900	4	3451	SWI/SNF chromatin remodeling
		ARID1A	614607			
Coffin–Siris syndrome-1,2,3,4; Nicolaides–Baraitser syndrome (CSS12,3,4; NCBRS)	BAFopathy	*ARID1B*	135900	124	1015	SWI/SNF chromatin remodeling
		ARID1A	614607			
		SMARCB1	614608			
		SMARCA4	614609			
		SMARCA2	601358			
Coffin–Siris syndrome-4 (CSS4)	CSS4_c.2650[a]	*SMARCA4*	614609	3	464	SWI/SNF chromatin remodeling
Coffin–Siris syndrome-9 (CSS9)	CSS9	*SOX11*	615866	13	430	Transcription factor
Cohen–Gibson syndrome; Weaver syndrome (COGIS; WVS)	PRC2	*EED*	617561	8	2444	Histone deacetylase Histone methyltransferase
		EZH2	277590			
Cornelia de Lange syndromes 1,2,3,4 (CDLS1,2,3,4)	CdLS	*NIPBL*	122470	70	3623	Chromosome cohesion/ condensation; DNA repair (RAD21)
		SMC1A	300590			
		SMC3	610759			
		RAD21	614701			
Down syndrome	Down	Chr21 trisomy	190685	40	24,712	CNV
Dystonia 28, childhood-onset (DYT28)	DYT28	*KMT2B*	617284	10	25,260	Histone methyltransferase
Epileptic encephalopathy, childhood-onset (EEOC)	EEOC	*CHD2*	615369	9	5284	Transcription factor
Floating–Harbor syndrome (FLHS)	FLHS	*SRCAP*	136140	21	26,811	SWI/SNF chromatin remodeling
Gabriele–de Vries syndrome (GADEVS)	GADEVS	*YY1*	617557	10	4380	Transcription factor
Genitopatellar syndrome (see also Ohdo syndrome, SBBYSS variant) (KAT6B)	GTPTS	*KAT6B*	606170	4	3008	Histone acetyltransferase
Helsmoortel–van der Aa syndrome (HVDAS)	HVDAS_C[a]	*ADNP*	615873	14	6986	Transcription factor

(Continues)

TABLE 1 (Continued)

Syndrome	Signature abbreviation	Underlying gene or region	OMIM (2022)	Samples	Probes	Category
Helsmoortel–van der Aa syndrome (HVDAS)	HVDAS_T[a]	*ADNP*	615873	21	16,756	Transcription factor
Hunter McAlpine craniosynostosis syndrome	HMA	Chr5q35-qter dup	601379	8	17,948	CNV
Immunodeficiency-centromeric instability-facial anomalies syndrome 1 (ICF1)	ICF_1	*DNMT3B*	242860	8	38,656	DNA methyltransferase
Immunodeficiency-centromeric instability-facial anomalies syndromes 2,3,4 (ICF2,3,4)	ICF_2_3_4	*ZBTB2* *CDCA7* *HELLS*	614069 616910 616911	7	66,568	Transcription factorc -Myc responsive gene SWI/SNF chromatin remodeling
Intellectual developmental disorder with seizures and language delay (IDDSELD)	IDDSELD	*SETD1B*	619000	11	5264	Histone methyltransferase
Kabuki syndromes 1, 2 (KABUK1,2)	Kabuki	*KMT2D* *KDM6A*	147920 300867	191	3749	Histone methyltransferase Histone demethylase
KDM2B-related syndrome	KDM2B	*KDM2B*	Unofficial	9	3632	Histone demethylase
Autosomal dominant intellectual developmental disorder-65 (MRD65)	KDM4B	*KDM4B*	619320	6	279	Histone demethylase
Kleefstra syndrome 1 (KLEFS1)	Kleefstra	*EHMT1*	610253	32	4124	Histone methyltransferase
Koolen de Vreis syndrome (KDVS)	KDVS	*KANSL1*	610443	16	6490	Histone acetylation
Luscan–Lumish syndrome (LLS)	LLS	*SETD2*	616831	4	2405	Histone methyltransferase
Menke–Hennekam syndromes 1,2 (MKHK1,2)	MKHK_ID4[a]	*CREBBP* *EP300*	618332 618333	13	2570	Histone acetyltransferase
Intellectual developmental disorder, X-linked, syndromic, Armfield type (MRXSA)	MRXSA	*FAM50A*	300261	6	4618	mRNA splicing
Mental retardation, autosomal dominant 23 (MRD23)	MRD23	*SETD5*	615761	25	2795	Histone methyltransferase
Mental retardation, autosomal dominant 51 (MRD51)	MRD51	*KMT5B*	617788	7	19,803	Histone methyltransferase
Intellectual developmental disorder, X-linked 93 (MRX93)	MRX93	*BRWD3*	300659	11	16,894	Transcription factor
Intellectual developmental disorder, X-linked 97 (MRX97)	MRX97	*ZNF711*	300803	18	3770	Transcription factor
Intellectual developmental disorder, X-linked syndromic, Nascimento-type (MRXSN)	MRXSN	*UBE2A*	300860	4	6065	Enzyme
Intellectual developmental disorder, X-linked, Snyder–Robinson type (MRXSSR)	MRXSSR	*SMS*	309583	17	4062	Enzyme
Intellectual developmental disorder, X-linked, syndromic, Claes–Jensen type (MRXSCJ)	MRXSCJ	*KDM5C*	300534	58	5013	Histone demethylase
Myopathy, lactic acidosis, and sideroblastic anemia 2 (MLASA2)	MLASA2	*YARS2*	613561	11	2304	tRNA synthesis
Ohdo syndrome, SBBYSS variant (SBBYSS)	SBBYSS	*KAT6B*	603736	9	1956	Histone acetyltransferase

TABLE 1 (Continued)

Syndrome	Signature abbreviation	Underlying gene or region	OMIM (2022)	Samples	Probes	Category
Phelan–McDermid syndrome (PHMDS)	PHMDS	Chr22q13.3del	606232	11	17,581	CNV
Rahman syndrome (RMNS)	RMNS	*HIST1H1E*	617537	9	26,101	Linker histone
Renpenning syndrome (RENS1)	RENS1	*PQBP1*	309500	8	5228	mRNA splicing
Rubinstein–Taybi syndrome 1 (RSTS1)	RSTS1	*CREBBP*	180849	37	5279	Histone acetyltransferase
Rubinstein–Taybi syndrome 2 (RSTS2)	RSTS2	*EP300*	613684	29	7998	Histone acetyltransferase
Sotos syndrome 1 (SOTOS1)	Sotos	*NSD1*	117550	69	43,022	Histone methyltransferase
Tatton–Brown–Rahman syndrome (TBRS)	TBRS	*DNMT3A*	615879	30	35,130	DNA methyltransferase
Velocardiofacial syndrome (VCFS)	VCFS	Chr22q11.2del	192430	47	4134	CNV
Wiedemann–Steiner syndrome (WDSTS)	WDSTS	*KMT2A*	605130	52	4777	Histone methyltransferase
Williams–Beuren deletion syndrome (WBS)	Williams	Chr7q11.23del	194050	22	13,131	CNV
Williams–Beuren duplication syndrome (Chr7q11.23 duplication syndrome)	Dup7	Chr7q11.23dup	609757	13	6963	CNV
Wolf–Hirschhorn syndrome (WHS)	WHS	Chr4p16.13del	194190	17	7838	CNV

Abbreviations: CNV, copy number variation; SWI/SNF, switch/sucrose non-fermentable.
[a]Episignatures that encompass a specific region or variant within a gene.

containing methylated and unmethylated signal intensity were imported into R 4.1.0 for analysis. Normalization was performed using the Illumina normalization method with background correction using the minfi package, with the same EPIC control sample used as the reference array for all samples. Probes with a detection $p > 0.01$, probes located on the X and Y chromosomes, probes that contained single nucleotide polymorphisms at the CpG interrogation or single nucleotide extension sites, and probes which are known to cross-react with other genomic locations were removed (Chen et al., 2013; Pidsley et al., 2016). Probes with beta values of 0 and the top 1% most variable (variance) probes within the case or control samples were removed. For each cohort, a set of controls was chosen matched for age, sex, and array type, using the R package matchit version 4.2.0. For each case sample, 1–10 controls were used (case: control ratio of 1:1–1:10). Mean, median, minimum, and maximum control cohort sizes were 60, 56, 30, and 191 (Supporting Information: Table S2).

Methylation levels (beta values) were used for linear regression modeling using the limma package version 3.48.0 (Ritchie et al., 2015). Estimated blood cell proportions (Houseman et al., 2012) were added to the model matrix as confounding variables. The generated p values were moderated using the eBayes function. To facilitate comparisons between samples processed using 450K and EPIC arrays, only probes found on both arrays were used for analysis. Probes that had a mean methylation difference of less than 5% between the case and control samples were removed and Benjamini–Hochberg adjusted p values were calculated for the remaining probes. Probes with an adjusted p value less than 0.01 were selected as DMPs for analysis, except for cohorts KDM4B and CSS4_c.2650, which had too few probes and probes with a nonadjusted $p < 0.001$ were used.

2.4 | Identification of differentially methylated regions

For each cohort, genome-wide DMR analysis was performed to identify genomic regions with differential DNA methylation between cases and matched controls. Methylation beta values equal to 1 were initially shifted by a very small value (1e−10) to avoid infinite M-values during conversion implemented using minfi. DMR analysis on the matrix of M values were identified using the R packages DMRcate version 2.6.0 (Peters et al., 2015), and regions were defined to have at least five CpG probes within 1000 bp of each other. Minimum absolute mean methylation difference between cases and controls was set to 0.1 and significant results were chosen using a Fisher p value cut-off of 0.01.

2.5 | Cohort comparisons and data visualizations

Circos-style plots were made using the R package circlize version 0.4.14 (Gu et al., 2014). DMRs and DMPs were annotated in relation to CpG islands (CGIs) and genes using the R package annotatr version 1.18.1 (Cavalcante & Sartor, 2017) with AnnotationHub version 3.0.0 and annotations hg19_cpgs, hg19_basicgenes, hg19_genes_intergenic, and hg19_genes_intronexonboundaries. CGI annotations included CGI shores from 0 to 2 kb on either side of CGIs, CGI

shelves from 2 to 4 kb on either side of CGIs, and inter-CGI regions encompassing all remaining regions. For gene annotations, promoters included up to 1 kb upstream of the transcription start site (TSS) and promoter+ the region 1–5 kb upstream of the TSS. Annotations to untranslated regions (5′-UTR and 3′-UTR), exons, introns, and exon/intron boundaries were combined into the "gene body" category. Enrichment of annotation terms were tested using hypergeometric test. Heatmaps were made using the R package pheatmap version 1.0.12.

Distance and similarities between cohorts were analyzed using agglomerative clustering. A tree diagram was made by first aggregating methylation levels of each DMP using its median value across samples with the same condition, resulting in a beta value matrix with rows and columns equal to the total number of all DMPs, and the number of cohorts, respectively. Euclidean distance between cohorts were computed using the generated beta matrix and clustering using Ward's method on the distances was then implemented. Initial analysis using all identified DMPs from all cohorts showed that the number of DMPs affected the clustering results. Therefore, for the final analysis, the top 500 DMPs ranked by *p* values for each cohort were initially selected before generating the beta value matrix. For cohorts with fewer than 500 DMPs, all of their DMPS were used. This resulted in a combined set of 20,904 probes across all groups. Clustering results on the distances computed using the new beta matrix were visualized as a tree-and-leaf diagram using the R package TreeAndLeaf version 1.4.2 to incorporate additional information such as global mean methylation difference and total number of DMPs identified for each cohort.

Two-dimensional and three-dimensional representations of the topological structure of the entire cohort database were analyzed using unified mapping approximation and projection (UMAP) using the R umap package version 0.2.7.0. The global structure approximated by UMAP was obtained by using 210 probes that were most differentiating across all cohorts selected by random forest feature importance as described below. The UMAP parameter for the number of nearest neighbors was set to 10 and minimum distance in final layout set to 0.99, and results were visualized in 3D using the R package plot3D version 1.4.

2.6 | Selection of most differentiating probes across all cohorts

Probes evaluated to be most discriminating of the 56 cohorts were identified by random forest using the R package randomForest version 4.6.14. Feature importance was computed for the selected 20,904 probes as previously described. Random forest multiclassification models were trained using the 20,904 probes and 1381 samples, and variable importance measured as mean accuracy decrease was computed for each probe. Due to the randomness of the model, we repeated the procedure 1000 times and summed all variable importance values across all sets. Finally, probes that ranked in the top one percentile were selected (210 probes). For each trial, 100 trees were fitted using 145 (default: sqrt(# of features)) randomly sampled probes at each split. Down-sampling was incorporated to account for the sample imbalance across cohorts, and the number of samples drawn per group at each split was set to the minimum number of samples among all cohorts to ensure an identical value.

2.7 | Functional annotation of genes overlapping selected DMPs and DMRs

Gene Ontology (GO) and KEGG pathways associated with DMPs and DMRs were identified using the R package missMethyl version 1.23.1 (Phipson et al., 2016). Enrichment analysis was performed using either all DMPs, or using DMPs that were found in more than five cohorts. For DMRs, term enrichment analysis was implemented using either all probes within all DMRs, or using probes that were found in at least two DMRs. The background list was generated by combining all probes remaining postfiltering for all cohorts which were used for the differential methylation analysis.

2.8 | Network diagrams

A network diagram was made by first determining the number of shared probes between each pair of cohorts. Each probe in a pair of cohorts was categorized depending on the direction of the probe's change in methylation: hyper–hyper (probe had increased methylation in both cohorts), hypo–hypo (probe had decreased methylation in both cohorts), hypo–hyper, or hyper–hypo (probe's methylation was increased in one cohort and decreased in the other). The obtained data matrix was visualized using Cytoscape version 3.9 (Su et al., 2014) in which nodes represent cohorts, the edges connecting the nodes represent hyper or hypomethylated probes, and the weight of the edge is proportional to the absolute count of the probes shared by the two nodes.

3 | RESULTS

3.1 | Detection of DMPs and regions

We generated lists of DMPs for each cohort (for the full list of cohort names and abbreviations, see Table 1). CSS4_c.2650, which only has three samples, had zero significant DMPs using an adjusted *p* value of less than 0.01, and KDM4B, which has six samples but more mild methylation changes had 77 DMPs. For these two cohorts we, therefore, used a nonadjusted *p* value resulting in 464 and 279 DMPs, respectively, for all subsequent analysis. The 56 cohorts, therefore, ranged from a minimum of 279 DMPs for KDM4B to a maximum of 151,848 DMPs for ADCADN, with a mean of 13,427 and a median of 5272 (Table 1, Supporting Information: Figure S1A and Table S3).

We next searched for DMPs recurring in more than one cohort. The 56 cohorts included a total of 253,431 unique DMPs. 113,911

(55%) were unique to a specific cohort, while 139,520 (45%) recurred in two or more cohorts. Most of the unique DMPs were found in the cohorts with the largest number of total DMPs: ADCADN and ICF2_3_4 accounted for 85% of the unique probes (Supporting Information: Figure S1B). All other cohorts shared at least 85% of their DMPs with at least one other cohort, with all 1015 BAFopathy DMPs found in at least one other cohort (Supporting Information: Figure S1B). The cohorts with the largest number of DMPs generally also had the largest number of shared DMPs (Figure 1, Supporting Information: Figures S1A,B). Among the 139,520 DMPs found in two or more cohorts, 46,635 (33%) were found in exactly two cohorts, while one DMP was found in each of 27 and 28 cohorts (Supporting Information: Figure S1C).

We then used all probes to identify DMRs. Forty-eight cohorts returned significant results ranging from one to 1384 DMRs, with the median of the DMR counts being eight and mean being 89. Based on the parameters used for DMR detection, seven cohorts did not have any significant DMRs: AUTS18, BAFopathy, CSS4_c.2650, CSS9, Kabuki, KDM4B, MRX93. Two of the cohorts with no DMRs, CSS4_c.2650 and KDM4B, had the fewest number of DMPs when using an adjusted *p* value cut-off, explaining the lack of identified DMRs. Most cohorts with no significant DMRs had either mild changes in methylation (Kabuki, MRX93) or a relatively small number of identified DMPs (AUTS18, CSS4_c.2650), or both (BAFopathy, CSS9, KDM4B). Therefore, as expected, cohorts that were highly hypo/hypermethylated and with a high number of DMPs also had the highest number of identified DMRs, such as ADCADN (1384 DMRs), ICF2_3_4 (851 DMRs), Sotos (809 DMRs) and ICF1 (514 DMRs) (Supporting Information: Figure S2).

3.2 | Genomic context of DMPs

We next examined the genomic locations of the probes. First, we assessed locations in relation to CGIs. CpG annotations were available for 3,137,161,264 nucleotides divided into CGI (0.7%), CGI shores (3.2%), CGI shelves (2.8%), and inter-CGI regions (93.3%). However, since CGI are enriched for DNA methylation CpGs, they are over-represented on the DNA methylation microarrays. After initial filtering to remove chromosomes X and Y and certain other probes as described in Section 2.3, 30.7% of microarray probes overlapped with CGI and 23.8% overlapped CGI shores. This represents the "background" or "default" distribution of probes on the microarray used for analysis (Figure 2a). We compared the distributions of DMPs to the background distribution of probes. Twelve of the 56 cohorts (21.4%) had probes overrepresented at CGI and 43 (76.8%) cohorts were overrepresented at shores. When considered together, 47 (84%) of cohorts had DMPs overrepresented at, or within 2 kb of, a CGI. Eight cohorts (14.3%) had probes overrepresented within CGI shelves and 18 cohorts (32.1%) at inter-CGI regions (Figure 2a).

Similar analysis was then performed for the 49 cohorts which had at least one DMR. Since the microarrays contain probes and not DMRs a default distribution for DMRs cannot be generated. Sixty-four percent of the 5221 total DMRs overlapped CGIs, and 38 of the 49 DMR cohorts (79%) had 50% or more of their DMRs overlapping CGIs. There was variability in results between cohorts with several having all DMRs overlapping CGIs and the lowest (cohort WHS) having only 1/13 (7.7%) of its DMRs overlapping CGIs (Figure 2b).

DMPs and DMRs were then annotated in relation to genes. The background microarray distribution of probes included 29.4% at promoters, 4.9% promoter+, 49.7% in gene bodies, and 16.0% intergenic (Figure 2c). Seven (12.5%) of the cohorts had DMPs overrepresented at the 1 kb promoters and all 56 had DMPs overrepresented when the larger 1–5 kb promoter+ region was included. No cohorts were overrepresented within gene bodies and 49 (88%) were overrepresented at intergenic regions (Figure 2c). 3033 (58.1%) of the 5221 total DMRs overlapped the promoter regions, and 29 of the 49 DMR cohorts (59.2%) had 50% or more of their DMRs overlapping the extended promoter regions (Figure 2d).

To identify categories of genes and molecular pathways with changes in DNA methylation we performed enrichment analysis using all DMPs, and then using the subset of probes that were within DMRs. To identify terms that were more representative of the full set of cohorts we repeated this analysis using "repeatedly selected" DMPs and probes within DMRs: for DMP analysis, we used probes identified in more than five cohorts, for DMRs analysis, we used DMR probes that were identified in at least two cohorts. When using all DMRs, 158 GO terms were found enriched in the identified regions ($p.adjust < 0.05$). However, only 12 terms were significant when using genes overlapping recurring probes in the DMRs, all of which were also in the 158 GO terms in the initial analysis. The top 10 most significant terms for the DMR results in both tests are shown in Table 2. Gene set ratio values indicate the ratio of the number of genes overlapping the DMRs to the number of genes overlapping all probes in the background list in the corresponding annotation. At the probe level, when using all DMPs 9 GO terms were identified, when using duplicated DMPs 20 GO terms were identified, one of which is common to both analyses. The top most significant terms are shown in Table 3. Results demonstrate enrichment in functional categories related to biological processes involved in nervous system development, neurogenesis, synaptic transmission, and synaptic signaling. Subsequent KEGG pathway enrichment using all DMR results revealed enrichment of genes related to the neuroactive ligand-receptor interaction pathway ($p.adjust = 3.15E{-}08$), nicotine addiction pathway ($p.adjust = 2.17E{-}04$), and steroid hormone biosynthesis pathway (p.adjust = 0.04), while using duplicated DMPs indicated enrichment in calcium signaling pathway ($p.adjust = 0.04$) [data not shown].

3.3 | Relationships between cohorts

All DMPs were used to calculate mean and median beta values for each cohort to identify overall trends in hypo- and hypermethylation. Thirty-seven (66.1%) cohorts had mean hypomethylation

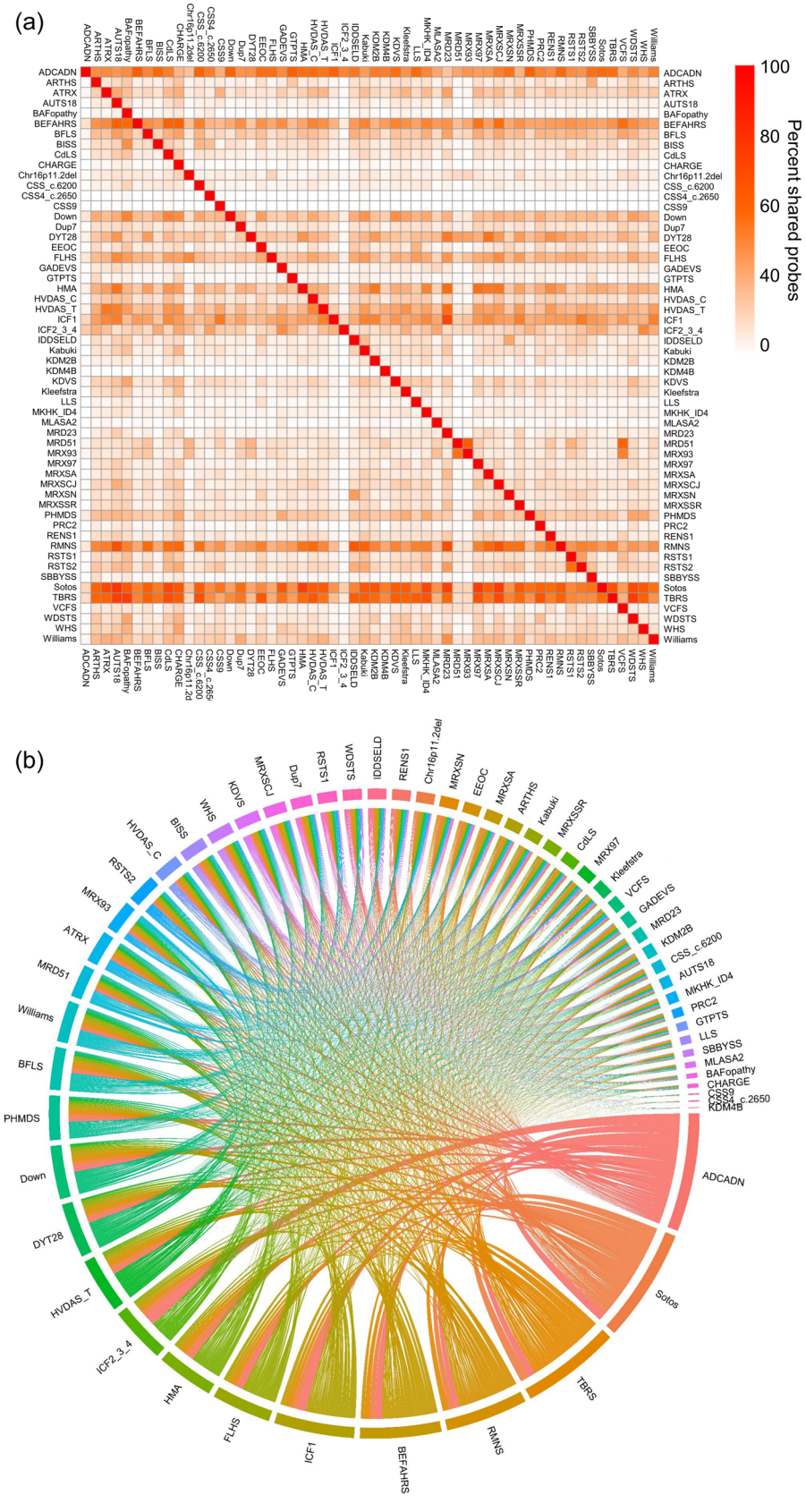

FIGURE 1 Differentially methylated probes found shared between multiple cohorts. (a) Percent of probes that are shared between each pair of cohorts. For each pair, the colors indicate the percent of the top/bottom cohort's probes that are also found in the left/right cohort's probes. (b) Probes that are shared between each pair of cohorts. Each labeled and colored sector represents one cohort. The thickness of connecting lines represents the number of probes shared between the two cohorts.

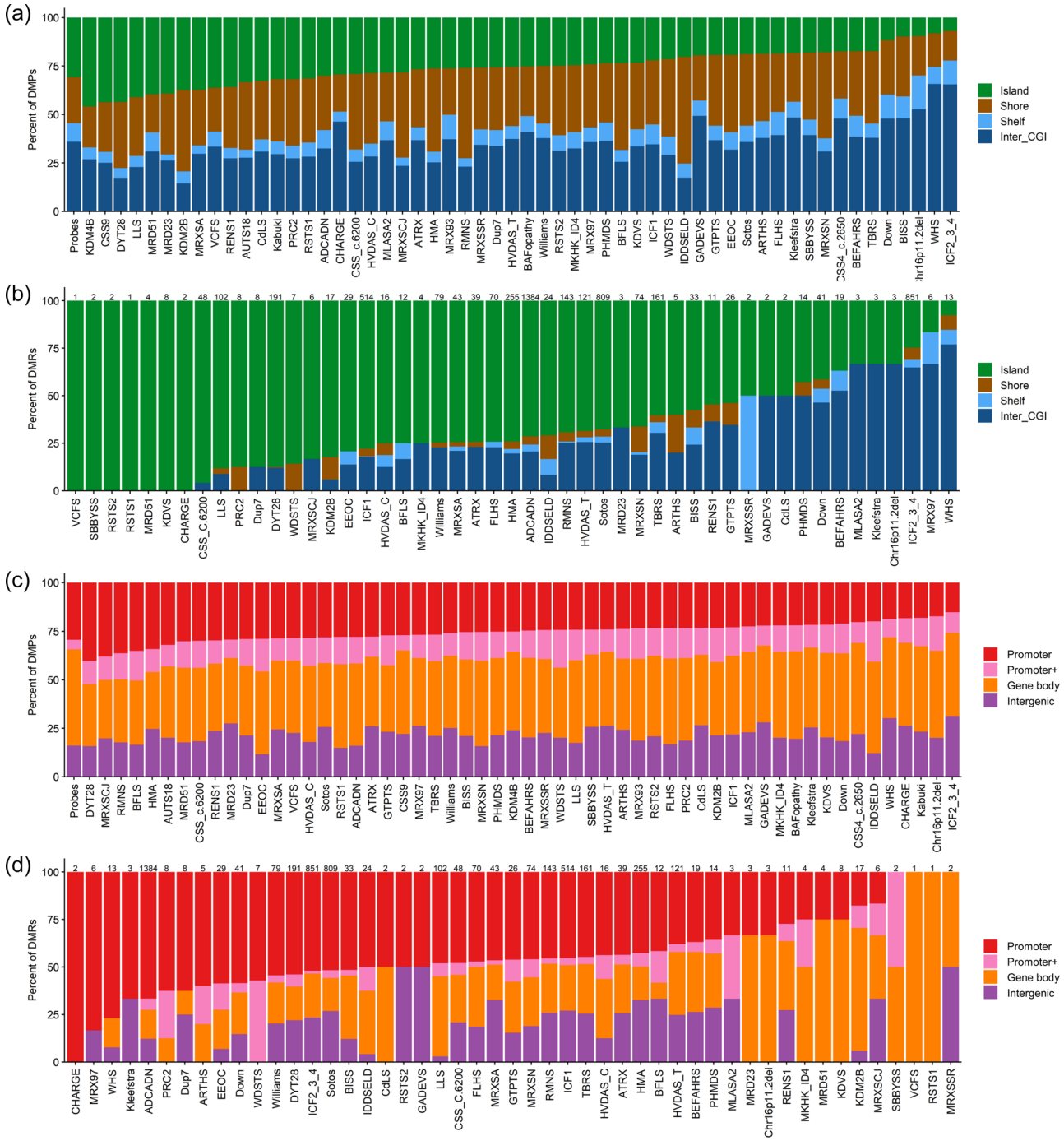

FIGURE 2 DMPs and DMRs annotated in the context of CpG islands and genes. (a) DMPs annotated in the context of CpG islands. (b) DMRs annotated in the context of CpG islands. (c) DMPs annotated in the context of genes. (d) DMRs annotated in the context of genes. For CpG plots: Island, CpG islands; Shore, within 0–2 kb of a CpG island boundary; Shelf, within 2–4 kb of a CpG island boundary; Inter_CGI, all other regions in the genome. For gene context plots: Promoter, 0–1 kb upstream of the transcription start site; Promoter+, 1–5 kb upstream of the transcription start site. For DMP plots, the Probes column represents the "background" or "default" distribution of all 450K array probes after initial filtering and used as input for DMP analysis. For DMR analysis, the numbers above each bar indicate the number of DMRs identified for each cohort. The following cohorts had no detected DMRs: AUTS18, BAFopathy, CSS4_c.2650, CSS9, Kabuki, KDM4B, MRX93. CGI, CpG island; DMP, differentially methylated probe; DMR, differentially methylated region.

(mean methylation below zero) and 19 (33.9%) cohorts had mean hypermethylation (mean methylation above zero) (Figure 3a). Using a stricter cut-off of at least a 5% change in mean methylation there were 12 (21.4%) hypomethylated cohorts and 10 (17.9%) hypermethylated, while the remaining 22 cohorts had mean methylation changes of less than 5% (Figure 3a).

To investigate relationships across all cohorts without bias caused by the number of DMPs selected, clustering analysis was

TABLE 2 Top significant GO terms from enrichment analysis of DMRs

Ontology	ID	Description	Gene set ratio	Adjusted *p* value
Using all CpG sites in the selected DMRs of all signatures				
BP	GO:0007156	Homophilic cell adhesion via plasma membrane adhesion molecules	75/159	3.55E−17
BP	GO:0098742	Cell–cell adhesion via plasma-membrane adhesion molecules	98/260	8.22E−15
CC	GO:0005887	Integral component of plasma membrane	342/1537	1.35E−13
CC	GO:0031226	Intrinsic component of plasma membrane	355/1612	2.16E−13
BP	GO:0032501	Multicellular organismal process	1204/7241	1.43E−09
CC	GO:0071944	Cell periphery	933/5464	2.50E−08
BP	GO:0003008	System process	390/2068	7.13E−08
CC	GO:0031224	Intrinsic component of membrane	806/4959	1.89E−06
MF	GO:0005509	Calcium ion binding	155/666	4.95E−06
BP	GO:0003002	Regionalization	94/315	3.55E−17
Using duplicated CpG sites in the selected DMRs of all signatures				
BP	GO:0007156	Homophilic cell adhesion via plasma membrane adhesion molecules	44/159	1.07E−21
BP	GO:0098742	Cell–cell adhesion via plasma-membrane adhesion molecules	52/260	3.08E−19
BP	GO:0098609	Cell–cell adhesion	79/837	1.56E−10
CC	GO:0005887	Integral component of plasma membrane	112/1537	7.71E−10
MF	GO:0005509	Calcium ion binding	65/666	2.65E−09
CC	GO:0031226	Intrinsic component of plasma membrane	114/1612	2.96E−09
BP	GO:0022610	Biological adhesion	101/1410	7.17E−07
BP	GO:0007155	Cell adhesion	100/1404	1.05E−06
BP	GO:0016339	Calcium-dependent cell–cell adhesion via plasma membrane cell adhesion molecules	11/42	3.81E−04
CC	GO:0071944	Cell periphery	244/5464	7.97E−03
BP	GO:0019953	Sexual reproduction	49/769	0.02
CC	GO:0031224	Intrinsic component membrane	211/4959	0.04

Abbreviations: DMR, differentially methylated region; GO, Gene Ontology.

performed on the combined top N DMPs for each cohort where N was either 500 or the total number of DMPs as detailed in the Methods section. Results were visualized using a binary tree with each node corresponding to a cohort (Figure 3b). The 56 cohorts were clustered into three groups: one group along the branch of ADCADN (upper left), a second group with the branch on the lower left (containing the ICF2_3_4 sub-branch), and the rest of the tree branches on the right as the third group. Subclustering of the second and third groups is also apparent. Some patterns are evident in the clustering as most of the highly hypo/hypermethylated cohorts are close together. Furthermore, for the other subclusters, cohorts on the same branch are either in the same range of mean methylation difference or number of DMPs due to the similarities in either node color or node size. We also see groupings consistent with our previous analysis of these cohorts where conditions sharing similarities, phenotypically or genetically, were clustered together: such as Sotos, ICF, RMNS, BFLS, and TBRS (Aref-Eshghi et al., 2020), and RSTS1 and RSTS2 (Levy et al., 2022). Cohort pairs were also observed generating terminal branches suggesting high level of similarity. Some of these cohort couples include BAFopathy and CSS9, which are both included in the BAF complex, ARTHS and SBBYSS, which are caused by pathogenic variants in KAT6 genes, and RSTS1 and RSTS2. To visualize global structure, we analyzed all cohorts using the most differentiating probes identified by random forest feature selection. Topological structures were approximated by UMAP and projected into two-dimensional and three-dimensional spaces as seen in Figure 4. Results of this analysis were concordant with the clustering analysis. Cohorts that are more alike are closer together, such as RSTS1 and RSTS2 (Figure 4b), and ARTHS and GTPTS (Figure 4c), which is also associated with variants in a KAT6 gene. While we can see a large degree of overlap for several cohorts in the 2D projection, we also observe locally condensed independent groupings of the same cohorts in the 3D projection. This demonstrates the level of complexity of the overall structure of the data and the effectiveness of a small set of probes to distinguish them to a certain degree.

TABLE 3 Top significant GO terms from enrichment analysis of DMPs

Ontology	ID	Description	Gene set ratio	Adjusted *p* value
Using all selected DMPs in all signatures				
CC	GO:0005737	Cytoplasm	10,892/10,994	3.87E−06
MF	GO:0005515	Protein binding	12,677/12,810	6.32E−05
BP	GO:0007399	Nervous system development	2235/2243	1.94E−03
MF	GO:0043167	Ion binding	5539/5583	3.88E−03
CC	GO:0005829	Cytosol	4940/4978	6.05E−03
BP	GO:0016043	Cellular component organization	5963/6013	6.05E−03
MF	GO:0046872	Metal ion binding	3900/3929	0.03
BP	GO:0022008	Neurogenesis	1544/1549	0.03
BP	GO:0071840	Cellular component organization or biogenesis	6147/6202	0.04
Using DMPs selected in more than five signatures.				
BP	GO:0007156	Homophilic cell adhesion via plasma membrane adhesion molecules	128/159	3.53E−05
BP	GO:0098742	Cell–cell adhesion via plasma-membrane adhesion molecules	192/260	7.34E−04
CC	GO:0030054	Cell junction	1277/1934	7.94E−04
CC	GO:0005911	Cell–cell junction	323/460	0.02
BP	GO:0007268	Chemical synaptic transmission	454/657	0.02
BP	GO:0098916	Anterograde trans-synaptic signaling	454/657	0.02
BP	GO:0099536	Synaptic signaling	471/683	0.02
BP	GO:0099537	Trans-synaptic signaling	458/663	0.02
CC	GO:0120025	Plasma membrane-bounded cell projection	1318/2059	0.02
BP	GO:0032989	Cellular component morphogenesis	512/732	0.02

Abbreviations: DMP, differentially methylated probe; GO, Gene Ontology.

We used network analysis to investigate the relationships of shared DMPs between cohorts. Using this analysis, we visualized shared probes versus non-shared probes ("self-loops"), and the proportions of and directions (hyper- or hypomethylated) of the shared DMPs. (Figure 5). Three observations are notable: First, the probes unique to the cohort (indicated by the self-loops) for the majority of the cohorts are hypermethylated, while the probes shared between the cohorts are all hypomethylated, except for some relationships which show mixed status (hypo/hyper in one cohort and the opposite in the second cohort): BEFAHRS shared hyper-hypo probes with MRD51, MRX93, GADEVS, BISS and DYT28, and Chr16p11.2del shared hyper-hypo probes with DYT28. Second, the ADCADN cohort (which had the largest number of DMPs) did not share a large proportion of probes with any other syndrome. Other cohorts with a high number of DMPs tended to share a larger proportion of their probes. Third, although this is fully connected network in which each cohort shares at least one probe with all other syndromes, it is possible to distinguish groups of cohorts that share a substantial number of the DMPs with each other. One such "triangle" is between Sotos, RMNS, and TBRS (Figure 5). While sharing a small number of probes by the cohorts can happen by chance, a substantial number of shared DMPs may indicate an underlying biological process that is common to all cohorts.

4 | DISCUSSION

4.1 | Significant overlap in DMPs between disorders

DNA methylation episignature analysis can be used as a primary screen for patients with a suspected genetic disorder, or as a reflex test for patients with a variant of unknown significance or with no variant identified but a clinical presentation suggestive of a syndrome with a known episignature (Kerkhof et al., 2021; Sadikovic et al., 2021). "EpiSigns" are sensitive and specific diagnostic biomarkers that can be gene, region, or even nucleotide specific (Levy et al., 2022). To generate a disorder-specific classifier, individual EpiSigns are optimized by selecting the most significant DMPs in relation to controls as well as all other episignature cohorts (Levy et al., 2022). In this study, we analyzed the functional biological aspects of global methylation changes in these disorders. We demonstrate a significant overlap in global DNA methylation profiles across all syndromes.

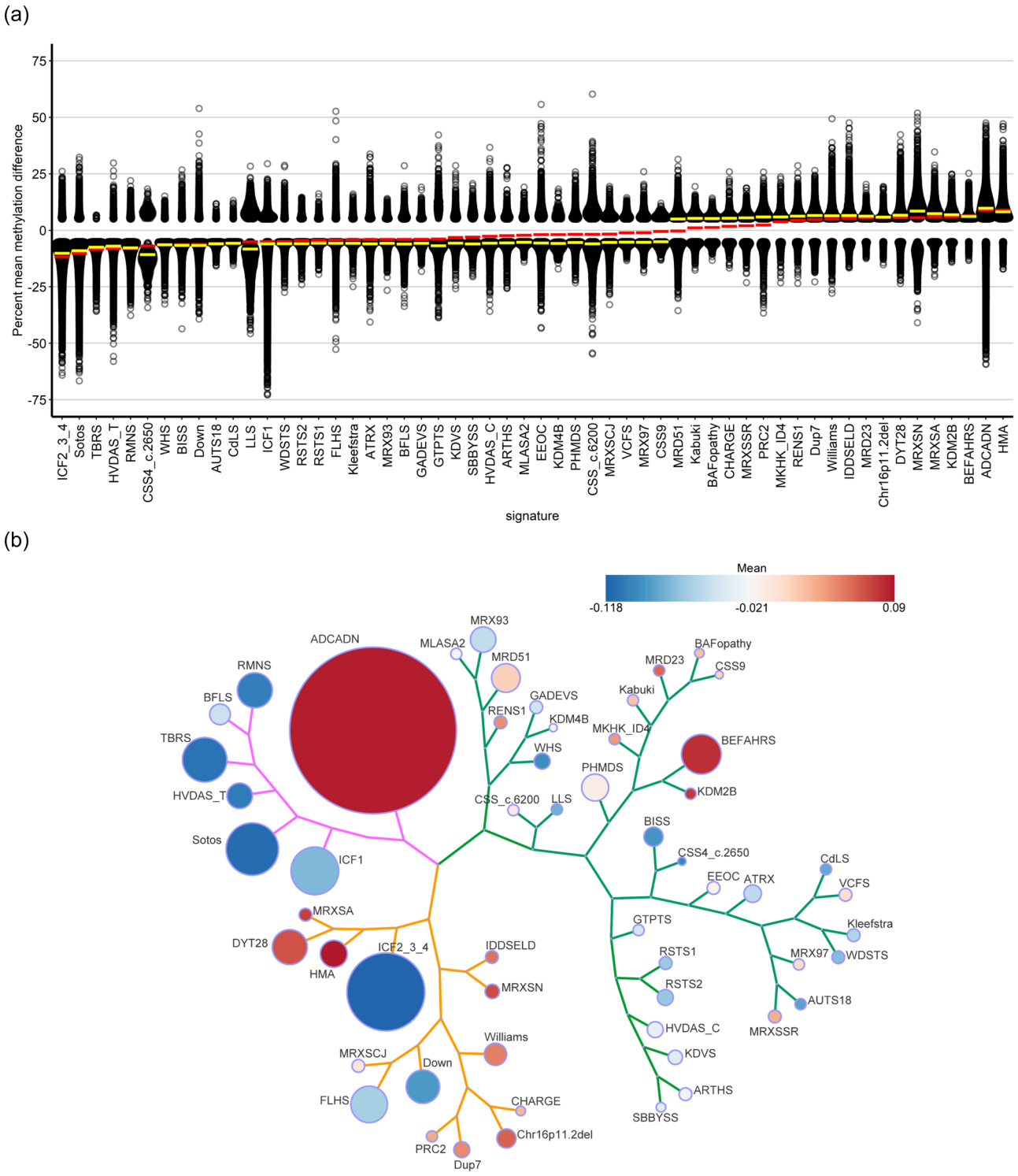

FIGURE 3 Relationships between cohorts. (a) Methylation differences of all differentially methylated probes for each cohort, sorted by mean methylation. Each circle represents one probe. Red lines indicate mean methylation, yellow lines indicate median methylation. (b) Tree and leaf visualization of Euclidean clustering of 56 cohorts using the top n DMPs for each group, where n = min (# of DMPs, 500). Cohort samples were aggregated using the median value of each probe within a group. A leaf node represents a cohort, with node sizes illustrating relative scales of the number of selected DMPs for the corresponding cohort, and node colors are indicative of the global mean methylation difference. Clustering of cohorts are represented by branches containing a subset of the cohorts. The colored branches indicate the main division of the three nodes as described in the text. DMP, differentially methylated probe.

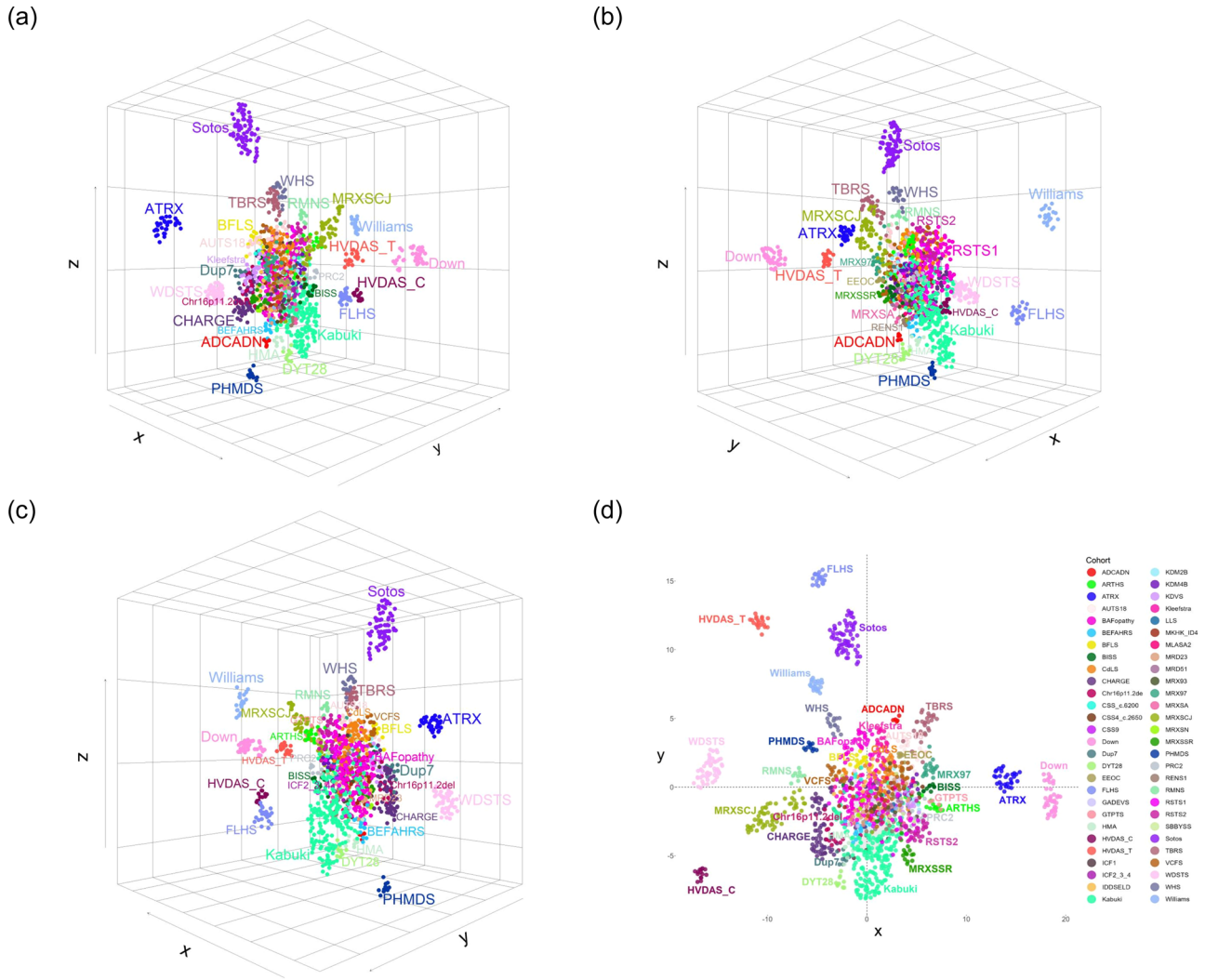

FIGURE 4 UMAP visualization of 56 cohorts using the most differentiating probes. (a–c) UMAP results projected to three-dimensional space and snapshot from different perspectives. (d) UMAP results projected to two-dimensional plane. For each view, a subset of cohort labels has been added dependent on cluster density and to optimize readability. UMAP, unified mapping approximation and projection.

This highlights the importance of training diagnostic EpiSign classifiers against other disorders to achieve diagnostic specificity. Overlap in methylation profiles may reflect similarity in the clinical presentations across the various genetic neurodevelopmental syndromes. Hence the clinical utility of DNA methylation episignatures is directly related to their continuing refinement (Aref-Eshghi et al., 2020; Levy et al., 2022).

A significant proportion of the DNA methylation profile in any disorder is shared with other cohorts. Several disorders share a high percentage of their DMPs including ADCADN, BEFAHRS, RMNS, Sotos, ICF1, and TBRS (Figure 1). These disorders exhibit among the highest number of DMPs, and involve genes related to various aspects of chromatin remodeling including DNA methylation (ADCADN BEFARHS, ICF1, TBRS), histone methylation (Sotos), or linker histones (RMNS). Other disorders involving chromatin remodeling genes demonstrated significant overlap and a high number of DMPs, including FLHS, ICF_2_3_4, HVDAS_T, DYT28, and BFLS. Copy number variant disorders that include chromatin remodeling genes within the deletion and duplications (HMA, Sotos, Dup7, Williams) also demonstrated high degrees of overlap and number of DMPs, and reciprocating deletion and duplication syndromes (HMA vs. Sotos and Dup7 vs. Williams) showed some overlap in probes but were dissimilar in the UMAP clustering.

4.2 | Sotos, TBRS, and RMNS show high overlap in probes

As observed in both Figures 1b and 5, three disorders, Sotos (caused by pathogenic variants in *NSD1)*, RNMS (caused by pathogenic variants in *HIST1H1E*), and TBRS (caused by pathogenic variants in *DNMT3A*), show significant overlap in DMPs. Our group has previously described the relatedness of the episignature probes for these three disorders (Aref-Eshghi et al., 2020), and this similarity is

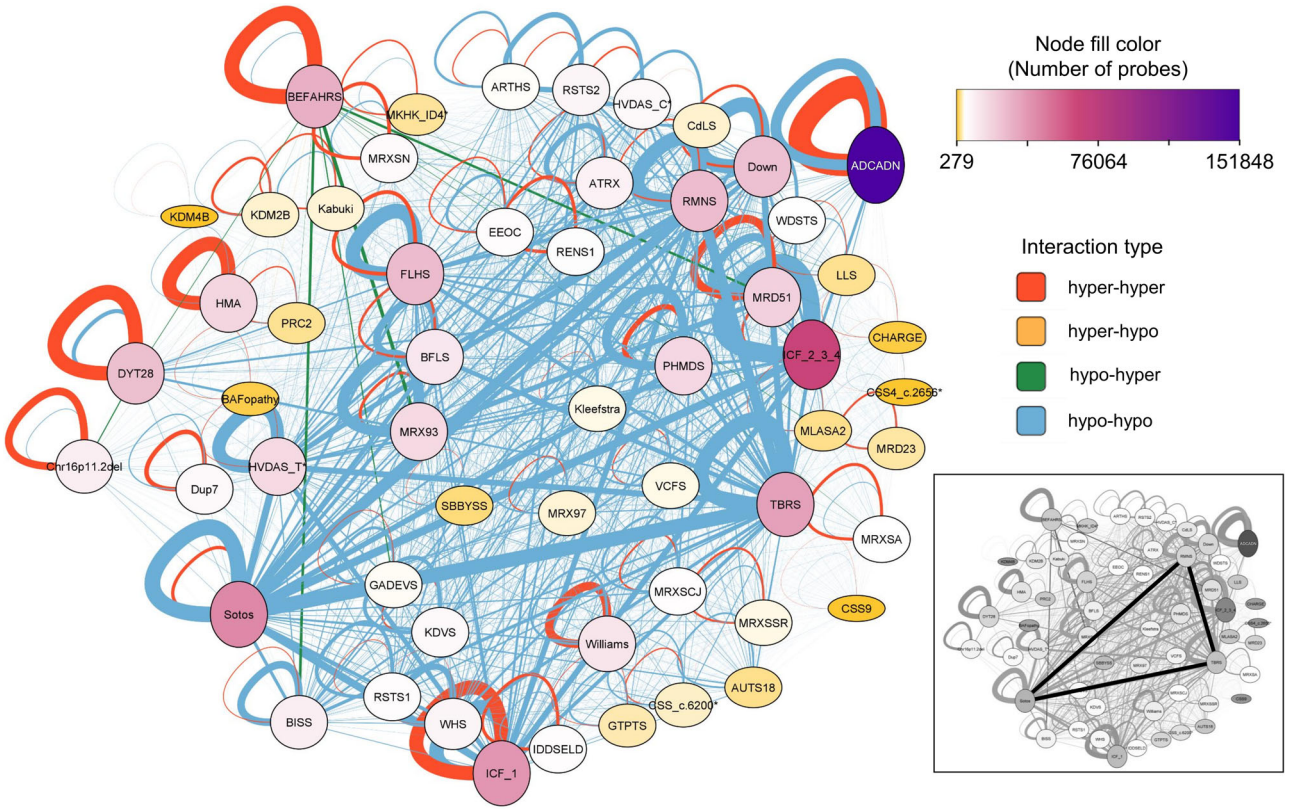

FIGURE 5 Differentially methylated probe sharing between the 56 cohorts. Network diagram showing cohorts connected by edges representing probes shared between them. Edge colors represent hyper–hyper, hyper–hypo, hypo–hyper, and hypo–hypo connections, edge width is proportional to the total number of probes shared (range from 1 to 138,727 shared probes). Probes unique to a cohort are represented by a self-loop. Node colors represent the total number of the differentially methylated probes in a cohort. The inset shows the triangle described in the text.

also observed in our current analysis that looks at all DMPs beyond the small subset of probes used for diagnostic EpiSign purposes. All three genes contribute to overgrowth phenotypes, either throughout childhood (Sotos and TBRS) or in infancy (RMNS), with direct roles in chromatin remodeling (Tatton-Brown et al., 2017). NSD1 is a histone methyltransferase and functional studies have shown that loss of NSD1 results in redistribution of DNMT3A and reduced methylation at the expected regions (Weinberg et al., 2019). Therefore, hypomethylation at shared probes may be a consequence of either lack of NSD1 recruitment of DNMT3A or the loss of DNMT3A altogether. The linker histone H1.4, encoded by the *HIST1H1E* gene, has key roles in chromatin accessibility and compaction (Flex et al., 2019), aligning gene function with other chromatin remodelers histone methyltransferase NSD1 and DNA methyltransferase DNMT3A. Additionally, a study found that these three genes, as well as three others (*CHD8*, *EED*, and *EZH2*), accounted for the pathogenic variants in 44% of patients in a large cohort of patients with overgrowth and intellectual disability (Tatton-Brown et al., 2017), however, a more recent study classifying RMNS further highlights that overgrowth is observed in infancy and that patient growth became progressively closer to average over time (Flex et al., 2019). Therefore, the observed overlap in all probes, as well as their relatedness when assessing only the top 500 probes, is another layer of functional evidence indicating these syndromes may have similar molecular etiology. We also assessed the methylation patterns for the three episignature-causing genes assessed in the study by Tatton-Brown et al. (2017). EZH2 and EED are components of the polycomb repressive complex 2 (PRC2) and samples from patients with pathogenic variants in either gene were included in the PRC2 cohort. CHD8 is an ATP-dependent chromatin-remodeling factor, and variants in this gene cause AUST18. Both the PRC2 and AUST18 cohorts exhibited small numbers of DMPs (less than 3000). However, a large number of their DMPs (between 38% and 74%) are present in TBRS, Sotos, or RMNS probe lists, indicating common regions are impacted in these other overgrowth syndromes.

4.3 | Differences in methylation profiles in paralogous genes

Of the 56 cohorts assessed, two sets of paralogous genes are involved in multiple syndromes. First, KAT6A and KAT6B are paralogous lysine acetyltransferases that form a complex with other proteins to control gene expression by histone acetylation (Wiesel-Motiuk & Assaraf,

2020). Truncating mutations in the C-terminal transactivation domain of *KAT6A* (as observed in our cohort) as well as missense variants, cause ARTHS (Kennedy et al., 2019; Tham et al., 2015). Truncating mutations in the proximal portion of the last exon of its paralog, *KAT6B*, lead to a protein with no transactivation domain and cause GTPTS (Campeau et al., 2012). *KAT6B* pathogenic variants can also lead to another syndrome SBBYSS. SBBYSS-related *KAT6B* variants can result in nonsense-mediated decay leading to a milder phenotype caused by haploinsufficiency (Campeau et al., 2012; Lonardo et al., 2019). SBBYSS and ARTHS cluster more closely on the leaf and tree diagram, while GTPTS is a few branches away. A study suggests that truncating mutations in the proximal portion of the last exon of *KAT6B*, when compared to SBBYSS mutations that occur more distally, lead to a gain of function in the protein (Campeau et al., 2012). This hypothesis proposes that the possible gain of function causes the phenotypes present in GTPTS but not in SBBYSS, with the shared clinical presentation a result of the haploinsufficiency of the transactivation domain (Campeau et al., 2012). This provides a possible reason as to why ARTHS and SBBYSS group more closely when compared to GTPTS. Pathogenic variants causing GTPTS and ARTHS fall in similar regions in the two genes (*KAT6B* and *KAT6A*, respectively), however, the two genes only share 60% sequence homology. Further investigations assessing the protein changes caused by variants may provide further insight as to why ARTHS and SBBYSS are more similar to each other than GTPTS, and to confirm that gain of function mutations in GTPTS underpin these phenotypic differences.

Two other paralogs, *CREBBP* and *EP300*, are associated with three cohorts that were assessed. CREBBP and EP300 are transcriptional coactivators and histone acetyltransferases that interact with over 400 proteins (Bedford et al., 2010). Pathogenic variants in *CREBBP* cause RSTS1 and pathogenic variants in *EP300* cause RSTS2, whereas variants in exon 30 and 31 of either gene can cause MKHK 1 and 2, respectively. Our MKHK cohort contains pathogenic variants in both genes that fall in the intrinsically disordered linker (ID4) region of these proteins (MKHK_ID4). One hypothesis is that missense mutations observed in MKHK patients result in gain of function of the proteins, resulting in a different phenotype compared to the loss of function observed in RSTS (Menke et al., 2018). Our data provides further functional evidence that these two syndromes have different pathological mechanisms with RSTS 1 and 2 showing high similarity to each other and exhibiting mean hypomethylation, while MKHK_ID4 exhibits overall hypermethylation and methylation dissimilarity from RSTS1 and 2, as observed by the tree and leaf diagram (Figure 3b). Gene expression analysis and functional assessment MKHK variants will provide more insight on the molecular mechanisms of these two syndromes.

4.4 | Hypomethylated probes are most commonly shared between disorders

The conditions with the greatest number of DMPs also had the highest numbers of unique probes, as outlined in the network diagram (Figure 5). The overlap in DMPs among all disorders is clear, however, the vast majority of overlapping probes among conditions are hypomethylated. Epigenetic changes in both DNA and histones, both transient and inherited, are essential to proper development, allowing for proper DNA expression that is cell-specific and temporal. Promoter hypomethylation may be indicative of common gene activation across the various syndromes, while hypermethylation may be related to the disorder-specific gene inactivation (Figure 5).

4.5 | GO analysis identifies enrichment in developmental and neurological pathways

The development of specific epigenetic biomarkers aids in the diagnosis of conditions with nonspecific clinical presentations, which include a spectrum of neurodevelopmental delay and dysmorphic features (Kvarnung & Nordgren, 2017). In line with the common clinical features, the significant GO terms for DMPs and DMRs include neurologic processes, such as chemical synaptic transmission, trans-synaptic signaling, synapse assembly, and glutamatergic synaptic transmission. Enrichment was also observed in terms involved in developmental pathways and morphology, such as anatomical structure morphogenesis, cell–cell adhesion pathways, nervous system development. Differential methylation within these genes, or near their promoters and CpG islands warrants further investigation in gene expression of these pathways.

Enrichment of neurodevelopmental pathways points towards the possible inappropriate expression of genes required for proper cortical development. Spatial and temporal control of gene expression through DNA methylation is a highly dynamic process during development and many of the genes represented by the cohorts studied are involved in DNA methylation regulation. A recent review highlights the importance of DNA methylation in neuronal development within a set of neurodevelopmental syndromes, many of which are represented by our cohorts (Ciptasari & van Bokhoven, 2020). Further analysis of specific genes impacted, as well as direction of methylation change in the context of a given disorder, will provide further insight into possible underlying biological pathways that may contribute to a given syndrome phenotype. Gene expression analysis would also further solidify the impact of these methylation changes on the genes in question.

5 | CONCLUSIONS

DNA methylation episignatures are providing very useful diagnostic biomarkers in an expanding number and scope of Mendelian neurodevelopmental disorders. To achieve high levels of accuracy, sensitivity, and specificity, these EpiSigns are limited to a small subset of genomic CpGs providing the most optimal power to differentiate between different disorders. The broader genome-wide DNA methylation profiles provide insights into the complex molecular etiology and pathophysiology of the associated conditions. In this

study, we assessed the relatedness of the global DNA methylation profiles across the majority of human Mendelian episignature disorders described to date. We created a map of the disorder-specific and recurring genome-wide DNA methylation profiles, demonstrating a substantial level of overlap in methylation profiles involving functional genomic elements among these conditions. We provide evidence for an enrichment of DMRs and DMPs across gene promoters and CpG islands suggesting functional roles for the related epigenetic changes. Enrichment analysis further demonstrates disrupted methylation associated with various neurodevelopmental pathways and mechanisms as part of the common etiology across these disorders. The degree of relatedness in the overlapping methylation profiles reflects the similarities in both gene function as well as the clinical presentations, suggesting a strong functional role of DNA methylation in the disease etiology. This study provides the foundation for future work including integration of gene expression and other omic analyses as part of a more comprehensive map of genetic/epigenetic interactions in Mendelian neurodevelopmental disorders.

ACKNOWLEDGMENTS

Funding for this study was provided in part by the London Health Sciences Molecular Diagnostics Development Fund. This work was funded by the government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-188) The research conducted at the Murdoch Children's Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program. The Chair in Genomic Medicine awarded to JC is generously supported by The Royal Children's Hospital Foundation. Funding was provided in part from the Italian Ministry of Health (5 × 1000, CCR-2017-23669081 and RCR-2020-23670068_001, to M. T.) and the Italian Ministry of Research (FOE 2019, to M. T.). Funding was provided in part by a grant from the South Carolina Department of Disabilities and Special Needs. In memory of Ethan Francis Schwartz, 1996–1998.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Gerarda Cappuccio http://orcid.org/0000-0003-3934-2342
Angus Clarke http://orcid.org/0000-0002-1200-9286
Cristina Gervasini http://orcid.org/0000-0002-1165-7935
Zandra Jenkins http://orcid.org/0000-0003-4628-9789
Stephen Robertson http://orcid.org/0000-0002-5181-7809
Gijs W. E. Santen http://orcid.org/0000-0003-1959-3267
Jozef Gecz http://orcid.org/0000-0002-7884-6861
Kym Boycott http://orcid.org/0000-0003-4186-8052
Nicola Brunetti-Pierri http://orcid.org/0000-0002-6895-8819
Philippe M. Campeau http://orcid.org/0000-0001-9713-7107
John Christodoulou http://orcid.org/0000-0002-8431-0641
Mark D. Fleming http://orcid.org/0000-0003-0948-4024
Marco Tartaglia http://orcid.org/0000-0001-7736-9672
Bekim Sadikovic http://orcid.org/0000-0001-6363-0016

REFERENCES

Aref-Eshghi, E., Bend, E. G., Hood, R. L., Schenkel, L. C., Carere, D. A., Chakrabarti, R., Nagamani, S., Cheung, S. W., Campeau, P. M., Prasad, C., Siu, V. M., Brady, L., Tarnopolsky, M. A., Callen, D. J., Innes, A. M., White, S. M., Meschino, W. S., Shuen, A. Y., Paré, G., … Sadikovic, B. (2018). BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin-Siris and Nicolaides-Baraitser syndromes. *Nature Communications*, *9*(1), 4885. https://doi.org/10.1038/s41467-018-07193-y

Aref-Eshghi, E., Kerkhof, J., Pedro, V. P., Barat-Houari, M., Ruiz-Pallares, N., Andrau, J. C., Lacombe, D., Van-Gils, J., Fergelot, P., Dubourg, C., Cormier-Daire, V., Rondeau, S., Lecoquierre, F., Saugier-Veber, P., Nicolas, G., Lesca, G., Chatron, N., Sanlaville, D., Vitobello, A., … Sadikovic, B. (2020). Evaluation of DNA methylation episignatures for diagnosis and phenotype correlations in 42 Mendelian neurodevelopmental disorders. *American Journal of Human Genetics*, *106*(3), 356–370. https://doi.org/10.1016/j.ajhg.2020.01.019

Aref-Eshghi, E., Rodenhiser, D. I., Schenkel, L. C., Lin, H., Skinner, C., Ainsworth, P., Paré, G., Hood, R. L., Bulman, D. E., Kernohan, K. D., Boycott, K. M., Campeau, P. M., Schwartz, C., & Sadikovic, B. (2018). Genomic DNA methylation signatures enable concurrent diagnosis and clinical genetic variant classification in neurodevelopmental syndromes. *American Journal of Human Genetics*, *102*(1), 156–174. https://doi.org/10.1016/j.ajhg.2017.12.008

Aref-Eshghi, E., Schenkel, L. C., Lin, H., Skinner, C., Ainsworth, P., Paré, G., Rodenhiser, D., Schwartz, C., & Sadikovic, B. (2017). The defining DNA methylation signature of Kabuki syndrome enables functional assessment of genetic variants of unknown clinical significance. *Epigenetics*, *12*(11), 923–933. https://doi.org/10.1080/15592294.2017.1381807

Bedford, D. C., Kasper, L. H., Fukuyama, T., & Brindle, P. K. (2010). Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases. *Epigenetics*, *5*(1), 9–15. https://doi.org/10.4161/epi.5.1.10449

Bend, E. G., Aref-Eshghi, E., Everman, D. B., Rogers, R. C., Cathey, S. S., Prijoles, E. J., Lyons, M. J., Davis, H., Clarkson, K., Gripp, K. W., Li, D., Bhoj, E., Zackai, E., Mark, P., Hakonarson, H., Demmer, L. A., Levy, M. A., Kerkhof, J., Stuart, A., … Sadikovic, B. (2019). Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome. *Clinical Epigenetics*, *11*(1), 64. https://doi.org/10.1186/s13148-019-0658-5

Campeau, P. M., Lu, J. T., Dawson, B. C., Fokkema, I. F., Robertson, S. P., Gibbs, R. A., & Lee, B. H. (2012). The KAT6B-related disorders genitopatellar syndrome and Ohdo/SBBYS syndrome have distinct clinical features reflecting distinct molecular mechanisms. *Human Mutation*, *33*(11), 1520–1525. https://doi.org/10.1002/humu.22141

Cavalcante, R. G., & Sartor, M. A. (2017). annotatr: Genomic regions in context. *Bioinformatics*, *33*(15), 2381–2383. https://doi.org/10.1093/bioinformatics/btx183

Chen, Y. A., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., Gallinger, S., Hudson, T. J., & Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the illumina infinium HumanMethylation450 microarray. *Epigenetics*, *8*(2), 203–209. https://doi.org/10.4161/epi.23470

Ciolfi, A., Aref-Eshghi, E., Pizzi, S., Pedace, L., Miele, E., Kerkhof, J., Flex, E., Martinelli, S., Radio, F. C., Ruivenkamp, C., Santen, G., Bijlsma, E., Barge-Schaapveld, D., Ounap, K., Siu, V. M., Kooy, R. F., Dallapiccola, B., Sadikovic, B., & Tartaglia, M. (2020). Frameshift mutations at the C-terminus of HIST1H1E result in a specific DNA

hypomethylation signature. *Clinical Epigenetics*, *12*(1), 7. https://doi.org/10.1186/s13148-019-0804-0

Ciolfi, A., Foroutan, A., Capuano, A., Pedace, L., Travaglini, L., Pizzi, S., Andreani, M., Miele, E., Invernizzi, F., Reale, C., Panteghini, C., Iascone, M., Niceta, M., Gavrilova, R. H., Schultz-Rogers, L., Agolini, E., Bedeschi, M. F., Prontera, P., Garibaldi, M., … Sadikovic, B. (2021). Childhood-onset dystonia-causing KMT2B variants result in a distinctive genomic hypermethylation profile. *Clinical Epigenetics*, *13*(1), 157. https://doi.org/10.1186/s13148-021-01145-y

Ciptasari, U., & van Bokhoven, H. (2020). The phenomenal epigenome in neurodevelopmental disorders. *Human Molecular Genetics*, *29*(R1), R42–R50. https://doi.org/10.1093/hmg/ddaa175

Flex, E., Martinelli, S., Van Dijck, A., Ciolfi, A., Cecchetti, S., Coluzzi, E., Pannone, L., Andreoli, C., Radio, F. C., Pizzi, S., Carpentieri, G., Bruselles, A., Catanzaro, G., Pedace, L., Miele, E., Carcarino, E., Ge, X., Chijiwa, C., Lewis, M., … Tartaglia, M. (2019). Aberrant function of the C-terminal tail of HIST1H1E accelerates cellular senescence and causes premature aging. *American Journal of Human Genetics*, *105*(3), 493–508. https://doi.org/10.1016/j.ajhg.2019.07.007

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, *20*(10), 590–607. https://doi.org/10.1038/s41580-019-0159-6

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). circlize implements and enhances circular visualization in R. *Bioinformatics*, *30*(19), 2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Haghshenas, S., Levy, M. A., Kerkhof, J., Aref-Eshghi, E., McConkey, H., Balci, T., Siu, V. M., Skinner, C. D., Stevenson, R. E., Sadikovic, B., & Schwartz, C. (2021). Detection of a DNA methylation signature for the intellectual developmental disorder, X-linked, syndromic, arm-field type. *International Journal of Molecular Sciences*, *22*(3), 1111. https://doi.org/10.3390/ijms22031111

Hood, R. L., Schenkel, L. C., Nikkel, S. M., Ainsworth, P. J., Pare, G., Boycott, K. M., Bulman, D. E., & Sadikovic, B. (2016). The defining DNA methylation signature of Floating-Harbor syndrome. *Scientific Reports*, *6*, 38803. https://doi.org/10.1038/srep38803

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., Wiencke, J. K., & Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, *13*, 86. https://doi.org/10.1186/1471-2105-13-86

Janssen, S. M., & Lorincz, M. C. (2021). Interplay between chromatin marks in development and disease. *Nature Reviews Genetics*, *23*, 137–153. https://doi.org/10.1038/s41576-021-00416-x

Kennedy, J., Goudie, D., Blair, E., Chandler, K., Joss, S., McKay, V., Green, A., Armstrong, R., Lees, M., Kamien, B., Hopper, B., Tan, T. Y., Yap, P., Stark, Z., Okamoto, N., Miyake, N., Matsumoto, N., Macnamara, E., Murphy, J. L., … Newbury-Ecob, R. (2019). KAT6A syndrome: Genotype-phenotype correlation in 76 patients with pathogenic KAT6A variants. *Genetics in Medicine*, *21*(4), 850–860. https://doi.org/10.1038/s41436-018-0259-2

Kerkhof, J., Squeo, G. M., McConkey, H., Levy, M. A., Piemontese, M. R., Castori, M., Accadia, M., Biamino, E., Della Monica, M., Di Giacomo, M. C., Gervasini, C., Maitz, S., Melis, D., Milani, D., Piccione, M., Prontera, P., Selicorni, A., Sadikovic, B., & Merla, G. (2021). DNA methylation episignature testing improves molecular diagnosis of Mendelian chromatinopathies. *Genetics in Medicine*, *24*, 51–60. https://doi.org/10.1016/j.gim.2021.08.007

Krzyzewska, I. M., Maas, S. M., Henneman, P., Lip, K., Venema, A., Baranano, K., Chassevent, A., Aref-Eshghi, E., van Essen, A. J., Fukuda, T., Ikeda, H., Jacquemont, M., Kim, H. G., Labalme, A., Lewis, S., Lesca, G., Madrigal, I., Mahida, S., Matsumoto, N., … Mannens, M. (2019). A genome-wide DNA methylation signature for SETD1B-related syndrome. *Clinical Epigenetics*, *11*(1), 156. https://doi.org/10.1186/s13148-019-0749-3

Kvarnung, M., & Nordgren, A. (2017). Intellectual disability & rare disorders: A diagnostic challenge. *Advances in Experimental Medicine and Biology*, *1031*, 39–54. https://doi.org/10.1007/978-3-319-67144-4_3

Levy, M. A., Beck, D. B., Metcalfe, K., Douzgou, S., Sithambaram, S., Cottrell, T., Ansar, M., Kerkhof, J., Mignot, C., Nougues, M. C., Keren, B., Moore, H. W., Oegema, R., Giltay, J. C., Simon, M., van Jaarsveld, R. H., Bos, J., van Haelst, M., Motazacker, M. M., … Fahrner, J. A. (2021). Deficiency of TET3 leads to a genome-wide DNA hypermethylation episignature in human whole blood. *NPJ Genomic Medicine*, *6*(1), 92. https://doi.org/10.1038/s41525-021-00256-y

Levy, M. A., McConkey, H., Kerkhof, J., Barat-Houari, M., Bargiacchi, S., Biamino, E., Bralo, M. P., Cappuccio, G., Ciolfi, A., Clarke, A., DuPont, B. R., Elting, M. W., Faivre, L., Fee, T., Fletcher, R. S., Cherik, F., Foroutan, A., Friez, M. J., Gervasini, C., … Sadikovic, B. (2022). Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. *HGG Advances*, *3*(1), 100075. https://doi.org/10.1016/j.xhgg.2021.100075

Lonardo, F., Lonardo, M. S., Acquaviva, F., Della Monica, M., Scarano, F., & Scarano, G. (2019). Say-Barber-Biesecker-Young-Simpson syndrome and genitopatellar syndrome: Lumping or splitting? *Clinical Genetics*, *95*(2), 253–261. https://doi.org/10.1111/cge.13127.

Menke, L. A., Ddd, s, Gardeitchik, T., Hammond, P., Heimdal, K. R., Houge, G., Hufnagel, S. B., Ji, J., Johansson, S., Kant, S. G., Kinning, E., Leon, E. L., Newbury-Ecob, R., Paolacci, S., Pfundt, R., Ragge, N. K., Rinne, T., Ruivenkamp, C., Saitta, S. C., … Hennekam, R. C. (2018). Further delineation of an entity caused by CREBBP and EP300 mutations but not resembling Rubinstein-Taybi syndrome. *American journal of medical genetics. Part A*, *176*(4), 862–876. https://doi.org/10.1002/ajmg.a.38626

Online Mendelian Inheritance in Man, OMIM®. (2022). McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). www.omim.org

Peters, T. J., Buckley, M. J., Statham, A. L., Pidsley, R., Samaras, K., V Lord, R., Clark, S. J., & Molloy, P. L. (2015). De novo identification of differentially methylated regions in the human genome. *Epigenetics & Chromatin*, *8*, 6. https://doi.org/10.1186/1756-8935-8-6

Phipson, B., Maksimovic, J., & Oshlack, A. (2016). missMethyl: An R package for analyzing data from illumina's HumanMethylation450 platform. *Bioinformatics*, *32*(2), 286–288. https://doi.org/10.1093/bioinformatics/btv560

Pidsley, R., Zotenko, E., Peters, T. J., Lawrence, M. G., Risbridger, G. P., Molloy, P., Van Djik, S., Muhlhausler, B., Stirzaker, C., & Clark, S. J. (2016). Critical evaluation of the illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology*, *17*(1), 208. https://doi.org/10.1186/s13059-016-1066-1

Radio, F. C., Pang, K., Ciolfi, A., Levy, M. A., Hernández-García, A., Pedace, L., Pantaleoni, F., Liu, Z., de Boer, E., Jackson, A., Bruselles, A., McConkey, H., Stellacci, E., Lo Cicero, S., Motta, M., Carrozzo, R., Dentici, M. L., McWalter, K., Desai, M., … Tartaglia, M. (2021). SPEN haploinsufficiency causes a neurodevelopmental disorder overlapping proximal 1p36 deletion syndrome with an episignature of X chromosomes in females. *American Journal of Human Genetics*, *108*(3), 502–516. https://doi.org/10.1016/j.ajhg.2021.01.015

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

Rooney, K., Levy, M. A., Haghshenas, S., Kerkhof, J., Rogaia, D., Tedesco, M. G., Imperatore, V., Mencarelli, A., Squeo, G. M., Di Venere, E., Di Cara, G., Verrotti, A., Merla, G., Tedder, M. L.,

DuPont, B. R., Sadikovic, B., & Prontera, P. (2021). Identification of a DNA methylation episignature in the 22q11.2 deletion syndrome. *International Journal of Molecular Sciences*, *22*(16), 8611. https://doi.org/10.3390/ijms22168611

Sadikovic, B., Aref-Eshghi, E., Levy, M. A., & Rodenhiser, D. (2019). DNA methylation signatures in Mendelian developmental disorders as a diagnostic bridge between genotype and phenotype. *Epigenomics*, *11*(5), 563–575. https://doi.org/10.2217/epi-2018-0192

Sadikovic, B., Levy, M. A., Kerkhof, J., Aref-Eshghi, E., Schenkel, L., Stuart, A., McConkey, H., Henneman, P., Venema, A., Schwartz, C. E., Stevenson, R. E., Skinner, S. A., DuPont, B. R., Fletcher, R. S., Balci, T. B., Siu, V. M., Granadillo, J. L., Masters, J., Kadour, M., … Alders, M. (2021). Clinical epigenomics: Genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. *Genetics in Medicine*, *23*(6), 1065–1074. https://doi.org/10.1038/s41436-020-01096-4

Schenkel, L. C., Aref-Eshghi, E., Rooney, K., Kerkhof, J., Levy, M. A., McConkey, H., Rogers, R. C., Phelan, K., Sarasua, S. M., Jain, L., Pauly, R., Boccuto, L., DuPont, B., Cappuccio, G., Brunetti-Pierri, N., Schwartz, C. E., & Sadikovic, B. (2021). DNA methylation epi-signature is associated with two molecularly and phenotypically distinct clinical subtypes of Phelan-McDermid syndrome. *Clinical Epigenetics*, *13*(1), 2. https://doi.org/10.1186/s13148-020-00990-7

Schenkel, L. C., Aref-Eshghi, E., Skinner, C., Ainsworth, P., Lin, H., Paré, G., Rodenhiser, D. I., Schwartz, C., & Sadikovic, B. (2018). Peripheral blood epi-signature of Claes-Jensen syndrome enables sensitive and specific identification of patients and healthy carriers with pathogenic mutations in KDM5C. *Clinical Epigenetics*, *10*, 21. https://doi.org/10.1186/s13148-018-0453-8

Schenkel, L. C., Kernohan, K. D., McBride, A., Reina, D., Hodge, A., Ainsworth, P. J., Rodenhiser, D. I., Pare, G., Bérubé, N. G., Skinner, C., Boycott, K. M., Schwartz, C., & Sadikovic, B. (2017). Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome. *Epigenetics & Chromatin*, *10*, 10. https://doi.org/10.1186/s13072-017-0118-4

Siu, M. T., Butcher, D. T., Turinsky, A. L., Cytrynbaum, C., Stavropoulos, D. J., Walker, S., Caluseriu, O., Carter, M., Lou, Y., Nicolson, R., Georgiades, S., Szatmari, P., Anagnostou, E., Scherer, S. W., Choufani, S., Brudno, M., & Weksberg, R. (2019). Functional DNA methylation signatures for autism spectrum disorder genomic risk loci: 16p11.2 deletions and CHD8 variants. *Clinical Epigenetics*, *11*(1), 103. https://doi.org/10.1186/s13148-019-0684-3

Su, G., Morris, J. H., Demchak, B., & Bader, G. D. (2014). Biological network exploration with Cytoscape 3. *Current Protocols in Bioinformatics*, *47*(8. 13), 11–24. https://doi.org/10.1002/0471250953.bi0813s47

Tatton-Brown, K., Loveday, C., Yost, S., Clarke, M., Ramsay, E., Zachariou, A., Elliott, A., Wylie, H., Ardissone, A., Rittinger, O., Stewart, F., Temple, I. K., Cole, T., Childhood Overgrowth, C., Mahamdallie, S., Seal, S., Ruark, E., & Rahman, N. (2017). Mutations in epigenetic regulation genes are a major cause of overgrowth with intellectual disability. *American Journal of Human Genetics*, *100*(5), 725–736. https://doi.org/10.1016/j.ajhg.2017.03.010

Tham, E., Lindstrand, A., Santani, A., Malmgren, H., Nesbitt, A., Dubbs, H. A., Zackai, E. H., Parker, M. J., Millan, F., Rosenbaum, K., Wilson, G. N., & Nordgren, A. (2015). Dominant mutations in KAT6A cause intellectual disability with recognizable syndromic features. *American Journal of Human Genetics*, *96*(3), 507–513. https://doi.org/10.1016/j.ajhg.2015.01.016

Velasco, G., & Francastel, C. (2019). Genetics meets DNA methylation in rare diseases. *Clinical Genetics*, *95*(2), 210–220. https://doi.org/10.1111/cge.13480

Villicana, S., & Bell, J. T. (2021). Genetic impacts on DNA methylation: Research findings and future perspectives. *Genome Biology*, *22*(1), 127. https://doi.org/10.1186/s13059-021-02347-6

Weinberg, D. N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K. N., Horth, C., McGuire, J. T., Xu, X., Nikbakht, H., Lemiesz, A. E., Marchione, D. M., Marunde, M. R., Meiners, M. J., Cheek, M. A., Keogh, M. C., Bareke, E., Djedid, A., Harutyunyan, A. S., … Lu, C. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature*, *573*(7773), 281–286. https://doi.org/10.1038/s41586-019-1534-3

Wiesel-Motiuk, N., & Assaraf, Y. G. (2020). The key roles of the lysine acetyltransferases KAT6A and KAT6B in physiology and pathology. *Drug Resistance Updates*, *53*, 100729. https://doi.org/10.1016/j.drup.2020.100729

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Levy, M. A., Relator, R., McConkey, H., Pranckeviciene, E., Kerkhof, J., Barat-Houari, M., Bargiacchi, S., Biamino, E., Palomares Bralo, M., Cappuccio, G., Ciolfi, A., Clarke, A., DuPont, B. R., Elting, M. W., Faivre, L., Fee, T., Ferilli, M., Fletcher, R. S., Cherick, F., … Sadikovic, B. (2022). Functional correlation of genome-wide DNA methylation profiles in genetic neurodevelopmental disorders. *Human Mutation*, 43, 1609–1628. https://doi.org/10.1002/humu.24446