

1 **TITLE**

2 *Diagnostic Utility of Genome-wide DNA Methylation Analysis in Genetically Unsolved*
3 *Developmental and Epileptic Encephalopathies and Refinement of a CHD2 Episignature*

4
5 **AUTHORS AND AFFILIATIONS**

6 Christy W. LaFlamme^{1,2*}, Cassandra Rastin^{3,4*}, Soham Sengupta¹, Helen E. Pennington^{1,5},
7 Sophie J. Russ-Hall⁶, Amy L. Schneider⁶, Emily S. Bonkowski¹, Edith P. Almanza Fuerte¹,
8 Miranda Galey⁷, Joy Goffena⁷, Sophia B. Gibson^{7,8}, Talia J. Allan⁶, Denis M. Nyaga⁹, Nico
9 Lieffering⁹, Malavika Hebbar⁷, Emily V. Walker¹⁰, Daniel Darnell¹⁰, Scott R. Olsen¹⁰, Pandurang
10 Kolekar¹¹, Nahdir Djekidel¹², Wojciech Rosikiewicz¹², Haley McConkey⁴, Jennifer Kerkhof⁴,
11 Michael A. Levy⁴, Raissa Relator⁴, Dorit Lev¹³, Tally Lerman-Sagie^{14,15}, Kristen L. Park¹⁶, Marielle
12 Alders¹⁷, Gerarda Cappuccio^{18,19}, Nicolas Chatron^{20,21}, Leigh Demain²², David Genevieve²³,
13 Gaetan Lesca^{20,21}, Tony Roscioli²⁴⁻²⁷, Damien Sanlaville^{20,21}, Matthew L. Tedder²⁸, Monika Weisz
14 Hubshman^{29,30}, Shamika Ketkar²⁹, Hongzheng Dai²⁹, Kim Carlyle Worley²⁹, Jill A. Rosenfeld²⁹,
15 Hsiao-Tuan Chao^{29,31-35}, Undiagnosed Diseases Network, Geoffrey Neale¹⁰, Gemma L. Carvill³⁶,
16 University of Washington Center for Rare Disease Research, Zhaoming Wang^{11,37}, Samuel F.
17 Berkovic⁶, Lynette G. Sadleir⁹, Danny E. Miller^{7,38-39}, Ingrid E. Scheffer^{6,40-41}, Bekim Sadikovic^{3,4#},
18 Heather C. Mefford^{1#}

19
20 ¹Center for Pediatric Neurological Disease Research, St. Jude Children's Research Hospital,
21 Memphis, TN 38105, USA

22 ²Graduate School of Biomedical Sciences, St. Jude Children's Research Hospital, Memphis, TN
23 38105, USA

24 ³Department of Pathology & Laboratory Medicine, Western University, London, Ontario, N5A
25 3K7, Canada

26 ⁴Verspeeten Clinical Genome Centre, London Health Science Centre, London, Ontario, N6A
27 5W9, Canada

28 ⁵Department of Mathematics & Statistics, Rhodes College, Memphis, TN 38112, USA

29 ⁶Epilepsy Research Centre, Department of Medicine, University of Melbourne, Austin Health,
30 Heidelberg, Victoria 3084, Australia

31 ⁷Division of Genetic Medicine, Department of Pediatrics, University of Washington and Seattle
32 Children's Hospital, Seattle, WA 98195, USA

33 ⁸Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA
34 98195, USA

35 ⁹Department of Paediatrics and Child Health, University of Otago, Wellington 6242, New
36 Zealand

37 ¹⁰Hartwell Center for Bioinformatics and Biotechnology, St. Jude Children's Research Hospital
38 Memphis, TN 38105, USA

39 ¹¹Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN,
40 38105, USA

41 ¹² Center for Applied Bioinformatics, St. Jude Children's Research Hospital, Memphis, TN,
42 38105, USA

43 ¹³Institute of Medical Genetics, Wolfson Medical Center, Holon 58100, Israel

44 ¹⁴Fetal Neurology Clinic, Pediatric Neurology Unit, Wolfson Medical Center, Holon 58100, Israel

45 ¹⁵Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel

46 ¹⁶Departments of Pediatrics and Neurology, University of Colorado School of Medicine, Aurora,
47 CO 80045, USA

48 ¹⁷Amsterdam UMC, University of Amsterdam, Department of Human Genetics, Amsterdam
49 Reproduction and Development Research Institute, Meibergdreef 9, 1105 AZ Amsterdam,
50 Netherlands

51 ¹⁸Telethon Institute of Genetics and Medicine, Pozzuoli, Italy

52 ¹⁹Department of Translational Medicine, Federico II University of Naples, Naples, Italy

53 ²⁰Department of Medical Genetics, Member of the ERN EpiCARE, University Hospital of Lyon
54 and Claude Bernard Lyon I University, Lyon France

55 ²¹Pathophysiology and Genetics of Neuron and Muscle (PNMG), UCBL, CNRS UMR5261 -
56 INSERM U1315, Lyon, France

57 ²²Manchester Centre for Genomic Medicine, United Kingdom

58 ²³Montpellier University, Reference Center for Rare Disease, Medical Genetic Department for
59 Rare Disease and Personalize Medicine, Inserm Unit 1183, CHU Montpellier, Montpellier,
60 France

61 ²⁴Neuroscience Research Australia (NeuRA), Sydney, Australia

62 ²⁵Prince of Wales Clinical School, Faculty of Medicine, University of New South Wales, Sydney,
63 Australia

64 ²⁶New South Wales Health Pathology Randwick Genomics, Prince of Wales Hospital, Sydney,
65 Australia

66 ²⁷Centre for Clinical Genetics, Sydney Children's Hospital, Sydney, Australia

67 ²⁸Greenwood Genetic Center, Greenwood, SC, 29646, USA

68 ²⁹Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX
69 77030, USA

70 ³⁰Texas Children's Hospital, Genetic Department, Houston, TX 77030, USA

71 ³¹Department of Pediatrics, Section of Neurology and Developmental Neuroscience, Baylor
72 College of Medicine, Houston, TX 77030, USA

73 ³²Cain Pediatric Neurology Research Foundation Laboratories, Jan and Dan Duncan
74 Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA, 77030

75 ³³Texas Children's Hospital, Houston, TX 77030, USA

76 ³⁴Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA

77 ³⁵McNair Medical Institute, The Robert and Janice McNair Foundation, Houston, TX 77030,
78 USA

79 ³⁶Ken and Ruth Davee Department of Neurology, Northwestern University Feinberg School of
80 Medicine, Chicago, IL, USA

81 ³⁷Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital,
82 Memphis, TN 38105, USA

83 ³⁸Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA
84 98195, USA

85 ³⁹Brotman Baty Institute for Precision Medicine, University of Washington, Seattle, WA 98195,
86 USA

87 ⁴⁰Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Melbourne,
88 Australia

89 ⁴¹Florey Institute and Murdoch Children's Research Institute, Melbourne, Australia

90

91 *These authors contributed equally.

92 #Correspondence: Heather.Mefford@stjude.org and Bekim.Sadikovic@lhsc.on.ca

93

94 **Keywords:** DNA methylation, unsolved, epilepsy, long-read sequencing, nanopore

95 **ABSTRACT**

96 Sequence-based genetic testing currently identifies causative genetic variants in ~50% of
97 individuals with developmental and epileptic encephalopathies (DEEs). Aberrant changes in DNA
98 methylation are implicated in various neurodevelopmental disorders but remain unstudied in
99 DEEs. Rare epigenetic variations (“epivariants”) can drive disease by modulating gene expression
100 at single loci, whereas genome-wide DNA methylation changes can result in distinct
101 “episignature” biomarkers for monogenic disorders in a growing number of rare diseases. Here,
102 we interrogate the diagnostic utility of genome-wide DNA methylation array analysis on peripheral
103 blood samples from 516 individuals with genetically unsolved DEEs who had previously
104 undergone extensive genetic testing. We identified rare differentially methylated regions (DMRs)
105 and explanatory episignatures to discover causative and candidate genetic etiologies in 10
106 individuals. We then used long-read sequencing to identify DNA variants underlying rare DMRs,
107 including one balanced translocation, three CG-rich repeat expansions, and two copy number
108 variants. We also identify pathogenic sequence variants associated with episignatures; some had
109 been missed by previous exome sequencing. Although most DEE genes lack known
110 episignatures, the increase in diagnostic yield for DNA methylation analysis in DEEs is
111 comparable to the added yield of genome sequencing. Finally, we refine an episignature for *CHD2*
112 using an 850K methylation array which was further refined at higher CpG resolution using bisulfite
113 sequencing to investigate potential insights into *CHD2* pathophysiology. Our study demonstrates
114 the diagnostic yield of genome-wide DNA methylation analysis to identify causal and candidate
115 genetic causes as ~2% (10/516) for unsolved DEE cases.

116 **KEYWORDS:** DNA methylation; epivariant; episignature; molecular diagnostics; developmental
117 and epileptic encephalopathies

118

119 **INTRODUCTION**

120 The developmental and epileptic encephalopathies (DEEs) are the most severe group of
121 epilepsies, defined by frequent epileptiform activity associated with developmental slowing or
122 regression¹. While each genetic etiology is rare, with more than 825 genes implicated², the
123 cumulative incidence of DEEs overall is 1 in 590 children³. Currently, *de novo*, X-linked, or
124 recessively inherited pathogenic germline variants are found in ~50% of individuals with DEEs
125 who undergo genetic testing⁴. These are identified by gene panels, exome sequencing (ES), and
126 now, genome sequencing (GS)⁵⁻⁷. A smaller subset is explained by copy number variants
127 (CNVs)⁸. Understanding the etiology guides management, such as clinical trial participation,
128 informs accurate reproductive counseling, enables families to join gene-based support groups,
129 and facilitates the development of targeted therapies⁹⁻¹². This, in turn, improves outcomes but is
130 not possible when the etiology is unknown (“unsolved”).

131 Epigenetic modifications, which alter the DNA without changing the DNA nucleotide
132 sequence, determine the etiology of some individuals with neurodevelopmental disorders but
133 have not yet been studied in the DEEs. DNA methylation is an essential epigenetic modification
134 that regulates cellular gene expression by adding a methyl (CH₃) group to a DNA strand, typically
135 at CpG sites. This can occur through methylation of promoter CpGs, genomic imprinting, and X-
136 chromosome inactivation¹³. Rare epigenetic variations (“epivariants”) disrupt normal methylation
137 and cause disease. While DNA methylation does not change the DNA sequence itself, epivariants
138 are often perpetrated by underlying in-cis DNA changes, such as rare sequence variants,
139 structural alterations, and CG-rich repeat expansions¹⁴ that are difficult to identify by standard
140 sequencing. One example is the methylation of the 5’ untranslated region (5’UTR) of *FMR1*
141 (MIM:309550) that represses gene expression and causes Fragile X syndrome (MIM:300624).
142 Similarly, hypermethylation of the 5’UTR of Xylosyltransferase 1 (*XYLT1*, MIM:608124), leading
143 to gene silencing, may identify the “missing” allele in the recessive disease Baratela-Scott

144 syndrome (BSS [MIM:615777])¹⁵. In both Fragile X and BSS, the aberrant methylation is due to
145 the expansion of a CG-rich repeat that is difficult to reliably detect using short-read sequencing.
146 Rare epivariants, also called rare differentially (hyper- and hypo-) methylated regions (DMRs),
147 are enriched in individuals with neurodevelopmental disorders and congenital anomalies (ND-CA)
148 compared to controls¹⁶.

149 In contrast to rare DMRs, which represent discrete genomic regions with outlier
150 methylation changes, genome-wide epigenetic profiles identify a collection of distinct individual
151 CpG site methylation changes across the genome. A growing number of rare diseases exhibit
152 these methylation patterns, or episignatures, that are reproducible among individuals with
153 pathogenic variants within the same protein domain, gene, or protein complex, yielding highly
154 sensitive and specific biomarkers^{17,18}. Since episignatures in diagnostics were first clinically
155 validated and implemented with the EpiSign™ assay in 2019¹⁹, episignatures for nearly 70 rare
156 diseases have been published. Episignatures provide strong evidence for genetic diagnosis,
157 regardless of whether an underlying pathogenic DNA variant is identified, and to resolve variants
158 of uncertain significance (VUS). Episignatures have been found for neurodevelopmental
159 disorders where epilepsy is part of the phenotype, but the diagnostic yield for DEEs has not been
160 determined. Furthermore, how these clinically relevant episignatures might be harnessed to
161 inform underlying disease biology and give insights into potential distinct and overlapping
162 pathogenic mechanisms among disorders is just beginning to be explored²⁰.

163 Both rare DMRs and episignatures can be detected in peripheral blood samples. Rare
164 DMRs derived from individuals with ND-CA are recapitulated across multiple tissue types,
165 including blood and fibroblasts¹⁶. Episignature classifiers are trained on data obtained from blood-
166 derived DNA and are, therefore, blood-specific.

167 Here, we assessed rare outlier DMRs and DNA methylation episignatures in peripheral
168 blood-derived DNA from 516 individuals with genetically unsolved DEEs. We report our

169 methylation array data processing pipeline, MethylMiner, which automates quality control,
170 normalization, and implementation of an algorithm that mines rare DNA methylation events¹⁴ in
171 addition to interactive data visualization. Using a combination of short- and long-read sequencing
172 (LRS), we identify variants underlying rare epivariants and episignatures. Finally, we refine the
173 robust episignature for the DEE gene *CHD2* (MIM:602119)¹⁸ to explore how clinically relevant
174 episignatures may give insights into underlying biology. For individuals with unsolved DEEs, we
175 show that rare epivariants and episignatures uncover molecular causes missed using standard
176 sequence-based approaches.

177

178 **MATERIALS AND METHODS**

179 **Cohorts**

180 Our cohorts consist of 530 affected individuals (43% female) with unsolved DEEs and 478 healthy
181 controls (46% female) (Figure S1, Table S1, Supplemental Methods). An additional 146 analytical
182 controls (60% female) were included for validation. Individuals with DEEs were recruited from
183 investigators' research and clinical programs^{21,22}. Methylation array data for healthy controls were
184 drawn from a public database²³ (n=111), an internal institutional database (n=337), and 30
185 unaffected parents of probands with DEEs (Supplemental Methods). Eight family members with
186 epilepsy were studied to identify familial methylation patterns (shared rare DMRs or
187 episignatures). Analytical controls, including i) six individuals each with a disease-associated rare
188 DMR, ii) 24 individuals with a pathogenic variant in a gene or CNV associated with an
189 episignature, and iii) 116 individuals with a pathogenic variant in a gene without a known
190 episignature, were used to validate positive and negative rare DMR and episignature findings in
191 the DEE cohort. After quality control and normalization (described below), there were 516
192 remaining individuals with unsolved DEEs who had undergone extensive molecular testing: 80%

193 had a gene panel, 40% microarray analysis, 76% ES, and 38% GS. Collectively, 98% had at least
194 one sequence-based investigation (gene panel, ES, or GS). There were also 464 healthy controls,
195 141 analytical controls, and eight affected family members for DNA methylation analysis. This
196 study was approved by the Institutional Review Board (IRB) of St. Jude Children's Research
197 Hospital (SJCRH). Written informed consent was provided by parents or legal guardians of
198 individuals with DEEs with local IRB approval from SJCRH, Austin Health (Australia), the
199 University of Washington (UW), and the National Institutes of Health (NIH).

200

201 **Methylation Array**

202 All data were derived from peripheral blood-derived DNA, except for five analytical control
203 samples used for outlier DMR analysis: saliva-derived DNA from one female individual with BSS
204 and her parent and lymphoblastoid cell line (LCL)-derived DNA from three individuals, including
205 two males and one female, with Fragile X syndrome (Coriell). These samples were used as
206 positive controls to validate the outlier analysis, and then removed from the final analysis to
207 minimize potential cell type differences. DNA was extracted from peripheral blood samples using
208 standard protocols, with approximately 250-500ng of DNA bisulfite converted. The Illumina
209 Infinium MethylationEPIC v1.0 (850K array) bead chip arrays (processed according to the
210 manufacturer's protocol, Supplemental Methods) interrogate >850,000 individual CpG sites,
211 including CpG islands, promoter regions, gene bodies, FANTOM5 enhancers, and proximal
212 ENCODE regulatory elements²⁴.

213 Of 1,162 individuals included, three individuals were run in triplicate, and 29 were run in
214 duplicate across different batches to produce a total of 1,197 blood-derived DNA methylation
215 array samples before quality control and processing. Each sample consisted of data for >850,000
216 probes that were rigorously quality-controlled for the removal of outlier samples as opposed to

217 outlier regions of interest. All data were combined and loaded into the R package minfi²⁵ for quality
218 control and normalization and the R package SVA²⁶ for batch correction using the ComBat
219 method²⁷. Individual CpG probes that failed (detection $p > 0.01$) in $>10\%$ of samples and probes
220 overlapping with common SNPs were removed. Samples judged to be of poor quality ($>1\%$ of
221 probes that failed) and samples that were deemed outliers based on manual inspection of the
222 principal component analysis (PC1 and PC2), using β values for probes located on chromosome
223 (chr) 1, were removed. We estimated blood cell type composition for six cell types (CD8T, CD4T,
224 NK, B-cell, monocytes, and granulocytes) from β values for each sample²⁸. Samples containing
225 outlier cellular fractions defined as $\geq 99^{\text{th}}$ percentile $+2\%$ or $\leq 1^{\text{st}}$ percentile -2% for at least two of
226 the six cell types were also removed. Methylation array intensity values on the sex chromosomes
227 (X, Y) were used to infer the sample sex and compared to the clinically reported sex. Samples
228 with sex mismatch were removed. Samples were separated into inferred sex (males and females)
229 for all downstream analyses of sex chromosomes. This quality control and filtering left 1,161
230 samples across 1,129 individuals (26 individuals in duplicate and three individuals in triplicate
231 across batches) assayed by the 850K array and 833,834 probes (814,945 autosomal probes and
232 18,889 sex chromosome probes) (Table S1).

233

234 **Identification and Annotation of Rare Epivariants**

235 To identify outlier DMRs, we used a sliding window approach as previously described¹⁴. In brief,
236 this algorithm employs user-defined quantile thresholds to determine outlier β values across
237 multiple CpG sites. Per 1Kb window, at least three consecutive CpG sites must exhibit outlier β
238 values in the same direction (hyper or hypo) for a sample compared to the rest of the cohort to be
239 considered an outlier DMR. We considered β values above the 99.25th percentile plus 0.15 as
240 hypermethylated, and those below the 0.75th percentile minus 0.15 as hypomethylated for
241 analysis of the autosomes (chr1-chr22). Since samples were split into inferred sex (males and

242 females) for analysis of the sex chromosomes, the stringency was adjusted accordingly to 99th
243 plus 0.15 for hypermethylated and 1st percentile minus 0.15 for hypomethylated. DMRs were then
244 annotated to inform functional interpretation using HOMER²⁹ and including overlap with UCSC
245 RefSeq gene bodies and promoter regions, defined as ± 2 Kb of the transcription start sites (TSS),
246 known CpG islands (CGIs), repetitive-element information (RepeatMasker and SimpleRepeats),
247 imprinting control centers³⁰, CTCF-binding sites³¹, gene molecular function information²⁹, OMIM
248 phenotype³², average brain expression using bulk RNA-seq data from the GTEx Portal, and in-
249 house epilepsy- and candidate-gene lists to prioritize candidates but not as exclusion criteria.
250 Additionally, a recent study delineated the rare DMR landscape in the human population by
251 examining 450K methylation array data from >23,000 individuals¹⁴. Regions from those data were
252 checked against our DMRs where possible to determine the frequency at which each DMR occurs
253 in the population. Based on this annotation information, DMRs were prioritized by four features:
254 (1) a low or negligible population frequency; (2) a well-annotated genomic location, such as in or
255 near known epilepsy and candidate genes; (3) recurrence in multiple individuals; and (4) manual
256 inspection of DMRs, including flanking regions.

257

258 **Development of a DNA Methylation Array Analysis and Visualization Pipeline**

259 We developed MethylMiner, a methylation array analysis pipeline tailored toward discovering rare
260 epivariants with interactive data visualization. The pipeline requires standard input files, raw signal
261 .idat files containing each sample's green and red channels, and a metadata sheet including
262 sample names, sentrix IDs, reported sample sex, and sample group (if applicable). In brief, the
263 pipeline performs quality control and normalization as described to derive output files, including
264 quality control reports, β values, M-values, and bigWig files for quick and convenient visualization
265 in the integrative genomics viewer (IGV)³³. The pipeline then performs the outlier DMR analysis
266 (using scripts derived from: https://github.com/AndyMSSMLab/Methylation_script) based on user-

267 defined quantile thresholds and outputs the DMRs and annotations into a tabulated sheet. This
268 annotated list of DMRs is then used as input for the interactive data visualization in JupyterDash,
269 which allows users to interact with plots for quality control metrics, DMR annotations, and DMR
270 genomic tracks. The pipeline is hosted on our GitHub page ([https://github.com/stjude-](https://github.com/stjude-biohackathon/MethylMiner)
271 [biohackathon/MethylMiner](https://github.com/stjude-biohackathon/MethylMiner)).

272

273 **Validation of Outlier DMRs Using Enzymatic Methyl-Sequencing**

274 We performed targeted Enzymatic Methyl-sequencing (targeted EM-seq) enriched with the Twist
275 Human methylome panel targeting 3.98M CpGs through 123 Mb of genomic content. Targeted
276 EM-seq of peripheral blood-derived DNA was used to validate a subset of outlier DMRs, including
277 n=2 positive control DMRs (*XYLT1* and *FMR1*) and n=29 DMRs-of-interest called amongst n=3
278 individuals with unsolved DEEs. EM-seq library preparation, target enrichment, and sequencing
279 were performed using standard protocols³⁴. Reads were processed using the “nf-core/methyseq”
280 pipeline with the '--emseq' flag. For detailed EM-seq methods, please refer to Supplemental
281 Methods.

282

283 **Identification of Structural Variants with Long-Read Sequencing**

284 We used both targeted and whole-genome LRS on the Oxford Nanopore Technologies (ONT)
285 platform to validate rare DMRs and identify candidate disease-causing variants at or near the site
286 of interest (Table S2A). Targeted LRS using the “read-until” function was performed on an ONT
287 GridION using a single R9.4.1 flowcell as described previously³⁵. At least 100Kb of sequence was
288 added to either side of the target region for capture. Libraries for GS were prepared using the
289 ligation sequencing kit (SQK-LSK110) following the manufacturer’s instructions, then loaded onto
290 a single flowcell (FLO-PRO110, R9.4.1) on a PromethION and run for 72 hours with one wash

291 and reload. All data were base called using Guppy 6.3.2 (ONT) with the superior model including
292 5mC methylation. Reads were aligned to GRCh38/hg38 using minimap2³⁶, SNP and indel
293 variants were called using Clair3³⁷, structural variants were called using Sniffles³⁸, SVIM³⁹, and
294 CuteSV⁴⁰, and phasing was performed using LongPhase⁴¹. Aligned and phased bam files were
295 visualized in IGV³³.

296

297 **Episignature Testing**

298 Data were blinded and submitted to the clinical bioinformatics laboratory [Molecular Diagnostics
299 Laboratory, London Health Sciences Centre (LHSC), Western University, London, Canada]
300 through a secure file transfer protocol and stored on encrypted servers. DNA methylation data for
301 each sample were compared to clinically validated DNA methylation episignatures for all disorders
302 which are part of the EpiSignTM v4 clinical test⁴². The reference database EpiSignTM Knowledge
303 Database (EKD) includes thousands of clinical, peripheral blood DNA methylation profiles from
304 disorder-specific reference and normal controls (general population samples of various ages and
305 racial backgrounds). Individual DNA methylation data for each individual were compared with the
306 EKD using the support vector machine (SVM) based classification algorithm for EpiSignTM
307 disorders. A Methylation Variant Pathogenicity (MVP) score between 0 and 1 was generated to
308 represent the confidence of prediction for the specific disorder the SVM was trained to detect.
309 Conversion of SVM decision values to these scores was carried out according to the Platt scaling
310 method⁴³.

311 Classification for a specific EpiSignTM disorder included a combination of MVP score,
312 hierarchical clustering, multidimensional scaling (MDS) of an individual's methylation data relative
313 to the disorder-specific EpiSignTM probe sets and controls. MVP score assessment had a scale
314 with thresholds of >0.5 for positive, <0.1 negative, 0.1–0.5 inconclusive or moderate confidence.

315 A detailed description of this analytics protocol was described previously^{18,44}. Possible types of
316 results included: positive (matching an EpiSign™ disorder), negative (not matching any EpiSign™
317 disorder), and inconclusive (described in detail in results).

318

319 **Exome and Genome Sequencing**

320 If sequencing data were already available for the individual on a collaborative research basis,
321 these data were reviewed. If the data were unavailable, ES or GS was performed on peripheral
322 blood-derived DNA using standard Illumina short-read sequencing techniques and bioinformatic
323 approaches (Supplemental Methods). We validated potentially pathogenic variants with Sanger
324 sequencing and confirmed sample identity and relatedness (e.g. trios) using Powerplex Short-
325 Tandem Repeat (STR) Identification analysis.

326

327 **RNA-sequencing and Gene Expression Analysis**

328 RNA was extracted from dermal fibroblasts established from skin punch biopsies for Family 2
329 (n=2) and Family 3 (n=3) described in the results. RNA-seq was performed using standard
330 Illumina short-read sequencing practices (Supplemental Methods), and the reads were processed
331 using the “nfcorn/rnaseq” pipeline. Removal of the adapter sequences was performed using Trim
332 Galore!, and low-quality reads were eliminated with FastQC⁴⁵. Subsequently, reads were aligned
333 to a reference genome using the STAR aligner⁴⁶. Gene expression quantification was performed
334 using Salmon⁴⁷, which estimates transcript abundance. To determine gene “dropout,” the
335 OUTRIDER algorithm⁴⁸ was applied to RNA-seq data for Family 2 (proband and parent), Family
336 3 (proband and parent 2), and Family 3 (parent 1 and parent 2) against a publicly available dataset
337 of n=139 fibroblast samples⁴⁹. PCA displayed no batch groupings, and genes with Fragments Per

338 Kilobase of transcript per Million mapped reads (FPKM) <1 were removed as lowly expressed
339 genes. Results were considered significant if they had a $padj < 0.05$ and a z-score cutoff of ± 2 .

340

341 **Refinement of a CHD2 Episignature**

342 A total of 17 females and 12 males with genetic variants in *CHD2* and clinical features consistent
343 with *CHD2*-epileptic encephalopathy of childhood (EEOC) were included in this expanded 850K
344 cohort. The detailed list of genetic variants classified as pathogenic or likely pathogenic according
345 to the American College of Medical Genetics guidelines is in Table S2B. All samples and records
346 were deidentified.

347 Details of the methylation data analysis and episignature refinement are as previously
348 described^{18,50-52}. Briefly, methylation signal intensities were imported into R 4.1.3 for analysis.
349 Normalization was performed by the Illumina normalization method with background correction
350 using minfi²⁵. Probes located on X and Y chromosomes, known SNPs, or probes that cross-react
351 (as reported by Illumina) were excluded. Samples containing failed probes of more than 5%
352 ($p > 0.1$, calculated by the minfi package) were also removed. The genome-wide methylation
353 density of all samples was examined, and principal component analysis (PCA) was performed to
354 visualize the overall data structure of the batches and to identify outlier samples. All 29 samples
355 passed and were used for probe selection. The MatchIt package was used to randomly select
356 controls, which were matched for age, sex, and array type from the EKD at the LHSC, as
357 previously described^{18,53}. The methylation level of each probe was calculated as the ratio of
358 methylated signal intensity over the sum of methylated and unmethylated signal intensities (β -
359 values), ranging between 0 (completely unmethylated) and 1 (fully methylated). β -values were
360 then converted to M-values by logit transformation using the formula $\log_2(\beta/(1-\beta))$ to perform linear
361 regression modeling, which was used to identify the differentially methylated probes (DMPs), via
362 the R package limma⁵⁴. The analysis was also adjusted for blood cell-type compositions, using

363 the Houseman algorithm⁵⁵. The estimated blood cell proportions were added to the model matrix
364 of the linear models as confounding variables. The generated p -values were moderated using the
365 eBayes function in the limma package and were corrected for multiple testing using the Benjamini
366 and Hochberg (BH) method.

367 Following this, probe selection was performed in three steps. Firstly, 1000 probes were
368 selected, which had the highest product of methylation difference means between case and
369 control samples and the negative of the logarithm of multiple-testing corrected p values derived
370 from the linear modeling. Secondly, a receiver's operating characteristic (ROC) curve analysis
371 was performed, and 200 probes with the highest area under the ROC curve (AUC) were retained.
372 Lastly, probes having pair-wise Pearson's correlation coefficient greater than 0.85 within case
373 and control samples separately were removed (none of the selected 200 probes met this criteria).
374 This resulted in the identification of 200 DMPs. These probes were used for the construction of a
375 hierarchical clustering model using Ward's method on Euclidean distance, as well as a MDS
376 model by scaling of the pairwise Euclidean distances between samples.

377

378 **Functional Annotation and Correlation of the *CHD2* Episignature**

379 Functional annotation and episignature cohort comparisons were performed according to our
380 published methods⁴⁴. Briefly, to assess the percentage of DMPs shared between the *CHD2*
381 episignature and other neurodevelopmental conditions on the EpiSign™ clinical classifier,
382 heatmaps and circos plots were produced. Heatmaps were plotted using the R package
383 pheatmap (version 1.0.12) and circos plots using the R package circlize (version 0.4.15)⁵⁶. To
384 determine the genomic location of the DMPs, probes were annotated in relation to CGIs and
385 genes using the R package annotatr⁵⁷ with AnnotationHub and annotations hg19_cpgs,
386 hg19_basicgenes, hg19_genes_intergenic, and hg19_genes_intronexonboundaries. CGI
387 annotations included CGI shores from 0–2Kb on either side of CGIs, CGI shelves from 2–4Kb on
388 either side of CGIs, and inter-CGI regions encompassing all remaining regions. A chi-squared

389 goodness of fit test was performed in R to investigate the significance between background DMP
390 annotation distribution and the CHD2 cohort annotation distribution. *P* values were obtained for
391 both annotation categories (gene and CGIs). To assess the relationship between the expanded
392 850K only CHD2 cohort and other EpiSign™ disorders, the distance and similarities between
393 cohorts were analyzed using clustering methods and visualized on a tree and leaf plot. This
394 assessed the top 500 DMPs for each cohort, ranked by *p*-value. For cohorts with less than 500
395 DMPs, all DMPs were used. Tree and leaf plots, generated using the R package TreeAndLeaf⁶⁸,
396 illustrated additional information, including global mean methylation difference and total number
397 of DMPs identified for each cohort.

398

399 **Whole-Genome Bisulfite Sequencing**

400 Genomic peripheral blood-derived DNA from n=3 CHD2 trios (proband and parents) and n=1
401 CHD2 singleton (proband) (total n=10 samples) were bisulfite-converted and then underwent
402 whole-genome bisulfite sequencing (WGBS) using standard Illumina short-read sequencing
403 processing methods (Supplemental Methods). Reads were trimmed by Trim Galore! and aligned
404 to the GRCh38/hg38 human genome reference using BSMAP2.74. The methylation ratios from
405 BSAMP mapping results were extracted using methratio.py. Duplicated reads were removed and
406 CpG methylation from both strands was combined. The methylation ratios were also corrected
407 according to the C/T SNP information estimated by the G/A counts on reverse strand.

408

409 **DMR Calling of DNA Methylation Array and WGBS**

410 We performed DMR analysis on Illumina 850K EPIC methylation array data for 16 individuals with
411 DEEs harboring pathogenic variants in *CHD2* compared to 18 controls. The data were normalized
412 using the minfi package's functional normalization algorithm⁵⁹, and we employed two independent
413 R packages to call DMRs, bumhunter⁶⁰ and DMRcate⁶¹. DMRs were defined as those passing

414 a significance threshold of $p < 0.05$ for bumhunter and Fisher's multiple comparison $P < 0.05$ for
415 DMRcate. A minimum of three CpGs and mean methylation difference between CHD2 and
416 controls of at least 5% was also required (bumhunter "cutoff" and DMRcate "betacutoff" = 0.05)
417 in either the hyper or hypo direction. For bumhunter, smoothing was used, and the number of
418 permutations for each condition was set to $B = 1000$. For DMRcate, default settings were used,
419 and the Gaussian kernel bandwidth for smoothed-function estimation was set to $\lambda = 1000$, meaning
420 that significant CpGs further than 1000 nucleotides were in separate DMRs.

421 The methylCall data from WGBS, which consists of the total number of reads covered for
422 each CpG site and the number of methylated C's at each CpG site, was used for calling DMRs
423 between four individuals with DEEs caused by pathogenic *CHD2* variants and six unaffected
424 parents. Firstly, CpG sites with less than 10X coverage and those on the sex chromosomes were
425 removed. DMRs were called from WGBS methylCall data using two independent R packages,
426 DMRcate⁶² and DSS⁶³. DMRcate identifies and ranks the most differentially methylated regions
427 across the genome, while DSS detects differentially methylated loci or regions from WGBS. For
428 DMRcate, the scaling factor for bandwidth "C" was set to 50, as recommended for WGBS. DSS
429 was run with default parameters. DMRs were defined by each algorithm (with smoothing) as
430 regions of a minimum of five CpGs with significance (Fisher's multiple comparison P value < 0.05)
431 and minimum methylation differences of 5% in either the hyper or hypo direction (DSS "delta" and
432 DMRcate "betacutoff" = 0.05) between cases and controls.

433 The genomic locations of output DMR calls were intersected between both callers
434 requiring a minimum overlap of 50bp in the same direction to reduce the false positive rate. This
435 resulted in high-confidence list of DMRs predicted by two independent callers for array
436 (bumhunter and DMRcate) and WGBS (DMRcate and DSS). The methylation difference
437 between CHD2 and control was averaged between both callers for the final DMR list. DMRs were
438 segmented by mean methylation difference between CHD2 and control (5%, 10%, 15%, and 20%)
439 for visualization and annotation with CpG elements (islands, shores, shelves) and gene regions

440 (1-5Kb upstream TSS, promoters as <1Kb upstream TSS, 5'UTRs, exons, introns, and 3'UTRs)
441 using annotator⁵⁷. To get adequate CpG element counting (i.e. a DMR spanning both a shore and
442 shelf would not get counted twice), CpG annotations were adjusted for DMR size by calculating
443 representation across CpG elements as a fraction of the total DMR length.

444

445 **RESULTS**

446 **Discovery and Validation of DMRs in Unsolved DEEs**

447 To determine the ability of our analysis pipeline to robustly detect rare, outlier DMRs, we included
448 DNA from six positive controls with genetic diseases: three individuals with heterozygous or
449 homozygous hypermethylation of *XYLT1*, and three individuals (two males and one female) with
450 hypermethylation of *FMR1*. The outlier DMR analysis detected both rare DMRs (Figure S2, Table
451 S3A). Additionally, we identified an *XYLT1* heterozygous hypermethylation carrier in our DEE
452 cohort. Targeted X-chromosome analysis in males identified complete methylation at the *FMR1*
453 locus in both Fragile X males compared to the remaining cohort, all of which were completely
454 unmethylated at *FMR1*. *FMR1* hypermethylation was also higher (~75%) in the Fragile X female
455 sample compared to the other females with 25-50% methylation, likely due to random X-
456 inactivation. Thus, our methylation array analysis approach detects outlier DMRs at known
457 disease loci for the autosomes and sex chromosomes.

458 Next, we assessed outlier DMRs in our cohort of 1,124 individuals (516 unsolved DEE)
459 across 1,156 array samples. We predicted n=2,140 total DMRs for the autosomes, n=59 DMRs
460 for males on chromosome X, n=42 DMRs for females on chromosome X, and no DMRs on
461 chromosome Y (Table S3B, S3D, S3F). After accounting for DMRs overlapping across samples
462 ($\geq 50\%$ probe overlap in the same direction of DNA methylation hyper- or hypo-methylation), we
463 derived n=1,540 unique DMRs for the autosomes (917 hyper, 623 hypo), n=45 for males on chrX
464 (26 hyper, 19 hypo), and n=34 for females on chrX (20 hyper, 14 hypo) (Table S3C, S3E, S3G).

465 Of the samples with one or more outlier DMRs, the majority had only a single outlier DMR (Figure
466 S3).

467 To determine the robustness of our DMR calling algorithm, we (i) assessed the
468 reproducibility of DMR calls in a subset of samples and (ii) performed validation of DMRs using
469 targeted EM-seq (Supplemental Methods). Using replicate array data for 29 individuals, we found
470 that 80% of DMRs were replicated across different batches for an individual (Supplemental
471 Methods). We then used targeted EM-seq, a non-bisulfite approach, to validate a subset of DMRs.
472 We confirmed that our positive control DMRs (*XYLT1* and *FMR1*) could be detected in the
473 targeted EM-seq data (Figure S4). We then validated 30 outlier DMRs by targeted EM-seq in four
474 individuals with unsolved DEEs (Figure S5 and S6). In addition to DMR validation, targeted EM-
475 seq provides much higher resolution of the extent of differential methylation than the methylation
476 array (e.g. >100 methylated CpG sites for the *XYLT1* DMR by targeted EM-seq compared to eight
477 representative probes on array; Table S4). Thus, we detected and validated outlier DMRs at
478 higher resolution using an orthogonal approach.

479

480 **Rare Outlier DMRs in Unsolved DEEs**

481 We narrowed down outlier DMR calls for individuals with unsolved DEEs to determine high-priority
482 candidates for further study based on DMR recurrence across multiple individuals, population
483 frequency¹⁴, functional annotations (Methods), and manual inspection of DMR plots for each
484 DMR. We identified 11 individuals with unsolved DEEs with one or more rare, potentially disease-
485 associated DMRs and performed follow-up studies (Table 1, Table S2A). One individual had
486 multiple DMRs due to a balanced translocation between chrX and chr13, four individuals had a
487 DMR due to expanded CG-rich repeats, and six individuals had DMRs due to underlying CNVs.

488

489 **Rare Outlier DMR Analysis Detects Hypermethylation of chr13 Due to X;13 Translocation**

490 One female with the DEE syndrome, epilepsy of infancy with migrating focal seizures (EIMFS),
491 had 27 rare outlier hypermethylated DMRs across chr13 (Figure 1A, Figure S7), none of which
492 were present in >23,000 controls¹⁴. The DMRs were replicated on a second, independent
493 methylation array from the same individual and validated using targeted EM-seq (Figure S6).
494 Methylation array analysis of both parents revealed that all rare hypermethylated DMRs occurred
495 *de novo* in the proband (Figure 1B). Whole-genome ONT long-read sequencing also confirmed
496 the hypermethylated DMRs and identified a balanced translocation between chrX and chr13
497 (Figure 1C), annotated as 46,XX,t(X;13)(q28;q14.2). The translocation provides a mechanism
498 whereby random X-inactivation induces hypermethylation on the portion of chr13q attached to the
499 large piece of the X chromosome. The translocation breakpoints were confirmed by PCR and
500 Sanger sequencing of peripheral blood-derived DNA as chrX:152,092,342 to chr13:47,005,269
501 and chr13:47,005,271 to chrX:152,092,344 (GRCh38/hg38). Parental methylation studies and
502 short-read GS confirmed that the translocation occurred *de novo*. The translocation is likely
503 causative in this individual given the *de novo* occurrence, absence of clearly pathogenic sequence
504 variants by trio sequence (Table S5), and report of a similar translocation in a female individual
505 with intellectual disability and bilateral retinoblastoma⁶⁴.

506

507 **Rare Outlier DMR Analysis Detects Hypermethylation Caused by Underlying Triplet Repeat** 508 **Expansions**

509 We detected two individuals with unsolved DEEs and one control individual with hypermethylation
510 spanning the 5'UTR and intron 1 of the Casein kinase 1 isoform epsilon (*CSNK1E*, MIM:600863,
511 Figure 2A) gene. Although present in one control and reported in 6/23,116 controls¹⁴, an individual
512 with DEE and probable haploinsufficiency due to a *de novo* splicing variant (c.885+1G>A) in

513 *CSNK1E* has been reported⁶⁵, suggesting further study is warranted to determine if variation in
514 this gene causes DEE. Segregation analysis revealed that the hypermethylation in one proband
515 was inherited (Family 1, Figure S8), whereas the other arose *de novo* (Family 2). After validation
516 of hypermethylation with targeted EM-seq for both probands (Figure S5), long-read sequencing
517 of the proband (whole-genome) and parent (targeted) from Family 1 confirmed the presence of
518 an expanded CGG motif in both (Figure 2B), as previously reported in individuals with
519 hypermethylation of *CSNK1E* at fragile site FRA22A and reduced expression in lymphoblastoid
520 cells¹⁴. Through GeneMatcher⁶⁶, we identified Family 3 consisting of a proband with the same
521 *CSNK1E* hypermethylated DMR inherited from his parent, who is mildly affected by learning,
522 speech, and sleep difficulties (Supplemental Phenotype data). Expression analysis in available
523 fibroblasts from Families 2 and 3 showed that individuals with *CSNK1E* hypermethylation had
524 decreased expression of *CSNK1E* compared to hypermethylation-negative controls (Figure 2C).
525 Analysis using the OUTRIDER algorithm⁴⁸ confirmed “drop-out” of *CSNK1E*
526 (ENSG00000213923) expression compared to publicly available fibroblast controls⁴⁹ (Figure 2C,
527 Figure S9). Thus, we report 3 individuals with unsolved DEEs harboring inherited and *de novo*
528 *CSNK1E* hypermethylation due to an underlying repeat expansion (n=4 LRS) that leads to
529 approximately 50% reduction in *CSNK1E* expression (n=3 RNA-seq drop-out). No other
530 candidate gene variants for these 3 probands were found by trio GS analysis. However, due to
531 finding this abnormality in seemingly unaffected individuals, one control and one parent (Family
532 1) in our cohort and others¹⁴, further work is required to determine whether variations in *CSNK1E*
533 cause or contribute to the DEEs.

534 A male individual with unsolved DEE displayed inherited hypermethylation of the *DIP2B*
535 (MIM:611379) promoter region and exon 1 (Figure S10), which is due to an underlying CGG-
536 repeat expansion and fragile site FRA12A⁶⁷. Loss of *DIP2B* is associated with an autosomal

537 dominant neurodevelopmental disorder (NDD) with variable penetrance, including a *DIP2B* repeat
538 expansion in an individual with epilepsy⁶⁷.

539 We detected a rare hypermethylated DMR on the X chromosome in exon 1 of an
540 uncharacterized gene (*BCLAF3/CXorf23*) in a male with unsolved DEE (Figure S10), that was
541 absent in >23,000 unaffected controls (>8,000 males)¹⁴. We validated hypermethylation using
542 targeted EM-seq (Figure S5), and ONT long-read sequencing of the proband and his parent
543 revealed a novel CGG repeat expansion in the proband (~2,500-3,000bp, Figure S11) inherited
544 from his parent, who had a smaller expansion (~1,700-1,900bp). LRS and standard X-inactivation
545 studies⁶⁸ show that the parent has skewed X-inactivation (Table S6) of the allele with the
546 expansion, which explains why outlier hypermethylation is not detected from her methylation array
547 data. There are no other candidate variants for the proband's DEE by trio GS. Collectively, these
548 results highlight the detection of repeat-expansion-associated loci based on outlier DMR analysis
549 of DNA methylation array in individuals with unsolved DEEs.

550

551 **Rare Outlier DMR Analysis Detects Copy Number Variants**

552 Six individuals displayed DMRs that were found to be due to underlying CNVs. One control and
553 an individual with unsolved DEE had ~10-15 hypomethylated DMRs along chr2 spanning ≥144Kb
554 (Figure S12A). Short and long-read sequencing analysis revealed this “DMR” was due to a
555 homozygous ~182Kb deletion encompassing outlier DMRs (Figure S12C). Segregation testing
556 found that the proband inherited the deletion from both parents, who were heterozygous carriers.
557 The CNV was also found on DNA methylation array using the R tool *conumee*⁶⁹ (Figure S12B).

558 Four individuals with unsolved DEEs and one control had a 686bp hypomethylated DMR
559 in intron 2 of the gene *LINGO1* (MIM:609791). DNA methylation array analysis for a proband's
560 parent found that the hypomethylation was at least in part inherited, and short and long-read

561 sequencing revealed that hypomethylation was caused by an underlying ~4Kb inherited deletion
562 (Figure S13).

563 Another individual with unsolved DEE had hypermethylation in the 5'UTR of
564 *CFAP36/CCDC104* (Figure S14A), which was not present in >23,000 controls¹⁴. DNA methylation
565 array analysis of both parents indicated it was inherited (Figure S14B), and targeted ONT long-
566 read sequencing revealed a ~500Kb tandem duplication from chr2:55,034,228-55,536,971
567 (GRCh38/hg38). Collectively, these results indicate that outlier DNA methylation can be due to
568 underlying CNVs and that the 850K methylation array may not have sufficient coverage to detect
569 smaller CNVs. Due to the high population frequencies and inheritance status of the CNVs we
570 found, we determined they are unlikely to contribute to the individuals' phenotypes. Still, these
571 findings illuminate ways in which detected DNA methylation changes are influenced by underlying
572 DNA variation.

573

574 **Episignature Screening Validates Pathogenicity of Genetic Diagnoses and Resolves** 575 **Variants of Uncertain Significance**

576 We next performed episignature analysis, using the EpiSignTM v4 classifier, including 70
577 conditions associated with 96 genes/genomic regions (Figure 3). To validate our approach, we
578 included several individuals with causal variants in episignature genes or CNVs and an individual
579 with a VUS. These included sixteen individuals with variants in *CHD2* (n=15 pathogenic, n=1
580 VUS) and one individual each with a pathogenic variant in *KDM5C*, *SETD1B*, *KMT2A*, or
581 *SMARCA2* (Table S2B). We also included two individuals with CNVs, including chr17p11.2
582 deletion and duplication. Fifteen of the individuals with variants in *CHD2* were positive for the
583 epileptic encephalopathy of childhood (EEOC) episignature¹⁸, also known as the developmental
584 and epileptic encephalopathy 94 (DEE94) episignature. However, one individual with a VUS in

585 *CHD2* was negative for the epesignature, and in combination with other clinical evidence the VUS
586 was reclassified as likely benign (Supplemental Phenotype data). The individuals with variants in
587 *KDM5C* (MIM:314690), *SETD1B* (MIM:611055), *KMT2A* (MIM:159555), and *SMARCA2*
588 (MIM:600014) were all positive for the epesignatures associated with their disorders. While these
589 individuals were considered solved before epesignature screening, the finding was used to support
590 the genetic diagnosis of the individual with a *KDM5C* variant.

591 Additionally, we identified two individuals with inconclusive results for epesignatures
592 despite definitive genetic and clinical findings for the associated syndromes. Inconclusive findings
593 are caused by methylation profiles that partially overlap existing signatures but are not a definitive
594 match. This included an individual with a 17p11.2 deletion inconclusive for the Smith-Magenis
595 syndrome epesignature (SMS_del) and a female individual with a 17p11.2 duplication inconclusive
596 for the Potocki-Lupski syndrome epesignature (PTLS, Figure S15). In each case, the inconclusive
597 epesignature finding is concordant with the genetic diagnosis but yields an inconclusive result
598 potentially attributable to variability introduced by differential CNV breakpoints. Because of this
599 and other factors, inconclusive EpiSign™ results are reported with the caveat that further follow-
600 up or investigation may be warranted if there is a clinical phenotype consistent with the
601 inconclusive epesignature in question.

602

603 **Epesignature Screening Solves Genetically Unsolved DEEs**

604 We then tested our cohort of 516 individuals with unsolved DEEs for 70 clinically validated
605 epesignatures, leading to a likely diagnosis in five individuals (Table 2). All methylation variant
606 pathogenicity (MVP) scores for epesignatures and detailed genomic variant information are in
607 Table S2C. Two unrelated individuals with unsolved DEEs were positive for the KGB syndrome
608 epesignature (KGBS_MRD23) caused by pathogenic variants in *ANKRD11* (Figures S16 and

609 S17). Exome or genome sequencing analysis revealed *de novo* pathogenic stop-gain variants in
610 both individuals, and phenotypes for each individual are consistent with the diagnosis
611 (Supplemental Phenotype data). One proband had affected siblings and family members (n=8,
612 Figure S17). However, none harbored the *ANKRD11* epismature and neither affected sibling
613 harbored the variant, indicating that there is likely a different explanation for this familial epilepsy.
614 One individual with unsolved DEE was positive for the epismature associated with *SETD1B*
615 (Figure S18). Exome sequencing revealed a pathogenic stop-gain variant in *SETD1B*. Another
616 individual with unsolved DEE harbored the epismature for *TET3* and had an inherited pathogenic
617 stop-gain variant in *TET3* on GS (Figure S19). This remains the likely cause of the individual's
618 DEE as the parent has a milder phenotype including macrocephaly and learning difficulties
619 (Supplemental Phenotype data). One male individual with unsolved DEE was positive for the
620 *UBE2A* epismature (Figure S20). Through exome sequencing, we identified a predicted
621 damaging inherited missense variant absent in gnomAD (c.376G>A, p.Ala126Thr). Although the
622 variant does not reach likely pathogenic classification using existing ACMG criteria, the prediction
623 scores (REVEL=.776, CADD=26.4, and PolyPhen-2=1.00) support pathogenicity; the variant is
624 inherited in an X-linked intellectual disability disorder; and the individual shares multiple
625 phenotypic features with *UBE2A* disorder. Thus, the variant has been determined to be the most
626 likely genetic cause of disease.

627 Of the high-confidence epismature findings, only one individual had an established
628 genetic diagnosis in another gene. This individual harbored a *de novo* variant in *PTEN* with a
629 consistent phenotype of macrocephaly and focal epilepsy but also had the epismature for
630 *KDM2B*. Further analysis identified an inherited missense variant in *KDM2B*. We performed
631 methylation array analysis for the unaffected parent who also harbored the *KDM2B* epismature.
632 This variant is predicted to be likely pathogenic (LP) by ACMG criteria due to its putative effect on
633 splicing regulation, though assessment of this variant with SpliceAI predicts that it does not have

634 a high likelihood of affecting splicing (Δ score for Donor Gain:0.01). When this criterion is taken
635 away, the designation of LP is reduced to a VUS; other computational predictors assess the
636 impact to be uncertain (REVEL=0.517). Thus, while it is unlikely that this *KDM2B* variant explains
637 the individual's phenotype, it still represents an underlying DNA change detected through
638 episignature screening. Collectively, we have identified positive episignatures and causal genetic
639 etiologies in five previously unsolved individuals with DEEs through episignature screening.

640 An additional 40 individuals with DEEs (80% unsolved) and nine controls had inconclusive
641 results for episignatures (Figure S20). Of the individuals with DEEs, 4/40 were run across multiple
642 methylation array batches. Three individuals did not reproduce their inconclusive episignature
643 result in the other sample(s). While one individual's inconclusive result did replicate across the
644 different batches, no pathogenic variants were found by GS in the associated genes(s). Of all the
645 individuals with available sequencing data (n=27), none harbored pathogenic variants in the
646 genes associated with episignature findings. While some had overlapping clinical features, most
647 were discordant with the described phenotypes for their inconclusive episignature finding.
648 Additional follow-up will be required to determine whether these inconclusive results are due to
649 array artifacts or have underlying biological or disease-associated meaning. If technical artifacts
650 are ruled out, an inconclusive result may be caused by episignatures in other genes that are yet
651 to be defined and trained against for specificity of the classifier.

652

653 **Redefining the CHD2 Episignature on the 850K EPIC Array**

654 While episignatures are proven to be clinically useful for diagnosis, little work has been done to
655 investigate how episignatures may inform disease biology by studying DMRs that may impact
656 gene expression. Here, we performed refinement and in-depth analysis of the episignature for the
657 DEE gene *CHD2*. The *CHD2* episignature was originally derived using overlapping 450K and

658 850K DNA methylation array probes representing individual CpG sites in n=9 individuals with
659 pathogenic *CHD2* variants¹⁸. We refer to this signature as the CHD2 450K episignature (Figure
660 4A upper, Figure S21A, Table S7). Here, we refine the CHD2 episignature exclusively on 850K
661 EPIC methylation array probes with data from a cohort of n=29 individuals with pathogenic *CHD2*
662 variants (Figure 4A lower, Figure S21B, Table S7). We refer to this signature as the CHD2 850K
663 episignature. Of the 200 probes included in the CHD2 850K episignature, 79/200 are specific to
664 the 850K EPIC array.

665

666 **Comparison of the CHD2 Episignature to 55 Other Clinically Validated Episignatures**

667 We then compared the CHD2 450K and 850K episignatures to 55 other NDD episignatures (57
668 total including CHD2)²⁰ by examining shared probes (Figure 4B, Figure S22), Euclidean clustering
669 (Figure 4C), probe mean methylation differences (Figure S23), and functional annotations (Figure
670 S24). As expected, the CHD2 850K episignature shares the most probe overlap with the CHD2
671 450K episignature (86/200 or 43%, Figure 4B, Figure S22). Euclidean clustering was used to
672 examine the relatedness of the episignatures by probe overlap and directionality. The CHD2 850K
673 episignature shares the closest branchpoint with the MRXSCJ episignature for *KDM5C* of which
674 it shares 7% of its top 500 DMPs. Collectively, both 450K and 850K episignatures do not share
675 immediate branches (other than the primary branchpoint) with many other episignatures. This
676 may indicate different sets of predominant pathways underlying *CHD2* pathophysiology compared
677 to the other episignatures. Additionally, the CHD2 850K episignature represents more
678 hypermethylated regions than the CHD2 450K episignature, as depicted by the mean methylation
679 differences in Figure 4C and Figure S23. We also performed functional annotation of episignature
680 probes for CpG characteristics and gene regions in relation to the 55 other NDD episignatures
681 (Figure S24). We found that both CHD2 850K and 450K DMPs map to predominately the coding
682 regions of genes (46% and 41%, respectively) with a significant difference in the distribution of

683 DMPs in these regions compared with the background probe distribution ($P < 9.06 \times 10^{-69}$ and
684 $P < 2.02 \times 10^{-79}$, respectively). Though the CHD2 850K episignature represents a higher portion of
685 interCGI regions compared with the 450K episignature (43% vs. 31%, respectively), both are
686 enriched in interCGI regions relative to background probe distribution ($P < 2.26 \times 10^{-121}$ and
687 $P < 9.17 \times 10^{-144}$).

688

689 **The CHD2 Episignature is Associated with Differentially Methylated Regions**

690 Since *CHD2* encodes a chromatin remodeler that has been shown to regulate gene
691 expression^{70,71}, we investigated whether individual episignature probes are contained within larger
692 DMRs between cases and controls. DMRs could potentially provide a link to downstream gene
693 expression. We first investigated DMRs in an unbiased genome-wide manner by calling DMRs
694 from the 850K DNA methylation array data (n=16 CHD2, n=18 controls) using bumhunter and
695 DMRcate. We predicted 1,684 DMRs from bumhunter and 963 DMRs from DMRcate. These
696 DMRs were intersected, requiring an overlap in the same direction (hyper/hypo) of at least 50bp,
697 to derive a high-confidence DMR list of 712 overlapping regions (349 hyper, 363 hypo).
698 Representative images of these DMRs are shown in Figure S25. These DMRs directly coincide
699 with 86/200 (43%) CHD2 450K episignature probes and an increased 90/200 (45%) CHD2 850K
700 episignature probes (Figure S26, Table S8). Thus, the CHD2 episignature is characterized by
701 DMRs, and this overlap increases by four probes for the CHD2 850K episignature.

702

703 **Increased CpG Resolution and Genomic Coverage of Differentially Methylated Regions** 704 **Using Whole Genome-Bisulfite Sequencing**

705 Due to limited genomic coverage, DNA methylation arrays can be skewed in their representation
706 of CpGs across the genome, as evidenced by their tendency to bias gene set analyses⁷². To
707 better understand the DMR landscape of *CHD2* and investigate DMRs at higher CpG resolution,

708 we performed WGBS with coverage of >20,000,000 CpGs on three CHD2 trios and one singleton.
709 We derived 11,019 DMRs from DSS, 4,078 DMRs from DMRcate, and 3,665 DMRs that overlap
710 between both callers (2420 hyper, 1235 hypo). To determine the robustness of this approach, we
711 manually inspected DMRs with a methylation difference of at least 20% (n=207 DMRs, 146 hyper,
712 61 hypo) by examining the reads in all 3 trios in IGV and confirmed 169/207 DMRs, yielding a
713 true call rate of 81.6%. Representative DMRs called from WGBS are shown in Figure S27. We
714 then investigated the overlap of episignature probes with the WGBS DMRs with a methylation
715 difference of at least 5% and found direct overlap with 76/200 (38%) CHD2 450K episignature
716 probes and an increased 94/200 (47%) CHD2 850K episignature probes (Figure 4D, Figure S26).
717 Thus, considering the increased genomic coverage afforded by WGBS and increased DMRs, it
718 is unsurprising that a higher proportion of CHD2 850K episignature probes overlap with DMRs
719 (Figure S26, Table S8). Notably, for nearly all probes found within DMRs, those DMRs could be
720 better visualized from the WGBS data due to the lack of probe coverage on the array. Thus, we
721 have confirmed using an orthogonal approach with higher CpG coverage that the CHD2
722 episignature is characterized by DMRs.

723 We further investigated DMR calls by functionally annotating them using the annotatr. We
724 first examined the representation of CpG islands, CpG shores, CpG shelves, and interCpG Island
725 (interCGI) regions for DMRs (Figure S28). We find that most DMRs called exclusively from WGBS
726 are located at interCGI regions compared to DMRs called from the array or overlap of both, likely
727 due to the bias of gene-enriched regions on the array compared with increased genomic coverage
728 of WGBS. We also annotated DMRs with gene annotations (Figure S29) and found similar
729 patterns across DMRs called by the 850K array, WGBS, or both, especially for DMRs called with
730 a methylation difference of at least 5% between CHD2 and controls. Notably, we show how the
731 global CHD2 episignature is characterized by DMRs (Figure S30) that correspond to gene
732 regulatory regions and therefore, likely affect underlying disease biology.

733

734 **DISCUSSION**

735 A major challenge in rare disease genetics is determining molecular causes in unsolved cases.
736 Even if ES or comprehensive GS of trios identifies all *de novo* and recessively inherited coding
737 and noncoding variants, prioritizing and functionally interpreting candidate variants is challenging.
738 In the case of the DEEs, this difficulty is further compounded by immense phenotypic and genetic
739 heterogeneity. Genome-wide DNA methylation analysis represents an innovative approach to
740 discovering genetic etiologies by investigating rare DMRs and screening for DNA methylation
741 signatures. Notably, rare DMRs and episignatures can be assessed with cost-effective, high-
742 throughput DNA methylation arrays using blood-derived DNA. Here, we performed genome-wide
743 DNA methylation analysis on 516 individuals with unsolved DEEs and identified causal or
744 candidate etiologies in 10 individuals: five from rare DMR analysis (Table 1) and five from
745 episignature screening (Table 2). Thus, the diagnostic yield of genome-wide methylation analysis
746 in individuals with unsolved DEEs is nearly 2%, similar to the added diagnostic yield of GS after
747 ES or gene panel^{73,74}. A study of unsolved ND-CA showed a similar 2-3% increase in diagnostic
748 yield using episignature analysis⁵².

749 We have performed rare outlier DMR analysis of methylation array data for a cohort of
750 individuals with unsolved DEEs and uncovered various underlying DNA variants using ONT long-
751 read sequencing. These include a X;13 translocation, CGG repeat expansions, and copy number
752 variants. We first validated a subset of outlier DMRs using targeted EM-seq enriched for 3.98M
753 CpGs, a highly effective bisulfite-free, enzyme-based conversion method for detecting CpG
754 methylation by sequencing. Targeted EM-seq has several advantages to bisulfite-based array
755 approaches, including minimizing DNA damage, lowering input requirements (picograms of DNA),
756 and detecting more CpGs³⁴. We found that all DMRs were confirmed using the EM-seq approach,
757 and the greater number of CpGs detected compared to the methylation array afforded higher

758 resolution to interpret DMRs. Future high-throughput DNA methylation analyses could consider
759 using EM-seq for validation or discovery.

760 We report an individual with 27 outlier hypermethylation events along chr13q detected
761 through the rare DMR analysis. Using ONT whole-genome long-read sequencing, we identified a
762 *de novo* X;13 translocation showing that the hypermethylation identified the likely cause of
763 disease. This discovery was enabled without the need for live cellular material, which is typically
764 required by classical cytogenetics approaches. This child passed away in infancy due to the
765 severity of the disease, and this approach provided a diagnosis postmortem using banked
766 genomic material.

767 We also found that several individuals displayed hypermethylation of loci associated with
768 known or novel CG-rich repeat expansions. These regions include the 5'UTR and intron 1 of the
769 epilepsy candidate gene *CSNK1E*, the 5'UTR of the neurodevelopmental disorder gene *DIP2B*,
770 and the 5'UTR of the uncharacterized gene *BCLAF3*. We report the occurrence of
771 hypermethylation, a CGG repeat expansion, and reduced expression of *CSNK1E* among three
772 unrelated individuals with unsolved DEEs and a mildly affected parent. *CSNK1E* has been
773 implicated in the circadian rhythm^{75,76}, and variation causes a familial advanced sleep phase
774 syndrome (FASPS)⁷⁷. Variation also produces a rapid eye movement phenotype in a knockout
775 mouse model⁷⁸. Interestingly, all our probands with DEEs and the mildly affected parent with
776 *CSNK1E* hypermethylation and a repeat expansion report sleep-related phenotypes
777 (Supplemental Phenotype data). Our results indicate that there is an enrichment of *CSNK1E*
778 hypermethylation in individuals with DEE compared to controls in our cohort combined with those
779 previously reported¹⁴ (Fisher's Exact $P=0.0185$), suggesting that further studies to determine if
780 *CSNK1E* variation contributes to DEEs are warranted.

781 One male proband with unsolved DEE displayed *de novo* outlier hypermethylation in a
782 region annotated as intergenic on the GRCh37/hg19 genome build and at the 5'UTR of *BCLAF3*

783 on the GRCh38/hg38 genome build. Using ONT long-read sequencing, we discovered a novel
784 CGG repeat expansion in exon 1 of *BCLAF3* in this proband inherited from his unaffected parent.
785 The parent's long-read data displayed skewed X-inactivation against the expanded allele, and
786 this was confirmed to be more global using an enzyme-based DNA methylation assay to profile
787 the *AR* and *HUMARA* loci (Table S6)⁶⁸. Skewed X-inactivation may explain why the parent does
788 not have a detectable DNA methylation abnormality at this locus and could provide a mechanism
789 to circumvent any functional consequences of the *BCLAF3* abnormality. While *BCLAF3* has been
790 previously predicted to be a potential disorder-associated gene on chrX⁷⁹, little is known about its
791 function or disease associations. Thus, further work is needed to investigate whether this
792 abnormality is present in other individuals and if loss of this gene on chrX in males could cause a
793 DEE.

794 We performed episignature screening of our unsolved DEE cohort using the EpiSign™ v4
795 classifier, which contains 90 episignatures representing 70 disorders encompassing 96
796 genes/genomic regions. We found six individuals with unsolved DEEs harbored positive
797 episignatures concordant with their phenotypes. We reviewed or reanalyzed available or newly
798 generated ES or GS data and identified pathogenic variants in the episignature-associated genes
799 in 5/6 individuals. In the individual with a pathogenic *SETD1B* variant, one parent was unavailable
800 for genetic testing to segregate the sequence variant. Thus, the positive episignature finding
801 provided supportive information for genetic diagnosis in lieu of inheritance data. Episignatures
802 can serve to screen for disorders that have broad, overlapping phenotypes and identify individuals
803 who may not have the classical features of a specific neurodevelopmental syndrome or DEE. For
804 instance, most DEEs have a phenotypic spectrum, so individuals with different etiology,
805 developmental trajectories, or subtle dysmorphic features may escape diagnosis until a molecular
806 etiology is found.

807 The top 27 most implicated genetic causes of DEEs explain 80% of DEEs⁷. However, only
808 1/27 genes (*CHD2*) has a clinically validated episignature. Like *CHD2*, 58/59 genes with robust
809 episignatures localize to the nucleus and are associated with DNA binding, transcriptional
810 regulation, and histone interactions. Since DNA methylation occurs in the nucleus, most genes
811 for which episignatures have been derived are directly or indirectly involved in the epigenetic and
812 transcriptional machinery. Whereas the top 27 DEE genes are associated with a range of cellular
813 processes⁵, only a minority are associated with direct DNA interactions, and only 10 of the top 27
814 most frequent DEE genes are annotated to localize to the nucleus at least partially. The only gene
815 with a clinically validated episignature not involved in any nuclear activity is *SLC32A1*, which
816 encodes solute carrier family 32 member 1 (*SLC32A1*, MIM:616440) responsible for inhibitory
817 neurotransmission, and variants in this gene cause a DEE⁸⁰. Unfortunately, *SLC32A1* is not
818 among the most common ~60 DEE genes. Therefore, the diagnostic utility of episignatures for
819 DEEs would increase when we can confidently derive episignatures for more DEE genes, such
820 as ion channel, synaptic transmission, and metabolic genes.

821 Episignature derivation is further complicated by the existence of variant-specific
822 episignatures that exist for a subgroup of variants within a gene (e.g. *SMARCA2*^{50,81}) or a set of
823 common genes within similar pathways (e.g. Coffin-Siris syndrome episignature, due to variants
824 in *ARID1A* (MIM: 603024), *ARID1B* (MIM:614556), *SMARCB1* (MIM:601607), and *SMARCA4*
825 (MIM:603254), and *SOX11* (MIM:600898)⁵⁰. Thus, there is not only a need to derive episignatures
826 for more epilepsy-related genes but also to analyze variants for testing based on variant type (i.e.
827 missense, nonsense) and protein domain, which may segregate with phenotypes. For instance,
828 our cohort included two females with solved DEEs and pathogenic truncating variants in the
829 *SMC1A* gene located on chromosome X. Neither had a positive episignature for *SMC1A* for
830 Cornelia de Lange syndrome (CdLS), which is usually due to missense or in-frame small indels
831 proposed to have a dominant negative effect. Truncating, loss-of-function variants, however, are

832 found exclusively in girls with DEEs. The difference in underlying disease mechanism likely
833 impacts the composition of the distinct probe sets contained within the episignatures. Discordant
834 or unusual findings like this example underscore additional considerations when deriving and
835 interpreting episignatures. We came across five individuals reported as male whose methylation
836 pattern on the X chromosome suggested two X chromosomes. Of 2/5 of these individuals who
837 had LRS, a genotype of XXY was confirmed, which is consistent with a diagnosis of Klinefelter
838 syndrome. More unexpected and incidental findings will arise as a greater number of
839 episignatures are derived, and methylation testing becomes more routine.

840 Episignatures for many epilepsy-related genes are currently in development. As more
841 episignatures are clinically validated, re-analysis of previously generated methylation array data
842 from unsolved individuals will identify pathogenic findings, akin to re-analysis of exome
843 sequencing data for new epilepsy genes years after initial sequencing was performed⁸². We found
844 that episignature analysis was useful for clarifying VUSs, including an individual annotated as
845 solved for *CHD2* displaying a VUS, which was re-assessed as benign based on a negative *CHD2*
846 episignature result. We anticipate that episignatures will also be useful for interpreting the impact
847 of noncoding variants.

848 There are additional considerations when determining the utility of DNA methylation
849 analysis for the molecular diagnosis of individuals with DEEs. Firstly, the diagnostic utility will vary
850 depending on when the individual receives the test relative to other genetic testing modalities. In
851 our study, we analyzed DNA from individuals with DEEs who had remained unsolved after
852 undergoing extensive genetic testing, including gene panels, microarrays, exome, and genome
853 sequencing. As DNA methylation testing becomes increasingly accessible to newly diagnosed
854 individuals with DEEs and as the number of epilepsy-relevant genes with robust episignatures
855 grows, the utility of DNA methylation analysis in unsolved DEEs may increase and guide which
856 regions should be sequenced to identify causal variants.

857 DNA methylation information can be readily assessed from both ONT long-read
858 sequencing and PacBio long-read sequencing data. Therefore, when long-read sequencing
859 becomes more available, there is potential for an “all-in-one” approach to genetic testing whereby
860 individuals can simultaneously be assessed for sequence variants, structural abnormalities, and
861 rare DNA methylation changes. While it is advantageous to study rare DMRs and their potential
862 underlying DNA defects using the same technology, applying episignatures to long-read
863 sequencing data is uncertain and may require new computational approaches to re-derive and
864 validate episignatures on each platform. As long-read sequencing produces far more data than
865 arrays (>20,000,000 CpGs versus ~850,000 CpGs), this will offer an opportunity to interrogate
866 DNA methylation more broadly and deeply.

867 As advances in sequencing technologies allow DNA methylation datasets to get larger,
868 there will be a need to analyze comparative data from controls to generate population-level
869 reference information. For our DMR analysis, we leveraged 450K DNA methylation array outlier
870 DMR calls generated from peripheral blood-derived DNA for >23,000 control individuals¹⁴. Where
871 possible, we used these data to approximate population frequencies for the DMRs we derived.
872 However, this reference information is not available for 850K exclusive DMRs or whole-genome
873 sequencing DMRs. Thus, interpreting DNA methylation data for unsolved DEEs and other
874 unsolved genetic disorders will improve as we understand more of the methylome, including
875 regions that were only recently resolved on the T2T genome build⁸³, using appropriate reference
876 datasets from diverse populations.

877 While episignatures provide a robust readout of the genetic etiology, they are composed
878 of individual array probes representing singular CpG sites that may not contribute to
879 understanding the underlying disease mechanism. Given that *CHD2* is the most frequent DEE
880 genes with a robust episignature and has a biological role as a chromatin remodeler, we were
881 interested to use the episignature to understand how DNA methylation relates to underlying *CHD2*

882 pathophysiology. First, we re-defined the episignature on exclusively 850K array probes with an
883 increased sample size from n=9 to n=29 individuals with *CHD2* pathogenic variants. Using DNA
884 methylation array and WGBS, we show that the *CHD2* episignature is associated with DMRs
885 between cases and controls. In a recent study, investigators derived DMRs for individuals with
886 pathogenic *HNRNPU* (MIM:617391) variants versus controls in methylation array data from
887 peripheral blood-derived DNA and reported 19 DMRs called with DMRcate (Fischer $P < 0.01$,
888 betacutoff=0.05, minCpG=5)⁸⁴. The comparative number of DMRs we derived for *CHD2* versus
889 control methylation array data under the same conditions using DMRcate is 474 DMRs. This
890 increased number of DMRs may represent the inherent function of *CHD2* as a chromatin
891 remodeler that interacts directly with the DNA, whereas *HNRNPU* forms complexes with RNA.
892 Furthermore, a subset of *CHD2* episignature probes overlap with DMRs in the TSS/5'UTR of
893 developmentally relevant genes and might regulate expression (Table S7). For instance, a cluster
894 of hypermethylated episignature probes for the *CHD2* 450K and 850K episignatures are
895 contained within a larger hypermethylated DMR in the TSS and 5'UTR of *HOXA4* (Figure S31).
896 However, *HOXA4* is not expressed in the blood, and, therefore, would not be expected to be
897 impacted by differential methylation. Thus, we have shown that *CHD2* is associated with DMRs
898 in the blood that correspond with the episignature. Our work suggests that future studies should
899 investigate the *CHD2* episignature in disease-relevant tissue types where DMRs are likely to
900 contribute directly to gene dysregulation and disease pathogenesis.

901 Here, we have utilized various DNA methylation analyses to identify causative and
902 candidate etiologies in 2% of our cohort of 516 individuals with unsolved DEEs. While DNA
903 methylation does not explain the majority of DEEs, methylation array yield is comparable to the
904 current added utility of GS^{4,73} and remains a low-cost approach that can detect missed genetic
905 etiologies and propose new molecular candidates. Importantly, this yield is expected to increase
906 over time as we interrogate the functional consequences of rare DMRs and better understand

907 which genes and pathways exhibit epesignatures, including unraveling inconclusive epesignature
908 results. We have also investigated the epesignature for the DEE gene *CHD2* in-depth and have
909 provided evidence that the CHD2 epesignature is associated with DMRs. DMRs may affect gene
910 expression, especially in disease-relevant tissue types. Furthermore, CHD2 epesignatures and
911 associated DMRs may have potential as a biomarker readout for therapeutic testing, as the DNA
912 methylation might potentially be reversed with targeted treatment. Thus, our work highlights the
913 impact of investigating DNA methylation in DEEs, both for the genetic diagnosis of unsolved cases
914 and to augment our understanding of underlying disease function toward the future development
915 of targeted therapies.

916

917 **Declaration of Interests**

918 B.S. is a shareholder in EpiSign Inc, a company involved in commercialization of EpiSign™
919 software. D.E.M is on a scientific advisory board at ONT and has received travel support from
920 ONT to speak on their behalf. D.E.M is engaged in a research agreement with ONT. D.E.M holds
921 stock options in MyOme. I.E.S. has served on scientific advisory boards for BioMarin, Chiesi,
922 Eisai, Encoded Therapeutics, GlaxoSmithKline, Knopp Biosciences, Nutricia, Rogcon, Takeda
923 Pharmaceuticals, UCB, Xenon Pharmaceuticals, Cerecin; has received speaker honoraria from
924 GlaxoSmithKline, UCB, BioMarin, Biocodex, Chiesi, Liva Nova, Nutricia, Zuellig Pharma, Stoke
925 Therapeutics and Eisai; has received funding for travel from UCB, Biocodex, GlaxoSmithKline,
926 Biomarin, Encoded Therapeutics Stoke Therapeutics and Eisai; has served as an investigator for
927 Anavex Life Sciences, Cerevel Therapeutics, Eisai, Encoded Therapeutics, EpiMinder Inc,
928 Epygenyx, ES-Therapeutics, GW Pharma, Marinus, Neurocrine BioSciences, Ovid Therapeutics,
929 Takeda Pharmaceuticals, UCB, Ultragenyx, Xenon Pharmaceuticals, Zogenix and Zynerva; has
930 consulted for Care Beyond Diagnosis, Epilepsy Consortium, Atheneum Partners, Ovid
931 Therapeutics, UCB, Zynerva Pharmaceuticals, BioMarin, Encoded Therapeutics and Biohaven
932 Pharmaceuticals; and is a Non-Executive Director of Bellberry Ltd and a Director of the Australian
933 Academy of Health and Medical Sciences and the Australian Council of Learned Academies
934 Limited. I.E.S. may accrue future revenue on pending patent WO61/010176 (filed: 2008):
935 Therapeutic Compound; has a patent for *SCN1A* testing held by Bionomics Inc and licensed to
936 various diagnostic companies; has a patent molecular diagnostic/theragnostic target for benign
937 familial infantile epilepsy (BFIE) [PRRT2] 2011904493 & 2012900190 and PCT/AU2012/001321
938 (TECH ID:2012-009). L.G.S. receives funding from the Health Research Council of New Zealand
939 and Cure Kids New Zealand, is a consultant for the Epilepsy Consortium, and has received travel
940 grants from Seqirus and Nutricia. L.G.S. has received research grants and consultancy fees from
941 Zynerva Pharmaceuticals and has served on and Takeda and Eisai Pharmaceuticals scientific

942 advisory panels. The Department of Molecular and Human Genetics at Baylor College of Medicine
943 receives revenue from clinical genetic testing conducted at Baylor Genetics Laboratories.

944

945 **Acknowledgments**

946 We thank all the individuals and their families for participating in this research. Major funding for
947 this project was provided by a grant (#631106) from Citizens United for Research in Epilepsy
948 (CURE). A subset of DNA methylation arrays was provided by the University of Washington
949 Center for Rare Disease Research (UW-CRDR), formerly known as the Center for Mendelian
950 Genomics (CMG), with support from NHGRI grants U01 HG011744, UM1 HG006493 and U24
951 HG011746 and with enthusiastic support from the late Debbie Nickerson. We gratefully
952 acknowledge support from the Australian Epilepsy Research Foundation grant, Australian
953 National Health and Medical Research Council (NHMRC) Centre for Research Excellence Grant
954 (GNT2006841), NHMRC Synergy Grant (GNT2010562), the Health Research Council of New
955 Zealand, Cure Kids New Zealand, and the Estate of Ernest Hyam Davis and the Tedd and Mollie
956 Carr Endowment Trust. We acknowledge the Epi25 Consortium, which provided exome sequence
957 data for review for a subset of individuals. C.W.L. has been funded through the American Epilepsy
958 Society (AES) predoctoral fellowship and the St. Jude Graduate School of Biomedical Sciences.
959 We would also like to acknowledge the inaugural St. Jude Biohackathon 2022 for coordinating
960 the event that led to a team comprised of C.W.L., P.K., M.D., and W.R., who assembled the
961 MethylMiner pipeline described here. D.E.M. is supported by NIH grant DP5OD033357. Research
962 reported in this manuscript by M.W.H., S.K., H.D., K.C.W., J.A.R., and H.T.C., was supported by
963 the NIH Common Fund, through the Office of Strategic Coordination/Office of the NIH Director
964 under Award Numbers U01HG007709 and U01HG007942. K.L.P. has been funded through the
965 *GRIN2B* foundation and CURE. I.E.S. is also supported by a NHMRC Senior Investigator

966 Fellowship (GNT1172897). We acknowledge Pratibha Kottapalli and Sanchit Trivedi from the St.
967 Jude Hartwell Center who performed Illumina sequencing for this project. The content is solely
968 the responsibility of the authors and does not necessarily represent the official views of the NIH.

969

970 **Data and Code Availability**

971 Methylation array data for individuals with unsolved DEEs and those with pathogenic variants in
972 *CHD2* who have given consent for data sharing will be made available through dbGaP. Additional
973 data requests can be directed to H.C.M. The methylation array analysis pipeline used in part of
974 this study for epivariant detection can be accessed on GitHub: [https://github.com/stjude-](https://github.com/stjude-biohackathon/MethylMiner)
975 [biohackathon/MethylMiner](https://github.com/stjude-biohackathon/MethylMiner). EpiSign™ is proprietary commercial software and is not publicly
976 available.

977

978 **Supplemental Information**

979 Supplemental phenotype data, methods, and figures are in the supplemental documents section.

980

981 **REFERENCES**

- 982 1 Scheffer, I. E. *et al.* ILAE classification of the epilepsies: Position paper of the ILAE
983 Commission for Classification and Terminology. *Epilepsia* **58**, 512-521 (2017).
984 <https://doi.org/10.1111/epi.13709>
- 985 2 Oliver, K. L. *et al.* Genes4Epilepsy: An epilepsy gene resource. *Epilepsia* **64**, 1368-1375
986 (2023). <https://doi.org/10.1111/epi.17547>
- 987 3 Poke, G., Stanley, J., Scheffer, I. E. & Sadleir, L. G. Epidemiology of Developmental and
988 Epileptic Encephalopathy and of Intellectual Disability and Epilepsy in Children.
989 *Neurology* **100**, e1363-e1375 (2023). <https://doi.org/10.1212/WNL.0000000000206758>
- 990 4 Palmer, E. E. *et al.* Integrating exome sequencing into a diagnostic pathway for epileptic
991 encephalopathy: Evidence of clinical utility and cost effectiveness. *Mol Genet Genomic*
992 *Med* **6**, 186-199 (2018). <https://doi.org/10.1002/mgg3.355>
- 993 5 McTague, A., Howell, K. B., Cross, J. H., Kurian, M. A. & Scheffer, I. E. The genetic
994 landscape of the epileptic encephalopathies of infancy and childhood. *The Lancet*
995 *Neurology* **15**, 304-316 (2016). [https://doi.org/10.1016/s1474-4422\(15\)00250-1](https://doi.org/10.1016/s1474-4422(15)00250-1)
- 996 6 Sanchez Fernandez, I., Loddenkemper, T., Gainza-Lein, M., Sheidley, B. R. & Poduri, A.
997 Diagnostic yield of genetic tests in epilepsy: A meta-analysis and cost-effectiveness
998 study. *Neurology* (2019). <https://doi.org/10.1212/WNL.0000000000006850>
- 999 7 Symonds, J. D. & McTague, A. Epilepsy and developmental disorders: Next generation
1000 sequencing in the clinic. *Eur J Paediatr Neurol* **24**, 15-23 (2020).
1001 <https://doi.org/10.1016/j.ejpn.2019.12.008>
- 1002 8 Mefford, H. C. *et al.* Rare copy number variants are an important cause of epileptic
1003 encephalopathies. *Ann Neurol* **70**, 974-985 (2011). <https://doi.org/10.1002/ana.22645>
- 1004 9 A roadmap for precision medicine in the epilepsies. *The Lancet Neurology* **14**, 1219-
1005 1228 (2015). [https://doi.org/10.1016/s1474-4422\(15\)00199-4](https://doi.org/10.1016/s1474-4422(15)00199-4)
- 1006 10 Bayat, A., Bayat, M., Rubboli, G. & Moller, R. S. Epilepsy Syndromes in the First Year of
1007 Life and Usefulness of Genetic Testing for Precision Therapy. *Genes (Basel)* **12** (2021).
1008 <https://doi.org/10.3390/genes12071051>
- 1009 11 D'Gama, A. M. *et al.* Evaluation of the feasibility, diagnostic yield, and clinical utility of
1010 rapid genome sequencing in infantile epilepsy (Gene-STEPS): an international,
1011 multicentre, pilot cohort study. *Lancet Neurol* **22**, 812-825 (2023).
1012 [https://doi.org/10.1016/S1474-4422\(23\)00246-6](https://doi.org/10.1016/S1474-4422(23)00246-6)
- 1013 12 Sheidley, B. R. *et al.* Genetic testing for the epilepsies: A systematic review. *Epilepsia*
1014 (2021). <https://doi.org/10.1111/epi.17141>
- 1015 13 Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function.
1016 *Neuropsychopharmacology* **38**, 23-38 (2013). <https://doi.org/10.1038/npp.2012.112>
- 1017 14 Garg, P. *et al.* A Survey of Rare Epigenetic Variation in 23,116 Human Genomes
1018 Identifies Disease-Relevant Epivariations and CGG Expansions. *Am J Hum Genet* **107**,
1019 654-669 (2020). <https://doi.org/10.1016/j.ajhg.2020.08.019>
- 1020 15 LaCroix, A. J. *et al.* GGC Repeat Expansion and Exon 1 Methylation of XYLT1 Is a
1021 Common Pathogenic Variant in Baratela-Scott Syndrome. *Am J Hum Genet* **104**, 35-44
1022 (2019). <https://doi.org/10.1016/j.ajhg.2018.11.005>
- 1023 16 Barbosa, M. *et al.* Identification of rare de novo epigenetic variations in congenital
1024 disorders. *Nat Commun* **9**, 2064 (2018). <https://doi.org/10.1038/s41467-018-04540-x>
- 1025 17 Levy, M. A. *et al.* Novel diagnostic DNA methylation epigenatures expand and refine the
1026 epigenetic landscapes of Mendelian disorders. *Human Genetics and Genomics*
1027 *Advances* (2021). <https://doi.org/10.1016/j.xhgg.2021.100075>

- 1028 18 Aref-Eshghi, E. *et al.* Evaluation of DNA Methylation Episignatures for Diagnosis and
1029 Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *Am J Hum*
1030 *Genet* **106**, 356-370 (2020). <https://doi.org/10.1016/j.ajhg.2020.01.019>
1031 19 Sadikovic, B. *et al.* Clinical epigenomics: genome-wide DNA methylation analysis for the
1032 diagnosis of Mendelian disorders. *Genet Med* **23**, 1065-1074 (2021).
1033 <https://doi.org/10.1038/s41436-020-01096-4>
1034 20 Levy, M. A. *et al.* Functional correlation of genome-wide DNA methylation profiles in
1035 genetic neurodevelopmental disorders. *Hum Mutat* **43**, 1609-1628 (2022).
1036 <https://doi.org/10.1002/humu.24446>
1037 21 Carvill, G. L. *et al.* Targeted resequencing in epileptic encephalopathies identifies de
1038 novo mutations in CHD2 and SYNGAP1. *Nat Genet* **45**, 825-830 (2013).
1039 <https://doi.org/10.1038/ng.2646>
1040 22 Scheffer, I. E. *et al.* Exome sequencing for patients with developmental and epileptic
1041 encephalopathies in clinical practice. *Dev Med Child Neurol* **65**, 50-57 (2023).
1042 <https://doi.org/10.1111/dmcn.15308>
1043 23 Parkinson Progression Marker, I. The Parkinson Progression Marker Initiative (PPMI).
1044 *Prog Neurobiol* **95**, 629-635 (2011). <https://doi.org/10.1016/j.pneurobio.2011.09.005>
1045 24 Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip
1046 microarray for whole-genome DNA methylation profiling. *Genome Biol* **17**, 208 (2016).
1047 <https://doi.org/10.1186/s13059-016-1066-1>
1048 25 Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the
1049 analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-1369 (2014).
1050 <https://doi.org/10.1093/bioinformatics/btu049>
1051 26 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package
1052 for removing batch effects and other unwanted variation in high-throughput experiments.
1053 *Bioinformatics* **28**, 882-883 (2012). <https://doi.org/10.1093/bioinformatics/bts034>
1054 27 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression
1055 data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
1056 <https://doi.org/10.1093/biostatistics/kxj037>
1057 28 Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-
1058 wide association studies. *Genome Biology* **15**, R31 (2014). [https://doi.org/10.1186/gb-](https://doi.org/10.1186/gb-2014-15-2-r31)
1059 [2014-15-2-r31](https://doi.org/10.1186/gb-2014-15-2-r31)
1060 29 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime
1061 cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-
1062 589 (2010). <https://doi.org/10.1016/j.molcel.2010.05.004>
1063 30 Court, F. *et al.* Genome-wide parent-of-origin DNA methylation analysis reveals the
1064 intricacies of human imprinting and suggests a germline methylation-independent
1065 mechanism of establishment. *Genome Res* **24**, 554-569 (2014).
1066 <https://doi.org/10.1101/gr.164913.113>
1067 31 Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and
1068 innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45**,
1069 e22 (2017). <https://doi.org/10.1093/nar/gkw967>
1070 32 Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online
1071 Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic
1072 disorders. *Nucleic Acids Res* **33**, D514-517 (2005). <https://doi.org/10.1093/nar/gki033>
1073 33 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24-26
1074 (2011). <https://doi.org/10.1038/nbt.1754>
1075 34 Vaisvila, R. *et al.* Enzymatic methyl sequencing detects DNA methylation at single-base
1076 resolution from picograms of DNA. *Genome Res* **31**, 1280-1289 (2021).
1077 <https://doi.org/10.1101/gr.266551.120>

- 1078 35 Miller, D. E. *et al.* Targeted long-read sequencing identifies missing disease-causing
1079 variation. *Am J Hum Genet* **108**, 1436-1449 (2021).
1080 <https://doi.org/10.1016/j.ajhg.2021.06.006>
- 1081 36 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-
1082 3100 (2018). <https://doi.org/10.1093/bioinformatics/bty191>
- 1083 37 Zheng, Z. *et al.* Symphonizing pileup and full-alignment for deep learning-based long-
1084 read variant calling. *Nature Computational Science* **2**, 797-803 (2022).
1085 <https://doi.org/10.1038/s43588-022-00387-x>
- 1086 38 Sedlazeck, F. J. *et al.* Accurate detection of complex structural variations using single-
1087 molecule sequencing. *Nat Methods* **15**, 461-468 (2018). <https://doi.org/10.1038/s41592-018-0001-7>
- 1089 39 Heller, D. & Vingron, M. SVIM: structural variant identification using mapped long reads.
1090 *Bioinformatics* **35**, 2907-2915 (2019). <https://doi.org/10.1093/bioinformatics/btz041>
- 1091 40 Jiang, T. *et al.* Long-read-based human genomic structural variation detection with
1092 cuteSV. *Genome Biol* **21**, 189 (2020). <https://doi.org/10.1186/s13059-020-02107-y>
- 1093 41 Lin, J. H., Chen, L. C., Yu, S. C. & Huang, Y. T. LongPhase: an ultra-fast chromosome-
1094 scale phasing algorithm for small and large variants. *Bioinformatics* **38**, 1816-1822
1095 (2022). <https://doi.org/10.1093/bioinformatics/btac058>
- 1096 42 *EpiSign v4 Menu*, https://episign.lhsc.on.ca/img/EpiSign_v4_Menu.pdf.
- 1097 43 Platt, J. C. Probabilistic outputs for support vector machines and comparisons to
1098 regularized likelihood methods. *Advances in Large Margin Classifiers* **10** (2000).
- 1099 44 Sadikovic, B., Levy, M. A. & Aref-Eshghi, E. Functional annotation of genomic variation:
1100 DNA methylation epigenatures in neurodevelopmental Mendelian disorders. *Hum Mol*
1101 *Genet* **29**, R27-R32 (2020). <https://doi.org/10.1093/hmg/ddaa144>
- 1102 45 FastQC: A Quality Control Tool for High Throughput Sequence Data (2010).
- 1103 46 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
1104 (2013). <https://doi.org/10.1093/bioinformatics/bts635>
- 1105 47 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast
1106 and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419 (2017).
1107 <https://doi.org/10.1038/nmeth.4197>
- 1108 48 Brechtmann, F. *et al.* OUTRIDER: A Statistical Method for Detecting Aberrantly
1109 Expressed Genes in RNA Sequencing Data. *Am J Hum Genet* **103**, 907-917 (2018).
1110 <https://doi.org/10.1016/j.ajhg.2018.10.025>
- 1111 49 Yépez, V. A., Murdock, D. R. & Lee, B. Gene expression counts from fibroblast, strand-
1112 specific, BCM UDN. *Zenodo* (2020). <https://doi.org/10.5281/zenodo.3963474>
- 1113 50 Aref-Eshghi, E. *et al.* BAFopathies' DNA methylation epi-signatures demonstrate
1114 diagnostic utility and functional continuum of Coffin-Siris and Nicolaides-Baraitser
1115 syndromes. *Nat Commun* **9**, 4885 (2018). <https://doi.org/10.1038/s41467-018-07193-y>
- 1116 51 Aref-Eshghi, E. *et al.* Genomic DNA Methylation Signatures Enable Concurrent
1117 Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental
1118 Syndromes. *Am J Hum Genet* **102**, 156-174 (2018).
1119 <https://doi.org/10.1016/j.ajhg.2017.12.008>
- 1120 52 Aref-Eshghi, E. *et al.* Diagnostic Utility of Genome-wide DNA Methylation Testing in
1121 Genetically Unsolved Individuals with Suspected Hereditary Conditions. *Am J Hum*
1122 *Genet* **104**, 685-700 (2019). <https://doi.org/10.1016/j.ajhg.2019.03.008>
- 1123 53 Ho, D. E., Imai, K., King, G. & Stuart, E. A. Matching as Nonparametric Preprocessing
1124 for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* **15**,
1125 199-236 (2017). <https://doi.org/10.1093/pan/mpl013>
- 1126 54 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing
1127 and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
1128 <https://doi.org/10.1093/nar/gkv007>

- 1129 55 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture
1130 distribution. *BMC Bioinformatics* **13**, 86 (2012). <https://doi.org/10.1186/1471-2105-13-86>
- 1131 56 Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances
1132 circular visualization in R. *Bioinformatics* **30**, 2811-2812 (2014).
1133 <https://doi.org/10.1093/bioinformatics/btu393>
- 1134 57 Cavalcante, R. G. & Sartor, M. A. annotatr: genomic regions in context. *Bioinformatics*
1135 **33**, 2381-2383 (2017). <https://doi.org/10.1093/bioinformatics/btx183>
- 1136 58 TreeAndLeaf: Displaying binary trees with focus on dendrogram leaves (R package
1137 version 1.12.0., 2023).
- 1138 59 Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves
1139 replication in large cancer studies. *Genome Biology* **15**, 503 (2014).
1140 <https://doi.org/10.1186/s13059-014-0503-2>
- 1141 60 Jaffe, A. E. *et al.* Bump hunting to identify differentially methylated regions in epigenetic
1142 epidemiology studies. *Int J Epidemiol* **41**, 200-209 (2012).
1143 <https://doi.org/10.1093/ije/dyr238>
- 1144 61 Peters, T. J. *et al.* De novo identification of differentially methylated regions in the human
1145 genome. *Epigenetics & Chromatin* **8**, 6 (2015). <https://doi.org/10.1186/1756-8935-8-6>
- 1146 62 Peters, T. J. *et al.* Calling differentially methylated regions from whole genome bisulphite
1147 sequencing with DMRcate. *Nucleic Acids Res* **49**, e109 (2021).
1148 <https://doi.org/10.1093/nar/gkab637>
- 1149 63 Wu, H. *et al.* Detection of differentially methylated regions from whole-genome bisulfite
1150 sequencing data without replicates. *Nucleic Acids Res* **43**, e141 (2015).
1151 <https://doi.org/10.1093/nar/gkv715>
- 1152 64 Tsutsumi, M. *et al.* A female patient with retinoblastoma and severe intellectual disability
1153 carrying an X;13 balanced translocation without rearrangement in the RB1 gene: a case
1154 report. *BMC Med Genomics* **12**, 182 (2019). <https://doi.org/10.1186/s12920-019-0640-2>
- 1155 65 Chen, X. *et al.* A de novo pathogenic CSNK1E mutation identified by exome sequencing
1156 in family trios with epileptic encephalopathy. *Hum Mutat* **40**, 281-287 (2019).
1157 <https://doi.org/10.1002/humu.23690>
- 1158 66 Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for
1159 connecting investigators with an interest in the same gene. *Hum Mutat* **36**, 928-930
1160 (2015). <https://doi.org/10.1002/humu.22844>
- 1161 67 Winnepeninckx, B. *et al.* CGG-repeat expansion in the DIP2B gene is associated with
1162 the fragile site FRA12A on chromosome 12q13.1. *Am J Hum Genet* **80**, 221-231 (2007).
1163 <https://doi.org/10.1086/510800>
- 1164 68 Kiedrowski, L. A. *et al.* DNA methylation assay for X-chromosome inactivation in female
1165 human iPS cells. *Stem Cell Rev Rep* **7**, 969-975 (2011). <https://doi.org/10.1007/s12015-011-9238-6>
- 1166 69 conumee: Enhanced copy-number variation analysis using Illumina DNA methylation
1167 arrays.
- 1168 70 Harada, A. *et al.* Chd2 interacts with H3.3 to determine myogenic cell fate. *EMBO J* **31**,
1170 2994-3007 (2012). <https://doi.org/10.1038/emboj.2012.136>
- 1171 71 Lamar, K. J. & Carvill, G. L. Chromatin Remodeling Proteins in Epilepsy: Lessons From
1172 CHD2-Associated Epilepsy. *Front Mol Neurosci* **11**, 208 (2018).
1173 <https://doi.org/10.3389/fnmol.2018.00208>
- 1174 72 Geeleher, P. *et al.* Gene-set analysis is severely biased when applied to genome-wide
1175 methylation data. *Bioinformatics* **29**, 1851-1857 (2013).
1176 <https://doi.org/10.1093/bioinformatics/btt311>
- 1177 73 Alfares, A. *et al.* Whole-genome sequencing offers additional but limited clinical utility
1178 compared with reanalysis of whole-exome sequencing. *Genet Med* **20**, 1328-1333
1179 (2018). <https://doi.org/10.1038/gim.2018.41>

- 1180 74 Palmer, E. E. *et al.* Diagnostic Yield of Whole Genome Sequencing After Nondiagnostic
1181 Exome Sequencing or Gene Panel in Developmental and Epileptic Encephalopathies.
1182 *Neurology* **96**, e1770-e1782 (2021). <https://doi.org/10.1212/WNL.00000000000011655>
1183 75 Vielhaber, E., Eide, E., Rivers, A., Gao, Z. H. & Virshup, D. M. Nuclear entry of the
1184 circadian regulator mPER1 is controlled by mammalian casein kinase I epsilon. *Mol Cell*
1185 *Biol* **20**, 4888-4899 (2000). <https://doi.org/10.1128/mcb.20.13.4888-4899.2000>
1186 76 Lee, C., Weaver, D. R. & Reppert, S. M. Direct association between mouse PERIOD
1187 and CKepsilon is critical for a functioning circadian clock. *Mol Cell Biol* **24**, 584-594
1188 (2004). <https://doi.org/10.1128/mcb.24.2.584-594.2004>
1189 77 Toh, K. L. *et al.* An hPer2 phosphorylation site mutation in familial advanced sleep phase
1190 syndrome. *Science* **291**, 1040-1043 (2001). <https://doi.org/10.1126/science.1057499>
1191 78 Zhou, L. *et al.* The circadian clock gene Csnk1e regulates rapid eye movement sleep
1192 amount, and nonrapid eye movement sleep architecture in mice. *Sleep* **37**, 785-793,
1193 793A-793C (2014). <https://doi.org/10.5665/sleep.3590>
1194 79 Leitao, E. *et al.* Systematic analysis and prediction of genes associated with monogenic
1195 disorders on human chromosome X. *Nat Commun* **13**, 6570 (2022).
1196 <https://doi.org/10.1038/s41467-022-34264-y>
1197 80 Platzer, K. *et al.* De Novo Missense Variants in SLC32A1 Cause a Developmental and
1198 Epileptic Encephalopathy Due to Impaired GABAergic Neurotransmission. *Ann Neurol*
1199 **92**, 958-973 (2022). <https://doi.org/10.1002/ana.26485>
1200 81 Cappuccio, G. *et al.* De novo SMARCA2 variants clustered outside the helicase domain
1201 cause a new recognizable syndrome with intellectual disability and blepharophimosis
1202 distinct from Nicolaides-Baraitser syndrome. *Genet Med* **22**, 1838-1850 (2020).
1203 <https://doi.org/10.1038/s41436-020-0898-y>
1204 82 Liu, P. *et al.* Reanalysis of Clinical Exome Sequencing Data. *N Engl J Med* **380**, 2478-
1205 2480 (2019). <https://doi.org/10.1056/NEJMc1812033>
1206 83 Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44-53 (2022).
1207 <https://doi.org/10.1126/science.abj6987>
1208 84 Rooney, K. *et al.* DNA methylation epigenature and comparative epigenomic profiling of
1209 HNRNPU-related neurodevelopmental disorder. *Genet Med*, 100871 (2023).
1210 <https://doi.org/10.1016/j.gim.2023.100871>
1211
1212

1213

Tables and Table Legends

Table 1. Summary of epivariants and underlying DNA defects identified in this study.

| Location | Gene | Direction | Underlying DNA Defect | Inheritance | Probands (n) |
|----------|-----------------------|-----------|------------------------|----------------|--------------|
| chr13 | chr13:multiple | Hyper | X;13 translocation (p) | <i>De novo</i> | 1 |
| Xp22 | <i>BCLAF3</i> | Hyper | CGG repeat (c) | <i>De novo</i> | 1 |
| 22q13 | <i>CSNK1E</i> | Hyper | CGG repeat (c) | Inherited | 2 + 1 match |
| 12q13 | <i>DIP2B</i> | Hyper | CGG repeat (c) | Inherited | 1 |
| 2p16 | <i>CFAP36/CCDC104</i> | Hyper | Tandem duplication (b) | Inherited | 1 |
| 2q37.3 | chr2:multiple | Hypo | Deletion (b) | Inherited | 1 |
| 15q24 | <i>LINGO1</i> | Hypo | Deletion (b) | Inherited | 4 |

List of epivariation findings from screening cohort with unsolved DEEs for rare outlier DNA methylation changes. Molecular findings are considered p=pathogenic, c=candidate, and b=benign.

1214

Table 2. Summary of episignatures and causative sequence variants identified in this study.

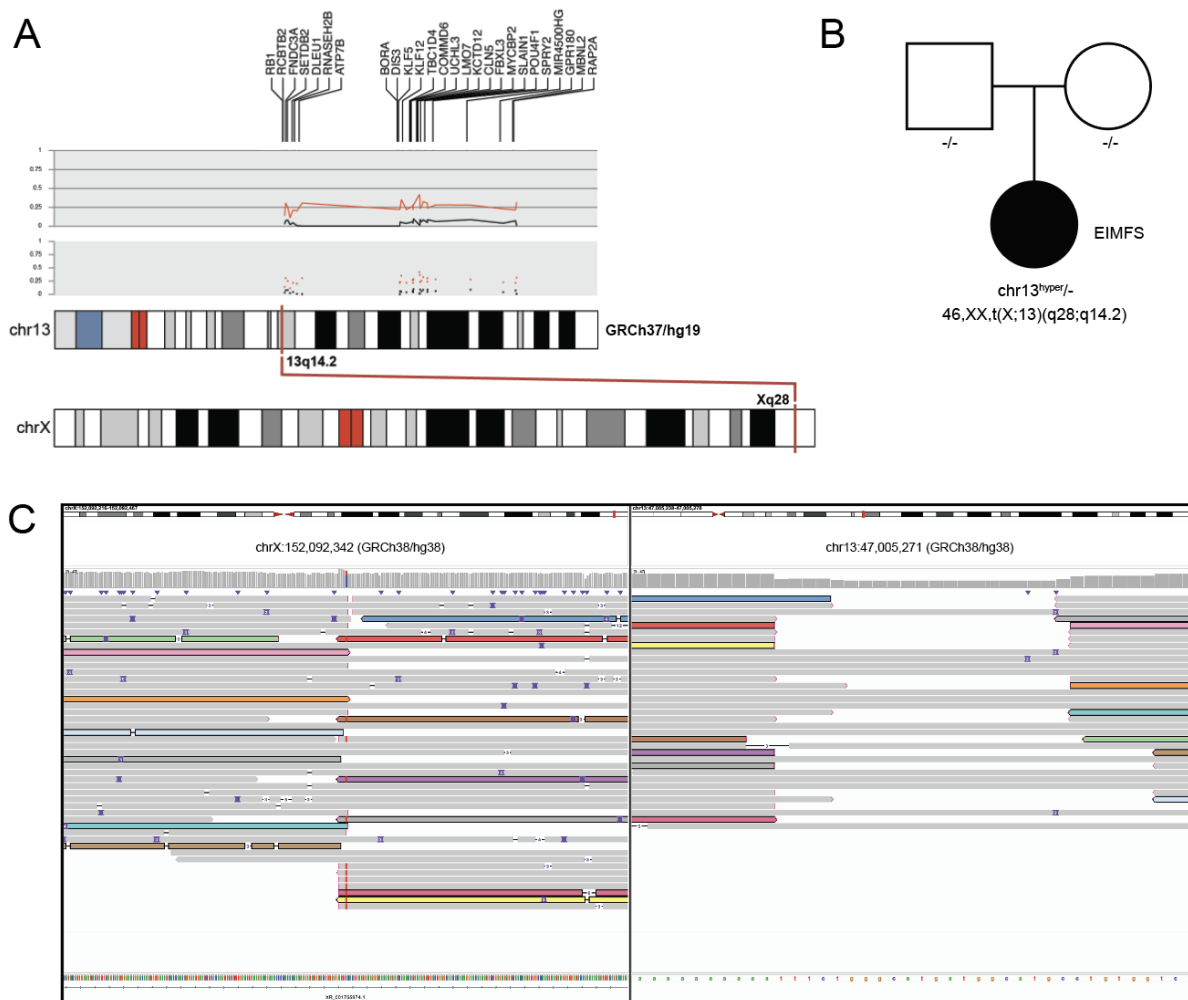
| Gene | Signature | MVP | Genomic Variant (hg38) | Consequence | Inheritance |
|----------------|----------------|-------|------------------------|------------------------|----------------|
| <i>ANKRD11</i> | KGBS/ MRD23 | 0.854 | chr16:89284030 G>A | p.Arg838Ter (p) | <i>De novo</i> |
| | | 0.989 | chr16:89279671 C>A | p.Glu2291Ter (p) | <i>De novo</i> |
| <i>SETD1B</i> | IDDSELD | 0.763 | chr12:121822939 C>T | p.Arg1454Ter (p) | Unknown |
| <i>TET3</i> | BEFAHRS | 0.327 | chr2:74102031 dup | p.Thr1749HisfsTer5 (p) | Inherited |
| <i>UBE2A</i> | MRXSN | 0.149 | chrX:119583172 G>A | p.Ala126Thr (p) | Inherited |
| <i>KDM2B</i> | KDM2B | 0.982 | chr12:121520986 C>G | p.Arg349Pro (u) | Inherited |

List of episignature findings from screening unsolved. Methylation Variant Pathogenicity (MVP) scores shown. Molecular findings are considered p=pathogenic or u=variant of uncertain significance.

1215

1216

Figures and Figure Legends

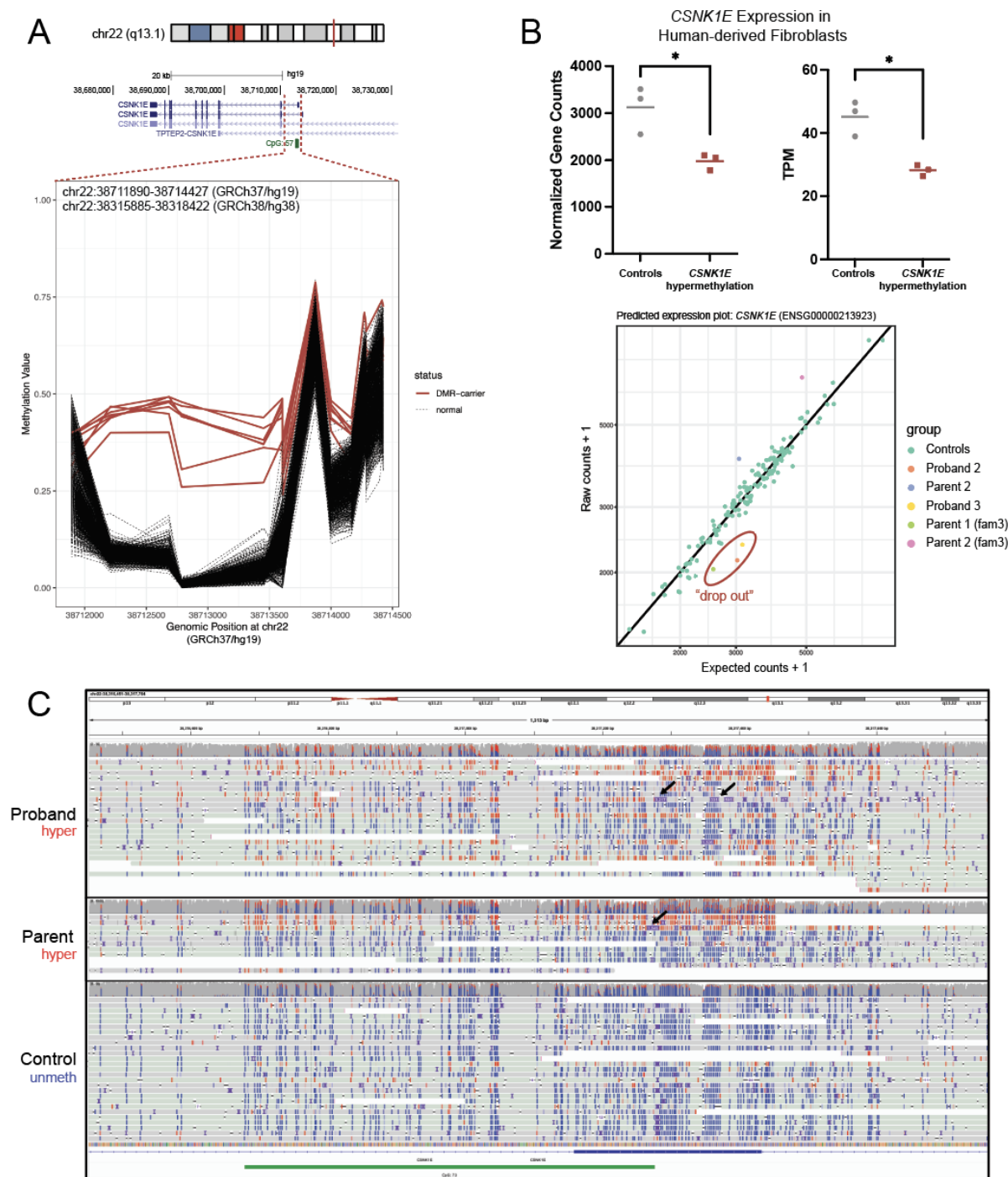


1217

1218 **Figure 1:** Rare outlier DMR analysis identifies chromosomal hypermethylation caused by X;13
1219 translocation. **A.** Graphical representation of chr13 rare hypermethylation events in a proband
1220 with unsolved DEE. The upper portion of the track displays the genes on chr13 for which
1221 hypermethylation events were called. The two grey panels (upper=line, lower=dot) depict β -
1222 values for the average of the proband's array replicates (red) and the average of the parents'
1223 array data (black) for a representative probe within each DMR ($n=27$). Subtle hypermethylation
1224 hovering around $\sim 25\%$ can be seen for the proband compared to the parents. The lower track
1225 shows chromosomal locations of the DMRs of the X;13 translocation. **B.** Pedigree showing that

1226 chr13 hypermethylation events and the X;13 translocation occurred *de novo*. **C.** IGV view of ONT
1227 LRS data for chrX (left) and chr13 (right). Some, but not all, reads spanning the translocation are
1228 colored to show that they span the breakpoint.

1229



1230

1231 **Figure 2:** Rare outlier DMR analysis identifies tandem repeat expansions. **A.** DMR plot depicting
 1232 outlier hypermethylation of the *CSNK1E* 5'UTR and intron 1 in two probands with unsolved DEEs
 1233 (three replicates across both for a total of five samples), one parent, and one unaffected control

1234 (total n=7 red lines) detected through epivariation analysis. **B.** The upper panel shows expression
1235 values from RNA-seq of human-derived fibroblasts for individuals with *CSNK1E* hypermethylation
1236 compared to control methylation levels. Significance between groups was determined by a two-
1237 tailed paired t-test ($p=0.029$ for gene counts and $p=0.0169$ for transcripts per million or TPM). A
1238 representative predicted expression plot from drop-out analysis using the OUTRIDER algorithm
1239 is shown at the lower portion of the panel. See Figure S9 for the individual OUTRIDER plots for
1240 each family and significance information. **C.** Unphased IGV view of LRS data showing CpG sites
1241 that are methylated (red) and unmethylated (blue). The CGG repeat expansion seen in the
1242 proband was inherited from the parent and is shown as purple squares denoting insertions in the
1243 reads (black arrows); not all reads that are methylated show the insertion as they terminated
1244 within the inserted sequence and are clipped by the alignment process.

1245

1246

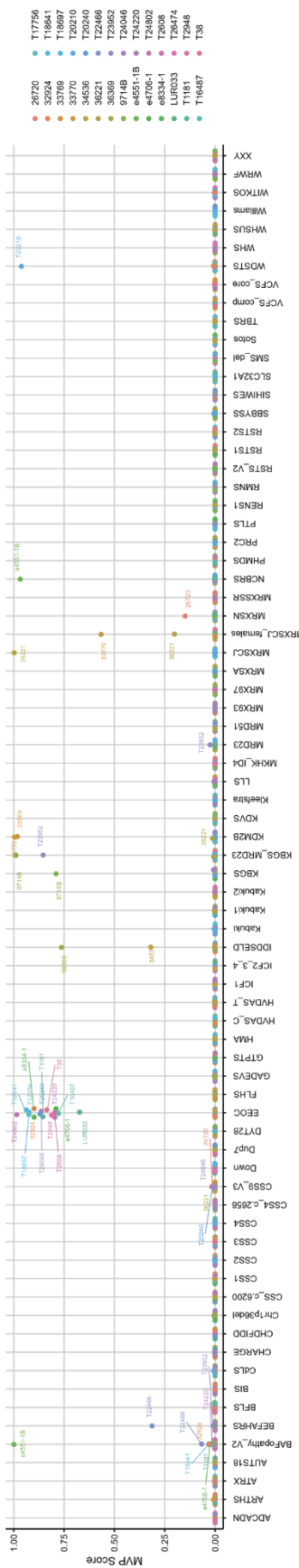
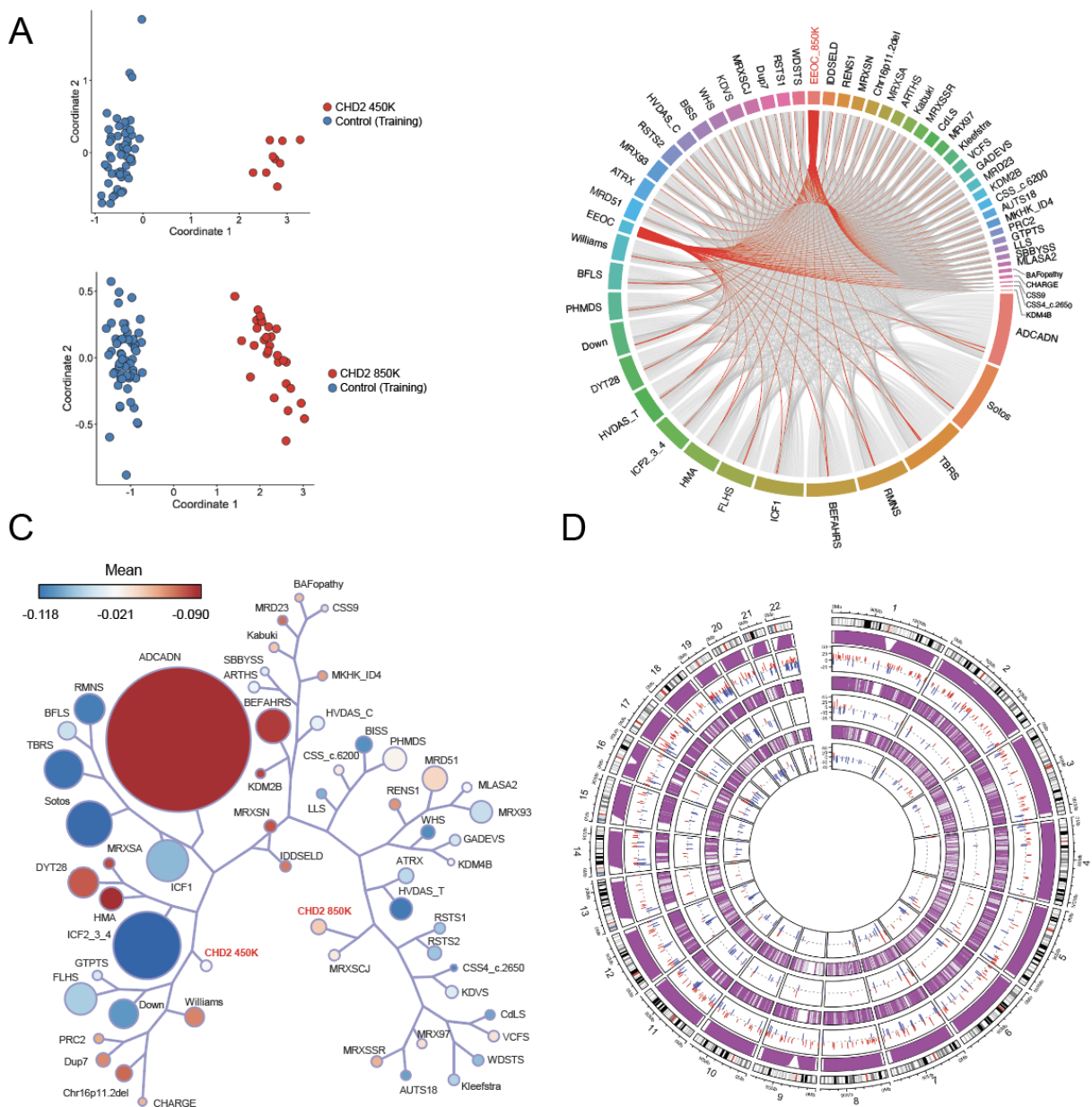


Figure 3: Summary Methylation Variant Pathogenicity (MVP) score for all individuals positive for EpiSign™ v4 Episignature analysis. A Methylation Variant Pathogenicity (MVP) score (between 0 and 1) was generated to represent the confidence of prediction for the specific episignature on the EpiSign™ v4 clinical classifier that the SVM was trained to detect. Each colored circle represents a different individual and its associated MVP score for each of the episignatures on the EpiSign™ v4 clinical classifier. Final classification for a specific EpiSign™ disorder includes a combination of MVP score, hierarchical clustering, and multidimensional scaling (MDS) review.



1247

1248 **Figure 4:** Insights from the *CHD2* Episignature. **A.** Multidimensional scaling (MDS) plot showing

1249 clustering of individuals with pathogenic *CHD2* variants (red, upper) for the previously described

1250 *CHD2* 450K (n=9) episignature with shared 450K and 850K array probes clusters away from the

1251 controls (blue) The refined *CHD2* 850K (n=29) episignature (red, lower) clusters away from

1252 unaffected controls (blue). **B.** Circos plot representing shared probes between episignatures.

1253 Differentially methylated probes (DMPs) shared between the *CHD2* 850K cohort (bold red), *CHD2*

1254 450K cohort (red), and 55 other epesignatures on EpiSign with functional correlation analysis
1255 previously published²⁰. The thickness of the connecting lines corresponds to the number of probes
1256 shared between the cohorts. **C.** Tree and leaf visualization of Euclidean clustering of
1257 epesignatures. Tree and leaf visualization for all 57 cohorts using the top 500 DMPs for each group
1258 (for cohorts with less than 500 DMPs, all DMPs were used). Cohort samples were aggregated
1259 using the median value of each probe within a group. A leaf node represents a cohort, with node
1260 sizes illustrating relative scales of the number of selected DMPs for the respective cohort, and
1261 node colors are indicative of the global mean methylation difference, a gradient of
1262 hypomethylation (blue) or hypermethylation (red). **D.** Circular karyotype plot showing overlap of
1263 *CHD2* 450K epesignature probes (inner circle, n=200), with *CHD2* 850K epesignature probes
1264 (middle circle, n=200), and WGBS DMRs derived with at least a 15% methylation difference for
1265 the condensed visual representation (outer circle, n=411). Each line depicts a probe or DMR
1266 where red denotes hypermethylation and blue denotes hypomethylation. The purple tracks depict
1267 coverage of the 450K array probes (inner), 850K EPIC array probes (middle), and WGBS reads
1268 (outer). Refer to Figure S29 for linear karyotype DMR plots for chr1-22.