Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions

Erfan Aref-Eshghi,[1,2] Eric G. Bend,[3] Samantha Colaiacovo,[4] Michelle Caudle,[4] Rana Chakrabarti,[4] Melanie Napier,[4] Lauren Brick,[5] Lauren Brady,[5] Deanna Alexis Carere,[2] Michael A. Levy,[1,2] Jennifer Kerkhof,[2] Alan Stuart,[2] Maha Saleh,[4] Arthur L. Beaudet,[6] Chumei Li,[5] Maryia Kozenko,[5] Natalya Karp,[4] Chitra Prasad,[4] Victoria Mok Siu,[4] Mark A. Tarnopolsky,[5] Peter J. Ainsworth,[1,2] Hanxin Lin,[1,2] David I. Rodenhiser,[7] Ian D. Krantz,[8] Matthew A. Deardorff,[8] Charles E. Schwartz,[3] and Bekim Sadikovic[1,2,*]

Conventional genetic testing of individuals with neurodevelopmental presentations and congenital anomalies (ND/CAs), i.e., the analysis of sequence and copy number variants, leaves a substantial proportion of them unexplained. Some of these cases have been shown to result from DNA methylation defects at a single locus (epi-variants), while others can exhibit syndrome-specific DNA methylation changes across multiple loci (epi-signatures). Here, we investigate the clinical diagnostic utility of genome-wide DNA methylation analysis of peripheral blood in unresolved ND/CAs. We generate a computational model enabling concurrent detection of 14 syndromes using DNA methylation data with full accuracy. We demonstrate the ability of this model in resolving 67 individuals with uncertain clinical diagnoses, some of whom had variants of unknown clinical significance (VUS) in the related genes. We show that the provisional diagnoses can be ruled out in many of the case subjects, some of whom are shown by our model to have other diseases initially not considered. By applying this model to a cohort of 965 ND/CA-affected subjects without a previous diagnostic assumption and a separate assessment of rare epi-variants in this cohort, we identify 15 case subjects with syndromic Mendelian disorders, 12 case subjects with imprinting and trinucleotide repeat expansion disorders, as well as 106 case subjects with rare epi-variants, a portion of which involved genes clinically or functionally linked to the subjects' phenotypes. This study demonstrates that genomic DNA methylation analysis can facilitate the molecular diagnosis of unresolved clinical cases and highlights the potential value of epigenomic testing in the routine clinical assessment of ND/CAs.

Introduction

Constitutional disorders that present with developmental delay (DD), intellectual disability (ID), and/or congenital anomalies (CA) are associated with variable, complex, and overlapping clinical features. Current methodologies for the screening and diagnosis of neurodevelopmental syndromes with or without congenital anomalies (ND/CA) involve clinical microarray testing for detection of copy number variants (CNVs),[1] along with targeted gene panel and exome sequencing to identify rare pathogenic sequence variants.[2] Despite exhaustive assessments, in a large proportion of case subjects, the underlying genetic etiology is not identified or the clinical assessment does not indicate a diagnosis.[3] This is in part due to the inadequate assessment of non-coding and more complex sequence variants (e.g., long indels) through conventional approaches,[4] as well as limited knowledge of the genetic etiology of many clinically defined conditions. Furthermore, a large proportion of the genetic changes identified through genetic testing in a potentially relevant gene are not clinically interpretable, i.e., variants of unknown clinical significance (VUS).

One critical aspect of the molecular etiology of ND/CA that is not currently screened for in the clinical settings involves heritable changes that do not entail alterations in the DNA sequence, also known as epigenetics.[5] Defects in DNA methylation, the most commonly studied epigenetic phenomena, have long been known to be involved in a certain group of constitutional syndromes such as imprinting conditions and trinucleotide repeat expansion disorders.[6,7] Epi-variants, a class of genetic variants that involve a change in the DNA methylation patterns of a limited number of CpGs at a specific locus, have recently been proposed as a potential contributor to hereditary conditions.[8,9] It has been observed that rare *de novo* epi-variants are found at a higher frequency in unexplained ND/CA-affected case subjects, and their impact on gene expression has been shown to be comparable to those involving loss-of-function sequence variants, providing a rationale that some of these epi-variants may contribute to the pathogenesis of some of the unexplained ND/CAs.[9] On

[1]Department of Pathology and Laboratory Medicine, Western University, London, ON N6A 3K7, Canada; [2]Molecular Genetics Laboratory, Molecular Diagnostics Division, London Health Sciences Centre, London, ON N6A 5W9, Canada; [3]Greenwood Genetic Center, Greenwood, SC 29646, USA; [4]Medical Genetics Program of Southwestern Ontario, London Health Sciences Centre, London, ON N6A 5W9, Canada; [5]Department of Pediatrics, McMaster University, Hamilton, ON L8S 4L8, Canada; [6]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [7]Department of Pediatrics, Biochemistry and Oncology, Western University, London, ON N6A 3K7, Canada; [8]The University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA
*Correspondence: bekim.sadikovic@lhsc.on.ca
https://doi.org/10.1016/j.ajhg.2019.03.008.

the other hand, we and others have shown that many genetic syndromes exhibit unique combinations of DNA methylation changes at multiple loci across the genome, i.e., epi-signatures, which can be used to resolve ambiguous clinical cases and classify those carrying VUSs in relevant genes.[10–14] Since both epi-variants and epi-signatures reflect a fundamental defect, they are detectable to variable extents across multiple tissues, including peripheral blood, which is easily accessible and remains a common source of clinical diagnostic specimens.[9,15]

In the present study, using genome-wide DNA methylation analysis, we develop a computational model for the concurrent identification of 14 ND/CA syndromes currently known to be associated with an epi-signature. Using this model, we demonstrate the ability to resolve 67 subjects with uncertain clinical diagnoses, some of whom had VUSs in clinically relevant genes. We show that the provisional diagnoses can be ruled out for most of these subjects, a few of whom are shown by our model to have other diseases initially not considered. Application of this model to a cohort of 965 ND/CA-affected case subjects with no diagnosis despite a previously extensive genetic work-up enables the identification of 15 case subjects with one of the above syndromes. We also identify 12 individuals with imprinting and trinucleotide repeat expansion disorders, as well as 106 individuals with rare epi-variants, a subset of which occurs at genes with a clinical/functional link to the subjects' phenotypes.

Subjects and Methods

Subjects and Cohorts

To design a computational model for the detection of congenital syndromes, peripheral blood DNA samples from subjects diagnosed with 14 syndromes known to date to harbor a DNA methylation epi-signature in the blood were collected. These included individuals with autosomal-dominant cerebellar ataxia with deafness and narcolepsy (ADCADN [MIM: 604121]), ATRX syndrome (MIM: 300032), BAFopathies (Coffin-Siris [MIM: 614607, 135900, 618027, 617808, 616938, 614609, and 614608], Nicolaides-Baraitser [MIM: 601358], and Chr6q25del [MIM: 612863] syndromes), CHARGE syndrome (MIM: 214800), Claes-Jensen syndrome (MIM: 300534), genitopatellar syndrome (GTPTS [MIM: 606170]), Floating-Harbor syndrome (FHS [MIM: 136140]), Kabuki syndrome (MIM: 147920), and Sotos syndrome (MIM: 117550), all of which were described in our previous studies[10,11,13,14,16–18] and were originally recruited from the Greenwood Genetic Center (Greenwood, SC, USA), Care4Rare Canada Consortium (Ottawa, ON, Canada), and London Health Sciences Center (London, ON, Canada). To this cohort, we added subjects with Down syndrome (MIM: 190685), Cornelia de Lange syndrome (CdLS [MIM: 122470, 614701, 610759, 300590, and 300882]), and ADNP syndrome (MIM: 615873), which were not previously published, obtained from McMaster University Medical Centre (Hamilton, ON, Canada) and Greenwood Genetic Center. Samples with Williams syndrome (MIM: 194050) and Chr7q11.23 duplication (Dup7 [MIM: 609757]) syndromes were downloaded from gene expression omnibus (GEO: GSE66552).[15,19] We supplemented the cohort of subjects with CHARGE syndrome, Sotos syndrome, Kabuki syndrome, and Down syndrome with publicly available DNA methylation data from GEO (GSE74432, GSE97362, GSE116300, GSE52588).[11,15,20,21]

Samples from other ND/CA disorders which were used only to measure the specificity of the classification model designed later, including those with Angelman syndrome (MIM: 105830), Prader-Willi syndrome (MIM: 176270), Beckwith-Wiedemann syndrome (MIM: 130650), Coffin-Lowry syndrome (MIM: 303600), Rett syndrome (MIM: 312750), Saethre-Chotzen syndrome (MIM: 101400), Fragile X syndrome (MIM: 300624), autism spectrum disorders, and RASopathies, were described by our previous studies[6,7,11,13,14] which were collectively obtained from the Greenwood Genetic Center. DNA methylation data files from subjects with Silver-Russell syndrome (MIM: 180860) and Weaver syndrome (MIM: 277590) were obtained from GEO (GSE104451, GSE55491, GSE74432).[22,23]

Any subject used herein to represent a genetic condition had a confirmed clinical diagnosis of the aforementioned syndrome and was screened for mutations in the related genes. The mutation report for every subject was reviewed according to the American College of Medical Genetics (ACMG) guidelines for interpretation of genomic sequence variants,[24] and only individuals confirmed to carry a pathogenic or likely pathogenic mutation and to have the clinical diagnoses were used to represent a syndrome.

Controls comprised healthy subjects with no developmental delay, intellectual disability, or congenital anomalies. The first set of controls that were used for feature selection and model training was collected from the Greenwood Genetic Center, the reference cohort in the LHSC laboratory, and two publicly available databases of healthy pediatric cohorts (GEO: GSE104812, GSE97362).[21,25] Control subjects that were later added to measure the specificity of our classifier and to map the rare DNA methylation epi-variants were compiled from five large databases of general population samples with various age distributions and racial backgrounds (GEO: GSE42861, GSE85210, GSE87571, GSE87648, and GSE99863).[26–29]

Samples with uncertain diagnoses as well as unsolved case subjects that were screened in this study for the detection of potentially affected individuals were collected from all of the sources above over a period of 4 years. These samples were supplemented with publicly available DNA methylation files from GEO for case subjects suspected to have Kabuki syndrome (GEO: GSE116300)[20] as well as a cohort of unsolved ND/CA-affected subjects (GEO: GSE89353).[9]

Methylation Experiment and Quality Controls

Peripheral whole-blood DNA was extracted using standard techniques. Following bisulfite conversion, DNA methylation analysis of the samples was performed using the Illumina Infinium methylation 450k or EPIC bead chip arrays, according to the manufacturer's protocol. This array covers above 860,000 human genomic methylation CpG sites, including 99% of RefSeq genes, all of the known disease-associated imprinted loci in humans, and 96% of CpG islands. The resulting methylated and unmethylated signal intensity data were imported into R 3.5.1 for analysis. Normalization was performed according to the Illumina normalization method with background correction using the minfi package.[30] Probes with detection p values > 0.01, those located on chromosomes X and Y, those known to contain a SNP at the CpG interrogation or single nucleotide extension, and probes known

to cross-react with chromosomal locations other than their target regions were removed. Arrays with more than 5% failure probe rates were excluded from the analysis. All of the samples were examined for genome-wide methylation density, and those deviating from a bimodal distribution were excluded. Factor analysis using a principal component analysis (PCA) was performed to examine the batch effect and identify the outliers. Since samples were assayed using two different platforms (450k and EPIC), after normalization and quality controls the downstream analyses were restricted to the probes shared across the two array types to maintain consistency in the computational workflow.

Selection of Case Subjects and Matched Control Subjects

We selected a random 75% subset of all case subjects with syndromes known to have DNA methylation epi-signatures as a training cohort for the purpose of feature selection and model training. The remaining 25% was set aside as a testing dataset to be used for the assessment of the performance of the classification model developed later. It was ensured that all of the syndromes and subtypes are equally represented in both of the training and testing cohorts. No division to training/testing cohort was performed for conditions with low sample sizes including ADCADN ($n = 5$) and genitopatellar syndrome ($n = 3$), and the samples from these diseases were only used for training. For every syndrome in the training cohort, an age-matched group of control subjects was selected using the MatchIt package. In the case of BAFopathies and CHARGE syndrome, in which the samples were processed using two different platforms (450k and EPIC), control subjects were also matched to the case subjects by the array type to avoid any experimental bias during the feature selection. The sample size of the controls was increased until both the matching quality and sample size were at their optimum and consistent across all diseases. This led to the determination of a control sample size four times larger than the case group in every comparison. Increasing the sample size beyond this value impaired the matching quality. After every matching trial, a PCA was performed to detect outliers and examine the data structures. Outlier samples and those with aberrant data structures were removed before a second matching trial was conducted. The iteration was repeated until no outlier sample was detected in the first two components of the PCA.

Feature Selection for Model Training

The analysis was performed according to our previously published protocol.[11,13,14,31] The methylation level for each probe was measured as a beta value, calculated from the ratio of the methylated signal intensity versus the total sum of unmethylated and methylated signal intensities for that probe, ranging between 0 (no methylation) and 1 (full methylation). Feature selection was performed in two steps separately for every disease. First, a multivariate limma regression modeling[32] was conducted to prioritize the differentially methylated probes that were not confounded by blood cell type compositions. The estimation of blood cell mixture was performed according to the algorithm developed by Houseman et al.[33] The estimated values for each cell component were incorporated into the model matrix of the regression analysis as confounding variables. The analysis was restricted to probes showing a minimum methylation difference between the case subjects and control subjects. This value was determined separately for each condition since every disease is associated with a different effect size in methylation change. Using the examination of the volcano plots of the regression modeling, this effect size was determined to be 5% for BAFopathies, CdLS, and CHARGE syndromes, 15% for Down syndrome, ADNP syndrome, ATRX syndrome, Williams syndrome, and Dup7, 20% for GTPTS and ADCADN, 25% for Sotos syndrome, and 10% for others. No probe with less than 5% methylation difference was used since microarray technology is not sensitive enough in detecting this level of variation in DNA methylation. For linear regression modeling, beta values were logit transformed to M-values using the following equation: log2 (beta/(1 − beta)). The obtained p values were moderated using the eBayes function. The CpGs were sorted according to the moderated p values (ascending), and the top 1,000 were retained for feature selection. In the second step, we performed a receiver's operating curve characteristics analysis for every probe and identified those with an area under the curve (AUC) > 0.80. Finally, we measured the pairwise correlations across the probes and removed the highly correlated features with R-squared > 0.80.

Clustering and Dimension Reduction

The selected probes were examined using a hierarchical clustering analysis to ensure their ability in separating the case subjects from control subjects and to examine the similarity between various subtypes of the studied disorders. Hierarchical clustering was performed using Ward's method on Euclidean distance by the gplots package. A dimension reduction on the combination of the probes identified in each comparison was performed using t-distributed stochastic neighbor embedding (t-SNE) to examine the degree of overlap across different syndromes. The analysis was performed using the Rtsne package according to the default parameters to reduce the dimensions of the data to two.[34] Different values for the t-SNE perplexity parameter were explored to ensure the consistency of the findings. No change in the outcome was observed at different trials; therefore, the default perplexity parameter in the Rtsne package (perplexity = 30) was used for the results presented in the manuscript.

Construction of a Multi-class Prediction Model

To design a classification model for the concurrent detection of the syndromes, a multiclass support vector machine (SVM) with linear kernel was trained on the training dataset using all of the selected probes. Training was done using the e1071 R package. Every disease was used in the training as a distinct class, except for BAFopathies in which the clustering analysis indicated that they may not be completely differentiated from each other. Healthy control subjects were considered as an additional class. To determine the best hyperparameter used in linear SVM (cost), and to measure the accuracy of the model, 10-fold cross-validation was performed during the training. In this process, the training set was randomly divided into ten folds. Nine folds were used for training the model and one fold for testing. After repeating this iteration for all of the ten folds, the mean accuracy was calculated, and the hyperparameters with the optimal performance were selected. For every sample, the model was set to generate multiple scores ranging from 0 to 1, representing the confidence of prediction for each class. Conversion of SVM decision values to these scores was done according to the Platt's scaling method.[35] The final model was applied to the testing dataset to ensure the success of the training. The class receiving the greatest score determined the predicted phenotype. Samples with diseases other than those used in the training are expected to be classified as control subjects.

Validation of the Classification Model

We ensured that the constructed model is not sensitive to the batch structure of the methylation experiment by applying it to all of the samples assayed on the same batch as the cases in the training dataset. To confirm that the classifier is not sensitive to the blood cell type compositions, we downloaded methylation data from isolated blood cell populations of healthy individuals from GEO (GSE35069)[36] and supplied them to our model for prediction and examined the degree to which the resulting scores were varied across different blood cell types. Next, the model was applied to the testing cohort (25% subset of the affected case subjects not used for feature selection or training) to evaluate the predictive ability of the model on affected subjects. To determine the specificity of the model, we supplied a large number of DNA methylation arrays from healthy subjects to the model. To understand whether this model was sensitive to other medical conditions presenting with ND/CA, we tested a large number of subjects with a confirmed clinical and molecular diagnosis of such syndromes by the model.

Screening of Undiagnosed and Uncertain Cases

The final model was used to classify subjects suspected of having any of the conditions used in the training including those with no sequence variant information available, with inconclusive clinical assessment, or with VUS. In addition, we used the model to screen a large group of individuals with various presentations of ND/CA but no established diagnosis despite udergoing routine clinical and molecular assessments including microarray CNV testing or exome sequencing. The subjects that were predicted as having any of the syndromes above were evaluated based on the available clinical and molecular information.

Exome Sequencing and Structural Variant Detection

Subjects with no mutation report, predicted herein to have a defined syndrome, were sequenced using a custom designed next-generation sequencing panel developed using SeqCap EZ MedExome probes covering the coding sequences of ~4,600 genes. These genes are identified to be medically relevant by the consortium of the Emory Genetics Lab, Harvard Laboratory of Molecular Medicine, and Children's Hospital of Philadelphia. The raw sequence data files (FASTQ) were mapped to the GRCh37 (hg19) reference sequence using Burrows-Wheeler aligner with default parameters.[37] The sorted and indexed binary alignment map (BAM) files were used by VarDict for calling single-nucleotide variants and small to moderate size indels.[38] Detection of large deletions, duplications, and insertions was performed using Pindel.[39] Copy number changes in the size of an exon and above were investigated using the ExomeDepth algorithm,[40] which is shown to have outstanding performance in the detection of rare CNVs involved in Mendelian disorders.[41] Variant annotation was performed using Alamut Batch v.1.5.2 (Interactive Biosoftware). To rule out that the identified individuals do not have other conditions explaining their clinical presentations, we did not limit the analysis to the disease candidate genes and evaluated the entire set of the 4,600 genes for a potentially causative variant.

Identification of Rare DNA Methylation Epi-variants

Individuals for whom the conducted analyses did not propose a potential syndrome were screened for a single high-penetrant DNA methylation defect at a single locus with an outlier DNA methylation pattern as compared with a large cohort of controls (epi-variant). First, we used a bump-hunting approach by the bumphunter package[42] to identify regions with ≥ 3 consecutive probes no more than 500 bp apart with a minimum average regional methylation difference of 0.1 between the case subject and the control subjects. Detection of regions on X and Y chromosomes was performed separately using control subjects of the same sex as each case subject. For every identified region, we measured an average ratio of between-group variance to within-group variance (F-statistic). A higher F-statistic indicates a greater deviation of the regional methylation level in the screened individual from a more tightly methylated region in control subjects. Regions with an F-statistic < 20 were excluded. From the remaining regions, those with either an F-statistic > 40 or with a methylation difference > 0.2 were retained. Next, we plotted the methylation levels in every individual against all of the control subjects for each prioritized region. Following a manual review of the plots, we selected regions showing an outlier pattern that was completely absent or found in ≤ 5 of the control subjects. The final list of regions was annotated to genes within 5 kb of the region, gene promoters (defined as 5 kb upstream to 1 kb downstream the transcription start site), and CpG islands, as well as to the OMIM gene and phenotype identifiers. Annotation was performed using the Bioconductor package AnnotationHub (see Web Resources). Using the geneimprint database, we annotated our gene list for an overlap with the known imprinted loci in the *Homo sapiens*. A comprehensive literature search was then performed for every region to identify those with a potential involvement in the clinical presentations of the unresolved individuals.

Technical Validation of Epi-variants

To ensure that the outlier methylation patterns in the identified regions are not due to experimental problems, several of the regions were subject to technical validation using either a repeat microarray experiment of another platform (EPIC versus 450k arrays) or bisulfite sequencing. The latter was performed using NimbleGen SeqCap Epi Target Enrichment probes, custom designed to cover the candidate regions and their 1 kb flanking. The raw sequence data files (FASTQ) were mapped to the GRCh37 (hg19) reference sequence using BSMAP mapper,[43] followed by the trimming of the adaptor sequences. Duplicate sequences were removed, the BAM files were filtered for properly paired reads, and the overlapping reads were clipped. BSMAP was then used to estimate a methylation ratio for every CpG dinucleotide in the target regions. None of the regions selected to be used in this experiment failed the confirmation, indicating the accuracy of microarray in the detection of epi-variants.

Ethics Statement

The study protocol has been approved by the Western University Research Ethics Board (REB 106302), McMaster University, and the Hamilton Integrated Research Ethics Boards (REB 13-653-T). All of the participants provided informed consent prior to sample collection. All of the samples and records were de-identified before any experimental or analytical procedures. The research was conducted in accordance with all relevant ethical regulations.

Results

Development of a Computational Model for Concurrent Classification of Constitutional Syndromes using Genome-wide DNA Methylation Data

We compiled peripheral blood DNA methylation data, generated using Infinium methylation arrays for 383

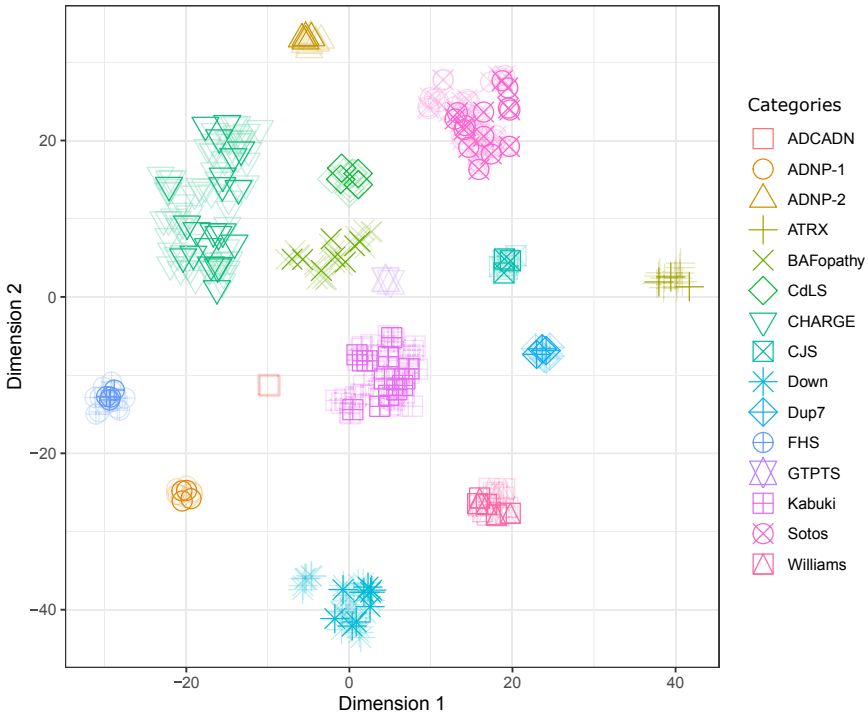

Figure 1. Dimension Reduction of DNA Methylation Data in Constitutional Syndromes

Using t-distributed stochastic neighbor embedding (t-SNE), we reduced the dimension of the DNA methylation data of all of the prioritized features (856 CpGs) in all of the individuals with the constitutional syndromes in the study. The data were reduced to two dimensions as presented on the x and y axes. Every point represents one subject. Subjects in the training cohort who were used for feature selection are shown in lighter colors, and those from the testing cohort who were not used for the prioritization of features are shown in darker colors. The plot shows that every disease generates a distinct cluster from the others and the samples in the testing cohort cluster within the correct disease group. BAFopathies, regardless of the subtype, generate one cluster, while those with ADNP syndrome, depending on the mutation coordinates, generate two distinct groups (ADNP-1 and ADNP-2). The observed pattern could not be explained by the experimental batch structure of the data. Abbreviations: CdLS, Cornelia de Lange syndrome; CJS, Claes-Jensen syndrome; FHS, Floating-Harbor syndrome; GTPTS, genitopatellar syndrome.

subjects who had confirmed clinical and molecular diagnoses of 14 syndromes with previously described epi-signatures, including ADCADN, ADNP, ATRX, BAFopathies, CHARGE, CJS, CdLS, Down, Dup7, FHS, GTPTS, Kabuki, Sotos, and Williams syndromes (Table S1). Following quality controls, ~400k probes (CpG sites) were retained for analysis in each condition. The entire cohort of the affected subjects was randomly divided into two sets of training (75% subset, n = 287) and testing (the remaining 25%, n = 96), ensuring every disease/subtype was equally represented in both subsets. For each disease group in the training cohort, an age-matched cohort of control subjects with a sample size four times larger was selected from a pool of healthy control subjects (n = 565) for the purpose of feature selection and model training. Using moderated p values adjusted for blood cell type compositions (limma multivariate regression modeling), we initially prioritized the top 1,000 differentially methylated probes. Subsequently, this list was narrowed down to the most differentiating features (area under the receiver operating characteristic curve > 0.8) followed by the removal of the redundant probes (pairwise correlation coefficient > 0.8). Hierarchical clustering of the final probes in each disease confirmed that they completely separated the case subjects from control subjects in most of the situations. The only exceptions were noticed for ADNP syndrome and BAFopathies. In ADNP syndrome, subjects with mutations between c.2000 and c.2340 (11 of 23 ADNP-affected case subjects; Table S1) generated an opposite pattern to the rest; whereas in BAFopathies, different subtypes showed overlapping DNA methylation patterns. Thus, subjects with BAFopathies were treated as a single entity, while those with ADNP were divided into two distinct categories (ADNP-1 [those with mutations in the entire *ADNP* gene (MIM: 611386) outside c.2000–2340] and ADNP-2 [those with mutations in *ADNP* between c.2000 and c.2340]). Feature selection for these 15 categories retained on average ~50 probes per condition, the intersections between which were negligible (Table S2). Dimension reduction on the methylation data of all of these probes (n = 856, Table S2) using t-SNE demonstrated that each disorder was distinct from the others (Figure 1). The addition of the testing subset to this analysis revealed that every testing sample clustered with members of the same class and did not mix with others (Figure 1). This indicated that the selected probes were robust in the differentiation of the syndromes, both from each other as well as from control subjects, and thus, it would be possible to generate one model for concurrent classification of all of the syndromes. We trained a multi-class support vector machine (SVM), composed of 16 classes (15 disease and 1 control categories), using all of the prioritized probes (n = 856) and samples in the training set (287 affected case subjects and 565 healthy control subjects). For every subject, the model was set to generate 16 scores ranging 0–1, representing the confidence in predicting whether they have a DNA methylation profile similar to each of the 16 classes in the model. The class obtaining the greatest score determined the predicted phenotype of a subject. 10-fold cross-validation during the training process resulted in an average accuracy of 99.3% (model details in Table S3).

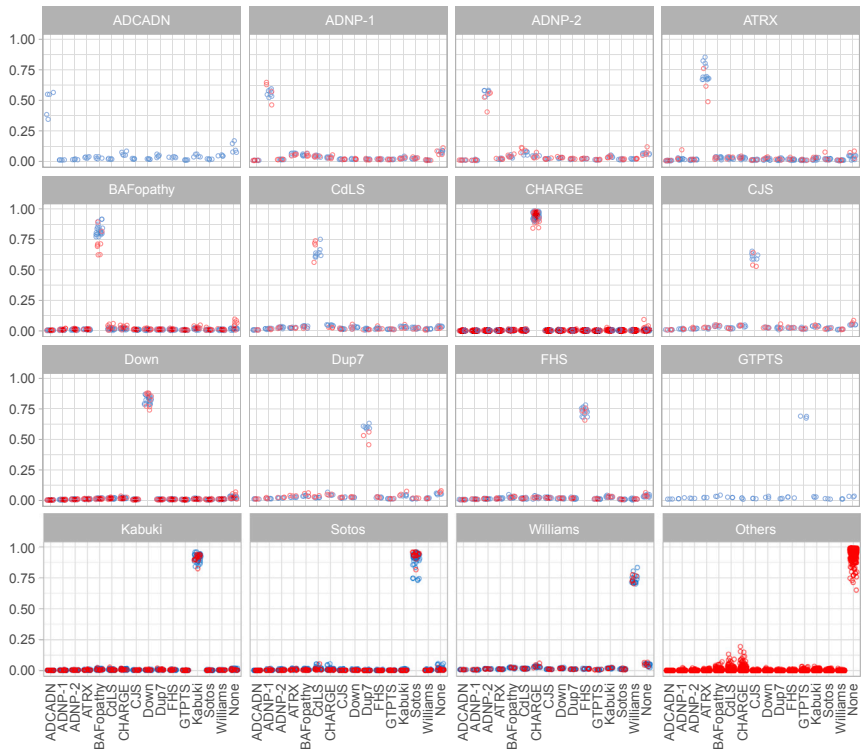

Figure 2. Concurrent Classification of Constitutional Syndromes using DNA Methylation Data

A multi-class SVM classifier generates 16 scores ranging from 0 to 1 for every given individual (y axes), as the level of confidence for having any of the 16 phenotypes for which the model is trained for (x axes). The first 15 classes represent different syndromes and the last one indicates a state that the DNA methylation profiling is not indicative of any of the disorders in the model (healthy control subjects or other diseases). The figure shows trials performed for individuals (represented with hollow points) known to have one of the constitutional syndromes indicated on the x axes, from both the training (blue) and testing (red) cohorts (Table S1). Every panel illustrates trials for a group of subjects with a distinct phenotype as indicated on top of the panel. The first 15 panels show that every affected subject has received the highest score in the correct disease category. Subjects from the testing cohort (red) have also received scores similar to those in the training cohort (blue) with the highest scores in the correct categories. The last panel represents a total of 3,307 subjects composed of 2,880 healthy control subjects and 442 case subjects with confirmed diagnoses of constitutional disorders other than those in the model, including imprinting defect disorders (Angelman syndrome [n = 14], Prader-Willi syndrome [n = 7], Silver-Russell syndrome [n = 64], and Beckwith-Wiedemann syndrome [n = 9]), Weaver syndrome (n = 7), Saethre-Chotzen syndrome (n = 25), Rett syndrome (n = 28), Coffin-Lowry syndrome (n = 11), RASopathies (Noonan syndrome [n = 69], cardiofaciocutaneous syndrome [n = 15], and LEOPARD syndrome [n = 2]), fragile X syndrome (n = 50), and autism spectrum disorders (n = 141). All of these subjects are correctly classified into the final category. Due to small sample sizes, no testing cohort is available for subjects with ADCADN and GTPTS. Abbreviations: CdLS, Cornelia de Lange syndrome; CJS, Claes-Jensen syndrome; FHS, Floating-Harbor syndrome; GTPTS, genitopatellar syndrome.

To control for the success of the procedure, the entire training cohort was supplied to the final model, which assigned correct classifications to all of the case and control subjects used for training (Figure 2). Every sample was classified into the category it belonged to, with scores significantly greater from the others. We were also able to confirm that the model was not sensitive to the batch structure of the data by applying it to other samples processed on the same batch as the case subjects, all of which were classified as control. Additionally, we evaluated the extent to which this model was sensitive to variations in blood cell type compositions by applying it to a total of 60 methylation array data files from 6 healthy individuals, each being assayed separately for whole blood, peripheral blood mononuclear cells, and granulocytes, as well as for 7 isolated cell populations ($CD4^+$ T, $CD8^+$ T, $CD56^+$ NK, $CD19^+$ B, $CD14^+$ monocytes, neutrophils, and eosinophils). All of these samples were classified as controls with <5% average inter-cell-type variability in the scores (Table S4).

We next applied the model to the entire testing cohort, composed of 96 samples that were not used for feature selection or model training. All of these samples were assigned the expected class with scores similar to those of the training dataset, confirming that the model was robust in the classification of the 16 categories (Figure 2). To measure the specificity of our classifier, we tested a total of 2,315 whole-blood methylation data from healthy subjects of various racial backgrounds (aged 0–94 years) obtained from GEO (GSE42861, GSE85210, GSE87571, GSE87648, and GSE99863), all of which were confidently classified as control (Figure 2). We also questioned whether the model could differentiate the above syndromes from other congenital or Mendelian disorders not included in the training. DNA methylation profiles of a total of 442 subjects diagnosed with this type of syndromic condition (details in Figure 2) were supplied to the model for classification, all of which were classified as control, further confirming the specificity of this classifier.

Classification of Subjects with an Uncertain Diagnosis of a Constitutional Syndrome

A common challenge in medical genetics practice involves case subjects presenting with non-specific clinical features or carrying a VUS in a potentially clinically relevant gene. We compiled a total of 67 subjects who, based on clinical or mutational assessments, had been suspected to have one of the syndromes noted above (Table S5 and Figure 3). Our trial utilizing the constructed model revealed that 21

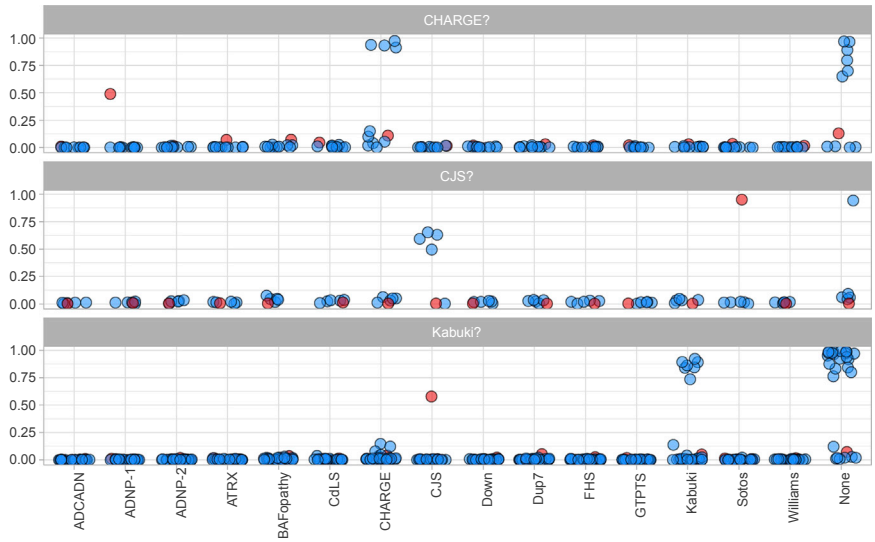

Figure 3. Classification of Uncertain Clinical Case Subjects

The model presented in Figure 2 was used to assess subjects with uncertain clinical/molecular diagnoses. The above figure shows three groups of uncertain case subjects for at least one of whom in every category the prediction was found to be different from what was initially suspected. The first panel shows 11 subjects suspected to have CHARGE syndrome, out of which 5 are classified to have CHARGE syndrome and another 5 are predicted to have none of the syndromes in the model. The final subject (shown in red), however, has received the highest score for the ADNP-1 category and low scores for all others. The second panel represents trials performed for those suspected of having Claes-Jensen syndrome (n = 6) among whom 4 case subjects of CJS and 1 case subject of Sotos syndrome (red) were identified. Similarly, in the last panel, among 26 subjects being assessed for Kabuki syndrome, 1 case of Claes-Jensen syndrome is found in addition to 8 case subjects with Kabuki syndrome (details in Table S5). Abbreviations: CdLS, Cornelia de Lange syndrome; CJS, Claes-Jensen syndrome; FHS, Floating-Harbor syndrome; GTPTS, genitopatellar syndrome.

(31%) of these case subjects showed a methylation profile consistent with the queried syndromes. In many instances, this occurred in individuals for whom a causative sequence variant could not be identified. This included a subject with a confident clinical assessment for CHARGE syndrome but no candidate coding sequence variant in *CHD7* (MIM: 608892), who was found by our model to have a methylation profile consistent with CHARGE syndrome; or three subjects with a clinical consideration of Down syndrome but with normal CNV microarray testing, one of whom was classified by our model as Down syndrome resulting from an apparently mosaic event. The majority of the assessed case subjects, however, did not have the initially assumed disorders. Of interest, few of these case subjects were classified into categories other than their primary suspected phenotypes, including a 9-year-old male with a VUS in *KDM5C* (MIM: 314690, responsible for Claes-Jensen syndrome) who was classified by our model as Sotos syndrome; a person referred for assessment of a VUS in *CHD7* (CHARGE syndrome) who was classified as ADNP syndrome (ADNP-1 subcategory, later confirmed to carry a frameshift mutation within the ADNP-1 region: GenBank: NM_015339.2; c.2491_2494del [p.Leu831Ilefs*82]); and a subject being queried for Kabuki syndrome who was found to be a case of Claes-Jensen syndrome (Figure 3). To support these predictions, wherever possible, secondary clinical and genetic variant assessments were performed in which the results were found to be in compliance (Table S5).

Screening of Unsolved ND/CA-Affected Individuals for Constitutional Syndromes

We assessed whether our model can be used to screen for syndromes with defined epi-signatures in two cohorts of unresolved subjects with ND/CA presentations (total n = 965). The first cohort was composed of 476 subjects, the majority of who had undergone CNV microarray testing as part of the standard clinical workup along with additional genetic testing in some instances, including targeted gene/panel or exome sequencing. The second cohort was, obtained from GEO (GSE89353),[9] comprising 489 subjects with both CNV and exome sequencing assessments. These individuals presented with various forms of ND/CA, including facial dysmorphism, DD/ID, degenerative neural disease, autism, and congenital heart and other organ defects, though none were suspected of having any of the syndromes above, or had a causative genetic etiology identified. Applying our classification model to these subjects identified individuals with DNA methylation patterns consistent with ATRX syndrome (n = 1), ADNP syndrome (n = 2), BAFopathies (n = 2), CdLS (n = 1), CHARGE syndrome (n = 1), CJS (n = 1), GTPTS (n = 1), Kabuki syndrome (n = 4), Sotos syndrome (n = 1), and Williams syndrome (n = 1, Table S6). Individual examination of the DNA methylation patterns of the identified case subjects using disease-specific probes was consistent with the findings (Figure 4). Wherever possible, clinical and molecular assessments were performed to confirm the diagnosis of these subjects (Table S6). DNA sequencing of the candidate genes identified pathogenic variants in a person predicted to have a BAFopathy (Coffin-Siris syndrome, *ARID1B* [MIM: 614556]; GenBank: NM_001346813.1; c.5271del [p.Lys1718Argfs*48]) and in another classified as Kabuki syndrome (*KMT2D* [MIM: 602113]; GenBank: NM_003482.3; c.13890_13894del [p.Pro4631Ilefs*14]). In a second subject predicted to have Kabuki syndrome, a VUS was found two base-pairs upstream of the fourth exon-intron junction (*KMT2D*; GenBank: NM_003482.3; c.509A>G [p.Gln170Arg]). Clinical assessments for all of these subjects were found to be consistent with the

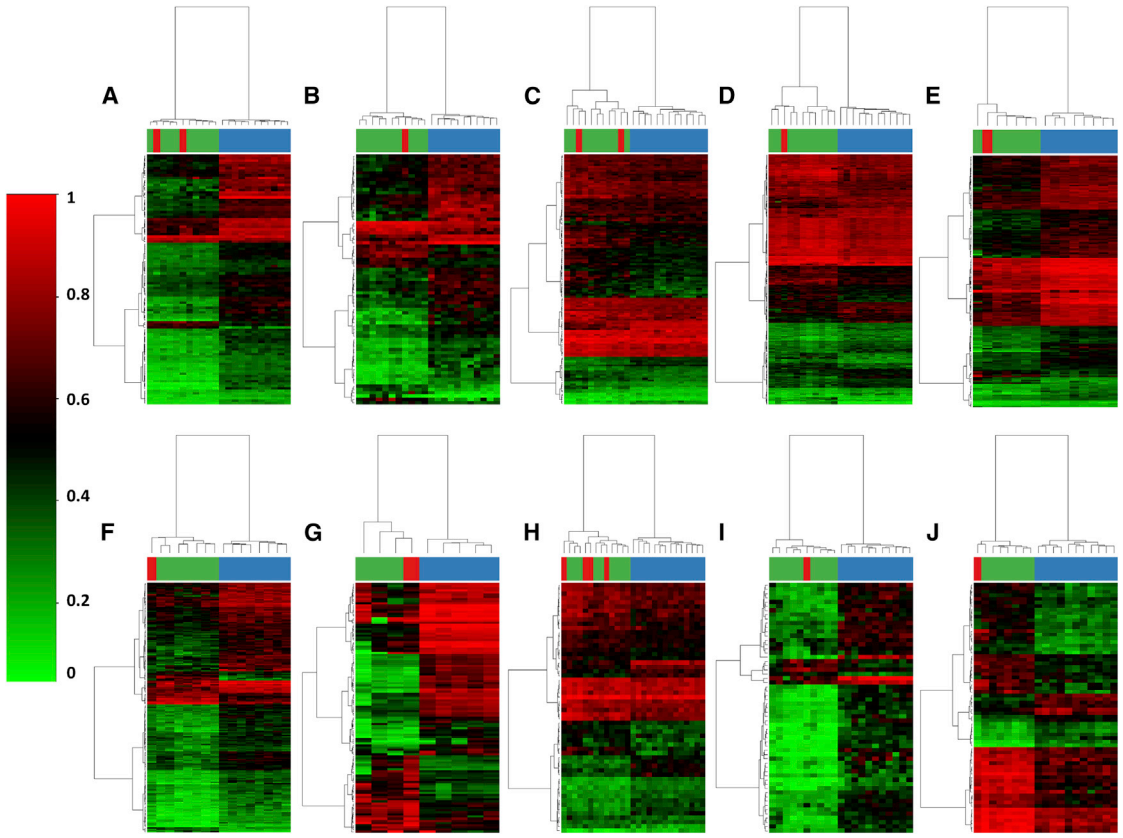

Figure 4. Examination of DNA Methylation Profiles in Unresolved Case Subjects
Screening of 965 unresolved ND/CA-affected case subjects using the model presented in Figure 2 identified a total of 15 subjects (Table S6) as potential cases of 10 of the syndromes in the model. Using the CpG probes specific to every condition, we compared the DNA methylation profiles of the identified subjects to a few representative diagnosed case subjects and healthy control subjects. The analysis was done using hierarchical clustering by Ward's method on Euclidian distance. In every panel above, rows represent the CpG probes and columns represent the subjects. The heatmap panes indicate the phenotype. Blue, healthy control subjects; green, individuals with the confirmed diagnosis of the syndromes; red, the candidate subject(s) under assessment. The heatmap color scale from green to red represents the range of the methylation levels (beta values) between 0 and 1. The analysis indicates that every candidate case subject presents a DNA methylation pattern consistent with the predicted phenotype, i.e., ADNP (A), ATRX (B), BAFopathies (C), CHARGE (D), Cornelia de Lang (E), Claes-Jensen (F), genitopatellar (G), Kabuki (H), Sotos (I), and Williams (J) syndromes.

predictions, including those for whom a pathogenic variant was not identified in sequence assessment (Table S6). The only exception was the subject predicted to have a methylation profile similar to Williams syndrome. This syndrome is caused by contiguous gene deletions at Chr7q11.23 (>1.5 Mb), a segment containing ~28 genes, and variable length deletion can result in phenotypic variations. It is not known which of these genes has the prominent role in establishing the DNA methylation profile of this syndrome. Our case was a 19-year-old male with intellectual disability, atrial septal defect, and dysplastic tricuspid valve without any facial dysmorphism. He did not present the typical features of the Williams syndrome including distinct facial features and connective tissue and endocrine abnormalities. He also did not have supravalvular aortic stenosis, the typical congenital heart malformation associated with Williams syndrome due to the deletion of the elastin gene.[44–47] Further assessment is needed to investigate the presence of a genetic defect within the Chr7q11.23 region.

Detection of Imprinting Defects and Trinucleotide Repeat Expansion Conditions using DNA Methylation Data

We have previously demonstrated that Infinium arrays can be used to detect DNA methylation abnormalities in imprinting defect disorders.[6] Imprinted regions present a hemi-methylated pattern in healthy individuals, whereas in the affected persons they show a near complete gain or loss of methylation. For every unresolved subject, we investigated genomic segments of ≥ 3 consecutive probes showing an outlier methylation pattern relative to 2,880 healthy control subjects (all of the healthy cohorts used in various stages of this manuscript). The identified regions were mapped against the known imprinted loci in humans. This led to the identification of two subjects with abnormal hypermethylation in the promoter of *SNRPN* (MIM: 182279), an indication of Prader-Willi syndrome (PWS) as well as one subject with hypomethylation in the same region, consistent with the diagnosis of Angelman syndrome (AS). The latter case subject had originally

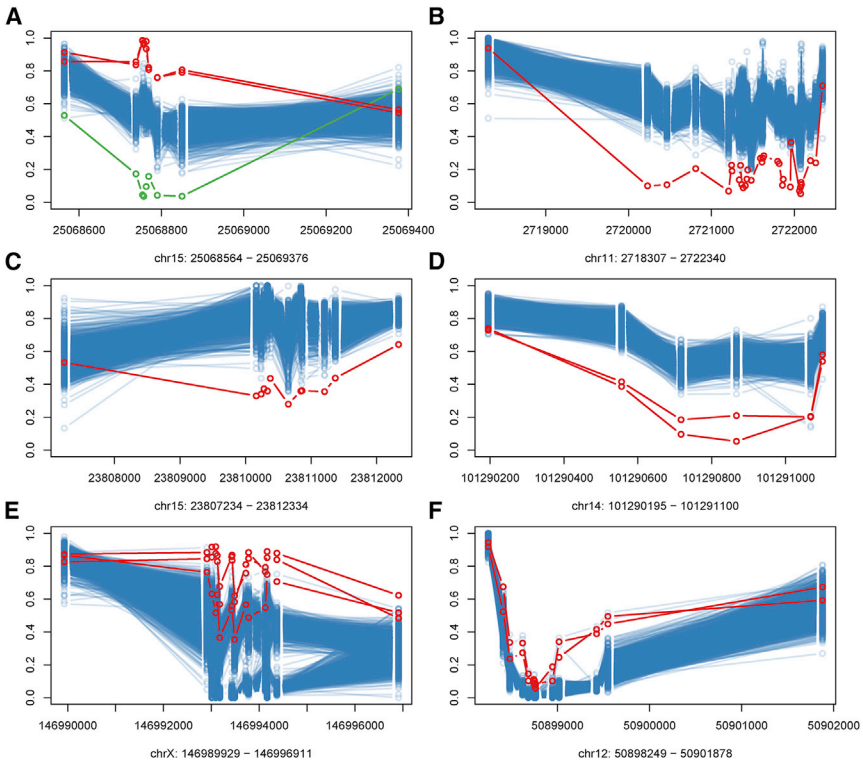

Figure 5. Detection of Imprinting Defects and Trinucleotide Repeat Expansion Disorders using DNA Methylation Analysis
The figure shows the DNA methylation pattern of four imprinted regions (A–D) and two coordinates with trinucleotide repeat elements (E and F) in 2,880 healthy individuals (blue) along with subjects identified with an imprinting or a trinucleotide repeat disorder (red and green). The x axis indicates the genomic coordinates and the y axis represents the DNA methylation levels between 0 and 1. Circles indicate the DNA methylation level for every individual at one CpG site. Neighboring CpG sites are connected with a line.
(A) Two subjects with hypermethylation (red) and one subject with hypomethylation (green) in the promoter of *SNRPN* are shown, indicating a deviation from the normal imprinting pattern (hemi-methylation) in Prader-Willi and Angelman syndromes, respectively.
(B) An extensive hypomethylation in the promoter of the imprinted gene *KCNQ1OT1*, indicating Beckwith-Wiedemann syndrome.
(C) Hypomethylation in the imprinted region mapping to the promoter of *MKRN3*.
(D) Hypomethylation of the promoter of *MEG3* in two subjects.
(E) Healthy males and females in the promoter of *FMR1* show a hypo- and hemi-methylated pattern (blue), respectively. Three male subjects are shown with a hemi-methylation pattern (n = 1, mosaic event), and a full hypermethylated pattern (n = 2), an indication for the expansion of the CGG repeat beyond the normal count (fragile X syndrome).
(F) Two individuals show increased methylation in the promoter of *DIP2B*, indicating CGG repeat expansion associated with mental retardation, FRA12A type. The same increased DNA methylation pattern is also observed in two apparently healthy subjects. Notably, the extended repeat in this region is not 100% penetrant.[49]

been assumed to have Rett syndrome, a known differential diagnosis of AS, based on the finding of a variant in *MeCP2* (MIM: 300005, IVS2-19delT) at age 3 years. However, her clinical course was different from Rett syndrome when re-assessed at 18 years, and the updated finding changed her ultimate clinical diagnosis. Subsequent CNV assessment confirmed copy number alterations at 15q11.2 in all of these case subjects. In another individual with global developmental delay, autism, and ptosis, we identified a hypomethylated region in the promoter of *MKRN3* (MIM: 603856), a maternally imprinted locus located within the AS/PWS region; however, this person did not have an abnormal DNA methylation pattern in the promoter of *SNRPN*. Deletions in the paternal chromosome 15 that do not include *MKRN3* have been documented in PWS, indicating that the loss of imprinting in this gene is not required for PWS.[48] However, a single case subject with a similar clinical presentation to our subject, including DD, obesity, mild infantile hemiplegia, facial dysmorphism, and mild ptosis has been reported in the literature whose deletion overlaps *MKRN3*, but similar to our case subject she did not have PWS or abnormal DNA methylation in *SNURF/SNRPN*.[48] This suggested the abnormal imprinting of the *MKRN3* promoter as a possible causative event in our case subject. We also identified a subject with abnormal hypomethylation in *KCNQ1OT1* (MIM: 604115), indicating a diagnosis of Beckwith-Wiedemann syndrome, and two subjects with hypomethylation in the 5′ UTR of *MEG3* (MIM: 605636), a paternally imprinted region where the loss of the imprinted allele is known to cause Temple (MIM: 616222) syndrome (Table S6, Figure 5).

A different cause of ND/CA that can be identified through DNA methylation analysis is the DNA methylation defects associated with the pathological expansion of trinucleotide repeats.[7] In cases where CpG dinucleotides are part of the repeat element, these expansions can result in an abnormal methylation pattern in the region and, if located in the promoter, can also lead to transcription disruption. We examined genes known to have such pathogenic repeat expansions in congenital disorders.[50] We identified three male subjects with excessive hypermethylation of the *FMR1* (MIM: 309550) promoter (CGG repeat expansion, Figure 5). All three had clinical features consistent with a diagnosis of fragile X syndrome (Table S6). Two of them had been previously assessed by PCR where no abnormality was detected. Subsequent Southern blotting confirmed the triplet repeat expansion of the *FMR1* promoter in these case subjects. The remaining subject was not available for molecular re-assessment. In addition, we

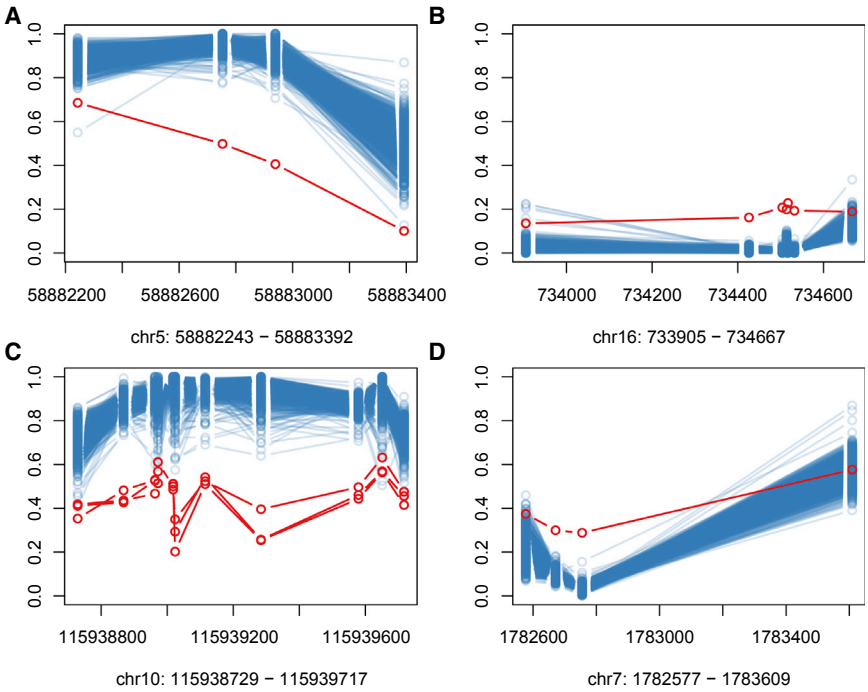

Figure 6. Detection of Rare Epi-variants with Potential Explanation of the Unresolved ND/CA-Affected Case Subjects' Phenotypes

The figure shows the DNA methylation pattern of 4 selected rare epi-variants that are likely explanatory of the phenotypes observed in 6 subjects (red) compared with 2,880 healthy individuals (blue). The x axis represents the genomic coordinates and the y axis shows the DNA methylation levels between 0 and 1. Circles indicate the DNA methylation level for every individual at one CpG site. Neighboring CpG sites are connected with a line.

(A) Hypomethylation in the promoter of *PDE4D*.

(B) Hypermethylation in the tightly unmethylated promoter of *STUB1*.

(C) Three individuals showing a hemi-methylation pattern in the normally hypermethylated promoter of *CDC186*.

(D) Hypermethylation in the promoter of *ELFN1*.

found two case subjects with intellectual disability who had promoter hypermethylation of the *DIP2B* (MIM: 611379) gene (Figure 5, Table S6), in which the CGG-repeat expansion has been shown to be associated with the transcriptional silencing of the brain-expressed gene *DIP2B*, resulting in the neurocognitive problems associated with mental retardation, FRA12A type (MIM: 136630).[49]

Exploration of Rare DNA Methylation Epi-variants in Undiagnosed Individuals

As previously demonstrated,[9] the presence of ND/CA in some case subjects can be attributed to highly penetrant methylation defects at isolated loci across the genome (epi-variants). Our screening of the epigenome for these variants among the unresolved case subjects identified 136 rare epi-variants in 109 individuals showing an outlier methylation pattern compared to 2,880 normal control subjects (Table S7). The coding genes overlapping or proximal to the identified regions were individually reviewed for possible clinical associations with the individuals' clinical presentations (Tables S6 and S7).

Assessment of regions mapping to genes known to cause Mendelian disorders identified a subject with intellectual disability and short stature with a hemi-methylation in the *PDE4D* (MIM: 600129) promoter (Figure 6A). This gene is responsible for acrodysostosis (MIM: 614613), characterized by skeletal abnormality and developmental delay. In the normal population, this gene promoter is always hyper-methylated, potentially downregulating the gene. Consistent with this finding, the mechanism of involvement for *PDE4D* mutations in acrodysostosis is proposed to be a gain-of-function effect.[51] Given that the major phenotype of our case subject, i.e., short stature, is present in nearly all of the case subjects with acrodysostosis, this methylation finding was considered a potential candidate for this case. In another subject with a progressive neurodegenerative disorder with features including ataxia, dysarthria, spasticity, dystonia, and brain atrophy, we identified a gain of methylation in the promoter of *STUB1* (MIM: 607207), a region remaining unmethylated in the normal control subjects (Figure 6B). Loss-of-function mutations in *STUB1* are associated with autosomal-recessive spinocerebellar ataxia (MIM: 615768),[52] which virtually matched the clinical presentations of our subject. In another case subject with neuromuscular symptoms, including hypokinesia, spasticity, and dystonia, we observed a hemi-methylation pattern in the promoter of *NDUFB8* (MIM: 602140), compared to a completely hypomethylated status in all of the control subjects. Loss-of-function mutations in *NDUFB8* result in mitochondrial complex I deficiency (MIM: 618252).[53] Follow-up genetic testing for the evidence of disruption of the second allele in the context of these recessive disorders and the subjects' presentations was not performed, but may have further diagnostic utility in these two latter cases.

The majority of the findings, however, occurred at genes that have not yet been established to cause a Mendelian phenotype. This included six individuals with autism spectrum disorders (ASD) who were identified to have a promoter hypermethylation in genes previously known to be downregulated (i.e., *PLXNA4* [MIM: 604280] and *RASL10B* [MIM: 612128]), to harbor CNV breakpoints (*USP15* [MIM: 604731] and *FAM19A2* [MIM: 617496]), to show CNV amplification (*PLCD3* [MIM: 608795]), or to have *de novo* mutations reported in other ASD-affected case subjects (*PPP2R1B* [MIM: 603113]). We also identified

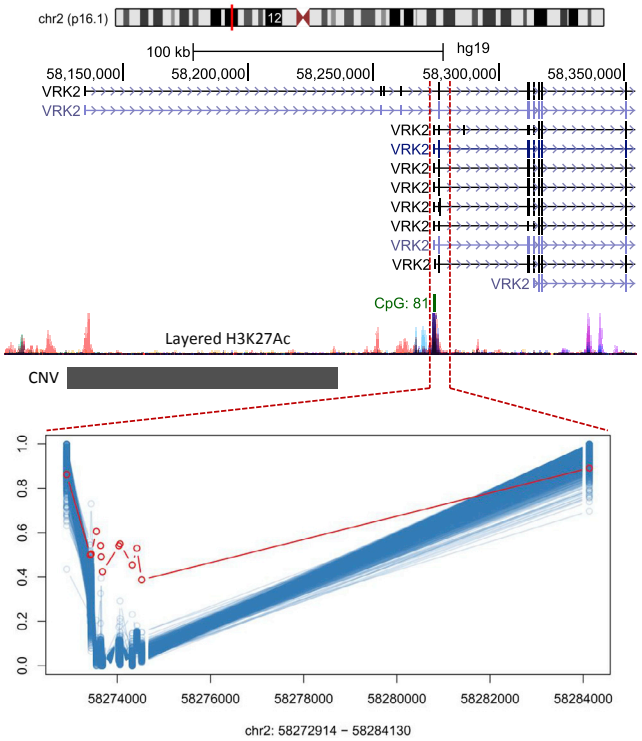

Figure 7. Hypermethylation of the *VRK2* Promoter Secondary to a 30 kb Upstream Deletion

The figure shows a hemi-methylation pattern in a tightly unmethylated CpG island in the promoter of *VRK2*, secondary to a microdeletion ~30 kb upstream. The deleted element is shown to harbor a site enriched for Histone 3 lysine 27 acetylation (H3K27ac), a marker of active regulatory elements. Affected case subjects, red; 2,880 healthy individuals, blue; x axis, genomic coordinate; y axis, DNA methylation levels between 0 and 1; circles, DNA methylation level for every individual at one CpG site.

three unrelated individuals with various presentations of developmental delay and autism, all of whom displayed an extensive hemi-methylation in the promoter of *CCDC186*, a region fully hypermethylated in the healthy population (Figure 6C). A nonsense mutation in this gene (GenBank: NM_018017; c.610C>T [p.Gln204*]) has previously been reported in a case subject with DD/ID,[54] suggesting that the observed event may be involved in their clinical presentations. We also identified a hemi-methylation pattern in the gene body of *MAD1L* (MIM: 602686) in two individuals with ID, which is fully methylated in the normal population. Partial CNV deletions encompassing *MAD1L* are reported in case subjects with DD/ID and congenital abnormalities.[55] Despite recurrences in these subjects, these events were not observed within the 2,880 healthy individuals. Other identified regions that provided an area for exploration of disease etiology included epi-variants in candidate genes for ID (*MTA1* [MIM: 603526]),[56] neurodevelopmental disorders (*PTPRN2* [MIM: 601698]),[57] and neural developments and ID (*PTPRO* [MIM: 600579]),[58] as well as a hypermethylation event in an individual with seizures and ID in the *ELFN1* (MIM: 614964) promoter (Figure 6D), a gene whose knock-out in mice induces seizures and ID.[59] Finally, in an 11-year-old female with DD and seizures, for whom CNV microarray testing had reported a single copy deletion (~113 kb) of uncertain clinical significance in Chr2p16.1, we identified a hemi-methylation status ~30 kb downstream of the deleted segment, at which coordinates all controls were hypomethylated (Figure 7). This region is annotated to the promoter of *VRK2* (MIM: 602162), a gene expressed in the brain in all stages of life, whose variants have been repeatedly identified in genome-wide association studies to be associated with idiopathic generalized epilepsy, as well as other neurological disorders.[60] Consistent with the hypermethylation event observed here, *VRK2* has been found to be downregulated in the blood of individuals with epilepsy as well as in those with schizophrenia, bipolar disorder, and multiple sclerosis.[61,62] Our finding provided a functional measure for clinical interpretation of a CNV change of unknown clinical significance. Further descriptions of the clinically significant epi-variants and the clinical/molecular assessment findings of the carrying unresolved case subjects are provided in Table S6.

Discussion

The clinical genetic testing of developmental delay, intellectual disability and congenital anomalies (ND/CA) began with clinical karyotyping in the late 1950s. Since then, the evolution of cytogenetic technologies including the advent of Giemsa banding, *in situ* hybridization, and microarrays continually increased the diagnostic yield of ND/CA.[63] In 2010, a consensus statement was issued to consider clinical copy number testing using microarray as the first-tier clinical test in individuals with ND/CA, since it resulted in a higher diagnostic rate (15%–20%) compared to G-banded karyotyping alone (3%–5%).[64] The introduction of clinical genome and targeted gene panel sequencing in the past decade has resulted in further improvement in the cumulative diagnostic yield to 42%–62%.[3] However, these assays suffer from several limitations. Targeted approaches such as gene panel sequencing are a powerful tool for rule-out assessments, but their utility in screening is limited to a set of specific genes. Genomic approaches including exome and genome sequencing and cytogenetic microarray analysis enable the screening of broader clinical genetic etiologies; yet, they suffer from analytical sensitivity for all types of genetic variations (screening as opposed to rule-out testing). Both of these approaches suffer from limitations in the interpretability of the detected genetic variants (i.e., VUSs), restricted knowledge in the genetic etiology of a significant proportion of clinical conditions, and technical limitations in the detection of complex or balanced genomic rearrangements. As a result, a substantial proportion of case subjects with ND/CA still remain unsolved.

In this study, we have shown that genomic DNA methylation analysis can be used to assign a diagnosis to

a significant subset of individuals with ND/CA, who may not be identifiable through conventional approaches (Tables S5 and S6). The accumulating reports of DNA methylation epi-signatures in the peripheral blood of constitutional disorders provide grounds for the development of novel methods for clinical screening and assessment. Our results indicate that unresolved ND/CA-affected subjects could have a wide range of conditions, from those relatively common and well-characterized such as Down syndrome to others that are rarer and difficult to recognize clinically such as Coffin-Siris syndrome. Overlapping clinical features with other conditions and non-specific presentations limit the clinical detection of these syndromes. In a large proportion of these cases, sequence analysis does not provide a conclusive result. In 30%–40% of individuals presenting with Coffin-Siris and Kabuki syndromes, for instance, a causative variant is not identified (see GeneReviews in Web Resources).[65] While some of these case subjects might have mutations in yet-to-be identified genes, others carry VUSs or variants not sequenced in routine screening, such as large indels or noncoding variants. In some other situations, the routine methods are not sensitive enough in the detection of a mosaic causative event as noticed in a subject with Down syndrome and another one with fragile X syndrome here. In the case of imprinting disorders, CNV analysis cannot detect a portion of the defects since some of them are caused by an aberrant establishment of imprinting with a normal allelic balance in the region. We demonstrated that DNA methylation profiling provides a powerful solution to these challenges. Our study shows that syndrome-specific DNA methylation epi-signatures can be reliably utilized to resolve ambiguous clinical case subjects, to reclassify those presenting with VUS, to re-assign novel diagnoses for whom the initial clinical assumption is not correct, and to identify cases of a disease through screening of an unresolved population with ND/CA.

In addition to disorders with well-defined DNA methylation signatures, we have also identified rare gene-specific epi-variants among the unresolved ND/CA-affected population (Table S6). These included the relatively common imprinting and trinucleotide repeat expansion disorders, as well as rare and previously unreported epi-variants, some of which overlapped genes with a biological/clinical relevance to the individuals' clinical presentations. However, the majority of the rare epi-variants that we found were not clinically interpretable as they either occurred at genes with no known function related to the disease or in intragenic regions. Our evaluation of the epi-variants was therefore restricted to those occurring at the gene promoters, assuming an inverse association between the gene expression and promoter methylation. It is acknowledged that DNA methylation changes can alter gene expressions through other mechanisms. Gene body methylation changes are proposed to have a direct association with gene expression. Hypermethylation of the repetitive elements in non-genic regions is known to have a role in the stability of the genome and silencing of transposons, the abnormal regulation of which can disrupt the function of other genes. DNA methylation alterations at regulatory elements such as those in enhancers and transcription factor binding sites can theoretically influence the expression of genes both in *cis* and *trans*. However, currently, the above-mentioned mechanisms are challenging to investigate in a clinical setting due to the lack of information on the functional consequence of a methylation change in non-promoter regions.

Adaptation of epi-variant assessment in disease screening would require building clinical databases of epi-variants observed in ND/CAs and consensus guidelines for their interpretation, similar to what is currently available for sequence variants and CNVs. In addition, functional genomic studies would be required to reveal the consequence of epi-variants on the genes whose promoter is not influenced by the event but have the potential to be affected in *trans*. Population epigenome projects can reveal genomic coordinates tightly regulated across the healthy population (i.e., a reference epigenome), pinpointing the genomic regions that are less tolerant of epigenetic variations. Construction of consensus guidelines for interpretation of epi-variants would proceed following the completion of these steps and particularly after the conduct of clinical trials to evaluate cost-effectiveness, feasibility, and diagnostic yield of epigenomic testing in constitutional disorders. Engagement of the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) would be required for the development and standardization of clinical and laboratory guidelines for clinical epigenomic testing. Identified epi-variants may be further investigated based on the allelic involvement, effects on the gene expression, and mode of inheritance. Finally, unlike epi-signatures that due to the genome-wide distribution of their CpG sites are less likely to be fully affected by technical problems, epi-variants warrant technical validations using orthogonal assays before reporting.

Adaptation of epi-signature assessment in the clinical setting, however, would require a different set of considerations. Epi-signatures are composed of a combination of DNA methylation changes across multiple loci, mapping and utilizing which might require compilation of reference methylation databases of diagnosed cases and healthy controls and development of computational tools. Clinical laboratories may choose to participate in lager laboratory networks with shared reference databases and bioinformatics resources similar to the current noninvasive prenatal testing (NIPT) model, or by generating data on-site and accessing centralized bioinformatics and database facilities. It will be important to ensure that such clinical diagnostic services meet performance indicators, which may involve adaptation of proficiency testing models possibly through the College of American Pathologists (CAP). Identification of an epi-signature as part of a clinical screen may warrant development of guidelines for subsequent confirmatory

targeted genetic testing as well as guidelines for the use of epi-signatures for re-classification of genetic VUSs. Recognizing that there will be individuals with disorder-specific epi-signatures but with no identifiable genetic variation in the known disease-related genes, the establishment of genetic diagnosis in the absence of sequence findings will warrant further investigations.

A current limitation of this assay to consider is that it is designed for peripheral blood. It is possible that some disorders do not exhibit epi-signatures or epi-variants in blood. Also, germ-line mosaicism restricted to tissues other than peripheral blood may decrease the clinical sensitivity of this technology in some cases. Further investigations and development of reference datasets in other accessible tissue types such as buccal epithelium or fibroblasts will be needed to extend the utility of this assay from peripheral blood to other tissues.

This study demonstrates that genome-wide DNA methylation testing can be successfully applied to enable molecular diagnosis of genetically unresolved ND/CA-affected case subjects. We propose that genomic DNA methylation assessment has a clinical utility in a broad range of constitutional disorders. The present technical cost of a genomic DNA methylation experiment is comparable to the cost of individual tests currently in use for the detection of imprinting or trinucleotide repeat expansion disorders. Both of these procedures could be replaced at a competitive cost by epigenomic analysis, a single test that provides much more information regarding the underlying genetic etiology of ND/CA. As more syndromes are being mapped for DNA methylation epi-signatures and more clinical DNA methylation data are becoming available for interpretation of rare epi-variants, epigenomic testing is expanding the potentials to significantly improve the current diagnostic yield of the ND/CA.

Data Availability

In addition to the publicly available data sources mentioned in the study, which can be obtained from GEO, DNA methylation microarray data from individuals with BAFopathies have been deposited to GEO with accession number GSE116992. The remaining datasets are not publicly available at the time of the study due to the institutional ethics restrictions.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.03.008.

Acknowledgments

We thank the staff, molecular geneticists, and clinical geneticists at London Health Science Centre, Greenwood Genetic Center, and Care4Rare Canada Consortium for the identification, evaluations, and diagnosis of the individuals with neurodevelopmental conditions presented in this study. We also thank the families of the affected subjects for providing consent and information. We thank Cindy Skinner and Heather Davis at the Greenwood Genetic Center for assisting with the collection of samples for the study. The study was supported in part by a grant from the South Carolina Department of Disabilities and Special Needs (SCDDSN), as well as Illumina Medical Research Grant and McMaster New Investigator Grant awarded to B.S. E.A.-E. was supported by Children's Health Research Institute Epigenetics and Bioinformatics Trainee Award, funded by the Children's Heart Foundation, London, ON, Canada. Dedicated to the memory of Ethan Francis Schwartz, 1996–1998.

Declaration of Interests

The authors declare no competing interests.

Received: November 30, 2018
Accepted: March 5, 2019
Published: March 28, 2019

Web Resources

AnnotationHub, https://bioconductor.org/packages/release/bioc/html/AnnotationHub.html
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
Geneimprint, http://www.geneimprint.com/site/genes-by-species
GeneReviews, Vergano, S.S., Santen, G., Wieczorek, D., Wollnik, B., Matsumoto, N., and Deardorff, M.A. (2016). Coffin-Siris syndrome, https://www.ncbi.nlm.nih.gov/books/NBK131811/
GEO, https://www.ncbi.nlm.nih.gov/geo/
Interactive Biosoftware, https://www.interactive-biosoftware.com
OMIM, http://www.omim.org/

References

1. Rodriguez-Revenga, L., Mila, M., Rosenberg, C., Lamb, A., and Lee, C. (2007). Structural variation in the human genome: the impact of copy number variants on clinical diagnosis. Genet. Med. *9*, 600–606.
2. Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N. Engl. J. Med. *369*, 1502–1511.
3. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., et al. (2014). Genome sequencing identifies major causes of severe intellectual disability. Nature *511*, 344–347.
4. Eisenberger, T., Neuhaus, C., Khan, A.O., Decker, C., Preising, M.N., Friedburg, C., Bieg, A., Gliem, M., Charbel Issa, P., Holz, F.G., et al. (2013). Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. PLoS ONE *8*, e78496.
5. Aref-Eshghi, E., Schenkel, L.C., Carere, D.A., Rodenhiser, D.I., and Sadikovic, B. (2018). Epigenomic Mechanisms of Human Developmental Disorders. In Epigenetics in Human Disease (Second Edition), pp. 837–859.
6. Aref-Eshghi, E., Schenkel, L.C., Lin, H., Skinner, C., Ainsworth, P., Paré, G., Siu, V., Rodenhiser, D., Schwartz, C., and

Sadikovic, B. (2017). Clinical validation of a genome-wide DNA methylation assay for molecular diagnosis of imprinting disorders. J. Mol. Diagn. *19*, 848–856.
7. Schenkel, L.C., Schwartz, C., Skinner, C., Rodenhiser, D.I., Ainsworth, P.J., Pare, G., and Sadikovic, B. (2016). Clinical validation of fragile X syndrome screening by DNA methylation array. J. Mol. Diagn. *18*, 834–841.
8. Banno, K., Kisu, I., Yanokura, M., Tsuji, K., Masuda, K., Ueki, A., Kobayashi, Y., Yamagami, W., Nomura, H., Tominaga, E., et al. (2012). Epimutation and cancer: a new carcinogenic mechanism of Lynch syndrome (Review). Int. J. Oncol. *41*, 793–797.
9. Barbosa, M., Joshi, R.S., Garg, P., Martin-Trujillo, A., Patel, N., Jadhav, B., Watson, C.T., Gibson, W., Chetnik, K., Tessereau, C., et al. (2018). Identification of rare de novo epigenetic variations in congenital disorders. Nat. Commun. *9*, 2064.
10. Schenkel, L.C., Aref-Eshghi, E., Skinner, C., Ainsworth, P., Lin, H., Paré, G., Rodenhiser, D.I., Schwartz, C., and Sadikovic, B. (2018). Peripheral blood epi-signature of Claes-Jensen syndrome enables sensitive and specific identification of patients and healthy carriers with pathogenic mutations in *KDM5C*. Clin. Epigenetics *10*, 21.
11. Aref-Eshghi, E., Bend, E.G., Hood, R.L., Schenkel, L.C., Carere, D.A., Chakrabarti, R., Nagamani, S.C.S., Cheung, S.W., Campeau, P.M., Prasad, C., et al. (2018). BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin-Siris and Nicolaides-Baraitser syndromes. Nat. Commun. *9*, 4885.
12. Jones, M.J., Farré, P., McEwen, L.M., Macisaac, J.L., Watt, K., Neumann, S.M., Emberly, E., Cynader, M.S., Virji-Babul, N., and Kobor, M.S. (2013). Distinct DNA methylation patterns of cognitive impairment and trisomy 21 in Down syndrome. BMC Med. Genomics *6*, 58.
13. Aref-Eshghi, E., Schenkel, L.C., Lin, H., Skinner, C., Ainsworth, P., Paré, G., Rodenhiser, D., Schwartz, C., and Sadikovic, B. (2017). The defining DNA methylation signature of Kabuki syndrome enables functional assessment of genetic variants of unknown clinical significance. Epigenetics *12*, 923–933.
14. Aref-Eshghi, E., Rodenhiser, D.I., Schenkel, L.C., Lin, H., Skinner, C., Ainsworth, P., Paré, G., Hood, R.L., Bulman, D.E., Kernohan, K.D., et al.; Care4Rare Canada Consortium (2018). Genomic DNA methylation signatures enable concurrent diagnosis and clinical genetic variant classification in neurodevelopmental syndromes. Am. J. Hum. Genet. *102*, 156–174.
15. Choufani, S., Cytrynbaum, C., Chung, B.H.Y., Turinsky, A.L., Grafodatskaya, D., Chen, Y.A., Cohen, A.S.A., Dupuis, L., Butcher, D.T., Siu, M.T., et al. (2015). NSD1 mutations generate a genome-wide DNA methylation signature. Nat. Commun. *6*, 10207.
16. Kernohan, K.D., Cigana Schenkel, L., Huang, L., Smith, A., Pare, G., Ainsworth, P., Boycott, K.M., Warman-Chardon, J., Sadikovic, B.; and Care4Rare Canada Consortium (2016). Identification of a methylation profile for DNMT1-associated autosomal dominant cerebellar ataxia, deafness, and narcolepsy. Clin. Epigenetics *8*, 91.
17. Schenkel, L.C., Kernohan, K.D., McBride, A., Reina, D., Hodge, A., Ainsworth, P.J., Rodenhiser, D.I., Pare, G., Bérubé, N.G., Skinner, C., et al. (2017). Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome. Epigenetics Chromatin *10*, 10.
18. Hood, R.L., Schenkel, L.C., Nikkel, S.M., Ainsworth, P.J., Pare, G., Boycott, K.M., Bulman, D.E., and Sadikovic, B. (2016). The defining DNA methylation signature of Floating-Harbor Syndrome. Sci. Rep. *6*, 38803.
19. Strong, E., Butcher, D.T., Singhania, R., Mervis, C.B., Morris, C.A., De Carvalho, D., Weksberg, R., and Osborne, L.R. (2015). Symmetrical dose-dependent DNA-methylation profiles in children with deletion or duplication of 7q11.23. Am. J. Hum. Genet. *97*, 216–227.
20. Sobreira, N., Brucato, M., Zhang, L., Ladd-Acosta, C., Ongaco, C., Romm, J., Doheny, K.F., Mingroni-Netto, R.C., Bertola, D., Kim, C.A., et al. (2017). Patients with a Kabuki syndrome phenotype demonstrate DNA methylation abnormalities. Eur. J. Hum. Genet. *25*, 1335–1344.
21. Butcher, D.T., Cytrynbaum, C., Turinsky, A.L., Siu, M.T., Inbar-Feigenberg, M., Mendoza-Londono, R., Chitayat, D., Walker, S., Machado, J., Caluseriu, O., et al. (2017). CHARGE and Kabuki syndromes: gene-specific DNA methylation signatures identify epigenetic mechanisms linking these clinically overlapping conditions. Am. J. Hum. Genet. *100*, 773–788.
22. Muurinen, M., Hannula-Jouppi, K., Reinius, L.E., Söderhäll, C., Merid, S.K., Bergström, A., Melén, E., Pershagen, G., Lipsanen-Nyman, M., Greco, D., and Kere, J. (2017). Hypomethylation of HOXA4 promoter is common in Silver-Russell syndrome and growth restriction and associates with stature in healthy children. Sci. Rep. *7*, 15693.
23. Prickett, A.R., Ishida, M., Böhm, S., Frost, J.M., Puszyk, W., Abu-Amero, S., Stanier, P., Schulz, R., Moore, G.E., and Oakey, R.J. (2015). Genome-wide methylation analysis in Silver-Russell syndrome patients. Hum. Genet. *134*, 317–332.
24. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al.; ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–424.
25. Shi, L., Jiang, F., Ouyang, F., Zhang, J., Wang, Z., and Shen, X. (2018). DNA methylation markers in combination with skeletal and dental ages to improve age estimation in children. Forensic Sci. Int. Genet. *33*, 1–9.
26. Kular, L., Liu, Y., Ruhrmann, S., Zheleznyakova, G., Marabita, F., Gomez-Cabrero, D., James, T., Ewing, E., Lindén, M., Górnikiewicz, B., et al. (2018). DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. Nat. Commun. *9*, 2397.
27. Su, D., Wang, X., Campbell, M.R., Porter, D.K., Pittman, G.S., Bennett, B.D., Wan, M., Englert, N.A., Crowl, C.L., Gimple, R.N., et al. (2016). Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. PLoS ONE *11*, e0166486.
28. Johansson, A., Enroth, S., and Gyllensten, U. (2013). Continuous aging of the human DNA methylome throughout the human lifespan. PLoS ONE *8*, e67378.
29. Ventham, N.T., Kennedy, N.A., Adams, A.T., Kalla, R., Heath, S., O'Leary, K.R., Drummond, H., Wilson, D.C., Gut, I.G., Nimmo, E.R., Satsangi, J.; IBD BIOM consortium; and IBD CHARACTER consortium (2016). Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. Nat. Commun. *7*, 13507.

30. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369.
31. Aref-Eshghi, E., Schenkel, L.C., Ainsworth, P., Lin, H., Rodenhiser, D.I., Cutz, J.C., and Sadikovic, B. (2018). Genomic DNA methylation-derived algorithm enables accurate detection of malignant prostate tissues. Front. Oncol. *8*, 100.
32. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–e47.
33. Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics *13*, 86.
34. Maaten, L.V.D., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.
35. Platt, J.C. (2000). Probabilities for support vector machines. In Advances in Large Margin Classifiers, A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, eds. (Cambridge: MIT Press), pp. 61–74.
36. Reinius, L.E., Acevedo, N., Joerink, M., Pershagen, G., Dahlén, S.E., Greco, D., Söderhäll, C., Scheynius, A., and Kere, J. (2012). Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS ONE *7*, e41361.
37. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, 1303.3997.
38. Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., McEwen, R., Johnson, J., Dougherty, B., Barrett, J.C., and Dry, J.R. (2016). VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. *44*, e108–e108.
39. Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*, 2865–2871.
40. Plagnol, V., Curtis, J., Epstein, M., Mok, K.Y., Stebbings, E., Grigoriadou, S., Wood, N.W., Hambleton, S., Burns, S.O., Thrasher, A.J., et al. (2012). A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics *28*, 2747–2754.
41. Ellingford, J.M., Campbell, C., Barton, S., Bhaskar, S., Gupta, S., Taylor, R.L., Sergouniotis, P.I., Horn, B., Lamb, J.A., Michaelides, M., et al. (2017). Validation of copy number variation analysis for next-generation sequencing diagnostics. Eur. J. Hum. Genet. *25*, 719–724.
42. Jaffe, A.E., Murakami, P., Lee, H., Leek, J.T., Fallin, M.D., Feinberg, A.P., and Irizarry, R.A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int. J. Epidemiol. *41*, 200–209.
43. Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics *10*, 232.
44. Changela, V., Maheshwari, S., and Bhat, M. (2009). Williams syndrome and Ebstein's anomaly: A rare association. Ann. Pediatr. Cardiol. *2*, 165–166.
45. Metcalfe, K., Rucka, A.K., Smoot, L., Hofstadler, G., Tuzler, G., McKeown, P., Siu, V., Rauch, A., Dean, J., Dennis, N., et al. (2000). Elastin: mutational spectrum in supravalvular aortic stenosis. Eur. J. Hum. Genet. *8*, 955–963.
46. Nakamoto, S., Saga, T., and Shinohara, T. (2003). Williams syndrome associated with complete atrioventricular septal defect. Heart *89*, e15–e15.
47. Yuan, S.M. (2017). Congenital heart defects in Williams syndrome. Turk. J. Pediatr. *59*, 225–232.
48. Kanber, D., Giltay, J., Wieczorek, D., Zogel, C., Hochstenbach, R., Caliebe, A., Kuechler, A., Horsthemke, B., and Buiting, K. (2009). A paternal deletion of MKRN3, MAGEL2 and NDN does not result in Prader-Willi syndrome. Eur. J. Hum. Genet. *17*, 582–590.
49. Winnepenninckx, B., Debacker, K., Ramsay, J., Smeets, D., Smits, A., FitzPatrick, D.R., and Kooy, R.F. (2007). CGG-repeat expansion in the DIP2B gene is associated with the fragile site FRA12A on chromosome 12q13.1. Am. J. Hum. Genet. *80*, 221–231.
50. Darling, A.L., and Uversky, V.N. (2017). Intrinsic disorder in proteins with pathogenic repeat expansions. Molecules *22*, 2027.
51. Lee, H., Graham, J.M., Jr., Rimoin, D.L., Lachman, R.S., Krejci, P., Tompson, S.W., Nelson, S.F., Krakow, D., and Cohn, D.H. (2012). Exome sequencing identifies PDE4D mutations in acrodysostosis. Am. J. Hum. Genet. *90*, 746–751.
52. Shi, C.-H., Schisler, J.C., Rubel, C.E., Tan, S., Song, B., McDonough, H., Xu, L., Portbury, A.L., Mao, C.-Y., True, C., et al. (2014). Ataxia and hypogonadism caused by the loss of ubiquitin ligase activity of the U box protein CHIP. Hum. Mol. Genet. *23*, 1013–1024.
53. Piekutowska-Abramczuk, D., Assouline, Z., Mataković, L., Feichtinger, R.G., Koňařiková, E., Jurkiewicz, E., Stawiński, P., Gusic, M., Koller, A., Pollak, A., et al. (2018). NDUFB8 mutations cause mitochondrial complex I deficiency in individuals with Leigh-like encephalomyopathy. Am. J. Hum. Genet. *102*, 460–467.
54. Monies, D., Abouelhoda, M., AlSayed, M., Alhassnan, Z., Alotaibi, M., Kayyali, H., Al-Owain, M., Shah, A., Rahbeeni, Z., Al-Muhaizea, M.A., et al. (2017). The landscape of genetic diseases in Saudi Arabia based on the first 1000 diagnostic panels and exomes. Hum. Genet. *136*, 921–939.
55. Vanzo, R.J., Martin, M.M., Sdano, M.R., Teta, K., Aggarwal, V., and South, S.T. (2014). SNX8: A candidate gene for 7p22 cardiac malformations including tetralogy of fallot. Am. J. Med. Genet. A. *164A*, 554–556.
56. Nevado, J., Mergener, R., Palomares-Bralo, M., Souza, K.R., Vallespín, E., Mena, R., Martínez-Glez, V., Mori, M.Á., Santos, F., García-Miñaur, S., et al. (2014). New microdeletion and microduplication syndromes: A comprehensive review. Genet. Mol. Biol. *37* (1, Suppl), 210–219.
57. Jarick, I., Volckmar, A.L., Pütter, C., Pechlivanis, S., Nguyen, T.T., Dauvermann, M.R., Beck, S., Albayrak, Ö., Scherag, S., Gilsbach, S., et al. (2014). Genome-wide analysis of rare copy number variations reveals PARK2 as a candidate gene for attention-deficit/hyperactivity disorder. Mol. Psychiatry *19*, 115–121.
58. Beltran, P.J., Bixby, J.L., and Masters, B.A. (2003). Expression of PTPRO during mouse development suggests involvement in axonogenesis and differentiation of NT-3 and NGF-dependent neurons. J. Comp. Neurol. *456*, 384–395.
59. Dolan, J., and Mitchell, K.J. (2013). Mutation of Elfn1 in mice causes seizures and hyperactivity. PLoS ONE *8*, e80491.
60. Helbig, I., and Lowenstein, D.H. (2013). Genetics of the epilepsies: where are we and where are we going? Curr. Opin. Neurol. *26*, 179–185.

61. Azimi, T., Ghafouri-Fard, S., Davood Omrani, M., Mazdeh, M., Arsang-Jang, S., Sayad, A., and Taheri, M. (2018). Vaccinia Related Kinase 2 (VRK2) expression in neurological disorders: schizophrenia, epilepsy and multiple sclerosis. Mult. Scler. Relat. Disord. *19*, 15–19.
62. Tesli, M., Wirgenes, K.V., Hughes, T., Bettella, F., Athanasiu, L., Hoseth, E.S., Nerhus, M., Lagerberg, T.V., Steen, N.E., Agartz, I., et al. (2016). VRK2 gene expression in schizophrenia, bipolar disorder and healthy controls. Br. J. Psychiatry *209*, 114–120.
63. Martin, C.L., and Warburton, D. (2015). Detection of chromosomal aberrations in clinical practice: from karyotype to genome sequence. Annu. Rev. Genomics Hum. Genet. *16*, 309–326.
64. Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., et al. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am. J. Hum. Genet. *86*, 749–764.
65. Bögershausen, N., Gatinois, V., Riehmer, V., Kayserili, H., Becker, J., Thoenes, M., Simsek-Kiper, P.Ö., Barat-Houari, M., Elcioglu, N.H., Wieczorek, D., et al. (2016). Mutation update for Kabuki syndrome genes KMT2D and KDM6A and further delineation of X-linked Kabuki syndrome subtype 2. Hum. Mutat. *37*, 847–864.