Episignatures Stratifying Helsmoortel-Van Der Aa Syndrome Show Modest Correlation with Phenotype

Michael S. Breen,[1,2,3,4,18] Paras Garg,[3,18] Lara Tang,[1,2] Danielle Mendonca,[1,2,4] Tess Levy,[1,2] Mafalda Barbosa,[1,4,5] Anne B. Arnett,[6] Evangeline Kurtz-Nelson,[6] Emanuele Agolini,[7] Agatino Battaglia,[8] Andreas G. Chiocchetti,[9] Christine M. Freitag,[9] Alicia Garcia-Alcon,[10] Paola Grammatico,[11] Irva Hertz-Picciotto,[12,13] Yunin Ludena-Rodriguez,[13] Carmen Moreno,[10] Antonio Novelli,[7] Mara Parellada,[10] Giulia Pascolini,[11] Flora Tassone,[12,14] Dorothy E. Grice,[1,2,4,15] Daniele Di Marino,[16] Raphael A. Bernier,[6] Alexander Kolevzon,[1,2,4] Andrew J. Sharp,[3,4] Joseph D. Buxbaum,[1,2,3,4,15,17] Paige M. Siper,[1,2,4,*] and Silvia De Rubeis[1,2,4,15,*]

Summary

Helsmoortel-Van der Aa syndrome (HVDAS) is a neurodevelopmental condition associated with intellectual disability/developmental delay, autism spectrum disorder, and multiple medical comorbidities. HVDAS is caused by mutations in activity-dependent neuroprotective protein (*ADNP*). A recent study identified genome-wide DNA methylation changes in 22 individuals with HVDAS, adding to the group of neurodevelopmental disorders with an epigenetic signature. This methylation signature segregated those with HVDAS into two groups based on the location of the mutations. Here, we conducted an independent study on 24 individuals with HVDAS and replicated the existence of the two mutation-dependent episignatures. To probe whether the two distinct episignatures correlate with clinical outcomes, we used deep behavioral and neurobiological data from two prospective cohorts of individuals with a genetic diagnosis of HVDAS. We found limited phenotypic differences between the two HVDAS-affected groups and no evidence that individuals with more widespread methylation changes are more severely affected. Moreover, in spite of the methylation changes, we observed no profound alterations in the blood transcriptome of individuals with HVDAS. Our data warrant caution in harnessing methylation signatures in HVDAS as a tool for clinical stratification, at least with regard to behavioral phenotypes.

Helsmoortel-Van der Aa syndrome (HVDAS) is an autosomal-dominant neurodevelopmental disorder (NDD) caused by *de novo* mutations in *ADNP* (MIM: 615873).[1] The syndrome was first described in ten individuals by Helsmoortel and colleagues in 2014.[1] The clinical presentation included intellectual disability/developmental delay (ID/DD), autism spectrum disorder (ASD), facial dysmorphisms, hypotonia, and congenital heart disease (CHD).[1] A recent study collating clinical information from medical records for 78 participants across 16 countries has further expanded the clinical spectrum and demonstrated high prevalence of CHD, visual problems, and gastrointestinal problems.[2]

ADNP encodes a multi-functional protein harboring eight zinc fingers, eight low complexity regions, a homeobox domain for the binding to the DNA,[3] a binding site for the heterochromatin protein 1 (HP1),[4] and a nuclear localization sequence[3,5] (Supplemental Notes). In mouse embryonic stem (ES) cells, ADNP interacts with the chromatin remodeler CHD4 and the chromatin architectural protein HP1 to form a complex called ChAHP.[6,7] The ChAHP complex binds to specific DNA regions and represses gene expression by locally rendering chromatin inaccessible.[7] Additionally, the ChAHP modulates the 3D chromatin architecture by antagonizing chromatin looping at CTCF sites.[6] Within the ChAHP, ADNP mediates the binding to specific DNA sites.[6,7] The genes transcriptionally repressed by the ChAHP in ES cells control cell-fate decisions.[7] In fact, *Adnp* null ($Adnp^{-/-}$) ES cells undergo spontaneous differentiation, and neuronal progenitors derived from them express mesodermal markers instead of neural markers.[7] In line with the critical role in cell-fate specification, $Adnp^{-/-}$

[1]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [4]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [5]Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [6]Department of Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA 98195, USA; [7]Laboratory of Medical Genetics Unit, Bambino Gesù Children's Hospital, 00145 Rome, Italy; [8]Department of Developmental Neuroscience, IRCCS "Stella Maris Foundation," 56128 Pisa, Italy; [9]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Autism Research and Intervention Center of Excellence, University Hospital Frankfurt Goethe University, Deutschordenstr. 50, 60528 Frankfurt am Main, Germany; [10]Child and Adolescent Psychiatry Department, Hospital General Universitario Gregorio Marañón, School of Medicine, Universidad Complutense, IiSGM, CIBERSAM, Madrid 28007, Spain; [11]Medical Genetics, Department of Molecular Medicine, Sapienza University, San Camillo-Forlanini Hospital, 00152 Rome, Italy; [12]MIND Institute, School of Medicine, University of California, Davis, Davis, CA 95817, USA; [13]Department of Public Health Sciences, School of Medicine, University of California, Davis, Davis, CA 95616, USA; [14]Department of Biochemistry and Molecular Medicine, University of California, Davis, Davis, CA 95817, USA; [15]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [16]Department of Life and Environmental Sciences, New York-Marche Structural Biology Center (NY-MaSBiC), Polytechnic University of Marche, 60131 Ancona, Italy; [17]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[18]These authors contributed equally to this work
*Correspondence: paige.siper@mssm.edu (P.M.S.), silvia.derubeis@mssm.edu (S.D.R.)
https://doi.org/10.1016/j.ajhg.2020.07.003.

embryos have impaired tissue specification, especially of the neural tube, and die *in utero*.[8,9] Heterozygous mice, on the other hand, are viable but show developmental delays and behavioral deficits.[10–13] Of note, ADNP has been described to have extra-nuclear functions critical for synaptic formation and maturation.[12,14]

Three recent studies—two on an overlapping sample of 22 individuals with HVDAS[15,16] and a third on an additional ten individuals[17]—reported a genome-wide DNA methylation signature associated with this syndrome. This observation adds to the body of evidence showing characteristic patterns of methylation changes in the peripheral blood of individuals with NDDs caused by mutations in chromatin factors, including Kabuki (MIM: 147920),[15,18,19] Claes-Jensen (MIM: 300534),[15,20] Sotos (MIM: 117550),[15,21] Coffin-Siris (MIM: 614607, 135900, 618027, 617808, 616938, 614609, and 614608),[15,22] CHARGE (MIM: 214800),[15,19] and Floating-Harbor syndrome (MIM: 136140).[15,23] Importantly, these episignatures are syndrome specific and might have diagnostic value in guiding the interpretation of genetic variants with unclear pathogenicity.[15,17,18,24] In addition to the signatures found in genetically-defined NDDs, we recently reported an enrichment of epigenetic changes (epimutations) in individuals diagnosed with NDDs and congenital anomalies,[25] ASD,[26] and schizophrenia,[26] suggesting that sporadic epimutations might contribute to the etiology of these disorders. Unlike in other syndromes,[15,18–23] the methylation changes observed in HVDAS clearly stratify mutations into two classes: class I for individuals with mutations located outside a region between nucleotides 2000 and 2340 of the *ADNP* coding sequence (GenBank: NM_015339.4) and class II for individuals with mutations between nucleotides 2000 and 2340, including the recurrent mutation p.Tyr719* (c.2156dupA, c.2157C>A, or c.2157C>G).[16] The two episignatures can be harnessed to predict the syndrome and the specific mutational class,[15–17] but whether they correlate with clinical outcomes or gene-expression changes in blood has not been studied.

To address these questions, we studied 43 individuals with a genetic diagnosis of HVDAS from four cohorts: cohort A, collected as part of the Autism Sequencing Consortium;[27,28] cohort R, collected at the Medical Genetics Laboratory of Ospedale San Camillo-Forlanini[29] or Ospedale Pediatrico Bambino Gesù, Rome, Italy; cohort S, prospectively assessed at the Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai; and cohort W, prospectively assessed at the University of Washington (Supplemental Notes). Mutations in cohort A were identified through research-based whole-exome sequencing[27,28] and validated by Sanger sequencing. Mutations in cohorts R, S, and W were identified and validated in Clinical Laboratory Improvement Amendments (CLIA)-certified laboratories. Participation was approved by the institutional review boards of participating sites. All caregivers provided informed written consent and assent was obtained when appropriate. We analyzed 24 samples (cohorts A, R, and part of S) for methylation analyses, prospective phenotype data from 32 individuals (cohorts S and W) for genotype-phenotype analyses, and 17 samples (part of cohort S) for RNA-seq analyses (Supplemental Notes).

We first performed a genome-wide methylation analysis by using Illumina EPIC 850K methylation arrays on peripheral blood DNA isolated from 24 individuals with HVDAS, 19 unaffected age-matched controls, and 14 unaffected siblings from cohorts S, A, and R (Supplemental Notes). The 24 HVDAS-affected individuals split evenly into class I and class II mutations (Figure 1A, Table S1). The methylation assays were performed in three batches, each including individuals from multiple cohorts and controls (Table S1, Supplemental Notes). We carefully measured the influence of batch on methylation profiles and modeled batch as a potential confounder (Supplemental Notes, Figure S1). We verified correspondence between nominal and inferred sex and age[30] and predicted the fraction of $CD4^+$ T cells, $CD8^+$ T cells, natural killer cells, B lymphocytes, monocytes, and granulocytes in the samples[31] (Supplemental Notes). The chronological and inferred epigenetic ages between individuals with class I and class II mutations, unaffected controls, and unaffected siblings were comparable (Supplemental Notes, Figure S2). Subsequently, principal component analysis (PCA) confirmed the clustering of *ADNP* episignatures into two groups according mutation class (Figure 1B). These two groups were well separated from unaffected age-matched or sibling controls. Specifically, class II mutations were located between nucleotides 2156 and 2317, within the interval defined by Bend and colleagues.[16] Eight class II participants harbored the most common variant found in ADNP, p.Tyr719* (Figure 1A, Table S1; note that another six participants with p.Tyr719* were included in the genotype-phenotype studies). We then performed linear regression using batch, age, sex, predicted blood cell composition, and *ADNP* mutation status as independent variables; we did this separately for class I and class II mutations. We selected probes associated with disease status at 1% FDR and with $\geq$10% mean methylation difference between affected individuals and age-matched and unaffected sibling controls (Supplemental Notes).

For class I individuals, we identified 6,448 autosomal CpGs that were differentially methylated compared to controls (Figure 1C, Table S2, Figure S3A). Of these differentially methylated sites, 4,143 overlapped with the 5,987 previously identified[16] (~69%). Of the 4,143 overlapping CpGs, we observed complete consistency in the direction of effect: 3,974 CpGs hypomethylated in both studies and the remaining hypermethylated in both studies (Figure S3A, Supplemental Notes). The overall difference in methylation between affected individuals and controls for these 4,143 overlapping CpGs was significantly greater compared to unique CpGs from Bend and colleagues[16] (Figure S3A). These sites mapped to the gene promoter

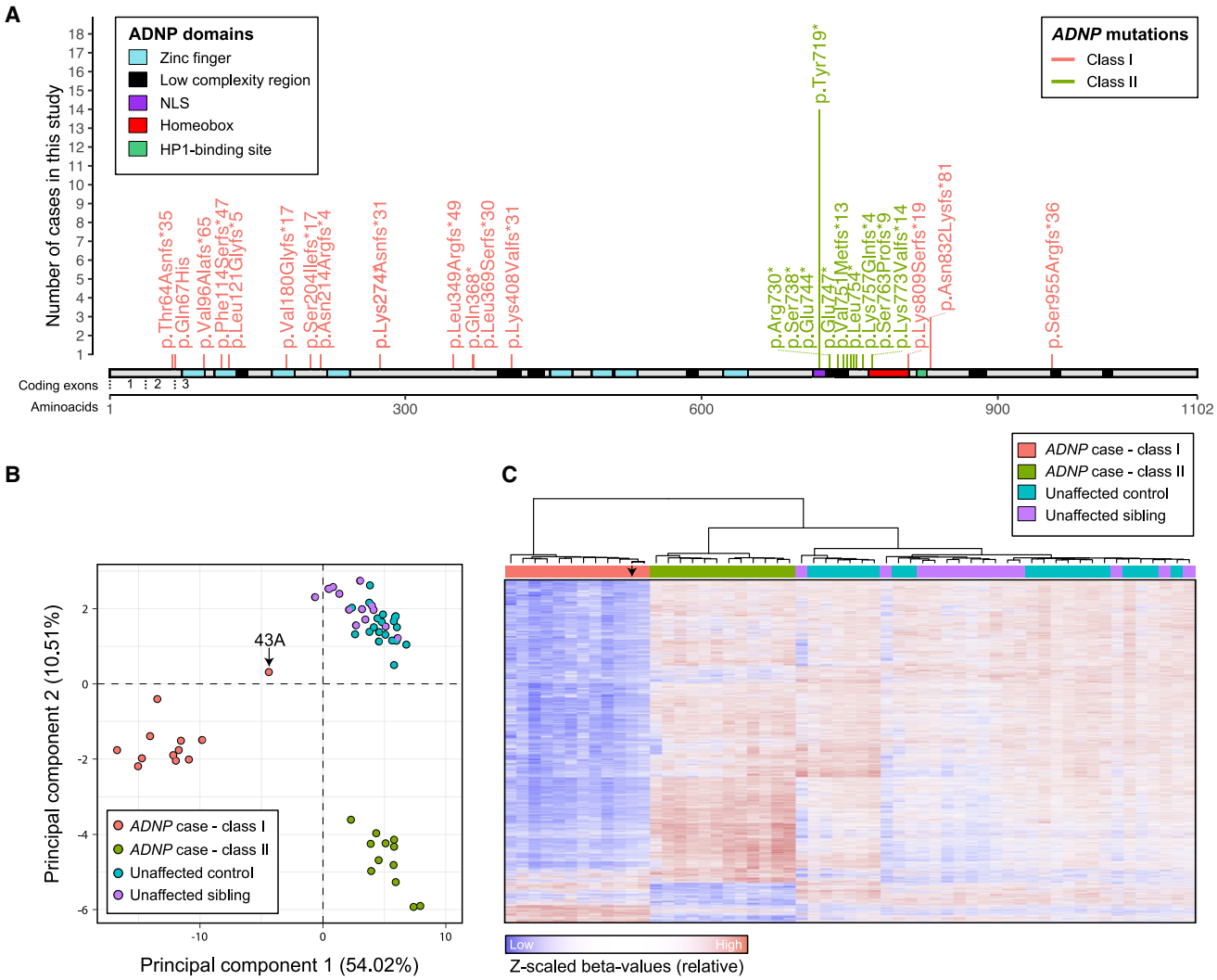


Figure 1. Two Distinct Methylation Signatures in Individuals with HVDAS

(A) Lollipop plot of the *ADNP* mutations in the 42 individuals with point mutations included in the study, including the 24 with methylation data (individual 1S, who carried a deletion, is not shown, Table S1). The amino acid positions are annotated according to ADNP RefSeq protein sequence NP_056154.1. The exon annotation refers to the coding exons in NM_015339.4. The Supplemental Notes contain information on the prediction and re-annotation of linear motifs and structural domains. Mutations in the coding region from nucleotide 2156 to 2317 are classified as class II and are shown in green; mutations outside of this interval are classified as class I and are depicted in pink.

(B) Representation of the two principal dimensions in a principal component analysis of the methylation data from 12 affected individuals in class I, 12 affected individuals in class II, 19 unaffected age-matched controls, and 14 unaffected siblings. The sample indicated by an arrow (43A) harbors the most terminal class I mutation, as shown in (A).

(C) Hierarchical clustering of 12 class I affected individuals, 12 class II affected individuals, 19 unaffected age-matched controls, or 14 unaffected siblings for the 6,448 autosomal CpGs found differentially methylated in individuals with *ADNP* mutations belonging to class I. The arrow points to data for individual 43A, who carried the most terminal class I mutation, as shown in panel A.

(defined here as ±2 kb from the transcriptional start sites) and/or gene body (transcription start to transcription end) of 2,802 autosomal RefSeq genes (Table S2). The distribution of these sites on gene promoters or gene body was also comparable to that found by Bend and colleagues[16] (Supplemental Notes). Differentially methylated genes in class I were enriched in neuronal and synaptic genes: the top three Gene Ontology (GO) processes at FDR < 0.05 were "neuron cell-cell adhesion" (fold enrichment = 5.59), "receptor localization to synapse" (fold enrichment = 3.83), and "neuronal action potential" (fold enrichment = 3.70). This dataset was also enriched in risk genes for ASD[28] (26 genes, $p < 0.0001$), ID/DD[32] (119 genes, $p = 4 \times 10^{-4}$), and CHD (24 genes, $p = 0.005$) (Supplemental Notes).

For class II individuals, we observed a more modest episignature in line with earlier findings[16] with 2,582 differentially methylated autosomal CpGs (Figure 1C, Table S2), 1,007 of which overlapped with the 1,374 previously described[16] (~73%) (Figure S3B, Supplemental Notes). Notably, the difference in methylation between affected individuals and controls for these 1,007 overlapping sites

was larger than for discordant sites (Figure S3B), as we observed for class I (Figure S3A). Strong and consistent direction of effect was observed for the 1,007 overlapping sites (R^2 = 0.91): 771 CpGs were hypermethylated in both studies and the remaining hypomethylated in both studies. Also, the distribution of these sites on gene promoters or gene body was comparable to that found by Bend and colleagues[16] (Supplemental Notes). The 1,442 unique RefSeq genes with altered methylation in the promoter and/or gene body were enriched in genes associated with ID/DD[32] (69 genes, $p = 5 \times 10^{-4}$) or ASD[28] (11 genes, p = 0.013) but not CHD (10 genes, p = 0.13). They were also enriched in neuronal and synaptic genes: the top three GO processes at FDR < 0.05 were "positive regulation of excitatory postsynaptic potential" (fold enrichment = 5.15), "neuromuscular junction development" (fold enrichment = 4.42), and "regulation of synapse assembly" (fold enrichment = 2.95). By selecting clusters comprising three or more differentially methylated CpGs located within 1 kb of each other, we found that class I and class II were associated with 1,398 and 461 differentially methylated regions, respectively (Table S2).

Interestingly, class I and class II episignatures shared 888 differentially methylated probes (1% FDR and with ≥10% mean methylation difference from controls for both classes). However, the direction of change for these CpGs was inversely correlated: there was hypomethylation in class I and hypermethylation in class II (Pearson, $R = -0.31$, $p = 2.2 \times 10^{-16}$) (Figure S4). This inverse correlation pattern was preserved when we repeated the analyses on the 472 CpG sites shared between the two mutations classes and overlapping with those found by Bend and colleagues[16] (Pearson, $R = -0.14$, p = 0.0019) (Figure S4).

Given the striking divergence between the episignatures in class I and class II (Figures 1B and1C, Figure S4) and the clustering of class II mutations (Figure 1A), we hypothesized that the two types of mutations might differ in their functional impact. Most pathogenic or likely pathogenic mutations in *ADNP* are nonsense or frameshift mutations. Most mutations, including 40 out of the 43 in this study, fall in exon 5 (coding exon 3) after the last coding exon-exon junction (Figure 1A) and are predicted to escape nonsense-mediated decay (NMD). A previous study has in fact shown that mutant RNAs can be detected in the blood of individuals with HVDAS[1] and that mutant *ADNP* ectopically expressed in HEK293T cells is translated into truncated proteins that can undergo proteasome-mediated degradation or mislocalize in the cytoplasm.[5] We therefore asked whether class I and class II mutations located in exon 5 (coding exon 3) differ in NMD escape. To test this hypothesis, we performed Sanger sequencing on the cDNA amplified from RNA isolated from peripheral blood of three individuals with class I mutations [p.Phe114Serfs*47 (c.339delC), p.Leu349Argfs*49 (c.1046_1047delTG), and p.Leu369-Serfs*30 (c.1106_1108delTACinsCTGT)] and two individuals with class II mutations [p.Tyr719* (c.2157C>A) and p.Glu747* (c.2239G>T)] (Table S1, Figure S5A). In all samples examined, we observed bi-allelic expression of *ADNP* (i.e., both reference and mutant alleles were expressed), confirming that mutant *ADNP* alleles escape NMD (Figure S5A). We observed no differences between class I and class II mutations, indicating that a differential NMD escape is most likely not the mechanism underlying the divergent episignature. We replicated this observation by examining mutation-mapping reads in genome-wide RNA-seq from 17 individuals with HVDAS (Table S1, Figures S5B–S5D). Notably, all mutations had non-zero RNA-seq read coverage and the *ADNP* locus was expressed at low abundance (fragments per kilobase million [FPKM], 3.03 ± 0.44). When we compared the overall expression of mutant alleles to the expression of reference alleles, we observed lower expression among the mutant reads on average across all samples (p value for difference in log median expression = 0.013; Figure S5D). Our data, alongside previous evidence,[1] indicate that the *ADNP* pathogenic mutations are not leading to classical haploinsufficiency but rather to the expression of mutant proteins that might or might not gain additional functions. Interestingly, one of our affected participants (1S) harbored a deletion extending from the 5′UTR to the second coding exon, suggesting that class I mutations are more likely resulting from a loss of function. Also, one class I participant showed attenuated genome-wide methylation changes (43A; Figures 1B and 1C). Notably, this individual carries the most distal frameshift mutation (Table S1, Figure 1A). Similarly, Bend and colleagues[16] found an individual with a p.Tyr780* (c.2340T>G) mutation having more modest changes compared to other class II mutations. These independent observations suggest that more terminally truncated ADNP mutant proteins might retain partial function associated with an attenuated epigenetic signature. The parsimonious hypothesis that individuals with mild epigenetic signatures (i.e., methylation similar to controls) have milder phenotypic severity remains to be tested.

Genotype-phenotype correlations have begun to emerge in HVDAS. Previous retrospective review of the medical records of 78 participants reported that individuals with p.Tyr719* have a higher rate of high pain thresholds and walked significantly later than individuals with other types of *ADNP* mutations.[2] Individuals with p.Tyr719* or adjacent variants might also have a higher risk of blepharophimosis.[29,33,34] An initial analysis of Bend and colleagues[16] on three class I and six class II individuals did not identify differences between the two groups, although the analysis had minimal power. To further probe potential phenotypic differences and relate them to epigenetic changes, we used extensive, prospectively collected clinical data for two cohorts, cohorts S (class I n = 10, class II n = 12) and W (n = 5/class) (Supplemental Notes). Given the profound methylation changes in the blood of individuals with class I mutations, we initially conjectured that individuals with class I mutations would have more severe clinical manifestations. To test this hypothesis, we contrasted data for 70 variables in the cohorts S and W separately and combined them by calculating individual effect

sizes and both unadjusted and Bonferroni-corrected p values (Table 1, Table S3). Although we did not perform methylation analysis on cohort W, we classified samples of cohort W in class I and class II on the basis of the mutation they presented, as supported by the 100% accuracy of the two episignatures in predicting class mutation.[16] Remarkably, both groups displayed similar frequency of ID, language impairment, ADHD diagnoses, and medical problems. However, we replicated earlier findings[2] of a larger delay in first walking independently in class II individuals (class I = 22 ± 2.67 months, class II = 36.33 ± 6.45 months, d = 2.90, adjusted p = 1.4×10^{-5}) (Table 1). We also observed some evidence for a higher prevalence and severity of ASD in class II (class I = 4/15, class II = 12/17, d = 1.02) (Table 1). There was also a considerable effect size for rate of self-injurious behavior in class II (d = 0.77), but this failed to reach statistical significance (Table S3). Importantly, these differences also hold up when we considered class II without the 11 individuals with a p.Tyr719* variant, indicating that p.Tyr719* is not acting as an outlier that skews the results. Overall, the striking divergent methylation patterns in class I and class II individuals did not translate into robust phenotypic differences between the two groups, but there is evidence for subtler associations.

We next asked whether changes at the gene expression level could be predicted on the basis of differentially methylated CpGs and also cluster HVDAS-affected individuals from controls. PCA was performed on RNA-seq data from 17 individuals with HVDAS and 19 unaffected siblings via genes harboring differentially methylated CpGs within their respective promoters; no distinct stratification was observed between samples of HVDAS-affected individuals and unaffected siblings (Figure S6A). Subsequently, differential gene expression analysis tested for genes that were over- or under-expressed in affected individuals relative to unaffected siblings and identified two genes associated with class I mutations and nine genes associated with class II mutations under an FDR < 0.05; each had small individual effect sizes (Table S2). A competitive gene-set ranking approach was used to functionally annotate genome-wide trends in gene expression. First, we integrated differentially methylated CpGs from gene promoters and observed a poor consistency between significantly hypermethylated and hypomethylated CpGs relative to genes that are under-expressed and over-expressed in HVDAS-affected individuals, respectively (Figures S6B and S6C). Second, we performed GO enrichment analysis by using our previously identified biological processes found to be significantly enriched for differentially methylated CpGs and confirmed a significant overexpression of genes mapping to gated and ion channel activity (Figure S6D). Notably, there was no significant enrichment for ASD, ID/DD, or CHD genes despite a trend of these genes to be overexpressed in affected individuals. Finally, exploratory GO enrichment analysis was performed and significant enrichment for ephrin, SMAD, and PTK signaling was observed in individuals with class I mutations. Processes related to DNA methylation and replication were enriched in individuals with class II mutations (Figures S6D and S6E). Overall, despite the critical role of ADNP on chromatin accessibility and architecture[6,7] and the methylation changes found in individuals with pathogenic *ADNP* mutations, we did not observe profound alterations in corresponding gene expression profiles.

Here, we replicated and expanded earlier findings[15–17] of DNA methylation changes in HVDAS-affected individuals. The mechanisms leading to these methylation changes are still unknown. Although initial observations suggested an association of ADNP with subunits of the chromatin remodeling complex called SWI/SNF,[9] a recent study failed to replicate this finding in mouse ES cells[7] and found instead that ADNP interacts with CHD4 and HP1 to form a complex (ChAHP) that restricts local chromatin accessibility[7] and controls chromatin looping.[6] Defective ChAHP complexes formed by mutant ADNP might potentially change accessibility for DNA methyltransferases and/or demethylases and explain the methylation changes found here. It is still puzzling, however, why mutations affecting residues 719–773 (or 719–780[15,16]) have a different effect than all other mutations. In fact, even within the groups of 888 differentially methylated CpGs shared between the two groups, the direction of change is opposite: there was more hypomethylation in class I and more hypermethylation in class II. This indicates that the two groups of mutant ADNP affect DNA methylation via different mechanisms. The first important observation is that even frameshift and nonsense mutations do not lead to NMD[1] (Figure S4), suggesting that these mutations lead to the expression of truncated proteins. These truncated proteins might act as hypomorphic or gain-of-function alleles. The fact that an individual carrying a deletion extending from the 5′UTR to coding exon 2 clusters with class I mutations (1S) and that the most terminal class I mutation has an attenuated signature [p.Ser955Argfs*36 (c.2865_2868delTGAG) in 43A] would suggest that this group of mutations might act as hypomorphic alleles. Experiments introducing GFP-tagged ADNP mutants in HEK293T cells have suggested a distinct pattern of expression and localization based on the location of the mutation, dividing the mutations into N-terminal mutations (in the first 412 residues), central mutations (in the 473–719 region), and C-terminal mutations (after 719).[5] N-terminal mutants were targeted to proteasomal degradation, expression of central mutations was restricted to the cytoplasm, and expression of C-terminal mutants were expressed in the nucleus but showed reduced co-localization with pericentromeric heterochromatin.[5] These data would suggest that N-terminal and C-terminal mutations reduce protein activity, whereas mutations in the central domain are more detrimental. Although our study and Bend et al. (2019)[16] do not include mutations in the 473–718 range, the study

Table 1. Phenotypic Comparison between the Two HVDAS Groups

Table 1	Class I (n = 15)	Class II (n = 17)	Effect	p Value	Adj. p Value
Sex	10 females, 5 males	5 females, 12 males	0.37	0.080	1
Age (years)	7.33 ± 3.33	7.18 ± 4.56	−0.04	0.910	1
Autism spectrum disorder	4	12	−0.48	0.020	1
ASD severity	4.27 ± 1.83	6.4 ± 2.32	1.02	0.010	0.70
Full scale IQ or developmental quotient (DQ)	36.23 ± 13.82	35.72 ± 17.55	−0.03	0.930	1
Nonverbal IQ or DQ	9	13	−0.23	0.365	1
Verbal IQ or DQ	53 ± 13.11	48.59 ± 12.86	−0.34	0.350	1
Motor delays	14	17	−0.19	0.949	1
Gait abnormalities	13	15	−0.27	0.464	1
First walked independently	22 ± 2.67	36.33 ± 6.45	2.90	2.1e−07	1.4e−5
Language delay	15	17	0.00	1.000	1
Receptive vocabulary (standard score)	60.5 ± 14.7	45.75 ± 20.41	−0.83	0.060	1
Expressive vocabulary (standard score)	53.25 ± 14.9	45.14 ± 18.33	−0.49	0.370	1
Nonverbal or minimally verbal	37.68 ± 14.34	37.97 ± 19.04	0.02	0.960	1
First phrase	80 ± 23.51	60 ± 15.74	−1.00	0.094	1
First single word	38.71 ± 20.53	33.6 ± 12.47	−0.30	0.457	1
Intellectual disability	15	16	0.00	1.000	1
Attention deficit/hyperactivity disorder	6	6	0.03	1.000	1
Visual motor integration (standard score)	58.17 ± 15.99	48.7 ± 11.24	−0.69	0.240	1
Vision problems	10	16	−0.35	0.126	1
Vineland adaptive behavior composite	35.92 ± 14.53	34.33 ± 17.97	−0.10	0.800	1
Sensory symptoms	15	17	0.00	1.000	1
High pain threshold	10	15	−0.3	0.252	1
Seizures	3	7	−0.23	0.364	1
GI problems	11	14	−0.11	0.851	1
Hypothyroidism	4	2	0.18	0.587	1
Hypotonia	11	15	0.01	1.000	1
Sleep disturbance	11	12	−0.02	1.000	1
Obstructive sleep apnea	5	1	0.44	0.185	1
Congenital heart defect	7	5	0.18	0.522	1
Early tooth eruption	8	7	0.27	0.388	1
Ptosis	1	3	−0.22	0.534	1
Hearing problems	6	3	0.25	0.313	1
Neonatal intensive care unit stay at birth	6	8	−0.07	0.964	1
Recurrent infections	6	7	−0.01	1.000	1
Feeding issues	3	9	−0.34	0.120	1
Structural brain changes	3	7	−0.3	0.369	1

Comparison between clinical measures collected prospectively from 15 class I and 17 class II HVDAS-affected individuals in cohorts S and W. Table S3 presents the analysis of additional measures, broken down by cohort. The Supplemental Notes describe all the measures listed (e.g., ASD severity refers to Comparison Score of the Autism Diagnostic Observation Schedule, 2nd Edition). The Shapiro-Wilk test was used to assess normality of all continuous variables, and either a one-way analysis of variance (ANOVA) or Wilcoxon rank sum test was implemented accordingly. Chi-square tests with Yates correction were used to test discrete variables. We adjusted all tests for multiple testing by using the Bonferroni method to control the false discovery rate. Effect sizes were estimated for discrete measures with the phi coefficient and for continuous measures with Cohen's d.

by Aref-Eshghi et al. (2020)[17] includes the p.Ala430Cysfs*10 (c.1287dupT) and p.Cys627Gly variants, both of which are associated with class I episignature. Overall, the methylation data support a model whereby mutations affecting residues 719–780 (and not residues 473–719) behave differently from all others. Also, independent evidence in mouse ES cells homozygous for p.Tyr718* (corresponding to human p.Tyr719*) show that the mutant ADNP reaches the nucleus.[7] Interestingly, this mutant ADNP binds CHD4 and DNA, despite the loss of the homeobox, but cannot interact with HP1.[7] Since HP1 mediates the repression activity of the ChAHP complexes, ADNP target genes are overexpressed in the p.Tyr718* mutant ES cells.[7] Although the analyses from Cappuyins et al.[5] and Ostapcuk et al.[7] begin to explore the functional impact of different *ADNP* mutations, relevance to the clinical syndrome is complicated by the use of overexpression systems[5] or the use of homozygous targeting.[7] Future studies on patient-derived cells, as well as advances in solving 3D structure of ADNP, would be important to better dissect the differential impact of the clinical mutations.

Prior analyses on three class I and six class II individuals failed to detect phenotypic differences between the two groups.[16] Our prospectively collected clinical data on 15 class I and 17 class II individuals show that the two mutational groups differ in the mean age of first walking and rates of ASD (Table 1): both were more severe in the class with less profound methylation changes. The lack of correspondence between molecular signatures, including methylation and gene expression, and clinical manifestations caution against making phenotypic inferences on the basis of the blood-based methylation profile. This is important to consider when evaluating the use of these episignatures as biomarkers for patient stratification and/or response to pharmacological agents in clinical trials. In the specific case of HVDAS, these observations are timely as potential experimental therapeutics emerge. In particular, pre-clinical data suggest that an eight-amino-acid ADNP peptide called NAP (NAPVSIPQ) might hold promises for treatment. NAP has been shown to have broad neuroprotective effects *in vitro* and *in vivo*,[35] and *in vivo* administration of NAP or its derivatives in a mouse model of HVDAS has been shown to ameliorate synaptic and behavioral defects.[10–13] In conclusion, although the episignatures in HVDAS have proven very valuable for the purpose of complementing clinical genetics and enhancing accuracy of diagnosis, their value as biomarkers for clinical stratification in clinical trials should be carefully evaluated.

Data and Code Availability

The methylation and RNA sequencing data are available at the Gene Expression Omnibus under accession number GEO: GSE152428 (GEO: GSE152427 for the methylation data and GEO: GSE152421 for the RNA sequencing data).

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2020.07.003.

Acknowledgments

This work was supported by grants from the Beatrice and Samuel A. Seaver Foundation and the ADNP Kids Research Foundation. M.S.B. is a Seaver Foundation Faculty Scholar. S.D.R. is a Fascitelli Research Scholar. The collection of samples in cohort A and some of the molecular and analytical studies were supported by the NIMH (MH111661). The collection and processing of samples, confirmation of diagnoses, and management of data for part of cohort A were also funded by the NIH (R01ES015359, P01ES011269, and UH3OD023365) and the EPA (STAR R829388, R833292, and RD83543201). The work was also supported by the Spanish Ministry of Science and Innovation, Instituto de Salud Carlos III, and CIBERSAM and co-financed by ERDF funds and European Union Seventh Framework Program, the European Union H2020 Program under the Innovative Medicines Initiative 2 Joint Undertaking (grant agreement 777394, project AIMS-2-TRIALS), and Fundación Familia Alonso. The work was partially supported by funding to D.D.M. by Fondazione Marche. Research reported in this paper was supported by the Office of Research Infrastructure of the National Institutes of Health under award number S10OD018522. This work was supported in part through the computational resources and staff expertise provide by Scientific Computing at the Icahn School of Medicine at Mount Sinai. We are grateful to all the families and individuals who participated in this research.

Declaration of Interests

S.D.R., P.M.S., J.D.B., and A.K. are supported by the ADNP Kids Research Foundation. A.K. receives research support from AMO Pharma and is on the scientific advisory board of Ovid Therapeutics. A.K. has also consulted to Takeda, Acadia, and Sema4 in the past year. M.P. has consulted for Servier and Exeltis. C.M. has acted as consultant or participated in the Data Monitoring Committee for Janssen, Servier, Lundbeck, Nuvelution, Angelini, and Otsuka and has received grant support from European Union Funds and Instituto de Salud Carlos III, Spanish Ministry of Economy and Competitiveness. All other authors declare no competing interests.

Received: March 29, 2020
Accepted: July 7, 2020
Published: August 5, 2020

Web Resources

Online Mendelian Inheritance in Man, https://www.omim.org

References

1. Helsmoortel, C., Vulto-van Silfhout, A.T., Coe, B.P., Vandeweyer, G., Rooms, L., van den Ende, J., Schuurs-Hoeijmakers, J.H., Marcelis, C.L., Willemsen, M.H., Vissers, L.E., et al. (2014). A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. Nat. Genet. *46*, 380–384.

2. Van Dijck, A., Vulto-van Silfhout, A.T., Cappuyns, E., van der Werf, I.M., Mancini, G.M., Tzschach, A., Bernier, R., Gozes, I., Eichler, E.E., Romano, C., et al.; ADNP Consortium (2019). Clinical Presentation of a Complex Neurodevelopmental Disorder Caused by Mutations in ADNP. Biol. Psychiatry *85*, 287–297.
3. Zamostiano, R., Pinhasov, A., Gelber, E., Steingart, R.A., Seroussi, E., Giladi, E., Bassan, M., Wollman, Y., Eyre, H.J., Mulley, J.C., et al. (2001). Cloning and characterization of the human activity-dependent neuroprotective protein. J. Biol. Chem. *276*, 708–714.
4. Mandel, S., Rechavi, G., and Gozes, I. (2007). Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis. Dev. Biol. *303*, 814–824.
5. Cappuyns, E., Huyghebaert, J., Vandeweyer, G., and Kooy, R.F. (2018). Mutations in ADNP affect expression and subcellular localization of the protein. Cell Cycle *17*, 1068–1075.
6. Kaaij, L.J.T., Mohn, F., van der Weide, R.H., de Wit, E., and Buhler, M. (2019). The ChAHP Complex Counteracts Chromatin Looping at CTCF Sites that Emerged from SINE Expansions in Mouse. Cell *178*, 1437–1451.e14.
7. Ostapcuk, V., Mohn, F., Carl, S.H., Basters, A., Hess, D., Iesmantavicius, V., Lampersberger, L., Flemr, M., Pandey, A., Thomä, N.H., et al. (2018). Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. Nature *557*, 739–743.
8. Pinhasov, A., Mandel, S., Torchinsky, A., Giladi, E., Pittel, Z., Goldsweig, A.M., Servoss, S.J., Brenneman, D.E., and Gozes, I. (2003). Activity-dependent neuroprotective protein: a novel gene essential for brain formation. Brain Res. Dev. Brain Res. *144*, 83–90.
9. Mandel, S., and Gozes, I. (2007). Activity-dependent neuroprotective protein constitutes a novel element in the SWI/SNF chromatin remodeling complex. J. Biol. Chem. *282*, 34448–34456.
10. Vulih-Shultzman, I., Pinhasov, A., Mandel, S., Grigoriadis, N., Touloumi, O., Pittel, Z., and Gozes, I. (2007). Activity-dependent neuroprotective protein snippet NAP reduces tau hyperphosphorylation and enhances learning in a novel transgenic mouse model. J. Pharmacol. Exp. Ther. *323*, 438–449.
11. Amram, N., Hacohen-Kleiman, G., Sragovich, S., Malishkevich, A., Katz, J., Touloumi, O., Lagoudaki, R., Grigoriadis, N.C., Giladi, E., Yeheskel, A., et al. (2016). Sexual divergence in microtubule function: the novel intranasal microtubule targeting SKIP normalizes axonal transport and enhances memory. Mol. Psychiatry *21*, 1467–1476.
12. Hacohen-Kleiman, G., Sragovich, S., Karmon, G., Gao, A.Y.L., Grigg, I., Pasmanik-Chor, M., Le, A., Korenková, V., McKinney, R.A., and Gozes, I. (2018). Activity-dependent neuroprotective protein deficiency models synaptic and developmental phenotypes of autism-like syndrome. J. Clin. Invest. *128*, 4956–4969.
13. Sragovich, S., Malishkevich, A., Piontkewitz, Y., Giladi, E., Touloumi, O., Lagoudaki, R., Grigoriadis, N., and Gozes, I. (2019). The autism/neuroprotection-linked ADNP/NAP regulate the excitatory glutamatergic synapse. Transl. Psychiatry *9*, 2.
14. Oz, S., Kapitansky, O., Ivashco-Pachima, Y., Malishkevich, A., Giladi, E., Skalka, N., Rosin-Arbesfeld, R., Mittelman, L., Segev, O., Hirsch, J.A., and Gozes, I. (2014). The NAP motif of activity-dependent neuroprotective protein (ADNP) regulates dendritic spines through microtubule end binding proteins. Mol. Psychiatry *19*, 1115–1124.
15. Aref-Eshghi, E., Bend, E.G., Colaiacovo, S., Caudle, M., Chakrabarti, R., Napier, M., Brick, L., Brady, L., Carere, D.A., Levy, M.A., et al. (2019). Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. Am. J. Hum. Genet. *104*, 685–700.
16. Bend, E.G., Aref-Eshghi, E., Everman, D.B., Rogers, R.C., Cathey, S.S., Prijoles, E.J., Lyons, M.J., Davis, H., Clarkson, K., Gripp, K.W., et al. (2019). Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome. Clin. Epigenetics *11*, 64.
17. Aref-Eshghi, E., Kerkhof, J., Pedro, V.P., Barat-Houari, M., Ruiz-Pallares, N., Andrau, J.C., Lacombe, D., Van-Gils, J., Fergelot, P., Dubourg, C., et al.; Groupe DI France (2020). Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. Am. J. Hum. Genet. *106*, 356–370.
18. 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.
19. 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.
20. 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.
21. Choufani, S., Cytrynbaum, C., Chung, B.H., Turinsky, A.L., Grafodatskaya, D., Chen, Y.A., Cohen, A.S., Dupuis, L., Butcher, D.T., Siu, M.T., et al. (2015). NSD1 mutations generate a genome-wide DNA methylation signature. Nat. Commun. *6*, 10207.
22. 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.
23. 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.
24. 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.
25. 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.
26. Garg, P., and Sharp, A.J. (2019). Screening for rare epigenetic variations in autism and schizophrenia. Hum. Mutat. *40*, 952–961.
27. De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al.; DDD Study; Homozygosity Mapping Collaborative for Autism; and UK10K Consortium (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215.
28. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell *180*, 568–584.e23.
29. Pascolini, G., Agolini, E., Majore, S., Novelli, A., Grammatico, P., and Digilio, M.C. (2018). Helsmoortel-Van der Aa Syndrome as emerging clinical diagnosis in intellectually disabled children with autistic traits and ocular involvement. Eur. J. Paediatr. Neurol. *22*, 552–557.
30. Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biol. *14*, R115.
31. 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.
32. Deciphering Developmental Disorders, S.; and Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature *542*, 433–438.
33. Takenouchi, T., Miwa, T., Sakamoto, Y., Sakaguchi, Y., Uehara, T., Takahashi, T., and Kosaki, K. (2017). Further evidence that a blepharophimosis syndrome phenotype is associated with a specific class of mutation in the ADNP gene. Am. J. Med. Genet. A. *173*, 1631–1634.
34. Krajewska-Walasek, M., Jurkiewicz, D., Piekutowska-Abramczuk, D., Kucharczyk, M., Chrzanowska, K.H., Jezela-Stanek, A., and Ciara, E. (2016). Additional data on the clinical phenotype of Helsmoortel-Van der Aa syndrome associated with a novel truncating mutation in ADNP gene. Am. J. Med. Genet. A. *170*, 1647–1650.
35. Magen, I., and Gozes, I. (2013). Microtubule-stabilizing peptides and small molecules protecting axonal transport and brain function: focus on davunetide (NAP). Neuropeptides *47*, 489–495.

The American Journal of Human Genetics, Volume *107*

Supplemental Data

Episignatures Stratifying Helsmoortel-Van Der Aa Syndrome Show Modest Correlation with Phenotype

Michael S. Breen, Paras Garg, Lara Tang, Danielle Mendonca, Tess Levy, Mafalda Barbosa, Anne B. Arnett, Evangeline Kurtz-Nelson, Emanuele Agolini, Agatino Battaglia, Andreas G. Chiocchetti, Christine M. Freitag, Alicia Garcia-Alcon, Paola Grammatico, Irva Hertz-Picciotto, Yunin Ludena-Rodriguez, Carmen Moreno, Antonio Novelli, Mara Parellada, Giulia Pascolini, Flora Tassone, Dorothy E. Grice, Daniele Di Marino, Raphael A. Bernier, Alexander Kolevzon, Andrew J. Sharp, Joseph D. Buxbaum, Paige M. Siper, and Silvia De Rubeis

Figure S1

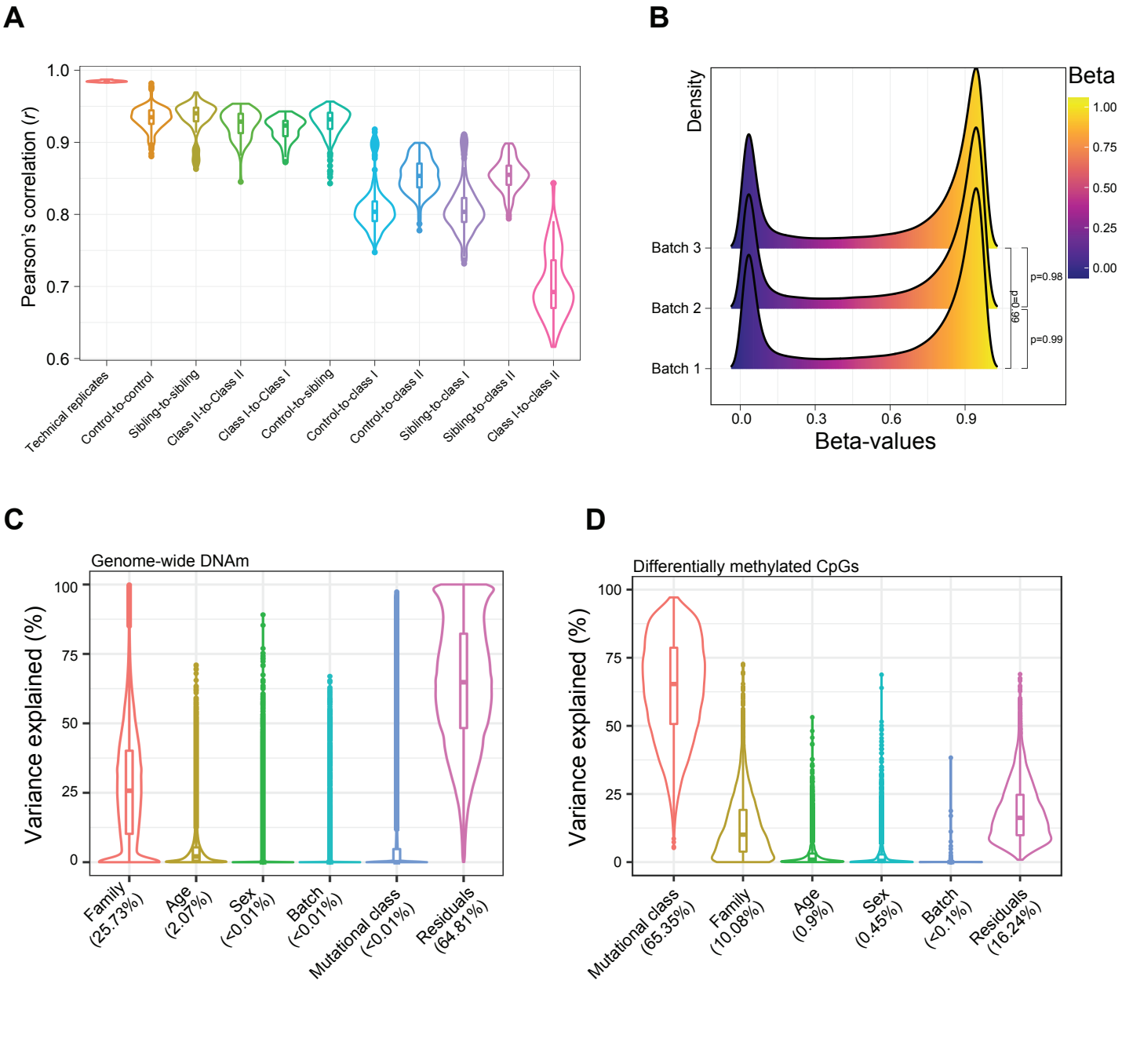
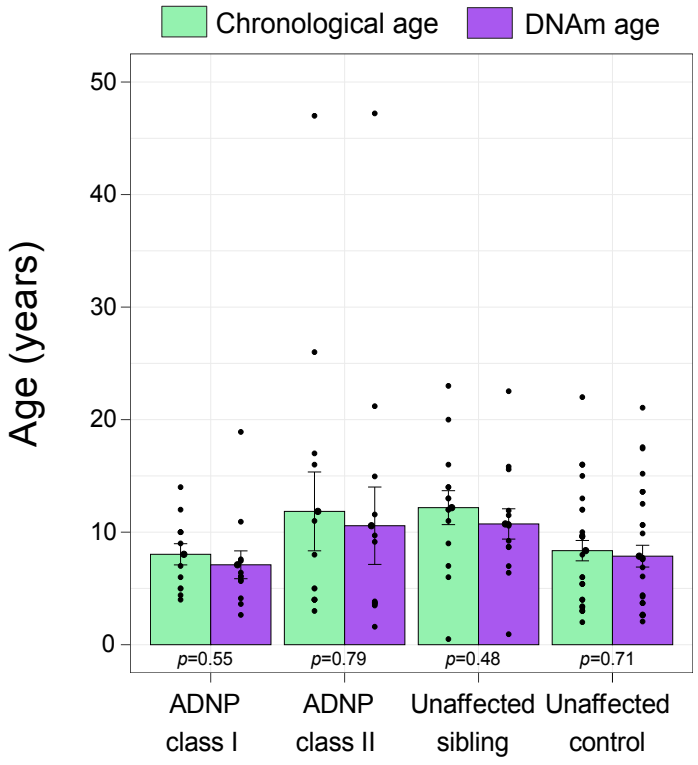

A
Pearson's correlation (r)
1.0
0.9
0.8
0.7
0.6
Technical replicates
Control-to-control
Sibling-to-sibling
Class II-to-Class II
Class I-to-Class I
Control-to-sibling
Control-to-class I
Control-to-class II
Sibling-to-class I
Sibling-to-class II
Class I-to-class II
B
Density
Beta
1.00
0.75
0.50
0.25
0.00
Batch 3
Batch 2
Batch 1
p=0.98
p=0.99
0.0
0.3
0.6
0.9
Beta-values
C
Genome-wide DNAm
Variance explained (%)
100
75
50
25
0
Family (25.73%)
Age (2.07%)
Sex (<0.01%)
Batch (<0.01%)
Mutational class (<0.01%)
Residuals (64.81%)
D
Differentially methylated CpGs
Variance explained (%)
100
75
50
25
0
Mutational class (65.35%)
Family (10.08%)
Age (0.9%)
Sex (0.45%)
Batch (<0.1%)
Residuals (16.24%)

Figure S2

A


Chronological age
DNAm age
Age (years)
0
10
20
30
40
50
p=0.55
p=0.79
p=0.48
p=0.71
ADNP class I
ADNP class II
Unaffected sibling
Unaffected control

B

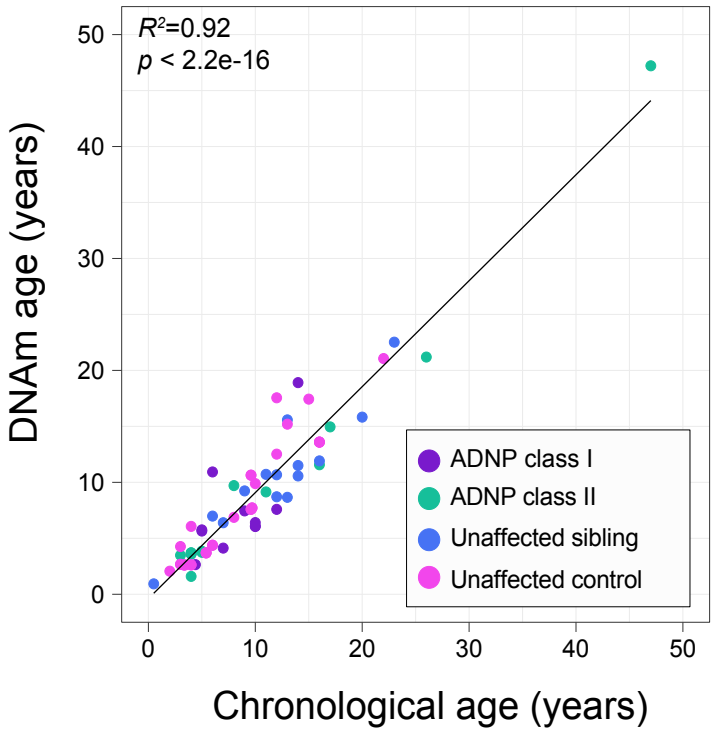

R²=0.92
p < 2.2e-16
DNAm age (years)
Chronological age (years)
0
10
20
30
40
50
ADNP class I
ADNP class II
Unaffected sibling
Unaffected control

Figure S3

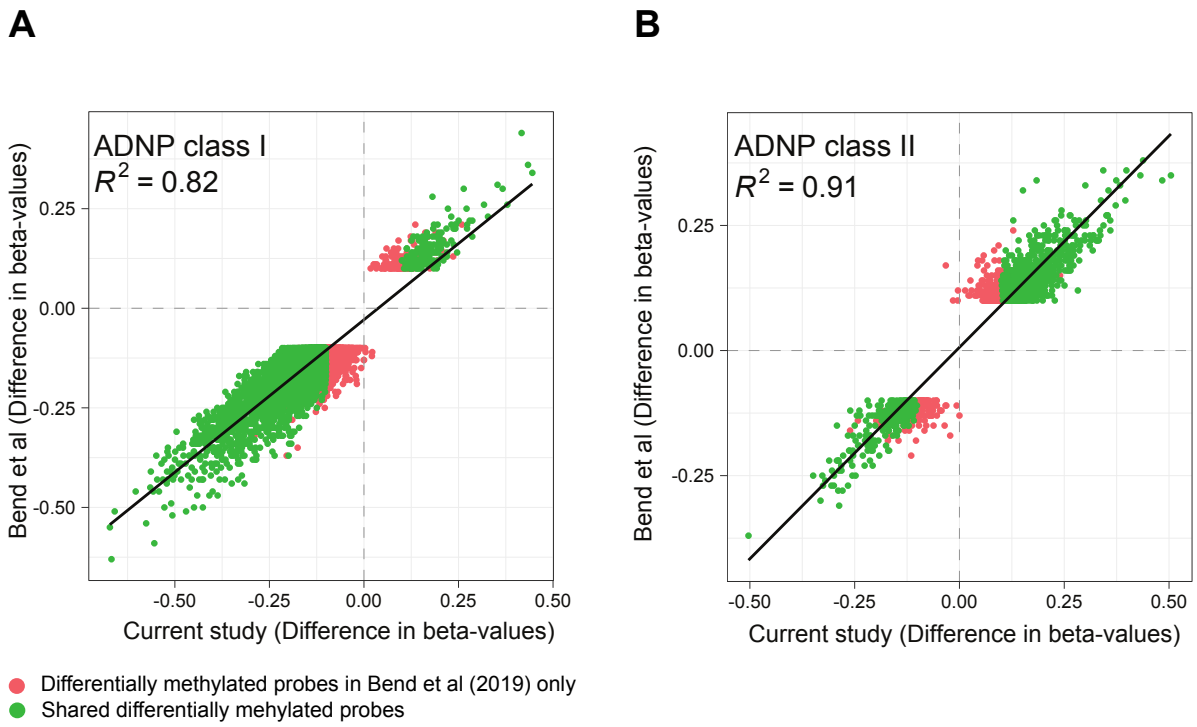
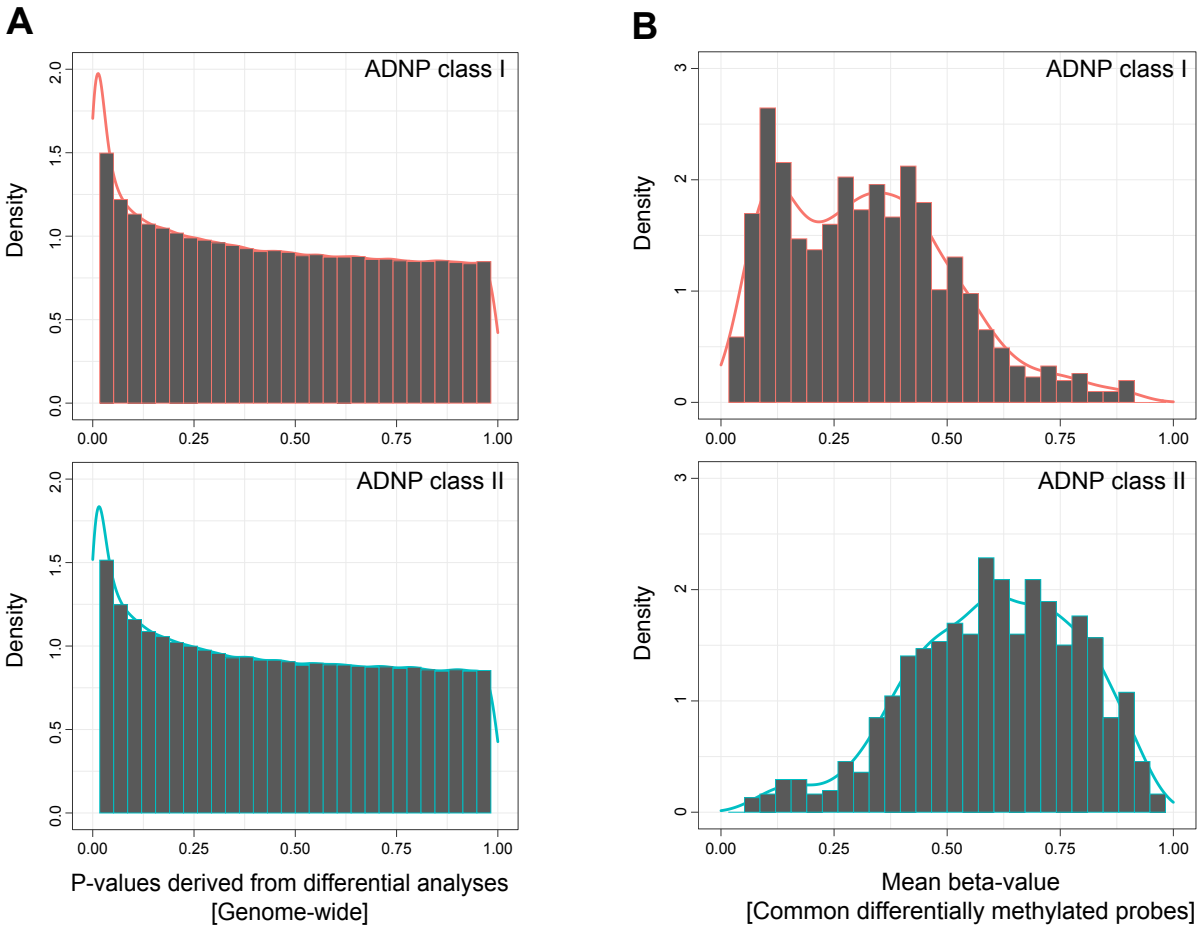

A
ADNP class I
R^2 = 0.82
Bend et al (Difference in beta-values)
Current study (Difference in beta-values)
0.25
0.00
-0.25
-0.50
-0.50
-0.25
0.00
0.25
0.50
B
ADNP class II
R^2 = 0.91
Bend et al (Difference in beta-values)
Current study (Difference in beta-values)
0.25
0.00
-0.25
-0.50
-0.25
0.00
0.25
0.50
Differentially methylated probes in Bend et al (2019) only
Shared differentially mehylated probes

Figure S4


A
ADNP class I
ADNP class II
Density
P-values derived from differential analyses
[Genome-wide]
B
ADNP class I
ADNP class II
Density
Mean beta-value
[Common differentially methylated probes]

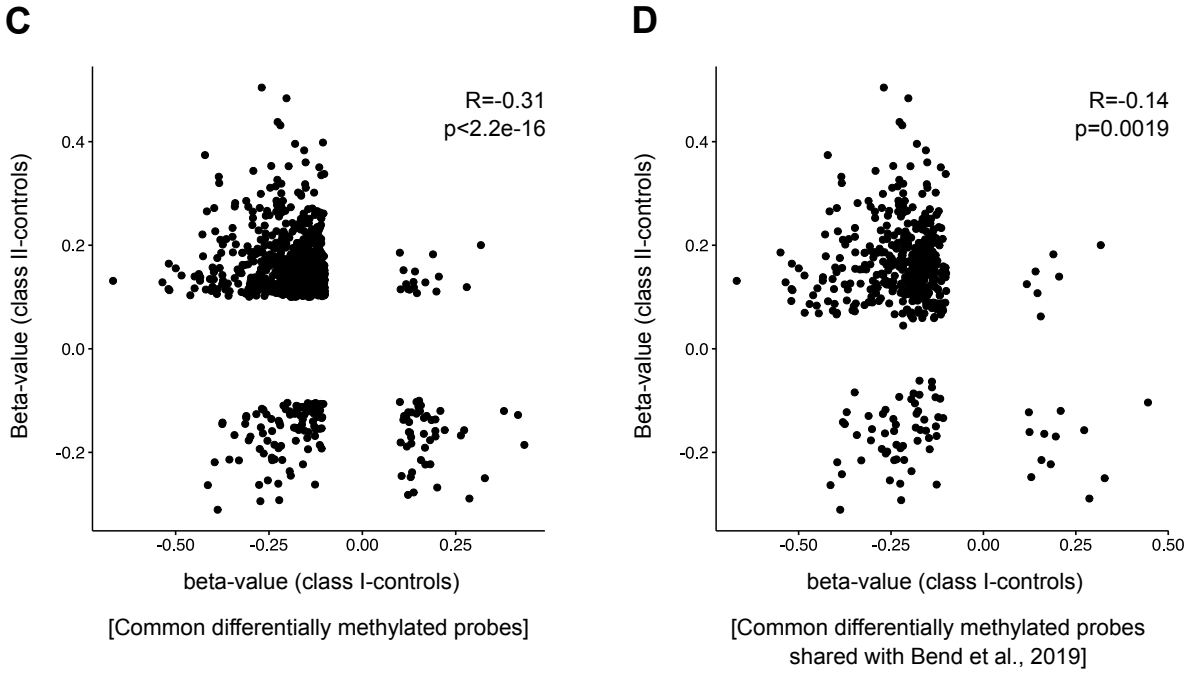

C
R=-0.31
p<2.2e-16
Beta-value (class II-controls)
beta-value (class I-controls)
[Common differentially methylated probes]
D
R=-0.14
p=0.0019
Beta-value (class II-controls)
beta-value (class I-controls)
[Common differentially methylated probes
shared with Bend et al., 2019]

Figure S5

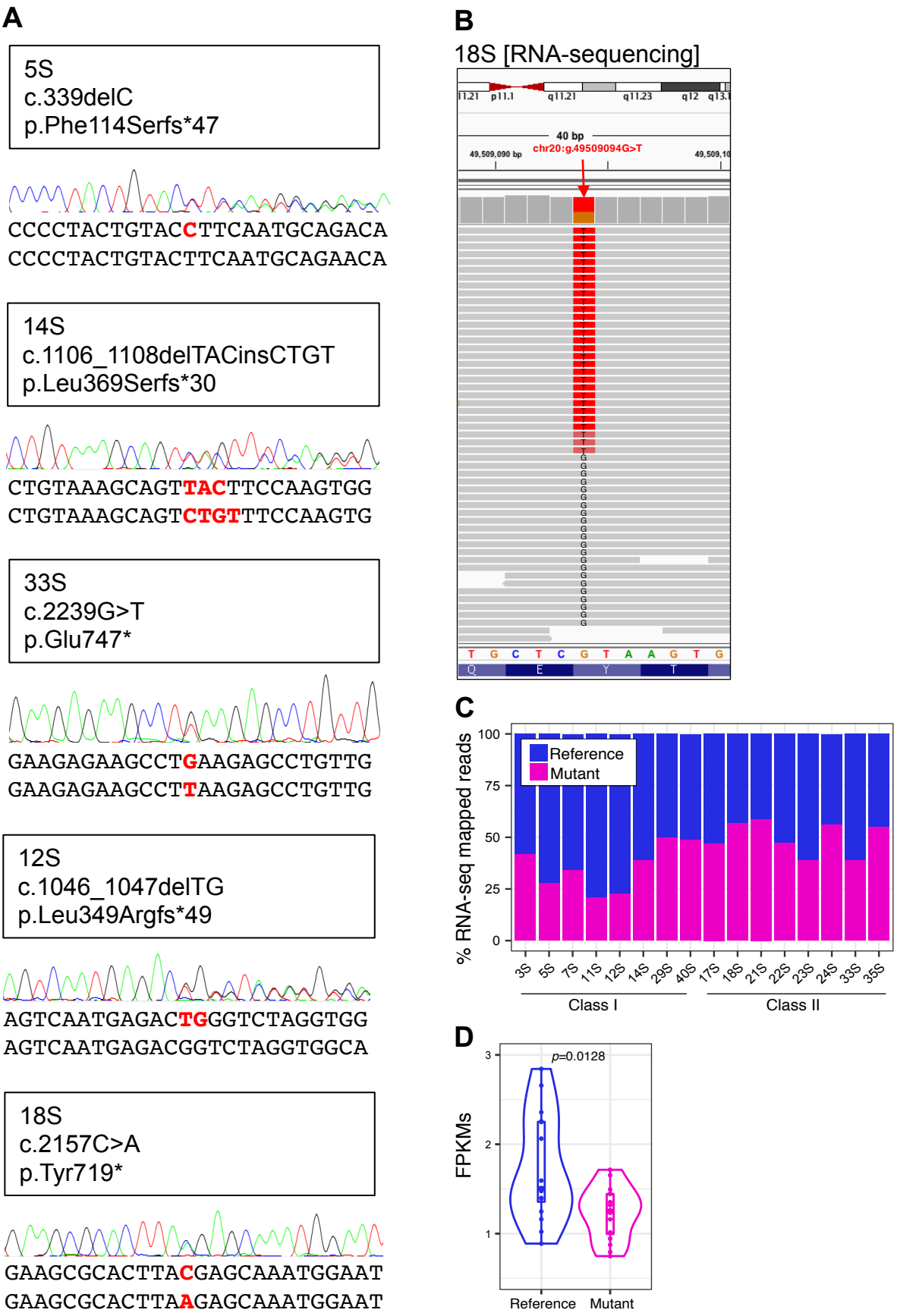
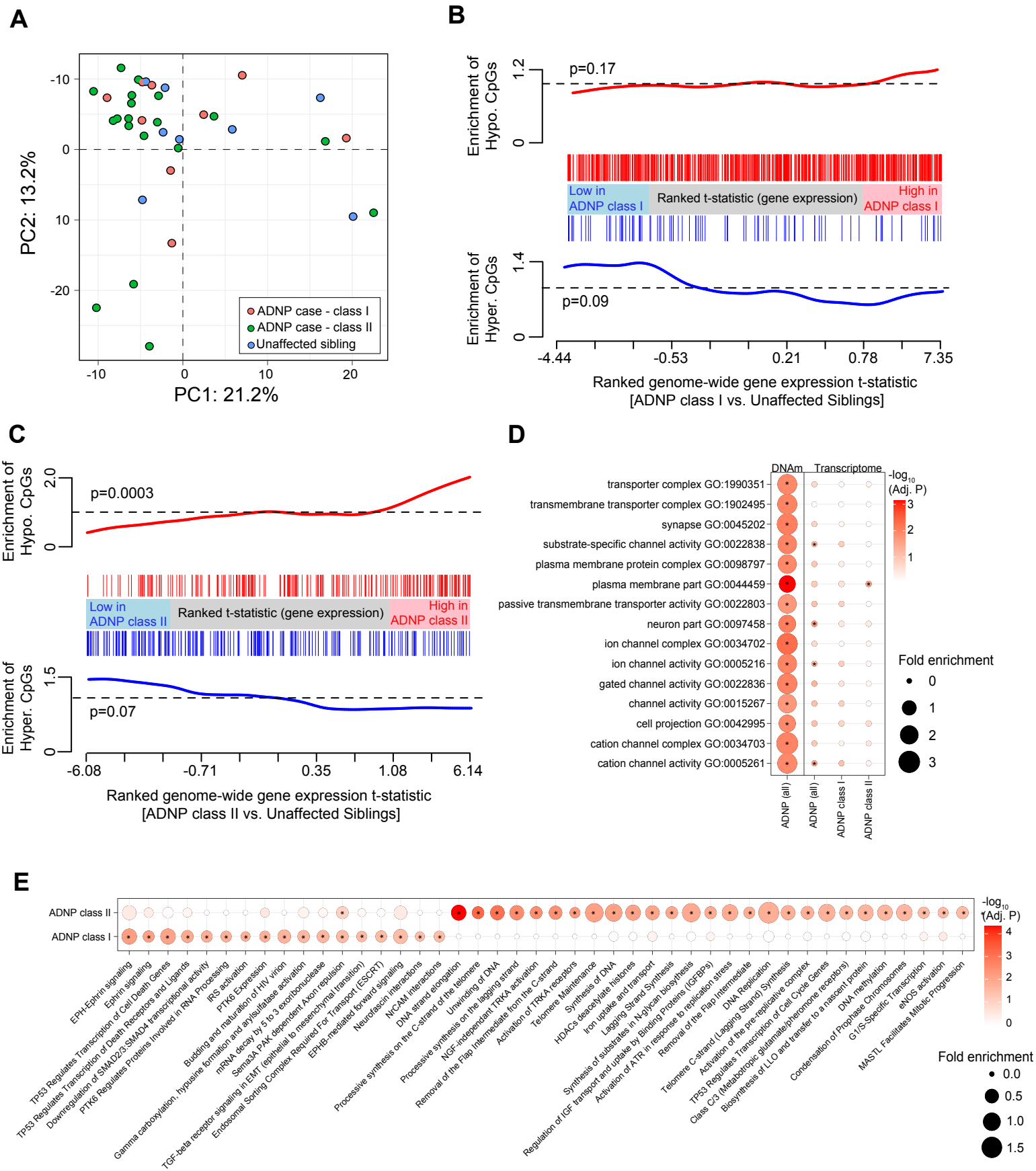

A
5S
c.339delC
p.Phe114Serfs*47
CCCCTACTGTACCTTCAATGCAGACA
CCCCTACTGTACTTCAATGCAGAACA
14S
c.1106_1108delTACinsCTGT
p.Leu369Serfs*30
CTGTAAAGCAGTTACTTCCAAGTGG
CTGTAAAGCAGTCTGTTTCCAAGTG
33S
c.2239G>T
p.Glu747*
GAAGAGAAGCCTGAAGAGCCTGTTG
GAAGAGAAGCCTTAAGAGCCTGTTG
12S
c.1046_1047delTG
p.Leu349Argfs*49
AGTCAATGAGACTGGGTCTAGGTGG
AGTCAATGAGACGGTCTAGGTGGCA
18S
c.2157C>A
p.Tyr719*
GAAGCGCACTTACGAGCAAATGGAAT
GAAGCGCACTTAAGAGCAAATGGAAT
B
18S [RNA-sequencing]
40 bp
chr20:g.49509094G>T
C
% RNA-seq mapped reads
Reference
Mutant
Class I
Class II
D
FPKMs
p=0.0128
Reference
Mutant

Figure S6


A
PC2: 13.2%
PC1: 21.2%
ADNP case - class I
ADNP case - class II
Unaffected sibling
B
Enrichment of Hypo. CpGs
p=0.17
Low in ADNP class I
Ranked t-statistic (gene expression)
High in ADNP class I
Enrichment of Hyper. CpGs
p=0.09
Ranked genome-wide gene expression t-statistic [ADNP class I vs. Unaffected Siblings]
C
Enrichment of Hypo. CpGs
p=0.0003
Low in ADNP class II
Ranked t-statistic (gene expression)
High in ADNP class II
Enrichment of Hyper. CpGs
p=0.07
Ranked genome-wide gene expression t-statistic [ADNP class II vs. Unaffected Siblings]
D
DNAm
Transcriptome
transporter complex GO:1990351
transmembrane transporter complex GO:1902495
synapse GO:0045202
substrate-specific channel activity GO:0022838
plasma membrane protein complex GO:0098797
plasma membrane part GO:0044459
passive transmembrane transporter activity GO:0022803
neuron part GO:0097458
ion channel complex GO:0034702
ion channel activity GO:0005216
gated channel activity GO:0022836
channel activity GO:0015267
cell projection GO:0042995
cation channel complex GO:0034703
cation channel activity GO:0005261
ADNP (all)
ADNP (all)
ADNP class I
ADNP class II
-log10 (Adj. P)
Fold enrichment
E
ADNP class II
ADNP class I
EPH-Ephrin signaling
Ephrin signaling
TP53 Regulates Transcription of Cell Death Genes
TP53 Regulates Transcription of Death Receptors and Ligands
Downregulation of SMAD2/3:SMAD4 transcriptional activity
PTK6 Regulates Proteins Involved in RNA Processing
IRS activation
PTK6 Expression
Budding and maturation of HIV virion
Gamma carboxylation, hypusine formation and arylsulfatase activation
mRNA decay by 5 to 3 exoribonuclease
Sema3A PAK dependent Axon repulsion
TGF-beta receptor signaling in EMT (epithelial to mesenchymal transition)
Endosomal Sorting Complex Required For Transport (ESCRT)
EPHB-mediated forward signaling
Neurotrascin interactions
NrCAM interactions
DNA strand elongation
Processive synthesis on the C-strand of the telomere
Unwinding of DNA
Processive synthesis on the lagging strand
NGF-independant TRKA activation
Removal of the Flap Intermediate from the C-strand
Activation of TRKA receptors
Telomere Maintenance
Synthesis of DNA
HDACs deacetylate histones
Iron uptake and transport
Lagging Strand Synthesis
Synthesis of substrates in N-glycan biosynthesis
Regulation of IGF transport and uptake by Binding Proteins (IGFBPs)
Activation of ATR in response to replication stress
Removal of the Flap Intermediate
DNA Replication
Telomere C-strand (Lagging Strand) Synthesis
Activation of the pre-replicative complex
TP53 Regulates Transcription of Cell Cycle Genes
Class C/3 (Metabotropic glutamate/pheromone receptors)
Biosynthesis of LLO and transfer to a nascent protein
Condensation of Prophase Chromosomes
DNA methylation
G1/S-Specific Transcription
eNOS activation
MASTL Facilitates Mitotic Progression
-log10 (Adj. P)
Fold enrichment

Supplemental Figures and Tables legend

Figure S1. Strong reproducibility across methylation batches. (A) Sample-to-sample Pearson's correlation across the three methylation batches. **(B)** Frequency distribution of β-values for all probes across the three batches. Significance was tested with a Wilcoxon rank sum test. **(C)** The linear mixed modeling frame work of the variancePartition R package was used to partitioned the fraction of genome-wide methylation variance explained by known factors in the dataset, including family (for probands and siblings), age, sex, methylation batch, and mutational class (class I vs. class II). Median variance explained in displayed below each factor in brackets. **(D)** Fraction of variance explained on differentially methylated probes that can be attributed to key variables in the dataset, including family (for probands and siblings), age, sex, methylation batch, and mutational class (class I vs. class II).

Figure S2. Strong correlation between chronological age and methylation-inferred (DNAm) age. (A) Chronological (green) and DNAm age (purple) across the four sample groups: *ADNP* class I mutations, *ADNP* class II mutations, unaffected siblings, and unaffected age-matched controls. **(B)** Correlation between chronological and DNAm age across the four sample groups: *ADNP* class I cases (purple), *ADNP* class II cases (green), unaffected siblings (blue), and unaffected age-matched controls (pink).

Figure S3. Robust convergence between the episignatures found in this study and in a previous study. (A) Scatterplot showing the methylation difference (mean $\beta\text{-values}_{cases}$ – mean $\beta\text{-values}_{controls}$) detected by Bend and colleagues (2019) [1] (Y-axis) and by this study (X-axis) for the probes found differentially methylated in class I individuals only by Bend and colleagues (2019) [1] (red) or by both studies (green). **(B)** Scatterplot showing the methylation difference (mean $\beta\text{-values}_{cases}$ – mean $\beta\text{-values}_{controls}$) detected by Bend and colleagues (2019) [1] (Y-axis) and by this study (X-axis) for the probes found differentially methylated in class II individuals only by Bend and colleagues [1] (red) or by both studies (green).

Figure S4. (A) Frequency distribution of the p-values for all the probes detected in class I cases (top panel) and class II cases (bottom panel). **(B)** Frequency distribution of the mean β-values for the 888 differentially methylated CpGs shared between the two classes for individuals with class I mutations (top panel) or class II mutations (bottom panel). **(C)** Correlation between (mean $\beta\text{-values}_{cases}$ – mean $\beta\text{-values}_{controls}$) class I (X-axis) and (mean $\beta\text{-values}_{cases}$ – mean $\beta\text{-values}_{controls}$) class II (Y-axis) cases for the 888 probes found shared between class I and class II in this study. **(D)** Correlation between (mean $\beta\text{-values}_{cases}$ – mean $\beta\text{-values}_{controls}$) in class I (X-axis) and (mean $\beta\text{-values}_{cases}$ – mean $\beta\text{-values}_{controls}$) in class II (Y-axis) for the 472 probes found shared between class I and class II in both this study and Bend and colleagues (2019) [1].

Figure S5. Biallelic expression of *ADNP* mutant mRNAs for (**A**) three class I mutations (p.Phe114Serfs*47, p.Leu349Argfs*49, and pLeu369Serfs*30) and two class II mutations (p.Tyr719* and p.Glu747*). The Sanger sequencing traces, alongside the cDNA sequence for reference (top) and mutant (bottom) alleles are shown. (**B**) RNA-sequencing pile-up of class II mutation p.Tyr719* (sample 18S) displaying coverage and number of mutant reads (red) and reference (*i.e.* healthy) reads (grey). (**C**) RNA-sequencing quantified the total fraction of *ADNP* mutant alleles and reference alleles present for each sample. Note, sample 1S was not included in this analysis (Table S1). (**D**) The distribution of expression (FPKM) for all mutated reads aggregated together relative to reference reads of *ADNP* across all samples. We observed lower expression of the mutant allele (Mann-Whitney U Test, P=0.01).

Figure S6. (**A**) Principal component analysis (PCA) on the RNAseq data from 17 *ADNP* cases (class I, red; class II green) and 19 unaffected sibling controls (blue), showing no distinct separation between mutational classes or controls. PCA was constructed using only genes harboring significant differentially methylated CpGs within their respective gene body. Competitive gene-set enrichment analysis examined over- and under-expression of genes harboring differentially methylated CpGs in their respective promoters. The analysis is partitioned to evaluate hypomethylated CpGs relative to over-expressed genes and hypermethylated CpGs relative to under-expressed genes, respectively for (**B**) class I cases and (**C**) and class II cases. Hypomethyated CpGs mapped to over-expressed genes in individuals with class II mutations (P=0.0003). No other significant associations were detected. (**D**) Supervised GO enrichment of pathways found to be significantly enriched for differentially methylated CpGs (labeled as DNAm) assessed three ways using

transcriptome data: *i)* using all cases versus controls (all); *ii)* class I mutations versus controls (class I); and, *iii)* class II mutations versus controls (class II). (**E**) Exploratory GO enrichment analysis and the top enrichment terms for individuals with class I and class II mutations.

Table S1. Genetic and demographic information about the 43 *ADNP* individuals included in the study. For each patient, the position of the genetic mutation on cDNA, protein and genomic DNA, coding exon location, effect, inheritance, classification of pathogenicity, mutation class, data availability, and the methylation batch are shown. Demographic information includes sex and country of origin.

Table S2. Differentially methylated CpGs, differentially methylated regions (defined as three or more differentially methylated CpGs within 1 kb of each other), and differentially expressed genes in individuals with class I and class II *ADNP* mutations.

Table S3. Extended version of Table 1, with separate and combined analyses of cohorts S and W across all clinical measures.

Supplemental Subjects and Methods

Cohorts, patients and genetic information

The study comprises four cohorts. Cohort A was collected as part of the Autism Sequencing Consortium [2; 3]. Cohort R was collected at the Medical Genetics Laboratory of Ospedale San Camillo-Forlanini [4] or Ospedale Pediatrico Bambino Gesù, Rome, Italy. Cohort S was collected at the Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai. Cohort W was collected at the University of Washington. Mutations in Cohort A were identified through research-based whole-exome sequencing [2; 3] and validated by Sanger sequencing. Mutations in Cohort R, S, and W were identified and validated in Clinical Laboratory Improvement Amendments (CLIA)-certified laboratories. Participation was approved by the Institutional Review Boards of participating sites. All caregivers provided informed written consent and assent was obtained when appropriate. The methylation analyses were performed on cohorts A, R and part of S because of DNA sample availability. The transcriptomic analyses were performed on part of S because this was the only cohort with RNA samples available. The phenotypic analyses were performed on cohorts S and W because they were the only two collected prospectively. Columns M-O of Table S1 reports the analyses performed on each of the 43 individuals. Specifically, we used 24 cases (cohorts A, R, and part of S) for methylation analyses; 30 cases (S and W) for phenotypic analyses; and, 17 cases (part of cohort S) for RNA-sequencing analysis. The 19 unaffected age-matched and 14 unaffected sibling controls were from cohort S. All mutations are described according to the Human Genome Variation Society (HGVS) guidelines for mutation nomenclature. The cDNA and amino acid positions are annotated according to RefSeq mRNA and protein sequence (NM_015339.4 and NP_056154.1). Nucleotide numbering referring to cDNA uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1. *ADNP* contains five exons, of which three are coding (Fig. 1A, Table S1). Our dataset includes a p.Gln67His missense mutation (individual 3S, see Table 1) that was clinically evaluated as likely pathogenic by Ambry Genetics. In addition to being *de novo*, absent in population databases, and located at a conserved amino acid and nucleotide position, Ambry's Translational Genomics laboratory pursued RNA studies that showed that the variant impacts splicing and causes the in-frame skipping of coding exon 2, leading to a p.Glu37_Gln67del in-frame deletion.

Annotation of ADNP protein domains

To verify that the ADNP domains annotated in the literature have strong support and probe the existence of additional domains, we used the ELM resource [5] to predict linear motifs. We found that the proteins contain eight zinc finger domains instead of the nine originally annotated using Pfam in 2001 [6]. All eight zinc fingers in ADNP are of the cysteine-cysteine-histidine-histidine (C2H2) type. Also, ELM re-mapped the boundaries of the homeobox domain (see table below) and identified eight low complexity regions. Information on the re-annotation used in Figure 1A is in the table below:

Domain	Start	End	Method
Zinc Finger 1	74	97	ELM prediction
Zinc Finger 2	107	129	ELM prediction
Low Complexity Region	130	141	ELM prediction
Zinc Finger 3	165	188	ELM prediction
Zinc Finger 4	221	244	ELM prediction
Low Complexity Region	393	418	ELM prediction
Low Complexity Region	424	441	ELM prediction
Zinc Finger 5	447	469	ELM prediction
Zinc Finger 6	489	510	ELM prediction
Zinc Finger 7	512	535	ELM prediction
Low Complexity Region	584	597	ELM prediction

Zinc Finger 8	622	647	ELM prediction
Nuclear Localization Sequence (NLS)	714	733	Literature [7]
Low Complexity Region	725	749	ELM prediction
Homeobox	769	810	ELM prediction
HP1-binding site	820	824	Literature [7; 8]
Low Complexity Region	871	889	ELM prediction
Low Complexity Region	953	964	ELM prediction
Low Complexity Region	1006	1017	ELM prediction

Clinical evaluation

Prospective clinical and psychological characterization was completed for 22 individuals seen at the Seaver Autism Center (cohort S) and 10 individuals seen at the University of Washington (cohort W). A battery of standardized assessments was used to examine ASD, intellectual functioning, adaptive behavior, language, motor skills, and sensory processing (see below). The medical evaluation included psychiatric, neurological, and clinical genetics examinations, and medical record review.

ASD phenotype. Gold-standard ASD diagnostic testing included the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) [9], the Autism Diagnostic Interview-Revised (ADI-R) [10], and a clinical evaluation to assess Diagnostic and Statistical Manual for Mental Disorders, Fifth Edition (DSM-5) criteria for ASD [11]. A consensus diagnosis was determined based on results from the ADOS-2, ADI-R, and the clinical evaluation. The ADOS-2 and ADI-R were administered and scored by research reliable raters and the psychiatric evaluation was completed by a board-certified child and adolescent psychiatrist or licensed psychologist. The ADOS-2 is a semi-structured observational assessment that provides scores in the domains of Social Affect, Restricted and Repetitive Behavior, and a total score. A comparison score ranging from 1-10, with higher scores reflecting a greater number of symptoms, was calculated to examine symptom severity within each ADOS-2 domain and in total [12]. Twenty-one individuals received Module 1 of the ADOS, for children who are nonverbal or communicate using single words. Eight individuals received Module 2, for individuals who communicate using phrase speech. One individual received Module 3, for children who are verbally fluent. The ADI-R is a structured caregiver interview that assesses ASD symptomatology within the domains of socialization, communication, and repetitive and restricted interests and behavior. A consensus diagnosis was determined for each participant based on results from the ADOS-2, ADI-R and clinical evaluation using DSM-5.

Intellectual functioning. Global cognitive ability was measured using the Mullen Scales of Early Learning [13], the Stanford Binet Intelligence Scales, Fifth Edition [14], or the Differential Ability Scales, Second Edition (DAS-II) [15], depending on age and verbal ability. The Mullen is validated for children from birth to 68 months, but is commonly used for older individuals with ID [16]. Developmental quotients were calculated using age equivalents divided by chronological age as has been done in previous studies [17]. For example, a nonverbal developmental quotient was computed by dividing the mean age equivalents on the visual reception and fine motor scales by the child's chronological age and then multiplying by 100. The DAS-II is a measure of cognitive functioning that assesses a child's verbal reasoning, nonverbal reasoning, and spatial abilities. A general conceptual ability index can be calculated to assess overall intellectual functioning. The Stanford-Binet Intelligence Scales, Fifth Edition is an intelligence test that produces a nonverbal intellectual quotient (IQ), verbal IQ, and full scale IQ based on performance across five scales: fluid reasoning, knowledge, quantitative reasoning, visual-spatial, and working memory.

Adaptive behavior. The Vineland Adaptive Behavior Scales, Second Edition, Survey Interview Form (Vineland-II) [18] and the Vineland Adaptive Behavior Scales, Third Edition, Comprehensive Interview Form (Vineland-3) are clinician-administered interviews that assesses adaptive behavior in the domains of communication, daily living skills, socialization, and motor skills. The Vineland-II was completed for 13 individuals. The Vineland-3 was completed for 19 individuals. The motor domain is intended for children ages six years and under, but was

assessed in 19 individuals given significant motor delays in this population. The Vineland-II and Vineland-3 were also used in conjunction with cognitive testing to identify the presence and severity of ID.

Language skills. Language milestones were assessed during the ADI-R and the psychiatric evaluation. Current expressive and receptive language abilities were assessed using the Mullen, Vineland-II, MacArthur-Bates Communicative Development Inventories [19], Peabody Picture Vocabulary Test, Fourth Edition [20], and Expressive Vocabulary Test [21] as appropriate.

Motor skills. Motor milestones were assessed during the ADI-R and the psychiatric evaluation. Current motor skills were assessed using the Vineland-II and Mullen fine and gross motor skills domains. The Beery Visual-Motor Integration Test, 6th Edition [22] was completed when appropriate.

Caregiver questionnaires. To further assess everyday behavior, caregiver questionnaires were completed in the domains of sensory processing (Sensory Profile [23]), aberrant behavior (Aberrant Behavior Checklist [24]), repetitive behavior (Repetitive Behavior Scale-Revised [25]), socialization (Social Responsiveness Scale, 2nd Edition [26]), and coordination (Developmental Coordination Disorder Questionnaire [27]).

Genome-wide methylation arrays and data quality control

For the methylation analyses, genomic DNA was isolated from the peripheral blood of 24 *ADNP* cases, 19 unaffected age-matched, and 14 unaffected siblings from cohorts S, A and R. DNA methylation analysis was performed using the Ilumina EPIC 850K methylation arrays, according to the manufacturer's protocol. Whole-blood genomic DNA samples for methylation analyses were prepared at each collecting site of the four cohorts described above and then further harmonized at the Seaver Autism Center. The methylation assays were performed in three batches, each including cases from multiple cohorts and controls. Column P of Table S1 reports the batch number for each sample. Batches 1 and 2 were run at the New York Genome Center, while batch 3 was run at NXT-Dx. In each batch, we included a number of controls present in previous batches as internal controls. All the methylation data in this paper are available at the Gene Expression Omnibus under accession number GSE152428. To ensure that batch effects have minimal influence on the current analysis, we performed a standardized multistep analytic approach (Figure S1). First, we computed genome-wide pairwise Pearson correlation coefficients across all possible pairs of samples and confirm that the highest correlations are amongst technical replicates included in separate batches ($R > 0.98$) (Figure S1A). Second, we evaluated whether any large shifts in the distribution of normalized β-values were observed across batches, and found no significant differences in the distributions of β-values across batches (Figure S1B). Third, we applied a mixed linear model from the variancePartition R package [28] to compute the fraction of genome-wide methylation variance explained by differences in batch, mutational class, family status, sex and age (Figure S1C). Collectively, these factors explained ~35% of methylation variation, with family as a repeated measure having the largest genome-wide effect that explained a median 25.7% of the observed variation. The remaining factors, including batch, had a limited effect on methylation variance. Fourth, we analyzed all batches together for the joint analysis presented in this study and included batch as a covariate in our regression model. By properly covarying for known sources of methylation variation, it is possible to (partially) correct for some variables. Finally, following differential methylation analyses, we once again applied a mixed linear model to quantify variance explained using only differentially methylated probes that were found to be significantly associated with *ADNP* mutational classes. As expected, the differences in mutational classes had the largest effect and explained a median 65.35% of the variance, whereas batch had a minor contribution to the variance on these probes (< 0.001 median variance) (Fig. S1D).

The sex of the individuals included in the study was inferred both by the number of probes on chromosome Y (chrY) with detection p>0.01, and the mean β-value of probes on chromosome X (chrX) per sample. Samples with high failure rate for chrY probes and high mean β-value for chrX probes were inferred as females, while samples with low failure rate for chrY probes and low mean β-value for chrX probes were inferred as males.

Epigenetic age (DNAm age) was predicted using the online tool published by Horvath, 2013 [29]. The chronological and DNAm age for cases and controls are as follows:

Group	Chronological age (yrs)	Age inferred from methylation data (DNAm age) (yrs)
ADNP class I cases	8.03 ± 0.94	7.10 ± 1.24
ADNP class II cases	11.85 ± 3.50	10.57 ± 3.44
Unaffected siblings	12.18 ± 1.51	10.73 ± 1.35
Unaffected controls	8.36 ± 0.91	7.87 ± 0.97

The fraction of circulating peripheral blood cell types were predicted for each sample using the method of Houseman et al. 2012 [30]. Predictions for the following cell types were obtained: CD4+ T cells, CD8+ T cells, natural killer cells, B lymphocytes, monocytes and granulocytes.

Data from *ADNP* cases and controls were normalized as described in previous studies [31; 32]. In short, 862,927 probe sequences (50-mer oligonucleotides) were remapped to the reference human genome hg19 (NCBI37) using BSMAP, allowing up to 2 mismatches and 3 gaps, to retain uniquely mapped autosomal probes. We removed any probe that overlapped SNPs with MAF ≥5% identified by the 1000 Genomes Project within 5 bp upstream of the targeted CpG. Further, we removed probes with a detection p>0.01 in each individual. After filtering, we retained 820,167 autosomal probes, which were subjected to background correction, two color channel normalization and quantile normalization using the *lumi* package in R [33]. The distributions of Infinium I and Infinium II probes were adjusted using *BMIQ* [34]*.* Probes were then annotated based on their position relative to RefSeq genes using BEDTools v2.17 [35]. We defined promoter regions as ±2 kb from transcriptional start sites, gene body regions as transcription start to transcription end, and intergenic regions not annotated by the preceding categories. We also annotated individual CpG based on their overlap with CpG islands based on annotations in the UCSC genome browser, CpG shore (±2 kb of island), CpG shelf (±2 kb of shore), and CpG sea (regions outside the previous three categories).

Identification of an episignature in *ADNP* cases

We performed linear regression using age, sex, predicted blood cell composition and *ADNP* mutation status as independent variables (Test Model: Methylation ~ Disease status + Age + Sex + CD4T + natural killer cells + B cells + Monocytes + Granulocytes + Batch). Regression analysis was completed separately for class I and class II *ADNP* mutations. We did not include CD8T cell composition in the model due to very low abundance across all samples. We selected probes associated with disease status at 1% FDR and with minimum β-value difference between *ADNP* cases and controls ≥0.1. Principal component analysis and unsupervised clustering of episignatures was performed on methylation data following Combat batch adjustment [36] to remove systematic sources of variability related to batches without introducing false signal. The differentially methylated probes are listed in Table S2. To identify differentially methylated regions, we have selected differentially methylated regions with three or more probes within 1 kb of each other. These differentially methylated regions are also listed in Table S2.

Comparison of episignatures found here and in Bend et al (2019).

We extracted the 5,987 and 1,374 differentially methylated CpGs (class I and class II, respectively) in Bend et al (2019) [1] and crossed with our lists of differentially methylated CpGs. Of the 6,448 sites we detected as differentially methylated in *ADNP* cases in class I, 4,143 were also found differentially methylated in Bend et al (2019) [1]. Of the 2,582 sites we detected as differentially methylated in *ADNP* cases in class II, 1,007 were also found differentially methylated in Bend et al (2019) [1].

Also, the differentially methylated probes overlapping between our study and Bend et al (2019) [1] completely agree in terms of directionality of change. Of the 4,143 class I probes in common between the two analyses,

3,974 were hypomethylated in both studies and the remaining were hypermethylated in both studies. Of the 1,007 overlapping probes for class II, 771 were found hypermethylated in both studies and the remaining were hypomethylated in both studies.

Further, we examined the number of differentially methylated probes that lie in the gene promoter (± 2 kb from the transcriptional start sites) or in the gene body (transcription start to transcription end) across the two studies. The stratification is as follows:

Mutational class	Genomic location	Bend et al (2019)	This study	Overlapping probes
Class I	Promoter	1,662 (27.7%)	1,839 (28.5%)	1,144 (27.6%)
	Gene body	2,953 (49.3%)	3,282 (50.9%)	2,089 (50.4%)
Class II	Promoter	449 (32.7%)	897 (34.7%)	322 (31.9%)
	Gene body	856 (62.3%)	1,558 (60.3%)	647 (64.3%)

RNA isolation, library preparation, and quantification of gene expression

Blood was collected for 17 *ADNP* cases and 19 unaffected siblings using PAXgene RNA tubes (Qiagen, Valencia, CA, USA) and total RNA was extracted and purified in accordance with the PAX gene RNA kit per manufacturer's instructions. Globin mRNA was depleted from samples using the GLOBINclear Human Kit (Life Technologies, Carlsbad, CA, USA). The quantity of purified RNA was measured on a Nanodrop 2000 Spectrophotomerter (Thermo Scientific; 61.4 ± 24.1 ng μl^{-1}) and RNA integrity numbers (RIN) measured with the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA; 8.31 ± 0.68). The Illumina TruSeq Total RNA kit (Illumina, San Diego, CA, USA) was used for library preparation accordingly to manufacturer instructions without any modifications. The 36 indexed RNA libraries were pooled and sequenced using long paired-end chemistry (2x150 bp) at an average read depth of 10M reads per sample using the Illumina HiSeq2500. Illumina adapter sequences were trimmed from all fragmented reads using TrimGalore (options –paired –illumina) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). All high-quality trimmed reads were mapped to UCSC *Homo sapiens* reference genome (build hg37) using default STAR v2.4.0 parameters [37]. Samtools was used to convert bamfiles to samfiles and featureCounts [38] was used to quantify gene expression levels for each individual sample using default paired-end parameters.

RNA-seq data quality control

Raw count data measured 56,632 genes across 36 subjects. Unspecific filtering removed lowly expressed genes that did not meet the requirement of a minimum of 1 count per million (cpm) in at least 8 subjects (40% of subjects). A total of 20,491 genes were retained, then subjected to edgeR VOOM normalization [39], a variance-stabilization transformation method. Normalized data were inspected for outlying samples using unsupervised hierarchical clustering of subjects (based on Pearson coefficient and average distance metric) and principal component analysis to identify potential outliers outside two standard deviations from these averages. No outliers were present in these data.

Mutant allele abundance

We took measures to reduce the likelihood of false positives and biases in quantification of relative allele abundances at each mutation location. RNA-seq reads were filtered and mapped to retain only uniquely mapped reads, using STAR [35] and SAMtools, and then further removed potential PCR duplicates using the MarkDuplicates function in Picard tools with default parameters (http://broadinstitute.github.io/picard/faq.html). Allelic counts were computed at each single-nucleotide variant position in each sample using the SAMtools mpileup function. Next, we recorded the number of overlapping sequences containing the mutant and reference allele. The mutant allele frequency was defined as the number of covering RNA-seq reads

containing the mutant allele at that position, divided by the total number of RNA-seq reads overlapping that position.

Differential gene expression and concordance with methylation changes

A moderated *t*-test, implemented through the *limma* package [39], assessed differential gene expression between unaffected siblings and *ADNP* cases with class I and class II mutations, respectively. The analysis covaried for the possible influence of sex, gender, RIN and sequencing batch on gene expression differences. Significance threshold was set to a Benjamini-Hochberg multiple test corrected *P*-value <0.05. Correlation adjusted mean rank (CAMERA) gene set enrichment [40] was performed using the two sets of resulting summary statistics for *ADNP* cases with class I and class II mutations. CAMERA performs a competitive gene set rank test to assess whether the genes in a given set are highly ranked in terms of differential expression relative to genes that are not in the set. The test ranks gene expression differences in *ADNP* cases with class I and class II mutations relative to unaffected siblings, respectively, to test whether gene sets are over-represented towards the extreme ends of these ranked lists. It uses *limma*'s linear model framework and accommodates the observational-level weights from *voom* in the testing procedure. After adjusting the variance of the resulting gene set test statistic by a variance inflation factor that depends on the gene-wise correlation (which we set to default parameters, 0.01) and the size of the set, a *P*-value is returned and adjusted for multiple testing. We used this function to test whether *ADNP*-associated changes in gene expression were indeed enriched for genes with significant differentially methylated CpGs. We specifically focused on the enrichment of *i)* hyper-methylated CpGs among the reported under-expressed genes and *ii)* hypo-methylated CpGs among the reported over-expressed genes in *ADNP* cases. In this fashion, we can further gauge the distribution of differentially methylated CpGs across a ranked list of differentially expressed genes.

Enrichment analyses

We assessed enrichment of differentially methylated CpG and differentially expressed genes using five gene lists: *a)* 914 genes implicated in developmental delay/intellectual disability (DD/ID), based on the Developmental Disorders Genotype-Phenotype Database (DDG2P) (https://decipher.sanger.ac.uk/info/ddg2p); *b)* 102 ASD risk genes defined by the Autism Sequencing Consortium [3]; *c)* 145 risk genes for CHD built in house; *d)* Gene ontology (GO) molecular functions and cellular processes; and *e)* REACTOME pathway gene sets. For DNA methylation analyses, we utilized a background gene set comprising 25,141 genes that overlapped 606,904 probes covered by the Illumina 850K EPIC methylation arrays. For transcriptome analyses, we utilized a background gene set of 20,491 genes that were defined as 'expressed' in the current data set (described above). For each comparison between the differentially methylated and differentially expressed genes in the input set, we first constructed the empirical distribution by randomly sampling the same number of genes as in the input set from the background gene set 10,000 times, using a custom script in R. Enrichment *P*-values were computed by calculating the number of sampled gene lists that had at least as many overlapping genes with the target sets as the input set, divided by 10,000 permutations.

Data availability

All methylation and RNA-sequencing raw data files are made available at the Gene Expression Omnibus (GEO) under SuperSeries accession number GSE152428.

RT-PCR and Sanger sequencing

Reverse transcription was performed on DNAseI-treated 50ng of RNA using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, California, USA) according to the manufacturer's protocol. PCR was performed using FastStart PCR Master (Roche, Mannheim, Germany) or Phusion High-Fidelity PCR Kit (New England BioLabs, Ipswich, Massachusetts), using primers indicated below:

Sample	Forward primer	Reverse primer
5S	ACGAAAAACCAGGACTATCGGA	AAACAGCTTGCTCTACACTGTCA
12S	ATCGGTTCCCTTGCTTCTGG	TGGCCCGATGAGAGAGAAGA
14S	TGCAGCAGAACAACTATGGAGT	CTGCAGCAGGTTTGGAACTG
18S	ATACCAGCAACATGACCGCC	TGGTGGGATAGGGCTGTTTG
33S	CACCCTCTCGGCTTAATCAGT	TAAACTGGCTGCTAGCTTCTCAA

PCR products were purified using a MultiScreen PCR filter plate (Millipore Sigma, Burlington, Massachusetts) and submitted for Sanger Sequencing at Genewiz with the primers indicated above.

Supplemental references

1. Bend, E.G., Aref-Eshghi, E., Everman, D.B., Rogers, R.C., Cathey, S.S., Prijoles, E.J., Lyons, M.J., Davis, H., Clarkson, K., Gripp, K.W., et al. (2019). Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome. Clin Epigenetics 11, 64.
2. De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 515, 209-215.
3. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell 180, 568-584 e523.
4. Pascolini, G., Agolini, E., Majore, S., Novelli, A., Grammatico, P., and Digilio, M.C. (2018). Helsmoortel-Van der Aa Syndrome as emerging clinical diagnosis in intellectually disabled children with autistic traits and ocular involvement. Eur J Paediatr Neurol 22, 552-557.
5. Gouw, M., Michael, S., Samano-Sanchez, H., Kumar, M., Zeke, A., Lang, B., Bely, B., Chemes, L.B., Davey, N.E., Deng, Z., et al. (2018). The eukaryotic linear motif resource - 2018 update. Nucleic Acids Res 46, D428-D434.
6. Zamostiano, R., Pinhasov, A., Gelber, E., Steingart, R.A., Seroussi, E., Giladi, E., Bassan, M., Wollman, Y., Eyre, H.J., Mulley, J.C., et al. (2001). Cloning and characterization of the human activity-dependent neuroprotective protein. J Biol Chem 276, 708-714.
7. Cappuyns, E., Huyghebaert, J., Vandeweyer, G., and Kooy, R.F. (2018). Mutations in ADNP affect expression and subcellular localization of the protein. Cell Cycle 17, 1068-1075.
8. Mosch, K., Franz, H., Soeroes, S., Singh, P.B., and Fischle, W. (2011). HP1 recruits activity-dependent neuroprotective protein to H3K9me3 marked pericentromeric heterochromatin for silencing of major satellite repeats. PLoS One 6, e15894.
9. Lord, C., Rutter, M., DiLavore, P.S., Risi, S., Gotham, K., and Bishop, D. (2012). Autism Diagnostic Observation Schedule, 2nd edition (ADOS-2) Manual (Part I): Modules 1-4.(Torrance, CA: Western Psychological Services).
10. Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 24, 659-685.
11. American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders, 5th edition, text revision.(Washington, DC: American Psychiatric Association).
12. Hus, V., Gotham, K., and Lord, C. (2014). Standardizing ADOS domain scores: separating severity of social affect and restricted and repetitive behaviors. J Autism Dev Disord 44, 2400-2412.
13. Mullen, E.M. (1995). Mullen Scales of Early Learning.(Circle Pines, MN: American Guidance Services).
14. Roid, G.H. (2003). Stanford Binet Intelligence Scales (5th edition).(Itasca, IL: Riverside Publishing).
15. Elliot, C.D. (2007). Differential Ability Scales—Second edition: Introductory and technical manual.(San Antonio, TX: Harcourt Assessment).
16. Bishop, S.L., Guthrie, W., Coffing, M., and Lord, C. (2011). Convergent validity of the Mullen Scales of Early Learning and the differential ability scales in children with autism spectrum disorders. American journal on intellectual and developmental disabilities 116, 331-343.
17. Akshoomoff, N. (2006). Use of the Mullen Scales of Early Learning for the assessment of young children with Autism Spectrum Disorders. Child Neuropsychol 12, 269-277.
18. Sparrow, S.S., Cicchetti, D.V., and Balla, D.A. (2005). Vineland Adaptive Behavior Scales: Second edition (Vineland II), Survey Interview Form/Caregiver Rating Form.(Livonia, MI: Pearson Assessments).
19. Fenson, L., Marchman, V.A., Thal, D.J., Dale, P.S., Reznick, J.S., and Bates, E. (2007). MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual, Second Edition.(Baltimore, MD: Brookes Publishing Co.).
20. Dunn, L.M., and Dunn, D.M. (2007). PPVT-4: Peabody picture vocabulary test.(Bloomington, MN: Pearson Assessments).
21. Williams, K.T. (2007). The Expressive Vocabulary Test (2nd edition).(Circle Pines, MN: AGS Publishing).
22. Beery, K.E., Buktenica, N.A., and Beery, N.A. (2010). Beery-Buktenica Developmental Test of Visual-Motor Integration, Sixth Edition.(Minneapolis, MN: Pearson).

23. Dunn, W., and Westman, K. (1997). The Sensory Profile: The performance of a national sample of children without disabilities. American Journal of Occupational Therapy 51, 25-34.
24. Aman, M.G., and Singh, N.N. (1986). Aberrant Behavior Checklist (ABC) Manual.(East Aurora: Slosson Educational Publications, Inc).
25. Lam, K.S., and Aman, M.G. (2007). The Repetitive Behavior Scale-Revised: independent validation in individuals with autism spectrum disorders. Journal of autism and developmental disorders 37, 855-866.
26. Constantino, J.N., and Gruber, C.P. (2012). Social Responsiveness Scale, Second Edition (SRS-2).(Torrance: Western Psychological Services).
27. Wilson, B.N., Dewey, D., and Campbell, A. (1998). Developmental coordination disorder questionnaire (DCDQ).(Canada: Alberta Children's Hospital Research Center).
28. Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. BMC bioinformatics 17, 483.
29. Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biol 14, R115.
30. 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.
31. 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. Nature communications 9, 2064.
32. Garg, P., and Sharp, A.J. (2019). Screening for rare epigenetic variations in autism and schizophrenia. Hum Mutat.
33. Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. Bioinformatics 24, 1547-1548.
34. Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., and Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 29, 189-196.
35. Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics 47, 11 12 11-34.
36. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28, 882-883.
37. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21.
38. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923-930.
39. 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 research 43, e47.
40. Wu, D., and Smyth, G.K. (2012). Camera: a competitive gene set test accounting for inter-gene correlation. Nucleic acids research 40, e133.