DNA methylation episignature and comparative epigenomic profiling of HNRNPU-related neurodevelopmental disorder

Target Journal: Genetics in medicine

DNA methylation episignature and comparative epigenomic profiling of *HNRNPU*-related neurodevelopmental disorder

Kathleen Rooney[1,2]†, Liselot van der Laan[3]†, Slavica Trajkova[4,], Sadegheh Haghshenas[2], Raissa Relator[1], Peter Lauffer[3], Niels Vos[3], Michael A. Levy[1], Brunetti-Pierri Nicola[5,6], Gaetano Terrone[6], Cyril Mignot[7], Boris Keren[7], Thierry Billette[8], Catharina M.L. Volker-Touw[9], Nienke Verbeek[9], Jasper J. van der Smagt[9], Renske Oegema[9], Alfredo Brusco[4,10], Giovanni Battista Ferrero[11], Mala Misra-Isrie[3], Ron Hochstenbach[3], Mariëlle Alders[3], Marcel M.A.M. Mannens[3], Bekim Sadikovic[1,2]‡, Mieke M. van Haelst[3]‡ and Peter Henneman[3]‡

† and ‡These authors contributed equally to this work

* Author for correspondence: p.henneman@amsterdamumc.nl and bekim.sadikovic@lhsc.on.ca

1. Verspeeten Clinical Genome Centre, London Health Science Centre, London, ON N6A 5W9, Canada

2. Department of Pathology and Laboratory Medicine, Western University, London, ON N5A 3K7, Canada

3. Department of Human Genetics, Amsterdam Reproduction & Development Research Institute, Amsterdam University Medical Centers, Meibergdreef 9, Amsterdam, AZ, 1105, The Netherlands

4. Department of Medical Sciences, University of Torino, 10124 Torino TO, Italy

5. Telethon Institute of Genetics and Medicine (TIGEM), 80078 Pozzuoli NA, Italy

6. Department of Translational Medicine, Federico II University, 80138 Naples NA, Italy

7. Assistance Publique-Hôpitaux de Paris, Service de Neuropédiatrie, Hôpital Armand Trousseau, Paris, France

8. Assistance Publique-Hopitaux de Paris, Sorbonne Université, Departement de Génétique, Groupe Hospitalier Pitie-Salpetriere et Hopital Trousseau, Paris, 75651, France

9. Department of Genetics, University Medical Centre Utrecht, 3584 CX Utrecht, the Netherlands.

10. Medical Genetics Unit, Città della Salute e della Scienza University Hospital, 10126 Torino TO, Italy

11. Department of Clinical and Biological Science, University of Torino, Torino, Italy.

Abstract

Purpose: *HNRNPU* haploinsufficiency is associated with Developmental and Epileptic Encephalopathy 54. This neurodevelopmental disorder is characterized by developmental delay, intellectual disability, speech impairment, and early onset epilepsy. We performed genome-wide DNA methylation (DNAm) analysis in a cohort of participants to develop a diagnostic biomarker and gain functional insights into the molecular pathophysiology of *HNRNPU*-related disorder.

Methods: DNAm profiles of participants carrying *HNRNPU* variants, identified through an international multi-center collaboration, were assessed using Infinium Methylation EPIC arrays. Statistical and functional correlation analyses were performed comparing the *HNRNPU* cohort to 56 rare disorders with previously reported DNAm episignatures.

Results: A robust and reproducible DNAm episignature and a global DNAm profile were identified. Correlation analysis identified partial overlap and similarity of the global *HNRNPU* DNAm profile to several other rare disorders.

Conclusion: This study demonstrates new evidence of a specific and sensitive DNAm episignature associated with pathogenic heterozygous *HNRNPU*-variants, establishing its utility as a clinical biomarker for the expansion of the EpiSign™ diagnostic test.

Keywords; *HNRNPU*, Intellectual disability, Neurodevelopmental Disorder, DNA methylation, Epigenetics, Episignature, CNV

Introduction

HNRNPU (Heterogeneous Nuclear Ribonucleoprotein U; OMIM #602869) haploinsufficiency has been associated with a neurodevelopmental disorder (NDD) referred to as Developmental and Epileptic Encephalopathy 54 (DEE54; OMIM #617391) [1, 2]. Patients with DEE54 present developmental delay, moderate to severe intellectual disability (ID), marked speech impairment and early onset epilepsy, often refractory to treatment [3]. Autistic features, microcephaly, (nonspecific) dysmorphic facial features, nonspecific brain MRI findings include ventriculomegaly and thinning of the corpus callosum, cardiac and renal abnormalities have also been reported [1-5].

Heterogeneous Nuclear Ribonucleoproteins (HNRNP) are part of an RNA binding protein family containing multiple extraordinarily complex and versatile proteins that are involved in controlling the maturation of newly formed nuclear RNAs into messenger RNAs. They also play a role in RNA splicing, polyadenylation, capping, modification, export, localization, translation, and turnover [4]. HNRNPs have interactions with RNPs (ribonucleoproteins) and are directly involved in every stage of mRNA processing and formation, and hence are essential in early development [5]. The observed high variety of HNRNPs functions results from their ability to produce different alternatively spliced isoforms that form many distinct protein complexes with other *HNRNP* genes [6]. As successful RNA regulation is particularly important in neuronal cells, many *HNRNPs* have also been linked to other diseases such as cancer, spinal muscular atrophy, amyotrophic lateral sclerosis, congenital myasthenic syndrome, multiple sclerosis, Alzheimer's disease, and frontotemporal lobe dementia [6-8]. HNRNPU is the largest among the HNRNP proteins and forms a complex capable of functioning in organizing and stabilizing nuclear chromatin, regulating gene transcription, RNA splicing, and RNA stability [6-8]. It has also been shown that *HNRNPU* regulates topologically associated domain (TAD) boundaries, which are linked to the 3D chromatin organization [9].

Genes involved in epigenetic machinery have been categorized as readers, writers, erasers and remodelers [10]. The phenotypes that result from aberrations linked to the epigenetic machinery are mostly linked to ID, delayed growth, and dysmorphic features [10, 11] . Proteins that function in

epigenetic regulation have critical roles during embryonic and fetal development. Germline mutations in these genes result in a spectrum of phenotypes including distinct DNA methylation patterns referred to as episignatures [12].

Given its role in epigenetic machinery and chromatin regulation [11] , we hypothesized that individuals with pathogenic *HNRNPU* variants would exhibit a specific DNA methylation episignature. DNA methylation episignature assessment using the EpiSign™ classifier can detect more than 56 episignatures associated with 65 disorders [13]. Episignature mapping may be applied in genome diagnostics to reclassify previously identified VUS (variants of unknown significance) in genes related to rare genetic disorders and can detect imprinted disorders such as Angelman and Beckwith-Wiedemann syndromes [14]. It has also been shown that copy number variants (CNVs) such as the 22q11.2 deletion syndrome show a unique episignature [15]. CNVs are variable in size and may involve multiple genes. The clinical outcome of CNVs can therefore be the result of the combined effects of disturbances to multiple genes some of which have been shown to influence the global DNA methylation patterns and episignature biomarkers in patients [16, 17].

In this study, we aimed to (1) find a DNAm episignature in patients with heterozygous pathogenic single nucleotide *HNRNPU* variants or deletions spanning *HNRNPU*, (2) validate it using an independent set of cases with pathogenic and VUS *HNRNPU* variants, and (3), compare the global DNA methylation profile associated with *HNRNPU* to previously described episignatures.

Methods

Subjects and study cohort

The study cohort includes ten participants (eight males and two females) with *HNRNPU* variants, four of which (case 3, 4, 8 and 9) have been previously described [2, 18, 19]. All participants were identified in a clinical diagnostic setting. Variants have been identified through chromosome microarray (CMA), whole exome, or targeted gene panels and were classified according to the guidelines of the American College of Medical Genetics (ACMG) [20]. Eight participants were identified with a pathogenic *HNRNPU* variant: (i) four had a frameshift variant, (ii) two a large deletion spanning several genes flanking

HNRNPU, (iii) one a deletion limited to *HNRNPU* only, (iv) and one had a splice-site variant. We also included one participant with an in-frame deletion (VUS), and a previously unsolved case from the EpiSign Knowledge Database (EKD) (case 10). For the in-frame deletion, bioinformatic protein structure analyses was based on different *in silico* tools; SIFT [21], MutPred-indel score [22]and Mutation taster [23].

Methylation analysis

DNA isolation from peripheral blood was performed according to standard techniques. DNA methylation analysis on the DNA samples were carried out using the Illumina Infinium methylation EPIC bead chip arrays (San Diego, CA, USA) according to manufacturer's protocols. Analysis and discovery of episignatures were carried out based on our laboratory's previously published protocols [13, 24, 25]. In brief, intensity data files (IDATs) containing methylated and unmethylated signal intensities were analyzed in R (version 4.1.1). Methylation data normalization was performed using the Illumina normalization method with background correction using the minfi package (version 1.40.0) in R [26]. The following probes were eliminated; probes with detection p value > 0.1, probes located on chromosomes X and Y, probes containing single nucleotide polymorphisms (SNPs) at or near the CpG interrogation or single nucleotide extension sites and probes that cross react with other genomic regions. Probes with beta values of 0 and the top 1% most variable probes were removed. DNA methylation assessment was performed 3 times; twice for episignature detection in biomarker discovery and once to assess the global *HNRNPU* DNA methylation profile as described in the results. Principal component analysis (PCA) was performed each time to examine batch structure and identify case or control outliers. Matched controls were randomly selected at a ratio of 1:5 from the EKD [14], matched by age, sex and array type using the MatchIt package (version 4.3.4) [27]. Methylation levels for each probe (beta values) were converted to M-values by logit transformation and subsequently applied in linear regression analysis (limma package, v3.50.0) to identify differentially methylated probes [28]. Estimated blood cell proportions were incorporated into the model matrix as confounding variables [29]. P-values were moderated using the eBayes function in the limma package [28].

Probe selection and training of the machine classifier

Selection of probes for the discovery and final episignatures were performed in three steps. First, 900 and 1000 probes (respectively) were retained with the highest product of absolute methylation differences between cases and controls and the negative of the logarithm of p-values. This was followed by a receiver's operating characteristic (ROC) curve analysis, and 450 and 333 probes were retained with the highest area under the ROC curve (AUC). Probes with pair-wise correlation greater than 0.65 and 0.70 measured using Pearson's correlation coefficients for all probes were eliminated. Unsupervised clustering models were applied using the remaining probes, including hierarchical clustering (heatmap) using Ward's method on Euclidean distance in the ggplot2 package in R (v3.1.1) and multidimensional scaling (MDS) by scaling of the pair-wise Euclidean distances between samples. To assess the robustness of the episignatures, multiple rounds of leave-1-out cross validation were performed: in each round a single *HNRNPU* sample was used as a testing sample and the remainder used for probe selection. The corresponding unsupervised clustering plots were visualized. The e1071 R package (version 1.7-9) was used to train a support vector machine (SVM) classifier and to construct a multi-class prediction model as previously described [13, 24].

Assessment of differentially methylated regions (DMRs)

Differentially methylated regions (DMRs) were detected using the DMRcate package in R (v 2.8.3), with p-cutoff set to default (FDR) [30]. Regions containing at least 5 significantly different CpGs within 1kb, with a minimum mean methylation difference of 5% and a Fisher's multiple comparison p-value < 0.01 were considered significant. DMRs were annotated using the UCSC Genome Browser Data Integrator with GENCODE V3lift37 comprehensive annotations (https://genome.ucsc.edu) and further characterized using the following USCS Genome Browser tracks; UCSC Genes, CpG Islands, H1-hESC Genome Segmentation by Combined Segway+ChromHMM selection from the Genome Segments track and the H3K27Ac Mark from the ENCODE regulation track.

Functional annotation of the global *HNRNPU* DNA methylation profile and correlation to the 56 EpiSign™ conditions

Functional annotation and EpiSign™ cohort comparisons were performed according to our previously published methods [25]. In short, heatmaps and circos plots were produced to assess the percentage of differentially methylated probes (DMPs) shared between the *HNRNPU* episignature and the 56 other neurodevelopmental conditions on the EpiSign™ clinical classifier. Heatmaps were plotted using the R package pheatmap (version 1.0.12) and circos plots using the R package circlize (version 0.4.15) [31]. To investigate relationships across all 57 cohorts without bias due to the number of DMPs selected, clustering analysis was performed on the combined top N DMPs for each cohort as described previously by Levy *et al* [25]. This assessed the top 500 DMPs for each cohort, ranked by p-value. For cohorts with less than 500 DMPs, all DMPs were used. The distance and similarities between cohorts were visualized on a tree and leaf plot. Tree and leaf plots were generated using the R package TreeAndLeaf (version 1.6.1) showing additional information including global mean methylation difference and total number of DMPs identified for each cohort.

To determine the genomic location of the DMPs, probes were annotated in relation to CpG islands (CGIs) and genes using the R package annotatr (version 1.20.0) [32] with AnnotationHub (version 3.2.2) and annotations hg19_cpgs, hg19_basicgenes, hg19_genes_intergenic, and hg19_genes_intronexonboundaries. CGI annotations included CGI shores from 0-2kb on either side of CGIs, CGI shelves from 2-4kb on either side of CGIs, and inter-CGI regions encompassing all remaining regions. For gene annotations, promoters included up to 1Kb upstream of the transcription start site (TSS) and promoter+ the region 1-5Kb upstream of the TSS. Annotations to untranslated regions (5'-UTR and 3' UTR), exons, introns, and exon/intron boundaries were combined into the "gene body" category.

Results

Identification and assessment of the *HNRNPU* episignature

The molecular details of our study cohort are summarized in Table 1 and Figure 1. All participants carried an *HNRNPU* variant (frameshift, in-frame deletion, deletion and splice) or large CNVs including *HNRNPU*.

Table 1 Molecular details of our cohort

Participant	ClinVar SCV code	Cohort type	Variant type	Nucleotide change	Amino acid change	Classification	Inheritance	Test
1		discovery	frameshift	c.906_907 del	p.(Asp304Serfs*33)	P	dn	genpanel
2		discovery	in-frame deletion	c.216_2478+8418del	-	P	dn	ES
3		discovery	splice	c.2425-3C>A	p.(?)	LP	dn	ES
4		discovery	frameshift	c.2304_2305del	p.(Gly769Glufs*83)	P	dn	ES
5		discovery	deletion	1q43q44(242045197_249212668)x1	-	P	dn	CMA
6		discovery	deletion	1q43q44(244571975_246704522)x1	-	P	dn	CMA
7		validation	frameshift	c.2072del	p.(Asn691Iefs*143)	P	not maternal	ID-panel
8		validation	frameshift	c.16delinsATT	p.(Val6Ilefs*4)	P	paternal mosaicism	ES
9		Validation_VUS	in-frame deletion	c.837_839del	p.(Glu279del)	VUS	dn	ES
10		unresolved case	in-frame deletion	c.1720_1722del	p.(Lys574del)	P	dn	ES

P pathogenic, LP likely pathogenic, dn de novo, CMA chromosome micro array, ES exome sequencing. Variants are based on (NM_031844.2/GRCh37)

Commentato [KR1]: Slavica to provide codes when ready

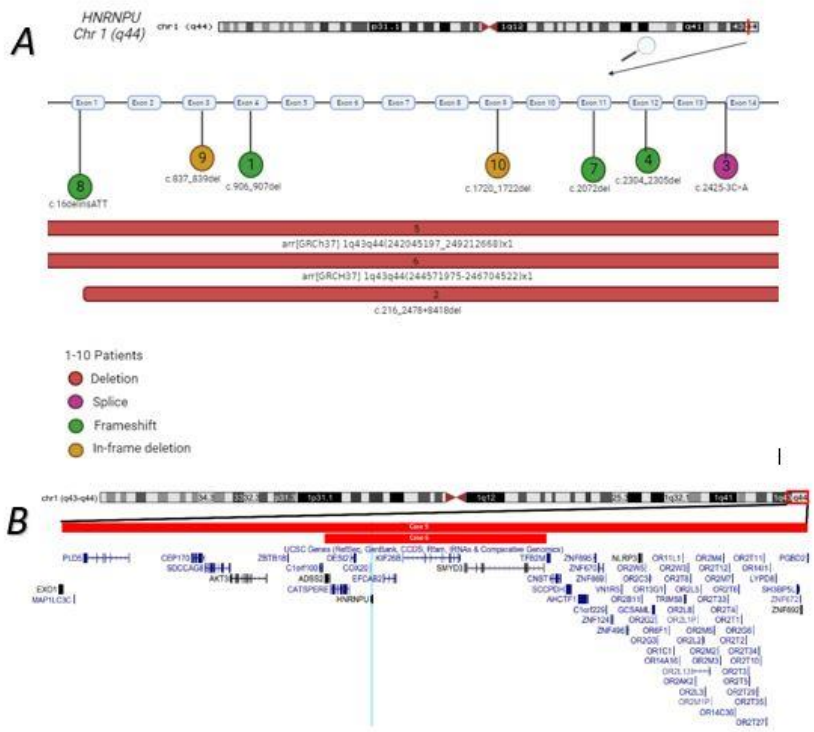


Figure 1 ***A.*** *Participants genetic information. The numbers match with the numbers in the table and figures. Comparison between the participants with deletions (red square), splice (purple circle), frameshift (green circle) and in-frame deletion (yellow circle) variants. Alamut Visual version NM_031844.2 HNRNPU. Created with BioRender.com* ***B.*** *Large deletions of chromosome region 1q43q44 in Cases 5 and 6 are represented by the horizontal red bars and the genes contained within the region listed below. Also highlighted in light blue is the location of HNRNPU. Cytogenetic bands and known genes are presented in this figure using the UCSC genome browser 2009 (GRCh37/hg19) genome build [33].*

Table 2 Clinical details of the HNRNPU cohort

Participant	1	2	3	4	5	6	7	8	9	10
General information										
Gender	F	M	M	F	M	M	M	M	M	M
Age (years)	6	31	28	11	20	28	29	14	17	3
Clinical information										
Growth										
Age at assessment (years)	11 m	28	21	9	8	28	17	14	15	2y11m
Microcephaly	-	-	-	+	+	NA	NA	+	-	-
Macrocephaly	-	+	-	-	-	NA	NA	-	-	-
Development/ behaviour										

intellectual dysfunction	+	+	+	+	+	+	+	+	+	+
Developmental delay	+	+	+	-	+	+	+	+	+	+
Motor delay	+	+	+	-	+	+	+	+	+	+
Language delay	+	+	+	+	+	+	+	+	+	+
Behavior abnormalities	+	+	NA	-	NA	+	+	-	-	+
Autistic features	+	+	+	-	NA	+	+	NA	-	+
ASD diagnosis	-	NA	NA	-	NA	-	+	-	NA	-
Neurologic symptoms										
Hypotonia	+	+	-	-	NA	+	+	+	-	+
Gait disturbance	+	NA	NA	-	NA	+	+	NA	-	NA
Seizures	+	+	-	+	NA	+	+	+	+	-
Epilepsy	+	+	+	+	NA	+	+	+	+	-
MRI abnormalities	+	+	NA	NA	NA	-	+	+	-	+
Dysmoprhism										
Craniofacial dysmoprhism	+	+	+	-	NA	+	+	+	+	+
Bulbous nasal tip	+	-	+	+	NA	+	NA	+	-	-
Anteverted nares	+	-	+	-	NA	+	NA	-	-	+
Short nose	+	-	+	+	NA	-	NA	+	-	+
Hypertelorism	NA	-	+	-	NA	-	NA	+	-	-
Deep set eyes	+	+	+	-	NA	-	NA	-	-	+
Limbs/ hands	NA	+	+	+	NA	-	NA	-	+	+
Other										
Cardiac anomalies	NA	-	-	-	NA	-	+	+	NA	-
Musculoskeletal anomalies	NA			-	NA	+	NA	+	NA	-
Dental anomalies	NA	+	-	-	NA	-	+	NA	-	NA
Cleft lip/palate	+	NA	-	-	NA	-	NA	-	-	-
Eye/vision anomalies	NA	-	-	-	NA	-	NA	-	-	-
Cutaneous anomalies	+	+	+	-	NA	-	NA	+	+	-
Perinatal complications	+	+	+	-	NA	-	NA	-	+	-
Other	+	-	-		NA	+	NA	+	-	-

NA not available, F Female, M Male, P pathogenic, ASD autism spectrum disorder.

Clinical summary

The clinical details of our cohort are summarized in Table 2. All patients presented with Intelectual dysfunction, language delay and dysmoprhism. Below are short summaries of our participants, more detailed reports are given in the supplemental clinical information document. The available facial photographs are shown in Figure 2.

Participant 1

Commentato [LLvd(2]: Kathleen;
Maybe a short clinical summary paragraph on each patient that we have more details for? This is just a thought, if we have information like age of presentation, presenting features etc.

Commentato [LLvd(3]: Peter zou jij kunnen kijken of dit een beetje goed is, Kathleen en ik dachten dat dit zo mooi was, niet te uitgebreid. Feel free to change.

She was referred to the clinical geneticist at 11 months old for evaluation of hypotonia, gait disturbance and seizures. Her medical history was notable for pyelonephritis with vesicoureteral reflux and a right-sided double collecting system, for which she was admitted to a pediatric clinic at 11 months old. Electroencephalography recorded multifocal epileptic abnormalities. An MRI of the brain showed a very prominent sagittal sinus and CT scan of the brain revealed a hypoplastic jugular foramen.). A pathogenic de novo c.906_907del p.(Asp304Serfs*33) *HNRNPU* variant was identified at genetic investigation, employing ID (intellectual disability) gene panel sequencing.

Participant 2

He was referred to the clinical geneticist at 28 years old for evaluation of ID. His medical history noted intra-uterine growth retardation and feeding difficulties. A ventriculo-peritoneal drain was placed at 3 months as treatment of hydrocephalus. At age 15 years he was treated by percutaneous epiphysiodesis of the right knee because of a leg length discrepancy. An MRI of the brain showed hydrocephalus, possibly due to aqueductal stenosis and a periventricular cyst, possibly post-hemorrhagic. A pathogenic c.216_2478+8418del p.(?) *HNRNPU* variant was identified with an ID gene panel sequencing.

Participant 3

He was referred to the clinical geneticist at 21 years old for evaluation of ID and generalized tonic clonic seizures. A likely pathogenic de novo c.2425-3C>A p.(?) *HNRNPU* variant was identified with ES (exome sequencing).

Participant 4

She was referred to the clinical geneticist at 15 months old for evaluation of ID and tonic-clonic generalized seizures with an abnormal electroencephalogram (EEG). ES revealed a pathogenic de novo variant c.2304_2305del p.(Gly769Glufs*83) *HNRNPU* variant.

Participant 5

He was referred to the clinical geneticist at 8 years old for evaluation of ID and microcephaly. A pathogenic *de novo* (1q43q44(242045197-249212668)x1) *HNRNPU* CNV was found with chromosome micro-array analysis (CMA).

Participant 6

He was referred to the clinical geneticist at 28 years for evaluation of ID, dysmorphic features and febrile generalized tonic clonic seizures. A pathogenic de novo (1q43q44(244571975-246704522)x1) *HNRNPU* CNV was found with CMA analysis.

Participant 7

His medical history noted a heart murmur that was detected after birth and then a VSD/ASD on ultrasound. She was referred to the clinical geneticist at 17 years for evaluation of ID, seizures, and epilepsy. A pathogenic c.2072del p.(Asn691Iefs*143) *HNRNPU* variant was found with ID gene panel sequencing. The variant was not maternal, and father was not tested

Participant 8

He was referred to the clinical geneticist at 6 years old for evaluation of ID and febrile seizures. An MRI of the brain revealed enlarged lateral ventricles. A pathogenic c.16delinsATT, p.(Val6Ilefs*4) *HNRNPU* variant was found with ES. The variant was identified in father in a mosaic state.

Participant 9

She was referred to the clinical geneticist at 15 years old for evaluation of ID, seizures and dysmorphic features. ES revealed a de novo variant of uncertain significance (VUS) in *HNRNPU*: c.837_839del p.(Glu279del).

Participant 10

He was referred to the clinical genetics at 2 years old for ID, dysmorphism and hypotonia. An MRI indicated prominent perivascular spaces, and a mildly abnormal dens. A pathogenic *de novo* c.1720_1722del p.(Lys574del) *HNRNPU* variant was found by ID gene panel sequencing.

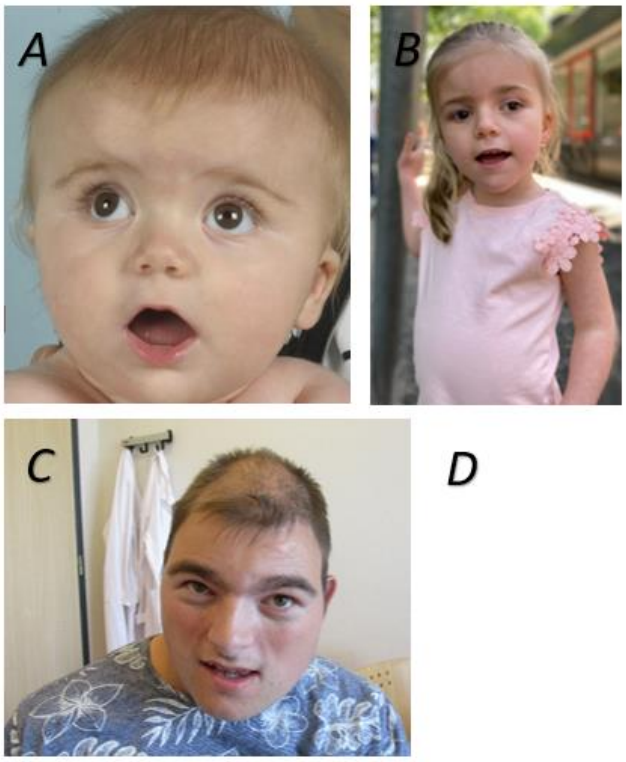


Figure 2. *Facial appearance of our participants with variants in HNRNPU. (**A**) Participant 1 at 11 months old. (**B**) Participant 1 at age 6 yrs. (**C**) Participant 2 at age 28 yrs. (**D**) Participant 7 at age 17 yrs.*

Identification of a DNA methylation episignature

We set out to determine if individuals with pathogenic variants in *HNRNPU* would exhibit a differential DNA methylation pattern compared to unaffected controls. For the discovery episignature, 123

differentially methylated CpG probes were retained that could be used as a biomarker to distinguish between affected and unaffected samples (Supplementary Table 1). Hierarchical clustering (heatmap) and multidimensional scaling methods confirmed that *HNRNPU* cases clustered apart from controls (Figure 2 A-B). Next, based on multiple rounds of leave-1-out cross validation, we observed a robust episignature visualized using unsupervised clustering methods, i.e., heatmap and MDS plots (Supplementary Figure 1). Next, we constructed a support vector machine (SVM) classifier model, including 56 other neurodevelopmental disorder conditions, that indicated a high level of specificity of the *HNRNPU* episignature. Herein, all *HNRNPU* cases showed an MVP score close to 1, compared to all 56 others that showed an MVP score at or close to 0 (Figure 3 C).

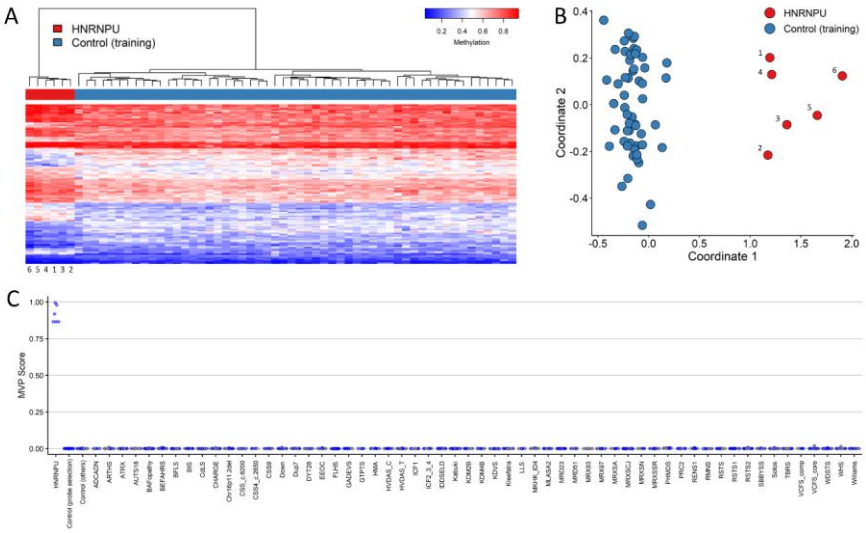


Figure 3. *HNRNPU episignature.* ***A.*** *Heatmap shows clustering of the HNRNPU training cases (red) away from age and sex matched controls (blue); each column represents a single case or control sample, and each row represents one of the 123 CpG probes selected for the episignature.* ***B.*** *Multidimensional scaling (MDS) plot shows clustering of HNRNPU cases from controls. The x and y axis represent the first and second dimension of the output (Coordinate 1 and 2 respectively).* ***C.*** *Support Vector Machine*

(SVM) classifier model, the x-axis represents an episignature on the EpiSign™ classifier and the y-axis a probability score, referred to as a Methylation Variant Pathogenicity score (MVP). The model was trained using the 123 selected HNRNPU episignature probes, 75% of controls and 75% of other neurodevelopmental disorder samples (blue). The remaining 25% controls and 25% of other disorder samples were used for testing (grey). Plot shows the HNRNPU discovery cases with an MVP score close to 1 compared with all other samples that are at or close to 0, showing the specificity of the classifier and episignature.

Validation of the *HNRNPU* episignature

To validate *HNRNPU*-associated episignature, we evaluated two additional cases with pathogenic variants in this gene (case 7 and 8) and a participant carrying the c.837_839del p.(Glu279del) VUS (case 9).

We visualized these results using unsupervised hierarchical and MDS clustering methods. We confirmed that both samples carrying pathogenic variants clustered with the *HNRNPU* episignature training cases and away from controls (Figure 4 A-B). In contrast, the assessment of the sample carrying a VUS in *HNRNPU* (case 9) showed that it did not cluster with the episignature cases. Indeed, a re-evaluation of c.837_839del showed two cases in the Genome Aggregation Database (gnomAD ver 2.1.1), and discordant pathogenic *in silico* predictions (SIFT: damaging, score 0.667; MutPred-indel score 0.33463; MutationTaster2021 benign). Thus, this variant was reclassified as likely benign.

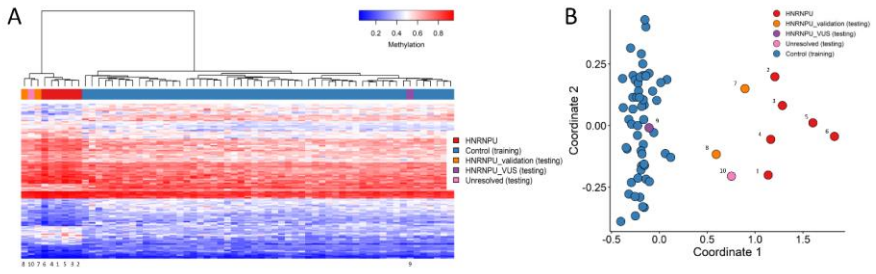


Figure 4. *Validation of the HNRNPU episignature.* ***A.*** *Heatmap; each column represents a single HNRNPU case or control, each row represents one of the 123 CpG probes selected for the episignature. This heatmap shows clustering of the 2 HNRNPU validation cases (orange) and a previously unresolved case (pink) with the 6 HNRNPU training cases (red) from controls (blue). The VUS sample (purple) is shown to cluster with controls.* ***B.*** *Multidimensional scaling (MDS) plot confirming the clustering of all pathogenic HNRNPU cases (training and validation) from controls. X and y axis represent the first and second dimension of the output (Coordinate 1 and 2 respectively).*

Screening of an unresolved case using the *HNRNPU* episignature

Using the SVM classifier constructed in Figure 2 C, unresolved cases in the EKD were screened using the *HNRNPU* episignature. Here, we identified a case (case 10) with an MVP score close to 1 that clustered with *HNRNPU* cases in both hierarchical clustering and MDS plots (Figure 4 A-B). Through subsequent follow up with the submitting clinical center we were able to confirm that the patient carried a variant in *HNRNPU* (c.1720_1722del p.(Lys574del)) (Table 1).

To improve the specificity of the detected episignature and determine a final episignature to be used as a biomarker for the EpiSign™ test, we repeated the mapping steps using the 9 confirmed pathogenic *HNRNPU* cases (6 training, 2 validation, 1 unresolved) as training samples against age and sex matched controls. We retained 106 differentially methylated CpG probes as clinical biomarkers. Results were visualized using the same clustering, SVM and cross validation methods and showed strong specificity and sensitivity (Supplementary Figure 2).

Genome-wide DNA methylation profiles of *HNRNPU* samples show an overall increase in DNA methylation

Next, we assessed genome-wide DNA methylation changes in participants with pathogenic *HNRNPU* variants and correlated these to the 56 other disorders reported by Levy *et al* [25]. Here we compared genome-wide DNA methylation profiles of the 9 training cases to episignature-negative, age and sex

matched controls from the EKD. We detected in total 4,780 differentially methylated probes (FDR <0.05) that predominantly showed a global hypermethylation profile (Supplementary File 1). In the 56 disorders in the Levy *et al* study, 66% (n=37) showed hypomethylation, in contrast to 34% (n=19) that showed global hypermethylation (Figure 5) [25].

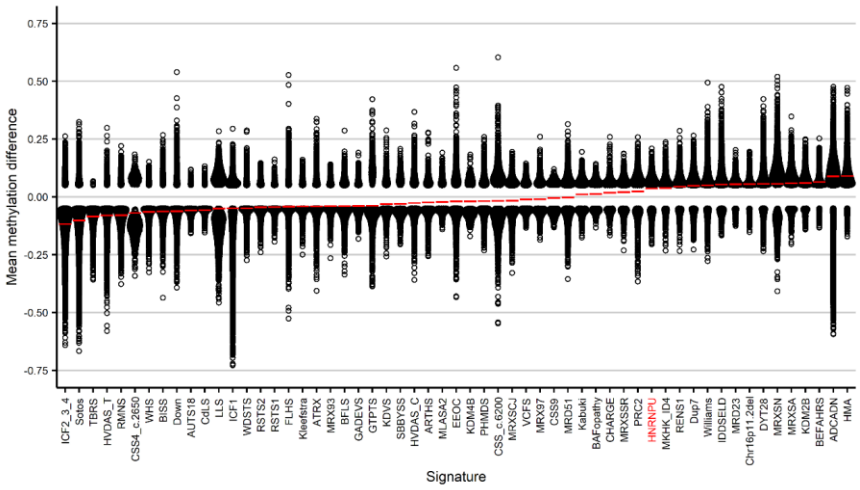


Figure 5. *Genome-wide DNA methylation profiles of the HNRNPU cohort and 56 disorders on EpiSign™. Global Methylation differences of all differentially methylated probes (DMPs, FDR < 0.05) for each cohort, sorted by mean methylation. Each circle represents one probe. Red lines indicate mean methylation. The x-axis represents one of the 57 episignatures and the y-axis is the mean methylation difference.*

Evaluation of the HNRNPU episignature compared to that of 56 NDDs

We compared the afore described list of DMPs from the *HNRNPU* cohort to the same DMP lists generated by Levy *et al* for 56 other EpiSign™ disorders [25]. Using a correlation matrix of DMPs showing the percentage of probes shared between each paired cohort, we observed that the *HNRNPU* cohort showed the highest overlap (~15%) with the Velocardiofacial syndrome (VCFS) and the BAFopathy cohorts. In addition, *HNRNPU* showed overlap with several other disorders, including ~13%

with Duplication 7q11.23 syndrome (Dup 7q11.23) and Luscan-Lumish syndrome (LLS; *SETD2*), ~12% in CHARGE syndrome (*CHD7*), and ~11% in Cornelia de Lange (CdLS; *NIPBL, SMC1A, SMC3, RAD21*), Intellectual developmental disorder X-linked 97 (MRX97), and Intellectual developmental disorder X-linked syndromic Claes-Jensen type (MRXSCJ) syndromes (Supplementary Figure 3). All other disorders showed a 10% or less overlap. These results were also visualized in a circos plot (Figure 6).

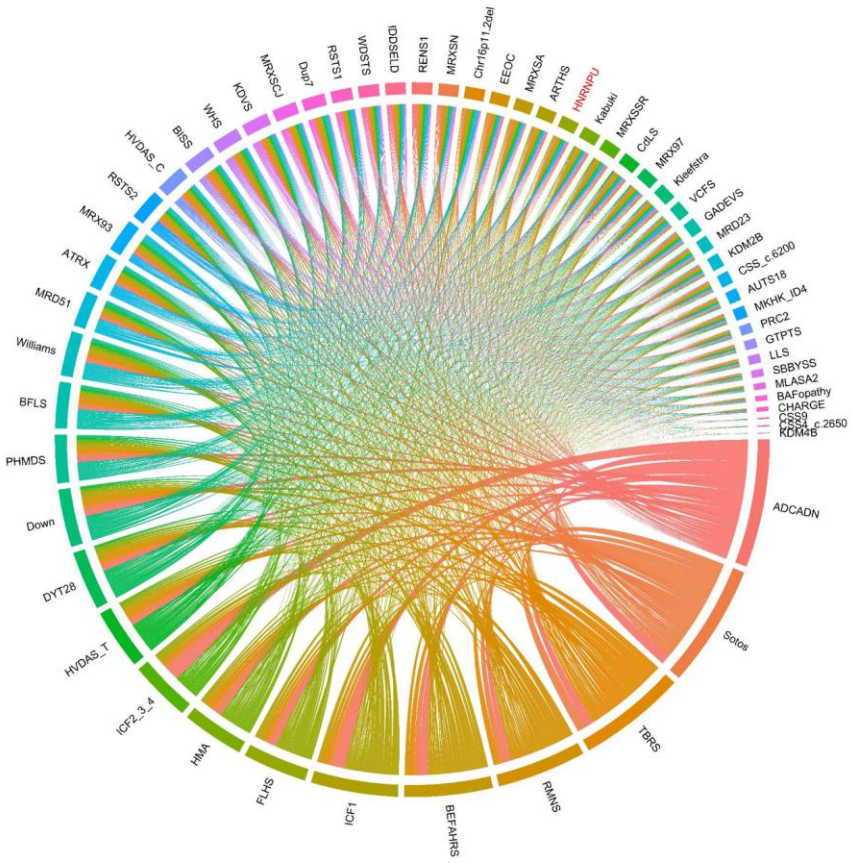


Figure 6. *Differentially methylated probes (DMPs) shared between the HNRNPU cohort and 56 other EpiSign™ disorders. Circos plot representing the probes shared between each pair of cohorts. The thickness of the connecting lines indicates the number of probes shared between the two cohorts.*

To determine the relatedness of each of the disorders, we visualized the DMP overlap as well as directionality of the change (hypo or hypermethylation) using a binary tree with each node corresponding to a cohort as described in the methods. Herein, we observed that *HNRNPU* clustered within a hypermethylation branch alongside Beck-Fahrner syndrome (BEFAHRS; *TET3*), KDM2B-related syndrome (*KDM2B*), Menke-Hennekam syndromes 1 and 2 (MKHK_ID4; *CREBBP, EP300*), Kabuki syndromes 1 and 2 (Kabuki; *KMT2D, KDM6A*), Intellectual developmental disorder autosomal dominant 23 (MRD23; *SETD5*), BAFopathy and Coffin-Siris syndrome-9 (CSS9; SOX11) (Figure 7). These shared DMPs may indicate an underlying biological process that is common between cohorts (disorders).

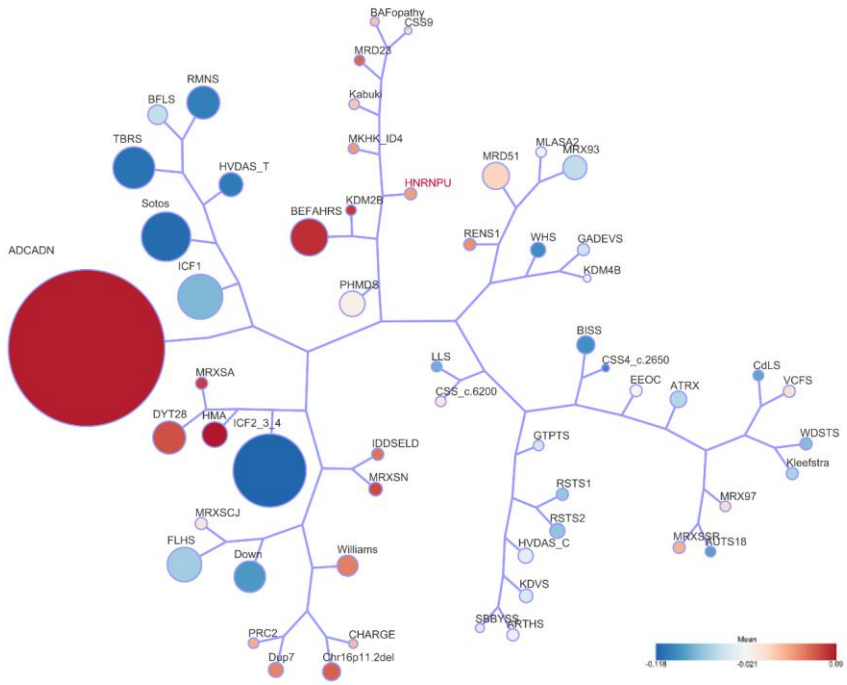


Figure 7. *Tree and leaf visualization of Euclidean clustering of all 57 cohorts using the top 500 DMPs for each group, (for cohorts with less than 500 DMPS all DMPS were used). Cohort samples were aggregated using the median value of each probe within a group. A leaf node represents a cohort, with node sizes illustrating relative scales of the number of selected DMPs for the corresponding cohort, and node colors are indicative of the global mean methylation difference (i.e. the overall methylation trend, hypomethylation (blue) or hypermethylation (red)).*

Lastly, we annotated the genomic locations of all the DMPs for all 57 cohorts in relation to genes and CpG islands. This demonstrated that the *HNRNPU* DMPs predominantly map within coding regions of genes (Figure 8 A) and almost equally to CpG island shore regions (within 0-2kb of a CpG island boundary) and regions outside of CpG islands (Figure 8 B). Similar to the other disorders, *HNRNPU* DMPs also map to intergenic regions.

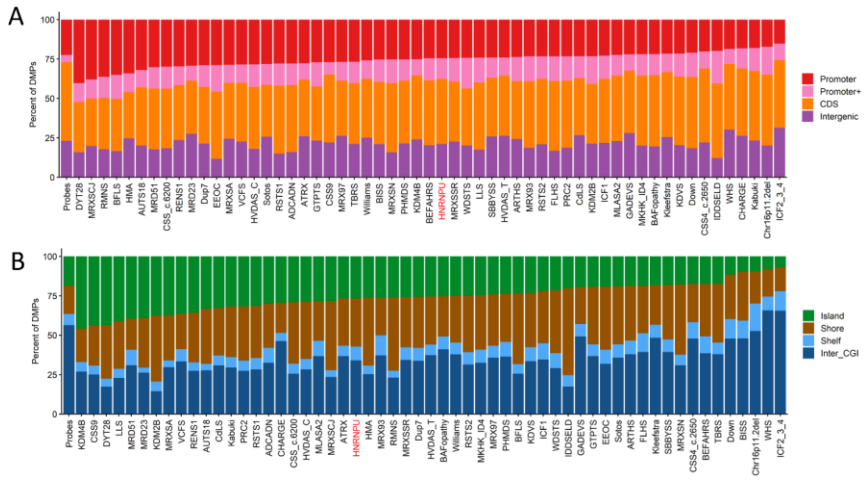


Figure 8. *Differentially methylated probes (DMPs) of HNRNPU cohort (red) and 56 EpiSign™ disorders. X-axis represents one of the 57 disorders (HNRNPU + 56 EpiSign™ disorders) and the y-axis the percentage of DMPs.* ***A.*** *DMPs annotated in the context of genes. Promoter, 0-1kb upstream of the*

transcription start site (TSS); Promoter+, 1-5kb upstream of the TSS; CDS, coding sequence; Intergenic, all other regions of the genome. ***B.*** *DMPs annotated in the context of CpG islands. Island, CpG islands; Shore, within 0-2kb of a CpG island boundary; Shelf, within 2-4kb of a CpG island boundary; Inter_CGI, all other regions in the genome. The Probes column in both A and B represents the background distribution determined in the Levy et al study [13] of all array probes after initial filtering and used as input for DMP analysis.*

Differentially methylated regions (DMRs)

We identified 18 DMRs (Supplementary Table 2), of which 12 were hypermethylation events (67%), in line with the overall global mean methylation difference and 6 were hypomethylation events (33%). Twelve DMRs were in regions with CpG islands (67%) (Supplementary Figure 4), 9 of these were hypermethylation events and 3 hypomethylation. Eight DMRs (44%) were annotated to predicted promoters with transcription start sites, 7 of these were in regions with CpG islands. Nine DMRs mapped to enhancers (50%), and 5 mapped to regions with no predicted regulatory elements (28%). Two of the DMRs (11%) were predicted to overlap CTCF binding sites suggestive of possible disruption to chromatin loop formation and TADs associated with these regions. Five of the DMRs were located on chromosome 19 (28%), two DMRs each on chromosomes 1, 2 and 13 (11% each), and one DMR on chromosomes 4, 5, 9, 11, 14, 16 and 22. One hypermethylated DMR containing an enhancer and CTCF binding site overlaps the *CHKB* gene that is associated with congenital muscular dystrophy megaconial type (OMIM #602541) (Supplementary Figure 4). This disorder shares several phenotypic manifestations similar to DEE54 including microcephaly, hypotonia, ID, delayed motor development, poor speech and seizures.

~~Bioinformatics predictions for the c.837_839del variant~~

~~Variant c.837_839del p.(Glu279del) (case 9), previously reported as VUS, was reclassified as likely benign. The variant was also in two cases of thee Genome Aggregation Database (gnomAD ver 2.1.1),~~

~~and *in silico* predictions were not supportive for a pathogenic role (SIFT: damaging, score 0.667 [21], MutPred indel score 0.33463 [22], MutationTaster2021 benign [23]).~~

Discussion

The aberrant DNA methylation patterns, as consequence of genetics defect, are thought to be established during early embryonic development and can be detected in blood, making them easily accessible as biomarkers in clinical diagnostics [12]. DNA methylation profiles can be used to help identify episignatures associated with Mendelian neurodevelopmental disorders, diagnose unsolved cases with unexplained intellectual deficit, and reclassify VUSs [24, 34, 35].

The aim of this study was to detect and validate a DNA episignature for variants of *HNRNPU* and to further explore the overlap with other Mendelian disorders that have known episignatures. We assessed DNA methylation profiles from peripheral blood of nine individuals with confirmed pathogenic variants in *HNRNPU*, including SNV and deletions spanning *HNRNPU.* The classification model was later validated in a separate cohort. During cross-validation of the final episignature, all the cases clustered together with the training cases and showed that *HNRNPU* episignature was robust and reproducible. The SVM model showed that the selected probes represented strong biomarkers enabling molecular diagnosis of DEE54. This model assessed whether the episignature was highly specific and sensitive and confirmed that our *HNRNPU* episignature was different from other EpiSign™ neurodevelopmental disorders.

Because episignatures have been shown to aid in the classification of variants of uncertain significance [36], we tested an individual with a *HNRNPU* VUS c.837_839del p.(Glu279del) (case 9) to determine if it mapped to the *HNRNPU* episignature. The methylation profile did not cluster within the *HNRNPU* episignature in the MDS plot and yielded a MVP score near zero, indicating the absence of a *HNRNPU* episignature. Indeed, this in-frame deletion was later reclassified as likely benign.

Commentato [AB4]: This part does not read fluently here. I added a couple of sentences above where you discussed Episign validation.

Commentato [AB5]: This part is in results.

Then the SVM classifier was applied to the unresolved cases in the EKD. We identified one case (case 10) with MVP score close to one that clustered with the *HNRNPU* cases in both hierarchical clustering and MDS plots. After subsequent follow up with the clinical center we were able to confirm that an *HNRNPU* variant was subsequently found after ES and also the phenotype of this participant was in line with this finding. The variant; c.1720_1722delAAG p.(Lys574del) was labeled as likely pathogenic according to the ACMG guidelines [20]. This study establishes *HNRNPU* episignature as a powerful tool that can be applied in diagnostics, screening, and variant interpretation. [14].

This cohort included two participants with large CNVs (case 5 and 6) involving multiple genes in addition to *HNRNPU*. Those participants were diagnosed with autosomal dominant Intellectual developmental disorder 22 also named chromosome 1q43-44 deletion syndrome (OMIM #612337). However, both cases clustered together with participants carrying deleterious single nucleotide variants (SNV) in *HRNNPU*. Case 5 involves a deletion of 7 Mb, involving also *ZBTB18* (OMIM #608433), one of the candidate genes for the NDD phenotype in 1q43-44 deletion syndrome (OMIM #612337) [18]. Case 6 had a smaller deletion which did not encompass *ZBTB18*. Taken together, this may indicate that *HNRNPU* is one of the syndrome-causing genes in the 1q43-44 deletion region [1, 18, 37, 38] and that the aberrant methylation is solely driven by *HNRNPU. HNRNPU* is known to be the main candidate for the epilepsy phenotype of patients with 1q43q44 deletion syndrome region [18] and ES analysis showed that *HNRNPU* variants are also responsible for the neurodevelopmental phenotypes [39].

We identified eighteen DMRs that were predominantly hypermethylation events. Nine hypermethylation events occurred across CpG islands in regions containing predicted promoters or enhancers. Hypermethylation of a region containing a CpG island, predicted enhancer and the *CHKB* gene was observed. *CHKB* is associated with congenital muscular dystrophy megaconial type (OMIM #602541), a disorder with several overlapping clinical features with DEE54 including microcephaly, hypotonia, ID, delayed motor development, poor speech, and seizures. Hypermethylation was also observed across the predicted promoters of several other genes for which there is currently no

associated Mendelian disorders. Further work would be required to investigate the potential for these regions and genes to be contributing to disease.

When we compared the DMPs of the previously mapped EpiSign™ disorders, the global *HNRNPU* DNA methylation profile showed the highest overlap with the VCFS and BAFopathy cohorts. Another report showed that patients with *HNRNPR* variants, another gene of the hnRNP family, exhibited significant expression changes to patients with *TBX1* mutations; a gene encompassed in the critical region of VCFS [40]. Other authors showed that *BRM* and *BRG1*, (aliases *SMARCA2* and *SMARCA4*) [41], important subunits of SWI/SNF complex, are involved in the splicing machinery by interacting with several RNA binding factors such as hnRNPU [42]. These studies could provide some possible functional insights into the DMP overlap between *HNRPNU* and the genes involved in VCFS or BAF complexes observed in this study. Examining the directionality of the *HNRNPU* methylation and its connection with *BEFAHRS*, *KDM2B* and other related disorders that showed a predominantly hypermethylation pattern (Figure 7), it was hard to draw conclusions, especially for genes coding for proteins with such pleiotropic and orchestrated functions [43]. Further data are needed to explain this conundrum.

A possible limitation of this study is the relatively small sample size. Due to the rarity of Mendelian neurodevelopmental disorders, episignatures are established first using a small number of affected individuals [13, 44]. Should more patients with specific genetic variants in *HNRNPU* become available in the future we can use those to increase the sensitivity and selectivity, by means of the detection of possible nested signatures in the case of CNVs that also may include additional genes that encode epigenetically active proteins.

Conclusion

In this study we show a specific and sensitive DNA methylation episignature for individuals carrying pathogenic variants in *HNRNPU* or a CNV including *HNRNPU* that can be used to assess and reclassify variants in *HNRNPU*. Global DNA methylation changes in patients with *HNRNPU* variants provide insights into the molecular pathways of epigenetic disruptions in this disorder.

Declarations

Acknowledgements

We would like to thank the participants and their families described in this study.

Ethics approval and consent to participate

The study was conducted in accordance with the regulations of the Western University Research Ethics Board (REB116108, and REB106302) and The Medical Ethical Committee (METC) of the Amsterdam UMC, location AMC. METC approval waived (anonymous study, further study in line with a clinical question).

Funding

Funding for this study is provided in part by the London Health Sciences Molecular Diagnostics Development Fund and Genome Canada Genomic Applications Partnership Program awarded to Bekim Sadikovic, Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) PRIN2020 code 20203P8C3X.

Consent for publication

We obtained written informed consent from the participants or their substitute decision maker to publish patients' clinical and genetic information.

Availability of data and materials

Raw DNA methylation data are not available due to institutional and ethics restrictions.

Competing interests

The authors declare no conflict of interest.

Authors' contributions

Conceptualization; KR, LVDL, MA, MMAM, BS, MMVH and PH.; Data Curation; KR, LVDL, SH, RR, MAL.; Formal analysis; KR, LVDL, SH, RR, MAL.; Investigation; ST, NV, PL, BPN, GT, CM, BK, TB, CMLVT, NV, JJVDS, RO, AF, GBF, MMI, RH, MA and MMVH.; Methodology; MMAM, BS, PH and MMVH.; Project administration; MMAM, BS, MMVH and PH.; Supervision; MMAM, BS, MMVH and PH.; Validation; KR and LVDL.; Visualization; KR and LVDL.; Writing-original draft; KR and LVDL.; Writing-review and editing; KR, LVDL, ST, PL, RH, MMAM, BS, MMVH and PH.

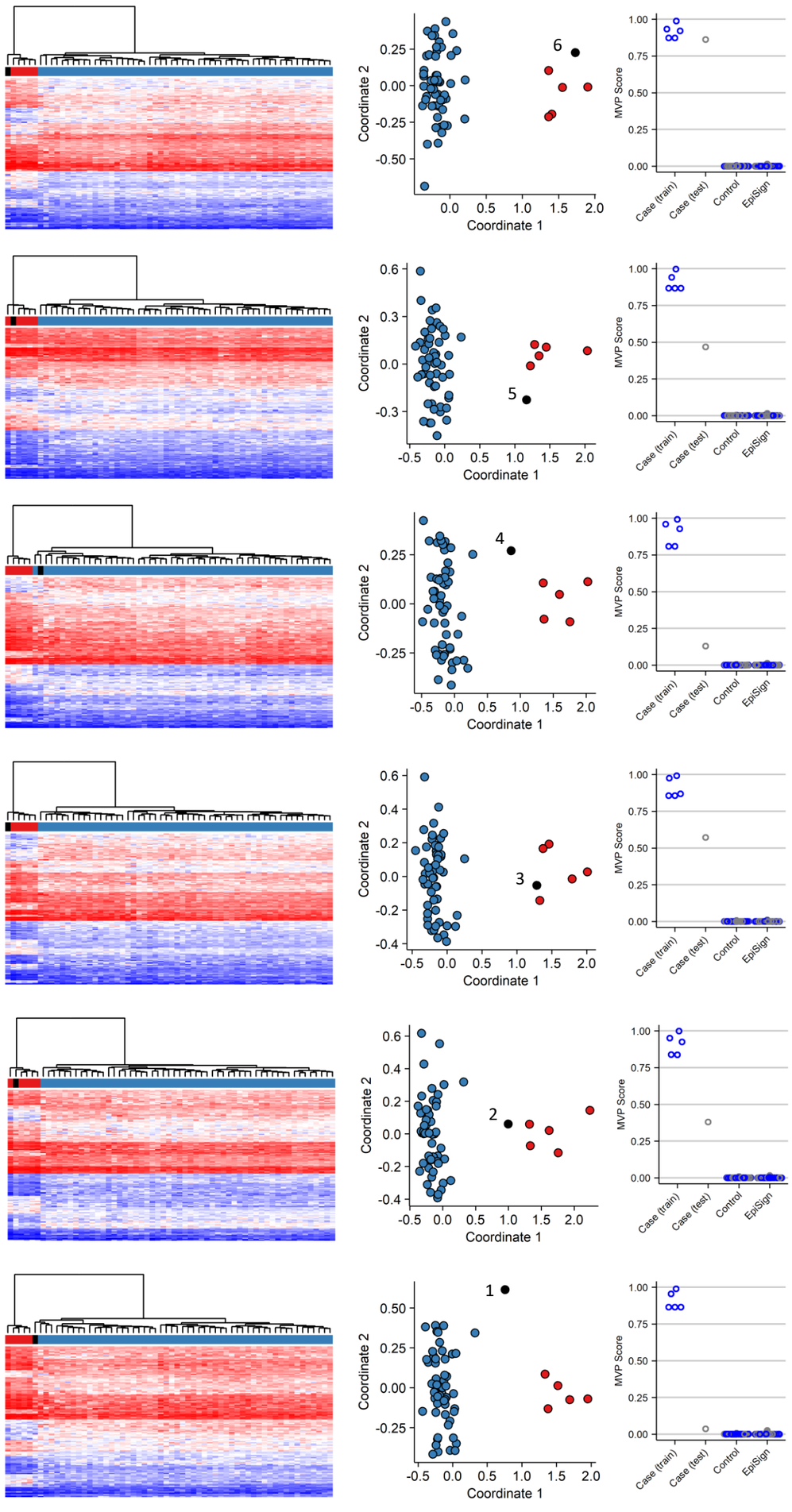
References

1. Bramswig NC, Lüdecke HJ, Hamdan FF, Altmüller J, Beleggia F, Elcioglu NH, et al. Heterozygous HNRNPU variants cause early onset epilepsy and severe intellectual disability. Hum Genet. 2017;136(7):821-34.
2. Taylor J, Spiller M, Ranguin K, Vitobello A, Philippe C, Bruel AL, et al. Expanding the phenotype of HNRNPU-related neurodevelopmental disorder with emphasis on seizure phenotype and review of literature. Am J Med Genet A. 2022;188(5):1497-514.
3. Balasubramanian M. HNRNPU-Related Neurodevelopmental Disorder. In: Adam MP, Everman DB, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle; 1993.
4. Wu B, Su S, Patil DP, Liu H, Gan J, Jaffrey SR, et al. Molecular basis for the specific and multivariant recognitions of RNA substrates by human hnRNP A2/B1. Nat Commun. 2018;9(1):420.
5. Lim I, Jung Y, Kim DY, Kim KT. HnRNP Q Has a Suppressive Role in the Translation of Mouse Cryptochrome1. PLoS One. 2016;11(7):e0159018.
6. Geuens T, Bouhy D, Timmerman V. The hnRNP family: insights into their role in health and disease. Hum Genet. 2016;135(8):851-67.
7. Sapir T, Kshirsagar A, Gorelik A, Olender T, Porat Z, Scheffer IE, et al. Heterogeneous nuclear ribonucleoprotein U (HNRNPU) safeguards the developing mouse cortex. Nat Commun. 2022;13(1):4209.
8. Yugami M, Okano H, Nakanishi A, Yano M. Analysis of the nucleocytoplasmic shuttling RNA-binding protein HNRNPU using optimized HITS-CLIP method. PLoS One. 2020;15(4):e0231450.
9. Zhang L, Song D, Zhu B, Wang X. The role of nuclear matrix protein HNRNPU in maintaining the architecture of 3D genome. Semin Cell Dev Biol. 2019;90:161-7.
10. Bjornsson HT. The Mendelian disorders of the epigenetic machinery. Genome Res. 2015;25(10):1473-81.
11. Fahrner JA, Bjornsson HT. Mendelian disorders of the epigenetic machinery: tipping the balance of chromatin states. Annu Rev Genomics Hum Genet. 2014;15:269-93.
12. Sadikovic B, Aref-Eshghi E, Levy MA, Rodenhiser D. DNA methylation signatures in mendelian developmental disorders as a diagnostic bridge between genotype and phenotype. Epigenomics. 2019;11(5):563-75.
13. Levy MA, McConkey H, Kerkhof J, Barat-Houari M, Bargiacchi S, Biamino E, et al. Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. HGG Adv. 2022;3(1):100075.
14. Aref-Eshghi E, Rodenhiser DI, Schenkel LC, Lin H, Skinner C, Ainsworth P, et al. Genomic DNA Methylation Signatures Enable Concurrent Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental Syndromes. Am J Hum Genet. 2018;102(1):156-74.
15. Rooney K, Levy MA, Haghshenas S, Kerkhof J, Rogaia D, Tedesco MG, et al. Identification of a DNA Methylation Episignature in the 22q11.2 Deletion Syndrome. Int J Mol Sci. 2021;22(16).
16. Rooney K, Sadikovic B. DNA Methylation Episignatures in Neurodevelopmental Disorders Associated with Large Structural Copy Number Variants: Clinical Implications. International Journal of Molecular Sciences. 2022;23(14):7862.
17. van der Laan L, Rooney K, Trooster TM, Mannens MM, Sadikovic B, Henneman P. DNA methylation episignatures: insight into copy number variation. Epigenomics. 2022;14(21):1373-88.
18. Depienne C, Nava C, Keren B, Heide S, Rastetter A, Passemard S, et al. Genetic and phenotypic dissection of 1q43q44 microdeletion syndrome and neurodevelopmental phenotypes associated with mutations in ZBTB18 and HNRNPU. Hum Genet. 2017;136(4):463-79.
19. Durkin A, Albaba S, Fry AE, Morton JE, Douglas A, Beleza A, et al. Clinical findings of 21 previously unreported probands with HNRNPU-related syndrome and comprehensive literature review. Am J Med Genet A. 2020;182(7):1637-54.

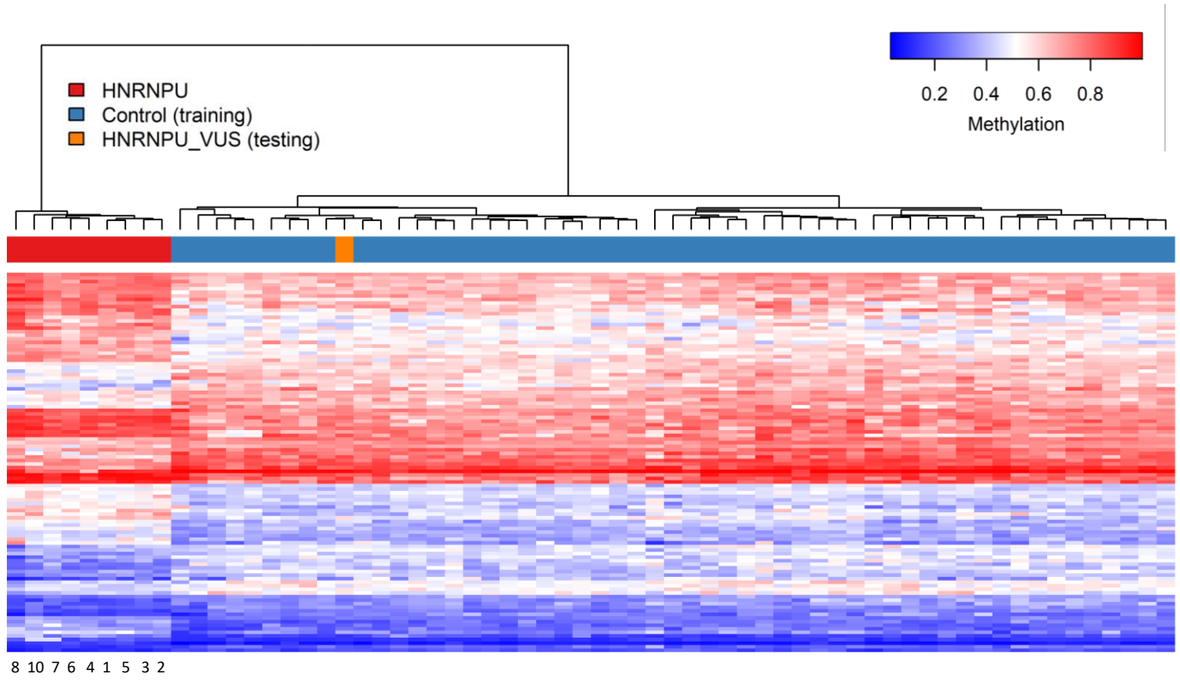
20. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.
21. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nat Protoc. 2016;11(1):1-9.
22. Pagel KA, Antaki D, Lian A, Mort M, Cooper DN, Sebat J, et al. Pathogenicity and functional impact of non-frameshifting insertion/deletion variation in the human genome. PLoS Comput Biol. 2019;15(6):e1007112.
23. Steinhaus R, Proft S, Schuelke M, Cooper DN, Schwarz JM, Seelow D. MutationTaster2021. Nucleic Acids Res. 2021;49(W1):W446-w51.
24. van der Laan L, Rooney K, Alders M, Relator R, McConkey H, Kerkhof J, et al. Episignature Mapping of TRIP12 Provides Functional Insight into Clark–Baraitser Syndrome. International Journal of Molecular Sciences. 2022;23(22):13664.
25. Levy MA, Relator R, McConkey H, Pranckeviciene E, Kerkhof J, Barat-Houari M, et al. Functional correlation of genome-wide DNA methylation profiles in genetic neurodevelopmental disorders. Hum Mutat. 2022.
26. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-9.
27. Ho D, Imai K, King G, Stuart EA. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software. 2011;42(8):1 - 28.
28. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
29. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.
30. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, R VL, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics Chromatin. 2015;8:6.
31. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. Bioinformatics. 2014;30(19):2811-2.
32. Cavalcante RG, Sartor MA. annotatr: genomic regions in context. Bioinformatics. 2017;33(15):2381-3.
33. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12(6):996-1006.
34. Aref-Eshghi E, Bend EG, Colaiacovo S, Caudle M, Chakrabarti R, Napier M, et al. Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. Am J Hum Genet. 2019;104(4):685-700.
35. Bend EG, Aref-Eshghi E, Everman DB, Rogers RC, Cathey SS, Prijoles EJ, et al. Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome. Clin Epigenetics. 2019;11(1):64.
36. Sadikovic B, Levy MA, Kerkhof J, Aref-Eshghi E, Schenkel L, Stuart A, et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. Genet Med. 2021;23(6):1065-74.
37. van der Schoot V, de Munnik S, Venselaar H, Elting M, Mancini GMS, Ravenswaaij-Arts CMA, et al. Toward clinical and molecular understanding of pathogenic variants in the ZBTB18 gene. Mol Genet Genomic Med. 2018;6(3):393-400.
38. Caliebe A, Kroes HY, van der Smagt JJ, Martin-Subero JI, Tönnies H, van 't Slot R, et al. Four patients with speech delay, seizures and variable corpus callosum thickness sharing a 0.440 Mb deletion in region 1q44 containing the HNRPU gene. Eur J Med Genet. 2010;53(4):179-85.
39. Zhu X, Petrovski S, Xie P, Ruzzo EK, Lu YF, McSweeney KM, et al. Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios. Genet Med. 2015;17(10):774-81.

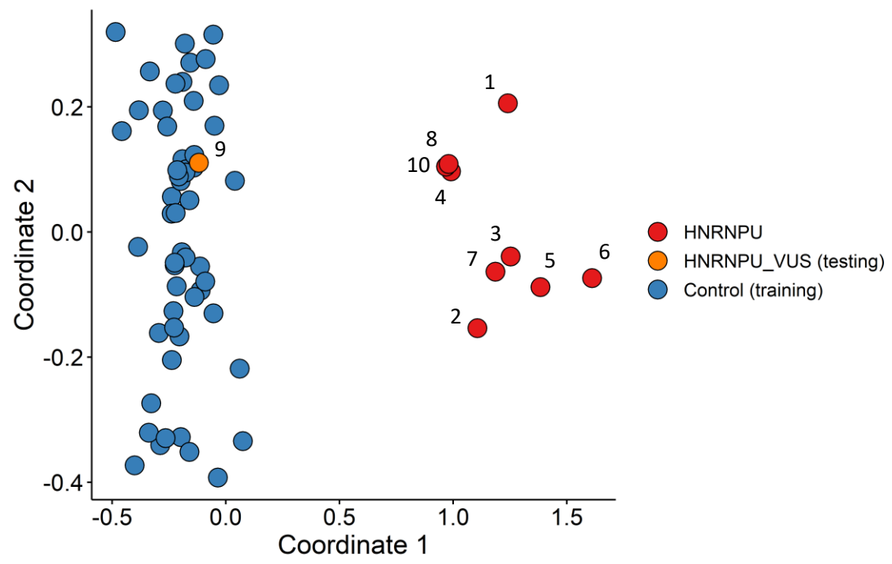
40. Duijkers FA, McDonald A, Janssens GE, Lezzerini M, Jongejan A, van Koningsbruggen S, et al. HNRNPR Variants that Impair Homeobox Gene Expression Drive Developmental Disorders in Humans. Am J Hum Genet. 2019;104(6):1040-59.
41. Centore RC, Sandoval GJ, Soares LMM, Kadoch C, Chan HM. Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. Trends Genet. 2020;36(12):936-50.
42. Gañez-Zapater A, Mackowiak SD, Guo Y, Tarbier M, Jordán-Pla A, Friedländer MR, et al. The SWI/SNF subunit BRG1 affects alternative splicing by changing RNA binding factor interactions with nascent RNA. Mol Genet Genomics. 2022;297(2):463-84.
43. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol. 2019;20(10):590-607.
44. Verberne EA, van der Laan L, Haghshenas S, Rooney K, Levy MA, Alders M, et al. DNA Methylation Signature for JARID2-Neurodevelopmental Syndrome. International Journal of Molecular Sciences. 2022;23(14):8001.

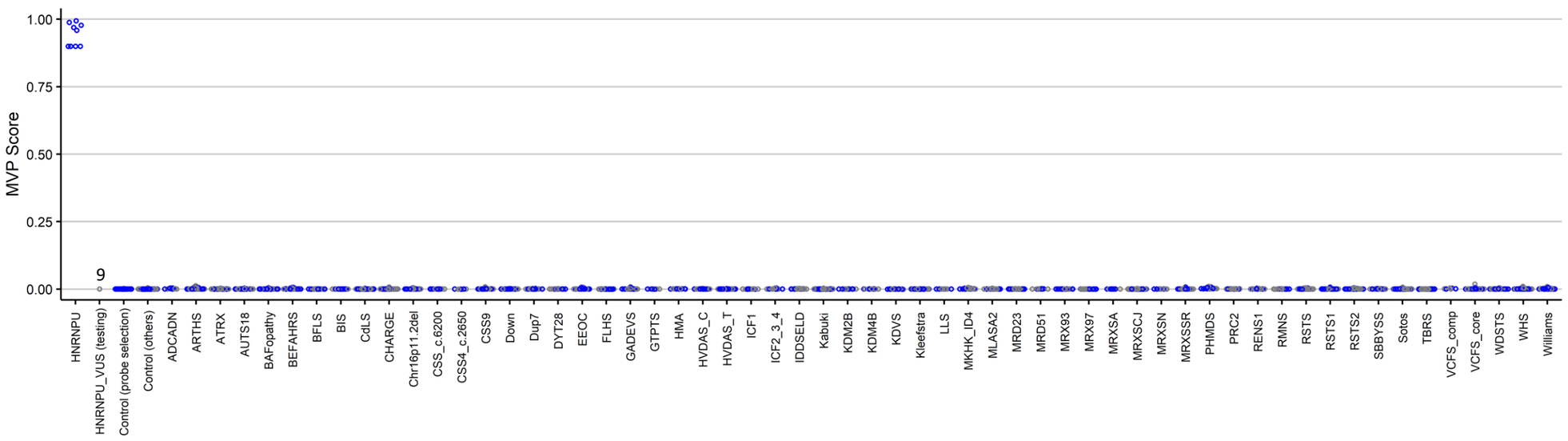
Supplementary Figure 1: Leave-1-out cross validation

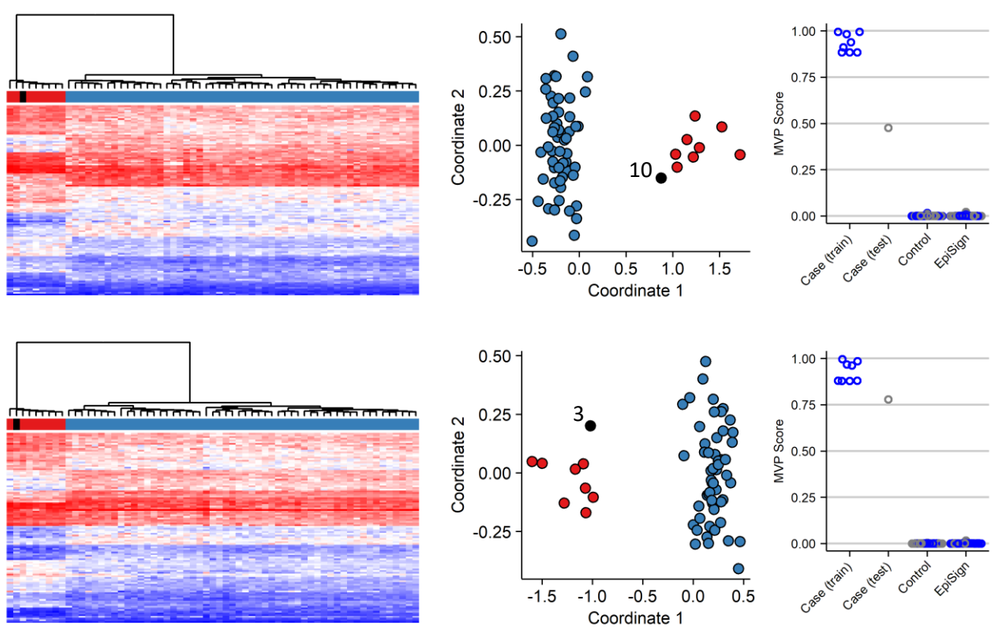


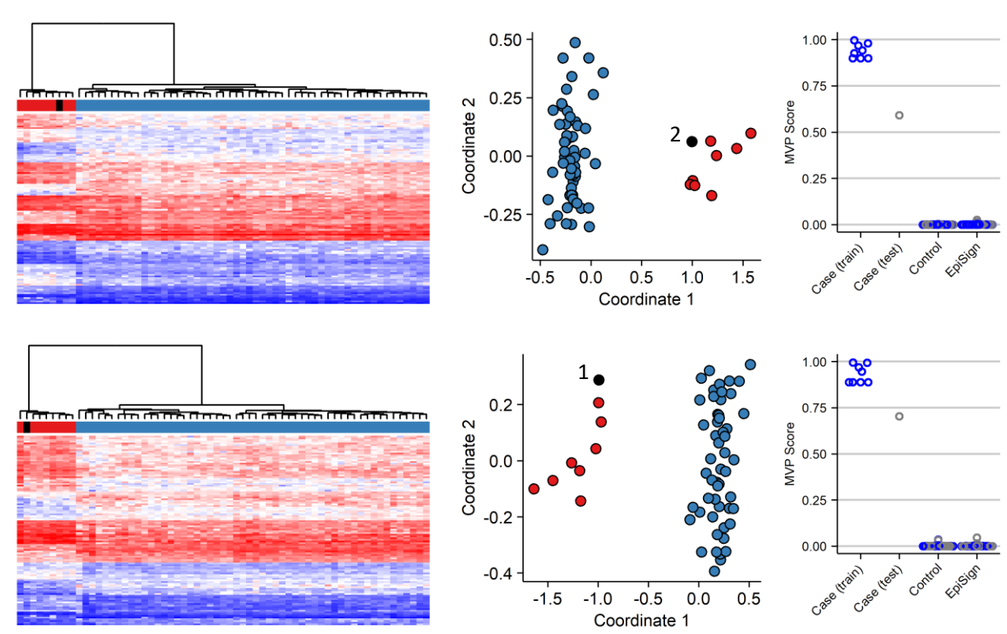
Supplementary Figure 2: *HNRNPU* Episignature mapping

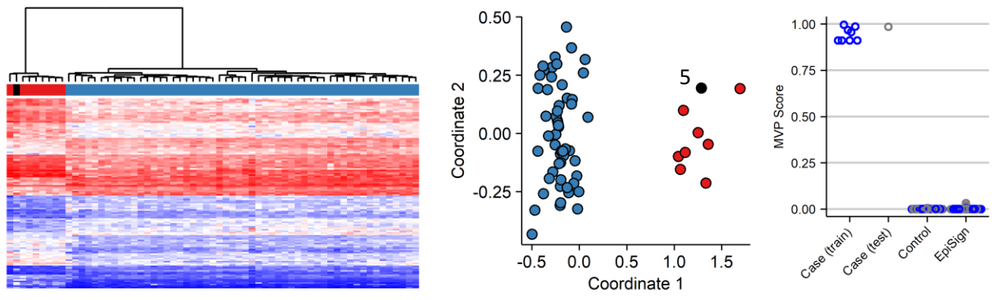


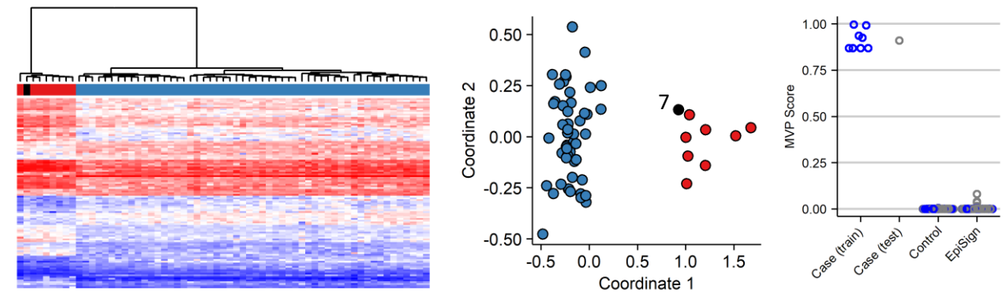


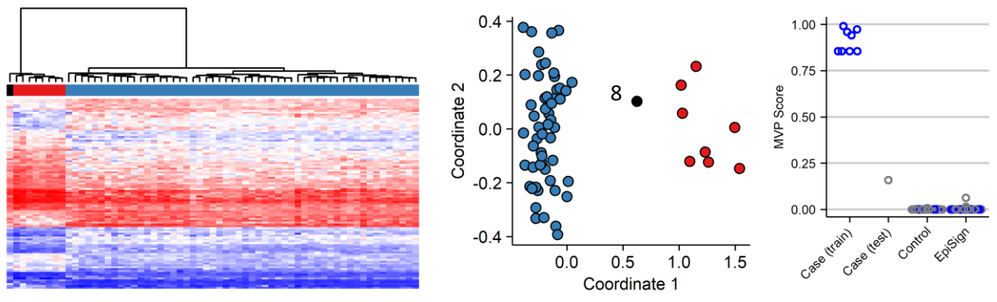


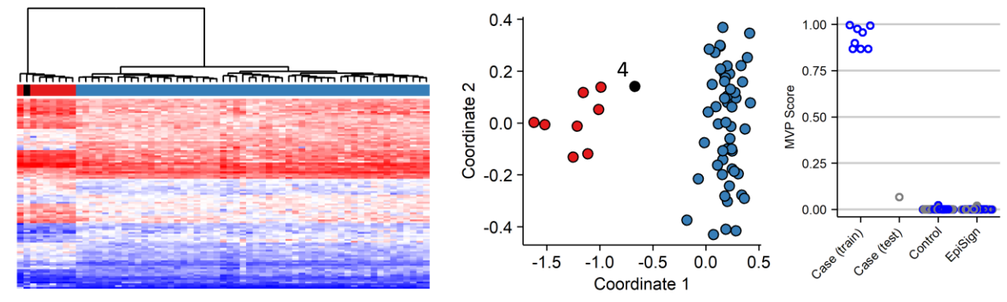


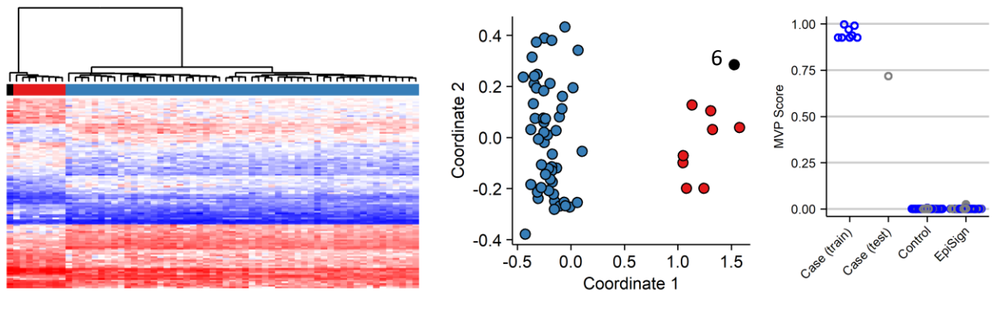




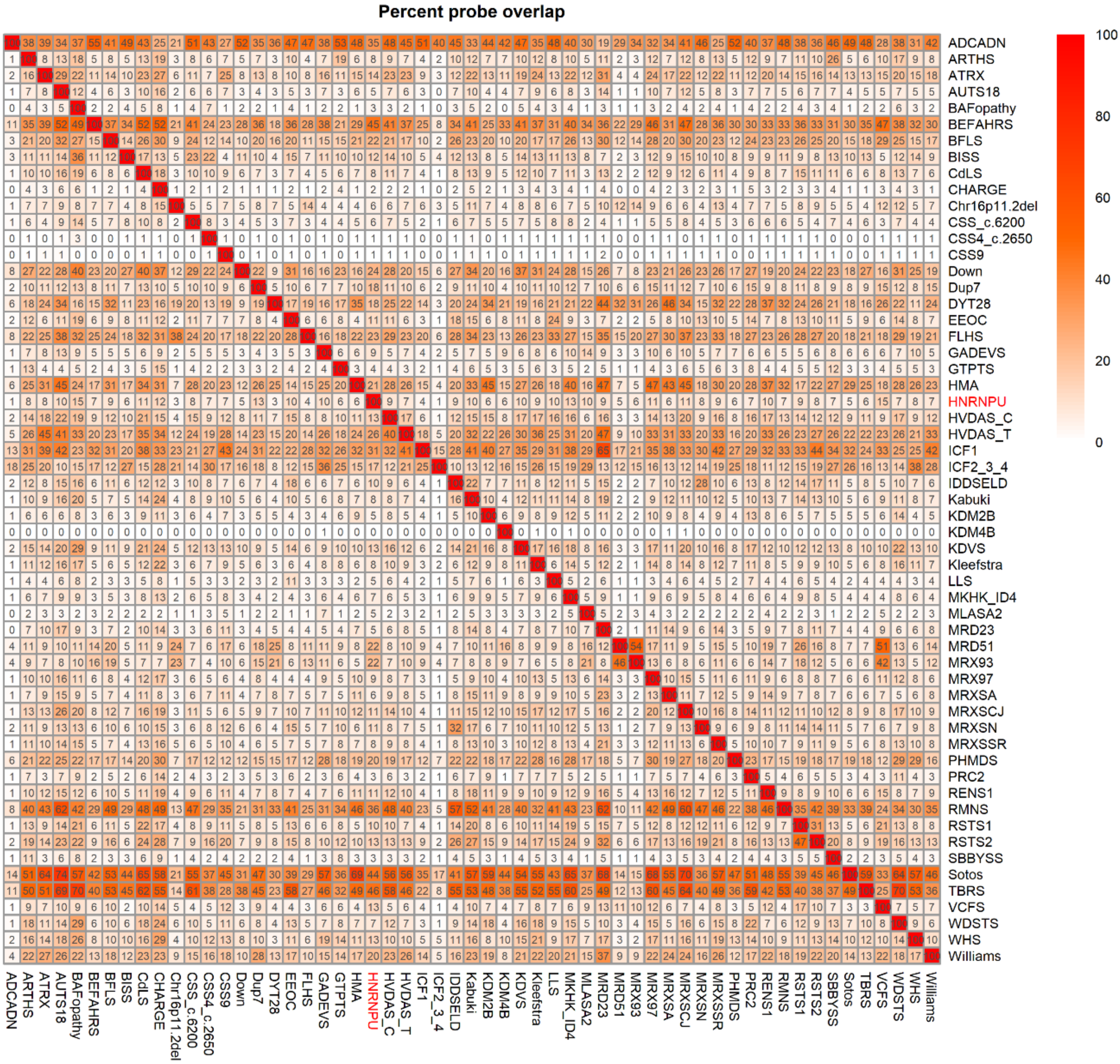




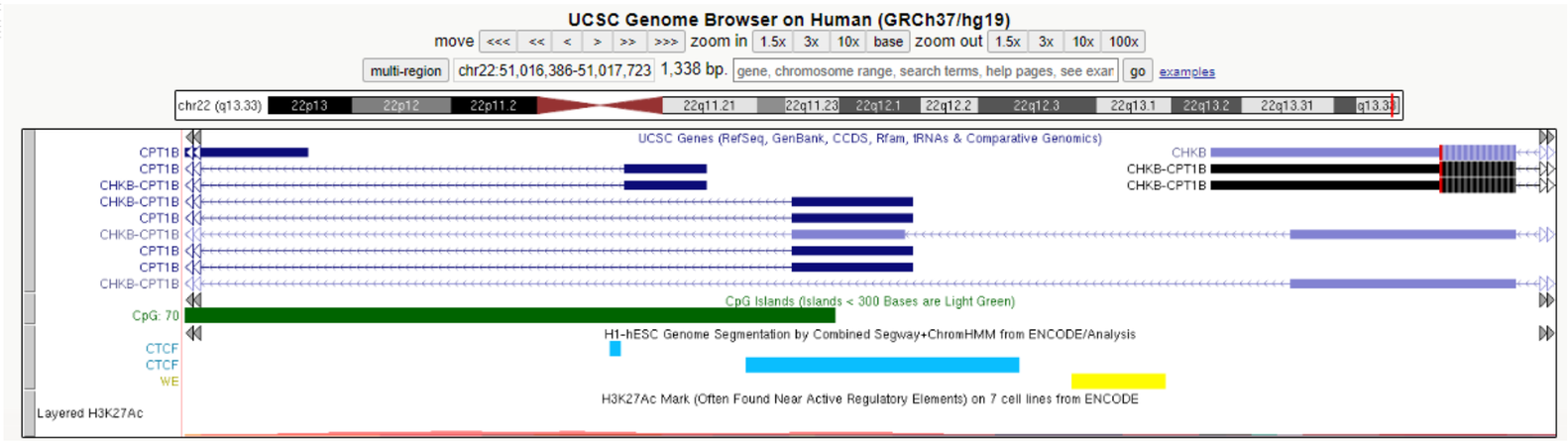


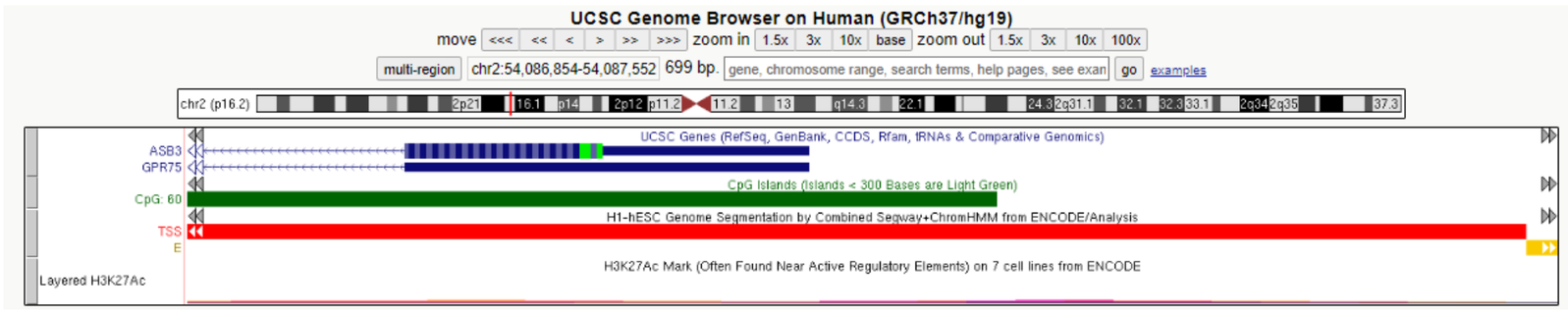


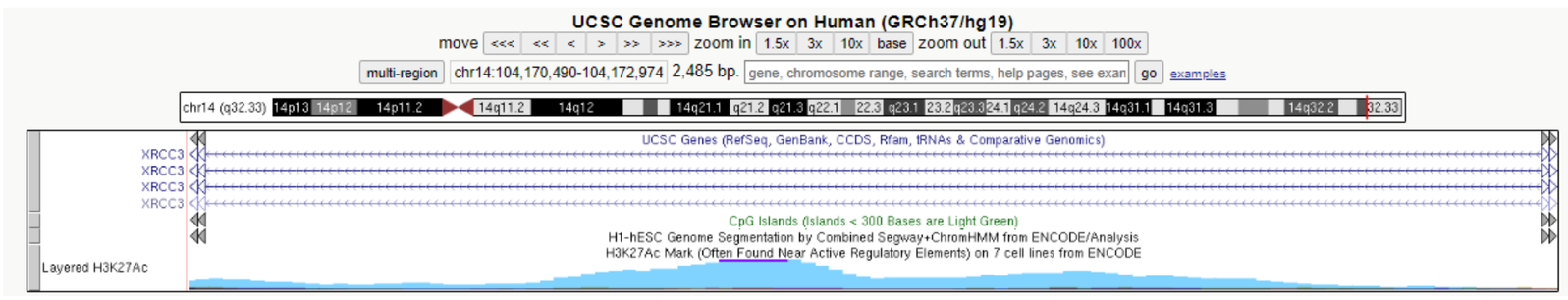
Supplementary Figure 3: *Correlation matrix of DMPs.*

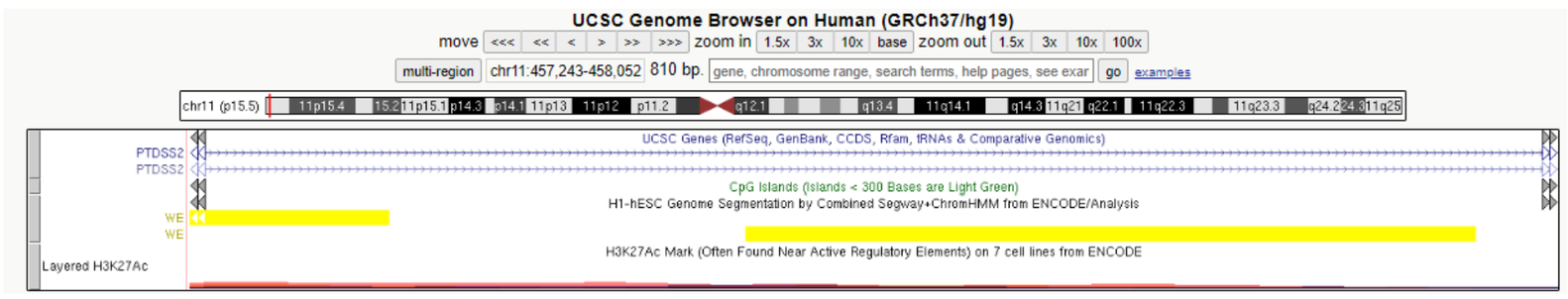


Supplementary Figure 4: *DMR location and regulatory elements*

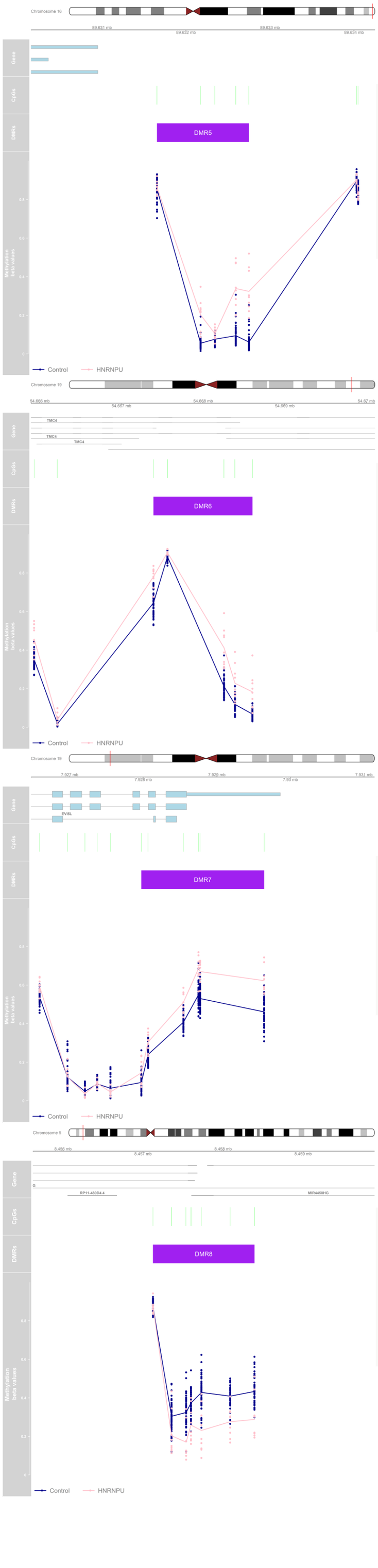


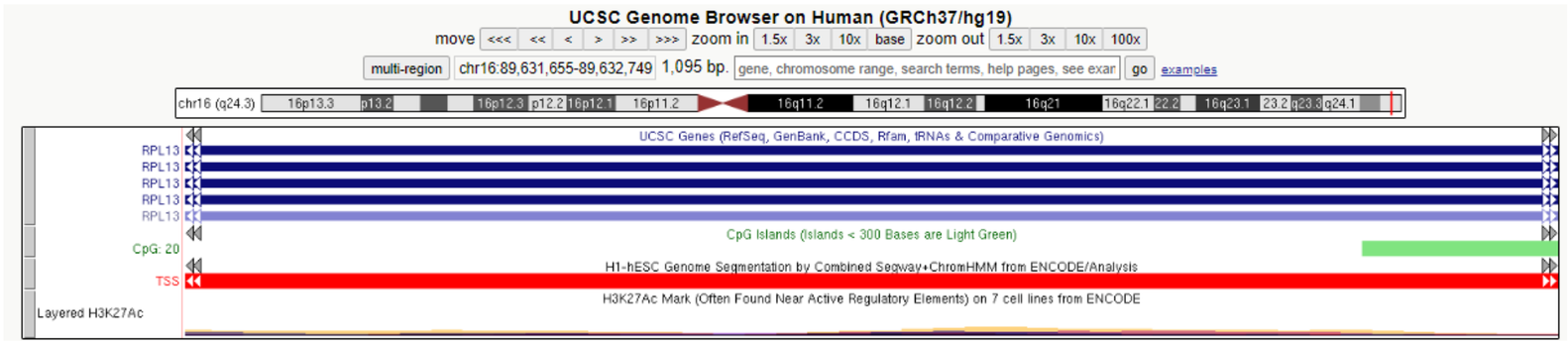


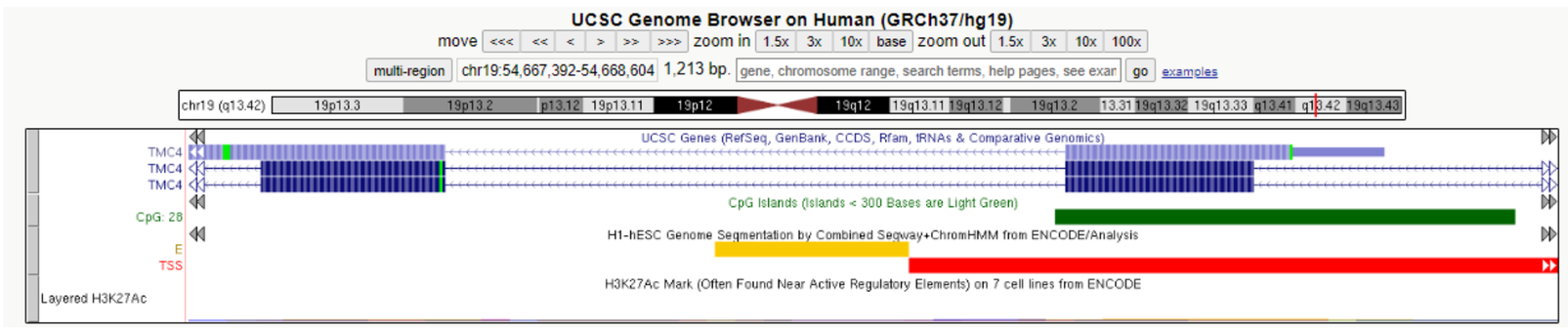


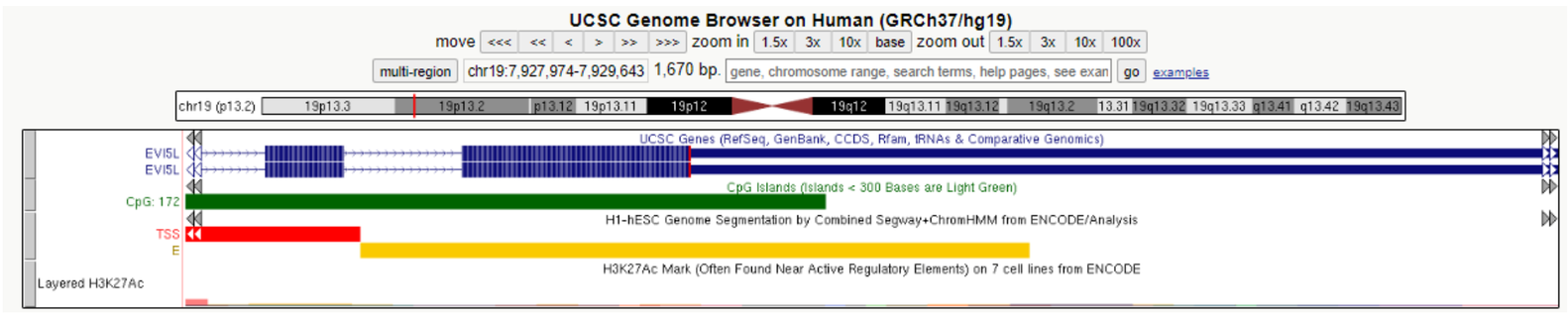


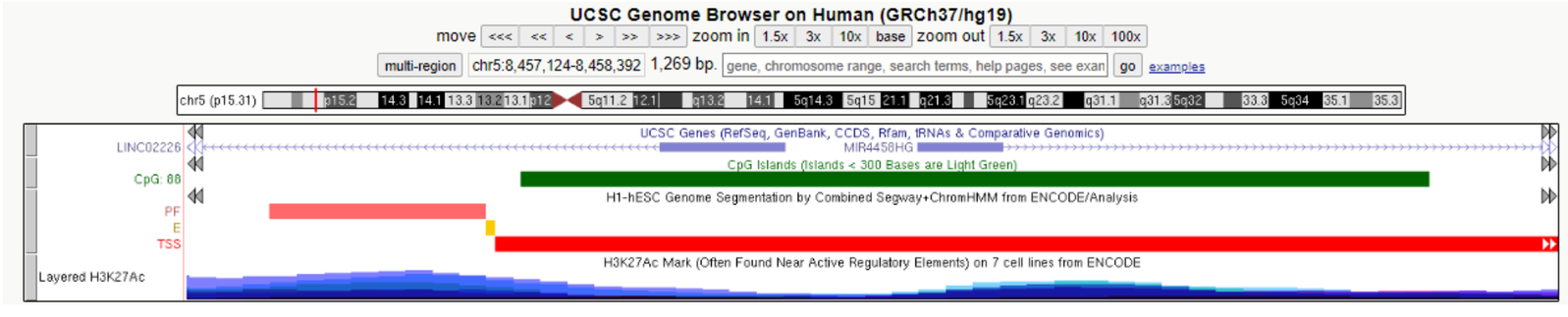
Supplementary Figure 4: *DMR location and regulatory elements*



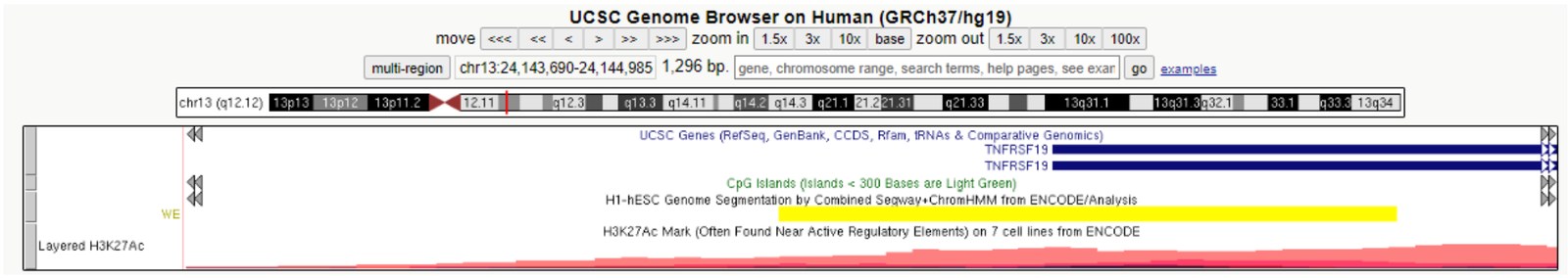


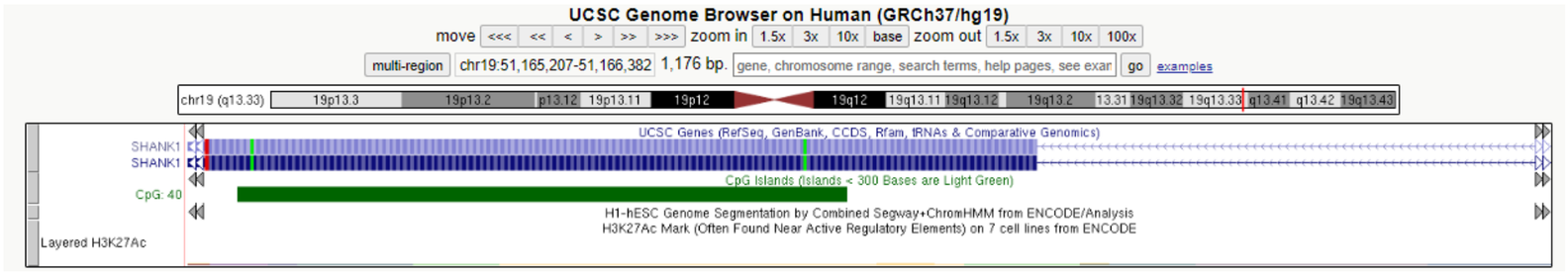


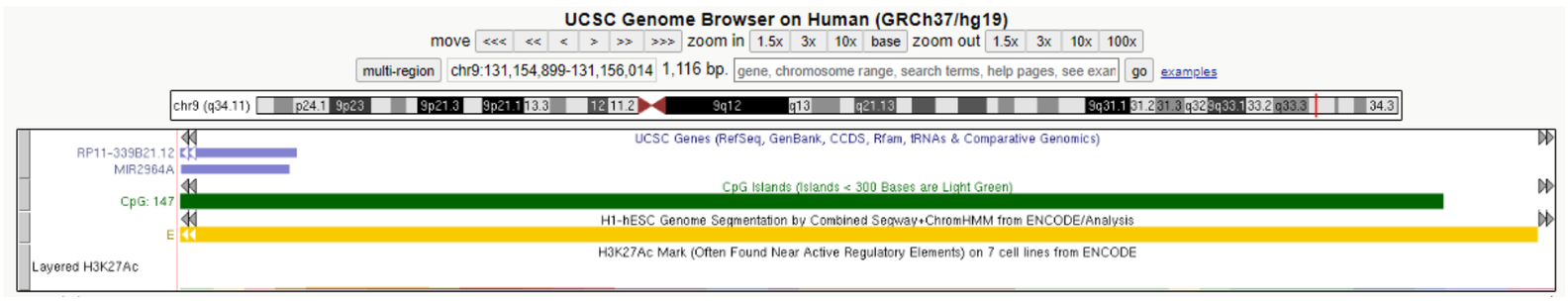


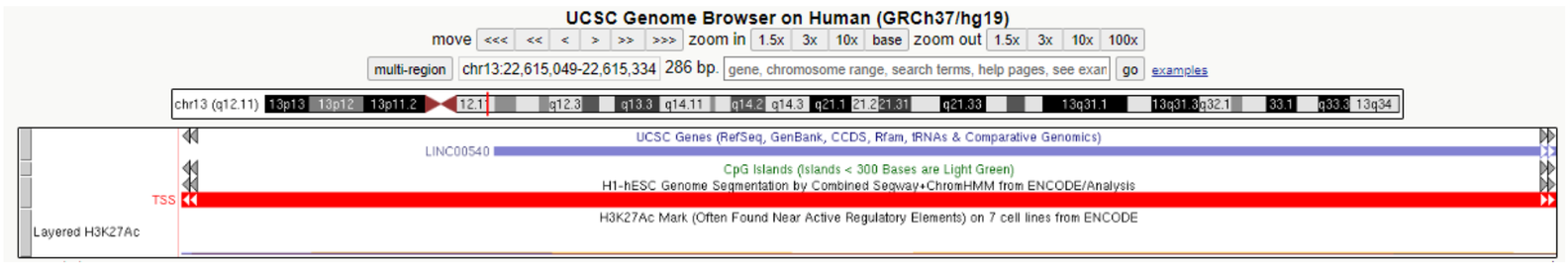


Supplementary Figure 4: *DMR location and regulatory elements*

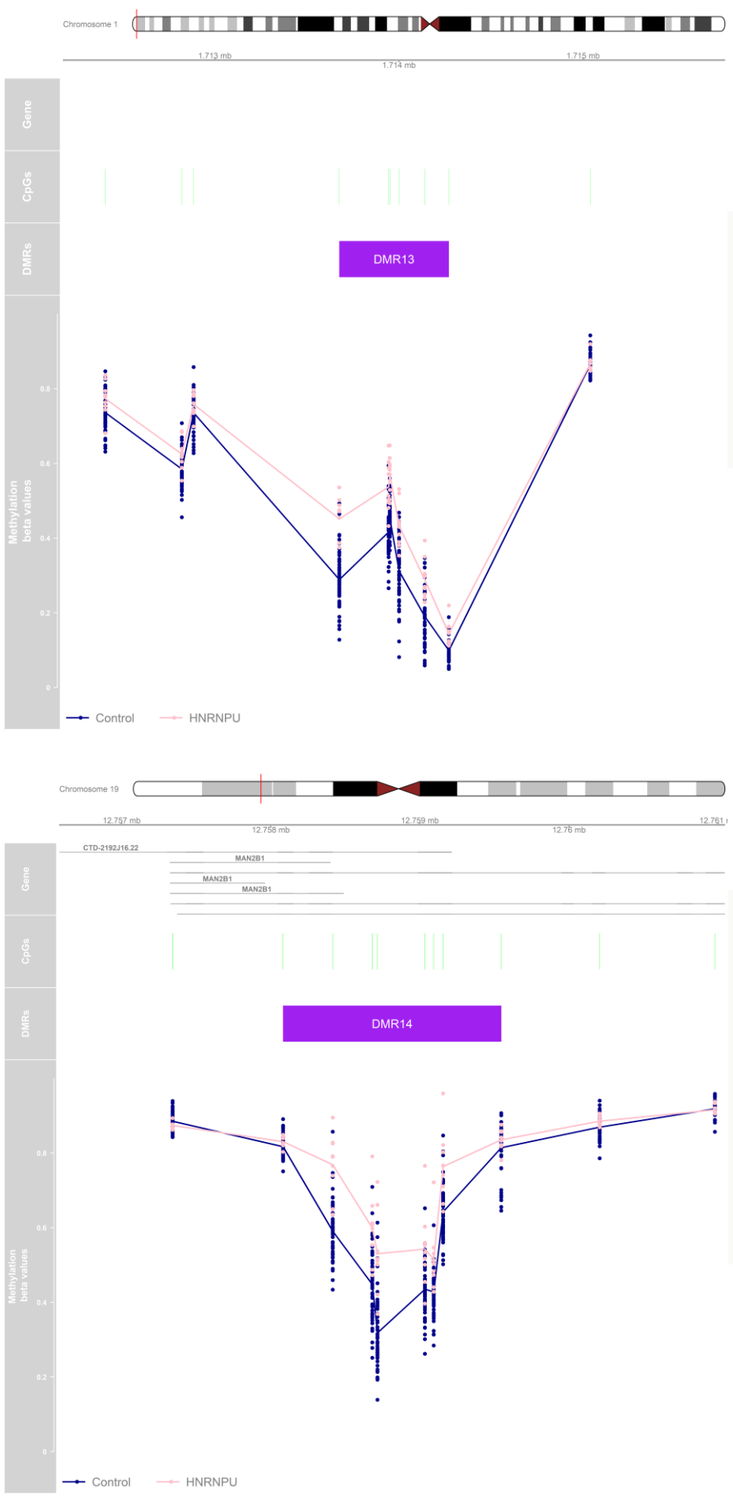


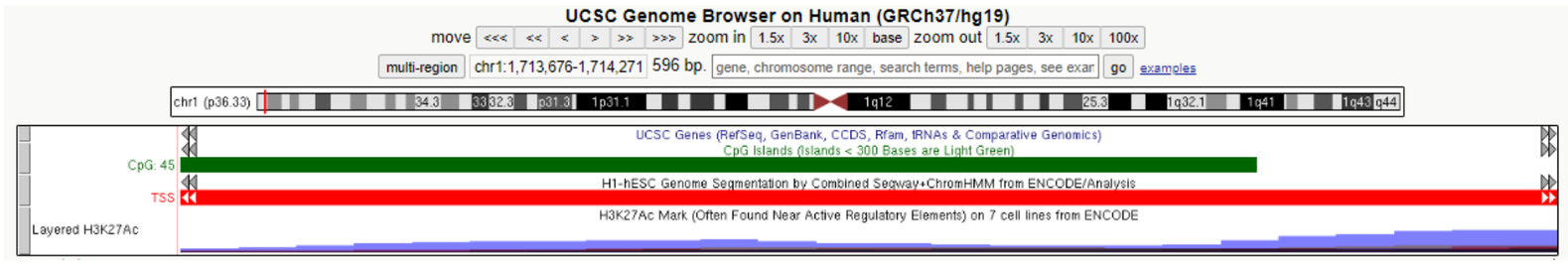


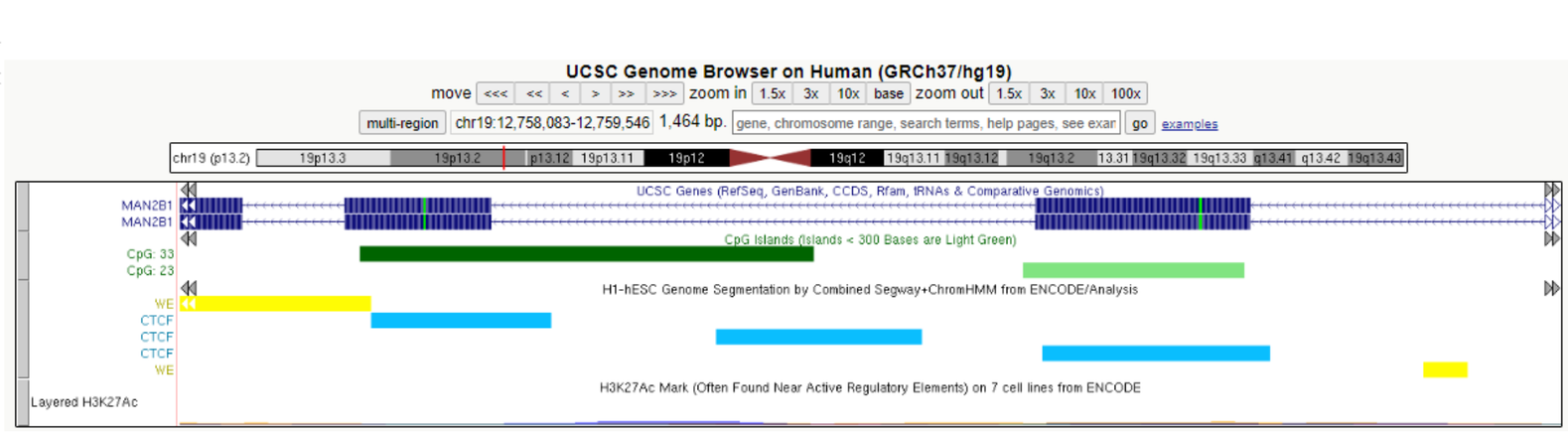


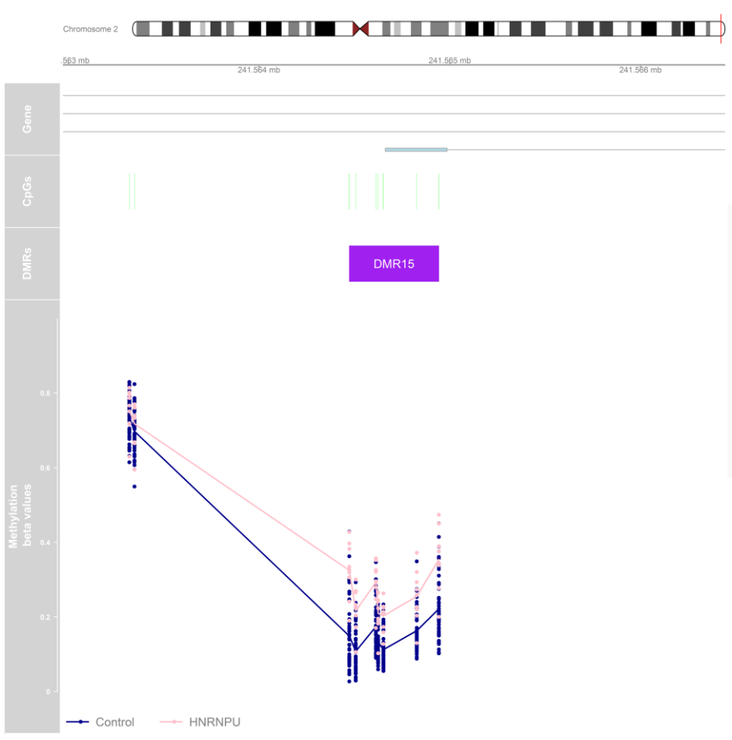


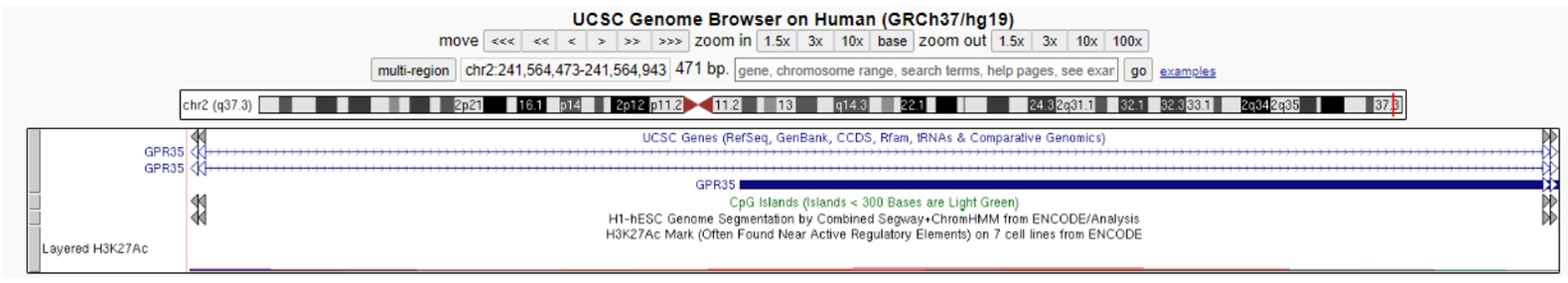
Supplementary Figure 4: *DMR location and regulatory elements*

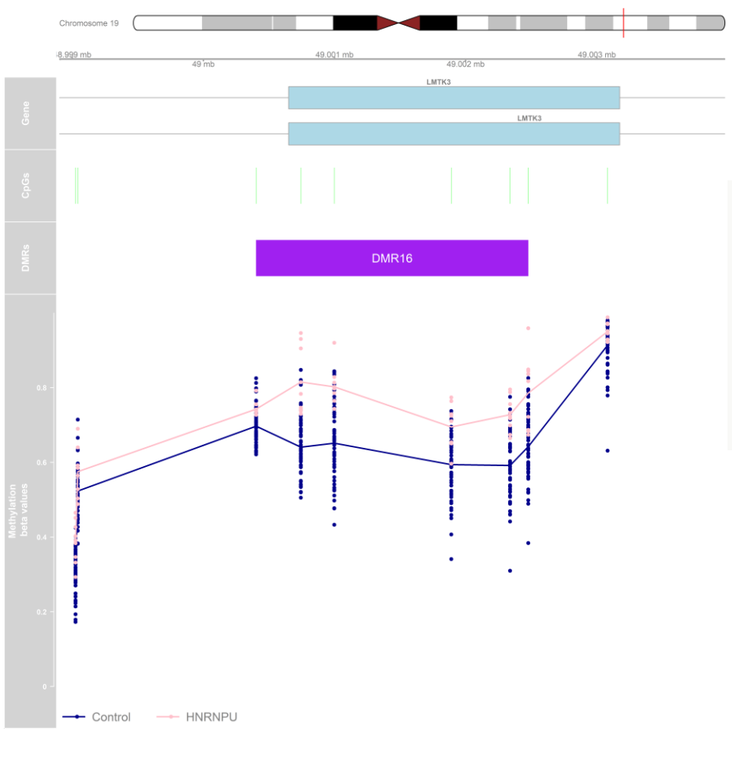


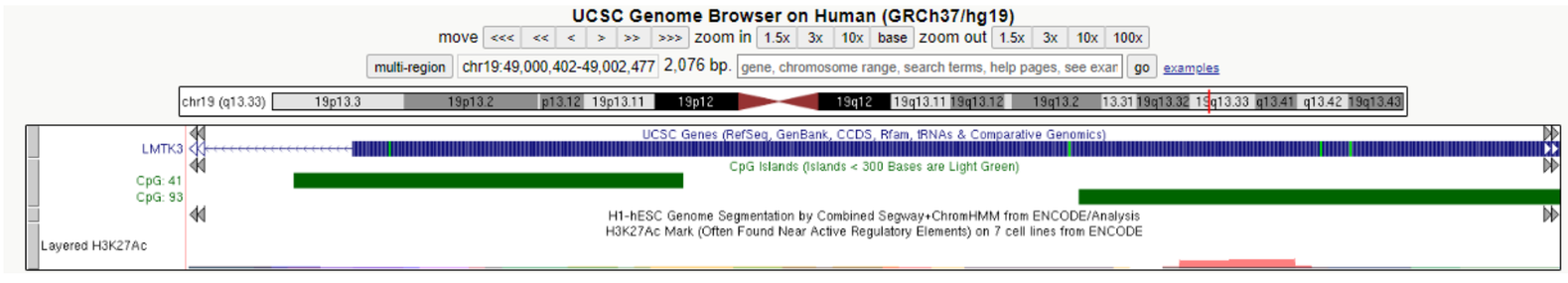












Supplementary Figure 4: *DMR location and regulatory elements*