

Supplementary methods

Genome sequencing and data processing

Generated raw sequencing data was processed either through Illumina DRAGEN Bio-IT Platform (v4.0-v4.3) or through benchmarked standard analysis pipelines that used BWA-MEM or BWA-MEM2^{1,2} for read alignment to version 38 of the human genome and GATK HaplotypeCaller (v4.1-4.3) for the detection of single nucleotide variants (SNV) and short insertions and deletions in line with the GATK Best Practices workflow for germline short variant Discovery³. Structural variants were detected using Manta, Control-FREEC or GRIDSS⁴⁻⁶ and expansions of short tandem repeats (STRs) were identified via ExpansionHunter⁷. All runs of homozygosity larger than 500kb in length were identified, and read depth calculated across the target region of interest to provide a measure of data quality. In addition to comprehensive individual analysis and reanalysis, genomic data of the entire dataset and/or subsets with similar phenotypes can also be analyzed to identify all individuals with variants in specific genes or regions.

Analysis of candidate variants in snRNA genes

We screened the three cohorts of undiagnosed individuals in the Spanish URD Programs for potentially pathogenic variants in the *RNU4-2* (NR_003137.2) and *RNU2-2* (NR_199791.1) genes, previously associated with NDDs. We also analyzed structural variants and small insertions or deletions (indels) within a ± 2000 base-pair window flanking either of the two loci. Given that the previously reported pathogenic variants in these genes were found to cluster in regions depleted of variation in the population, we

extended our analysis to other small nuclear RNA (snRNA) genes of the major spliceosome that are similarly constrained. A threshold of -0.25 in the normalized observed proportion of single nucleotide variants (SNVs) was used to define such regions of constraint ⁸. Only two additional snRNA genes, *RNU5B-1* (NR_002757.3) and *RNU4-1* (NR_003925.3), contained regions meeting this criterion. Variants were annotated with allele counts (AC) from gnomAD v4.1.0, retaining those with AC < 40 for further analysis. All candidate variants were visually reviewed using the Integrative Genome Viewer (IGV) ⁹. Selected variants were categorized by type, population frequency, and inheritance pattern.

To evaluate the potential contribution of oligogenic mechanisms, we assessed the enrichment of digenic variant combinations in NDD probands compared to unaffected individuals. Digenic combinations were defined as the presence of rare variants (AF < 0.001 in gnomAD v4.1.0) in two or more distinct snRNA subunits within the same individual. The NDD cohort comprised 1,708 probands enrolled in the Spanish URD Programs. The control group consisted of 2,121 unaffected individuals: 1,021 unaffected relatives from the URD programs (parents or siblings with no reported neurodevelopmental phenotype) and 1,100 individuals from an independent cancer genomics project, included as additional controls given the absence of reported neurodevelopmental phenotypes. Statistical enrichment was assessed using a two-sided Fisher's exact test.

Sanger sequencing following PCR amplification of the target regions was performed in trios to confirm inheritance. PCR primers used to amplify genomic DNA are detailed in Table S11.

RNA folding prediction

NR_199791.1 (*RNU2-2*), NR_003925.2 (*RNU4-1*), NR_003137.3 (*RNU4-2*) and NR_002757.3 (*RNU5B-1*) sequences were used as templates to predict the RNA secondary structure of the respective snRNAs via RNAfold¹⁰. Predictions were generated for both the wild-type RNA and mutant versions carrying selected putative pathogenic variants identified in our cohort. Default parameters were applied: Turner model 2024, avoidance of isolated base pairs, and dangling energies applied on both sides of a helix. A folding constraint was applied only for *RNU5B-1* to preserve known structural features, specifically preventing base pairing at the nine nucleotides of the Sm-binding site¹¹.

Face-mask analysis

Frontal facial photographs of individuals with *de novo* dominant variants in *RNU4-2*, *RNU2-2* and *RNU5B-1* genes were analyzed using the Face2Gene Research application tool (FDNA Inc., USA)¹² (Figure S17). Two main algorithms were employed in this analysis: DeepGestalt and GestaltMatcher^{12,13}. The DeepGestalt algorithm preprocesses facial images by detecting landmarks, aligning the face, and analyzing specific regions through a Deep Convolutional Neural Network (DCNN). It generates a softmax vector indicating correspondence to known syndromes and creates a ranked list of possible syndrome matches, providing a DeepGestalt (DG) score for each suggested syndrome. Additionally, DeepGestalt enables binary comparisons between cohorts, calculating Area Under the Curve (AUC) values for the ROC curve, to quantify group distinguishability. The second algorithm applied is GestaltMatcher, which encodes individual photos into 320-dimensional Facial Phenotype Descriptors (FPDs) and compares the resulting vectors to each other creating a "Clinical Face Phenotype

Space", such that the distances between photos define syndromic similarity. GestaltMatcher can determine similarity between query images or to individually embedded images, provide syndrome suggestions for single query images, and visualize clustering patterns using t-distributed Stochastic Neighbor Embedding (t-SNE) plots.

Genome-wide DNA methylation profiling

DNA samples were extracted from whole blood of 17 individuals with *RNU4-2* variants, 5 individuals with *RNU2-2* variants and 4 individuals with *RNU5B-1* variants using standard protocols. Individuals' samples, along with DNA samples from 47 age- and sex-matched controls were processed and hybridized to the Illumina Infinium MethylationEPICv2 BeadChip arrays (San Diego, CA, USA) following the manufacturer's protocol. The resulting intensity data files were processed in R (version 4.2.3) using the meffil package¹⁴. Additional control data were obtained from GSE accession number GSE246337 (45 samples) (Table S3). Stringent quality control (QC) measures were applied to ensure data integrity. Samples and CpG sites were filtered based on bead count and detection *p-value* thresholds (both set at 0.05). To ensure data integrity, we used the xreactive_probes function from the minfi package¹⁵ to identify and exclude probes according to established stringent quality-control criteria. In addition probes that met some of this criteria were excluded: 1) probes located on sex chromosomes (chrX or chrY), 2) probes containing common SNPs at the CpG site (population frequency >1%), 3) probes located at single nucleotide extension sites (masked probe list from <https://github.com/zhou-lab/InfiniumAnnotationV1>), 4) probes non-CpG targeting and 5) probes known to cross-hybridize with off-target chromosomal sites¹⁵. Sex prediction was assessed with meffil package using a standard deviation threshold of 3. Normalization was carried out using functional normalization¹⁶. A selection of 3-5

principal components was determined based on residual analysis to optimize data normalization. Every pool of samples was normalized separately. Finally, to address technical variations, batch effect correction was performed using the ComBat function from the sva package ¹⁷, effectively mitigating array-related batch effects.

Mapping of DNA methylation epesignatures

For *RNU4-2*, a subset of 10 affected individuals with de novo *RNU4-2* variants was selected as the training cohort to map DNA methylation signatures and train classification models. The remaining 6 individuals with de novo variants were used as a testing dataset to evaluate the classification model's performance (Table S3). A matched control group was selected using the MatchIt package for the training cohort ¹⁸. Matching was based on predicted age (calculated with the DNAmAge function from the methylclock package ¹⁹ and sex ¹⁴, with the control sample size being four times larger than the case group. Additionally, the single individual carrying biallelic *RNU4-2* variants (a deletion in trans with n.119A>G; U4-p32) was subsequently applied to the trained classifier to explore whether the biallelic configuration would be classified within the de novo epesignature. For *RNU2-2* and *RNU5B-1*, due to the small sample size, no formal distinction between training and validation groups was made, and all individuals were analyzed together.

Beta values, ranging from zero (no methylation) to one (full methylation), were transformed into M-values using a logit transformation. Linear regression analysis was performed using the limma package ²⁰ to detect differentially methylated probes (DMPs). Estimated blood cell proportions (blood gse167998 ¹⁴, age and sex were included in the model matrix to account for confounding variables. P-values were moderated using the eBayes function and corrected with the Benjamini and Hochberg (BH) algorithm.

121 We employed a three-step probe selection strategy, adapted from Vos *et al.* 2024²¹:

122 1. Probes were ranked based on the product of mean methylation difference ($\Delta\beta$)

123 and the negative logarithm of the corrected p-value, using the formula $-|\Delta\beta| \log$

124 p, with a pre-selection that retained only probes passing an adjusted $p < 0.05$ and

125 displaying an absolute $\Delta\beta \geq 0.05$, thus ensuring both statistical significance and a

126 biologically meaningful effect size.

127 2. AUROC values were calculated for the top 1000 probes, and the 500 probes

128 with the highest classification performance were retained.

129 3. Highly correlated probes with pairwise Pearson correlation coefficients > 0.75

130 were removed to avoid redundancy.

131 The final selection of probes was assessed using hierarchical clustering and

132 multidimensional scaling (MDS) to confirm robust differentiation between cases and

133 controls. The reproducibility of the episinature was further validated using leave-one-

134 out cross-validation (LOOCV).

135 A support vector machine (SVM) classifier was trained using the final probe set. The

136 model was trained using a radial basis function (RBF) kernel. The SVM implementation

137 was performed using the e1071 package²² with a fixed random seed to ensure

138 reproducibility.

139

Detection of Differentially Methylated Regions (DMRs)

Differentially methylated regions (DMRs) were identified using the DMRcate package²³. Regions containing at least 5 significantly different CpGs within 2.5 kb, with a minimum mean methylation difference of 10% and a Fisher's multiple comparison *p-value* <1e-4, were considered significant. To further characterize these DMRs, we used UCSC Genome Browser tracks: Refseq Genes, CpG Islands, the H3K27Ac Mark from the ENCODE regulation track, and ENCODE GeneHancer regulatory elements.

Functional enrichment analysis of DMPs

For enrichment analysis, we used the top 500 probes with the highest AUROC performance, as identified in the feature selection pipeline. This approach ensured that the probes included in the functional analysis were those with the strongest discriminatory power between cases and controls, maximizing the detection of biologically relevant pathways. To assess the biological relevance of differentially DMPs and near-by genes, we conducted gene-set enrichment analysis (GSEA) using the missMethyl package²⁴. The full set of analyzed CpGs (647,153 probes) was used as the background to ensure robust statistical inference. Gene Ontology (GO) enrichment was performed using the gometh function, analyzing Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) terms. Additionally, the enrichment of DMPs in six classic histone modifications was assessed using ChIP-seq data from 18 blood cell type-related epigenomes, sourced from the Roadmap Epigenomics Consortium. Chromatin state annotations for these same samples were derived using the NIH Roadmap's ChromHMM expanded 18-state model. To evaluate DMP enrichment within transcription factor binding sites (TFBS), data from human meta-

clusters available in the Gene Transcription Regulation Database (GTRD) were employed (v18.06). Statistical significance of the observed enrichments was determined using one-sided Fisher's exact tests (*adjusted p-value* < 0.05), comparing the overlap of DMPs with the relevant dataset against a background set of 647,153 filtered probes from the HumanMethylationEPICv2 array.

SUPPLEMENTAL FIGURES

Figure S1: Expression heatmap of snRNA genes across human tissues based on GTEx data. For each snRNA subfamily (*RNU2*, *RNU4*, *RNU5*, *RNU6*), the top 10 subunits with the highest expression in brain tissues are displayed, along with selected non-neural tissues (liver, pancreas, and whole blood) for comparison. The color scale represents relative expression levels, with blue indicating higher expression and yellow indicating lower expression.

Figure S2. Pedigree of family U4-p30 carrying compound heterozygous variants in *RNU4-2*. The proband is indicated with an arrow. Individuals shaded in black are affected

Figure S3: Visualization of digenic variant combinations observed in the cohort. A. Each circle represents a pairwise combination of snRNA subunits. Circles are divided into two coloured segments: red for individuals with neurodevelopmental disorders (NDD) and blue for unaffected individuals. The number inside each segment indicates the number of individuals carrying the respective digenic combination in each group. B. Comparison of the overall number of individuals carrying digenic combinations between NDD and unaffected groups. A Fisher's exact test was applied to assess statistical significance ($p < 0.05$; indicated by *).

Figure S4: Predicted RNA secondary structures for wild-type vs. mutant *RNU5B-1* using RNAfold. Minimum free energy (MFE) and centroid structure drawings are shown along mountain plot representation of MFE, thermodynamic ensemble of RNA structures (pf) and centroid structure. Structure drawings are coloured by base-pairing probabilities, and mountain plot includes positional entropy for each nucleotide. Location of the variants is indicated by a blue arrow. Folding constraint of sm domain 9 positions not to be paired was specified to create the prediction.

Figure S5: Predicted RNA secondary structures for wild-type vs. mutant *RNU4-1* (panel A) and *RNU4-2* (panel B) and *RNU2-2* (panel C), using RNAfold. Minimum free energy (MFE) and centroid structure drawings are shown along mountain plot representation of

MFE, thermodynamic ensemble of RNA structures (pf) and centroid structure. Structure drawings are coloured by base-pairing probabilities, and mountain plot includes positional entropy for each nucleotide. Location of the variants is indicated by a blue arrow. Folding constraint of sm domain 9 positions not to be paired was specified to create the prediction.

Figure S6: Supplementary figures of *RNU4-2* episignature analysis. A. MDS plots from the LOOCV approach for U4 episignature identification, with case, control, and test samples shown in red, blue, and yellow, respectively. B. Probability classification results from the constructed SVM model for the U4 episignature. Samples are categorized by genes, with colors indicating specific variants. C. Correlation plot comparing methylation differences ($\Delta\beta$) for episignature-specific probes previously reported. Each point represents a CpG probe from the *RNU4-2* episignature described in the previous study. The x-axis shows the $\Delta\beta$ values from our current analysis, and the y-axis shows the $\Delta\beta$ values reported in the original publication. Colour indicates whether the direction of methylation change is concordant between studies. Pearson correlation coefficient is also calculated D. Heatmap showing histone mark enrichment analysis of the top 500 CpGs derived from the U4 episignature analysis. Enrichments were calculated using ChIP-seq data from 18 blood cell type-related epigenomes provided by the Roadmap Epigenomics Consortium. The color scale represents the odds ratio of significant CpG sites, as determined by one-sided Fisher's exact tests using filtered EPICv2 probes as background. E. Heatmap illustrating chromatin state enrichment for the same top 500 CpGs, based on annotations from the ChromHMM expanded 18-state model (NIH Roadmap Epigenomics). Color intensity corresponds to the odds ratio of significant enrichments across 18 chromatin states. F. Dot plot displaying the enrichment of transcription factor binding sites (TFBS) in the top 500 CpGs using data from the GTRD database (v18.06). Statistical significance was assessed using one-sided Fisher's exact tests with EPICv2 CpGs as background.

Figure S7: *RNU2-2* and *RNU5B-1* episignature analysis A. Hierarchical clustering using U2-specific probes. Rows correspond to selected probes, while columns represent samples. The color scale indicates methylation levels (blue: hypomethylated; red: hypermethylated). U2, U4, U5, and control groups are labeled in blue, pink, yellow, and gray, respectively. B. MDS plot from the LOOCV approach for U2 episignature identification. Samples are categorized into red (controls), blue (cases), and yellow (testing samples). C. Hierarchical clustering using U5-specific probes, following the same layout as panel A. D. MDS plot from the LOOCV approach for U5 episignature identification, with samples colored as in panel B.

Figure S8-14. Differentially methylated regions (DMRs). Upper panel: UCSC Genome Browser screenshot displaying the genomic context of the DMR, including annotated genes (blue), CpG islands (green), H3K27Ac marks, ENCODE cis regulatory elements (transcription start sites (TSS; red), enhancers (E; yellow), and weak enhancers (WE; pale yellow)) , GeneHancer annotations and JASPAR Transcription factor binding site. Lower panel: plot showing methylation levels (mean $\pm$ confidence intervals) for each CpG probe within the DMR in U4 cases (pink) and others (gray).

Figure S15: Heatmap from hierarchical clustering of phenotypic profiles based on HPO terms. The horizontal axis represents individual patient cases, along with their corresponding ages and pathogenic variants in the genes *RNU2-2*, *RNU4-2*, and *RNU5B-1*. The vertical axis categorizes the HPO terms for each patient into four groups: facial dysmorphologies, musculoskeletal and growth abnormalities, nervous system physiology, and other systems. Clustering was performed using the Euclidean method, indicating the presence (red), absence (blue), or missing data (white) of HPO terms. The analysis reveals patterns of similarity, highlighting correlations between phenotype and genotype.

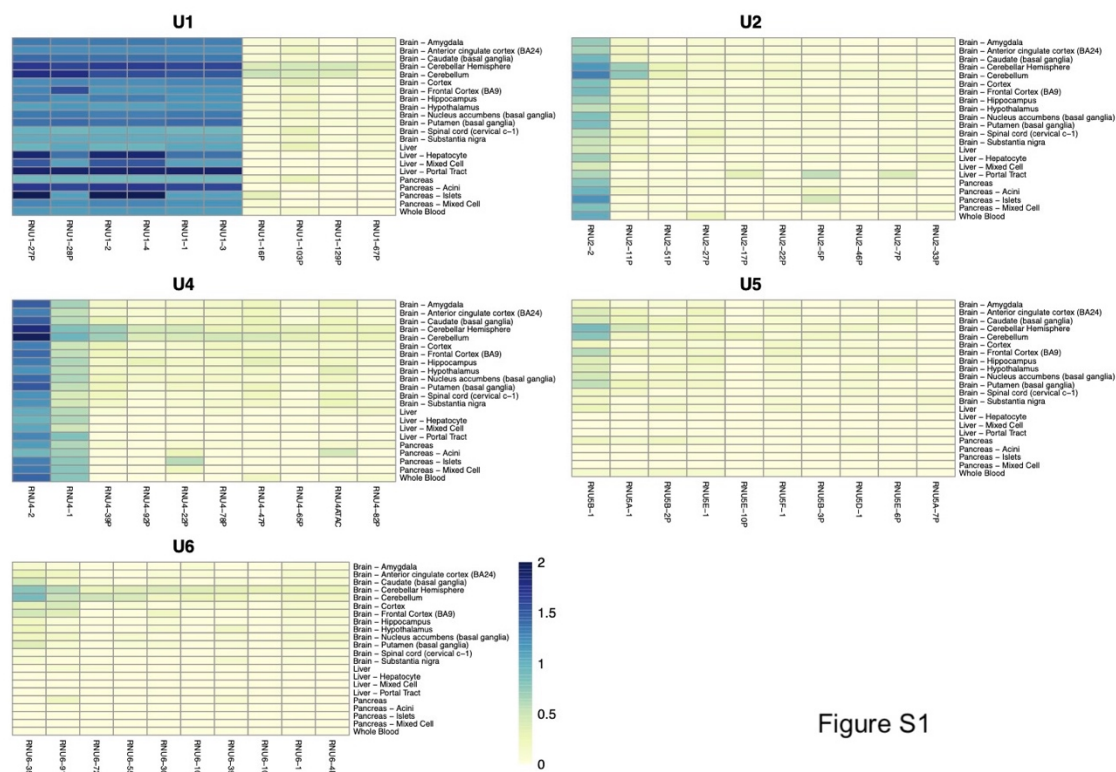


Figure S1

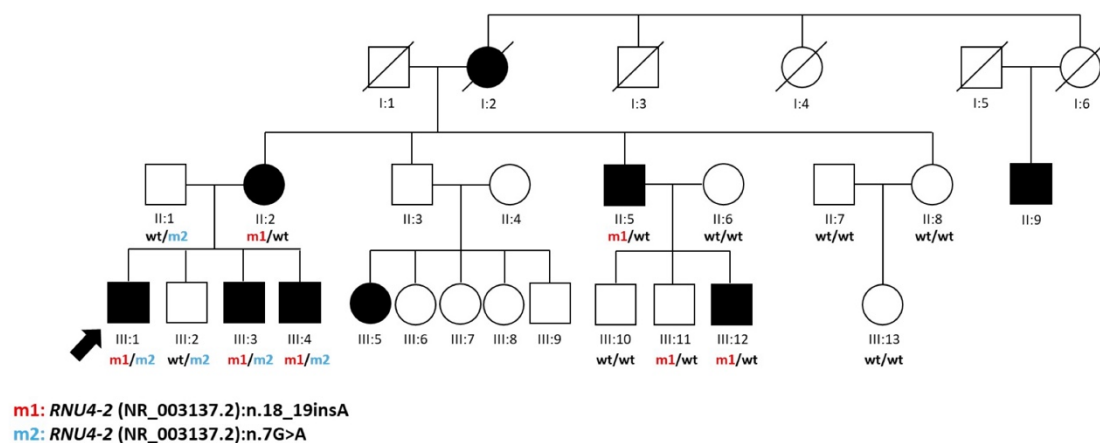
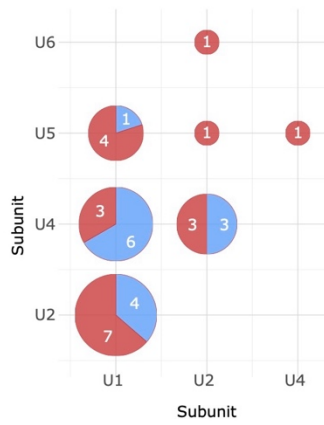


Figure S2

A.



B.

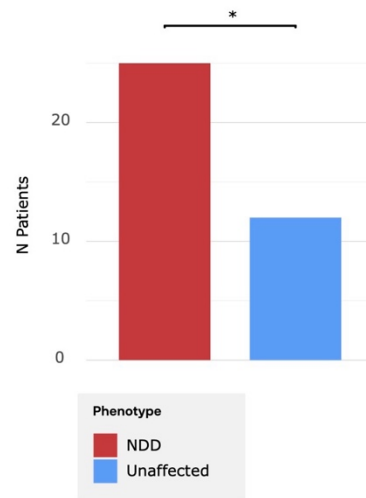


Figure S3

252

RNU5B-1

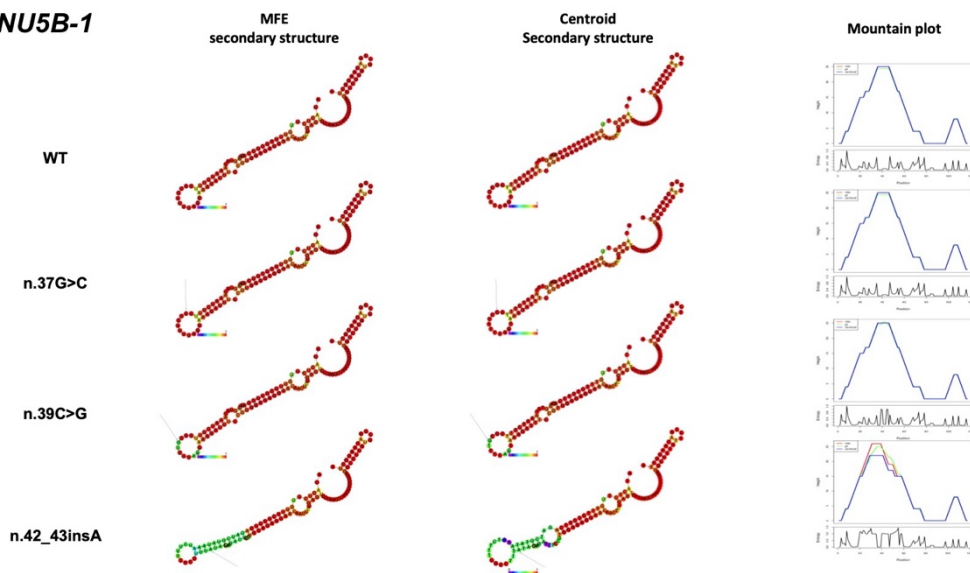


Figure S4

253

RNU4-1

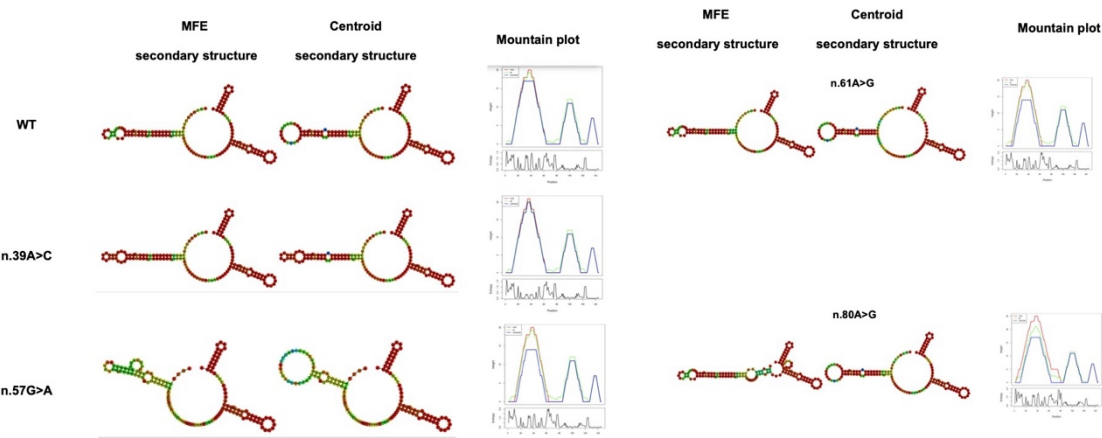


Figure S5a

RNU4-2

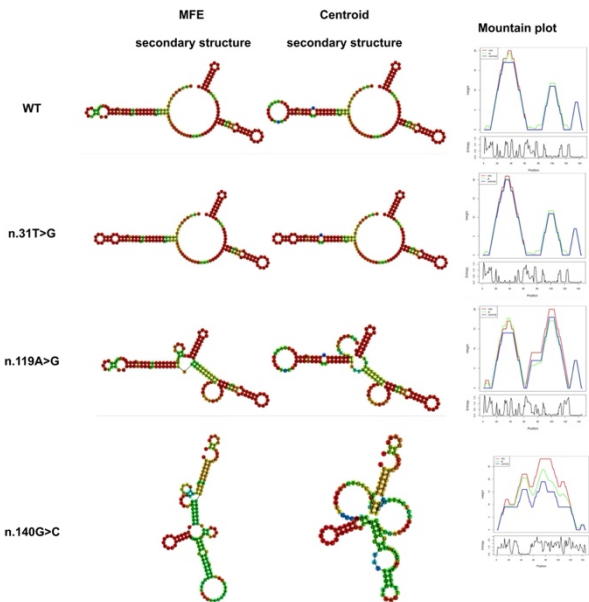


Figure S5b

RNU2-2

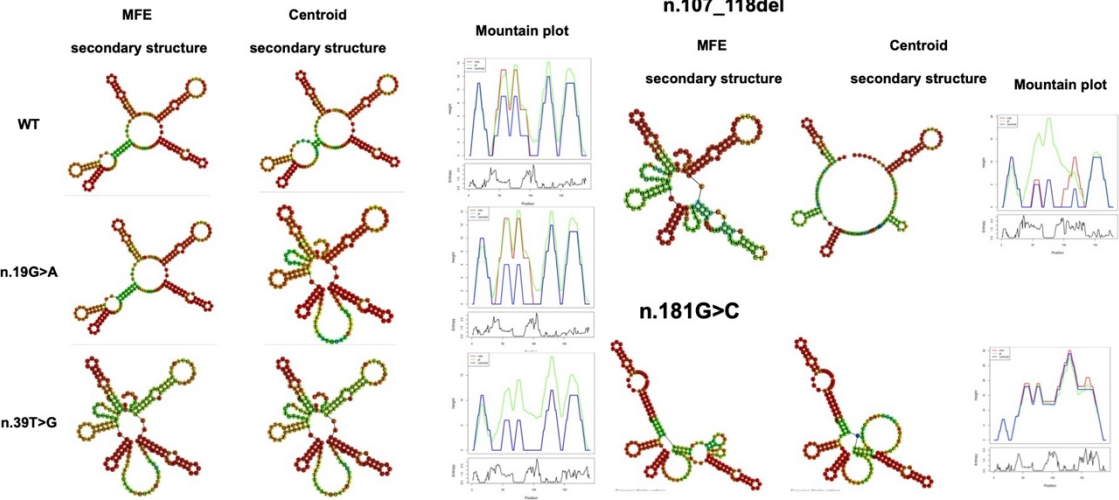


Figure S5c

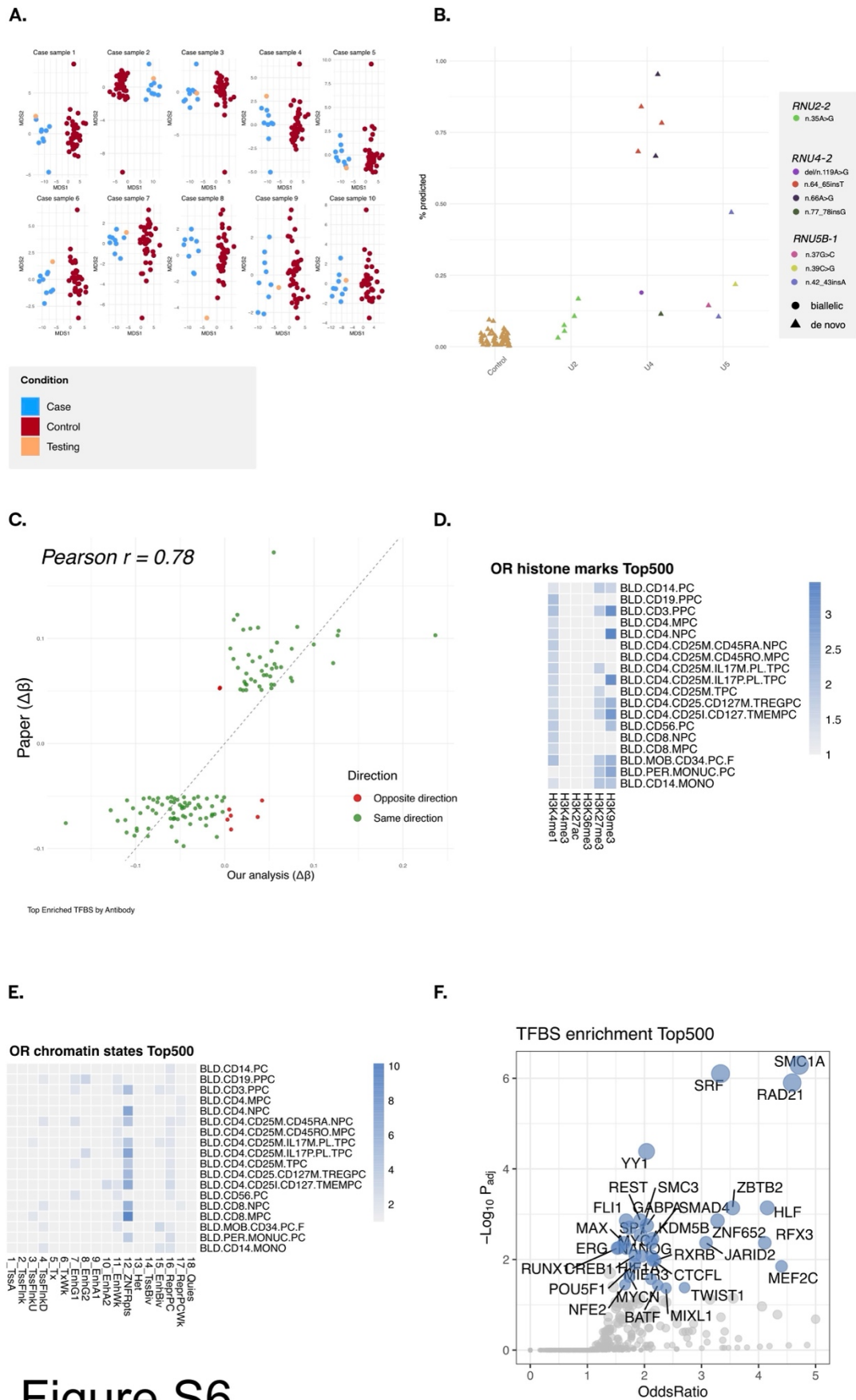


Figure S6

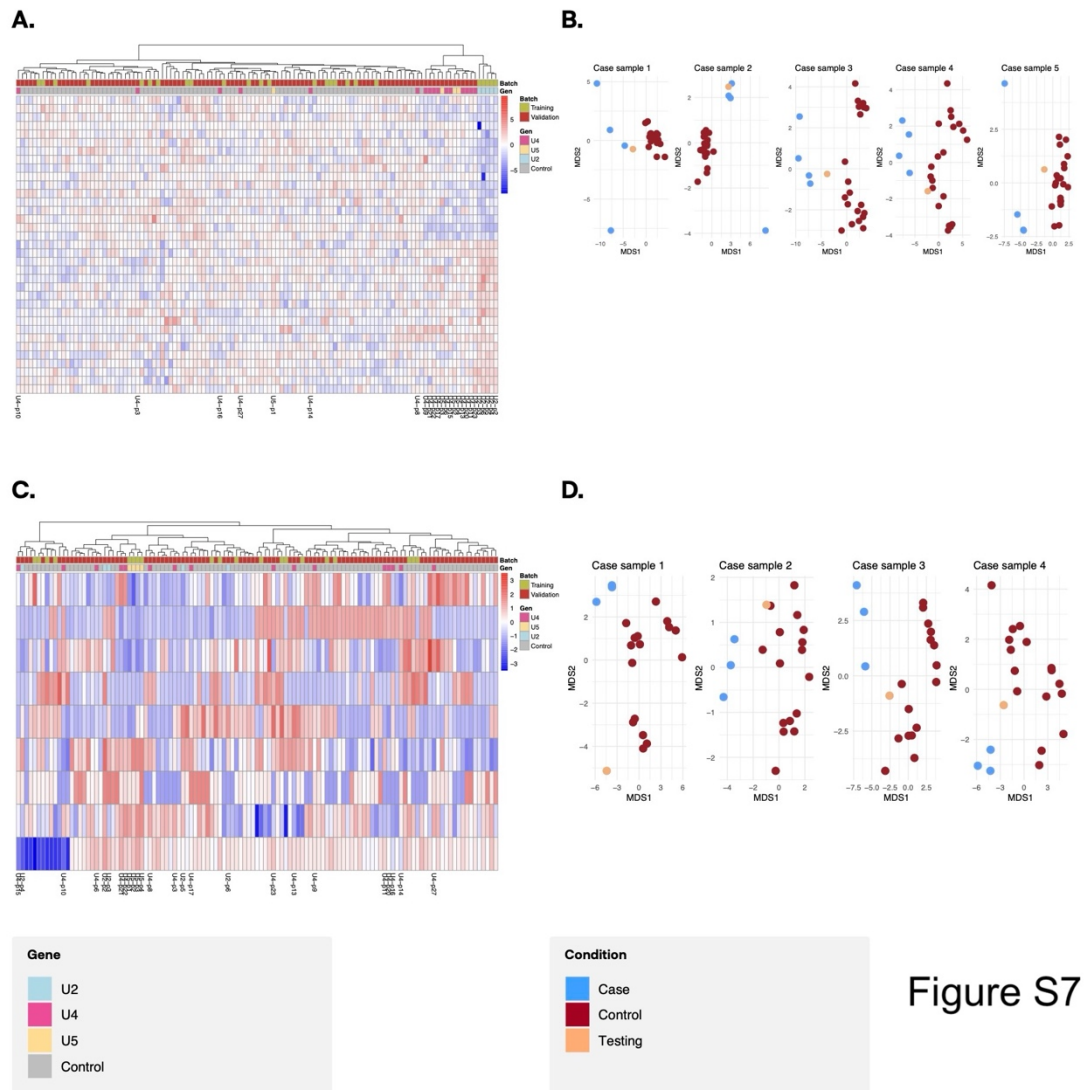


Figure S7

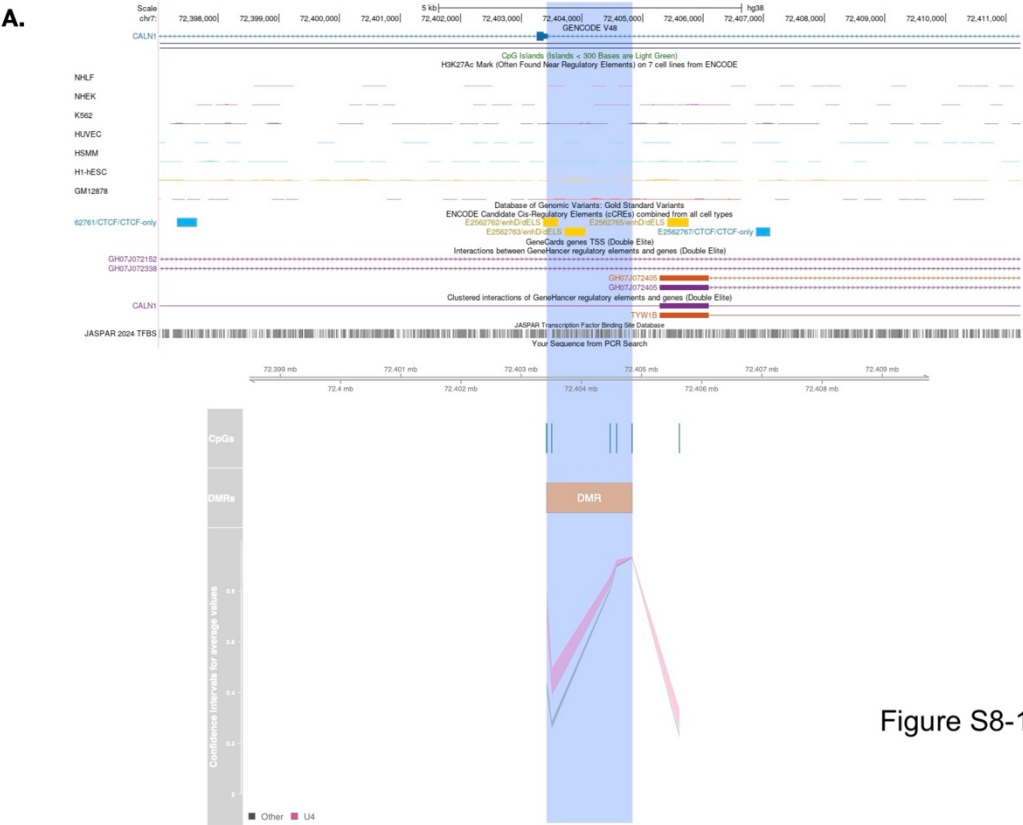


Figure S8-14

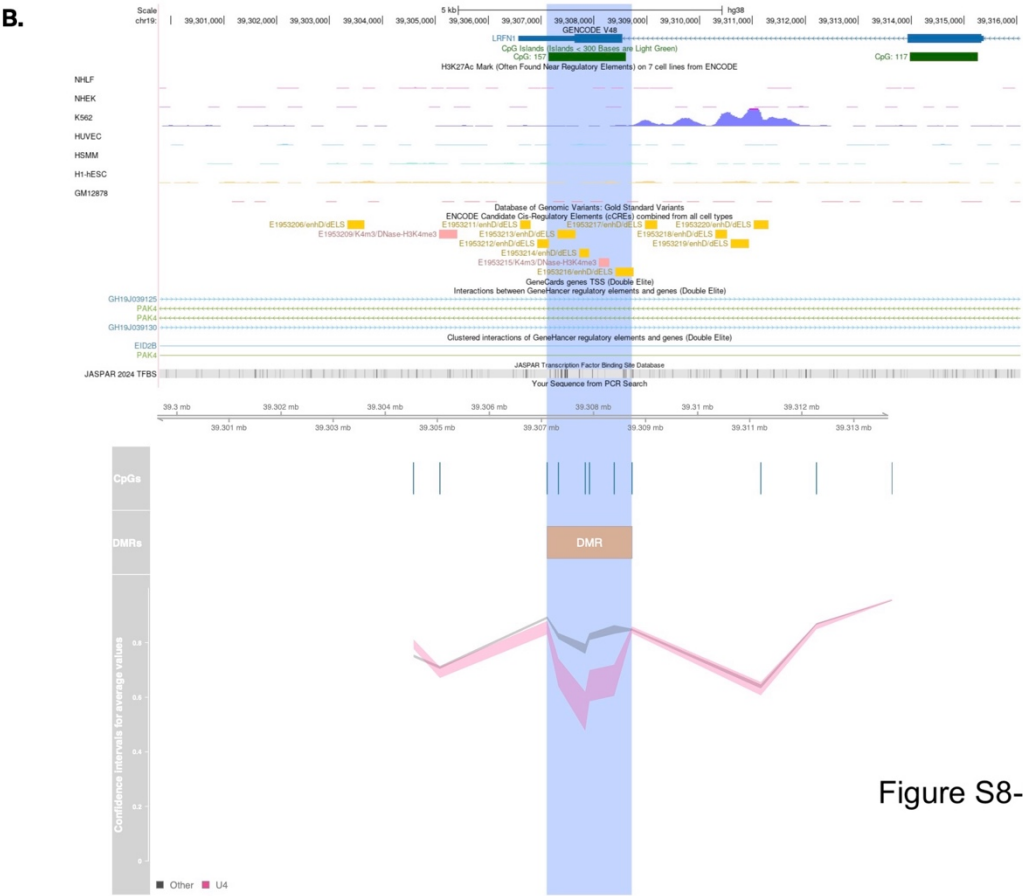


Figure S8-14

C.



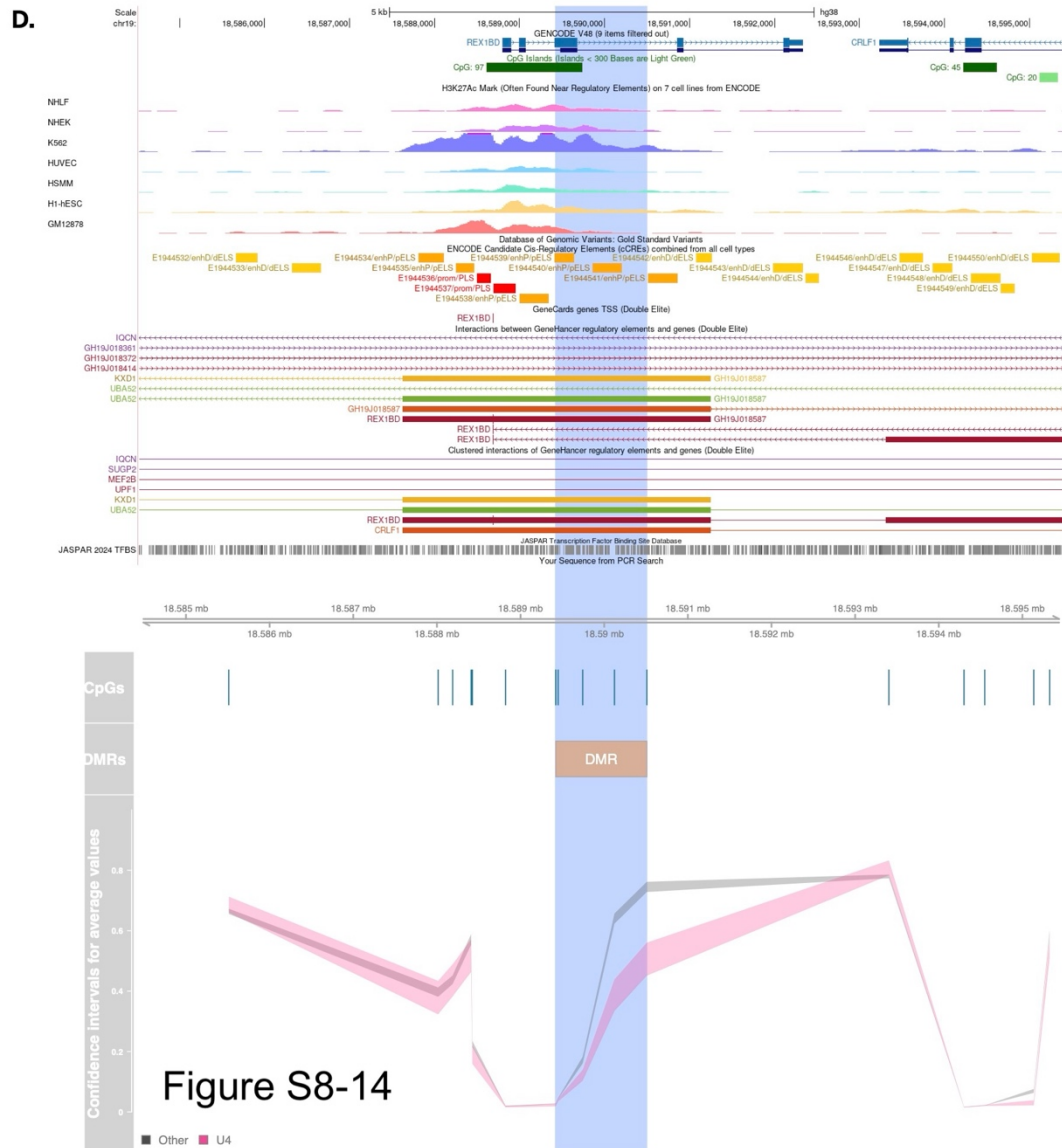


Figure S8-14

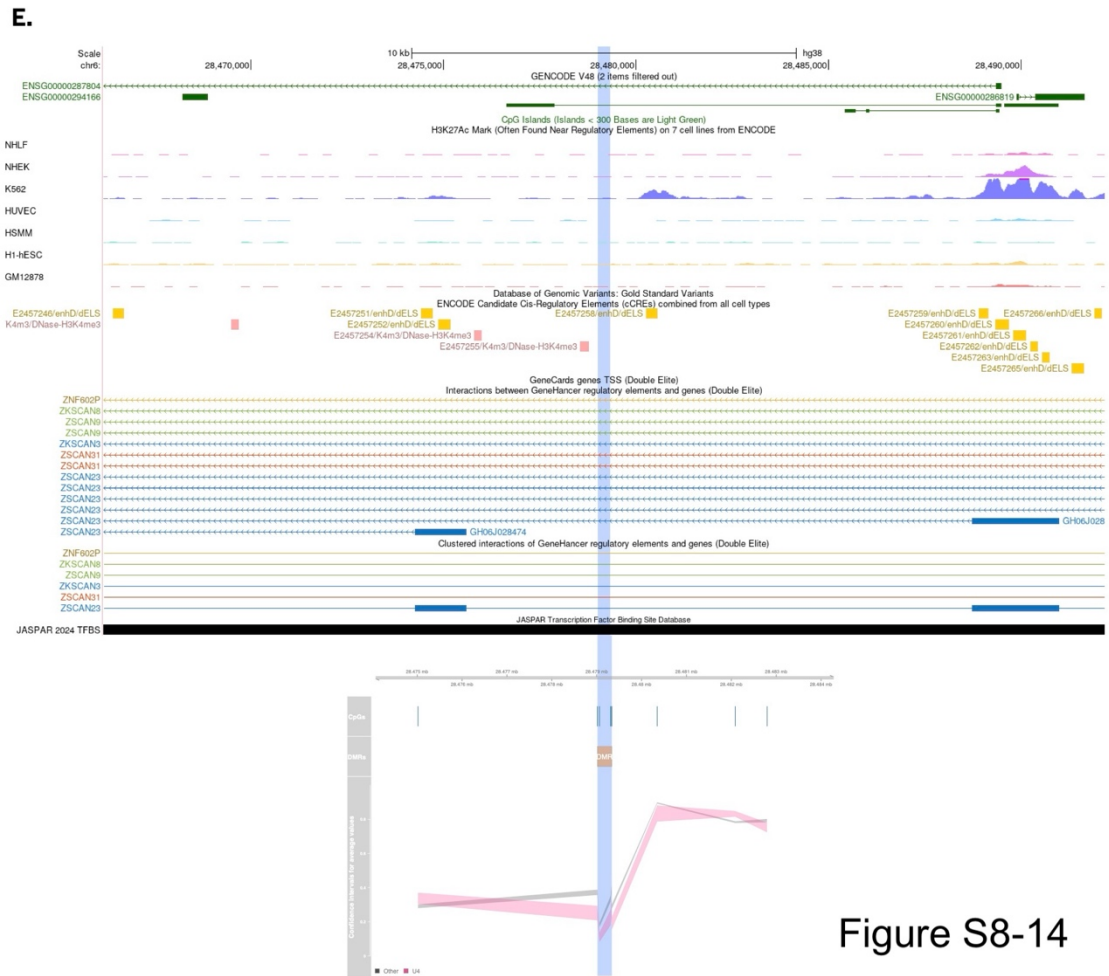


Figure S8-14

F.

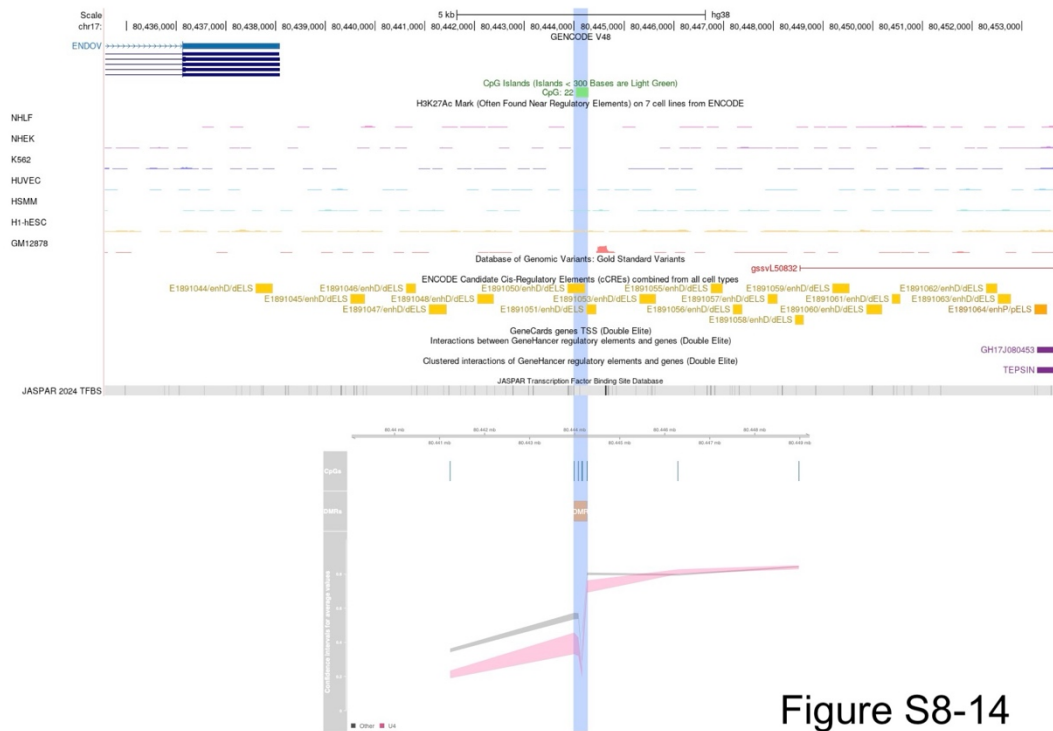


Figure S8-14

G.



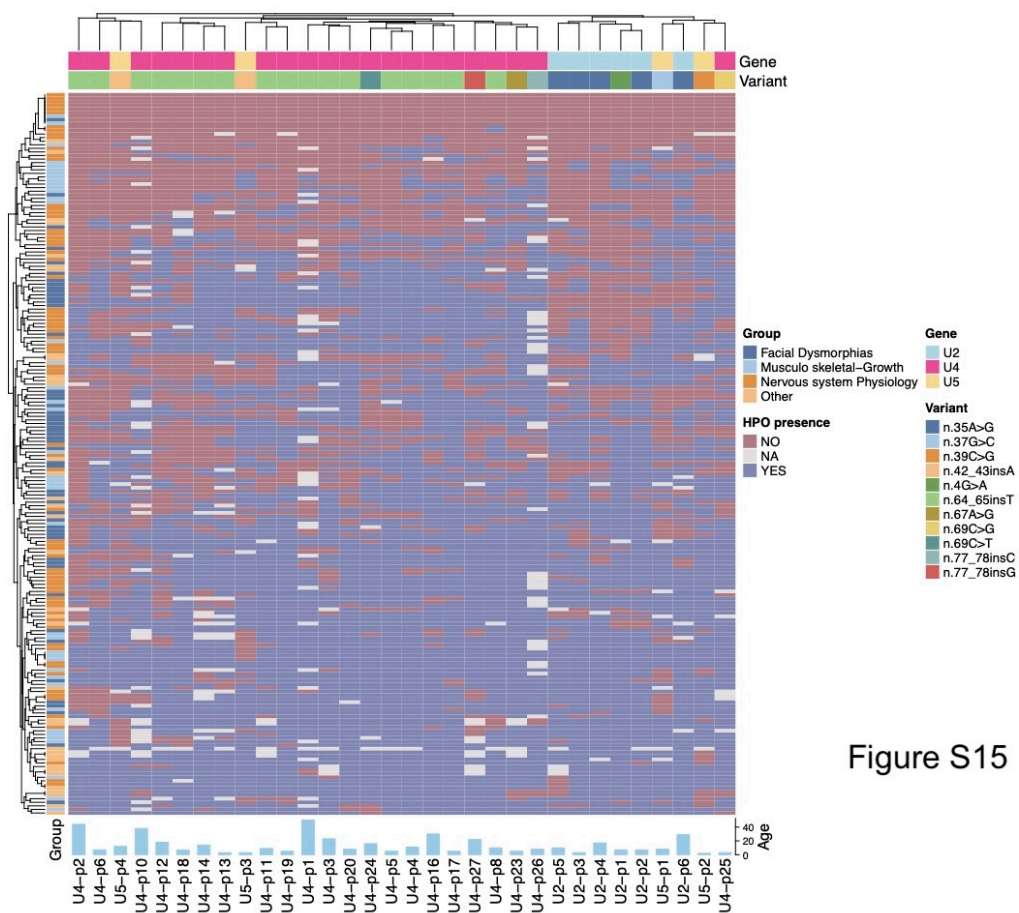


Figure S15

SUPPLEMENTAL TABLES

Table S1: List of patients included in the Face2Gene analysis.

Table S2: List of patients' samples used for methylation analysis, including GSE accession numbers and matching parameters. Column Batch refers to patient designation as training or validation for the analysis.

Table S3: Metadata of patients with de novo variants in *RNU4-2*, *RNU2-2*, and *RNU5B-1*. 1 with structural domain annotation.

Table S4: Rare variants identified in *RNU4-2*, *RNU2-2*, and *RNU5B-1*, in our cohort with inheritance confirmation, gnomAD frequencies, variant classification, domain annotation and clinical classification.

Table S5: List of snRNA genes of the major spliceosome included in the analysis of biallelic variants. Genes were selected based on their expression levels in brain tissue, according to GTEx data.

Table S6: Detailed list of biallelic snRNA variants identified in cases and controls. Information on gnomAD frequencies, zygosity, and associated phenotype is included.

Table S7: List of differentially methylated probes (DMPs) in *RNU4-2* patients.

Table S8: Summary of differentially methylated regions (DMRs), including genomic coordinates and overlapping genes.

Table S9: Clinical features of patients with de novo snRNA variants, annotated with Human Phenotype Ontology (HPO), grouped by major HPO categories. These terms were used for reverse phenotyping and subsequently for clustering analysis. Numbers in parentheses represent total responses per HPO term.

Table S10: Detailed prenatal history of patients with variants in *RNU2-2*, *RNU4-2* and *RNU5B-1*. * One case conceived by egg donation resulting in twin pregnancy with uterine single-twin intrauterine loss in the first trimester. ** Two male patients who are a

294 monozygotic twin pair. *** A male patient large for gestational age (birth weight of
295 4,535
296 grams) was excluded due to maternal gestational diabetes. Prenatal diagnosis in the
297 *RNU4-2* group included two cases with prenatal array CGH and one chorionic villus
298 sampling performed due to increased nuchal translucency and ventriculomegaly. All
299 results showed a normal karyotype (46,XX) and normal array CGH findings. Fetal mild
300 ventriculomegaly was detected in five patients, and in one case, an MRI at 32 weeks
301 also revealed a cavum septum pellucidum cyst and flattened temporal lobe
302 morphology.
303 Table S11: Primer sequences used for Sanger validation of *RNU4-2*, *RNU2-2*, and
304 *RNU5B-1* variants.

307 **References**

- 308 1. Li H. Aligning sequence reads, clone sequences and assembly contigs with
309 BWA-MEM. Published online March 16, 2013. Accessed July 23, 2025.
310 <https://arxiv.org/pdf/1303.3997>
- 311 2. Vasimuddin M, Misra S, Li H, Aluru S. Efficient Architecture-Aware
312 Acceleration of BWA-MEM for Multicore Systems. In: *2019 IEEE International
313 Parallel and Distributed Processing Symposium (IPDPS)*. 2019:314-324.
314 doi:10.1109/IPDPS.2019.00041
- 315 3. Poplin R, Ruano-Rubio V, DePristo MA, et al. Scaling accurate genetic
316 variant discovery to tens of thousands of samples. *bioRxiv*. Published online
317 January 1, 2018:201178. doi:10.1101/201178
- 318 4. Boeva V, Popova T, Bleakley K, et al. Control-FREEC: a tool for assessing
319 copy number and allelic content using next-generation sequencing data.
320 *Bioinformatics*. 2012;28(3):423-425. doi:10.1093/bioinformatics/btr670
- 321 5. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of
322 structural variants and indels for germline and cancer sequencing
323 applications. *Bioinformatics*. 2016;32(8):1220-1222.
324 doi:10.1093/bioinformatics/btv710
- 325 6. Cameron DL, Schröder J, Penington JS, et al. GRIDSS: sensitive and specific
326 genomic rearrangement detection using positional de Bruijn graph
327 assembly. *Genome Res*. 2017;27(12):2050-2060.
328 doi:10.1101/GR.222109.117
- 329 7. Dolzhenko E, Deshpande V, Schlesinger F, et al. ExpansionHunter: a
330 sequence-graph-based tool to analyze variation in short tandem repeat
331 regions. *Bioinformatics*. 2019;35(22):4754-4756.
332 doi:10.1093/bioinformatics/btz431
- 333 8. Chen Y, Dawes R, Kim HC, et al. De novo variants in the *RNU4-2* snRNA
334 cause a frequent neurodevelopmental syndrome. *Nature*.
335 2024;632(8026):832-840. doi:10.1038/s41586-024-07773-7
- 336 9. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics
337 viewer. *Nat Biotechnol*. 2011;29(1):24-26. doi:10.1038/nbt.1754
- 338 10. Lorenz R, Bernhart SH, Höner Zu Siederdisen C, et al. ViennaRNA Package
339 2.0. *Algorithms Mol Biol*. 2011;6:26. doi:10.1186/1748-7188-6-26
- 340 11. Turunen JJ, Niemelä EH, Verma B, Frilander MJ. The significant other:
341 splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA*. 2013;4(1):61-
342 76. doi:10.1002/wrna.1141
- 343 12. Gurovich Y, Hanani Y, Bar O, et al. Identifying facial phenotypes of genetic
344 disorders using deep learning. *Nat Med*. 2019;25(1):60-64.
345 doi:10.1038/s41591-018-0279-0
- 346 13. Hsieh TC, Bar-Haim A, Moosa S, et al. GestaltMatcher facilitates rare
347 disease matching using facial phenotype descriptors. *Nat Genet*.
348 2022;54(3):349-357. doi:10.1038/s41588-021-01010-x
- 349 14. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M. Meffil: efficient
350 normalization and analysis of very large DNA methylation datasets.
351 *Bioinformatics*. 2018;34(23):3983-3989. doi:10.1093/bioinformatics/bty476

15. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049
16. Fortin JP, Labbe A, Lemire M, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol*. 2014;15(11):503. doi:10.1186/s13059-014-0503-2
17. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882-883. doi:10.1093/bioinformatics/bts034
18. Ho D, Imai K, King G, Stuart E, Greifer N. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *CRAN: Contributed Packages*. Published online January 6, 2005. doi:10.32614/CRAN.PACKAGE.MATCHIT
19. Pelegí-Sisó D, de Prado P, Ronkainen J, Bustamante M, González JR. methylclock: a Bioconductor package to estimate DNA methylation age. *Bioinformatics*. 2021;37(12):1759-1760. doi:10.1093/bioinformatics/btaa825
20. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47-e47. doi:10.1093/NAR/GKV007
21. Vos N, Haghshenas S, van der Laan L, et al. The detection of a strong episinature for Chung-Jansen syndrome, partially overlapping with Börjeson-Forssman-Lehmann and White-Kernohan syndromes. *Hum Genet*. 2024;143(6):761-773. doi:10.1007/s00439-024-02679-w
22. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien [R package e1071 version 1.7-16]. *CRAN: Contributed Packages*. Published online September 16, 2024. doi:10.32614/CRAN.PACKAGE.E1071
23. Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin*. 2015;8(1):1-16. doi:10.1186/1756-8935-8-6/TABLES/1
24. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*. 2016;32(2):286-288. doi:10.1093/bioinformatics/btv560