ONLINE SUPPLEMENTARY MATERIAL

Supplementary methods

Study of SOX11 expression in fetal brain

Expression of SOX11 transcripts in fetal structures were evaluated using RNAscope in situ hybridization (ISH) assay and compared with the expression pattern of gonadotropin-releasing hormone receptor (GnRHR). Fetal tissue from Carnegie Stage (CS) 19, 20, 21 and 23 were evaluated. Embryos were collected by the Human Developmental Biology Resource (https://www.hdbr.org) with ethics approval and following appropriate consent. 8µm tissue sections were taken through the brain and the slides were baked for 1 h at 60°C before the paraffin was removed in xylene and the sections were dehydrated in two changes of 100% ethanol. 1 x target retrieval was performed by heating the sections for 20 min at 95°C, followed by protease treatment for 15 min at 40°C. RNAscope probes Hs-GnRHR (ID 553421-C1) and Hs-SOX11-CDS-C2 (ID 443871-C2) were hybridised to the tissue for 2 h at 40°C followed by multiple signal amplification steps. Probe hybridisation was detected using Fast Red (C1) and Fast Green (C2); negative control sections were counter-stained with methyl green for 30 s at room temperature.

SOX11 episignature

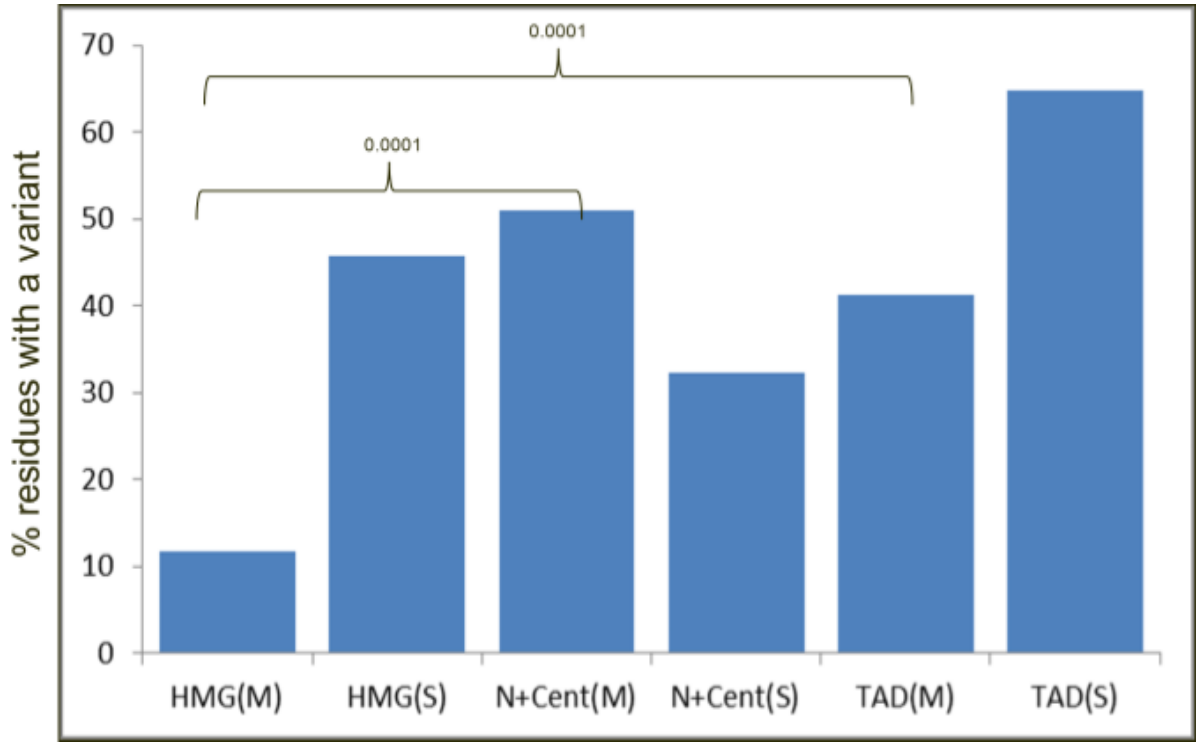
Peripheral blood DNA was extracted using standard techniques. Bisulfite conversion was performed, and samples were analyzed using Illumina Infinium MethylationEPIC BeadChips, according to the manufacturer's protocol. Details of DNA methylation data analysis and episignature discovery were previously described[1–3]. Briefly, IDAT files containing methylated and unmethylated signal intensity were imported into R V.4.0.2 for analysis following normalization with background correction using the minfi package[4]. Probes located on X and Y chromosomes, contained single nucleotide polymorphisms (SNPs) at or near the CpG interrogation or single nucleotide extension, or were cross-reactive with other genomic regions were eliminated from the analysis, in order to ensure that the differences observed between the case and control groups are solely based on methylation changes rather than other potentially confounding factors. Moreover, microarrays with a probe failure rate higher than 5% were removed. The genome-wide methylation density was examined for all the samples, and those deviating from a bimodal distribution were excluded from the analysis. Where indicated, age of the DNA specimens was predicted using the wateRmelon package[5].

Principal component analysis (PCA) was also performed in order to observe the overall batch structure, as well as to detect outlier samples.

For mapping the episignature (probe and feature selection), MatchIt package[6] was used to randomly select controls matched for age, sex, and array type from the EKD, providing a control sample size five times larger than that of the cases, resulted in 50 controls. Increasing the sample size beyond this value impaired the matching quality. After selection of matched controls, PCA was performed to detect outliers but no outlier sample was detecetd.

Methylation levels (β-values) were then transformed into M-values, which were used for linear regression modelling. Using the limma package, linear regression modeling was performed for the purpose of calculating the methylation differences between the case and control groups, along with the corresponding p-value for each probe. Blood cell type compositions, estimated using the algorithm developed by Houseman[7], were also entered to the model matrix of the regression analysis as confounding variables. Subsequently, selection of the significant probes was performed in a four-step process. First, probes with a methylation difference below 5% between the case and control groups were removed. Then, 1000 probes with the highest value obtained from multiplication of the mean methylation difference between the case and control groups by the negative of the logarithm of the p-values were selected. Among these probes, 500 probes with the highest area under the receiver's operating characteristic curve (AUROC) were retained. Finally, probes with pairwise Pearson's correlation coefficients over 0.6, calculated between case and control groups, were eliminated, leaving 224 probes for the rest of the analysis. The methylation levels at these 224 differentially methylated probes (DMPs), considered as the *SOX11* episignature, were utilized to construct unsupervised models including hierarchical clustering in heatmap using Ward's method on Euclidean distance, as well as multidimensional scaling (MDS) by scaling of the pair-wise Euclidean distances between samples. Then, 10 rounds of cross-validation were performed on MDS plot from the 10 *SOX11* samples, of which 9 samples were used as the training set and a single sample was used as the testing set.

Using the 224 DMPs, two binary support vector machine (SVM) classifiers with a linear kernel were constructed using the e1071 package as described previously[1,2]. The first classifier was trained using only the *SOX11* samples against the control samples, and then samples from 38 other Mendelian neurodevelopmental disorders with an established episignature from the EKD were supplied into the model in order to assess the specificity of the model. In order to increase the model's specificity, the second classifier was constructed, using all *SOX11* samples against 75% of control samples and patients from the other 38 Mendelian neurodevelopmental disorders for training, and the remaining 25% for testing. Using the Platt's scaling method, the classifiers generate a methylation variant pathogenicity (MVP) score ranging from 0 to 1 for each sample, where a score near 1 is indicative of similarity to the identified *SOX11*-syndrome episignature, while a score near 0 demonstrates that the sample has a methylation profile different from the *SOX11*-syndrome episignature.



Supplementary Figure 1. Illustration of the percentage of amino acids with a missense variant in each domain of the *SOX11* protein from the Gnomad database.

The percentage of amino acids with a missense variant (M) was significantly lower in the HMG than the N+Cent or TAD domains. (M)=missense, (S)=synonymous variant. HMG = high mobility group, N+cent = n-terminal and central domain, TAD = transactivating domain. A chi-squared test was used to compare the percentage of amino acids with a missense variant between domains.

p.(Ser80Phe)

p.(Trp87Arg)

p. (Arg64Cys)

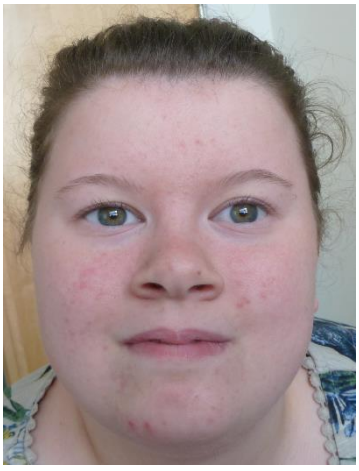
Intragenic SOX11 deletion

p.(Arg51Gly)

p. (Trp87Arg)

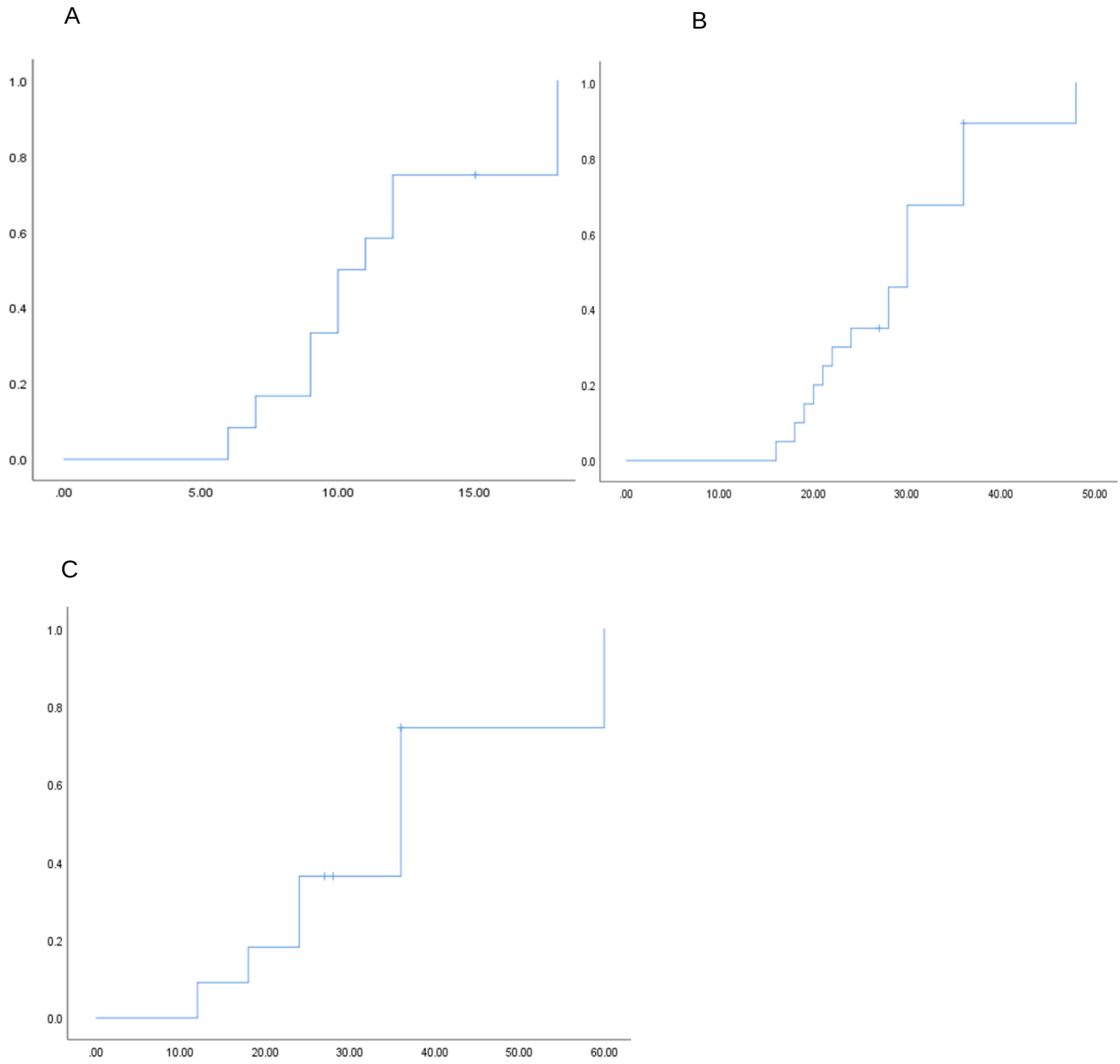
p. (His75Asp)

p. (Ala55Thr)

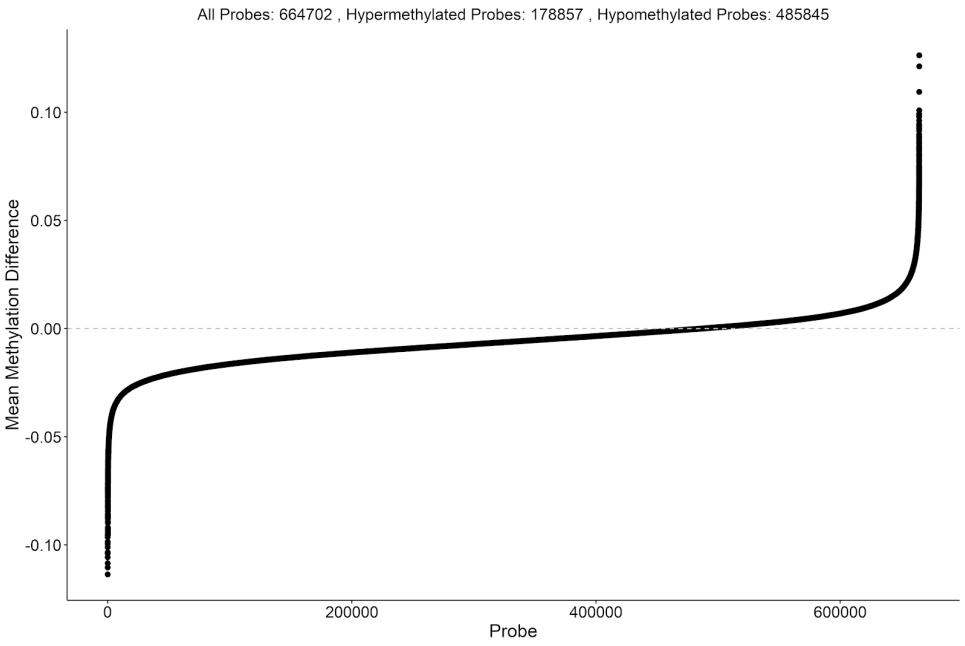


p.(Arg64Pro)

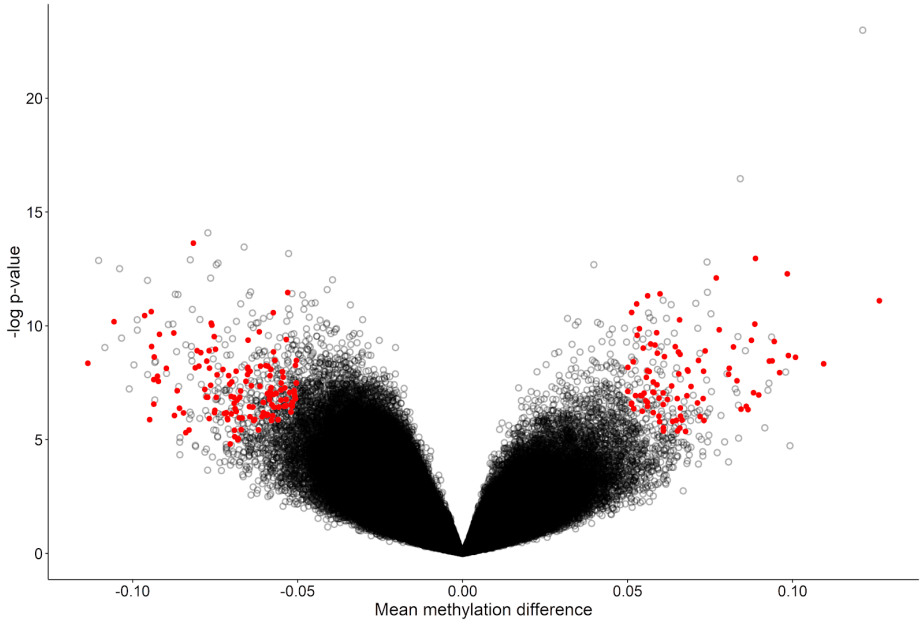
Supplementary Figure 2. Clinical Photographs.
The *SOX11* variant is shown in a text box above the photograph.



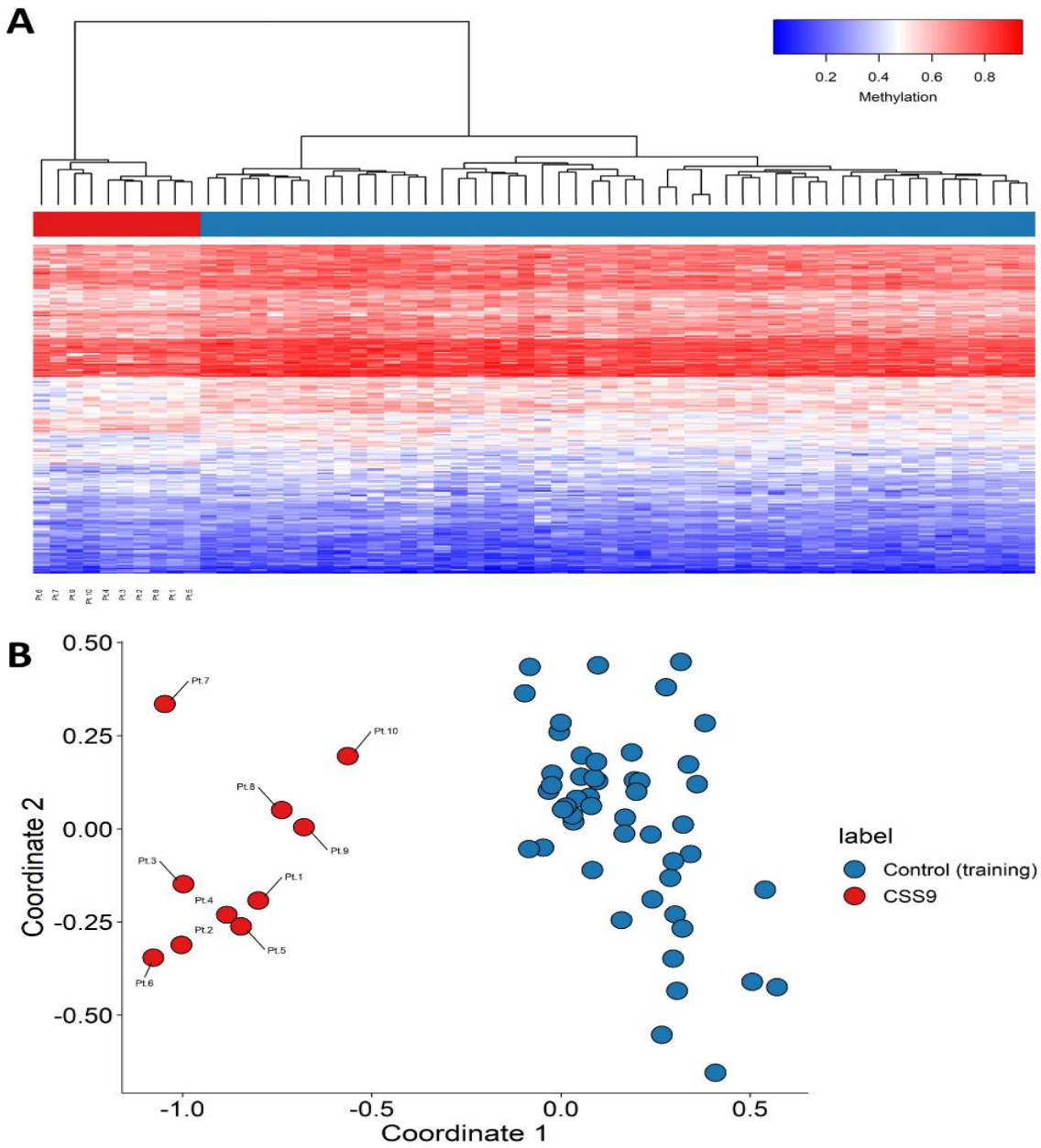
Supplementary Figure 3. Kaplan-Meier analysis of fevelopmental milestones in *SOX11*-syndrome.
S3A. Age in months at which independent sitting achieved.
S3B. Age in months at which independent walking attained.
S3C. Age in months at which first word spoken.



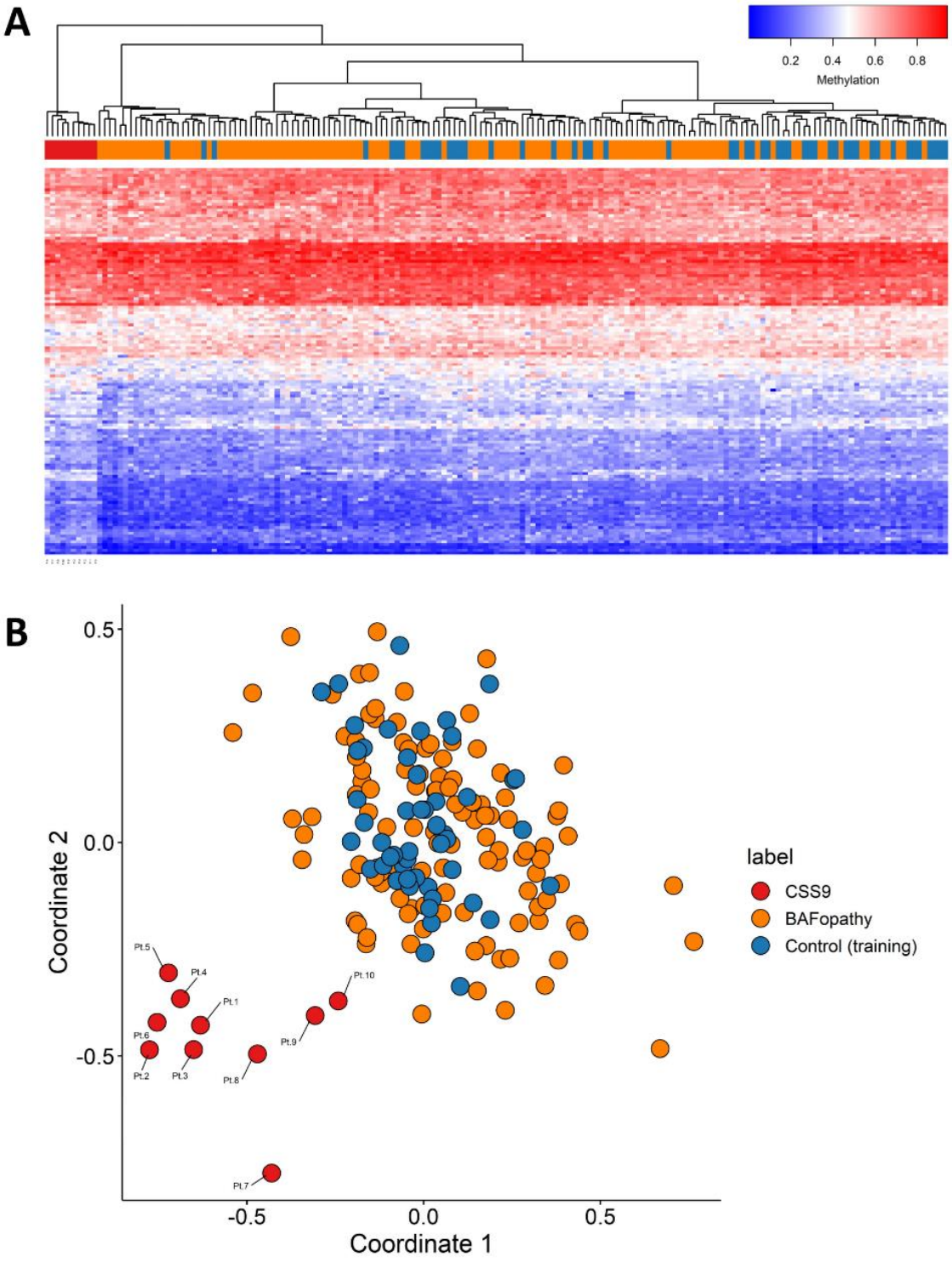
Supplementary Figure 4. Mean methylation difference between 10 *SOX11* and control samples versus individual probes.



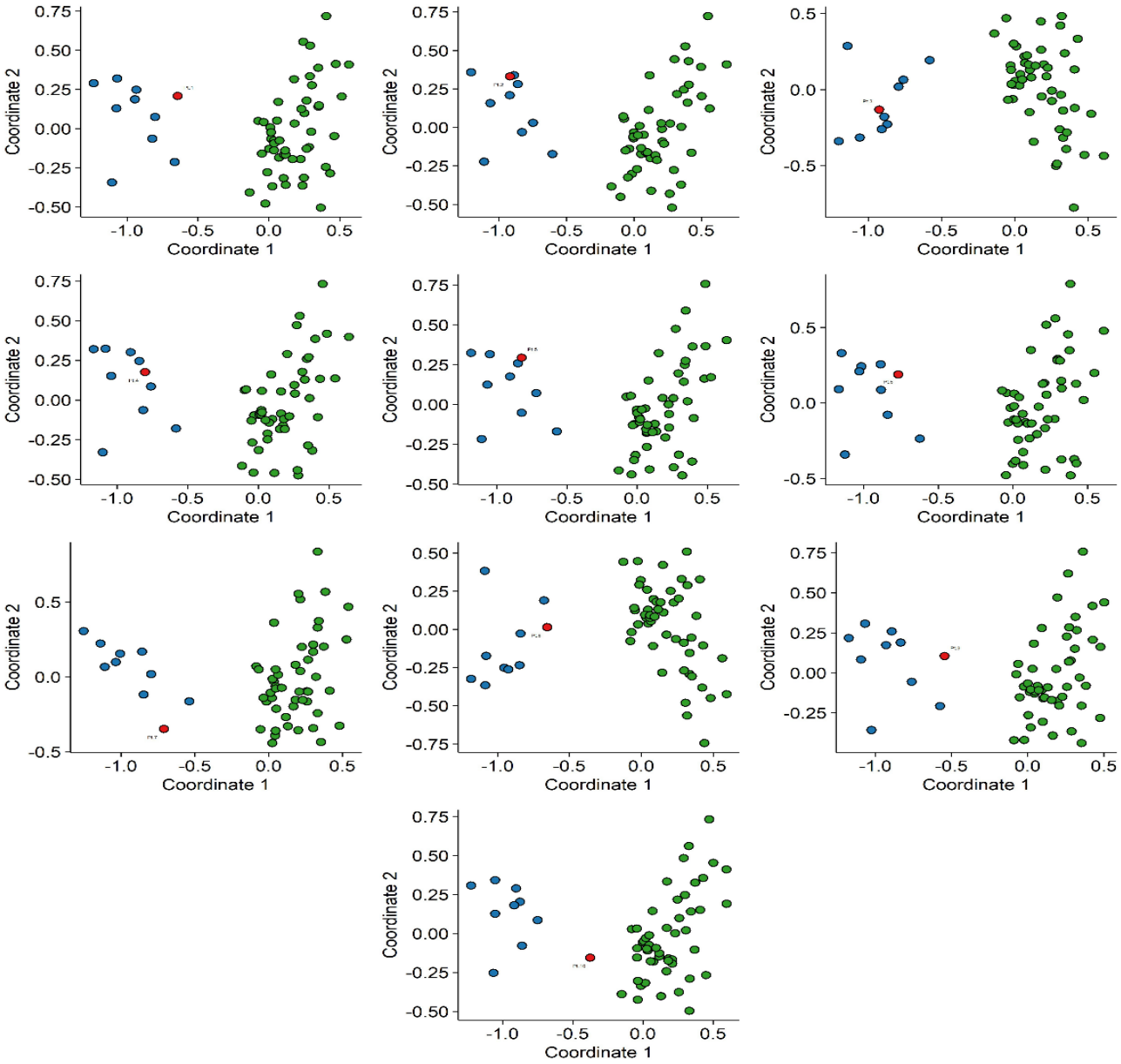
Supplementary Figure 5. Volcano plot of methylation difference between 10 *SOX11* samples and controls versus statistical significance (-log p-value) of individual probes. Red dots represent selected, significant differentially methylated probes (PMDs). Posative and negative mean methylation difference show hypermethylation and hypomethylation, respectively.



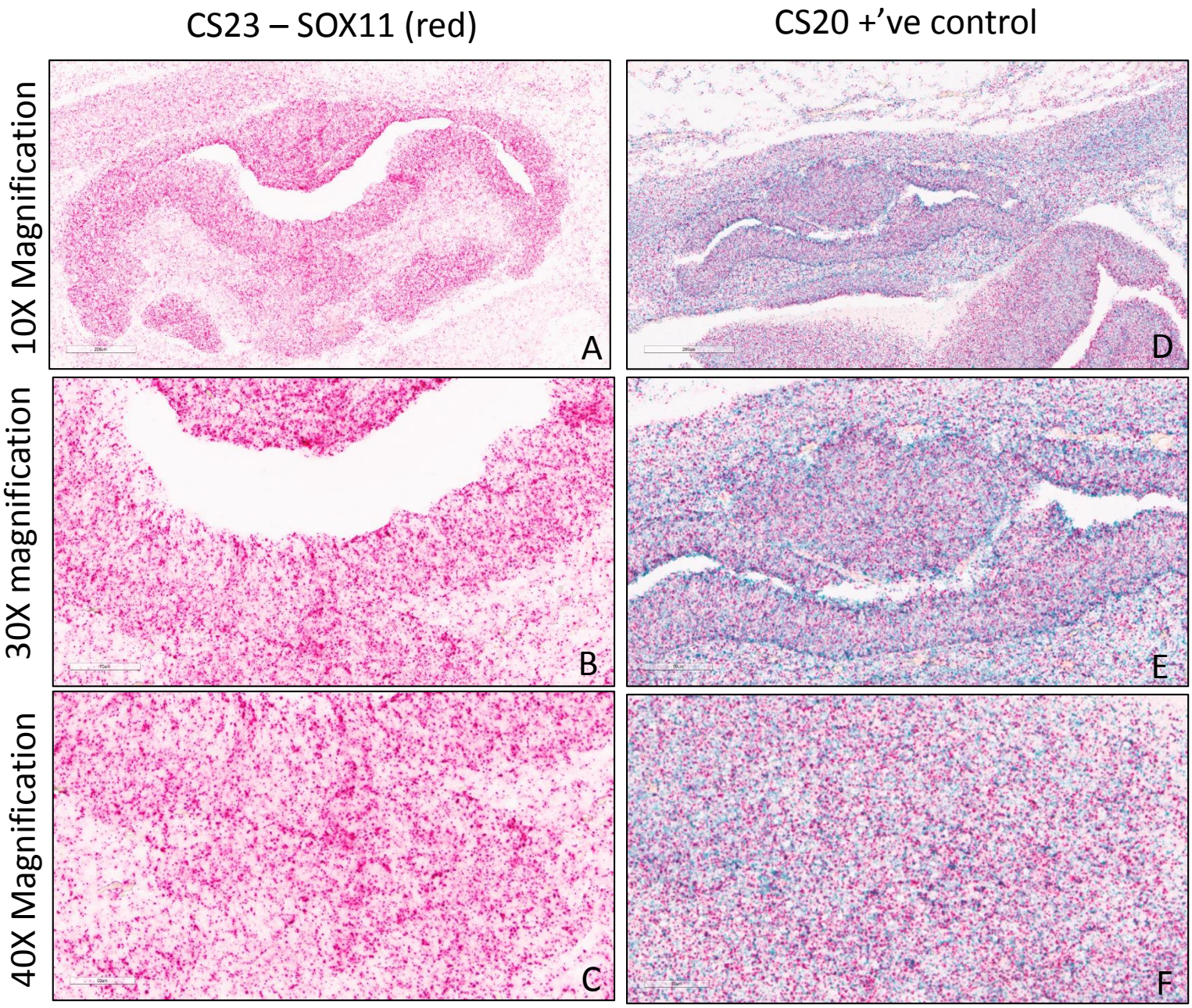
Supplementary Figure 6. Identification of the *SOX11* episignature. A) Hierarchical clustering with Ward's method on Euclidean distance was performed. In the heatmap plot, each row illustrates a selected CpG sate, and each column is related to a sample. The heatmap color scale indicates the range of methylation level; from blue (no methylation or 0) to red (full methylation or 1). This plot conveys that the detected episignature clearly differentiates between 10 *SOX11* samples and controls. B) Multidimensional scaling plot using the selected probes, illustrating the power of the



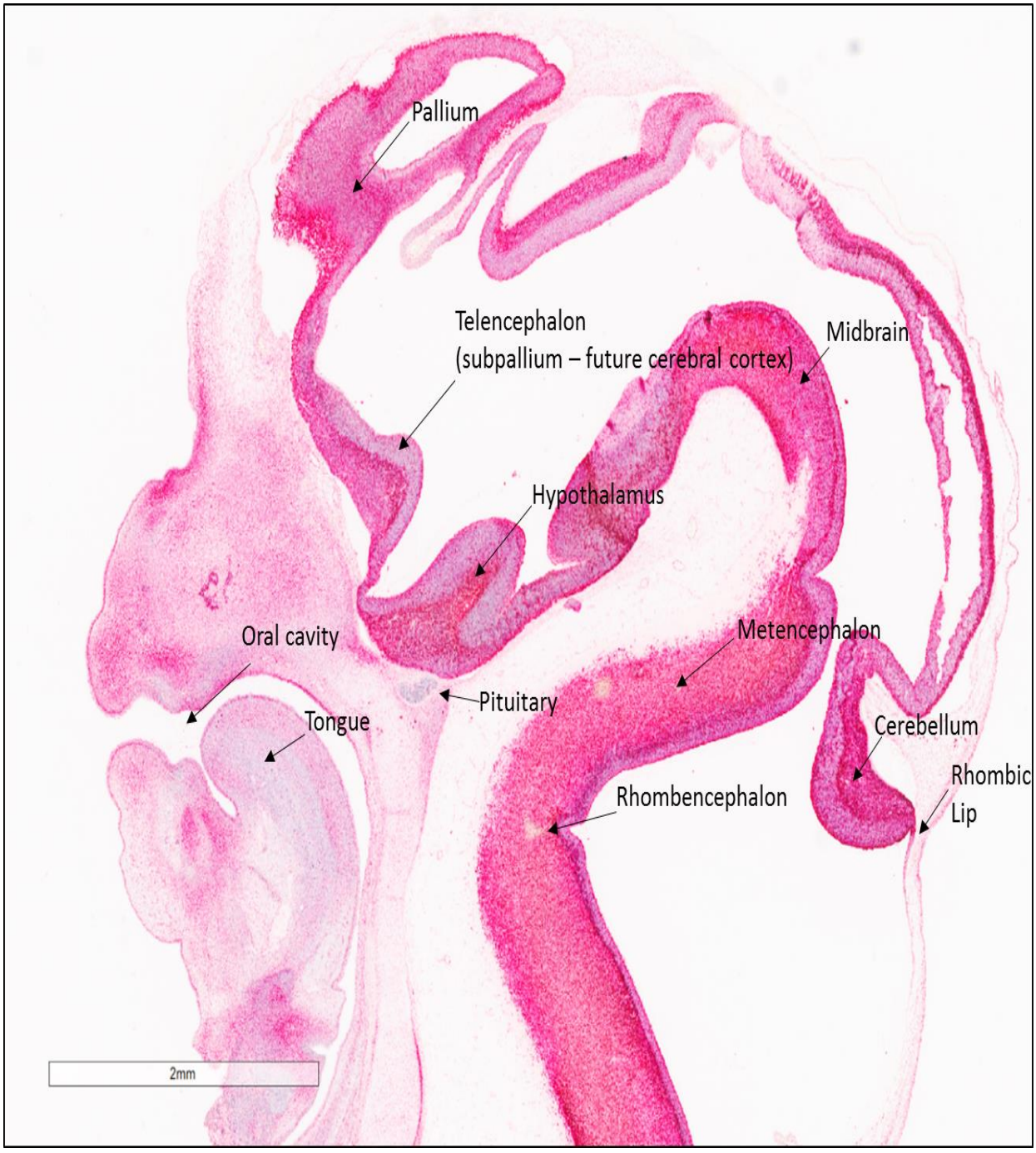
Supplementary Figure 7 Adding BAFopathy complex samples to the *SOX11* episignature. A) Hierarchical clustering, B) Multidimensional scaling. Red, blue, and orange colors represent *SOX11*, control, and BAFopathy complex subjects, respectively.



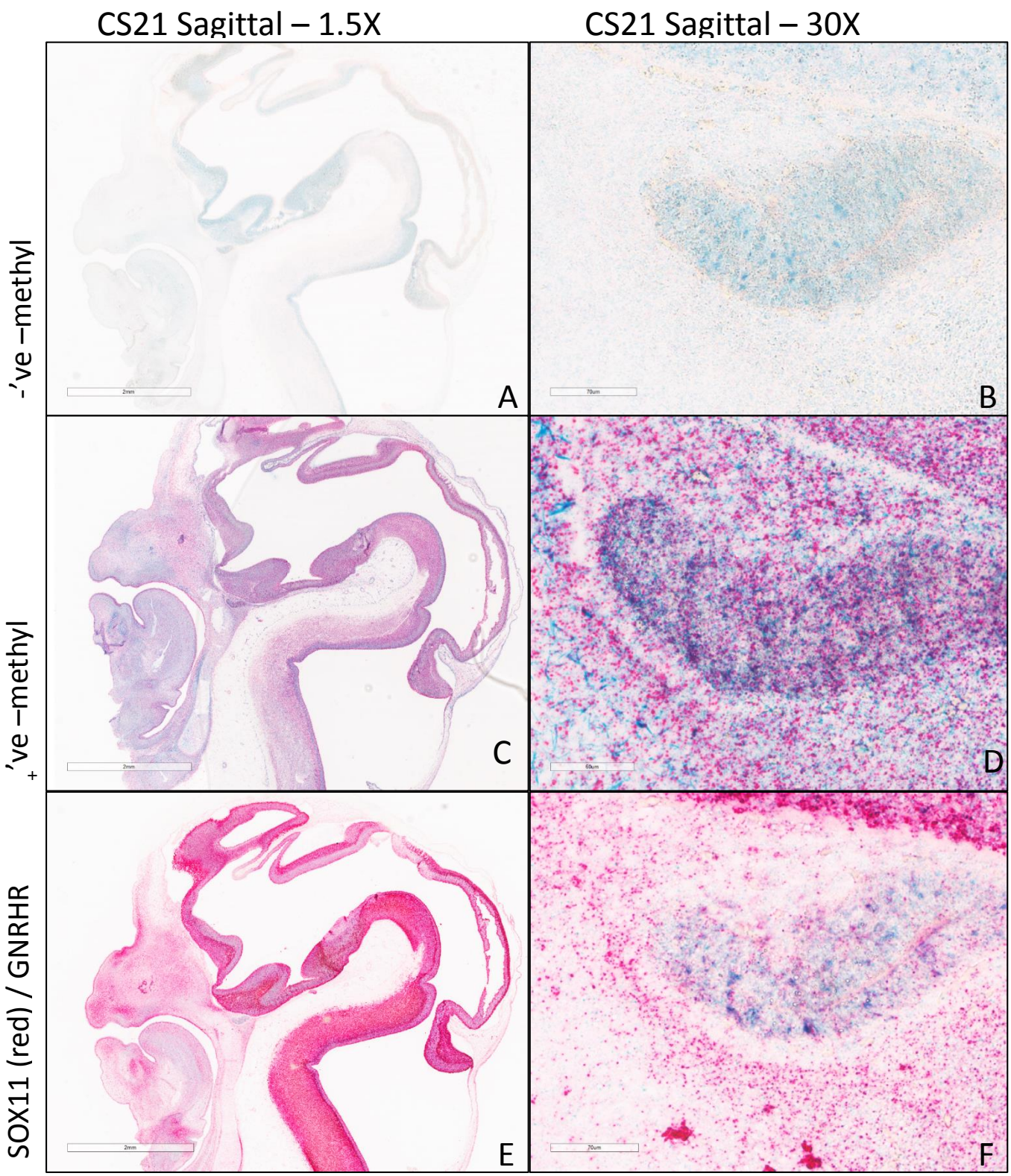
Supplementary Figure 8. Ten rounds of cross-validation were done on a multidimensional scaling plot.



Supplementary Figure 9. Transverse sections (A-C) at increasing magnification through a CS23 head. Images show widespread expression of SOX11 (red), without counterstain, in the developing pituitary gland. (D-F) positive controls at increasing magnifications.



Supplementary Figure 10. Annotated sagittal section of CS21. SOX11 (red) / GNRHR (green). No counterstain.



Supplementary Figure 11. Sagittal sections of CS21. (A-B) negative control (methyl green stain) (C-D) positive control, (E-F) SOX11 (red) and GnRHR (green), showing widespread expression of in fetal cranial structures.

Supplementary Table 1. Summary of clinical data for *SOX11* variant heterozygotes

Please see separate excel file.

Supplementary Table 2. Endocrine test results for SOX11 variant heterozygotes with hypogonadotropic hypogonadism.

Case number (from supplementary table 1)	Endocrine phenotype	Endocrine tests
Case 1	Delayed puberty	LH (u/L) <0.2*, FSH (u/L) <0.03**
Case 9	Delayed puberty	LH (u/L)<0.2*, FSH (u/L) 0.7**
Case 14	Delayed puberty	LH (u/L) <0.03*, FSH (u/L) <0.03**
Case 15	Hypoplastic genitals, intra abdominal testes	LH (u/L) 0.2*, FSH (u/L) 0.7**
Case 17	Delayed puberty	GnRH stimulation test: no rise in LH or FSH
Case 25	Delayed puberty	LH (u/L) <0.2*, FSH (u/L) 0.4 **
Case 37	Congenital hypogonadotrophic hypogonadism, anosmia	LH (u/L) 0.36*, FSH (u/L) 1.6**. GnRH stimulation test: no rise in LH or FSH.
Case 38	Delayed puberty	LH (u/L) <0.05*, FSH (u/L) 0.8**.

LH = luteinising hormone, FSH = follicle stimulating hormone, GnRH = gonatotrophin releasing hormone. Endocrine evaluation was undertaken by Consultant Endocrinologists as part of routine clinical care leading to a diagnosis of hypogonadotrophic hypogonadism being made. Blood LH and FSH levels were diagnostic of hypogonadotrophic hypogonadism.

* LH reference range (u/L). Follicular phase 1.9-12.5, Luteal phase 0.5-16.9.

** FSH reference range (u/L). Follicular phase 2.5-10.2, Luteal phase 1.5-9.1, mid cycle 3.4-33.4, post menopause 23.0-116.3.

Supplementary Table 3. Genomic data for *SOX11* variant heterozygotes.

Please see separate excel file.

Supplementary Table 4. Pathogenic missense variants affecting equivalent residues in *SOX10* and *SOX11* (DECIPHER, ClinVar).

Amino Acid in SOX11	*SOX11* variant	*SOX10* variant
Lys50	Lys50Gln Lys50Asn	Lys105Gln
Arg51	Arg51Leu Arg51Gln Arg51Gly Arg51Trp	Arg106Gly
Pro52	Pro52Ser Pro52Leu	Pro107Arg
Met53	Met53Arg Met53Ile Met53Val	Met108Thr
Ser80	Ser80Phe	Ser135Thr Ser135Asn Ser135Arg Ser135Gly
Trp87	Trp87Arg	Trp142Arg

HMG box for *SOX10* and *SOX11* as defined in PFAM database. Pathogenic *SOX10* variants from DECIPHER and ClinVar.

Supplementary Table 5. Cases for DNA methylation study.

Case ID	LHSC ID	Sex	Age	SOX11 Variant
Case 7	MS2568	m	7	c.239C>T, p.(Ser80Phe)
Case 44	MS3889	m	6.5*	c.305C>T, p.(Ala102Val)
Case 32	MS3890	f	6*	c.152G>A, p.(Arg51Gln)
Case 43	MS3891	f	12*	c.347A>G, p.(Tyr116Cys)
Case 45	MS3892	f	13*	c.154C>T, p.(Pro52Ser); c.235A>G, p.(Ile79Val), Both are on the same allele – in cis.
Case 10	MS2567	m	3	c.250G>A, p.(Gly84Ser)
Pt.7	MS4468	f	26	c.49del, p.(Glu17Argfs*37)
Case 8	MS3003	f	7	c.259T>C, p.(Trp87Arg)
Pt.9	MS2155	f	0.3	c.145A>C p.(Ile49Leu)
Case 5	MS2566	f	19*	c.159G>A, p.(Met53Ile)

Supplementary Table 6. Differentially Methylated Probes.

Please see separate excel file.

Supplementary References

1. Aref-Eshghi E, Kerkhof J, Pedro VP, et al. Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *Am J Hum Genet*. 2020;106(3):356-370. doi:10.1016/j.ajhg.2020.01.019

2. Sadikovic B, Levy MA, Kerkhof J, et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. *Genet Med*. 2021;23(6):1065-1074. doi:10.1038/s41436-020-01096-4

3. Aref-Eshghi E, Bend EG, Colaiacovo S, 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. doi:10.1016/j.ajhg.2019.03.008

4. 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). doi:10.1093/bioinformatics/btu049

5. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013;14(1). doi:10.1186/1471-2164-14-293

6. Wiewel MA, Huson MA, van Vught LA, et al. Impact of HIV infection on the presentation, outcome and host response in patients admitted to the intensive care unit with sepsis; a case control study. *Crit Care*. 2016;20(1). doi:10.1186/s13054-016-1469-0

7. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13(1). doi:10.1186/1471-2105-13-86

8. Hempel A, Pagnamenta AT, Blyth M, et al. Deletions and de novo mutations of SOX11 are associated with a neurodevelopmental disorder with features of Coffin-Siris syndrome. *J Med Genet*. 2016;53(3):152-162. doi:10.1136/jmedgenet-2015-103393

9. Tsurusaki Y, Koshimizu E, Ohashi H, et al. De novo SOX11 mutations cause Coffin–Siris syndrome. *Nat Commun*. 2014;5(1). doi:10.1038/ncomms5011

10. Okamoto N, Ehara E, Tsurusaki Y, Miyake N, Matsumoto N. Coffin-Siris syndrome and cardiac anomaly with a novel SOX11 mutation. *Congenit Anom (Kyoto)*. Published online 2017. doi:10.1111/cga.12242

11. Sekiguchi F, Tsurusaki Y, Okamoto N, et al. Genetic abnormalities in a large cohort of Coffin–Siris syndrome patients. *J Hum Genet*. 2019;64(12):1173-1186. doi:10.1038/s10038-019-0667-4

12. Khan U, Study D, Baker E, Clayton-Smith J. Observation of cleft palate in an individual with SOX11 mutation: Indication of a role for SOX11 in human palatogenesis. *Cleft Palate-Craniofacial J*. 2018;55(3):456-461. doi:10.1177/1055665617739312

13. Cho CY, Tsai WY, Lee CT, et al. Clinical and molecular features of idiopathic hypogonadotropic hypogonadism in Taiwan: A single

center experience. *J Formos Med Assoc*. Published online 2021. doi:10.1016/j.jfma.2021.03.010

14. Diel H, Ding C, Grehn F, Chronopoulos P, Bartsch O, Hoffmann EM. First observation of secondary childhood glaucoma in Coffin-Siris syndrome: a case report and literature review. *BMC Ophthalmol*. 2021;21(1). doi:10.1186/s12886-020-01788-0

15. Hanker B, Gillessen-Kaesbach G, Hüning I, Lüdecke HJ, Wieczorek D. Maternal transmission of a mild Coffin–Siris syndrome phenotype caused by a SOX11 missense variant. *Eur J Hum Genet*. Published online 2021. doi:10.1038/s41431-021-00865-2

16. Wakim V, Nair P, Delague V, et al. SOX11-related syndrome: Report on a new case and review. *Clin Dysmorphol*. Published online 2020:44-49. doi:10.1097/MCD.0000000000000348