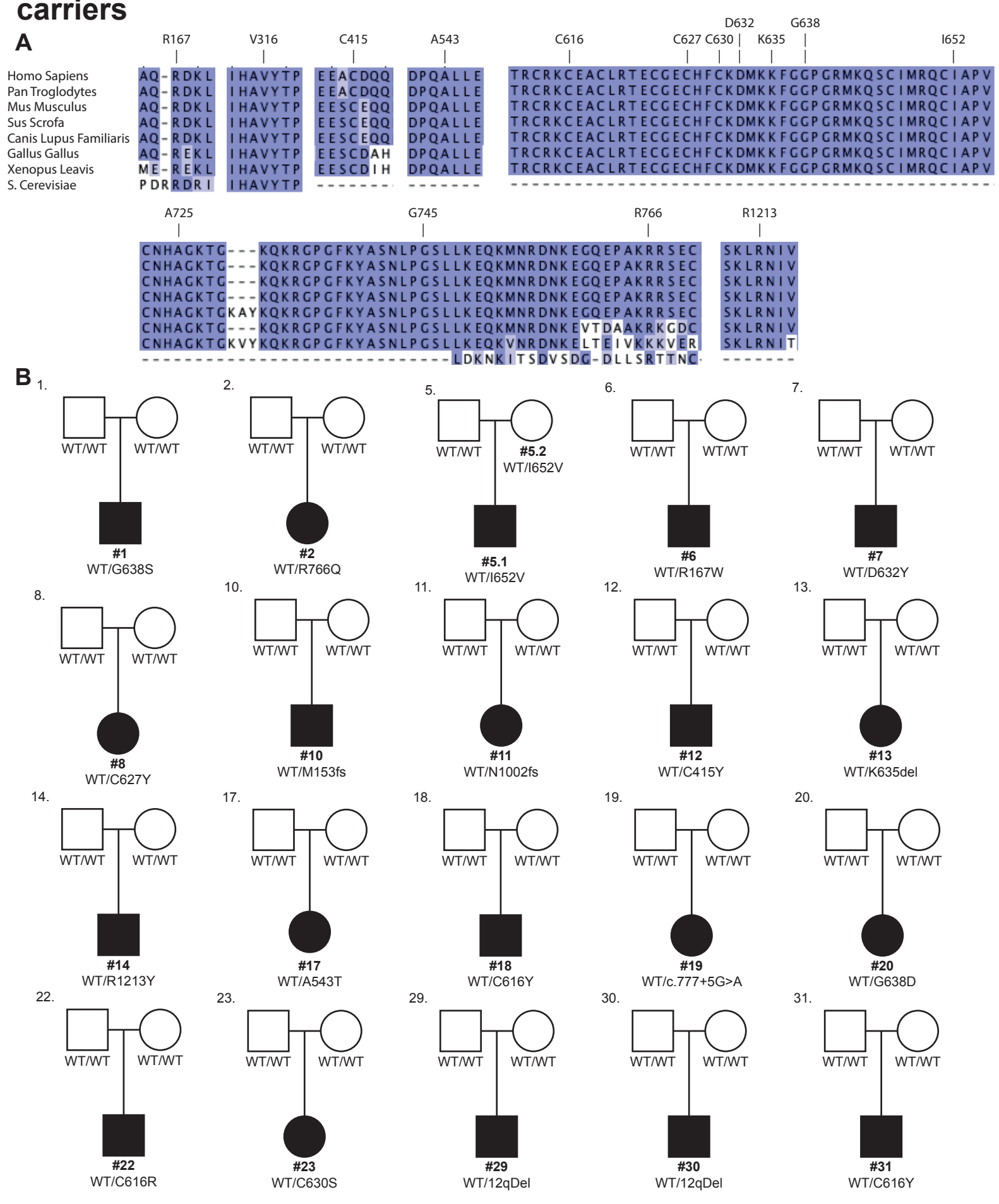
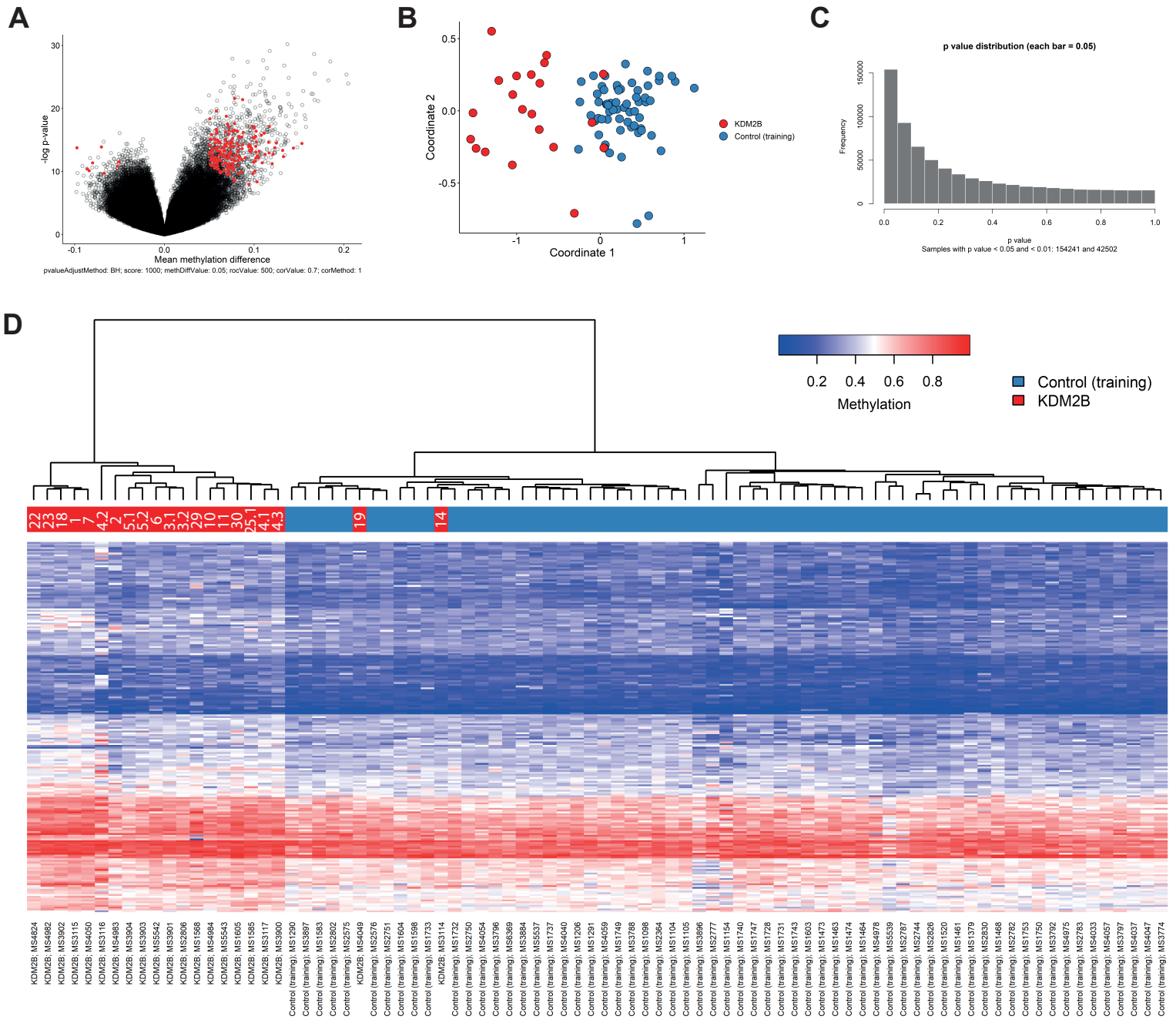
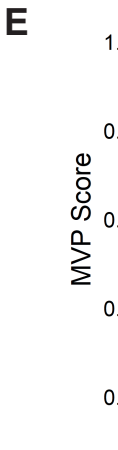
Supplementary Figure 1: A cohort of heterozygous *KDM2B* variant carriers



Supplementary Figure 1: A cohort of individuals with heterozygous KDM2B variants. A) Conservation of KDM2B residues affected by coding altering variants in the cohort. Purple color intensity is indicative of conservation across the species included in the figure. B) Pedigrees for the individuals not included in Figure 1A and/or not published before.

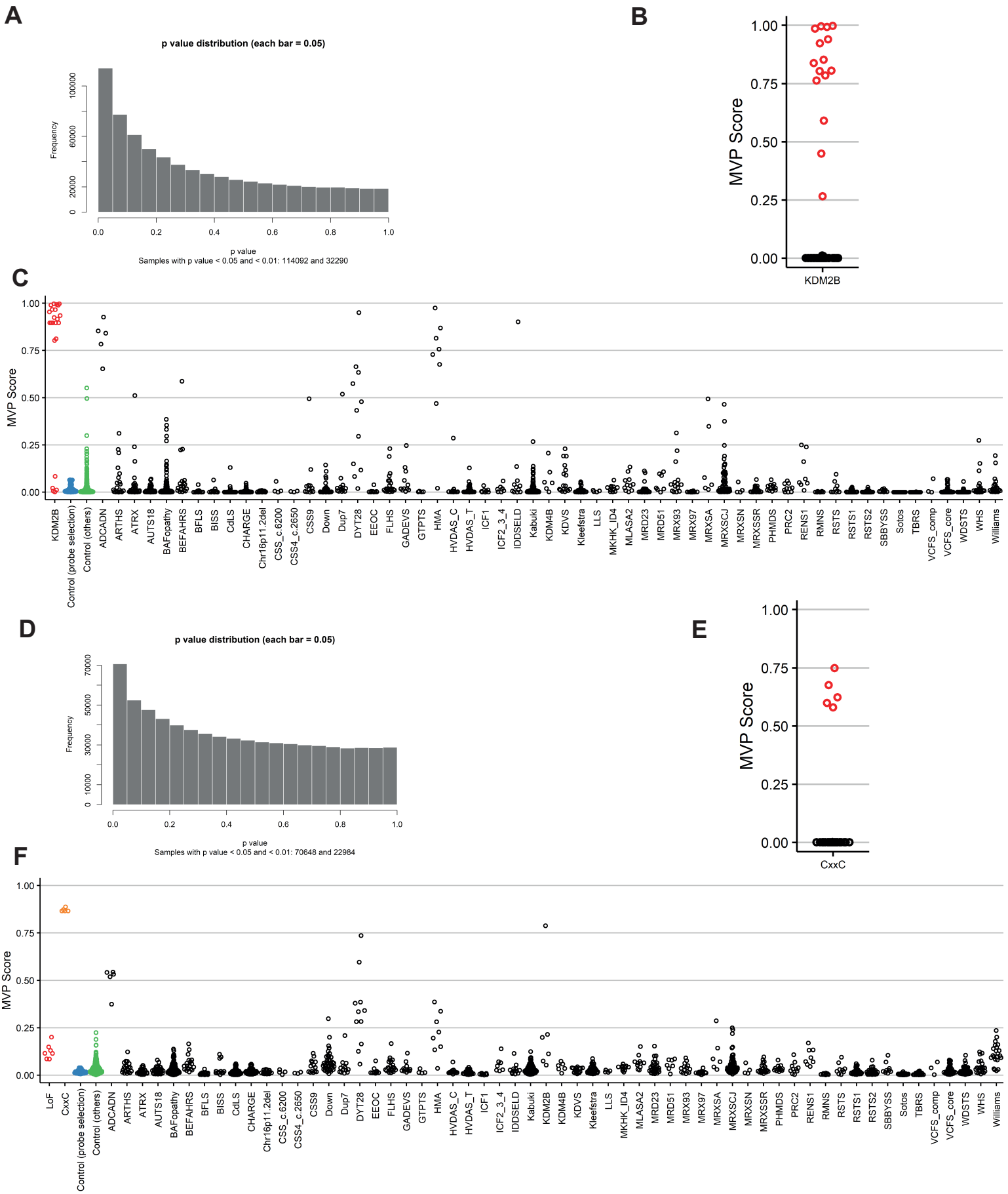
Supplementary Figure 2: Identification of a KDM2B associated episignature





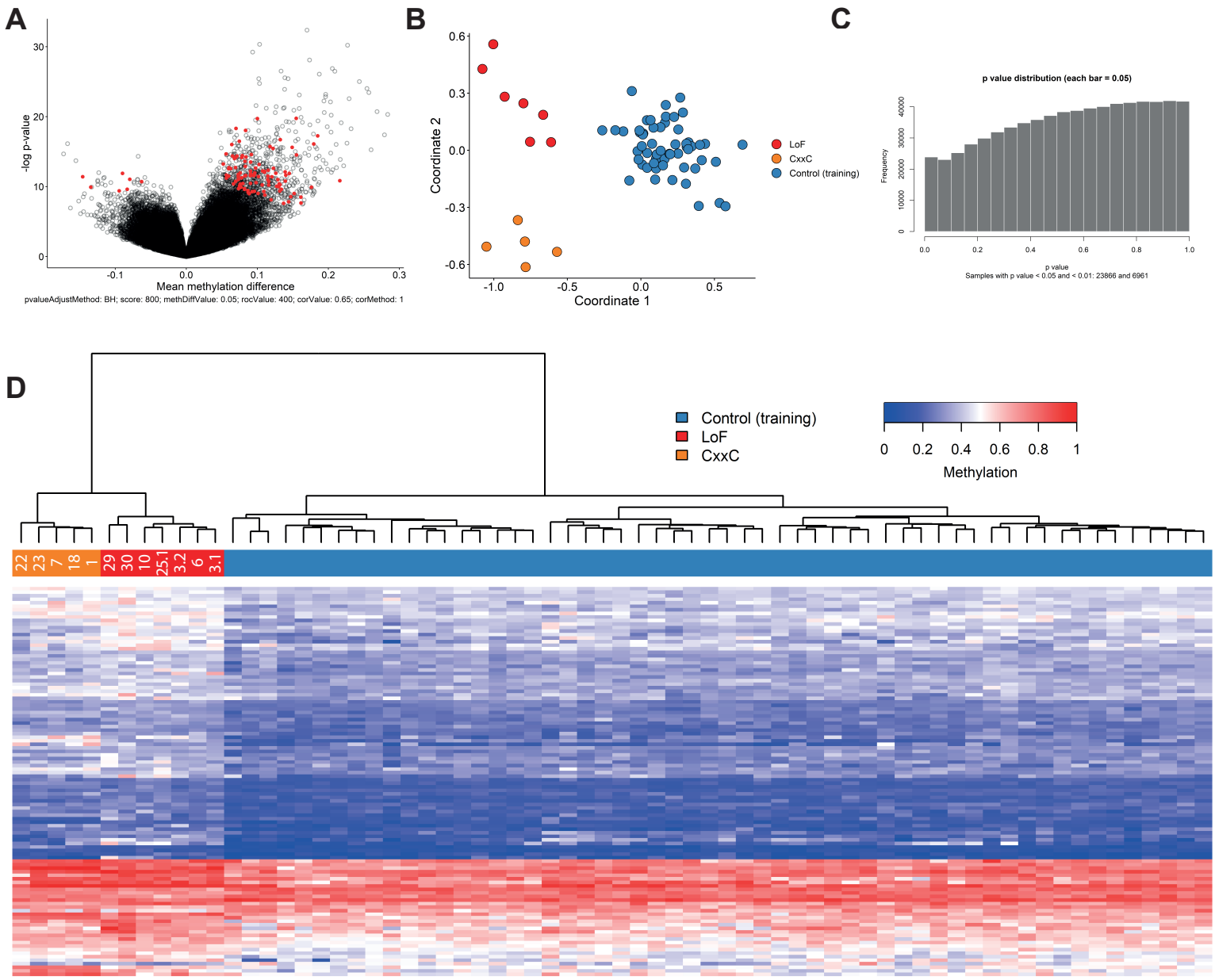
Supplementary Figure 2: Identification of a KDM2B associated episignature. All samples for which methylation array data was available were used to train an episignature. A) Volcano plot indicating all selected probes (red) included in the signature. B) Multidimensional scaling (MDS) plot for selected probes, representing the pairwise distance across samples (red) and controls (blue), based on the top two dimensions. C) Distribution of the (raw) p-values of all probes. D) Heatmap of selected probes and unsupervised hierarchal clustering results indicating the clustering of the majority of variants (red) apart from controls (blue). E) Cross-validation of the episignature. Each test sample (red) was removed from the training cohort and was subsequently tested against the resulting classifier.

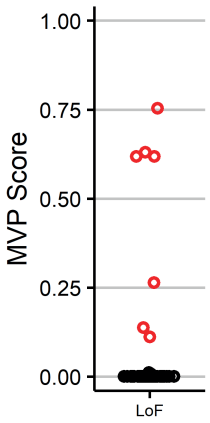
Supplementary Figure 3: Supporting data for figures 2 and 3



Supplementary Figure 3: Supporting data for figures 2 and 3. A) Distribution of the (raw) p-values of all probes for the samples used to train the KDM2B episignature (Figure 2). B) Cross-validation of the KDM2B episignature. Each test sample (red) was removed from the training cohort and was subsequently tested against the resulting classifier. C) Support Vector Machine (SVM) model trained on the samples included in Figure 2. All case samples and matched controls were used for training. Next, each sample was tested against the provided classifier. D) As (A), E) as (B) and F) as (C) for the CxxC episignature (Figure 3).

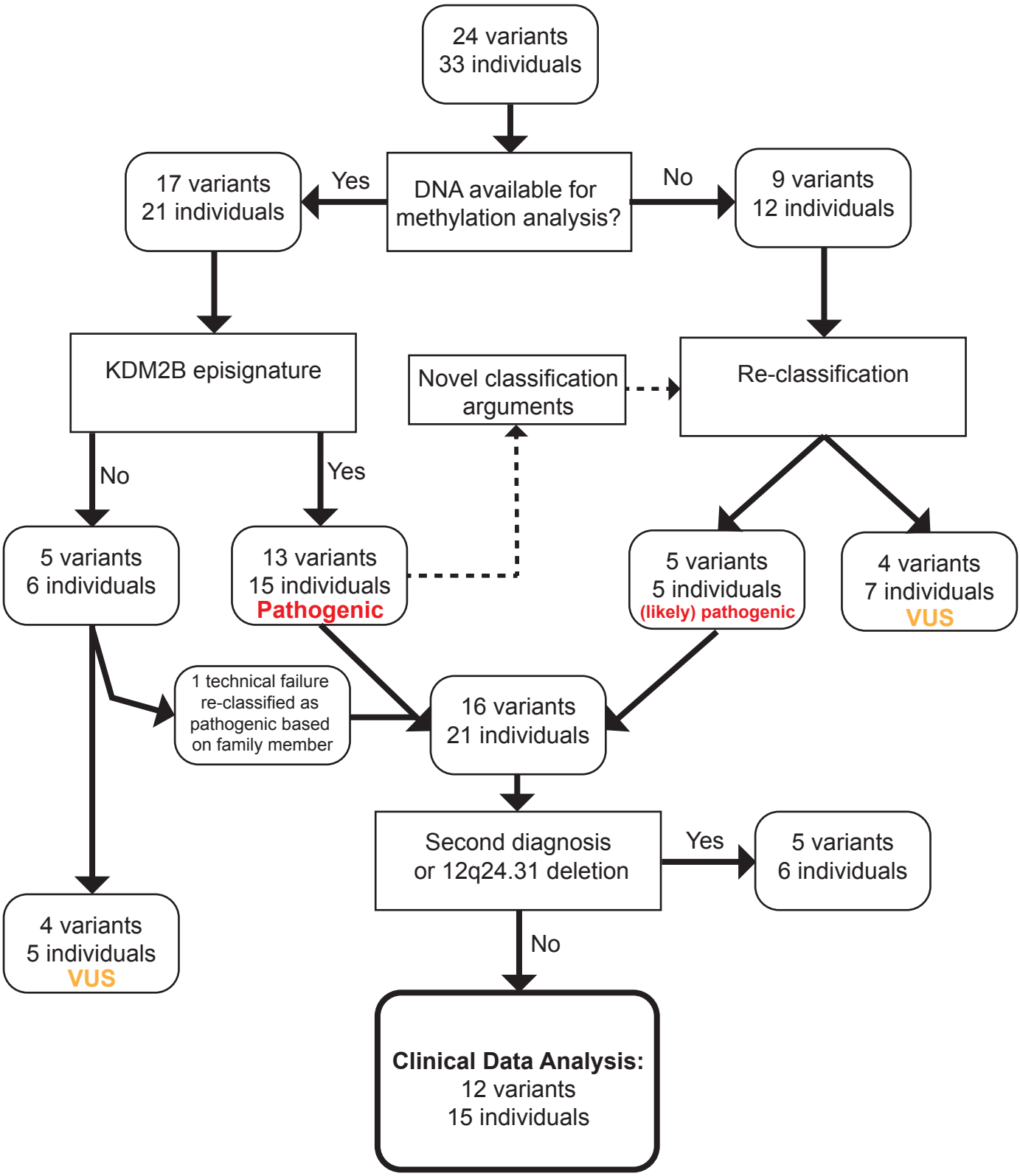
Supplementary Figure 4: A KDM2B-LoF associated episignature





Supplementary Figure 4: Identification of a LoF associated episignature. All samples representing a loss-of-function (LoF) variant and included in the KDM2B episignature, were used to train a LoF-variant specific episignature. A) Volcano plot indicating all selected probes (red) included in the LoF episignature. B) Multidimensional scaling (MDS) plot for selected probes, representing the pairwise distance across CxxC variants (orange), LoF variants (red) and controls (blue), based on the top two dimensions. C) Distribution of the (raw) p-values of all probes. D) Heatmap of selected probes and unsupervised hierarchal clustering results indicating the episignature's ability to decipher KDM2B variants (red and orange) from controls (blue), and to decipher CxxC variants (orange) from LoF variants (red). E) Cross-validation of the LoF episignature. Each test LoF sample (red) was removed from the training cohort and was subsequently tested against the resulting classifier.

Supplementary Figure 5: Patient inclusion flow chart



Supplementary Figure 5: Patient inclusion flow chart. Flow chart indicating which samples/variants were used in the different sub-cohorts used throughout this study.

Delineation of a KDM2B-related neurodevelopmental disorder and its associated DNA methylation signature

Clinical histories

Pathogenic variants

KDM2B_1 p.(Gly638Ser)

This boy presented at age 5 to the department of Clinical Genetics. He was referred to us because of developmental issues and congenital abnormalities.

He was born at term with BW 3545 grams to non-consanguineous parents. During pregnancy at 34 GW the fetus was diagnosed with an atrial flutter and the mother was treated with sotalol.

After birth bilateral postaxial polydactyly of both hands and feet were noted. This segregated as an autosomal dominant trait with reduced penetrance in the mother's family. He had normal sinus rhythm at birth, cardiac echo showed two small ventricular septal defects (VSD), open ductus botalli and patent foramen ovale/ atrial septal defect (ASD) type 2. His neonatal period was normal without feeding difficulties or hypotonia. At follow up the VSDs closed spontaneously. The atrial septal defect was successfully corrected at age 3.5 years. Motor development was near normal with crawling at 11 months and independent walking at 18months, parents noted he was always a bit clumsy. His speech development was delayed with uttering first comprehensive words after his second birthday. Due to persistent speech delay and behavioral abnormalities he was diagnosed with autism spectrum disorder at age 4 and below-average intelligence SON-IQ 86. Physical examination at 5 years: height 114.2 cm (-0,1 SD), OFC 52 cm (+0,3 SD). Friendly and cooperative boy. He has a rounded face, short palpebral fissures, normal palate, large ears with flattened antihelix with absent crus superior and prominent ear lobes. Two small hyperpigmentations on the chest. Surgical scars on the fifth fingers from removal of 6th digits. Bilateral poly-syndactyly of the feet. He had normal male genitalia. Renal ultrasound was normal. At age 7 years he is a happy boy doing well. He swims and rides a bicycle. He attends special schooling for children with challenging behavior, with a maximum of 15 children in 1 group. He is developing his social skills. Illumina® CytoSNP-850K SNP-array showed no abnormal CNVs.

KDM2B_3.1, KDM2B_3.2 p.(Arg1124*)

This girl was born at 37 GW with ceasarian section due to oligohydramnios and breech presentation. She was diagnosed with a ventricular septal defect (VSD) and double chambered right ventricle (DCRV) after birth. The VSD closed spontaneously, the DCRV was surgically corrected at 10 days after birth/ 5 months of age. Feeding difficulties resolved after the cardiac surgery. She was hypotonic and her development was globally delayed. She walked independently at 30 months of age. She babbles but has no comprehensive speech.

At 3 years 4 months she was extensively evaluated. Her height was 102 cms (0.5 SD), weight 18 kgs (+1.4 SDS) and OFC 51.7 cms (+1.7 SDS). She had a high forehead, arched eyebrows, epicanthal folds, full upper eye lids and cheeks, heterochromia of the irides, anteverted nares, deep philtrum, and bilateral hockey-stick creases on the palms. Absence of eye contact and speech, stereotypic movements were present. She had a broad-based gait, normal muscle tone, normal reflexes. Bayley-

III-NL at 40 months verbal production equivalent to 6 months, verbal comprehension of 11 months and cognition of 11 months. She had hypermetropia (+3 dpt) and astigmatism. Brain MRI showed global atrophy, underdeveloped hippocampi, delayed myelination, cavum septum pellucidum et vergae and a small plexus choroid cyst. Metabolic screening was normal.

Bayley-lll-NL at age 6 years 11 months verbal production equivalent to 12 months, verbal comprehension of 15 months and cognition of 24 months. Fine motor skills were equivalent to 25 months and gross motor skills to 21 months. She is overall a friendly and timid girl. She is interested in other children. She needs clear routines and structures, she is easily overwhelmed by environmental stimuli.

SNP array and trio exome sequencing were performed simultaneously. SNP array showed a terminal 5.8 Mb 22q13 deletion, leading to a diagnosis of Phelan-McDermid syndrome. Subsequent karyotype showed a ring chromosome 22 (46,XX,r(22)p11q13.3)).

Family history: Her father (KDM2B_3.2) has had learning difficulties, and has only attended primary education. He has been diagnosed with ADD. He has COPD. His height is 188 cms and OFC 54.5 cms (-1.9 SDS). Cardiac examination showed normal echo and ECG. He has a normal male karyotype (46,XY). His son from a previous relationship has ADHD and learning difficulties, but he was not available for this study.

The mother of 3.1 has mild intellectual disability – IQ 60. She has a normal female karyotype (46,XX), normal SNP array result and a normal ID gene panel.

KDM2B_4.1, KDM2B_4.2, KDM2B_4.3 p.(Val316Ile)

The index patient is a 15 year-old boy, presenting with epilepsy, PDD-NOS, ADHD and mild intellectual disability.

He was born the first of a dizygotic twin pregnancy, after 33+5 gestational weeks and with a birth weight of 2200 gram (0 SD). He had a global developmental delay (independent walking at 18 months), mild intellectual disability, ADHD, PDD-NOS and behavioral difficulties. At the age of 7 years, total IQ was 67, VIQ 71, PIQ 68, VS 72 (WISC-III). He received physical therapy to support his motor development and institutional support for his behavioral difficulties. He developed epilepsy when he was 14 years old, with several different seizure types, including tonic-clonic and more migraine-like with aura, and seizures with blinking and lip smacking. 24-hr video EEG showed an normal background pattern, with centro-parietal functional abnormalities which increased when eating. Focal epileptic activity occured during sleep. MRI was normal. Physical examination showed normal growth parameters, plagiocephaly, low hanging columella, thin lips and square chin.

Metabolic testing was normal. Array-CGH showed a 334kb interstitial duplication on 7q36.2 (including part of the DPP6-gene). Segregation analysis showed that father (Individual KDM2B_4.2) has the same duplication. Trio-exome sequencing identified a paternal missense variant in KDM2B.

Family history: He has a healthy twin sister, attending a normal secondary school. His older sister (KDM2B_4.3) was born premature and small for gestational age, further neonatal history was uncomplicated. As a 3-year old, she presented with severe behavioral problems and later she was diagnosed with developmental delay. EEG made because of presumed drop attacks, was normal. Brain MRI at age 5 years and metabolic testing were normal. She has moderate intellectual disability (WISC-III total IQ 48) and autism spectrum disorder. Seizures were never noted. Physical examination shows height and weight above average (179.3 cm (+ 1.5 SD), 78.2 kg (+1.75 SD)). Facial features include high anterior hairline, deeply set eyes, horizontal eyebrows, bifid nasal tip, thin upper lip and

broad chin with vertical dimple. Targeted sequencing revealed she has the same KDM2B variant as her brother.

Their father (KDM2B_4.2) attended special schooling because of learning difficulties. Facial features include low hanging columella, underdeveloped alae nasi, thin lips and prominent, square chin. The mother did not report learning difficulties. Several of her family members have learning difficulties and/or behavioral problems.

KDM2B_6 p.(Arg167Trp)

The index individual is a male, second child of non-consanguineous parents. He was born prematurely at 36+1 weeks with birth weight 2925g. Right after the birth, right-sided ptosis was noted, but neurological and ophthalmological investigations did not reveal any abnormalities. Development was considered normal, until he went to school, where speech/language delay was noted. He attended a speech therapist due to these problems. At that time, IQ was tested and showed non-verbal IQ=97. He was diagnosed with ADHD, for which he receives Ritalin, with good effect. He also has problems with sleep, for which he receives melatonin. Due to the common ear infections, he had tympanostomy tubes placed bilaterally. Additionally, he had adenotonsillectomy, after which he had a significantly prolonged bleeding. He was investigated for bleeding causes, but no coagulation defects were identified. He had normal vision and hearing. No data regarding abdominal or heart ultrasound were available. Clinical evaluation revealed mild facial dysmorphism, and hyperpigmentation of skin. His height, weight, and OFC were within normal range (131.8cm (-0.1 SD), 24.1kg (-1 SD), and 50cm (-1.5 SD)). From the family history, older brother and sister of mother were known to have “language problems”, but no other features. The proband was tested for Noonan syndrome gene panel, as well as chromosomal microarray was performed, but the results were normal. Therefore, clinical trio WES was performed which revealed *de novo* VUSes in three genes: *CDH11* p.(Tyr273*), *KDM5D* p.(Arg769Trp), and *KDM2B* p.(Arg167Trp).

KDM2B_7 p.(Asp632Tyr)

The index individual is a male, second child of non-consanguineous parents. He was born at term (41+3) with good start (Apgar score 9/10) and 4370g birth weight. Soon after birth, generalized hypertonia was noted and brain MRI was performed, but no abnormalities were found. Additionally, he had neonatal feeding difficulties with regurgitation. He had global developmental delay with more pronounced speech/language delay, speaking <10 words at the 3 years old age, but with good understanding of spoken language. He attended speech and physical therapist due to the mentioned problems. He is social, plays with other children, but behavior sometimes is impulsive and even aggressive. Further, due to the hypertonia, at the age of 3-3.5 years he developed multiple progressive contractures, including in elbows, knees, and feet, with exorotation of feet and specific (penguin-like) gait, as well as he had femoral fracture due to a fall (related to the unstable gait). Because of the new clinical features, primary arthrogryposis, or a syndromic skeletal dysplasia was suspected, but leg and arm X-ray did not show any abnormalities. He has no scoliosis or kyphosis, or other skeletal abnormalities and normal height, weight, and OFC (100cm (-0.5SD), 16.5kg (+0.1SD), and 49cm (-1SD), respectively). On neurological examination, he had a bit increased muscle tonus, but no paresis, or abnormal reflexes. He had no history of seizures. He had good hearing and vision (with glasses to correct myopia). IQ was not tested. A heart murmur was later noted, and on a further examination ASD was identified, which required surgical closure. He was investigated for metabolic and lysosomal storage disorders, as well as for chromosomal aberrations by chromosomal microarray, but no abnormalities were identified. Therefore, clinical trio WES was performed, which revealed *de novo KDM2B* variant p.(Asp632Tyr).

KDM2B_8 p.(Cys627Tyr)

She is now 6.5 years of age. Her main phenotype is single kidney and congenital heart defect: atrial septal defect –ASD- and pulmonic stenosis.

During pregnancy- normal follow up including all ultrasound scans beside:

- single kidney that was observed in her scan at 16 GW.
- Mild polyhydramnion in 36 GW.

Amniocentesis during pregnancy with normal chromosomal micro array.

She was born with caesarian section, her birth weight was 3.7 Kg. She was hospitalized for 11 days after birth with no need for ventilation, but than was diagnosed with the heart murmur and the ASD. She was treated (ASD closure + pulmonic valve dilatation) at the age of 6 with cardiac catheterization. After birth her singe kidney was diagnosed with cortical cysts and calcifications. She has normal kidney function at of 6.5 years of age. She has short stature (below the 3rd percentile) and mild language/ developmental delay.

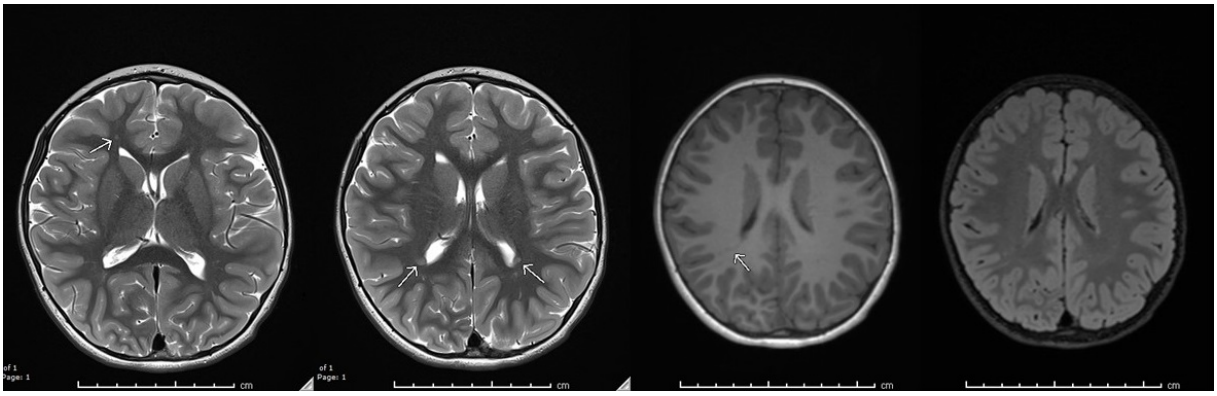
Normal eye and ear exam.

KDM2B_10 p.(Met153fs)

Clinical summary not available.

KDM2B_11 p.(Asn1002Sfs35)

She was born at term with BW 3583 grams to non-consanguineous parents via uncomplicated pregnancy. No complications after birth until 5 months of age with the development of infantile spasms. Her MRI at the time was notable for periventricular nodular heterotopia and multiple focal cortical dysplasias. She began walking around 18mos of age, first words around 15mo. Her developmental progress has slowed. She has global developmental delay and around age 6 she was noted to have an age equivalency of less than 2.4 yo on standardized testing and moderate intellectual disability. She has autism spectrum disorder and hyperactivity. She is now 7.5 yo at last follow-up with refractory epilepsy with tonic seizures and epileptic spasms despite multiple medications and undergoing corpus callosotomy. She was noted to have hypopigmented macules on her skin and retina. Otherwise she has had a normal echocardiogram. The genetic diagnosis was made on clinical exome sequencing from trio of proband and both parents. Her most recent measurements included weight of 18.2 kg (19%ile)*; (-0.87), height of 112 cm (28%ile)*; (-0.59), and head circumference of 48.25 (-1.45SD).



KDM2B_18 p.(Cys616Tyr)

Clinical summary not available.

KDM2B_20 p.(Gly638Asp)

Patient born from healthy non-consanguineous parents. She developed with speech delay around age 5. Normal primary school until 10 years old then adapted classes towards job learning. Psychometric evaluation WISC 5 noted below average skills with difficulties with verbal comprehension but absence of ID. Speech therapist until now. Scoliosis treated by physiotherapist. Some autistic features but mild. No hearing loss.

KDM2B_22 p.(Cys616Arg)

The proband is a male, second child of unrelated parents. Anamnesis was uninformative. He was born at 41 weeks of an uneventful pregnancy. At birth, his weight was 4,650 g (3.1 SD), length 53 cm (1.4 SD), and OFC 37.5 cm (2.4 SD); Apgar score was 8-10. Clinical evaluation documented left eye anophthalmia, bilateral hypoacusia, which required prosthesis implantation, and left kidney agenesis. During the neonatal period, hypotonia of upper limbs, hypertonia lower limbs, and sucking difficulties were noted. MRI highlighted a somewhat formed eye and agenesis of optic nerve on the left side. Left lateral ventricle dilation, and moderate temporal dilation of the subarachnoid spaces bilaterally with enlarged Rolandic opercula were also observed. Ultrasonography documented left kidney agenesis and increased size of the right kidney, likely due to functional compensation. Echocardiogram revealed patent foramen ovale. Audiometric evaluation (10 months) revealed a general improvement with slight and moderate hypoacusia at the right and left ear, respectively. Developmental milestones were overall delayed: seated position 10 months, standing position 14 months, babbling 12 months. At last clinical evaluation (16 months), the child presented with language retardation and craniofacial dysmorphisms including facial asymmetry, posterior left plagiocephaly, broad and bulging forehead, left eye anophthalmia with homolateral hypoplastic eyebrow, slightly asymmetrical lips with smaller left side, eversion of lower lip, and retrognathia. No epilepsy nor behavioural issues were observed. Anthropometric assessment showed a height of 79.5 cm (-0.3 SD), weight of 10,930 g (0.3 SD), and OFC of 47.3 cm (0.2 SD). CGH array and karyotype were normal.

KDM2B_23 p.(Cys630Ser)

A 3-year-old girl was referred to our institution for further investigation of multi organ lesions including intellectual disability, agenesis of left kidney and atrial septal defect. She was the second child of healthy non-consanguineous parents. She and her mother have no complications in the pregnancy period. She was born at 38 weeks of gestation and presented weight of 2954 g (+0.4SD), length of 49.5 cm (+0.7SD), and head circumference of 32 cm (−0.7SD). Arterial septal defect (ASD) was detected at 3 weeks after birth, and she underwent operations for ASD at age of 1 year 9 months. Frequent urinary tract infections were observed. Renal imaging studies revealed vesicoureteral reflux (VUR) in the right kidney and the defect of the left kidney. She developed severe developmental delay (DQ: 20 at age of 2 years and 1 month), intellectual disability and hypotonia. A facial finding includes mild hypertelorism. At the age of 2 years and 10 months, she underwent operation for congenital nasolacrimal duct obstruction. At the age of 3 years 5 months, her body height and weight were 86.7 cm (−2.5 SD), 9.6 kg (−3.0 SD). She cannot speak even a word. Muscle weakness in lower limb was prominent. She cannot walk by her own but can keep sitting position.

Her VUR was improved with age. She lives a stable life without medication receiving rehabilitation for developmental delay.

KDM2B_25

25.1 This girl has been published by Chouery et al in 2013. She is now 12.6 years old, her height is 150 cm; her OFC: 47.6cm and her weight: 40 Kgs. She has not developed speech and she never learned to walk. She has a severe ID. She likes to listen to music all day long and when she doesn't like a song she express her anger by throwing the phone on the floor. No menarche yet. Very small breasts (p1 stage). She has no more constipation. Lab exam showed normal renal function. She had a bilateral hip luxation for which she had been treated. She is well controlled for her seizures with Keppra, Lamictal, and Fycompa.

25.2 Her father is 196 cm tall. He has normal intelligence and no general health problems except for his insulin dependent diabetes diagnosed at 14 years.

KDM2B_29 12q24.31 deletion (including KDM2B & SETD1B)

Described in https://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-019-0749-3

KDM2B_30 12q24.31 deletion (including KDM2B & SETD1B)

Described in https://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-019-0749-3

KDM2B_31 p.(Cys616Tyr)

The case is a 4-year old boy presenting with a mild ID. Pre- and perinatal life were marked by a growth retardation (birth weight and length <3p). He walked independently at 14 months, but exhibited a mild developmental delay with language delay. Brain MRI showed a thin corpus callosum, delayed frontal myelination and a posterior fossa cyst. His clinical examination at 4 years showed microcephaly, dysmorphic features, stereotypic movements, and hyperpigmentation on the shoulder. Other clinical features include atrial septal defect, bilateral talus pes, unilateral cryptorchidism, kyphosis and congenital obstruction of ductus nasolacrimalis. Array-CGH was performed and did not show any abnormality. Trio WES identified a de novo missense variation in KDM2B (c.1847G>A, p.(Cys616Tyr)).

Likely pathogenic variants (sample not available for analysis)

KDM2B_13 p.(Lys635del)

This girl first presented to Clinical Genetics at 1 year of age. She was referred to us because of developmental issues and dysmorphism. Family history is unremarkable. She was born at term by C-section due to oligohydramnios with BW 3827 grams to non-consanguineous parents. Pregnancy was complicated by maternal hypertension. On day 1 of life, patient was noted to have poor suck and feeding, hemangiomas on abdomen and back, and brachycephaly. Postnatal course was complicated by jaundice, hyperbilirubinemia, and poor weight gain. She had a percutaneous gastrostomy tube placed at 1 month of life. At 6 months of age she began developing pubic hair without other secondary sex characteristics. At 4 years of age, she was noted to have two atrial septal defects and pulmonic stenosis by echocardiogram. During her genetics evaluation that year, she was noted to

have hypotonia, microcephaly, precocious breast development, and mild dysmorphism of synophrys, arched eyebrows, narrow bifrontal diameter, mild microstomia, pectus excavatum and left 5th finger contracture. Her motor development was mildly delayed and she started walking at 17 months of age but was noted to have an unsteady gait. She had marked speech delay with a vocabulary of about 3-4 words at age 4. Chromosome microarray and trio exome sequencing done. Microarray identified a maternally inherited 22q13.31->q13.31 duplication (568 Kb). Exome sequencing identified a de novo variant in KDM2B and maternally inherited variants in EZH2 and DISP1.

VUS (sample not available for analysis)

KDM2B_17 p.(Ala543Thr)

A 7-year-old female presented for history of developmental delay, autism, apraxia, and polymicrogyria. (Patient currently 9 yrs-old. MRI was not available for review.) Family history is unrevealing. She is an only child, born to mother of Irish descent and father of Indian descent.

She was born at 39 weeks gestation via vaginal delivery with vacuum assistance. Birth weight 2522g (-2.0SD), Length 48.26cm (-1.5SD). Nuchal cord was present. Seizures were reported during the newborn period. Apnea followed an attempt at feeding at 3-1/2 hours after birth. She remained in the hospital for 2 months. She was described with failure to thrive until 2-1/2 years of age due to feeding difficulties. Over time, she gained weight and then subsequently became overweight. At age 7, she measured +2.1 SD for height, +3.0 SD for weight, and +1.1 SD for head circumference. She has hypotonia and poor balance and coordination. She had frequent ear infections and has had myringotomy tubes placed, tonsillectomy and adenoidectomy.

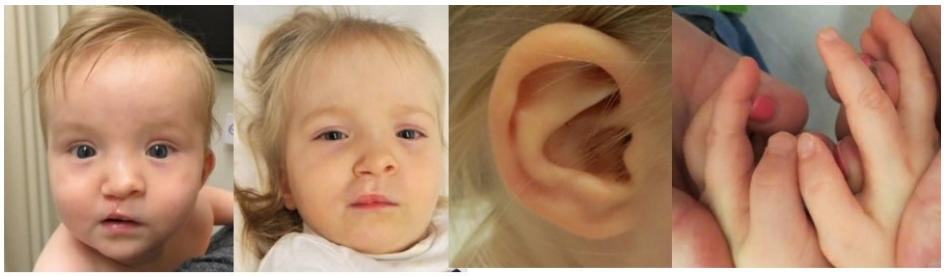
She rolled over at 12 months, sat at 18 months, crawled at 18 months, walked at 2 years. She has significant speech delays and is primarily nonverbal. She uses a few signs and is able to use an IPad. Receptive language skills are better than expressive. She was diagnosed with autism at age 6 years; ADOS score 21.

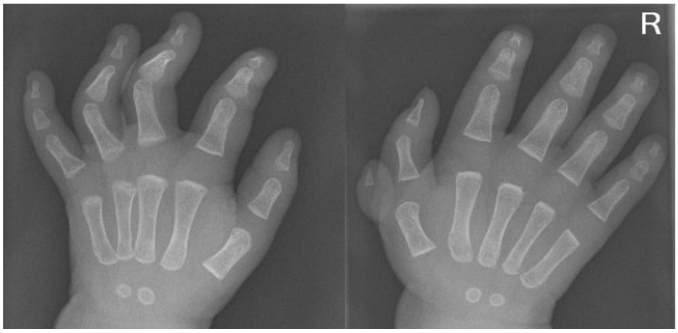
Diagnostic testing was completed on DNA from saliva swab. First, Cytoscan HD array revealed a 9p23 deletion (57kb) that was maternally inherited and felt to be unrelated. Whole-exome trio analysis was performed and identified a de novo c.1627G>A (p.Ala543Thr) variant of uncertain clinical significance in KDM2B (NM_032590.4). Two additional findings were reported: a heterozygous PEX26 (NM_017929.5) c.34dup (p.Leu12Profs*103) likely pathogenic variant (carrier status for peroxisome biogenesis disorder) and a heterozygous CNTNAP5 (NM_130773.3) c.1357G>A (p.Val453Ille) variant of uncertain clinical significance. The patient's clinical course has not been felt to follow that of a typical peroxisomal disorder and therefore the PEX26 was not felt to provide support of an explanation for her features. In addition, the CNTNAP5 variant was detected in father, so also less likely to provide a clear cause for her symptoms.

VUS (Variants not showing KDM2B specific episignature)

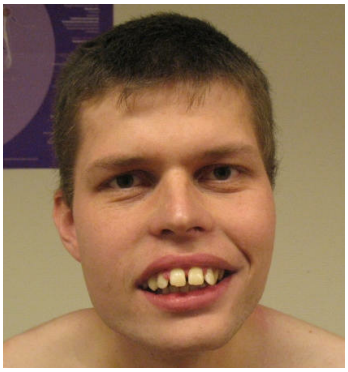
KDM2B_2 p.(Arg766Gln)

This girl was born with multiple congenital abnormalities, cleft lip/palate, congenital contracture PIP 3rd and 4th digit of the left hand, pre-axial polydactyly of the right hand with hypoplastic thumb grade II-III. At age 2, her psychomotor development is normal.



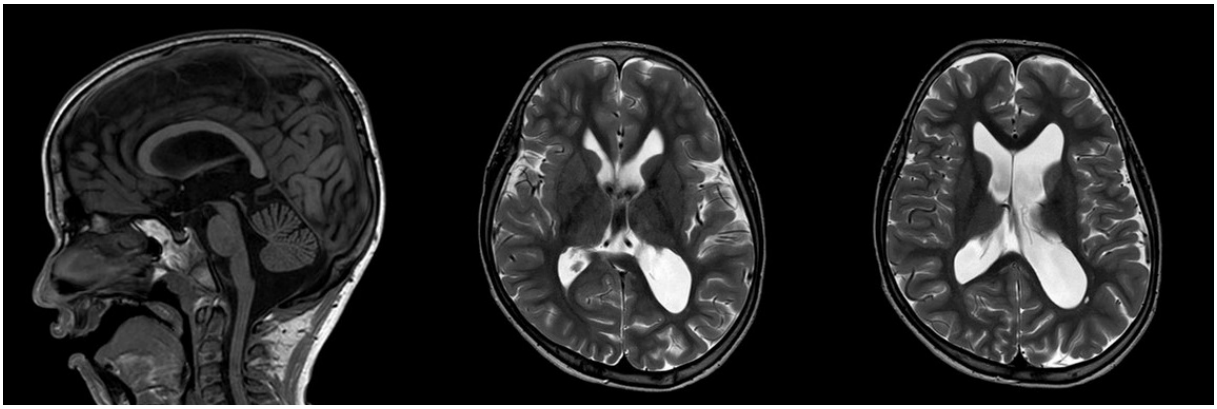


KDM2B_5 p.(Ile652Val)



This is an adult male with mild-moderate ID and autism. The variant was inherited from his normally developed mother.

KDM2B_14 p.(Arg1213Trp)



KDM2B_19 c.777+5G>A

The patient was born from the 3th pregnancy of not consanguineous parents. First pregnancy ended with a spontaneous miscarriage; an healthy male was born from the second pregnancy. The patient has an heterozygous, healthy, twin sister. No cases of genetic syndromes or neurodevelopmental disorders are known in her family. Cesarean delivery was performed after 36 weeks. Weight was 2780 g (+0.45 SD), length 47 cm (+0.1 SD), head circumference 32 cm (-0.35 SD). After 3 days of life the child underwent seizures, treated with Levetiracetam. In the sequent days she suffered of severe anemia, thrombotic angiopathy and intraventricular hemorrhage. Several blood clots were observed by imaging, the most evident one in the inferior vena cava. She received a diagnosis of atypical hemolytic uremic syndrome. Auditory Brainstem Response showed a severe bilateral neurosensorial hypoacusis. US showed an atrial septal defect, not requesting surgery. Kidneys are normal.

At 10 months she suffered a Pneumomediastinum. At the age of 12 months her weight was 6000 g (-3 DS), the length 65 cm (-3 DS), the head circumference 41,5 cm (-4 DS), she showed epicanthus, low-set ears, thin eyebrows, mild micrognathia. She has a severe developmental delay. She is not able to sit. Verbal language is absent. The variant was detected by trio/WES.

Episignature methods

Methylation array and quality control

DNA methylation analysis and Episignature classifier development was performed using previously established protocol (Aref-Eshghi et al., 2019; Aref-Eshghi, Bend, et al., 2018; Aref-Eshghi, Rodenhiser, et al., 2018; Bend et al., 2019). Stored genomic DNA samples extracted from peripheral blood, previously used for genomic sequencing, were used for bisulfite conversion and hybridization to the Illumina infinium methylation EPIC bead chip arrays, according to manufacturer's protocol. Idat files, containing methylated and unmethylated signal intensity plots (beta values) were produced from these microarrays, and used for analysis in R 4.0.2. Normalization was performed using the Illumina Infinium methylation EPIC array with background correction from the minfi package (Aryee et al., 2014). Previously defined exclusion criteria (Aref-Eshghi, Rodenhiser, et al., 2018; Bend et al., 2019) were used to exclude probes with detection p values >0.01, probes on the x and y chromosomes, probes known to contain SNPs at the site of CpG interrogation or single nucleotide extension, and probes known to cross react with chromosomal locations other than their target regions were removed. All samples were examined for genome wide methylation distribution and those deviating from a bimodal distribution were excluded. Factor analysis using a principal component analysis (PCA) was performed to examine batch effect and identify outliers.

DNA methylation profiling

Probe methylation levels (beta values), were calculated as the ratio of signals intensity in methylated probes vs total sum of unmethylated and methylated probes, resulting in values ranging from zero to one. To allow for linear regression modeling, beta values were logit transformed using the limma package (Ritchie et al., 2015), allowing for the identification of

differentially methylated probes. Data was adjusted for the blood cell type composition as per Houseman et al (Houseman et al., 2012). Estimated blood cell proportion was added to the model matrix of the linear models as confounding variables (Reinius et al., 2012). Using the eBayes function in the limma package (Smyth, 2004), p values were moderated and corrected for multiple testing using the Benjamini Hochberg method. Probes with the most significant methylation differences are selected using two facts from this dataset, the level of methylation difference (relative methylation signal intensity), and the probability that an observed difference is due to random chance (p values). Evaluation of this interaction is carried out by multiplying the absolute methylation difference between affected cases and controls by the negative value of the log transformed p values, and ranking the top 1000 probes with the highest values from this transformation. Next, receiver operating characteristic analysis (ROC) is performed on each probe, to measure the pairwise correlation coefficient between probes. Probes with low area under curve values from ROC analysis are removed, as well as highly correlated probes, eliminating probes with low sensitivity and specificity, and probes with highly correlated characteristics using Pearson's correlation coefficient. This ensures that the final probeset contains the most differentiating, non-redundant probes that are not influenced by random data structures. Only probes with a methylation difference greater than 5% were included in this analysis. This probe filtering process was designed to avoid reporting of probes with low effect size, and those influenced by technical or random variations as conducted in previous studies (Aref-Eshghi et al., 2019; Aref-Eshghi, Rodenhiser, et al., 2018).

Selection of matched controls for methylation profiling

For episignature characterization, mapping of probes and feature selection, matched controls were randomly selected from the LHSC EpiSign Knowledge Database (EKD) (Aref-Eshghi,

Rodenhiser, et al., 2018). All of the KDM2B samples were assayed, therefore all the controls selected for episignature identification were analyzed using the same array type. Samples were matched by age, sex and batch using the MatchIt package. A 4:1 ratio of controls to cases was deemed optimal for this analysis, as previously described (Aref-Eshghi et al., 2019). PCA analysis was performed after each attempt at matching to detect outliers and determine data structures for the presence of batch effect. Outlier samples, and those with highly aberrant data structures were removed, and subsequent matching trials were performed until consistent iterations with no outliers in the first two components of the PCA were derived.

Clustering and dimension reduction

Hierarchical clustering and multidimensional scaling were used after each iteration of analysis to examine the data structure of the identified episignature. Hierarchical clustering was performed using Ward's method on Euclidean distance by the ggplot package (Ward, 1963; Wickham, 2009). Multi-dimensional scaling provides a visual representation of sample methylation profile similarity based on the scaling of the pairwise Euclidean distances between each sample. Observations of study samples' methylation profiles at this stage allowed for further refinement of the cohort used for probe selection training.

Discovery/training cohort selection

Identification of disease specific episignatures was performed using a randomly selected sub-setting of the database, on a 75:25 ratio of discovery:training, using the caTools package in R. Testing samples were used to assess the performance of the classification model developed later in the study. For every disease group in the discovery cohort, a sex and age matched control group with a sample size at least 4 times larger was selected from the reference

control group using the MatchIT package, and methylation profiles were compared between the two.

Cross validation

For each round of validation, one of the selected KDM2B samples was removed from probe selection, alongside matched controls. The remaining KDM2B samples were designated as testing samples, and all three groups were modeled using multidimensional scaling to determine how they cluster/segregate with one another. This process was repeated with different combinations of assigned training and testing samples until all cases had been removed from probe selection and used for testing once.

Episignature classification model

Specificity of the episignature was assessed using the Methylation Variant Pathogenicity (MVP) score, using all the identified probes. A support vector machine (SVM) used a linear kernel for training on KDM2B cases and controls. Once again, a 4:1 ratio of controls to cases was used to divide both the case and control samples previously matched and used for probe selection into training and testing cohorts for the SVM. Furthermore, the remaining unselected samples from the EKD were also divided similarly (75% training, 25% testing) to allow for comparison and testing of signature robustness against all of the samples in the EKD. Using the e1071 R package, we performed 10-fold cross validation to determine hyperparameters optimal for episignature classification. In this process, the training set was divided into ten folds by random assignment, where the first nine are used for training, and the last used for testing the accuracy of the model. The mean accuracy over all rounds, was then calculated, and hyperparameters with the best performance by this metric were selected. The model provides a score ranging from 0-1 for each subject, representing the model's confidence in predicting whether the subject has a DNA methylation profile similar

to the KDM2B probe set or not. Conversion of these SVM decision values was done using Platt's scaling method (AJ & PJ, 2000), and the class obtaining the greatest score determined the predicted phenotype. A classification as KDM2B was made when a sample received the greatest score for that class (normally greater than 0.5). Finally, the model was applied to both a training set of a large cohort of individuals with clinical and molecular diagnoses of neurodevelopmental disorders, as well as a group of healthy controls to determine its effective specificity.

Validation of Episign classification

To ensure the model is not susceptible to the batch structure of the methylation experiment, the classifier was applied to samples assayed on the same batch as the cases used for training. Using methylation data from individuals without a confirmed diagnosis of KDM2B within the EKD assayed on the same microarray chip as case samples, methylation profiles were modeled to ensure the classifier is not confounded by technical artifacts unique to the given microarray. Specificity was determined by supplying a large number of DNA methylation arrays from unaffected subjects to the model. To further assess the specificity of the KDM2B classifier relative to other neurodevelopmental disorder we applied it to cases with other patient cohorts exhibiting distinct episignatures within the EKD.

References

AJ, S., & PJ, B. (2000). *Advances in Large-Margin Classifiers*: MIT Press.

Aref-Eshghi, E., Bend, E. G., Colaiacovo, S., Caudle, M., Chakrabarti, R., Napier, M., . . . Sadikovic, B. (2019). Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. *Am J Hum Genet, 104*(4), 685-700. doi:10.1016/j.ajhg.2019.03.008

Aref-Eshghi, E., Bend, E. G., Hood, R. L., Schenkel, L. C., Carere, D. A., Chakrabarti, R., . . . Sadikovic, B. (2018). BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin-Siris and Nicolaides-Baraitser syndromes. *Nat Commun, 9*(1), 4885. doi:10.1038/s41467-018-07193-y

Aref-Eshghi, E., Rodenhiser, D. I., Schenkel, L. C., Lin, H., Skinner, C., Ainsworth, P., . . . Sadikovic, B. (2018). Genomic DNA Methylation Signatures Enable Concurrent

Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental Syndromes. *Am J Hum Genet, 102*(1), 156-174. doi:10.1016/j.ajhg.2017.12.008

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics, 30*(10), 1363-1369. doi:10.1093/bioinformatics/btu049

Bend, E. G., Aref-Eshghi, E., Everman, D. B., Rogers, R. C., Cathey, S. S., Prijoles, E. J., . . . Sadikovic, B. (2019). Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome. *Clin Epigenetics, 11*(1), 64. doi:10.1186/s13148-019-0658-5

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., . . . Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics, 13*, 86. doi:10.1186/1471-2105-13-86

Reinius, L. E., Acevedo, N., Joerink, M., Pershagen, G., Dahlen, S. E., Greco, D., . . . Kere, J. (2012). Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One, 7*(7), e41361. doi:10.1371/journal.pone.0041361

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res, 43*(7), e47. doi:10.1093/nar/gkv007

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol, 3*, Article3. doi:10.2202/1544-6115.1027

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association, 58*(301), 236-244. doi:10.1080/01621459.1963.10500845

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.