

Supplemental information

SRSF1 haploinsufficiency is responsible

for a syndromic developmental disorder

associated with intellectual disability

Elke Bogaert, Aurore Garde, Thierry Gautier, Kathleen Rooney, Yannis Duffourd, Pontus LeBlanc, Emma van Reempts, Frederic Tran Mau-Them, Ingrid M. Wentzensen, Kit Sing Au, Kate Richardson, Hope Northrup, Vincent Gatinois, David Geneviève, Raymond J. Louie, Michael J. Lyons, Lone Valentin Laulund, Charlotte Brasch-Andersen, Trine Maxel Juul, Fatima El It, Nathalie Marle, Patrick Callier, Raissa Relator, Sadegheh Haghshenas, Haley McConkey, Jennifer Kerkhof, Claudia Cesario, Antonio Novelli, Nicola Brunetti-Pierri, Michele Pinelli, Perrine Pennamen, Sophie Naudion, Marine Legendre, Cécile Courdier, Aurelien Trimouille, Martine Doco Fenzy, Lynn Pais, Alison Yeung, Kimberly Nugent, Elizabeth R. Roeder, Tadahiro Mitani, Jennifer E. Posey, Daniel Calame, Hagith Yonath, Jill A. Rosenfeld, Luciana Musante, Flavio Faletra, Francesca Montanari, Giovanna Sartor, Alessandra Vancini, Marco Seri, Claude Besmond, Karine Poirier, Laurence Hubert, Dimitri Hemelsoet, Arnold Munnich, James R. Lupski, Christophe Philippe, Christel Thauvin-Robinet, Laurence Faivre, Bekim Sadikovic, Jérôme Govin, Bart Dermaut, and Antonio Vitobello

Supplemental Note

Case reports

The antenatal period was uneventful for 11/17 individuals. Pregnancy was marked by the detection of intrauterine growth restriction (IUGR) with normal prenatal karyotype in individual 1 (I1 - F1-II-1), poor active fetal movements in individual 2 (I2 - F2-II-2), pre-eclampsia in individual 14 (F13-II-3), premature rupture of membranes in individual 7 (F6-II-1), a positive non-invasive prenatal testing for trisomy 21 leading to amniocentesis with normal karyotype in individual 15 (46,XY, F14-II-1) and an increased nuchal translucency observed at the first trimester ultrasound in individual 17 (F16-II-1) (**Table 3, Table S1**). Neonatal growth parameters were in the normal range except for individual 1 (F1-II-1) and individual 4 (F3-II-2) who had presented IUGR during pregnancy. Neonatal complications were observed in 10/15 individuals: individual 1 (F1-II-1) required oxygen therapy for 3 weeks because of pulmonary hypoplasia; individual 3 (F2-II-3) presented congenital torticollis; individual 14 (F13-II-3) had neonatal hypotonia and sucking difficulties; individual 6 (F5-II-1) had feeding difficulties and jaundice, also present in individual 13 (F12-II-1) and individual 16 (F15-II-3); individual 11 (F10-II-1) had transient tachypnea and increased muscle tone; individual 4 (F3-II-2) was monitored for intrauterine growth restriction and had neonatal hypocalcemia; individual 17 (F16-II-1) had transient tachypnea. Congenital heart defect, including ventricular septal defect (VSD), right aortic arch, and aberrant subclavian artery were detected in individual 9 (F8-II-1). The majority of individuals presented growth parameters in the normal range. However, individuals 8 (F7-II-1), 10 (F9-II-3) and 17 (F16-II-1) had short stature with low body weight for individuals 8 and 10. Individual 8 (F7-II-1) also had postnatal microcephaly with OFC at -2.9 SD at the last evaluation at the age of 3.1 years. On the contrary, individual 4 (F3-II-2) had postnatal macrocephaly with OFC at +2.7 SD at the last evaluation at the age of 34 years. Failure to thrive was reported for 6/16 individuals. All of them presented ID or DD ranging from mild to severe. Individual 4 (F3-II-2) had learning disabilities with an intellectual quotient (IQ) of 76. Motor delay was observed in 13/17 individuals and speech delay in 16/17. Hypotonia was reported in ten individuals and oral motor hypotonia was observed in one individual. Behavioral disorders were described in 13/17 individuals, including outbursts of anger (8/13), social difficulties (5/13), stereotypies (4/13), hyperactivity (2/13), hetero-aggressive behavior (1/13), and anxiety (1/13). Three individuals had seizures. Among them, two individuals had a single episode of febrile seizure (I14 - F3-II-2; I10 - F9-II-3) and one individual had recurrent episodes of fixed and deviant gaze (I16 - F15-II-3). I7 (F6-II-1) had some EEG anomalies while falling asleep without clinical signs of seizures.

Cerebral MRI was performed in 13 individuals. Among them, 6 were abnormal, including deficient or delayed myelination in three individuals. Additional MRI anomalies were observed in 6 individuals (**Table S1**, **Table 3**). No hearing loss was observed in the cohort. Vision problems were present in 8/15 individuals, including hypermetropia (2/8), myopia (4/8), astigmatism (3/8), strabismus (3/8) and nystagmus (1/8). Skeletal anomalies were observed in ten individuals: four had pectus deformity, pectus excavatum in 2 individuals and pectus carinatum in 2 individuals; five had scoliosis and individual 1 (I1-II-1) needed scoliosis brace and arthrodesis at the age of 13 years. Among the five adults, three of them were referred for marfanoid features with dolichostenomelia (I14 - F13-II-3), arachnodactyly, pectus deformity, scoliosis (I4 - F3-II-2; I14 - F13-II-3; I15 - F14-II-1), and highly arched palate (I4 - F3-II-2; I14 - F13-II-3). One individual with marfanoid features was part of PHRC national 2008-A00515-50, funded by the French Ministry of Health.¹ Congenital malformations were observed in nine individuals. Five individuals had urogenital malformations including cryptorchidism (3/5), hypospadias (2/5), hydrocele (1/5). Nonspecific facial features were observed in the cohort including downslanted palpebral fissures, low-set ears, smooth philtrum, thin upper lip and high arched palate were reported (**Figure 1C**, **Table S1**).

Supplemental Figures and legends

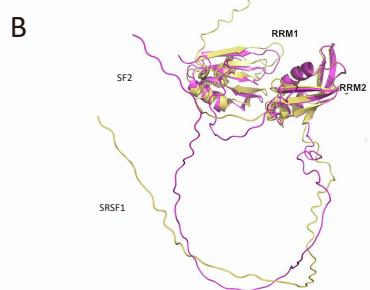
A



B

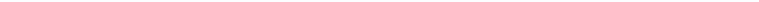


Figure S1. 17q22 deletions observed in individuals 6 and individual 17. (A) All genes present in the chromosomal deletion 17q22 (558065434-5654497)x1, and (B) (55442363-56309063)x1 are listed in the box.



C

Legend: SRSF1 (Yellow), SF2 (Red)



Events

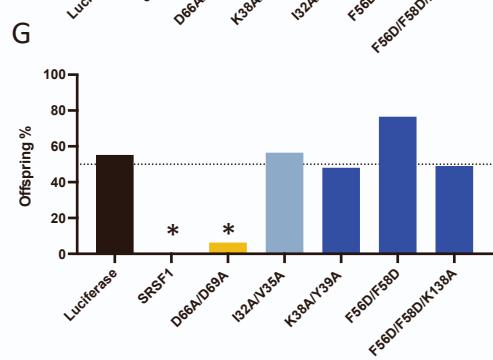
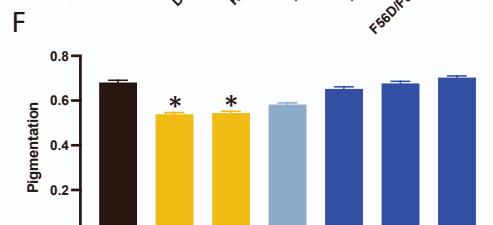
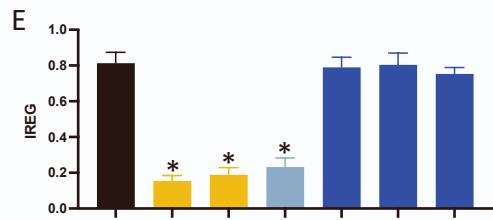
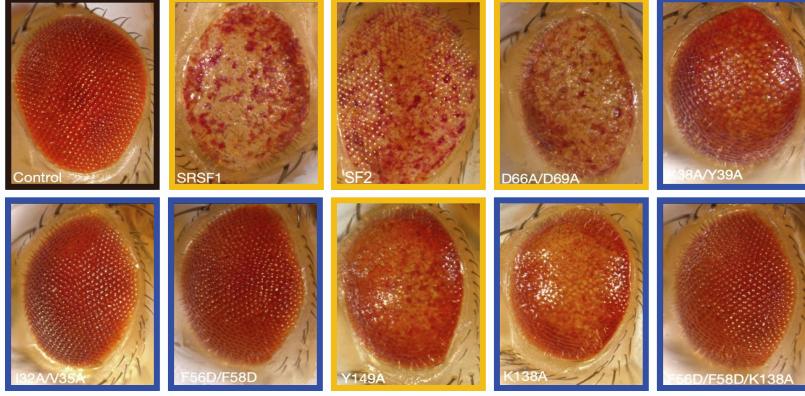
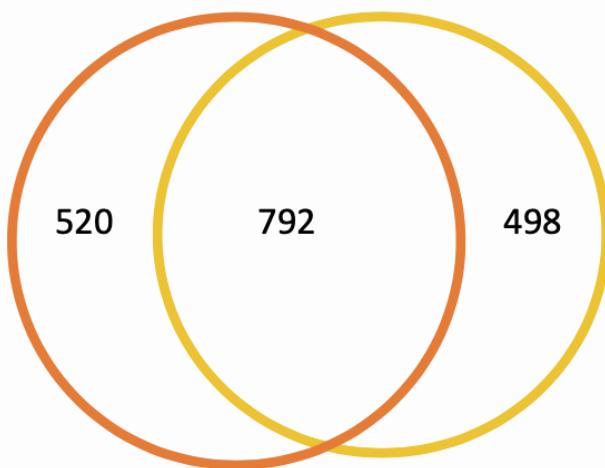


Figure S2: Functional splicing read-outs of SRSF1 in the eye and nervous system of *Drosophila*. (A) Schematic presentation of the sequence conservation of humans SRSF1 and SF2, the *Drosophila* orthologue. The sequence of the RRM1 and RRM2 domains is 81% and 76% identical respectively, the RS domain is less conserved with a sequence identity of 45%. **(B)** Structural conservation of SRSF1 and SF2 protein. **(C)** RNA splicing analysis indicates functional conservation of SRSF1 and SF2. Five different alternative splicing events were measured compared to luciferase control flies: Skipped Exon (SE), Retained Intron (RI), Alternative 5' Splice Site (A5SS), and Alternative 3' Splice Site. **(D)** Representative pictures of flies expressing Luciferase, SRSF1, SF2, D66A/D69A, K38A/Y39A, I32A/V35A, F56D/F58D, Y149A, K138A, F56D/F58D/K138A. Amino acid substitutions corresponding to previously reported biochemically characterized SRSF1 are abbreviated using one letter codes to distinguish them from clinical variants (in three letter codes). **(E)** The irregularity score or regularity index of flies expressing luciferase, and different splicing controls of RRM1 are quantified. ($n > 10$ and * indicates p -value < 0.01 , data are represented as mean \pm SEM). **(F)** Pigmentation score measuring the depigmentation in the flies expressing luciferase, SRSF1 and SRSF1 splicing controls. ($n > 10$ and * indicates p -value < 0.01 , data are represented as mean \pm SEM). **(G)** Offspring frequencies were measured in flies expressing Luciferase, SRSF1, and splicing controls ($n > 10$ and * indicates p -value < 0.01).

A



B

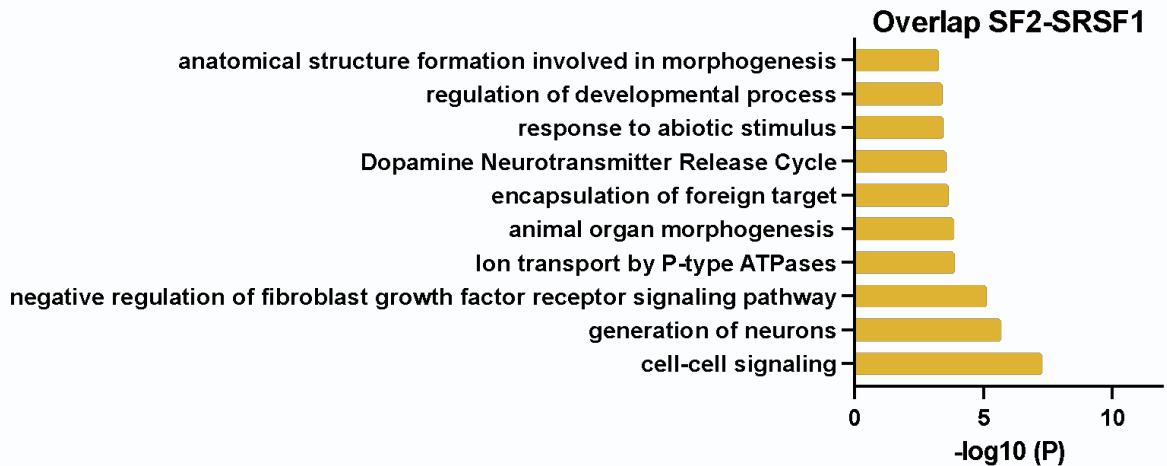
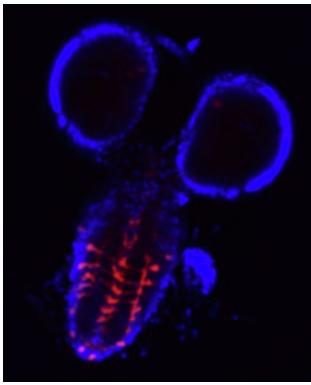


Figure S3: Splicing role of SRSF1 in *Drosophila*. **(A)** Venn diagram of RNA-splicing analysis depicted in **Figure S2C**, representing the alternatively spliced genes due to overexpression of SRSF1 (yellow) or SF2 (orange). The numbers indicate the total number of alternatively spliced genes. **(B)** Metascape analysis for gene ontology on biological processes on the overlapping genes that were alternatively spliced by SF2 and SRSF1 overexpression.

A



B

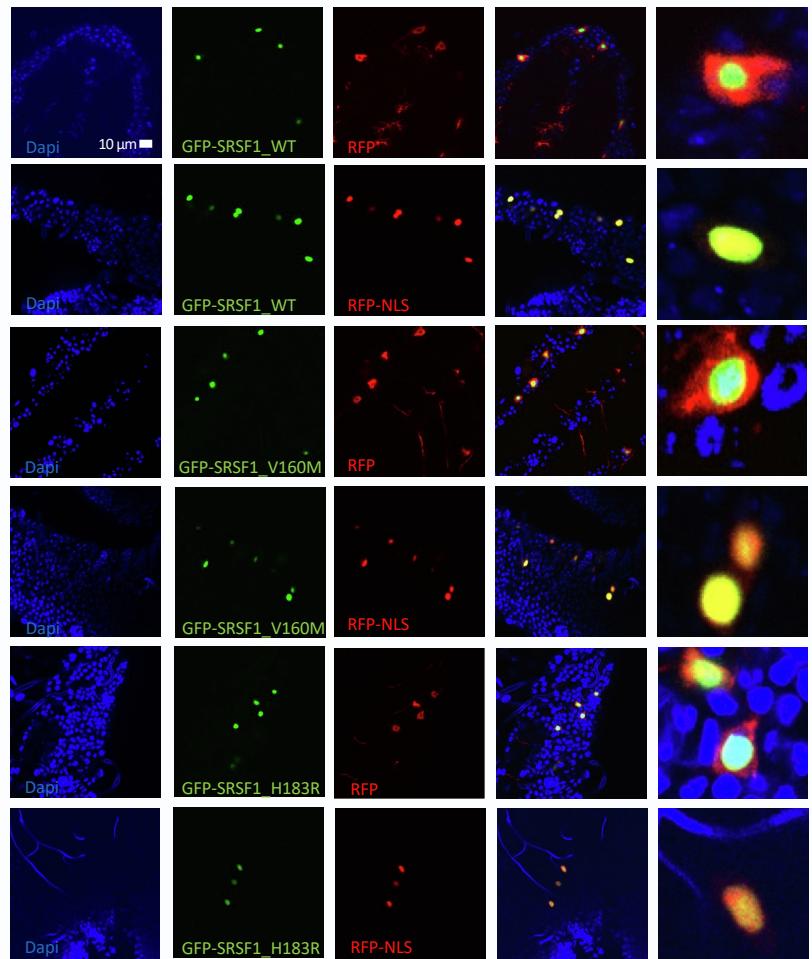


Figure S4. Subcellular localization of SRSF1 and V160M and H183R. **(A)** Confocal image of a ventral nerve cord of third instar larvae expressing RFP in CCAP neurons. **(B)** Confocal image of CCAP-neurons in the ventral nerve cord of flies expressing RFP to mark the CCAP-neuron or RFP-NLS to mark the nucleus of CCAP-neurons. GFP-tagged SRSF1 protein variants were imaged in green. In the last lane, a zoomed image of representative CCAP-neurons is shown.

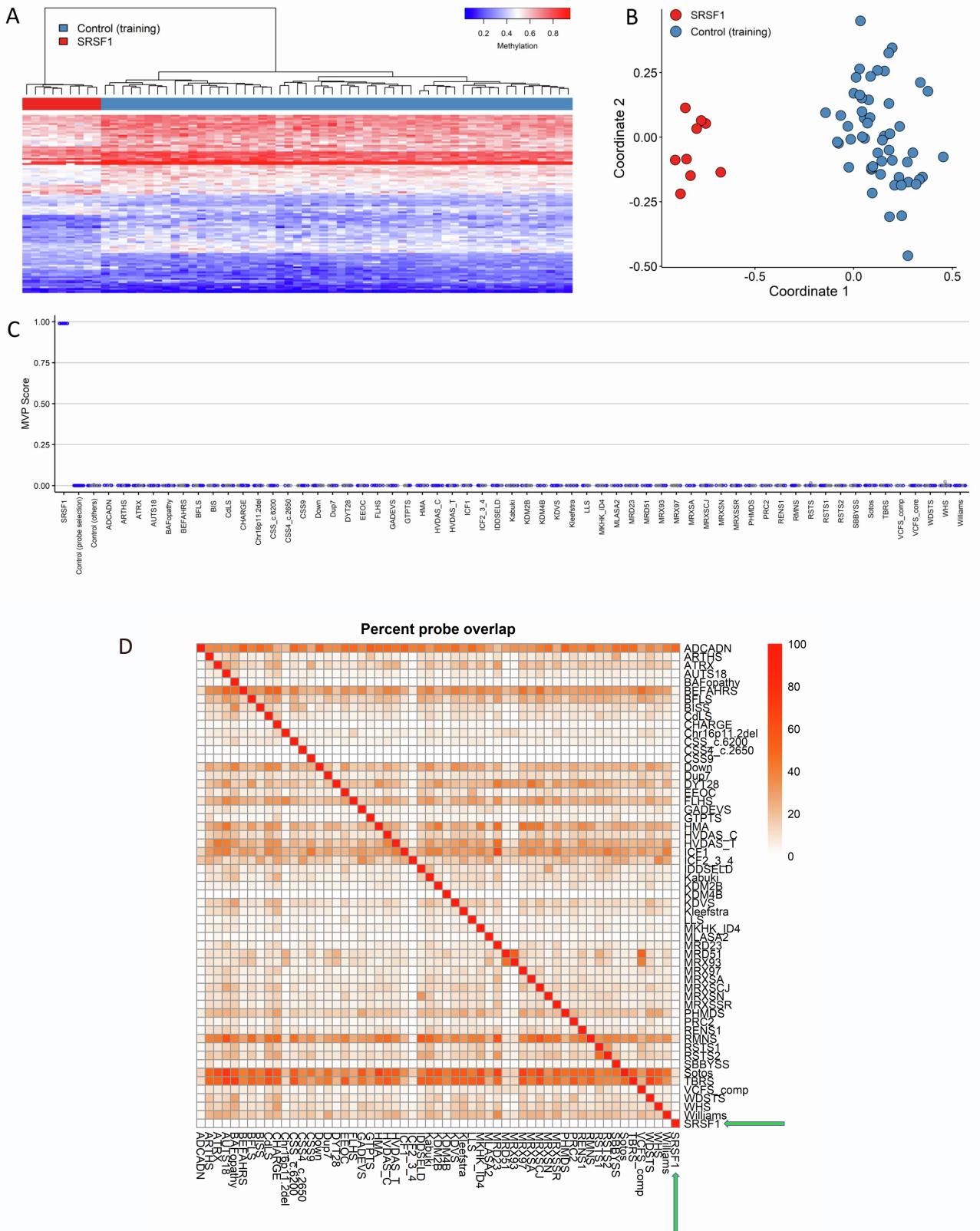


Figure S5. SRSF1 episignature. **(A)** Heatmap indicates clear separation of the 9 SRSF1 samples (red) from age and sex-matched controls (blue). Each row represents one of the 107 probes selected as the episignature and each column represents either an SRSF1 individual or a control. **(B)** Multidimensional scaling plot (MDS) shows clustering of the SRSF1 samples together and away from controls. **(C)** Support vector machine classifier model (SVM). The model is trained using the 107 selected SRSF1 episignature probes and 75% of controls and other neurodevelopmental disorder samples on EpiSign™ (blue circles). The 25% remaining are used as testing samples (grey circles). The SRSF1 samples have a probability score (methylation variant pathogenicity score, MVP) of close to 1 compared with all other samples. The SVM shows the specificity of the SRSF1 episignature in classifying the SRSF1 samples. **(D)** Heatmap of differentially methylated probes (DMPs) shared between the SRSF1 cohort and 56 other EpiSign™ disorders with known genome-wide methylation profiles; The heatmap shows the percentage of probes shared between each paired cohort. Colours indicate the percent of the y-axis cohort's DMPs that are also found in the x-axis cohort's DMPs. The SRSF1 cohort is highlighted by the green arrows. SRSF1 DMPs show the largest overlap with ADCADN (~49%) and the next most overlapping are several cohorts with ~9-11% overlap (Hunter McAlpine syndrome, HMA; Tatton-Brown Rahman syndrome, TBRS; Sotos syndrome, Sotos; and Rahman syndrome, RMNS).

A

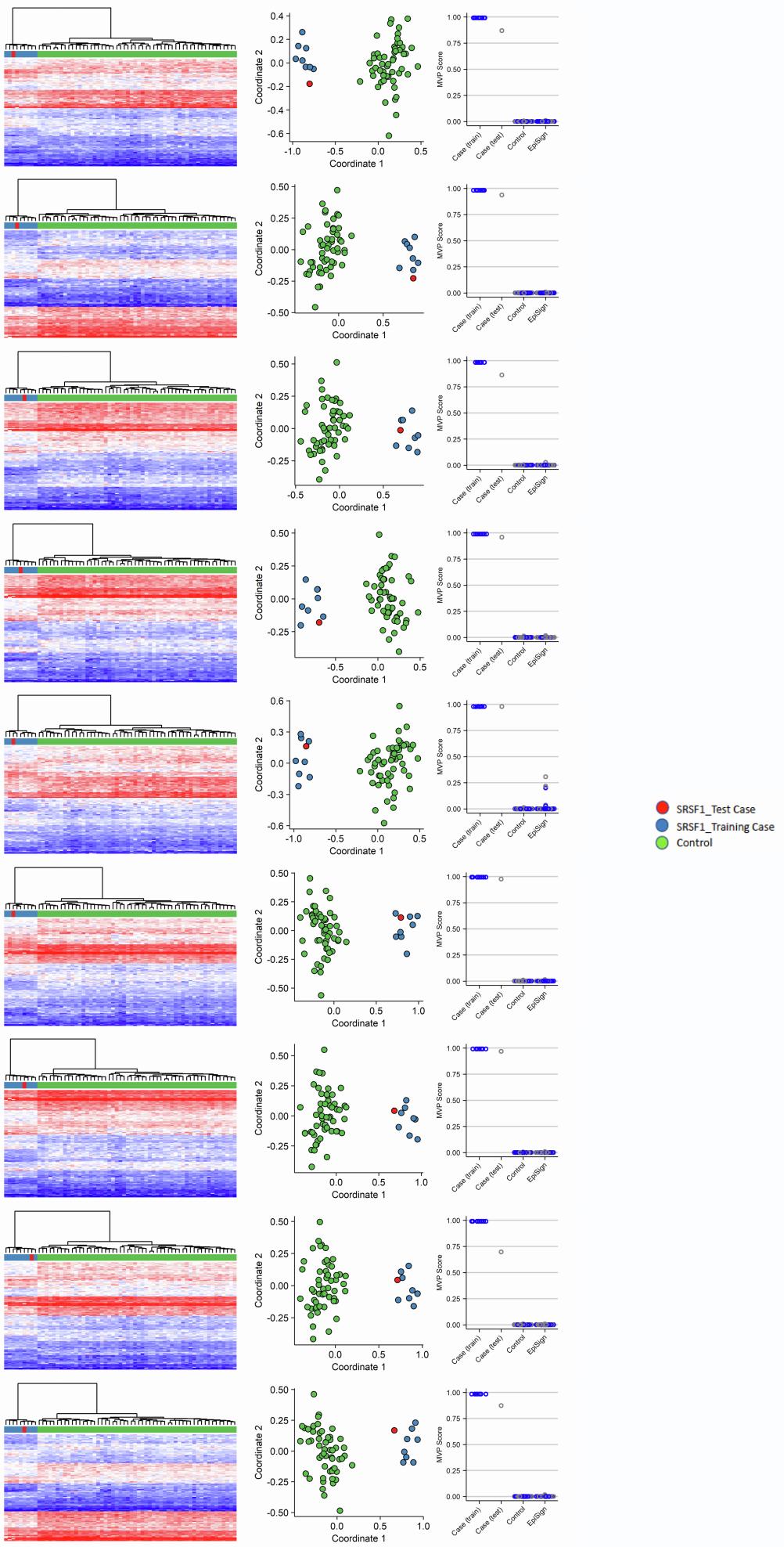


Figure S6. DNA methylation data leave-1-out cross-validation results. In each round a single *SRSF1* sample with confirmed pathogenic variant was used for testing (red) and plotted with the remaining eight *SRSF1* samples (those used for episignature training) (blue) and controls (green). The heatmap and MDS clustering methods confirmed the robustness and sensitivity of the *SRSF1* episignature and the high MVP scores for training and testing cases in the SVM confirm the specificity.

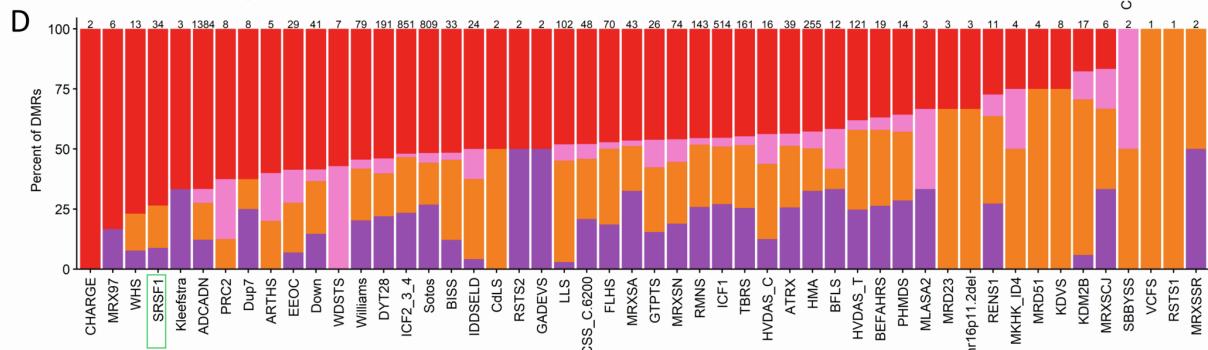
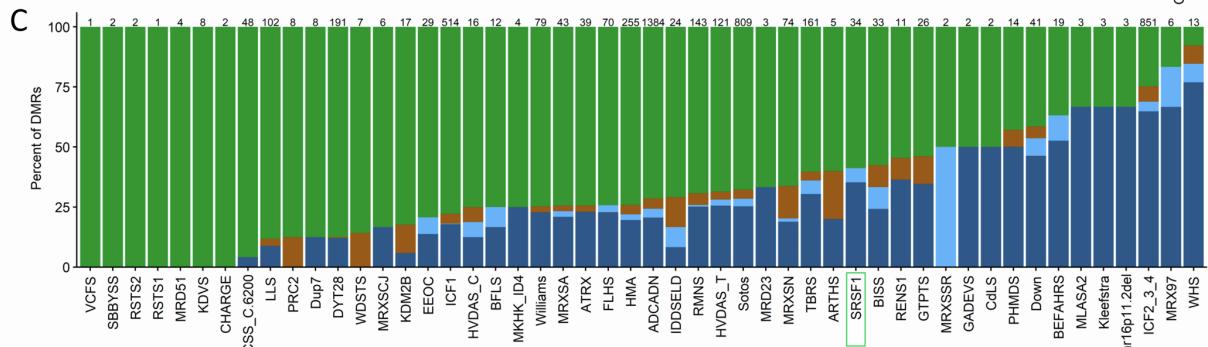
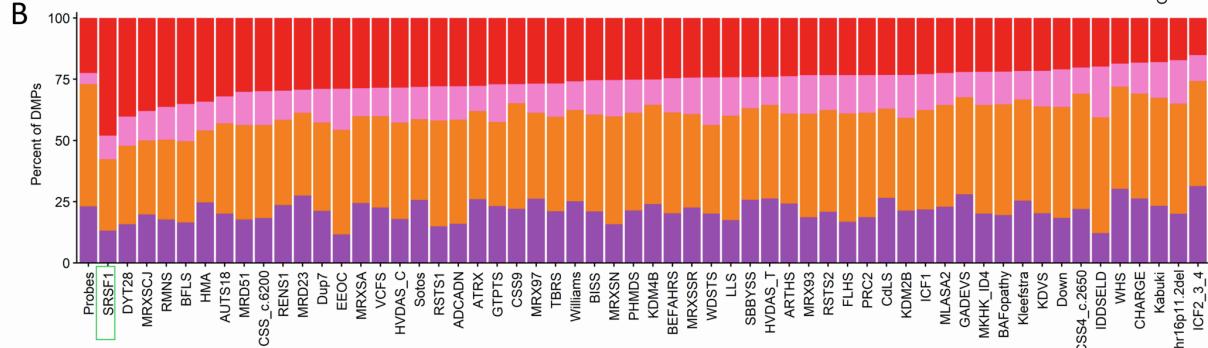
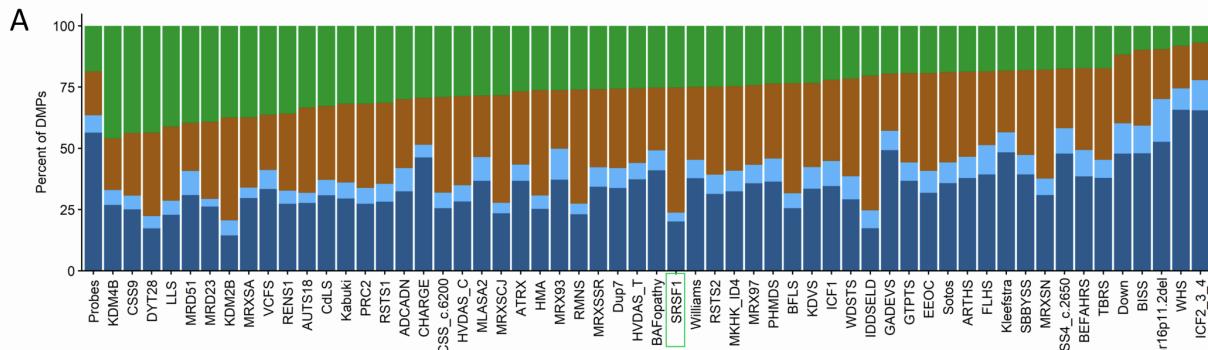


Figure S7. Differentially methylated probes (DMPs) and regions (DMRs) annotated in the context of CpG islands and genes. **(A)** DMPs annotated in relation to CpG islands. Island, CpG island; Shore, within 0-2kb of a CpG island; Shelf, within 2-4kb of a CpG island boundary; Inter_CGI, all other regions in the genome. **(B)** 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. The probes column in A and B represent the background distribution of probes. **(C)** DMRs annotated in relation to CpG islands. **(D)** DMRs annotated in the context of genes. For DMR analysis, the numbers above indicate the number of DMRs identified in each cohort. The *SRSF1* cohort is outlined in green boxes in A-D.

A

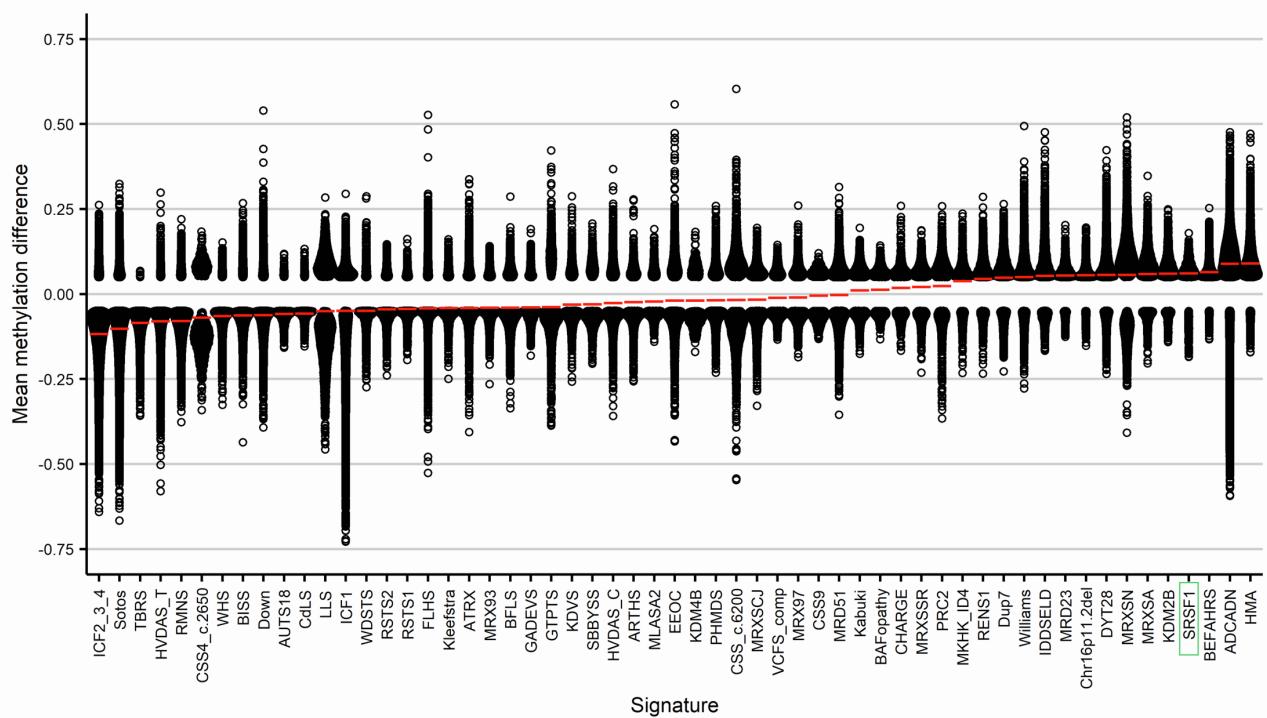


Figure S8. Overall methylation trend of each cohort. Methylation differences of all differentially methylated probes (DMPs) for each cohort, sorted by mean methylation. Each circle represents one probe. Red lines indicate the mean methylation values. SRSF1 cohort is outlined in the green box.

Supplemental tables

Table S1 Clinical features of the SRSF1 cohort.

Subject	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5	Individual 6	Individual 7	Individual 8	Individual 9	Individual 10	Individual 11	Individual 12	Individual 13	Individual 14	Individual 15	Individual 16	Individual 17	
Sex	Female	Female	Male	Male	Female	Female	Male	Female	Male	Male	Female	Female	Female	Male	Male	Female	Female	
Molecular finding																		
SRSF1 Genomic coordinates (hg19)	NC_000017.10:g_56083708_56083236C>T 3709del	NC_000017.10:g_56083236C>T	NC_000017.10:g_56083237C>T	NC_000017.10:g_56083236C>T	NC_000017.10:g_56082935dup	NC_000017.10:g_(55806534_5654 0597)del	NC_000017.10:g_56083166T>C	NC_000017.10:g_56084417G>A	NC_000017.10:g_56083166T>C	NC_000017.10:g_56084380C>A	NC_000017.10:g_56084402C>A	NC_000017.10:g_56082914del	NC_000017.10:g_56083875C>T	NC_000017.10:g_56084428G>A	NC_000017.10:g_56083852A>C	NC_000017.10:g_6083832A>C	NC_000017.10:g_56084369C>T	NC_000017.10:g_(55442363-56309063)del
Transcript (NM_006924.5)	c.377_378del	c.478G>A	c.478G>A	c.478G>A	c.579dup	-	c.82C>T	c.548A>G	c.119G>T	c.97G>T	c.601del	c.208G>A	c.71C>T	c.231T>G	c.251T>G	c.130G>A	-	
Protein	p.(Ser126Trpfs*17)	p.(Val160Met)	p.(Val160Met)	p.(Val160Met)	p.(Val194Serfs*2)	-	p.(Arg28*)	p.(His183Arg)	p.(Gly40Val)	p.(Glu33*)	p.(Ser201Valfs*87)	p.(Ala70Thr)	p.(Pro24Leu)	p.(Tyr77*)	p.(Leu84Arg)	p.(Asp44Asn)	-	
Mode of inheritance	Undetermined	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>	<i>De novo</i>		
Pregnancy / Birth																		
Pregnancy complications	IUGR	Poor active fetal movements	C-section	No	Ovarian stimulation C-section for breech positioning	No	Premature rupture of membranes at 31w+2d	No	C-section for fetal distress	No follow up	No	No	No	Toxemia C-section	Screening test positive: amniocentesis with karyotype 46,XY	No	Increased nuchal translucency	
Gestational weeks (weeks + days)	36	40	38+4	Full term	39	37+1	31+2	38	Full term	ND	Full Term	41+2	39+2	41	39	38+6	ND	
Birth weight g (SD)	1700 (-2.6)	3300 (-0.5)	3180 (M)	2920 (-1.5)	3100 (-0.5)	2550 (-0.6)	1765 (+0.6)	2790 (-0.6)	3827 (+0.5)	ND	3430 (0)	3700 (+0.6)	3997 (+1.7)	ND	2980 (-1)	3065 (+0.1)	ND	
Birth length cm (SD)	39 (-3.6)	49.5 (-0.4)	51 (+1)	45.5 (-2.8)	51 (0.6)	46 (-0.7)	40.5 (-0.2)	45 (-1.8)	53.5 (+1.3)	ND	50 (0)	34 (-0.5)	50.8 (+0.9)	ND	50 (0)	51 (+1.3)	ND	
Birth OFC cm (SD)	30.5 (-1.5)	35 (+0.4)	35 (+0.4)	35 (-0.1)	34 (-0.3)	32 (-1)	30 (+0.4)	33.5 (-0.2)	35 (-0.2)	ND	34 (-0.1)	34 (-0.5)	ND	ND	35 (+0.2)	34 (+0.2)	ND	
Neonatal complications	Oxygen requirement because of pulmonary hypoplasia	No	Congenital torticoli	RCIU, hypocalcemia	No	Feeding difficulties Icterus	No	No	Congenital heart defect	No	Transient tachypnea Increased muscle tone	No	Icterus	Neonatal hypotonia Sucking difficulties	No	Icterus	Transient tachypnea of the newborn	
Growth																		
Age at last examination (Year + month)	18 years	4 years and 9 months	2 years	34 years	6 years and 8 months	4 years	5 years and 2 months	3 years and 1 month	8 years and 5 months	23 years and 9 months	5 years	3 years and 8 months	2 years and 2 months	28 years	18 years	14 months	13 years and 6 months	
Height at last visit cm (SD)	152 (-1.8)	103 (-0.5)	86 (-0.4)	189.5 (+2)	119.5 (+0.1)	98.3 (-0.6)	112.8 (+0.7)	84 (-2.7)	128 (+0.1)	156 (-2.9)	109 (+0.3)	99 (-0.7)	84.2 (+0.9)	174 (-0.4)	180.5 (+0.7)	77.4 (+0.5)	144 (-2 SD)	
Weight at last visit kg(SD)	54.2 (-0.3)	17.5 (+0.2)	11.6 (-0.8)	65.6 (-0.4)	20 (-0.5)	14.5 (-0.8)	19 (+0.1)	10 (-2.8)	28 (+0.4)	41 (-3.3)	19.7 (+0.5)	17.8 (+0.9)	12.7 (+0.2)	57.7 (-1.4)	57.6 (-1)	8.9 (-1.2)	ND	
BMI at last visit kg/m2 (SD)	23,5 (+0,6)	16,5 (+0,8)	15,7 (-0,2)	18,3 (-1,6)	14 (-0,9)	15 (-0,2)	14,9 (-0,3)	14,2 (-1)	17,1 (+0,7)	16,8 (-2,4)	16,6 (+0,8)	18,2 (+1,8)	17,9 (+1,6)	19,1 (-0,9)	17,7 (-1,4)	14,9 (-0,9)	ND	
OFC at last visit cm (SD)	55	51 (+0.6)	50 (+1)	59 (+2.7)	52.5 (+0.6)	48.5 (-0.7)	51 (-0.1)	45.5 (-2.9)	51.8 (-0.4)	54.5 (-0.4)	48.7 (-1,1)	49 (-1)	47.8 (+0.1)	56 (+0.6)	55 (-0.1)	47 (-1.1)	50 (-2,6 SD)	
Failure to thrive	Yes	No	No	No	No	ND	No	Yes	No	Yes Chewing difficulty	No	Infant reflux	No	Yes	No	No	Yes	
Truncal overweight	Yes	No	Trunk adiposity	No	No	No	No	No	No	No	No	Mildly overweight	No	No	No	No	No	

Facial features	Arched eyebrow High and prominent forehead Frontal bossing Asymmetric occipital plagiocephaly Triangular face Low anterior hairline Prominent eyes Small, low-set ears Wide nasal bridge Short philtrum High-arched palate Downturned of corners of mouth Prominent chin	Thin upslanted eyebrows Uplanted palpebral fissures Deep eyes, epicanthus Low-set ears Long philtrum Downturned of corners of mouth Buccal hypotonia High-arched palate Serious gaze Chin crease	Uplanted palpebral fissures Deep eyes, epicanthus Protuding ears High arched palate Malar hypoplasia	Aquiline nose Retrognathia Protuding ears High arched palate Curved upper lip Buccal hypotonia High arched palate	Long face Horizontal eyebrows Bilateral epicanthus Posterior rotating ears Smooth philtrum Low-set ears Exaggerated cupid's bow	Downslanting palpebral fissures Smooth philtrum Low-set ears	Hypertelorism Mild micrognathia Smooth philtrum Thin lips Gingival hyperplasia Wide nasal bridge Low set ears	No dysmorphic - appropriate for family	Brachycephaly Malar hypoplasia Left preauricular pit Narrow ear canals Smooth philtrum Thin upper lip vermillion Mildly broad/low posterior hairline	Downslanting palpebral fissures Strabismus Maxillary hypoplasia	Synophrys Highly arched eyebrow Tubular nose with bulbous tip Thin upper lips	Thin upper lip Hypertelorism Short chin	Full cheeks Epicantal folds Telecanthus Thin upper lip	Oval face Thick lips Pointed chin Long and thin nose Microretrognathism Downslanted palpebral fissures	Triangular face with facial asymmetry Upstanging palpebral fissures Depressed nasal bridge Midly downturned corners of the mouth	Downslanted palpebral fissures Depressed nasal bridge Midly downturned corners of the mouth	Elongated face Low hair line Ptosis
Other features																	
Cutaneous abnormalities	Stretch marks	ND	No	No	Truncal and upper limb hypertrichosis Palmar erythema	Hirsutism	No	No	Light skin color Multiple freckles	No	No	No	-	No	No	No	
Additional clinical features	Missing teeth Endometriosis Mixed malignant germ cell tumor of the left ovary	Relative pain insensitivity	Relative pain insensitivity	No	Unexplained episodes of thrombocytosis	Sacro-coccygeal dimple	No	Severe constipation	No	No	Sleeping problems	Sparse hair 2 hemangiomas	Moebius syndrome	Bifid uvula	No	No	
Additional molecular features	-	-	-	-	-	-	-	-	NM_000195.3: c.973_974insC, p.Met325Thrfs*128 Diagnosed with Hermansky Pudlak Syndrome	-	MBD6 variant NM_052897.3:c.2337dupT	-	-	-	-	15q11.2 BP1-BP2 microdeletion	

a ND, no data;

b OFC, occipito-frontal circumference;

c SD, standard deviation;

d +, present;

e -, not present

Supplemental methods

Exome sequencing

Representative exome sequencing analyses are reported here below.

Individual 8: whole exome sequencing and data processing were performed by the Genomics Platform at the Broad Institute of MIT and Harvard with an Illumina Nextera or Twist exome capture (~38 Mb target) and sequenced (150 bp paired reads) to cover >80% of targets at 20x and a mean target coverage of >100x. Exome sequencing data was processed through a pipeline based on Picard and mapping done using the BWA aligner to the human genome build 38. Variants were called using Genome Analysis Toolkit (GATK) HaplotypeCaller package version 3.5.

Individual 9: Using genomic DNA from the proband and parents, the exonic regions and flanking splice junctions of the genome were captured using the IDT xGen Exome Research Panel v1.0 (Integrated DNA Technologies, Coralville, IA). Massively parallel (NextGen) sequencing was done on an Illumina system with 100bp or greater paired-end reads. Reads were aligned to human genome build GRCh37/UCSC hg19, and analyzed for sequence variants using a custom-developed analysis tool. Reported variants were confirmed, if necessary, by an appropriate orthogonal method in the proband and, if submitted, in selected relatives. Additional sequencing technology and variant interpretation protocol has been previously described.² The general assertion criteria for variant classification are publicly available on the GeneDx ClinVar submission page (<http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/>)

Individual 15: WES methods and data processing, including sequence alignment to GRCh37/hg19, variant filtering and prioritization, predicted functional impact, inheritance and validation, were performed as previously reported.³ In Brief: WES sequencing was performed on the trio (patient and both parents) was performed at the IRCCS Burlo Garofolo (Trieste, Italy). According to the manufacturer, DNA was processed for library enrichment with the Twist Human Core for Enrichment-Exome panel (Twist). Sequencing was performed on NextSeq 500 (Illumina, San Diego, USA) instrument. Sequencing reads were aligned to the human reference genome (GRCh37/hg19) employing the BWA-mem tool, and variant calling were performed with GATK v4.1.2 HaplotypeCaller (EnGenome, Pavia, Italy). Annotation and prioritization of the variants were conducted with eVAI-enGenome software based on the American College of Medical Genetics and Genomics (ACMG) guidelines.⁴ SNVs and INDELs were filtered referring to public databases (dbSNP build150, NHLBI Exome Sequencing Project (ESP), Exome Aggregation Consortium (ExAC) database, Genome Aggregation Database (gnomAD)) and led to ruling out those variants previously reported as polymorphism. In particular, a minor allele frequency (MAF) cut-off of ≤0.01% was utilized. The trio was examined for the following inheritance patterns: de novo, homozygous recessive, compound

heterozygous, and hemizygous. The pathogenicity of variants was assessed with the ClinVar, The Human Gene Mutation Database (HGMD), Online Mendelian Inheritance in Man (OMIM), and DECIPHER. Several in silico tools, such as PolyPhen-2,⁵ Sorting Intolerant from Tolerant (SIFT),⁶ PaPI,⁷ DANN,⁸ dbscSNV score,⁹ and Combined Annotation Dependent Depletion (CADD) score,¹⁰ were applied to establish the pathogenicity of novel variants. The diagnostic procedure adopted included discussion of WES data in the context of phenotypic data at interdisciplinary meetings; the interpretation of WES data was supported by systematic bibliographic review and public database consultation.¹¹ Variants' validation and segregation analyses were carried out by Sanger sequencing.

A quantitative *in vivo* splicing assay in *Drosophila* to model human SRSF1 activity

To develop an *in vivo* readout for human SRSF1 function, we used *Drosophila* as a model system. Human SRSF1 and its fly ortholog, SF2 (DIOPT weighted score of 12.87, https://www.flrnai.org/cgi-bin/DRSC_orthologs.pl), are evolutionarily highly conserved and show an overall identity of 63% increasing to 81% and 76% in RRM1 and RRM2 respectively (**Figure S2A**). In accordance, when we aligned the human and fly protein models from the AlphaFold Protein Structure Database, we observed that both RRM domains adopt a nearly identical folding. Even the prediction for the flexible parts outside the RRMs was very similar (**Figure S2B**). These data support structural and functional conservation of both proteins.

We therefore generated transgenic fly lines expressing wild-type (WT) and mutant human SRSF1 and fly SF2. We used PhiC31 integrase-mediated insertion to minimize the influence of genomic background on expression levels. We performed a transcriptome analysis in our overexpression lines (**Figure S2C**). The alternative splicing events induced by SRSF1 and SF2 proteins were very similar, confirming the functional conservation of the two proteins at the molecular level. Alternative 5' splicing and exon skipping events were the splicing events with the main alterations, in line with the reported splicing activity of SRSF1.^{12–14} Furthermore, more than 60% of the altered genes are shared between both proteins and pathway analysis on the differentially spliced genes shows enrichment for genes involved in morphogenesis and development (**Figure S4**).

We therefore ectopically expressed SRSF1 and SF2 in the visual system. Eye-specific expression of human SRSF1 and fly SF2 resulted in a strong rough eye phenotype characterized by disorganized ommatidia, misorientation of the interommatidial bristles, heterogenous depigmentation and necrotic spots (**Figure S2D**), supporting the functional homology of both proteins. To causally link the splicing alterations with the observed eye phenotype, we generated *SRSF1* constructs harboring biochemically characterized amino acid changes (**Figure S2D**).^{14–16}

SRSF1 p.Ile32Ala/p.Val35Ala and p.Phe56Asp/p.Phe58Asp (hereafter indicated as I32A/V35A and F56D/F58D respectively) completely destabilize U1snRNP recruitment to the ESE binding side. Therefore, they can be considered splicing-defective variants. SRSF1 I32A/V35A specifically intervenes with the U1-70K binding sites of RRM1 and F56D/F58D abolishes the RNA binding capacity of RRM1.¹⁶ Both splicing deficient SRSF1 proteins lost their capacity to induce an eye phenotype showing that recruitment and stabilization of the U1 snRNP at the 5'SS is necessary to induce the phenotype. Furthermore, a control variant p.Asp66Ala/p.Asp69Ala (D66A/D69A) with preserved splicing activity retained this capacity.¹⁶ SRSF1 p.Lys38Ala/p.Tyr39Ala (K38A/Y39A) with limited splicing activity presented with an intermediate phenotype accordingly.¹⁶ The phenotype-inducing capacity of SRSF1 p.Lys138Ala (K138A), a variant abolishing specifically the functionality of the RRM2 or pseudo RRM, was not altered compared to a control variant p.Tyr149Ala (Y149A) with retained capacity.¹⁵ We thus quantified the severity of the eye phenotype based on the roughness (IREG-score, **Figure S2E**) and depigmentation level of all RRM1 variants (**Figure S2F**). We observed a clear difference in both readouts between splicing active (D66A/D69A) and inactive species (I32A/V35A; F56D/F58D and F56D/F58D/K138A). To confirm that these results were not tissue-dependent, we tested the effect of overexpression of SRSF1 active and splicing deficient species in the nervous system (**Figure S2G**). Misexpression of wild-type SRSF1 in the nervous tissue resulted in pharate adult lethality, which is absent in flies overexpressing a splicing-deficient SRSF1. Together these results indicate that the modeling mainly addresses the stabilization of the U1 snRNA on the pre-mRNA by SRSF1 and that this model can be used to study the splicing activity of human SRSF1 *in vivo* in a quantitative manner.

Acknowledgements

We thank the University of Burgundy “Centre de Calcul” (CCuB) for technical support and management of the informatics platform. We also thank the Undiagnosed Diseases Network International (UDNI) for contributing to affected individual identification. This work was made possible only thanks to international data sharing through GeneMatcher.¹⁷ Our cohort is composed of 17 individuals from 13 different medical centers and 6 different countries. We submitted our first individual with a *SRSF1* variant in 2019 and we have been able to gather 16 additional individuals in three years. GeneMatcher is a powerful tool to create interactions between researchers, and has been essential to accelerate the identification of novel disease genes in the last years.^{18,19}

Supplemental references

1. Chevarin, M., Duffourd, Y., A. Barnard, R., Moutton, S., Lecoquierre, F., Daoud, F., Kuentz, P., Cabret, C., Thevenon, J., Gautier, E., et al. (2020). Excess of de novo variants in genes involved in chromatin remodelling in patients with marfanoid habitus and intellectual disability. *J. Med. Genet.* *57*, 466–474. 10.1136/jmedgenet-2019-106425.
2. Retterer, K., Juusola, J., Cho, M.T., Vitazka, P., Millan, F., Gibellini, F., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K.G., et al. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genet. Med. Off. J. Am. Coll. Med. Genet.* *18*, 696–704. 10.1038/gim.2015.148.
3. Musante, L., Faletra, F., Meier, K., Tomoum, H., Najarzadeh Torbati, P., Blair, E., North, S., Gärtnner, J., Diegmann, S., Beiraghi Toosi, M., et al. (2022). TTC5 syndrome: Clinical and molecular spectrum of a severe and recognizable condition. *Am. J. Med. Genet. A.* *188*, 2652–2665. 10.1002/ajmg.a.62852.
4. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). 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. Off. J. Am. Coll. Med. Genet.* *17*, 405–424. 10.1038/gim.2015.30.
5. Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr. Protoc. Hum. Genet.* *76*. 10.1002/0471142905.hg0720s76.
6. Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* *40*, W452–457. 10.1093/nar/gks539.
7. Limongelli, I., Marini, S., and Bellazzi, R. (2015). PaPI: pseudo amino acid composition to score human protein-coding variants. *BMC Bioinformatics* *16*, 123. 10.1186/s12859-015-0554-8.
8. Quang, D., Chen, Y., and Xie, X. (2015). DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinforma. Oxf. Engl.* *31*, 761–763. 10.1093/bioinformatics/btu703.
9. Jian, X., Boerwinkle, E., and Liu, X. (2014). In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res.* *42*, 13534–13544. 10.1093/nar/gku1206.
10. Kircher, M., Witten, D.M., Jain, P., O’Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* *46*, 310–315. 10.1038/ng.2892.

11. Musante, L., Costa, P., Zanus, C., Faletra, F., Murru, F.M., Bianco, A.M., La Bianca, M., Ragusa, G., Athanasakis, E., d'Adamo, A.P., et al. (2022). The Genetic Diagnosis of Ultrarare DEEs: An Ongoing Challenge. *Genes* **13**, 500. 10.3390/genes13030500.
12. Paz, S., Ritchie, A., Mauer, C., and Caputi, M. (2021). The RNA binding protein SRSF1 is a master switch of gene expression and regulation in the immune system. *Cytokine Growth Factor Rev.* **57**, 19–26. 10.1016/j.cytofr.2020.10.008.
13. Jobbins, A.M., Campagne, S., Weinmeister, R., Lucas, C.M., Gosliga, A.R., Clery, A., Chen, L., Eperon, L.P., Hodson, M.J., Hudson, A.J., et al. (2022). Exon-independent recruitment of SRSF1 is mediated by U1 snRNP stem-loop 3. *EMBO J.* **41**. 10.15252/embj.2021107640.
14. Cho, S., Hoang, A., Chakrabarti, S., Huynh, N., Huang, D.-B., and Ghosh, G. (2011). The SRSF1 linker induces semi-conservative ESE binding by cooperating with the RRM. *Nucleic Acids Res.* **39**, 9413–9421. 10.1093/nar/gkr663.
15. Cléry, A., Sinha, R., Anczuków, O., Corrionero, A., Moursy, A., Daubner, G.M., Valcárcel, J., Krainer, A.R., and Allain, F.H.-T. (2013). Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E2802-2811. 10.1073/pnas.1303445110.
16. Cho, S., Hoang, A., Sinha, R., Zhong, X.-Y., Fu, X.-D., Krainer, A.R., and Ghosh, G. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 8233–8238. 10.1073/pnas.1017700108.
17. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Hum. Mutat.* **36**, 928–930. 10.1002/humu.22844.
18. Bruel, A.-L., Vitobello, A., Mau-Them, F.T., Nambot, S., Duffourd, Y., Quéré, V., Kuentz, P., Garret, P., Thevenon, J., Moutton, S., et al. (2019). 2.5 years' experience of GeneMatcher data-sharing: a powerful tool for identifying new genes responsible for rare diseases. *Genet. Med.* **21**, 1657–1661. 10.1038/s41436-018-0383-z.
19. Julkowska, D., Austin, C.P., Cutillo, C.M., Gancberg, D., Hager, C., Halftermeyer, J., Jonker, A.H., Lau, L.P.L., Norstedt, I., Rath, A., et al. (2017). The importance of international collaboration for rare diseases research: a European perspective. *Gene Ther.* **24**, 562–571. 10.1038/gt.2017.29.