Musante *et al. Genome Medicine* (2026) 18:34
https://doi.org/10.1186/s13073-026-01610-4

Genome Medicine

RESEARCH Open Access

A novel spliceosomopathy caused by *de novo* *SF3B3* variants

Luciana Musante[1*], Pavel Janos[2†], Giulia Pianigiani[1†], Sara Cappelli[3], Alessandra Longo[3], Carolina Alves[4], Eva MC Schwaibold[5], Matias Wagner[6,7,8], Gregory Costain[9,10], Run Fridriksdottir[11], Kari Stefansson[11], Patrick Sulem[11], Klaske D Lichtenbelt[12], Ellen van Binsbergen[12], Richard H van Jaarsveld[12], Alfredo Brusco[13,14], Lisa Pavinato[13,15,16], Elisa Biamino[17], Alessandra Spano[17], Clara C Hildebrandt[18], Yee-Ming Chan[19,20,21], Emily Groopman[21,22], Michal Berkenstadt[23], Daniel Koboldt[24], Rachel Williamson[25], Han G Brunner[26,27], Lisenka ELM Vissers[26], Pernille M Torring[28], Qin Hao[28], Bruce D Gelb[29], Elizabeth Goldmuntz[30,31], Kristen Reed[30], Emma C Bedoukian[32], Davide Vecchio[33], Emanuela Salzano[34], Maria Piccione[34,35], Caterina Zanus[1], Catia Mio[36], Evan E Eichler[37,38], Tianyun Wang[39,40,41], Wesley G Patterson[42], Kameryn M Butler[42], Mattie Piotrowski[42], Sandra Mercier[43,44], Benjamin Cogné[43,44,45], Ingrid M. Wentzensen[46], Emanuele Buratti[3], Alessandra Magistrato[2] and Flavio Faletra[36,47*]

Abstract

Background Spliceosomopathies are syndromes caused by pathogenic variants in genes involved in splicing and mRNA metabolism. Here, we report a novel spliceosomopathy caused by *de novo* variants in *SF3B3*, encoding a subunit of the spliceosomal SF3b complex.

Methods We performed genomic, clinical, computer-aided gestalt analysis, molecular dynamics simulations, and functional studies using patient-derived fibroblasts.

Results Through international data sharing, we collected clinical and molecular data from 24 unrelated individuals with heterozygous *SF3B3* variants, mostly missense, consistent with autosomal dominant inheritance. Individuals exhibited a congruent phenotype including autism spectrum disorder (ASD), developmental delay (DD), intellectual disability (ID), language and motor delay, multiple congenital anomalies, and distinctive craniofacial features, confirmed by GestaltMatcher analysis. In patient fibroblasts, *SF3B3* mRNA was within the normal range, whereas protein levels were reduced by approximately 15–30% depending on the variant. All-atom simulations revealed impaired interactions of mutant SF3B3 with SF3b components. Transcriptome profiling revealed widespread gene expression changes, including genes involved in cell-cycle regulation, urogenital development, heart morphogenesis, neural crest differentiation, and neurogenesis. Alternative splicing analyses demonstrated specific alterations,

†Pavel Janos and Giulia Pianigiani contributed equally to this work.

*Correspondence:
Luciana Musante
luciana.musante@burlo.trieste.it
Flavio Faletra
flavio.faletra@asufc.sanita.fvg.it

Full list of author information is available at the end of the article

including increased retained intron events. Functional assays confirmed cell-cycle abnormalities in patient-derived fibroblasts.

Conclusions *SF3B3* variants cause a novel spliceosomopathy with a continuous clinical spectrum, from extremely severe prenatal forms with perinatal lethality to a milder form with autism ASD and DD/ID. These variants alter both stability and function of the SF3b complex, resulting in dysregulated transcriptome, alternative splicing defects, and downstream cellular consequences such as cell-cycle perturbation.

Keywords *SF3B3*, *De novo*, Spliceosomopathy, SF3b complex, All-atom simulation, RNA sequencing

Background

Neurodevelopmental disorders (NDDs) are complex conditions and a significant public health concern, affecting approximately 3% of the global population [1]. Defined as "behavioural and cognitive disorders that arise during the developmental period that involve significant difficulties in the acquisition and execution of specific intellectual, motor, language, or social functions" (ICD-11, https://icd.who.int/browse11/l-m/en), they frequently manifest pleiotropic or complex syndromic features, involving other organs in addition to the central nervous system [2]. Although NDDs are largely of genetic origin, the clinical and genetic heterogeneity of these conditions represents a challenge to a syndrome-based approach to diagnosis. The advent of next-generation sequencing (NGS) technologies and an increasing number of bioinformatic tools coupled with phenotypic and genomic data-sharing platforms has dramatically improved the diagnostic process. This has resulted in an enhanced possibility of establishing molecular causes and identifying novel disease genes. Moreover, it has also highlighted the importance of *de novo* variants in syndromic forms of NDDs [3, 4].

Concurrently, an increasing body of evidence has emerged indicating that pathogenic variants in genes involved in mRNA splicing and RNA metabolism play a role in the pathogenesis of NDDs [5, 6]. Indeed, pathogenic variants in several core components of the spliceosome have been identified as the underlying cause of a group of syndromes collectively known as spliceosomopathies [7, 8]. The spliceosome is one of the most complex machineries of eukaryotic cells. It is responsible for identifying noncoding intronic sequences within mRNA precursors and facilitating the precise splicing process that generates functional messengers in eukaryotes and long non coding RNAs. Furthermore, the variation in the patterns of intron removal (alternative splicing) contributes to proteome diversification. Indeed, higher eukaryotes contain a U2-dependent (major) splicing pathway and a less abundant U12-dependent (minor) one, which splice-specific pre-mRNA introns (U2-type and U12-type) differing in the splice site and branch point consensus sequences. The core of the major spliceosome is composed of small nuclear RNAs (U1, U2, U4, U5 and U6) and more than 100 proteins, forming five ribonucleoprotein complexes (snRNPs) [9]. The spliceosome assembles on pre-mRNAs to establish a dynamic cascade of RNA and protein interactions that facilitate removing the introns and ligation of the exons [10]. The conformation and composition of the spliceosome, characterised by multiple intermediary complexes (E, A, B, B^{act}, B*, C and P), are highly dynamic, allowing the machinery to be accurate and flexible [11]. SF3b, an intrinsic component of the functional U2 snRNP, plays a pivotal role in recognition of the branch point sequence, facilitating the assembly of the spliceosome and the conversion into the active form (B^{act} complex) in a process called activation [12]. SF3b is a dynamic and integral component of four intermediary complexes (A, B, and B^{act}, B*) and comprises multiple proteins with varying molecular sizes and distinct domains. The fundamental core of SF3b is structured with the SF3B1 protein associated with PHF5A, SF3B3, SF3B5, SF3B2, SF3B6 and SF3B4 [13]. Recent studies have indicated that mutations in several components of the SF3b complex and other U2-associated splicing factors are involved in the aetiology of rare human syndromes. These include: (a) haploinsufficiency of *SF3B2*, which is responsible for the craniofacial microsomia (MIM: #164210) [14], (b) heterozygous *SF3B4* variants causing Nager syndrome (MIM: #154400) [15] (c) somatic variants in *SF3B1*, *U2AF1*, and *SRSF2* associated with myelodysplastic syndromes (MIM: #614286) [16, 17] and (d) *PHF5A*, *RNF113A*, and *PUF60* pathogenic alterations detected in syndromic NDDs (MIM: #300953, MIM: #615583) [18–20]. Somatic variants in *SF3B3* have been documented in several tumours (COSMIC v97), with overexpression of the protein being a common feature. In accordance with these findings, *SF3B3* knockdown induced cell death in both in vitro and in vivo cancer models through autophagy [21]. However, to date, no evidence suggests that *SF3B3* germline variants are associated with human disease.

SF3B3 is not only a core component of the spliceosome, but also a stable subunit of the transcriptional coactivator SAGA/TFTC, which is a histone acetyltransferase complex implicated in the regulation of gene expression via chromatin remodelling. SF3B3 thus represents a versatile scaffold protein that acts in various cellular contexts,

coupling transcription and splicing processes [22–24]. Additionally, SF3B3 modulates the pre-mRNA splicing of the Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit (*EZH2*) [25], encoding a histone methyltransferase and a component of the polycomb repressive complex 2 (PRC2), which is mutated in patients with Weaver syndrome (MIM: #277590) [26]. Moreover, the Lysine-Specific Demethylase 4B (*KDM4B*), mutated in patients with a dominant form of NDD (MIM: #619320), regulates the chromatin accessibility and couples the spliceosome to the chromatin itself, eliciting its binding to the splicing factor SF3B3 and thereby regulating the alternative splicing of target RNAs [27].

In this study, we present clinical and molecular data, collected through an international data-sharing initiative, from 24 patients with *de novo* variants in *SF3B3*. These variants are associated with a recognisable syndromic NDD characterised by DD, ID, ASD, language and motor delay, multiple congenital anomalies and a distinctive gestalt. Furthermore, we elucidate the impact of SF3B3 protein variants on the functional dynamics of the SF3b complex via all-atom simulation. Finally, we identify transcriptomic and alternative splicing alterations as well as functional cell-cycle dysregulation in patient-derived fibroblasts. Taken together, these results suggest that SF3B3 variants disrupt core SF3b functions, leading to widespread molecular dysregulation that underlies a novel autosomal dominant neurodevelopmental disorder.

Methods

Identification of patients

The study cohort comprised 24 patients. Subject 1 was enrolled at the Institute for Maternal and Child Health - IRCCS "Burlo Garofolo" - Trieste, Italy. The other individuals were identified through multiple sources: GeneMatcher (subjects 2–10, 12, 23–24) [28], the Pediatric Cardiac Genomics Consortium (subjects 13–14), research and clinical diagnostic laboratories (subjects 11 and 19), and published literature (subjects 15–18, 20–22). Sources of ascertainment were not mutually exclusive, and each patient was included once in the final cohort. We collected clinical and genetic details via a standardised form from their clinicians/geneticists. All the investigations were conducted according to the Declaration of Helsinki principles and approved by the local institutional ethics committees. We obtained informed consent from all participants and permission to publish patient photographs.

GestaltMatcher analysis

The GestaltMatcher algorithm [29] is a computational tool that converts facial characteristics captured by a photograph into numerical descriptors, allowing for quantitative comparison and clustering of facial phenotypes. The tool is part of Face2Gene (FDNA Inc., USA), a HIPAA-compliant platform, which enables the interaction of different algorithms to analyse patients' facial phenotypes and can be leveraged for syndrome classification (lumping and splitting), phenotype delineation, and patient clustering based on facial phenotype [30, 31]. The GestaltMatcher analysis conducted in this study is the Pairwise Comparison Matrix (PCM), in which images from a cohort of patients were compared with the FDNA's GestaltMatcher gallery (4300 images), yielding degrees of similarity. The number in each of the matrix cells represents the similarity rank achieved. Dark green values (low rank) indicate higher similarity in facial phenotypic features within the test cohort.

WES/WGS methods, data processing and validation

Unless otherwise indicated, trio or singleton whole exome sequencing (WES) or whole genome sequencing (WGS) was performed at the respective institutions. Genomic DNA was extracted from peripheral blood. The target enrichment kits and WES/WGS statistics are reported in Additional file 1: Table S1A. Data processing, including sequence alignment to GRCh37/GRCh38, variant filtering and prioritisation, were performed as previously reported (Additional file 1: Table S1A and references herein).

Confirmation and segregation analysis of variants has been performed by Sanger sequencing. The *SF3B3* RefSeq NM_012426.5 transcript was used as a reference.

We predicted the variants' functional impact through a series of in silico prediction tools as detailed in Additional file 1: Table S1A. All the variants were interpreted using VarSome [32] (https://varsome.com/), generating a pathogenicity recommendation based on the American College of Medical Genetics and Genomics (ACMG) guidelines [33] (Additional file 1: Table S1A). We evaluated the clinical relevance of the identified variants by consulting the following databases (last accession on 10 March 2025): ClinVar [34], Human Gene Mutation Database (HGMD) Professional (Qiagen) [35], Online Mendelian Inheritance in Man (OMIM) [36], and DECIPHER [37].

The MetaDome web tool (v.1.0.1) [38] (https://stuart.radboudumc.nl/metadome, last accession on 10 March 2025), a free available web server, was employed to further interpret the SF3B3 variants. The MetaDome tool [38] assesses the tolerance of mutations at each position in a human protein. It combines data from genomics and proteomics and maps variants from the gnomAD database [39] and pathogenic ClinVar variants onto Pfam protein domains, and uses this information to produce a map of quantified tolerance to variation at every position in the predicted protein.

The algorithms I-Mutant2.0 [40] (https://folding.biofold.org/i-mutant/i-mutant2.0.html) and MUpro [41] (https://mupro.proteomics.ics.uci.edu/) were used to assess the protein's stability (last accession on 23 November 2025) (Additional file 1: Table S1B).

Structural model building and computational methods

The major component of the SF3b complex is the SF3B1 protein. The C-terminal region of SF3B1 comprises 22 HEAT repeats capable of undergoing a conformational change from an open (RNA-free) to a closed (RNA-bound) state. SF3B1 assumes an open conformation at an early stage of the splicing cycle prior to the binding of the RNAs [42]. Subsequently, upon recognition of the pre-mRNA, SF3B1 is "clamped" on the bound U2 snRNA and pre-mRNA strands, thereby assuming a closed state. Hence, to study the effects of the SF3B3 variants, we have built three different models of the SF3b complex. The first model (termed Closed) was based on the cryoEM structure of the human mature B^{act} complex (PDB ID: 6FF4) [43]. In this structure, missing protein parts were modelled using other B^{act} structures deposited in the PDB (PDB IDs: 5Z58, 7DVQ) [44, 45]. The second model (termed Open_1) was based on the x-ray structure of the SF3b core with an inhibitor bound (PDB ID: 6EN4) [46]. The third model (Open_2) was built on the cryo-EM structure of the human 17S U2 snRNP (PDB ID: 7Q3L) [47]. The Open models differ from the Closed one in the protein composition (for a comprehensive list of proteins, please refer to Additional file 2: Table S2), the SF3B1 conformation and the relative position of SF3B1 and SF3B3. The Open_2 model, based on a more recent cryo-EM structure, contains more protein components than Open_1, and was built to monitor the impact of SF3B3 variants on specific protein-protein interactions present in the newly solved regions.

We modelled missing loops in the structures using Modeller [48]. We prepared them in Schrodinger's Protein Preparation Wizard [49, 50], including prediction of the protonation states of ionizable residues using propKa [51, 52]. Models were then solvated in a cubic water box with a minimal distance from the solute to the box edge of 12 Å. NaCl 150 mM was added to the solvent. We calculated the amount of ions required to reach the concentration employing SLTCAP [53]. The topologies of the model were built using AmberTools' leap program [54]. We included the mutations by deleting the side chain atoms of the wild-type residues and building the corresponding mutated structure using the leap module of the AmberTools. We converted the Amber topologies and coordinates into the Gromacs format using parmed of AmberTools [55] and described the system using following force-field parameters: FF14SB for the protein [56], OL3 for the RNA [57], TIP3P for the water molecules [58], Joung-Chetham for the Na^+ and Cl^- ions [59] and a dummy atom model of Zn^+ ions [60]. All systems were minimised, brought up to 300 ^oK and gently equilibrated, applying the protocol detailed in Additional file 2: Table S3. Production molecular dynamics simulations, performed using Gromacs 2022 [61, 62], were run at 300 ^oK and 1 bar; temperature and pressure were maintained using Bussi thermostat and Parrinello-Rahman barostat, respectively [63, 64]. A 2-fs integration time step was used with all covalent bonds involving hydrogen atoms constrained using the LINCS algorithm [65]. We treated electrostatic interactions using the Particle Mesh Ewald approach with a space cutoff of 10 Å [66]. We conducted analyses with cpptraj from AmberTools [67]. Binding free energies were calculated with the Molecular Mechanics Generalized Born Surface Area (MM/GBSA) approach using the MMPBSA.py contained in the AmberTools [68]. Structures were visualised using VMD [69].

RNA extraction

Total RNA was extracted from patient and age- and sex-matched control fibroblast cell lines using the RNeasy Mini Kit (Qiagen). DNase I treatment was performed to ensure the absence of any DNA contamination, in accordance with the instructions provided by the manufacturer.

Gene expression analysis and validation using quantitative real-time PCR (qRT-PCR)

Total RNA (1 µg) was reverse transcribed into cDNA using the iScript Advanced cDNA Synthesis Kit for RT-qPCR (Bio-Rad). qRT-PCR was performed under standard PCR conditions using the CFX96 Real-Time PCR System (Bio-Rad) with iTaq Universal SYBR Green Supermix (Bio-Rad) and custom-designed primers (listed in Additional file 2: Table S4). The amplification protocol consisted of an initial denaturation at 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s, and a final melting curve analysis from 65 °C to 95 °C with 0.5 °C increments every 5 s. *GAPDH* was used as the housekeeping gene for normalization. Relative gene expression levels and fold changes were determined using the 2^−ΔΔCt method [70]. All experiments were performed in triplicate.

RNA sequencing

Library preparation and RNA sequencing were performed by Novogene (https://en.novogene.com/). Sequencing was performed on an Illumina Novaseq platform, and 150-bp, paired-end reads were generated. Briefly, the quantity and quality of the RNA was evaluated with the RNA Nano 6000 Assay Kit on the Bioanalyzer 2100 system (Agilent Technologies). The messenger RNA (mRNA) was purified from the total RNA using

poly-T oligo-attached magnetic beads. The fragmentation was carried out using divalent cations under elevated temperature in the First Strand Synthesis Reaction Buffer (5X). First-strand cDNA was synthesised using random hexamer primer and M-MuLV reverse transcriptase (RNase H). Subsequently, second-strand cDNA synthesis was performed using DNA polymerase I and RNase H. The remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of the 3' ends of the DNA fragments, adaptors with a hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 370 ~ 420 bp in length, the library fragments were purified using the AMPure XP system. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. Finally, the PCR products were purified (AMPure XP system) and the quality of the library was assessed using the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) following the manufacturer's instructions. Raw reads containing adapter, reads containing poly-N and low-quality reads were successively removed through Novogene in-house Perl scripts. Pair-ends clean reads were aligned to GRChg38 human reference genome using Hisat2 v.2.0.5 and assembled by StringTie v1.3.3b [71]. Differential gene expression analysis was carried out using DEseq2 R packages (v1.20.0). The overall distribution of differentially expressed genes (DEGs) was evaluated using the following cut-off: up-regulated genes Fold Change (FC) > 2 and adjusted p-value (padj) < 0.05; down-regulated genes FC < 0.5 and padj < 0.05. ClusterProfiler package (v4.6.2) from R was used for gene set enrichment (GSE) and over-representation gene ontology (GO) analysis of the RNA sequencing data. Graphical representation of the RNA sequencing data was realised using the ggplot2 R package (v3.4.2).

Differential alternative splicing (DAS) events were analyzed using the rMATS algorithm. Significant events were defined as those with a False Discovery Rate (FDR) < 0.05 and an absolute Inclusion Level Difference (|IncLevelDifference|) ≥ 0.2.

To characterize splicing alterations, we first assessed differences in the global distribution of event types between patients and controls using the Wilcoxon rank-sum test. Since patients carried distinct mutations, Z-scores were calculated for each individual relative to the control distribution to quantify deviations across event types. To determine whether specific AS categories were selectively enriched among significant events, a Chi-square test was used to compare the proportional distribution of event types between all detected and significant events.

Representative events were visualized as grouped sashimi plots using the rmats2sashimiplot package on Python (v3.9.23).

We used the publicly available SRplot platform (https://www.bioinformatics.com.cn/srplot) to prepare bubble graphs, GO chords, and alluvial plots.

Cell culture and western blot

HEK293T were cultured in Dulbecco's modified Eagle medium (DMEM) with high glucose, sodium pyruvate and L-glutamine (Euroclone), supplemented with 10% v/v fetal bovine serum (Euroclone) and 50 units penicillin/ 50 µg streptomycin (Euroclone). Fibroblast cell lines were cultured in DMEM with high glucose, sodium pyruvate and L-glutamine (Euroclone), supplemented with 20% v/v fetal bovine serum (Euroclone) and 50 units penicillin/ 50 µg streptomycin (Euroclone). Cells were maintained in a humidified incubator at 37 °C, 5% CO_2.

For protein extraction, patients and controls' fibroblasts were harvested and lysed in a lysis buffer (65 mM Tris-HCL pH6.8, 26% w/v glycerol, 2% SDS, 0.01% bromophenol blue). Subsequently, protein lysates were purified using the QIAshredder (QIAGEN). Protein concentration was determined by Nanodrop One (Thermo Fisher Scientific). We separated the protein lysates into 4–20% SDS-PAGEs and transferred them to PVDF membranes. We blocked the membranes in 5% milk in TBST to prevent non-specific antibody binding. We conducted the immunoblotting with the primary antibody incubated overnight at 4 °C, followed by three washing steps with TBST and by the secondary antibody conjugated with horseradish peroxidase (1 h at room temperature), developed using Clarity™ Western ECL Substrate (BioRad). We captured images using the ChemiDoc MP imaging system (BioRad) and quantified bands with ImageJ software version 1.53e [72]. The following primary antibodies were employed: rabbit polyclonal anti-SF3B3 (A302-508 A, Bethyl Laboratories), rabbit polyclonal anti-SF3B1 (27684-1-AP, Proteintech), rabbit polyclonal anti-SF3B2 (10919-1-AP, Proteintech), rabbit polyclonal anti-SF3B4 (A303-950 A, Bethyl Laboratories), rabbit polyclonal anti-SF3B14 (SF3B6) (12378-1-AP, Proteintech), rabbit polyclonal anti-PHF5A (15554-1-AP, Proteintech), rabbit polyclonal anti-RNF113A (27018-1-AP, Proteintech), mouse monoclonal anti-GAPDH (0411) (sc-47724, Santa Cruz Biotechnology Inc). We utilised the following secondary antibodies: goat anti-rabbit IgG (H + L)-HRP (#31460, Invitrogen) and donkey anti-mouse IgG (H + L)-HRP (715-035-150, Jackson ImmunoResearch).

Cell fractionation assay

Patient-derived and control fibroblasts were plated at 5×10^5 cells per well in 6-well plates. After 24 h, cells were harvested using trypsin-EDTA and subjected to

nuclear and cytoplasmic fractionation using the NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Protein concentrations were determined, and equal amounts of lysates were analyzed by Western blot using anti-SF3B3 antibody (A302-508 A, Bethyl Laboratories). Anti-Lamin A/C (SC-7292, Santa Cruz) and anti-Tubulin antibodies (ab6160, Abcam) were used to assess the purity of nuclear and cytoplasmic fractions, respectively.

MG132 treatment

Patient-derived and control fibroblasts were plated in 6-well plates at a density of 5×10^5. After 24 h, cells were treated with 10 μM MG132 or DMSO for 6 h at 37 °C. After treatment, the cells were lysed with a lysis buffer (65 mM Tris-HCL pH6.8, 26% w/v glycerol, 2% SDS, 0.01% bromophenol blue). Samples were purified using the QIAshredder (QIAGEN). For all samples protein concentration was measured by Nanodrop. Lysates were separated in SDS-PAGE and transferred to PVDF membrane. Then western blot analyses were performed as described in previous section. The membranes were incubated with primary antibodies overnight at 4 °C.

Cloning of *SF3B3* constructs and transient transfection

HEK293T cells were plated in 6-well plates at a density of 5×10^5 cells per well one day prior to transfection. Cells were co-transfected with 375 ng of either wild-type or mutant *SF3B3* constructs together with 150 ng of pcDNA3.1-GFP using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's instructions. GFP was used to control for transfection efficiency. After 24 h, cells were harvested and lysed in lysis buffer (65 mM Tris-HCl pH 6.8, 26% w/v glycerol, 2% SDS, 0.01% bromophenol blue). Lysates were clarified using QIAshredder columns (QIAGEN) and subjected to western blot analysis as described above. Membranes were incubated overnight at 4 °C with primary antibodies against FLAG (F3165, Sigma-Aldrich) and GFP (11814460001, Roche). GAPDH was used as a loading control, detected with mouse monoclonal anti-GAPDH (sc-47724, Santa Cruz Biotechnology Inc).

Cell cycle analysis

In each experiment, 1×10^6 cells were collected, washed with PBS, and fixed in ice-cold 70% ethanol. After centrifugation, cell pellets were washed and resuspended in PI/RNase Staining Buffer (BD Biosciences) for 15 min at room temperature. For all samples, cellular debris were excluded based on FSC-A/SSC-A gating, and singlets were identified using PI-W versus PI-A dot plots. Cell-cycle phase distribution was subsequently quantified from PI-A histograms. Samples were acquired on a FACSMelody flow cytometer (BD Biosciences) and analyzed using FlowJo v10.10.0 (BD Biosciences).

Developmental expression in human brain cortex

To extrapolate the developmental expression trend of SF3b complex components and DEGs in the human brain, we downloaded the developing human brain transcriptome dataset from the BrainSpan databases (http://www.brainspan.org) [73]. We used RNA sequencing data from 8 to 11 neocortical regions from the 8th post-conceptional week to the 40th postnatal year. The gene-level RPKM expression values were further transformed to log2 (RPKM + 1) values. To explore the temporal expression pattern in the cortex, the samples from the same developmental time point were grouped together and the expression values across those grouped samples were averaged, as previously reported [74].

Data analysis and statistics

Quantitative data are presented using Microsoft Excel as the mean ± standard deviation (SD). Two-tailed homoscedastic Student's t-test was used for comparison. A p value < 0.05 was considered statistically significant ($^{*}p \le 0.05$; $^{**}p \le 0.01$; $^{***}p \le 0.001$). Shaded area error bar plots have been created using R version 4.4.2 (R Foundation for Statistical Computing).

Results

Through international collaborations, the GeneMatcher [28], the Paediatric Cardiac Genomics Consortium (PCGC), and from data from large cohorts of individuals with NDDs previously published [75–77], we collected a total of 24 individuals with a NDD and a heterozygous *SF3B3* variant.

Clinical description

We sent a standardised form to the clinicians to obtain clinical details, including prenatal findings, growth, neurodevelopment, congenital malformations, skeletal abnormalities and facial features. For a few individuals (15–18 and 20–21), we collected the available clinical data from published descriptions [75–77].

The phenotypic characteristics of all individuals are presented in Table 1. A more detailed description of each individual and their clinical history is provided in the supplementary information (Additional file 2: Case Reports).

The clinical spectrum observed in individuals with *SF3B3* variants is consistent with a syndromic neurodevelopmental condition with disease severity ranging from: (1) severe, with two newborns dying a few days after birth and one foetus (individuals 9–11); (2) multisystemic, with multiple congenital malformations and a typical gestalt (individuals 1–8, 12–18, and 24); (3) mild,

Table 1 Molecular and clinical details of individuals with *SF3B3* variants

Features \ Patient_ID*	TOTAL	SEVERE	9	11	10	MULTISYSTEMIC	15	18	6	5	7	24	17	4	3
SF3B3 (NM_012426.5): nucleotide change			c.452G>A	c.914C>T	c.1349C>T		c.376A>G	c.637C>A	c.647G>A	c.656A>C	c.1480G>A	c.1574G>A	c.1762_1764del	c.1922G>A	c.2012A>G
SF3B3 (NP_036558.3): amino acid change			p.Arg151Gln	p.Thr305Ile	p.Ser450Phe		p.Lys126Glu	p.Leu213Ile	p.Gly216Asp	p.His219Pro	p.Val494Met	p.Arg525Gln	p.Val588del	p.Cys641Tyr	p.Asn671Ser
SF3B3 variant origin			*de novo*	*de novo*	*de novo*		*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*
Gender	11M - 12F	1M-1F	N/A	F	M	6M-10F	F	M	F	F	F	M	M	F	F
Last examination (y.m.)			Prenatal Foetus	Deceased at 1 day	Deceased at 20 days		N/A	N/A	6 y 5 m	7 y 10 m	6 y	4 y 10 m	N/A	6 y 5 m	6 y 5 m
Prenatal Findings	10/15 (67%)	3/3 (100%)	yes	yes	yes	7/10 (70%)	N/A	N/A	yes	yes	yes	N/A	N/A	no	yes
Pre or postnatal microcephaly	4/15 (27%)	3/3 (100%)	yes	yes	yes	1/10 (10%)	N/A	N/A	yes	no	N/A	no	N/A	no	no
Intrauterine growth restriction / Small for Gestational Age (<10°)	7/14 (50%)	2/2 (100%)	N/A	yes	yes	5/10 (50%)	N/A	N/A	no	yes	N/A	no	N/A	no	yes
Short Stature	6/13 (46%)		N/A	N/A	N/A	6/11 (54%)	N/A	N/A	yes	yes	no	no	N/A	yes	no
Global Neurodevelopmental delay / Intellectual disability	15/17 (88%)		N/A	N/A	N/A	13/15 (87%)	yes	yes	yes	yes	yes	yes	yes	no	yes
Speech delay	12/13 (92%)		N/A	N/A	N/A	10/11 (91%)	N/A	N/A	yes	yes	yes	yes	N/A	yes	yes
Motor delay	12/13 (92%)		N/A	N/A	N/A	10/11 (91%)	N/A	N/A	yes	yes	no	yes	N/A	yes	yes
Autistic spectrum disorders / autistic findings	11/16 (69%)		N/A	N/A	N/A	6/11 (55%)	N/A	N/A	yes	yes	yes	yes	N/A	no	no
Brain and Spinal Bone Marrow MRI/Ultrasound findings	10/11 (91%)	2/2 (100%)	yes	N/A	yes	8/9 (89%)	N/A	N/A	no	yes	N/A	yes	N/A	yes	yes
Ophtalmological anomalies	6/13 (46%)		N/A	N/A	N/A	6/11 (55%)	N/A	N/A	no	no	yes	yes	N/A	no	yes
ENT involvement	5/11 (45%)		N/A	N/A	N/A	5/9 (55%)	N/A	N/A	yes	no	N/A	yes	N/A	no	no
Genitourinary involvement	4/9 (44%)	1/1 (100%)	N/A	N/A	yes	3/6 (50%)	N/A	N/A	N/A	yes	N/A	no	N/A	N/A	N/A
Cardiac involvement	11/15 (73%)	1/1 (100%)	N/A	N/A	yes	10/12 (83%)	N/A	N/A	yes	no	yes	yes	N/A	yes	yes
Gastrointestinal involvement	6/13 (46%)		N/A	N/A	N/A	5/11 (46%)	N/A	N/A	no	yes	no	no	N/A	no	yes
Skeletal involvement	9/13 (69%)	3/3 (100%)	yes	yes	yes	6/9 (67%)	N/A	N/A	no	yes	no	N/A	N/A	N/A	yes
Skin abnormalities	8/11 (73%)		N/A	N/A	N/A	7/10 (70%)	N/A	N/A	yes	yes	yes	N/A	N/A	yes	yes
OTHER			Atrophic changes in skeletal muscle fibers						Mild hyperestensible joints; frequent upper airways infections; sleep disturbances; auto and hetero aggression	Bone age delay	Severe aggression, sleep disturbance	Unilateral single palmar crease; brachycephaly			Recurrent aspiration pneumonia

Table 1 (continued)

Patient_ID* / Features	TOTAL	MULTISYSTEMIC	1	13	16	2	12	14	8	MILD	21	20	19	23	22
SF3B3 (NM_012426.5): nucleotide change			c.2066C>T	c.2066C>T	c.2066C>T	c.2183G>C	c.2364dupT	c.2506G>A	c.3224C>A		c.88A>G	c.1205A>C	c.1242_1243insCTGGC	c.1390C>T	c.2384A>G
SF3B3 (NP_036558.3): amino acid change			p.Thr689Ile	p.Thr689Ile	p.Thr689Ile	p.Arg728Pro	p.Val789Cysfs*5	p.Ala836Thr	p.Ala1075Asp		p.Ile30Val	p.Asp402Ala	p.Asn417Trpfs*21	p.Arg464Ter	p.Asn795Ser
SF3B3 variant origin			*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*		*de novo*	*de novo*	*de novo*	Not inherited by the mother. Father not tested	*de novo*
Gender	11M - 12F	6M-10F	M	F	F	M	M	F	F	4M-1F	F	M	M	M	M
Last examination (y.m.)			19 y	9 y 5 m	N/A	13 y 10 m	15y	2 y 9 m	19 y 1 m		5 y 7 m	N/A	4 y 11 m	5 y	23 y
Prenatal Findings	10/15 (67%)	7/10 (70%)	yes	yes	N/A	no	no	N/A	yes	0/2 (0%)	N/A	N/A	no	no	N/A
Pre or postnatal microcephaly	4/15 (27%)	1/10 (10%)	no	no	N/A	no	no	N/A	no	0/2 (0%)	N/A	N/A	no	no	N/A
Intrauterine growth restriction / Small for Gestational Age (<10°)	7/14 (50%)	5/10 (50%)	yes	yes	N/A	yes	no	N/A	no	0/2 (0%)	N/A	N/A	no	no	N/A
Short Stature	6/13 (46%)	6/11 (54%)	yes	yes	N/A	yes	no	no	no	0/2 (0%)	N/A	N/A	no	no	N/A
Global Neurodevelopmental delay / Intellectual disability	15/17 (88%)	13/15 (87%)	yes	yes	yes	yes	yes	N/A	no	2/2 (100%)	N/A	N/A	yes	yes	N/A
Speech delay	12/13 (92%)	10/11 (91%)	yes	yes	N/A	yes	yes	N/A	no	2/2 (100%)	N/A	N/A	yes	yes	N/A
Motor delay	12/13 (92%)	10/11 (91%)	yes	yes	N/A	yes	yes	N/A	yes	2/2 (100%)	N/A	N/A	yes	yes	N/A
Autistic spectrum disorders / autistic findings	11/16 (69%)	6/11 (55%)	yes	no	N/A	yes	no	N/A	no	5/5 (100%)	yes	yes	yes	yes	yes
Brain and Spinal Bone Marrow MRI/Ultrasound findings	10/11 (91%)	8/9 (89%)	yes	yes	N/A	yes	N/A	N/A	yes	NA	N/A	N/A	N/A	N/A	N/A
Ophtalmological anomalies	6/13 (46%)	6/11 (55%)	yes	yes	N/A	yes	no	N/A	no	0/2 (0%)	N/A	N/A	no	no	N/A
ENT involvement	5/11 (45%)	5/9 (55%)	yes	NA	N/A	yes	yes	N/A	no	0/2 (0%)	N/A	N/A	no	no	N/A
Genitourinary involvement	4/9 (44%)	3/6 (50%)	yes	NA	N/A	no	no	N/A	yes	0/2 (0%)	N/A	N/A	no	no	N/A
Cardiac involvement	11/15 (73%)	10/12 (83%)	yes	yes	N/A	no	yes	Yes	yes	0/2 (0%)	N/A	N/A	no	no	N/A
Gastrointestinal involvement	6/13 (46%)	5/11 (46%)	yes	no	N/A	no	yes	N/A	yes	1/2 (50%)	N/A	N/A	no	yes	N/A
Skeletal involvement	9/13 (69%)	6/9 (67%)	yes	yes	N/A	no	yes	N/A	yes	0/1 (0%)	N/A	N/A	N/A	no	N/A
Skin abnormalities	8/11 (73%)	7/10 (70%)	no	no	N/A	yes	no	N/A	yes	1/1 (100%)	N/A	N/A	yes	N/A	N/A
OTHER			Relative macrocepahly; complex partial seizures with clonic seizures; recurrent airways infections	Seizure; iris coloboma; neonatal relative macrocephaly		Female distribution of adipose tissue; in the past 9 generalized tonic seizures, stop under treatment; now no seizures any more for 6 y without treatment	Degenerative changes in the neck and tense muscles		Extreme tightness through the lower extremities causing flexion at the knee joints						

N/A Not available

* Patient IDs reflect the order of recruitment and are consistently used throughout the manuscript. In the table, patient are listed according to clinical severity categories (Severe, Multisystemic, Mild) and within each category by the position of the *SF3B3* variant on the nucleotide and aminoacid levels

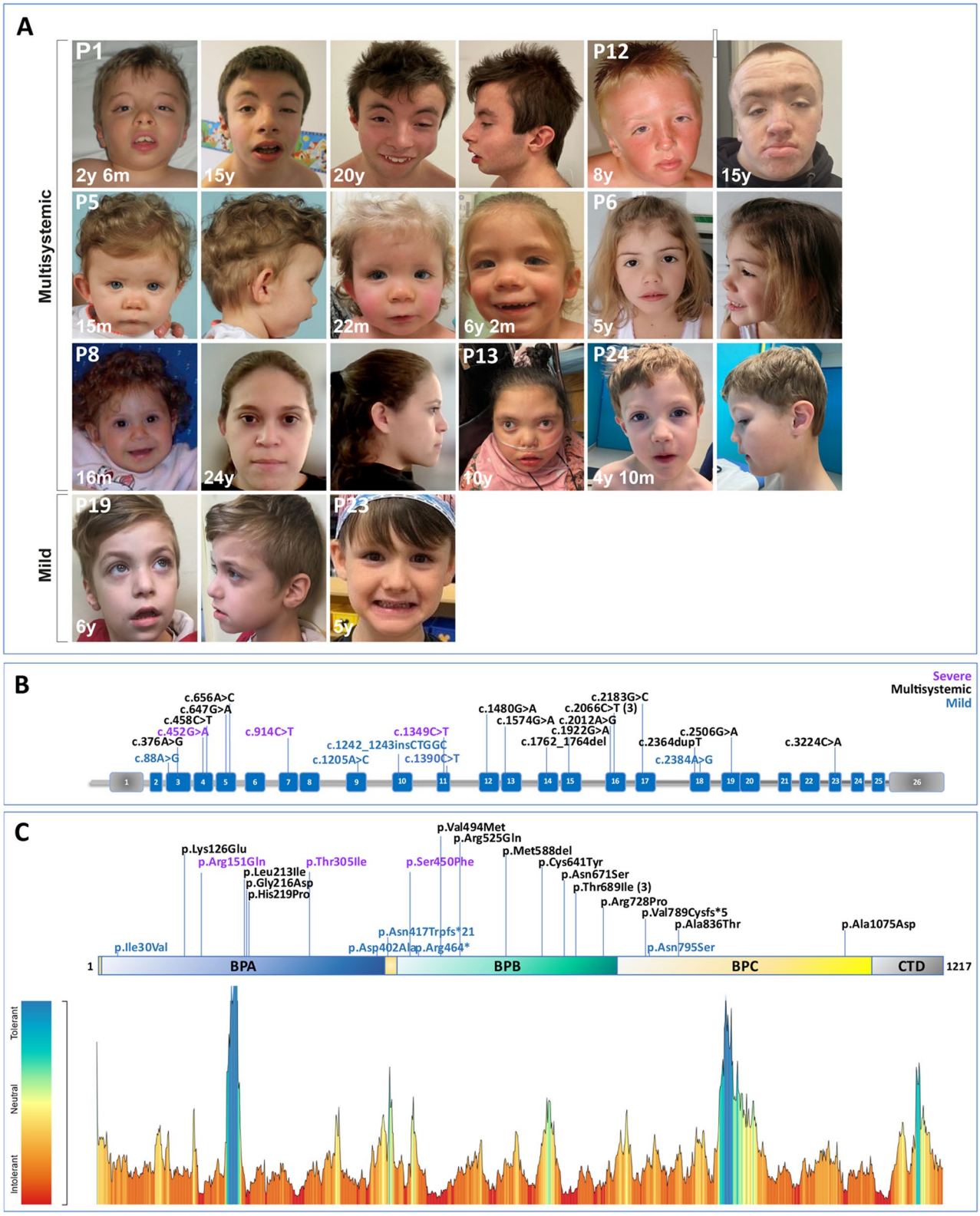

Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Genetic analysis of *SF3B3* variants. (**A**) Facial features of subjects with *de novo SF3B3* variants. Individuals P1 to P13 and P24 present with a multisystemic form of the *SF3B3*-associated disorder. It is noteworthy that these subjects exhibit a number of distinctive characteristics, including light hair colour, a large/high forehead, hypertelorism, deeply set eyes with down slanted palpebral fissures, a typical nose characterised by a depressed nasal bridge with a bulbous nasal tip, a smooth/broad philtrum and tented upper lip vermilion. P19 and P23 present with a milder form with ASD and DD/ID. (**B**) Schematic representation of *SF3B3* (NM_012426). Variants and their positions are shown. Lines and boxes indicate introns and exons, respectively. Exons are numbered with respect to the exon upstream of the first coding exon. (**C**) Schematic representation of the arrangement of the SF3B3 domains (top panel). The protein contains three β-propeller domains: BPA (aa 11–404, blue), BPB (aa 442–772, green) and BPC (aa 1–10, 405–441, 773–1132, yellow), followed by an α-helical C-terminal domain (CTD, aa 1133–1217, grey). The positions of the protein changes are indicated. The tolerance landscape of the *SF3B3* gene is shown in the lower panel (MetaDome web server [38]). The MetaDome analysis revealed that the mutations are located in regions of high intolerance (low missense over synonymous variant count ratio). Variants associated with the severe, multisystemic and mild presentation are shown in violet, black and blue, respectively. The c.2066C>T, p.(Thr689Pro) variant was identified in three unrelated individuals

with ASD and DD/ID always present (100%) but without congenital anomalies or other features (individuals 19–23). Notably, 67% of the individuals exhibited at least one prenatal finding.

The core clinical characteristics observed in individuals with *SF3B3* variants included DD/ID, ranging from mild to severe, and language delay (88% and 92%, respectively). A high proportion of probands (69%) exhibited behavioural features and/or ASD. Motor delay was also frequently observed (92%). Spinal and brain MRI or ultrasound was available for 11 individuals, revealing anomalies in 10 of them (91%). The most frequently observed abnormalities were those affecting the corpus callosum and enlargement of the ventricles. We noted a recognisable pattern of human malformations, including cardiac malformations (i.e., septal defects, PFO, PVS, and tetralogy of Fallot) in 73% of the individuals, genitourinary defects in 44% and skeletal abnormalities in 69%. Additionally, 46% (6 out of 13) of the subjects exhibited short stature. We requested frontal and lateral photographs of each individual, and we obtained images from nine probands with informed consent for publication (Fig. 1A). Moreover, the HPO terms for phenotypic facial abnormalities were collected for two additional subjects. The majority of patients exhibited a similar phenotypic spectrum characterised by a large/high forehead (9/10), hypertelorism (8/11), deeply set eyes (7/10) with downslanting palpebral fissures (6/10), a typical nose characterised by a depressed nasal bridge (10/11) with a bulbous nasal tip (10/11), low-set ears (7/10), a smooth or broad philtrum (7/9) and tented upper lip vermilion (6/10). Another common feature was the horizontal crease on the chin (8/10) (Additional file 2: Table S5). A consistent feature was the colour of the hair, which ranged from light blonde to light brown. In a few cases, paediatric and adult-age photographs of the same individual were available, showing that the facial dysmorphisms became more pronounced with age (Fig. 1A).

In the most severely affected cases (individuals 9–11), the clinical presentation showed a severe microcephaly with a sloping forehead, growth retardation, multiple anomalies and malformation.

Frontal facial photographs of nine probands were analysed with the GestaltMatcher algorithm to investigate the occurrence of facial gestalt clustering in our cohort. The photograph of subject P24 was not available at the time of the analysis. The pairwise ranks of the patients' photographs demonstrated remarkable similarities among seven individuals affected by the multisystemic form of the condition who were grouped in one cluster with a gradient of similarity as indicated by the scores (Additional file 2: Fig S1). Notably, individuals P19 and P23, who manifested the milder clinical phenotype (ASD and DD), clustered away from all the others. Thus, the results, suggested the presence of a specific facial gestalt characteristic for this new syndrome.

Molecular findings

In total, we observed 22 different variants, including 18 missense changes, a one-amino-acid in-frame deletion, one nonsense and two frameshifts (Fig. 1B; Table 1). The missense variant (c.2066C>T, p.(Thr689Ile)) was recurrent in three unrelated individuals. With the exception of individual 23, for whom the paternal sample was unavailable for testing, all changes occurred *de novo*. All but two of the single-nucleotide variants (SNVs) and insertion-deletion (indels) variants were absent from the gnomAD population database [39]; the two variants observed there – c.88A>G (one heterozygote; allele frequency: 6.210e-7) and c.2506G>A (one heterozygote; allele frequency: 6.196e-7) – were extremely rare. Most variants are predicted to be damaging by in silico tools, with an average CADD score of 27.1 (Additional file 1: Table S1A). The deleterious nature of the changes is supported by the high intolerance to both missense (Z score: 9.84) and loss-of-function (LOF) variants (pLI:1; LOUEF = 0.28) of the *SF3B3* gene (gnomAD v.4.1.0 [39]), . The majority of the identified single amino acid changes and the one-amino-acid in-frame deletion were located in two of the three intertwined seven-bladed β-propeller domains (BPA: aa 11–404 and BPB: aa 442–772) of SF3B3 (Fig. 1C upper panel), which together with the third β-propeller domain (BPC (aa 1–10, 405–441, and 773–1132)) contact the α-helical C-terminal domain (CTD; aa 1133–1217), resulting in a fold that is essential for the binding

of interacting partners [45]. The MetaDome analysis revealed their location in regions of high intolerance of the protein (Fig. 1C, bottom panel; Additional file 1: Table S1A) [38]. I-Mutant2.0 and MuPro algorithms predicted a reduction in protein stability for the majority of missense changes (Additional file 1: Table S1B) [40, 41].

We investigated the SF3B3 protein levels in the whole-cell lysates from control and patient-derived fibroblasts of individuals 1, 2, 6 and 8, carrying the variants c.2066C>T p.(Thr689Ile), c.2183G>C p.(Arg728Pro), c.647G>A p.(Gly216Asp) and c.3224C>A p.(Ala1075Asp) respectively, by immunoblotting. In cell lines from all patients, SF3B3 protein levels were significantly decreased by approximately 15–30% compared with controls (Fig. 2A-B).

Because SF3B3 protein levels were reduced, we next assessed whether this decrease originated at the transcript level. qRT-PCR analysis of fibroblast-derived cDNA from individuals 1, 2, 6, and 8 demonstrated comparable *SF3B3* mRNA levels between patients and controls (Fig. 2C). Thus, the observed reduction in SF3B3 protein is not caused by altered transcript abundance or transcript instability but instead likely reflects post-transcriptional mechanisms. We next quantified several additional components of the SF3b complex and one associated protein (SF3B1, SF3B2, SF3B4, SF3B6, PHF5A, and RNF113A). All but one were maintained at control-like levels; SF3B1 was significantly reduced in fibroblasts from patients 1, 2, and 6 (Fig. 2A–B). However, *SF3B1* mRNA abundance was comparable between patient and control samples (Additional file 2: Fig S2), again supporting a post-transcriptional effect.

To assess whether SF3B3 variants affect subcellular localization, we performed cell fractionation followed by western blotting. SF3B3 was detected exclusively in the nuclear fraction in both control and patient fibroblasts (Additional file 2: Fig S3). Subtle changes within nuclear subcompartments were not investigated. Having ruled out major mislocalization, we next examined protein stability. Treatment of fibroblasts with the proteasome inhibitor MG132 resulted in a significant increase in SF3B3 levels in patients 2 and 6, suggesting enhanced proteasome-mediated degradation of the mutant protein. Patient 8 showed a non-significant trend toward increased SF3B3 after MG132 treatment (Additional file 2: Fig S4A-B). These findings indicate that reduced SF3B3 protein levels in some patient fibroblasts are at least partially due to increased proteasomal degradation.

To further examine intrinsic protein stability outside the patient fibroblast context, FLAG-tagged wild-type and mutant SF3B3 constructs carrying the p.(Thr689Ile), p.(Arg728Pro), p.(Gly216Asp) and p.(Ala1075Asp) respectively, were transiently transfected in HEK293T cells together with a construct expressing GFP, used as transfection efficiency control. Western blot analysis after 24 h revealed that mutant proteins were consistently less abundant than the wild-type protein, supporting the interpretation that the amino acid substitutions themselves can compromise SF3B3 stability. GAPDH was used as loading control (Additional file 2: Fig S5).

To evaluate *SF3B3* contribution to neurodevelopmental traits, we performed bioinformatic analyses to assess its expression trajectory along neocortex development, employing the developing human brain transcriptome dataset [73]. We observed that *SF3B3* exhibits the highest expression during the first 24 postconceptional weeks, followed by a rapid decline before birth until 4 months of age when it reaches a relatively constant level in late childhood and adult stages (Additional file 2: Fig S6). The same analysis, performed for other SF3b components and *RNF113A*, showed that the expression of *SF3B1-4* is higher in the prenatal stages and that *SF3B4* has a time-dependent expression profile similar to *SF3B3*. The expression of *RNF113A* is similar in the prenatal, postnatal and later stages (Additional file 2: Fig S6).

All-atom simulations

To elucidate the impact of *SF3B3* missense variants and to gain insight into the specificity and severity of pathogenic mechanisms associated with them, we used all-atom molecular dynamic simulations. To this end we have built three models of the SF3b complex (Additional file 2: Fig S7). We mapped all amino acid changes onto the structure of the SF3b core splicing factor, which contained the SF3B1 protein in its closed conformation, as observed in the B^{act} state of the spliceosome (Fig. 3A). Most of the variants were located on the periphery/surface of the SF3B3 protein, where they presumably affected SF3B3 interactions with other proteins within the spliceosome complex, thereby altering its function.

We then considered SF3b complex with SF3B3, either in the wild type (WT) form or in the presence of six variants: N671S, T689I, R728P, G216D, H219P and A1075D. The latter are designated in this section employing single amino acid nomenclature. Although none of the aforementioned variants was directly involved in the interactions with RNA strands, they were selected because all resided at the interface with different proteins forming the SF3b complex. We then performed 500 ns of all atom molecular dynamics simulation for each model. A detailed inspection of the system in the presence of the N671S, T689I and R728P variants (Fig. 3B-C) revealed that these variants disrupt the interaction between SF3B3 and RNF113A, triggering RNF113A dissociation or significantly altering its positioning and/or its conformation (Fig. 3B). Briefly, the N671 residue was located at the interface between the SF3B3 and RNF113A proteins and its amide moiety was directly involved in a hydrogen

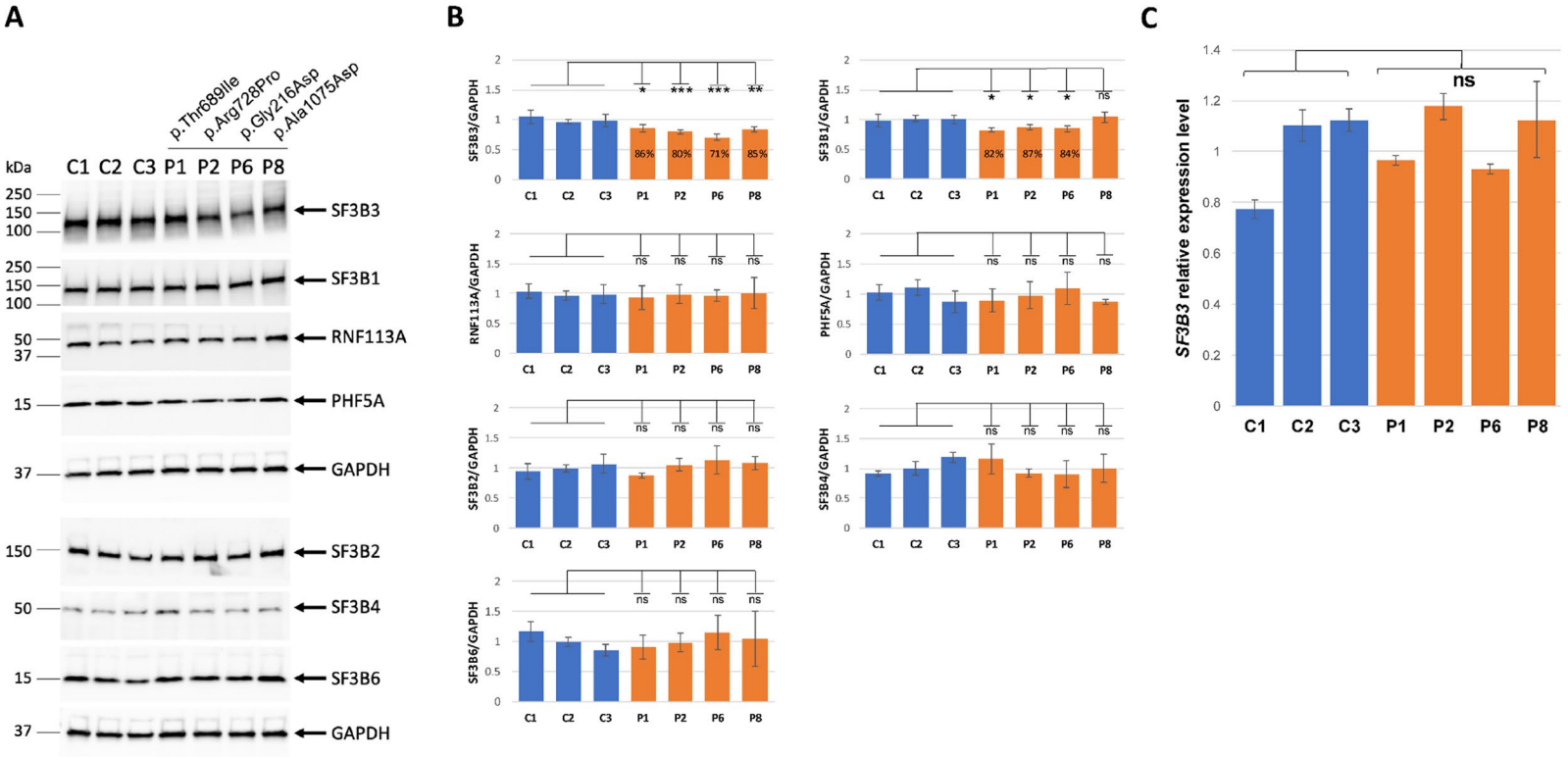

Fig. 2 *SF3B3* mRNA expression and protein levels of SF3B1-4, SF3B6, PHF5A, and RNF113A in patient-derived and control fibroblasts. (**A**) Representative immunoblots of whole cell lysates from subject (P1, P2, P6 and P8) and control (C1-3) fibroblast cultures. Levels of the investigated proteins were monitored with the indicated antibodies. Band intensity was quantified using ChemiDoc MP imaging system. SF3B3, SF3B1-2, SF3B4, SF3B6, PHF5A, and RNF113A protein levels were normalized to GAPDH used as loading control. (**B**) Histograms showing the relative amount of SF3B3 (mean ± SD of four experiments), SF3B1-2 and 4 (mean ± SD of three experiments), SF3B6 and RNF113A (mean ± SD of five experiments) and PHF5A (mean ± SD of 6 experiments); t test (two-tailed, homoscedastic): *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. (**C**) Histogram showing the expression level of *SF3B3* mRNA, assessed by qRT-PCR, in the patients' fibroblast cell lines compared with controls ($n = 3$; t test (two-tailed, homoscedastic); ns: non-significant). *GAPDH* was used as the housekeeping gene for normalization

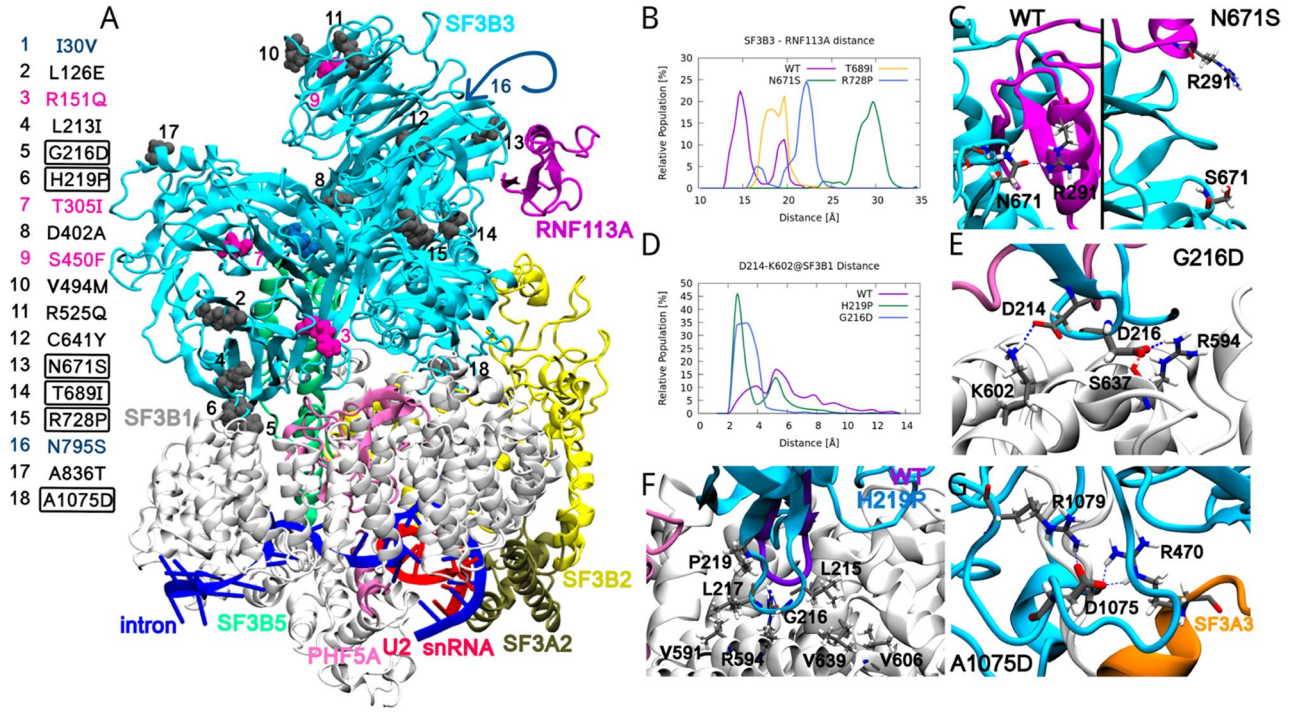

Fig. 3 Binding mode of SF3B3 variants in the SF3b model. (**A**) Overview of the SF3b model (corresponding to the B^{act} state of the spliceosome, bearing the SF3B1 protein in the closed conformation). The SF3B1, SF3B2, SF3B3, SF3B5, SF3A2, PHF5A and RNF113A proteins are shown as white, yellow, cyan, green, tan, pink and magenta cartoons, respectively. The pre-mRNA strand and U2 snRNA are represented as blue and red cartoons, respectively. The single amino acid variants numbered 1 to 18, are shown in van der Waals representation coloured according to severity: purple for the lethal, black for syndromic and blue for non-syndromic ASD DD/ID. SF3B3 variants that were investigated via MD simulations are encircled by black boxes. (**B**) Distribution of the distances between the centre of mass of the RNF113A protein and the RNF113A-binding interface of SF3B3 (formed by residues 629–636, 669–696, 1108–1114, 1200–1217) for the wild type (WT) SF3B3 protein and in the presence of three variants (N671S, T689I, R728P). (**C**) Close-up of the interactions established by the WT N671 and the mutated S671 residue. (**D**) Distance distribution of salt bridge interactions between the D241@SF3B3 and K602@SF3B1 in WT, G216D and H219P SF3B3 variants. (**E**) Close-up of the interactions formed by D214 and D216 in the G216D-containing SF3B3 model. The amino acids are indicated with single letter nomenclature. (**F**) Close-up of the conformational change of the SF3B3 loop (residues 214–219) upon insertion of H219P variant. The loop established new interactions with SF3B1. (**G**) Close-up of the interactions formed by D1075 after the SF3B3 A1075D variant was introduced

bond with the R291 residue of RNF113A. When mutated to S671, the hydroxyl group of the serine residue disrupted the hydrogen bond to a nearby SF3B3 loop (Fig. 3C). Similarly, when the T689 residue, located at the SF3B3/RNF113A interface, was mutated to I689 its interactions with the RNF113A loop were lost (Additional file 2: Fig S8). Conversely, the R728 residue, buried inside the SF3B3 core, formed an intramolecular salt bridge interaction with E1180 (Additional file 2: Fig S8). This interaction vanished upon the R728P mutation, thus triggering the remodelling of E1180 to form an alternative salt bridge interaction with the R690 residue at the interacting interface with RNF113A (Additional file 2: Fig S8). The G216D and H219P mutations were located in a loop formed by residues 214–219, which directly interacted with the SF3B1 protein. A comparison of different cryo-EM structures (PDB: 6EN4 and 6FF4) showed that the SF3B3 changed position with respect to the SF3B1 protein in the spliceosome A and B^{act} state. Indeed, the solenoidal structure of SF3B1 underwent a conformational transition from the open to the closed conformation. The loop containing the G216 and H219 residues contributed to locking the different positions. Indeed, after the G216D and H219P variants were introduced, the interaction of this loop with SF3B1 in the open conformation was strengthened. In particular, the G216D variant established a salt bridge between D214@SF3B3 and K602@SF3B1 and, in addition, a new salt bridge interaction was formed between the mutated D216@SF3B3 and R594@SF3B1 (Fig. 3D-E). The H219P mutation, in contrast, had a more subtle effect; the rigid, unique backbone properties of P219 caused changes in the conformation of the loop (Fig. 3F), thus allowing other residues in the loop to interact with SF3B1. The strengthening of the interaction between SF3B1 and SF3B3 was supported by a calculation of the binding free energy differences of the SF3B3 loop 214–219 to the SF3B1 (Additional file 2: Table S6), which became more negative (corresponding to a stronger interaction) by 16.4 kcal/mol and 12.1 kcal/mol, for G216D and H219P variants, respectively. We

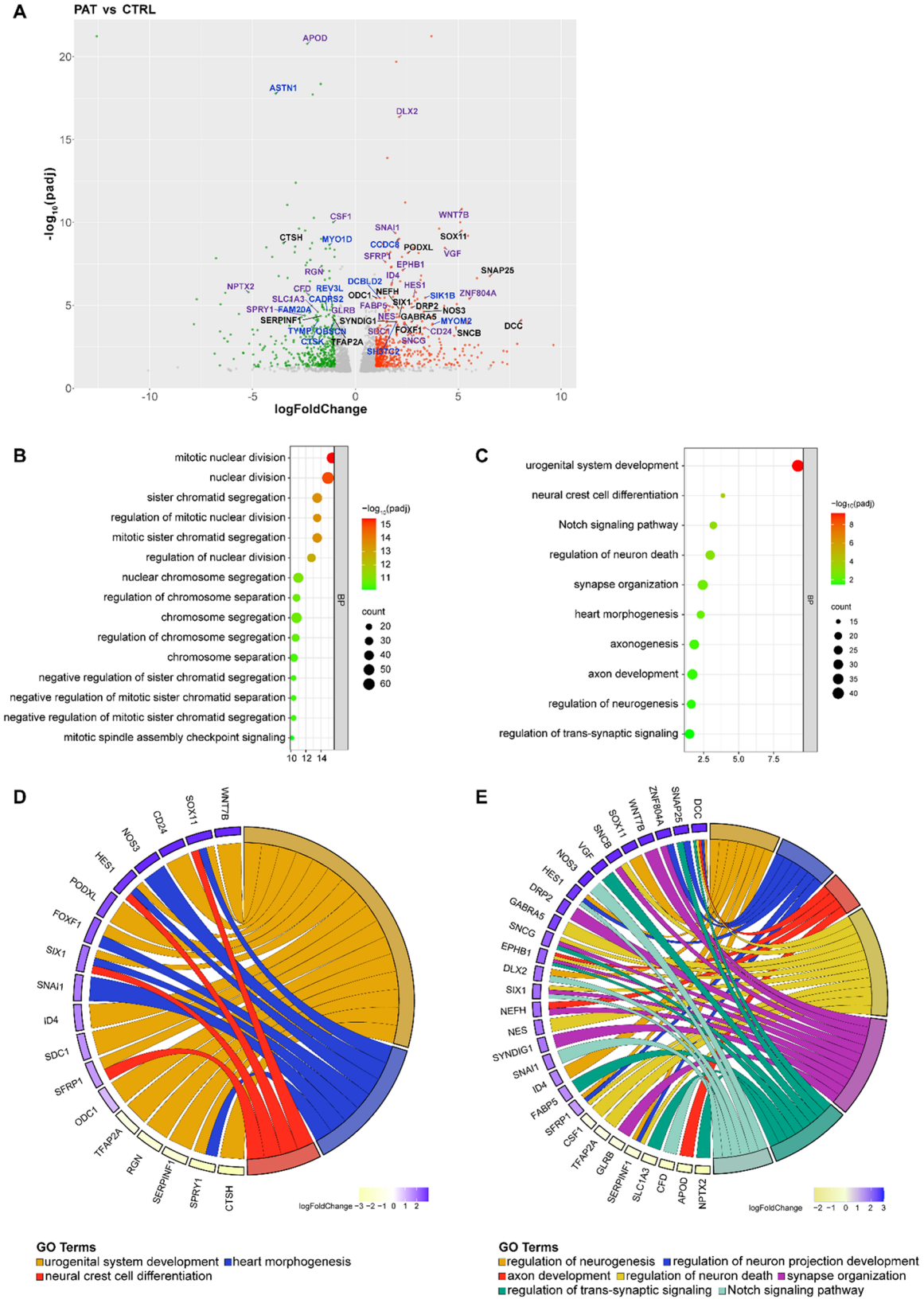

Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Identification of differentially expressed genes (DEGs) in the transcriptome sequencing data of fibroblasts from subjects 1, 2, and 6 in comparison with control fibroblasts. (**A**) Volcano plot of DEGs in *SF3B3* patient-derived fibroblasts compared to controls. A total of 887 genes exhibited significant differential expression between the patient and control groups, comprising 494 genes that were upregulated (red dots) and 393 genes that were downregulated (green dots). The filtering criteria applied were an adjusted p-value (padj) less than 0.05 and a cutoff of fold changes (FC) > 2 for upregulated transcripts and FC < 0.5 for downregulated transcripts. The grey dots represent genes that did not meet the criteria for statistical significance. Of the top 100 significantly upregulated and downregulated genes (selected as arbitrary cutoff), 38 DEGs, highlighted in black and violet, are involved in Biological Processes (BPs) relevant to the *SF3B3*-associated clinical presentation. These include urogenital system development, heart morphogenesis, and neural crest differentiation. Furthermore, 29 of the top 100 significantly upregulated and downregulated genes are disease-causing genes as classified in the HGMD database [35], related to conditions with clinical features similar to those observed in individuals with *SF3B3* variants (highlighted in blue and violet). The genes that are common to both the strategies used to interpret the biological relevance of the DEGs identified, are indicated in violet. (**B**) Bubble plot illustrating the top 15 Gene Ontology (GO) BP significantly enriched in *SF3B3* patients-derived fibroblasts in comparison to controls. These are related to cell cycle regulation. (**C**) Bubble plot illustrating significant over-represented BPs related to the *SF3B3*-associated clinical presentation in *SF3B3* patients-derived fibroblasts. The colour of the nodes displayed in a gradient from red to green, corresponds to the decreasing order of the padj value. The size of each node is proportional to the number of gene counts. (**D-E**) GO Chord plots: representation of the relationship between the 38 significant DEGs, listed in Table 2, and their associated GO terms. The left of the plots shows the genes contributing to the enrichment, arranged in order of their logFC, which is displayed in descending intensity of blue squares for the upregulated genes and yellow squares for the downregulated ones. The genes are linked to their assigned terms via coloured ribbons chord

also investigated the effect of the A1075D variant, which enabled the disordered loop formed by residues 1069–1079 to acquire a partial α-helical character (Additional file 2: Fig S9). Due to the formation of an intra-loop hydrogen bond with R1079@SF3B3 in the open state, this mutation additionally led to the formation of an inter-protein hydrogen bond with R470@SF3A3 (Fig. 3G), potentially increasing the interaction of SF3A3 with the spliceosome.

Transcriptome analysis

Given that our data indicated impaired binding of mutant SF3B3 proteins to other spliceosome components, we investigated the transcriptome profiles of individuals carrying *SF3B3* variants. We performed RNA sequencing on patient-derived fibroblast cell lines from P1 (c.2066C>T, p.(Thr689Ile)), P2 (c.2183G>C, p.(Arg728Pro)), and P6 (c.647G>A, p.(Gly216Asp)), which were the only patient samples available at the time the experiments were conducted, as well as three unrelated controls.

We first compared gene expression distributions across samples using boxplots, which showed similar gene density and number, and correlation analysis revealed a coefficient (R^2) ≥ 0.9 (Additional file 2: Fig S10A-B). Principal component analysis (PCA) indicated that controls and patient samples clustered separately, with each group showing similar pattern of variability (Additional file 2: Fig S10C). Differential expression analysis identified 887 genes significantly altered in patients compared to controls, with 494 upregulated and 393 downregulated (Fig. 4A).

To understand the biological significance of these DEGs, we performed Gene Ontology (GO) enrichment analysis. The top 15 significantly enriched biological processes (BP) were related to cell cycle regulation (Fig. 4B, Additional file 1: Table S7A), consistent with KEGG pathway analysis, which highlighted the cell cycle (hsa04110) as highly enriched (Additional file 1: Table S7B). Moreover, the analysis revealed over-representation of BPs related to *SF3B3*-associated clinical features, including urogenital system development, heart morphogenesis, neural crest differentiation, neurogenesis, neuron projection and axon development, neuron death, and synapse organization (Fig. 4C, Additional file 1: Table S7C). Notch signaling and regulation of trans-synaptic signaling were also significantly enriched (Fig. 4C, Additional file 1: Table S7C).

Among the top 100 significantly up- and downregulated genes (arbitrary cutoff), 38 DEGs encoded for proteins involved in at least one of these processes, including nine transcription factors or regulators: DLX2, SOX11, SNAI1, ID4, HES1, ZNF804A, SIX1, FOXF1, and TFAP2A (Fig. 4A, D-E; Table 2, Additional file 1: Table S7C). Notably, *DLX2*, *WNT7B*, *SOX11*, and *SNAI1*, which are implicated in early heart, nervous system, and urogenital development, ranked among the top ten most upregulated genes. DLX2 and SNAI1 are also components of the Notch signaling pathway (Fig. 4A, D-E; Table 2, Additional file 1: Table S7C).

Next, we analyzed *SF3B3* and neurodevelopmentally relevant DEGs using the developing human brain RNA dataset [73]. Interestingly, trajectories and differences in the incline, decline and shape were observed (Additional file 2: Fig S11A-E). It was observed that *WNT7B* and *SOX11* exhibited higher expression in the prenatal period, which decreased rapidly between 24 weeks gestation and 4 months after birth as was also the case for *SF3B3*. Conversely, expression of *VGF*, *SNAP25* and *SERPINF1* (also known as *PEDF*) exhibited a rapid increase trajectory between the same time frame (Additional file 2: Fig S11F).

To further dig out for clinically relevant DEGs, we adopted a second strategy. We systematically searched for disease-causing genes, as classified in the Human Genome Mutation Database (HGMD) Professional as selection criteria [35], assuming that there might be

Table 2 Genes within the top 100 significantly up- and downregulated in patients' samples, involved in biological processes related to the *SF3B3*-associated clinical presentation and overrepresented in patients-derived fibroblasts

Up-regulated			
Gene symbol	**Description**	**Fold Change**	**padj**
DLX2	Distal-Less Homeobox 2	4.37	4.31E-17
WNT7B	Wnt Family Member 7B	36.08	1.54E-11
SOX11	SRY-Box Transcription Factor 11	35.86	3.11E-10
SNAI1	Snail Family Transcriptional Repressor 1	3.85	4.33E-10
VGF	VGF Nerve Growth Factor Inducible	20.35	3.37E-09
PODXL	Podocalyxin Like	5.86	6.84E-09
SFRP1	Secreted Frizzled Related Protein 1	2.75	2.35E-08
EPHB1	EPH Receptor B1	4.95	7.41E-08
SNAP25	Synaptosome Associated Protein 25	92.67	1.69E-07
ID4	Inhibitor Of DNA Binding 4	3.43	4.62E-07
HES1	Hes Family BHLH Transcription Factor 1	7.14	2.91E-06
ODC1	Ornithine Decarboxylase 1	2.04	3.63E-06
ZNF804A	Zinc Finger Protein 804A	46.40	3.96E-06
FABP5	Fatty Acid Binding Protein 5	3.20	1.12E-05
DRP2	Dystrophin Related Protein 2	6.66	1.34E-05
NEFH	Neurofilament Heavy Chain	4.31	1.49E-05
NOS3	Nitric oxide synthase 3	10.00	2.40E-05
NES	Nestin	4.10	3.29E-05
SIX1	SIX Homeobox 1	4.36	5.01E-05
DCC	DCC Netrin 1 Receptor	268.89	7.76E-05
GABRA5	Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha5	6.62	9.91E-05
SYNDIG1	Synapse Differentiation Inducing 1	3.95	9.91E-05
FOXF1	Forkhead Box F1	5.32	1.25E-04
SNCG	Synuclein Gamma	5.12	1.32E-04
SDC1	Syndecan 1	3.00	1.59E-04
SNCB	Synuclein Beta	29.72	2.09E-04
CD24	CD24 Molecule	10.44	2.16E-04
Down-regulated			
APOD	Apolipoprotein D	0.20	1.63E-21
CSF1	Colony Stimulating Factor 1	0.47	9.63E-11
CTSH	Cathepsin H	0.09	1.73E-09
RGN	Regucalcin	0.32	4.23E-08
NPTX2	Neuronal pentraxin 2	0.03	1.60E-06
CFD	Complement factor D	0.21	2.25E-06
SLC1A3	Solute carrier family 1 member 3	0.28	2.46E-05
GLRB	Glycine Receptor Beta	0.41	3.73E-05
SPRY1	Ornithine Decarboxylase 1	0.18	4.12E-05
SERPINF1	Serpin Family F Member 1	0.31	4.37E-05
TFAP2A	Transcription Factor AP-2 Alpha	0.47	5.19E-05

common features between the genetic regulatory networks of different diseases with an overlapping clinical presentation. Focusing on the top 100 DEGs, we identified 29 disease-causing genes related to conditions with clinical features similar to those presented by the *SF3B3* patients (Fig. 4A; Table 3). Thus, combining the two strategies, we identified 16 clinically and biologically relevant DEGs, of which 13 were significantly up-regulated and 3 down-regulated (Fig. 4A and Additional file 2: Fig S12).

We also leveraged the RNA-seq dataset from Pervolaraki et al., which charts human heart development across 9, 12 16 weeks of gestation [78–80] which represent the timeline of the human fetal heart development that is reasonably affected in the *SF3B3* patients presenting with cardiac malformations. Comparing the 887 DEGs in SF3B3 patient fibroblasts with genes differentially expressed at 12 and 16 weeks gestation revealed 64 overlapping genes, 20 of which were among the top 100 DEGs in SF3B3 samples, including *SOX11*, *SNAI1*, and *ID4* (Table 4).

Targeted analysis of spliceosomal components, based on relevant GO terms, did not reveal any significant DEGs, suggesting no compensatory transcriptional effects. In particular, mRNA levels of *SF3B3* and *SF3B1* were similar between patients and controls, confirming qRT-PCR results (Fig. 2C, Additional file 2: Fig S2), although protein levels were altered. Levels of *SF3B2-6*, *PHF5A*, and *RNF113A* were consistent across patient and control samples (Additional file 1: Table S7D).

Finally, qRT-PCR validation on a subset of DEGs (*WNT7B*, *SOX11*, *SNAI1*, *APOD*, *SERPINF1*, *CSF1*) in patients 1, 2, 6, and 8 confirmed RNA-seq results (Additional file 2: Fig S13).

Alternative splicing analysis

We evaluated RNA sequencing data to assess global and individual patient alterations in alternative splicing (AS) events.

Global analysis including all three patients versus controls did not reveal statistically significant differences for most event types (Additional file 2: Fig S14). Given that the patients carry distinct variants, we next analyzed each individually versus controls. Z-score analysis indicated that retained intron (RI) events showed the largest deviations from control values. Two patients had positive Z-scores (~1.5–1.9), whereas one patient showed negative Z-scores across all event types, indicating patient-specific differences in splicing, particularly for RI events (Additional file 2: Fig S15).

Analysis of AS events revealed 332 significantly enriched. Skipped exon (SE) events are the predominant class (~72% of all events), a trend maintained even among significantly altered events (72% of significant events). Notably, RI events were overrepresented among significantly altered events (8% vs 5% of all events), while mutually exclusive exon (MXE) events were relatively underrepresented (7% vs 11%). Proportions of alternative 3' splice site (A3SS) and alternative 5' splice site (A5SS) events remained largely unchanged (6–7%) (Additional file 2: Fig S16). Chi-square tests comparing the proportions of total versus significantly altered events confirmed significant overrepresentation for RI ($p = 9.82E-03$) and underrepresentation for MXE ($p = 4.11E-02$), but not for SE, A3SS, or A5SS ($p > 0.05$) (Additional file 1: Table S8A). These results indicate distinct splicing signatures potentially associated with the patient condition.

Restricting the global analysis to patients 1 and 2, who carry functionally related variants, revealed a trend toward enrichment of RI events ($p = 0.083$), consistent with the individual Z-score analysis and suggesting a shared mechanistic effect (Additional file 2: Fig S17).

Filtered rMATS events (FDR < 0.05, |IncLevelDifference| ≥ 0.2) generate a list of genes associated with significantly altered splicing. Among these, 26 RI events were significantly enriched in patients compared to controls. Sashimi plots illustrating these RI events are shown in Additional file 2: Fig S18.

Gene Ontology enrichment analysis of significantly altered events highlighted cellular components (Additional file 2: Fig S19, Additional file 1: Table S8B), indicating potential structural or compartment-specific involvement. Targeted analysis of spliceosomal components, based on relevant GO terms, identified five genes affected by significant altered alternative splicing (Additional file 2: Table S9 and Additional file 2: Fig S20).

KEGG pathway analysis highlighted two pathways (Additional file 1: Table S8C), potentially related to the observed phenotype, although further investigation is required.

Cell cycle analysis

To further investigate the cellular consequences of the identified *SF3B3* variants, we analysed cell cycle distribution in four patient-derived fibroblast lines using flow cytometry. The analysis revealed significant alterations in cell cycle distribution in patients 1 and 8, carrying the variants c.2066C>T, p.(Thr689Ile) and c.3224C>A, p.(Ala1075Asp), respectively. Fibroblasts from patient 1 showed a significant accumulation of cells in the G0/G1 phase (79.9% vs. ~73% in controls; $p < 0.05$), accompanied by a clear reduction in the S-phase population (7.6% vs. ~11%). In contrast, fibroblasts from patient 8 exhibited a highly significant increase in the G2/M population (20.5% vs. ~14–16% in controls; $p < 0.01$), suggesting impaired progression through the G2/M checkpoint (Additional file 2: Fig S21). The remaining two patient lines did not show significant deviations from control distributions. Together, these findings indicate that SF3B3 variants

Table 3 Disease-causing genes (HGMD) [35] within the top 100 significantly up- and downregulated genes in the *SF3B3* patient-derived fibroblasts. The conditions associated with these genes, resulting in an overlapping clinical presentation to the *SF3B3*-syndrome, are indicated. AR: autosomal recessive; AD: autosomal dominant; XL: X-linked

Up-regulated				
Gene symbol	**Description**	**Fold Change**	**padj**	**Inheritance, disease, reference**
CCDC8	Coiled-Coil Domain Containing 8	4.36	9.55E-10	AR, 3M syndrome, PMID: 21737058
SOX11	SRY-Box Transcription Factor 11	35.86	3.11E-10	AD, Coffin Siris/ID, PMID: 24886874
PODXL	Podocalyxin Like	5.86	6.84E-09	AD, focal segmental glomerulosclerosis, PMID: 30523047
SNAP25	Synaptosome Associated Protein 25	92.67	1.69E-07	AD, Epilepsy and intellectual disability, PMID: 25003006
SIK1B	Salt Inducible Kinase 1B	10.53	3.46E-06	AD, Developmental epilepsy, PMID: 25839329
DCBLD2	Discoidin, CUB And LCCL Domain Containing 2	2.28	3.61E-06	AR, Intellectual disability syndrome, PMID: 34145321
ODC1	Ornithine Decarboxylase 1	2.04	3.63E-06	AD, Syndromic neurodevelopmental disorder, PMID: 30475435
DRP2	Dystrophin Related Protein 2	6.66	1.34E-05	XL, Charcot-Marie-Tooth neuropathy, PMID: 26227883
NEFH	Neurofilament Heavy Chain	4.31	1.49E-05	AD, Charcot-Marie-Tooth neuropathy, PMID: 32780247
NOS3	Nitric Oxide Synthase 3	9.99	2.40E-05	AD, Primary ovarian insufficiency, PMID: 34480478
SIX1	SIX Homeobox 1	4.36	5.01E-05	AD, Branchio-oto-renal syndrome, PMID: 15141091
DCC	DCC Netrin 1 Receptor	268.89	7.76E-05	AD, Corpus callosum agenesis, PMID: 28250454
GABRA5	Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha5	6.62	9.91E-05	AD, Seizure, PMID: 30033060
SH3TC2	SH3 Domain And Tetratricopeptide Repeats 2	3.98	9.91E-05	AR, Charcot-Marie-Tooth neuropathy, PMID: 20220177
SYNDIG1	Synapse Differentiation Inducing 1	3.95	9.91E-05	AD, Lower plasma triglyceride level, PMID: 26934567
FOXF1	Forkhead Box F1	5.32	1.25E-04	AD, alveolar capillary dysplasia and other malformations, PMID: 19500772
MYOM2	Myomesin 2	13.32	1.48E-04	AD, Conotruncal heart defects, PMID: 35872890; AD, Tetralogy of Fallot PMID: 35872890
SNCB	Synuclein Beta	29.72	2.09E-04	AD, Dementia with Lewy bodies, PMID: 15365127
Down-regulated				
ASTN1	Astrotactin 1	0.07	1.65E-18	AD, Autism/CC anomalies, PMID: 24381
CTSH	Cathepsin H	0.09	1.73E-09	AR, Severe myopia, PMID: 23830514
MYO1D	Myosin ID	0.42	2.23E-09	AD, Visceral heterotaxy, PMID: 34589502
REV3L	REV3 Like, DNA Directed Polymerase Zeta Catalytic Subunit	0.46	9.19E-06	AR, Syndromic DD, PMID: 33474647
CADPS2	Calcium Dependent Secretion Activator 2	0.38	1.93E-05	AD, Autism, PMID: 24737869
FAM20A	FAM20A Golgi Associated Secretory Pathway Pseudokinase	0.21	2.78E-05	AR, Enamel-renal syndrome, PMID: 23468644
SERPINF1	Serpin Family F Member 1	0.31	4.37E-05	AR, Osteogenesis imperfecta, PMID: 21353196
TFAP2A	Transcription Factor AP-2 Alpha	0.47	5.19E-05	AD, Branchio-oculo-facial syndrome, PMID: 18423521
TYMP	Thymidine Phosphorylase	0.28	8.30E-05	AR, Mitochondrial neurogastrointestinal encephalomyopathy, PMID: 9924029
OBSCN	Obscurin, Cytoskeletal Calmodulin And Titin-Interacting RhoGEF	0.48	1.12E-04	AD, Cardiomyopathy, PMID: 34601892
CTSK	Cathepsin K	0.27	2.36E-04	AR, Pycnodysostosis, PMID: 8703060

can perturb cell cycle progression in a patient-specific manner, either by promoting G0/G1 accumulation or by delaying progression through G2/M, consistent with the transcriptomic evidence of dysregulated cell cycle control (Additional file 1: Table S7B).

Discussion

The present study demonstrates that heterozygous germline variants in *SF3B3* cause a novel syndrome. The *SF3B3*-associated clinical features are consistent with those described for other spliceosomopathies, with the

Table 4 Genes, within the top 100 significantly up- and downregulated in *SF3B3* patient-derived fibroblasts, which are also differentially expressed at different stages of human heart development. w9: 9 weeks gestational age; w12: 12 weeks gestational age; w16: 16 weeks gestational age

Up-regulated				
Gene symbol	**Description**	**Fold Change**	**padj**	**differential expression during human heart development**
SOX11	SRY-box 11	35.86	3.11E-10	down-regulated in w16 vs w9
SNAI1	snail family transcriptional repressor 1	3.85	4.33E-10	down-regulated in w12 and w16 vs w9
FSD1	fibronectin type III and SPRY domain containing 1	3.31	5.06E-08	down-regulated in w16 vs w9
ID4	inhibitor of DNA binding 4, HLH protein	3.43	4.62E-07	down-regulated in w16 vs w9
B4GALNT4	beta-1,4-N-acetyl-galactosaminyltransferase 4	3.84	1.91E-06	down-regulated in w12 and w16 vs w9
NEFH	neurofilament heavy polypeptide	4.31	1.49E-05	down-regulated in w12 and w16 vs w9
ADGRL4	adhesion G protein-coupled receptor L4	2.91	3.93E-05	up-regulated in w16 vs w9
SH3TC2	SH3 domain and tetratricopeptide repeats 2	3.98	9.91E-05	up-regulated in w12 vs w9
MYOM2	myomesin 2	13.32	1.48E-04	up-regulated in w16 vs w9
Down-regulated				
HLA-C	major histocompatibility complex, class I, C	0.32	9.58E-10	up-regulated in w16 vs w9
CBX7	chromobox 7	0.42	3.66E-09	up-regulated in w16 vs w9
TXNIP	thioredoxin interacting protein	0.22	3.71E-09	up-regulated in w16 vs w9
DKK2	dickkopf WNT signaling pathway inhibitor 2	0.22	6.15E-09	down-regulated in w12 vs w9
COL23A1	collagen type XXIII alpha 1 chain	0.01	1.42E-08	down-regulated in w16 vs w9
NEAT1	nuclear paraspeckle assembly transcript 1	0.34	8.18E-08	up-regulated in w16 vs w9
CFB	complement factor B	0.33	1.12E-05	up-regulated in w16 vs w9
AKAP6	A-kinase anchoring protein 6	0.25	5.86E-05	up-regulated in w16 vs w9
PODN	podocan	0.28	6.59E-05	up-regulated in w12 vs w9
SLC27A1	solute carrier family 27 member 1	0.45	1.65E-04	up-regulated in w16 vs w9
CTSK	cathepsin K	0.27	2.36E-04	up-regulated in w16 vs w9

majority of affected organ systems being ectodermal-derived nervous system and craniofacial skeleton, as well as mesodermal-derived ones. In particular, patients with *SF3B3* variants exhibited a consistent phenotype, characterised by ASD, DD or ID, language and motor delay, and multiple congenital defects, including cardiac and urogenital malformations. Deep-phenotyping highlighted the existence of three levels of severity in a broad continuous spectrum ranging from the most severe form characterised by microcephaly with sloping forehead, intrauterine growth retardation, major malformations and early death after birth, to a multisystemic form, presenting with mild to severe intellectual disability, along with abnormal behaviour, congenital malformations and typical facial gestalt. The third form, which is milder, is associated with ASD without malformations (Fig. 5A). Patients affected from the severe form deeply differ from the others in terms of clinical picture. In this regard, the severe microcephaly, jointly with the distinctive facies (resembling a form of microcephalic dwarfism) of these individuals, appears to delineate a condition of marked severity.

SF3B3 encodes the splicing factor 3B subunit 3, a component of the U2 snRNP complex of the spliceosome. This complex is responsible for accurate splicing by removing noncoding sequences (introns) and ligating coding sequences (exons).

It has been demonstrated that mutations in more than two dozen of the components of this machinery cause human diseases known as spliceosomopathies, which result in a severely compromised accuracy of splicing and altered gene expression [7, 8, 81–84]. Although pre-mRNA splicing is a global process, the associated disorders tend to affect specific cell types and the derived tissues due to the tissue specificity of several spliceosome components [7, 8, 72, 73]. The majority of spliceosomopathies, including those linked to variants in key components of the SF3b complex PHF5A, SF3B2 and SF3B4, exhibit craniofacial abnormalities [8, 14, 18, 85]. The *SF3B3* individuals who presented with the multisystemic form shared several craniofacial features, including a large/high forehead, hypertelorism, deeply set eyes with downslanting palpebral fissures, a typical nose characterised by a depressed nasal bridge with a bulbous nasal tip, low-set ears, a smooth/broad philtrum, a tented upper lip vermilion and a horizontal crease on the chin. Furthermore, all the most severe cases presented with microcephaly. As the recognition of facial features may be challenging and ultimately constrained by individual human expertise, we employed the GestaltMatcher (GM), a tool that was effective in reaching a correct diagnosis of ultra-rare syndromes [29, 86]. The computer-assisted analysis of the facial images of nine individuals revealed that the *SF3B3* patients with the multisystemic

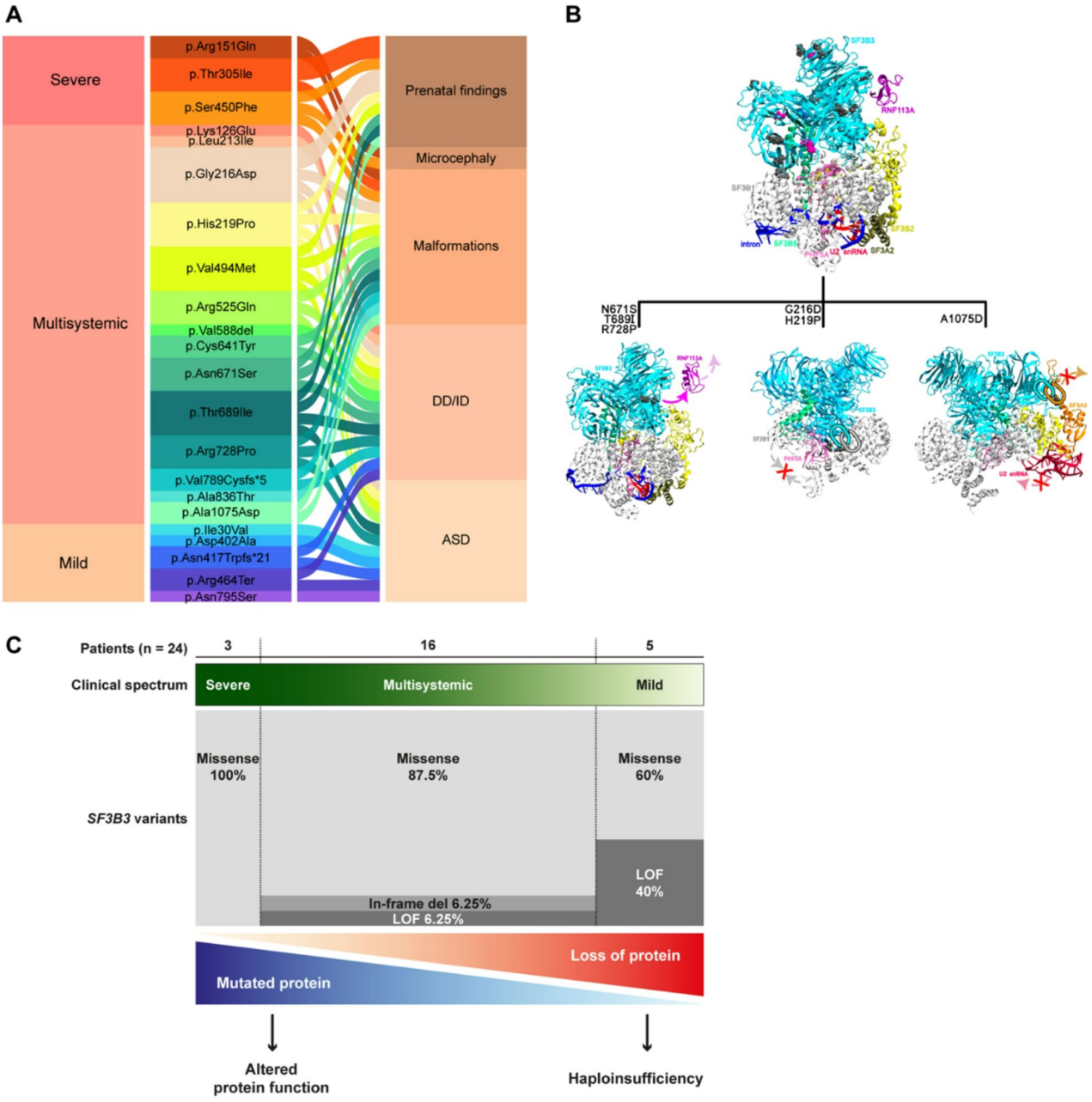

Fig. 5 Clinical spectrum of the *SF3B3*-associated condition: consequences of genetic variants and pathogenic mechanisms. (**A**) The alluvial plot illustrates, from left to right, the clinical presentation of the condition, which ranges from a severe form, through a multisystemic form and ending with a milder form (column 1); SF3B3 protein variants (column 2); and presentation according to the major five phenotype categories (column 3). (**B**) Graphical summary of three clusters of missense variants and sketch of their effect on the SF3b complex. The SF3b complex components and single amino acid variants are depicted and colored as in Fig. 3A. First cluster of variants, including p.Asn671Ser, p.Thr689Ile, and p.Arg728Pro, disrupts the interaction between SF3B3 and RNF113A causing RNF113A dissociation from the SF3b complex. Second cluster, including variants p.Gly216Asp and p.His219Pro, strengthens the interactions between SF3B3 and SF3B1 hindering the SF3B1 conformational transition from open to closed state. The third cluster, represented by p.Ala1075Asp variant, strengthens SF3B3's interactions with SF3A3, thus preventing its dissociation required to progress in the splicing cycle. (**C**) Different *SF3B3* variants result in distinct molecular mechanisms underlying a phenotypic continuum. Percentage of variants' type identified in the three forms: severe, multisystemic and milder are shown in degrees of grey. LOF: loss of function. Different *SF3B3* variants cause the clinical spectrum of the condition, likely through distinct molecular mechanisms. Missense changes can lead to amino acid substitutions that alter protein-protein binding, reducing splicing efficiency or fidelity. In addition, single amino acid changes can also result in a partial loss of the SF3B3, presumably due to an increase in the instability of the protein itself upon a specific mutation. Rare null mutations, such as nonsense/frameshift mutations that result in premature terminations and large deletions (DECIPHER acc. no.s 393580 and 501047) leading to no SF3B3 production, appear to be associated with a milder *SF3B3* phenotype

form cluster within a group, indicating the existence of distinctive craniofacial characteristics for this condition. While the analysis would benefit from including additional patients to further define the syndrome's gestalt phenotype, the results suggest that GM might be a valuable tool in accelerating an early diagnosis.

In the present study, we have identified 22 variants in the *SF3B3* gene, comprising 18 missense changes, a one-amino-acid in-frame deletion, one nonsense variant, and two frameshifts. The analysis with several prediction tools indicated a gradient of pathogenicity for the different variants, consistent with the clinical presentation's severity (Additional file 1: Table S1A).

In patients' fibroblasts we observed a reduction in SF3B3 protein levels ranging from approximately 15% to 30%, while homeostatic levels of SF3B2, SF3B4, SF3B6, PHF5A, and RNF113A were maintained. In contrast, SF3B1 levels were decreased in fibroblasts from patients 1, 2, and 6, suggesting individual differences in the ability to maintain normal levels of this critical splicing factor. Despite well-established autoregulatory and feedback mechanisms of other spliceosomal components [18], our data do not support the presence of such regulation for *SF3B3*-related condition. This reinforces the idea that even modest reductions in SF3B3 expression may have pronounced, cell-type–specific consequences during embryonic development—particularly in vulnerable populations such as neurons and neural crest cells. Evidence from other syndromes supports the plausibility of such cellular vulnerabilities: conditional knockout of factors such as *Sf3b4*, *Eftud2* and *Snrpb* in mouse embryos leads to aberrant splicing and defects in neuronal and neural crest cell development, resulting in microcephaly or craniofacial malformations [81, 87–89], phenotypes that closely mirror the human syndromic conditions associated with pathogenic variants in these genes [15, 90, 91]. Our findings suggest that impaired SF3B3 stability—together with variant-specific functional defects—may similarly disrupt RNA processing during critical developmental windows, contributing to the neurodevelopmental and craniofacial features observed in affected individuals.

The all-atom molecular dynamics simulations have made a unique contribution to the decryption of the complexity of spliceosome assembly along the splicing cycle. They have shed light on the role of disease-causing variants in genes coding its components [92, 93] and in this study revealed a remarkable impact of the *SF3B3* missense variants on complex assembly and function. The analysis identified three clusters of variants, each exhibiting a distinctive impact on the molecular and functional dynamics of SF3b (Fig. 5B). The first cluster (p.(Asn671Ser), p.(Thr689Ile), and p.(Arg728Pro)) weakened the interactions with RNF113A. This protein is a component of the B^{act} complex and is involved in recognising the guanine nucleotide at the 5'-end of the 5′-splice site (5′SS). It directly interacts with several splicing factors within the spliceosome complex, including SF3B2 [94]. RNF113A modulates the structural rearrangements that facilitate the relocation of the 5'SS into the catalytic centre for the first splicing reaction [43]. Furthermore, RNF113A interacts and polyubiquitinates the U5 snRNP BRR2 DExD/H-box ATPase, a splicing helicase responsible for U4/U6 unwinding during spliceosomal activation, which is involved in the DNA alkylation damage repair process [95]. In this scenario, the investigated variants, by disrupting the interactions between RNF113A and SF3B3, could hinder the formation of the B^{act} complex, thus preventing the first catalytic step, ultimately resulting in reduced splicing efficiency or fidelity.

It is noteworthy that a reduction in RNF113A activity is linked to neurodevelopmental disorders, including autism and X-linked trichothiodystrophy syndrome, characterised by intellectual disability, ataxia, seizures and microcephaly [96].

The second cluster of variants (p.(Gly216Asp) and p.(His219Pro)) may hinder the conformational transition from the open to the closed state, which is essential for the formation of the B^{act} spliceosome complex. It may result in the stalling of the splicing cycle and a subsequent compromise in splicing efficiency. In light of the experimental evidence on cancer-associated *SF3B1* variants, we speculate that these mutations may have a more pronounced impact on the recognition of 'weak' suboptimal intronic sequences (i.e., sequences which have a reduced base pair complementarity to U2 snRNA). It may result in greater difficulties to 'outcompete' the variants' effect in comparison to 'strong' optimal pre-mRNA sequences, which may instead be able to trigger the formation of a closed SF3B1 structure even in the presence of the variants. Finally, the majority of spliceosome proteins contain intrinsically disordered regions (IDRs) presumably to promiscuously interact with other proteins and thus facilitate the transitions between different stages of the splicing cycle. As such, the p.(Ala1075Asp) variant, representing the third cluster, may hinder the possibility of interacting with and quickly exchanging binding partners. This could affect the progression of the spliceosome along the splicing cycle, thus lowering its efficiency. This hypothesis is corroborated by the observation that Ala1075 interacts with different partners in the Open/Closed SF3b models. Specifically, it forms an electrostatic salt bridge with SF3A3, part of the spliceosome in E, A and B states. However, this interaction is lost during the transition into the B^{act} complex, when SF3B1 assumes a closed conformation. Here, the p.(Ala1075Asp) variant interacts with the SF3B2 protein.

Our findings suggest that *SF3B3* variants may impair the complex formation and progression through the

different phases of the splicing cycle or lead to altered splicing kinetics. While the spliceosome is active in all cell types, with the majority of the proteins presumably expressed ubiquitously throughout development, there are some expression biases towards specific tissue types and developmental time points [97]. Concerning this matter, we know that certain cell types are more susceptible to splicing inhibition and that the splicing kinetics of individual introns is variable depending on the origin of the cells [98]. Furthermore, cells with elevated splicing kinetics, such as neural progenitor cells in the developing embryo, may be more susceptible even to minor perturbation of the spliceosomal function.

With regard to the *SF3B3* variants associated with the severe form, identified in the probands P9, P10, and P11, it is only possible to speculate about their functional consequences. While no direct contacts between the mutated residue and the protein partners can be observed in any of the available structures, all contain intrinsically disordered regions (IDRs) for which it is difficult to extract structural data by cryo-EM studies. The presence of IDRs is a common feature of most splicing factors or proteins that are involved in pre-mRNA selection and/or proofreading since they help to mediate transition through different stages of the splicing cycle [99].

In our cohort, we have identified three LOF *SF3B3* variants in the individuals P12, P19 and P23, exhibiting different degrees of severity. Two variants (c.1242_1243insCTGGC, p.(Asn417Trpfs*21) and c.1390C>T, p.(Arg464*)), associated with milder phenotype with ASD and DD/ID in P19 and P23, are located closer to the 5′ end of the gene and their transcripts likely undergo nonsense-mediated decay (NMD). This hypothesis is corroborated by the identification of two individuals in the DECIPHER database [37] with small deletions of the 16q22.1, encompassing the *SF3B3* gene. The aforementioned deletions result in a haploinsufficiency, with one at least confirmed to be *de novo*. Such variants are associated with a phenotype characterised by ASD or/and ID (DECIPHER acc. no.s 393580 and 501047). Notably, the only other gene with a high pHaplo score in these deletions is *AP1G1*. Nevertheless, heterozygous pathogenic variants of this gene are associated with a syndromic phenotype (MIM: #619467). Furthermore, while the homozygous knockout of the *Sf3b3* mouse ortholog impacts lethality at an early stage of embryonic development (prior to organogenesis), the heterozygous mouse exhibited behavioural abnormalities, hyperactivity, abnormal coat pigmentation, and corneal morphology without any other prominent anomalies or congenital malformations [100] (MGI: 6159653). These observations suggest that haploinsufficiency of *SF3B3* is associated with a milder presentation with DD/ID and ASD.

Notably, the third LOF variant, c.2364dup, p.(Val789Cysfs*5), was identified in individual P12, exhibiting a phenotype consistent with a syndromic presentation. Given that this mutation occurs more proximal to the 3' end of the gene, it may escape NMD and result in a mutant protein retaining some altered functional protein activity or acting in a dominant-negative fashion. This hypothesis has not been investigated due to the unavailability of biological materials and thus still requires confirmation. Nevertheless, studies have demonstrated that nonsense SNVs and frameshift indels variants are enriched towards the 3′ end of the affected gene in the control population, resulting in aberrant mRNAs expressed at levels comparable to wild-type ones, indicating a greater tolerance to truncation close to the end of the coding sequence [101]. In this regard, it has been demonstrated in patient-derived fibroblast cell lines that some transcripts harbouring PTCs in *SF3B4* and *PHF5A* escape NMD, resulting thereby in stable mutant transcripts [18, 102].

The analysis of RNA sequencing data from patient-derived fibroblasts revealed transcriptional and post-transcriptional dysregulation. Differential expression (DE) analysis identified cell cycle regulators as the most prominently affected gene class, consistent with altered cell cycle profiles observed by flow cytometry. These findings indicate that SF3B3 variants disrupt core transcriptional programs controlling proliferation and cell cycle progression, a process increasingly recognized as relevant to neurodevelopmental disorders [103]. Proper cell cycle regulation is crucial for growth, differentiation, morphogenesis, and organogenesis, with rapidly proliferating cells such as neural precursors being particularly sensitive to spliceosomal perturbations [81, 104]. Coordinated cell proliferation is required for correct tissue organization, and disruptions in division or cell death can cause severe developmental defects and contribute to central nervous system pathophysiology [105–107]. Because fibroblast transcriptomes cannot fully recapitulate tissue-specific developmental processes, these expression changes should be interpreted as hypothesis-generating rather than as direct mechanistic evidence.

Given the pivotal role of splicing in gene expression, mutations, altered expression, or post-translational modifications in over 30 core spliceosomal components—including PHF5A, a SF3b complex component linked to a novel spliceosomopathy overlapping *SF3B3*-associated phenotypes—can disrupt cell cycle progression [18, 81, 108]. In line with the essential roles of spliceosomal components, including SF3B3, in cell progression, survival, and differentiation—as evidenced by early embryonic lethality in knockout mice [81]—the transcriptome analysis revealed significant enrichment of DEGs involved in urogenital system development, heart morphogenesis, neural crest differentiation, neuronal death,

neurogenesis, and Notch signaling. Key DEGs included genes that encode for transcription factors SOX11, HES1, SIX1, ID4, and SNAI1, which orchestrate early embryonic development, organogenesis, and neural differentiation. These factors interact with Notch signaling, with ID4 modulating HES1 activity to control neural stem/progenitor cell proliferation and differentiation [109–117]. Dysregulation of this network in *SF3B3* patient fibroblasts suggests disrupted spatiotemporal control of cell fate, consistent with the observed neural and organ developmental defects and the novel neurodevelopmental phenotype associated with *SF3B3* variants.

AS analysis revealed patient-specific alterations, notably the overrepresentation of RI events, suggesting a potential disease-specific splicing signature. RI events can modulate mRNA stability and translation, mechanisms critical during embryonic development [118, 119]. Consistent with our observations, mutations in the spliceosomal component EFTUD2 cause extensive intron retention, illustrating that RI can reflect core spliceosomal dysfunction [120]. Other spliceosome components, including SF3B1 and ZRSR2, are also associated with intron retention in disease contexts [121, 122]. Across all significantly altered AS events—not limited to RI—Gene Ontology enrichment highlighted ionotropic glutamate and neurotransmitter receptor complexes, NuA4/H4-H2A histone acetyltransferases, nuclear matrix components, and adhesion/ECM pathways, indicating potential impacts on synaptic excitability, transcriptional regulation, and structural programs relevant to congenital malformations [123–125]. Together, these findings suggest that SF3B3 could coordinate cell cycle progression at the transcriptional level while controlling isoform-specific programs affecting for synaptic function, chromatin organization, and cell architecture. The RI signature might link post-transcriptional regulation to transcriptional and cell cycle alterations, providing a mechanistic framework connecting aberrant splicing, DEGs, and the neurodevelopmental phenotypes characteristic of this novel *SF3B3*-related condition.

Finally, analysis of expression trajectories during human neurodevelopment revealed potential co-regulation modules between *SF3B3* and DEGs enriched in neurobiologically relevant GO processes. For example, *WNT7B*, *DLX2*, and *SOX11* show higher expression until late mid-fetal development, followed by a rapid decrease between 24 and 37 weeks of gestation, paralleling *SF3B3* patterns. Similarly, genes such as *DCX*, *VGS*, *SNAP25*, and *SERPINF1* display temporal expression consistent with neuronal progenitor activity, dendrite formation, and synapse development [73]. These observations suggest that aberrant SF3B3 levels may disrupt finely controlled spatiotemporal gene expression during brain development, potentially affecting neuronal maturation. Further studies are needed to validate this hypothesis.

Conclusions

We report on 24 individuals with *de novo* heterozygous *SF3B3* variants responsible for a novel syndromic neurodevelopmental disorder characterised by DD, ID, language and motor delay, multiple congenital anomalies, and distinctive craniofacial gestalt. The clinical severity of the condition ranges from an extremely severe prenatal form, which may result in perinatal lethality, to a milder presentation with ASD and DD/ID.

The phenotypic spectrum associated with *SF3B3* variants likely results from distinct molecular mechanisms, which contribute to varying extents depending on the specific variant (Fig. 5C). A gradient of pathogenicity is consistent with the severity of clinical presentation. Missense variants, which account for 100%, 86%, and 60% of all identified variants in the most severe, multisystemic, and milder forms, respectively, lead to amino acid substitutions that alter SF3B3 function, including protein-protein interactions within the SF3b complex, resulting in reduced splicing efficiency or fidelity. These variants also produce partial loss of SF3B3, likely due to conformational changes that increase intrinsic protein instability. Proteasome inhibition experiments suggest that reduced protein stability may contribute to the phenotype, although variant-specific effects appear complex. Rare null variants (nonsense and frameshift), predominantly associated with milder phenotypes, as well as large deletions (reported in DECIPHER [37]), likely cause SF3B3 haploinsufficiency.

Importantly, patient-derived fibroblasts carrying *SF3B3* variants exhibit widespread transcriptomic changes, including dysregulation of cell cycle genes and a distinct alternative splicing profile (e.g., retained introns and other splicing alterations). Functionally, clear cell cycle perturbations are observed in some patients, highlighting the interplay between quantitative (protein level) and qualitative (functional) effects of SF3B3 variants. These observations indicate that multiple mechanisms—including altered protein stability and impaired function—converge to affect transcriptome fidelity, alternative splicing, and downstream cellular processes, collectively shaping the clinical spectrum.

We note that while these findings help explain how variants in *SF3B3* can produce a continuous spectrum of clinical phenotypes, the diversity of phenotypes across SF3b-related syndromes likely reflects differences in expression patterns, variant-specific effects, and tissue-specific splicing programs. Our study is also limited by the use of fibroblasts as a surrogate model, the relatively small sample size, and the lack of in vivo validation. Nonetheless, these results underscore the importance of variant-specific effects in *SF3B3*-related neurodevelopmental disorders and provide a framework for future functional studies to dissect the underlying pathogenic mechanisms.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-026-01610-4.

Additional file 1. Table S1A: NGS info and statistics, prediction tools. Table S1B: Protein stability prediction upon mutation. Table S7A: Gene Ontology enrichment analysis of genes significantly differentially expressed between subjects and controls. Top 15 most significantly enriched GO terms. Table S7B: KEGG pathway enrichment analysis highlighted the cell cycle (hsa04110) as the most significant enriched pathway. Table S7C: Significant enriched BPs (in patients vs controls) related to the SF3B3-associated clinical presentation. Table S7D: mRNA levels of SF3b complex components in the patients and control fibroblasts. Table S8A: Distribution of alternative splicing event types significantly altered in patients compared to controls. Table S8B: Gene Ontology (GO) enrichment analysis of genes associated with significantly altered splicing events in patients (FDR < 0.05, |IncLevelDifference| ≥ 0.2). Table S8C: KEGG Pathway Enrichment Analysis of Genes with Significantly Altered Splicing Events.

Additional file 2: Case Reports (Subjects 1-24), Fig S1: Pairwise Comparison Matrix. Fig S2: Expression level of SF3B1 mRNA. Fig S3: Western blot analysis of cytosolic and nuclear fractions from control and patient-derived fibroblasts. Fig S4: Control and patient-derived fibroblasts treatment with the proteasome inhibitor MG132. Fig S5: Intrinsic protein stability. Fig S6: Expression trajectories of SF3B3 and other SF3b complex components during human neocortical development. Fig S7: SF3b models used in the study. Fig S8: Close-up view of interactions. Fig S9: Close-up view of interactions. Fig S10: Gene expression level analysis. Fig S11: Expression trajectories of SF3B3 and biologically relevant DEGs during human neocortical development. Fig S12: Volcano plot of 16 DEGs clinically and biologically relevant. Fig S13: Validation of RNA sequencing data using qRT-PCR. Fig S14: Number of splicing events in fibroblast RNA sequencing data. Fig S15: Z-score heatmap of alternative splicing events in patients versus controls. Fig S16: Distribution of alternative splicing event types in patient fibroblast RNA sequencing data. Fig S17: Number of splice events in RNA sequencing data of patients 1 and 2, who carry functionally related variants. Fig S18: Sashimi plots of the 26 retained-intron (RI) events that showed significant enrichment in patients. Fig S19: Bubble plot illustrating over-represented Gene Ontology (GO) terms in alternative splicing events significantly altered in SF3B3 patient-derived fibroblasts. Fig S20: Sashimi plots of the alternative spliced events affecting splicing and spliceosome-related genes that showed significant enrichment in patients. Fig S21: Flow cytometry-based cell-cycle analysis of control and patient-derived fibroblasts. Table S2: Composition of the SF3b model used in this study. Table S3: Equilibration protocol. Table S4: Oligonucleotides used for qRT-PCR experiments. Table S5: Detailed craniofacial dysmorphic features of individuals with SF3B3 variants. Table S6: Binding free energy (ΔG) and binding free energy difference (ΔΔG). Table S9: Splicing and spliceosome-related genes with significantly altered alternative splicing in patients compared to controls.

Acknowledgements

The authors would like to thank all participants and families included in this study. The Canadian contributor thanks the SickKids Complex Care Program and many other clinicians involved in the participant's care. The authors thank Martina Bradaschia for the English review of the manuscript. The authors sincerely acknowledge Alessandro Pecori for guidance and assistance using R. The authors would like to thank the SeqOIA-IT for their bioinformatics expertise during data processing and analysis. The authors thank Nicole Fleischer for her support in leveraging the Face2Gene platform for the GestaltMatcher analysis.

Authors' contributions

Author Contributions: Conceptualization L.M., F.F.; data curation: L.M., E.M.C.S., M.W., G.C., R.F., K.S., P.S., K.D.L., E.vB., R.H.vJ., A.B., L.P., E.B., C.C.H., Y.-M.C., E.G. M.B., D.K., R.W., H.G.B., L.E.L.M.V., P.M.T., Q.H., B.D.G., E.G., K.R.L., E.C.B., D.V., E.S., M.P., C.Z., C.M., E.E.E., T.W., W.G.P., K.M.B., M.P., S.M., B.C., I.M.W., E.B., A.M., F.F.; formal analysis L.M., P.J., G.P., S.C. A.L.; writing original draft L.M., P. J., A.M., G.P., S.C., A.L., F.F.; writing-review and editing: L.M., C.M., F.F.

Funding

This work was supported by the Italian Ministry of Health through the contribution given to the Institute for Maternal and Child Health - IRCCS "Burlo Garofolo" - Trieste, Italy (RC12/22; RCR-2019-23669117_001). This work was supported, in part, by US National Institutes of Health (NIH) grant R01MH101221 to EEE. EEE is an investigator of the Howard Hughes Medical Institute. The Canadian contributor acknowledges support from the University of Toronto McLaughlin Centre. PJ and AM thank the financial support of PNRR: National Center for Gene Therapy and Drugs based on RNA Technology CUP B83C22002860006 CN_0000004. AM thanks the Italian Association for Cancer Research (AIRC) for financial support (AIRC IG 24514).
This work was supported, in part, by grants U01HL131003, UM1HL098147, UM1HL098123, UM1HL128761, UM1HL128711, UMHL098153, UMHL098163, UMHL153009 and UM1HL098162 in support of the Pediatric Cardiac Genomics Consortium from the National Heart, Lung, and Blood Institute and the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute, Eunice Kennedy Shriver Institute of Child Health and Development, or the National Institutes of Health. Italian Ministry for Education, University and Research (Ministero dell'Istruzione, dell'Università e della Ricerca - MIUR) PRIN2020 code 20203P8C3X to AB. This work was also supported, in part, by the National Natural Science Foundation of China (82201314 and 82471194) and the "Fundamental Research Funds for the Central Universities" starting fund (BMU2022RCZX038) to T.W. Trio exome sequencing for family/subject 2 was performed within the project Translate NAMSE funded by German Innovation Fund. This work was also supported, in part, by the Nationwide Foundation Innovation Fund. LELMV is a member of the European Reference Network on Rare Congenital Malformations and Rare Intellectual Disability ERN-ITHACA [EU Framework Partnership Agreement ID: 3HP-HPFPA ERN-01-2016/739516]. Work at Boston Children's Hospital was supported by Eunice K. Shriver National Institute of Child Health and Human Development grant R01HD089521 and sequencing and analysis were provided by the Broad Institute of MIT and Harvard Center for Mendelian Genomics (Broad CMG) and was funded by the National Human Genome Research Institute, the National Eye Institute, and the National Heart, Lung and Blood Institute grant UM1HG008900, and in part by National Human Genome Research Institute grant R01HG009141. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data availability

NGS data generated from the patients included in this study cannot be made publicly available due to legal restrictions and patient confidentiality. However, RNA sequencing data may be made available upon reasonable request and will be shared through a data transfer and use agreement. Requests for access to the data should be sent to the corresponding authors. The dataset will be shared following approval and completion of the data transfer and use agreement within three months.

Declarations

Ethics approval and consent to participate

The study has obtained the institutional review board approval (IRB-BURLO 01/2022 09.02.2022). We obtained informed consent from all participants.

Consent for publication

Written informed consent has been obtained for publication for all participants. We obtained permission to publish patient photographs.

Competing interests

E.E.E. is a scientific advisory board (SAB) member of Variant Bio, Inc. C.C.H. one-time consulting for Ultragenyx Pharmaceutical Inc. (2024). I.M.W. is an employee of and may own stock in GeneDx, LLC. Y.-M.C. receives royalties from UpToDate on topics related to differences of sex development. All the authors declare that they have no competing interests.

Author details

[1]Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy

[2]CNR - Istituto Officina dei Materiali (IOM) c/o International School for Advanced Studies (SISSA), Via Bonomea 265, Trieste 34136, Italy
[3]Molecular Pathology Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste 34149, Italy
[4]FDNA, Boston, MA, USA
[5]Institute of Human Genetics, Heidelberg University, Heidelberg, Germany
[6]Institute of Human Genetics, School of Medicine and Health, Technical University of Munich, Munich, Germany
[7]Institute for Neurogenomics, Helmholtz Zentrum München, Neuherberg, Germany
[8]Division of Paediatric Neurology, Developmental Neurology, and Social Pediatrics, Dr von Hauner Children's Hospital, Munich, Germany
[9]Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, and Program in Genetics and Genome Biology, SickKids Research Institute, Toronto, ON, Canada
[10]Departments of Paediatrics and of Molecular Genetics, University of Toronto, Toronto, ON, Canada
[11]deCODE genetics/Amgen Inc, Reykjavik, Iceland
[12]Department of Genetics, University Medical Center Utrecht, Utrecht, the Netherlands
[13]Department of Neurosciences Rita Levi-Montalcini, University of Turin, Turin 10126, Italy
[14]Unit of Medical Genetics, Città della Salute e della Scienza University Hospital, Turin 10126, Italy
[15]Institute of Oncology Research (IOR), Bellinzona Institutes of Science (BIOS+), Bellinzona 6500, Switzerland
[16]Faculty of Biomedical Sciences, Università della Svizzera Italiana, Lugano 6900, Switzerland
[17]Department of Pediatrics, Regina Margherita Children's Hospital, Turin, Italy
[18]Department of Pediatrics, Division of Genetics and Genomics, University of North Carolina, Chapel Hill, NC, USA
[19]Division of Endocrinology, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA
[20]Department of Pediatrics, Harvard Medical School, Boston, MA, USA
[21]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[22]Division of Genetics and Genomics, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA
[23]The Danek Gertner Institute of Human Genetics, Sheba Medical Center, Tel Hashomer, Israel
[24]The Steve and Cindy Rasmussen Institute for Genomic Medicine, Nationwide Children's Hospital, 700 Children's Drive, Columbus, OH 43205, USA
[25]Akron Children's Hospital, One Perkins Square, Akron, OH 44308, USA
[26]Radboudumc, Department of Human Genetics, Nijmegen, the Netherlands
[27]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, the Netherlands
[28]Department of Clinical Genetics, Odense University Hospital, Odense, Denmark
[29]Mindich Child Health and Development Institute, Departments of Pediatrics and Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[30]Division of Cardiology, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[31]Department of Pediatrics, Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA, USA
[32]Roberts Individualized Medical Genetics Center, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[33]Rare Disease and Medical Genetics Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy
[34]Division of Medical Genetics, AOOR Villa Sofia-Cervello, Palermo, Italy
[35]Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties, University of Palermo, Palermo, Italy
[36]Department of Medicine (DMED), University of Udine, Udine, Italy
[37]Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA
[38]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA
[39]Department of Medical Genetics, Center for Medical Genetics, School of Basic Medical Sciences, Peking University, Beijing 100191, China
[40]Neuroscience Research Institute, Key Laboratory for Neuroscience, Peking University, Ministry of Education of China & National Health Commission of China, Beijing 100191, China
[41]Autism Research Center, Peking University Health Science Center, Beijing 100191, China
[42]Greenwood Genetic Center, Greenwood, SC, USA
[43]Service de Génétique médicale, Nantes Université, CHU de Nantes, Nantes F-44000, France
[44]Nantes Université, CHU de Nantes, CNRS, INSERM, l'institut du thorax, Nantes F-44000, France
[45]Laboratoire de Biologie Médicale Multi-Sites SeqOIA (laboratoire-sequoia.fr), Paris, France
[46]GeneDx LLC, Gaithersburg, MD 20877, USA
[47]Institute of Medical Genetics, Azienda Sanitaria Universitaria Friuli Centrale, Udine, Italy

Received: 26 April 2025 / Accepted: 9 February 2026
Published online: 19 February 2026

References

1. Deciphering Developmental Disorders Study. Prevalence and architecture of de Novo mutations in developmental disorders. Nature. 2017;542:433–8. https://doi.org/10.1038/nature21062.
2. Homsy J, Zaidi S, Shen Y, Ware JS, Samocha KE, Karczewski KJ, et al. De Novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science. 2015;350:1262–6. https://doi.org/10.1126/science.aac9396.
3. Fernandez-Marmiesse A, Gouveia S, Couce ML. NGS technologies as a turning point in rare disease Research, diagnosis and treatment. Curr Med Chem. 2018;25:404–32. https://doi.org/10.2174/0929867324666170718101946.
4. Gallagher A, Bulteau C, Cohen D, Michaud JL. Neurocognitive development: normative Development. [Internet]. San Diego: Elsevier; 2020. https://www.sciencedirect.com/handbook/handbook-of-clinical-neurology/vol/173/suppl/C.
5. Linder B, Fischer U, Gehring NH. mRNA metabolism and neuronal disease. FEBS Lett. 2015;589:1598–606. https://doi.org/10.1016/j.febslet.2015.04.052.
6. Rogalska ME, Vivori C, Valcárcel J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. Nat Rev Genet. 2023;24:251–69. https://doi.org/10.1038/s41576-022-00556-8.
7. Griffin C, Saint-Jeannet J, Spliceosomopathies. Diseases and mechanisms. Dev Dyn. 2020;249:1038–46. https://doi.org/10.1002/dvdy.214.
8. Beauchamp M, Alam SS, Kumar S, Jerome-Majewska LA. Spliceosomopathies and neurocristopathies: two sides of the same coin? Dev Dyn. 2020;249:924–45. https://doi.org/10.1002/dvdy.183.
9. Papasaikas P, Valcárcel J. The spliceosome: the ultimate RNA chaperone and sculptor. Trends Biochem Sci. 2016;41:33–45. https://doi.org/10.1016/j.tibs.2015.11.003.
10. Matera AG, Wang Z. A day in the life of the spliceosome. Nat Rev Mol Cell Biol. 2014;15:108–21. https://doi.org/10.1038/nrm3742.
11. Will CL, Luhrmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol. 2011;3:a003707–003707. https://doi.org/10.1101/cshperspect.a003707.
12. Sun C. The SF3b complex: splicing and beyond. Cell Mol Life Sci. 2020;77:3583–95. https://doi.org/10.1007/s00018-020-03493-z.
13. Rakesh R, Joseph AP, Bhaskara RM, Srinivasan N. Structural and mechanistic insights into human splicing factor SF3b complex derived using an integrated approach guided by the cryo-EM density maps. RNA Biol. 2016;13:1025–40. https://doi.org/10.1080/15476286.2016.1218590.
14. Timberlake AT, Griffin C, Heike CL, Hing AV, Cunningham ML, Chitayat D, et al. Haploinsufficiency of SF3B2 causes craniofacial microsomia. Nat Commun. 2021;12:4680. https://doi.org/10.1038/s41467-021-24852-9.
15. Bernier FP, Caluseriu O, Ng S, Schwartzentruber J, Buckingham KJ, Innes AM, et al. Haploinsufficiency of SF3B4, a component of the Pre-mRNA spliceosomal Complex, causes Nager syndrome. Am J Hum Genet. 2012;90:925–33. https://doi.org/10.1016/j.ajhg.2012.04.004.
16. Zhang J, Ali AM, Lieu YK, Liu Z, Gao J, Rabadan R, et al. Disease-Causing mutations in SF3B1 alter splicing by disrupting interaction with SUGP1. Mol Cell. 2019;76:82–e957. https://doi.org/10.1016/j.molcel.2019.07.017.
17. Pellagatti A, Boultwood J. Splicing factor mutations in the myelodysplastic syndromes: role of key aberrantly spliced genes in disease pathophysiology

and treatment. Adv Biol Regul. 2023;87:100920. https://doi.org/10.1016/j.jbior.2022.100920.
18. Harms FL, Dingemans AJM, Hempel M, Pfundt R, Bierhals T, Casar C, et al. De Novo PHF5A variants are associated with craniofacial abnormalities, developmental delay, and hypospadias. Genet Med. 2023;25:100927. https://doi.org/10.1016/j.gim.2023.100927.
19. Corbett MA, Dudding-Byth T, Crock PA, Botta E, Christie LM, Nardo T, et al. A novel X-linked trichothiodystrophy associated with a nonsense mutation in RNF113A. J Med Genet. 2015;52:269–74. https://doi.org/10.1136/jmedgenet-2014-102418.
20. El Chehadeh S, Kerstjens-Frederikse WS, Thevenon J, Kuentz P, Bruel A-L, Thauvin-Robinet C, et al. Dominant variants in the splicing factor PUF60 cause a recognizable syndrome with intellectual disability, heart defects and short stature. Eur J Hum Genet. 2017;25:43–51. https://doi.org/10.1038/ejhg.2016.133.
21. Zhang S, Zhang J, An Y, Zeng X, Qin Z, Zhao Y, et al. Multi-omics approaches identify SF3B3 and SIRT3 as candidate autophagic regulators and druggable targets in invasive breast carcinoma. Acta Pharm Sin B. 2021;11:1227–45. https://doi.org/10.1016/j.apsb.2020.12.013.
22. Brand M. UV-damaged DNA-binding protein in the TFTC complex links DNA damage recognition to nucleosome acetylation. EMBO J. 2001;20:3187–96. https://doi.org/10.1093/emboj/20.12.3187.
23. Martinez E, Palhan VB, Tjernberg A, Lymar ES, Gamper AM, Kundu TK, et al. Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo. Mol Cell Biol. 2001;21:6782–95. https://doi.org/10.1128/MCB.21.20.6782-6795.2001.
24. Stegeman R, Spreacker PJ, Swanson SK, Stephenson R, Florens L, Washburn MP, et al. The spliceosomal protein SF3B5 is a novel component of drosophila SAGA that functions in gene expression independent of splicing. J Mol Biol. 2016;428:3632–49. https://doi.org/10.1016/j.jmb.2016.05.009.
25. Chen K, Xiao H, Zeng J, Yu G, Zhou H, Huang C, et al. Alternative splicing of EZH2 pre-mRNA by SF3B3 contributes to the tumorigenic potential of renal cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 2017;23:3428–41. https://doi.org/10.1158/1078-0432.CCR-16-2020.
26. Cyrus S, Burkardt D, Weaver DD, Gibson WT. PRC2-complex related dysfunction in overgrowth syndromes: A review of *EZH2*, *EED*, and *SUZ12* and their syndromic phenotypes. Am J Med Genet C Semin Med Genet. 2019;181:519–31. https://doi.org/10.1002/ajmg.c.31754
27. Duan L, Chen Z, Lu J, Liang Y, Wang M, Roggero CM, et al. Histone lysine demethylase KDM4B regulates the alternative splicing of the androgen receptor in response to androgen deprivation. Nucleic Acids Res. 2019;gkz1004. https://doi.org/10.1093/nar/gkz1004.
28. Sobreira N, Schiettecatte F, Valle D, Hamosh A, GeneMatcher:. A matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015;36:928–30. https://doi.org/10.1002/humu.22844.
29. Hsieh T-C, Bar-Haim A, Moosa S, Ehmke N, Gripp KW, Pantel JT, et al. GestaltMatcher facilitates rare disease matching using facial phenotype descriptors. Nat Genet. 2022;54:349–57. https://doi.org/10.1038/s41588-021-01010-x.
30. Wadman E, Fernandes E, Muss C, Powell-Hamilton N, Wojcik MH, Madden JA, et al. A novel syndrome associated with prenatal Fentanyl exposure. Genet Med Open. 2023;1:100834. https://doi.org/10.1016/j.gimo.2023.100834.
31. Pascolini G, Gaudioso F, Baldi M, Alario D, Dituri F, Novelli A, et al. Facial clues to the photosensitive trichothiodystrophy phenotype in childhood. J Hum Genet. 2023;68:437–43. https://doi.org/10.1038/s10038-023-01134-4.
32. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, et al. VarSome: the human genomic variant search engine. Bioinf Oxf Engl. 2019;35:1978–80. https://doi.org/10.1093/bioinformatics/bty897.
33. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. Genet Med. 2015;17:405–24. https://doi.org/10.1038/gim.2015.30.
34. Landrum MJ, Chitipiralla S, Kaur K, Brown G, Chen C, Hart J, et al. ClinVar: updates to support classifications of both germline and somatic variants. Nucleic Acids Res. 2025;53:D1313–21. https://doi.org/10.1093/nar/gkae1090.
35. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The human gene mutation database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet. 2014;133:1–9. https://doi.org/10.1007/s00439-013-1358-4.
36. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. Nucleic Acids Res. 2019;47:D1038–43. https://doi.org/10.1093/nar/gky1151.
37. Foreman J, Perrett D, Mazaika E, Hunt SE, Ware JS, Firth HV. DECIPHER: improving genetic diagnosis through dynamic integration of genomic and clinical data. Annu Rev Genomics Hum Genet. 2023;24:151–76. https://doi.org/10.1146/annurev-genom-102822-100509.
38. Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C, MetaDome. Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. Hum Mutat. 2019. https://doi.org/10.1002/humu.23798. humu.23798.
39. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genomic mutational constraint map using variation in 76,156 human genomes. Nature. 2024;625:92–100. https://doi.org/10.1038/s41586-023-06045-0.
40. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005;33:W306–310. https://doi.org/10.1093/nar/gki375.
41. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins. 2006;62:1125–32. https://doi.org/10.1002/prot.20810.
42. Tholen J, Galej WP. Structural studies of the spliceosome: bridging the gaps. Curr Opin Struct Biol. 2022;77:102461. https://doi.org/10.1016/j.sbi.2022.102461.
43. Haselbach D, Komarov I, Agafonov DE, Hartmuth K, Graf B, Dybkov O, et al. Structure and conformational dynamics of the human spliceosomal Bact complex. Cell. 2018;172:454–e46411. https://doi.org/10.1016/j.cell.2018.01.010.
44. Zhang X, Yan C, Zhan X, Li L, Lei J, Shi Y. Structure of the human activated spliceosome in three conformational States. Cell Res. 2018;28:307–22. https://doi.org/10.1038/cr.2018.14.
45. Bai R, Wan R, Wang L, Xu K, Zhang Q, Lei J, et al. Structure of the activated human minor spliceosome. Science. 2021;371:eabg0879. https://doi.org/10.1126/science.abg0879.
46. Cretu C, Schmitzová J, Ponce-Salvatierra A, Dybkov O, De Laurentiis EI, Sharma K, et al. Molecular architecture of SF3b and structural consequences of its Cancer-Related mutations. Mol Cell. 2016;64:307–19. https://doi.org/10.1016/j.molcel.2016.08.036.
47. Tholen J, Razew M, Weis F, Galej WP. Structural basis of branch site recognition by the human spliceosome. Science. 2022;375:50–7. https://doi.org/10.1126/science.abm4245.
48. Webb B, Sali A. Comparative protein structure modeling using MODELLER. Curr Protoc Bioinforma. 2016;54:5.6.1–5.6.37 https://doi.org/10.1002/cpbi.3.
49. Schrödinger R. 2023-4: Protein Preparation Wizard (2023) [Internet]. https://www.schrodinger.com/life-science/download/release-notes/release-2023-4/
50. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013;27:221–34. https://doi.org/10.1007/s10822-013-9644-8.
51. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput. 2011;7:525–37. https://doi.org/10.1021/ct100578z.
52. Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. J Chem Theory Comput. 2011;7:2284–95. https://doi.org/10.1021/ct200133y.
53. Schmit JD, Kariyawasam NL, Needham V, Smith PE. SLTCAP: A simple method for calculating the number of ions needed for MD simulation. J Chem Theory Comput. 2018;14:1823–7. https://doi.org/10.1021/acs.jctc.7b01254.
54. Case DA, Aktulga HM, Belfon K, Ben-Shalom IY, Berryman JT, Brozell SR et al. Amber 2022, University of California, San Francisco, 2022. https://ambermd.org/doc12/Amber22.pdf
55. Swails J, Hernandez C, Mobley DL, Nguyen H, Wang L-P, Janowski P. ParmEd [Internet]. 2010. https://github.com/ParmEd/ParmEd
56. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput. 2015;11:3696–713. https://doi.org/10.1021/acs.jctc.5b00255.
57. Zgarbová M, Otyepka M, Sponer J, Mládek A, Banáš P, Cheatham TE et al. Refinement of the Cornell Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. J Chem Theory Comput. 2011;7:2886–902. https://doi.org/10.1021/ct200162x

58. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79:926–35. https://doi.org/10.1063/1.445869.
59. Joung IS, Cheatham TE. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J Phys Chem B. 2008;112:9020–41. https://doi.org/10.1021/jp8001614.
60. Pang Y-P. Novel zinc protein molecular dynamics simulations: steps toward antiangiogenesis for cancer treatment. J Mol Model. 1999;5:196–202. https://doi.org/10.1007/s008940050119.
61. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001.
62. Bauer P, Hess B, Lindahl EGROMACS, Manual. Zenodo; 2022 [cited 2025 Feb 26]; https://doi.org/10.5281/ZENODO.6103568
63. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;126:014101. https://doi.org/10.1063/1.2408420.
64. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys. 1981;52:7182–90. https://doi.org/10.1063/1.328693.
65. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J Comput Chem. 1997;18:1463–72. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H.
66. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys. 1995;103:8577–93. https://doi.org/10.1063/1.470117.
67. Roe DR, Cheatham TE, PTRAJ. Software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput. 2013;9:3084–95. https://doi.org/10.1021/ct400341p.
68. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: an efficient program for End-State free energy calculations. J Chem Theory Comput. 2012;8:3314–21. https://doi.org/10.1021/ct300418h.
69. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14:33–8. https://doi.org/10.1016/0263-7855(96)00018-5.
70. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta delta C(T)) method. Methods San Diego Calif. 2001;25:402–8. https://doi.org/10.1006/meth.2001.1262.
71. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5. https://doi.org/10.1038/nbt.3122.
72. Schneider CA, Rasband WS, Eliceiri KW. NIH image to imageJ: 25 years of image analysis. Nat Methods. 2012;9:671–5. https://doi.org/10.1038/nmeth.2089.
73. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. Spatio-temporal transcriptome of the human brain. Nature. 2011;478:483–9. https://doi.org/10.1038/nature10523.
74. Radio FC, Pang K, Ciolfi A, Levy MA, Hernández-García A, Pedace L, et al. SPEN haploinsufficiency causes a neurodevelopmental disorder overlapping proximal 1p36 deletion syndrome with an episignature of X chromosomes in females. Am J Hum Genet. 2021;108:502–16. https://doi.org/10.1016/j.ajhg.2021.01.015.
75. Kaplanis J, Samocha KE, Wiel L, Zhang Z, Arvai KJ, Eberhardt RY, et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. Nature. 2020;586:757–62. https://doi.org/10.1038/s41586-020-2832-5.
76. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An J-Y, et al. Large-Scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell. 2020;180:568–e58423. https://doi.org/10.1016/j.cell.2019.12.036.
77. Zhou X, Feliciano P, Shu C, Wang T, Astrovskaya I, Hall JB, et al. Integrating de Novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. Nat Genet. 2022;54:1305–19. https://doi.org/10.1038/s41588-022-01148-2.
78. Pervolaraki E, Dachtler J, Anderson RA, Holden AV. The developmental transcriptome of the human heart. Sci Rep. 2018;8:15362. https://doi.org/10.1038/s41598-018-33837-6.
79. Pervolaraki E, Anderson RA, Benson AP, Hayes-Gill B, Holden AV, Moore BJR, et al. Antenatal architecture and activity of the human heart. Interface Focus. 2013;3:20120065. https://doi.org/10.1098/rsfs.2012.0065.
80. Pervolaraki E, Dachtler J, Anderson RA, Holden AV. Ventricular myocardium development and the role of connexins in the human fetal heart. Sci Rep. 2017;7:12272. https://doi.org/10.1038/s41598-017-11129-9.
81. Olthof AM, White AK, Kanadia RN. The emerging significance of splicing in vertebrate development. Dev Camb Engl. 2022;149:dev200373. https://doi.org/10.1242/dev.200373.
82. Love SL, Emerson JD, Koide K, Hoskins AA. Pre-mRNA splicing-associated diseases and therapies. RNA Biol. 2023;20:525–38. https://doi.org/10.1080/15476286.2023.2239601.
83. Deutsch HM, Song Y, Li D. Spliceosome complex and neurodevelopmental disorders. Curr Opin Genet Dev. 2025;93:102358. https://doi.org/10.1016/j.gde.2025.102358.
84. Griffin C. Modeling craniofacial spliceosomopathies: a pathway toward Deciphering disease mechanisms. Front Cell Dev Biol. 2025;13:1624043. https://doi.org/10.3389/fcell.2025.1624043.
85. Merkuri F, Fish JL. Developmental processes regulate craniofacial variation in disease and evolution. Genes N Y N 2000. 2019;57:e23249. https://doi.org/10.1002/dvg.23249
86. Schmidt A, Danyel M, Grundmann K, Brunet T, Klinkhammer H, Hsieh T-C, et al. Next-generation phenotyping integrated in a National framework for patients with ultrarare disorders improves genetic diagnostics and yields new molecular findings. Nat Genet. 2024;56:1644–53. https://doi.org/10.1038/s41588-024-01836-1.
87. Alam SS, Kumar S, Beauchamp M-C, Bareke E, Boucher A, Nzirorera N, et al. Snrpb is required in murine neural crest cells for proper splicing and craniofacial morphogenesis. Dis Model Mech. 2022;15:dmm049544. https://doi.org/10.1242/dmm.049544.
88. Kumar S, Bareke E, Lee J, Carlson E, Merkuri F, Schwager EE, et al. Etiology of craniofacial and cardiac malformations in a mouse model of *SF3B4* -related syndromes. Proc Natl Acad Sci. 2024;121:e2405523121. https://doi.org/10.1073/pnas.2405523121.
89. Beauchamp M-C, Djedid A, Bareke E, Merkuri F, Aber R, Tam AS, et al. Mutation in *Eftud2* causes craniofacial defects in mice via mis-splicing of *Mdm2* and increased P53. Hum Mol Genet. 2021;30:739–57. https://doi.org/10.1093/hmg/ddab051.
90. Lines MA, Huang L, Schwartzentruber J, Douglas SL, Lynch DC, Beaulieu C, et al. Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly. Am J Hum Genet. 2012;90:369–77. https://doi.org/10.1016/j.ajhg.2011.12.023.
91. Lynch DC, Revil T, Schwartzentruber J, Bhoj EJ, Innes AM, Lamont RE, et al. Disrupted auto-regulation of the spliceosomal gene SNRPB causes cerebrocosto-mandibular syndrome. Nat Commun. 2014;5:4483. https://doi.org/10.1038/ncomms5483.
92. Borišek J, Saltalamacchia A, Gallì A, Palermo G, Molteni E, Malcovati L, et al. Disclosing the impact of carcinogenic SF3b mutations on Pre-mRNA recognition via All-Atom simulations. Biomolecules. 2019;9:633. https://doi.org/10.3390/biom9100633.
93. Spinello A, Janos P, Rozza R, Magistrato A. Cancer-Related mutations alter RNA-Driven functional Cross-Talk underlying Premature-Messenger RNA recognition by splicing factor SF3b. J Phys Chem Lett. 2023;14:6263–9. https://doi.org/10.1021/acs.jpclett.3c01402.
94. Hegele A, Kamburov A, Grossmann A, Sourlis C, Wowro S, Weimann M, et al. Dynamic protein-protein interaction wiring of the human spliceosome. Mol Cell. 2012;45:567–80. https://doi.org/10.1016/j.molcel.2011.12.034.
95. Brickner JR, Soll JM, Lombardi PM, Vågbø CB, Mudge MC, Oyeniran C, et al. A ubiquitin-dependent signalling axis specific for ALKBH-mediated DNA dealkylation repair. Nature. 2017;551:389–93. https://doi.org/10.1038/nature24484.
96. Tsampoula M, Tarampoulous I, Manolakou T, Ninou E, Politis PK. The neurodevelopmental disorders associated gene Rnf113a regulates survival and differentiation properties of neural stem cells. Stem Cells Dayt Ohio. 2022;40:678–90. https://doi.org/10.1093/stmcls/sxac030.
97. Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, et al. Gene expression across mammalian organ development. Nature. 2019;571:505–9. https://doi.org/10.1038/s41586-019-1338-5.
98. Bedi K, Magnuson BR, Narayanan I, Paulsen M, Wilson TE, Ljungman M. Co-transcriptional splicing efficiencies differ within genes and between cell types. RNA N Y N. 2021;27:829–40. https://doi.org/10.1261/rna.078662.120.
99. Korneta I, Bujnicki JM. Intrinsic Disorder in the Human Spliceosomal Proteome. Iakoucheva LM, editor. PLoS Comput Biol. 2012;8:e1002641. https://doi.org/10.1371/journal.pcbi.1002641
100. Adams DJ, Barlas B, McIntyre RE, Salguero I, van der Weyden L, Barros A, et al. Genetic determinants of micronucleus formation in vivo. Nature. 2024;627:130–6. https://doi.org/10.1038/s41586-023-07009-0.

101. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A systematic survey of Loss-of-Function variants in human Protein-Coding genes. Science. 2012;335:823–8. https://doi.org/10.1126/science.1215040.
102. Marques F, Tenney J, Duran I, Martin J, Nevarez L, Pogue R et al. Altered mRNA Splicing, Chondrocyte Gene Expression and Abnormal Skeletal Development due to SF3B4 Mutations in Rodriguez Acrofacial Dysostosis. Bateman JF, editor. PLOS Genet. 2016;12:e1006307. https://doi.org/10.1371/journal.pgen.1006307
103. Shigenaka A, Nitta E, Nakagawa T, Nakagawa M, Hosoi T. Defective neural stem and progenitor cell proliferation in neurodevelopmental disorders. J Dev Biol. 2025;13:40. https://doi.org/10.3390/jdb13040040.
104. Beemster GTS, Mironov V, Inzé D. Tuning the cell-cycle engine for improved plant performance. Curr Opin Biotechnol. 2005;16:142–6. https://doi.org/10.1016/j.copbio.2005.01.006.
105. Matthew J, Vishwakarma V, Le TP, Agsunod RA, Chung S. Coordination of cell cycle and morphogenesis during organ formation. eLife. 2024;13:e95830. https://doi.org/10.7554/eLife.95830.
106. Godard BG, Heisenberg C-P. Cell division and tissue mechanics. Curr Opin Cell Biol. 2019;60:114–20. https://doi.org/10.1016/j.ceb.2019.05.007.
107. Wang W, Bu B, Xie M, Zhang M, Yu Z, Tao D. Neural cell cycle dysregulation and central nervous system diseases. Prog Neurobiol. 2009;89:1–17. https://doi.org/10.1016/j.pneurobio.2009.01.007.
108. Hluchý M, Blazek D. CDK11, a splicing-associated kinase regulating gene expression. Trends Cell Biol. 2024. https://doi.org/10.1016/j.tcb.2024.08.004. S0962-8924(24)00161-2.
109. Niessen K, Karsan A. Notch signaling in cardiac development. Circ Res. 2008;102:1169–81. https://doi.org/10.1161/CIRCRESAHA.108.174318.
110. Luxán G, D'Amato G, MacGrogan D, de la Pompa JL. Endocardial Notch signaling in cardiac development and disease. Circ Res. 2016;118:e1–18. https://doi.org/10.1161/CIRCRESAHA.115.305350.
111. Zhou B, Lin W, Long Y, Yang Y, Zhang H, Wu K, et al. Notch signaling pathway: architecture, disease, and therapeutics. Signal Transduct Target Ther. 2022;7:95. https://doi.org/10.1038/s41392-022-00934-y.
112. Boareto M, Iber D, Taylor V. Differential interactions between Notch and ID factors control neurogenesis by modulating Hes factor autoregulation. Dev Camb Engl. 2017;144:3465–74. https://doi.org/10.1242/dev.152520.
113. Roschger C, Cabrele C. The Id-protein family in developmental and cancer-associated pathways. Cell Commun Signal CCS. 2017;15:7. https://doi.org/10.1186/s12964-016-0161-y.
114. Kobayashi T, Kageyama R. Expression dynamics and functions of Hes factors in development and diseases. Curr Top Dev Biol. 2014;110:263–83. https://doi.org/10.1016/B978-0-12-405943-6.00007-5.
115. Hirata H, Tomita K, Bessho Y, Kageyama R. Hes1 and Hes3 regulate maintenance of the isthmic organizer and development of the mid/hindbrain. EMBO J. 2001;20:4454–66. https://doi.org/10.1093/emboj/20.16.4454.
116. Kageyama R, Ohtsuka T, Kobayashi T. Roles of Hes genes in neural development. Dev Growth Differ. 2008;50(Suppl 1):S97–103. https://doi.org/10.1111/j.1440-169X.2008.00993.x.
117. Sang L, Coller HA, Roberts JM. Control of the reversibility of cellular quiescence by the transcriptional repressor HES1. Science. 2008;321:1095–100. https://doi.org/10.1126/science.1155998.
118. Porras-Tobias AL, Caldera A, Castro-Piedras I, Intron Retention. A reemerging paradigm in RNA biology and Post-Transcriptional gene regulation. Genes. 2025;16:986. https://doi.org/10.3390/genes16080986.
119. Monteuuis G, Wong JJL, Bailey CG, Schmitz U, Rasko JEJ. The changing paradigm of intron retention: regulation, ramifications and recipes. Nucleic Acids Res. 2019;47:11497–513. https://doi.org/10.1093/nar/gkz1068.
120. Liu Y, Chen Z, Duan Y, Shao Z, Chen Y, Yang N, et al. Spliceosomal GTPase EFTUD2 mediates DDX41 intron retention to promote the malignant progression of ovarian cancer. Br J Cancer. 2025;133:508–23. https://doi.org/10.1038/s41416-025-03079-1.
121. Kashyap MK, Karathia H, Kumar D, Vera Alvarez R, Forero-Forero JV, Moreno E, et al. Aberrant spliceosome activity via elevated intron retention and upregulation and phosphorylation of SF3B1 in chronic lymphocytic leukemia. Mol Ther Nucleic Acids. 2024;35:102202. https://doi.org/10.1016/j.omtn.2024.102202.
122. Inoue D, Polaski JT, Taylor J, Castel P, Chen S, Kobayashi S, et al. Minor intron retention drives clonal hematopoietic disorders and diverse cancer predisposition. Nat Genet. 2021;53:707–18. https://doi.org/10.1038/s41588-021-00828-9.
123. Rekad Z, Izzi V, Lamba R, Ciais D, Van Obberghen-Schilling E. The alternative matrisome: alternative splicing of ECM proteins in development, homeostasis and tumor progression. Matrix Biol J Int Soc Matrix Biol. 2022;111:26–52. https://doi.org/10.1016/j.matbio.2022.05.003.
124. Pollina EA, Gilliam DT, Landau AT, Lin C, Pajarillo N, Davis CP, et al. A NPAS4-NuA4 complex couples synaptic activity to DNA repair. Nature. 2023;614:732–41. https://doi.org/10.1038/s41586-023-05711-7.
125. Santos JR, Park J. MATR3's role beyond the nuclear matrix: from gene regulation to its implications in amyotrophic lateral sclerosis and other diseases. Cells. 2024;13:980. https://doi.org/10.3390/cells13110980.

Publisher's Note