



## OPEN Alternative splicing analysis in a Spanish ASD (Autism Spectrum Disorders) cohort: in silico prediction and characterization

S. Dominguez-Alonso<sup>1</sup>, M. Tubío-Funqueiriño<sup>1</sup>, J. González-Peñas<sup>3</sup>, M. Fernández-Prieto<sup>1,2</sup>, M. Parellada<sup>3</sup>, C. Arango<sup>3</sup>, A. Carracedo<sup>1,2</sup> & C. Rodríguez-Fontenla<sup>1</sup>✉

Autism Spectrum Disorders (ASD) are complex and genetically heterogeneous neurodevelopmental conditions. Although alternative splicing (AS) has emerged as a potential contributor to ASD pathogenesis, its role in large-scale genomic studies has remained relatively unexplored. In this comprehensive study, we utilized computational tools to identify, predict, and validate splicing variants within a Spanish ASD cohort (360 trios), shedding light on their potential contributions to the disorder. We utilized SpliceAI, a newly developed machine-learning tool, to identify high-confidence splicing variants in the Spanish ASD cohort and applied a stringent threshold ( $\Delta \geq 0.8$ ) to ensure robust confidence in the predictions. The in silico validation was then conducted using SpliceVault, which provided compelling evidence of the predicted splicing effects, using 335,663 reference RNA-sequencing (RNA-seq) datasets from GTEx v8 and the sequence read archive (SRA). Furthermore, ABRSplice was employed for additional orthogonal in silico confirmation and to elucidate the tissue-specific impacts of the splicing variants. Notably, our analysis suggested the contribution of splicing variants within *CACNA1I*, *CBLB*, *CLTB*, *DLGAP1*, *DVL3*, *KIAA0513*, *OFD1*, *PKD1*, *SLC13A3*, and *SCN2A*. Complementary datasets, including more than 42,000 ASD cases, were employed for gene validation and gene ontology (GO) analysis. These analyses revealed potential tissue-specific effects of the splicing variants, particularly in adipose tissue, testis, and the brain. These findings suggest the involvement of these tissues in ASD etiology, which opens up new avenues for further functional testing. Enrichments in molecular functions and biological processes imply the presence of separate pathways and mechanisms involved in the progression of the disorder, thereby distinguishing splicing genes from other ASD-related genes. Notably, splicing genes appear to be predominantly associated with synaptic organization and transmission, in contrast to non-splicing genes (i.e., genes harboring *de novo* and inherited coding variants not predicted to alter splicing), which have been mainly implicated in chromatin remodeling processes. In conclusion, this study advances our comprehension of the role of AS in ASD and calls for further investigations, including in vitro validation and integration with multi-omics data, to elucidate the functional roles of the highlighted genes and the intricate interplay of the splicing process with other regulatory mechanisms and tissues in ASD.

Autism Spectrum Disorders (ASD) encompass a group of phenotypically and genetically heterogeneous neurodevelopmental disorders (NDDs) characterized by difficulties in social interaction and communication, repetitive behavior, and restricted interests<sup>1</sup>. ASD have a strong heritability, estimated around 80%<sup>2</sup>, although their complex genetic etiology has limited progress towards understanding their molecular basis. Genomic approaches, including genome-wide association studies (GWAS)<sup>3–5</sup>, whole exome (WES) and whole-genome sequencing (WGS) of families<sup>6–9</sup>, as well as transcriptome analyses by RNA-sequencing (RNA-seq) and microarray techniques<sup>10–12</sup>, have yielded association with hundreds of genes over the last decades. However, a

<sup>1</sup>Grupo de Medicina Xenómica, Center for Research in Molecular Medicine and Chronic Diseases (CiMUS), Universidad de Santiago de Compostela, Santiago de Compostela, Spain. <sup>2</sup>Grupo de Medicina Xenómica, Center for Research in Molecular Medicine and Chronic Diseases (CiMUS), Fundación Instituto de Investigación Sanitaria de Santiago de Compostela (FIDIS), Universidad de Santiago de Compostela, Santiago de Compostela, Spain. <sup>3</sup>Centro De Investigación Biomédica en Red de Salud Mental (CIBERSAM), School of Medicine, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Universidad Complutense, IISGM, Madrid, Spain. ✉email: mariacristina.rodriguez.fontenla@usc.es

large portion of their genetic architecture remains uncharacterized, and ASD diagnosis continues to be a major challenge in both clinical and research settings.

Emerging advances have implicated alternative splicing (AS) as yet another process influencing the development of the disease. AS is the mechanism by which introns are excised from the pre-mRNA primary transcript, and exons are selected and concatenated in different arrangements, generating multiple transcript isoforms and, consequently, protein products, from a single gene. More than 95% of multiexon human genes that encode proteins undergo AS, and most mRNA splice patterns exhibit tissue- and cell-type-specificity<sup>13</sup>. Consequently, AS is mainly accountable for generating the vast proteomic diversity observed in complex organisms and has been increasingly linked to functional intricacies within the central nervous system. It plays crucial roles in nervous system development, neuronal differentiation and maturation, as well as complex neuronal processes, such as the control of synaptic plasticity associated with cognition<sup>14,15</sup>. Some neuronal genes, like neurexins, possess the ability to produce hundreds of mRNA isoforms. This phenomenon stands as one of the most extensive cases of AS regulation observed to date and plays a pivotal role in optimal neuronal function<sup>14</sup>.

Notably, aberrations in AS have been associated with several NDDs, including schizophrenia and bipolar disorders<sup>16,17</sup>, Rett syndrome<sup>18,19</sup>, fragile X syndrome<sup>20,21</sup> and ASD<sup>16,22–29</sup>. These aberrations may occur at two different levels: cis-acting motifs (exonic/intronic splicing enhancers and silencers) and trans-acting factors that regulate the assembly of the spliceosome by recognition of proximal cis-elements, such as RNA-binding proteins (RBPs)<sup>14,30,31</sup>.

The majority of AS studies in ASD patients pertain to individual RBPs that govern the inclusion/exclusion of microexons (3–27 nt). Microexons constitute the most conserved component of the neural-regulated AS program. They generally reside on protein surfaces, specifically in domains with important roles in shaping protein-protein interactions (PPIs). These sequences are significantly enriched in genes with neuronal function which have been genetically linked to ASD<sup>25,32</sup>, predominantly being neuronal-included<sup>33</sup>. RNA-Seq analysis of post-mortem brain samples in idiopathic ASD individuals shows missplicing in 30–40% of brain-specific microexons<sup>25</sup>. Analyses of larger cohorts of ASD individuals corroborated these observations and demonstrated that most of the differential splicing events involve the exclusion of these microexons<sup>16,22</sup>. Most microexons are controlled by the neural-specific Ser/Arg-related splicing factor, nSR100/SRRM4<sup>26,27</sup>, and its misregulation is linked to reduced expression levels of nSR100/SRRM4 in autistic brains. Employing heterozygous mutant mice expressing approximately 50% wild-type levels of nSR100.

Microexons are also related not only to PPIs but also with disrupting synaptic protein networks critical for brain development and function. Microexons encode surface-accessible residues that fine-tune protein-protein interactions (PPIs)<sup>34</sup>. Thus, the inclusion of a microexon in *CPEB4* prevents irreversible protein aggregation, enabling dynamic regulation of mRNA translation in response to neuronal activity. In addition, microexons can regulate synaptic plasticity, with exclusion increasing synaptic NMDA receptor levels (e.g., GluN1) and triggering ASD-like hyperexcitability<sup>33</sup>.

Quesnel-Valli  res et al., 2016<sup>27</sup> demonstrated that a single variant in a splicing regulator, and thus the disruption of its target splicing program, suffice to reproduce hallmark features of ASD, including altered social behaviour, synaptic transmission and neuronal excitability. Additional examples of RBPs implicated in the differential splicing changes observed in ASD brain are Rbfox1<sup>35–37</sup> and *PTBPI*<sup>22,35,38</sup>.

Given the mounting evidence of splicing disruptions in ASD patients, we sought to investigate the possible effects of *de novo* mutations in AS within a Spanish cohort of 360 ASD trios. Here, we employed SpliceAI, a machine-learning tool that robustly predicts splice sites and splice-disrupting variants, which has exhibited notable performance when compared to previously developed in silico detection tools<sup>39</sup>. Moreover, we utilized SpliceVault<sup>40</sup> and ABSsplice<sup>41</sup>, two newly developed tools that assess AS, in order to bolster robustness of the in silico splicing prediction and delve into the specific effects within human tissues. This paper provides initial insights into the potential functional roles of genes harboring splicing variants, thereby highlighting molecular pathways and biological processes in which they are implicated.

## Methods

### Subjects

The analysis described herein builds upon the complete sample set examined in Alonso Gonzalez et al. 2021<sup>42</sup>. DNA extraction from the Spanish ASD samples, consisting of 360 trios (unaffected parents and affected proband), was performed using the GentraPuregene blood kit (Qiagen Inc., Valencia, CA, USA) from peripheral blood.

Participants from Santiago ( $n=136$ ) were recruited from the Complejo Hospitalario Universitario de Santiago de Compostela and Galician ASD organizations. Meanwhile, subjects from Madrid ( $n=224$ ) were enrolled through the AMITEA program at the Child and Adolescent Department of Psychiatry, Hospital General Universitario Gregorio Mara  n. Inclusion criteria stipulated that only individuals aged 3 years or older were included in the study.

Enrolled participants received a clinical diagnosis of ASD from trained pediatric neurologists or psychiatrists, following the criteria outlined in both the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Text Revision (DSM-IV-TR) and Fifth Edition (DSM-5). Additionally, when deemed necessary, the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R) were administered.

All participants, along with their parents or legal representatives, provided written informed consent, and the study was conducted in accordance with the principles outlined in the Declaration of Helsinki.

The Galician Committee of Research Ethics (Xunta de Galicia) has approved this study under the Register 2020/400.

Ninety samples from the Spanish cohort (360 trios), were already analyzed by Lim et al., 2017<sup>43</sup>. The entire Spanish cohort was included in Satterstrom et al, 2020<sup>44</sup> as part of the Autism Sequencing Consortium (ASC), a large-scale international genomic consortium integrating ASD cohorts and sequencing data from over one hundred investigators. Samples were all assessed for *de novo* mutations in exons, both postzygotic and germinal, as described by Alonso González et al. 2021<sup>42</sup>. However, the analysis using CNV (Copy Number Variation) array did not reveal any significant findings for these mutations, indicating negative results in the context of the studied samples. All data generated as part of the ASC was transferred to dbGaP with Study Accession: phs000298.v4.p3.

### In silico annotation and prioritization of splicing variants

To perform subsequent analyses, we leveraged previously reported biallelic *de novo* variants (DNVs) from our previous work. DNVs were defined as those present exclusively in probands and not in parents (i.e., genotypes 1/0 or 1/1 in probands and 0/0 in parents). The filtering steps carried out in Alonso-Gonzalez et al., 2021<sup>42</sup> were not used with SpliceAI. Instead, appropriate quality filtering for splicing variants was applied to the raw *de novo* variants of the analysis, as detailed below.

To perform *in silico* annotation and prioritization of splice variants, we used SpliceAI (<https://github.com/Illumina/SpliceAI>), a machine-learning tool that robustly predicts splice sites and splice-disrupting variants. SpliceAI has been already utilized to assess the clinical impact of non-coding mutations that act through altered splicing in patients with ASD<sup>39</sup>.

SpliceAI was executed with default parameters (-D: maximum distance between the variant and gained/lost splice site (default: 50); -M: mask scores representing annotated acceptor/donor gain and unannotated acceptor/donor loss (default: 0)). For the gene annotation file, we used the GENCODE V24 canonical annotation files included in the package.

SpliceAI annotation was available for 252,520 *de novo* variants (72.91% of 346,352 initial variants). Variants with delta scores  $\Delta \geq 0.8$  (indicating confidently predicted splice-altering effects, high-precision cut-off as described by Janagathan et al. 2019) were retained ( $n = 1,836$ ). The delta score indicates the location of splicing changes relative to the variant position, with positive values indicating downstream and negative values indicating upstream changes. In addition, we considered only *de novo* variants with genome quality (GQ)  $\geq 20$  and alternate read depth (AD)  $\geq 7$  and removed any call if it had an allele frequency  $> 0.1\%$  across the samples in our dataset or in the non-psychiatric subset of gnomAD (1,793 putative *de novo* variants excluded).

### In silico variant characterization

#### SpliceVault

The tool SpliceVault<sup>40</sup> was employed to accurately predict the exact manner in which genetic variants affect the splicing process. It ranks the four most common unannotated splicing events across 335,663 reference RNA-seq samples (300 K-RNA Top-4) from GTEx v8 (Genotype-Tissue Expression project) and SRA (Sequence Read Archive; an online archive of high-throughput RNA sequencing data). 300 K-RNA is built in hg38 (GRCh38), so genome coordinates were converted between assemblies GRCh37 and GRCh38 using the UCSC web server's LiftOver tool (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>).

Splice variants were then interrogated against the SpliceVault web portal (<https://kidsneuro.shinyapps.io/splicevault/>) using default settings recommended for clinical use: top 4 events, skipping of  $\leq 2$  exons and cryptic positions within  $\pm 600$  nt. A variant was considered validated if it appeared in the top 4 events of SpliceVault. The default settings were chosen as they gave a sensitivity of 90% and PPV of 31% for predicting exon skipping and cryptic activation events.

#### Tissue specificity: ABSplice

Each variant was examined for its specific effect in any given human tissue. ABSplice<sup>41</sup> (<https://github.com/gauneurlab/absplice>), a model that maps acceptor and donor splice sites and quantifies their usage in 49 human tissues, was run with default parameters. For each variant, we recorded the tissue with the highest score for AS outputted by ABSplice. This score indicates the likelihood that a specific genetic variant causes abnormal splicing in a particular tissue. ABSplice thresholds are defined as 0.01 (low), 0.05 (intermediate), and 0.2 (high), which have approximately the same recalls as the high, medium, and low cutoffs of SpliceAI. A variant was considered validated if it appeared in the ABSplice dataset with a score higher than 0.2.

Additionally, to enhance our understanding of the variant effects across various tissues, we annotated our variants by cross-referencing them with precomputed ABSplice-DNA scores for all tissues, not just the highest scoring one. The precomputed ABSplice-DNA scores for 49 human tissues and all possible SNVs genome-wide for hg38 were made available at Zenodo (<https://zenodo.org/records/7871809>).

SpliceAI, SpliceVault, and ABSplice do not apply standard multiple testing corrections. However, the use of respective scores thresholds helps reduce false positives by focusing on high-precision predictions.

### Complementary datasets

Comparing functional profiles can reveal functional consensus and differences among different experiments and helps in identifying differential functional modules in different datasets.

For further validation of hereafter interrogated genes harboring splice variants, analysis of gene ontology (GO) enrichments and comparison of enriched terms in each dataset (see next section), we utilized previously reported ASD-related sets of genes (Table 1).

After *in silico* variant confirmation, genes harboring the validated variants were contrasted against different complementary gene datasets. These included genes associated with ASD and implicated in AS<sup>25,28,29,36,39</sup>, as

Description	Study type	Samples	Data availability	Reference
The Human Gene module of the SFARI database (up-to-date reference for all known human genes associated with autism ASD) was accessed (SFARI 07-17-2023 release) and queried against 1–2 scoring genes (high confidence and strong evidence of association with ASD, respectively).	Mutation screening, family-based association, case-control, WES, WGS and CNV array	-	<a href="https://gene.sfari.org/database/human-gene/">https://gene.sfari.org/database/human-gene/</a>	-
Highly conserved program of neural microexons primarily regulated by the neuronal-specific splicing factor nSR100/SRRM4.	RNA-seq custom pipeline	22 autistic individuals, 20 controls	Supplementary Table 2: Neural-regulated AS events in human	25
Unique patterns of AS and gene co-expression in ASD-affected dizygotic twins compared to their parents.	AS and co-expression analyses	Two pairs of DZ twins and their parents	Supplementary Table 1: Differential AS events	28
Distinct AS patterns in the blood of patients with ASD compared to typically developing individuals.	Whole genome exon arrays	30 ASD patient, 20 controls	Additional file 2	29
Dysregulated splicing pattern of <i>RBFOX1</i> -dependent alternative exons in the ASD brain.	Post-mortem brain tissue	19 autism samples, 17 controls	Supplementary Data: Differential Splicing Events	36
Detection of a 1.30-fold enrichment of <i>de novo</i> splicing mutations in ASD ( $p=0.0203$ ) compared to healthy controls when employing SpliceAI.	High-depth mRNA sequencing	36 autism samples	Supplementary Table 3	39
Analysis of <i>de novo</i> and inherited variants identifies 60 genes with exome-wide significance implicated in ASD, including five new risk genes ( <i>NAV3</i> , <i>ITSN1</i> , <i>MARK2</i> , <i>SCAF1</i> and <i>HNRNPUL2</i> ).	WES/WGS	42,607 autism cases	Supplementary Table 1	45

**Table 1.** Complementary datasets included in gene validation and GO analysis. The first 6 studies were used for gene validation and the last6, were used for comparison of enriched terms. CNV, copy number variant; DZ, dizygotic; AS, alternative splicing, WES, whole exome sequencing, WGS, whole genome sequencing.

well as the SFARI database. Genes that overlapped with these datasets were considered validated and were subsequently included in further GO analysis.

Additionally, we utilized the same gene datasets (ASD-related genes implicated in AS) along with a study integrating *de novo* and inherited variants, not predicted to affect AS, in 42,607 ASD cases<sup>45</sup>. This information was employed for a detailed comparison of enriched terms among different categories of genetic variants, namely, *de novo* and inherited coding variants not predicted to alter splicing *versus* splicing variants.

Gene ontology (GO) analysis  
Gene network analysis

HumanBase (<https://hb.flatironinstitute.org/>) tool was used to build a gene network for the genes already associated, and thus, validated, in the aforementioned complementary datasets. HumanBase serves as a comprehensive resource for biological research and offers data-driven predictions related to gene expression, function, regulation, and interactions within the human domain, with a particular focus on specific cell types, tissues, and diseases.

In order to capture tissue-specific gene function, we used the “tissue specific gene networks: GIANT” analysis tool from HumanBase. The GIANT analysis tool, constructs comprehensive genome-scale functional maps for various human tissues by integrating extensive datasets from over 14,000 distinct publications, covering thousands of experiments. The platform automatically evaluates the relevance of each dataset to 144 tissue- and cell lineage-specific functional contexts. The resulting functional gene maps offer detailed insights into protein function and interactions in specific human tissues and cell lineages.

*CACNA1I*, *CBLB*, *CLTB*, *DLGAP1*, *DVL3*, *KIAA0513*, *OFD1*, *PKD1*, *SLC13A3* and *SCN2A* (i.e., the validated genes) were selected as the input genes along with brain tissue in the 5 existing data types (co-expression, transcription factor binding, interaction, gene set enrichment analysis (GSEA) microRNA targets, and GSEA perturbations). The resultant network (henceforth designated as the splicing gene list, Supplementary Table 2, Supplementary Fig. 2) contains the subset of functionally related genes specific to brain tissue, capturing tissue-specific gene function, all of which were used to test for functional enrichment using genes annotated to GO biological process (BP), cellular component (CC) and molecular function (MF) terms.

Enrichment analysis

ClusterProfiler (<https://github.com/YuLab-SMU/clusterProfiler>)<sup>46</sup>, an R package tailored for contrasting biological themes among gene groups, was harnessed to perform both GO over-representation test and to deduce enriched functional profiles on separate gene clusters (i.e., gene sets).

GO enrichment analysis

The package *org.Hs.eg.db*, provided by Bioconductor, was used as the genome wide annotation for Human. We employed the *bitr* tool (Biological Id TRanslator), already implemented in the *clusterProfiler* package (with parameters: *fromType* = “SYMBOL”, *toType*=“ENTREZID”, *OrgDb*=“*org.Hs.eg.db*”), to obtain Entrez Gene identifiers for the genes of interest. For genes failing conversion to an Entrez ID, we further employed the *mygene* module in Python (<https://pypi.org/project/mygene/>), which obtains the gene annotation data from several public data resources (NCBI Entrez, Ensembl, Uniprot, NetAffx, PharmGKB, UCSC, and CPDB) and keep them up-to-date.



GO enrichment analysis was performed with specific significance thresholds ( $p\text{-valueCutoff}=0.01$ ,  $q\text{-valueCutoff}=0.05$ ) adjusted by Benjamini-Hochberg procedure. Highly similar GO terms (e.g.,  $>0.25$ ) were removed by applying the “simplify” function to retain the most representative terms (i.e., the most significant) with parameters:  $\text{cutoff}=0.25$ ,  $\text{by} = “p.adjust”$ , and  $\text{select\_fun} = \text{min}$ .

#### Cluster comparer

In order to perform a biological theme comparison between the aforementioned sets of ASD-related genes, we used the “compareCluster” function, which calculates enriched functional profiles of each gene dataset and aggregates the results into a single object. For visualization purposes, the “showCategory” parameter, indicating the display of the topmost significant categories, was set to 5.

This tool was utilized to compare enrichments between: (i) the splicing gene list *versus* previously ASD-associated genes implicated in AS, and (ii) the splicing gene list *versus* genes harboring *de novo*/inherited variants with no predicted roles in AS.

#### Gene expression analysis

The GTEx Multi Gene Query tool of the GTEx Project version 8 (<https://www.gtexportal.org/home/>) was employed to carry out the gene expression heatmaps. Those genes harboring in silico validated splice variants were used as an input.

Expression values are represented as TPM (Transcripts Per Million), calculated as the number of reads for a gene and normalized by gene length. Additionally, different transcripts for each gene are collapsed during the normalization process.

Heatmaps display the average expression per tissue. Darker blue means higher relative expression of that gene in each label (tissue type), compared to a yellow/light-green color in the same label. Genes and tissues are ordered by cluster.

#### UCSC cell browser

The UCSC Cell Browser is an interactive tool for visualizing single-cell genomic data, including a specific dataset ASD (<https://autism.cells.ucsc.edu>). scRNA seq data were used to test if the genes carrying splicing variants with the highest ABSplice scores (e.g.  $\geq 0.2$ ) to understand if the genes containing variants with the highest ABSplice scores were differentially expressed in specific cell types, compared to genes containing variants with the lowest ABSplice scores (e.g.  $<0.2$ ). Gene expression plots showing multi-gene comparisons were constructed for each type of neuronal cell.

## Results

### In silico variant prediction and validation

Using previously reported *de novo* mutations in a Spanish cohort of 360 ASD trios<sup>42</sup>, several variant-level quality control filters were implemented. This process led to the identification of 43 high-confidence splicing variants. These variants demonstrated a SpliceAI  $\Delta \geq 0.8$  in at least one of the interrogated splice sites, resulting in four predictions: acceptor gain (AG), acceptor loss (AL), donor gain (DG), and donor loss (DL) (Fig. 1, Supplementary Table 1).

It is worth noting that achieving ideal in vitro validation would necessitate access to the tissue of relevance (presumably developing brain), which was not feasible. Consequently, we have undertaken validation through the utilization of diverse methods and supplementary RNA-seq datasets from brain and other tissues.

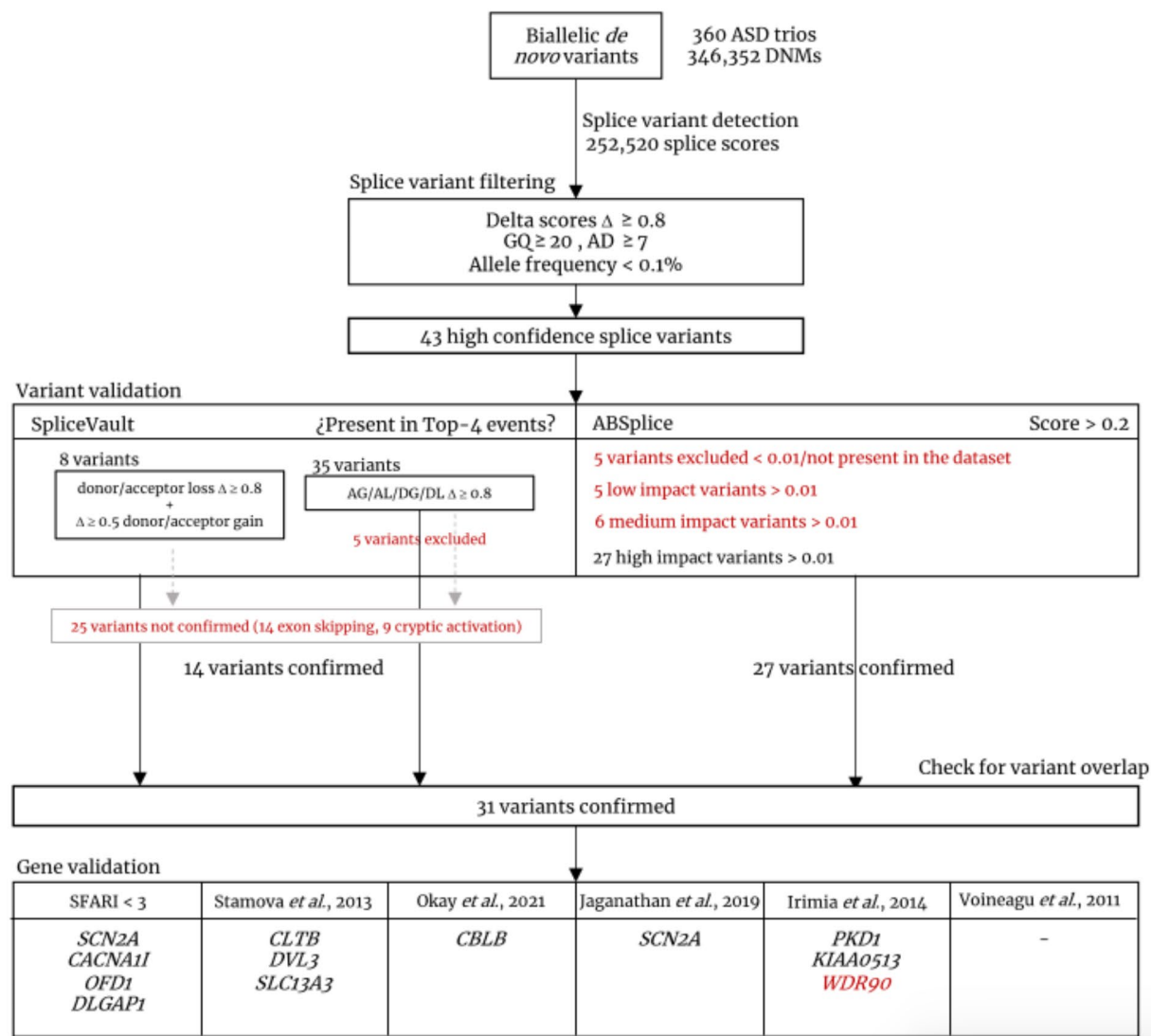
Several procedures were followed in order to ensure robustness in the in silico prediction of the splicing effects of these variants. Although the recommended threshold for splice-altering variants is  $\Delta \geq 0.5$ , we adopted a much more conservative threshold of  $\Delta \geq 0.8$ , which yields higher precision. This cutoff showed the highest validation rate and outperformed other popular classifiers that have been referenced in the literature for rare genetic disease diagnosis (GeneSplicer, MaxEntScan and NNSplice)<sup>39</sup>.

In addition, we reassessed all variants using SpliceVault, which quantifies natural variation in splicing and potentially predicts variant-related splicing changes (i.e., exon-skipping events and cryptic splice sites). Our dataset included 8 variants exhibiting cryptic donor/acceptor sites scores of  $\Delta \geq 0.8$  for site loss and  $\Delta \geq 0.5$  for site gain. Among these, 87.5% of the cryptic activation variants were validated ( $n=7$  present in the Top-4 events ranked by SpliceVault, Table 2). The remaining variant (chr1-16895732-C-T) resulted in exon 23 skipping in 51.9% of unannotated splice sites (Table 2). In order to detect single exon skipping events, we would need to observe: (i) one single variant with AL and DL  $\Delta \geq 0.8$ , or (ii) one individual harboring two different variants flanking the same exon with AL and DL  $\Delta \geq 0.8$ , respectively. Methodological limitations prevented us from validating this phenomenon. The use of exome sequencing data introduces the potential limitation that deep intronic variants, not detectable through this method, might be associated with exon exclusion.

Furthermore, 35 variants yielded  $\Delta \geq 0.8$  in only one out of four scored positions by SpliceAI (AG/AL/DG/DL). Excluding 5 variants with (i) no annotated splicing, (ii) gene was not present in SpliceVault server (FAM27B), or (iii) no cryptic annotation (non annotated splicing events); 30 variants were queried against the SpliceVault server (Table 3).

One variant with AG  $\Delta \geq 0.8$  and 3 with DL  $\Delta \geq 0.8$  were confirmed by SpliceVault to be correctly predicted. For the rest of the variants ( $n=26$ ), 50% ( $n=13$ ) Top-1 event resulted in exon skipping (11 single exon skipping and 2 double-exon skipping), while 13 variants resulted in cryptic activation, not detected in our method.

Further on, we sought to revalidate predicted variants against ABSplice<sup>41</sup>, with 60.46% ( $n=26$ ) of the variants yielding scores  $\geq 0.2$  (equivalent to the high precision cutoff  $\Delta \geq 0.8$  in SpliceAI) (Table 4), and were thus, confirmed.



**Fig. 1.** Workflow for in silico variant prediction, variant validation, and gene validation. Red font color is used to indicate variants that have not been confirmed or genes harboring unconfirmed variants. AD, alternate read depth; AG, acceptor gain; AL, acceptor loss; DG, donor gain; DL, donor loss; GQ, genome quality.

Variant (GRCh37)	Gene	AG	AL	DG	DL	SpliceVault check?
chr1-155981618-G-A	SSR2	0 (33)	0 (-47)	0.98 (-2)	0.92 (-17)	Y
chr7-1538341-A-C	INTS1	0 (-26)	0 (39)	0.90 (34)	1.00 (2)	Y
chr9-114176268-T-C	KIAA0368	0.66 (20)	0.98 (-2)	0 (21)	0 (-2)	Y
chr9-139407471-A-C	NOTCH1	0 (28)	0 (34)	0.90 (34)	1.00 (2)	Y
chr12-3649770-A-C	PRMT8	0.76 (10)	0.98 (2)	0 (8)	0 (1)	Y
chr16-85105388-G-T	KIAA0513	0.92 (5)	1.00 (1)	0 (2)	0 (-32)	Y
chrX-47,003,870-A-C	NDUFB11	0 (45)	0 (32)	0.71 (12)	1.00 (2)	Y
chr1-16895732-C-T	NBPF1	0.74 (-2)	0.94 (-1)	0 (-2)	0 (-1)	N (exon 23 skipping)

**Table 2.** Variants with an AL/DL  $\Delta \geq 0.8$  and AG/DG  $\Delta \geq 0.5$ . Variants are in GRCh37. Columns AG, AL, DG and DL show spliceai  $\Delta$  scores. Delta position (i.e., the location where splicing changes in relation to the variant's position) is shown between parenthesis (negative numbers refer to positions upstream of the variant while positive numbers refer to downstream positions). AG, acceptor gain; AL, acceptor loss; DG, donor gain; DL, donor loss.

Variant (GRCh37)	Gene	AG	AL	DG	DL	SpliceVault check?	Top 1 non-annotated event
chr1-20650027-T-C	VWA5B1	0 (-35)	0 (-41)	0 (-8)	0.94 (-2)	Y	NA
chr1-67242087-G-A	TCF7L1	0 (3)	0 (-50)	0 (3)	0.98 (-1)	Y	NA
chr5-843723-C-A	ZDHHC11	0 (37)	0 (-1)	0 (-8)	0.81 (0)	Y	NA
chr16-5141894-G-C	EEF2KMT	0.94 (1)	0 (45)	0 (-1)	0 (-32)	Y	NA
chr2-95539855-T-G	TEKT4	0 (-38)	0 (-40)	0.04 (-40)	0.98 (-2)	N	cryptic activation + 512
chr2-166170276-G-A	SCN2A	0 (6)	0 (-30)	0 (10)	0.85 (-5)	N	cryptic activation + 213
chr3-122629685-A-C	SEMA5B	0 (-32)	0 (49)	0.48 (24)	1 (2)	N	cryptic activation + 343
chr8-91033285-G-T	DECR1	0 (-2)	0 (-48)	0 (-2)	0.98 (-1)	N	cryptic activation - 51
chr10-118620666-A-G	ENO4	0 (-1)	0 (35)	0.90 (-1)	0.40 (35)	N	cryptic activation - 39
chr11-65784647-T-G	CATSPER1	0.44 (-9)	0.96 (-2)	0 (43)	0 (-6)	N	cryptic activation + 31
chr13-114005162-A-C	GRTP1	0 (1)	0 (2)	0 (-44)	0.88 (2)	N	cryptic activation - 52
chr16-711712-C-T	WDR90	0 (-2)	0 (-38)	0.92 (-2)	0 (-29)	N	cryptic activation - 77
chr16-2163160-A-C	PKD1	0 (7)	0 (25)	0.03 (-2)	0.86 (2)	N	cryptic activation + 5
chr16-20638576-A-T	ACSM1	0 (47)	0 (2)	0 (-25)	0.98 (2)	N	cryptic activation - 67
chr16-29473043-G-A	SULT1A4	0 (2)	0 (-43)	0.81 (1)	0.02 (16)	N	cryptic activation - 25
chr17-40835837-A-G	CNTNAP1	0.02 (16)	1 (2)	0 (21)	0 (-49)	N	cryptic activation - 158
chr19-10572358-T-G	PDE4A	0 (-14)	0 (-2)	0.18 (-14)	0.92 (-2)	N	cryptic activation + 35
chr3-105421304-C-A	CBLB	0.17 (-21)	0.92 (-1)	0 (-21)	0 (-15)	N	exon skipping (12)
chr4-110749291-T-G	RRH	0 (32)	0 (-2)	0.37 (12)	0.90 (-2)	N	exon skipping (4-5)
chr5-176958524-T-G	FAM193B	0.17 (-22)	0.98 (-2)	0 (-22)	0 (-2)	N	exon skipping (5)
chr5-179133258-G-A	CANX	0 (2)	0.94 (1)	0 (-50)	0 (29)	N	exon skipping (3)
chr9-78711019-G-A	PCSK5	0 (17)	0 (-1)	0 (10)	0.96 (-1)	N	exon skipping (8)
chr11-376072-A-C	B4GALNT4	0 (37)	0.96 (2)	0 (-36)	0 (1)	N	exon skipping (12)
chr15-42168847-T-G	SPTBN5	0.15 (-8)	0.85 (-2)	0 (47)	0 (-2)	N	double exon skipping (19-20)
chr16-22269096-G-T	EEF2K	0 (31)	0 (-37)	0 (50)	0.81 (-5)	N	exon skipping (9)
chr16-30910856-T-G	CTF1	0 (14)	0 (-16)	0 (-30)	0.81 (-2)	N	exon skipping (2)
chr16-56904007-G-A	SLC12A3	0.11 (2)	1.00 (1)	0 (-22)	0 (0)	N	exon skipping (5)
chr19-7686019-A-C	XAB2	0 (0)	0 (35)	0.05 (13)	0.96 (2)	N	exon skipping (9)
chr20-3641171-A-C	GFRA4	0 (-30)	0 (15)	0 (23)	0.98 (2)	N	exon skipping (3)
chrX-13,767,653-G-C	OFD1	0 (3)	0 (-45)	0.46 (3)	1 (-1)	N	exon skipping (9)
chr22-40060742-A-C	CACNA1I	0.14 (6)	1 (2)	0 (36)	0 (1)	*	*
chr11-118938598-C-G	VPS11	0 (-1)	0 (-18)	0.96 (-1)	0.01 (-13)	*	*
chr10-51130591-A-C	PARG	0 (-38)	0 (30)	0 (-43)	0.98 (2)	*	*
chr9-67793896-C-A	FAM27B	0 (-47)	0 (-42)	0.08 (-46)	0.96 (1)	**	**
chr18-3502489-A-G	DLGAP1	0 (-17)	0 (2)	0.04 (-17)	0.98 (2)	***	***

**Table 3.** Variants with AG/AL/DG/DL  $\Delta \geq 0.8$ . \*Non annotated splicing, \*\*gene not present in the dataset, \*\*\* no cryptic annotation. Variants are in GRCh37. Columns AG, AL, DG and DL shows SpliceAI  $\Delta$  scores. Delta position (i.e., the location where splicing changes in relation to the variant's position) is shown between parenthesis (negative numbers refer to positions upstream of the variant while positive numbers refer to downstream positions). For exon skipping events, the skipped exon is shown in parenthesis. For cryptic activation events in SpliceVault, the cryptic position is depicted. AG, *acceptor gain*; AL, *acceptor loss*; DG, *donor gain*; DL, *donor loss*.

After the validation process, 75.61% ( $n=31$ ) of the initially predicted splicing variants (excluding those variants not present in any of the complementary datasets ( $n=2$ )) were confirmed (Supplementary Table 1).

### Tissue specificity of predicted splice variants

Following cross-referencing with SpliceVault and ABRSplice, we further evaluated tissue-specific effects of the in silico validated variants ( $n=31$ ), albeit the score provided by ABRSplice. This approach allowed us to globally assess tissue-specific effects of all validated variants, acknowledging that some may not have high confidence in ABRSplice. Thus, variants that did not reach the 0.2 impact score threshold in ABRSplice but were validated in SpliceVault were also included in this analysis (Supplementary Table 1).

After removing one variant not present in the dataset, we evaluated tissue-specific effects in 26 variants with score  $\geq 0.2$  (high impact), 2 variants with score  $\geq 0.05$  (medium impact) and 2 with score  $\geq 0.01$  (low impact). Notably, adipose tissue yielded the highest scores for 38.7% of the variants ( $n=12$ ), followed by brain with 32.1% ( $n=9$ ), testis and adrenal gland with 9.6% each ( $n=3$ ), and the remaining 3 variants (each comprising 3.6% of the total) were distributed among nerve tibial, kidney cortex, and artery coronary.

Variant (GRCh37)	Gene	ABSplice score	ABSplice tissue
chr22-40060742-A-C	<i>CACNA1I</i>	<b>0.43</b>	Brain Cerebellum
chr12-3649770-A-C	<i>PRMT8</i>	<b>0.4</b>	Brain Nucleus accumbens basal ganglia
chr1-67242087-G-A	<i>TCTEX1D1</i>	<b>0.38</b>	Brain Frontal Cortex BA9
chr8-91033285-G-T	<i>DECRI</i>	<b>0.36</b>	Adipose Subcutaneous
chr9-139407471-A-C	<i>NOTCH1</i>	<b>0.36</b>	Adipose Subcutaneous
chrX-13,767,653-G-C	<i>OFD1</i>	<b>0.36</b>	Adipose Subcutaneous
chr9-78711019-G-A	<i>PCSK5</i>	<b>0.35</b>	Adipose Visceral Omentum
chr16-2163160-A-C	<i>PKD1</i>	<b>0.34</b>	Adipose Subcutaneous
chr16-85105388-G-T	<i>KIAA0513</i>	<b>0.34</b>	Adrenal Gland
chr2-95539855-T-G	<i>TEKT4</i>	<b>0.34</b>	Testis
chr7-1538341-A-C	<i>INTS1</i>	<b>0.34</b>	Adipose Subcutaneous
chr16-22269096-G-T	<i>EEF2K</i>	<b>0.33</b>	Adipose Subcutaneous
chr3-122629685-A-C	<i>SEMA5B</i>	<b>0.31</b>	Artery Coronary
chr5-179133258-G-A	<i>CANX</i>	<b>0.29</b>	Brain Amygdala
chr13-114005162-A-C	<i>GRTP1</i>	<b>0.28</b>	Adrenal Gland
chr16-56904007-G-A	<i>SLC12A3</i>	<b>0.28</b>	Kidney Cortex
chr2-166170276-G-A	<i>SCN2A</i>	<b>0.28</b>	Brain Cerebellar Hemisphere
chr16-30910856-T-G	<i>CTF1</i>	<b>0.27</b>	Adrenal Gland
chr19-7686019-A-C	<i>XAB2</i>	<b>0.26</b>	Adipose Subcutaneous
chr9-114176268-T-C	<i>KIAA0368</i>	<b>0.26</b>	Adipose Subcutaneous
chr11-376072-A-C	<i>B4GALNT4</i>	<b>0.24</b>	Brain Amygdala
chr17-40835837-A-G	<i>CNTNAP1</i>	<b>0.23</b>	Brain Anterior cingulate cortex BA24
chr3-105421304-C-A	<i>CBLB</i>	<b>0.23</b>	Adipose Subcutaneous
chr11-65784647-T-G	<i>CATSPER1</i>	<b>0.21</b>	Testis
chr18-3502489-A-G	<i>DLGAP1</i>	<b>0.21</b>	Brain Anterior cingulate cortex BA24
chr15-42168847-T-G	<i>SPTBN5</i>	<b>0.2</b>	Nerve Tibial
chr1-20650027-T-C	<i>VWA5B1</i>	0.18	Testis
chr1-16895732-C-T	<i>NBPF1</i>	0.18	Brain Cerebellar Hemisphere
chr5-176958524-T-G	<i>FAM193B</i>	0.17	Adipose Visceral Omentum
chr16-29473043-G-A	<i>SULT1A4</i>	0.09	Brain Cerebellum
chr16-5141894-G-C	<i>EEF2KMT</i>	0.07	Adipose Visceral Omentum
chr16-20638576-A-T	<i>ACSM1</i>	0.055	Testis
chr5-843723-C-A	<i>ZDHHC11</i>	0.045	Brain Cerebellar Hemisphere
chr1-155981618-G-A	<i>SSR2</i>	0.04	Adipose Subcutaneous
chr20-3641171-A-C	<i>GFRA4</i>	0.04	Brain Amygdala
chr10-118620666-A-G	<i>ENO4</i>	0.032	Testis
chr16-711712-C-T	<i>WDR90</i>	0.032	Adipose Subcutaneous
chr19-10572358-T-G	<i>PDE4A</i>	0.021	Testis
chr4-110749291-T-G	<i>RRH</i>	<0.01	NA
chr11-118938598-C-G	<i>VPS11</i>	<0.01	NA
chr9-67793896-C-A	<i>FAM27B</i>	*	NA
chr10-51130591-A-C	<i>PARG</i>	*	NA
chrX-47,003,870-A-C	<i>NDUFB11</i>	*	NA

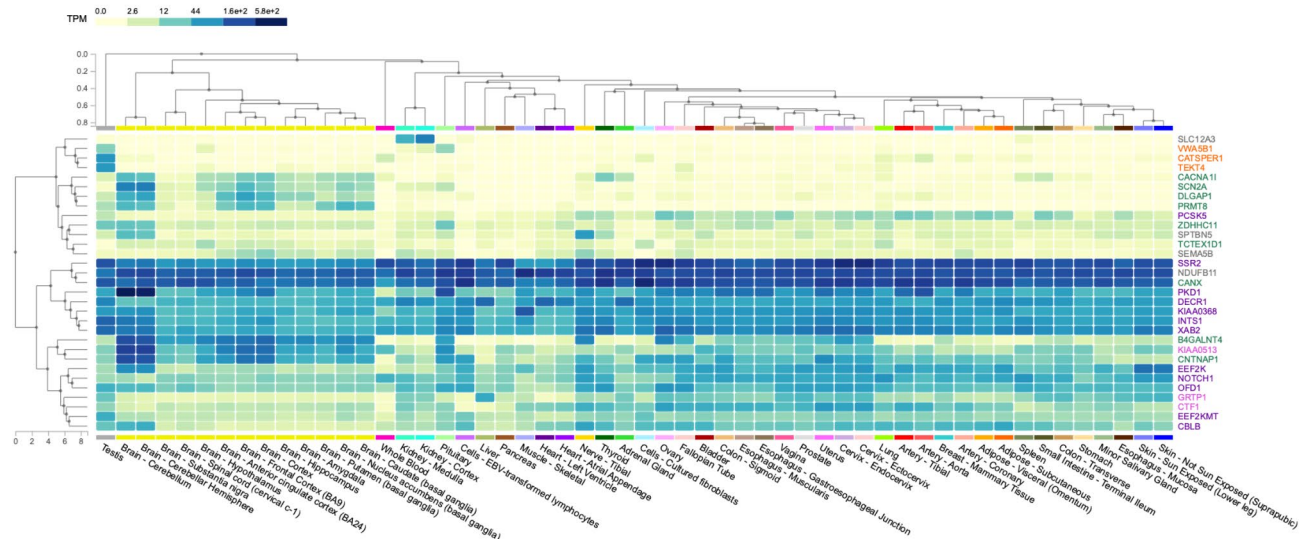
**Table 4.** ABSplice prediction. Variants are sorted by absplice scores. Variants with scores  $\geq 0.2$  (bold font) were confirmed. \*Variant not present.

Then, genes harboring the in silico validated splice altering variants were queried against the GTEx portal to assess whether the predicted tissue-specific effects were attributable to gene expression restricted to that particular tissue.

Overall, genes like *CACNA1I*, *SCN2A*, *DLGAP1* and *PRMT8*, which harbor variants predicted to have their highest impact in brain tissue, do show higher expression values restricted to brain tissues (Fig. 2, gene name in green). Only one of these genes (namely, *CANX*), with a validated variant predicted to have the highest impact in amygdala, did not show a high expression limited to the brain.

However, genes that host splicing variants with the highest impact in adipose tissue (Fig. 2, gene name in purple) do not exhibit expression limited to any specific tissue. This is in contrast to the expectation, as a specific expression limited to adipose tissue would provide a logical rationale for the increased burden of variants yielding the highest scores in adipose tissue.





**Fig. 2.** Tissue specific expression of genes harboring 31 in silico validated variants. Gene expression heatmap for genes harboring splice variants with predicted tissue-specific impact. Genes and tissues are ordered by cluster. Genes are highlighted based on the tissue where their splicing variant yields the highest score: green for the brain, purple for adipose tissue, orange for testis, and gray for other tissues.

Nonetheless, variants were checked against the whole set of tissues, and it was observed that most of the variants with predicted highest scores in brain, and all the variants with predicted highest effect in adipose tissue and adrenal gland, yielded the same high scores in other tissues (i.e., ABSplice scores were not exclusive for that tissue) (data not shown).

In contrast, 3 variants with the highest score identified in testis, exhibited a clear tissue-specificity (i.e., the ABSplice scores retrieved for the remaining GTEx tissues of tissues were notably lower) (Supplementary Fig. 1). However, genes harboring these variants (*CATSPER1*, *TEKT4*, *VWA5B1*) had tissue specific expression (Fig. 2, gene name in orange) restricted to testis.

### Cluster enrichment

Genes harboring high-confidence splice variants predicted by SpliceAI and validated with SpliceVault and ABSplice ( $n=31$  variants, one variant per gene) were cross-referenced with: (i) previously reported genes associated with ASD AS<sup>25,28,29,36,39</sup>, to check for similarities in splicing relevant pathways and perform gene validation, (ii) ASD-associated genes in the SFARI Gene (category 1: high confidence, category 2: strong candidate), for gene validation only, and (iii) genes harboring *de novo* and inherited coding mutations<sup>45</sup>, to see if different types of mutations act through distinct mechanisms.

*CACNA11*, *CBLB*, *CLTB*, *DLGAP1*, *DVL3*, *KIAA0513*, *OFD1*, *PKD1*, *SLC13A3* and *SCN2A* were present in at least one of the above-mentioned datasets and were thus used to construct a brain-specific network of functionally related genes (Supplementary Fig. 2, Supplementary Table 2,  $n=60$ ). The resultant gene network was interrogated for enrichment in the GO categories of BP, MF, and CC.

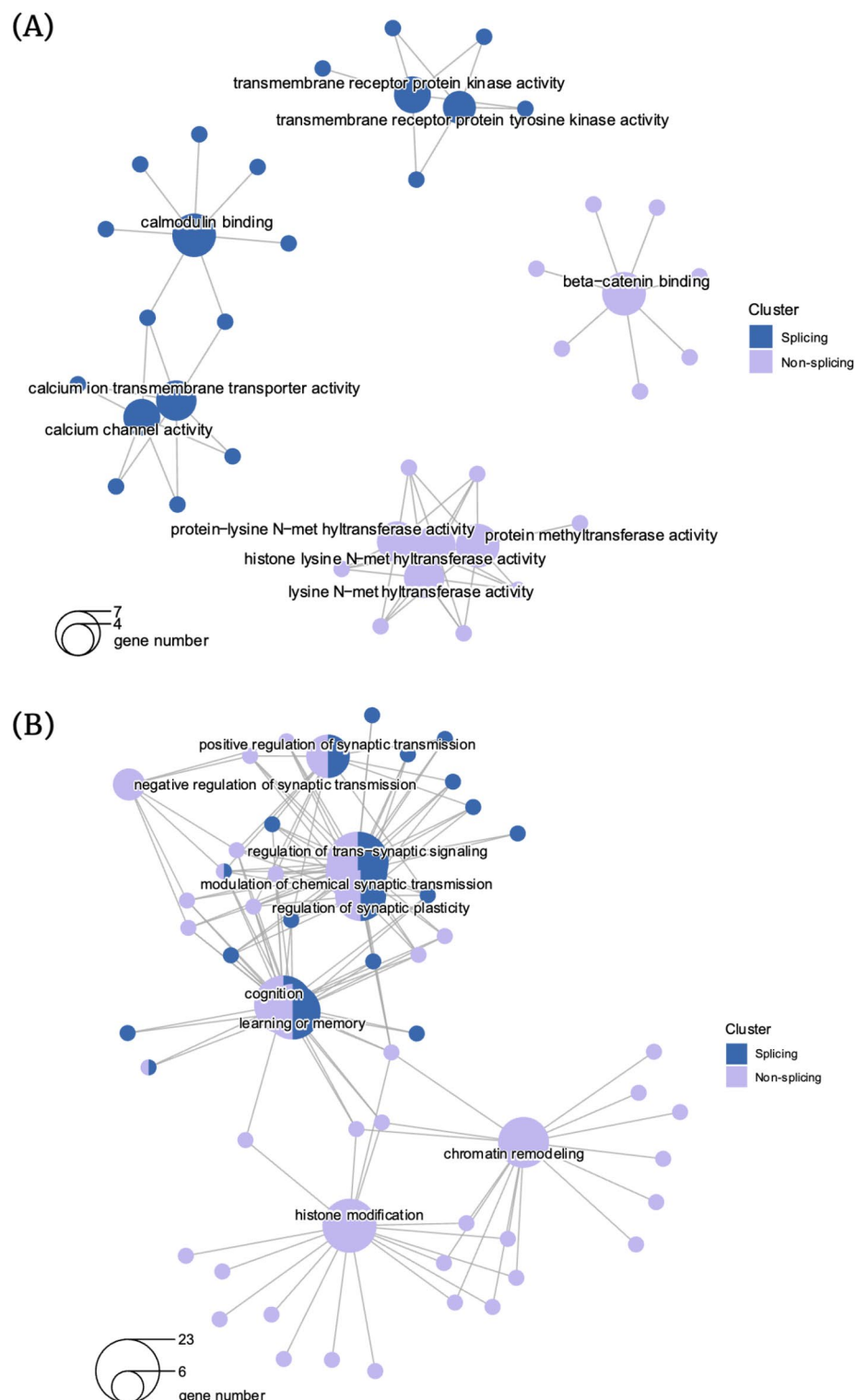
The analysis revealed that these genes were significantly enriched for biological processes related to proper neuronal functioning (modulation of chemical synaptic transmission (gene ratio 12/60,  $q$ -value =  $1.68 \times 10^{-5}$ ), trans-synaptic signaling (gene ratio 12/60,  $q$ -value =  $1.68 \times 10^{-5}$ ), synaptic plasticity (gene ratio 8/60,  $q$ -value =  $1.07 \times 10^{-4}$ ), cognition (gene ratio 8/60,  $q$ -value =  $1.07 \times 10^{-4}$ ), and memory (gene ratio 9/60,  $q$ -value =  $1.68 \times 10^{-4}$ ).

Enriched CC terms were all related to the synapse, with top significant findings including postsynaptic specialization (gene ratio 12/60,  $q$ -value =  $1.12 \times 10^{-7}$ ), synaptic membrane (gene ratio 10/60,  $q$ -value =  $7.84 \times 10^{-6}$ ), glutamatergic synapse (gene ratio 10/60,  $q$ -value =  $1.03 \times 10^{-5}$ ), and presynaptic synapse (gene ratio 12/60,  $q$ -value =  $3.34 \times 10^{-4}$ ), among others.

Top enriched MF terms were associated with calmodulin binding channels (gene ratio 7/60,  $q$ -value =  $3.25 \times 10^{-4}$ ), transmembrane receptor protein kinase activities (gene ratio 5/60,  $q$ -value =  $3.25 \times 10^{-4}$ ), calcium ion channels (gene ratio 6/60,  $q$ -value =  $3.25 \times 10^{-4}$ ), and tyrosine activities (gene ratio 4/60,  $q$ -value =  $1.67 \times 10^{-4}$ ), among others.

Furthermore, we compared functional profiles amongst the different datasets and calculated enriched functional profiles of each gene cluster. In analyzing datasets for genes implicated in AS in ASD, we found that our gene network clusters together in terms of BP, CC and MF (Supplementary Figs. 3–5). Examples of common significantly enriched terms included: (i) protein autophosphorylation and modulation of chemical synaptic transmission, for BP, (ii) postsynaptic specialization and cell leading edge, for CC, and (iii) acting/calmodulin binding, for MF.

However, when contrasting these findings with genes harboring coding variants (henceforth designated as the non-splicing gene list), the overlap between enriched categories is notably dissipated (Fig. 3, Supplementary



**Fig. 3.** Cluster enrichment analysis for splicing vs. inherited/de novo datasets. Graphs depicting the number of genes of each list included in (a) molecular functions, (b) biological processes. The circle size is proportional to the number of genes included in each category, and not significant as depicted in the legend (both panels have different size ratios).

Fig. 6). CC terms did not exhibit a clear separation between datasets, with all common enriched terms relating to synapse components or postsynaptic density (Supplementary Fig. 6). While some BP (e.g., cognition, learning and memory) were significantly enriched in both datasets, others (e.g., histone modification (gene ratio 17/72,  $q$ -value =  $3.77 \times 10^{-9}$ ) and chromatin remodeling (gene ratio 15/72,  $q$ -value =  $2.45 \times 10^{-8}$ )) were specifically

enriched in the non-splicing gene list, absent in our 60-gene list (Table 5). On the other hand, genes from both datasets were incorporated into categories associated with the proper function and organization of the synapse. However, the majority of significant enrichments in the splicing gene list (48 out of 62 enriched BP terms) were exclusively identified within that particular dataset. Some examples of the top enrichments are provided in Table 5.

The most pronounced difference emerged when analyzing enriched terms in the MF category: genes with predicted splice variants were significantly enriched in terms such as calmodulin binding (gene ratio 7/60, q-value =  $3.25 \times 10^{-4}$ ), calcium ion channel/transporter activity (gene ratio 5/60, q-value =  $3.25 \times 10^{-4}$ ), and transmembrane receptor protein kinase/tyrosine activity (gene ratio 6/60, q-value =  $3.25 \times 10^{-4}$ ), while genes in the non-splicing list were specific to histone lysine N-methyltransferase activity (gene ratio 6/72, q-value =  $3.85 \times 10^{-6}$ ) and beta-catenin binding (gene ratio 7/72, q-value =  $3.85 \times 10^{-6}$ ) (Fig. 3).

scRNA-seq analysis

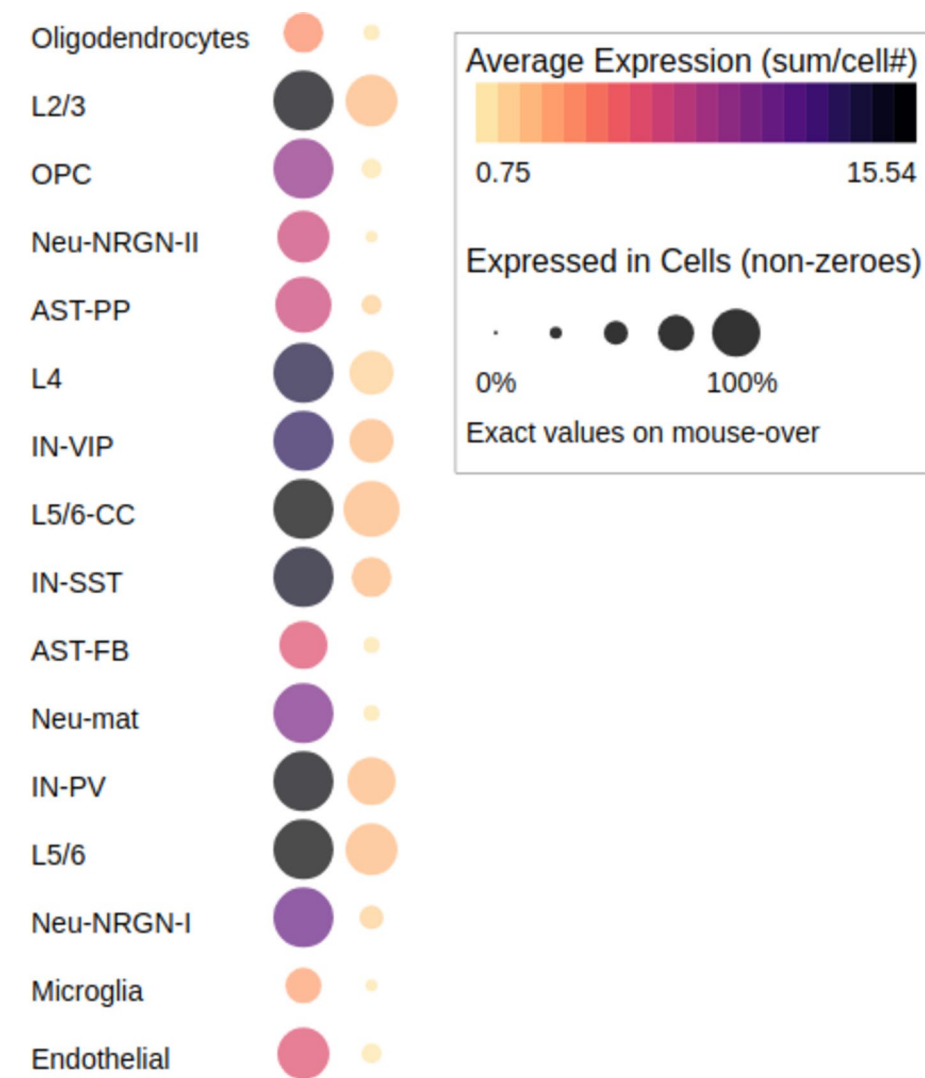
Regarding the scRNA seq analysis we have added information separating genes with ABSplicescore >= 0.2 and presenting a lower score, to understand if the genes containing variants with the highest ABSplice scores are differentially expressed in specific cell types, compared to genes containing variants with the lowest ABSplice scores (Fig. 4).

We found that overall genes containing variants with the highest ABSplice scores are overexpressed among all neuronal cell types and genes containing variants with the lowest ABSplice scores exhibit an underexpression among all neuronal cell types. The findings from the UCSC Browser regarding ABSplice scores and gene expression in neuronal cell types are consistent with current understanding of alternative splicing (AS) in the context of neuronal diversity and function. Genes containing variants with the highest ABSplice scores being overexpressed among all neuronal cell types suggests that alternative splicing plays a crucial role in neuronal gene regulation and function and these genes likely contribute to neuronal cell type-specific properties and may be involved in critical neuronal processes.

Conversely, the underexpression of genes containing variants with the lowest ABSplice scores across all cell types indicates that these genes might be less important for neuronal-specific functions or may have more generalized roles across various cell types. Low ABSplice scores could suggest less complex splicing patterns, which may not be as critical for neuronal diversity.

Gene list	Description	Gene ratio	q-value
Splicing gene list	regulation of signaling receptor activity	6/60	$2.79 \times 10^{-3}$
	regulation of neurotransmitter receptor activity	4/60	$6.00 \times 10^{-3}$
	protein autophosphorylation	6/60	$8.17 \times 10^{-3}$
	regulation of JNK cascade	5/60	$8.17 \times 10^{-3}$
	dendrite development	6/60	$8.17 \times 10^{-3}$
	regulation of protein catabolic process	7/60	$1.02 \times 10^{-2}$
	positive regulation of MAPK cascade	8/60	$1.08 \times 10^{-2}$
	JNK cascade	5/60	$1.34 \times 10^{-2}$
	regulation of phosphatidylinositol 3-kinase signaling	4/60	$2.15 \times 10^{-2}$
	regulation of synapse assembly	4/60	$2.15 \times 10^{-2}$
	calcium ion transport	7/60	$2.25 \times 10^{-2}$
	excitatory postsynaptic potential	4/60	$2.46 \times 10^{-2}$
Non-splicing gene list	histone modification	17/72	$3.77 \times 10^{-9}$
	chromatin remodeling	15/72	$2.45 \times 10^{-8}$
	histone lysine methylation	7/72	$4.11 \times 10^{-5}$
	histone methylation	7/72	$1.20 \times 10^{-4}$
	histone H3-K4 methylation	5/72	$3.40 \times 10^{-4}$
	regulation of histone methylation	3/72	$2.07 \times 10^{-2}$
	regulation of histone modification	4/72	$2.57 \times 10^{-2}$
	positive regulation of histone H3-K4 methylation	2/72	$3.05 \times 10^{-2}$
	histone lysine demethylation	2/72	$3.50 \times 10^{-2}$
	histone demethylation	2/72	$3.67 \times 10^{-2}$
Common	learning or memory (splicing gene list)	8/60	$4.34 \times 10^{-4}$
	learning or memory (non-splicing gene list)	2/72	$2.52 \times 10^{-7}$
	cognition (splicing gene list)	9/60	$1.68 \times 10^{-4}$
	cognition (non-splicing gene list)	13/72	$1.24 \times 10^{-7}$

**Table 5.** Common/unique enriched biological processes in splicing-related gene lists and the non-splicing gene list.



**Fig. 4.** Gene expression plots from UCSC browser using scRNAseq data from Velmeshev et al. Science. 2019. First column includes all genes from Table 4 with  $ABScore > 0.2$  (bold). Second column includes all genes from Table 4 with  $ABScore < 0.2$  except *WDR90* not found on the scRNAseq dataset.

## Discussion

### Context of the study and limitations

#### *Genetic data and in silico analysis*

The multifaceted genetic etiology of ASD, characterized by substantial phenotypic and genetic heterogeneity, has long been a challenge in unraveling its underlying molecular basis. The identification of splicing variants has not been included in the major WGS or WES genetic studies involving large ASD cohorts. However, AS, an intricate mechanism that diversifies protein isoforms from a single gene, has recently garnered attention as a potential contributor to ASD pathogenesis.

We acknowledge that the sample size of the study is limited. A significant challenge in our research was the restricted access to certain ASD datasets, even those under controlled access, which are intended to be available to qualified researchers. While these restrictions are meant to ensure responsible data use, excessive bureaucratic barriers hinder scientific progress. Additionally, we note that ensuring all exome datasets are unified in terms of sequencing platforms and other quality control parameters could also be a limitation, which may have contributed to maintaining a more limited sample size for this study.

Although our study focused on *de novo* biallelic variants, we acknowledge that inherited variants also play a significant role in the genetic architecture of ASD. While *de novo* variants, particularly those affecting splicing, have been strongly associated with ASD and other NDDs, inherited variants contribute substantially to ASD susceptibility through diverse mechanisms, including recessive inheritance, compound heterozygosity, and polygenic risk. Family-based studies have shown that inherited rare variants can impact ASD risk, particularly in multiplex families, where the recurrence rate is higher than in sporadic cases. Moreover, polygenic risk scores highlight the contribution of common inherited variants to ASD susceptibility. Our study was designed

to prioritize de novo biallelic variants due to their higher predicted impact and lower background genetic variability. Future studies incorporating inherited splicing variants, particularly in familial ASD cases, could provide a more comprehensive understanding of the contribution of splicing alterations to ASD etiology.

The present study delved into the intricate landscape of AS in ASD through in silico prediction and validation of splicing variants. However, the conservative threshold (SpliceAI  $\Delta \geq 0.8$ ) chosen for splice-altering variant prediction<sup>39</sup> may, in turn, result in elevated numbers of false negatives. While this approach was necessary, in vitro confirmation of the predicted variants (e.g., by Sanger sequencing), validation of the predicted alterations on AS (by reverse transcription polymerase chain reaction (RT-PCR)) and functional analysis of their molecular impacts (RNA-Seq and/or minigene reporter systems<sup>47</sup>), would prove much more adequate and sensitive. Assessing the functional impact of these variants using RT-PCR would allow for direct measurement of the resulting splicing products, confirming whether the variants lead to exon skipping, activation of cryptic splice sites, or other forms of aberrant splicing. Beyond validating the presence of splicing events, functional analyses are critical to understanding the broader consequences on protein function. Techniques such as RNA sequencing (RNA-Seq) or minigene reporter systems could be employed to evaluate how these splicing alterations affect the expression, stability, and activity of the encoded proteins. This could provide insights into how specific splicing changes disrupt cellular processes that are crucial for neurodevelopment and contribute to ASD pathogenesis. For example, abnormal splicing in genes involved in synaptic function, cell signaling, or neuronal differentiation could lead to defective protein-protein interactions, altered synaptic plasticity, or impaired neuronal connectivity, all of which are hallmarks of ASD.

Still, in vitro validation was unfeasible due to the lack of sample availability and the difficulty in contacting participants for resampling. Thus, our validation strategy using SpliceVault and ABSplice, at least partially, provided further evidence on the robustness of the predicted splicing effects. On the one hand, SpliceVault exhibited superior sensitivity and positive predictive value than SpliceAI when it comes to exon- and double-exon skipping predictions or cryptic splice site activation, and represents the first evidence-based method for predicting the nature of variant-associated mis-splicing<sup>40</sup>. On the other hand, another study demonstrated that applying SpliceAI on the tissue-specific splice sites defined by SpliceMap (integrated into ABSplice) increased the precision of SpliceAI to 22% at 20% recall, with a significantly higher auPRC consistently across tissues<sup>41</sup>. Notably, our validation approach revealed that the majority of cryptic activation events were successfully corroborated when leveraging evidence-based data (Supplementary Table 1). This further demonstrates the role of these predicted cryptic sites in ASD-associated splicing perturbations. Furthermore, 35% of cryptic splice variants with weak and intermediate predicted scores ( $\Delta$  0.35–0.8) exhibit significant differences in the fraction of normal and aberrant transcripts produced across tissues. Variants with high-predicted scores are significantly less likely to produce tissue-specific effects<sup>39</sup>. Therefore, being able to choose a less conservative threshold and perform in vitro validation, would be tremendously helpful in gaining insight into tissue-specific effects. However, the tissue-specific effects of splicing variants gained prominence through our assessment using ABSplice.

Another limitation of our study is the use of WES data, which primarily targets coding regions and proximal intronic areas, potentially overlooking deep intronic variants. These deep intronic variants may play a crucial role in gene regulation and splicing, yet they are not captured by WES. This limitation could result in the underrepresentation of important non-coding variants that may contribute to neurodevelopmental disorders like ASD. Nevertheless, splicing branch points present an additional source of potentially damaging non-coding variants, which are amenable to systematic analysis in WGS data<sup>48</sup>, but remain undetectable in WES data. This represents another methodological constraint in our approach. Of note, the availability of WGS data could enhance our understanding of the splicing landscape in ASD by enabling the detection of intron retention, a splicing aberration already associated with ASD and other NDDs<sup>49,50</sup>.

### Tissue-specificity in alternative splicing (AS) in ASD

Research in ASD has primarily focused on neurological aspects, looking at factors such as brain structure and function, and neurotransmitter systems. However, it is worth noting that there is ongoing research in the field of neuroimmunology and the gut-brain axis, which explores the connections between the gut and the brain. Recently, GI dysfunction has been described in various neurodevelopmental and psychiatric disorders including ASD<sup>51,52</sup>. Moreover, some studies have suggested a possible link between GI (gastrointestinal) tissue, adipose tissue and brain, with accumulating evidence suggesting that the communication pathways linking them might be promising intervention points for metabolic disorders<sup>53</sup>. However, since adipose tissue can produce certain signaling molecules, such as adipokines, it is possible that there could also be indirect connections between adipose tissue and neurodevelopmental conditions like ASD<sup>54</sup>, but this area of research is still emerging and not yet well-understood. There are some studies regarding the role of adipokines in neurogenesis, neuroprotection, synaptogenesis, synaptic plasticity, and even neurodegenerative diseases such as Alzheimer's disease<sup>55–58</sup>. However, we urge caution in interpreting our results, as they are based on in silico validation with RNA-seq data from GTEx and not from ASD patients. Functional in vivo analyses (e.g., iPSCs, organoids) are necessary to confirm the potential connections between adipose tissue and NDDs.

Importantly, preliminary findings of this study indicated a role of adipose tissue in ASD (Table 4). Yet, upon examining the expression of genes containing variants with adipose tissue-specific effects we noted a relatively uniform expression across all tissues (Fig. 2). Notably, these genes exhibited elevated expression levels in the brain, and the associated splicing variants also yielded high scores in brain tissues. Consequently, the function of adipose tissue remains unclear. However, considering this in silico evidence alongside previous studies, further research is necessary. Further studies could test this hypothesis by interrogating whether these genes are driving pleiotropic effects in both sets of tissues, or by performing overall comparisons between splice site usage in neuronal *versus* adipose tissue.



Moreover, three variants in the final set of splicing validated variants show unique values of tissue-specificity in testis (Supplementary Fig. 1). Nonetheless, when comparing transcriptomes, it has been observed that the brain and testis significantly surpass other tissues in terms of the diversity of expressed splice variants<sup>59</sup>. Consequently, our findings may lack sufficient power to attribute specific significance to testis in the context of ASD risk.

### Biological underpinnings of alternative splicing (AS)

On another note, the convergence of genes harboring validated splicing variants with previously reported ASD-associated genes from various datasets substantiates the potential significance of AS in ASD. Our creation of a brain-specific network encompassing functionally related genes (Supplementary Fig. 2, Supplementary Table 2) demonstrated enrichment in BP intricately tied to neuronal functioning, synaptic transmission, synaptic plasticity, cognition, and memory (Fig. 3). Although we identified a relatively small number of genes with in silico validated splice variants, these findings align with previous studies showcasing aberrant splicing patterns in genes critical to neural development, which may collectively contribute to the complex ASD phenotype.

Additional support for this evidence includes: (i) analyzing larger ASD cohorts under the same criteria, both to augment our splicing gene-list (and thus provide more statistical support for enrichments in GO categories) and to perform a burden test analysis of numbers of mutations (are the numbers of splicing variants per gene consistent with gene length, conservation, and the number of different isoforms? ), (ii) testing whether the splicing genes carry an excess of non-splice DNVs in autism probands to further correct this measure, and (iii) using multiplex family/case-control cohorts to check if splice DNVs are enriched in affected individuals when compared with healthy siblings.

In addition, we cross-referenced our gene list with genes that have exome-wide significance when combining evidence from both coding DNVs and rare inherited variants (non-splicing gene list), to encompass a broader spectrum of the disorder. Interestingly, our gene list shows significant enrichment in MF different from those enriched in the non-splicing gene list (Fig. 3A). Moreover, none of these non-splicing enriched terms were found in any other splicing complementary dataset analyzed in this study. Similar results were observed for the category of BP (Fig. 3B), where cognition and learning or memory terms are common amongst both lists (splicing *versus* non-splicing), but chromatin remodeling and histone modification are specific to the non-splicing gene list, and most synaptic-related terms are specific to the splicing gene list (Table 5).

The fact that GO enrichment analysis points out to different MF and BP in genes harboring splicing variants and in the non-splicing gene list, might suggest a divergence of affected pathways and mechanisms, thus pointing to different mechanisms in which they participate in the development of the disease.

In fact, a previous study on the full-length isoform transcriptome of the developing human brain<sup>15</sup> has shown that differentially expressed isoforms (DEIs) reveal distinct signals relative to differentially expressed genes (DEGs). The GO enrichment analysis demonstrated stronger enrichment of DEI in neurodevelopment-relevant processes compared with DEGs. In contrast, DEGs were enriched in basic biological function-related processes, such as mitotic cell cycle, metabolic processes, protein targeting, and localization. Also, molecular functions such as kinase activity (one out of three most significant molecular functions strictly associated in our gene list) were solely linked to DEIs. In vitro studies using human cortical neurons treated with the anticonvulsant valproic acid (VPA) have shown that differentially expressed genes (DEGs) exhibit enrichment in distinct molecular functions compared to genes with differential transcript isoform usage (DTU)<sup>60</sup>. These findings align with those presented in this study, suggesting that the full-length isoform transcriptome provides better biological insights into brain development than the gene transcriptome. However, when comparing these results at the level of molecular functions specific to DEGs or DTUs, caution should be exercised, since these in vitro experiments involve specific lines of neurons from ASD patients, unlike our in silico experiments, using GTEx RNA-seq data from brain and other tissues. Therefore, further studies utilizing transcriptomic methods in brain samples from ASD patients and controls, combined with WGS data, are needed to address this question.

In general, this situation mirrors the still unresolved question of whether *de novo* and inherited variations affect the same biological pathways. Ruzzo et al. illustrated that inherited variations cluster in specific biological pathways, introducing novel pathways linked to ion transport, the cell cycle, and the microtubule cytoskeleton<sup>61</sup> distinct from those enriched for *de novo* variants<sup>6</sup>. In a separate study, Wilfert and colleagues stated that DNVs and transmitted LGD variants converge on the same pathway but may be targeting distinct sets of genes<sup>62</sup>.

Our investigation into the cell-type-specific expression of genes implicated in adiposity and ASD revealed interesting patterns that warrant further exploration. Using publicly available autism-related data (<https://cells.ucsc.edu/>), we assessed the expression of all the related adipose genes except *WRD60* (not found in the database) in various neuronal cell types. Notably, we observed significant overexpression of these genes in Layer 5/6 excitatory cortical neurons, a cell population previously involved in ASD. In contrast, there was no detectable overexpression in oligodendrocytes or microglia, which suggests that the genes associated with adiposity may not be directly involved in oligodendrocyte/myelin development in the context of ASD. This finding highlights the possibility that these genes may contribute to ASD pathophysiology through alterations in excitatory neuronal circuits, rather than through mechanisms related to myelin formation or glial cell function. Future studies will need to further investigate the role of these genes in neuronal development and function, as well as their potential interactions with other ASD-related pathways.

### Final remarks and future perspectives

In essence, this study exemplifies the intricate genetic landscape of ASD and aims to raise new questions regarding the involvement of AS in this neurodevelopmental disorder, proposing novel avenues for future research. The comprehensive in silico validation pipeline employed here showcases the potential of such methods in deciphering splicing perturbations.

However, as with any computational approach, in vitro validation is paramount to fully comprehend the functional consequences of these predicted splicing changes. Further studies should expand on the identified splicing events' downstream effects on protein function, signaling pathways, and cellular processes. Integration of our findings with multi-omics data, such as full-length isoform transcriptomics, may provide a more holistic view of the intricate ASD molecular network. Moreover, examining the potential interplay between splicing and other regulatory mechanisms, such as epigenetics, could elucidate additional layers of complexity in ASD etiology.

While numerous questions remain unanswered, and the functional validation of most predicted splice-disrupting variants is still necessary to affirm a molecular diagnosis, in silico tools' predictions can serve as supportive evidence in variant classification. Commonly used variant annotation tools are not designed to assess the deleterious impact of splicing variants and their predictions are largely restricted to canonical splice sites. However, if multiple computational sources, such as the framework presented here, indicate that a variant has a deleterious effect, these predictions can be employed. Furthermore, they can aid in prioritizing splice-disrupting variants for subsequent functional testing or experimental validation.

Our results emphasize the overlap of splicing-related variants with synaptic signaling pathways, especially concerning microexons and neuron-specific splicing control, which have surfaced as crucial mechanisms in ASD. This finding corresponds with the increasing literature highlighting the significance of synaptic signaling in the development of ASD. Additionally, our splicing predictions emphasize the significance of neuron-specific splice isoforms in the development of ASD, providing new perspectives on how alternative splicing could influence the disease's phenotype.

In comparison, variants that appeared not to affect splicing were significantly enriched in chromatin remodeling genes, aligning with earlier research that has recognized chromatin regulation as another important biological category in the genetics of ASD. The stochastic tendency towards chromatin-related genes in the non-splicing group may be affected by constraints in sample size and statistical power, challenges that often arise when researching rare variants. We take into account Zhou et al. (2022) results<sup>45</sup>, which combined both de novo and inherited variants in an extensive ASD cohort to pinpoint moderate-risk genes. Even though these variants were not expected to influence splicing, their role in chromatin remodeling processes emphasizes a significant aspect of ASD that deserves additional exploration.

This study underscores the utility of computational predictions in identifying splicing variants. It is important to clarify that the main objective of this study is to investigate the molecular mechanisms underlying ASD, rather than to identify diagnostic markers or therapeutic targets.

However, to comprehensively address the involvement of AS processes in ASD etiology, several functional and in vitro studies are needed in the near future to solve some of the limitations previously described in this discussion. These include: (i) the validation of splicing variants (by RT-PCR or RNA-seq), (ii) functional characterization, performing functional assays to elucidate how splicing variants influence protein function and pathway activity (e.g., iPSCs and/or animal models), (iii) integrative omics approaches, integrating splicing variant data with other omics data (e.g., genomic, epigenomic, proteomic) to gain a comprehensive understanding of the molecular mechanisms underlying ASD, and (iv) to deeper explore the clinical implications of splicing variants in ASD, including potential biomarker discovery and therapeutic targets. These future studies will aim, together with the in silico workflow using AI tools as presented in this study, to advance our understanding of splicing perturbations in ASD and their broader implications.

## Data availability

WES (Whole Exome Sequencing) data from the Spanish cohort were generated as part of the ASC and are transferred to dbGaP with Study Accession: phs000298.v4.p3. Previously published in Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, Peng M, Collins R, Grove J, Klei L, Stevens C, Reichert J, Mulhern MS, Artomov M, Gerges S, Sheppard B, Xu X, Bhaduri A, Norman U, Brand H, Schwartz G, Nguyen R, Guerrero EE, Dias C; Autism Sequencing Consortium; iPSYCH-Broad Consortium; Betancur C, Cook EH, Gallagher L, Gill M, Sutcliffe JS, Thurm A, Zwick ME, Børghlum AD, State MW, Cicek AE, Talkowski ME, Cutler DJ, Devlin B, Sanders SJ, Roeder K, Daly MJ, Buxbaum JD. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*. 2020 Feb 6;180(3):568–584.e23. doi: 10.1016/j.cell.2019.12.036. Epub 2020 Jan 23. PMID: 31981491; PMCID: PMC7250485. Splicing variants identified in this research are provided within the manuscript and the supplementary information.

Received: 23 September 2024; Accepted: 21 March 2025

Published online: 28 March 2025

## References

1. American Psychiatric Association, American Psychiatric Association, eds. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5* 5th edn (American Psychiatric Association, 2013).
2. Sandin, S. et al. The heritability of autism spectrum disorder. *JAMA* **318** (12), 1182. <https://doi.org/10.1001/jama.2017.12141> (2017).
3. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51** (3), 431–444. <https://doi.org/10.1038/s41588-019-0344-8> (2019).
4. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism*. **8** (1), 21. <https://doi.org/10.1186/s13229-017-0137-9> (2017).
5. Gaugler, T. et al. Most genetic risk for autism resides with common variation. *Nat. Genet.* **46** (8), 881–885. <https://doi.org/10.1038/ng.3039> (2014).

6. De Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515** (7526), 209–215. <https://doi.org/10.1038/nature13772> (2014).
7. Sanders, S. J. et al. De Novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485** (7397), 237–241. <https://doi.org/10.1038/nature10945> (2012).
8. Trost, B. et al. Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature* **586** (7827), 80–86. <https://doi.org/10.1038/s41586-020-2579-z> (2020).
9. Werling, D. et al. Limited contribution of rare, noncoding variation to autism spectrum disorder from sequencing of 2,076 genomes in quartet families. *Eur. Neuropsychopharmacol.* **29**, S784–S785. <https://doi.org/10.1016/j.euroneuro.2017.08.010> (2019).
10. Sun, Y. et al. Target genes of autism risk loci in brain frontal cortex. *Front. Genet.* **10**, 707. <https://doi.org/10.3389/fgene.2019.00707> (2019).
11. Arpi, M. N. T. & Simpson, T. I. SFARI genes and where to find them; modelling autism spectrum disorder specific gene expression dysregulation with RNA-seq data. *Sci. Rep.* **12** (1), 10158. <https://doi.org/10.1038/s41598-022-14077-1> (2022).
12. Quesnel-Vallières, M., Weatheritt, R. J., Cordes, S. P. & Blencowe, B. J. Autism spectrum disorder: insights into convergent mechanisms from transcriptomics. *Nat. Rev. Genet.* **20** (1), 51–63. <https://doi.org/10.1038/s41576-018-0066-2> (2019).
13. Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456** (7221), 470–476. <https://doi.org/10.1038/nature07509> (2008).
14. Raj, B. & Blencowe, B. J. Alternative splicing in the mammalian nervous system: recent insights into mechanisms and functional roles. *Neuron* **87** (1), 14–27. <https://doi.org/10.1016/j.neuron.2015.05.004> (2015).
15. Chau, K. K. et al. Full-length isoform transcriptome of the developing human brain provides further insights into autism. *Cell. Rep.* **36** (9). <https://doi.org/10.1016/j.celrep.2021.109631> (2021).
16. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362** (6420), eaat8127. <https://doi.org/10.1126/science.aat8127> (2018).
17. Reble, E., Dineen, A. & Barr, C. L. The contribution of alternative splicing to genetic risk for psychiatric disorders. *Genes Brain Behav.* **17** (3), e12430. <https://doi.org/10.1111/gbb.12430> (2018).
18. Li, R. et al. Misregulation of alternative splicing in a mouse model of Rett syndrome. *PLOS Genet.* **12** (6), e1006129. <https://doi.org/10.1371/journal.pgen.1006129> (2016).
19. Osenberg, S. et al. Activity-dependent aberrations in gene expression and alternative splicing in a mouse model of Rett syndrome. *Proc. Natl. Acad. Sci. U S A.* **115** (23), E5363–E5372. <https://doi.org/10.1073/pnas.1722546115> (2018).
20. Shah, S., Richter, J. D. & Do Fragile, X. Syndrome and other intellectual disorders converge at aberrant Pre-mRNA splicing?? *Front. Psychiatry.* **12**, 715346. <https://doi.org/10.3389/fpsy.2021.715346> (2021).
21. Shah, S. et al. FMRP control of ribosome translocation promotes chromatin modifications and alternative splicing of neuronal genes linked to autism. *Cell. Rep.* **30** (13), 4459–4472e6. <https://doi.org/10.1016/j.celrep.2020.02.076> (2020).
22. Parikshak, N. N. et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540** (7633), 423–427. <https://doi.org/10.1038/nature20612> (2016).
23. Xiong, H. Y. et al. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347** (6218), 1254806. <https://doi.org/10.1126/science.1254806> (2015).
24. Smith, R. M. & Sadee, W. Synaptic signaling and aberrant RNA splicing in autism spectrum disorders. *Front. Synaptic Neurosci.* **3**. <https://doi.org/10.3389/fnsyn.2011.00001> (2011).
25. Irimia, M. et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159** (7), 1511–1523. <https://doi.org/10.1016/j.cell.2014.11.035> (2014).
26. Quesnel-Vallières, M. et al. Misregulation of an Activity-Dependent splicing network as a common mechanism underlying autism spectrum disorders. *Mol. Cell.* **64** (6), 1023–1034. <https://doi.org/10.1016/j.molcel.2016.11.033> (2016).
27. Quesnel-Vallières, M., Irimia, M., Cordes, S. P. & Blencowe, B. J. Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev.* **29** (7), 746–759. <https://doi.org/10.1101/gad.256115.114> (2015).
28. Okay, K. et al. Alternative splicing and gene co-expression network-based analysis of dizygotic twins with autism-spectrum disorder and their parents. *Genomics* **113** (4), 2561–2571. <https://doi.org/10.1016/j.ygeno.2021.05.038> (2021).
29. Stamova, B. S. et al. Evidence for differential alternative splicing in blood of young boys with autism spectrum disorders. *Mol. Autism.* **4** (1), 30. <https://doi.org/10.1186/2040-2392-4-30> (2013).
30. Wang, Y. & Wang, Z. Systematical identification of splicing regulatory cis-elements and cognate trans-factors. *Methods San Diego Calif.* **65** (3), 350–358. <https://doi.org/10.1016/j.ymeth.2013.08.019> (2014).
31. Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell. Biol.* **10** (11), 741–754. <https://doi.org/10.1038/nrm2777> (2009).
32. Gonatopoulos-Pournatzis, T. et al. Genome-wide CRISPR-Cas9 interrogation of splicing networks reveals a mechanism for recognition of Autism-Misregulated neuronal microexons. *Mol. Cell.* **72** (3), 510–524e12. <https://doi.org/10.1016/j.molcel.2018.10.008> (2018).
33. Gonatopoulos-Pournatzis, T. & Blencowe, B. J. Microexons: at the nexus of nervous system development, behaviour and autism spectrum disorder. *Curr. Opin. Genet. Dev.* **65**, 22–33. <https://doi.org/10.1016/j.cde.2020.03.007> (2020).
34. Garcia-Cabau, C. et al. Mis-splicing of a neuronal microexon promotes CPEB4 aggregation in ASD. *Nature* **637**, 496–503. <https://doi.org/10.1038/s41586-024-08289-w> (2025).
35. Li, Y. I., Sanchez-Pulido, L., Haerty, W. & Ponting, C. P. RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res.* **25** (1), 1–13. <https://doi.org/10.1101/gr.181990.114> (2015).
36. Voineagu, I. et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474** (7351), 380–384. <https://doi.org/10.1038/nature10110> (2011).
37. Sebat, J. et al. Strong association of de Novo copy number mutations with autism. *Science* **316** (5823), 445–449. <https://doi.org/10.1126/science.1138659> (2007).
38. Gueroussov, S. et al. An alternative splicing event amplifies evolutionary differences between vertebrates. *Science* **349** (6250), 868–873. <https://doi.org/10.1126/science.aaa8381> (2015).
39. Jaganathan, K. et al. Predicting splicing from primary sequence with deep learning. *Cell* **176** (3), 535–548e24. <https://doi.org/10.1016/j.cell.2018.12.015> (2019).
40. Dawes, R. et al. SpliceVault predicts the precise nature of variant-associated mis-splicing. *Nat. Genet.* **55** (2), 324–332. <https://doi.org/10.1038/s41588-022-01293-8> (2023).
41. Wagner, N. et al. Aberrant splicing prediction across human tissues. *Nat. Genet.* **55** (5), 861–870. <https://doi.org/10.1038/s41588-023-01373-3> (2023).
42. Alonso-Gonzalez, A. et al. Exploring the biological role of postzygotic and germinal de Novo mutations in ASD. *Sci. Rep.* **11** (1), 319. <https://doi.org/10.1038/s41598-020-79412-w> (2021).
43. Lim, E. T. et al. Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. *Nat. Neurosci.* **20** (9), 1217–1224. <https://doi.org/10.1038/nn.4598> (2017).
44. Satterstrom, F. K. et al. Large-Scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180** (3), 568–584e23. <https://doi.org/10.1016/j.cell.2019.12.036> (2020).
45. Zhou, X. et al. Integrating de Novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. *Nat. Genet.* **54** (9), 1305–1319. <https://doi.org/10.1038/s41588-022-01148-2> (2022).

46. Wu, T. et al. ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov.* **2** (3), 100141. <https://doi.org/10.1016/j.xinn.2021.100141> (2021).
47. Lord, J. & Baralle, D. Splicing in the diagnosis of rare disease: advances and challenges. *Front. Genet.* **12**, 689892. <https://doi.org/10.3389/fgene.2021.689892> (2021).
48. Blakes, A. J. M. et al. A systematic analysis of splicing variants identifies new diagnoses in the 100,000 genomes project. *Genome Med.* **14** (1), 79. <https://doi.org/10.1186/s13073-022-01087-x> (2022).
49. Ong, C. T. & Adusumalli, S. Increased intron retention is linked to Alzheimer's disease. *Neural Regen. Res.* **15** (2), 259–260. <https://doi.org/10.4103/1673-5374.265549> (2019).
50. Zhang, R. et al. An intronic variant of CHD7 identified in autism patients interferes with neuronal differentiation and development. *Neurosci. Bull.* **37** (8), 1091–1106. <https://doi.org/10.1007/s12264-021-00685-w> (2021).
51. Rodriguez-Fontenla, C. & Carracedo, A. UTMOST, a single and cross-tissue TWAS (Transcriptome wide association Study), reveals new ASD (Autism Spectrum Disorder) associated genes. *Transl Psychiatry.* **11** (1), 1–11. <https://doi.org/10.1038/s41398-021-01378-8> (2021).
52. Niesler, B. & Rappold, G. A. Emerging evidence for gene mutations driving both brain and gut dysfunction in autism spectrum disorder. *Mol. Psychiatry.* **26** (5), 1442–1444. <https://doi.org/10.1038/s41380-020-0778-5> (2021).
53. Yi, C. X. & Tschöp, M. H. Brain–gut–adipose-tissue communication pathways at a glance. *Dis. Model. Mech.* **5** (5), 583–587. <https://doi.org/10.1242/dmm.009902> (2012).
54. Puente-Ruiz, S. C. & Jais, A. Reciprocal signaling between adipose tissue depots and the central nervous system. *Front. Cell. Dev. Biol.* **10**, 979251. <https://doi.org/10.3389/fcell.2022.979251> (2022).
55. Ge, T., Fan, J., Yang, W., Cui, R. & Li, B. Leptin in depression: a potential therapeutic target. *Cell. Death Dis.* **9** (11), 1–10. <https://doi.org/10.1038/s41419-018-1129-1> (2018).
56. Bouret, S. G. Neurodevelopmental actions of leptin. *Brain Res.* **1350**, 2–9. <https://doi.org/10.1016/j.brainres.2010.04.011> (2010).
57. Beccano-Kelly, D., Harvey, J. & Leptin A novel therapeutic target in Alzheimer's disease?? *Int. J. Alzheimer's Dis.* **2012**, e594137. <https://doi.org/10.1155/2012/594137> (2012).
58. McGregor, G. & Harvey, J. Leptin regulation of synaptic function at hippocampal TA-CA1 and SC-CA1 synapses: implications for health and disease. *Neurochem Res.* **44** (3), 650–660. <https://doi.org/10.1007/s11064-017-2362-1> (2019).
59. Naro, C., Cesari, E. & Sette, C. Splicing regulation in brain and testis: common themes for highly specialized organs. *Cell. Cycle.* **20** (5–6), 480–489. <https://doi.org/10.1080/15384101.2021.1889187> (2021).
60. Leung, C. S. et al. Dysregulation of the chromatin environment leads to differential alternative splicing as a mechanism of disease in a human model of autism spectrum disorder. *Hum. Mol. Genet.* **32** (10), 1634–1646. <https://doi.org/10.1093/hmg/ddad002> (2023).
61. Ruzzo, E. K. et al. Inherited and de Novo genetic risk for autism impacts shared networks. *Cell* **178** (4), 850–866e26. <https://doi.org/10.1016/j.cell.2019.07.015> (2019).
62. Wilfert, A. B. et al. Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nat. Genet.* **53** (8), 1125–1134. <https://doi.org/10.1038/s41588-021-00899-8> (2021).

## Acknowledgements

We would like to warmly thank the ASC (Autism Sequencing Consortium) (<https://genome.emory.edu/ASC/>) that has sequenced the Spanish trios.

## Author contributions

M Tubio-Funqueirino, M. Fernandez-Prieto, J Gonzalez-Peñas, A. Carracedo, C. Arango and M.Parellada participated in the recruitment of samples. S.Dominguez-Alonso has carried out the analyses and wrote the paper. A.Carracedo., C.Rodriguez-Fontenla, M.Parellada and C.Arango participated in the design and coordination of this study. A.Carracedo. and C.Rodriguez-Fontenla critically revised the work and approved the final content.

## Funding

This study has been funded by Instituto de Salud Carlos III (ISCIII) through the project “PI22/00208” and co-funded by the European Union.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-95456-2>.

**Correspondence** and requests for materials should be addressed to C.R.-F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025