Aberrant splicing prediction during human organ development

Nils Wagner[1,2], Aleksandr M. Neverov[1], Alexandra C Martin-Geary[3,4], Shubhankar Londhe[1,2,7], Carina Schröder[1], Vicente A. Yépez[1], Nicola Whiffin[3,4,5], Julien Gagneur[1,2,6,7]

[1] Department of Informatics, Technical University of Munich, Garching, Germany
[2] Helmholtz Association - Munich School for Data Science (MUDS), Munich, Germany
[3] Big Data Institute, University of Oxford, Oxford, UK
[4] Centre for Human Genetics, University of Oxford, Oxford, UK
[5] Broad Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[6] Institute of Human Genetics, School of Medicine, Technical University of Munich, Munich, Germany
[7] Computational Health Center, Helmholtz Center Munich, Neuherberg, Germany

Developmental disorders constitute a major class of genetic diseases, yet tools to identify splicing-disruptive variants during development are lacking. To address this need we extended the AbSplice framework to incorporate splicing dynamics across developmental stages. Moreover, we introduced several improvements including a refined ground truth from the aberrant splicing caller FRASER2, a continuous representation of splice site usage, and integration of the rich set of predictions from the sequence-based model Pangolin. These advances double the precision and recall of the original model at predicting aberrant splicing events and enable predictions across embryonic, childhood, and adult tissues. Genome-wide scores for all single-nucleotide variants and a web interface to score indels are available to facilitate the exploration of the predictions. Our genome-wide predictions reveal a class of variants with splicing-disruptive effects confined to early development, particularly abundant in the brain (>18,000 variants). Analysis of genomes of individuals with a suspected Mendelian disorder from Genomics England's National Genomics Research Library identified 26 unique variants in disease-linked genes, with stronger predicted effects during development than in adulthood, including a candidate new diagnosis in the gene *FGFR1*. Altogether, these results improve the accuracy of splice-disruptive variant prediction and provide developmental context to aid interpretation.

Introduction

Aberrant splicing can disrupt gene function by altering RNA isoforms, leading to outcomes such as frameshifts, transcript degradation, or the loss of critical protein domains. If a variant strongly affects splicing isoform choice, the remaining abundance of functional RNA isoforms may be insufficient to maintain normal cellular function of the gene. Predicting such splicing aberrations

is therefore critical for variant interpretation, particularly in the context of rare disease diagnostics and oncology[1–8].

Aberrant splicing events are rare large alterations of splice isoform usage and can be detected from RNA-seq data from human tissues[9–13]. Recently, we introduced AbSplice[14], a model to predict the probability that a rare variant leads to aberrant splicing in a specific human tissue. AbSplice was trained on rare variants associated with aberrant splicing events for 49 tissues from the GTEx dataset[15]. The model combines tissue-agnostic sequence-based variant effect predictions from state-of-the art deep learning models with tissue-specific SpliceMaps derived from RNA-seq data quantifying splice site usage. Notably, SpliceMaps can be generated from healthy and cancerous tissues[16] and can be obtained from a handful of RNA-seq samples[14], highlighting the framework's adaptability. Aberrant splicing predictions from AbSplice are valuable not only in isolation but also as input for downstream tasks such as aberrant gene expression prediction[17] and rare variant association testing[18].

The initial implementation of AbSplice used aberrant splicing events detected by FRASER[11] as its ground truth for model training and evaluation. However, the more recently developed FRASER2[12], based on a more robust intron-centric metric, offers a more accurate ground truth by reducing false-positive calls, enabling a more reliable assessment of model performance.

Advances in deep learning models for splicing prediction have further refined our ability to predict and interpret splicing alterations caused by variants (reviewed in [19–21]). For instance, Pangolin[22] shares the same context window and model architecture as SpliceAI[1] yet demonstrated superior performance due to enhancements in both the quantity and quality of its training data. Specifically, Pangolin incorporates splicing quantifications derived from RNA-seq data across four tissues (heart, liver, brain, testis) and four mammalian species (human, rhesus macaque, mouse, rat). In contrast, SpliceAI was trained on binary labels of splice sites derived from the human standard genome annotation GENCODE[23] as well as novel splice junctions commonly observed in the GTEx cohort. The initial implementation of AbSplice did not include Pangolin as the two tools were developed in parallel.

During organ development, splicing events can undergo dynamic changes in inclusion levels, with brain and testis showing up to 20% of cassette exons being developmentally regulated[24]. Developmental alternative splicing events are highly conserved across evolution, underscoring their functional significance[24]. Variants impacting splicing at developmentally regulated sites may cause developmental disorders. However, their effects on splicing may be limited to specific stages, such as embryogenesis, and not manifest in adults. Since current splicing annotations, including the SpliceMaps[14] used in the AbSplice model, are derived from adult RNA-seq data, the impact of variants on developmentally regulated splicing events is potentially underestimated. Thus, integrating developmental splicing dynamics into splicing prediction models should be considered to accurately assess variant effects across the full spectrum of an organism's life stages.

Here, we present an enhanced version of AbSplice (referred to as AbSplice2) that addresses these issues. We leverage the improved ground truth from FRASER2, integrate state-of-the-art splicing predictions from Pangolin and extend our analyses to include developmental splicing dynamics (Fig. 1). Together, this led to improved detection of aberrant splicing causing variants while providing developmental and tissue-specific context.

Results

An improved ground truth of aberrantly spliced events leads to a 2-fold improvement in recall

AbSplice was originally trained and evaluated using the aberrant splicing caller FRASER applied to the Genotype-Tissue Expression (GTEx) dataset[11,14]. FRASER2, a refined version of FRASER, was shown to reduce false positives in outlier detection by about one order of magnitude at minimal loss of sensitivity[12]. The AbSplice pipeline[14] had already incorporated a rare-variant-based post-filter to suppress likely false positive outliers from the original FRASER. Nonetheless, using FRASER2 outliers as ground truth in our benchmark led to a 2-fold improvement in recall for AbSplice (Fig. 2a). This improvement was independent of retraining the AbSplice model with the FRASER2 outlier labels (SI Fig. 1). While a reduction of false outlier labels directly improved the recall of the model it had a minor effect on the learned feature contributions which were nearly identical for an AbSplice model trained on FRASER compared to FRASER2 outlier labels (SI Fig. 1).

Replacing SpliceAI with Pangolin leads to improved precision

AbSplice1 was based on variant effect predictions of the tissue-agnostic deep learning models MMSplice[2] and SpliceAI[1] which were combined with tissue-specific splicing annotations (SpliceMaps)[14]. Pangolin is a deep learning model that uses a similar model architecture as SpliceAI. While SpliceAI was purely trained on genomic annotations, Pangolin additionally used quantitative usage levels of splicing in 4 different tissues and was trained across 4 mammalian species. Pangolin's additional training data led to a better predictive performance than SpliceAI[22,25,26]. Both SpliceAI and Pangolin provide scores for splice site usage increase and decrease. We had originally used the absolute maximum variant effect score of SpliceAI. However, the direction of change in splicing could also be informative (SI Fig. 2) as it enhances the model's ability to learn interactions among other signed feature scores, such as the predicted variant effect of MMSplice, providing a more reliable basis for predictions when models agree on the direction of change. Indeed, replacing the maximum Delta Score of SpliceAI with the four predicted scores for acceptor gain, acceptor loss, donor gain, and donor loss improved the average precision by about 12% ($P = 3.3\times10^{-6}$, SI Fig. 2). Furthermore, replacing SpliceAI with Pangolin and using its two predicted scores of splice site usage gain and loss led to an additional improvement in average precision by about 9% ($P = 1.3\times10^{-8}$, Fig.2a, SI Fig. 2).

Continuous splice site usage levels lead to improvements in precision compared to binarized splice site usage

To discriminate between genuine splicing events and splicing noise, we had previously employed a binary threshold on the median number of split reads supporting a splice site across samples capturing splice site usage. However, thresholding can be highly sensitive to small variations in the data near the threshold and lead to drastically distinct predictions and misleading interpretations. By learning a continuous sigmoid-shaped effect of splice site usage instead of a binary threshold, we achieved an improved average precision by about 41% ($P = 1.8\times10^{-15}$, Fig. 2a,b, SI Fig. 3a,b). This adjustment reduced abrupt prediction shifts due to minor fluctuations in splice site usage, enhancing model reliability. The contribution scores on the plateaus were similar to what the binarized model had learned (SI Fig. 3b), giving evidence for the robustness of the entire framework. Comparing model predictions from the binarized and continuous models showed that the improvement in precision was mostly due to a reranking of false positive variant predictions which were in the proximity of splice sites with intermediate usage values (SI Fig. 3c). While the two models gave similar predicted values to variants that disrupt highly expressed splice sites, the continuous model predicted a lower probability for variants disrupting splice sites with low or intermediate usage levels. Stratifying the benchmark to different levels of gene expression showed that the model trained on the continuous usage levels outperformed all other models (including a model trained on gene TPM values instead of splice site usage levels) in all analyzed gene expression regimes (SI Fig. 4).

Our final model led to a 2-fold improvement in precision compared to AbSplice1. In the following, we will refer to this model as AbSplice2.

AbSplice2 achieves 2-fold improvement in precision over AbSplice1

As we have done for AbSplice1, we created AbSplice2-RNA, a model that can further incorporate aberrant splicing events detected from RNA-seq of clinically accessible tissues. This reflects the clinical rare disease diagnostics use case in which RNA-seq could be collected from accessible tissues such as blood or skin fibroblasts, but not from the suspected disease tissue. AbSplice2-RNA further improved the average precision by about 21% (Fig. 2a,b).

We provide genome-wide predictions for all possible single-nucleotide variants (SNVs) in protein-coding genes for Pangolin and AbSplice2. Additionally, we offer a web interface (https://absplice.cmm.cit.tum.de)[27] that enables scoring any variant of interest, including insertions and deletions.

Improvements of AbSplice2 are confirmed on independent benchmarks

Having established our model in GTEx, we assessed it in another independent cohort consisting of 230 individuals from the BrainGVEx study[28] with paired whole genome sequence and RNA-seq data from the brain frontal cortex. We applied AbSplice2 using the SpliceMap from GTEx brain frontal cortex. Aberrant splicing calls from FRASER2 were used to evaluate the predictions. While the model reached similar precision levels as in the corresponding GTEx tissue, the maximum recall of the models in BrainGVEx was reduced compared to GTEx, perhaps due to different sequencing quality (Fig.2c). Nevertheless, the relative performance of all models was replicated, with AbSplice2 performing best.

In addition to aberrant splicing prediction, we evaluated our model predictions on the task of predicting aberrant gene underexpression in the GTEx dataset. Aberrant splicing often leads to the expression of isoforms containing premature termination codons which can be recognized by the nonsense-mediated decay machinery for degradation. As a result, aberrant splicing predictors are informative about aberrantly underexpressed genes. Therefore, we further evaluated the models at predicting aberrant underexpression, using a benchmark of expression outliers across GTEx that we established earlier[17]. We found that AbSplice2 led to a 2-fold improvement in average precision over AbSplice1. Consequently, replacing AbSplice1 with AbSplice2 in the input features of the aberrant gene underexpression predictor AbExp[17] led to an improved average precision by about 7% ($P = 9.3\times10^{-4}$, Fig. 2d).

Dynamically changing splice sites lead to changing predicted variant effects during development

Having established an improved AbSplice model, we next considered extending its application to developmental stages. To this end, we derived SpliceMaps for human RNA-seq data from 7 tissues (brain, cerebellum, heart, kidney, liver, ovary and testis) gathered from 15 different developmental stages spanning embryogenesis to the adult with the earliest recorded stage being 4 weeks post conception[29]. With 1 to 5 samples per tissue-time point pair, the sample size of the developmental dataset was strongly reduced compared to the average 331 samples per tissue in GTEx. Downsampling investigation of the GTEx dataset indicated that SpliceMaps largely agreeing with the original SpliceMaps could still be derived with as few as 4 samples (SI Fig. 5a) and would not affect AbSplice performance (SI Fig. 5b). To reach a similar minimal sample size while preserving temporal resolution, we pooled data of each developmental time point and its adjacent time points. Moreover, we confirmed the agreement of the tissue-specific SpliceMaps between the adult stage of the developmental dataset and GTEx by comparing their SpliceMaps (SI Fig. 6c,d, Jaccard similarity typically above 0.75).

We computed AbSplice2 scores for all tissues and developmental stages for all possible SNVs within a 100-bp window of all 47,769 previously annotated developmental alternative splicing (devAS) sites[24]. One such devAS site is visualized in Fig. 3a, where exon 6 of the gene *SSBP3*

which is known to play a role in embryonic stem cell differentiation and head development[30,31], is included in an early developmental stage (4 weeks post conception) at a high inclusion level in the brain, but is removed from the spliced transcript at the point of birth and is subsequently missing in the adult brain. In the vicinity of the splice sites of *SSBP3* exon 6 we predicted a significant difference in variant effects between early and late stages of development (Fig. 3a). While there are variants with a high predicted probability to disrupt splicing in early stages, AbSplice2 predicted exclusively low variant impacts in adult stages. One example variant within this region is displayed in Fig. 3b. While the deep learning models Pangolin and MMSplice provided a variant score that was independent of the developmental stage, the splicing metrics from the SpliceMaps captured how the usage of each splice junction changed during development. The drop in inclusion of *SSBP3* exon 6 was captured in the developmental SpliceMaps which was reflected in AbSplice2 predictions being strong in the early stages only. Moreover, the continuous representation of the median splice site usage as opposed to the binary representation used in AbSplice1 led to smoother and more robust predictions over development time (SI Fig. 6 for an example).

We predicted for each tissue the number of variants genome-wide that led to a splice-disrupting effect in an early stage of development (from conception to toddler age) and no effect in the adult (Fig. 3c, SI Fig. 7). The brain, which is the tissue with most devAS events[24] showed the highest number of such variants (N=18,206 which is 10% of all high impact variants in the brain across all developmental stages around devAS sites, SI Fig. 7). This example showcases that while SpliceAI and Pangolin can capture variants affecting splicing, AbSplice2 adds the developmental and tissue context of these effects. These variants may be of clinical importance as developmentally changing splicing events were shown to be highly conserved across evolution[24]. Fitting this hypothesis, these development-specific effect variants are enriched in loss-of-function (LoF) intolerant genes, with genes in the 10% most LoF-intolerant group being 6.8-fold more likely to harbor a development-specific effect variant compared to the 10% most LoF-tolerant genes (Fisher's exact test $P = 6.7\times10^{-65}$) (Fig. 3d).

Application to rare disease samples

To assess the clinical potential of predicting developmental aberrant splicing, we applied AbSplice2 to 64,736 rare disease participants from the UK National Genomics Research Library[32] (NGRL). We focussed on variants located within 100 bp of devAS splice sites and for which AbSplice2 predicted an effect during early development (AbSplice2 score above the cutoff 0.05) that is stronger than in adult. We further restricted to variants in 'Green' genes confidently linked with genetic disorders from PanelApp[33], identifying a total of 26 unique variants in 47 individuals (Table 1, SI Table 1). One example is a splice acceptor variant in *ARFGEF1* (chr8:67203252:C>T, NM_006421.5:c.4960-1G>A) that had already been identified as the causal variant in an individual recruited with intellectual disability. This gene is associated with developmental delay, epilepsy, and intellectual disability. AbSplice2 predicted this variant to activate a cryptic splice site within the adjacent exon causing a frameshift with an effect that is strongest in the brain between 12 and 16 weeks post-conception (SI Fig. 8a). ADP ribosylation factor guanine nucleotide exchange factor (*ARFGEF*) genes are involved in cellular trafficking

and neuronal development in the hippocampus, and play an important role in neurite development and axon elongation[34] with evidence suggesting that neurite development spans from early development into adolescence, with the possibility of different tissues having greater or lesser neurite densities at different timepoints[35].

Three rare (gnomAD allele frequency < 10^{-4}) variants in three individuals were prioritised as potential diagnostic candidates after reviewing phenotypic fit and mode of inheritance (Table 1). The first variant (chr8:38428099:C>T, NM_023110.3:c.449-6G>A) lies in *FGFR1* in an individual with hypertrophic hypogonadism. This variant introduces an AG dinucleotide in the AG exclusion zone upstream of the canonical acceptor site[36], causing a frameshift and introduction of a premature termination codon (Fig. 4a). This variant is classified in ClinVar as having 'conflicting interpretations of pathogenicity' with both likely pathogenic, and uncertain significance classifications. *FGFR1* plays a crucial role in the development of the neuroendocrine hypothalamus and is central to the genesis of gonadotropin-releasing hormone (GnRH) secreting neurons in the olfactory bulb, and their migration into the hypothalamus during development[37]. Variants that result in a loss of function in the *FGFR1* gene have been found to cause a failure of neuronal migration and reduced GnRH secretion[37] leading to hypogonadotropic hypogonadism and anosmia (Kallmann syndrome)[38]. AbSplice2 predicted embryogenesis-specific activation of this acceptor site in early forebrain development (Fig. 4b) and in the kidney, but not after birth. The effect in the forebrain is consistent with the disease mechanism.

The second candidate variant was identified in *FN1*, a gene associated with skeletal dysplasia. Whilst we identified an elevation in AbSplice2 scores at early developmental stages across tissues, the specific tissues with the greatest relevance to the disorder (skeletal dysplasia) were unavailable within our data (SI Fig. 8b). For *NDUFS2*, a mitochondrial complex 1 member associated with Leigh syndrome, the clinical relevance in the NGRL participant remains uncertain, as deleterious variants in this gene are typically associated with early-onset fatal outcomes in childhood, while this participant is an adult. It is possible that this variant may be a hypomorphic non-coding variant, and therefore result in a less severe phenotype, however we are unable to confirm this currently. All remaining variants did not represent good candidates for the individual's disorder when mode of inheritance or phenotypic matching were considered, including 8 instances where the participants' case had been solved, or partially solved with the causal variant being found in a different gene (Table S1).

Discussion

Here, we have developed a model for tissue-specific aberrant splicing prediction across developmental stages. Thereby, our updated AbSplice predicts a class of variants whose effect on splicing changes over the course of development. Additionally, we improved on the previous AbSplice model by replacing SpliceAI with Pangolin, leveraging the directionality of predicted splicing changes as well as the predicted affected position, which led to improved precision. Second, we transitioned from a binary quantification of splice site usage to continuous usage

levels which led to a two-fold improvement in the precision of AbSplice as well as providing more robust predictions. We demonstrated that these enhancements replicated in independent benchmarks. We provide the code and developmental SpliceMaps spanning stages from embryogenesis to adulthood. For convenience, we also provide precomputed scores for AbSplice2 and Pangolin for all possible SNVs.

In our previous publication[14] we used FRASER aberrant splicing calls as ground truth to train and evaluate AbSplice, as we found FRASER to be more robust and precise than other outlier callers[9,10]. In this study we employed an improved ground truth based on FRASER2 and achieved a two-fold increase in recall without altering the model's feature importance scores upon retraining. This suggests that the original model effectively captured the key signal from true positive outliers, even with a noisier ground truth. Importantly, these findings validate the reliability of the previously published AbSplice model.

The developmental predictions were based on SpliceMaps constructed from limited sample sizes and only for seven tissues. While we showed that the available sample sizes led to reasonably robust estimates, there are potential benefits of larger cohorts for more robust and comprehensive splicing quantifications. We anticipate that efforts such as the upcoming developmental data from GTEx will enable the construction of more reliable SpliceMaps and facilitate the detection of weakly spliced junctions, which are often present in only a few individuals, but can lead to drastic aberrant splicing events when activated by a variant[39,40]. Moreover, an update would only require computing SpliceMaps from such data owing to the modular structure of AbSplice. Our evaluations in this study and our previous study[14] showed that the AbSplice framework is very robust to changes in the SpliceMaps while keeping the model unchanged.

We applied AbSplice to rare disease participants from the UK National Genomics Research Library (NGRL) and prioritized variants predicted to have a developmental splicing effect, resulting in 26 candidate variants in known disease-associated genes. While AbSplice predicted developmental splicing disruptions, not all identified variants were in genes clearly related to the individual’s specific phenotypes. Many variants were heterozygous in genes that require a biallelic impact to cause disease. While these were not considered to be causal for the individual’s phenotypes, this does not imply that the underlying splicing predictions were incorrect. It remains possible that the predicted splicing alterations are real, but that disrupted splicing is not the primary mechanism underlying the observed phenotype — or that it contributes without being sufficient on its own. For some variants we observed evidence of weak splicing at the affected site exclusively in the developmental dataset, but not in adult GTEx tissues, demonstrating the importance of considering temporally dynamic regulatory patterns in variant interpretation. These findings underscore the added value of incorporating developmental context into variant interpretation.

AbSplice helps narrow down splicing-relevant variants to specific tissues and developmental stages for which SpliceMaps are available. Consequently, using the maximum score across all tissues is not the intended usage and leads to reduced precision in tissue-specific tasks (SI Fig. 9). Moreover, since not all tissues or developmental stages are covered, variants might still have

effects in unobserved contexts. Therefore, we see AbSplice as a companion tool to link predicted variant effects to the observed phenotypes by analyzing the developmental and tissue-specific predictions[41–43].

One limitation of our approach is that it depends on the accuracy of the underlying deep learning model. If the model fails to predict a splicing event, even in regions where a tissue-specific splice site is annotated, our framework cannot recover the prediction, resulting in low model scores. This limitation is particularly problematic for tissue-specific splice sites, which may be underrepresented in models trained on standard genome annotations. Such models, like SpliceAI, are optimized for canonical splice sites used across multiple tissues but may inadequately capture truly tissue-specific events. While Pangolin was trained on RNA-seq data from four tissues, its training data was limited by small cohort sizes and did not leverage the full diversity of tissue types or developmental stages. This restricts its ability to accurately predict splicing events in tissues with high variability, such as the brain, or in developmental stages where splicing patterns differ significantly from those in adults. To this end, the combination of new modeling techniques, for instance building on self-supervised[44–48] or supervised[49–51] foundation models, along with expanded training datasets covering more tissues, cell types, and developmental stages could address these limitations and improve model performance.

Figures

Main figures

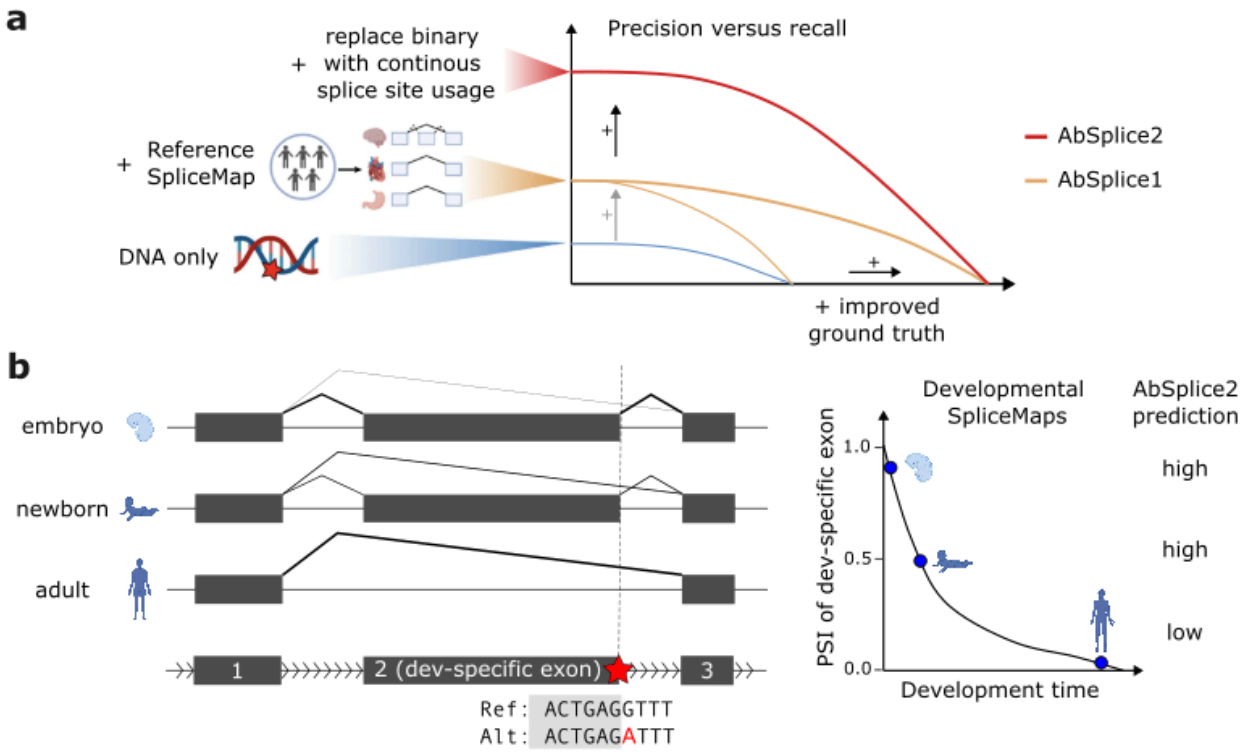


Fig.1: Improvements in the AbSplice framework and extension to variant effect prediction around developmentally regulated splice sites. a) We evaluated the performance of splicing prediction tools on GTEx splicing outliers[11]. AbSplice1 already improved sequence-only predictors by leveraging tissue-specific SpliceMaps. An improved ground truth based on the aberrant splicing caller FRASER2[12] led to a two-fold increase in recall. Replacing binary estimates of splice site usage with a continuous metric led to a further two-fold precision increase. **b)** Variants (red star) in the vicinity of developmentally regulated alternative splice sites exhibit changing effects over the course of organ development. These dynamic changes are captured by developmental SpliceMaps and reflected in AbSplice2 predictions.

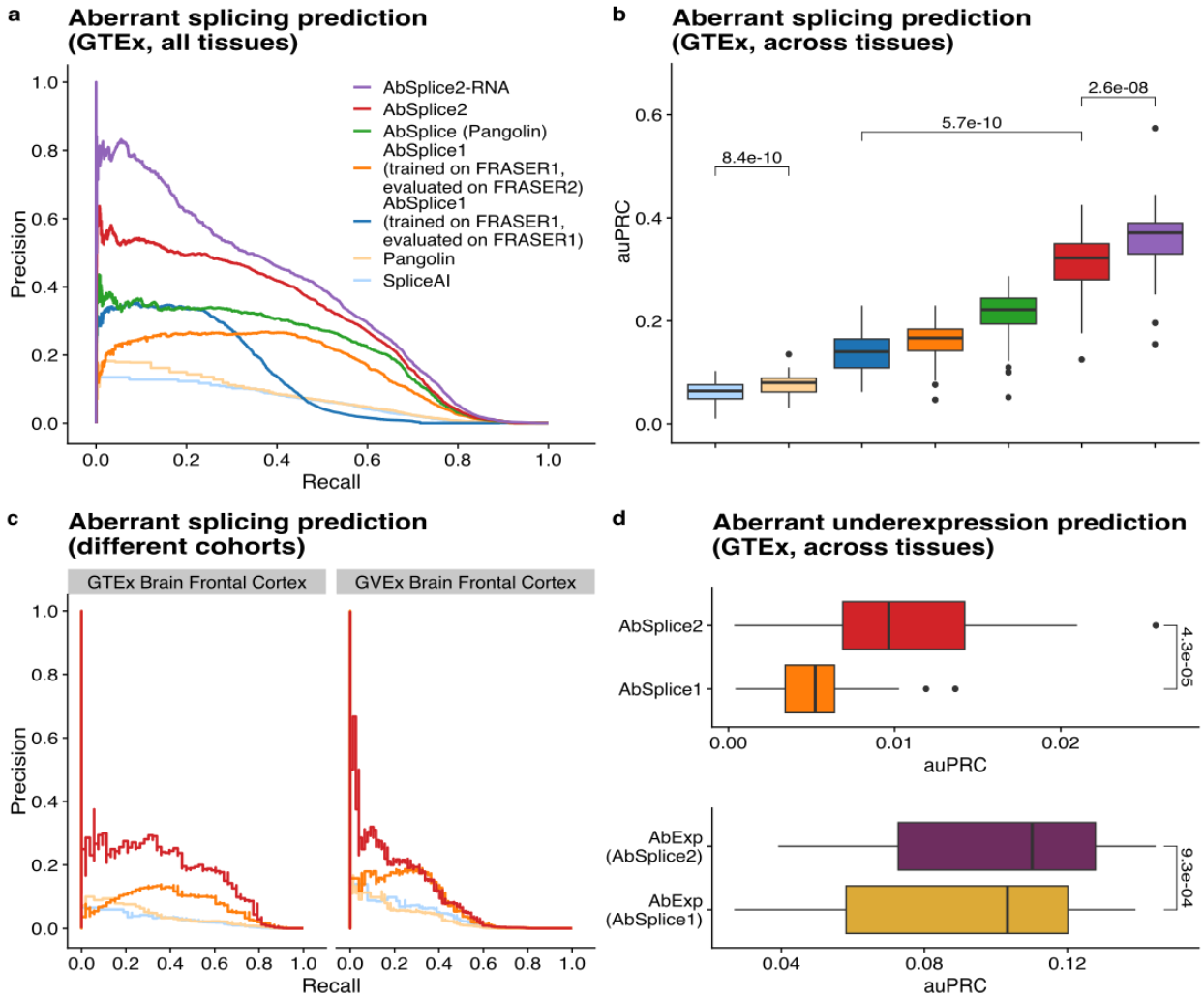


Fig. 2: Improvements for prediction of aberrant splicing and underexpression. a) Precision–recall curve comparing the overall prediction performance on all GTEx tissues of the models SpliceAI, Pangolin, AbSplice1, AbSplice (Pangolin) which replaces the single maximum SpliceAI Delta score for the two Pangolin gain and loss scores, and AbSplice2 which further replaces a binarized with a continuous splice site usage. AbSplice1 was evaluated on splicing outliers detected by FRASER1 and FRASER2. All other models are evaluated on the FRASER2 outlier calls. The color legend in (**a**) is shared across (**a**-**c**). **b)** Distribution of the area under the precision–recall curve (auPRC) of the models in (**a**) across tissues (n = 49). Center line, median; box limits, first and third quartiles; whiskers span all data within 1.5 interquartile ranges of the lower and upper quartiles. P values were computed using the paired one-sided Wilcoxon test. **c)** Comparison of precision–recall curves for the models in (**a**) using SpliceMaps from the tissue 'Brain Frontal Cortex BA9' from the GTEx dataset to predict aberrant splicing in GTEx and the independent dataset BrainGVEx. **d)** Distribution of the auPRC for AbSplice1 and AbSplice2 as well as the aberrant gene underexpression prediction model AbExp (using AbSplice1 or AbSplice2 as input features) on the task of predicting aberrant gene underexpression in the GTEx dataset.

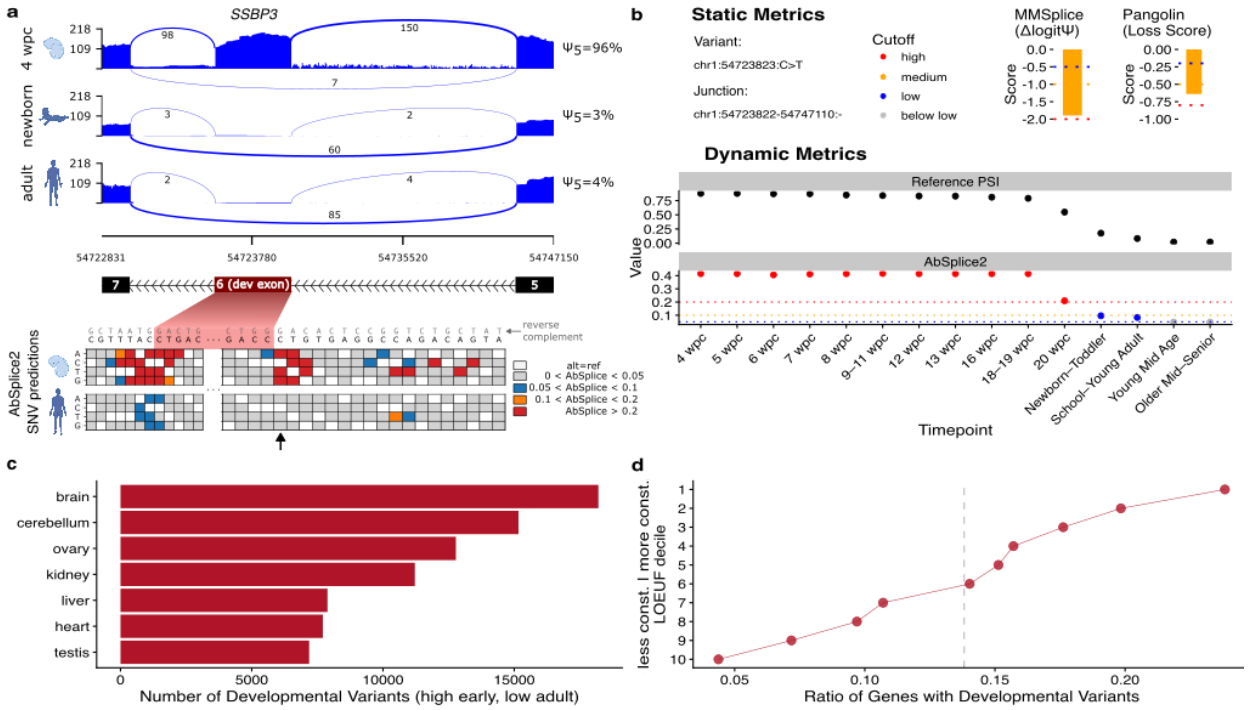


Fig. 3: Variants around developmentally dynamic splice sites exhibit developmentally dynamic effects. a) Exon 6 in the gene *SSBP3* is expressed in the brain in early stages of development during pregnancy, but not expressed at birth and afterwards. Possible SNVs in the proximity of the donor and acceptor sites of exon 6 lead to high predicted variant effects for early stages of development and low or absent effects for adult stages. 4 wpc: 4 weeks post conception. **b)** Predicted effects of the variant chr1:54723823:C>T (NM_145716.4:c.367-1G>A). The genomic location is indicated in (**a**) with a black arrow (below the heatmap). The deep learning models MMSplice and Pangolin predict a variant effect independent of the developmental timepoint. AbSplice2 predictions change dynamically over time. The drastic drop in the score from early to late developmental stages is caused by the change in the inclusion level of the exon (captured by the SpliceMap metric 'Reference PSI'). Timepoints are labelled according to the developmental study by Mazin et al.[24]. Scores are colored based on applied cutoffs: high (AbSplice2: 0.2, Pangolin: 0.8, MMSplice: 2), medium (AbSplice2: 0.1, Pangolin: 0.5, MMSplice: 1), low (AbSplice2: 0.05, Pangolin: 0.2, MMSplice: 0.5). **c)** Number of genome-wide possible SNVs with a predicted effect above the high AbSplice2 cutoff (0.2) at any developmental stage pre-birth up to toddler and below the low AbSplice2 cutoff (0.05) for adult stages (based on GTEx). **d)** Ratio of genes containing a developmental variant (variants with AbSplice2 above the high cutoff (0.2) at any developmental stage pre-birth up to toddler and AbSplice2 below the low cutoff (0.05) for adult stages) as a function of LOEUF decile. Genes in lower LOEUF deciles are more constrained. The dotted grey line indicates the genome-wide baseline ratio of genes with developmental variants.

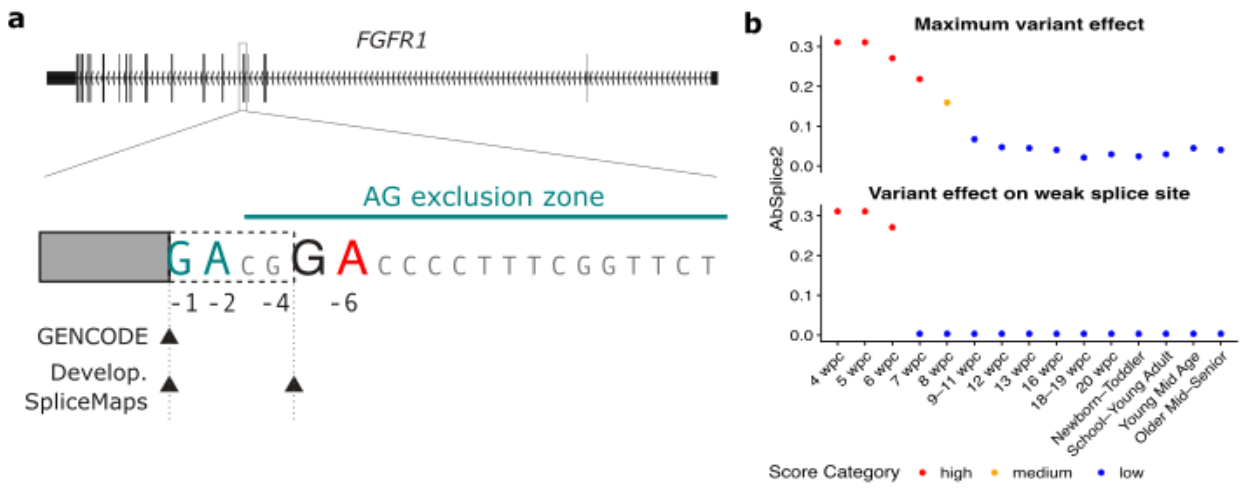


Fig 4. AbSplice2 predicts the activation of a weak splice site in early development providing a potential diagnostic candidate variant for an undiagnosed individual. a) Acceptor-site creating variant (chr8:38428099:C>T, NM_023110.3:c.449-6G>A) in the AG exclusion zone of exon 6 of the *FGFR1* gene that was identified in an individual within the NGRL recruited with hypertrophic hypogonadism. Variants that result in a loss of function in *FGFR1* have been found to lead to the disease phenotype (Kallmann syndrome). The variant affects a weak splice site identified during embryogenesis (Develop. SpliceMaps, lower track) but not in the GENCODE annotation, which would result in a frameshift. **b**) Top. Maximum AbSplice2 effect predictions of the variant regarding loss of the canonical site and the gain of the weak splice site. Bottom. AbSplice2 effect predictions of the variant on the weak splice site. The effect is predicted to be strong only in the weeks 4-6 post conception.

Tables

Gene	Gene Id	Variant Id	Condition	Comments	HPO Matches
ARFGEF1	ENSG00000066777	chr8:67203252C>T (NM_006421.5:c.4960-1G>A)	Intellectual disability	This variant is in a family whose case is marked within the NGRL as “solved” with this as the associated variant. This variant does not appear in gnomAD or ClinVar. *ARFGEF1* is associated with Intellectual disability and is a good match to the participant’s phenotype.	Intellectual disability Autistic behaviour Delayed speech and language Global developmental delay Small for gestational age Delayed gross motor development Delayed fine motor development Delayed fine motor development
FGFR1	ENSG00000077782	chr8:38428099C>T (NM_023110.3:c.449-6G>A)	Idiopathic hypogonadotropic hypogonadism	ClinVar Likely pathogenic (single submission) linked to condition “Hypogonadotropic hypogonadism 2 with or without anosmia; Pfeiffer syndrome" – Strong match	Idiopathic Hypogonadotropic hypogonadism Abnormal spermatogenesis Decreased serum testosterone level.
FN1	ENSG00000115414	chr2:215372382G>C (NM_212482.4:c.6248-7C>G)	Paediatric disorders	Strong potential phenotypic match *FN1* – Unsolved. Suggested	Optic atrophy Abnormality of the hip bone

				“Spondylometaphyseal dysplasia, 'corner fracture' type” – mechanism undetermined	
NDUFS2	ENSG00000158864	chr1:161213926G>A (NM_001377299.1:c.1354+5G>A)	Epilepsy plus other features	*NDUFS2* Biallelic variants. The gene partially fits the participant’s phenotype. This participant also has another candidate heterozygous missense variant (chr1:161213401C>G). The splice variant has conflicting interpretations, but also some indication of a reduced product. The mechanism is GoF so both variants look like good candidates, pending confirmation of compound heterozygosity. ClinVar entry (VUS) from Illumina flagged this variant in a population with Mitochondrial complex I deficiency, nuclear type 1. No direct evidence was found, but they were also not able to rule it out.May be a partial explanation for participant’s phenotype.	Abnormality of forebrain morphology, Other retinal disorders, Ischaemic heart diseases Abnormal corpus callosum morphology, Abnormal cortical gyration, abnormality of the cerebral cortex, Seizures

Table 1: Prioritized variants by applying AbSplice2 on the NGRL rare disorder cohorts. HPO matches: overlapping Human Phenotype Ontology terms between the participant and the gene associated with the AbSplice2 candidate variant. Variant coordinates are in hg38.

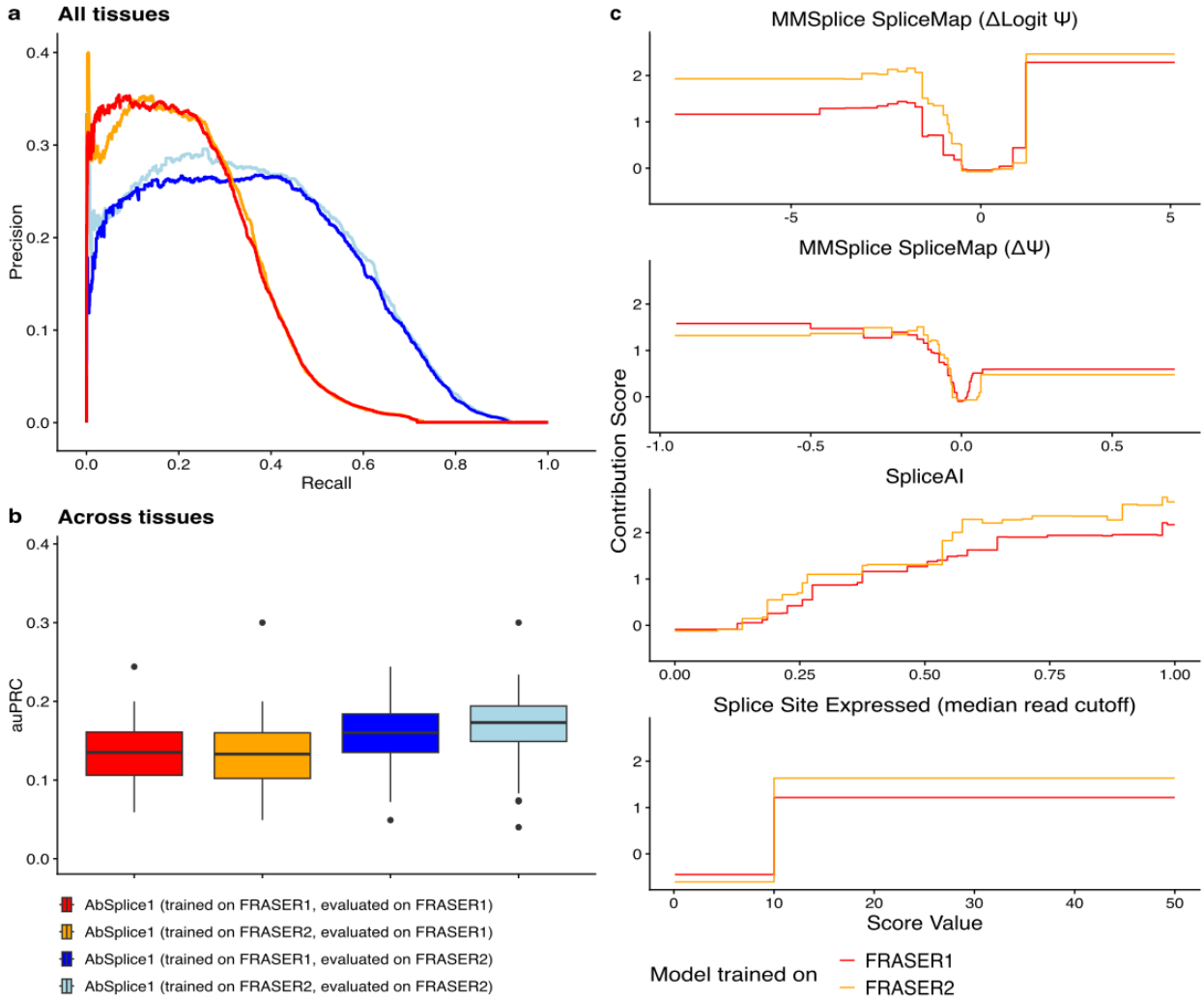
Variant	Participant Count	Gene	Mode of Inheritance	gnomAD AC/AN/AF	Participant recruited disorder(s)	Gene associated disorder	Notes	AbSplice2 observations
chr1:181793663A>C (NM_001205293.3:c.5899-2A>C, hg19: chr1:181762799A>C)	1	*CACNA1E*	Monoallelic	2/1608198/1.24E-06	Paediatric disorders	*CACNA1E*-related epileptic encephalopathy with contractures, macrocephaly, and dyskinesia (G2P)	Heterozygous. Case Solved CHD7	Strong dev effects in brain and cerebellum at junction chr1:181759692-181762800:+
chr2:99554536T>C (NM_001386135.1:c.3336-2A>G, hg19: chr2:100170998T>C)	1	*AFF3*	Monoallelic	22/1611672/1.37E-05	Coarse facial features including Coffin-Siris-like disorders	*AFF3*-related KINSSHIP syndrome, *AFF3*-related intellectual disability (G2P)	Heterozygous. Case solved AUTS2	Strong dev effects in brain and kidney at junction chr2:100170996:100171144:-
chr5:179623635C>G (NM_001257293.2:c.-72G>C, hg19: chr5:179050636C>G)	1	*HNRNPH1*	Monoallelic	Not in gnomAD	Intellectual disability	*HNRNPH1*-related neurodevelopmental disorder (G2P)	Heterozygous. MYT1L VUS is better match	Strong dev effects in brain and testis at junction chr5:179044933-179050646:-
chr11:57806454C>G (NM_001085458.2:c.1877-7C>G, hg19: chr11:57573926C>G)	1	*CTNND1*	Monoallelic	7/1601964/4.37E-06	Intellectual disability	*CTNND1*-related blepharo-cheiro-dontic syndrome (G2P)	Heterozygous. Unclear phenotype match	Strong dev effects in brain, cerebellum, heart, kidney, ovary and testis at multiple junctions. Affected junctions: chr11:57573507:57573932:+ (brain, cerebellum, heart, kidney, ovary, testis), chr11:57572252:57573932:+ (brain), chr11:57559145:57573932:+ (brain), chr11:57569668-57573932:+ (cerebellum), chr11:57509426-57573932:+ (cerebellum)

chr12:57582586A>G (NM_004984.4:c.2993-16A>G, hg19: chr12:57976369A>G)	1	*KIF5A*	Monoallelic	5/1608500/3.11E-06	Noonan syndrome	*KIF5A*-related severe neonatal myoclonus (G2P)	Heterozygous. Case partially solved USP7 variant is better match	Strong dev effects in brain and cerebellum at multiple junctions
chr15:57252494T>A (NM_207037.2:c.1260+2T>A, hg19: chr15:57544692T>A)	1	*TCF12*	Monoallelic	7/1612868/4.34E-06	Paediatric disorders	*TCF12*-related neurodevelopmental disorder with coronal craniosynostosis (G2P)	Heterozygous. Case Solved HNRNPU	Strong dev effects in all tissues at junction chr15:57544690:57545459:+
chr17:31253005G>A (NM_001042492.3:c.4173+5G>A, hg19: chr17:29580023G>A)	1	*NF1*	Monoallelic	1/1610582/6.21E-07	Paediatric disorders	*NF1*-related neurofibromatosis	Heterozygous, Actionable variant identified in CACNA1F is better match	Strong dev effects in brain at junction chr17:29580018-29585361:+ and chr17:29580018:29685497:+
chr22:23766253T>G (NM_213720.3:c.284A>C, hg19: chr22:24108440T>G)	2	*CHCHD10*	Monoallelic	8894/988434/9.00E-03	IUGR and IGF abnormalities, Early onset dystonia	Frontotemporal dementia and/or amyotrophic lateral sclerosis 2, Mitochondrial disease (ClinGen)	Heterozygous, ClinVar Likely Benign	Strong dev effects in heart at junction chr22:24108429-24109560:-
chr15:89303919A>G (NM_001113378.2:c.3058+4A>G, hg19: chr15:89847150A>G)	1	*FANCI*	Biallelic	39/1613850/2.42E-05	Intellectual disability	*FANCI*-related Fanconi anemia (G2P)	Pathogenic variant found in LZTR1 is better match	Strong dev effects in multiple junctions in brain and cerebellum
chr16:75542598T>C (NM_001077418.3:c.664+4A>G, hg19: chr16:75576496T>C)	3	*TMEM231*	Biallelic	78/1613500/4.83E-05	Retinal disorders, Cerebellar hypoplasia, Rod-cone dystrophy	*TMEM231*-related Joubert syndrome (G2P)	Heterozygous, no clear candidate compound heterozygous variant identified.	Strong dev effects in brain and cerebellum at junction chr16:75575353:75576499:-
chr16:75542601C>T (NM_001077418.3:c.664+1G>A, hg19: chr16:75576499C>T)	1	*TMEM231*	Biallelic	4/1613706/2.48E-06	Cystic renal disease	*TMEM231*-related Joubert syndrome (G2P)	Heterozygous, no clear candidate compound heterozygous variant identified.	Strong dev effects in brain and cerebellum at junction chr16:75575353:75576499:-
chr9:13147659C>G (NM_001378778.1:c.3631-1G>C, hg19: chr9:13147658C>G)	1	*MPDZ*	Biallelic	21/1607054/1.31E-05	Hereditary ataxia	*MPDZ*-related nonsyndromic hydrocephalus (G2P)	Heterozygous, no clear candidate compound heterozygous variant identified.	Strong dev effects in brain, heart, kidney, ovary and cerebellum at junction chr9:13147657-13150509:-
chr9:89334732T>C (NM_024077.5:c.1089+2T>C, hg19: chr9:91949647T>C)	1	*SECISBP2*	Biallelic	1/1608240/6.22E-07	Paediatric disorders	*SECISBP2*-related thyroid hormone metabolism, abnormal (G2P)	Heterozygous, no clear candidate compound heterozygous variant identified.	Strong dev effects in brain at junction chr9:91949564:91953372:+
chr10:60166590C>A (NM_020987.5:c.2614+1G>T, hg19: chr10:61926348C>A)	1	*ANK3*	Biallelic	11/1612514/6.82E-06	Intellectual disability	Intellectual disability (ClinGen)	Heterozygous. Does not segregate in family	Strong dev effects in heart at junction chr10:61898845-61926348:-

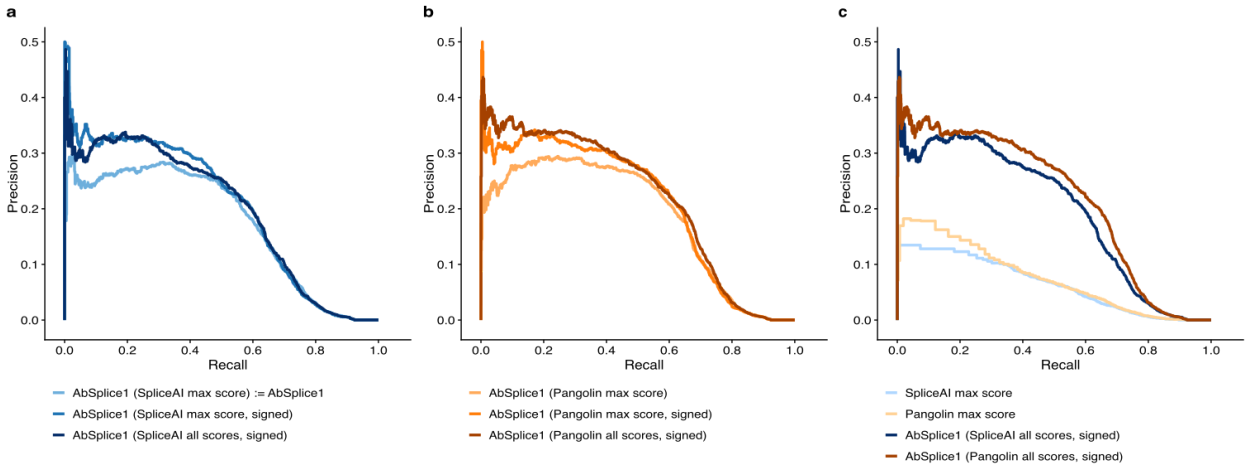
chr7:108237751C>T (NM_001037132.4:c.124+1G>A, hg19: chr7:107878195C>T)	1	*NRCAM*	Biallelic	21/1592558/ 1.32E-05	Paediatric disorders	*NRCAM*-related neurodevelopmental disorder with dysmorphic features, hypotonia, and spasticity (G2P)	Heterozygous, no clear candidate compound heterozygous variant identified.	Strong dev effects in brain and cerebellum at junction chr7:107875132-107878195:-,
chr8:144085277T>G (NM_003801.4:c.1261-12T>G, hg19: chr8:145140180T>G)	2	*GPAA1*	Biallelic	37/1588380/ 2.329e-5	Hereditary ataxia, Ultra-rare undescribed monogenic disorders	*GPAA1*-related developmental delay, epilepsy, cerebellar atrophy, and osteopenia (G2P)	Heterozygous. ClinVar Likely Benign	Strong dev effects in heart and liver at junction chr8:145140041-145140191:+, Some dev effects in brain at junction chr8:145139512-145140191:+
chr1:46190809T>G (NM_017739.4:c.1540-25A>C, hg19: chr1:46656481T>G)	1	*POMGNT1*	Biallelic	22/1584768/ 1.39E-05	Congenital myopathy	*POMGNT1*-related muscular dystrophy-dystroglycanopathy congenital with brain and eye anomalies/ intellectual developmental disorder, type B, 3/ limb-girdle (G2P)	Heterozygous, no clear candidate compound heterozygous variant identified.	Strong dev effects in multiple junctions in all tissues
chr2:206149017T>C (NM_005006.7:c.338+3A>G, hg19: chr2:207013741T>C)	1	*NDUFS1*	Biallelic	42/1601574/ 2.62E-05	Epilepsy plus other features	*NDUFS1*-related Leigh syndrome, *NDUFS1*-related mitochondrial respiratory chain complex I deficiency (G2P)	Heterozygous, no clear candidate compound heterozygous variant identified.	Strong dev effects in brain at junction chr2:207012558:207013743:-
chr4:127891683T>C (NM_014264.5:c.2038+2T>C, hg19: chr4:128812838T>C)	1	*PLK4*	Biallelic	0/1276152/0 .00E+00	Paediatric disorders	*PLK4*-related microcephaly, growth failure and retinopathy (G2P)	Heterozygous. Case Solved SOS1	Strong dev effects in brain, cerebellum, ovary, heart and liver at junction chr4:128812836-128813519:+. In testis opposite dev effects.
chr5:6606872C>T (NM_017755.6:c.1549G>A, hg19: chr5:6606985C>T)	1	*NSUN2*	Biallelic	Not in gnomAD	Intellectual disability	*NSUN2*-related intellectual disability (G2P)	Heterozygous. Poor match with specific phenotype	Dev effects in brain and cerebellum at multiple junctions
chr3:51987627A>T (NM_000666.3:c.921+3A>T, hg19: chr3:52021643A>T)	1	*ACY1*	Biallelic	1/1613852/6 .20E-07	Intellectual disability	*ACY1*-related aminoacylase 1 deficiency (G2P)	Heterozygous, no clear candidate compound heterozygous variant identified.	Minor dev effects
chr1:21838825C>T (NM_005529.7:c.10150G>A, hg19: chr1:22165318C>T)	1	*HSPG2*	Biallelic	31/1611812/ 1.92E-05	Likely inborn error of metabolism	*HSPG2*-related Schwartz-Jampel syndrome, *HSPG2*-related dyssegmental dysplasia, Silverman-Handmaker type (G2P)	Heterozygous. Not good phenotypic match	Minor dev effects

Supplementary Table 1: 22 candidate variants identified within the UK National Genomics Research Library (NGRL), including the number of individuals in the NGRL observed with that variant (GEL Count); the allele frequency in gnomAD v.4, the mode of inheritance and gene-associated disorder as recorded in DECIPHER, the disorder for which the participant was recruited, the disorder most closely linked to the gene as provided by DECIPHER from the Gene2Phenotype (G2P) and ClinGen databases, notes on prioritisation including existing identified alternate candidate genes, and how well HPO terms recorded for the AbSplice2 variant gene match with the individual's phenotype, as well as information on AbSplice2 predictions across tissues and developmental stages. Where an existing variant has been identified, AbSplice variants were assessed with reference to whether they could further contribute to an explanation of the individual's phenotype. Junction coordinates in AbSplice2 observations are in hg19.

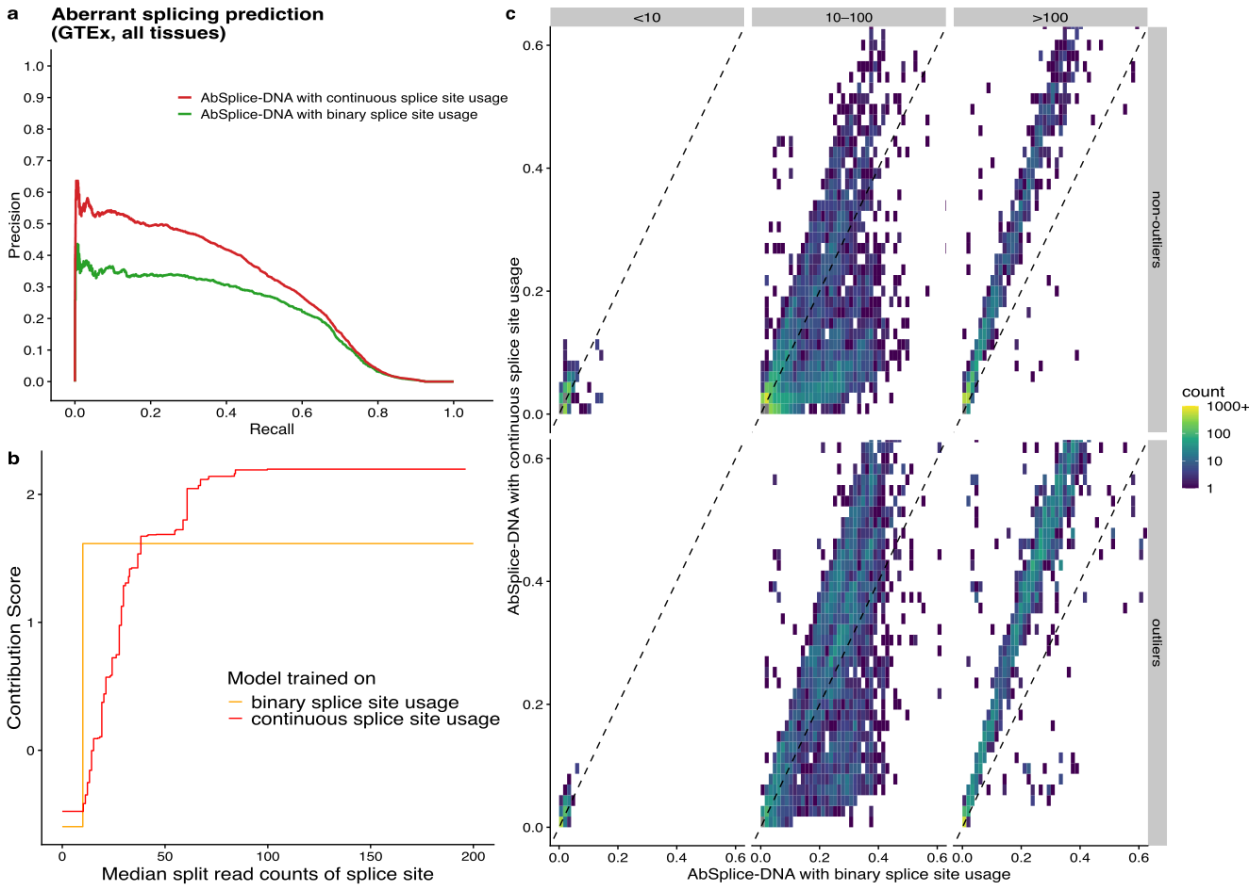
Supplementary Figures



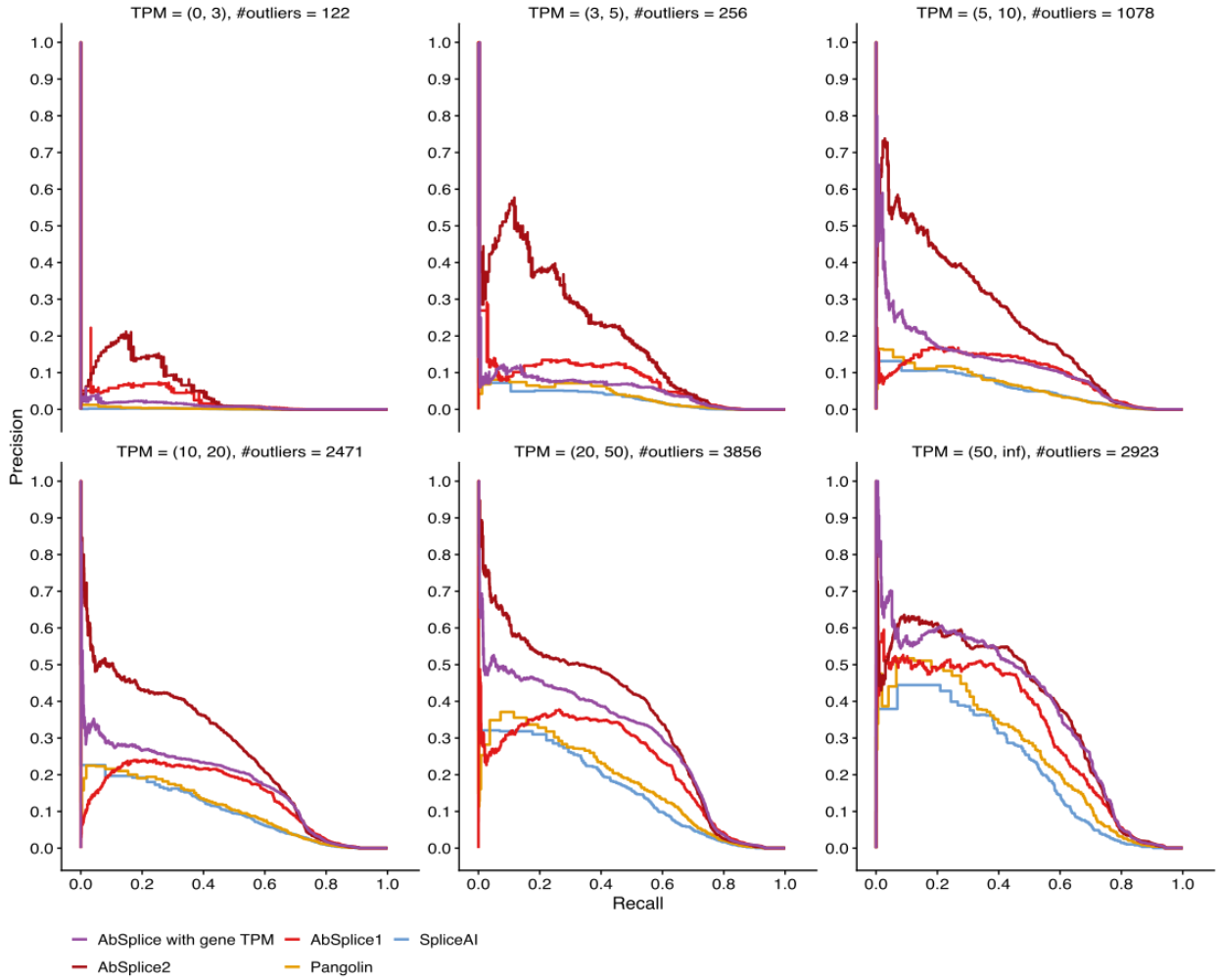
Supplementary Fig. 1: Retraining on an improved ground truth does not alter the feature contributions. a) Precision–recall curve comparing the overall prediction performance on all GTEx tissues of AbSplice models which were trained and evaluated on FRASER1 and FRASER2 splicing outlier ground truth respectively. The color legend is shared across (**a**,**b**). **b)** Distribution of the auPRC of the models shown in (**a**) across tissues (n = 49). Center line, median; box limits, first and third quartiles; whiskers span all data within 1.5 interquartile ranges of the lower and upper quartiles. **c)** Feature contribution functions (the higher the more important to the overall probability prediction[52]) of AbSplice models trained on FRASER1 and FRASER2 ground truth. The feature contributions did not change strongly while changing the ground truth.



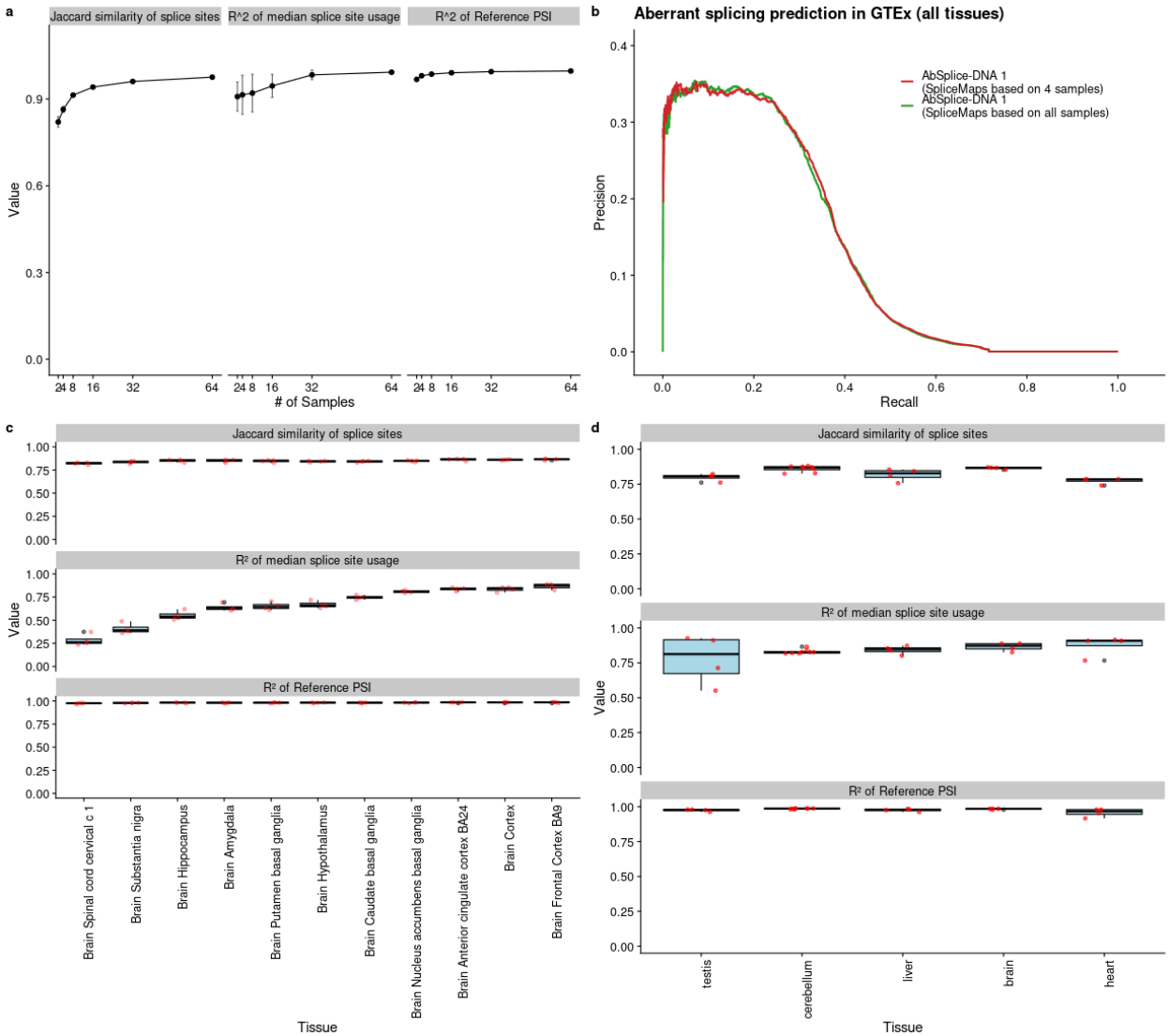
Supplementary Fig. 2: Using the directionality of predicted splicing changes improves model performance. Precision–recall curve comparing the overall prediction performance on all GTEx tissues of AbSplice models using different sets of input features from the deep learning models SpliceAI and Pangolin. All displayed models in (**a,b**) are variations of AbSplice1. **a**) AbSplice1 (light blue). Maximum SpliceAI Delta score is replaced by the signed maximum SpliceAI Delta score (blue) or by the four SpliceAI Delta scores of acceptor gain, acceptor loss, donor gain and donor loss (dark blue). **b**) Maximum SpliceAI Delta score is replaced by the maximum absolute value of the Pangolin scores (light orange) or by the Pangolin score with largest absolute value (orange) or the two Pangolin scores (dark orange). **c)** Performance comparison of Pangolin and SpliceAI as well as the best performing AbSplice models trained on Pangolin and SpliceAI scores (from models shown in (**a**,**b**)).



Supplementary Fig. 3: Using continuous splice site usage improves model performance. **a)** Precision–recall curve comparing the overall prediction performance on all GTEx tissues of AbSplice models trained on a binary feature from SpliceMap indicating if the splice site is expressed in the target tissue (using a cutoff of 10 reads for the median number of split reads sharing the splice site) as well as on a continuous feature (the median number of split reads sharing the splice site). **b)** Comparison of the learned feature contribution functions of AbSplice models shown in (**a**) for the binary and continuous splice site usage feature. **c)** Comparison of AbSplice model scores shown in (**a**) for outliers (bottom row) and non-outliers (top row) in different splice site usage regimes (columns, median split read counts with the two cutoffs 10 and 100 match the break points in the binary splice site usage function depicted in (**b**)). Non-outliers were randomly subsampled to match 100 times the number of outliers for visualization purposes.

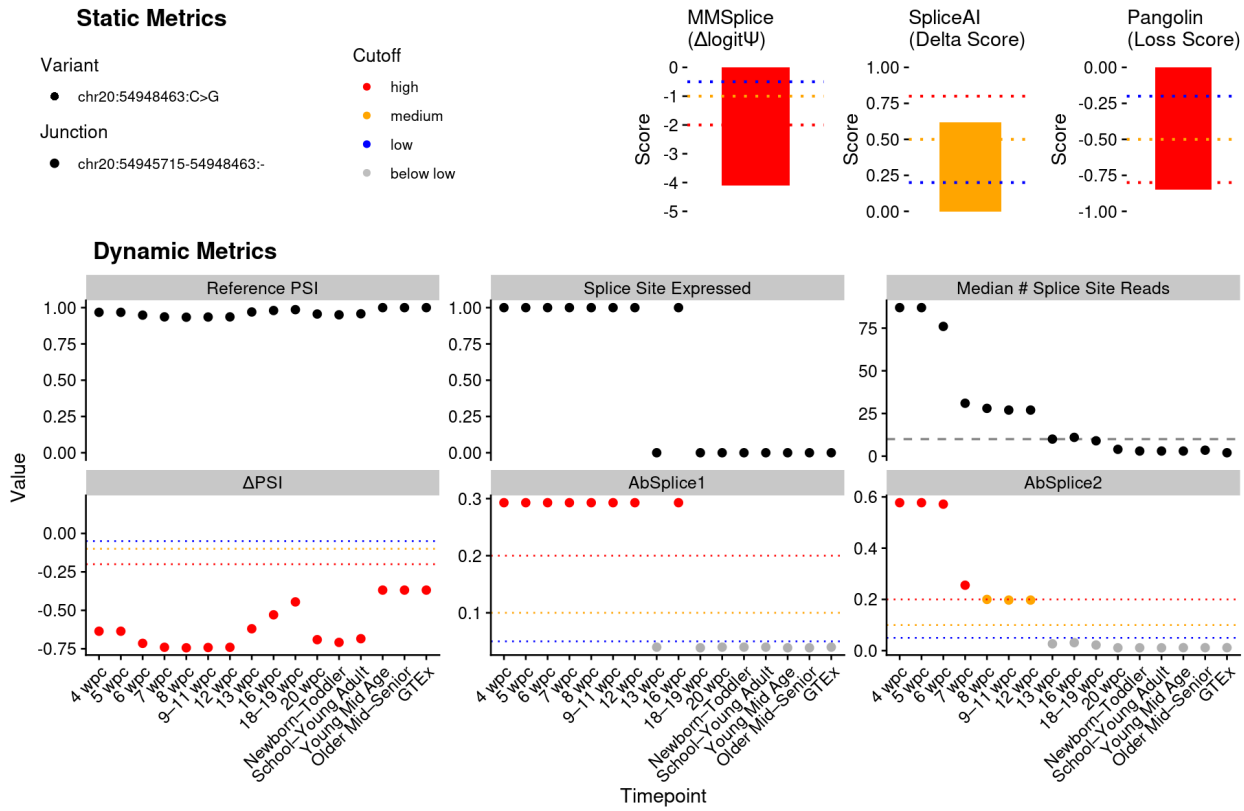


Supplementary Fig. 4: AbSplice model trained on continuous splice site usage levels outperforms other models in all gene expression regimes. a) Precision–recall curves comparing the overall prediction performance on all GTEx tissues of different models constrained for different gene expression regimes. Displayed are SpliceAI, Pangolin, AbSplice1, AbSplice2 and AbSplice2 with gene TPM which replaces the continuous splice site usage feature in AbSplice2 with the TPM value of the gene.

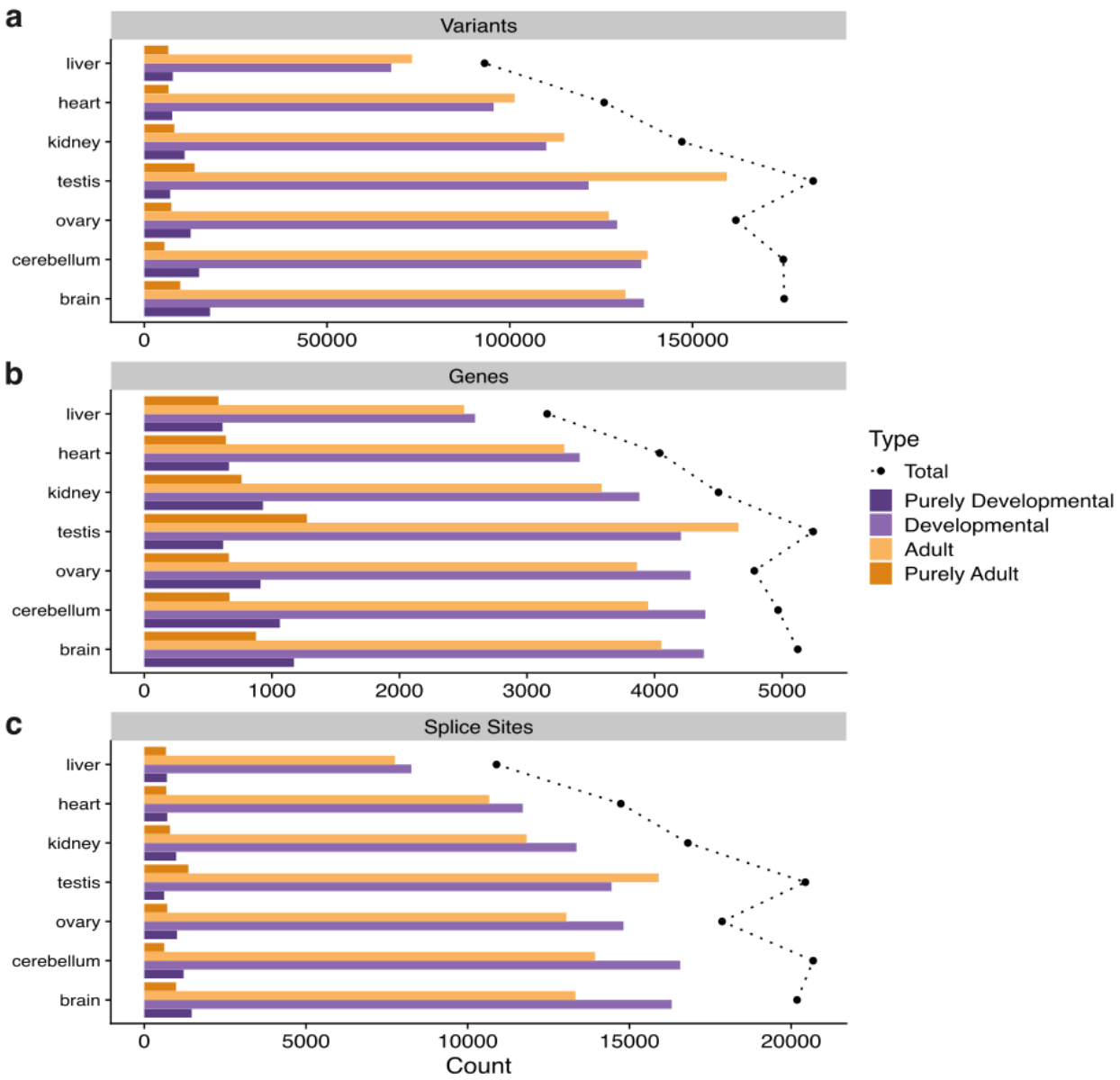


Supplementary Fig. 5: Robust SpliceMaps can be constructed from even a few RNA-seq samples. a) Comparison between SpliceMaps created from all available samples (N = 203) and different sample subsets in the GTEx tissue "Liver". Displayed metrics are the R^2 of Ψ_{ref} (computed on shared introns in the SpliceMaps), the R^2 of the median splice site usage (computed on shared introns in the SpliceMaps) and the Jaccard similarity of annotated splice sites. Error bars represent one standard deviation estimated with 10 random samplings with replacement. **b)** Precision-recall curve in the GTEx benchmark dataset comparing AbSplice predictions based on SpliceMaps created from all available samples in each tissue and predictions based on SpliceMaps created from only 4 random samples in each tissue. **c)** Comparison between SpliceMaps in brain tissues from developmental stages after birth (infant, teen, adult, senior) created from RNA-seq data of Mazin et al.[24] (N ≤ 8) and GTEx SpliceMaps from all sub tissues in the brain (N ≥ 121). Displayed metrics are the R^2 of Ψ_{ref} (computed on shared introns in the SpliceMaps), the R^2 of the median splice site usage (computed on shared introns in the SpliceMaps) and the Jaccard similarity of annotated splice sites. Center line, median (across developmental stages); box limits, first and third quartiles; whiskers span all

data within 1.5 interquartile ranges of the lower and upper quartiles. **d)** Comparison between SpliceMaps in all available tissue types from developmental stages after birth (infant, teen, adult, senior) created from RNA-seq data of Mazin et al.[24] ($N \leq 8$) and GTEx SpliceMaps from the same tissue type ($N \geq 121$). For the brain and the heart in GTEx the most similar sub-tissue was chosen (brain: Brain Frontal Cortex BA9, heart: Heart Left Ventricle). Center line, median (across developmental stages); box limits, first and third quartiles; whiskers span all data within 1.5 interquartile ranges of the lower and upper quartiles.

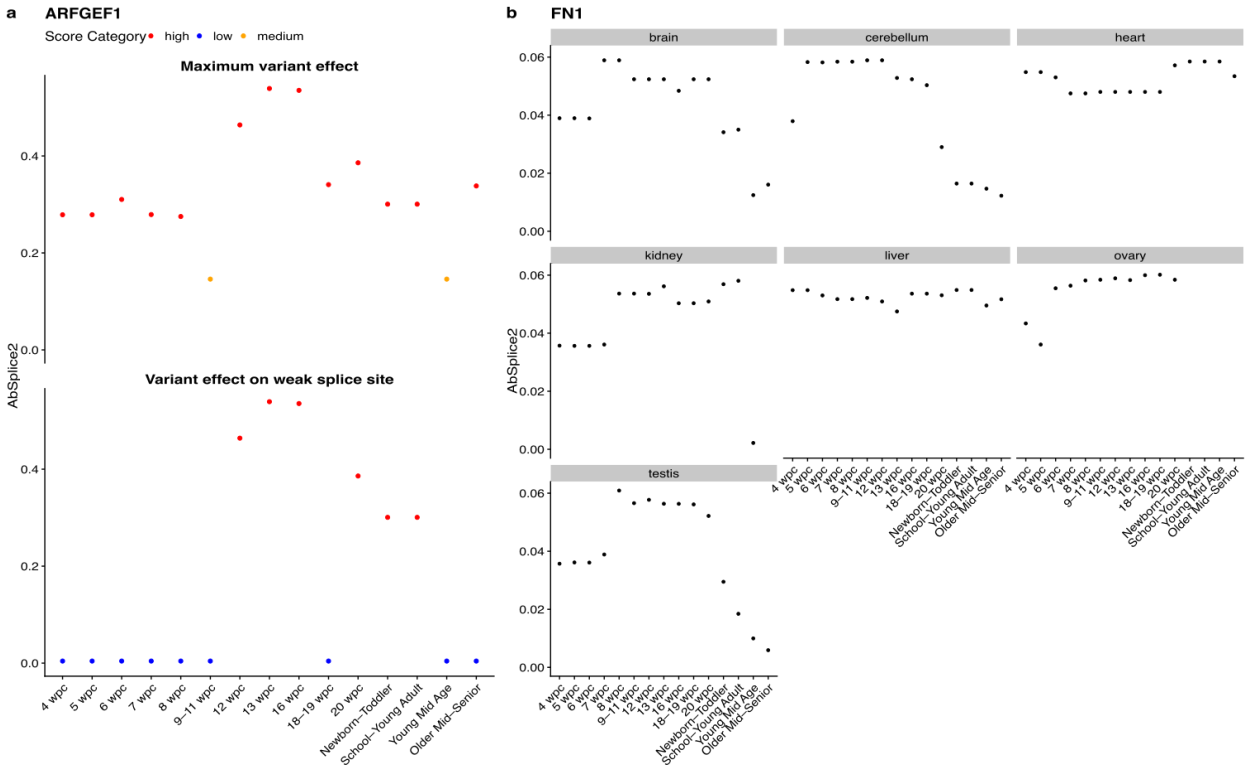


Supplementary Fig. 6: AbSplice2 provides smooth predictions around splice sites with low splice site usage. Predicted variant effects of the variant chr20:54948463:C>G (NM_198437.3:c.854+1G>C) affecting the junction chr20:54945715-54958463:- at different developmental timepoints in the brain. The deep learning models MMSplice, Pangolin and SpliceAI predict a variant effect independent of the developmental timepoint. The splice site usage of the junction changes over the course of development and is captured in the metric Ψ_{ref} and the median number of split reads of the splice site annotated in the SpliceMaps of the respective timepoints. One of the input features for AbSplice1 was the binarized splice site usage with a cutoff of 10 split reads (highlighted with a dashed grey line). This binarization led to abrupt prediction shifts based on minor changes in splice site usage. AbSplice2 which made use of the non-binarized splice site usage produced much smoother prediction changes better reflecting the usage changes. Genomic coordinates are in hg19.

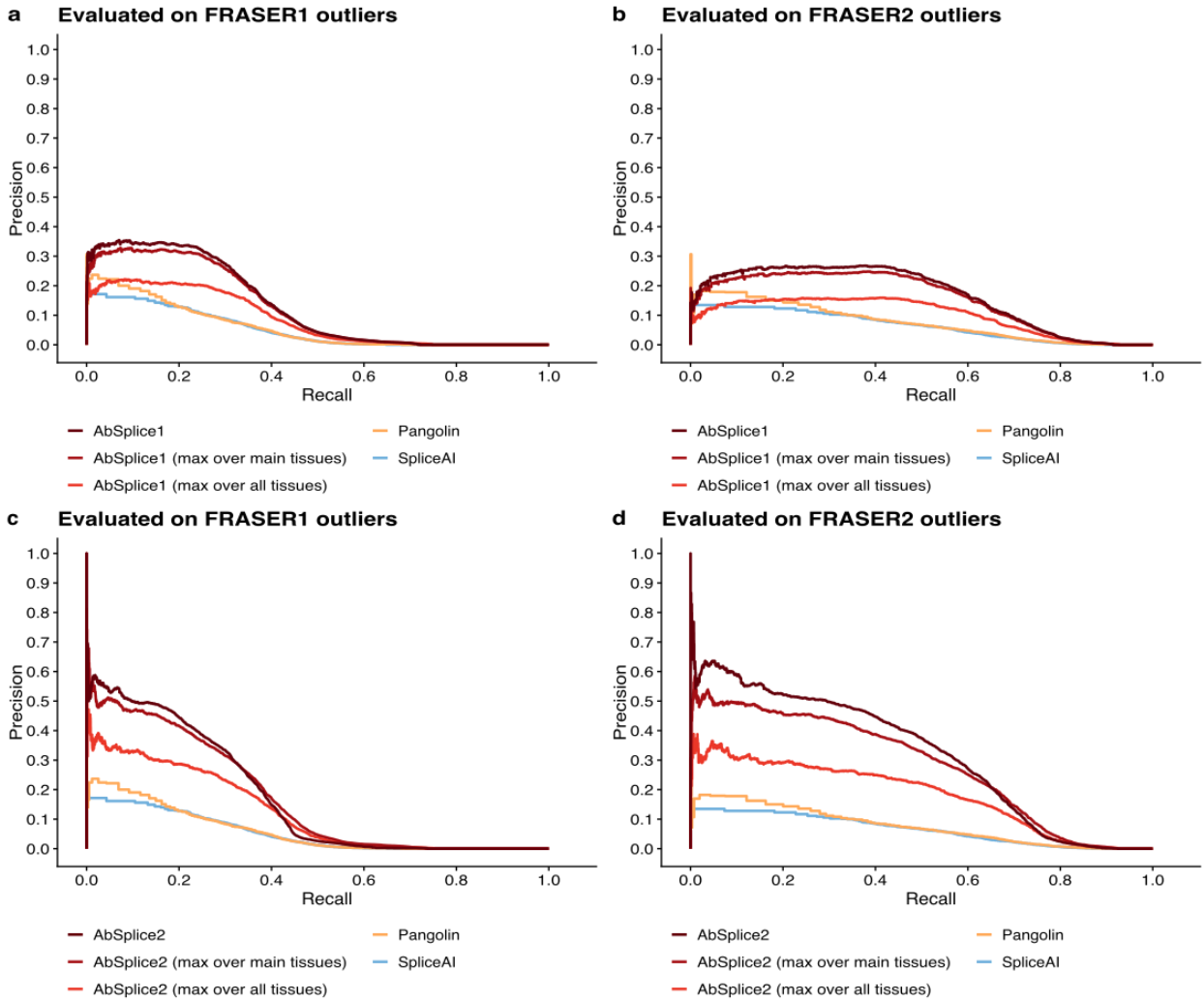


Supplementary Fig. 7: Statistics of variants with a high AbSplice2 score around devAS sites. All possible SNVs around devAS sites were scored with AbSplice2 for 15 developmental stages (including early development and adult stages) and 7 tissues. Variant categories were defined based on AbSplice2 scores. The category 'Total' contains all variants with a score above 0.2 in at least one of the 15 developmental stages. The category 'Developmental' contains all variants with a score above 0.2 in at least one early developmental stage (from embryogenesis to toddler). The category 'Purely Developmental' contains all variants with a score above 0.2 in at least one early developmental stage (from embryogenesis to toddler) and a score below 0.05 for adult stages. The category 'Adult' contains all variants with a score above 0.2 in at least one adult stage. The category 'Purely Adult' contains all variants with a score above 0.2 in at least one adult stage and a score below 0.05 for all early developmental stages (from embryogenesis to toddler). **a)** The number of variants per tissue for the different

variant categories. **b)** The number of unique genes affected by the variants in (a). **c)** The number of unique splice sites affected by the variants in (a).



Supplementary Fig. 8: Application to rare disease diagnostics. a) The variant chr8:67203252:C>T (NM_006421.5:c.4960-1G>A) in the gene *ARFGEF1* has been identified as the causal variant in an individual recruited with intellectual disability within the NGRL. The variant activates a weak splice site within the exon causing a frameshift. Top. Maximum AbSplice2 effect predictions of the variant regarding loss of the canonical site and the gain of the weak splice site. Bottom. AbSplice2 effect predictions of the variant on the weak splice site. The effect is predicted to be strong in the brain between 12 and 16 weeks post-conception. **b)** The variant chr2:215372382:G>C (NM_212482.4:c.6248-7C>G) in the gene *FN1* was identified in an individual recruited to the NGRL with a diagnosis of skeletal dysplasia and developmental phenotypes. While the most relevant tissue for this phenotype was not available in our dataset, AbSplice2 predicted developmentally dynamic splicing effects across multiple tissues, with higher predicted impact during early development. This pattern supports the variant as a strong candidate, given the early clinical manifestation.



Supplementary Fig. 9: Using the maximum across tissues significantly decreases performance in predicting tissue-specific aberrant splicing. Precision–recall curves comparing the overall prediction performance on all GTEx tissues of different models evaluated on FRASER1 and FRASER2 outlier ground truth. For AbSplice models the tissue-specific scores are compared to the maximum scores across all tissues and main tissue types. **a)** SpliceAI, Pangolin and AbSplice1 models evaluated on FRASER1 ground truth. **b)** SpliceAI, Pangolin and AbSplice1 models evaluated on FRASER2 ground truth. **c)** SpliceAI, Pangolin and AbSplice2 models evaluated on FRASER1 ground truth. **d)** SpliceAI, Pangolin and AbSplice2 models evaluated on FRASER2 ground truth.

Methods

Datasets

GTEx

We downloaded the RNA-seq read alignment files (BAM files) and the variant calling files (VCF files) from WGS from GTEx v8p (hg38) from the database of Genotypes and Phenotypes (dbGaP) (study accession: phs000424.v8.p2). We used data from 946 individuals with paired WGS and RNA-seq measurements (n = 16,213) in at least one tissue.

BrainGVEx

We downloaded the RNA-seq read alignment files (BAM files) and the variant calling files (VCF files) from WGS of the BrainGVEx dataset from the PsychENCODE portal (SynID: syn4590909[53]). We used a total of 230 samples isolated from the Brain Frontal Cortex of individuals with bipolar disorder or schizophrenia as well as from neurotypical controls.

Developmental dataset

Annotations for devAS events were taken from Supplementary files of [24]. The RNA-seq data for all developmental stages and tissues was downloaded from[29].

Data preprocessing

Most data preprocessing was done analogously to our previous publication[14]. We describe here only the differences, namely replacing our ground truth of aberrantly spliced events called by FRASER with events called by FRASER2.

Splicing outlier detection

FRASER2 (v1.99.4) was run in the different datasets via the Detection of RNA Outliers Pipeline[54] (DROP, v.1.4.0) with the default parameters and cut-offs. The main difference between FRASER2 and FRASER is the use of a new intron-specific metric (called intron Jaccard index) replacing the previous splice-site specific metrics. Significant junctions were those with an absolute splice deviation (delta Jaccard) > 0.1 and a FDR < 0.1.
We additionally applied the rare variant filter as described in[14], filtering for splicing outliers containing a rare variant in the vicinity of ±250 bp of every splice site based on RNA-seq from the sample. Importantly, this filter was applied to all splice sites identified by FRASER2, which includes both annotated splice sites as well as cryptic splice sites.

Aberrant splicing prediction benchmark

The benchmark was done analogously to our previous study[14], with the task being to predict whether a protein coding gene with one or more rare variants within the gene body is aberrantly spliced in a given tissue of an individual. The performance of models was evaluated using precision-recall curves and the average precision score.

Tissue-specific SpliceMap

SpliceMaps were created analogously to our previous study[14].

Developmental SpliceMaps

We created developmental SpliceMaps by using RNA-seq data from[29] in 7 human tissues (brain, cerebellum, heart, liver, testis, ovary, kidney) and for 15 different developmental timepoints (from 4 weeks post conception to adult stages). Due to small cohort sizes, each SpliceMap for a given developmental stage was created from samples corresponding to the same tissue using a sliding window of size 1 for the developmental stage.

Aberrant splicing prediction models

Some aberrant splicing prediction models (MMSplice, AbSplice1) were used analogously to our previous study[14]. We highlight here only the changes.

Pangolin

Pangolin is a deep learning model that predicts the effect of a variant on the splice site usage. It computes a gain and a loss score for every position within a user-defined window around the variant that represents the increase and decrease in the usage of a potential splice site at the respective positions. Pangolin outputs the maximum gain and the maximum loss scores within the window together with the corresponding positions. It also provides an option to mask scores when a genome annotation is provided to the model, which sets those scores to zero if Pangolin predicts activation for annotated splice sites and deactivation for unannotated splice sites. We ran Pangolin with the default settings, i.e. a window of length 50 nt and masking using the Gencode annotation (release 38). We computed genome-wide predictions for all possible SNVs in protein coding genes which are made publicly available (https://doi.org/10.5281/zenodo.15649338).
We intersected the genomic coordinates of the predicted affected splice site position with splice sites annotated in the tissue-specific SpliceMaps. The splice site usage of the affected splice site along with the variant effect score were provided as input features to the AbSplice2 model.

AbSplice2

Training of the AbSplice2 model was analogous to our previous study[14] changing the input features of the model and using FRASER2 outlier labels as a ground truth. The features of the

final AbSplice2 model were the prediction score from MMSplice + SpliceMap, MMSplice + SpliceMap + ψ_ref, the predicted loss and gain scores from Pangolin and a continuous feature from SpliceMap quantifying splice site usage in the target tissue (the median number of split reads per sample sharing the same acceptor/donor site, depending on which splice junction was used to score the variant in MMSplice + SpliceMap) as well as the splice site usage of the predicted affected site by Pangolin.
We computed genome-wide predictions for all possible SNVs in protein coding genes which are made publicly available (https://doi.org/10.5281/zenodo.15644548).

Developmental variants genome-wide

We computed all possible SNVs within a 100bp window from all splice sites related to devAS events annotated in the Supplementary files from Mazin et al.[24]. We used SpliceMaps from all 15 developmental stages and all 7 tissues to compute AbSplice2 scores for each developmental stage.

Analysis of variants in the NGRL

We identified 39,915 participants from the Genomics England 100,000 Genomes Project (100k) and 24,821 participants from the Genomic Medicine Service (GMS) datasets within the UK NGRL. Consenting probands were selected where they had a genome sequence obtained from blood samples, aligned to GRCh38, and where variants were called using the 'V4' pipeline. Where multiple alignment files existed, the most recent was selected, and for any remaining duplicates, the specific variant file used in the GEL interpretation pipeline was chosen.

Candidate variants were filtered to those that were fully genotyped and passed GEL's 'Stringent'[55] filters (n=552 unique variants in 100k, 272 unique variants in GMS). Variants were further filtered to those that are in 'Green' PanelApp genes for each participant's phenotype (n=22 unique variants in 100k, and 17 unique variants in GMS).

We further subset variants to those where the AbSplice2 score (in any of the seven tissues with developmental data) was greater at early developmental stages compared to the adult prediction (n= 15 unique variants in 100k, 13 unique variants in GMS). 26 unique variants in 47 individuals were represented in total.

One variant (chr8:67203252 C/T, NM_006421.5:c.4960-1G>A) was identified in a 'solved' case in the NGRL Exit Questionnaire table, with this variant provided as the diagnostic variant (Table 1) (positive controls). Eight additional participants were flagged as 'solved', with an alternative variant recorded. We compared these participants' Human Phenotype Ontology (HPO) terms with the HPO terms associated with both the gene in the exit questionnaire, and the gene identified using AbSplice2. None were supportive of an expansion to the existing diagnosis.

Each potential novel diagnostic variant was then assessed based on:

- Allele frequency from gnomAD

- Presence in ClinVar - removed if benign, no conflicts
- The genotype of the variant matched the reported gene-disease mechanism in DECIPHER
- Phenotypic overlap was assessed by comparing HPO terms provided for each participant, with HPO terms associated with the gene in the Human Phenotype Ontology[56], accounting for frequency of association where available.

For plausible variants in genes with likely biallelic inheritance (n=23), we searched for a clear second pathogenic variant within the MANE transcript coordinates. Any additional variants were passed to Ensembl's VEP and assessed based on gnomAD (AF < 0.05), predicted consequence (missense, splice region, frameshift, stop-gained, start-lost) and LOFTEE high confidence LoF annotation (in line with known mechanisms for that gene as defined in DECIPHER), SpliceAI score > 0.2 CADD (> 25.3) and PhyloP (> 7.367) scores, and if present in ClinVar that they had not been classified as benign).

After this assessment, we prioritised three candidate variants (Table 1). For the purpose of anonymisation any combination of HPO terms for a given participant that were shared by fewer than 5 participants in the 100,000 genomes project were adjusted to parent/grandparent HPO terms where necessary, to ensure participants were not identifiable.

AbSplice web interface

A web interface (https://absplice.cmm.cit.tum.de) allows users to run AbSplice on input SNVs and indels. As input, it requires a list of up to 10 variants and the genome build (hg19 or hg38). In addition, the user can select the tissue(s) of interest out of 49 GTEx tissues (the default is all) as well as tissues corresponding to 15 different developmental timepoints (spanning embryogenesis to adult stages) in brain, cerebellum, heart, kidney, liver, ovary and testis. The user has the option to display only the maximum score across the selected tissues. The output displays the variant, Ensembl gene ID, HGNC Symbol, tissue, and AbSplice2 score, as well as Pangolin and MMSplice scores. The output can be downloaded as a CSV file.

Data availability

No new data was generated for this study. SpliceMaps for all 49 GTEx tissues (hg19 and hg38) and all 15 developmental timepoints (hg19) in brain, cerebellum, heart, kidney, liver, ovary and testis are available at Zenodo; hg19 SpliceMaps: https://zenodo.org/records/7821508, hg38 SpliceMaps: https://doi.org/10.5281/zenodo.6387937. Precomputed Pangolin scores (hg38) are available at Zenodo, https://doi.org/10.5281/zenodo.15649337. Precomputed AbSplice2 scores (hg38) in all 49 GTEx tissues are available at Zenodo, https://doi.org/10.5281/zenodo.15644548. Precomputed AbSplice2 scores within 100 bp of devAS sites for all 15 developmental timepoints (hg19) are available at Zenodo, https://doi.org/10.5281/zenodo.15674612. Due to potential donor re-identification when revealing rare variants, the GTEx benchmark dataset cannot be shared without restrictions.

Users with access to the GTEx data can reproduce the benchmark using the code repository below. The GTEx v8 dataset is available at (under dbGaP protection) https://gtexportal.org/home. The BrainGVEx dataset is available at the PsychENCODE portal (SynID: syn459090949). Research on the de-identified patient data used in this publication can be carried out in the Genomics England Research Environment subject to a collaborative agreement that adheres to patient led governance. All interested readers will be able to access the data in the same manner that the authors accessed the data. For more information about accessing the data, interested readers may contact research-network@genomicsengland.co.uk or access the relevant information on the Genomics England website: https://www.genomicsengland.co.uk/research.

Code availability

AbSplice2 predictions using the tissue-specific SpliceMap annotations can be performed with the custom-written python package 'absplice2' (publicly available at: https://github.com/gagneurlab/absplice2). The analyses are available at: https://github.com/gagneurlab/AbSplice2_analysis.

References

1. Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548.e24 (2019).
2. Cheng, J. *et al.* MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol.* **20**, 48 (2019).
3. Yeo, G. & Burge, C. B. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *J. Comput. Biol.* **11**, 377–394 (2004).
4. Rosenberg, A. B., Patwardhan, R. P., Shendure, J. & Seelig, G. Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences. *Cell* **163**, 698–711 (2015).
5. Xiong, H. Y. *et al.* The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
6. Rentzsch, P., Schubach, M., Shendure, J. & Kircher, M. CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome*

Med. **13**, 31 (2021).

7. Danis, D. *et al.* Interpretable prioritization of splice variants in diagnostic next-generation sequencing. *Am. J. Hum. Genet.* **108**, 2205 (2021).
8. Cheng, J., Çelik, M. H., Kundaje, A. & Gagneur, J. MTSplice predicts effects of genetic variants on tissue-specific splicing. *Genome Biol.* **22**, 94 (2021).
9. Jenkinson, G. *et al.* LeafCutterMD: an algorithm for outlier splicing detection in rare diseases. *Bioinformatics* **36**, 4609–4615 (2020).
10. Ferraro, N. M. *et al.* Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science* **369**, eaaz5900 (2020).
11. Mertes, C. *et al.* Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat. Commun.* **12**, 529 (2021).
12. Scheller, I. F., Lutz, K., Mertes, C., Yépez, V. A. & Gagneur, J. Improved detection of aberrant splicing using the Intron Jaccard Index. 2023.03.31.23287997 Preprint at https://doi.org/10.1101/2023.03.31.23287997 (2023).
13. Segers, A. *et al.* saseR: Juggling offsets unlocks RNA-seq tools for fast and Scalable differential usage, Aberrant Splicing and Expression Retrieval. 2023.06.29.547014 Preprint at https://doi.org/10.1101/2023.06.29.547014 (2024).
14. Wagner, N. *et al.* Aberrant splicing prediction across human tissues. *Nat. Genet.* **55**, 861–870 (2023).
15. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
16. Cao, X. *et al.* Analysis of 3,760 hematologic malignancies reveals rare transcriptomic aberrations of driver genes. 2023.08.08.23293420 Preprint at https://doi.org/10.1101/2023.08.08.23293420 (2024).
17. Hölzlwimmer, F. R. *et al.* Aberrant gene expression prediction across human tissues. *Nat. Commun.* **16**, 3061 (2025).

18. Clarke, B. *et al.* Integration of variant annotations using deep set networks boosts rare variant association testing. *Nat. Genet.* **56**, 2271–2280 (2024).

19. Rowlands, C. F., Baralle, D. & Ellingford, J. M. Machine Learning Approaches for the Prioritization of Genomic Variants Impacting Pre-mRNA Splicing. *Cells* **8**, 1513 (2019).

20. Hwang, H., Jeon, H., Yeo, N. & Baek, D. Big data and deep learning for RNA biology. *Exp. Mol. Med.* **56**, 1293–1321 (2024).

21. Capitanchik, C., Wilkins, O. G., Wagner, N., Gagneur, J. & Ule, J. From computational models of the splicing code to regulatory mechanisms and therapeutic implications. *Nat. Rev. Genet.* **26**, 171–190 (2025).

22. Zeng, T. & Li, Y. I. Predicting RNA splicing from DNA sequence using Pangolin. *Genome Biol.* **23**, 103 (2022).

23. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).

24. Mazin, P. V., Khaitovich, P., Cardoso-Moreira, M. & Kaessmann, H. Alternative splicing during mammalian organ development. *Nat. Genet.* **53**, 925–934 (2021).

25. Smith, C. & Kitzman, J. O. Benchmarking splice variant prediction algorithms using massively parallel splicing assays. *Genome Biol.* **24**, 294 (2023).

26. Charles, M. *et al.* Functional impact of splicing variants in the elaboration of complex traits in cattle. *Nat. Commun.* **16**, 3893 (2025).

27. Neverov, A. & Wagner, N. AbSplice web interface. https://absplice.cmm.cit.tum.de.

28. Akbarian, S. *et al.* The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).

29. Cardoso-Moreira, M. *et al.* Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019).

30. Enkhmandakh, B., Makeyev, A. V. & Bayarsaihan, D. The role of the proline-rich domain of Ssdp1 in the modular architecture of the vertebrate head organizer. *Proc. Natl. Acad. Sci.* **103**, 11631–11636 (2006).

31. Liu, J. *et al.* Single-stranded DNA binding protein Ssbp3 induces differentiation of mouse embryonic stem cells into trophoblast-like cells. *Stem Cell Res. Ther.* **7**, 79 (2016).

32. Genomics England. The National Genomic Research Library v5.1. https://doi.org/doi:10.6084/m9.figshare.4530893/7. 2020.

33. Martin, A. R. *et al.* PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat. Genet.* **51**, 1560–1565 (2019).

34. Thomas, Q. *et al.* Haploinsufficiency of ARFGEF1 is associated with developmental delay, intellectual disability, and epilepsy with variable expressivity. *Genet. Med.* **23**, 1901–1911 (2021).

35. Kember, J., Gracia-Tabuenca, Z., Patel, R., Chakravarty, M. & Chai, X. J. Neurite development varies across the hippocampus and covaries with the cellular composition of hippocampal tissue. *Commun. Biol.* **8**, 1–10 (2025).

36. Bryen, S. J. *et al.* Prevalence, parameters, and pathogenic mechanisms for splice-altering acceptor variants that disrupt the AG exclusion zone. *Hum. Genet. Genomics Adv.* **3**, (2022).

37. Tsai, P.-S., Brooks, L. R., Rochester, J. R., Kavanaugh, S. I. & Chung, W. C. J. Fibroblast growth factor signaling in the developing neuroendocrine hypothalamus. *Front. Neuroendocrinol.* **32**, 95–107 (2011).

38. Jarzabek, K., Wolczynski, S., Lesniewicz, R., Plessis, G. & Kottler, M. Evidence that *FGFR1* loss-of-function mutations may cause variable skeletal malformations in patients with Kallmann syndrome. *Adv. Med. Sci.* **57**, 314–321 (2012).

39. Kremer, L. S. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* **8**, 15824 (2017).

40. Dawes, R. *et al.* SpliceVault predicts the precise nature of variant-associated mis-splicing. *Nat. Genet.* **55**, 324–332 (2023).

41. Kopp, J. *et al.* Loss-of-function variants affecting the STAGA complex component

SUPT7L cause a developmental disorder with generalized lipodystrophy. *Hum. Genet.* **143**, 683–694 (2024).

42. Li, K. *et al.* Prioritizing de novo potential non-canonical splicing variants in neurodevelopmental disorders. *eBioMedicine* **99**, (2024).

43. Dominguez-Alonso, S. *et al.* Alternative splicing analysis in a Spanish ASD (Autism Spectrum Disorders) cohort: in silico prediction and characterization. *Sci. Rep.* **15**, 10730 (2025).

44. Ji, Y., Zhou, Z., Liu, H. & Davuluri, R. V. DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. *Bioinformatics* **37**, 2112–2120 (2021).

45. Chen, K. *et al.* Self-supervised learning on millions of primary RNA sequences from 72 vertebrates improves sequence-based RNA splicing prediction. *Brief. Bioinform.* **25**, bbae163 (2024).

46. Karollus, A. *et al.* Species-aware DNA language models capture regulatory elements and their evolution. *Genome Biol.* **25**, 83 (2024).

47. Wu, D. *et al.* Generative modeling for RNA splicing predictions and design. 2025.01.20.633986 Preprint at https://doi.org/10.1101/2025.01.20.633986 (2025).

48. Tang, X., Lei, H., Guo, J., Shen, Y. & Zhang, M. Spliformer-v2 predicts multi-tissue RNA splicing and reveals functional genomic links with neurodegenerative diseases. 2025.06.28.662167 Preprint at https://doi.org/10.1101/2025.06.28.662167 (2025).

49. Celaj, A. *et al.* An RNA foundation model enables discovery of disease mechanisms and candidate therapeutics. 2023.09.20.558508 Preprint at https://doi.org/10.1101/2023.09.20.558508 (2023).

50. Linder, J., Srivastava, D., Yuan, H., Agarwal, V. & Kelley, D. R. Predicting RNA-seq coverage from DNA sequence as a unifying model of gene regulation. *Nat. Genet.* **57**, 949–961 (2025).

51. Avsec, Ž. *et al.* AlphaGenome: advancing regulatory variant effect prediction with a unified DNA sequence model. 2025.06.25.661532 Preprint at https://doi.org/10.1101/2025.06.25.661532 (2025).

52. Nori, H., Jenkins, S., Koch, P. & Caruana, R. InterpretML: A Unified Framework for Machine Learning Interpretability. Preprint at https://doi.org/10.48550/arXiv.1909.09223 (2019).

53. BrainGVEX. https://doi.org/10.7303/syn4590909.

54. Yépez, V. A. *et al.* Detection of aberrant gene expression events in RNA sequencing data. *Nat. Protoc.* **16**, 1276–1296 (2021).

55. Aggregated Variant Calls (AggV2) - Genomics England Research Environment User Guide. https://re-docs.genomicsengland.co.uk/aggv2/.

56. Gargano, M. A. *et al.* The Human Phenotype Ontology in 2024: phenotypes around the world. *Nucleic Acids Res.* **52**, D1333–D1346 (2024).

Acknowledgements

We thank all the scientists who isolated and sequenced the data which builds the basis for our research. We thank the authors of Pangolin for making their code easily available and the authors of the developmental splicing study for providing such complete and well documented Supplementary data. We thank Alex Karollus, Pedro Tomaz da Silva, Johannes Hingerl, Laura Martens and Muhammed H. Çelik for insightful discussions and advice. We thank Christian Mertes for his insights on implementing the web interface. The Ethics Committee of the Technical University of Munich issued a statement of no objection regarding this study (reference #2025-263-S-KK). This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 461264291. This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via the project NFDI 1/1 "GHGA - German Human Genome-Phenome Archive" (#441914366 to N.Wagner, AMN, VAY, JG), the German Bundesministerium für Bildung und Forschung (BMBF) through the ERA PerMed project PerMiM (01KU2016B to N.Wagner, VAY, JG), the Helmholtz Association under the joint research school Munich School for Data Science - MUDS (to N.Wagner and SL). This study was funded by the European Research Council (ERC) (EPIC, Grant number: 101118521) funded by the European Union. ERDERA has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement N°101156595. Views and opinions expressed are those of the author(s) only and do not necessarily reflect

those of the European Union or any other granting authority, who cannot be held responsible for them. Parts of Fig. 1 and Fig. 3 were created with BioRender.com. The Genotype-Tissue Expression (GTEx) project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by the NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. This study was supported by data provided by the Answer ALS Consortium, administered by the Robert Packard Center for ALS at Johns Hopkins. The funders (of GTEx) had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. N.Whiffin is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant no. 220134/Z/20/Z) and a Lister Institute Research Prize.

This research was made possible through access to data in the National Genomic Research Library, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.

Author information

Contributions

N.Wagner and J.G. conceived the study. N.Wagner and A.M.N. designed the model and implemented the framework. N.Wagner, A.M.N., A.M.G., S.L., C.S. and V.A.Y. analyzed the data. A.M.G. performed the Genomics England NGRL analysis. A.M.N. implemented the web interface. N.Wagner and J.G. wrote the manuscript with input from all the authors. J.G. and N.Whiffin supervised the project.