Ancestral intronic splicing regulatory elements in the SCNα gene family

Ekaterina Chernyavskaya[1], Margarita Vorobeva[2], Sergei A. Spirin[1,3], Dmitry A. Skvortsov[1,2], and Dmitri D. Pervouchine[1,*]

[1]Center for Molecular and Cellular Biology, Moscow 143026, Russia

[2]Moscow State University, Faculty of Chemistry, Moscow 119991, Russia

[3]Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow 119991, Russia

[*]e-mail: pervouchine@gmail.com; tel/fax +7 (916) 220 69 99

August 16, 2025

Abstract

SCNα genes encode components of voltage-gated sodium channels that are crucial for generating electrical signals in many organisms. Humans have ten paralogous SCNα genes, some of which contain duplicated mutually exclusive exons 5a and 5b (MXE). In reconstructing their evolutionary history, we report multiple unannotated copies of exon 5 in dis-

tant species and show that exon 5 duplication likely goes back to a common ancestor of all SCNα genes. We provide evidence from RNA-seq experiments that the nonsense mediated decay (NMD) system plays a role in maintaining mutually exclusive choice of human exons 5a and 5b, and characterize their splicing patterns across tissues and developmental stages. Comparative analysis of nucleotide sequences among paralogs identified two highly conserved complementary intronic elements, IRE1 and IRE2, possibly forming RNA structure that impacts splicing. Indeed, blocking of IRE2 by antisense oligonucleotides significantly changed the 5a/5b ratio without the loss of mutually exclusive pattern. IRE1 and IRE2 are conserved from human to Elephant shark and may represent an ancient, evolutionarily conserved regulatory mechanism that is shared across many SCNα genes. These results demonstrate the power of comparative analysis of paralogs for elucidating conserved splicing regulatory mechanisms.

Keywords: SCN, mutually exclusive exons, splicing, RNA structure, duplication.

Introduction

Voltage-gated sodium channels are crucial for generation and propagation of action potentials in excitable cells including neurons, muscle cells, and some glial cells. The α-subunits of these channels, encoded by the SCNα gene family, form the principal ion-conducting pore [1]. Each α-subunit consists of four homologous domains, with each domain containing six transmembrane segments (S1–S6). The voltage-sensing module includes segments S1–S4 and undergoes conformational changes in response to membrane potential alterations, particularly through the movement of charged residues in the S4 segment, while segments S5–S6 form the

sodium-selective pore [2, 3, 4].

The human SCNα family comprises ten paralogs (*SCN1A–SCN11A*) which exhibit distinct tissue-specific expression patterns [1]. For instance, *SCN1A–SCN3A* and *SCN9A* are predominantly expressed in the nervous system, whereas *SCN4A* is found to be activated in skeletal muscle [5, 6]. A notable feature of *SCN1–3A*, *SCN5A*, *SCN8A* and *SCN9A* is the presence of splice isoforms with mutually exclusive exons 5a and 5b (MXE), the expression of which is tightly regulated during embryonic development [7, 8]. Imbalanced MXE isoform composition causes altered sodium channel permeability that has been linked to various pathologies, including cardiac-conduction delay, neuropathic pain, and epilepsy [9, 10, 11]. Similar alterations in the relative abundances of splice isoforms of other genes, particularly, voltage-gated calcium channels are involved in a broad group of disorders termed channelopathies [12].

Exons 5a and 5b are both 92-nt-long, hence their coordinated inclusion or skipping would cause a frameshift and generate a transcript isoform with a premature termination codon. Elimination of such transcripts by the nonsense mediated decay (NMD) pathway has been proposed as a plausible mechanism underlying mutually exclusive splicing in clusters with exactly two exons [13]. However, insights from other MXE systems suggest a potential role of RNA structure. A well-characterized example is the *Drosophila Dscam* gene, in which large clusters of MXEs are controlled by competing RNA structures [14, 15, 16]. It has been proposed that the acquisition of competing RNA structures and MXEs is a natural consequence of tandem genomic duplications which, on the one hand, generate a group of homologous exons and, on the other hand, produce a group of sequences that compete for complementary base pairings with another sequence [17, 18, 19].

Regulatory RNA structures are anecdotally found in paralogues [20]. A notable example is the group of RNA structures in the fibroblast growth factor receptor (FGFR) gene family, which control mutually exclusive splicing of exons IIIb and IIIc [21]. Complementary base pairings between intronic elements IAS2 and ISAR in the *FGFR2* gene have been conserved from sea urchin to humans [22], while sharing a common sequence with complementary intronic elements ISE-2 and ISAR in its paralog *FGFR1* [23]. In this case, regulatory RNA structures probably had been present in the common ancestor of *FGFR1* and *FGFR2* before gene duplication occurred. However, the example of the BET gene family demonstrates that RNA structures regulating alternative splicing in paralogs can also be acquired independently in the course of convergent evolution [24].

The SCNα genes represent one of the most studied gene families with respect to their function [25, 26], structure [27, 28], evolution [29, 30], role in disease and drug resistance [31, 32, 33]. Their common ancestry and, furthermore, a remarkable similarity of exon 5a and 5b sequences motivated us to examine this family in searching for intronic regulatory elements (IRE) involved in alternative splicing regulation. In this work, we characterize two IRE shared by most SCNα genes with exon 5 duplications, which are capable of forming RNA structures that impact splicing.

Results

Exon 5 duplications in the SCNα gene family

The human genome contains 10 paralogous SCNα genes distributed across different chromosomes, with five genes (*SCN1A*, *SCN2A*, *SCN3A*, *SCN7A*, and *SCN9A*) clustered on chromosome 2, three genes (*SCN5A*, *SCN10A*, and *SCN11A*) on chromosome 3, and the remaining two (*SCN8A* and *SCN4A*) on chromosomes 12 and 17, respectively. Genes sharing chromosomal locations arose through tandem duplications following the initial dispersion of ancestral genes to different chromosomes, as evidenced by their tandem genomic arrangement.

All ten human SCNα genes share a common exon structure in their core protein-coding part with minor variations in exon lengths, however some of them contain duplicated exon 5 variants (5a and 5b) showing on average 91% sequence identity (Figure 1A). At that, exons 5a from different genes are more similar to each other than they are to the respective exons 5b from the cognate gene (Figure 1B). This observation indicates that exon 5 duplication must have occurred in the common ancestor prior to gene duplication events that generated the current genomic organization of the SCNα gene family.

To reconstruct the evolutionary history of exon 5 duplications, we applied comparative sequence analysis to 76 orthologous genes from the SCNα family. Exon 5 duplications were widely distributed among jawed vertebrates and included several instances of unannotated duplications (Table 1). In particular, a previously unknown duplication of exon 5 was observed in *SCN2A* homologs in sheep and tropical clawed frog, as well as in *SCN1A* homologs in zebrafish and elephant shark, and in the *SCN5A* homolog in zebrafish. As might be expected,

exon 5 evolutionary dynamics involved not only acquisitions but also subsequent losses. For instance, exon 5 duplication was absent in the *SCN5A* homolog in the tropical clawed frog. Rodents (mouse, chinese hamster, guinea pig) showed the loss of the duplicated exon in *SCN1A* homolog, while their closest phylogenetic relatives, rabbits and american pikas, retained two copies of exon 5 (Figure 2A).

The phylogenetic trees of SCNα genes in jawless vertebrates and cephalochordata species yielded intriguing results (Figure 2B, the full tree in Figure S1). While no exon 5 duplications were found in jawless vertebrates, potentially due to genome assembly limitations, two out of the five SCNα genes in the *B. lanceolatum* contained duplications of the exon homologous to the human exon 5. The gene LOC136428574 (sodium channel protein 1 brain-like) contained an annotated exon 5 duplication, while LOC136422135 (sodium channel protein type 4 subunit alpha B-like) showed an unannotated exon 5 duplication with a remarkable level of sequence identity (Table 1). In tunicates, represented by *Ciona* with three annotated SCNα genes, only a partial duplication of exon 5 in the *Cinav1a* gene was detected, but the level of sequence identity unambiguously supports the occurrence of an ancestral duplication event (Figure 2B, bottom panel). While no duplications of exon 5 homolog were detected in protostomes (arthropods and mollusks), duplications of other exons (such as exon 19, which is duplicated in *D. melanogaster*) were still present. Previous studies attributed such duplication patterns to convergent evolution in different lineages [34].

In sum, our findings suggest that exon 5 duplications are characteristic of most human SCNα genes and likely existed in the common ancestor of all chordate species. This duplication may have originated in a common ancestor of chordates, and one of whose genes gave rise to all

human SCNα genes. The subsequent evolutionary processes involved multiple independent losses of duplicated exons, particularly in *SCN4A*, *SCN7A*, *SCN10A*, and *SCN11A*.

Splicing patterns of duplicated exons

In challenging the assumption that exons 5a and 5b are mutually exclusive, we reanalyzed RNA-seq experiments from the GTEx project focusing on six SCNα genes with duplicated exons (*SCN1A*, *SCN2A*, *SCN3A*, *SCN5A*, *SCN8A*, and *SCN9A*). Indeed, these genes exhibit distinct splicing patterns, with *SCN1A*, *SCN2A*, and *SCN3A* primarily expressing exon 5b, and *SCN9A* primarily expressing exon 5a (Figure 3A). The 5a/5b transcript ratio shows a moderate degree of variation across tissues, particularly across brain regions, but in *SCN8A* we observed almost a complete reversal of 5a/5b ratio in pituitary and testes, a pattern not observed earlier in any other SCNα gene. Simultaneous inclusion or skipping of both exons were rare, e.g., in *SCN2A* they were both completely skipped in pancreas. However, the pancreatic expression of *SCN2A* was remarkably low, consistent with the degradation of a transcript containing a premature stop codon by the NMD pathway.

To examine whether mutually exclusive splicing of exons 5a and 5b is regulated by NMD, we analyzed RNA-seq data from A549 cells treated with translation inhibitor cycloheximide, in which the activity of the NMD pathway was suppressed. Among SCNα genes with exon 5 duplications, only *SCN9A* showed a measurable level of expression. In this gene, we observed a substantial upregulation of isoforms containing double exons, i.e., both 5a and 5b ($\Delta\Psi = 0.1$) and isoforms with exon 5b ($\Delta\Psi = 0.12$) accompanied by downregulation of isoforms with exon 5a ($\Delta\Psi = -0.12$). Taken together these results indicate that exons 5a and 5b are, indeed,

mutually exclusive and that NMD pathway plays a role in the suppression of double exons.

Previous studies reported significant changes of 5a/5b isoform ratio during development [7, 8]. For example, neonatal *SCN5A* transcripts include exon 5a, whereas adult isoforms predominantly contain exon 5b. To validate this, we reanalyzed RNA-seq experiments from a study of mammalian organ development [35] and mouse embryonic transcriptomes [36]. In all SCNα genes with exon 5 duplications and detectable expression levels, we observed a significant change of the 5a/5b isoform ratio across developmental stages, which was concordant between human and mouse (Figure 3B, Figure S2A). Interestingly, in *SCN9A* we observed a previously undocumented switch in hindbrain and neural tube, where the isoform with exon 5a was dominant in prenatal but not in postnatal samples. On the other hand, the 5a/5b isoform ratio in postnatal neural tube resembled that of *SCN2A* and *SCN3A* (Figure S2B). No such postnatal switching occurred in the forebrain or midbrain.

Growing evidence indicates that altered SCNα isoform expression is implicated in disease [37, 38, 39]. To assess splicing patterns of SCNα genes in tumors, we reanalyzed RNA-seq experiments from TCGA and found that tumors exhibited a broad spectrum of 5a/5b splicing aberrations: some cases showed complete exon switching, others co-occurred with dual-exon skipping, and a subset displayed the loss of tissue specificity (Figure S3). Striking examples include reduced exon 5b inclusion in *SCN2A* (ESCA, BRCA) and reduced exon 5a inclusion in *SCN9A* (LUAD, LUSC). In BRCA, an elevated fraction of *SCN2A* transcripts lacking both exons was observed suggesting production of nonfunctional mRNA. A similar upregulation was also found for *SCN9A* in LUSC. While these findings demonstrate tumor-associated disruption of mutually exclusive splicing in SCNα genes, they provide no evidence of direct implication

of these genes in cancer pathogenesis, being merely a manifestation of global alterations of splicing regulatory programs and NMD pathway efficiency in tumors.

Common intronic splicing regulatory elements in human SCNα genes

The mutually exclusive choice of exons 5a and 5b in SCNα genes hints at the existence of a common regulatory mechanism. On the one hand, such a mechanism is the NMD pathway as it selectively eliminates non-MXE transcripts. However, the examples of *FGFR1* and *FGFR2* genes show that the degradation by NMD doesn't preclude RNA structure-based splicing regulation. Since *FGFR1* and *FGFR2* share complementary IRE inherited from the common ancestor [23], we examined the introns of SCNα genes for presence of similar common regulatory sequences.

To do this, we constructed a multiple sequence alignment of the nucleotide sequences of introns flanking exons 5a and 5b in the six human SCNα genes with exon 5 duplications (*SCN1A*, *SCN2A*, *SCN3A*, *SCN5A*, *SCN8A*, and *SCN9A*). The intronic sequence of *SCN8A* was too divergent, hence it was analyzed separately. The introns upstream of exon 5a showed almost no similarity besides the expected homology at splice site sequences and polypyrimidine tracts (Figure S4A); the intervening introns between exons 5a and 5b of the remaining five genes were remarkably similar (Figure S4B); while the introns downstream of exon 5b contained a pair of IRE, termed IRE1 and IRE2, which shared nearly the same sequence and were complementary to each other (Figure S5). Both regions were evolutionarily conserved across 100 vertebrates, with IRE1 being located 10–15 nts downstream of exon 5b, and IRE2 positioned 89–533 nts downstream.

We next examined whether IRE1 and IRE2 were present in the genome-wide catalog of pairs of conserved complementary regions (PCCR) [40] or supported by proximity ligation assays targeting RNA structure (RIC-seq [41], PARIS [42], and KARR-seq [43]). Indeed, base pairings between IRE1 and IRE2 in *SCN1A*, *SCN2A*, *SCN3A*, *SCN5A*, and *SCN9A* were annotated as PCCRs with free energies (ΔG) ranging from -15 kcal/mol to -23.6 kcal/mol and confirmed by KARR-seq, and by both KARR-seq and RIC-seq in *SCN2A* and *SCN9A* (Figure 4A–D). Phylogenetically independent complementary substitutions in IRE1 and IRE2 across mammalian species were detected in *SCN3A* ($E = 0.45$) and *SCN9A* ($E = 0.06$). In *SCN1A*, the base pairings between IREs were the weakest compared to those in the other genes ($\Delta G = -15$ kcal/mol) due to nucleotide substitutions that distinguish its IRE1 from those of other genes, while IRE2 sequence remained the same (Figure S5).

No sequences similar to IRE1 or IRE2 were detected in *SCN8A*, however there was an alternative complementary base pairing between intron and exon 5b sequences, which was supported by RIC-seq (Figure S6). Given that the divergence of *SCN8A* and *SCN1A*, *SCN2A*, *SCN3A*, *SCN9A* occurred later than that of *SCN8A* and *SCN5A*, the lack of common secondary structure may reflect its loss or evolutionary modification.

In sum, IRE1 and IRE2 sequences are common to introns of most SCNα genes with exon 5 duplications and are capable of forming complementary base pairings. Additionally, they are supported by RNA proximity ligation assays and contain traces of complementary substitutions. As it appears highly unlikely to observe all this by pure chance, we conclude that IRE1 and IRE2 may serve the purpose of forming RNA secondary structure.

Validation of the IRE impact on splicing

To test whether the interaction between IRE influences MXE splicing, we chose to quantitatively assess the inclusion levels of exons 5a and 5b in *SCN9A* using human Huh7 cells because the expression level of this gene in this cell line was the most prominent [44]. In designing the antisense oligonucleotides (AON) to disrupt IRE1/IRE2 base pairing, we found that AON reliably blocking IRE1 could interfere with the neighbouring splice site. Therefore, we focused on the AON2 sequence that blocks IRE2 (Figure 5A). AON2 was transfected into Huh7 cells at three different concentration levels followed by quantitative assessment of splicing by qPCR in comparison to the non-treated control and the treatment with a control AON targeting an offtarget site.

Cells treated with AON2 showed a significant concentration-dependent decrease of exon 5a inclusion by approximately 15%, accompanied by a significant increase of exon 5b inclusion by approximately 10%, and statistically insignificant variations in the proportion of double and skipped exons (Figure 5B). These results indicate that the IRE2 sequence is important for maintaining the balance of splice isoforms with exons 5a and 5b without disrupting mutually exclusive splicing, resembling the pattern observed earlier for the ATE1 gene [45].

Evolutionary origin of IRE in SCNα genes

The remarkable similarity of IRE sequences in the human SCNα genes with exon 5 duplications suggests that these regulatory elements may have existed in their common ancestor, or at least in the ancestral form of *SCN1A*, *SCN2A*, *SCN3A*, and *SCN9A* genes. To test this hypothesis, we aligned intronic sequences downstream of exon 5b in all genes with annotated or predicted exon

5 duplications from human, mouse, sheep, chicken, frog, zebrafish, elephant shark, lancelet, and *Ciona*. In most species, the homologs of the human *SCN1A*, *SCN2A*, *SCN3A*, *SCN9A*, and *SCN5A* genes maintained IRE sequences, with the highest degree of conservation observed in amniotes (Figure 6A). The organization of these loci completely matched that of the human ones with the exception of chicken *SCN5A* and sheep *SCN1A*, both of which lost IRE2, and mouse *SCN1A*, which is lacking both duplication and IREs.

Notable patterns emerged in even more distant taxa. For example, in frogs, the gene cluster containing *SCN1A*, *SCN2A*, and *SCN3A* maintained exon 5 duplication and complete IRE sequences, while the *SCN5A* gene lacked both exon 5 duplication and IREs. In zebrafish, *SCN5A* showed an exon 5 duplication pattern homologous to that in humans but lacked IRE2. Among zebrafish *SCN1A* homologs, one maintained exon 5 duplication but lacked both IREs, while another had neither duplication nor conserved intronic regions. An interesting case was found in elephant shark (*C. milii*), where *SCN1A* homolog had both exon 5 duplication and IRE sequences, while other SCNα gene homologs in sharks lacked these features (Figure 6B). In lancelet and *Ciona*, exon duplications were found without IRE sequences.

These findings suggest that the elephant shark represents, perhaps, the most distant from the human living organism possessing both exon 5b duplication and IRE sequences capable of forming RNA structure, which potentially regulate alternative splicing of exons 5a/5b. These structural elements may have been present in the common ancestor of all vertebrates, whereas exon duplications appeared earlier in the common ancestor of all chordates. This evolutionary dynamics highlights the important role of complementary intronic sequences in regulating SCNα gene splicing throughout vertebrate history.

Discussion

Current evidence suggests that MXE splicing is crucial for shaping proteomic diversity and controlling transcript isoform expression in a tissue- and developmental stage-specific manner [46, 47]. It is a strictly coordinated mechanism, in which one and only one of two or more consecutive exons is included in the mature mRNA. In some cases, the mutually exclusive splicing pattern is characteristic to entire paralogous gene families, while in others each gene specifically conserves just one specific exon [48]. RNA structure is often associated with MXE regulation [49], but typically such regulatory mechanism is studied in just one paralog, e.g., in the *TPM2* gene in the tropomyosin gene family [50]. It is therefore natural to ask whether a group of paralogs possesses regulatory elements that was present in the ancestral gene and maintained by evolution after gene duplication. Here, we present a comprehensive analysis of splicing regulation in the SCNα gene family, shedding light on the evolution of mutually exclusive splicing and providing a framework for identifying conserved regulatory elements through paralogs comparison.

The intronic regulatory elements identified in this study are homologous across paralogs and exhibit multiple characteristic features of functional RNA structures. Their complementary interaction is supported by the combined evidence from proximity ligation sequencing data and compensatory substitutions. Similar to splicing regulatory elements in FGFR genes [23], IRE1 and IRE2 display evolutionarily conserved, non-random patterns, suggesting these regions were present in the common ancestor of SCNα genes. Furthermore, the disruption of IRE2 accessibility using antisense oligonucleotides altered the 5a/5b isoform ratio without the loss of mutual exclusivity, recapitulating observations in the ATE1 gene where RNA structure similarly

regulates the splicing outcome [45]. On the one hand, these findings support the hypothesis that IRE1 and IRE2 represent an evolutionarily ancient and functionally important mechanism for controlling the ratio of splice isoforms in the SCNα gene family through RNA structure-mediated regulation.

However, several factors challenge a straightforward interpretation of the IRE1 and IRE2 function in forming RNA structure. While these regulatory elements are complementary in most SCNα paralogs with exon 5 duplication, their complementarity in SCN1A is notably weaker due to sequence divergence of IRE1, a puzzling observation given high conservation of unpaired positions in IRE2. Unlike other known RNA structures with impact on splicing [51], the predicted IRE1/IRE2 interaction lacks extended continuous helices, instead forming shorter, more diffuse interactions. This leaves a possibility for other regulatory actors such as RNA-binding proteins or, considering known mechanisms of MXE splicing, competing secondary structures. Such alternative structures, however, remain undetected since no further stable base pairings could be identified in the gene sequences by in silico folding or RNA proximity ligation assays.

Comparative genomics is capable of revealing evolutionary trajectories of regulatory mechanisms in paralogs. While some gene families like dynamins show paralog-specific patterns, with RNA structure-mediated regulation of MXE splicing only in *DNM1* but not in *DNM2* [52], the SCNα family exhibits more ancient and conserved regulatory elements. Four of six SCNα paralogs share nearly the same ancestral sequences controlling 5a/5b exon inclusion, similar to the evolutionarily conserved mechanism observed in the FGFR genes [23]. However, unlike the well-characterized RNA structural regulation in FGFRs, our findings suggest the SCNα regula-

tory sequences may operate through both structural and non-structural mechanisms. While the conservation of regulatory elements across most SCNα paralogs points to their ancestral origin prior to gene duplication, their structural flexibility still may reflect adaptation to tissue-specific regulatory requirements.

Together our findings not only highlight the common mechanism of splicing regulation in the SCNα gene family, but they also demonstrate the power of comparative genomics in deciphering splicing regulatory mechanisms in paralogs. In most SCNα genes (with the exception of *SCN9A*), the complementarity between IRE1 and IRE2 cannot be found using single-sequence RNA folding tools such as MFOLD [53], while phylogenetic methods don't have enough statistical power due to extreme level of sequence conservation and absence of not only compensatory, but any variations at all. In contrast, paralogy-based approaches can underscore functional base-paired regions by contrasting sequence divergence with structural preservation. Future studies applying this strategy to other gene families may reveal similar hidden regulatory architectures. Moreover, experimental validation of possible protein-binding sites could clarify whether the observed IREs function autonomously or require trans-acting factors for splicing modulation.

Methods

Gene sequences and search for unannotated exon duplications

The nucleotide sequences and exon-intron boundaries of annotated transcript isoforms of 76 homologous SCNα genes across 13 different species were obtained from the Ensembl and

NCBI databases (Table S1). To identify unannotated SCNα genes in other organisms, tBlastN alignments were performed against the respective genomes using the annotated SCNα gene aminoacid sequence as a query. The MAFFT multiple alignment tool [54] was used to confirm the homology of the selected genes. Consecutive exons of the same gene with high aminoacid sequence identity (>30%) were classified as tandemly duplicated exons. In order to identify unannotated exon duplications, the translated sequence of each human exon (5a or 5b) was aligned against the intronic sequences flanking its homolog in the target species in all possible reading frames using tBlastN tool. Additionally, we aligned exons in the target species against their own flanking introns. To assess the significance of the predicted exon homology, we used the false positive rate (FPR) metric computed from realignments of the query sequence to 1000 random permutations of the target aminoacid sequence by counting the number of times when the random permutation had a higher alignment score than the actual alignment. Only exons with FPR below 5% are reported throughout this study.

The construction of phylogenetic trees

Multiple sequence alignments were constructed separately for two groups of amino acid sequences, the major protein isoform of each SCNα gene and for the translated sequences of annotated and newly discovered exons 5a and 5b using Muscle5 program with the default settings [55]. The major protein isoform was used to construct a phylogenetic tree. The optimal evolutionary model was selected using IQ-TREE ModelFinder, with the JTT+R8 model identified as most appropriate [56]. Phylogenetic trees were constructed in IQ-TREE using the selected evolutionary model. Branch support values were assessed through ultrafast bootstrap

analysis with 1000 replicates.

Identification and analysis of IRE

To identify IRE, we performed multiple sequence alignment of homologous introns flanking mutually exclusive exons. Blocks of homologous sequences of at least 10 nts in length were extracted from these alignments and subsequently analyzed using PrePH [40] and RNAup program from the Vienna RNA package [57] programs with default settings to predict complementary base pairings. For compensatory mutation analysis, we applied the R-scape v1.2.340 software [58] to blocks of multiple sequence alignments of 30 mammalian species after merging through a 10-nt spacer as explained earlier [40]. The E-value of the structure (E) was calculated as the product of E-values of its constituent base pairs that were provided by R-scape.

Percent-Spliced-In (PSI) metric

In order to characterize the exon 5a and 5b inclusion rates across human tissues and tumors, we used the Percent-Spliced-In (PSI, Ψ) metric computed from split read counts. For the GTEx dataset (8,555 samples), the TCGA dataset (731 tumor-normal pairs) and NMD inactivation RNA-seq experiments in A549 cell line [59], these counts were computed earlier [60, 59]. The poly(A)+ RNA-seq data from the study of mammalian organ development [35] and mouse embryonic transcriptomes [36] were downloaded in the BAM format and processed by the IPSA pipeline with the default settings [61]. The PSI metric was computed as the number of split reads supporting exon inclusion as a fraction of the collective number of split reads supporting exon inclusion and skipping.

Cell lines

The Huh7 cell line was used to determine the expression of splice isoforms with qPCR. The cell line was cultured in DMEM/F-12 (Paneco LLC, Russia) medium supplemented with 10% fetal bovine serum (INTL Kang, China), penicillin (50 IU/ml), and streptomycin (0.05 mg/ml) (Thermo Fisher Scientific, USA) at 37°C and 5% CO_2 in the atmosphere.

Antisense oligonucleotides (AON)

LNA/DNA mixmers (Syntol LLC, Russia) were used as antisense oligonucleotides to block complementary base pairs within RNA. AON had a phosphorothioated backbone to protect against cellular nucleases. The sequence of AON2 and the control sequence were +G*A*+T*T*+A*A*+A*C*+T*C*+A*A*+A and +T*G*+G*A*+A*G*+T*C*+T*T*+C*G*+T, respectively. In these sequences, G, A, T, and C denote DNA bases; +A and +G denote LNA bases; G*, A*, T*, and C* denote phosphorothioated DNA bases; +G*, +A*, +T*, and +C* denote phosphorothioated LNA bases.

AON transfection

Approximately 200,000 Huh7 cells were seeded into a 12-well plate and cultured for 24 hours. AON transfection into the cells was performed using Lipofectamine RNAiMAX (Invitrogen), at AON concentrations of 5, 25, and 100 nM. 3 µl of Lipofectamine and the appropriate concentration of LNA/DNA oligonucleotide were used per well, the lipid-RNA complex was dissolved in Opti-MEM Reduced Serum Medium (Gibco) and added to the cells according to the manufacturer's protocol. The cells were then incubated for 48 hours before harvesting.

RNA isolation

Total RNA was isolated using RUplus Column total RNA isolation kit (Biolabmix, Russia). One microgram of total RNA was first subjected to RNase-free DNase I digestion (Thermo Fisher Scientific, USA) at 37°C for 30 min to remove contaminating genomic DNA. Next, 500 ng of total RNA was used for complementary DNA (cDNA) synthesis using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Fisher Scientific, USA) to a final volume of 5 µl. cDNA was diluted 1:10 with nuclease-free water for quantitative PCR (qPCR) and reverse transcription PCR (RT-PCR) analysis.

RT-qPCR

RT-qPCR reactions were performed in triplicates in a final volume of 12 µl in 96-well plates with 420 nM gene-specific primers and 2 µl of cDNA using 5X qPCRmix-HS SYBR reaction mix (Evrogen, Russia). Primers for RT-qPCR are listed in Table S2. A sample without reverse transcriptase enzyme was included as control to verify the absence of genomic DNA contamination. Amplification of the targets was carried out on CFX96 Real-Time System (Bio-Rad, USA), with following parameters: 95°C for 2 min, followed by 42 cycles at 95°C for 30 s, 57°C for 30 s and 72°C for 30 s, ending at 72°C for 5 min. Gene and gene isoform expression change was calculated using an estimate of the amplification efficiency value.

Statistical tests

The data were analyzed using python version 3.8.2 and R statistics software version 3.6.3. Two-tailed t-test (with Bonferroni correction in the case of multiple comparisons) was performed if

no departure from population normality was observed. Non-parametric tests were performed using normal approximation with continuity correction. In all figures, the significance levels 5% and 0.1% are denoted by * and ***, respectively.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by Russian Science Foundation grant 21-64-00006.

Authors' contributions

EC and DP designed the study; EC performed data analysis; MV and DS performed the experiments; EC and DP wrote the first draft of the manuscript. All authors edited the final version of the manuscript.

Acknowledgments

The authors thank Prof. O.A. Dontsova for insightful discussions.

References

[1] Wang, J., Ou, S.-W., and Wang, Y.-J. (2017) Distribution and function of voltage-gated sodium channels in the nervous system. *Channels (Austin, Tex.),* **11**, 534–554.

[2] Stühmer, W., Conti, F., Suzuki, H., Wang, X. D., Noda, M., Yahagi, N., Kubo, H., and Numa, S. (1989) Structural parts involved in activation and inactivation of the sodium channel. *Nature,* **339**, 597–603.

[3] Catterall, W. A., Goldin, A. L., and Waxman, S. G. (2005) International Union of Pharmacology. XLVII. Nomenclature and structure-function relationships of voltage-gated sodium channels. *Pharmacological reviews,* **57**, 397–409.

[4] Alabi, A. A., Bahamonde, M. I., Jung, H. J., Kim, J. I., and Swartz, K. J. (2007) Portability of paddle motif function and pharmacology in voltage sensors. *Nature,* **450**, 370–5.

[5] Mantegazza, M., Cestèle, S., and Catterall, W. A. (2021) Sodium channelopathies of skeletal muscle and brain. *Physiological reviews,* **101**, 1633–1689.

[6] Heighway, J., Sedo, A., Garg, A., Eldershaw, L., Perreau, V., Berecki, G., Reid, C. A., Petrou, S., and Maljevic, S. (2022) Sodium channel expression and transcript variation in the developing brain of human, Rhesus monkey, and mouse. *Neurobiology of disease,* **164**, 105622.

[7] Liang, L., Fazel Darbandi, S., Pochareddy, S., Gulden, F. O., Gilson, M. C., Sheppard, B. K., Sahagun, A., An, J.-Y., Werling, D. M., Rubenstein, J. L. R., Sestan, N., Bender,

K. J., and Sanders, S. J. (2021) Developmental dynamics of voltage-gated sodium channel isoform expression in the human and mouse brain. *Genome medicine,* **13**, 135.

[8] Pang, P. D., Alsina, K. M., Cao, S., Koushik, A. B., Wehrens, X. H. T., and Cooper, T. A. (2018) CRISPR -Mediated Expression of the Fetal Scn5a Isoform in Adult Mice Causes Conduction Defects and Arrhythmias. *Journal of the American Heart Association,* **7**, e010393.

[9] Raymond, C. K., Castle, J., Garrett-Engele, P., Armour, C. D., Kan, Z., Tsinoremas, N., and Johnson, J. M. (2004) Expression of alternatively spliced sodium channel alpha-subunit genes. Unique splicing patterns are observed in dorsal root ganglia. *The Journal of biological chemistry,* **279**, 46234–41.

[10] Freyermuth, F., Rau, F., Kokunai, Y., Linke, T., Sellier, C., Nakamori, M., Kino, Y., Arandel, L., Jollet, A., Thibault, C., Philipps, M., Vicaire, S., Jost, B., Udd, B., Day, J. W., Duboc, D., Wahbi, K., Matsumura, T., Fujimura, H., Mochizuki, H., Deryckere, F., Kimura, T., Nukina, N., Ishiura, S., Lacroix, V., Campan-Fournier, A., Navratil, V., Chautard, E., Auboeuf, D., Horie, M., Imoto, K., Lee, K.-Y., Swanson, M. S., de Munain, A. L., Inada, S., Itoh, H., Nakazawa, K., Ashihara, T., Wang, E., Zimmer, T., Furling, D., Takahashi, M. P., and Charlet-Berguerand, N. (2016) Splicing misregulation of SCN5A contributes to cardiac-conduction delay and heart arrhythmia in myotonic dystrophy. *Nature communications,* **7**, 11067.

[11] Menezes, L. F. S., Sabiá Júnior, E. F., Tibery, D. V., Carneiro, L. D. A., and Schwartz, E. F. (2020) Epilepsy-Related Voltage-Gated Sodium Channelopathies: A Review. *Frontiers in*

pharmacology, **11**, 1276.

[12] Munyao, W., Rahman, M. M., Sabzanov, S. A., Chu, E. H., Wang, R., Wang, Z., Yu, Y., and Ruggiu, M. (2025) Alternative Splicing and CaV-Associated Channelopathies. *Wiley interdisciplinary reviews. RNA,* **16**, e70016.

[13] Jin, Y., Dong, H., Shi, Y., and Bian, L. (2018) Mutually exclusive alternative splicing of pre-mRNAs. *Wiley interdisciplinary reviews. RNA,* **9**, e1468.

[14] May, G. E., Olson, S., McManus, C. J., and Graveley, B. R. (2011) Competing RNA secondary structures are required for mutually exclusive splicing of the Dscam exon 6 cluster. *RNA (New York, N.Y.),* **17**, 222–9.

[15] Wang, X., Li, G., Yang, Y., Wang, W., Zhang, W., Pan, H., Zhang, P., Yue, Y., Lin, H., Liu, B., Bi, J., Shi, F., Mao, J., Meng, Y., Zhan, L., and Jin, Y. (2012) An RNA architectural locus control region involved in Dscam mutually exclusive splicing. *Nature communications,* **3**, 1255.

[16] Hong, W., Shi, Y., Xu, B., and Jin, Y. (2020) RNA secondary structures in ¡i¿Dscam1¡/i¿ mutually exclusive splicing: unique evolutionary signature from the midge. *RNA (New York, N.Y.),* **26**, 1086–1093.

[17] Hatje, K. and Kollmar, M. (2013) Expansion of the mutually exclusive spliced exome in Drosophila. *Nature communications,* **4**, 2460.

[18] Ivanov, T. M. and Pervouchine, D. D. (2018) An Evolutionary Mechanism for the Generation of Competing RNA Structures Associated with Mutually Exclusive Exons. *Genes,* **9**, 356.

[19] Ivanov, T. M. and Pervouchine, D. D. (2022) Tandem Exon Duplications Expanding the Alternative Splicing Repertoire. *Acta naturae,* **14**, 73–81.

[20] Parker, B. J., Moltke, I., Roth, A., Washietl, S., Wen, J., Kellis, M., Breaker, R., and Pedersen, J. S. (2011) New families of human regulatory RNA structures identified by comparative analysis of vertebrate genomes. *Genome research,* **21**, 1929–43.

[21] Soll, D. R. and Waddell, D. R. (1975) Morphogenesis in the slime mold Dictyostelium discoideum. 1. The accumulation and erasure of ”morphogenetic information”. *Developmental biology,* **47**, 292–302.

[22] Mistry, N., Harrington, W., Lasda, E., Wagner, E. J., and Garcia-Blanco, M. A. (2003) Of urchins and men: evolution of an alternative splicing unit in fibroblast growth factor receptor genes. *RNA (New York, N.Y.),* **9**, 209–17.

[23] Muh, S. J., Hovhannisyan, R. H., and Carstens, R. P. (2002) A Non-sequence-specific double-stranded RNA structural element regulates splicing of two mutually exclusive exons of fibroblast growth factor receptor 2 (FGFR2). *The Journal of biological chemistry,* **277**, 50143–54.

[24] Petrova, M., Margasyuk, S., Vorobeva, M., Skvortsov, D., Dontsova, O. A., and Pervouchine, D. D. (2024) BRD2 and BRD3 genes independently evolved RNA structures to control unproductive splicing. *NAR genomics and bioinformatics,* **6**, lqad113.

[25] Xu, L., Ding, X., Wang, T., Mou, S., Sun, H., and Hou, T. (2019) Voltage-gated sodium channels: structures, functions, and molecular modeling. *Drug discovery today,* **24**, 1389–1397.

[26] Gazina, E. V., Richards, K. L., Mokhtar, M. B. C., Thomas, E. A., Reid, C. A., and Petrou, S. (2010) Differential expression of exon 5 splice variants of sodium channel alpha subunit mRNAs in the developing mouse brain. *Neuroscience,* **166**, 195–200.

[27] de Lera Ruiz, M. and Kraus, R. L. (2015) Voltage-Gated Sodium Channels: Structure, Function, Pharmacology, and Clinical Indications. *Journal of medicinal chemistry,* **58**, 7093–118.

[28] Catterall, W. A. (2014) Structure and function of voltage-gated sodium channels at atomic resolution. *Experimental physiology,* **99**, 35–51.

[29] Zakon, H. H. (2012) Adaptive evolution of voltage-gated sodium channels: the first 800 million years. *Proceedings of the National Academy of Sciences of the United States of America,* **109 Suppl 1**, 10619–25.

[30] Liavas, A., Lignani, G., and Schorge, S. (2017) Conservation of alternative splicing in sodium channels reveals evolutionary focus on release from inactivation and structural insights into gating. *The Journal of physiology,* **595**, 5671–5685.

[31] Sanchez-Sandoval, A. L., Hernández-Plata, E., and Gomora, J. C. (2023) Voltage-gated sodium channels: from roles and mechanisms in the metastatic cell behavior to clinical potential as therapeutic targets. *Frontiers in pharmacology,* **14**, 1206136.

[32] Saavedra-Rodriguez, K., Maloof, F. V., Campbell, C. L., Garcia-Rejon, J., Lenhart, A., Penilla, P., Rodriguez, A., Sandoval, A. A., Flores, A. E., Ponce, G., Lozano, S., and Black, W. C. (2018) Parallel evolution of vgsc mutations at domains IS6, IIS6 and IIIS6 in pyrethroid resistant Aedes aegypti from Mexico. *Scientific reports,* **8**, 6747.

[33] Yuan, H., Shan, W., Zhang, Y., Yan, H., Li, Y., Zhou, Q., Dong, H., Tao, F., Liu, H., Leng, P., Peng, H., and Ma, Y. (2023) High frequency of Voltage-gated sodium channel (VGSC) gene mutations in Aedes albopictus (Diptera: Culicidae) suggest rapid insecticide resistance evolution in Shanghai, China. *PLoS neglected tropical diseases,* **17**, e0011399.

[34] Copley, R. R. (2004) Evolutionary convergence of alternative splicing in ion channels. *Trends in genetics : TIG,* **20**, 171–6.

[35] Mazin, P. V., Khaitovich, P., Cardoso-Moreira, M., and Kaessmann, H. (2021) Alternative splicing during mammalian organ development. *Nature genetics,* **53**, 925–934.

[36] He, P., Williams, B. A., Trout, D., Marinov, G. K., Amrhein, H., Berghella, L., Goh, S.-T., Plajzer-Frick, I., Afzal, V., Pennacchio, L. A., Dickel, D. E., Visel, A., Ren, B., Hardison, R. C., Zhang, Y., and Wold, B. J. (2020) The changing mouse embryo transcriptome at whole tissue and single-cell resolution. *Nature,* **583**, 760–767.

[37] Bahcheli, A. T., Min, H.-K., Bayati, M., Zhao, H., Fortuna, A., Dong, W., Dzneladze, I., Chan, J., Chen, X., Guevara-Hoyer, K., Dirks, P. B., Huang, X., and Reimand, J. (2024) Pan-cancer ion transport signature reveals functional regulators of glioblastoma aggression. *The EMBO journal,* **43**, 196–224.

[38] House, C. D., Vaske, C. J., Schwartz, A. M., Obias, V., Frank, B., Luu, T., Sarvazyan, N., Irby, R., Strausberg, R. L., Hales, T. G., Stuart, J. M., and Lee, N. H. (2010) Voltage-gated Na+ channel SCN5A is a key regulator of a gene transcriptional network that controls colon cancer invasion. *Cancer research,* **70**, 6957–67.

[39] Igci, Y. Z., Bozgeyik, E., Borazan, E., Pala, E., Suner, A., Ulasli, M., Gurses, S. A., Yumrutas, O., Balik, A. A., and Igci, M. (2015) Expression profiling of SCN8A and NDUFC2 genes in colorectal carcinoma. *Experimental oncology,* **37**, 77–80.

[40] Kalmykova, S., Kalinina, M., Denisov, S., Mironov, A., Skvortsov, D., Guigó, R., and Pervouchine, D. (2021) Conserved long-range base pairings are associated with pre-mRNA processing of human genes. *Nature communications,* **12**, 2300.

[41] Cai, Z., Cao, C., Ji, L., Ye, R., Wang, D., Xia, C., Wang, S., Du, Z., Hu, N., Yu, X., Chen, J., Wang, L., Yang, X., He, S., and Xue, Y. (2020) RIC-seq for global in situ profiling of RNA-RNA spatial interactions. *Nature,* **582**, 432–437.

[42] Lu, Z., Gong, J., and Zhang, Q. C. (2018) PARIS: Psoralen Analysis of RNA Interactions and Structures with High Throughput and Resolution. *Methods in molecular biology (Clifton, N.J.),* **1649**, 59–84.

[43] Wu, T., Cheng, A. Y., Zhang, Y., Xu, J., Wu, J., Wen, L., Li, X., Liu, B., Dou, X., Wang, P., Zhang, L., Fei, J., Li, J., Ouyang, Z., and He, C. (2024) KARR-seq reveals cellular higher-order RNA structures and RNA-RNA interactions. *Nature biotechnology,* **42**, 1909–1920.

[44] Jin, H., Zhang, C., Zwahlen, M., von Feilitzen, K., Karlsson, M., Shi, M., Yuan, M., Song, X., Li, X., Yang, H., Turkez, H., Fagerberg, L., Uhlén, M., and Mardinoglu, A. (2023)

Systematic transcriptional analysis of human cell lines for gene expression landscape and tumor representation. *Nature communications,* **14**, 5417.

[45] Kalinina, M., Skvortsov, D., Kalmykova, S., Ivanov, T., Dontsova, O., and Pervouchine, D. D. (2021) Multiple competing RNA structures dynamically control alternative splicing in the human ATE1 gene. *Nucleic acids research,* **49**, 479–490.

[46] Driscoll, S., Merkuri, F., Chain, F. J. J., and Fish, J. L. (2024) Splicing is dynamically regulated during limb development. *Scientific reports,* **14**, 19944.

[47] Tian, G. G., Li, J., and Wu, J. (2020) Alternative splicing signatures in preimplantation embryo development. *Cell & bioscience,* **10**, 33.

[48] Abascal, F., Tress, M. L., and Valencia, A. (2015) The evolutionary fate of alternatively spliced homologous exons after gene duplication. *Genome biology and evolution,* **7**, 1392–403.

[49] Yang, Y., Sun, F., Wang, X., Yue, Y., Wang, W., Zhang, W., Zhan, L., Tian, N., Shi, F., and Jin, Y. (2012) Conservation and regulation of alternative splicing by dynamic inter- and intra-intron base pairings in Lepidoptera 14-3-3ζ pre-mRNAs. *RNA biology,* **9**, 691–700.

[50] Clouet d'Orval, B., d'Aubenton Carafa, Y., Sirand-Pugnet, P., Gallego, M., Brody, E., and Marie, J. (1991) RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. *Science (New York, N.Y.),* **252**, 1823–8.

[51] Pervouchine, D. D. (2018) Towards Long-Range RNA Structure Prediction in Eukaryotic Genes. *Genes,* **9**, 302.

[52] Suyama, M. (2013) Mechanistic insights into mutually exclusive splicing in dynamin 1. *Bioinformatics (Oxford, England),* **29**, 2084–7.

[53] Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research,* **31**, 3406–15.

[54] Katoh, K. and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution,* **30**, 772–80.

[55] Edgar, R. C. (2022) Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nature communications,* **13**, 6968.

[56] Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., and Lanfear, R. (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular biology and evolution,* **37**, 1530–1534.

[57] Lorenz, R., Bernhart, S. H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011) ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB,* **6**, 26.

[58] Rivas, E., Clements, J., and Eddy, S. R. (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nature methods,* **14**, 45–48.

[59] Margasyuk, S., Kuznetsova, A., Zavileyskiy, L., Vlasenok, M., Skvortsov, D., and Pervouchine, D. D. (2024) Human introns contain conserved tissue-specific cryptic poison exons. *NAR genomics and bioinformatics,* **6**, lqae163.

[60] Mironov, A., Petrova, M., Margasyuk, S., Vlasenok, M., Mironov, A. A., Skvortsov, D., and Pervouchine, D. D. (2023) Tissue-specific regulation of gene expression via unproductive splicing. *Nucleic acids research,* **51**, 3055–3066.

[61] Pervouchine, D. D., Knowles, D. G., and Guigó, R. (2013) Intron-centric estimation of alternative splicing from RNA-seq data. *Bioinformatics (Oxford, England),* **29**, 273–4.

[62] Consortium, G. (2013) The Genotype-Tissue Expression (GTEx) project. *Nature genetics,* **45**, 580–5.

Figures and Tables

Species	Gene	Closest human homolog	Specie's exon 5 identity	Human exon 5 identity
Pika	SCN1A	SCN1A	56%	56%
Sheep	SCN2A	SCN2A	96%	96%
Frog	SCN2A	SCN2A	47%	50%
Zebrafish	scn1lab	SCN1A	90%	93%
Zebrafish	SCN8A	SCN8A	93%	97%
Zebrafish	SCN10A	SCN2A	80%	77%
Elephant shark	scn1lab	SCN1A	67%	65%
Lancelet	LOC136422135	SCN4A	96%	76%
Ciona	LOC100180733	SCN3A	50%	43%

Table 1: Unannotated exon duplication events in different species.

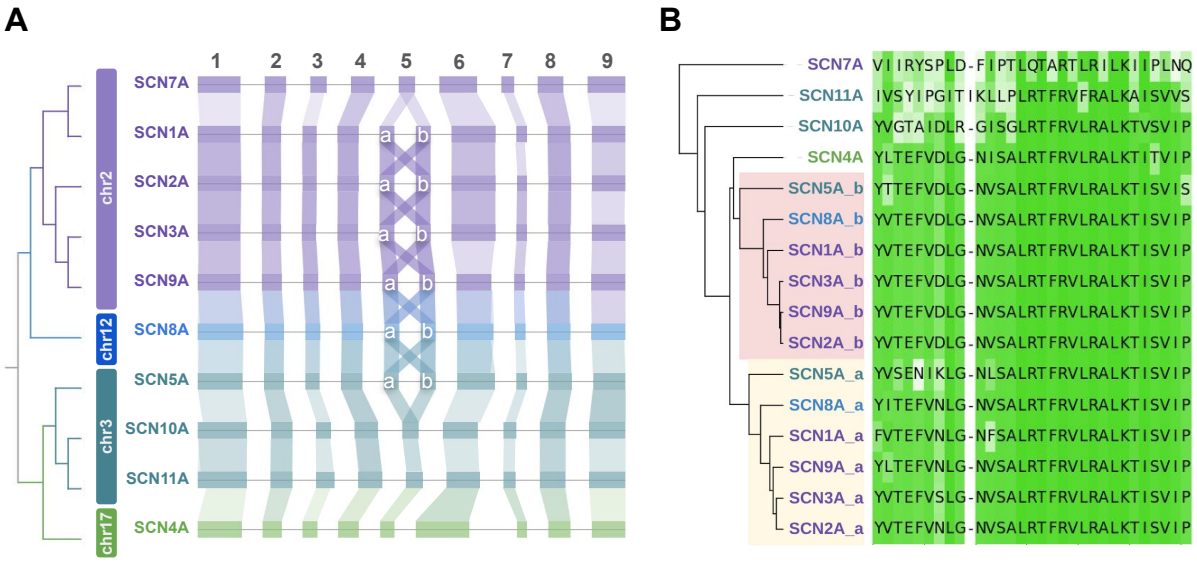

Figure 1: Transcript architecture in the human SCNα gene family **(A)** Exon-exon alignment of the human SCNα genes (exons 1-9). Duplications of exon 5 are marked as "a" and "b". (B) The alignment of aminoacid sequences of the human SCNα exons 5a and 5b.

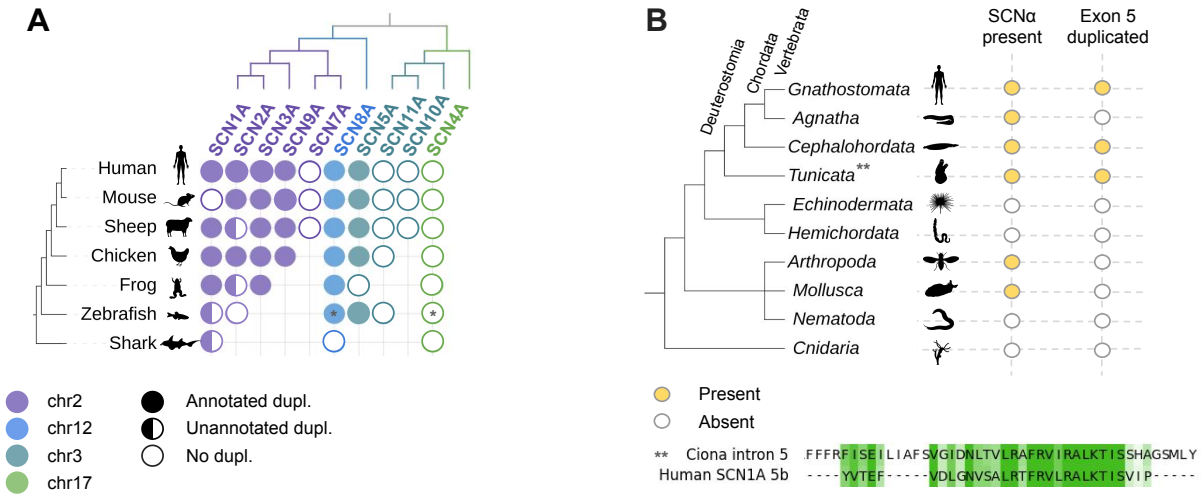

Figure 2: Exon 5 duplications in orthologs of human SCNα genes across species. **(A)** SCNα genes and exon 5 duplications in vertebrates. Asterisk indicates that zebrafish have an additional whole genome duplication, hence *SCN8A* and *SCN4A* orthologs have two duplicated genes, which are not shown here. **(B)** SCNα genes and their exon 5 duplications in animals. Bottom panel: Alignment of the *Cinav1a* gene intron 5 with human *SCN1A* exon 5b, indicating partial exon duplication in *Ciona*.

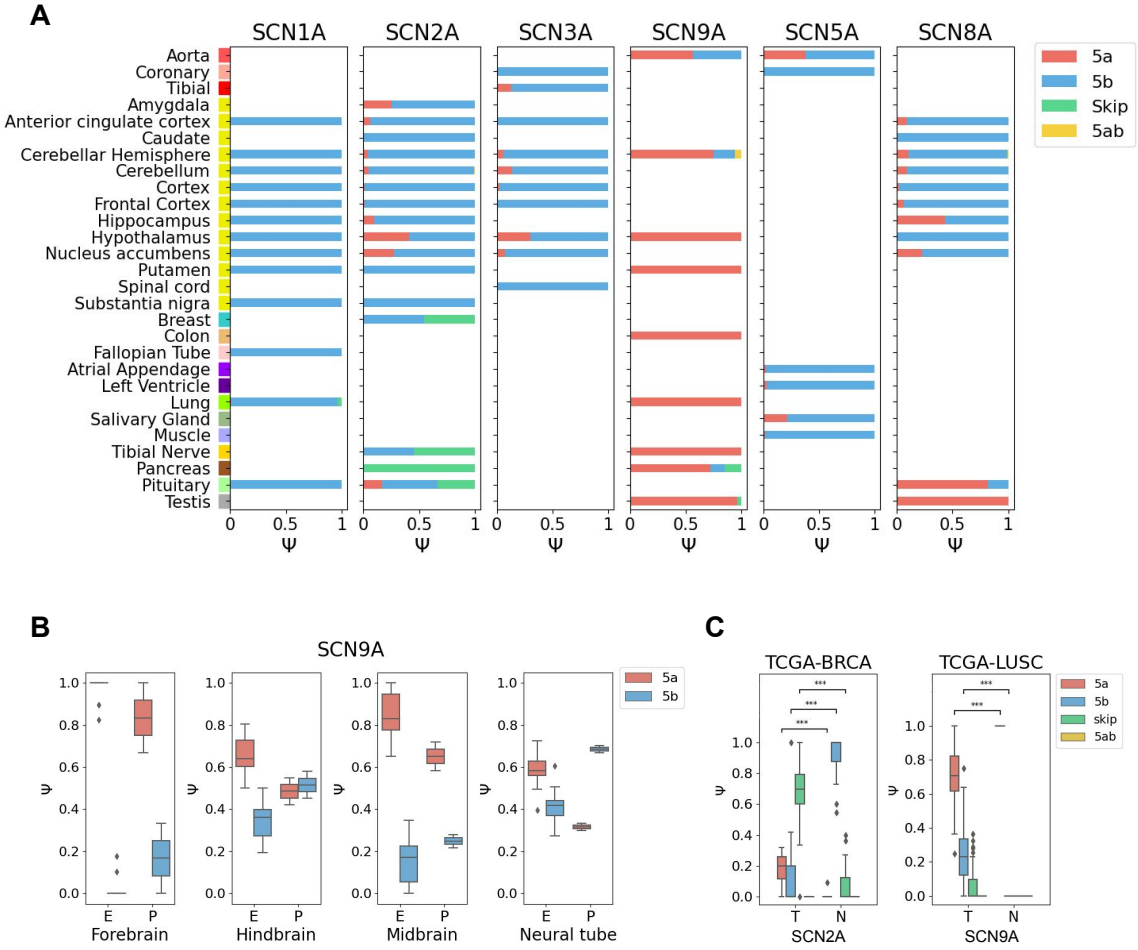

Figure 3: Splicing patterns in SCNα genes. **(A)** Splicing patterns in human tissues. The median Ψ value of each isoform (5a, 5b, 5a5b, and skip) is shown for each tissue. Tissue colors correspond to the standard GTEx color scheme [62]. **(B)** Developmental dynamics of *SCN9A* exons 5a and 5b in mouse. E — embryonic; P — postnatal. RNA-seq on postnatal stages were made for structures that develop from embryonic ones. **(C)** Cancer-specific splicing of exons 5a and 5b in the breast cancer (BRCA) and lung squamous cell carcinoma (LUSC) cohorts from TCGA. Statistically significant differences at the 0.1% significance level are denoted by *** (two-tailed t-test with Bonferroni correction).

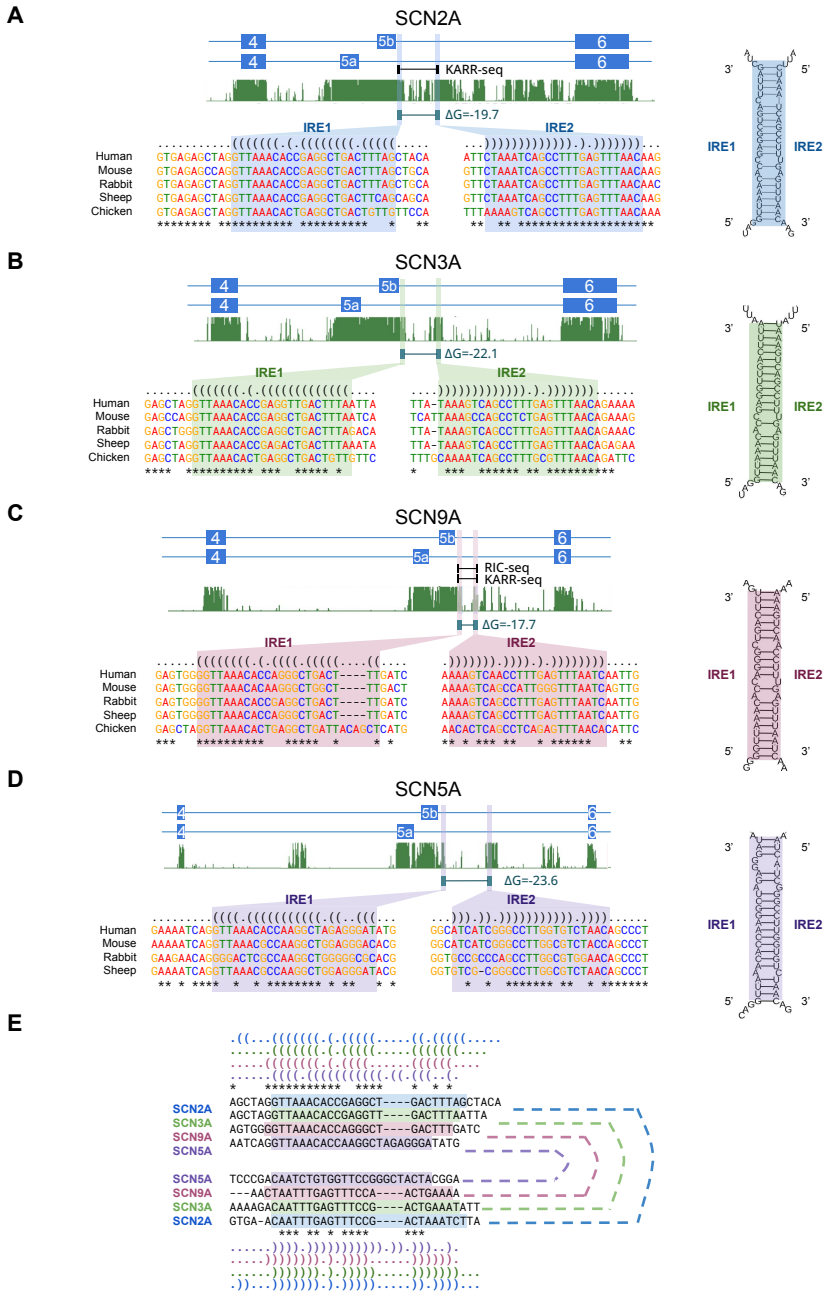

Figure 4: Intronic regulatory elements. **(A–D)** Left panels: genomic organisation of exons 4-6 in the human *SCN2A*, *SCN3A*, *SCN9A*, and *SCN5A* genes, the locations of IRE1 and IRE2, their free energy of hybridization (ΔG), support by RIC-seq or KARR-seq, 100 vertebrates conservation score by PhastCons (green track). Conserved nucleotides in multiple sequence alignments are indicated by asterisks. Right panels: RNA secondary structure formed by IRE1 and IRE2 binding. **(E)** Multiple sequence alignment of IRE1 and IRE2 sequences showing conserved nucleotides and RNA structures (dot-bracket representation).

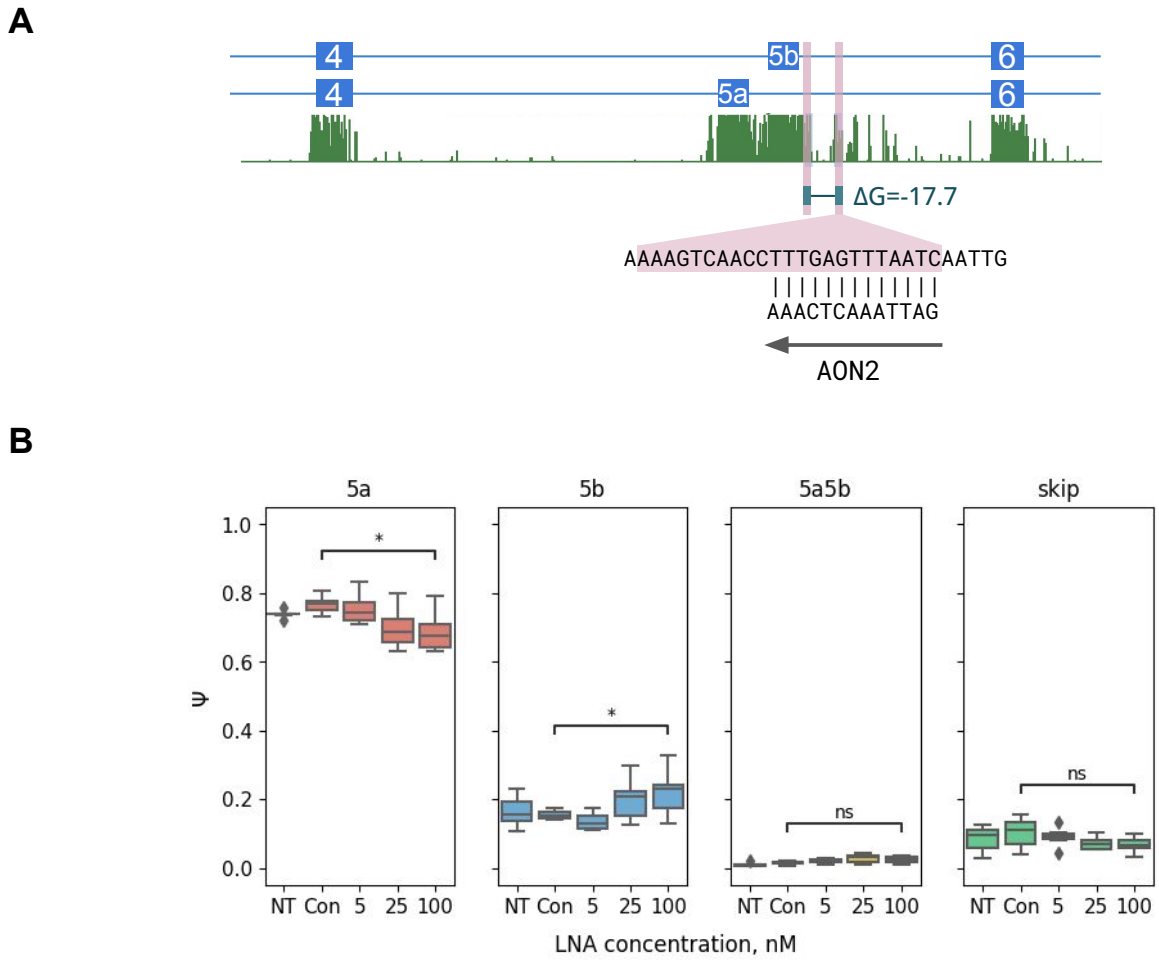

Figure 5: Experimental validation of the IRE2 impact on *SCN9A* splicing. **(A)** Genomic organisation of *SCN9A* exons 4–6, IRE2 location, sequence and its corresponding AON2. **(B)** Treatment with AON2 significantly reduces 5a inclusion and increases 5b inclusion. NT (nontreated control); Con (control AON). Asterisks (*) denote statistically significant differences at the 5% significance level; nonsignificant differences are labeled ”ns.”

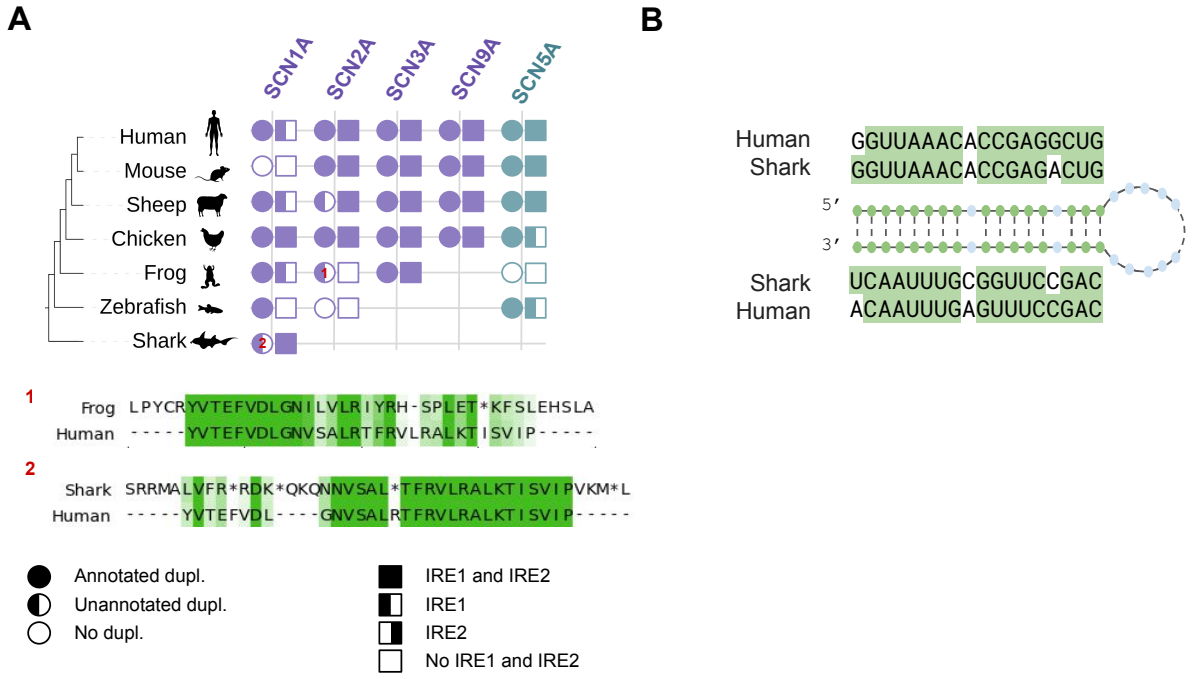

Figure 6: Common RNA structures in vertebrate SCNα genes. **(A)** SCNα genes, their exon 5 duplications and presence of IRE1/IRE2 sequences. **(B)** RNA structures formed by IRE1 and IRE2 are conserved between human and elephant shark.