

Exonic splicing code and coordination of divalent metals in proteins

Dara Bakhtiar¹, Katarina Vondraskova², Reuben J. Pengelly¹, Martin Chivers¹, Jana Kralovicova^{1,2} and Igor Vorechovsky^{1,*}

¹University of Southampton, Faculty of Medicine, Southampton SO16 6YD, UK

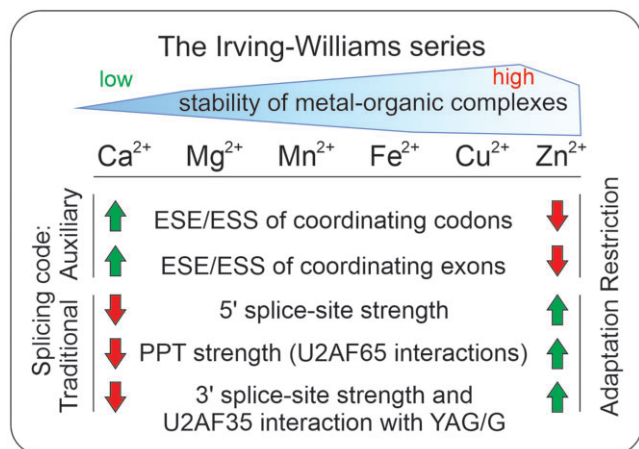
²Slovak Academy of Sciences, Centre of Biosciences, 840 05 Bratislava, Slovak Republic

*To whom correspondence should be addressed. Tel: +44 2381 206425; Fax: +44 2381 204264; Email: igvo@soton.ac.uk

Abstract

Exonic sequences contain both protein-coding and RNA splicing information but the interplay of the protein and splicing code is complex and poorly understood. Here, we have studied traditional and auxiliary splicing codes of human exons that encode residues coordinating two essential divalent metals at the opposite ends of the Irving–Williams series, a universal order of relative stabilities of metal–organic complexes. We show that exons encoding Zn²⁺-coordinating amino acids are supported much less by the auxiliary splicing motifs than exons coordinating Ca²⁺. The handicap of the former is compensated by stronger splice sites and uridine-richer polypyrimidine tracts, except for position –3 relative to 3' splice junctions. However, both Ca²⁺ and Zn²⁺ exons exhibit close-to-constitutive splicing in multiple tissues, consistent with their critical importance for metalloprotein function and a relatively small fraction of expendable, alternatively spliced exons. These results indicate that constraints imposed by metal coordination spheres on RNA splicing have been efficiently overcome by the plasticity of exon–intron architecture to ensure adequate metalloprotein expression.

Graphical abstract



Introduction

Metal-binding complements of eukaryotic proteomes, or metalloproteins, require metal ions to assist catalysis or impart structure (1). Metalloproteins account for a large proportion of gene products in many organisms (2–5). For example, the number of human proteins predicted to bind Ca²⁺, Mg²⁺ or Zn²⁺, the most abundant divalent ions in the human body, was estimated at ~10⁵ each (6), while almost half of the enzymes structurally characterized in the Protein Data Bank (PDB) need metals (1,4). Because proteins are flexible, steric selection of metals is imperfect, especially by nascent polypeptides first emerging from the ribosome (7). Under such conditions the relative affinities of divalent metals for proteins

have a tendency to follow ligand field stabilization energies of metals themselves, creating a universal order of binding preferences known as the Irving–Williams series (7–9), first proposed for a subset of transition metals 75 years ago (10). For essential divalent metals abundant in humans [ppm > 50 (9)], this order is Ca²⁺ ~ Mg²⁺ < Mn²⁺ < Fe²⁺ < Cu²⁺ ≥ Zn²⁺, with Ca²⁺ and Mg²⁺ forming the weakest and Cu²⁺ and Zn²⁺ forming the tightest complexes with organic ligands (1,5,7,8,11). In equimolar mixtures of these and other divalent metals in the series (Co²⁺ and Ni²⁺), proteins that require weaker binding ions preferentially bind tighter metals. Although metalloproteins can bind incorrect metals *in vitro*, metalation *in vivo* is usually accurate (12,13), implying that cells evolved

Received: June 16, 2023. Revised: November 15, 2023. Editorial Decision: November 16, 2023. Accepted: November 17, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

efficient strategies to overcome or reduce these constraints and avoid binding by tight non-cognate metals (mismetallation) to safeguard selective metal-binding properties of their polymers (7,14,15). Harmful mismetalation events can distort geometry of cognate metal-binding sites, recruit undesired ligands or exploit only a subset of native ligands (1). Hence, it is essential to understand how protein-coding genes and their products manage metal acquisition to avoid mismetalation. However, the underlying mechanisms and relative contributions of eukaryotic gene expression steps to strategies counteracting mismetalation and metal coordination constraints are poorly understood.

Eukaryotic gene expression involves a series of highly coordinated steps, starting from transcription and splicing of mRNA precursors (pre-mRNAs), followed by mRNA export, translation and decay (16). These steps require multi-component cellular machines that are tethered to each other (16). In humans, the most complex step is believed to be pre-mRNA splicing, which removes intervening sequences or introns and joins coding sequences or exons together with single-nucleotide precision. This process is orchestrated by the spliceosome, a 2.7 MDa RNA–protein complex consisting of small nuclear RNAs U1, U2 and U4–U6, and ~125 associated proteins in spliceosomal assembly intermediates (17). The spliceosome forms *ad hoc* on each intron to recognize 3' and 5' splice sites (3'ss and 5'ss), lariat branchpoints and polypyrimidine tracts (PPTs) (17). Although conserved, these traditional or core signals are insufficient for accurate splicing, which requires additional pre-mRNA motifs in exons and introns, particularly in higher vertebrates and plants (18). These auxiliary motifs have become known as exonic or intronic splicing enhancers or silencers (ESEs/ESSs and ISEs/ISSs) (19–23). Because most exons are protein coding, and ESEs/ESSs and amino acids are encoded by the same nucleotides, ESE/ESS evolution had to be shaped by both splicing and protein-coding restrictions (24,25). Indeed, a large proportion of the splicing information in ESEs, recently estimated at ~50%, is coincident with the protein-coding information (26) and can be characteristic of a protein domain (27). Many mutations that disrupt protein function were shown to alter pre-mRNA splicing (24). Although the exact relationship of the two selective forces remains obscure, the high number of metal-binding residues in the human proteome provides an opportunity to better understand the impact of metal coordination restrictions on the auxiliary splicing code in exons.

The ESE/ESS profiles have been recently linked to amino acid frequencies at binding sites for divalent metals (28). The weakest binders of the Irving–Williams series (Ca^{2+} , Mg^{2+}) were shown to preferentially bind residues encoded by splice-enhancing codons, whereas the tight binders (Cu^{2+} , Zn^{2+}) are coordinated to amino acids encoded by splice-repressing codons, with moderate binders in the series (Mn^{2+} , Fe^{2+}) exhibiting intermediate codon-specific ESE/ESS values (28) (Figure 1A). This splicing dichotomy suggested that protein-binding sites for weak divalent metals may have been promoted during evolution at the exon level, whereas sites for competitive metals may have been repressed, potentially reducing mismetalation. Notably, codons for residues that coordinate calcium in Ca^{2+} -binding proteins (CaBPs) have the capacity to promote exon inclusion in mature transcripts significantly more than the average (27). The high potential of codons for Ca^{2+} -coordinating residues to be retained in mature transcripts may have facilitated the expansion of

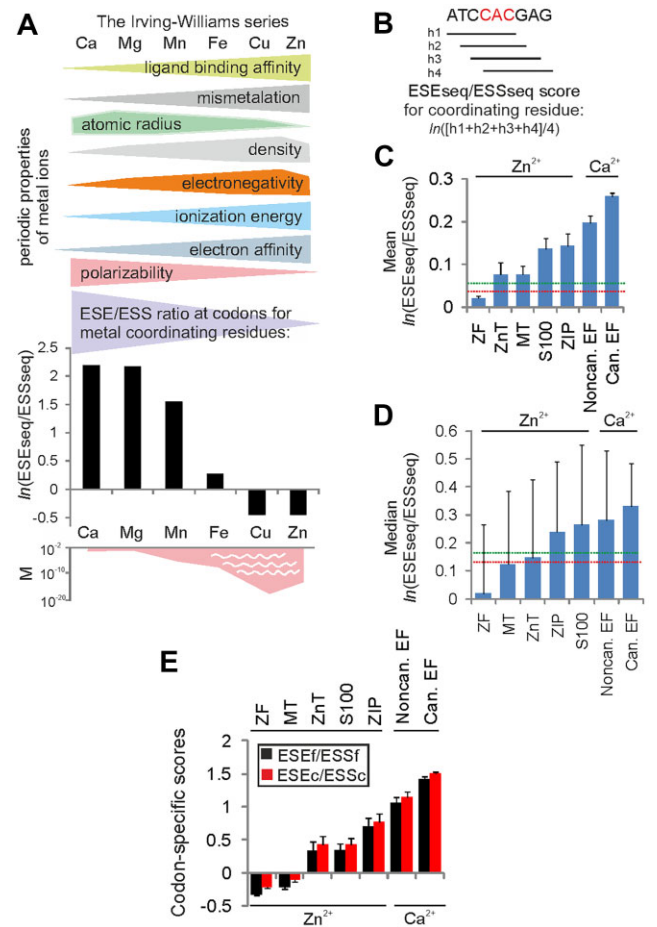


Figure 1. ESE/ESS profiles of codons for Zn^{2+} - and Ca^{2+} -coordinating residues in human proteins. **(A)** The Irving–Williams series and the auxiliary splicing code in exons: a hypothesis. The upper panel shows selected properties of six abundant and biologically important divalent metals in the human body and their non-linear trends in the Periodic Table of Elements. Atomic radius, the distance from atomic nucleus to outermost electron orbital; electronegativity, a tendency of an atom in a molecule to attract shared electrons; ionization energy, the amount of energy required to remove the first electron from neutral atoms; electron affinity, the energy released when an electron is added; polarizability, a tendency to acquire electric dipole moment in electric fields (133). The middle panel shows ESE/ESS profiles for codons encoding coordinating residues for the indicated metal [adapted from (28)]. The ESEseq/ESSseq values were weighted by amino acid frequencies at metal-binding sites estimated by fragment transformation methods (89). The bottom panel shows estimates of metal concentrations (mol/l) in ancient (sulfidic and anoxic) sea waters (>4000 million years ago). Concentrations were compiled from (11,134,135). **(B)** Average ESEseq/ESSseq score calculations across codons for Zn^{2+} - and Ca^{2+} -binding sites (red). ESEseq and ESSseq scores (22) for overlapping hexamers (h, horizontal bars) were averaged for each coordinating codon as indicated. Splicing-neutral hexamers (22) were ignored. **(C and D)** Average **(C)** and median **(D)** values of ESEseq/ESSseq ratios across metal-coordinating codons in major groups of Zn^{2+} - and Ca^{2+} -BPs: ZF, zinc fingers; ZnT, Zn^{2+} transporters; MT, metallothioneins; ZIP, Zrt/Irt-like proteins; EF, non-canonical and canonical EF-hand proteins. Error bars for means **(C)** are standard errors of the mean (SEMs), error bars for medians **(D)** are interquartile ranges. Horizontal dotted lines denote values for various control exon groups, with a maximum value denoted in green and minimum in red. Significant differences between groups were seen with both parametric [single-factor analysis of variance (ANOVA) followed by the Tukey–Kramer post-hoc test, F -value = 116.2, $P < 10^{-16}$] and non-parametric (Kruskal–Wallis one-way ANOVA, H statistics = 651.3, $P < 0.00001$) tests. **(E)** Codon-specific ESE/ESS scores of the same protein groups. Columns denote means, error bars are SEMs.

Ca²⁺-binding sites such as EF-hands during evolution (27). However, auxiliary splicing motifs in bona fide exons have not been studied for other metalloproteins, and it remains unclear if they have been influenced by constraints of the Irving–Williams series and periodic trends of divalent metal properties (Figure 1A).

In this study, we have characterized ESE/ESS profiles for sequences that encode protein-binding sites for Zn²⁺, a tight binder in the Irving–Williams series (1,7,8,11). As predicted (28), we demonstrate that codons for Zn²⁺-coordinating residues in human proteins confer a lower capacity to retain these coding sequences in mature transcripts than codons that coordinate weak Ca²⁺. We also show that entire exons encoding Zn²⁺-coordinating residues have significantly lower ESE/ESS ratios than exons encoding Ca²⁺-coordinating residues. However, both groups of exons show a relative scarcity of alternative splicing as compared with average exons, which was supported by RNA sequencing (RNA-seq) and expressed sequence tag (EST) datasets, and by polymerase chain reaction (PCR)-based validation of exon candidates for regulated splicing. We also demonstrate that the diminished inclusion capacity of Zn²⁺-coordinating codons and their exons to be included in mRNAs was compensated by the increased intrinsic strength of their splice sites, PPTs and the remaining ESE/ESS within their exons. In this respect, we have explored the role of pyrimidine transitions at position –3 relative to 3′ss, and provide evidence for their splicing preferences, supporting distinct interactions between each subunit of the auxiliary factor of U2 (U2AF) and exons encoding Zn²⁺- versus Ca²⁺-binding residues. Together, these results show how traditional and auxiliary splicing motifs in the pre-mRNA evolved to alleviate constraints imposed by coordination of essential divalent metals, providing new insights into the interplay of protein and splicing codes in human DNA.

Materials and Methods

Extraction and validation of exons that code for Zn²⁺-binding sites in proteins

A non-redundant set of DNA sequences encoding Zn²⁺-coordinating residues was compiled using the following resources: (i) PDB; (ii) UniprotID (29); (iii) ZincBind (30), (iv) the Database of Metal Binding Sites (31); and (v) GeneCards (<https://www.genecards.org>). We prepared two main datasets, termed stringent and extended (Table 1; Supplementary Datasets S1–S4). The stringent dataset contained a total of 3862 Zn²⁺-coordinating residues in ~990 Zn²⁺-binding sites in 437 mammalian proteins with structural evidence for Zn²⁺ binding, as determined by nuclear magnetic resonance (NMR) spectroscopy or X-ray crystallography (average resolution 2.3 Å). The dataset also included structures investigated by electron microscopy (SLC30A8, USP39, RPL37, COX5B, POLR3K, POLR2I, RPS29, POLR2L, POLR2B, RPAP2, POLR3B, FBXO5 and POLR2K). Most structures were derived from human proteins, with just ~19% from non-human mammals (Supplementary Table S1). To expand the sample size further while capitalizing on a high conservation of ESE/ESS motifs in higher vertebrates (18), we also compiled the extended dataset, which contained residues predicted to bind Zn²⁺ that were inferred from multiple sequence alignments of conserved groups of Zn²⁺-binding proteins (ZnBPs) (Table 1; Supplementary Figures S1–S4). In this dataset, amino

Table 1. Datasets of codons that encode Zn²⁺-coordinating residues

Supplementary Online Dataset	Extended through homology in multiple alignments	CMM filtered	No. of codons ^a	No. of proteins
S1 (stringent, without CMM filter)	No	No	3862	437
S2 (stringent, CMM filtered)	No	Yes	3348	395
S3 (extended, without CMM filter)	Yes	No	4465	482
S4 (extended, CMM filtered)	Yes	Yes	3951	440

^aResidues annotated in PDB as Zn²⁺ binding but not metal coordinating were not included.

Table 2. ZnBPs included in the study

ZnBP group	No. of proteins	No. of Zn ²⁺ -coordinating codons	Zn ²⁺ -coordinating codons in the vicinity of splice sites ^a
ZnTs	10	142	3
ZIPs	14	128	4
MTs	11	220	11
S100	20	193	0
ZFs and other ZnBPs	427	3782	149
Total	482	4465	167

^a< 3 nt from exon–intron junctions.

acids were considered Zn²⁺ coordinating if their alignment positions in other species shared identical residues with human proteins. If the alignment showed distinct residues at this position, we included only residues that were previously shown to coordinate Zn²⁺, as exemplified by substitutions: H113 > E and H115 > S in KLK2; H183 > Q in NLN; D20 > G in OMP; or H231 > S or E236 > L in KLK3. Other residues at shared alignment positions were not considered in subsequent analyses, such as D20 > G in OMP or E236 > L in KLK3. As compared with the stringent dataset, the extended dataset contained sequences encompassing additional 603 Zn²⁺-coordinating amino acids (Table 1).

Because metal-containing structures in PDB may contain incorrect metal assignments (32), PDB identifiers with evidence for Zn²⁺ binding were subjected to validation. We employed CheckMyMetal (CMM) (33), a sophisticated algorithm, to categorize Zn²⁺-binding sites into three groups (termed plausible, problematic and not validated; coloured in green, orange and red in Supplementary Datasets S1–S4). This classification was based on six validation parameters, including valence, geometry, ligand and vacancy (32). The CMM validation was carried out for both stringent and extended datasets to create filtered datasets, each without low resolution CMM adjustments (Table 1). Residues not validated as metal-binding sites by CMM for at least two validation parameters were excluded from ESE/ESS profiling (Table 1).

Finally, amino acids included in the four datasets and their flanking sequences were matched to human coding sequences in Ensembl (<http://www.ensembl.org>, build 104) to obtain non-redundant Zn²⁺-coordinating codons and exons

for ESE/ESS profiling, totalling 482 different ZnBPs (Supplementary Datasets S1–S4).

Characterization of ESEs/ESSs that encode Zn²⁺-coordinating residues in proteins

Spliceosomes of higher vertebrates recognize small exons in the sea of very long introns in a process known as exon definition (34). Their choice to include (exons) or exclude (introns) a pre-mRNA sequence in or from mature transcripts is strongly influenced by the balance of ESEs and ESSs, which exhibit a gradient in exon–intron definition on a continuous scale of exon inclusion capacity (19,20,22,35). To characterize the auxiliary splicing code underlying Zn²⁺-binding sites in proteins, we employed a comprehensive set of ESE and ESS hexamers that were previously derived by *ex vivo* splicing promotion or repression afforded by 4096 synthetic oligomers inserted into two model exons at five different positions (22). The resulting duplicate minigene libraries were used to obtain hexamer ESEseq and ESSseq scores, which were calculated independently of metal binding affinities and which provide good estimates of exon inclusion activities (22).

To assign ESEseq and ESSseq scores to codons for residues coordinating Zn²⁺, we computed average scores of four overlapping hexamers using custom Microsoft Excel functions/formulas (Figure 1B; Supplementary Figure S5). We applied the same procedure to each of the four datasets and also to codons encoding Ca²⁺-coordinating residues, which were ascertained previously (27). Apart from ESEseq/ESSseq scores, we calculated frequency ratios for a total of 4728 ESE codons and 4360 ESS codons, shown here as ln(ESEf/ESSf) (Supplementary Dataset S5). We also determined codon counts in 1182 high-confidence ESEs and 1090 high-confidence ESSs (22) to compute the ESEc/ESSc ratios. The two measures estimate codon-specific splicing activities, perhaps barring stop codons since these translation termination signals may induce capricious nonsense-mediated mRNA decay in transient transfections of plasmid DNA (36), although this was considered unlikely (22). For control datasets, we extracted RefSeq sequences of human protein-coding exons as defined by the UCSC Table Browser (<https://genome.ucsc.edu/cgi-bin/hgTables>), comprising ~35 million hexamers in ~200 000 exonic segments, as described (27). Mean ESEseq/ESSseq, ESEf/ESSf and ESEc/ESSc values were computed for exons devoid of the first and the last three nucleotides (nt) since these exonic positions shape the 3′ss and 5′ss consensus, respectively.

mRNA inclusion levels of exons that encode Zn²⁺- and Ca²⁺-coordinating residues

To determine PSI (percent spliced in) values (37) for our tested exons, we employed PSI tables (hg38) from the Vertebrate Alternative Splicing and Transcription Database (VastDB) (38). VastDB provides a comprehensive PSI resource across vertebrate exons in various tissues and developmental stages (38). In addition, we compared PSI values of tested and control exons using EST data in HEXEvent, a database of Human EXon splicing Events, which stores EST-derived inclusion levels for ~200 000 exons (39). HEXEvent shows alternative splicing information for human internal exons but, unlike VastDB, HEXEvent does not include intron retention events, avoiding a bias toward short retained introns (25,39).

Comparisons of traditional splicing signals

To determine the intrinsic strength of exon splice sites, we used maximum entropy scores defined previously (40), which successfully predict aberrant 3′ss and 5′ss induced by human pathogenic mutations (41,42). To compare the strength and location of lariat branchpoints of tested exons in ZnBPs and CaBPs, we employed BPP (43) and SVM-BPfinder (44), currently the best performing algorithms for *ab initio* branchpoint prediction that consider PPTs and AG dinucleotide exclusion zones (43,45). We ignored branchpoints outside the range between –13 and –60 nt relative to the 3′ss, which contains the vast majority of human branchpoints (44,46,47) but not distant branchpoints, which are rare (28,48). We used WebLogo (49) to display relative nucleotide frequencies of predicted branchpoints with the highest branchpoint scores in the two groups of exons and in controls. Location of branchpoint motifs in intronic sequence is shown in Supplementary Dataset S6. The dataset also compares BPP and SVM-BPfinder assignments with 59 359 high-confidence branchpoints identified by Mercer *et al.* (47) and identifies non-canonical introns, including AT–AC introns, whose branchpoint motifs may be recognized by U12 (50), and an AT–AA intron (51). The fraction of U12-type introns (52,53) was 0.2% in ZnBP exons and 3.5% in CaBP exons (Supplementary Dataset S6). Control introns were obtained from the UCSC browser using curated mRNA transcripts from NCBI RefSeq, totalling 195 903 unique entries (hg38).

Validation of alternative splicing of exons encoding Zn²⁺- and Ca²⁺-binding residues

Selection of candidate exons for experimental validation and regulated splicing was based on average VastDB PSI values lower than 90% in neural or muscle tissue, testis or embryonal stem cells. In addition, total RNA was prepared from exponentially growing cell lines HEK293 (54) and SH-SY5Y [ATCC CRL-2266 (55)] that were derived from human embryonal kidney and human neuroblastoma, respectively. In addition, we used total RNA isolated from 16 human tissues, including brain and skeletal muscles (Ambion, cat. # 0912010 and 0905009). Total RNA samples were reverse transcribed with the Moloney murine leukaemia virus reverse transcriptase (SuperScript III RT, Invitrogen) and oligo d(T) primers according to the manufacturer's recommendation. Reverse transcription–PCR (RT–PCR) primers (Supplementary Table S2) were designed to target exons adjacent to tested exons that had average PSI values below the indicated limit. PCRs were carried out with DreamTaq Green PCR Master Mix (Thermo Scientific™) with 1.5 mM Mg²⁺ and at two annealing temperatures. PCR products were separated on 1.5% agarose gels. Signal intensities of the spliced products were measured using the Amersham Imager 600 (GE Healthcare) to estimate exon inclusion levels.

Ex vivo splicing assays

Splicing reporters containing pyrimidine transitions at position –3 of 3′ss were first tested for three human genes: *F8* (56), *UBE2F* (57) and *HGD* (58). The reporters were selected from our minigene library owing to the presence of at least four contiguous pyrimidines at PPT, uridine or cytosine at position –3, and the presence of spliced products with and without the middle exon upon transient transfections. Plasmid mutagenesis was carried out using overlap extension PCR

with primers shown in [Supplementary Table S3](#) to obtain constructs with weaker and stronger PPTs and weaker and stronger exon positions +1 through +3 in the presence of CAG or TAG 3' ss. Mutated constructs were Sanger-sequenced to exclude undesired mutations (Eurofins). The minigene collection was then extended to include 19 additional constructs to test disease-associated single-nucleotide substitutions at position -3 reported in the literature and 24 additional reporters with mid-exons encoding coordinating residues for Zn²⁺ or Ca²⁺ ([Supplementary Tables S4 and S5](#)).

Plasmid DNA samples were transiently transfected into HEK293 or Chinese hamster ovary (CHO) cells. HEK293 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) bovine calf serum (Biosera) under standard conditions. CHO cells were grown in Ham's F12 supplemented with 2 mM glutamine and 10% bovine calf serum. Transfections were carried out in 12- or 24-well plates using 150 ng of reporter DNA and jetPRIME (Polyplus) according to the manufacturer's recommendations. The cells were lysed 24 h later for RNA extraction. Total RNA was isolated using TRI Reagent (Molecular Research Center) according to the manufacturer's protocol and used for the first-strand cDNA synthesis with the Moloney murine leukaemia virus reverse transcriptase (Promega) and oligo d(T) primers. RT-PCRs were carried out using vector-specific primers (59) to obtain spliced products from exogenous transcripts. PCR products were separated on agarose gels and the amount of exon inclusion and skipping was measured with the Amersham Imager 600 (GE Healthcare).

Results

Auxiliary splicing code in exons and protein-binding sites for Zn²⁺

To test the hypothesis shown in [Figure 1A](#) and to determine the extent to which codons that encode Zn²⁺-coordinating residues support exon inclusion in mature transcripts, we extracted and validated their nucleotide sequences from human proteomic, structure and genomic databases (see the Materials and Methods). To compare codons for tight (Zn²⁺) and weak (Ca²⁺) metal-binding sites using the same method ([Figure 1B](#)), we assigned hexamer ESEseq and ESSseq scores (22) to codons encoding Zn²⁺-coordinating residues ([Supplementary Datasets S1–S4](#)) and codons for Ca²⁺-coordinating residues obtained previously (27). [Figure 1C and D](#) shows that codons for Zn²⁺-coordinating residues have, in aggregate, a significantly lower ESEseq/ESSseq ratio than codons encoding Ca²⁺-coordinating residues, here exemplified by EF-hand proteins. The most reduced values were found for zinc fingers (ZFs), a copious group of human ZnBPs in which Zn²⁺ largely binds Cys or His residues (60). In contrast, ZnBPs where Zn²⁺ is increasingly coordinated by acidic amino acids showed higher predicted exon inclusion levels ([Figure 1C, D](#)). These ZnBPs included Zn²⁺ transporter proteins, which move Zn²⁺ out of (ZnTs) and into (ZIPs, Zrt/Irt-like proteins) the cell, acting in opposite directions and selectively binding Zn²⁺ to stabilize its cytosolic concentration (61,62), metallothioneins (MTs) and S100 proteins, which often bind divalent metals other than Zn²⁺, including Ca²⁺ (63) ([Table 2](#)). The exonization potential of Zn²⁺-coordinating codons in ZnBPs showed a hierarchy of ZIPs/S100 > ZnTs/MTs > ZFs ([Figure 1C, D](#)). This hierarchy did not change when removing Zn²⁺-coordinating codons lo-

cated at the start or end of exons, which constituted ~3.5% of the total sample size. Their replacement with extended native motifs yielded very similar values. As compared with the extended dataset, the stringent dataset devoid of ZIPs and many ZnTs, MTs and S100 proteins showed only a small reduction in mean and median values, consistent with a higher fraction of core Zn²⁺-coordinating residues (Cys and His) in this sample. Codon-specific scores computed for the same datasets confirmed the same hierarchy of ZnBPs and CaBPs ([Figure 1E](#)). Again, replacement of terminal triplets with extended native motifs or their removal did not significantly alter these values, or the order of the indicated protein groups.

Finally, to investigate the extent to which potentially incorrect assignment in a proportion of metal-containing structures in PDB can influence the hierarchy of ESE/ESS profiling, we removed PDB structures that could not be validated by CMM (32) from our datasets ([Table 1](#)). However, the hierarchy of the five groups of ZnBPs remained the same in the CMM filtered data, each group showing significantly lower ESEseq/ESSseq ratios than EF-hand proteins.

We conclude that codons coding for amino acids coordinating the two divalent metals in human proteins have the capacity to confer a significantly lower exon inclusion for tight Zn²⁺ than for weak Ca²⁺.

mRNA inclusion of exons that encode Ca²⁺- and Zn²⁺-coordinating residues

Next, we set out to test if the observed ESE/ESS dichotomy of codons for Zn²⁺- and Ca²⁺-coordinating amino acids can influence inclusion levels of entire exons. We disregarded exons >350 nt long to reduce a bias resulting from terminal or large exons where the fraction of coordinating codons was negligible. We also pooled ZnBPs other than ZFs into a single group to obtain a sufficient sample size. [Figure 2A](#) shows that exons that encode Ca²⁺-coordinating residues (termed Ca²⁺ exons) retained higher than average ESEseq/ESSseq scores as compared with those encoding Zn²⁺-coordinating residues (termed Zn²⁺ exons). Unlike the splice-inhibiting effect of codons for Zn²⁺-coordinating amino acids in ZFs ([Figure 1C–E](#)), the ESEseq/ESSseq values for entire ZF Zn²⁺ exons were not lower than the average but were somewhat elevated ([Figure 2A](#)). This suggests that the low exon inclusion capacity of codons encoding Zn²⁺-coordinating residues has been compensated by the auxiliary splicing code in the remaining portions of the same exons.

Next, we determined average PSI values for Zn²⁺ and Ca²⁺ exons and control exons by exploring existing, publicly available RNA-seq and EST data. Importantly, we found that average PSI values derived from RNA-seq experiments compiled in VastDB (38) were significantly higher for exons encoding residues that coordinate either metal than for controls ([Figure 2B](#)). The mean PSI values remained high across multiple tissues ([Supplementary Figure S6](#)). Because very small exons are generally less efficiently included in mature transcripts (25,34,64), we excluded small exons and microexons from VastDB PSI tables, but the difference in PSI values ([Figure 2B](#)) remained highly significant.

To further support these findings, we compared inclusion levels for the three exon groups and controls using independent EST data. We explored HEXEvent (39), which defines four types of EST-derived exon inclusion [defined as inclLevel, constitLevel, 3usageLevel and 5usageLevel in (39)],

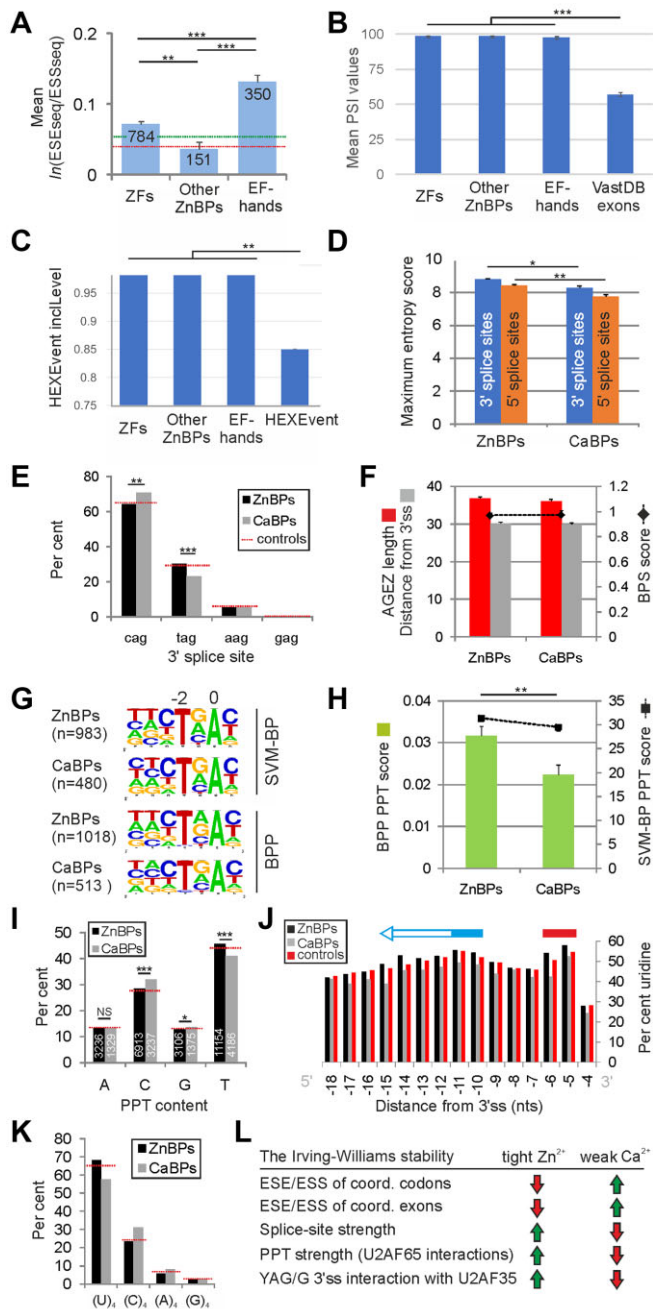


Figure 2. High inclusion of weak Zn^{2+} exons in mature transcripts is supported by both traditional and auxiliary splicing signals. **(A)** Mean $\ln(\text{ESEseq}/\text{ESSseq})$ values for full exons that encode Zn^{2+} - or Ca^{2+} -binding sites in ZFs, other ZnBPs and canonical EF-hands. Significant differences between groups were confirmed by ANOVA (F -value 44.3, $P < 0.0001$). The number of exons in each group is shown at the top of each column. Error bars are SEMs. $**P < 0.001$, $***P < 0.0001$. Horizontal dotted lines show the range for various control exon groups, demarcated by maximum (green) and minimum (red) values. **(B)** Mean PSI values in the same groups of exons. Human VastDB exons were used as controls. Error bars are SEMs. **(C)** Average EST-derived exon inclusion levels in the same groups. InclLevel values were defined previously (39). HEXEvent controls show mean inclLevel values for 227 179 human exons in the HEXEvent database. Error bars are SEMs. Additional EST-derived exon inclusion events are shown in Supplementary Figure S7. **(D)** Zn^{2+} exons have stronger splice sites than Ca^{2+} exons. The intrinsic strength of splice sites was determined by averaging maximum entropy scores (40). Error bars are SEMs. $*P = 0.05$, $**P < 0.001$ (Mann–Whitney U-test). **(E)** Frequency of (C/T/A/G)AG 3' ss

extending information from cassette exons to their major alternative splice sites. This comparison confirmed that for each exon inclusion category, Zn^{2+} and Ca^{2+} exons had higher average inclusion values than controls (Figure 2C; Supplementary Figure S7).

Together, these results reveal a close-to-constitutive inclusion and relative paucity of alternative splicing for exons that are critical for binding sites for the two arguably most important divalent metals in eukaryotic cell signalling.

Balancing auxiliary and traditional splicing motifs by exons encoding Zn^{2+} - and Ca^{2+} -coordinating residues

Next, we set out to identify pre-mRNA features that overcome the low capacity of Zn^{2+} -coordinating codons (Figure 1C–E) and Zn^{2+} exons (Figure 2A) to be spliced into mature transcripts in such a way that these exons are largely constitutive *in vivo* (Figure 2B, C). Comparison of the intrinsic strength of splice sites showed that Zn^{2+} exons in a pooled sample of all ZnBPs had, on average, significantly stronger 5' ss and 3' ss than Ca^{2+} exons (Figure 2D). We also observed an excess of CAG 3' ss in the latter group, largely at the expense of TAG 3' ss, while control introns showed intermediate values (Figure 2E). At the first exon position, Ca^{2+} exons had significantly fewer guanines and more adenines than Zn^{2+} exons ($\chi^2 = 16.1$, $P < 0.0001$; Supplementary Figure S8).

Because the difference in the intrinsic splice site strength between Zn^{2+} and Ca^{2+} exons appeared smaller for 3' ss than for 5' ss (Figure 2D), and T/C mutations at position -3 relative to 3' ss may alter branchpoint selection (65), we tested if an increased strength of lariat branchpoints and/or PPTs can compensate the reduced ESE/ESS profiles of Zn^{2+} exons. However, introns preceding Zn^{2+} and Ca^{2+} exons differed neither in average distances between the strongest branchpoints and 3' ss, nor in the length of AG dinucleotide exclusion zones (Figure 2F) or PPTs (Supplementary Figure S9). Moreover, there was no significant departure from the branchpoint consensus (Figure 2G; Supplementary Dataset S5). In contrast, we found significantly higher PPT scores for Zn^{2+} exons as compared with their Ca^{2+} counterparts (Figure 2H), which was corroborated by a clear shift to uridine-rich PPTs in the former group (Figure 2I), consistent with the maximum entropy scoring of partial PPTs (20 nt preceding the 3' ss, Figure 2D). The excess of uridine in PPTs of Zn^{2+} exons over Ca^{2+} exons could be seen at each position between -4 and -18 nt relative to the 3' ss,

in ZnBPs and CaBPs. χ^2 for the 2×4 table = 12.4. $**P < 0.001$, $***P < 0.0001$. **(F)** Average length of the AG dinucleotide exclusion zone and distance from 3' ss (left axis) for branchpoints with the highest SVM-BP scores (right axis) in introns that precede Zn^{2+} and Ca^{2+} exons. **(G)** Consensus branchpoint sequence for the two exon groups, as predicted by SVM-BP (44) and BPP (43). The number of introns is in parentheses. **(H)** Introns that precede Zn^{2+} exons have stronger PPTs than those preceding Ca^{2+} exons. Error bars are SEMs. $**P < 0.001$ (median test, both with SVM-BP and BPP). **(I)** Zn^{2+} exons are preceded by uridine-rich PPTs as compared with Ca^{2+} exons. The total number of nucleotides between the branchpoint and 3' ss is in white. $***P < 0.0001$, $**P < 0.001$ (χ^2 test). **(J)** Uridine content at positions -4 to -18 relative to 3' ss. Horizontal red and blue bars denote putative maxima of U2AF65 RRM1 and RRM2 interactions with the PPT. **(K)** Tetranucleotides in PPTs of introns preceding Zn^{2+} and Ca^{2+} exons (2×4 table $\chi^2 = 11.2$, $P < 0.0001$). **(L)** Summary of compensatory trends in auxiliary and traditional splicing signals of Zn^{2+} and Ca^{2+} exons.

with the first peak at position -5 and a second, broader peak extending toward branchpoints from position -11 in controls to position -15 upstream (blue arrow in Figure 2J).

The two uridine peaks at PPTs are thought to reflect interactions with RNA-recognition motifs RRM1 and RRM2 of the large U2AF subunit (U2AF65) (66, and references therein) and possibly with other RNA-binding proteins. We therefore compared the frequencies of $(U)_4$, an optimal binding platform for U2AF65 *in vitro* and *in vivo* (67, and references therein), in the two groups of PPTs downstream of predicted branchpoints. We found that introns preceding Zn^{2+} exons had a significantly higher fraction of PPTs with at least one $(U)_4$ motif than those preceding Ca^{2+} exons (47.5% versus 37.4%, $\chi^2 = 8.4$, $P < 0.0001$), largely at the expense of $(C)_4$ repeats (Figure 2K). In addition, distribution of distances between these motifs and 3'ss showed noticeable shifts at positions -6 through -15 for $(U)_4$ and -6 through -25 for $(C)_4$ (Supplementary Figure S10A, B). As compared with control introns, the fraction of PPTs with at least one $(U)_4$ between BPP branchpoints and 3'ss was lower for Ca^{2+} exons and slightly higher for Zn^{2+} exons, whereas the opposite was observed for $(C)_4$ motifs (Supplementary Figure S10C).

Finally, we also examined the overlap between Zn^{2+} and Ca^{2+} exons and those differentially expressed in cells depleted of the small subunit of U2AF (U2AF35) (54), and identified multiple genes and exons in common, exemplified by Ca^{2+} -binding *SDF4* or *MCFD2* or Zn^{2+} -binding *ZFAND1*. Figure 2L and the Graphical abstract summarize changes in traditional and auxiliary splicing signals for Zn^{2+} and Ca^{2+} exons and predicted U2AF interactions.

Taken together, these results suggest that core splicing signals for Zn^{2+} and Ca^{2+} exons evolved to compensate their metal-specific ESE/ESS signals. They also support distinct 3'ss interactions of the two exon groups with U2AF early during spliceosome assembly.

Alternative splicing of exons encoding coordinating residues for Ca^{2+} and Zn^{2+}

Although alternative splicing of Zn^{2+} and Ca^{2+} exons is reduced compared with the average (Figure 2B, C), we detected a small subset of exons with low mean PSI values. To identify those that may be regulated, we tested 38 exons that had average VastDB PSI values $< 90\%$ in at least one of four tissues important for Ca^{2+}/Zn^{2+} signalling. We designed RT-PCR primers (Supplementary Table S2) in exons adjacent to Zn^{2+} and Ca^{2+} exons and determined their mRNA inclusion in human neuroblastoma and embryonal kidney cell lines, and in human brain and skeletal muscle samples. The screening revealed exon skipping (red triangles in Figure 3A) for at least one transcriptome in 18 cases (47%; Figure 3B). Alternative splicing of these exons appeared to be more frequent in neural and muscle tissues than in the remaining tissues of the 16-tissue Ambion RNA panel. Only six exons revealed significant skipping in all four cell types. Alternative splicing of these exons was detected in other organisms for some but not all homologues or tissues (see example for the rat in Figure 3C and D; Supplementary Table S2), consistent with reduced transcriptomic diversity in less complex eukaryotes. Figure 3E shows examples of AlphaFold-predicted structures for two validated human exons in Zn^{2+} transporters, both with exposed alternatively spliced segments on the surface.

Together, these results identify exons encoding residues coordinating Zn^{2+} and Ca^{2+} that may be regulated by late-evolving alternative splicing events in a tissue- or developmental stage-specific manner.

Selection of pyrimidines at position -3 relative to 3'ss of Zn^{2+} and Ca^{2+} exons

Although the poorer ESE/ESS profiles of Zn^{2+} exons as compared with Ca^{2+} exons were compensated by traditional splicing motifs (Figure 2D, H, I–L), there was one exception. We were puzzled by the excess of CAG 3'ss of Ca^{2+} exons over TAG 3'ss (Figure 2E). Human CAG 3'ss are more abundant than TAG 3'ss by a factor of > 2 (2.23 in our sample of introns) (68), but it is unclear why they are preferred. The two 3'ss are not functionally equivalent (65). Published case reports suggest that mutation $-3C>T$ can induce exon skipping (65,69,70), promote exon inclusion (71,72), activate cryptic 3'ss (73) or remain silent (74–76). In line with these observations, mutation $-3C>T$ induced opposite shifts in open and closed conformations of U2AF65 in different RNA contexts (66). However, binding of short CAG 3'ss RNAs to the wild-type yeast orthologue of U2AF35, which stabilizes U2AF65 PPT interactions (77,78, and references therein), was weaker as compared with UAG 3'ss RNAs, irrespective of the nucleotide identity at the first exon position (78). Can the observed difference in uridine and cytosine content at position -3 of Zn^{2+} and Ca^{2+} exons still be compensatory?

To address the question, we first compiled sequences of 38 published cases (Supplementary Table S5) and 39 unpublished ClinVar (79) records (Supplementary Table S6) of human $-3C/T$ variants. In the latter collection, substitutions $-3C>T$ were more prevalent than substitutions $-3T>C$ (Supplementary Table S6, 32 versus 7, respectively, $P < 0.0001$, binomial test). This might be due to a possible ascertainment bias toward more severe phenotypes or increased mutability of cytosines by deamination (80) and a progressive loss of cytosines at this position over time, but the resulting TAG might still provide a superior 3'ss in some cases. However, the bias was not observed for published phenotype-associated cases (Supplementary Table S5, 20 versus 18, respectively, $P > 0.01$) where the link between phenotype and mutation is tighter, further suggesting that substitutions $-3T>C$ could create less efficient 3'ss in some transcripts.

To test this experimentally and to better understand the function of CAG and TAG 3'ss, we first explored our minigene library to select three informative reporters that produced skipping and inclusion of their mid-exons. We also weakened their PPTs in the presence of CAG or TAG 3'ss (Figure 4A; Supplementary Figure S11). Upon transient transfections into HEK293 or CHO cells, mutation $-3T>C$ promoted exon inclusion in each transcript. In a PUF60-sensitive *UBE2F*, which has a very long PPT, we were able to see the distinct outcome of the two alleles only upon co-expression with mutated PUF60 (Figure 4A), possibly due to weakening 3'ss through reduced PUF60 binding to RNA or impaired balance between PUF60 and other splicing factors (57,99). Uridine-richer PPTs partially or fully restored the splicing defects, but they did not alter the direction of $C>T$ mutations (Figure 4A). In *HGD*, weaker PPT was associated with the activation of upstream cryptic 3'ss, consistent with susceptibility of the *HGD*

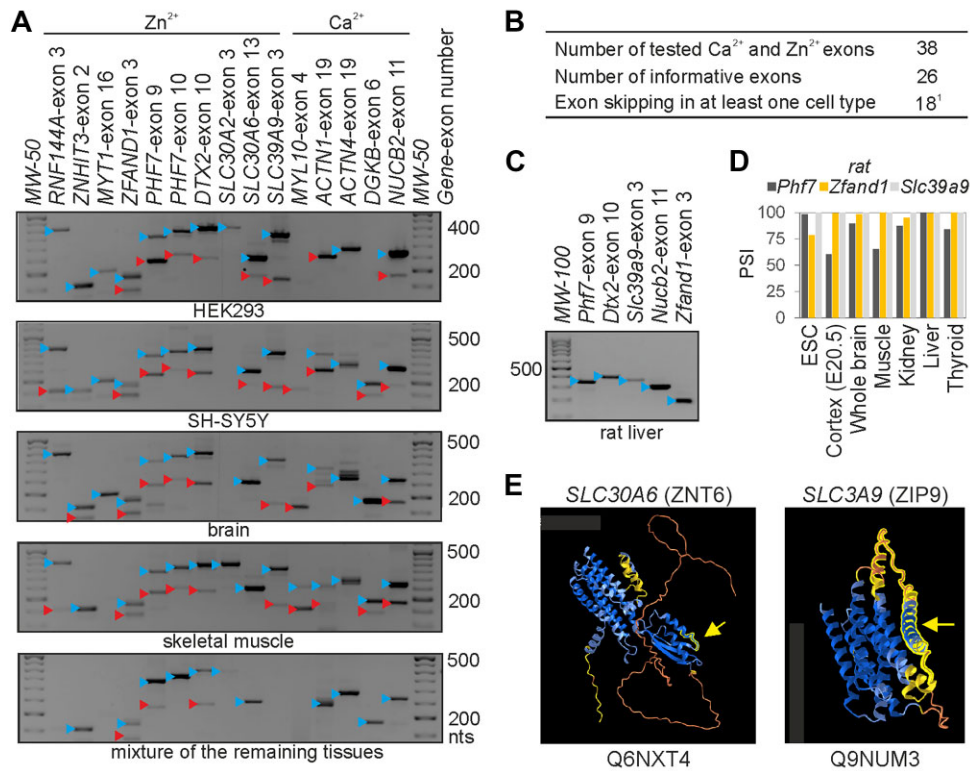


Figure 3. Validation of candidate Zn²⁺ or Ca²⁺ exons for regulated splicing. **(A)** Representative gel electrophoresis showing inclusion (blue triangles) or skipping (red triangles) of tested exons in the indicated cell lines and tissues. Sizes of spliced products (in nt) and primer sequences are given in [Supplementary Table S2](#). The bottom panel shows splicing in a mixture of RNAs from adipose, bladder, cervix, colon, oesophagus, kidney, lung, ovary, pancreas, placenta, prostate, small intestine, testes and thymus. **(B)** Summary of validated alternatively spliced exons. ¹Exons showing skipping in at least one tissue are identified in [Supplementary Table S2](#). **(C)** Examples of a lack of exon skipping of rat orthologues. **(D)** VastDB PSI values for two rat exon homologues in a subset of rat tissues. **(E)** AlphaFold-predicted (136) structures of an efflux and influx Zn²⁺ transporter with alternatively spliced Zn²⁺ exons. Arrows point to accessible peptides (wrapped in yellow) that are absent in isoforms lacking the alternatively spliced exon.

mid-exon to skipping in the presence of mutations in adjacent intronic sequences (58). The exon was fully skipped only in HEK293, but not in CHO cells (Figure 4A).

Positions within splice site consensus motifs are not independent (81). In the reference human genome, the -3T allele at 3'ss is positively associated with exon positions +1G, +2A, +3A, +4T, +5T and +6A (Figure 4B; [Supplementary Figure S12](#)). To test how the associated exon haplotypes influence the outcome of -3C/T variants and if exon mutations could reverse the observed differences in TAG and CAG 3'ss usage, we mutated minigenes *F8* and *HGD* at the first three exon positions in the presence of -3C or -3T (Figure 4C). In *F8*, the TAG 3'ss yielded stronger exon inclusion in the presence of the associated haplotype +G +2A +3A than the wild-type +1A +2T +3T or the CAG-associated haplotype +1C +2G +3C. Improved exon inclusion for TAG 3'ss in the presence of haplotype +1G +2A +3A was also found for *HGD*.

To expand the number of tested 3'ss, we then randomly selected a subset of disease-associated (DA) mutations -3C>T or -3T>C ([Supplementary Table S5](#)) and transcripts with Zn²⁺/Ca²⁺ exons ([Supplementary Table S4](#)). We cloned the tested exons and their flanking intronic sequences between two *U2AF1* exons to create hybrid reporters. We mutated the wild-type reporters to prepare an additional 18 DA and 24 Zn²⁺/Ca²⁺ minigene pairs that differed only by pyrimidine at position -3 (Figure 4D, E). Upon sequence validation and

transfections into HEK293 cells, we found that of 11 informative DA exons, CAG 3'ss gave higher exon inclusion than TAG 3'ss in 10 transcripts whereas only one TAG 3'ss was clearly superior over CAG (binomial test, $P = 0.005$, Figure 4D). For *SMN1* haplotypes, mutation -3C>T at 3'ss of exon 7 produced exon skipping irrespective of exon 7 T/C variants at position +6, but the *SMN1*-specific exon 7 allele (+6C) partially rescued this defect (Figure 4F). We did not observe the bias for Zn²⁺/Ca²⁺ exons (Figure 4D, E, G) but the fraction of informative cases was much lower than for DA exons (29% versus 59%), consistent with their high PSI values in the native context (Figure 2B, C). Overall, counting all informative minigene pairs ($n = 22$), there was a significant excess of -3T over -3C among mutations leading to exon skipping (18 versus 4, $P = 0.002$, binomial test, Figure 4G).

Taken together, although a more abundant CAG is not always a 3'ss of choice for the human spliceosome, CAG 3'ss are preferred as they usually confer higher exon inclusion than TAG 3'ss. This information is fully encoded by minigene mid-exons and their native intronic flanks. Figure 4G also suggests that decisions as to which pyrimidine at position -3 activates or represses 3'ss are likely to be influenced by the auxiliary splicing code in exons, including the first exon position(s). We therefore suggest that the excess of CAG 3'ss in Ca²⁺ exons (Figure 2E) may still compensate the weaker exon position +1 in this exon group ([Supplementary Figure S8](#)), ensuring optimal U2AF35 binding to 3'ss YAG/R motifs.

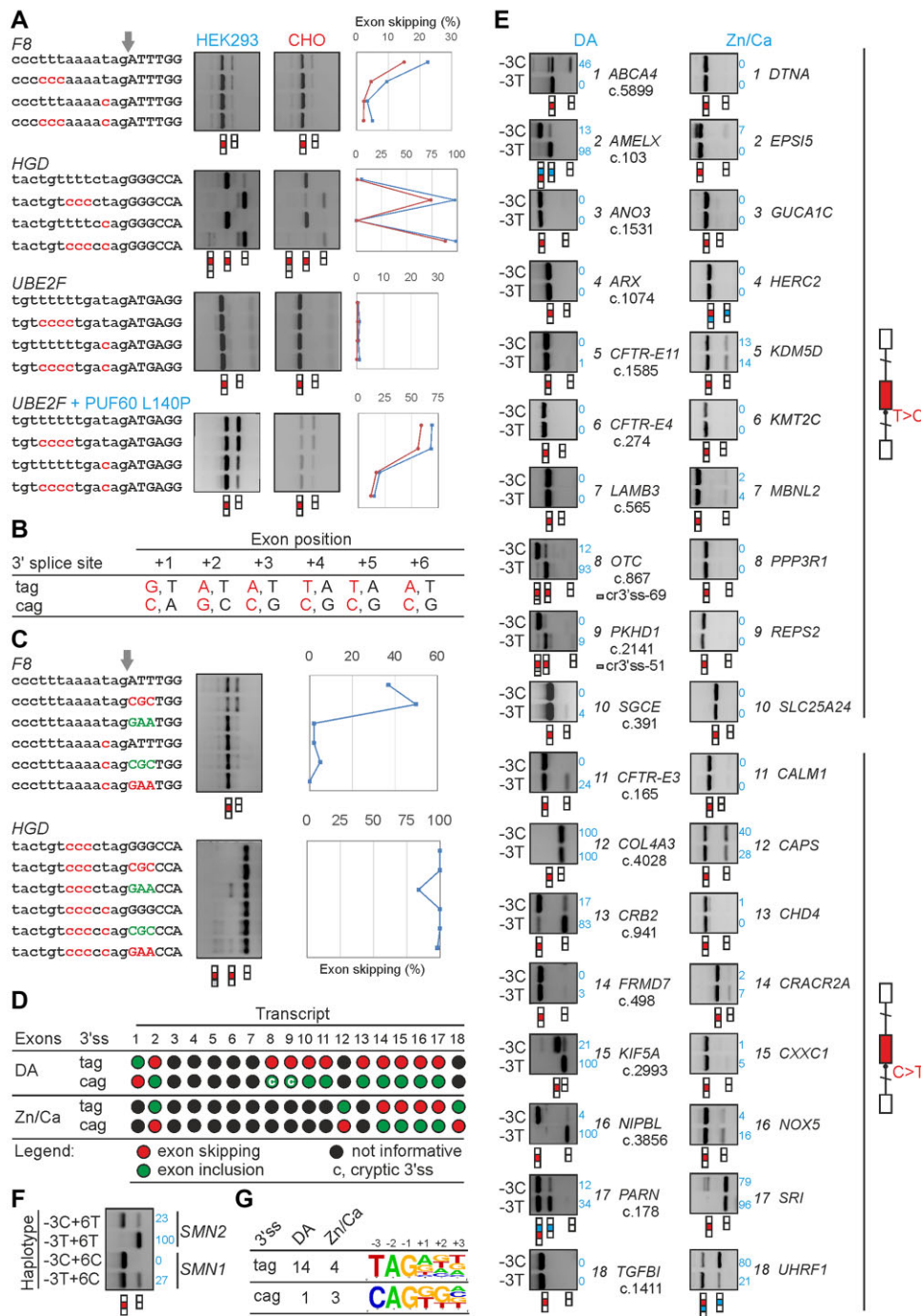


Figure 4. The human spliceosome prefers CAG 3' splice sites over TAG 3' splice sites. **(A)** Splicing outcome of mutations -3T>C and -3C>T in three minigene constructs in two cell lines. Arrow denotes 3' splice site. Mutations are in red. Spliced products are shown at the bottom; tested mid-exons are red boxes. *HGD* produced cryptic 3' splice sites (grey box) 176 nt upstream (58) when the PPT was weakened. Differential skipping of *UBE2F* mid-exons was unmasked by co-expressing the reporter plasmid with equimolar amounts of uridine-binding PUF60 that had RRM1 substitution L140P or H169Y (57,99). Right panels show quantitation of exon skipping. **(B)** Allelic association between human TAG/CAG 3' splice sites and the first six exon positions. The most associated alleles are shown in red. The table is based on haplotypes in ~195 000 introns of the reference human genome. **(C)** Exon skipping of reporters with TAG and CAG 3' splice sites carrying positively (green) and negatively (red) associated haplotypes at the first three exon positions. Right panels show exon skipping in HEK293 cells. **(D)** CAG 3' splice sites are preferred over TAG 3' splice sites in most informative transcripts. Supporting data are shown in (E). **(E)** Representative gels showing exon promotion or repression by mutations -3T>C (top) or -3C>T (bottom) introduced at the 3' splice sites of the indicated constructs. Wild-type and mutated constructs are schematically shown to the right: boxes are exons, vertical lines are introns, diagonal lines are restriction sites. The reporters were prepared for disease-associated (DA) mutations (left column) and Zn²⁺/Ca²⁺ exons (right column). Their numbering corresponds to (D). Spliced products are shown at the bottom of each panel: tested exons are in red, exons adjacent to tested exons in the pre-mRNA are in blue, exonic segments activated by cryptic 3' splice sites are in grey, and exon skipping measurements (%) are in blue. DA mutations are designated by c numbers that denote distances in nucleotides between the first position of the tested exon and adenine in the start codon of the longest open reading frame (Supplementary Table S5). **(F)** The impact of *SMN1/2* exon 7 variants at position +6 on the splicing outcome of closely linked C/T alleles at position -3 of exon 7 acceptor site. **(G)** Total number of informative TAG and CAG 3' splice sites that promoted exon skipping in tested constructs. The character height in logos depicts relative nucleotide frequencies at the indicated 3' splice sites positions.

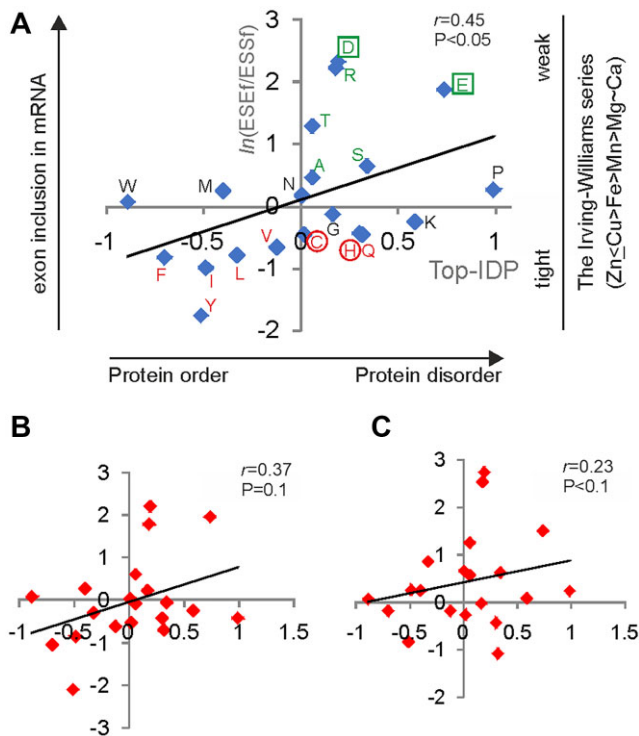


Figure 5. Auxiliary splicing code in exons, intrinsically disordered regions and the Irving–Williams series. **(A)** Correlation between the propensity to protein disorder and codon-specific exon inclusion levels assuming equal codon usage. Each diamond represents one amino acid. Codons for amino acids in green generally promote splicing, and codons for residues in red generally repress splicing on a continuous scale of $\ln(\text{ESEf}/\text{ESSf})$ values (y-axis). Residues in green boxes are preferred ligands for weak divalent metals, and residues in red circles preferentially bind tighter metals in the Irving–Williams series. Top-IDP values (x-axis) predict IDRs (87). The correlation coefficient (top right) was similar for other disorder propensity scales (Table 3). **(B)** and **(C)** Diminished correlation of Top-IDP and γ -weighted $\ln(\text{ESE}/\text{ESS})$ values, reflecting codon usage in dividing **(B)** and differentiated cells **(C)**. Codon usage data are from (88). The y- and x-axes are as in (A).

The Irving–Williams series, ESEs/ESSs and intrinsically disordered regions

Alternative splicing tends to avoid protein domains while preferentially affecting intrinsically disordered regions (IDRs), polypeptide segments that do not acquire a defined tertiary structure autonomously but adopt diverse interconverting conformational states (82–86). Because this association may include domains that bind Zn^{2+} and Ca^{2+} , we wished to explore the relationship between IDRs and codons for Zn^{2+} - and Ca^{2+} -coordinating residues. Under the assumption of an equal codon usage, codon-specific ESEf/ESSf values significantly correlated with various measures that predict protein disorder (Figure 5A). For example, one of the best performing measures, the TOP-IBD scale (87), showed a correlation coefficient of 0.45 while other measures of protein disorder yielded similar values (Table 3), confirming a general tendency of alternatively spliced regions to have higher ESE/ESS ratios and to be intrinsically disordered. The correlation is mainly driven by (i) exon-promoting codons for negatively charged amino acids (Asp, Glu), which coordinate weak divalent metals of the Irving–Williams series (Figure 1A); (ii) a subset of Arg, Thr, Ser and Ala codons, which were previously associated with protein disorder (87); and (iii) exon-repressing codons for Phe,

Table 3. Correlation between average $\ln(\text{ESEf}/\text{ESSf})$ values for 20 amino acids and four measures of protein disorder

Average $\ln(\text{ESEf}/\text{ESSf})^a$	Top-IDP	B-value	FoldUnfold	DisProt
Correlation coefficient	0.45	0.50	−0.56	0.55
P-value	0.047	0.020	0.009	0.011

^a Average values were computed assuming equal codon usage. Weighted values were associated with increased P-values and diminished statistical significance (Figure 5B, C).

Tyr and Ile/Leu/Val, which contribute to the protein order (Figure 5A). His and Cys residues, which coordinate tight metals, displayed inconsistent positions on some protein disorder scales (Supplementary Figure S13).

However, when the ESEf/ESSf values were weighted to reflect codon usage in gene categories involved in cell proliferation as opposed to differentiation (88), this correlation was reduced or disappeared (Figure 5B, C; Supplementary Figure S7). This observation is in line with a good correspondence between codon usage preferences and supply of tRNAs that were induced in proliferating cells and repressed in differentiating cells (88). There was no correlation between the molecular weight of amino acids and ESEf/ESSf values (Supplementary Figure S14A, B). Inclusion of both Zn^{2+} and Ca^{2+} exons is therefore influenced by codons encoding residues in the middle of the molecular weight spectrum and is largely supported by charged amino acids (D, E, R) while outermost residues tend to be splicing neutral (W, G).

Discussion

How splicing overcomes metal coordination constraints

This study demonstrates that exonic segments encoding Zn^{2+} -coordinating amino acids have, on average, a significantly lower capacity to promote inclusion in mRNAs than those encoding Ca^{2+} -coordinating residues. The two elements represent examples of tight and weak metals in the Irving–Williams series ($\text{Ca}^{2+} \sim \text{Mg}^{2+} < \text{Mn}^{2+} < \text{Fe}^{2+} < \text{Cu}^{2+} \geq \text{Zn}^{2+}$) (8). Because weak Mg^{2+} ions are preferentially coordinated to the same residues as Ca^{2+} and tight Cu^{2+} ions to the same residues as Zn^{2+} (89), it is reasonable to expect a similar ESE/ESS dichotomy for Mg^{2+} and Cu^{2+} (Figure 1A). However, exon datasets for protein-binding sites for these and other metals in the series are either much smaller or less well defined than those for Zn^{2+} or Ca^{2+} . For example, the number of exons that encode Cu^{2+} -coordinating residues in human copper proteins is currently an order of magnitude smaller than the number of Zn^{2+} exons in this study, but codon usage for their main coordinating amino acids is much more similar to Zn^{2+} than to Ca^{2+} (D.B. et al. ms in preparation).

These results support the existence of an ESE/ESS-mediated selection pressure that favours exons encoding binding sites for weak metals and disfavours exons encoding binding sites for tight metals. However, the extent to which such selection took place during evolution remains unknown. The Irving–Williams constraints were adopted by genetic codes of both prokaryotes and eukaryotes, which have similar residue preferences in their coordination spheres (His, Cys for tight metals, Glu, Asp for weak metals). However, pro- and eukaryotes have vastly distinct codon usage frequencies for these amino acids, which were linked to the expanding number of

introns during eukaryotic evolution and which could affect ESE/ESS profiles. In addition, divalent metal bioavailability and cellular requirements dramatically changed during evolution, largely driven by photosynthesis-mediated oxygenation of euxinic sea and atmosphere (11). Arrival of Ca^{2+} signalling systems uniquely suitable for life came early and, at the exon level, they probably remained protected for over a billion years (Figure 2A). In contrast, addition of Zn^{2+} and Cu^{2+} to the environment took place later (11). These considerations warrant more detailed studies into the role of universal stabilities of these metals and their properties in the Periodic Table of Elements in shaping ESE/ESS-mediated splicing responses in other species (Figure 1A).

The ‘natural defence against tight metals’ embedded in the exonic splicing code is efficiently defeated by combinatorial diversity and flexibility of the splicing process *in vivo* (Figures 1 and 2). This would be consistent with the ability to overpower the Irving–Williams affinity order *in vitro* by flexible design of mutually exclusive metal-dependent protein conformation states (92). PSI values of both Zn^{2+} and Ca^{2+} exons were higher than those of average exons (Figure 2B, C), indicating that the cells evolved adaptive mechanisms at the exon level to safeguard Zn^{2+} - and Ca^{2+} -binding sites in the proteome. These adaptations over-ride the low capacity of ‘tight’ codons to promote inclusion of their exons in most, albeit not in all, cases (Figures 2 and 3). They include selection for stronger 5′ and 3′ss/PPTs but also stronger ESE/ESS profiles in the remaining exonic segments, at least for ZFs (Figure 2D, H–K). The compensation by splice site strength gauged by maximum entropy scoring appeared greater for 5′ss than for 3′ss (Figure 2D), in line with the central role of donor sites in splicing networks (90); nevertheless, we detected a robust effect of PPT composition albeit not of PPT length (Figure 2H–J). This suggests that the PPT position between the branchpoint and 3′ss is more important than PPT length when the number of pyrimidines is limiting, in line with a previous observation (91). Only ~46% of control introns had optimal (U)₄ motifs for U2AF65 binding downstream of predicted branchpoints and upstream of 3′ss (Supplementary Figure S10C).

Candidate *trans*-acting factors involved in differential recognition of Zn^{2+} and Ca^{2+} exons

PPT/3′ss interact with U2AF, a tight and highly abundant heterodimer found in the early spliceosome assembly complex and required for recruitment of U2 small nuclear ribonucleoproteins (RNPs) (93,94). U2AF65 strongly favours uridine-rich sequences (94,95). Uridine excess and depletion preceding Zn^{2+} and Ca^{2+} exons, respectively, were observed for most PPT positions (Figure 2J). A significant depletion of (U)₄ in strong (Ca^{2+}) exons and over-representation in weak (Zn^{2+}) exons (Supplementary Figure S10) also argues for a role for U2AF65, assuming that the (U)₄ motif is the most favourable binding platform (67). In Zn^{2+} exons, the upstream peak of uridine frequencies extended towards branchpoints (Figure 2J) and a similar shift was observed as a compensation of poorly competing U2AF65 PPTs selected by enhanced cross-linking and immunoprecipitation (96). PPT uridines are preferred by both U2AF65 RRM1, but RRM1 appears to be more tolerant to cytidine and purine substitutions than RRM2, which is thought to interact with the upstream peak (97,98). Distinct PPTs preceding Zn^{2+} and Ca^{2+} exons may require dis-

tinct interactions with U2AF for accurate splicing, not excluding other factors involved in PPT-mediated splicing regulation, such as PUF60, TIA1/TIAR, PTB or hnRNP C (67,99, and references therein; FAS exon 6 in Supplementary Table S5). On the other hand, splicing *in vivo* may also benefit from mechanisms that delay multiple occupancy of candidate 3′ss by U2AF rather than from recruiting other factors (100). The compensation by intronic mutations improving PPTs of weak exons during evolution provides a more viable alternative to modifying exonic sequences. The auxiliary splicing code in exons *per se* may have been insufficient to compensate metal coordination constraints and ensure correct expression of mRNAs encoding metal-binding sites. This highlights thus far unknown function of introns in splicing of pre-mRNAs that encode metalloproteins.

Although 3′ss recognition depends, to a large extent, on uridine fractions in PPT (94), weak PPT may be partially overcome by strong branchpoints (95). Even if no obvious changes in branchpoint consensus and distance from 3′ss between ZnBPs and CaBPs were observed, we cannot formally exclude this possibility given the highly degenerate nature of a short branchpoint sequence consensus (101), suboptimal accuracy of branchpoint prediction tools (45), possible tissue-specific branchpoint use (102), discordant branchpoint assignments by computational versus experimental approaches (28) and the existence of multiple, distant and non-canonical branchpoints (28,47,48,103,104). The limited fraction of concordant branchpoints identified by the three methods (90 among 240 informative cases; Supplementary Dataset S6) reduces the exclusion power of branchpoint comparisons, but discordant branchpoint assignments did not invalidate the significant difference in uridine content in PPTs between Zn^{2+} and Ca^{2+} exons. On the other hand, a lack of evidence for a compensatory role for branchpoint motifs would be consistent with a lack of conformational shifts of U2AF65 in the presence of branchpoint-bound splicing factor 1 (77, and references therein).

Candidate *trans*-acting factors that recognize RNA motifs containing codons for main Zn^{2+} - and Ca^{2+} -coordinating residues are listed in Supplementary Table S8.

Together, we propose that U2AF65 interactions with stronger PPTs of Zn^{2+} exons may help achieve high exon inclusion and defy ESE/ESS restrictions imposed by Zn^{2+} coordination. Future studies should address if stronger splice sites of Zn^{2+} exons are sufficient to overcome their poor ESE/ESS profiles or if they require additional features in their intron–exon architecture, such as pre-mRNA structures.

Position –3 relative to 3′ splice sites in Zn^{2+} and Ca^{2+} exons

The increased binding affinity of U2AF65 for uridine-rich PPTs correlates with a shift from closed to open or active conformation required for spliceosome assembly, which is stabilized by U2AF35 (77, and references therein). U2AF35 recognizes both UAG and CAG 3′ss (78). However, the higher binding affinity of U2AF35 for UAG than for CAG or RAG RNAs (R = purine), which was detected by isothermal calorimetry irrespective of the nucleotide identity at the first exon position (78), seems incompatible with a clear preference for human CAG 3′ss (Figure 4D–F) and it is unclear if it could explain their excess in Ca^{2+} exons either (Figure 2E). The 3′ss YAG/R motifs are bound by both U2AF35 ZFs, but position –3 mainly

interacts with ZF1 via R28, H29, S34 and R35 (78). These residues are in the vicinity of Zn²⁺-coordinating cysteines C27 and C33. Adjacent S34 is a hot spot for cancer-associated substitutions S34F and S34Y (105). In the complex of wild-type *Saccharomyces cerevisiae* U2AF35 with UAG 3'ss, -3U is stacked with the imidazole ring of H29 and surrounded by S34, R35 and a short ZF1 helix, whereas the pocket accepting the -3 base is too shallow to accommodate purines (78). Replacement of -3U by -3C was associated with a loss of hydrogen bond to S34, with the S34 mutants showing less discrimination at position -3 (78). The distribution of CAG and UAG 3'ss was reported to be altered in tumours carrying mutation U2AF35 S34F, with the S34F mutant preferentially binding CAG 3'ss (106,107, and references therein). The higher affinity of the wild-type *S. cerevisiae* U2AF35 for UAG as compared with CAG RNAs (78) could reflect a predominance of UAG 3'ss in some yeasts, whereas multicellular organisms have been generally associated with a higher abundance of CAG 3'ss, including almost invariant CAG 3'ss in some nematodes (68,90,108,109). This notion is supported by a dramatically better splicing of UAG 3'ss over CAG 3'ss of a fission yeast *cdc2* intron (109). Although the excess of CAG 3'ss in human Ca²⁺ exons could compensate weaker adenines at position +1 (cf. Figure 2E and Supplementary Figure S8), adenine at this position was absent in a small number of CAG 3'ss that induced more exon skipping than UAG 3'ss (Figure 4G).

Prediction of splicing defects for pyrimidine transitions at position -3 relative to 3'ss

Our results improve prediction of splicing and phenotypic outcomes of T/C variants at position -3. First, their effect on splicing clearly depends on intron type. This is exemplified by a 2 nt insertion between adjacent *SCN1A* exons due to mutation c.4477-3T>C in the AT-AC type intron (Supplementary Table S5). Since this intron ends with AC and not AG (tttcta[-3t>c]/acTTTGGAGG, where / is a new 3'ss), the mutation replicates the TAC 3'ss 2 nt upstream, creating a frameshift by inserting an extra AC dinucleotide into the mRNA. In a canonical GT-AG intron, the same mutation could not create the new splice site. Second, mutations -3C>T are more likely to result in exon skipping than -3T>C (Figure 4D, E). Although it is unclear why it is not always the case, the compiled lists (Supplementary Tables S5 and S6) will be useful in future studies of functional differences of CAG and TAG 3'ss. Third, even if mutations -3T/C do not alter splicing of exogenous transcripts (Figure 4D, E), they may have other unforeseen effects *in vivo* that may be cell type specific, such as cryptic 3'ss activation (Figure 4A, E). Many disease-associated -3C/T mutations were found in the last introns (Supplementary Table S5), potentially affecting gene termination. Fourth, the PPT strength can partially or completely compensate for a weaker CAG or TAG 3'ss (Figure 4A, C). Fifth, the first position of exons as well as other exonic positions often modulate the splicing outcome of mutations -3T/C (Figure 4C). U2AF35 interactions with the 3'ss YAG/R consensus may extend to or depend on a downstream ESE (110), suggesting that exon variants closely linked to mutations -3C/T (111,112; Supplementary Table S5) could help identify U2AF35-mediated interactions between position -3 and exonic sequences. Although *ab initio* prediction of phenotypic consequences of pyrimidine transitions at position -3

relative to 3'ss is futile at present, this guidance should improve it.

Evolution of exons encoding binding sites for divalent metals

Given the high ESE/ESS conservation (18), differences in ESE/ESS profiles in human Zn²⁺ or Ca²⁺ exons are likely to be replicated in other vertebrates. Alternative splicing of selected human Zn²⁺ or Ca²⁺ exons was, however, only partially recapitulated in the rat (Figure 3C, D) even though each rat coordinating codon in tested exons was identical to human counterparts.

The total content of the six metal elements (Figure 1A) is similar in animals and plants (except for a higher Mn²⁺ in the latter) (11), raising questions about metal-dependent evolution of the splicing code in other multicellular organisms. The number of proteins that bind four of the six elements (Ca²⁺, Mg²⁺, Cu²⁺ and Zn²⁺) increased very rapidly in multicellular eukaryotes (11).

Compensatory interactions between ESEs/ESSs and splice sites of Zn²⁺ and Ca²⁺ exons (Figure 2) support exons as primary evolutionary units in eukaryotes (90). The greater promotion of exonic sequences for Ca²⁺-binding sites by their ESEs/ESSs than for Zn²⁺-binding sites may have contributed to their very rapid spread when the first eukarya emerged (113). Cell type-specific expression of CaBPs could be better regulated by a combination of weak traditional and strong auxiliary splicing signals (Figures 1 and 2). Eukaryotic proteomes contain a greater fraction of Zn²⁺- and Ca²⁺-binding structures than those binding iron, and the eschewal of iron for Zn²⁺ and Ca²⁺ during evolution was their defining feature (113).

The hypothesis that evolution of the auxiliary splicing code in exons was responsive to the Irving-Williams constraints (Figure 1A) relies on two assumptions. First, codon frequencies at coordination spheres for weak metals are distinct from those for tight metals (Figure 1A). Second, PDB structures for metalloproteins largely reflect Zn²⁺ and Ca²⁺ binding *in vivo*. While the first assumption appears to be solid (89, and references therein), there are multiple examples where the latter supposition does not hold. Existing structural models may have employed procedures that led to incorrect metal assignments, including the use of chelating agents, improper ion concentration ranges during protein purification procedures or artificial culture growth conditions (9). Apart from metal properties such as redox state, the metal content in a protein depends on species, cell type and subcellular location, restricting or promoting metal bioavailability (9). In eukarya, subcellular compartmentalization of protein interactions and their rewiring can be determined by alternative exon usage (114), but examples of functional isoforms that are alternatively spliced to include or exclude metal-binding sites have been scarce. Our screening of low-PSI Zn²⁺ and Ca²⁺ exons confirmed alternative splicing in 18 cases (Figure 3), suggesting that only a limited number of these exons is expendable without altering fitness. Known examples include alternative splicing of *SLC30A2* (ZnT2) exon 3 (Figure 3A), which generates two mRNA isoforms in mammary glands that differ by 49 amino acids and a transmembrane domain of the antiporter (115). Both isoforms are functional and have a distinct subcellular location (115). Alternative splicing regulates other antiporters and uniporters, such as KEA3 via the C-terminal

KTN domain (116) or MICU1 via a small microexon (117). Canonical or major isoforms are highly abundant for some proteins (115,116) but not for others (117). The importance of subcellular location in differential metal binding was also supported for Mia40 (*CHCHD4*) (118), MitoNEET (*CISD1*) (119) or Miner1 (*CISD2*) (120), but the overall number of these cases remains low.

A gradient of ESE/ESS profiles for codons encoding Zn²⁺-coordinating residues

A growing fraction of proteins has been reported to bind more than one metal element, and binding of one divalent metal can influence binding of another. For example, Zn²⁺-binding sites in S100 proteins can bind Cu²⁺, which could be displaced by Zn²⁺, but not by Ca²⁺ (121). Binding of tight Cu²⁺ to S100A5 impairs binding of the EF-hand to Ca²⁺ (122). High Zn²⁺ concentrations can displace Mg²⁺ from the EF-hands and alter stoichiometry of Ca²⁺ binding (123). Elevation of extracellular Zn²⁺ can increase intracellular Ca²⁺ and inhibit calmodulin (124, and references therein). Zn²⁺ binding may enhance the affinity for target peptides over the effect of Ca²⁺ alone (125). Zn²⁺ binding to S100 EF-hands can alter the geometry of the Ca²⁺-binding loop and Ca²⁺ affinity (126) although the effects of Zn²⁺ binding on the structure of S100 proteins appear to be generally more modest than those elicited by Ca²⁺ binding (121). Binding of both tight and weak metals by a single protein might reduce selection pressures to exclude tight codons from mature transcripts. The intermediate ESE/ESS score values for S100 and other proteins that bind non-Zn²⁺ metals (Figure 1C–E) suggest that codons for acidic residues compensate low exon inclusion levels typical of coordinating codons for Zn²⁺ (Figure 1C–E) and other tight binders (Figure 1A). During evolution, the proportion of ZFs in proteomes increased from ~0.2% in archaea to 1.9% in yeast and > 3% in eukarya (127), yet their low exon inclusion potential did not impede ZF spread in intron-containing organisms and was compensated to some degree at the ESE/ESS level (Figure 2A). Crystal structures of the best studied bacterial ZIP ancestor, ZIPB, showed coordination spheres distinct from a classical Zn²⁺ tetrahedral site, with a more complex binuclear site and coordination numbers ranging from 4 to 6 (62). The Zn²⁺ transport sites of ZIPB or YiiP employ acidic residues rather His/Cys (Supplementary Figures S2 and S3), bridging separate ions and giving the coordination spheres higher flexibility. In addition, alternative acidic residues may confer promiscuous metal selectivity (62). Thus, acidic residues have important roles in ZnBPs, and their codons would compensate low exon inclusion levels of associated Cys/His codons. Codons for acidic residues at Zn²⁺ interfaces may account for 18–19% of the total (128,129), a figure that may have a global impact on exon inclusion.

Does the exonic splicing code reduce mismetalation?

To safeguard codons for residues coordinating both weak and tight metals in mature transcripts, the splicing code evolved to repress codons for tighter metals and promote codons for weak metals, potentially reducing mismetalation in the cell. Mismetalation coincides with altered metal bioavailability, which is believed to be a crucial contributor to its avoidance (1, and references therein). Even in prokaryotes, bioavailability in cyto- and periplasmic compartments can restrict com-

petition between tighter binding Cu²⁺ and weaker Mn²⁺ cupins (130). In higher eukaryotes, bioavailability can be substantially controlled by regulated splicing decisions that restrict or permit protein–protein interactions by introducing distinct protein isoforms with distinct post-translation modification, folding or intrinsically disordered patterns (86) (Figure 3E). Zn²⁺ may bind differently to ordered and disordered regions (Supplementary Table S9), nevertheless it is as yet unclear if post-translational modifications of metal-coordinating residues such as His phosphorylation could play a role in reducing mismetalation. The location of protein folding can also determine the specificity of metalation, with correct or incorrect metals kinetically trapped within folded proteins (130). Importantly, metal availability follows, or tends toward, the inverse of the Irving–Williams series (131) and also appears to reflect metal concentrations estimated for euxinic sea waters (Figure 1A, bottom).

In conclusion, our results provide new insights into the intertwined protein-coding, splicing and metal codes in human DNA and most probably in DNA of other mammals and higher vertebrates. The universal order of metal binding preferences is difficult for evolution to circumvent (132), forcing the cell to exploit both the ‘invisible’ exonic splicing code and traditional splicing motifs in introns to ensure adequate expression of metalloproteins that bind weak or competitive divalent metals.

Data availability

The authors confirm that the data supporting the findings in this study are available within the article and its supplementary data.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We wish to thank staff members for technical help and useful discussions.

Funding

This work was supported by patent royalties (to I.V. and J.K.) personally contributed to the University of Southampton and administered as a research grant by the same institution; VEGA [2/0016/22 to J.K.] and the Slovak Research and Development Agency [APVV-18-0096 to J.K.].

Conflict of interest statement

None declared.

References

1. Foster, A.W., Young, T.R., Chivers, P.T. and Robinson, N.J. (2022) Protein metalation in biology. *Curr. Opin. Chem. Biol.*, **66**, 102095.
2. Holm, R.H., Kennepohl, P. and Solomon, E.I. (1996) Structural and functional aspects of metal sites in biology. *Chem. Rev.*, **96**, 2239–2314.

3. Waldron, K.J., Rutherford, J.C., Ford, D. and Robinson, N.J. (2009) Metalloproteins and metal sensing. *Nature*, **460**, 823–830.
4. Andreini, C., Bertini, I., Cavallaro, G., Holliday, G.L. and Thornton, J.M. (2008) Metal ions in biological catalysis: from enzyme databases to general principles. *J. Biol. Inorg. Chem.*, **13**, 1205–1218.
5. Clemens, S. (2022) The cell biology of zinc. *J. Exp. Bot.*, **73**, 1688–1698.
6. Brylinski, M. and Skolnick, J. (2011) FINDSITE-metal: integrating evolutionary information and machine learning for structure-based metal-binding site prediction at the proteome level. *Proteins*, **79**, 735–751.
7. Waldron, K.J. and Robinson, N.J. (2009) How do bacterial cells ensure that metalloproteins get the correct metal? *Nat. Rev. Microbiol.*, **7**, 25–35.
8. Frausto da Silva, J.J.R. and Williams, R.J.P. (2001) In: *The Biological Chemistry of the Elements*. 2nd edn., Oxford University Press, Oxford.
9. Maret, W. (2010) Metalloproteomics, metalloproteomes, and the annotation of metalloproteins. *Metallomics*, **2**, 117–125.
10. Irving, H. and Williams, R.J.P. (1948) Order of stability of metal complexes. *Nature*, **162**, 746–747.
11. Williams, R.J.P. and Fausto da Silva, J.J. (2006) In: *The Chemistry of Evolution: The Development of Our Ecosystem*. Elsevier, Amsterdam.
12. Foster, A.W., Osman, D. and Robinson, N.J. (2014) Metal preferences and metallation. *J. Biol. Chem.*, **289**, 28095–28103.
13. Osman, D., Martini, M.A., Foster, A.W., Chen, J., Scott, A.J.P., Morton, R.J., Steed, J.W., Lurie-Luke, E., Huggins, T.G., Lawrence, A.D., et al. (2019) Bacterial sensors define intracellular free energies for correct enzyme metalation. *Nat. Chem. Biol.*, **15**, 241–249.
14. Dudev, T. and Lim, C. (2003) Principles governing Mg, Ca, and Zn binding and selectivity in proteins. *Chem. Rev.*, **103**, 773–788.
15. Dudev, T. and Lim, C. (2014) Competition among metal ions for protein binding sites: determinants of metal ion selectivity in proteins. *Chem. Rev.*, **114**, 538–556.
16. Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature*, **416**, 499–506.
17. Wahl, M.C., Will, C.L. and Luhrmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701–718.
18. Yeo, G., Hoon, S., Venkatesh, B. and Burge, C.B. (2004) Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc. Natl Acad. Sci. USA*, **101**, 15000–15005.
19. Fairbrother, W.G. and Chasin, L.A. (2000) Human genomic sequences that inhibit splicing. *Mol. Cell. Biol.*, **20**, 6816–6825.
20. Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M. and Burge, C.B. (2004) Systematic identification and analysis of exonic splicing silencers. *Cell*, **119**, 831–845.
21. Ke, S., Zhang, X.H. and Chasin, L.A. (2008) Positive selection acting on splicing motifs reflects compensatory evolution. *Genome Res.*, **18**, 533–543.
22. Ke, S., Shang, S., Kalachikov, S.M., Morozova, I., Yu, L., Russo, J.J., Ju, J. and Chasin, L.A. (2011) Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res.*, **21**, 1360–1374.
23. Ke, S., Anquetil, V., Zamalloa, J.R., Maity, A., Yang, A., Arias, M.A., Kalachikov, S.M., Russo, J.J., Ju, J. and Chasin, L.A. (2018) Saturation mutagenesis reveals manifold determinants of exon definition. *Genome Res.*, **28**, 11–24.
24. Pagani, F., Raponi, M. and Baralle, F.E. (2005) Synonymous mutations in *CFTR* exon 12 affect splicing and are not neutral in evolution. *Proc. Natl Acad. Sci. USA*, **102**, 6368–6372.
25. Movassat, M., Forouzmand, E., Reese, F. and Hertel, K.J. (2019) Exon size and sequence conservation improves identification of splice-altering nucleotides. *RNA*, **25**, 1793–1805.
26. Rong, S., Buerer, L., Rhine, C.L., Wang, J., Cygan, K.J. and Fairbrother, W.G. (2020) Mutational bias and the protein code shape the evolution of splicing enhancers. *Nat. Comm.*, **11**, 2845.
27. Pengelly, R., Bakhtiar, D., Borovska, I., Kralovicova, J. and Vorechovsky, I. (2022) Exonic splicing code and protein binding sites for calcium. *Nucleic Acids Res.*, **50**, 5493–5512.
28. Kralovicova, J., Borovska, I., Pengelly, R., Lee, E., Abaffy, P., Sindelka, R., Grutzner, F. and Vorechovsky, I. (2021) Restriction of an intron size *en route* to endothermy. *Nucleic Acids Res.*, **49**, 2460–2487.
29. UniProt Consortium. (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480–D489.
30. Ireland, S.M. and Martin, A.C.R. (2019) ZincBind—the database of zinc binding sites. *Database (Oxford)*, **2019**, baz006.
31. Putignano, V., Rosato, A., Banci, L. and Andreini, C. (2018) MetalPDB in 2018: a database of metal sites in biological macromolecular structures. *Nucleic Acids Res.*, **46**, D459–D464.
32. Zheng, H., Chordia, M.D., Cooper, D.R., Chruszcz, M., Muller, P., Sheldrick, G.M. and Minor, W. (2014) Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. *Nat. Protoc.*, **9**, 156–170.
33. Zheng, H., Cooper, D.R., Porebski, P.J., Shabalin, I.G., Handing, K.B. and Minor, W. (2017) *CheckMyMetal*: a macromolecular metal-binding validation tool. *Acta Crystallogr.*, **D73**, 223–233.
34. Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.*, **270**, 2411–2414.
35. Kralovicova, J. and Vorechovsky, I. (2007) Global control of aberrant splice site activation by auxiliary splicing sequences: evidence for a gradient in exon and intron definition. *Nucleic Acids Res.*, **35**, 6399–6413.
36. Gerbracht, J.V., Boehm, V. and Gehring, N.H. (2017) Plasmid transfection influences the readout of nonsense-mediated mRNA decay reporter assays in human cells. *Sci. Rep.*, **7**, 10616.
37. Schafer, S., Miao, K., Benson, C.C., Heinig, M., Cook, S.A. and Hubner, N. (2015) Alternative splicing signatures in RNA-seq data: percent spliced in (PSI). *Curr. Protoc. Hum. Genet.*, **87**, 11.16.1–11.16.14.
38. Tapial, J., Ha, K.C., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., Quesnel-Vallieres, M., Permanyer, J., Sodaei, R., Marquez, Y., et al. (2017) An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.*, **27**, 1759–1768.
39. Busch, A. and Hertel, K.J. (2013) HEXEvent: a database of Human EXon splicing Events. *Nucleic Acids Res.*, **41**, D118–D124.
40. Yeo, G. and Burge, C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.*, **11**, 377–394.
41. Vorechovsky, I. (2006) Aberrant 3' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res.*, **34**, 4630–4641.
42. Buratti, E., Chivers, M.C., Kralovicova, J., Romano, M., Baralle, M., Krainer, A.R. and Vorechovsky, I. (2007) Aberrant 5' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res.*, **35**, 4250–4263.
43. Zhang, Q., Fan, X., Wang, Y., Sun, M., Shao, J. and Guo, D. (2017) BPP: a sequence-based algorithm for branch point prediction. *Bioinformatics*, **33**, 3166–3172.
44. Corvelo, A., Hallegger, M., Smith, C.W. and Eyras, E. (2010) Genome-wide association between branch point properties and alternative splicing. *PLoS Comput. Biol.*, **6**, e1001016.
45. Leman, R., Tubeuf, H., Radd, S., Tournier, J., Derambure, C., Lanos, R., Gildrat, P., Castelain, G., Hauchard, J., Killian, A., et al. (2020) Assessment of branch point prediction tools to predict physiological branch points and their alteration by variants. *BMC Genomics*, **21**, 86.

46. Harris,N.L. and Senapathy,P. (1990) Distribution and consensus of branch point signals in eukaryotic genes: a computerized statistical analysis. *Nucleic Acids Res.*, **18**, 3015–3019.
47. Mercer,T.R., Clark,M.B., Andersen,S.B., Brunck,M.E., Haerty,W., Crawford,J., Taft,R.J., Nielsen,L.K., Dinger,M.E. and Mattick,J.S. (2015) Genome-wide discovery of human splicing branchpoints. *Genome Res.*, **25**, 290–303.
48. Gooding,C., Clark,F., Wollerton,M., Grellscheid,S.-N., Groom,H. and Smith,C.W. (2006) A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. *Genome Biol.*, **7**, R1.
49. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
50. Wu,Q. and Krainer,A.R. (1999) AT–AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. *Mol. Cell. Biol.*, **19**, 3225–3236.
51. Levine,A. and Durbin,R. (2001) A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Res.*, **29**, 4006–4013.
52. Alioto,T.S. (2007) U12DB: a database of orthologous U12-type spliceosomal introns. *Nucleic Acids Res.*, **35**, D110–D115.
53. Olthoff,A.M., Hyatt,K.C. and Kanadia,R.N. (2019) Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. *BMC Genomics*, **20**, 686.
54. Kralovicova,J., Knut,M., Cross,N.C. and Vorechovsky,I. (2015) Identification of U2AF(35)-dependent exons by RNA-Seq reveals a link between 3' splice-site organization and activity of U2AF-related proteins. *Nucleic Acids Res.*, **43**, 3747–3763.
55. Ross,R.A., Spengler,B.A. and Biedler,J.L. (1983) Coordinate morphological and biochemical interconversion of human neuroblastoma cells. *J. Natl Cancer Inst.*, **71**, 741–747.
56. Alvarez,M.E.V., Chivers,M., Borovska,I., Monger,S., Giannoulatou,E., Kralovicova,J. and Vorechovsky,I. (2020) Transposon clusters as substrates for aberrant splice-site activation. *RNA Biol.*, **18**, 354–367.
57. Kralovicova,J., Borovska,I., Kubickova,M., Lukavsky,P.J. and Vorechovsky,I. (2020) Cancer-associated substitutions in RNA recognition motifs of PUF60 and U2AF65 reveal residues required for correct folding and 3' splice-site selection. *Cancers*, **12**, 1865.
58. Ascher,D.B., Spiga,O., Sekelska,M., Pires,D.E.V., Bernini,A., Tiezzi,M., Kralovicova,J., Borovska,I., Soltysova,A., Olsson,B., et al. (2019) Homogentisate 1,2-dioxygenase (HGD) gene variants, their analysis and genotype–phenotype correlations in the largest cohort of patients with AKU. *Eur. J. Hum. Genet.*, **27**, 888–902.
59. Kralovicova,J., Gaunt,T.R., Rodriguez,S., Wood,P.J., Day,I.N.M. and Vorechovsky,I. (2006) Variants in the human insulin gene that affect pre-mRNA splicing: is –23HpbI a functional single nucleotide polymorphism at IDDM2? *Diabetes*, **55**, 260–264.
60. Klug,A. (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.*, **79**, 213–231.
61. Baltaci,A.K. and Yuce,K. (2018) Zinc transporter proteins. *Neurochem. Res.*, **43**, 517–530.
62. Kambe,T., Taylor,K.M. and Fu,D. (2021) Zinc transporters and their functional integration in mammalian cells. *J. Biol. Chem.*, **296**, 100320.
63. Lin,H., Andersen,G.R. and Yatime,L. (2016) Crystal structure of human S100A8 in complex with zinc and calcium. *BMC Struct. Biol.*, **16**, 8.
64. Sterner,D.A., Carlo,T. and Berget,S.M. (1996) Architectural limits on split genes. *Proc. Natl Acad. Sci. USA*, **93**, 15081–15085.
65. Jaruzelska,J., Abadie,V., d'Aubenton-Carafa,Y., Brody,E., Munnich,A. and Marie,J. (1995) In vitro splicing deficiency induced by a C to T mutation at position –3 in the intron 10 acceptor site of the phenylalanine hydroxylase gene in a patient with phenylketonuria. *J. Biol. Chem.*, **270**, 20370–20375.
66. Warnasooriya,C., Feeney,C.F., Laird,K.M., Ermolenko,D.N. and Kielkopf,C.L. (2020) A splice site-sensing conformational switch in U2AF2 is modulated by U2AF1 and its recurrent myelodysplasia-associated mutation. *Nucleic Acids Res.*, **48**, 5695–5709.
67. Sutandy,F.X.R., Ebersberger,S., Huang,L., Busch,A., Bach,M., Kang,H.S., Fallmann,J., Maticzka,D., Backofen,R., Stadler,P.F., et al. (2018) In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. *Genome Res.*, **28**, 699–713.
68. Shapiro,M.B. and Senapathy,P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.*, **15**, 7155–7174.
69. Vezain,M., Gerard,B., Drunat,S., Funelot,B., Fehranbach,S., N'Guyen-Viet,V., Vallat,J.M., Frebourg,T., Tosi,M., Martins,A., et al. (2011) A leaky splicing mutation affecting SMN1 exon 7 inclusion explains an unexpected mild case of spinal muscular atrophy. *Hum. Mutat.*, **32**, 989–994.
70. Heintz,C., Dobrowolski,S.F., Andersen,N.S., Demirkol,M., Blau,N. and Andresen,B.S. (2012) Splicing of phenylalanine hydroxylase (PAH) exon 11 is vulnerable: molecular pathology of mutations in PAH exon 11. *Mol. Genet. Metab.*, **106**, 403–411.
71. Yuan,J., Ma,Y., Huang,T., Chen,Y., Peng,Y., Li,B., Li,J., Zhang,Y., Song,B., Sun,X., et al. (2018) Genetic modulation of RNA splicing with a CRISPR-guided cytidine deaminase. *Mol. Cell*, **72**, 380–394.
72. Weisschuh,N., Mazzola,P., Bertrand,M., Haack,T.B., Wissinger,B., Kohl,S. and Stingl,K. (2021) Clinical characteristics of POC1B-associated retinopathy and assignment of pathogenicity to novel deep intronic and non-canonical splice site variants. *Int. J. Mol. Sci.*, **22**, 5396.
73. Hata,Y., Oku,Y., Taneichi,H., Tanaka,T., Igarashi,N., Niida,Y. and Nishida,N. (2020) Two autopsy cases of sudden unexpected death from Dravet syndrome with novel *de novo* SCN1A variants. *Brain Dev.*, **42**, 171–178.
74. Smith,C.W., Chu,T.T. and Nadal-Ginard,B. (1993) Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol. Cell. Biol.*, **13**, 4939–4952.
75. Raynal,C., Baux,D., Theze,C., Bareil,C., Taulan,C., Roux,A.-F., Claustres,M., Tuffery-Giraud,S. and des Georges,M. (2013) A classification model relative to splicing for variants of unknown clinical significance: application to the CFTR gene. *Hum. Mutat.*, **34**, 774–784.
76. Yan,Y. and Yang,Y. (2015) c.2381-3T>C mutation of DMD gene: a rare SNP without significant pathogenicity. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi*, **32**, 77–80.
77. Voith von Voithenberg,L., Sanchez-Rico,C., Kang,H.S., Madl,T., Zanier,K., Barth,A., Warner,L.R., Sattler,M. and Lamb,D.C. (2016) Recognition of the 3' splice site RNA by the U2AF heterodimer involves a dynamic population shift. *Proc. Natl Acad. Sci. USA*, **113**, E7169–E7175.
78. Yoshida,H., Park,S.-Y., Sakashita,G., Nariai,Y., Kuwasako,K., Muto,Y., Urano,T. and Obayashi,E. (2020) Elucidation of the aberrant 3' splice site selection by cancer-associated mutations on the U2AF1. *Nat. Commun.*, **11**, 4744.
79. Landrum,M.J., Lee,J.M., Riley,G.R., Jang,W., Rubinstein,W.S., Church,D.M. and Maglott,D.R. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.*, **42**, D980–D985.
80. Duncan,B.K. and Miller,J.H. (1980) Mutagenic deamination of cytosine residues in DNA. *Nature*, **287**, 560–561.
81. Denisov,S., Bazykin,G., Favorov,A., Mironov,A. and Gelfand,M. (2015) Correlated evolution of nucleotide positions within splice sites in mammals. *PLoS One*, **10**, e0144388.

82. Kriventseva, E.V., Koch, I., Apweiler, R., Vingron, M., Bork, P., Gelfand, M.S. and Sunyaev, S. (2003) Increase of functional diversity by alternative splicing. *Trends Genet.*, **19**, 124–128.
83. Romero, P.R., Zaidi, S., Fang, Y.Y., Uversky, V.N., Radivojac, P., Oldfield, C.J., Cortese, M.S., Sickmeier, M., LeGall, T., Obradovic, Z., et al. (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl Acad. Sci. USA*, **103**, 8390–8395.
84. Pentony, M.M. and Jones, D.T. (2009) Modularity of intrinsic disorder in the human proteome. *Proteins*, **78**, 212–221.
85. Hegyi, H., Kalmar, L., Horvath, T. and Tompa, P. (2011) Verification of alternative splicing variants based on domain integrity, truncation length and intrinsic protein disorder. *Nucleic Acids Res.*, **39**, 1208–1219.
86. Buljan, M., Chalancon, G., Dunker, A.K., Bateman, A., Balaji, S., Fuxreiter, M. and Babu, M.M. (2013) Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Curr. Opin. Struct. Biol.*, **23**, 443–450.
87. Campen, A., Williams, R.M., Brown, C.J., Meng, J., Uversky, V.N. and Dunker, A.K. (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett.*, **15**, 956–963.
88. Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sorensen, K.D., et al. (2014) A dual program for translation regulation in cellular proliferation and differentiation. *Cell*, **158**, 1281–1292.
89. Lu, C.H., Lin, Y.F., Lin, J.J. and Yu, C.S. (2012) Prediction of metal ion-binding sites in proteins using the fragment transformation method. *PLoS One*, **7**, e39252.
90. Xiao, X., Wang, Z., Jang, M. and Burge, C.B. (2007) Coevolutionary networks of splicing cis-regulatory elements. *Proc. Natl Acad. Sci. USA*, **104**, 18583–18588.
91. Coolidge, C.J., Seely, R.J. and Patton, J.G. (1997) Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.*, **25**, 888–896.
92. Choi, T.S. and Tezcan, F.A. (2022) Overcoming universal restrictions on metal selectivity by protein design. *Nature*, **603**, 522–527.
93. Zamore, P.D. and Green, M.R. (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl Acad. Sci. USA*, **86**, 9243–9247.
94. Singh, R., Valcarcel, J. and Green, M.R. (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science*, **268**, 1173–1176.
95. Roscigno, R.F., Weiner, M. and Garcia-Blanco, M.A. (1993) A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. *J. Biol. Chem.*, **268**, 11222–11229.
96. Glasser, E., Maji, D., Biancon, B., Puthenpeedikakkal, A.M., Cavender, C.E., Tebaldi, T., Jenkins, J.L., Mathews, D.H., Halene, S. and Kielkopf, C.L. (2022) Pre-mRNA splicing factor U2AF2 recognizes distinct conformations of nucleotide variants at the center of the pre-mRNA splice site signal. *Nucleic Acids Res.*, **50**, 5299–5312.
97. Jenkins, J.L., Agrawal, A.A., Gupta, A., Green, M.R. and Kielkopf, C.L. (2013) U2AF65 adapts to diverse pre-mRNA splice sites through conformational selection of specific and promiscuous RNA recognition motifs. *Nucleic Acids Res.*, **41**, 3859–3873.
98. Agrawal, A.A., McLaughlin, K.J., Jenkins, J.L. and Kielkopf, C.L. (2014) Structure-guided U2AF65 variant improves recognition and splicing of a defective pre-mRNA. *Proc. Natl Acad. Sci. USA*, **111**, 17420–17425.
99. Kralovicova, J., Sevcikova, I., Stejskalova, E., Obuca, M., Hiller, M., Stanek, D. and Vorechovsky, I. (2018) PUF60-activated exons uncover altered 3' splice-site selection by germline missense mutations in a single RRM. *Nucleic Acids Res.*, **46**, 6166–6187.
100. Chen, L., Weinmeister, R., Kralovicova, J., Eperon, I.P., Vorechovsky, I., Hudson, A.J. and Eperon, I.C. (2017) Stoichiometries of U2AF35, U2AF65 and U2 snRNP reveal new early spliceosome assembly pathways. *Nucleic Acids Res.*, **45**, 2051–2067.
101. Gao, K., Masuda, A., Matsuura, T. and Ohno, K. (2008) Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.*, **36**, 2257–2267.
102. Pineda, J.M.B. and Bradley, R.K. (2018) Most human introns are recognized via multiple and tissue-specific branchpoints. *Genes Dev.*, **32**, 577–591.
103. Hartmuth, K. and Barta, A. (1988) Unusual branch point selection in processing of human growth hormone pre-mRNA. *Mol. Cell. Biol.*, **8**, 2011–2020.
104. Chebli, K., Gattoni, R., Schmitt, P., Hildwein, G. and Stevenin, J. (1989) The 216-nucleotide intron of the E1A pre-mRNA contains a hairpin structure that permits utilization of unusually distant branch acceptors. *Mol. Cell. Biol.*, **9**, 4852–4861.
105. Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., et al. (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*, **478**, 64–69.
106. Fei, D.L., Motowski, H., Chatrikhi, R., Prasad, S., Yu, J., Gao, S., Kielkopf, C.L., Bradley, R.K. and Varmus, H. (2016) Wild-type U2AF1 antagonizes the splicing program characteristic of U2AF1-mutant tumors and is required for cell survival. *PLoS Genet.*, **12**, e1006384.
107. Esfahani, M.S., Lee, L.J., Jeon, Y.J., Flynn, R.A., Stehr, H., Hui, A.B., Ishisoko, N., Kildebeck, E., Newman, A.R., Bratman, S.V., et al. (2019) Functional significance of U2AF1 S34F mutations in lung adenocarcinomas. **10**, 5712.
108. Schirman, D., Yakhini, Z., Pilpel, Y. and Dahan, O. (2021) A broad analysis of splicing regulation in yeast using a large library of synthetic introns. *PLoS Genet.*, **17**, e1009805.
109. Romfo, C.M. and Wise, J.A. (1997) Both the polypyrimidine tract and the 3' splice site function prior to the first step of splicing in fission yeast. *Nucleic Acids Res.*, **25**, 4658–4665.
110. Zuo, P. and Maniatis, T. (1996) The splicing factor U2AF35 mediates critical protein–protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.*, **10**, 1356–1368.
111. Tsujikawa, K., Tsujikawa, M., Yamamoto, S., Fujikado, T. and Tano, Y. (2002) Allelic homogeneity due to a founder mutation in Japanese patients with lattice corneal dystrophy type IIIA. *Am. J. Med. Genet.*, **113**, 20–22.
112. Menendez, M., Castellví-Bel, S., Pineda, M., de Cid, R., Munoz, J., Gonzalez, S., Teule, A., Balaguer, F., Ramon y Cajal, T., Maria, E.J., et al. (2010) Founder effect of a pathogenic *MSH2* mutation identified in Spanish families with Lynch syndrome. *Clin. Genet.*, **78**, 186–190.
113. Dupont, C.L., Butcher, A., Valas, R.E., Bourne, P.E. and Caetano-Anolles, G. (2010) History of biological metal utilization inferred through phylogenomic analysis of protein structures. *Proc. Natl Acad. Sci. USA*, **107**, 10567–10572.
114. Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T.A. and Soreq, H. (2005) Function of alternative splicing. *Gene*, **344**, 1–20.
115. Lopez, V. and Kelleher, S.L. (2009) Zinc transporter-2 (ZnT2) variants are localized to distinct subcellular compartments and functionally transport zinc. *Biochem. J.*, **422**, 43–52.
116. Armbruster, U., Leonelli, L., Galvis, V.C., Strand, D., Quinn, E.H., Jonikas, M.C. and Niyogi, K.K. (2016) Regulation and levels of the thylakoid K⁺/H⁺ antiporter KEA3 shape the dynamic response of photosynthesis in fluctuating light. *Plant Cell Physiol.*, **57**, 1557–1567.
117. Vecellio Reane, D., Vallese, F., Checchetto, V., Acquasaliente, L., Butera, G., De Filippis, V., Szabo, I., Zanotti, G., Rizzuto, R. and Raffaello, A. (2016) A MICU1 splice variant confers high

- sensitivity to the mitochondrial Ca(2+) uptake machinery of skeletal muscle. *Mol. Cell*, **64**, 760–773.
118. Daithankar,V.N., Farrell,S.R. and Thorpe,C. (2009) Augmenter of liver regeneration: substrate specificity of a flavin-dependent oxidoreductase from the mitochondrial intermembrane space. *Biochemistry*, **48**, 4828–4837.
119. Wiley,S.E., Murphy,A.N., Ross,S.A., van der Geer,P. and Dixon,J.E. (2007) MitoNEET is an iron-containing outer mitochondrial membrane protein that regulates oxidative capacity. *Proc. Natl Acad. Sci. USA*, **104**, 5318–5323.
120. Conlan,A.R., Axelrod,H.L., Cohen,A.E., Abresch,E.C., Zuris,J., Yee,D., Neschushtai,R., Jennings,P.A. and Paddock,M.L. (2009) Crystal structure of Miner1: the redox active 2Fe–2S protein causative in Wolfram syndrome 2. *J. Mol. Biol.*, **392**, 143–153.
121. Gilston,B.A., Skaar,E.P. and Chazin,W.J. (2016) Binding of transition metals to S100 proteins. *Sci. China Life Sci.*, **59**, 792–801.
122. Schafer,B.W., Fritschy,J.-M., Murmann,P., Troxler,H., Durusse,I., Heizmann,C.W. and Cox,J.A. (2000) Brain S100A5 is a novel calcium-, zinc-, and copper ion-binding protein of the EF-hand superfamily. *J. Biol. Chem.*, **275**, 30623–30630.
123. Tsvetkov,P.O., Roman,A.Y., Baksheeva,V.E., Nazipova,A.A., Shevelyova,M.P., Vladimirov,V.I., Buyanova,M.F., Zinchenko,D.V., Zamyatin,A.A. Jr, Devred,F, *et al.* (2018) Functional status of neuronal calcium sensor-1 is modulated by zinc binding. *Front. Mol. Neurosci.*, **11**, 459.
124. Beyersmann,D. and Haase,H. (2001) Functions of zinc in signaling, proliferation and differentiation of mammalian cells. *Biometals*, **14**, 331–341.
125. Wilder,P.T., Baldisseri,D.M., Udan,R., Valley,K.M. and Weber,D.J. (2003) Location of the Zn(2+)-binding site on S100B as determined by NMR spectroscopy and site-directed mutagenesis. *Biochemistry*, **42**, 13410–13421.
126. Baudier,J., Deloulme,J.C. and Shaw,G.S. (2020) The Zn²⁺ and Ca²⁺-binding S100B and S100A1 proteins: beyond the myths. *Biol. Rev. Camb. Philos. Soc.*, **95**, 738–758.
127. Decaria,L., Bertini,I. and Williams,R.J.P. (2010) Zinc proteomes, phylogenetics and evolution. *Metallomics*, **2**, 706–709.
128. Kocyla,A., Ba Tran,J. and Krezel,A. (2021) Galvanization of protein–protein interactions in a dynamic zinc interactome. *Trends Biochem. Sci.*, **46**, 64–79.
129. Thompson,M.W. (2022) Regulation of zinc-dependent enzymes by metal carrier proteins. *Biometals*, **35**, 187–213.
130. Tottey,S., Waldron,K.J., Firbank,S.J., Reale,B., Bessant,C., Sato,K., Cheek,T.R., Gray,J., Banfield,M.J., Dennison,C., *et al.* (2008) Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature*, **455**, 1138–1142.
131. Osman,D. and Robinson,N.J. (2022) Protein metalation in a nutshell. *FEBS Lett.*, **597**, 141–150.
132. Imlay,J.A. (2014) The mismetallation of enzymes during oxidative stress. *J. Biol. Chem.*, **289**, 28121–28128.
133. Mingos,D.M.P. (2019) In: *The Periodic Table I*. Springer Nature, Cham, Switzerland.
134. Mulikdjanian,A.Y. (2009) On the origin of life in the zinc world: 1. Photosynthesizing, porous edifices built of hydrothermally precipitated zinc sulfide as cradles of life on Earth. *Biol. Direct*, **4**, 26.
135. Saito,M.A., Sigman,D.M. and Morel,F.M.M. (2003) The bioinorganic chemistry of the ancient ocean: the co-evolution of cyanobacterial metal requirements and biogeochemical cycles at the Archean–Proterozoic boundary? *Inorg. Chim. Acta*, **356**, 308–318.
136. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Zidek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.