

# Copper-binding proteins and exonic splicing enhancers and silencers

Dara Bakhtiar and Igor Vorechovsky 

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

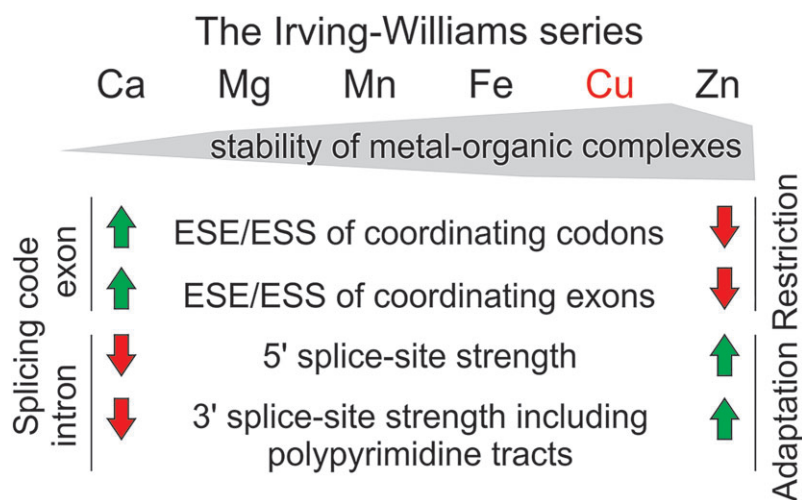
\*Correspondence: University of Southampton, Faculty of Medicine, Southampton SO16 6YD, UK. E-mail: [igvo@soton.ac.uk](mailto:igvo@soton.ac.uk)

## Abstract

Eukaryotic DNA codes not only for proteins but contains a wealth of information required for accurate splicing of messenger RNA precursors and inclusion of constitutively or alternatively spliced exons in mature transcripts. This “auxiliary” splicing code has been characterized as exonic splicing enhancers and silencers (ESE and ESS). The exact interplay between protein and splicing codes is, however, poorly understood. Here, we show that exons encoding copper-coordinating amino acids in human cuproproteins lack ESEs and/or have an excess of ESSs, yet RNA sequencing and expressed sequence tags data show that they are more efficiently included in mature transcripts by the splicing machinery than average exons. Their largely constitutive inclusion in messenger RNA is facilitated by stronger splice sites, including polypyrimidine tracts, consistent with an important role of the surrounding intron architecture in ensuring high expression of metal-binding residues during evolution. ESE/ESS profiles of codons and entire exons that code for copper-coordinating residues were very similar to those encoding residues that coordinate zinc but markedly different from those that coordinate calcium. Together, these results reveal how the traditional and auxiliary splicing motifs responded to constraints of metal coordination in proteins.

**Keywords:** metalloproteins, copper,  $Zn^{2+}$ ,  $Ca^{2+}$ , the Irving–Williams series, alternative splicing

## Graphical abstract



Differential response of traditional and auxiliary splicing motifs to weak and competitive divalent metals.

## Introduction

Evolution of life on Earth involved a complex mixture of organic and inorganic compounds.<sup>1</sup> Unlike enzymatically synthesized and degraded organic compounds, inorganic elements, once solubilized from the Earth's crust, are neither created nor destroyed, and must be under strict regulation to avoid metal toxicity and binding to organic molecules by noncognate metals (mismetallation).<sup>1</sup> This regulation evolved over billions of years

of life evolution, but is poorly understood, particularly in eukaryotic organisms with discrete and highly complex gene expression pathways.

Metal-binding complements of eukaryotic proteomes, or metalloproteins, require metal ions to assist catalysis or impart structure.<sup>2</sup> In complex organisms, metalloproteins account for a large proportion of gene products,<sup>3–6</sup> suggesting that evolution of metal interactions shaped not only protein coding but also regu-

Received: February 20, 2024. Accepted: April 21, 2024

© The Author(s) 2024. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

latory components in eukaryotic DNA. For example, when considering splicing of precursor messenger RNA (pre-mRNA), the most complex gene expression step, human exons encoding coordinating residues for calcium ( $\text{Ca}^{2+}$ ) are richer in splicing enhancers as compared to exons that code for residues coordinating zinc ( $\text{Zn}^{2+}$ ).<sup>7</sup> However, the lower capacity of “ $\text{Zn}^{2+}$  exons” to be included in mature transcripts can be compensated by intrinsically stronger splice sites, including stronger polypyrimidine tracts,<sup>7</sup> key recognition motifs of splice acceptor sites of higher eukaryotes.<sup>8,9</sup> Both exon groups showed higher than average exon inclusion levels in mature transcripts and a paucity of alternative splicing, ensuring efficient inclusion of codons for residues coordinating  $\text{Ca}^{2+}$  and  $\text{Zn}^{2+}$  in mature transcripts and their expression at the protein level.<sup>7</sup> Although these data strongly suggest that the cells evolved to alleviate constraints imposed by metal coordination, responses of exon-intron architecture to coding restrictions imposed by other metals remain unknown.

Metalloproteins *in vitro* coordinate divalent metal ions with affinities largely reflective of the Irving–Williams series ( $\text{Mn}^{2+} < \text{Fe}^{2+} < \text{Co}^{2+} < \text{Ni}^{2+} < \text{Cu}^{2+} \geq \text{Zn}^{2+}$ ). The series was first proposed by Harry Irving and Robert Williams who observed that the stability of complexes formed by divalent first-row transition metal ions generally increased across the period to a maximum stability at copper.<sup>10–13</sup> The series was later proposed to include additional metals such as divalent s-block elements represented by weak alkali earth metals and was suggested to be an all-embracing principle for the control of metal ions in biology, balancing ion affinity and availability in the cell and predicting free metal concentrations for identical ligands.<sup>11,12</sup> The series trend appears to operate within proteins as well as within small molecule chelators, contributing to their metal ion selectivity,<sup>14</sup> but the extent of the Irving–Williams series universality is not completely understood.

Copper in cuproproteins is largely coordinated by the side chains of histidine, cysteine, and methionine<sup>15</sup> and the frequencies of coordinating residues resemble those established for zinc.<sup>16</sup> Because copper is the most competitive metal in the Irving–Williams series, human exons encoding Cu-coordinating residues would be expected to possess splicing silencer and enhancer profiles that are more similar to  $\text{Zn}^{2+}$  than to weaker metals, such as  $\text{Mn}^{2+}$  or alkali earth metals. However, it is not known if these residue preferences had any bearing on the “invisible” splicing code in exons and exonic splicing enhancers (ESE) vs. exonic splicing silencers (ESS) profiles during evolution. With the increasing number of cuproproteins described to date,<sup>1,17</sup> we set out to test this question by comparing splicing enhancers and silencers in exons that encode Cu-coordinating residues with control exons and exons encoding binding sites for other metals. We have also determined their average inclusion levels in mature transcripts in human tissues and compared them to controls.

## Materials and methods

### Ascertainment of human exons that encode Cu-binding sites

We compiled (i) human proteins that were reported to bind copper, as supported by X-ray crystallography/solution nuclear magnetic resonance spectroscopy structures/electron microscopy and had Protein Data Bank (PDB) records (resolution lower than 3.6 Å). Residues were considered to coordinate copper if they were annotated in the PDB metal coordination tabs or specified as copper coordinating by the submitters in the supporting literature (Supplementary Table S1). Our sample of cuproproteins also included (ii) human proteins conserved between prokary-

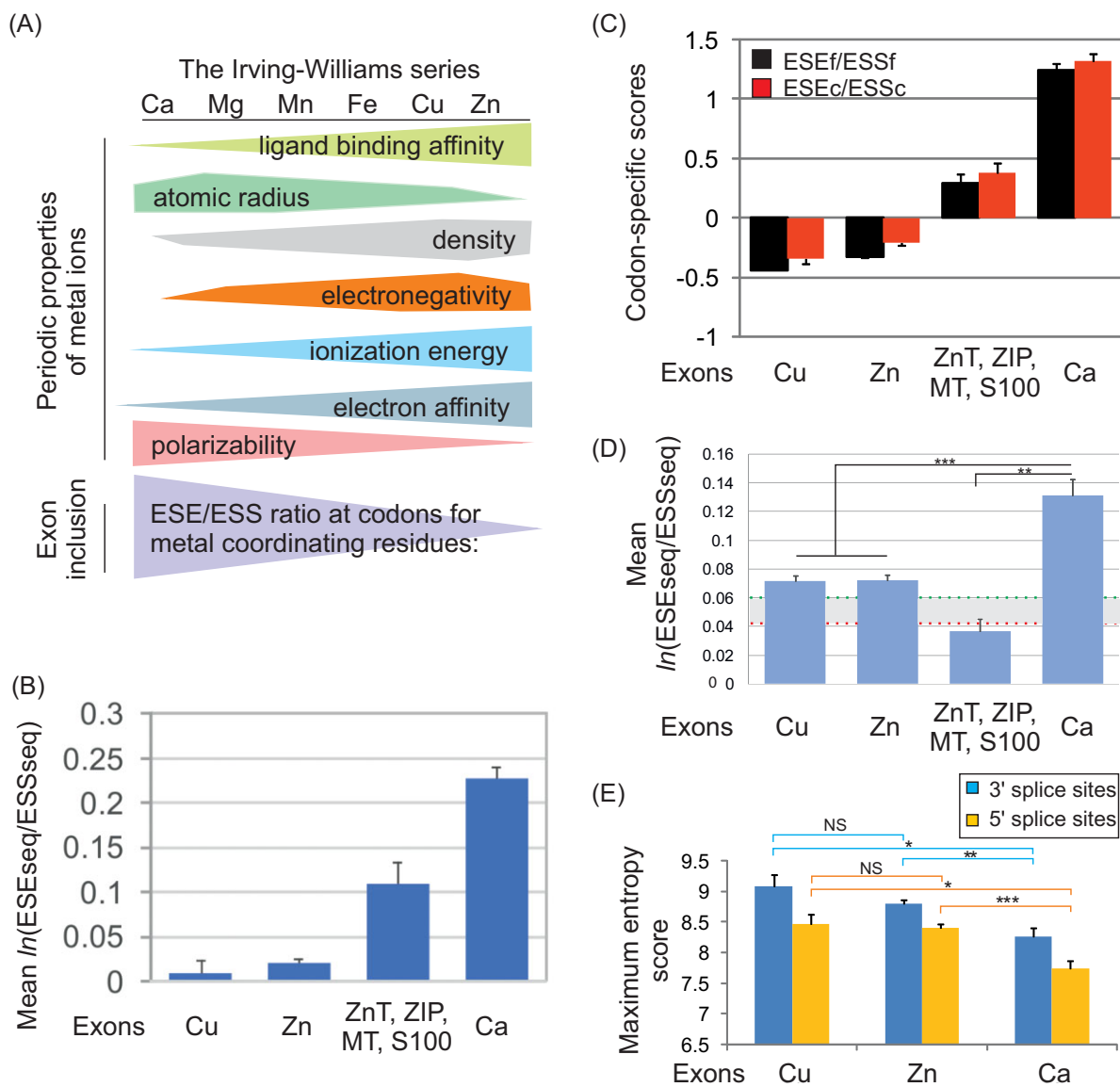
otes and humans, fungi and humans, or animals and humans if the copper-coordinating residues in PDB in lower organisms aligned with the same residue in human counterparts or with residues previously shown to coordinate this metal (Cys, His, Met, Tyr, Glu, Gly),<sup>15</sup> with a few exceptions (ATP7B, TYRP1, COMMD1, CTR1, LOXL2, XIAP; Supplementary Table S1).<sup>18</sup> Incomplete coordination spheres were included to maximize the codon pool size. Apart from PDB, the list of Cu-binding proteins was cross-checked against MetalPDB,<sup>19</sup> Uniprot,<sup>20</sup> and a previously published list of cuproproteins.<sup>17</sup> Human gene symbols for the selected proteins were then matched to the Human Gene Nomenclature and Genecards records. The search yielded a set of 57 human proteins containing 491 unique codons for Cu-coordinating residues (December 2023; Supplementary Table S1). Finally, exons and adjacent introns encoding Cu-binding residues were retrieved from Ensembl<sup>21</sup> and codons that encode Cu-coordinating residues (highlighted in red in Supplementary Table S1) and were used for ESE/ESS profiling as described.<sup>7,22</sup>

### Splicing enhancers and silencers for codons that encode Cu-coordinating residues

To characterize the auxiliary splicing code underlying Cu-binding sites in proteins, we employed a comprehensive set of human ESE and ESS hexamers derived by splicing promotion or repression *ex vivo*, which was measured for 4096 synthetic oligomers inserted into model WT1 and HBB exons at five different positions.<sup>23</sup> The resulting minigene libraries were used to acquire hexamer ESEseq and ESSseq scores defined previously.<sup>23</sup> The scores were obtained independently of the presence or absence of metal-binding sites encoded by the test exons and provide good overall estimates of exon inclusion activities mediated by exon themselves.<sup>23</sup> To assign ESEseq and ESSseq scores to Cu-coordinating codons, we computed average scores for overlapping hexamers using custom Microsoft Excel functions/formulas as described.<sup>7</sup> To estimate codon-specific splicing activities, we also calculated frequency ratios for a total of 4728 ESE codons and 4360 ESS codons, presented here as  $\ln(\text{ESEf}/\text{ESSf})$ , and we also determined codon counts in 1182 high-confidence ESEs and 1090 high-confidence ESSs<sup>23</sup> to compute the ESEc/ESSc ratios, as previously defined.<sup>7</sup> 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.<sup>7</sup> Apart from codons, mean ESEseq/ESSseq, ESEf/ESSf, and ESEc/ESSc values were also computed for entire exons that encode all Cu-coordinating residues. Exonic sequences were devoid of the first and the last three nucleotides (nts) since these exonic positions shape the 3'ss and 5'ss consensus, respectively.<sup>24</sup>

### Characterization of average mRNA inclusion levels

To determine PSI (% spliced in) values<sup>25</sup> for tested exons, we employed PSI tables (hg38) from the Vertebrate Alternative Splicing and Transcription Database (VastDB), which provides comprehensive PSI values across vertebrate exons in various tissues and developmental stages.<sup>26</sup> In addition, we compared PSI values of tested and control exons using expressed sequence tag (EST) data in HEXEvent, a database of Human EXon splicing Events, which stores EST-derived inclusion levels for ~200 000 exons.<sup>27</sup> HEXEvent shows additional alternative splicing information for human internal exons but unlike VastDB, HEXEvent does not include intron retention events, avoiding a bias toward short retained introns.<sup>27,28</sup>



**Fig. 1** Splicing enhancers and silencers in exons that encode coordinating residues for copper, zinc, and calcium. (A) The Irving-Williams series and the auxiliary splicing code in exons. Upper panel shows selected properties of six abundant and biologically important divalent metals in the human body and their nonlinear 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; and polarizability, tendency to acquire electric dipole moment in electric fields.<sup>40</sup> Lower panel shows expected ESE/ESS profiles for codons encoding coordinating residues for the indicated metal in the series.<sup>7,22,41</sup> (B, C) Estimates of exonic splicing enhancer and silencer strength for codons that encode coordinating residues for human cuproproteins ( $n = 57$ ), zinc-binding proteins ( $n = 427$ , of which 293 were ZFs) excluding Zrt/Irt-like proteins, zinc transporters, metallothioneins and S100 ( $n = 55$ ), calcium-binding EF-hand proteins ( $n = 189$ ) and exon controls (horizontal bar) using ESEseq/ESSseq ratios (B) or codon-specific scores (C). The ESEseq/ESSseq values for controls were determined as previously described.<sup>7,22</sup> (D) The mean ESEseq/ESSseq ratios in entire exons that encode coordinating residues for the same metals in the same protein groups.  $**P < 0.001$ ,  $***P < 0.0001$ . Horizontal bar denotes the range (maximum in green, minimum in red) of mean values for various control exon groups.<sup>7,22</sup> (E) The intrinsic splice site strength of Cu, Zn, and Ca exons.  $*P < 0.01$ ,  $**P < 0.001$ ,  $***P < 0.0001$  (unpaired t-tests corrected for multiple comparisons). The sample of Cu exons had 146 and 153 unique 5' and 3' splice sites, respectively. Error bars are SEMs. Data for Zn and Ca exons were taken from our previous study.<sup>7,22</sup>

## Results

### Auxiliary splicing code in exons that code for Cu-coordinating residues

To characterize ESE/ESS profiles of codons/exons that encode copper-binding function, we first obtained amino acid sequences surrounding copper-binding residues in 57 human cuproproteins (Supplementary Table S1). They were encoded by a total of 170 unique exons that contained 491 different codons for Cu-

coordinating amino acids. Protein Cu-binding sites were dominated by His, Cys, and Met, in accordance with previous studies.<sup>15</sup>

We first assigned hexamer ESEseq and ESSseq scores to each codon and compared them to controls and those previously ascertained for human calcium- and zinc-binding proteins.<sup>7,22</sup> For simplicity, we term these exons as Cu, Ca, and Zn exons. Figure 1B shows that ESEseq/ESSseq score ratios for Cu-coordinating codons resembled more Zn exons than Ca exons, as expected (see Introduction). The similarity was most pronounced

for zinc finger proteins (ZFs) whereas other Zn-binding proteins, particularly those coordinated by acidic residues, showed higher values, albeit not reaching levels observed for enhancer-rich Ca codons (Fig. 1B and Bakhtiar *et al.*<sup>7</sup>) A similar pattern was found for codon-specific ESEf/ESSf and ESEc/ESSc scores (Fig. 1C). Taking into account extensions of codons close to 3' or 5' splice sites to adjacent sequences (~6% of the total codon count, Supplementary Table S1) yielded very similar values and the same order of tested metal exons.

We then asked whether the observed splicing dichotomy for two tight and one weak divalent metals of the Irving–Williams series can be seen at the level of entire exons. Figure 1D shows a comparison of mean ESEseq/ESSseq ratios for full exons that encode Zn-, Ca-, and Cu-coordinating residues and for control exons. Importantly, the capacity of ESE and ESS to include pre-mRNA segments encoding Cu-binding sites in mature transcripts was similar for Zn exons but significantly lower than for Ca exons, despite a smaller sample size.

Finally, we compared the intrinsic splice site strength of the studied metal exons (Fig. 1E) by computing maximum entropy scores for their 3' and 5' splice sites. We found that both 3' and 5' splice sites of Cu exons were similar to Zn exons, but significantly stronger than for the ESE-rich Ca exons. This was reflected by a higher number of uridines in the input sequence for maximum entropy scoring of 3' splice sites [ $7.94 \pm 0.32$  ( $\pm$ SEM) for Cu exons,  $8.16 \pm 0.10$  for Zn exons and  $7.73 \pm 0.18$  for Ca exons], which provide rough estimates of the strength of polypyrimidine tracts.

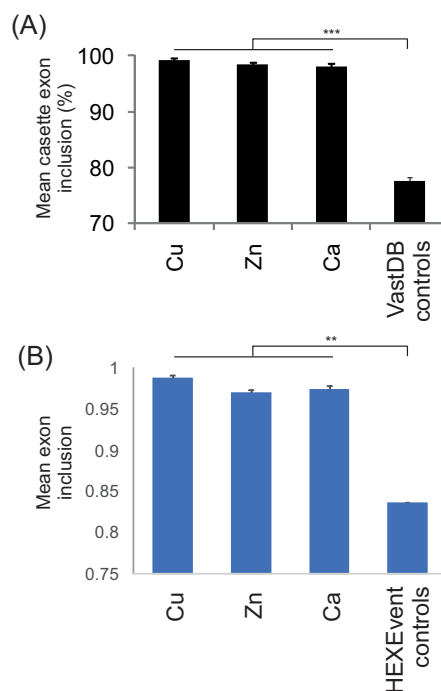
Taken together, these results indicate the auxiliary splicing code of Cu exons is similar to Zn exons, but markedly weaker than for Ca exons as if recognition of exons that encode coordinating residues for tight metals in the Irving–Williams series (Zn, Cu) was selected against whereas that of exons encoding binding sites for weak Ca was encouraged in the cells to avoid mismetalation. They also show that this dichotomy was compensated by both 5' and 3' splice sites and their polypyrimidine tracts.

### Exon inclusion in mature transcripts and Cu binding in proteins

Next, we explored the question of whether the lower ESEseq/ESSseq profiles of Cu exons were reflected in lower exon inclusion levels *in vivo*; in other words, to what extent the observed dichotomous ESE/ESS profiles may influence their alternative splicing in the cell. We addressed the question by comparing average PSI values for Cu exons with control human exons and other groups of metal-coordinating exons previously analysed.<sup>7,22</sup> Comparison of PSI values from the VastDB RNA-seq for Cu exons and control cassette exons showed that, similar to other exons that encode residues coordinating Zn and Ca,<sup>7,22</sup> Cu exons had significantly higher exon inclusion levels (Fig. 2A). A similar difference was seen with HEXEvent EST data (Fig. 2B). The overall paucity of their alternative splicing was observed across multiple tissues (Fig. 3). We identified only a few Cu exons that showed PSI values below 80% in multiple tissues or repeated samples from the same tissue (Table 1). Although physiological roles of the potentially regulated splicing events and associated transcripts will require experimental verification, the overall high inclusion levels of Cu, Ca, and Zn exons (Fig. 2) are consistent with their critical role in metalloprotein expression.

### Discussion

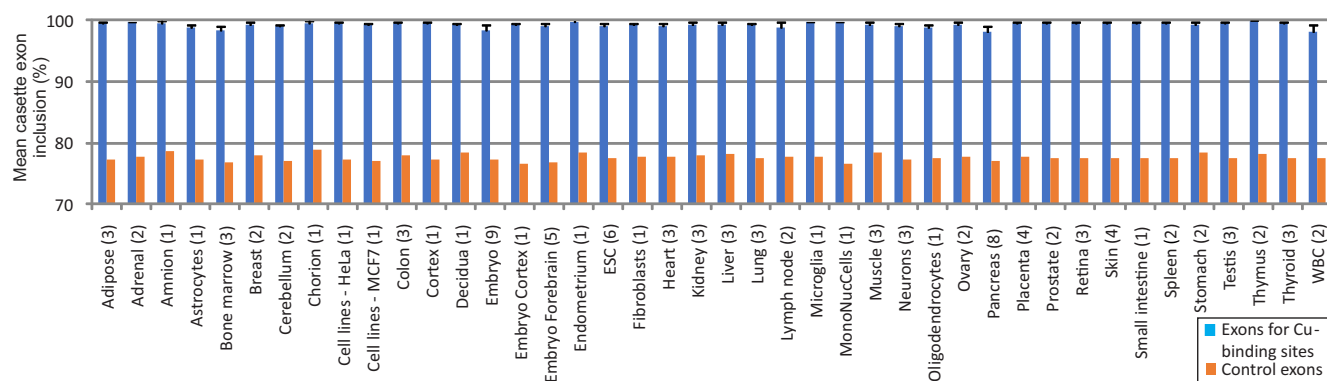
This study is the first to evaluate how copper-binding constraints in proteins influence the most complex gene expression step in



**Fig. 2** Mean *in vivo* inclusion of human exons that encode metal-coordinating residues in human proteins. (A) VastDB PSI values. (B) HEXEvent EST data. Error bars denote SEM, \*\* $P < 0.001$ , \*\*\* $P < 0.0001$ .

eukaryotes. Eukaryotic exons are recognized by the largest RNA-protein complex in the cell with a single-nucleotide accuracy in the sea of much longer introns.<sup>29</sup> To achieve this spectacular precision, the complex not only recognizes short motifs at splice sites but heavily relies on widespread ESEs, ESSs and additional intronic sequences in the pre-mRNA.<sup>29</sup> Our study uncovers important differences in ESE/ESS profiles between Cu, Zn, and Ca exons, raising a speculation that the splicing machinery has not been immune to inorganic components of complex eukaryotes. However, coding restrictions that favor Zn and Cu binding to His/Cys codons and Ca binding to acidic residues could not be overcome by the “invisible” splicing code in exons. Their ESE/ESS profiles per se do not explain inclusion levels measured *in vivo*. Notably, the average density of coordinating codons in Cu, Zn, and Ca exons was not identical, with a hierarchy of  $Ca > Zn > Cu$  exons (Fig. 4), which is reminiscent of the Irving–Williams order. However, their density cannot explain ESE/ESS profiles. The correlation between codon counts and ESEseq/ESSseq ratios was poor if any [ $r = 0.14$  ( $P$ -value = 0.03), 0.02 ( $P = 0.3$ ) and 0.15 ( $P = 0.07$ ) for Ca, Zn and Cu exons, respectively], consistent with only a marginal density contribution of coordinating residues in Ca exons coding for canonical EF hands.

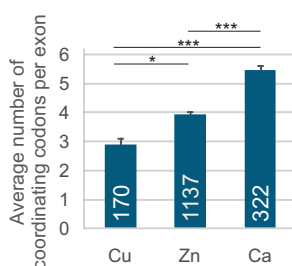
Apart from exonic sequences, the constitutive inclusion of exons coding for weak and tight metals in mature transcripts (Fig. 3) is likely to require compensation by intronic splicing motifs, including stronger splice sites and polypyrimidine tracts.<sup>7</sup> The tracts are highly variable traditional splicing recognition motifs that serve as more potent signals than ESEs/ESSs.<sup>8,9</sup> Selecting mutated introns adjacent to Cu exons during evolution would be easier for the cell than selecting mutations around metal-binding residues, which may not be compatible with accurate splicing. It has become increasingly apparent that the cells face an incessant conflict between ensuring splicing efficiency and preserving the coding capacity for optimal proteins and their interactions.<sup>30,31</sup> The fraction of synonymous mutations that affect RNA processing can be very high although it varies from exon to exon.<sup>30,31</sup>



**Fig. 3** Average inclusion levels of Cu exons and VastDB controls in mature transcripts across multiple tissues. Number of independent tissue samples is shown in parentheses.

**Table 1.** Selection of VastDB exons encoding copper-coordinating residues with low PSI values (<80%)

Cuproprotein	Gene symbol	Binding residue	VastDB exon ID	Exon size	Tissues
Copper-transporting ATPase $\beta$	ATP7B	C575	HsaEX0007040	162	Astrocytes, neurons, chorion
Hephaestin	HEPH	H965	HsaEX0029693	226	Embryonal forebrain, placenta, pancreas, ovary, spleen
Synthesis of cytochrome C oxidase 1	SCO1	C169	HsaEX0056613	198	Embryo
Laccase domain containing 1	LACC1	G416	HsaEX0009320	161	Embryo, forebrain, kidney, liver
Coeruloplasmin	CP	H656	HsaEX0016927	213	ESCs
Amyloid $\beta$ precursor like protein 1	APLP1	E362	HsaEX0005147	159	Heart, liver
Lysyl oxidase homolog 2	LOXL2	H626	HsaEX0036151	244	Pancreas



**Fig. 4** Mean number of codons for coordinating residues in Ca, Zn, and Cu exons. Number of exons is in white. Error bars denote SEM. Ca, codons coordinating Ca in binding loops of canonical EF-hand proteins. \* $P < 0.01$ , \*\*\* $P < 0.0001$ , corrected unpaired t-test.

Apart from chemical properties of metals themselves, biological utilization of metals has been influenced by geochemical pressures.<sup>1,15,32</sup> Prokaryotes, particularly anaerobic bacteria and archaea, have adopted a very limited role for Cu, focusing largely on protection from Cu through sensing, intracellular mobilization and export.<sup>1</sup> In contrast, the increasing mean intron per gene density in single-cell (0.1~1.5) and multi-cellular (up to 8 in mammals<sup>33</sup>) organisms has been associated with Cu delivery to expanding arrays of metalloproteins located within various organelles that facilitate tissue-specific regulation, indicative of tight control of Cu homeostasis both within and outside the cell.<sup>1</sup> For example, the high-affinity Cu importer Ctr1 in *Saccharomyces cerevisiae*, which lacks introns, has evolved into a primary transporter SLC31A1 (4 introns in zebrafish to humans) that is responsible for dietary Cu uptake in the intestinal epithelium.

Studying evolution of introns surrounding metal-coordinating exons in more detail should provide deeper insights into the interplay of the “metal code” and traditional and auxiliary splicing signals.

Phylogenetic studies suggested that the protein folds for binding Mn and Fe evolved prior to the global oxygenation processes whereas the folds for Cu and Zn evolved after.<sup>34</sup> Many cuproproteins have oxygen-related function, such as catalyzing redox reactions or dioxygen transport.<sup>15</sup> The percentage of metal-binding sites within a proteome remains relatively stable over time, implying that the inclusion of a new metal is usually at the expense of another.<sup>15,34</sup> It is still unclear, however, how released competitive ions (Cu and Zn) can target specific proteins in signaling pathways, such as Cu binding to ZFs that can alter their ZF function.<sup>32,35–38</sup>

Unlike Ca and Zn, Cu exists in oxidized ( $\text{Cu}^{2+}$ ) and reduced ( $\text{Cu}^+$ ) states in the cell. Free Cu is extremely limited within cells: the metal is highly toxic, particularly because of a high reactivity of  $\text{Cu}^{2+}$  and generation of OH radicals.<sup>32</sup> Soluble Cu proved hazardous to early life, possibly by replacing other metal cofactors in proteins.<sup>15</sup> As with oxygen, copper would eventually transform from poison to a necessity, with sophisticated networks for acquiring, transporting, sequestering, and exporting this metal.<sup>15</sup> Copper proteins can traffic the metal or utilize it as a cofactor.<sup>15</sup> The latter tend to use high-affinity sites of high coordination numbers that help prevent loss of the metal during redox reactions. The former, on the other hand, tend to use low-affinity sites with/or moderate coordination numbers.<sup>15</sup> It was hypothesized that various combination of Cu-binding residues (His, Cys, Met) can afford dynamic multifunctional domains that can facilitate Cu transfer

in various intra- and extracellular environments.<sup>15</sup> For example, Cys may be more suitable for Cu accumulation in hypoxic environments where the metal is scarce whereas Met is well suited for Cu accumulation in acidic, oxidative milieu, with His preferred in neutral environments where Met-rich coordination spheres may not be sufficiently tight.<sup>15</sup> Such preferences are likely to be reflected in ESE/ESS profiles that might in turn influence combinatorial diversity provided by the three ligands, which can offer a choice of neutral vs. charged environment, hydrophobic vs. hydrophilic, nitrogen vs. sulfur, and susceptibility to pH changes or oxidation.<sup>15</sup>

The number of cuproproteins in our study was comparable to a recent count,<sup>17</sup> but their final number in human cells is not yet known. The sample size of Cu exons had a smaller statistical power compared to Zn or Ca exons, preventing us to address intronic contributions in more detail or compare the three abundant metals, as accomplished for Zn<sup>2+</sup> and Ca<sup>2+</sup>.<sup>7</sup> The excess of Cu<sup>2+</sup> over Cu<sup>+</sup> in our study was by a factor of ~3, hampering detailed ESE/ESS comparisons of coordination spheres for cupric and cuprous ions, which may vary with oxidation states. It is not yet known which alternative splicing events control relatively high copper concentrations in copper storage vesicles that have been identified in the brain and other cell types<sup>39</sup> and how they are regulated at the RNA level. Not in the least, the identification of physiological roles of metals remains to be technically challenging, and has been hindered by examples of ion dissociation during protein or nucleic acid purification, limited detection in structural models, the impact of overexpressing metal-binding molecules on metal availability, and metal interchangeability.<sup>32</sup> Some of these questions may attract new investigators in the emerging field of alternative RNA processing in metal signaling.

## Acknowledgement

We wish to thank Dr Reuben Pengelly for technical help.

## Author contributions

Dara Bakhtiar—methodology, investigation, formal analysis, and writing original draft; Igor Vorechovsky—conceptualization, methodology, investigation, formal analysis, writing original draft, and supervision.

## Supplementary data

Supplementary data are available at *Metallomics* online.

## Conflict of interest

None declared.

## Data availability

The data underlying this article are available in the article in its online supplementary material.

## References

- R. A. Festa and D. J. Thiele. Copper: an essential metal in biology, *Curr. Biol.*, 2011, 21(21), R877–RR83. <https://doi.org/10.1016/j.cub.2011.09.040>
- A. W. Foster, T. R. Young, P. T. Chivers and N. J. Robinson. Protein metalation in biology, *Curr. Opin. Chem. Biol.*, 2022, 66, 102095. <https://doi.org/10.1016/j.cbpa.2021.102095>
- R. H. Holm, P. Kennepohl and E. I. Solomon. Structural and functional aspects of metal sites in biology, *Chem. Rev.*, 1996, 96(7), 2239–2314. <https://doi.org/10.1021/cr9500390>
- K. J. Waldron, J. C. Rutherford, D. Ford and N. J. Robinson. Metalloproteins and metal sensing, *Nature*, 2009, 460(7257), 823–830. <https://doi.org/10.1038/nature08300>
- C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday and J. M. Thornton. Metal ions in biological catalysis: from enzyme databases to general principles, *J. Biol. Inorg. Chem.*, 2008, 13(8), 1205–1218. <https://doi.org/10.1007/s00775-008-0404-5>
- S. Clemens. The cell biology of zinc, *J. Exp. Bot.*, 2022, 73(6), 1688–1698. <https://doi.org/10.1093/jxb/erab481>
- D. Bakhtiar, K. Vondraskova, R. Pengelly, M. Chivers, J. Kralovicova and I. Vorechovsky. Exonic splicing code and coordination of divalent metals in proteins, *Nucleic Acids Res.*, 2024, 52(3), 1090–1106. <https://doi.org/10.1093/nar/gkad116/>
- R. F. Roscigno, M. Weiner and M. A. Garcia-Blanco. A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing, *J. Biol. Chem.*, 1993, 268(15), 11222–11229. [https://doi.org/10.1016/S0021-9258\(18\)82114-7](https://doi.org/10.1016/S0021-9258(18)82114-7)
- C. J. Coolidge, R. J. Seely and J. G. Patton. Functional analysis of the polypyrimidine tract in pre-mRNA splicing, *Nucleic Acids Res.*, 1997, 25(4), 888–896. <https://doi.org/10.1093/nar/25.4.888>
- H. Irving and R. J. P. Williams. Order of stability of metal complexes, *Nature*, 1948, 162(4123), 746–747. <https://doi.org/10.1038/162746a0>
- J. J. R. Frausto da Silva and R. J. P. Williams, *The Biological Chemistry of the Elements*, 2nd edn, Oxford University Press, Oxford, 2001.
- R. J. P. Williams and J. J. Fausto da Silva, *The Chemistry of Evolution: The Development of Our Ecosystem*, Elsevier, Amsterdam, 2006.
- A. W. Foster, D. Osman and N. J. Robinson. Metal preferences and metallation, *J. Biol. Chem.*, 2014, 289(41), 28095–28103. <https://doi.org/10.1074/jbc.R114.588145>
- K. A. McCall and C. A. Fierke. Probing determinants of the metal ion selectivity in carbonic anhydrase using mutagenesis, *Biochemistry*, 2004, 43(13), 3979–3986. <https://doi.org/10.1021/bi0498914>
- J. T. Rubino and K. J. Franz. Coordination chemistry of copper proteins: how nature handles toxic cargo for essential function, *J. Inorg. Biochem.*, 2012, 107(1), 129–143. <https://doi.org/10.1016/j.jinorgbio.2011.11.024>
- C. H. Lu, Y. F. Lin, J. J. Lin and C. S. Yu. Prediction of metal ion-binding sites in proteins using the fragment transformation method, *PLoS One*, 2012, 7(6), e39252. <https://doi.org/10.1371/journal.pone.0039252>
- P. Wittung-Stafshede. Another pearl in the “copper-transport” necklace, *Biophys. J.*, 2021, 120(20), 4305–4306. <https://doi.org/10.1016/j.bpj.2021.08.030>
- T. E. Adams, M. F. Hockin, K. G. Mann and J. Everse. The crystal structure of activated protein C-inactivated bovine factor Va: implications for cofactor function, *Proc. Natl. Acad. Sci. U.S.A.*, 2004, 101(24), 8918–8923. <https://doi.org/10.1073/pnas.0403072101>
- V. Putignano, A. Rosato, L. Banci and C. Andreini. MetalPDB in 2018: a database of metal sites in biological macromolecular structures, *Nucleic Acids Res.*, 2018, 46(D1), D459–D464. <https://doi.org/10.1093/nar/gkx989>
- UniProt Consortium. UniProt: the universal protein knowledge-base in 2021, *Nucleic Acids Res.*, 2021, 49(D1), D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
- K. L. Howe, P. Achuthan, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K.

- Billis, S. Boddu, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, B. El Houdaigui, R. Fatima, A. Gall, C. Garcia Giron, T. Grego, C. Guijarro-Clarke, L. Haggerty, A. Hemrom, T. Hourlier, O. G. Izuogu, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. Gonzalez Martinez, J. C. Marugan, T. Maurel, A. C. McMahon, S. Mohanan, B. Moore, M. Muffato, D. N. Oheh, D. Paraschas, A. Parker, A. Parton, I. Prosovskaia, M. P. Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, E. Steed, M. Szpak, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, B. Walts, A. Winterbottom, M. Chakiachvili, A. Chaubal, N. De Silva, B. Flint, A. Frankish, S. E. Hunt, G. R. Ilsley, N. Langridge, J. E. Loveland, F. J. Martin, J. M. Mudge, J. Morales, E. Perry, M. Ruffier, J. Tate, D. Thybert, S. J. Trevanion, F. Cunningham, A. D. Yates, D. R. Zerbino and P. Flicek. Ensembl 2021, *Nucleic Acids Res.*, 2021, 49(D1), D884–D891. <https://doi.org/10.1093/nar/gkaa942>
22. R. Pengelly, D. Bakhtiar, I. Borovska, J. Kralovicova and I. Vorechovsky. Exonic splicing code and protein binding sites for calcium, *Nucleic Acids Res.*, 2022, 50(10), 5493–5512. <https://doi.org/10.1093/nar/gkac270>
23. S. Ke, S. Shang, S. M. Kalachikov, I. Morozova, L. Yu, J. J. Russo, J. Ju and L. A. Chasin. Quantitative evaluation of all hexamers as exonic splicing elements, *Genome Res.*, 2011, 21(8), 1360–1374. <https://doi.org/10.1101/gr.119628.110>
24. S. M. Mount. A catalogue of splice junction sequences, *Nucleic Acids Res.*, 1982, 10(2), 459–472. <https://doi.org/10.1093/nar/10.2.459>
25. S. Schafer, K. Miao, C. C. Benson, M. Heinig, S. A. Cook and N. Hubner, in *Current Protocols in Human Genetics*, vol. 87, John Wiley & Sons, Inc., Hoboken, New Jersey, 2015, pp. 11.6.1–11.6.4.
26. J. Tapial, K. C. Ha, T. Sterne-Weiler, A. Gohr, U. Braunschweig, A. Hermoso-Pulido, M. Quesnel-Vallieres, J. Permanyer, R. Sodaie, Y. Marquez, L. Cozzuto, X. Wang, M. Gomez-Velazquez, T. Rayon, M. Manzanares, J. Ponomarenko, B. J. Blencowe and M. Irimia. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms, *Genome Res.*, 2017, 27(10), 1759–1768. <https://doi.org/10.1101/gr.220962.117>
27. A. Busch and K. J. Hertel. HEXEvent: a database of Human EXon splicing Events, *Nucleic Acids Res.*, 2013, 41(D1), D118–D124. <https://doi.org/10.1093/nar/gks969>
28. M. Movassat, E. Forouzmand, F. Reese and K. J. Hertel. Exon size and sequence conservation improves identification of splice-altering nucleotides, *RNA*, 2019, 25(12), 1793–1805. <https://doi.org/10.1261/rna.070987.119>
29. M. C. Wahl, C. L. Will and R. Luhmann. The spliceosome: design principles of a dynamic RNP machine, *Cell*, 2009, 136(4), 701–718. <https://doi.org/10.1016/j.cell.2009.02.009>
30. F. Pagani, M. Raponi and F. E. Baralle. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, 102(18), 6368–6372. <https://doi.org/10.1073/pnas.0502288102>
31. M. Raponi, F. E. Baralle and F. Pagani. Reduced splicing efficiency induced by synonymous substitutions may generate a substrate for natural selection of new splicing isoforms: the case of CFTR exon 12, *Nucleic Acids Res.*, 2007, 35(2), 606–613. <https://doi.org/10.1093/nar/gkl1087>
32. D. G. J. Smethurst and N. Shcherbik. Interchangeable utilization of metals: new perspectives on the impacts of metal ions employed in ancient and extant biomolecules, *J. Biol. Chem.*, 2021, 297(6), 101374. <https://doi.org/10.1016/j.jbc.2021.101374>
33. E. V. Koonin. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?, *Biol. Direct*, 2006, 1(1), 22. <https://doi.org/10.1186/1745-6150-1-22>
34. C. L. Dupont, A. Butcher, R. E. Valas, P. E. Bourne and G. Caetano-Anolles. History of biological metal utilization inferred through phylogenomic analysis of protein structures, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, 107(23), 10567–10572. <https://doi.org/10.1073/pnas.0912491107>
35. K. A. Koch, M. M. Pena and D. J. Thiele. Copper binding motifs in catalysis. transport, detoxification and signaling, *Chem. Biol.*, 1997, 4(8), 549–560. [https://doi.org/10.1016/S1074-5521\(97\)90241-6](https://doi.org/10.1016/S1074-5521(97)90241-6)
36. P. F. Predki and B. Sarkar. Effect of replacement of “zinc finger” zinc on estrogen receptor DNA interactions, *J. Biol. Chem.*, 1992, 267(9), 5842–5846. [https://doi.org/10.1016/S0021-9258\(18\)42630-0](https://doi.org/10.1016/S0021-9258(18)42630-0)
37. K. Ok, M. R. Filipovic and S. L. Michel. Targeting zinc finger proteins with exogenous metals and molecules: lessons learned from tristetraproline, a CCCH type zinc finger, *Eur. J. Inorg. Chem.*, 2021, 2021(37), 3795–3805. <https://doi.org/10.1002/ejic.202100402>
38. Y. Wang, H. Hu, S. Yuan, L. Yunyan, K. Cao, H. Sun and Y. Liu. Cuprous ions can disrupt the structure and functions of the RING finger domain of RNF11, *Inorg. Chem. Front.*, 2022, 9(15), 3820. <https://doi.org/10.1039/D2QI00631F>
39. S. C. Leary and M. Ralle. Advances in visualization of copper in mammalian systems using X-ray fluorescence microscopy, *Curr. Opin. Chem. Biol.*, 2020, 55, 19–25. <https://doi.org/10.1016/j.cbpa.2019.12.002>
40. D. M. P. Mingos, *The Periodic Table I*, Springer Nature, Cham, Switzerland, 2019.
41. J. Kralovicova, I. Borovska, R. Pengelly, E. Lee, P. Abaffy, R. Sindelka, F. Grutzner and I. Vorechovsky. Restriction of an intron size en route to endothermy, *Nucleic Acids Res.*, 2021, 49(5), 2460–2487. <https://doi.org/10.1093/nar/gkab046>