

Research Article

Alternative Splicing of Transcription Factors' Genes: Beyond the Increase of Proteome Diversity

David Talavera,^{1,2,3} Modesto Orozco,^{1,4,5,6} and Xavier de la Cruz^{1,7,8}

¹ Barcelona Science Park (PCB), Molecular Modelling and Bioinformatics Unit, Institute for Research in Biomedicine (IRB), 08028 Barcelona, Spain

² European Bioinformatics Institute (EMBL), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

³ Faculty of Life Sciences, University of Manchester, Manchester M13 9PL, UK

⁴ Barcelona Science Park (PCB), Protein Structure and Modelling Node, The Spanish National Bioinformatics Institute, Genoma España, 08028 Barcelona, Spain

⁵ Departament de Bioquímica i Biologia Molecular, Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Spain

⁶ Department of Life Sciences, Barcelona Supercomputing Center, 08034 Barcelona, Spain

⁷ Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

⁸ Consejo Superior de Investigaciones Científicas, Institut de Biologia Molecular de Barcelona, 08028 Barcelona, Spain

Correspondence should be addressed to David Talavera, talavera@ebi.ac.uk and Xavier de la Cruz, xavier@mmb.pcb.ub.es

Received 25 December 2008; Revised 6 April 2009; Accepted 18 May 2009

Recommended by Antoine Danchin

Functional modification of transcription regulators may lead to developmental changes and phenotypical differences between species. In this work, we study the influence of alternative splicing on transcription factors in human and mouse. Our results show that the impact of alternative splicing on transcription factors is similar in both species, meaning that the ways to increase variability should also be similar. However, when looking at the expression patterns of transcription factors, we observe that they tend to diverge regardless of the role of alternative splicing. Finally, we hypothesise that transcription regulation of alternatively spliced transcription factors could play an important role in the phenotypical differences between species, without discarding other phenomena or functional families.

Copyright © 2009 David Talavera et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Some years ago the hypothesis that morphological differences between species are due to changes in the gene regulatory regions was proposed [1]. Recent advances in the development field are providing supporting evidences even when comparing relatively remote species [2] and their interpretation within an evolutionary context has led to creation of the evo-devo field. An important part of the research within this field is focused on the comparison of gene regulatory regions [2]. However, less attention has been devoted to explore the role of transcription factors (TFs). Regulation of the activity of transcription factors is a complex process [3] including a broad range of intrinsic and environmental factors. This is particularly relevant within the context of development and evolutionary research, since control of the amount of TFs, at precise locations and

times, may constitute a finer alternative to the more drastic presence/absence of TFs binding sites.

Among the mechanisms modulating the activity of TFs the role of alternative splicing (AS) has been well documented in the recent years [3–5]. Indeed, different studies have shown that AS of TFs results in regulatory isoforms [3, 6–9] that can be tissue- or development stage-specific [4, 10–13] and show cell distribution variation [14]. In general, the biological effect of AS on TFs can be easily interpreted if we consider the fact that TFs are commonly large proteins with a modular composition [4, 7]. TFs domains have different roles related to the main function of TFs: DNA binding, dimerization and function regulation. DNA binding domains are required to recognize target sequences; dimerization domains allow the building of dimers or oligomers which are the biological unit of many TFs; and regulatory domains are used to detect external

stimulus or signals from transduction pathways. Therefore, loss of one of these domains will be accompanied by the loss of one of these functional properties, thus resulting [3, 15–18] in transcripts that either (i) lack the original activity, (ii) show and increase or decrease in this activity, or (iii) act as dominant negative of the fully-functional isoform, having an antagonistic effect. Obviously, the nature of the regulation associated to AS depends on the domains involved. However, it is important to notice that a partial deletion or substitution can also lead to a loss or modification of domain functionality [19, 20].

Bioinformatics research has widely studied AS using several approaches, contributing to shape our present view of the functional changes caused by this phenomenon [21–29]. In the case of TFs, work from different authors has focused in the study of AS mechanisms of cancer-related genes [30], in the properties of specific domains in AS [31], and in the role of AS of TFs in different species [10]. Among other facts, it has been established that human and mouse have different TFs variants [31] and that TFs seem to be more frequently spliced, creating tissue-specific isoforms with different domain architectures [10]. In spite of their growing amount, the view provided by these studies is still incomplete, and important aspects related to the variability generated by AS and its interspecies conservation remain unclear.

In this article we compare how AS of TFs varies between human and mouse, two species with clear morphological differences, focusing on some characteristics related to the generation of regulatory mechanisms. In addition, we also compared the expression levels of orthologous TFs to see whether there are substantial expression differences between both species. Our results indicate that human and mouse use similar mechanisms to regulate the action of TFs. We also find, for a population of human-mouse orthologs, that TFs tend to show diverging gene expression changes not related to the presence of AS.

2. Materials and Methods

2.1. Data. TFs and enzymes sequences and position of variable regions were obtained from SwissProt [32]. The most important point in the bioinformatics studies is the election of the dataset. There are several options: (1) manually curated databases; (2) automatically-annotated databases and (3) mining of publicly available experimental data, all having strengths and weaknesses [33]. Curated databases contain the most reliable information, but have a small coverage. Automatic annotation of databases increases the amount of data, but as the annotations are transferred by homology their reliability depend on the used thresholds. Finally, the public repositories contain a huge amount of unprocessed data. Obviously, all achieved conclusions should be nuanced according to the used dataset. As our analysis relied in the correlation between alternative regions and functional domains, we thought that it was crucial to analyse real isoforms, which are provenly transcribed and translated, instead of putative peptides. Consequently, we chose the database containing less artefacts [34]. In addition, although

we missed several isoforms in our analyses, the general trends should not be importantly affected, as we found in a previous study [26].

2.2. Similarity of TFs. We compared all TFs isoforms using the CD-HIT software [35]. This tool clusters all sequences at a given identity threshold.

2.3. Domain Predictions. Functional domains from SMART [36] and Pfam [37] were identified with RPS-BLAST (a PSI-BLAST variant [38]) using 0.02 as *E*-value and the low complexity filter. When the domain annotation for these two databases overlapped without completely coinciding, we chose the longest domain assignment and discarded the shortest if at least 60% of it overlapped with the longest.

2.4. Precision of AS Effects on TFs. For each AS modification, we took the N- and C-terminal positions of the variable region and the N- and C-terminal boundaries of the related functional domains, if any. We ordered the four positions and calculated the precision as the ratio between the residues being both functional and alternatively spliced and all the residues being either in the functional domain or the AS region. A precision close to 1 meant that AS was almost coincident with functional domain boundaries, whereas a low precision showed a lower correlation.

2.5. TFs Expression Patterns. Expression data for human and mouse genes were retrieved from the SymAtlas server (<http://symatlas.gnf.org>, now at <https://biogps.gnf.org/>) (human: GNF1H; mouse: GNF1M) [39]. These data did not contain isoform expression, but all isoforms were grouped. This fact was useful to our goal, because we were not interested in the expression of the equivalent isoforms, but in the effect of the ability of alternatively splice genes on the expression pattern of orthologs.

We analysed 559 TFs from 30 common tissues (mouse spinal cord data was the average of upper and lower spinal chord ones). This dataset was divided in three sets, depending on the presence of AS, as follows: (a) a set of 123 TFs for which both orthologs had AS; (b) a set of 109 TFs which were alternatively spliced in one species but not in the other; and finally (c) a set of 327 TFs with no AS neither in human nor in mouse. In addition, we analysed the expression pattern of 1923 enzymes so as to study if TFs had some kind of specific features.

Gene expression tissue patterns were compared using the Pearson's correlation, following similar studies [40–43]. High correlations mean that genes have similar expression patterns, whereas low correlations indicate variations due to tissue-specific expression. As suggested by Liao and Zhang [43], we used relative abundance instead of signal intensity measured from the microarrays. Signal intensity does not quantify the abundance of mRNA in the sample, is different for each experiment and is influenced by many factors. On the contrary, the relative abundance normalises the values within each experiment: briefly, each expression level is divided by the sum of all the signals in the experiment.

TABLE 1: Summary of AS data for human and mouse TFs. The results for ALL GENES are those obtained analysing all entries in SwissProt. Each TFs dataset must be compared with the same species ALL GENES dataset.

	TFs		All genes	
	Human	Mouse	Human	Mouse
No. of genes	1077	737	12946	10031
% AS+	29.4	26.1	26.9	17.6
No. of isoforms/gene	3.0	2.9	2.8	2.7

This normalisation allows comparison of results of different experimental results. When several replicas were available for a given experiment we averaged the expression data. In addition, comparison between these replicas was utilized to obtain a control of the reproducibility of the results. In absence of alternative experimental validation, this control ensures that the possible differences are not due to the variance between experiments.

3. Results and Discussion

3.1. Diversity of TFs. Since the publication of the human genome it was suggested that rate differences in AS could be associated to phenotypic differences between organisms [44]. In the case of TFs, Taneri and colleagues [10] have recently shown that for mouse the percentage of genes with AS is higher for TFs than for other proteins. In our case (Table 1) we found that this was the case for mouse (T -test P -value ~ 0) as well as for human (T -test P -value ~ 0.04). This means that the higher rate of AS in TFs could be a general feature. In addition, the average number of isoforms per gene was similar in both species. The actual percentages of genes having AS were different from those obtained by Taneri and colleagues (62% and 29% of TFs loci and all loci, resp.) [10], because they build their database using a computational algorithm, whereas we relied only on experimentally validated protein isoforms. The important point is that the biases affecting the manual or automatic annotation of AS would affect all loci. Thus, independently of the presented percentages, the fact is that TFs genes have a significantly higher rate of AS. The rate differences between human and mouse could be real or simply due to the different coverage of the two species [45, 46].

Additionally, we looked at the similarity of TFs. Given the existence of structural and evolutionary restrictions [47, 48] and that the number of functional domains is limited; we expected that many TFs would share a big part of their sequence. However, we observed that very few TFs had a high identity percentage neither in human nor in mouse (Table 2). Moreover, just one third or one quarter of the isoforms was at least 40% identical, meaning that the majority of TFs probably did not have the same functionality [49]. The latter result was not different from that obtained with the control; however, control included all human non-TFs proteins, which could be very different proteins. According to the literature, no big differences would be expected if using mouse proteins instead of the human dataset as a control

TABLE 2: Similarity among TFs. The table shows the percentage of TFs isoforms which share an identity percentage. Control includes all the human non-TFs protein sequences.

	Control	Human	Mouse
No. of genes	10937	1077	737
No. of protein sequences	17026	1702	1093
90%	0.04	0.01	0.01
80%	0.07	0.02	0.02
70%	0.10	0.07	0.04
60%	0.16	0.13	0.08
50%	0.22	0.22	0.15
40%	0.31	0.34	0.23

[26, 27, 45, 50, 51]. These results suggest that the role of AS in the slight regulation of TFs is almost unique, as gene duplicates are so divergent that they hardly could play similar roles [26].

3.2. AS Effects on TFs Functional Domains. Next we studied how AS affects domain composition in TFs, a feature directly related to the regulation of protein function in general [26, 27, 52–54] and to that of TFs in particular [4, 5, 16, 55]. Analysing TFs with AS from our dataset, we found that in 28% and 18% of human and mouse TFs at least one isoform showed domain composition changes; and from the different domain types present, 59% in human and 68% in mouse were affected by alternative splicing. These data suggest that both organisms use AS to regulate the action of TFs in a similar way.

As the effect of AS upon the TFs depend on the affected functional domains, it is important to know whether the process is random-like. In our case we first observed that AS did not affect all possible domains. And second, we found that among those affected, not all were affected with the same frequency, in accordance with what was found by Liu and Altman [25] when considering the whole proteome. Moreover, we see that the most affected domains (HOX, HOLI, HLH, and PHD) are the same and with similar percentages in both species. The only exception was the (Krueppel-associated box) KRAB domain, which is frequently affected in human and not in mouse, probably due to the absence of mouse orthologs in our dataset. In accordance with Taneri and colleagues [10], we find that AS frequently affects DNA-binding domains in both species. For example, the most frequently affected domain common to both species is HOX, the homeobox domain responsible of DNA binding in transcription factors. However, we found that AS may also affect domains with other biochemical properties. This is the case of the common second most frequently affected domain, HOLI, a hormone-binding domain present in nuclear hormone receptors. Also a frequently spliced-out domain, PHD, is involved in protein-protein interactions and/or interactions with chromatin.

Previous studies on the effect of AS on functional domains indicate that AS shows a bias towards encompassing whole domains [29, 56]. However, as emphasized by Zavolan

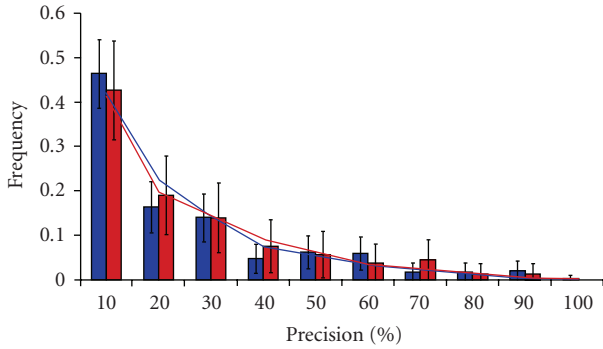


FIGURE 1: The four possible types of overlap between AS (coloured in red) and protein domains (boxes).

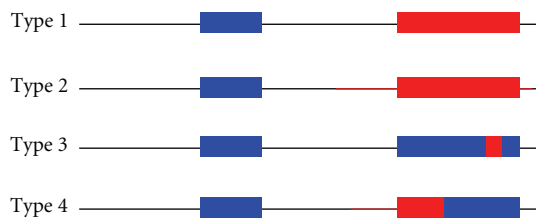


FIGURE 2: Precision of AS effect on TFs functional domains. Bar histograms show results for the observed frequencies, whereas the lines show distribution for the expected frequencies at random. Human (blue) and mouse (red) results are shown.

and van Nimwegen [57] there are also many cases in which AS only affects part of the protein domain. To explore what happened in the case of TFs, we measured the overlap between AS and domain boundaries. Firstly, we measured the precision of AS effect on TFs. Surprisingly; results show a random distribution either in human or mouse (Figure 1). As alternative splicing is determined by gene structure and cannot occur wherever by chance, we refined our analysis. More precisely we considered four different situations (Figure 2): (I) AS and domain boundaries coincide exactly; (II) the spliced-out region spreads over the whole domain and surrounding sequence; (III) the spliced-out region is completely included within the domain and (IV) there is only a partial overlap between the spliced-out region and the functional domain. A high frequency of type I cases would mean a positive selection towards the co-evolution of AS and functional domains. On the contrary, types II, III, and IV suggest that any kind of correlation, if existing, would be weak. The results obtained are shown in Table 3. We see that the first situation was never observed, in accordance with the very low probability of exact coincidence between AS and domain boundaries. However, the three remaining situations were found in both species with comparable ratios—including differences between observed and expected frequencies—indicating that the functional regulation of TFs by AS is reached by similar mechanisms in both species.

3.3. *Expression of Orthologous TFs.* All these results point to a coincidence between human and mouse in the regulation

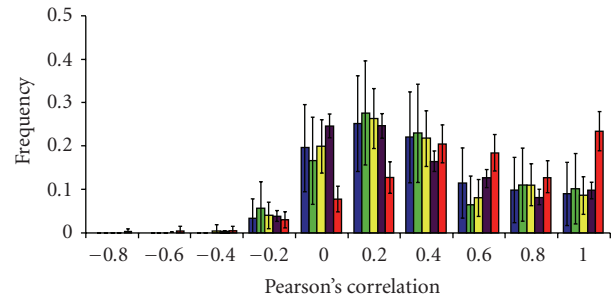


FIGURE 3: Comparison of the expression pattern of human and mouse ortholog TFs. The figure shows the distribution of Pearson's correlation values for the three following comparisons between human and mouse ortholog TFs: TFs having AS in both species (blue), TFs having AS in only one species (green) and TFs with no AS in any of the two species (yellow). In addition, it shows distribution for enzymes (purple) and replica's control (red). We calculated 95% confidence intervals for each bar on frequency histograms assuming the independence hypothesis [58].

TABLE 3: *AS effects on functional domains.* For each of the species are shown the observed and the expected (in parenthesis) frequencies.

	Human	Mouse
Type I	0.00 (0.00)	0.00 (0.00)
Type II	0.36 (0.20)	0.24 (0.14)
Type III	0.23 (0.42)	0.43 (0.47)
Type IV	0.41 (0.38)	0.33 (0.39)

of TFs function by AS. Finally, we studied whether the presence/absence of AS may result in differences in TFs expression pattern. In our study we considered three different situations: (a) the TFs gene has AS in both species, (b) the TFs gene has AS in only one species; and (c) no AS was observed for the TFs genes in neither species. We compared TFs expression with that from enzymes, in order to have an external group. Because array data may show large fluctuations, we also used as a control experimental replicas for each TFs. When looking at the results (Figure 3) we can see that the replica's control has a high correlation in the expression distribution meaning that the different experiments were consistent. More interestingly, we observe that the distributions for the three TFs human-mouse comparisons are clearly different from that of the replica's control, whereas they are similar to that of enzymes. This points that, in general, human and mouse orthologous genes tend to have divergent tissue expression patterns and that TFs are no exception, even if the majority of their equivalent functional domains (either constitutive or alternative) are identical. This means that human and species do not use comparatively the same TFs in the same tissues. Finally, the trend does not depend on whether the TFs has AS, since the distributions for the three possible situations are essentially the same; thus, discarding the exclusive role of the specific isoforms in the apparition of the expression divergence. The uncoupling between AS and transcription regulation suggests they provide two independent levels of control of TFs products.

Overall, our results indicate that rather than AS alone, a combination of AS and regulation at the transcription level may determine the nature and final amount of product for TFs [59, 60]. This suggests that in addition to changes at promoter regions, regulation of TFs activity might also play an important role in those processes that result in phenotypic differences between species. Importantly, this does not discard the role of other functional families or alternative phenomena, such as posttranslational modifications [61], in the apparition of interspecies differences. Additional experimental studies should be done in the future to test the validity of these hypotheses.

Acknowledgments

Xavier de la Cruz acknowledges funding from the Spanish government (Grant BIO2006-15557). Modesto Orozco acknowledges funding from the Spanish government (Grant BIO2006-01602). David Talavera acknowledges funding from the Spanish government (Grant BIO2006-01602) and the European Commission (Contract LHSG-CT-2003-503265).

References

- [1] M. C. King and A. C. Wilson, "Evolution at two levels in humans and chimpanzees," *Science*, vol. 188, no. 4184, pp. 107–116, 1975.
- [2] S. B. Carroll, "Evolution at two levels: on genes and form," *PLoS Biology*, vol. 3, no. 7, article e245, 2005.
- [3] D. S. Latchman, "The Oct-2 transcription factor," *International Journal of Biochemistry and Cell Biology*, vol. 28, no. 10, pp. 1081–1083, 1996.
- [4] A. J. Lopez, "Developmental role of transcription factor isoforms generated by alternative splicing," *Developmental Biology*, vol. 172, no. 2, pp. 396–411, 1995.
- [5] A. J. Lopez, "Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation," *Annual Review of Genetics*, vol. 32, pp. 279–305, 1998.
- [6] T. Hsu, J. A. Gogos, S. A. Kirsh, and F. C. Kafatos, "Multiple zinc finger forms resulting from developmentally regulated alternative splicing of a transcription factor gene," *Science*, vol. 257, no. 5078, pp. 1946–1950, 1992.
- [7] B. M. Laoide, N. S. Foulkes, F. Schlotter, and P. Sassone-Corsi, "The functional versatility of CREM is determined by its modular structure," *The EMBO Journal*, vol. 12, no. 3, pp. 1179–1191, 1993.
- [8] D. Duma, C. M. Jewell, and J. A. Cidlowski, "Multiple glucocorticoid receptor isoforms and mechanisms of post-translational modification," *Journal of Steroid Biochemistry and Molecular Biology*, vol. 102, no. 1–5, pp. 11–21, 2006.
- [9] M. Zou, Y. Guan, H. Ren, F. Zhang, and F. Chen, "Characterization of alternative splicing products of bZIP transcription factors OsABI5," *Biochemical and Biophysical Research Communications*, vol. 360, no. 2, pp. 307–313, 2007.
- [10] B. Taneri, B. Snyder, A. Novoradovsky, and T. Gaasterland, "Alternative splicing of mouse transcription factors affects their DNA-binding domain architecture and is tissue specific," *Genome Biology*, vol. 5, no. 10, article R75, 2004.
- [11] A. M. Cervera, N. Apostolova, F. Luna-Crespo, A. Sanjuan-Pla, R. Garcia-Bou, and K. J. McCreath, "An alternatively spliced transcript of the PHD3 gene retains prolyl hydroxylase activity," *Cancer Letters*, vol. 233, no. 1, pp. 131–138, 2006.
- [12] D. Ram, E. Ziv, F. Lantner, V. Lardans, and I. Schechter, "Stage-specific alternative splicing of the heat-shock transcription factor during the life-cycle of *Schistosoma mansoni*," *Parasitology*, vol. 129, no. 5, pp. 587–596, 2004.
- [13] E. Lamani, Y. Wu, J. Dong, M. S. Litaker, A. C. Acevedo, and M. MacDougall, "Tissue- and cell-specific alternative splicing of NFIC," *Cells Tissues Organs*, vol. 189, no. 1–4, pp. 105–110, 2008.
- [14] R. Caballero, F. Setien, L. Lopez-Serra, et al., "Combinatorial effects of splice variants modulate function of Aiolos," *Journal of Cell Science*, vol. 120, no. 15, pp. 2619–2630, 2007.
- [15] G. A. Robichaud, M. Nardini, M. Laflamme, M. Cuperlovic-Culf, and R. J. Ouellette, "Human Pax-5 C-terminal isoforms possess distinct transactivation properties and are differentially modulated in normal and malignant B cells," *The Journal of Biological Chemistry*, vol. 279, no. 48, pp. 49956–49963, 2004.
- [16] L. Waltzer, L. Bataille, S. Peyrefitte, and M. Haenlin, "Two isoforms of serpent containing either one or two GATA zinc fingers have different roles in *Drosophila* haematopoiesis," *The EMBO Journal*, vol. 21, no. 20, pp. 5477–5486, 2002.
- [17] H. L. Ford, "Homeobox genes: a link between development, cell cycle, and cancer?" *Cell Biology International*, vol. 22, no. 6, pp. 397–400, 1998.
- [18] R. P. Kuiper, M. Schepens, J. Thijssen, E. F. Schoenmakers, and A. G. van Kessel, "Regulation of the MiTF/TFE bHLH-LZ transcription factors through restricted spatial expression and alternative splicing of functional domains," *Nucleic Acids Research*, vol. 32, no. 8, pp. 2315–2322, 2004.
- [19] H.-S. Choi, M. Chung, I. Tzamelis, et al., "Differential transactivation by two isoforms of the orphan nuclear hormone receptor CAR," *The Journal of Biological Chemistry*, vol. 272, no. 38, pp. 23565–23571, 1997.
- [20] F. E. Chen, D.-B. Huang, Y.-Q. Chen, and G. Ghosh, "Crystal structure of p50/p65 heterodimer of transcription factor NF- κ B bound to DNA," *Nature*, vol. 391, no. 6665, pp. 410–412, 1998.
- [21] Q. Xu, B. Modrek, and C. Lee, "Genome-wide detection of tissue-specific alternative splicing in the human transcriptome," *Nucleic Acids Research*, vol. 30, no. 17, pp. 3754–3766, 2002.
- [22] A. R. R. Forrest, D. F. Taylor, M. L. Crowe, et al., "Genome-wide review of transcriptional complexity in mouse protein kinases and phosphatases," *Genome Biology*, vol. 7, no. 1, article R5, 2006.
- [23] K. Yura, M. Shionyu, K. Hagino, et al., "Alternative splicing in human transcriptome: functional and structural influence on proteins," *Gene*, vol. 380, no. 2, pp. 63–71, 2006.
- [24] A. Resch, Y. Xing, B. Modrek, M. Gorlick, R. Riley, and C. Lee, "Assessing the impact of alternative splicing on domain interactions in the human proteome," *Journal of Proteome Research*, vol. 3, no. 1, pp. 76–83, 2004.
- [25] S. Liu and R. B. Altman, "Large scale study of protein domain distribution in the context of alternative splicing," *Nucleic Acids Research*, vol. 31, no. 16, pp. 4828–4835, 2003.
- [26] D. Talavera, C. Vogel, M. Orozco, S. A. Teichmann, and X. de la Cruz, "The (in)dependence of alternative splicing and gene duplication," *PLoS Computational Biology*, vol. 3, no. 3, article e33, 2007.
- [27] A. Valenzuela, D. Talavera, M. Orozco, and X. de la Cruz, "Alternative splicing mechanisms for the modulation of protein function: conservation between human and other

- species," *Journal of Molecular Biology*, vol. 335, no. 2, pp. 495–502, 2004.
- [28] S. Lois, N. Blanco, M. Martínez-Balbás, and X. de la Cruz, "The functional modulation of epigenetic regulators by alternative splicing," *BMC Genomics*, vol. 8, article 252, 2007.
- [29] E. V. Kriventseva, I. Koch, R. Apweiler, et al., "Increase of functional diversity by alternative splicing," *Trends in Genetics*, vol. 19, no. 3, pp. 124–128, 2003.
- [30] P. Ferro, A. Forlani, M. Muselli, and U. Pfeffer, "Alternative splicing of the human estrogen receptor alpha primary transcript: mechanisms of exon skipping," *International Journal of Molecular Medicine*, vol. 12, no. 3, pp. 355–363, 2003.
- [31] K. Tajul-Arifin, R. Teasdale, T. Ravasi, et al., "Identification and analysis of chromodomain-containing proteins encoded in the mouse transcriptome," *Genome Research*, vol. 13, no. 6, pp. 1416–1429, 2003.
- [32] B. Boeckmann, A. Bairoch, R. Apweiler, et al., "The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003," *Nucleic Acids Research*, vol. 31, no. 1, pp. 365–370, 2003.
- [33] G. A. Reeves, D. Talavera, and J. M. Thornton, "Genome and proteome annotation: organization, interpretation and integration," *Journal of the Royal Society Interface*, vol. 6, no. 1, pp. 129–147, 2009.
- [34] L. Nagy, H. Hegyi, K. Farkas, et al., "Identification and correction of abnormal, incomplete and mispredicted proteins in public databases," *BMC Bioinformatics*, vol. 9, article 353, 2008.
- [35] W. Li, L. Jaroszewski, and A. Godzik, "Clustering of highly homologous sequences to reduce the size of large protein databases," *Bioinformatics*, vol. 17, no. 3, pp. 282–283, 2001.
- [36] I. Letunic, R. R. Copley, S. Schmidt, et al., "SMART 4.0: towards genomic data integration," *Nucleic Acids Research*, vol. 32, pp. D142–D144, 2004.
- [37] A. Bateman, L. Coin, R. Durbin, et al., "The Pfam protein families database," *Nucleic Acids Research*, vol. 32, pp. D138–D141, 2004.
- [38] S. F. Altschul, T. L. Madden, A. A. Schaffer, et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.
- [39] A. I. Su, T. Wiltshire, S. Batalov, et al., "A gene atlas of the mouse and human protein-encoding transcriptomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 16, pp. 6062–6067, 2004.
- [40] L. Huminiecki and K. H. Wolfe, "Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse," *Genome Research*, vol. 14, no. 10, pp. 1870–1879, 2004.
- [41] I. Yanai, D. Graur, and R. Ophir, "Incongruent expression profiles between human and mouse orthologous genes suggest widespread neutral evolution of transcription control," *OMICS*, vol. 8, no. 1, pp. 15–24, 2004.
- [42] I. K. Jordan, L. Marino-Ramirez, and E. V. Koonin, "Evolutionary significance of gene expression divergence," *Gene*, vol. 345, pp. 119–126, 2005.
- [43] B.-Y. Liao and J. Zhang, "Evolutionary conservation of expression profiles between human and mouse orthologous genes," *Molecular Biology and Evolution*, vol. 23, no. 3, pp. 530–540, 2006.
- [44] E. S. Lander, L. M. Linton, B. Birren, et al., "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.
- [45] D. Brett, H. Pospisil, J. Valcárcel, J. Reich, and P. Bork, "Alternative splicing and genome complexity," *Nature Genetics*, vol. 30, no. 1, pp. 29–30, 2002.
- [46] H. Kim, R. Klein, J. Majewski, et al., "Estimating rates of alternative splicing in mammals and invertebrates," *Nature Genetics*, vol. 36, no. 9, pp. 915–917, 2004.
- [47] C. Chothia and A. M. Lesk, "The relation between the divergence of sequence and structure in proteins," *The EMBO Journal*, vol. 5, no. 4, pp. 823–826, 1986.
- [48] V. Daburon, S. Mella, J.-L. Plouhinec, S. Mazan, M. Crozatier, and A. Vincent, "The metazoan history of the COE transcription factors. Selection of a variant HLH motif by mandatory inclusion of a duplicated exon in vertebrates," *BMC Evolutionary Biology*, vol. 8, no. 1, article 131, 2008.
- [49] W. Tian and J. Skolnick, "How well is enzyme function conserved as a function of pairwise sequence identity?" *Journal of Molecular Biology*, vol. 333, no. 4, pp. 863–882, 2003.
- [50] T. A. Thanaraj, F. Clark, and J. Muilu, "Conservation of human alternative splice events in mouse," *Nucleic Acids Research*, vol. 31, no. 10, pp. 2544–2552, 2003.
- [51] R. H. Waterston, K. Lindblad-Toh, E. Birney, et al., "Initial sequencing and comparative analysis of the mouse genome," *Nature*, vol. 420, no. 6915, pp. 520–562, 2002.
- [52] J. Stetefeld and M. A. Ruegg, "Structural and functional diversity generated by alternative mRNA splicing," *Trends in Biochemical Sciences*, vol. 30, no. 9, pp. 515–521, 2005.
- [53] F. A. Kondrashov and E. V. Koonin, "Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences," *Trends in Genetics*, vol. 19, no. 3, pp. 115–119, 2003.
- [54] M. Hiller, K. Huse, M. Platzer, and R. Backofen, "Creation and disruption of protein features by alternative splicing—a novel mechanism to modulate function," *Genome Biology*, vol. 6, no. 7, article R58, 2005.
- [55] K. Jackson Behan, J. Fair, S. Singh, et al., "Alternative splicing removes an Ets interaction domain from *Lozenge* during *Drosophila* eye development," *Development Genes and Evolution*, vol. 215, no. 8, pp. 423–435, 2005.
- [56] K. Homma, R. F. Kikuno, T. Nagase, O. Ohara, and K. Nishikawa, "Alternative splice variants encoding unstable protein domains exist in the human brain," *Journal of Molecular Biology*, vol. 343, no. 5, pp. 1207–1220, 2004.
- [57] M. Zavolan and E. van Nimwegen, "The types and prevalence of alternative splice forms," *Current Opinion in Structural Biology*, vol. 16, no. 3, pp. 362–367, 2006.
- [58] L. A. Goodman, "On simultaneous confidence intervals for multinomial proportions," *Technometrics*, vol. 7, pp. 247–254, 1965.
- [59] L. Chen and S. Zheng, "Studying alternative splicing regulatory networks through partial correlation analysis," *Genome Biology*, vol. 10, no. 1, article R3, 2009.
- [60] H. Vihma, P. Pruunsild, and T. Timmusk, "Alternative splicing and expression of human and mouse NFAT genes," *Genomics*, vol. 92, no. 5, pp. 279–291, 2008.
- [61] A. Nayak, J. Glockner-Pagel, M. Vaeth, et al., "SUMOylation of the transcription factor NFATc1 leads to its subnuclear relocalization and IL2 repression by HDAC," *The Journal of Biological Chemistry*, vol. 284, no. 16, pp. 10935–10946, 2009.