

Correspondence

Alternative Splicing at NAGNAG Acceptors: Simply Noise or Noise and More?

Michael Hiller, Karol Szafranski, Rolf Backofen, Matthias Platzer

Alternative splicing at pairs of acceptors in close proximity are one frequent cause of transcriptome complexity. In particular, acceptors with the pattern NAGNAG are widespread in several genomes [1–3]. When affecting the coding regions, alternative splicing at NAGNAGs mainly results in the insertion/deletion of one amino acid. While such subtle events are undoubtedly frequent, an important question arises: do they have functional consequences or are they simply noise tolerated by cells?

Zavolan and colleagues [3,4] suggest that these variations are the result of stochastic binding of the spliceosome at neighboring splice sites and do not discuss known functional implications. We previously found indications against a general noise assumption for NAGNAG splice events [1]: biases towards intron phase 1 and single amino acid insertions/deletions, correlation of amino acid variation and the peptide environment, enrichment of polar residues at NAGNAG exon–exon junctions, preference for protein–protein interactions and particular Pfam domains, human–mouse conservation of the intronic AG, and tissue-specific splicing at several NAGNAG acceptors. These findings indicate negative selection against NAGNAG-derived variability deleterious for certain protein regions, which agrees with the underrepresentation of NAGNAGs in coding regions detected by Zavolan and colleagues [4]. This does not rule out that variability may be advantageous for other proteins, but signs of positive selection are much harder to detect and remain to be shown.

Zavolan's finding that confirmed NAGNAGs (current mRNAs/expressed sequence tags do show alternative splicing) are not better conserved between human and mouse than unconfirmed ones may argue against functional implications. However, this result is probably biased by the unconfirmed dataset, which consists of ~60% NAGGAG whose GAG is part of the conserved exon. To avoid such a bias, we split confirmed NAGNAGs into those in which the “extra” AG is either intronic or exonic, according to the transcript annotation [1]. Interestingly, intronic but not exonic extra AGs have a significant conservation. Meanwhile, Akerman and Mandel-Gutfreund found a high conservation of the intronic flanking regions [5], typical for biologically meaningful alternative splicing [6].

The finding of Zavolan and colleagues that relative acceptor strength is predictive for confirmed and unconfirmed NAGNAGs refers to an accepted fact of splicing (for example, alternative exons have weaker splice sites than constitutive ones [7]). In tandems, the splice-site strength often determines the preferred acceptor, consistent with our earlier results (see Supplementary Notes in [1]). Thus, we agree that thermodynamic fluctuation plays an essential role during splice-site recognition at NAGNAG acceptors. This is in line with the finding that a single mutation is sufficient to

convert a normal acceptor into a NAGNAG tandem, enabling alternative splicing [8]. However, this useful model is not valid for all NAGNAGs. In particular, tissue-specific regulation of alternative NAGNAG splicing challenges this model [1,9]. Overrepresented sequence motifs found in the vicinity of confirmed NAGNAGs are likely to contribute to this regulation [5].

Moreover, some protein isoforms derived by alternative splicing at NAGNAG acceptors are known to be functionally different: IGF1R, signaling [10]; DRPLA, cellular localization [9]; mouse Pax3, DNA binding [11]; and *Arabidopsis thaliana* U11-35K, protein binding [12]. Alternative NAGNAG splicing in the untranslated region of mouse Ggt1 affects the translational efficiency [13]. Furthermore, a NAGNAG mutation in ABCA4 is relevant for Stargardt disease 1 [14]. For clarity, we did not claim that all alternative splice events at NAGNAGs serve as protein “fine-tuning” mechanism [1,8] (as misinterpreted by [4]). In our opinion, like genetic variants, splice variants may be neutral or result in phenotypic differences. Thus, they represent just another playground of molecular evolution [15,16]. The few currently evident cases of biologically different NAGNAG-derived isoforms may represent just the tip of an iceberg.

Finally, in the context of the problem discussed here, it has to be considered that noise is important for many biological processes [17], leading to the model of “cultivated noise” [18]. For example, splicing noise at the *Drosophila Dscam* gene is used for cell individualization [19]. Although it has yet to be proven, it is tempting to speculate that noise arising by splicing at NAGNAG acceptors provides another “cultivated” stochastic mechanism.

In conclusion, it remains unknown what fraction of the more than 1,900 currently confirmed human NAGNAGs play a role in biological functions. To facilitate further experimental and bioinformatics analyses, we developed a database, TassDB (<http://helios.informatik.uni-freiburg.de/TassDB>), that provides information and large collections of NAGNAG acceptors. ■

Michael Hiller
Rolf Backofen
Albert-Ludwigs-University Freiburg
Freiburg, Germany

Karol Szafranski
Matthias Platzer (mplatzer@fli-leibniz.de)
Leibniz Institute for Age Research Jena, Germany

References

- Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, et al. (2004) Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nat Genet* 36: 1255–1257.
- Sugnet CW, Kent WJ, Ares M Jr, Haussler D (2004) Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac Symp Biocomput* 2004: 66–77.
- Zavolan M, Kondo S, Schonbach C, Adachi J, Hume DA, et al. (2003) Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res* 13: 1290–1300.
- Chern TM, van Nimwegen E, Kai C, Kawai J, Carninci P, et al. (2006) A simple physical model predicts small exon length variations. *PLoS Genet* 2: doi:10.1371/journal.pgen.0020045
- Akerman M, Mandel-Gutfreund Y (2006) Alternative splicing regulation at tandem 3' splice sites. *Nucleic Acids Res* 34: 23–31.

6. Sorek R, Ast G (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res* 13: 1631–1637.
7. Sorek R, Shemesh R, Cohen Y, Basechess O, Ast G, et al. (2004) A non-EST-based method for exon-skipping prediction. *Genome Res* 14: 1617–1623.
8. Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, et al. (2006) Single-nucleotide polymorphisms in NAGNAG acceptors are highly predictive for variations of alternative splicing. *Am J Hum Genet* 78: 291–302.
9. Tadokoro K, Yamazaki-Inoue M, Tachibana M, Fujishiro M, Nagao K, et al. (2005) Frequent occurrence of protein isoforms with or without a single amino acid residue by subtle alternative splicing: The case of Gln in DRPLA affects subcellular localization of the products. *J Hum Genet* 50: 382–394.
10. Condorelli G, Bueno R, Smith RJ (1994) Two alternatively spliced forms of the human insulin-like growth factor I receptor have distinct biological activities and internalization kinetics. *J Biol Chem* 269: 8510–8516.
11. Vogan KJ, Underhill DA, Gros P (1996) An alternative splicing event in the Pax-3 paired domain identifies the linker region as a key determinant of paired domain DNA-binding activity. *Mol Cell Biol* 16: 6677–6686.
12. Lorkovic ZJ, Lehner R, Forstner C, Barta A (2005) Evolutionary conservation of minor U12-type spliceosome between plants and humans. *RNA* 11: 1095–1107.
13. Joyce-Brady M, Jean JC, Hughey RP (2001) gamma-glutamyltransferase and its isoform mediate an endoplasmic reticulum stress response. *J Biol Chem* 276: 9468–9477.
14. Maugeri A, van Driel MA, van de Pol DJ, Klevering BJ, van Haren FJ, et al. (1999) The 2588G-->C mutation in the ABCR gene is a mild frequent founder mutation in the Western European population and allows the classification of ABCR mutations in patients with Stargardt disease. *Am J Hum Genet* 64: 1024–1035.
15. Ast G (2004) How did alternative splicing evolve? *Nat Rev Genet* 5: 773–782.
16. Modrek B, Lee CJ (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat Genet* 34: 177–180.
17. Fedoroff N, Fontana W (2002) Genetic networks. Small numbers of big molecules. *Science* 297: 1129–1131.
18. Rao CV, Wolf DM, Arkin AP (2002) Control, exploitation and tolerance of intracellular noise. *Nature* 420: 231–237.
19. Neves G, Zucker J, Daly M, Chess A (2004) Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nat Genet* 36: 240–246.

Citation: Hiller M, Szafranski K, Backofen R, Platzer M (2006) Alternative splicing at NAGNAG acceptors: Simply noise or noise and more? *PLoS Genet* 2(11): e207. doi:10.1371/journal.pgen.0020207

Copyright: © 2006 Hiller et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors were supported by grants from the German Ministry of Education and Research (01GR0504 and 0313652D) as well as from the Deutsche Forschungsgemeinschaft (SFB604–02).

Competing Interests: The authors have declared that no competing interests exist.

Authors' Reply

That splice variation at tandem acceptor sites is frequent has been reported by several groups, including Zavolan et al. [1], Sugnet et al. [2], and Hiller et al. [3], and is uncontroversial. It is to be expected that at least some of these variations will affect protein function, and this is also beyond dispute, in spite of suggestions to the contrary in the letter of Hiller et al. [4]. The questions that are under discussion concern the mechanism that brings about these splice variations and their “functional consequences” or “role in biological functions.” The rather vague formulation of these questions has, in our opinion, given rise to much misunderstanding. Therefore, to be concrete, we list what we believe are the main relevant questions. (1) Why are these splice variations so common? By what mechanism are they brought about? (2) To what extent is the introduction of these variations controlled and regulated by the cell? (3) What

fraction of these variations is neutral and what fraction affects protein function? (4) To what extent are the non-neutral variations deleterious and to what extent are they beneficial?

With respect to the first question, we have shown [5] that one need not invoke a complicated mechanism for introducing these variations, but that a simple model of stochastic binding of the spliceosome to competing splice sites, in combination with nonsense-mediated decay, can fully explain the abundance of these variations. Moreover, this model accurately predicts the relative frequencies of all small length variations, not only at acceptor but also at donor splice sites. As Hiller et al. also stress in their letter, there is little doubt that thermodynamic fluctuations, i.e., noise, play a role in splice-site selection. The combination of these facts suggests to us that thermodynamic noise is responsible for introducing a large fraction of the observed alternative splicing events at tandem acceptors.

With respect to the second question, if the introduction of splice variation at NAGNAG acceptors were highly controlled by the cell, then one would not expect that they could be predicted from the local sequence at the splice site only. The fact that our same simple model successfully predicts which NAGNAG acceptors show splice variation and which do not suggests that at least a substantial fraction of all such splice variations are not tightly controlled by the cell. We agree with Hiller et al. that our model cannot explain the observed cases of variation in the relative proportion of the alternative splice forms across different tissues. We disagree, however, that this invalidates our model for these NAGNAGs. Just as different point mutations occur at different rates in different cellular states and sequence contexts, so may the relative probabilities with which the spliceosome binds to competing splice sites depend on details of the kinetics that may vary between tissues. It remains to be determined if the cells are able to actively regulate kinetic details so as to specifically regulate alternative splicing at tandem acceptor sites. In fact, we feel that one of the main uses of our model is to provide a baseline expectation under simple thermodynamic noise, allowing one to more effectively identify interesting cases that deviate significantly from this behavior.

With respect to questions 3 and 4, it is of course to be expected that some of the variations affect protein function. Indeed, Hiller et al. [3] have provided several lines of evidence that indicate a bias toward alternative NAGNAG acceptors that minimize the effect on the proteins. We agree with Hiller et al. that this strongly suggests that, at least in some cases, the effects of NAGNAG variations are deleterious and that selection acts to avoid them. We strongly disagree, however, that this argues against noise being responsible for *introducing* these variations. By the same reasoning one could argue that point mutations are not introduced by noise because one observes negative selection against certain single point mutants. Rather, the observed selection against NAGNAG motifs in locations where splice variation would deleteriously affect protein function suggests that the splice variation at NAGNAG acceptors is not under tight control of the cell, and supports the idea that these variations are mostly the result of uncontrollable noise. Finally, the fact that some variations deleteriously affect protein function does not imply that all these variations play a “role in biological function.” In many cases some amount of deleterious variation might just be tolerated.

How frequent are cases in which variations are beneficial for the cell, i.e., in which the cell uses both functionally different forms? We agree with Hiller et al. that such cases remain to be identified, but do not agree that the problem lies with the general difficulty of showing signs of positive selection. Positive selection is typically used to refer to cases where selection has favored *change* at particular positions. In contrast, to show that NAGNAG variations are beneficial, one would need to show only that there is clear selection for conserving the tandem acceptor property of variant NAGNAGs. This was in fact precisely the purpose of our test that compared the conservation of variant NAGNAG acceptors with that of invariant NAGNAG acceptors. Hiller et al. call this test “probably biased” due to a substantial fraction of NAGGAG tandem acceptors in which the GAG is part of the “conserved exon.” The point that we may not have stressed enough [5], and that is apparently not appreciated by Hiller et al., is that if there is selection for maintaining a NAGNAG acceptor that supports splice variation, then *both* AG dinucleotides need necessarily to remain conserved. This selection pressure is stronger even than the selection pressure on NAGs that are part of the exon, where selection will chiefly operate at the level of their coding potential, often allowing for neutral mutation of the AG dinucleotide. Thus, NAGNAGs at invariant acceptors must necessarily be under less selection to conserve both AG dinucleotides than beneficial variant NAGNAGs. If a substantial proportion of the variant NAGNAGs were under selection for their tandem acceptor property, then we would expect to see their NAGNAG property more often conserved than for invariant NAGNAGs. Since we do not observe this, we conclude that the fraction of NAGNAGs under selection for retaining their tandem acceptor function cannot be very large. Finally, Hiller et al. discuss the conservation test that they performed [3] and mention the conservation statistics obtained more recently by

Akerman and Mandel-Gutfreund [6]. In Text S1 we discuss our interpretation of both these conservation tests. ■

Erik van Nimwegen
Mihaela Zavolan (mihaela.zavolan@unibas.ch)
University of Basel
Basel, Switzerland

Supporting Information

Text S1. Conservation Patterns at NAGNAG Acceptor Sites

Found at doi:10.1371/journal.pgen.0020208.sd001 (92 KB PDF).

References

1. Zavolan M, Kondo S, Schonbach C, Adachi J, Hume D, et al. (2003) Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res* 13: 1290–1300.
2. Sugnet CW, Kent WJ, Ares M Jr, Haussler D (2004) Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac Symp Biocomput* 2004: 66–77.
3. Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, et al. (2004) Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nat Genet* 36: 1255–1257.
4. Hiller M, Szafranski K, Backofen R, Platzer M (2006) Alternative splicing at NAGNAG acceptors: Simply noise or noise and more? *PLoS Genet* 2: doi:10.1371/journal.pgen.0020207
5. Chern TM, van Nimwegen E, Kai C, Kawai J, Carninci P, et al. (2006) A simple physical model predicts small exon length variations. *PLoS Genet* 2: doi:10.1371/journal.pgen.0020045
6. Akerman M, Mandel-Gutfreund Y (2006) Alternative splicing regulation at tandem 3' splice sites. *Nucleic Acids Res* 34: 23–31.

Citation: van Nimwegen E, Zavolan M (2006) Alternative splicing at NAGNAG acceptors: Authors' reply. *PLoS Genet* 2(11): e.208. doi:10.1371/journal.pgen.0020208

Copyright: © 2006 van Nimwegen and Zavolan. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors received no specific funding for this article.

Competing Interests: The authors have declared that no competing interests exist.