Microexons Go Big

Li Yang[1,*] and Ling-Ling Chen[2,3,*]
[1]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology; CAS Center for Excellence in Brain Science
[2]State Key Laboratory of Molecular Biology, Shanghai Key Laboratory of Molecular Andrology, Institute of Biochemistry and Cell Biology Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[3]School of Life Science and Technology, ShanghaiTech University, Shanghai, 200031, China
*Correspondence: liyang@picb.ac.cn (L.Y.), linglingchen@sibcb.ac.cn (L.-L.C.)
http://dx.doi.org/10.1016/j.cell.2014.12.004

Microexons are frequently underestimated in transcriptome analyses. Two studies published in *Cell* and *Genome Research* now independently report the identification of hundreds of microexons. Alternative splicing of some microexons is regulated by neuronal-specific RNA-binding proteins and modifies the function of proteins involved in neurogenesis, with misregulation linked to autism.

Nearly all human multi-exonic genes undergo alternative splicing (AS) to produce more than one mature mRNA, thus greatly expanding transcriptomic complexity and functional diversity (Nilsen and Graveley, 2010). Precise annotation of all AS events is essential for better understanding this repertoire under both physiological and pathological conditions, but the combinatorial aspect of this problem has been a major challenge. This problem is exacerbated in the case of microexons, exons less than 51 nt, which have often been overlooked because their short length makes them computationally difficult to identify. Microexons are thought to be unfavorable for splicing because they lack sufficient exonic splicing enhancers and they are so short that the splicing machinery cannot physically assemble at both the 3′ and 5′ splice sites (Black, 1991; Blencowe, 2000; Fairbrother et al., 2002). Individual studies in mammals reported an important role for microexons in the brain (Carlo et al., 2000; Zibetti et al., 2010), but the wider role of microexons and the rules governing their splicing have remained unclear. Now, two independent papers by Irimia et al. (2014) in this issue of *Cell* and by Li et al. (2015) in *Genome Research* uncover hundreds of highly conserved microexons from RNA-seq data sets across species, outline the features regulating the inclusion of these microexons, and show that many of these impact neurogenesis and brain function.

To assess the contribution of microexons to the transcriptome, Irimia et al. (2014) develop a multi-module analysis pipeline to systematically define all neural-regulated AS patterns, especially microexons with very short lengths (3–15 nt), from more than 100 different human and mouse cell and tissue types. They show that the regulation of microexons is highly dynamic during neuronal differentiation (Figure 1). Strikingly, although microexons represent only 1% of AS observed, they constitute up to one-third of all conserved neural-regulated AS between human and mouse. The inclusion in the final transcript of most identified neural microexons is regulated by a brain-specific factor, nSR100, which binds to intronic enhancer UGC motifs close to the 3′ splice sites. Of particular interest, these microexons are enriched for lengths that are multiples of 3 nt and are thus highly likely to produce alternative protein isoforms if included or excluded from the final transcript. The authors further provide several lines of evidence, both computational and experimental, to demonstrate that inclusion of microexons can modulate the function of interaction domains of proteins involved in neurogenesis (Figure 1). Interestingly, misregulation of neural-specific AS microexons is observed in individuals with autism spectrum disorder (ASD).

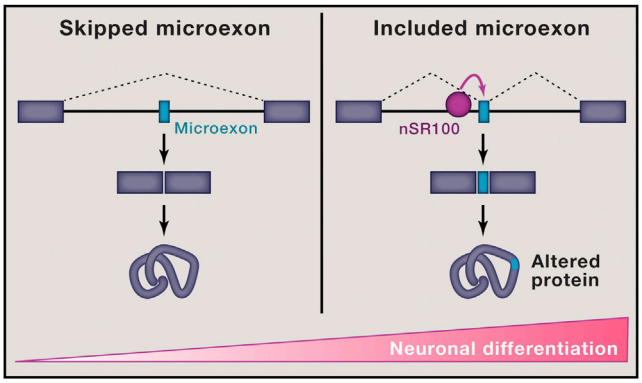

Figure 1. Alternatively Spliced Microexons Are Often Included during Neurogenesis along with the Expression of Neural-Specific Splicing Factors
Neural-specific microexons generally possess weak genomic features for splicing, such as unfavorable 3′ splice sites, which leads to skipping at early stages of neuronal differentiation (left). During neurogenesis, some neuronal-specific splicing regulators, such as nSR100, become highly expressed and are recruited to intronic splicing enhancer regions near suboptimal 3′ splice sites to promote inclusion of microexons. Inclusion leads to the addition of small numbers of amino acid residues in their protein products, which can alter protein-protein interactions (right).

Li et al. (2015) approach the issue in a different way. They treat microexons as insertions between annotated splice junctions to retrieve a set of microexons that are shorter than 51 nt, including both constitutively spliced (CS) and AS microexons, from more than 900 human and mouse samples, nearly half of which were from brain tissues. The authors find that AS microexons are evolutionarily conserved and exhibit tissue-specific inclusion. Furthermore, they show that AS of

microexons mediated by specific RNA-binding proteins (RBPs), such as RBFox and PTBP1, may alter protein sequences, thus leading to changes in protein-protein interactions.

In contrast to CS microexons that possess strong *cis* elements to enhance splicing (Li et al., 2015), AS microexons require additional interactions with the splicing machinery, which are usually enhanced by RBPs. Interestingly, many brain-specific microexons might be regulated by a single RBP acting as a master splicing regulator. For instance, Irimia et al. (2014) demonstrate that neural-specific factor nSR100 promotes the AS of very short microexons during neurogenesis (Figure 1), and Li et al. (2015) confirm that most brain-specific microexons are enhanced by tissue-specific RBFox proteins. It is noteworthy that RBFox1-dependent AS has been implicated in ASD (Voineagu et al., 2011). These two new reports focus on different *trans*-acting factors, and questions remain, including the extent of the overlap of these two data sets and whether these distinct splicing regulators act independently or in concert to regulate AS of microexons in the brain. In addition, it will be of interest to identify other RBPs or master RBPs that can regulate AS microexons in different tissues or under different conditions.

Beyond the discovery of these surprising rules for the regulation of microexon splicing, it is of particular significance that these two studies (Irimia et al., 2014; Li et al., 2015) demonstrate that alternative inclusion of microexons generates proteins with altered functions in neurogenesis. Whereas many microexons introduce short stretches of amino acids that alter protein-protein interactions, others may introduce novel charged regions or new platforms for post-translational modification. Not all lead necessarily to changes on the protein surface—one might envision that some microexon AS results in subtle alterations in protein folding or catalytic function. Furthermore, microexons can change the properties of the mRNA, altering its structure, stability, or subcellular location. Given the myriad ways that microexons can exert their influence, it is likely that they may have tissue-specific functions in other organs, and their mis-regulation may correlate with disease, as was observed for neuronal-specific microexons and ASD (Irimia et al., 2014). Also, as neural-specific microexon splicing is highly conserved during evolution at both the levels of genomic sequence and tissue-specific inclusion pattern (Irimia et al., 2014; Li et al., 2015), it will be of interest to study how selection acts on microexons. Together, these reports of the identification and impact of microexons demonstrate the feasibility of computationally probing transcriptome for previously hidden information and begin to outline the mechanisms used by the cell to achieve the rich complexity of protein-protein interactions that govern tissue-specific processes.

ACKNOWLEDGMENTS

L.Y. is supported by 31271390 from NSFC. L.-L.C. is supported by XDA01010206 from CAS and 31322018 from NSFC.

REFERENCES

Black, D.L. (1991). Genes Dev. *5*, 389–402.

Blencowe, B.J. (2000). Trends Biochem. Sci. *25*, 106–110.

Carlo, T., Sierra, R., and Berget, S.M. (2000). Mol. Cell. Biol. *20*, 3988–3995.

Fairbrother, W.G., Yeh, R.F., Sharp, P.A., and Burge, C.B. (2002). Science *297*, 1007–1013.

Irimia, M., Weatheritt, R.J., Ellis, J., Parikshak, N.N., Gonatopoulos-Pournatzis, T., Babor, M., Quesnel-Vallières, M., Tapial, J., Raj, B., O'Hanlo, D., et al. (2014). Cell *159*, this issue, 1511–1523.

Li, Y.I., Sanchez-Pulido, L., Haerty, W., and Ponting, C.P. (2015). Genome Res. Published online December 18, 2014. http://dx.doi.org/10.1101/gr.181990.114.

Nilsen, T.W., and Graveley, B.R. (2010). Nature *463*, 457–463.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Nature *474*, 380–384.

Zibetti, C., Adamo, A., Binda, C., Forneris, F., Toffolo, E., Verpelli, C., Ginelli, E., Mattevi, A., Sala, C., and Battaglioli, E. (2010). J. Neurosci. *30*, 2521–2532.