

## Gene regulation: The origin of neural microexons

Rupert Faraway<sup>1,2</sup> and Jernej Ule<sup>1,2</sup>

1 The Francis Crick Institute, London NW1 1AT, United Kingdom

2 Department for Neuromuscular Diseases, UCL Institute of Neurology, London WC1N 3BG, United Kingdom

Comparative analysis of microexons across bilaterians identified a new protein domain associated with the evolutionary origin of microexon inclusion in neural tissues.

Microexons, have played a central role in the evolution of biology's most impressive innovation: the vertebrate brain. They can be as short as 3 nucleotides and normally lead to addition of only one or a few amino acids into the encoded proteins, which have diverse cellular functions, including vesicle transport, cytoskeleton organization, and chromatin regulation<sup>1</sup>. Most microexons are exclusively included in neuronal tissues, where they play a crucial role in development and maintenance of neuronal functions<sup>1,2</sup>. Notably, misregulation of activity-dependent splicing of microexons contributes to autism<sup>1,3,4</sup>. Consistent with their functional importance, neural microexons are evolutionarily conserved, yet their evolutionary origin remains poorly understood.

Writing in *Nature Ecology & Evolution*, Torres-Méndez et al. report that the regulatory programs that determine the inclusion of microexons in the neuronal transcriptome is at least 600 million years old. Using tissue-specific RNA sequencing, the authors report increased inclusion of microexons in the transcriptomes of neural tissues in various bilaterians, including non-vertebrates fruitfly (*Drosophila melanogaster*), centipede (*Strigamia maritima*) and amphioxus (*Branchiostoma lanceolatum*). The authors identify several microexons that are shared by all phyla, as well as a multitude of shared sequence features that are shared between neural microexons between vertebrates and non-vertebrates.

Studies in mammals have previously shown a central role for the neuronal protein SRRM4 (also known as nSR100) in promoting the neuron-specific inclusion of microexons<sup>1,5</sup>. Torres-Méndez et al. identify a previously unannotated domain in the C-terminus of SRRM4, the 'enhancer of microexons' (eMIC), that is essential for its function (Fig. 1). Expression of the eMIC-containing C-terminus is sufficient to promote microexon inclusion when transfected into a non-neuronal human HEK293 cell line, while deletion of the eMIC domain abrogates the function of SRRM4. The authors identify the interacting partners of multiple SRRM4 domains via affinity purification coupled mass-spectrometry. Interestingly, the N-terminal end of SRRM4 interacts with members of the exon-junction complex, which have recently been shown to also play an important role in enhancing microexon inclusion<sup>6</sup>. The eMIC domain, however, primarily interacts with SF1 and the U2AF2 proteins, two factors that play a role in the initial steps of spliceosome assembly. *In vitro* splicing experiments showed that eMIC-containing peptides enhance the formation of earliest stages of splicing complex formation by promoting the recruitment of SF1 and U2AF2.

There are two vertebrate paralogues of the *Srrm4* gene: the broadly expressed *Srrm2* and the neuron-specific *Srrm3*. Torres-Méndez et al. trace this family back to a single pan-eukaryotic gene, referred to as *Cwc21* in yeast. The gene lacks the eMIC domain in all non-bilaterians, as well as in the most basal, non-nephrozoan bilaterian clade (*Xenacoelomorpha*), which lacks the brain. The eMIC domain first emerged in the non-vertebrate nephrozoan bilaterians, where it is present in the gene referred to as *Srrm2/3/4*. Multiple RNA isoforms of *Srrm2/3/4* gene are made in non-vertebrates, such that the isoforms containing the eMIC domain are expressed only in neuronal tissues. These isoforms are produced through a fascinating combination of alternative promoter usage, splicing and polyadenylation (Fig. 1). In vertebrates, however, the domain is primarily expressed in the brain due to transcriptional control of *Srrm3* and *Srrm4* genes, both of which contain the eMIC domain (Fig. 1). To demonstrate the universal function of the eMIC domain across evolution, the authors transfect the human HEK293 cell line with the various isoforms of *Srrm2/3/4* orthologues from fruitfly and amphioxus, showing that only the isoforms that contain the eMIC domain are capable of promoting microexon inclusion. While the *Nematostella Srrm2* lacks the eMIC domain, addition of the human eMIC domain from *SRRM4* converts it into a protein capable of promoting microexon inclusion.

Torres-Méndez et al. study microexons from a broad evolutionary perspective spanning hundreds of millions of years, and employ a detailed biochemical perspective of protein-protein interactions and spliceosomal recruitment. Combined, these approaches reveal not only a much wider role for microexons in the animal kingdom, but also an important lesson about the evolution of gene-regulatory programmes. The eMIC appears to have emerged through de novo evolution and is predicted to be intrinsically disordered, which is in line with previous observations that intrinsically disordered proteins and alternatively spliced disordered domains often drive the emergence and inheritance of biological traits<sup>7-9</sup>. It is likely that neuron-specific programme first evolved through regulation of alternative RNA isoforms of *Srrm2/3/4* containing the eMIC domain, which was then consolidated in vertebrates at the level of the genome through duplications of this gene. Ultimately, Torres-Méndez and colleagues' study shows how small changes in a tissue-specific splicing factor drove the evolution of many tiny microexons, which in turn led to major functional innovations in the brain proteome.

Figure 1: The evolutionary origins of neuronal microexons.

Torres-Mendez et al. have traced back the evolution of neuronal microexons to a common bilaterian ancestor that existed over 600 million years ago. This neuronal microexon programme is facilitated by the emergence of the previously uncharacterised eMIC domain, the expression of which is restricted to the brain. This tissue specificity is achieved first by neuron-specific alternative splicing, and subsequently in vertebrates by gene duplication and specialisation.

1. Irimia, M. *et al.* A highly conserved program of neuronal microexons is misregulated in

- autistic brains. *Cell* **159**, 1511–1523 (2014).
2. Li, Y. I., Sanchez-Pulido, L., Haerty, W. & Ponting, C. P. RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res.* **25**, 1–13 (2015).
  3. Quesnel-Vallières, M., Irimia, M., Cordes, S. P. & Blencowe, B. J. Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev.* **29**, 746–759 (2015).
  4. Quesnel-Vallières, M. *et al.* Misregulation of an Activity-Dependent Splicing Network as a Common Mechanism Underlying Autism Spectrum Disorders. *Mol. Cell* **64**, 1023–1034 (2016).
  5. Raj, B. *et al.* A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol. Cell* **56**, 90–103 (2014).
  6. Gonatopoulos-Pournatzis, T. *et al.* Genome-wide CRISPR-Cas9 Interrogation of Splicing Networks Reveals a Mechanism for Recognition of Autism-Misregulated Neuronal Microexons. *Mol. Cell* **72**, 510–524.e12 (2018).
  7. Buljan, M. *et al.* Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* **46**, 871–883 (2012).
  8. Ellis, J. D. *et al.* Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* **46**, 884–892 (2012).
  9. Chakrabortee, S. *et al.* Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. *Cell* **167**, 369–381.e12 (2016).

Figure 1

