Evolutionary genomics

https://doi.org/10.1038/s41559-022-01942-5

De novo gene increases brain size

April Rich & Anne-Ruxandra Carvunis

Comparative analysis of human and macaque brain transcripts together with experiments in mice and in a cortical organoid model show the de novo emergence of a hominoid-specific protein-coding gene implicated in brain development. The evolution of RNA nuclear export signals enabled a new protein to become translated from an ancestral long-noncoding RNA locus.

The de novo emergence of protein-coding genes from ancestrally noncoding sequences was long believed to be nearly impossible[1]. However, many de novo genes have now been discovered that are of high interest because of their potential contribution to novel species-specific traits[2]. With the advent of next-generation sequencing, it has become clear that most of the genome is pervasively transcribed[3] and riddled with random start and stop codons[4]. Any sequence of triplets of nucleotides contained between start and stop codons, or an open reading frame (ORF), can in theory be translated. In fact, ribosome sequencing studies have shown that many are[5–7]. Therefore, noncoding transcripts can, if they contain random ORFs, provide the seeds for de novo gene birth[7]. Many studies over the past decade have analysed how the evolution of ORFs and transcripts contribute to de novo gene emergence[2]. Yet a glaring question has not been addressed: why aren't all transcribed random ORFs translated? Writing in *Nature Ecology & Evolution*, An et al.[8] show that acquiring the ability to leave the nucleus is a major hurdle for transcripts to cross towards becoming a de novo protein-coding gene.

An et al.[8] identify 74 human- or hominoid-specific de novo genes that transitioned from long noncoding RNAs (lncRNA) to protein-coding transcripts (mRNAs) through the acquisition of nuclear export signals. Unlike mRNAs, lncRNAs tend to localize primarily in the nucleus where they cannot be translated into proteins. An et al.[8] show that ancestral lncRNAs acquired sequence substitutions that allowed efficient export to the cytoplasm, where the transcripts became accessible to the ribosome for translation. Specifically, these sequence substitutions either increased splicing efficiency or decreased U1 recognition sites that mediate transcript tethering to chromatin[9]. For each of the 74 de novo genes, an orthologous lncRNA exists in macaques that did not evolve the signals that enable efficient nuclear export and translation. Interestingly, these orthologous lncRNAs in macaques have high GC content compared to other lncRNAs and even to protein-coding genes. The nucleotide substitutions that enable active nuclear export of these transcripts in humans and hominoids, but not in macaques, are evolutionarily fixed under selective constraints. An et al.'s analyses[8] suggest a new mechanistic model through which atypical lncRNAs can cross the boundary between lncRNA and mRNA to give rise to de novo protein-coding genes (Fig. 1).

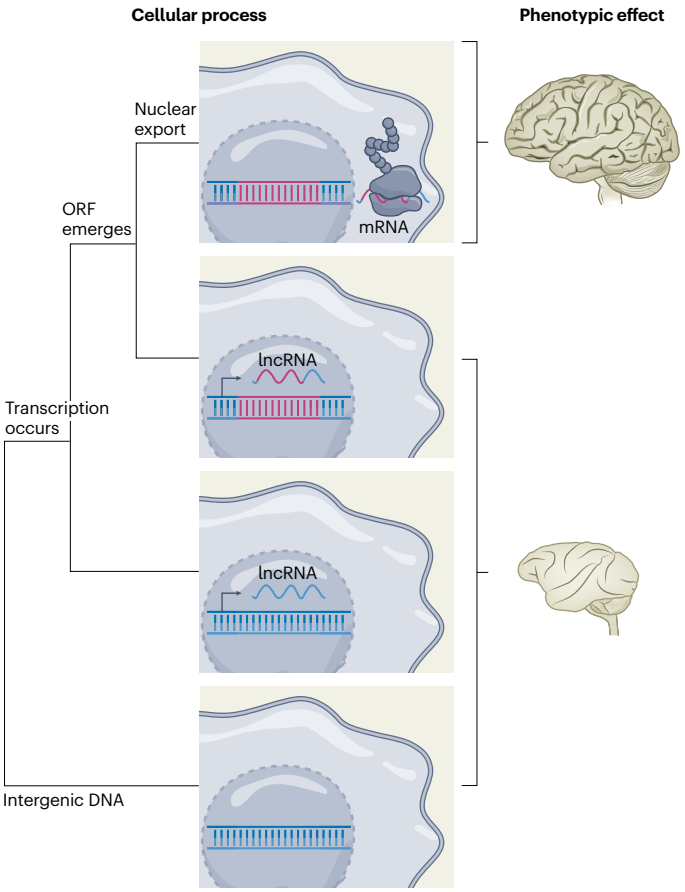

Fig. 1 | Model for de novo gene birth at the *ENSG00000205704* loci. Once the ancestral lncRNA locus has acquired the ability to efficiently exit the nucleus, the translation of the human de novo gene slows the rate of brain development and increases brain size.

An et al.[8] illustrate this mechanism by focusing on one hominoid de novo gene. The authors show experimentally that evolution of efficient nuclear export of *ENSG00000205704* may have had a role in human brain evolution. Using gene-editing tools in human neural progenitor cells, the authors show the importance of splicing and U1 recognition sites for cytoplasmic localization of the transcript. In a cortical organoid model, overexpression of this gene results in slower development times and increased organoid size, whereas deletion of this gene results in faster development and smaller size. Ectopic expression of *ENSG00000205704* in mice results in an enlarged cortex. Although many studies have shown that novel genes have biased expression towards brain regions compared to older genes[10,11], An et al.[8] provide evidence for phenotypic effects of a de novo gene on human brain growth and development.

The work presented by An et al.[8] is a pioneering illustration of the importance of RNA processing in de novo gene origination. Future studies investigating different RNA processing mechanisms, as well as later stages in the production of proteins, may yield further insights into how de novo genes emerge and fix. Once a transcript encoding a de novo gene exits the nucleus, it still faces many challenges. For example, might selection further refine how it associates with the ribosome; how the resulting polypeptide receives adequate modifications and proper folding; how it comes to localize to subcellular compartments; and how it influences phenotype — all while evading degradation by the quality control mechanisms of the cell?[12] The model provided by An et al.[8] paves the way towards further understanding of de novo gene birth from an evolutionary cell-biology perspective.

April Rich & Anne-Ruxandra Carvunis ✉
Department of Computational and Systems Biology, Pittsburgh Center for Evolutionary Biology and Medicine, University of Pittsburgh Medical School, Pittsburgh, PA, USA.
✉e-mail: anc201@pitt.edu

Published online: 02 January 2023

References

1. Jacob, F. *Science* **196**, 1161–1166 (1977).
2. Van Oss, S. B. & Carvunis, A.-R. *PLoS Genet.* **15**, e1008160 (2019).
3. Neme, R. & Tautz, D. *eLife* **5**, e09977 (2016).
4. Couso, J.-P. & Patraquim, P. *Nat. Rev. Mol. Cell Biol.* **18**, 575–589 (2017).
5. Mudge, J. M. et al. *Nat. Biotechnol.* **40**, 994–999 (2022).
6. Ruiz-Orera, J. & Albà, M. M. *Trends Genet.* **35**, 186–198 (2019).
7. Carvunis, A.-R. et al. *Nature* **487**, 370–374 (2012).
8. An, N. A. et al. *Nat. Ecol. Evol.* https://doi.org/10.1038/s41559-022-01925-6 (2023).
9. Yin, Y. et al. *Nature* **580**, 147–150 (2020).
10. Zhang, Y. E., Landback, P., Vibranovski, M. D. & Long, M. *PLoS Biol.* **9**, e1001179 (2011).
11. Wu, D.-D., Irwin, D. M. & Zhang, Y.-P. *PLoS Genet.* **7**, e1002379 (2011).
12. Parikh, S. B. et al. *Yeast* **39**, 471–481 (2022).

Competing interests

The authors declare no competing interests.