

Preview

Duplicate, diversify, repeat: The evolution of *NOTCH2NL*

Nachshon Egyes¹ and David Gokhman^{1,*}¹Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel*Correspondence: david.gokhman@weizmann.ac.il<https://doi.org/10.1016/j.xgen.2026.101274>

Segmental duplications are key drivers of evolutionary innovation but are also particularly challenging to study. In this issue, Real et al. resolve the genetic diversity, structural history, and regulatory landscape of the human-specific *NOTCH2NL* gene family, a likely contributor to human cortical brain expansion.

The expansion of the cerebral cortex is among the most distinctive features of human evolution, yet the genetic changes that enabled it—and the selective pressures that fixed them—remain poorly understood. Comparative genomics has increasingly implicated structural variation, and in particular segmental duplications, as a key substrate for the emergence of lineage-specific paralogs and gene regulation.¹ However, the forces driving and shaping these duplications, and their functional consequences, remain unclear.

A compelling case is the *NOTCH2NL* (*NOTCH2-N-terminus-like*) family. Fiddes and colleagues,² together with Suzuki and colleagues,³ previously showed that partial duplications of *NOTCH2* gave rise to four human-specific paralogs that interfere with Notch signaling, prolonging self-renewal of cortical progenitors and delaying their differentiation into neurons—a plausible mechanism for increasing neuronal output during fetal cortical neurogenesis.^{2,3} However, partial *NOTCH2* duplications are also present in nonhuman great apes (Figure 1). This leaves open questions regarding (1) when each paralog actually arose, (2) whether the underlying duplications occurred independently in different lineages, and (3) whether the paralogs differ in function, regulation, or patterns of variation within present-day humans or between humans and nonhuman apes. These questions have been difficult to address because the region is large, repetitive, and nearly identical across copies—an arena in which short-read sequencing struggles.

In this issue, Real and colleagues⁴ address these questions by combining

near-complete long-read assemblies of human and nonhuman ape genomes,^{5,6} long-read transcript sequencing, and Fiber-seq—a long-read method that resolves chromatin accessibility at the level of individual DNA molecules and can therefore disentangle nearly identical paralogs.⁷

To reconstruct the duplication history of the *NOTCH2NL* family, the authors compared 12 nonhuman ape haploid assemblies with the human T2T reference and built a maximum-likelihood phylogeny from alignments of *NOTCH2NL* intronic sequences across human, chimpanzee, bonobo, gorilla, and orangutan. The human paralogs form a monophyletic clade to the exclusion of all nonhuman ape sequences, with parallel lineage-specific expansions in gorilla and in the chimpanzee-bonobo branch—a topology most parsimoniously explained by recurrent, independent duplication (Figure 1). Reinforcing this picture, none of the nonhuman ape *NOTCH2* duplications carry the four-base-pair deletion in the terminal exon that is required for stable protein production in human paralogs^{2,4}; copies predicted to encode a stable protein are therefore confined to humans (Figure 1). Molecular dating placed the diversification of these protein-coding human copies at roughly 2.2 to 3.7 million years ago. This broadly coincides with the first documented increases in hominin endocranial volume⁸ and parallels the trajectories of other human-specific duplicated gene families implicated in cortical expansion. Nevertheless, inter-locus gene conversion (IGC) between paralogs—the unidirectional transfer of genetic material between homologous

sequences—can affect the reliability of these time estimates and thus of conclusions relating to the lineage specificity of these events.

To further elucidate the functional roles of the *NOTCH2NL* paralogs, the authors go on to investigate genetic variation and potential IGC events within this gene family in present-day humans. To this end, they leveraged 70 human haploid genome assemblies from diverse populations and developed a tripartite workflow based on both coding and noncoding sequence similarity among paralogs, as well as on similarity between the genomic regions flanking them. Disagreements between these three assignments were used to flag IGC events,⁹ which proved pervasive, affecting roughly 28% of haplotypes and giving rise to eleven distinct structural configurations (Figure 1). *NOTCH2* and *NOTCH2NLA* were the only paralogs present in every haplotype examined, hinting at a mechanistic or selective bias in conversion direction. The same workflow uncovered a previously undescribed version of *NOTCH2NLR*, termed *NOTCH2tv* (truncated version), in which a tract of ancestral *NOTCH2* sequence has overwritten part of the *NOTCH2NLR* pseudogene. Despite acquiring a *NOTCH2*-like promoter and N terminus, *NOTCH2tv* fails to produce a stable protein in cell-based assays—illustrating that IGC can remodel duplicated sequences but not necessarily resurrect them as functional genes.

To study the regulation of these paralogs, the authors generated Fiber-seq and full-length transcript data from dorsal forebrain organoids derived from a heterozygote individual carrying both the *NOTCH2tv* and *NOTCH2NLR* versions,



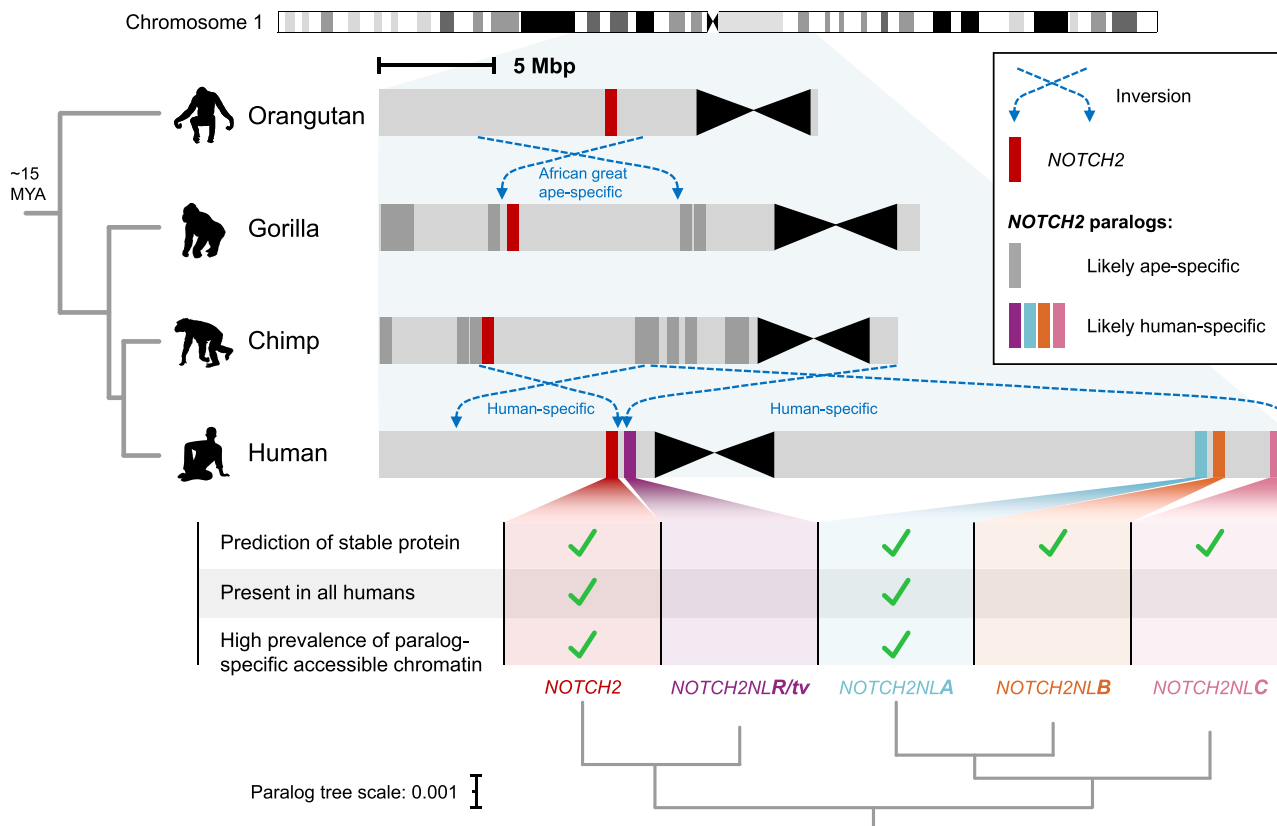


Figure 1. Evolutionary history and functional analysis of the *NOTCH2NL* gene family in great apes

Paralog tree scale bar indicates 0.001 substitutions per site, as inferred by Real et al.⁴ using 21 kbp from intron 2. MYA, million years ago.

allowing all paralogs to be compared within a single genomic background. Most accessible regulatory elements within several hundred kilobases of *NOTCH2NL* were shared across paralogs, but their degree of accessibility frequently differed, and a meaningful minority—roughly one in eight—were paralog specific. These paralog-specific elements were concentrated around the two paralogs present in all human haplotypes, *NOTCH2* and *NOTCH2NLA* (Figure 1). Predicted three-dimensional genome organization mirrored this pattern, with each paralog occupying a distinct local environment despite the high underlying sequence identity—echoing observations that human-accelerated regions tend to be associated with rewired topological domains and altered chromatin contacts.¹⁰

Taken together, the work casts *NOTCH2NL* as a vivid illustration of a broader theme: unstable, duplication-prone genomic regions can serve as engines of adaptive innovation while simultaneously predisposing to disease—the

same segmental duplications implicated in cortical expansion also underlie 1q21.1 distal duplication/deletion and neuronal intranuclear inclusion disease, which are characterized by neurodevelopmental abnormalities and chronic neurodegeneration, respectively.

This study skillfully delineates the genetic architecture at the *NOTCH2NL* loci within great apes and humans and establishes a framework for studying some of the most complex evolutionary dynamics in our genome. However, one important caveat is that IGC itself blurs duplication timing, since the divergence being measured may reflect the last conversion event rather than the original duplication, and equally complicates inference of independent versus shared duplication across ape lineages.

Lastly, several intriguing questions remain open. Specifically, we are yet to uncover the selective forces that shaped the apparent fixation of *NOTCH2NLA*, how the paralog-specific accessible chromatin regions affect regulation, and the extent to

which these paralogs perform genuinely distinct functions. Notwithstanding, the work provides a rare and comprehensive view of a locus that sits at the intersection of human evolution and disease, opening the door to systematic investigation of duplicated genomic regions that have long remained intractable.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used Claude (Opus 4.6, Anthropic) in order to aid in the editing and refinement of the manuscript text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

REFERENCES

1. Marques-Bonet, T., Girirajan, S., and Eichler, E.E. (2009). The origins and impact of primate

- segmental duplications. *Trends Genet.* 25, 443–454. <https://doi.org/10.1016/j.tig.2009.08.002>.
2. Fiddes, I.T., Lodewijk, G.A., Mooring, M., Bosworth, C.M., Ewing, A.D., Mantalas, G.L., Novak, A.M., van den Bout, A., Bishara, A., Rosenkrantz, J.L., et al. (2018). Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. *Cell* 173, 1356–1369.e22. <https://doi.org/10.1016/j.cell.2018.03.051>.
3. Suzuki, I.K., Gacquer, D., Van Heurck, R., Kumar, D., Wojno, M., Bilheu, A., Herpoel, A., Lambert, N., Cheron, J., Polleux, F., et al. (2018). Human-Specific *NOTCH2NL* Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. *Cell* 173, 1370–1384.e16. <https://doi.org/10.1016/j.cell.2018.03.067>.
4. Real, T.D., Hebbar, P., Yoo, D., Antonacci, F., Pačar, I., Dubocanin, D., Diekhans, M., Mikol, G.J., Popoola, O.G., Mallory, B.J., et al. (2026). Genetic diversity and regulatory features of human-specific NOTCH2NL duplications. *Cell Genom.* 6, 101194. <https://doi.org/10.1016/j.xgen.2026.101194>.
5. Yoo, D., Rhie, A., Hebbar, P., Antonacci, F., Logsdon, G.A., Solar, S.J., Antipov, D., Pickett, B.D., Safonova, Y., Montinaro, F., et al. (2025). Complete sequencing of ape genomes. *Nature* 641, 401–418. <https://doi.org/10.1038/s41586-025-08816-3>.
6. Liao, W.-W., Asri, M., Ebler, J., Doerr, D., Hukness, M., Hickey, G., Lu, S., Lucas, J.K., Monlong, J., Abel, H.J., et al. (2023). A draft human pangenome reference. *Nature* 617, 312–324. <https://doi.org/10.1038/s41586-023-05896-x>.
7. Stergachis, A.B., Debo, B.M., Haugen, E., Churchman, L.S., and Stamatoyannopoulos, J.A. (2020). Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* 368, 1449–1454. <https://doi.org/10.1126/science.aaz1646>.
8. Aiello, L., and Dean, C. (1990). *An Introduction to Human Evolutionary Anatomy* (Academic Press).
9. Chen, J.-M., Cooper, D.N., Chuzhanova, N., Férec, C., and Patrinos, G.P. (2007). Gene conversion: mechanisms, evolution and human disease. *Nat. Rev. Genet.* 8, 762–775. <https://doi.org/10.1038/nrg2193>.
10. Keough, K.C., Whalen, S., Inoue, F., Przytycki, P.F., Fair, T., Deng, C., Steyert, M., Ryu, H., Lindblad-Toh, K., Karlsson, E., et al. (2023). Three-dimensional genome rewiring in loci with human accelerated regions. *Science* 380, eabm1696. <https://doi.org/10.1126/science.abm1696>.