Cell Genomics

CellPress
OPEN ACCESS

Preview

Illuminating hidden genetic architecture in autism

Maria H. Chahrour[1,2,3,4,5,*]
[1]Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[2]Department of Neuroscience, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[3]Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[4]Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[5]Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
*Correspondence: maria.chahrour@utsouthwestern.edu
https://doi.org/10.1016/j.xgen.2026.101248

Structural and repeat variation relevant to autism has remained largely inaccessible to short-read sequencing. Using long-read genome sequencing, a recent study reveals previously hidden variants at scale, quantifies their contribution to autism risk, and links them to gene regulatory disruption in neurodevelopment, expanding the mutational and mechanistic landscape of autism.

Autism spectrum disorder (ASD) is among the most genetically complex neurodevelopmental conditions, with contributions spanning rare, highly penetrant mutations to common polygenic risk.[1–4] Advances in short-read genome sequencing have transformed gene discovery, particularly for single-nucleotide variants (SNVs) and small insertions or deletions.[1,2] However, much of the genome, rich in repetitive sequences and prone to large-scale rearrangements, has remained difficult to interrogate.[5] As a result, structural variants (SVs) and repeat expansions, despite established roles in neurological disease,[6,7] are likely underrepresented in our current understanding of ASD genetics.

A study by Mortazavi and colleagues leverages long-read genome sequencing to systematically uncover structural and repeat variation in individuals with ASD.[8] By overcoming limitations of short-read technologies, long-read sequencing enables more accurate mapping of complex genomic regions, revealing previously inaccessible or mischaracterized variants (Figure 1).[5] Mortazavi et al. not only expand the catalog of structural variation in ASD but also address a central challenge in human genetics: how to interpret functional consequences of these variants in a regulatory and developmental context.[6,7]

A strength of this work lies in its ability to detect a broader spectrum of variant classes, including large insertions, inversions, tandem repeat expansions, and complex rearrangements. The study also identifies recurrent classes of complex structural variation, including nested duplication-deletion rearrangements, revealing previously unrecognized patterns of mutational architecture.[8,9] These recurrent configurations suggest that specific genomic regions may be intrinsically prone to rearrangement,[7,10] consistent with locus-specific vulnerability in ASD. Such variants are enriched in genomic regions that are highly repetitive or structurally dynamic, features that have hindered alignment and variant calling using short reads.[5] By applying long-read sequencing, the authors identify previously undetected variants, including *de novo* and somatic-mosaic SVs,[5,7] several of which occur in genes and regulatory regions implicated in neurodevelopment.[8] This indicates that estimates of genetic contribution to ASD are constrained, in part, by prior technological limitations. Consistent with this, the authors estimate that rare variants identified through integrated long- and short-read sequencing explain 11.7% of the variance in ASD case status, corresponding to ~7.4% of heritability on the liability scale, with SVs (5.7%) and tandem repeats (3.2%) contributing alongside rare SNVs (4.6%).[8] Inclusion of polygenic risk modestly increases total variance explained to 13.8% (~8.9% heritability),[8] underscoring the extent to which previously inaccessible forms of variation contribute to ASD genetic architecture. Notably, the ability to phase long reads enables detection of somatic-mosaic events arising during early development,[7] which are largely inaccessible to short-read approaches,[5] and adds another layer of genetic heterogeneity.

Beyond detection, the study moves toward functional interpretation, where SVs have lagged behind SNVs. The authors integrate transcriptomic and epigenomic data, including direct measurement of DNA methylation from long-read sequencing, to assess how these variants influence gene expression and regulatory landscapes.[5,8] In some cases, SVs disrupt gene bodies or alter splicing, whereas in others they perturb noncoding regulatory elements, leading to more subtle but potentially significant effects on gene dosage or expression timing. In parallel, joint analysis of phased variants and DNA methylation demonstrates that repeat length can directly modulate local methylation states.[8] These observations link sequence variation and epigenetic regulation (Figure 1). More broadly, the study highlights cases in which structural variation reshapes local regulatory context, altering transcriptional output in genes involved in neuronal development and synaptic function.[8] Examples where repeat expansions or structural rearrangements are associated with coordinated changes in gene expression and chromatin state provide a mechanistic link between genome structure, epigenetic regulation, and ASD phenotypes. Collectively, these findings reinforce that, in ASD genetics, risk often arises through disruption of transcriptional regulation, alternative splicing, chromatin organization, and related gene-regulatory processes rather than coding sequence alone.[1–4]

This work also highlights the diversity of mutational mechanisms contributing to

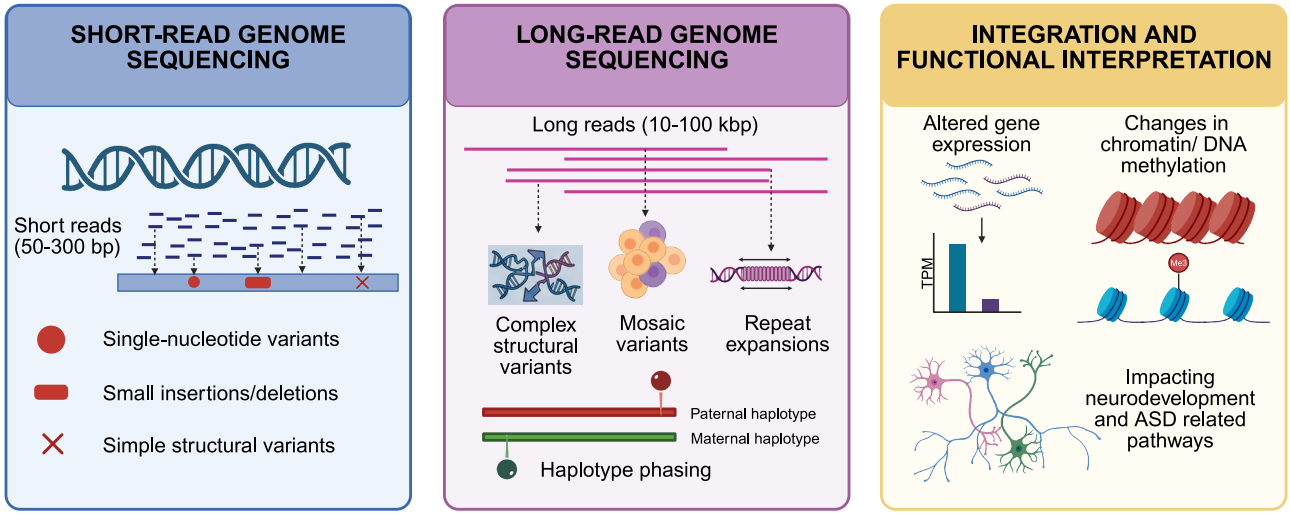

Figure 1. Long-read sequencing reveals multi-layered regulatory disruption from hidden genetic variation in autism

Schematic overview of how long-read genome sequencing expands detection and interpretation of genetic variation in autism. Left: short-read genome sequencing accurately detects single-nucleotide variants and small insertions/deletions but has limited ability to resolve repetitive regions, complex structural variants, and mosaic variation. Middle: long-read genome sequencing enables direct mapping across repetitive and structurally complex regions, revealing structural variants (e.g., insertions, inversions, and nested duplication-deletion rearrangements), repeat expansions, and somatic-mosaic variants, while enabling haplotype phasing. Right: integration with functional genomic data links these variants to gene regulatory disruption across multiple layers, including altered gene expression, changes in chromatin context, and DNA methylation remodeling (e.g., repeat-associated methylation changes), ultimately impacting neurodevelopment and ASD related pathways. This figure was created with BioRender.com.

ASD,[1–4] spanning germline and somatic structural variation, recurrent complex rearrangements, and epigenetic dysregulation linked to DNA methylation. While *de novo* point mutations in constrained genes have been a major focus of the field, these findings underscore that structural variation and repeat instability[6] represent parallel and potentially interacting layers of genetic risk.[1–4] The ability to capture these layers within a single experimental framework represents an advance and raises the possibility that a subset of unsolved ASD cases may be attributable to variants that have remained undetected using conventional approaches.

However, several challenges remain. The interpretation of structural and repeat variants is inherently more complex than that of coding point mutations. Predicting the impact of a large insertion or a repeat expansion on gene regulation requires a more complete understanding of chromatin architecture, three-dimensional genome organization,[7] and cell type-specific regulatory logic; areas that remain incompletely understood. In addition, scalability of long-read sequencing remains a practical consideration. While costs have decreased, the large cohort sizes required for robust gene discovery and genotype-phenotype correlations will necessitate additional technological and computational advances.

Another important question is how these newly identified variants fit into the broader genetic architecture of ASD. Whether structural and repeat variants converge on core ASD pathways or implicate additional, underappreciated processes remains unclear. Early indications suggest both convergence and extension,[8] reinforcing the idea that ASD arises from perturbations across multiple, interacting layers of genome regulation.

Finally, this work prompts a reevaluation of how we define and prioritize "risk variants" in neurodevelopmental disorders. As detection methods improve and larger, more ancestrally diverse cohorts are sequenced, the boundaries between rare and common variation, and between mutations with protein-coding versus regulatory or structural effects, become increasingly blurred. Integrative approaches that combine long-read sequencing with functional assays, particularly in relevant cell types and developmental contexts, will be essential for translating variant discovery into biological insight.

In sum, the application of long-read genome sequencing to ASD represents a technical advance that expands the mutational landscape for interrogation and interpretation.[8] By illuminating previously inaccessible forms of genetic variation, including somatic-mosaic variants, recurrent complex rearrangements, and methylation-associated repeat variation, and linking them to gene regulatory function, this study provides a more complete framework for understanding the molecular basis of ASD. As the field moves forward, incorporating these approaches at scale will be essential to build a more complete and mechanistically grounded model of neurodevelopmental disease.

ACKNOWLEDGMENTS

M.H.C. is grateful to members of the Chahrour laboratory for helpful discussions, to Ashlesha Gogate for the figure, and for funding support from NHGRI (UM1HG011996), NINDS (R01NS138271), the Walter and Lillian Cantor Foundation, the SETBP1 Society, the Chan Zuckerberg Initiative, the Peter O'Donnell Jr. Brain Institute, and the University of Texas Southwestern Medical Center.

DECLARATION OF INTERESTS

The author declares no competing interests.

REFERENCES

1. Fu, J.M., Satterstrom, F.K., Peng, M., Brand, H., Collins, R.L., Dong, S., Wamsley, B., Klei, L., Wang, L., Hao, S.P., et al. (2022). Rare

coding variation provides insight into the genetic architecture and phenotypic context of autism. Nat. Genet. *54*, 1320–1331. https://doi.org/10.1038/s41588-022-01104-0.

2. Zhou, X., Feliciano, P., Shu, C., Wang, T., Astrovskaya, I., Hall, J.B., Obiajulu, J.U., Wright, J.R., Murali, S.C., Xu, S.X., et al. (2022). Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. Nat. Genet. *54*, 1305–1319. https://doi.org/10.1038/s41588-022-01148-2.
3. Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. *51*, 431–444. https://doi.org/10.1038/s41588-019-0344-8.
4. Cirnigliaro, M., Chang, T.S., Arteaga, S.A., Pérez-Cano, L., Ruzzo, E.K., Gordon, A., Bicks, L.K., Jung, J.Y., Lowe, J.K., Wall, D.P., and Geschwind, D.H. (2023). The contributions of rare inherited and polygenic risk to ASD in multiplex families. Proc. Natl. Acad. Sci. USA *120*, e2215632120. https://doi.org/10.1073/pnas.2215632120.
5. Logsdon, G.A., Vollger, M.R., and Eichler, E.E. (2020). Long-read human genome sequencing and its applications. Nat. Rev. Genet. *21*, 597–614. https://doi.org/10.1038/s41576-020-0236-x.
6. La Spada, A.R., and Taylor, J.P. (2010). Repeat expansion disease: progress and puzzles in disease pathogenesis. Nat. Rev. Genet. *11*, 247–258. https://doi.org/10.1038/nrg2748.
7. Sreenivasan, V.K.A., Yumiceba, V., and Spielmann, M. (2025). Structural variants in the 3D genome as drivers of disease. Nat. Rev. Genet. *26*, 742–760. https://doi.org/10.1038/s41576-025-00862-x.
8. Mortazavi, M., Guevara, J., Diaz, J., Tran, S., Ziaei Jam, H., Reeves, C., Batalov, S., Jepsen, K., Bainbridge, M., Besterman, A.D., et al. (2026). Long-read genome sequencing improves detection and functional interpretation of structural and repeat variants in autism. Cell Genom. *6*, 101186. https://doi.org/10.1016/j.xgen.2026.101186.
9. Brandler, W.M., Antaki, D., Gujral, M., Kleiber, M.L., Whitney, J., Maile, M.S., Hong, O., Chapman, T.R., Tan, S., Tandon, P., et al. (2018). Paternally inherited cis-regulatory structural variants are associated with autism. Science *360*, 327–331. https://doi.org/10.1126/science.aan2261.
10. Schloissnig, S., Pani, S., Ebler, J., Hain, C., Tsapalou, V., Söylev, A., Hüther, P., Ashraf, H., Prodanov, T., Asparuhova, M., et al. (2025). Structural variation in 1,019 diverse humans based on long-read sequencing. Nature *644*, 442–452. https://doi.org/10.1038/s41586-025-09290-7.