Perspective | Open Access

Journal of Translational Genetics and Genomics

Luo *et al. J Transl Genet Genom.* 2026;10:442-7 **DOI:**10.20517/jtgg.2026.62

Deciphering the tandem repeat code in autism: landscape, parent-of-origin effects, and mechanistic diversity

Tengfei Luo[1,2,4,5], Lu Xia[1,3], Jinchen Li[1,2,3,7,8], Kun Xia[1,3,6,7,8]

Citation: Luo T, Xia L, Li J, Xia K. Deciphering the tandem repeat code in autism: landscape, parent-of-origin effects, and mechanistic diversity. *J Transl Genet Genom.* 2026;10:442-7. https://dx.doi.org/10.20517/jtgg.2026.62

Received: 22 May 2026 **First Decision:** 6 Jul 2026 **Revised:** 13 Aug 2026 **Accepted:** 14 Aug 2026 **Published:** 21 Aug 2026

Academic Editor: Harry Steinbusch **Copy Editor:** Ping Zhang **Production Editor:** Ping Zhang

INTRODUCTION

Autism spectrum disorder (ASD) is a highly heritable neurodevelopmental condition[1], yet a substantial proportion of its genetic architecture remains unresolved. Short-read whole-genome sequencing (SR-WGS) has successfully identified numerous *de novo* coding protein-truncating and large copy number variants, which account for only a fraction of the overall ASD risk[2]. This missing heritability reflects a complex genomic architecture involving rare non-coding variants, cryptic structural rearrangements, polygenic background, and repetitive sequences. Among these potential contributors, tandem repeat (TR) variants represent a highly variable genomic component that remained poorly understood during the SR-WGS era[3]. Advances in long-read whole-genome sequencing (LR-WGS) now allow us to interrogate these complex regions, offering a critical opportunity to investigate how TR variations[4] and their associated parent-of-origin transmission biases elucidate the diverse downstream molecular mechanisms that shape the broader inherited architecture of ASD [Figure 1].

[1]The Key Laboratory of Pediatric Rare Diseases of the Ministry of Education, Central South University, Changsha & University of South China, Hengyang 421001, Hunan, China.
[2]Bioinformatics Center, National Clinical Research Center for Geriatric Diseases, Department of Geriatrics, Xiangya Hospital & Center for Medical Genetics and Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha 410008, Hunan, China.
[3]Center for Medical Genetics and Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha 410078, Hunan, China.
[4]The Central Laboratory of Birth Defects Prevention and Control, The Affiliated Women and Children's Hospital of Ningbo University, Ningbo 315012, Zhejiang, China.
[5]Ningbo Key Laboratory of Genomic Medicine and Birth Defects Prevention, The Affiliated Women and Children's Hospital of Ningbo University, Ningbo 315012, Zhejiang, China.
[6]College of Basic Medical Sciences, Hengyang Medical School, University of South China, Hengyang 421001, Hunan, China.
[7]NHC Key Laboratory of Birth Defect for Research and Prevention, Changsha 410078, Hunan, China.
[8]Center for Computational Biology and Bioinformatics, Furong Laboratory, Central South University, Changsha 410078, Hunan, China.

Correspondence to: Prof. Kun Xia, The Key Laboratory of Pediatric Rare Diseases of the Ministry of Education, Central South University, Changsha & University of South China, Hengyang 421001, Hunan, China. E-mail: xiakun@sklmg.edu.cn

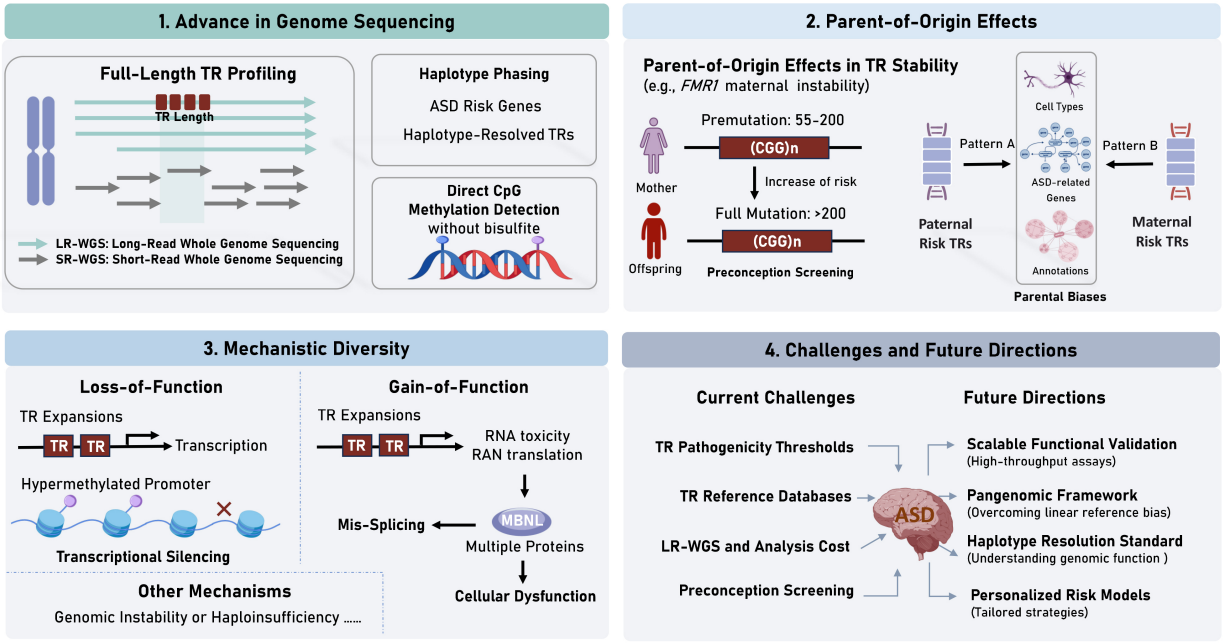

Figure 1. Overview of the tandem repeat landscape in ASD. (1) Sequencing advancements accelerate TR discovery in ASD; (2) Parent-of-origin effects clarify parental bias in TR transmission; (3) Mechanistic diversity shapes a multi-layered TR regulatory spectrum; (4) Current challenges and future directions in TR-ASD research. ASD: Autism spectrum disorder; TR: tandem repeat; SR-WGS: short-read whole-genome sequencing; LR-WGS: long-read whole-genome sequencing.

ADVANCES IN GENOME SEQUENCING: MAPPING THE TR-ASD LANDSCAPE

In the etiology of ASD, it is crucial to distinguish between highly penetrant Mendelian repeat-expansion disorders and the broader, genome-wide TR burden. While Mendelian expansions typically drive deterministic monogenic syndromes, the broader TR risk architecture encompasses diverse variants, ranging from rare large expansions to subtle non-coding variations, that collectively shape baseline developmental susceptibility. Driven by advances in sequencing technologies and bioinformatic tools, empirical evidence for this genome-wide impact has rapidly accumulated[5,6]. For instance, rare TR expansions detected by SR-WGS are estimated to account for approximately 2.6% of ASD risk, a proportion that likely remains underestimated[5]. This is because SR-WGS is inherently biased toward shorter, simpler repeats, making the accurate resolution of larger, structurally complex, or heavily expanded TR alleles a formidable challenge. Consequently, current clinical consensus guidelines emphasize that while SR-WGS serves as a powerful, cost-effective tool for high-throughput screening, bioinformatic estimates of repeat size should be confirmed by locus-specific molecular assays before being used as the basis for clinical diagnosis[7]. To address these constraints, LR-WGS technologies have emerged, generating reads spanning kilobases to megabases[8]. This technology can resolve complex structural variants, unambiguously determine parental origin, and capture previously inaccessible large repeat tracts[9,10]. Nevertheless, the widespread integration of LR-WGS into routine clinical diagnostics remains hindered by high sequencing and computational costs, which limit cohort scalability[11]. Furthermore, substantial technical hurdles persist, including the detection of low-frequency somatic mosaicism, the characterization of sequence interruptions within repeat tracts, and the difficulties that alignment algorithms face in precisely defining the boundaries of variable-number tandem repeat variants[12-14]. Ongoing innovations in both sequencing and computational frameworks are poised to overcome these barriers, thereby transforming our understanding of TR variations and enabling a new era of precise genetic counseling and targeted intervention.

PARENT-OF-ORIGIN EFFECTS: DECIPHERING THE PARENTAL BIAS IN TR TRANSMISSION

Parent-of-origin effects (POEs) refer to the phenomenon in which the phenotypic impact of a genetic variant is dictated by its parental mode of inheritance[15]. Although classically associated with genomic imprinting, POEs are recognized as an important aspect of mutational behavior and clinical penetrance in TR disorders. Fragile X syndrome (FXS), one of the most common inherited monogenic causes of ASD, results from a full mutation (> 200 CGG repeats) in the *FMR1* gene[16]. Maternal carriage of a premutation allele substantially elevates the risk of full mutation expansion in the offspring, whereas paternal transmission of a similarly sized allele rarely drives such progression[17]. Moreover, contraction of the premutation repeat displays a paternal bias. The full mutation expansion triggers aberrant methylation of the *FMR1* promoter and subsequent transcriptional silencing, ultimately causing a deficiency in fragile X messenger ribonucleoprotein essential for synaptic plasticity[18]. In contrast to *FMR1*, other disease-associated loci exhibit divergent dynamics, with paternal transmission driving greater instability. For example, expansion of the *HTT* CAG repeat occurs almost exclusively through the male germline, and a similar paternal bias has been observed within specific allelic ranges of the *DMPK* CTG repeat[19]. Intriguingly, recent genome-wide findings demonstrate that abnormal ASD-risk TR variants across the entire genome also display significant, though often heterogeneous, parental biases[20]. These biased yet divergent parental dynamics challenge the traditional assumption of parental equivalence in gene-disease associations. Therefore, deciphering POEs in ASD-associated TRs requires a conceptual shift: we should not merely evaluate the presence of these repeat variants but fundamentally characterize their parental lineage of transmission. For a highly heterogeneous condition such as ASD, this parental-stratification lens holds the potential to unveil critical signals that have historically been obscured by bulk, parent-agnostic genetic analyses.

MECHANISTIC DIVERSITY: ENCOMPASSING A MULTI-LAYERED REGULATORY SPECTRUM

Understanding how TR variants drive the complex neurodevelopmental phenotypes characteristic of ASD represents a major challenge in the field. Emerging evidence suggests that many TR expansions confer ASD risk through a loss-of-function mechanism driven by localized epigenetic collapse[21]. In individuals with FXS, expansions of the CGG repeat in the *FMR1* promoter trigger hypermethylation and subsequent chromatin compaction, leading to complete transcriptional silencing of the locus[22]. Similar epigenetic silencing has been documented in other loci associated with neurodevelopmental delay and ASD, including *AFF2*[23], *XYLT1*[24], *FRA10AC1*[25], *CBL*[26], and *DIP2B*[27]. Interestingly, a recent study integrating 18,236 SR-WGS and 950 LR-WGS samples revealed that massive CGG expansions and promoter hypermethylation at *DIP2B* are frequently observed in phenotypically normal individuals, showing no robust clinical association with neurological deficits[28]. These findings challenge the deterministic assumption that repeat expansion invariably leads to pathology, highlighting the critical influence of genetic modifiers or epigenetic thresholds on clinical penetrance. While these epigenetic collapse models primarily result in a loss of normal gene expression, other pathogenic TR variants operate via toxic gain-of-function routes[22]. For instance, the intronic GCC expansion in *AFF3* appears to drive pathogenesis through mechanisms such as RNA toxicity or repeat-associated non-AUG (RAN) translation[29]. The concept of dominant-negative interference is further exemplified by myotonic dystrophy type 1 (DM1), a condition frequently comorbid with ASD. In DM1, transcribed CUG repeats form stable secondary structures that sequester conserved RNA-binding proteins essential for alternative splicing in the brain, precipitating widespread mis-splicing across the transcriptome[30]. Although primarily a neuromuscular disorder, DM1 serves as a paradigmatic model for how repeat-induced RNA toxicity could similarly disrupt neurodevelopmental pathways in ASD. More broadly, these observations suggest that the molecular etiology of pathogenic TR variants extends well beyond simple promoter silencing, RNA toxicity, or RAN translation, encompassing a considerably broader spectrum of regulatory disruption. These diverse mechanisms represent frontier areas that warrant rigorous functional validation in future ASD cohort studies to elucidate the precise genotype-phenotype correlations.

CHALLENGES AND FUTURE DIRECTIONS

Translating the expanding landscape of TR variants into clinically actionable insights presents a formidable bottleneck in ASD genetics. The path from technological discovery to routine clinical implementation is hindered by several interconnected biological, technical, and ethical challenges. First, determining whether a specific TR expansion is benign or pathogenic remains exceptionally difficult. In contrast to single-nucleotide variants, TR variants lack robust, standardized pathogenicity thresholds. This uncertainty is compounded by incomplete or variable penetrance, as exemplified by the massive *DIP2B* expansions observed in phenotypically normal individuals[28]. Consequently, defining the boundary between normal polymorphic variation and disease-causing expansion remains an elusive endeavor. Beyond biological ambiguity, our current ability to interpret TR variants is severely restricted by incomplete reference databases. Traditional linear reference genomes fail to capture the complex, multi-allelic nature of repetitive loci and suffer from a profound ancestry bias. Transitioning to pangenome references, which utilize graph-based structures to represent diverse human genomic architectures, is critical to overcoming these gaps. By incorporating multi-ancestry data, pangenomics will enable the comprehensive mapping of population-specific TR alleles, thereby mitigating ancestry bias and establishing accurate baselines for diverse populations. Finally, if POEs and risk-associated TRs are integrated into preconception screening, genetic counselors will face unprecedented ethical challenges. Conveying risks characterized by highly variable phenotypes, unpredictable parental expansion dynamics, and poorly understood penetrance thresholds to prospective parents will require nuanced communication strategies. Such efforts are essential to prevent unnecessary anxiety and avoid guiding families toward ill-informed reproductive choices.

CONCLUSION

The inherited architecture proposed in this perspective constitutes a foundational framework rather than a theoretical abstraction, as TR variants, phasing, and POEs represent fundamental syntactic elements of the human genome. These components critically dictate how genetic information is transmitted, regulated, and expressed. While these layers of complexity remained largely illegible during the era of SR-WGS, LR-WGS has begun to decipher this nuanced framework within the context of ASD. Realizing the full clinical potential of this model hinges on prioritizing four actionable milestones: expanding large, family-based LR-WGS cohorts; establishing standardized benchmarks for TR calling; deploying high-throughput functional validation assays; and implementing ancestry-aware reference resources. Although deciphering this multidimensional landscape remains a formidable challenge, these advances lay the groundwork for integrating these genomic insights into more precise diagnostic frameworks and into informed clinical management of individuals with ASD.

DECLARATIONS

Acknowledgments

Some graphic elements, including the brain and cell illustrations, were created using BioGDP (https://BioGDP.com).

Authors' contributions

Drafted and revised the manuscript: Luo T
Conceptualized and revised the manuscript: Xia L, Li J, Xia K
All authors approved the final version for submission.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool Gemini (version 3.6 Flash, released 2026-05-20) was used solely for language editing. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy,

integrity, and final content of the manuscript.

Financial support and sponsorship

This work was supported by the Key Technology Breakthrough Program of Ningbo Sci-Tech Innovation YONGJIANG 2035 (No. 2025Z160) and the Zhejiang Clinovation Pride (No. CXRCTD202602008). The funders played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest

Xia K serves as an Editorial Board Member of the *Journal of Translational Genetics and Genomics*. He was not involved in any aspect of the editorial process for this manuscript, including reviewer selection, manuscript handling, or editorial decision-making. The other authors declare that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Hirota T, King BH. Autism spectrum disorder: a review. *JAMA.* 2023;329:157-68. DOI PubMed
2. Trost B, Thiruvahindrapuram B, Chan AJS, et al. Genomic architecture of autism from comprehensive whole-genome sequence annotation. *Cell.* 2022;185:4409-4427.e18. DOI
3. Tanudisastro HA, Deveson IW, Dashnow H, MacArthur DG. Sequencing and characterizing short tandem repeats in the human genome. *Nat Rev Genet.* 2024;25:460-75. DOI PubMed
4. Eisfeldt J, Ek M, Nordenskjöld M, Lindstrand A. Toward clinical long-read genome sequencing for rare diseases. *Nat Genet.* 2025;57:1334-43. DOI PubMed
5. Trost B, Engchuan W, Nguyen CM, et al. Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature.* 2020;586:80-6. DOI
6. Mitra I, Huang B, Mousavi N, et al. Patterns of de novo tandem repeat mutations and their role in autism. *Nature.* 2021;589:246-50. DOI PubMed PMC
7. Guha S, Rajan-Babu IS, Kesari A, et al. Detection of repeat expansion variants using next generation sequencing: a points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med.* 2026;28:102520. DOI PubMed PMC
8. Devaney JM, Chong JX, Lopes PC, et al. Sensitivity of HiFi long-read genome sequencing for difficult-to-detect pathogenic variants when applied to real-world clinical laboratory samples. *Am J Hum Genet.* 2026;113:1036-48. DOI PubMed PMC
9. Mortazavi M, Guevara J, Diaz J, et al. Long-read genome sequencing improves detection and functional interpretation of structural and repeat variants in autism. *Cell Genom.* 2026;6:101186. DOI PubMed PMC
10. de Bitter TJJ, van der Sanden B, Sagath L, et al. Clinical long-read genome sequencing for rare-disease diagnostics. *N Engl J Med.* 2026;395:405-8. DOI
11. Sabbagh Q, Gilissen C, Yntema HG, Vissers LELM, Hoischen A. Near-perfect genome sequencing in medical genetics. *Nat Genet.* 2026;58:1480-9. DOI PubMed
12. Lu TY, Smaruj PN, Fudenberg G, Mancuso N, Chaisson MJP. The motif composition of variable number tandem repeats impacts gene expression. *Genome Res.* 2023;33:511-24. DOI PubMed PMC
13. Dolzhenko E, English A, Dashnow H, et al. Characterization and visualization of tandem repeats at genome scale. *Nat Biotechnol.* 2024;42:1606-14. DOI PubMed PMC
14. Bai WY, Liu S, Duan Z, et al. Genome-wide associations of structural variants with human traits through imputation from long-read assemblies. *Nat Genet.* 2026;58:1258-67. DOI PubMed PMC
15. Ryan NM, Heron EA. Evidence for parent-of-origin effects in autism spectrum disorder: a narrative review. *J Appl Genet.* 2023;64:303-17. DOI PubMed PMC

16. Rajan-Babu IS, Dolzhenko E, Eberle MA, Friedman JM. Sequence composition changes in short tandem repeats: heterogeneity, detection, mechanisms and clinical implications. *Nat Rev Genet.* 2024;25:476-99. DOI PubMed
17. Nolin SL, Glicksman A, Tortora N, et al. Expansions and contractions of the FMR1 CGG repeat in 5,508 transmissions of normal, intermediate, and premutation alleles. *Am J Med Genet A.* 2019;179:1148-56. DOI
18. Penagarikano O, Mulle JG, Warren ST. The pathophysiology of fragile x syndrome. *Annu Rev Genomics Hum Genet.* 2007;8:109-29. DOI PubMed
19. Depienne C, Mandel JL. 30 years of repeat expansion disorders: what have we learned and what are the remaining challenges? *Am J Hum Genet.* 2021;108:764-85. DOI PubMed PMC
20. Li J, Luo T, Ren X, et al. Haplotype-resolved long-read sequencing reveals parent-of-origin effects of tandem-repeat variation in autism spectrum disorder. *Sci Bull.* 2026;71:3273-7. DOI
21. Annear DJ, Kooy RF. Unravelling the link between neurodevelopmental disorders and short tandem CGG-repeat expansions. *Emerg Top Life Sci.* 2023;7:265-75. DOI PubMed PMC
22. Zhao Y, Annear D, Allen EG, Kooy RF, Nelson DL, Jin P. Polymorphic CGG repeats in gene regulation and disease. *Am J Hum Genet.* 2026;113:1131-58. DOI PubMed PMC
23. Knight SJ, Flannery AV, Hirst MC, et al. Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell.* 1993;74:127-34. DOI
24. LaCroix AJ, Stabley D, Sahraoui R, et al. GGC repeat expansion and exon 1 methylation of XYLT1 is a common pathogenic variant in Baratela-Scott syndrome. *Am J Hum Genet.* 2019;104:35-44. DOI
25. Sarafidou T, Kahl C, Martinez-Garay I, et al. Folate-sensitive fragile site FRA10A is due to an expansion of a CGG repeat in a novel gene, FRA10AC1, encoding a nuclear protein. *Genomics.* 2004;84:69-81. DOI
26. Jones C, Penny L, Mattina T, et al. Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene CBL2. *Nature.* 1995;376:145-9. DOI
27. Winnepenninckx B, Debacker K, Ramsay J, et al. CGG-repeat expansion in the DIP2B gene is associated with the fragile site FRA12A on chromosome 12q13.1. *Am J Hum Genet.* 2007;80:221-31. DOI PubMed PMC
28. Ko CY, Schütz L, Braun T, et al. No evidence for an association between DIP2B repeat expansion and neurological disease. *Mov Disord.* 2026. DOI
29. Jadhav B, Garg P, van Vugt JJFA, et al. A phenome-wide association study of methylated GC-rich repeats identifies a GCC repeat expansion in AFF3 associated with intellectual disability. *Nat Genet.* 2024;56:2322-32. DOI PubMed PMC
30. Sznajder ŁJ, Khan M, Ciesiołka A, et al. Autism-related traits in myotonic dystrophy type 1 model mice are due to MBNL sequestration and RNA mis-splicing of autism-risk genes. *Nat Neurosci.* 2025;28:1199-212. DOI PubMed PMC