Review | Open Access

Journal of Translational Genetics and Genomics

Tang *et al. J Transl Genet Genom.* 2026;10:631-52 **DOI:**10.20517/jtgg.2026.58

Long-read sequencing: the resolution revolution in embryonic development and beyond

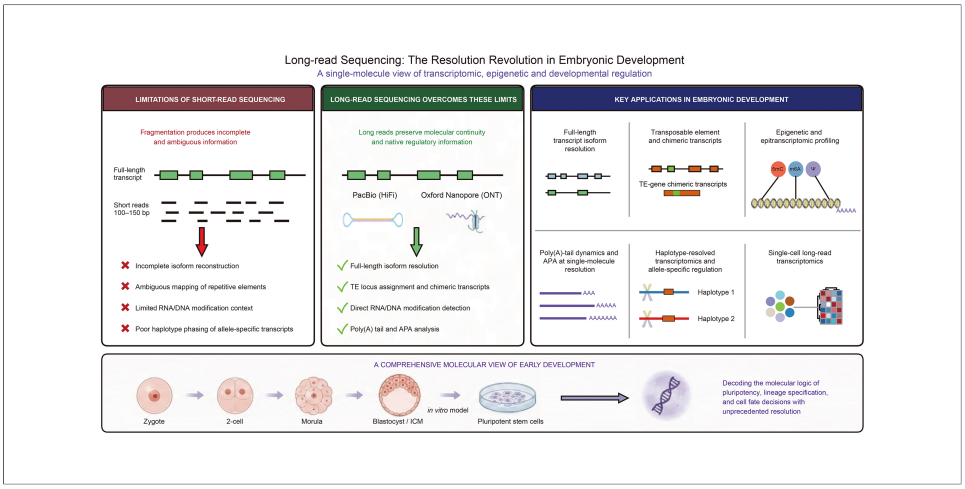
Huan Tang[1,#], Isaac A. Babarinde[2,#], Yingyu Mu[1], Qiaosen Shen[3], Yu Qiao[4], Zhen Xiao[4], Andrew P. Hutchins[4], Gang Ma[1]

Keywords:
Long-read sequencing, embryonic development, transposable elements, alternative splicing, single-cell sequencing

Citation: Tang H, Babarinde IA, Mu Y, Shen Q, Qiao Y, Xiao Z, Hutchins AP, Ma G. Long-read sequencing: the resolution revolution in embryonic development and beyond. *J Transl Genet Genom.* 2026;10:631-52. https://dx.doi.org/10.20517/jtgg.2026.58

Received: 12 May 2026 **First Decision:** 30 Sep 2026 **Revised:** 5 Nov 2026 **Accepted:** 12 Nov 2026 **Published:** 27 Nov 2026

Academic Editors: Junjiang Fu, Sheng-Ying Qin **Copy Editor:** Ping Zhang **Production Editor:** Ping Zhang



Abstract

Early embryonic development and stem cell pluripotency are controlled by the coordinated regulation of transcriptomic diversity and dynamic epigenetic states. Short-read sequencing (SRS) has provided foundational insights into developmental regulation, but the fragmentation of nucleic acids before sequencing limits the characterization of full-length transcript dynamics, epigenetic change, and transposable element (TE) activity, which are key features of early embryogenesis. Here we review recent advances in long-read sequencing (LRS) and evaluate their impact on the study of embryonic development and embryonic stem cells (ESCs). LRS enables the interrogation of full-length transcript isoforms, alternative splicing events, poly(A) tail dynamics, DNA and RNA modifications, TE-derived chimeric RNAs, and haplotype-resolved regulatory features. It also supports measurements that lie outside RNA biology, including the direct measurement of telomere

[1]Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, Guangdong, China.
[2]Laboratory of Inflammation and Vaccines, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, Guangdong, China.
[3]Fujian Metrology Institute, Fuzhou 350003, Fujian, China.
[4]Department of Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China.
[#]These authors contributed equally

Correspondence to: Prof. Andrew P. Hutchins, Department of Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China. E-mail: andrewh@sustech.edu.cn; Prof. Gang Ma, Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, Guangdong, China. E-mail: magang@smu.edu.cn

length, the detection of *de novo* retrotransposon insertions, and the profiling of chromatin accessibility on individual molecules. By preserving molecular continuity and regulatory information, LRS provides an integrated view of the transcriptional, epigenetic, and post-transcriptional mechanisms that coordinate pluripotency, lineage specification, and early cell fate decisions. We also discuss the technical and analytical limitations that still constrain the field and outline the directions in which it is moving, including single-cell, spatially resolved, and epitranscriptomic applications.

INTRODUCTION

The study of embryonic stem cells (ESCs) and early embryonic development is fundamental to understanding the molecular and regulatory principles that govern cell fate decisions, developmental programs, and regenerative medicine[1-5]. Unlike somatic cells, ESCs exist in a unique pluripotent state, characterized by a highly plastic and open chromatin landscape that permits broad transcriptional competence[6]. This biological complexity is not solely defined by the expression of core transcription factors such as OCT4, NANOG, and SOX2, but is driven by the coordinated regulation of thousands of transcript isoforms, non-coding RNAs, and the expression of transposable elements (TEs), many of which are embedded within or give rise to non-coding transcripts[7,8]. For instance, the transient expression of specific retrotransposons is recognized as a hallmark of early embryonic transitions, acting as alternative promoters and regulatory elements that shape the transcriptome during zygotic genome activation (ZGA)[9-11]. Previous studies demonstrated that retrotransposon activation is not merely a byproduct of transcriptional derepression but plays an active regulatory role in early development, influencing chromatin accessibility and transcriptional competence[12,13]. Mechanistically, studies have shown that LINE1 elements can interact with nuclear factors such as nucleolin to modulate chromatin organization and gene expression, thereby contributing to the establishment of totipotency and the progression of early embryogenesis[12]. Furthermore, the epigenetic remodeling and dynamic poly(A) tail regulation that occur during the maternal-to-zygotic transition create a complex regulatory layer[14]. Deciphering these multi-dimensional networks requires high-resolution genomic, epigenetic, and transcriptomic analyses to understand gene regulatory networks, repetitive genomic architecture, and the molecular switches that define cellular identity and developmental fidelity[15,16].

Sequencing technologies have revolutionized the study of genomics, largely driven by advances in short-read sequencing (SRS). Early SRS-based studies used reads shorter than 100 base pairs[17]. Subsequent improvements increased the read length modestly, to a common length of 150 bp. Although SRS has driven major advances, its limited read length imposes constraints on its ability to resolve complex biological features, particularly transcript splicing dynamics[18]. The median human protein-coding transcript length in version 115 of the Ensembl database[19] is 2,276 base pairs, with the longest transcripts exceeding 30,000 base pairs [Figure 1A and B]. Because full-length transcripts are typically longer than common SRS reads, transcripts are fragmented before sequencing. To complicate matters further, 219,583 of 221,639 human protein-coding transcripts (~99%) are multi-exonic [Figure 1C], and 19,911 of 22,847 annotated protein-coding genes (87%) have multiple transcript isoforms [Figure 1D]. It has long been clear that SRS is poor at assembling and quantifying transcripts accurately[20]. Assignment of fragmented reads to the transcript from which they originated requires extensive computational investment that does not guarantee accuracy. This limitation is exacerbated for TE-containing transcripts, in which fragmented reads can map to hundreds of genomic loci, making it difficult to identify the true genomic origin. Assessments of transcript assembly based purely on SRS, benchmarked against gold-standard transcripts, have routinely demonstrated that SRS is unreliable for transcript assembly[18,21,22].

A

Ensembl gene ID	Longest transcript	Gene name	Length (bp)	Exon count
ENSG00000155657	ENST00000412264	TTN	109488	365
ENSG00000291738	ENST00000711671	MUC16	46191	113
ENSG00000175387	ENST00000262160	SMAD2	34626	11
ENSG00000185070	ENST00000330753	FLRT2	33681	2
ENSG00000273079	ENST00000630791	GRIN2B	30670	15

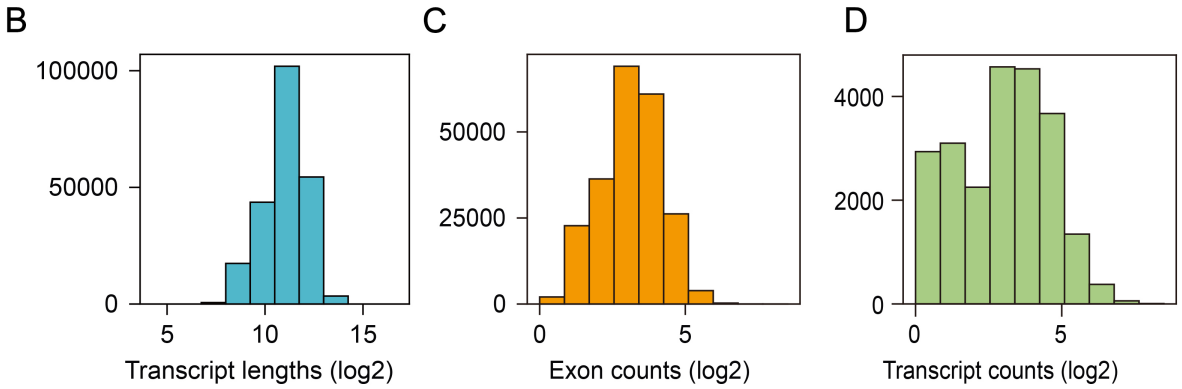


Figure 1. Characterization of human protein-coding gene transcripts. (A) The top five genes with the longest transcripts. Histograms show the distributions of transcript length (B), exon count per transcript (C), and transcript count per gene (D). Transcript information for human protein-coding genes was retrieved from version 115 of the Ensembl database. For visualization, all counts and lengths were log2-transformed. Panels A and B establish the scale mismatch between transcript length and short-read length: the median human protein-coding transcript is 2,276 bp, whereas the longest exceeds 30,000 bp. Panels C and D show that almost all protein-coding transcripts contain multiple exons and that the majority of protein-coding genes produce more than one isoform, so any approach that fragments RNA before sequencing must reconstruct both exon connectivity and isoform identity from partial information.

Recent advances in long-read sequencing (LRS) technologies[23-25] have begun to overcome these limitations. Two single-molecule platforms dominate current applications in developmental biology. Pacific Biosciences (PacBio) sequencing detects fluorescently labeled nucleotides during real-time DNA synthesis inside zero-mode waveguides, whereas Oxford Nanopore Technologies (ONT) detects base-specific disruptions in ionic current as a nucleic acid strand passes through a protein nanopore[26,27]. However, these are not the sole means of obtaining long-range sequence data. Linked-read library approaches preserve long-range information by barcoding high-molecular-weight DNA prior to short-read sequencing; optical genome mapping produces ultra-long-range physical maps of genome structures on the megabase scale; and newer single-molecule instruments are continually broadening the spectrum of available read lengths and cost options[26]. Our focus here is on PacBio and ONT, as they constitute the vast majority of published developmental applications, though we do highlight instances where the broader ecosystem of long-range technologies is pertinent.

Although PacBio and ONT both enable long-read, single-molecule analysis, their strengths in embryonic studies are complementary [Table 1]. PacBio sequencing, particularly Iso-Seq with high-accuracy consensus reads, is especially well suited to isoform-resolved transcriptome annotation, transcript-level abundance estimation, and the characterization of alternative splicing events[27]. In contrast, ONT-based approaches, especially direct RNA sequencing and end-capture strategies such as Nano3P-seq[40], preserve native RNA or complementary DNA (cDNA) continuity and can be used to interrogate poly(A) tail length, alternative polyadenylation, and native RNA modifications[34]. Whereas ONT supports direct sequencing of native RNA molecules, PacBio requires reverse transcription to cDNA and therefore cannot be used for direct RNA modification detection [Table 1]. This distinction should be kept in mind when interpreting studies of RNA modifications in early development.

Table 1. Comparison of PacBio and Oxford Nanopore long-read sequencing technologies and their applications in early embryonic development. Entries summarize information from previously published studies and benchmarking analyses, which are cited in the corresponding cells

Feature	PacBio (HiFi/SMRT)	Oxford Nanopore (ONT)
Sequencing mechanism	Single-molecule real-time (SMRT) sequencing in zero-mode waveguides (ZMWs); fluorescently labeled nucleotides are detected in real time during DNA synthesis, and circular consensus sequencing (CCS) generates HiFi reads by multiple passes of the same molecule[23,24]	Nanopore-based electrical sensing: a protein nanopore embedded in a membrane detects base-specific ionic current disruptions as a single-stranded DNA or RNA molecule passes through the pore. No DNA synthesis or optics are required[24]
Read length	HiFi reads typically 10-25 kb; ultra-long reads (> 50 kb) possible in non-HiFi modes	Routinely > 10 kb, and reads exceeding 2 Mb have been reported. Best suited to spanning full-length repetitive elements and structural variants
Raw read accuracy and error profile	HiFi reads: > 99.9% (Q30+). Errors are random and mostly base substitutions, which are readily corrected by consensus[23]	Raw accuracy 90%-98% depending on pore and basecaller, with errors predominantly insertions and deletions (indels); consensus accuracy can reach Q30+[28,29]
Direct RNA sequencing	Not supported. RNA must be reverse-transcribed to cDNA, and amplification may introduce bias and lose native modification information	Supported by direct RNA sequencing (dRNA-seq)[25,30]. Native RNA molecules are sequenced without reverse transcription or amplification, preserving natural RNA modifications and poly(A) tail lengths
Epigenetic and epitranscriptomic detection	Can detect DNA modifications (for example, 6mA and 5mC) through interpulse duration (IPD) kinetics during DNA synthesis, but sensitivity and specificity are lower than for ONT. Because direct RNA sequencing is not available, RNA modifications cannot be measured directly	Direct detection of DNA modifications (5mC, 6mA) and RNA modifications (m^6A, pseudouridine, inosine) from raw current signals, without chemical conversion. Suited to single-molecule methylome and epitranscriptome mapping[31-33]
Isoform detection and quantification	Superior for precise isoform annotation because of high per-base accuracy; resolves complex alternative splicing and similar isoforms with high confidence, and quantification is accurate given sufficient depth[23]	Ultra-long reads provide full 5′-3′ coverage, including distal alternative polyadenylation (APA) sites; the higher error rate may require reference-based polishing or hybrid short-read correction for accurate quantification, and benchmarking studies highlight these trade-offs[34]
Throughput and cost	High throughput per run; moderate to high cost per run, best suited to large-scale isoform cataloging	Flexible throughput (MinION, GridION, PromethION) and low entry cost with MinION, although reagent cost scales with yield
Typical applications in early embryogenesis	Building isoform catalogs of preimplantation embryos[35]; resolving TE-containing chimeric transcripts with base-level precision; allele-specific expression and splicing analysis[36,37]	Measuring poly(A) tail dynamics during the maternal-to-zygotic transition; profiling RNA modifications that regulate totipotency; ultra-long reads for locus-specific TE resolution; direct DNA methylation mapping at repetitive regions without bisulfite conversion[31,38,39]
Complementary use	The two platforms are often used together: PacBio for high-definition isoform discovery and ONT for native modification analysis and ultra-long structural resolution	

HiFi: High fidelity; cDNA: complementary DNA; 6mA: N^6-methyldeoxyadenosine in DNA; 5mC: DNA 5-methylcytosine; m^6A: N^6-methyladenosine in RNA; TE: transposable element; PacBio: Pacific Biosciences.

In this review, we highlight how LRS technologies are reshaping our understanding of early embryogenesis. We focus on the following themes: (i) the accurate reconstruction of full-length transcript isoforms and alternative splicing events that govern pluripotency and lineage specification; (ii) the direct characterization of DNA and RNA modifications that shape developmental regulatory programs; (iii) the identification of TE-derived and TE-containing transcripts and other repetitive genomic features that are difficult to resolve using conventional approaches; and (iv) the reconstruction of haplotype-resolved transcriptomes, including allele-specific isoforms, alternative polyadenylation and poly(A) tail dynamics. We then discuss emerging single-cell, spatial, and epitranscriptomic applications, and the technical challenges that remain. Together, these themes illustrate how LRS enables a comprehensive view of the molecular landscapes that drive early mammalian development.

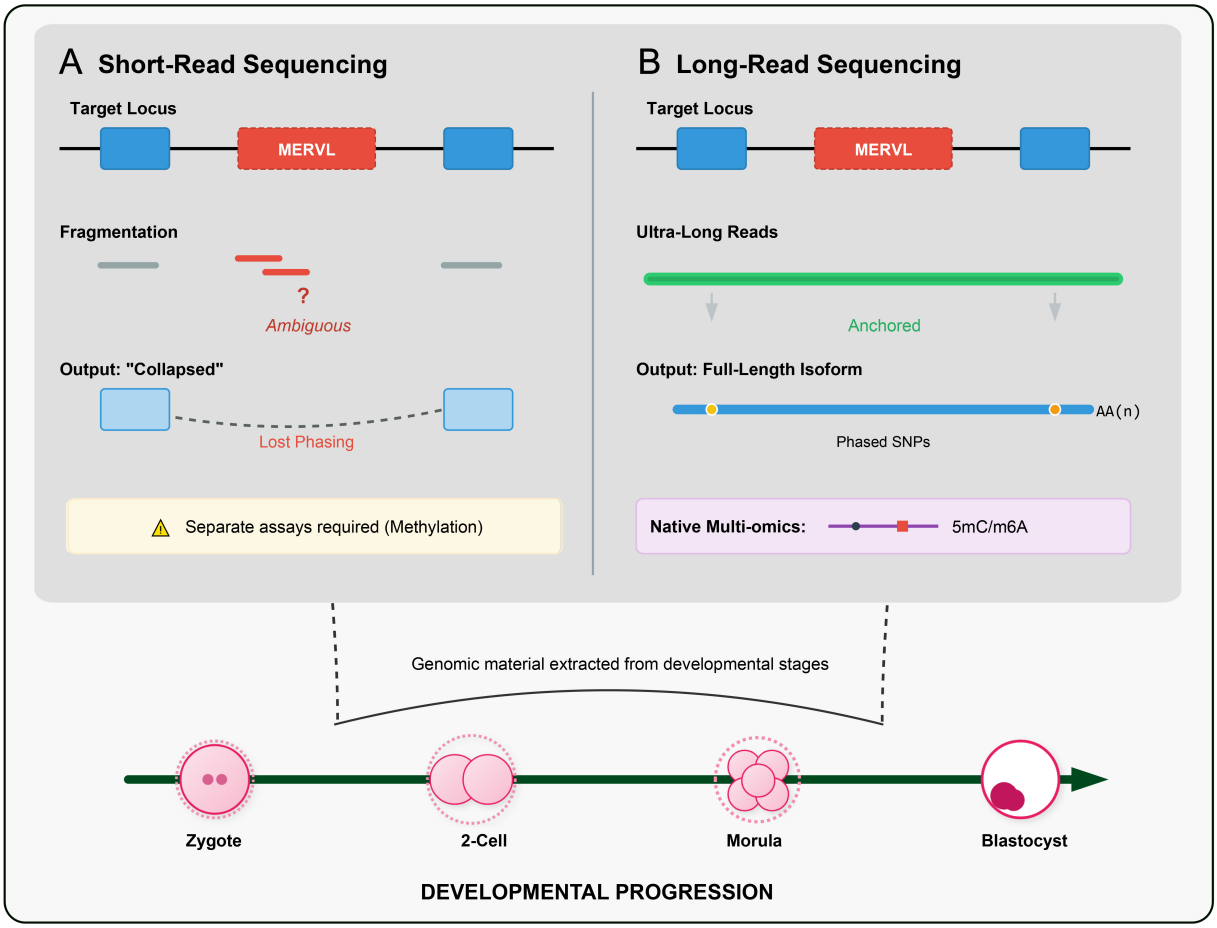


Figure 2. The resolution revolution: from fragmented snapshots to native holistic views. (A) Short-read sequencing requires fragmentation of the target molecule before sequencing. Reads derived from a repetitive target locus such as MERVL cannot be assigned unambiguously to their genomic origin, and the resulting output is a collapsed representation in which haplotype phasing is lost; DNA methylation, RNA modification, and poly(A) tail information must be obtained from separate assays; (B) Long-read sequencing sequences the whole molecule without fragmentation. Ultra-long reads span repetitive elements together with their unique flanking sequences, so that the transcript can be anchored to a single locus, single-nucleotide polymorphisms within the same molecule can be phased, and native modifications such as 5mC and m^{6}A are read from the same molecule that provides the sequence. Both panels are set against the same developmental continuum, from zygote to blastocyst, to emphasize that LRS preserves the molecular continuity of material extracted from these stages rather than sampling it in fragments. 5mC: DNA 5-methylcytosine; m^{6}A: N^6-methyladenosine in RNA; LRS: long-read sequencing; MERVL: murine endogenous retrovirus-L.

Long-read sequencing has also changed clinical and cancer genomics, immunology, and microbiology. Those applications lie outside the scope of this review, which focuses on early embryonic development.

LRS IN EARLY EMBRYONIC DEVELOPMENT

LRS capability is particularly transformative for developmental biology, where previously unresolved genomic "dark matter", including repetitive elements, non-coding transcripts and complex isoforms, often contributes to cell fate specification[18]. As illustrated in Figure 2, LRS allows researchers to bypass assembly-dependent biases, revealing the transcriptomic architecture of totipotent and pluripotent states[35,36]. In embryonic systems, this advantage is particularly important for resolving full-length isoforms, TE-derived chimeric transcripts, allele-specific transcripts, and transcript-end features that are difficult to reconstruct accurately from fragmented short reads[14,35,36,40,41].

Resolution of full-length transcripts and alternative splicing

The maintenance of pluripotency and cell differentiation depends on intricate gene regulatory networks, in which alternative splicing expands proteomic diversity without increasing genome size[42,43]. Alternative splicing can generate distinct protein variants from the same genomic locus, and isoform switching accompanies differentiation[44]. Crucially, isoform-dependent regulation can act as a switch for cell fate decisions. A well-characterized example is provided by the transcription factor FOXP1, for which an ESC-specific isoform (FOXP1-ES) confers distinct DNA-binding specificity that promotes the expression of pluripotency genes such as OCT4 and NANOG[45]. Upon differentiation, cells switch to the somatic isoform, which represses these same pluripotency factors and drives lineage commitment[45,46]. Another example is the *de novo* DNA methyltransferase 3 beta (DNMT3B), a protein with demonstrated roles in mouse development[47,48] and human embryonic stem cells[49]. The DNMT3B gene produces as many as 40 alternatively spliced transcripts, the majority of which are aberrant transcripts[50,51]. Several of these isoforms, including DNMT3B3, lack a functional catalytic domain, yet act as accessory proteins that stimulate gene-body methylation in mouse embryonic stem cells[52] and the human HCT116 colon cancer cell line[53]. Isoform choice can therefore modulate DNA methylation independently of catalytic activity, which makes the accurate assignment of DNMT3B transcripts a requirement for interpreting methylation dynamics rather than a descriptive exercise.

Early estimates based on short-read data indicated that more than 95% of multi-exon human genes are alternatively spliced[54]. Long-read sequencing has confirmed this general conclusion while showing that isoform diversity is considerably greater than these estimates suggested: isoform-resolved analyses of human tissues and cell lines have uncovered large numbers of previously unannotated transcripts and have revised the number of isoforms per gene upward[18,55,56]. Capturing this heterogeneity is often beyond the reach of SRS, particularly for low-abundance transcripts and complex splicing events that lack sufficient junction-spanning reads. Consequently, transcript assembly can generate fragmented or spurious transcript models, leading to inaccurate isoform reconstruction and quantification[21,22]. In contrast, by sequencing intact RNA molecules or full-length cDNA molecules, LRS reduces the dependence on transcript reconstruction and enables direct observation of full-length transcripts. This transition from probabilistic inference to direct molecular observation has led to numerous biological discoveries and substantially improved our understanding of transcript diversity during early development[23,35,57].

In mammalian systems, LRS has enabled high-resolution annotation of the preimplantation embryo transcriptome and has substantially expanded the known repertoire of developmental transcripts[41]. In mouse preimplantation embryos, PacBio Iso-Seq identified 2,280 previously unannotated transcripts and 6,289 novel splicing isoforms[41]. Extending this approach to human development, an isoform-resolved transcriptome of 73 human preimplantation embryos across six stages identified 110,212 previously unannotated isoforms from known genes and 17,964 isoforms from previously unannotated genes, many of which showed stage-specific expression patterns that were obscured in prior SRS datasets[35].

Nanopore-based LRS has enabled simultaneous measurement of transcript identity and poly(A) tail length during the mouse oocyte-to-embryo transition. Transcriptome-wide nanopore long-read RNA sequencing revealed an extended wave of global messenger RNA (mRNA) deadenylation during fertilization, establishing a switch in translational regulation across the oocyte-to-embryo transition and demonstrating that poly(A) tail length is coupled to translation efficiency[14]. Complementary work has shown that mature mRNA processing, which removes 3' end sequences, directs translational activation and is required for embryonic development[58]. Together, these studies demonstrate the power of nanopore LRS to connect transcript identity, poly(A) tail regulation, maternal RNA clearance and translational control during the earliest stages of embryogenesis[14,40,58].

Direct interrogation of epigenetic and post-transcriptional regulation

Epigenetic regulation, including both DNA and RNA modifications, plays a central role in maintaining pluripotency and directing embryonic development[3,13,59]. Traditional approaches for profiling these modifications, such as bisulfite sequencing, rely on chemical conversion steps that fragment nucleic acids and disrupt long-range molecular information[59]. In contrast, LRS enables direct detection of DNA methylation without chemical treatment while preserving long DNA molecules[59]. This capability allows single-molecule, locus-specific and allele-resolved characterization of methylation states across repetitive regions, differentially methylated regions and imprinted loci that are often difficult to resolve using short-read technologies[60]. As a result, LRS provides a more comprehensive view of epigenetic regulation during development and disease.

The length of the poly(A) tail is a second layer of post-transcriptional regulation that LRS has made directly measurable on individual transcripts. The initial evidence for an embryonic switch in translational control came from poly(A)-tail profiling by sequencing (PAL-seq), a short-read method that measures the poly(A) tail length of individual RNA molecules. In Xenopus and zebrafish embryos, PAL-seq showed that maternal transcripts are progressively deadenylated and that tail length correlates positively with translation efficiency[61]; related work in Drosophila showed that deadenylation-driven tail shortening broadly reshapes translation[62]. Because PAL-seq is a short-read method, it cannot sequence the poly(A) tail together with the full-length transcript. This is particularly an issue in genes with multiple isoforms that share the same transcript termination site, in which case it is difficult to resolve the poly(A) tails of different isoforms using PAL-seq. In addition, PAL-seq cannot simultaneously measure tail lengths and other epigenetic modifications of the same molecule. This process is regulated by factors controlling mRNA deadenylation[63]. For instance, members of the maternal mRNA translation regulator (MARTRE) family negatively regulate the CCR4-NOT deadenylase complex, thereby protecting maternal transcripts from excessive deadenylation and maintaining poly(A) tail length; loss of MARTRE function results in shortened poly(A) tails and reduced translation efficiency in mouse zygotes, without accelerating maternal mRNA decay[64]. LRS-based measurements have since resolved this process at the level of individual molecules. Nanopore sequencing of mouse oocytes and preimplantation embryos uncovered isoform-resolved poly(A) tail dynamics associated with 3′ untranslated region (3′ UTR) length, splice isoform usage and polyadenylation-site selection, and revealed substantial heterogeneity in tail length among molecules of the same transcript[14].

Transcripts associated with ZGA, including those driven by endogenous retroelements such as murine endogenous retrovirus-L (MERVL), display distinct poly(A) tail length distributions and temporal dynamics, suggesting coordinated regulation between transcriptional activation and post-transcriptional RNA processing[10,14]. These findings link RNA tailing dynamics to the selective translation of developmental regulators and to the establishment of totipotency during early embryogenesis.

DNA methylation during early development was first mapped with short-read, bisulfite-based methods, which established the overall dynamics of global erasure and re-establishment of methylation in human preimplantation embryos[65]. Because these methods depend on chemical conversion of fragmented DNA, they provide limited access to allele-resolved methylation at imprinted loci and to repetitive genomic regions. LRS-based methylome profiling, using either nanopore current signals or PacBio kinetics, now captures dynamic remodeling of DNA methylation across preimplantation development, including allele-specific methylation patterns at imprinted loci and repetitive genomic regions, without chemical conversion[31,38,60].

On the RNA side, the m^{6}A landscape of mouse oocytes and preimplantation embryos has been mapped using picogram-scale methylated RNA immunoprecipitation and sequencing (picoMeRIP-seq), a low-input antibody-based method that provides transcript-resolved modification profiles and revealed stage-specific

methylation of maternal and zygotic transcripts[39]. Antibody-based enrichment reports the relative modification status of transcript populations rather than of individual molecules. Methodological advances in signal-level interpretation have therefore focused on nanopore data, where deep learning approaches have improved the accuracy of transcriptome-wide m^{6}A detection[32,33,66]. Emerging studies that combine LRS with isoform-level m^{6}A characterization are beginning to reveal roles for RNA modifications in RNA surveillance and gene expression regulation[67].

Long-read sequencing extends to chromatin structure as well. Two complementary strategies have been developed. The first labels accessible DNA with a methyltransferase that methylates guanine followed by cytosine (5′-GpC-3′) (GpC) dinucleotides, after which both the introduced accessibility marks and endogenous cytosine followed by guanine (5′-CpG-3′) (CpG) methylation are read from the same long molecules by nanopore sequencing, so that nucleosome occupancy and methylation state are obtained simultaneously[68]. The second combines long reads with a single-cell assay for transposase-accessible chromatin (ATAC)-based readout: single-cell ATAC on Nanopore sequencing (scNanoATAC-seq) profiles chromatin accessibility and genetic variants together within an individual cell, which links accessible regions to the sequence variants that lie on the same molecule[69]. In early embryos, chromatin accessibility constitutes a pivotal characteristic marking the transition from a transcriptionally inert to an active genome. To date, this phenomenon has predominantly been characterized using short-read single-cell assays, which resolve accessibility but fail to elucidate the underlying sequence context[70,71]. The application of long-read accessibility assays to the limited cellular samples obtainable from preimplantation embryos remains a burgeoning field of research, representing a logical and promising extension of the methodologies discussed herein.

Single-molecule linkage of epigenetic state and transcript structure

A further advantage of LRS, beyond the recovery of longer sequences, is that it measures several regulatory features on the same physical molecule. Because a single read can carry a splice junction, a polyadenylation site, the poly(A) tail, and the modification status of individual nucleotides, relationships that were previously inferred by comparing averaged datasets can now be observed directly. This makes it possible to ask whether the epigenetic state of a locus is coupled in cis to the transcript structure that the locus produces[72,73].

Several observations illustrate this principle. Modification-aware long-read analyses of human cell lines have shown that m^{6}A marks are unevenly distributed across isoforms of the same gene and that modification clusters coincide with isoform-specific polyadenylation events, so that the choice of 3’ end and the modification status of the transcript are linked on the same molecule[72,73]. Allele-resolved long-read sequencing extends the same logic across parental genomes: m^{6}A modifications can be assigned to maternal or paternal alleles, revealing allele-specific modification patterns that population-level assays average away[37,74]. On the DNA side, long-read accessibility and methylation measurements resolve chromatin state along individual DNA molecules and can be related to the expression of the corresponding gene[68,75]. This is a locus-level comparison between chromatin state and expression rather than a measurement of both on the same molecule, and a direct link between the chromatin state of a given DNA molecule and the isoform it produces has not yet been established.

The developmental relevance of this cis-linkage is most apparent at the maternal-to-zygotic transition, where rapid clearance of maternal transcripts and activation of the zygotic genome require coordinated, multi-level control. During this window, transcripts that carry particular modification patterns are preferentially retained, translated, or degraded, and the corresponding poly(A) tail states change within hours[14,39]. Similarly, the choice of a distal rather than a proximal polyadenylation site changes 3’ UTR length, which in turn affects transcript stability and the availability of modification and microRNA sites; in single cells, m^{6}A is

distributed unevenly across isoforms of the same gene, so isoform identity and modification state need to be measured together[67,76]. Testing these relationships causally remains difficult, but the measurements themselves are now within reach, and they define a set of questions that short-read approaches cannot address at all.

LRS of repetitive sequences

Repetitive elements, including endogenous retroviral sequences (ERVs) and other TE families, undergo dynamic activation during early developmental transitions, particularly during ZGA[77]. Throughout preimplantation development, expression of repetitive sequences continues to fluctuate and is closely associated with cell fate transitions and lineage specification. In mouse embryos, the transition from the late 2-cell to the 4-cell stage is accompanied by a transcriptional burst that includes numerous novel transcript isoforms and activation of ERVs, such as the rodent-specific MERVL family[9,78]. These repetitive elements can function as species-specific enhancers and alternative promoters, thereby contributing to regulatory programs associated with totipotency and early embryonic cell fate transitions[10,77].

Chimeric transcripts that combine repetitive and unique sequences are a recurrent feature of early development, because TE-derived promoters and splice donors can be spliced onto exons of neighboring genes to produce composite transcripts. SRS can detect the overall upregulation of TE transcripts such as MERVL, but the repetitive nature of these sequences results in extensive multi-mapping across genomic loci, which limits locus-specific assignment, isoform reconstruction, and functional interpretation. Integrative pipelines that combine LRS with short-read sequencing have expanded the known landscape of TE-containing transcripts in human pluripotent stem cells (PSCs) and differentiation models, identifying lineage-specific TE transcripts and showing that the same TE family can be expressed from different transcripts in different lineages[79].

LRS enables assignment of transcripts to their genomic origins, improves transcript identification, isoform resolution and locus-specific quantification, and supports full-length reconstruction of TE-gene chimeric transcripts, including nested TE insertions that function as stage-specific promoters or cryptic splice donors[80-82]. Recent work has begun to apply this capability directly to embryos rather than to model systems, resolving TE-derived chimeric transcripts at ZGA in mammalian embryos[28]. Similarly, full-length transcriptome analyses of cellular reprogramming have uncovered an isoform-switching event involving a previously uncharacterized non-coding RNA that contributes to reprogramming efficiency[83]. Studies of human PSCs have also shown that repetitive sequence fragments are extensively incorporated into both coding and noncoding transcripts, where they act as functional modules that shape the cellular transcriptome and its regulatory networks[7].

Evidence from non-mammalian vertebrates shows that this is not a rodent-specific phenomenon. Zebrafish and Xenopus are separated from mammals by several hundred million years of evolution, so the recurrence of TE-driven transcriptomic plasticity in these species indicates that it is a deeply conserved strategy of early vertebrate embryogenesis rather than a peculiarity of mammalian embryos. In zebrafish embryos, PacBio full-length RNA sequencing across 21 developmental stages identified 2,113 previously unannotated genes and 33,018 novel isoforms, revealing extensive transcriptome complexity, dynamic alternative splicing, and locus-specific activation of TE families that were largely inaccessible to short-read analyses[84]. These findings indicate that early embryogenesis is characterized by extensive transcriptome remodeling at the isoform level rather than by changes in gene-level abundance alone. Likewise, transcriptomic analyses of Xenopus laevis embryos revealed dynamic TE expression and subgenome-specific TE regulation throughout embryogenesis, highlighting the contribution of repetitive elements to early developmental programs in a species whose genome contains two coexisting subgenomes[85].

Collectively, these studies demonstrate that the advantage of LRS lies in its ability to determine the architecture and genomic origins of TE-associated transcripts. By spanning repetitive elements together with unique flanking sequences within single molecules, LRS enables locus-specific quantification, full-length isoform reconstruction, and identification of TE-containing splice isoforms and chimeric transcripts. It therefore provides a framework for understanding how repetitive sequences are integrated into developmental gene regulatory networks and how they contribute to cell fate transitions during early embryogenesis.

Haplotype-resolved transcriptomics and transcript-end regulation

The complexity of the embryonic transcriptome arises not only from extensive isoform diversity and rapid developmental transitions but also from the coexistence of two genetically and epigenetically distinct parental genomes and the presence of structurally complex genomic regions[7,86]. Following fertilization, the maternal and paternal genomes undergo asynchronous epigenetic reprogramming while retaining parent-of-origin-dependent regulatory information at specific loci, giving rise to allele-specific patterns of gene expression and transcript regulation. A prominent example of such parent-of-origin-dependent regulation is genomic imprinting, in which either the maternal or the paternal allele is preferentially expressed. By preserving allele-specific DNA methylation and chromatin states through epigenetic reprogramming, imprinting regulates embryonic growth, placental development, lineage specification, and developmental competence, whereas its disruption can lead to abnormal embryogenesis and imprinting-related disorders. However, most studies have focused on gene-level parent-of-origin expression or imprinting status, leaving unresolved whether maternal and paternal alleles generate distinct transcript isoforms through differential promoter usage, alternative splicing, transcript termination, or alternative polyadenylation (APA).

Studies of mouse pluripotent states using allele-specific RNA-seq have shown that parent-of-origin-dependent expression contributes to transcriptional heterogeneity associated with pluripotency and developmental potential[86], while allele-specific binding of the imprinting regulator ZFP57 demonstrates that parent-of-origin epigenetic regulation extends beyond classical imprinted loci and participates in the establishment and maintenance of monoallelic expression programs during early development[87]. Similarly, APA generates transcript isoforms with distinct 3' UTR lengths that influence mRNA stability and translation efficiency, and single-cell 3' UTR profiling shows that these APA patterns are stage-specific in human and mouse preimplantation embryos[88-91]. Although these studies reveal important regulatory mechanisms, SRS cannot reliably determine whether maternal and paternal alleles produce distinct transcript architectures through differential promoter usage, alternative splicing, transcript termination, or APA, because full-length transcript structures are fragmented across multiple reads.

LRS is suited to resolving allele-specific transcript diversity because individual reads span multiple heterozygous variants together with complete transcript structures[92]. Recent work combining PacBio Iso-Seq with WhatsHap-based phasing in polymorphic F1 mouse hybrids demonstrated that full-length reads can be assigned to parental alleles using single-nucleotide polymorphisms, enabling isoform-level analysis of allele-specific loci, including imprinted coding and non-coding genes[37]. By preserving transcript structures, including promoter usage, splice-junction combinations, transcript termination sites, and alternative polyadenylation events, LRS can link parental origin to transcript architecture and enable direct assessment of how maternal and paternal genomes contribute to transcript diversity.

Beyond resolving allele-specific isoforms, LRS can also characterize transcript-end features, including APA, 3' UTR remodeling, and poly(A) tail dynamics, from individual RNA molecules[40,89]. When integrated with phased genomic variants, these measurements enable reconstruction of parent-of-origin-specific transcript isoforms while connecting transcript architecture with post-transcriptional regulation. Such

haplotype-resolved analyses provide an opportunity to investigate imprinting, parent-of-origin expression, and the relationship between genomic context and transcript diversity. Collectively, these advances are shifting embryonic transcriptomics from gene-level quantification toward haplotype-resolved, isoform-level characterization of the transcriptome. By integrating phased diploid genome assemblies with LRS and complementary epigenomic approaches, researchers can simultaneously connect parental origin with promoter usage, alternative splicing, transcript termination, APA, and poly(A) tail dynamics within full-length transcripts.

Haplotype-resolved genome assembly, structural variation and genome stability

Computational analyses of RNA-seq data depend on high-quality reference genomes. Extending telomere-to-telomere (T2T)-inspired strategies to embryonic samples[29] through haplotype-resolved *de novo* diploid assembly provides a more accurate framework for interpreting transcriptional outputs, because such assemblies enable the resolution of allele-specific expression, isoform usage, and parent-of-origin-dependent regulation. By reconstructing paternal and maternal haplotypes independently, LRS allows precise assignment of RNA reads to their allelic origins, which is essential for studying genomic imprinting during early development[93-95]. High-fidelity LRS further supports phased assemblies that separate paralogous regions and structural variants, improving the characterization of gene structure, promoter usage, and transcript isoform diversity, although robust assembly from low-input embryonic material remains technically challenging[95].

Beyond genome reconstruction, LRS detects forms of genomic variation with direct transcriptional consequences that are largely invisible to SRS. Long-read whole-genome sequencing has uncovered cryptic structural variants in ESCs and induced PSCs (iPSCs), including tandem repeat expansions and transposon insertions that can alter gene expression, introduce aberrant transcription start sites, or generate chimeric transcripts[96]. Large-scale iPSC resources further demonstrate the prevalence of such variants and their impact on regulatory regions and transcript structures[96], and direct comparisons confirm that long reads recover structural variants that short reads systematically miss[96-98].

Genome stability in pluripotent cells is closely tied to retrotransposon activity, and LRS provides the detail needed to study this relationship directly. Primed mouse PSCs show elevated retrotransposon expression together with increased genomic instability relative to naive cells, a difference that is difficult to characterize with short reads because newly inserted elements are repetitive and often structurally complex[99]. Nanopore sequencing captures transposable element insertions together with their flanking sequence and their DNA methylation state, enabling locus-resolved analysis of element structure and chromatin context in a single assay[100]. Applying these approaches to pluripotent cells allows *de novo* insertions to be cataloged against the genetic background of the parental line, which is relevant both to the genomic integrity of stem cell resources and to the interpretation of developmental phenotypes.

Ultra-long reads also make it possible to measure chromosome ends directly. PacBio HiFi sequencing has been used to measure telomere length at nucleotide resolution in high throughput, distinguishing individual chromosome arms rather than reporting a population average[101]. In stem cell and developmental contexts, this provides a way to track telomere maintenance across passaging and differentiation, and to relate chromosome-end structure to the replicative behavior of cells that must be expanded in culture before they can be studied.

Together, these advances illustrate the growing potential of LRS to resolve complex genomic architectures that remain largely invisible to conventional SRS-based approaches. At the same time, the routine deployment of *de novo* diploid assembly, comprehensive structural variant detection, and chromosome-end

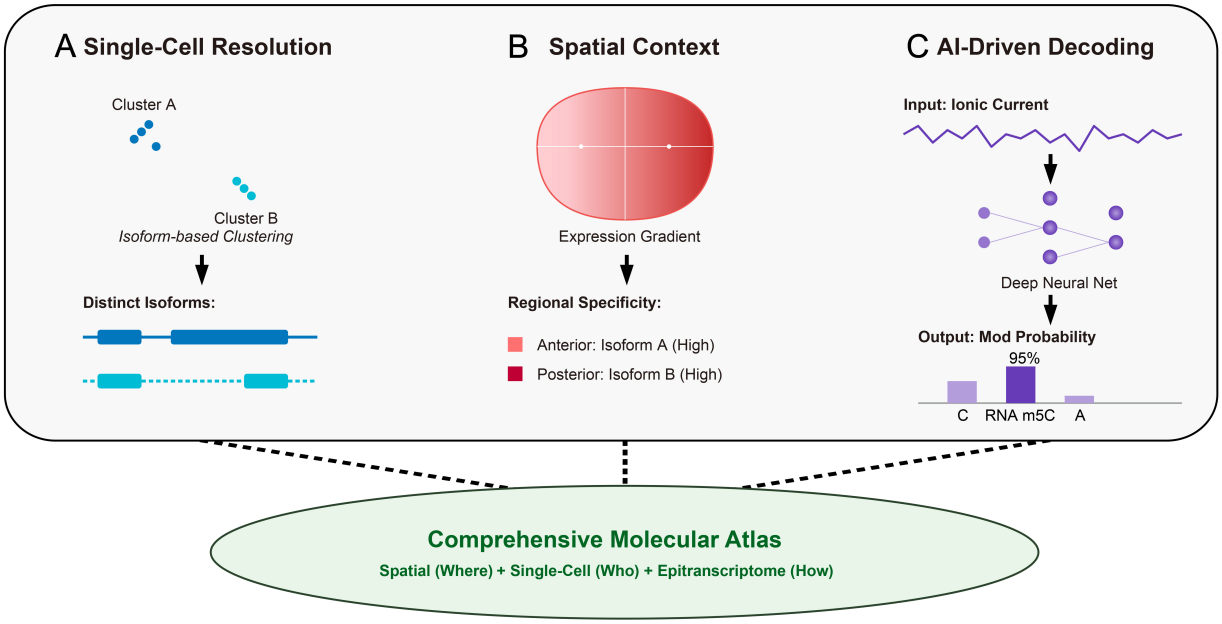


Figure 3. The future ecosystem of developmental genomics: converging frontiers. The next generation of LRS applications will be defined by three emerging pillars; (A) Single-cell resolution: moving from bulk to single-cell long-read sequencing (scLRS) to resolve isoform diversity at cellular granularity. Isoform-based clustering separates cell subpopulations (for example, Cluster A and Cluster B) that gene-level clustering merges; (B) Spatial context: spatial barcoding maps full-length isoforms to anatomical coordinates within developing tissue, so that expression gradients can be assigned to defined regions (for example, anterior versus posterior) and to specific cell types; (C) AI-driven decoding: deep neural networks interpret raw ionic current signal to estimate modification probabilities at individual positions, distinguishing, for example, RNA 5-methylcytosine (m5C) from unmodified cytosine without chemical conversion. The three pillars converge to produce a comprehensive molecular atlas that combines the spatial (where), single-cell (who), and epitranscriptomic (how) dimensions of development. The individual examples shown in the panel are illustrative of each pillar rather than observations from a single study. AI: Artificial intelligence.

measurement in embryonic and stem cell research will depend on continued improvements in sequencing accuracy, cost efficiency, and low-input library preparation. Addressing these challenges will be essential for translating high-resolution genome analysis from a specialized capability into a reliable foundation for developmental genomics and regenerative medicine.

EMERGING APPLICATIONS OF LRS

Three frontiers are actively shaping the development of LRS technologies: the integration of LRS with single-cell modalities, the extension of long-read measurements into spatial context, and the use of machine learning to decode epigenetic and epitranscriptomic information directly from raw signal. These three frontiers correspond to the three pillars of Figure 3, and they converge on a common objective: building quantitative, spatially resolved atlases of the molecular logic that governs cell fate. The genome assembly and structural variation applications discussed in Section "Haplotype-resolved genome assembly, structural variation and genome stability" underpin all three, because every downstream measurement inherits the quality of the reference to which it is aligned.

Advances in single-cell long-read sequencing

Single-cell RNA sequencing (scRNA-seq) has advanced our understanding of human embryo development by revealing cellular heterogeneity and developmental trajectories that are obscured in bulk analyses[95,102], and single-cell multi-omics has extended this to the chromatin and methylation layers[71]. Building on this foundation, recent advances in LRS are extending single-cell transcriptomics into an isoform- and modification-resolved dimension. Historically, scRNA-seq relied on SRS reads from fragmented transcripts,

often with 3'-end bias, to quantify gene-level expression, with limited ability to reconstruct full-length isoforms. Recent methodological innovations are overcoming throughput and cost limitations through concatenation-based approaches such as PacBio Multiplexed arrays sequencing (MAS-Seq), which concatenates cDNA molecules before sequencing and was reported to increase throughput by approximately 15-fold relative to standard Iso-Seq, yielding nearly 40 million cDNA reads per run[103].

The integration of LRS with single-cell technologies has revealed cell-to-cell variation in TE expression and transcript structure that is obscured in bulk datasets. The original single-cell long-read method, single-cell long-read RNA sequencing (CELLO-seq), enabled full-length sequencing of TE-containing transcripts in individual mouse embryonic cells and human PSCs, allowing precise locus-specific quantification and isoform reconstruction of transcripts that cannot be assigned to a single genomic origin with short reads[104]; a detailed protocol is now available[105]. Collectively, these advances demonstrate that LRS is shifting the study of repetitive genomic regions from cataloging TE expression to resolving their locus-specific regulatory functions. By enabling accurate locus assignment, full-length transcript reconstruction, and single-cell characterization of TE activity, LRS provides an integrated framework for understanding how repetitive elements shape gene regulation during ZGA, pluripotency maintenance, and lineage specification[106].

Single-cell LRS is also proving pivotal for resolving heterogeneity within cell populations at the isoform level, in which cells that appear transcriptionally similar at the gene level express distinct coding and non-coding isoforms. Single-cell implementations of LRS are also being extended to poly(A) tail and modification-aware measurements, although these remain less mature than isoform-level analysis; isoform-level data currently support developmental trajectories built on isoforms rather than genes. Two illustrative studies show what this resolution provides. In human cerebral organoids, single-cell LRS uncovered cell-type-specific and autism-associated exons that were completely masked in short-read data[107]; although organoids do not reproduce the preimplantation embryo, they demonstrate the gain in resolution that single-cell LRS provides for a developing human tissue. In preimplantation embryos, single-cell isoform and TE analysis has identified isoform-switching events associated with lineage specification[36].

Advances in spatial long-read transcriptome sequencing

Beyond dissociated single cells, an emerging frontier lies in preserving spatial context. By integrating spatial barcoding with long-read platforms, full-length transcript isoforms can be assigned to anatomical coordinates within developing tissues, so that splicing and polyadenylation decisions can be compared between adjacent cell populations that share a transcriptional program. Because these methods are recent, the published examples are few, and it is important to distinguish assays that generate long-read sequence data themselves from computational analyses of conventional spatial transcriptomics data that infer isoform structure without long-read support.

For spatial LRS in the strict sense, two studies illustrate the current state of the art. Foord *et al.* combined LRS with Slide-seqV2 to reveal layer-specific splicing regulation in the developing human cortex, linking alternative splicing to postsynaptic maturation in distinct neuronal layers, and to identify developmental regulation of polyadenylation sites between cortical cell types[108]. Lebrigand *et al.* mapped spatial isoform switching of Plp1 across olfactory bulb layers in the mouse brain and generated the first spatial A-to-I editing map[109]. Together, these studies demonstrate that spatial long-read approaches can resolve isoform diversity and RNA processing events within their native tissue context, revealing regulatory patterns that the conventional short-read spatial atlas of the mouse embryo cannot resolve[110]. Their value for developmental biology lies in the ability to relate isoform choice to position, which is precisely the relationship that dissociated single-cell methods discard.

Alongside these assays, computational frameworks have been developed that infer the isoform landscape and regulatory mechanisms from spatial transcriptomics data. Spatial Isoform Statistical Modeling (SPLISOSM), for example, identified more than 1,000 spatially variable transcript diversity events in the mouse brain, enriched in synaptic pathways conserved in the human prefrontal cortex[111]. SPLISOSM operates on spatial transcriptomics data generally, and so its isoform assignments are inferred from short-read data rather than measured on long reads; it is included here because it demonstrates the kinds of spatial isoform questions that long-read spatial assays can now address directly. The combination of long-read spatial assays with such computational frameworks, and the extension of both to embryonic tissue[112], is a promising direction. Isoform structure inferred from short-read spatial data is informative and agrees broadly with long-read measurements, but long-read data provide direct confirmation when isoform-level structure is the point of the experiment.

Artificial intelligence-driven decoding of the epitranscriptome

The epitranscriptome constitutes an additional regulatory layer that can shape RNA fate during development[113,114]. LRS, particularly nanopore-based direct RNA sequencing (dRNA-seq), offers a unique opportunity to interrogate this layer by sequencing native RNA molecules without reverse transcription or polymerase chain reaction (PCR) amplification. By contrast, PacBio requires reverse transcription to cDNA and therefore cannot be used for direct RNA modification detection [Table 1]. Importantly, this approach allows the quantitative estimation of modification stoichiometry rather than merely inferring the presence of a modification. Unlike antibody-based enrichment methods such as methylated RNA immunoprecipitation and sequencing (MeRIP-seq)[115], which provide transcript population-averaged and relative signals, LRS yields a digital, molecule-resolved readout. This quantitative capability is particularly relevant for developmental systems, where subtle, dynamic shifts in modification rates can fine-tune transcript stability, translation efficiency, and cell fate decisions. For instance, in human hematopoietic stem cells, the m^{6}A reader YTHDF2 targets TAL1 mRNA for degradation, and knocking down YTHDF2 stabilizes TAL1, thereby promoting ex vivo hematopoietic stem cell expansion rather than differentiation[116]. This example lies outside embryonic development but illustrates the general principle that a single modification reader can determine whether a transcript is translated, stored, or degraded.

Beyond modification detection, LRS enables simultaneous observation of multiple RNA features on the same molecule, providing a single-molecule link between site-specific m^{6}A modification patterns, alternative splicing patterns such as exon skipping versus inclusion, and poly(A) tail length dynamics[72,73]. Recent studies have shown that modification clusters are associated with isoform-specific polyadenylation events, and that these relationships are not captured by short-read or enrichment-based approaches[72,76]. Such multi-layered measurements are especially relevant during the maternal-to-zygotic transition, where rapid degradation of maternal transcripts and activation of the zygotic genome require precise, multi-level post-transcriptional coordination[14].

The principal bottleneck in epitranscriptomic long-read analysis currently lies in the accurate decoding of raw nanopore ionic current signals. Progress will depend on modification-aware basecallers built on deep learning architectures, including convolutional neural networks (CNNs) for local signal feature extraction and transformers or bidirectional long short-term memory networks (Bi-LSTMs) for capturing long-range signal dependencies. Training these models requires ground-truth benchmark datasets derived from *in vitro* transcription (IVT) systems or synthetic RNA oligonucleotides carrying modifications at precisely defined positions. An important direction is the creation of integrated analytical pipelines that use nanopore signal features, such as dwell time and current intensity deviations, to report sequence, splice isoform identity, poly(A) tail length, and site-specific modification probabilities jointly from the same RNA molecule. For developmental biology, the practical value of such pipelines will depend on whether they retain accuracy

when applied to the small numbers of cells available from preimplantation embryos, which remains an open question.

CURRENT CHALLENGES AND FUTURE DIRECTIONS

LRS is being adopted rapidly, but several technical and analytical constraints still limit what can be concluded from long-read data in developmental systems. The challenges below are grouped by whether they arise from the measurement itself, from the analysis of the resulting data, or from the inferential gap between a catalog of molecular features and a mechanistic explanation of development.

Technical constraints of current platforms

Input requirements remain the most immediate obstacle for embryonic work. Early embryos and individual blastomeres contain picogram quantities of nucleic acid, and although low-input protocols have improved substantially, DNA-based long-read assays generally require more material than RNA-based assays, which constrains the developmental stages and cell numbers that can be examined. Throughput and cost remain limiting for single-cell long-read sequencing. Compared with conventional short-read scRNA-seq, single-cell long-read sequencing (scLRS) generally provides lower throughput and carries higher sequencing cost per cell, which restricts the number of cells and developmental stages that can be profiled at sufficient depth. Most current scLRS workflows rely on reverse transcription and PCR amplification, introducing amplification bias and uneven isoform recovery, particularly for low-abundance transcripts, and limited reads per cell lead to substantial data sparsity that makes it difficult to distinguish rare isoforms or cell-state-specific splicing events from technical dropout[57]. Each platform also imposes a trade-off between read length and per-base accuracy: high-accuracy consensus approaches shorten effective read lengths, whereas the longest reads carry higher error rates, and the choice between these regimes determines which questions a study can address [Table 1]. Finally, quantitative comparisons between platforms in developmental contexts are still scarce; benchmarking of long-read RNA-seq methods has been carried out mainly in cell lines, which are themselves heterogeneous[117], and the extent to which those results transfer to low-input embryonic samples is not yet established[27,34,118].

Analytical and computational challenges

Analysis of long-read data is limited less by read generation than by the maturity of the software and the availability of appropriate reference resources. Isoform quantification, transcript discovery and differential splicing analysis each depend on assumptions that were developed for short reads, and their accuracy varies substantially between tools and datasets[34,118,119]. For repetitive elements, the central analytical difficulty is that reads or read fragments that lie entirely within a repeat family cannot be assigned to a locus; only reads that span a repeat and a unique flank can be anchored, which means that locus-specific quantification is systematically biased toward elements that happen to sit near unique sequence. Reference quality compounds this problem. Many developmentally important genomes remain incompletely assembled, and where structural variation, tandem repeats, or unplaced sequence are unresolved, read alignment is correspondingly unreliable; haplotype-resolved assemblies address part of this problem but are themselves difficult to produce from low-input material[93,94]. Modification analysis adds a further layer of uncertainty, because estimated modification probabilities depend on the training data and the basecalling model used, and comparisons across studies that used different models are not yet straightforward[32,33]. Standardized benchmarking datasets for embryonic material, together with shared standards for reporting modification calls, remain a high priority.

From molecular catalogs toward mechanism

The most significant conceptual limitation is inferential rather than technical. Long-read studies of development have so far produced catalogs: lists of novel isoforms, locus-assigned TE transcripts, allele-specific transcripts, and modification sites. Catalogs support hypotheses but do not test them, and the

demonstration that a transcript structure or modification correlates with a developmental transition does not establish that it causes the transition. Closing this gap will require methods that perturb specific isoforms, polyadenylation sites, or modifications without altering the underlying gene, and readouts that can measure the consequences at single-molecule resolution. Because regulatory features are often linked in cis, the natural experimental design is one in which a single molecule is perturbed and then re-measured, which is now technically feasible but has been demonstrated in only a small number of systems.

Several directions follow from this analysis. First, single-cell and spatial long-read approaches need to be combined within the same experiment, so that isoform identity, cell state, and position are measured together rather than inferred separately[57,108,112]. Second, developmental reference resources should be built with haplotype-resolved genomes and matched isoform annotations, so that allele-specific analyses rest on an accurate substrate[94,95]. Third, benchmarking and reporting standards specific to low-input samples and to modification detection should be established, so that results from different studies can be compared directly[34]. Fourth, the interpretability of modification-aware models needs to be addressed, because a predictive model that reports a modification probability without exposing the signal features that drive it cannot readily be used to design an experiment[66,76]. Progress on these four fronts would move the field from describing the molecular landscape of early development to explaining how it determines cell fate.

CONCLUSION

LRS, exemplified by PacBio and ONT, has substantially expanded the methodological repertoire available for studying ESC biology and early embryonic development. By enabling single-molecule resolution and preserving long-range continuity, LRS addresses longstanding limitations of SRS, particularly in resolving full-length transcript isoforms, DNA and RNA modifications, repetitive elements, and structurally complex genomic regions. These capabilities have provided a more reliable representation of the native genomic and transcriptomic landscape that underlies early developmental processes.

Beyond improving structural resolution, LRS offers a unique window into dynamic regulatory layers that are difficult to capture using conventional approaches. The direct measurement of alternative splicing, TE activity, allele-specific variation, and post-transcriptional features such as RNA methylation and poly(A) tail length has begun to reshape how the regulatory landscape of early development is interrogated. These advances extend beyond descriptive profiling and provide mechanistic insights that are increasingly relevant to both basic research and translational applications, including high-resolution genetic assessment in preimplantation contexts[120-122].

Because individual reads span several regulatory features at once, LRS also allows those features to be related to one another within a single molecule rather than across averaged datasets. LRS complements rather than replaces existing sequencing paradigms, and its value lies in this capacity to link sequence, structure, and molecular modification in one measurement. As the technical and analytical challenges discussed above are addressed, LRS is poised to play an increasingly central role in developmental genomics, offering a more integrated and information-rich perspective on the molecular logic that governs early life. The application of LRS extends beyond embryonic development. Indeed, LRS technology continues to provide great insights into different aspects of biology and medicine.

DECLARATIONS

Authors' contributions

Performed the literature review and revised the manuscript: Tang H, Babarinde IA, Mu Y, Shen Q, Qiao Y, Xiao Z

Conceived and supervised the project and drafted the manuscript: Hutchins AP, Ma G

All authors read and approved the final version.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool ChatGPT (version GPT-5, released 2025-08-07) was used in two ways: to check the language and the formatting of the reference list, and, at the earliest stage of the work, to help frame the initial concept of the Graphical Abstract and Figures 2-3. It was not used to generate, analyze, or interpret any data, and its contribution to the figures was confined to that initial concept and structure. The Graphical Abstract and Figures 1-3 were designed by the authors. Every change suggested by the tool was reviewed by the authors before it was adopted. 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 National Natural Science Foundation of China (32500456 to Ma G and W2512079 to Hutchins AP) and the Young Elite Scientist Sponsorship Program of the Guangdong Association for Science and Technology (SKXRC2025357 to Ma G).

Conflicts of interest

All authors declare that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. Shahbazi MN. Mechanisms of human embryo development: from cell fate to tissue shape and back. *Development.* 2020:147. DOI PubMed PMC
2. Alberio R. Regulation of cell fate decisions in early mammalian embryos. *Annu Rev Anim Biosci.* 2020;8:377-93. DOI PubMed
3. Sun L, Fu X, Ma G, Hutchins AP. Chromatin and epigenetic rearrangements in embryonic stem cell fate transitions. *Front Cell Dev Biol.* 2021;9:637309. DOI PubMed PMC
4. Ma Y, Li Y, Jin Z. Stem cell-based therapy for myopic maculopathy: a new concept. *J Transl Genet Genom.* 2022;6:179-203. DOI
5. Wang X, Xiao Z. Extraembryonic mesoderm in primate and rodent embryos and models. *J Transl Genet Genom.* 2026;10:262-88. DOI
6. Du P, Wu J. Hallmarks of totipotent and pluripotent stem cell states. *Cell Stem Cell.* 2024;31:312-33. DOI PubMed PMC
7. Babarinde IA, Ma G, Li Y, et al. Transposable element sequence fragments incorporated into coding and noncoding transcripts modulate the transcriptome of human pluripotent stem cells. *Nucleic Acids Res.* 2021;49:9132-53. DOI PubMed PMC
8. Ma G, Babarinde IA, Zhou X, Hutchins AP. Transposable elements in pluripotent stem cells and human disease. *Front Genet.* 2022;13:902541. DOI PubMed PMC
9. Sakashita A, Kitano T, Ishizu H, et al. Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nat Genet.* 2023;55:484-95. DOI PubMed PMC
10. Macfarlan TS, Gifford WD, Driscoll S, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature.* 2012;487:57-63. DOI PubMed PMC
11. De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet.* 2017;49:941-5. DOI PubMed PMC
12. Percharde M, Lin CJ, Yin Y, et al. A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell.* 2018;174:391-405.e19. DOI PubMed PMC
13. Burton A, Torres-Padilla ME. Epigenome dynamics in early mammalian embryogenesis. *Nat Rev Genet.* 2025;26:587-603. DOI PubMed

14. Lee K, Cho K, Morey R, Cook-Andersen H. An extended wave of global mRNA deadenylation sets up a switch in translation regulation across the mammalian oocyte-to-embryo transition. *Cell Rep.* 2024;43:113710. DOI PubMed PMC
15. Oomen ME, Rodriguez-Terrones D, Kurome M, et al. An atlas of transcription initiation reveals regulatory principles of gene and transposable element expression in early mammalian development. *Cell.* 2025;188:1156-1174.e20. DOI
16. Jin Y, Liang X, Wang X. Alternative splicing in stem cells and development: research progress and emerging technologies. *Cell Regen.* 2025;14:20. DOI PubMed PMC
17. Heather JM, Chain B. The sequence of sequencers: the history of sequencing DNA. *Genomics.* 2016;107:1-8. DOI PubMed PMC
18. Kuo RI, Cheng Y, Zhang R, et al. Illuminating the dark side of the human transcriptome with long read transcript sequencing. *BMC Genomics.* 2020;21:751. DOI PubMed PMC
19. Dyer SC, Austine-Orimoloye O, Azov AG, et al. Ensembl 2025. *Nucleic Acids Res.* 2025;53:D948-57. DOI
20. Lahens NF, Kavakli IH, Zhang R, et al. IVT-seq reveals extreme bias in RNA sequencing. *Genome Biol.* 2014;15:R86. DOI PubMed PMC
21. Babarinde IA, Hutchins AP. The effects of sequencing depth on the assembly of coding and noncoding transcripts in the human genome. *BMC Genomics.* 2022;23:487. DOI PubMed PMC
22. Babarinde IA, Li Y, Hutchins AP. Computational methods for mapping, assembly and quantification for coding and non-coding transcripts. *Comput Struct Biotechnol J.* 2019;17:628-37. DOI PubMed PMC
23. Monzó C, Liu T, Conesa A. Transcriptomics in the era of long-read sequencing. *Nat Rev Genet.* 2025;26:681-701. DOI PubMed
24. Gupta P, O'Neill H, Wolvetang EJ, Chatterjee A, Gupta I. Advances in single-cell long-read sequencing technologies. *NAR Genom Bioinform.* 2024;6:lqae047. DOI PubMed PMC
25. Ament IH, DeBruyne N, Wang F, Lin L. Long-read RNA sequencing: a transformative technology for exploring transcriptome complexity in human diseases. *Mol Ther.* 2025;33:883-94. DOI PubMed PMC
26. Marx V. Method of the year: long-read sequencing. *Nat Methods.* 2023;20:6-11. DOI PubMed
27. Wissel D, Mehlferber MM, Nguyen KM, et al. A systematic benchmark of high-accuracy PacBio long-read RNA sequencing for transcript-level quantification. *Genome Biol.* 2026;27:110. DOI PubMed PMC
28. Kawakami S, Kitao K, Ikeda S, Honda S. Long-read sequencing reveals transposable element-derived chimeric transcripts at zygotic genome activation in mammalian embryos. *bioRxiv.* 2026. DOI
29. Liehr T. Inclusion of human centromeres in genomic browsers - will it finally come? *J Transl Genet Genom.* 2026;10:156-59. DOI
30. Workman RE, Tang AD, Tang PS, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods.* 2019;16:1297-305. DOI PubMed PMC
31. Akbari V, Garant JM, O'Neill K, et al. Genome-wide detection of imprinted differentially methylated regions using nanopore sequencing. *Elife.* 2022;11:e77898. DOI PubMed PMC
32. Liu H, Begik O, Lucas MC, et al. Accurate detection of m^{6}A RNA modifications in native RNA sequences. *Nat Commun.* 2019;10:4079. DOI PubMed PMC
33. Pratanwanich PN, Yao F, Chen Y, et al. Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat Biotechnol.* 2021;39:1394-402. DOI
34. Pardo-Palacios FJ, Wang D, Reese F, et al. Systematic assessment of long-read RNA-seq methods for transcript identification and quantification. *Nat Methods.* 2024;21:1349-63. DOI
35. Torre D, Francoeur NJ, Kalma Y, et al. Isoform-resolved transcriptome of the human preimplantation embryo. *Nat Commun.* 2023;14:6902. DOI PubMed PMC
36. Wang C, Shi Z, Huang Q, et al. Single-cell analysis of isoform switching and transposable element expression during preimplantation embryonic development. *PLoS Biol.* 2024;22:e3002505. DOI PubMed PMC
37. Lemoine L, Hoelzl S, Hasenbein TP, Graf E, Andergassen D. Long-read sequencing disentangles isoform complexity at allele-specific loci. *Sci Rep.* 2025;15:39389. DOI PubMed PMC
38. Gigante S, Gouil Q, Lucattini A, et al. Using long-read sequencing to detect imprinted DNA methylation. *Nucleic Acids Res.* 2019;47:e46. DOI PubMed PMC
39. Wang Y, Li Y, Skuland T, et al. The RNA m^{6}A landscape of mouse oocytes and preimplantation embryos. *Nat Struct Mol Biol.* 2023;30:703-9. DOI PubMed PMC
40. Begik O, Diensthuber G, Liu H, et al. Nano3P-seq: transcriptome-wide analysis of gene expression and tail dynamics using end-capture nanopore cDNA sequencing. *Nat Methods.* 2023;20:75-85. DOI PubMed PMC
41. Qiao Y, Ren C, Huang S, et al. High-resolution annotation of the mouse preimplantation embryo transcriptome using long-read sequencing. *Nat Commun.* 2020;11:2653. DOI PubMed PMC

42. Jia YK, Yu Y, Guan L. Advances in understanding the regulation of pluripotency fate transition in embryonic stem cells. *Front Cell Dev Biol.* 2024;12:1494398. DOI
43. Ule J, Blencowe BJ. Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol Cell.* 2019;76:329-45. DOI PubMed
44. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010;28:511-5. DOI PubMed PMC
45. Gabut M, Samavarchi-Tehrani P, Wang X, et al. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell.* 2011;147:132-46. DOI
46. Han H, Irimia M, Ross PJ, et al. MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature.* 2013;498:241-5. DOI
47. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell.* 1999;99:247-57. DOI
48. Mukamel Z, Lifshitz A, Mittnenzweig M, et al. DNA methyltransferases 3A and 3B target specific sequences during mouse gastrulation. *Nat Struct Mol Biol.* 2022;29:1252-65. DOI
49. Liao J, Karnik R, Gu H, et al. Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nat Genet.* 2015;47:469-78. DOI PubMed PMC
50. Gopalakrishna-Pillai S, Iverson LE. A DNMT3B alternatively spliced exon and encoded peptide are novel biomarkers of human pluripotent stem cells. *PLoS One.* 2011;6:e20663. DOI PubMed PMC
51. Gopalakrishnan S, Van Emburgh BO, Shan J, et al. A novel DNMT3B splice variant expressed in tumor and pluripotent cells modulates genomic DNA methylation patterns and displays altered DNA binding. *Mol Cancer Res.* 2009;7:1622-34. DOI PubMed PMC
52. Zeng Y, Ren R, Kaur G, et al. The inactive Dnmt3b3 isoform preferentially enhances Dnmt3b-mediated DNA methylation. *Genes Dev.* 2020;34:1546-58. DOI PubMed PMC
53. Duymich CE, Charlet J, Yang X, Jones PA, Liang G. DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. *Nat Commun.* 2016;7:11453. DOI PubMed PMC
54. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet.* 2008;40:1413-5. DOI PubMed
55. Glinos DA, Garborcauskas G, Hoffman P, et al. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature.* 2022;608:353-9. DOI PubMed PMC
56. Chau KK, Zhang P, Urresti J, et al. Full-length isoform transcriptome of the developing human brain provides further insights into autism. *Cell Rep.* 2021;36:109631. DOI PubMed PMC
57. Wen L, Tang F. Single-cell omics sequencing technologies: the long-read generation. *Trends Genet.* 2026;42:46-62. DOI
58. Takada Y, Fierro L, Sato K, et al. Mature mRNA processing that deletes 3' end sequences directs translational activation and embryonic development. *Sci Adv.* 2023;9:eadg6532. DOI
59. Zhao LY, Song J, Liu Y, Song CX, Yi C. Mapping the epigenetic modifications of DNA and RNA. *Protein Cell.* 2020;11:792-808. DOI PubMed PMC
60. Urakawa T, Hattori A, Ogiwara Y, et al. A comprehensive long-read sequencing system to assess DNA methylation at differentially methylated regions and imprinting-disorder-related genes. *Genome Med.* 2025;17:144. DOI PubMed PMC
61. Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature.* 2014;508:66-71. DOI PubMed PMC
62. Eichhorn SW, Subtelny AO, Kronja I, Kwasnieski JC, Orr-Weaver TL, Bartel DP. mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in drosophila oocytes and early embryos. *Elife.* 2016;5:e16955. DOI PubMed PMC
63. Goldstrohm AC, Wickens M. Multifunctional deadenylase complexes diversify mRNA control. *Nat Rev Mol Cell Biol.* 2008;9:337-44. DOI PubMed
64. Yang J, Bu J, Liu B, et al. MARTRE family proteins negatively regulate CCR4-NOT activity to protect poly(A) tail length and promote translation of maternal mRNA. *Nat Commun.* 2025;16:248. DOI PubMed PMC
65. Smith ZD, Chan MM, Humm KC, et al. DNA methylation dynamics of the human preimplantation embryo. *Nature.* 2014;511:611-5. DOI PubMed PMC
66. Yu B, Nagae G, Midorikawa Y, et al. m6ATM: a deep learning framework for demystifying the m6A epitranscriptome with Nanopore long-read RNA-seq data. *Brief Bioinform.* 2024;25:bbae529. DOI PubMed PMC
67. Ren Z, He J, Huang X, et al. Isoform characterization of m^6A in single cells identifies its role in RNA surveillance. *Nat Commun.* 2025;16:5828. DOI PubMed PMC
68. Lee I, Razaghi R, Gilpatrick T, et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat Methods.* 2020;17:1191-9. DOI PubMed PMC

69. Hu Y, Jiang Z, Chen K, et al. scNanoATAC-seq: a long-read single-cell ATAC sequencing method to detect chromatin accessibility and genetic variants simultaneously within an individual cell. *Cell Res.* 2023;33:83-6. DOI PubMed PMC
70. Jachowicz JW, Bing X, Pontabry J, Bošković A, Rando OJ, Torres-Padilla ME. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat Genet.* 2017;49:1502-10. DOI PubMed
71. Li L, Guo F, Gao Y, et al. Single-cell multi-omics sequencing of human early embryos. *Nat Cell Biol.* 2018;20:847-58. DOI
72. Kim Y, Saville L, O'Neill K, et al. Nanopore direct RNA sequencing of human transcriptomes reveals the complexity of mRNA modifications and crosstalk between regulatory features. *Cell Genom.* 2025;5:100872. DOI PubMed PMC
73. Leger A, Amaral PP, Pandolfini L, et al. RNA modifications detection by comparative nanopore direct RNA sequencing. *Nat Commun.* 2021;12:7198. DOI PubMed PMC
74. Park D, Cenik C. Long-read RNA sequencing reveals allele-specific N^6-methyladenosine modifications. *Genome Res.* 2025;35:999-1011. DOI PubMed PMC
75. Wang Y, Wang A, Liu Z, et al. Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res.* 2019;29:1329-42. DOI PubMed PMC
76. Liu T, Conesa A. Profiling the epigenome using long-read sequencing. *Nat Genet.* 2025;57:27-41. DOI PubMed
77. Peaston AE, Evsikov AV, Graber JH, et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell.* 2004;7:597-606. DOI
78. Zhang W, Chen F, Chen R, et al. Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. *Nucleic Acids Res.* 2019;47:8485-501. DOI PubMed PMC
79. Babarinde IA, Fu X, Ma G, et al. Transposable element expression and sub-cellular dynamics during hPSC differentiation to endoderm, mesoderm, and ectoderm lineages. *Nat Commun.* 2025;16:7670. DOI
80. Zhang Y, She J, Hu X, et al. Transposable elements drive species-specific and tissue-specific transcriptomes in human development. *Genome Biol.* 2025;26:379. DOI PubMed PMC
81. Ho JSY, Douse CH, Marazzi I. Transposable element DNA and RNA: drivers of gene expression, evolution, and disease. *Cell.* 2026;189:3513-40. DOI PubMed PMC
82. Cheon Y, Alvstad EG, Torre D, et al. Transposable element-gene chimera cartography, origination and role in enhancing transcriptome plasticity. *Nat Struct Mol Biol.* 2026;33:448-63. DOI
83. Fort V, Khelifi G, He R, et al. Long-read RNA sequencing identifies non-coding isoform switching as a regulator of cell fate. *Nat Commun.* 2026:17. DOI PubMed PMC
84. Bo J, Fang W, Wang J, He S, Yang L. High resolution of full-length RNA sequencing deciphers massive transcriptome complexity during zebrafish embryogenesis. *BMC Biol.* 2025;23:155. DOI PubMed PMC
85. Tittarelli E, Carotti E, Carducci F, Barucca M, Canapa A, Biscotti MA. Transposable element dynamics in Xenopus laevis embryogenesis: a tale of two coexisting subgenomes. *Mob DNA.* 2025;16:17. DOI PubMed PMC
86. Dirks RAM, van Mierlo G, Kerstens HHD, et al. Allele-specific RNA-seq expression profiling of imprinted genes in mouse isogenic pluripotent states. *Epigenetics Chromatin.* 2019;12:14. DOI PubMed PMC
87. Strogantsev R, Krueger F, Yamazawa K, et al. Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biol.* 2015;16:112. DOI PubMed PMC
88. Zhang Y, Huang Z, Lu W, Liu Z. Alternative cleavage and polyadenylation: key regulatory mechanisms in health and disease. *RNA Biol.* 2025;22:1-33. DOI PubMed PMC
89. Zhang Y, Li A, Wu J, Tuzankina IA, Xu S, Xing Y. The dynamic landscape of alternative 3' UTR during mammalian preimplantation development. *Dev Biol.* 2026;534:101-14. DOI
90. Di Giammartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. *Mol Cell.* 2011;43:853-66. DOI PubMed PMC
91. Mayr C. What Are 3' UTRs doing? *Cold Spring Harb Perspect Biol.* 2019;11:a034728. DOI PubMed PMC
92. Wu J, Hu W, Li S. Long-read transcriptome sequencing reveals allele-specific variants at high resolution. *Trends Genet.* 2023;39:31-3. DOI
93. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* 2021;18:170-5. DOI PubMed PMC
94. Jarvis ED, Formenti G, Rhie A, et al.; Human pangenome reference consortium. Semi-automated assembly of high-quality diploid human reference genomes. *Nature.* 2022;611:519-31. DOI
95. Zhao C, Plaza Reyes A, Schell JP, et al. A comprehensive human embryo reference tool using single-cell RNA-sequencing data. *Nat Methods.* 2025;22:193-206. DOI PubMed PMC
96. Scheinfeldt L, Pompetti A, Calendo G, et al. New iPSC resource with long-read whole genome sequencing characterizations for enhanced in vitro modeling. *BMC Genomics.* 2026. DOI

97. Kosugi S, Terao C. Comparative evaluation of SNVs, indels, and structural variations detected with short- and long-read sequencing data. *Hum Genome Var.* 2024;11:18. DOI PubMed PMC
98. Stevanovski I, Chintalaphani SR, Gamaarachchi H, et al. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Sci Adv.* 2022;8:eabm5386. DOI PubMed PMC
99. Fu H, Zhang W, Li N, et al. Elevated retrotransposon activity and genomic instability in primed pluripotent stem cells. *Genome Biol.* 2021;22:201. DOI PubMed PMC
100. Ewing AD, Smits N, Sanchez-Luque FJ, et al. Nanopore sequencing enables comprehensive transposable element epigenomic profiling. *Mol Cell.* 2020;80:915-928.e5. DOI
101. Tham CY, Poon L, Yan T, et al. High-throughput telomere length measurement at nucleotide resolution using the PacBio high fidelity sequencing platform. *Nat Commun.* 2023;14:281. DOI PubMed PMC
102. Gul S, Zhang C. Single cell RNA sequencing and its impact on understanding human embryo development. *Int J Mol Sci.* 2025;26:7741. DOI PubMed PMC
103. Al'Khafaji AM, Smith JT, Garimella KV, et al. High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nat Biotechnol.* 2024;42:582-6. DOI PubMed PMC
104. Berrens RV, Yang A, Laumer CE, et al. Locus-specific expression of transposable elements in single cells with CELLO-seq. *Nat Biotechnol.* 2022;40:546-54. DOI PubMed PMC
105. Marlow SA, Deaville LA, Berrens RV. Long-read RNA sequencing of transposable elements from single cells using CELLO-seq. *Nat Protoc.* 2025;20:3070-95. DOI PubMed PMC
106. Deaville LA, Berrens RV. Technology to the rescue: how to uncover the role of transposable elements in preimplantation development. *Biochem Soc Trans.* 2024;52:1349-62. DOI PubMed PMC
107. Yang Y, Yang R, Kang B, Qian S, He X, Zhang X. Single-cell long-read sequencing in human cerebral organoids uncovers cell-type-specific and autism-associated exons. *Cell Rep.* 2023;42:113335. DOI PubMed PMC
108. Foord C, Prjibelski AD, Hu W, et al. A spatial long-read approach at near-single-cell resolution reveals developmental regulation of splicing and polyadenylation sites in distinct cortical layers and cell types. *Nat Commun.* 2025;16:8093. DOI PubMed PMC
109. Lebrigand K, Bergenstråhle J, Thrane K, et al. The spatial landscape of gene expression isoforms in tissue sections. *Nucleic Acids Res.* 2023;51:e47. DOI PubMed PMC
110. Sampath Kumar A, Tian L, Bolondi A, et al. Spatiotemporal transcriptomic maps of whole mouse embryos at the onset of organogenesis. *Nat Genet.* 2023;55:1176-85. DOI PubMed PMC
111. Su J, Qu Y, Schertzer M, et al. Mapping isoforms and regulatory mechanisms from spatial transcriptomics data with SPLISOSM. *Nat Biotechnol.* 2026. DOI
112. Belchikov N, Hsu J, Li XJ, et al. Understanding isoform expression by pairing long-read sequencing with single-cell and spatial transcriptomics. *Genome Res.* 2024;34:1735-46. DOI PubMed PMC
113. Li C, Chen K, Li X, Xiong X. Epitranscriptome-epigenome interactions in development and disease mechanisms. *Trends Genet.* 2025;41:691-705. DOI
114. El-Far A, Cheng J, Fu L, et al. Decoding the DHX36-m6A-G4 axis: exploring its role in health and disease through epigenetic and molecular insights. *J Transl Genet Genom.* 2026;10:89-118. DOI
115. Krusnauskas R, Stakaitis R, Steponaitis G, Almstrup K, Vaitkiene P. Identification and comparison of m6A modifications in glioblastoma non-coding RNAs with MeRIP-seq and nanopore dRNA-seq. *Epigenetics.* 2023;18:2163365. DOI PubMed PMC
116. Li Z, Qian P, Shao W, et al. Suppression of m^{6}A reader Ythdf2 promotes hematopoietic stem cell expansion. *Cell Res.* 2018;28:904-17. DOI
117. Zhu Q, Zhao X, Zhang Y, et al. Single cell multi-omics reveal intra-cell-line heterogeneity across human cancer cell lines. *Nat Commun.* 2023;14:8170. DOI PubMed PMC
118. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 2020;21:30. DOI PubMed PMC
119. Warburton PE, Sebra RP. Long-read DNA sequencing: recent advances and remaining challenges. *Annu Rev Genomics Hum Genet.* 2023;24:109-32. DOI PubMed
120. Zhang S, Liang F, Lei C, et al. Long-read sequencing and haplotype linkage analysis enabled preimplantation genetic testing for patients carrying pathogenic inversions. *J Med Genet.* 2019;56:741-9. DOI PubMed PMC
121. Shi Q, Xie T, Li J, et al. A novel targeted long-read sequencing-based preimplantation genetic testing method for α-thalassemia (tlrPGT-α-thal). *J Assist Reprod Genet.* 2025;42:2565-74. DOI PubMed PMC
122. Tsuiko O, El Ayeb Y, Jatsenko T, et al. Preclinical workup using long-read amplicon sequencing provides families with de novo pathogenic variants access to universal preimplantation genetic testing. *Hum Reprod.* 2023;38:511-9. DOI

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.



© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.