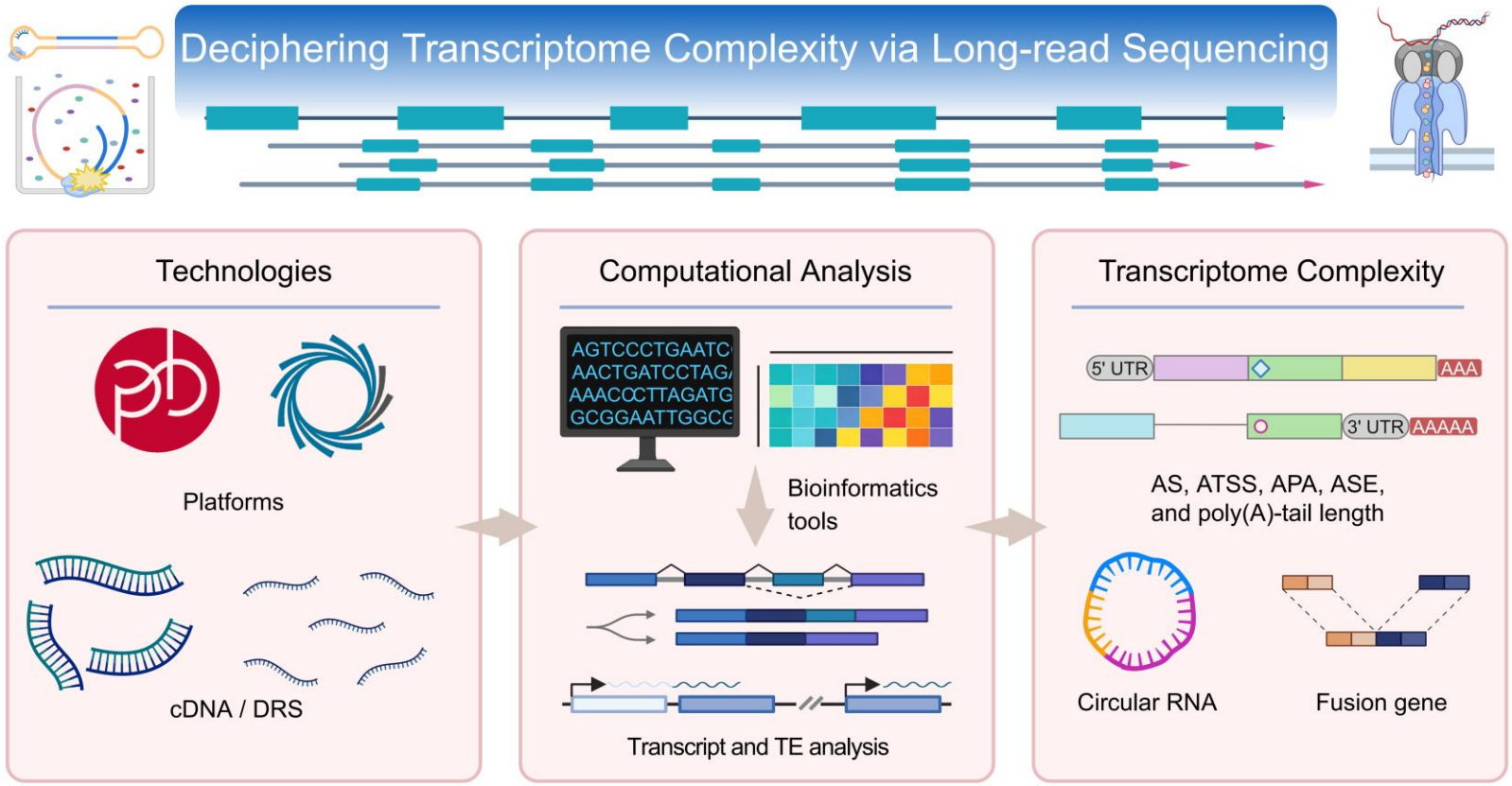
Deciphering Transcriptome Complexity via Long-read Sequencing

Chuwen Xu[1#], Jia Li[1#], Chenxi Yin[2#], Keying Li[2#], Tian Wang[2], Shiwen Gao[1], Yue Yu[1], Cheng Chang[3], Ye Wang[4], Guoqing Tong[2], Liang Gong[5],Wen Hu[6], Guoliang Chai[7,8,9], Yilai Han[8], Qian Qin[10], Suoqin Jin[1], Xiufen Zou[1], Bo Li[11], Tianyuan Zhang[12], Ana Conesa[13], Yunhao Wang[2*], Dingjie Wang[1,14*]

1 School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
2 Department of Reproductive Medicine, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi 710061, China
3 State Key Laboratory of Medical Proteomics, National Center for Protein Sciences (Beijing), Research Unit of Proteomics-driven Cancer Precision Medicine (Chinese Academy of Medical Sciences),Beijing 102206, China
4 Sichuan Key Laboratory of Conservation Biology on Endangered Wildlife, Chengdu Research Base of Giant Panda Breeding, Chengdu, Sichuan, China
5 Department of Urology of The First Affiliated Hospital & Liangzhu Laboratory, Zhejiang University School of Medicine, Hangzhou 310058, China
6 Brain Cognition and Brain Disease Institute, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
7 Chinese Institutes for Medical Research (CIMR), Beijing, China
8 Department of Neurology, Xuanwu Hospital Capital Medical University, National Center for Neurological Disorders, Beijing, China
9 Beijing Municipal Geriatric Medical Research Center, Beijing, China
10 School of Artificial Intelligence, Wuhan University, Wuhan 430072, China
11 Stanley and Judith Frankel Institute for Heart and Brain Health, University of Michigan Medical Center, Ann Arbor, MI, USA
12 Benagen Institute, Wuhan 430073, China
13 Institute for Integrative Systems Biology, Spanish National Research Council, Paterna, Valencia, Spain
14 Institute for Math & AI, Wuhan University, Wuhan 430072, China

Chuwen Xu, Jia Li, Chenxi Yin, Keying Li, Tian Wang, Shiwen Gao, Yue Yu, *et al.* 2026. Deciphering Transcriptome Complexity via Long-read Sequencing. ***iMeta*** 5: e70171. https://doi.org/10.1002/imt2.70171

Overview

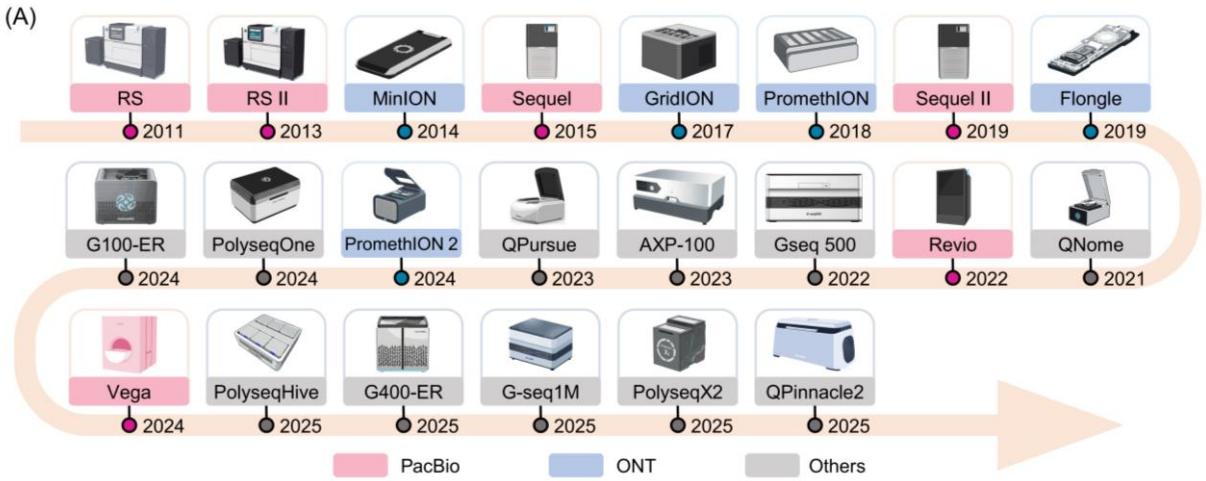


- Long-read sequencing enables isoform-resolved transcriptomics by directly capturing full-length transcript structures and linked molecular features.
- This review provides an end-to-end overview of sequencing platforms, library preparation, computational workflows, transcriptome complexity, and downstream analyses.
- We further discuss current challenges and emerging frontiers, including single-cell, spatial, RNA modification, and multi-omics applications.

Highlight

- Long-read sequencing enables isoform-resolved characterization of transcriptome complexity by directly resolving full-length transcript structures and linked regulatory features.
- The choice of sequencing platform and library preparation strategies critically influences transcript completeness, accuracy, and biological interpretation.
- Standardized computational workflows are needed for reproducible and scalable long-read transcriptomics studies.
- Single-cell, spatial, direct RNA sequencing, and multi-omics approaches expand long-read omics frontiers.

A panoramic overview of long-read transcriptomics workflow



(A) Developmental timeline of LRS platforms from PacBio (pink), ONT (blue), and others (grey).

(B) Overview of end-to-end workflow for LRS-based transcriptomics studies.

(C) Frontiers of long-read transcriptomics toward single-cell, spatial, direct RNA, and multi-omics integration applications.

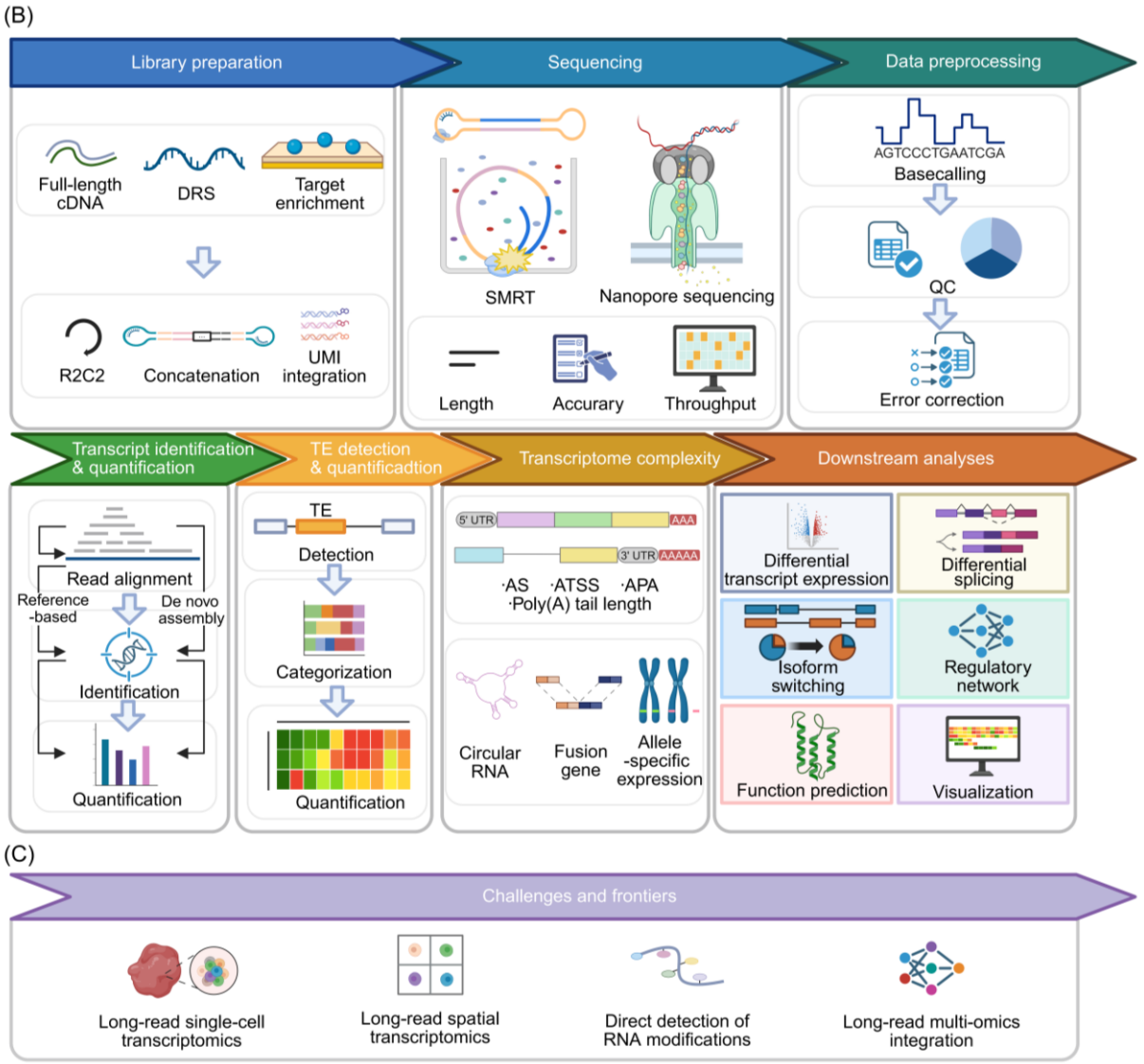
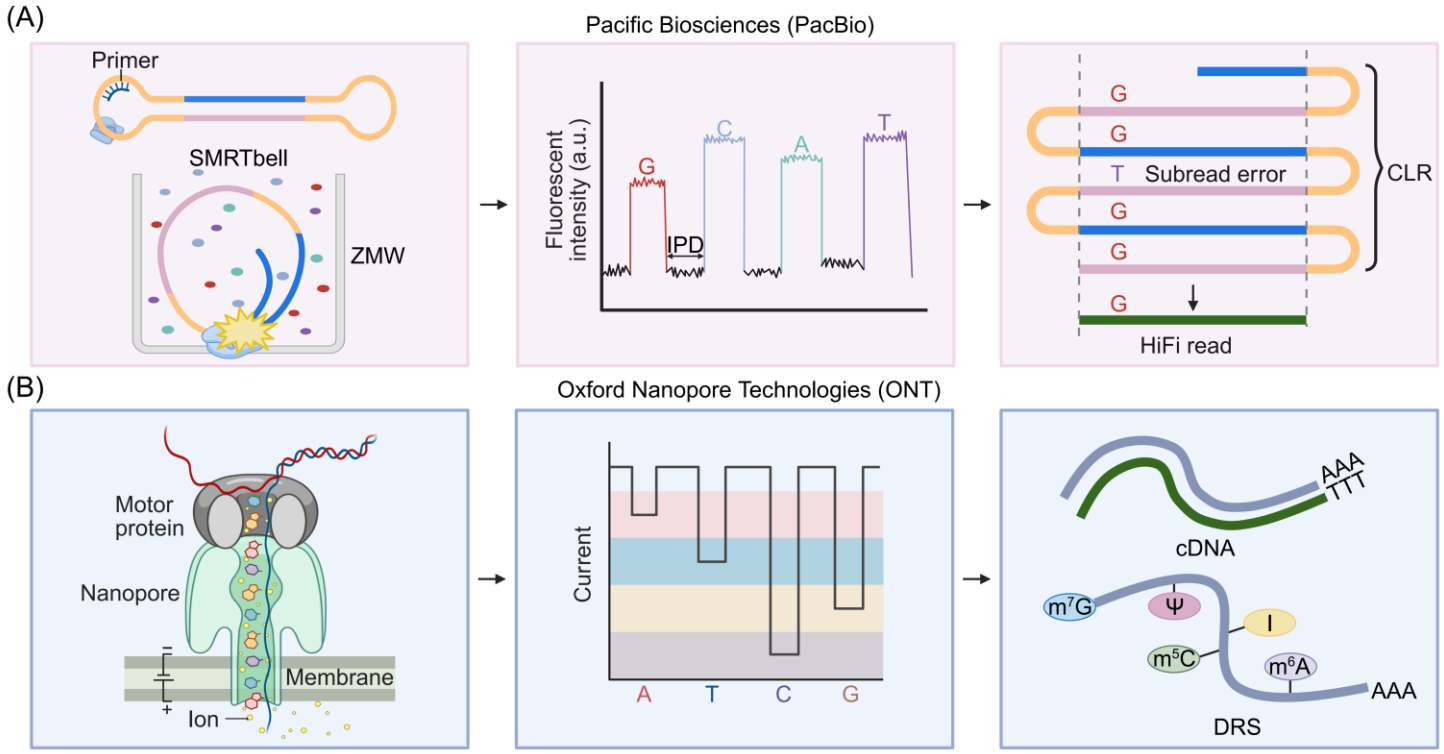


Figure 1. A panoramic overview of long-read transcriptomics workflow

Overview of LRS platforms and principles for transcriptomics studies

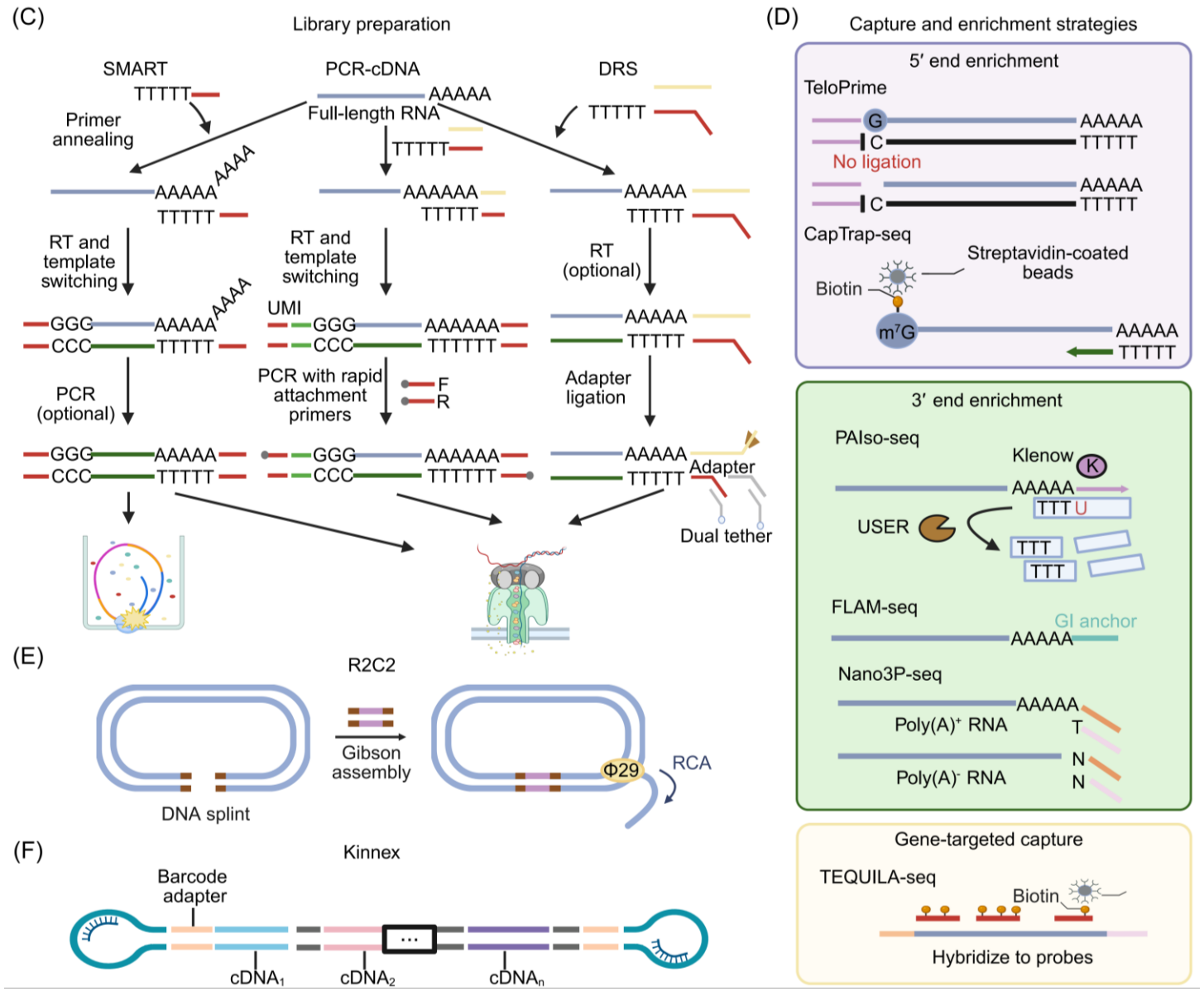


(A) For PacBio sequencing, a SMRTbell circular template is loaded into a ZMW and sequenced by polymerase; fluorescence pulses and IPD signals are recorded; and repeated passes produce multiple subreads to yield a consensus HiFi read.

(B) For ONT sequencing, a motor protein controls strand translocation through a nanopore; the pore generates an ionic-current "squiggle" determined by local sequence context; and the squiggle is basecalled into sequence, while RNA base modifications could be retained in the raw signal when DRS is performed.

Figure 2. Overview of LRS platforms and principles, and library preparation and enrichment strategies for transcriptomics studies

Overview of LRS library preparation and enrichment strategies



(C) Three major workflows for long-read RNA-seq library preparation: SMART, ONT's PCR-cDNA and DRS.

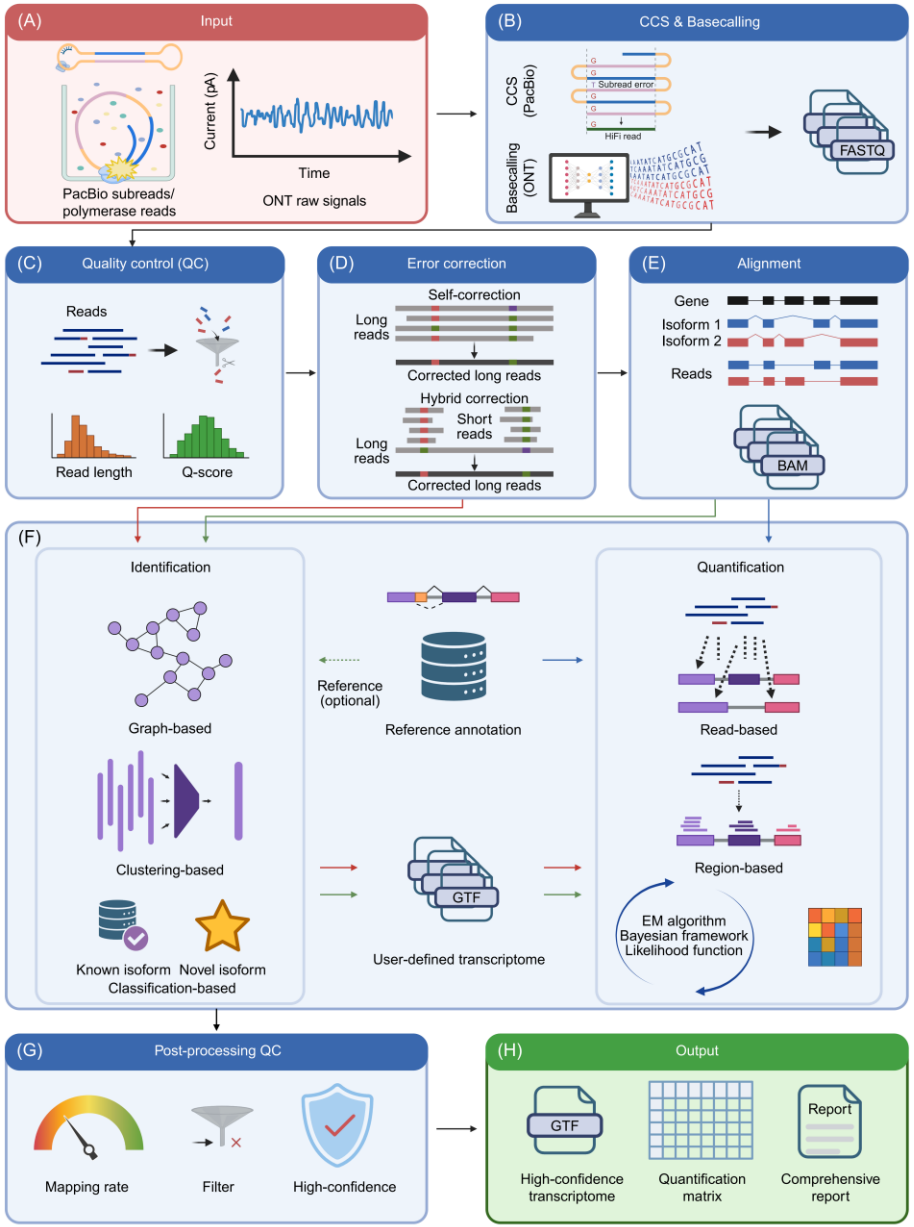
(D) Representative capture and enrichment strategies: 5′-end enrichment (e.g., CapTrap-seq), 3′-end enrichment approaches (e.g., PAIso-seq), and gene-targeted capture (TEQUILA-seq).

(E) R2C2: rolling-circle amplification creates repeated units that are sequenced and assembled into a higher-accuracy consensus sequence.

(F) Kinnex: concatenation of multiple cDNA fragments into a single long template increases sequencing throughput.

Figure 2. Overview of LRS platforms and principles, and library preparation and enrichment strategies for transcriptomics studies

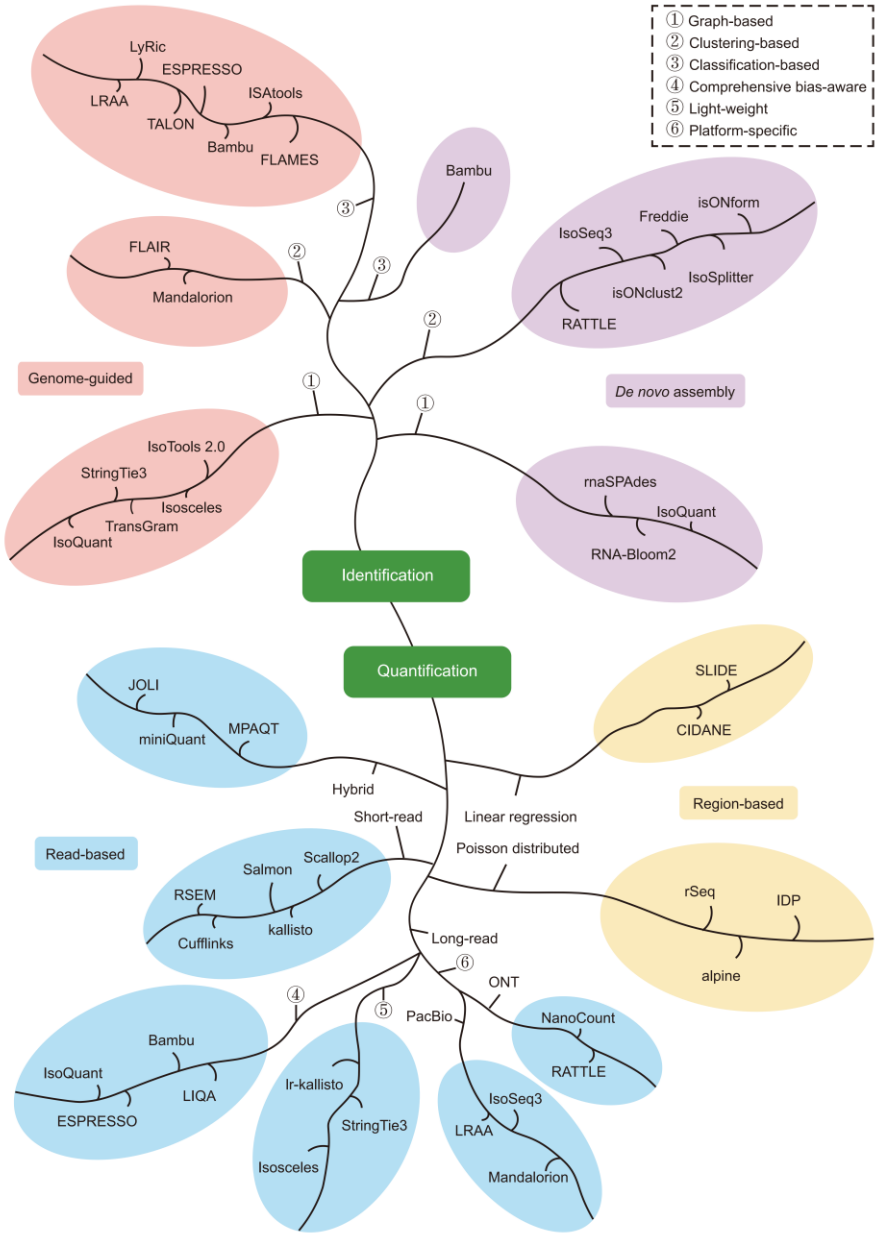
Overview of LRS platforms and principles for transcriptomics studies



- PacBio and ONT data undergo platform-specific preprocessing, followed by quality control and optional error correction.
- High-quality reads are aligned to a reference genome or directly used for transcript reconstruction.
- Transcript identification and quantification can follow reference-free, reference genome-guided, or annotation-driven strategies, producing transcript annotations and expression estimates.
- Post-analysis QC filters low-confidence results, yielding a high-confidence transcriptome, isoform quantification matrix, and analysis report.

Figure 3. Computational workflow for long-read transcriptome data analysis

Computational approaches for transcript identification and quantification

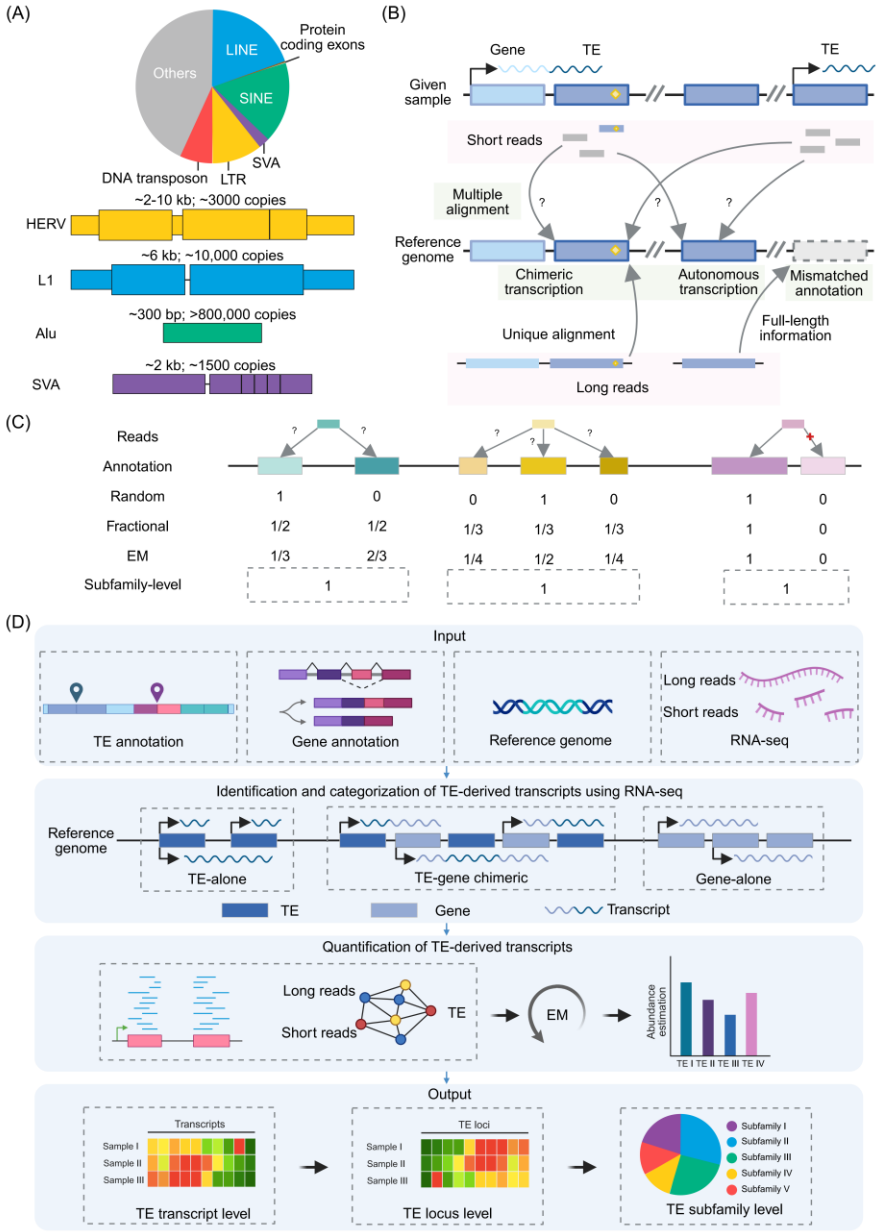


- **Isoform identification:** Methods are grouped into genome-guided and de novo approaches, with three major algorithmic paradigms: (1) graph-based, (2) clustering-based, and (3) classification-based frameworks.

- **Isoform quantification:** Methods are divided into read-based and region-based models across short-read, hybrid, and long-read data. Long-read methods are further classified as (4) comprehensive bias-aware, (5) lightweight, and (6) platform-specific approaches. Statistical assumptions, such as Poisson models and linear regression, are indicated where applicable. Representative tools are shown within each methodological class.

Figure 4. Methodological phylogeny of computational tools for transcript identification and quantification

Characterization of transposable element-derive transcripts



(A) Classification of TEs with representative structural features of major TE families.

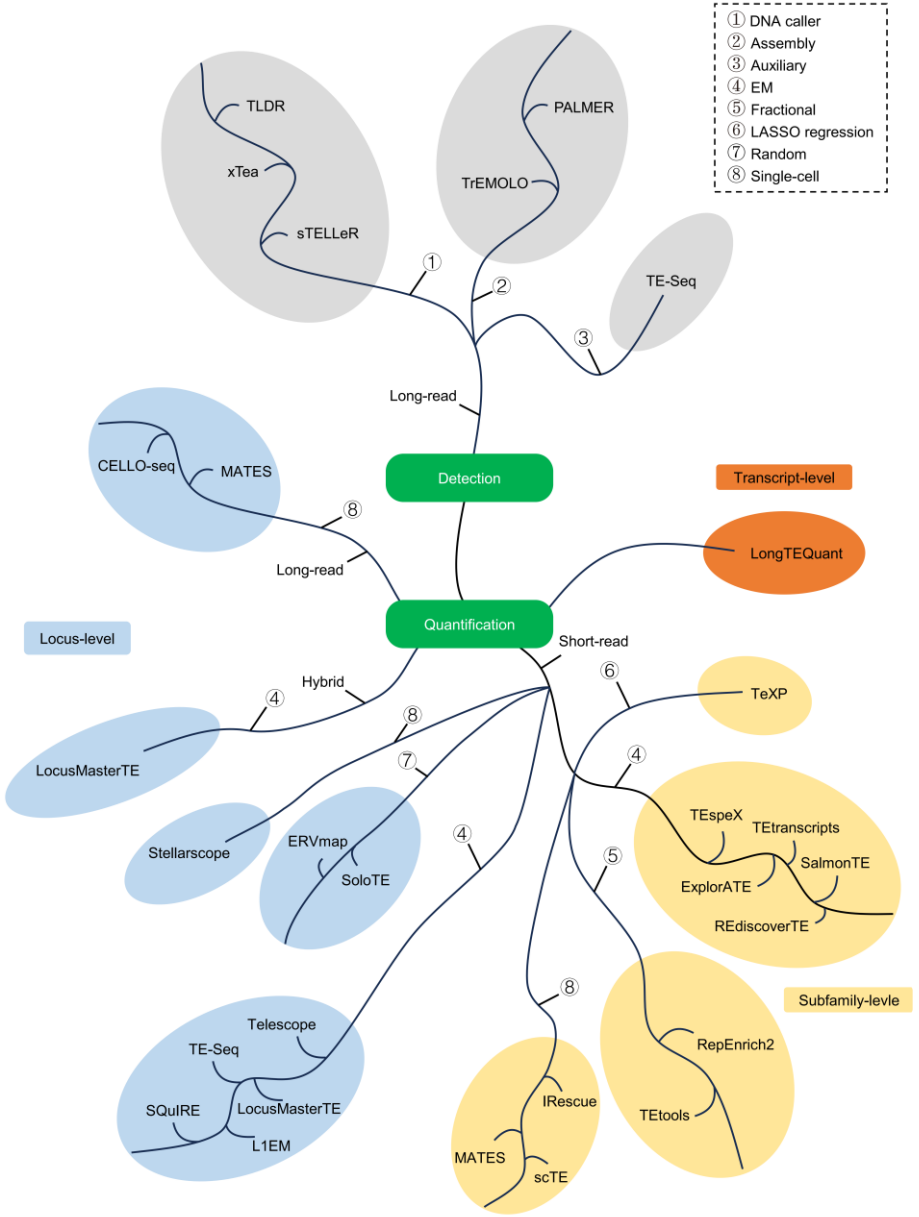
(B) Long reads improve TE transcript analysis by spanning repetitive regions, enabling unique alignment and better discrimination of TE-derived and TE-gene chimeric transcripts compared with short reads.

(C) Traditional TE quantification relies on RepeatMasker annotations and assigns multi-mapping reads using unique, fractional, or EM-based probabilistic strategies.

(D) LongTEQuant enables transcript-level TE quantification by integrating TE/gene annotations, genome references, and RNA-seq data. It identifies TE-alone, TE-gene chimeric, and gene-alone transcripts, and reports TE expression at transcript, locus, and subfamily levels.

Figure 5. Characterization of transposable element-derive transcripts

Computational approaches for TE-derived transcript quantification

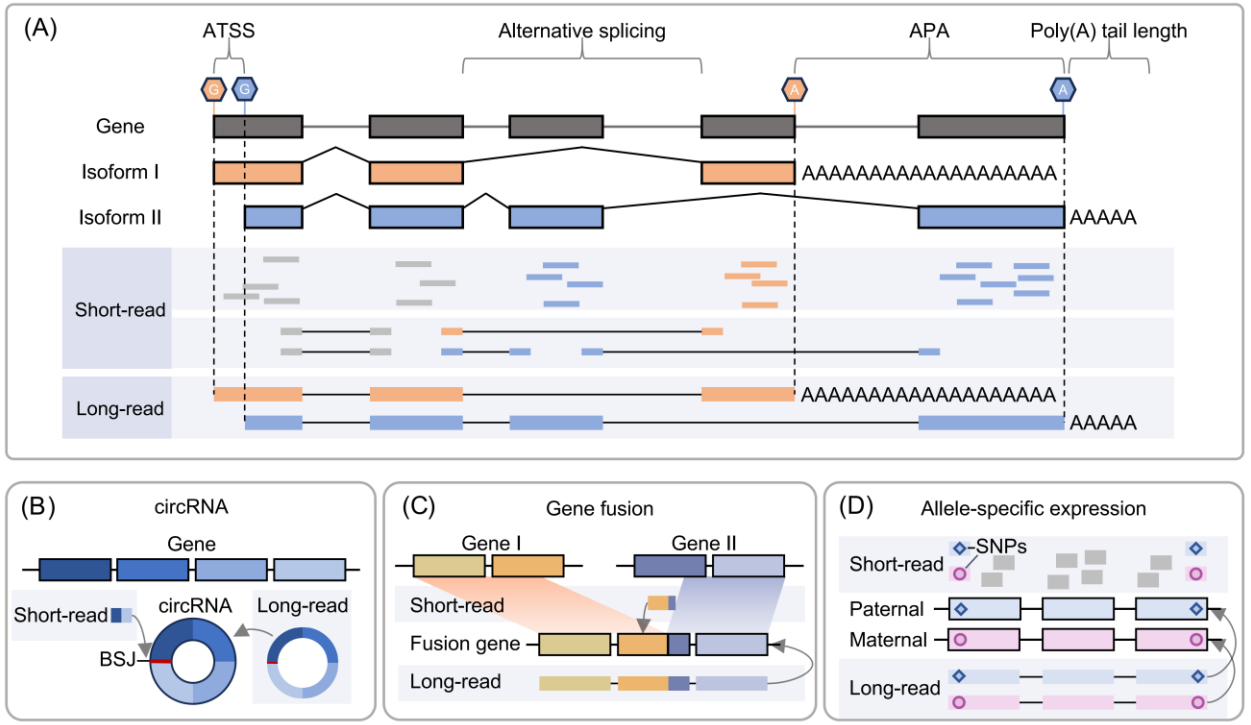


- **TE detection:** Long-read-based methods are grouped into (1) DNA callers, (2) assembly-based methods, and (3) auxiliary approaches.

- **TE quantification:** Methods are categorized into short-read, long-read, and hybrid approaches, with short-read methods further divided by subfamily- and locus-level resolution. Core strategies include (4) expectation-maximization (EM), (5) fractional assignment, (6) LASSO regression, (7) random assignment, and (8) single-cell-oriented approaches. Transcript-level frameworks directly quantify TE-derived transcripts, while long-read and hybrid methods leverage long-range information to improve mapping specificity and locus-level quantification.

Figure 6. Overview of LRS platforms and principles, and library preparation and enrichment strategies for transcriptomics studies

Long-read advantages for complex transcription



(A) Long reads resolve full-length isoforms, transcript ends, complex alternative splicing events, and poly(A) tail lengths, overcoming the ambiguous mapping limitations of short reads.

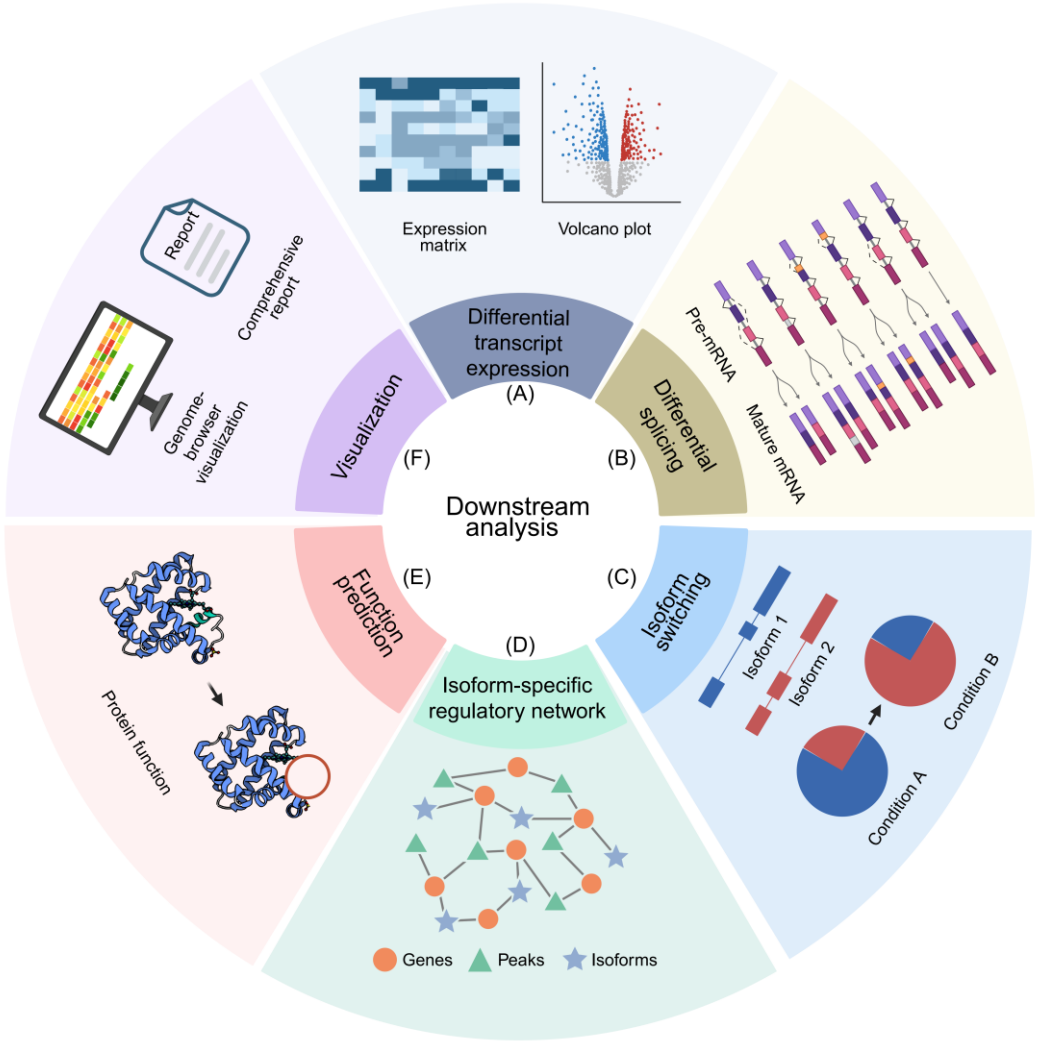
(B) Long reads provide improved resolution of full-length circular RNA structures compared with short reads.

(C) Long reads directly capture full-length fusion transcripts, whereas short reads typically provide only partial evidence.

(D) Long reads enable reliable parental allele assignment by linking multiple SNPs across transcripts, while short reads are limited by sparse informative SNP coverage.

Figure 7. Advantages of long reads over short reads in resolving complex transcription events

Downstream analyses for transcriptomics studies enabled by LRS



(A) Differential transcript expression analysis generates an expression matrix and identifies significantly changed isoforms between conditions.

(B) Differential splicing analysis detects condition-associated alternative splicing events with a full-length picture.

(C) Isoform switching analysis quantifies changes in the relative usage of isoforms from the same gene across conditions.

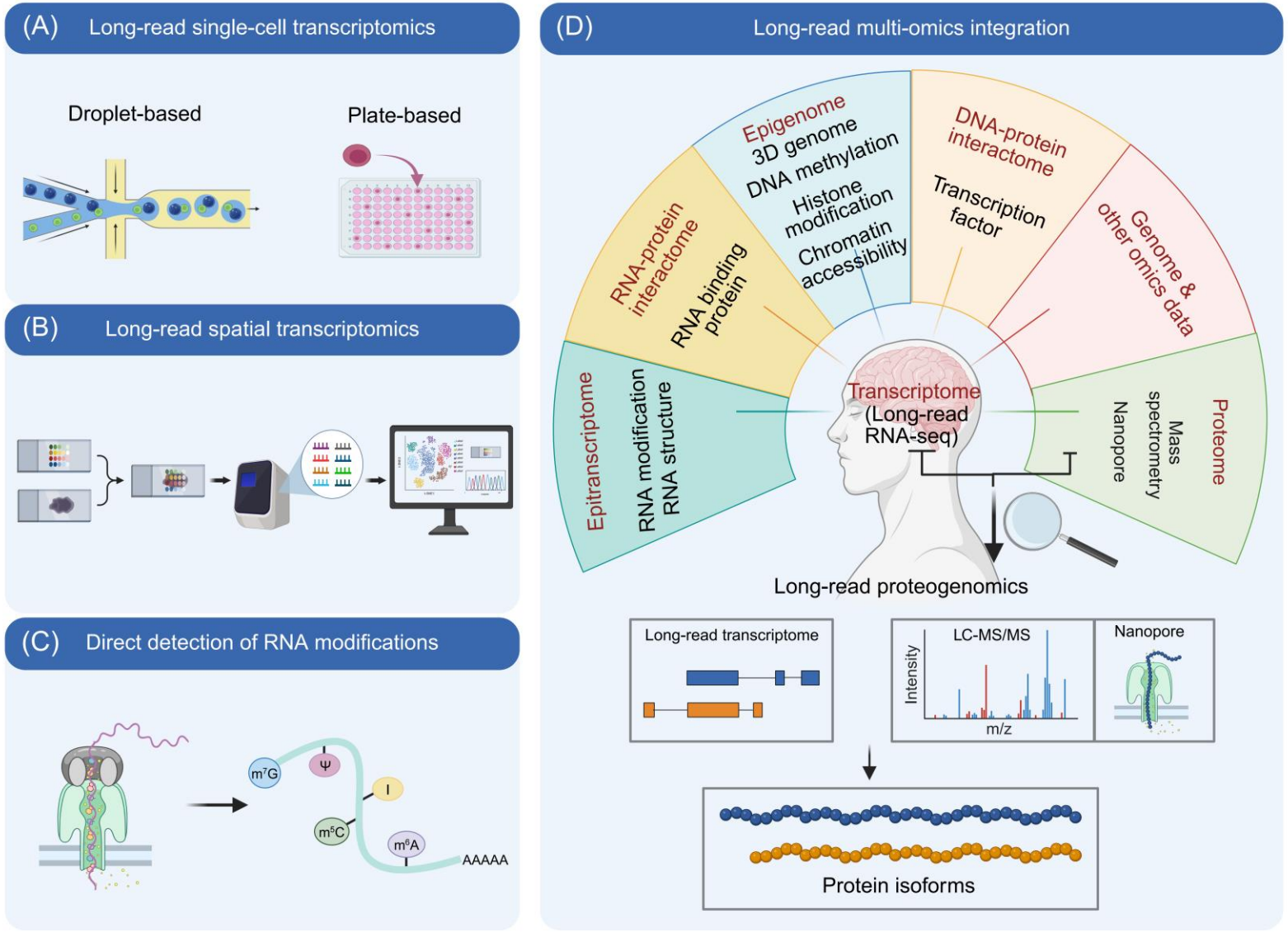
(D) Isoform-specific regulatory network inference links transcript isoforms to upstream regulators and downstream targets to reveal isoform-level regulatory programs.

(E) Functional prediction annotates isoforms and evaluates potential impacts on protein domains and biological functions.

(F) Visualization and reporting integrate results through genome browsers and comprehensive reports to facilitate interpretation and dissemination.

Figure 8. Downstream analyses for transcriptomics studies enabled by LRS

Emerging frontiers for long-read transcriptomics studies



(A) Long-read single-cell RNA-seq integrates LRS with single-cell resolution.

(B) Long-read spatial RNA-seq integrates spatial context with full-length transcript structures.

(C) Direct detection of RNA modifications by DRS.

(D) Long-read multi-omics integration.

Figure 9. Emerging frontiers for long-read transcriptomics studies

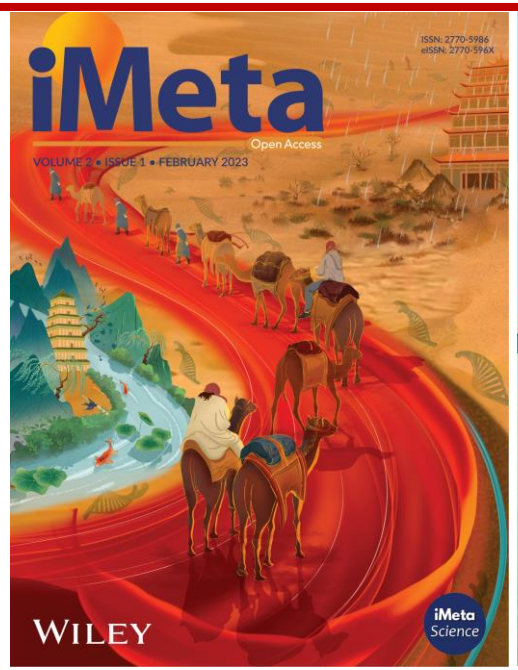
Summary

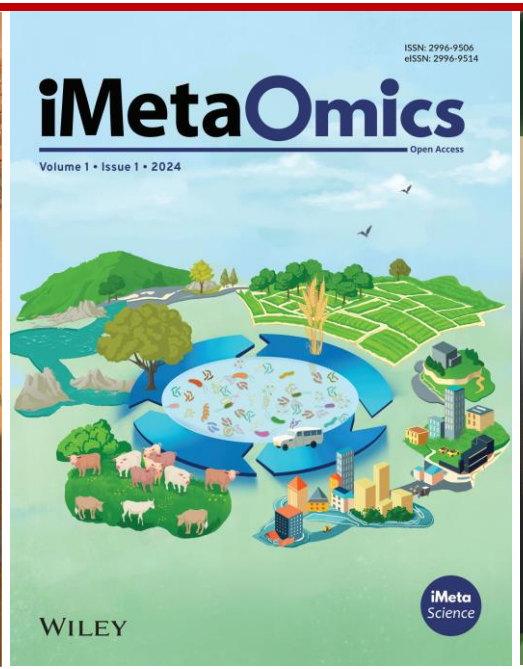
- Long-read sequencing is reshaping transcriptomics toward isoform-resolved characterization by directly resolving transcript structures, including exon connectivity, transcript boundaries, and complex events.
- Long-read transcriptomics enables deeper investigation of transcriptome complexity, isoform regulation, and multi-layer molecular interactions.
- Reliable LRS applications require optimized experimental design, suitable sequencing and library strategies, robust computational workflows, and comprehensive benchmarking. In practice, LRS should be evaluated from three perspectives: experimental suitability, analytical confidence, and biological validation.
- GitHub repository: https://github.com/jia10046/LRS_Doc

Chuwen Xu, Jia Li, Chenxi Yin, Keying Li, Tian Wang, Shiwen Gao, Yue Yu, *et al.* 2026. Deciphering Transcriptome Complexity via Long-read Sequencing. ***iMeta*** 5: e70171. https://doi.org/10.1002/imt2.70171

iMeta: Biology and Medicine in Info and Meta Era

WILEY





"***iMeta***" launched in 2022 by iMeta Science Society, impact factor (IF) **44.4**, ranking top 47/22643 in the world. It aims to publish innovative and high-quality papers with broad and diverse audiences. Its scope is similar to *Cell*, *Nature* and *Science*. Its unique features include video abstract, bilingual publication, and social media with 800,000 followers. Indexed by **SCIE/ESI**, **PubMed**, **Google Scholar** etc.

"***iMetaOmics***" launched in 2024, indexed by **ESCI**, **PubMed**, **Google Scholar**, with a **target IF>15, and its scope is similar to** *Nature Communications*, *Science Advances*, *Advanced Science*, *Nucleic Acids Research*, etc.

"***iMetaMed***" launched in 2025, with a target IF>15, similar to *Med, Cell Reports Medicine, eBioMedicine, eClinicalMedicine* etc.

Society: http://www.imeta.science
Publisher: https://wileyonlinelibrary.com/journal/imeta
Submission: iMeta: https://wiley.atyponrex.com/journal/IMT2
iMetaOmics: https://wiley.atyponrex.com/journal/IMO2
iMetaMed: https://wiley.atyponrex.com/journal/IMM3

iMetaScience
iMetaScience

office@imeta.science
imetaomics@imeta.science

Promotion Video

Update 2026/6/17