Profiling the Epigenome Using Long-Read Sequencing

Journal Club Presentation

Liu T, Conesa A. Nature Genetics. 2025;57:27-41

Journal Club Discussion
October 3, 2025

October 3, 2025

Genomics of Gene Expression Lab

Outline

Review Scope and Objectives

What This Review Covers

- Long-read sequencing (LRS) technologies for epigenomics research
- Experimental and computational strategies for characterizing chromatin states
- Advantages of LRS over short-read sequencing (SRS) methods
- Integration of epigenomic and transcriptomic data for multi-omics studies

Key Technologies

Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio)

Long-Read Sequencing: Key Advantages

Technical Advantages

- Direct DNA methylation detection
- Single-molecule resolution
- No PCR amplification needed
- Long reads (>10 kb)

Research Applications

- Haplotype-resolved analysis
- Highly repetitive regions (HRRs)
- Multi-epigenetic events on same molecule
- Complex genomic interactions

Overview of LRS Epigenomic Strategies

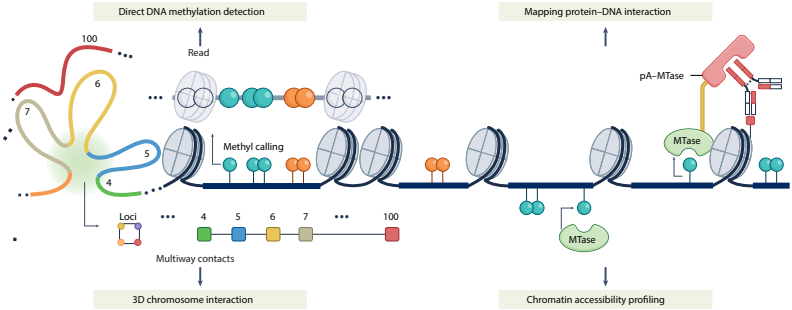

Figure 1 | Single-molecule LRS strategies for epigenomic profiling and chromatin interactions.

Four Main LRS Strategies

1. **Direct DNA methylation detection**: Sequencing native DNA molecules without bisulfite conversion to detect 5mC, 6mA, and other modifications
2. **3D chromosome interaction**: Coupling ligation of long-range interacting regions with LRS to resolve complex 3D topologies and multiway contacts
3. **Chromatin accessibility profiling**: Treatment with methyltransferases that mark open regions, followed by LRS to analyze accessibility and nucleosome positioning
4. **Protein-DNA interaction mapping**: Fusion of methyltransferases to antibodies for targeted modifications at specific chromatin sites to reveal histone marks and transcription factor binding

Traditional SRS Methods for DNA Methylation

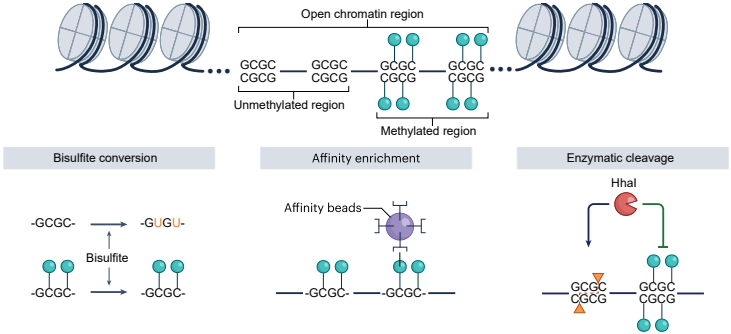

Figure 2a | SRS methods: bisulfite conversion, affinity enrichment, and enzymatic cleavage.

Limitations of Short-Read Methylation Methods

Key Limitations

- DNA degradation from bisulfite conversion
- Bias toward CpG-rich regions (affinity enrichment)
- Limited methylome coverage (restriction enzyme methods)
- Cannot distinguish modification types (5mC vs 5hmC vs 6mA)
- Poor performance in extreme GC content regions
- Difficulty mapping highly repetitive regions
- No allele-specific methylation (ASM) detection

LRS Direct Methylation Detection Workflow

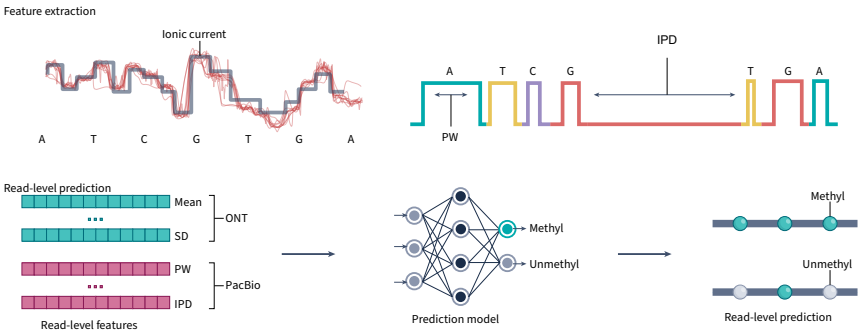

Figure 2b | LRS methylation detection: feature extraction, read-level prediction, and site-level calling.

Three-Step Methylation Calling Process

Genomics of Gene Expression Lab

1. **Feature extraction**
   - ONT: Electrical current signals through nanopore
   - PacBio: Interpulse distance and pulse width
2. **Read-level prediction**
   - Statistical tests or deep learning models
   - Tools: Nanopolish, Dorado/Remora, Fibertools, Primrose
3. **Site-level prediction**
   - Direct count or model-based approach

Key Methylation Calling Tools

Platform	Tool	Modification Types
ONT	Nanopolish	5mC
ONT	Dorado + Remora	5mC, 5hmC, 6mA, 4mC (combinations)
PacBio	Fibertools	6mA
PacBio	Primrose	5mC

Performance

Recent comparisons show both ONT and PacBio offer high-quality CpG methylation detection, with strong correlation to bisulfite sequencing

Traditional vs LRS Chromatin Accessibility Methods

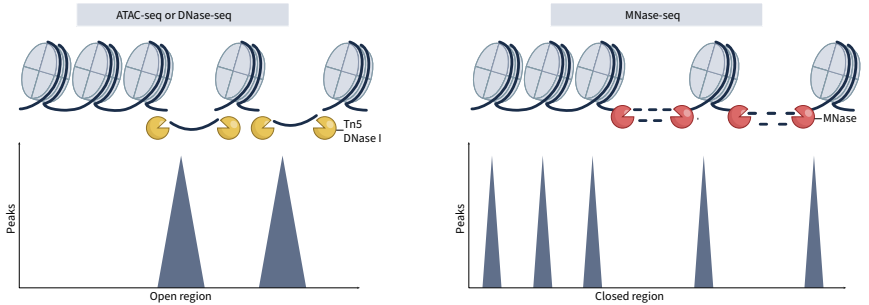

Figure 3a | Cleavage-based methods: ATAC-seq, DNase-seq, and MNase-seq.

Short-Read Chromatin Accessibility Limitations

Genomics of Gene Expression Lab

Cleavage-Based Methods (ATAC-seq, DNase-seq, MNase-seq)

- No coordination or co-occurrence of distal accessibility events
- Poor performance in segmental duplications and HRRs

Methyltransferase-Based Methods (NOMe-seq)

- Require bisulfite treatment with lower resolution
- Cannot distinguish endogenous from exogenous methylation

LRS Methyltransferase-Based Accessibility Assays

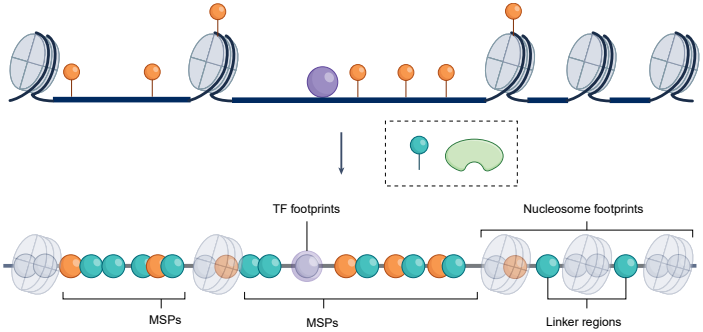

Figure 3b | DNA methyltransferase treatment followed by direct LRS methylation detection.

LRS Chromatin Accessibility Protocols

Method	Enzyme	Platform	Year
Fiber-seq	Hia5 (6mA)	PacBio, ONT	2020
SAMOSA	EcoGII (6mA) + MNase	PacBio	2020
nanoNOMe	CviPI (GpC)	ONT	2020
SMAC-seq	CviPI + SssI + EcoGII	ONT	2020
STAM-seq	EcoGII (6mA)	ONT	2023
SAM-seq	EcoGII (6mA)	ONT	2024

Key Advantage of 6mA-Based Methods

Higher resolution (shorter adenine distances) and no endogenous 6mA in most eukaryotes

Single-Molecule Chromatin Accessibility Analysis

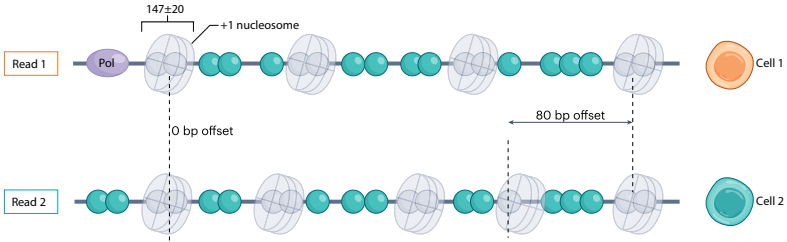

Figure 4 | Data analysis workflow for LRS chromatin accessibility profiling.

Analytical Workflow for Accessibility Data

1. **Nucleosome footprints**: Detect 147 bp inaccessible regions
2. **Methylase-sensitive patches (MSPs)**: Identify accessible regions between nucleosomes
3. **Machine learning classification**: FIRE tool distinguishes open chromatin from linker DNA
4. **Protein footprints**: FiberHMM identifies RNA Pol/TF occupancy
5. **Co-actuation analysis**: How one region's accessibility influences adjacent regions

Traditional SRS Protein-DNA Mapping Methods

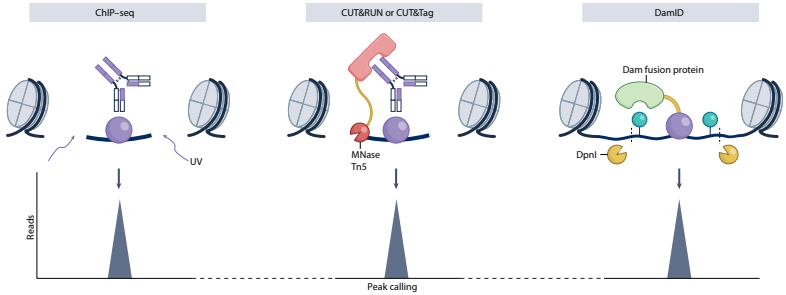

Figure 5a | SRS methods: ChIP-seq, CUT&RUN, CUT&Tag, and DamID.

Limitations of SRS Protein-DNA Methods

Major Limitations

- Cannot measure multiple protein-binding events on same DNA molecule
- No combinatorial binding pattern analysis
- Cannot phase haplotype-specific interactions
- Poor mapping in highly repetitive regions
- DNA methylation information lost during PCR amplification
- Limited to pairwise interaction detection

LRS Antibody-Targeted Methylation Methods

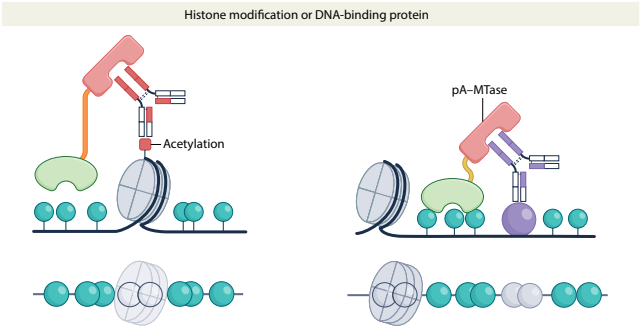

Figure 5b | LRS protein-DNA interaction mapping: DiMeLo-seq, nanoHiMe-seq, BIND&MODIFY.

LRS Protein-DNA Interaction Workflow

Protocol Steps

1. Permeabilize nuclei; add antibody and magnetic beads
2. Apply pA-methyltransferase fusion protein
3. Activate with SAM; methylate adenines near target sites
4. Extract DNA and sequence with ONT/PacBio
5. Detect methylation as binding markers

Signal Characteristics

Methylation signal decays exponentially (170 bp half-life)

Quantitative Advantages: DiMeLo-seq Example

Genomics of Gene Expression Lab

CENP-A Density in Centromeres

- >20-fold coverage in repetitive regions
- 26% ± 5% nucleosomes contain CENP-A
- ChIP-seq estimate: 4% (1 in 25)

Key Advantage

Absolute frequency measurements without PCR bias

Single-Molecule Histone Modification Analysis

Genomics of Gene Expression Lab

BIND&MODIFY Approach

Categorize into heavy (>75%), medium (25–75%), light (<25%) states

Key Findings

- Heavy/medium H3K27me3 = lower expression
- H3K27me3 + CpG methylation on same fibers

LRS-Based 3D Methods

Method	Platform	Capability
C-walk	PacBio	First LRS 3D genome
MC-4C	PacBio, ONT	Targeted loci
MC-3C	PacBio	All-vs-all
Pore-C	ONT	Multiway contacts
HiPore-C	ONT	Enhanced resolution

Major Advantage

Multiway chromatin interactions in single molecules

Pore-C Multiway Interaction Analysis

Genomics of Gene Expression Lab

Integration Potential

Transcription clusters: Colocalized loci coordinating expression

Current Limitations

- Tools not designed for single-molecule
- 3D loops are dynamic and short-lived

LRS in Transcriptomics

ONT

- >100M reads/flow cell
- Direct RNA-seq (m6A)
- Real-time acquisition

PacBio

- >99% HiFi accuracy
- 100M reads/run
- Isoform resolution

Emerging Applications

Nascent-seq, Ribo-STAMP, single-cell and spatial RNA-seq

Toward Multi-Omics Applications

Examples

- Hi-C + LRS + RNA-seq
- Synchronized genome, methylome, epigenome, transcriptome
- scNanoCOOL-seq: Single-cell profiling

Key Challenge

Cannot map modalities to same molecules

Analytical Challenges for Multi-Omics

Experimental Needs

Higher signal-to-noise, coverage, reduced costs

Computational Needs

- Benchmarks and signal extraction
- Co-occurrence detection
- Integration frameworks

LRS Achievements

- Direct detection of 5mC, 6mA, 4mC, 5hmC
- Single-molecule chromatin states
- Haplotype-resolved analysis
- Mapping repetitive regions
- Multiway chromatin interactions
- Quantitative protein-DNA without PCR bias

Gaps and Challenges

Genomics of Gene Expression Lab

Technical

- High costs, limited coverage
- Model retraining with updates
- Signal decay, limited standardization

Analytical

- TF motif degeneracy
- Lack of single-molecule 3D tools
- Cannot link DNA to RNA

Future Directions

Genomics of Gene Expression Lab

1. **Technology**: Universal models, cost reduction, enrichment
2. **Analytics**: Integration frameworks, co-occurrence detection, TF assignment
3. **Biology**: Link chromatin to transcription, combinatorial regulation, disease

Take-Home Messages

Key Points

1. Direct detection without conversion/amplification
2. Single-molecule, haplotype-resolved
3. Overcomes SRS limitations
4. Multi-omics profiling
5. Linking chromatin to transcription

Discussion Questions

Genomics of Gene Expression Lab

1. How might single-molecule data change our understanding of gene regulation?
2. What computational challenges need addressing?
3. Which biological questions could benefit from LRS?
4. How to balance cost vs coverage?
5. How to link DNA to RNA molecules?

Thank you for your attention!

Questions and Discussion?

References I