nature genetics

Review article

https://doi.org/10.1038/s41588-024-02038-5

Profiling the epigenome using long-read sequencing

Received: 3 July 2024

Accepted: 19 November 2024

Published online: 8 January 2025

Check for updates

Tianyuan Liu & Ana Conesa

The advent of single-molecule, long-read sequencing (LRS) technologies by Oxford Nanopore Technologies and Pacific Biosciences has revolutionized genomics, transcriptomics and, more recently, epigenomics research. These technologies offer distinct advantages, including the direct detection of methylated DNA and simultaneous assessment of DNA sequences spanning multiple kilobases along with their modifications at the single-molecule level. This has enabled the development of new assays for analyzing chromatin states and made it possible to integrate data for DNA methylation, chromatin accessibility, transcription factor binding and histone modifications, thereby facilitating comprehensive epigenomic profiling. Owing to recent advancements, alternative, nascent and translating transcripts can be detected using LRS approaches. This Review discusses LRS-based experimental and computational strategies for characterizing chromatin states and highlights their advantages over short-read sequencing methods. Furthermore, we demonstrate how various long-read methods can be integrated to design multi-omics studies to investigate the relationship between chromatin states and transcriptional dynamics.

Recent advancements in single-molecule, long-read sequencing (LRS) technologies, developed by companies such as Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio), have substantially transformed genomics research[1]. These technologies have a crucial role in the complete sequencing of the human genome from telomere to telomere[2] and are leading ambitious projects aimed at sequencing all species on Earth[3].

Beyond genomics, LRS offers several key advantages over traditional short-read sequencing (SRS) for epigenomics, primarily owing to its ability to generate substantially longer reads of single native DNA molecules. First, LRS can directly sequence native DNA molecules, enabling the detection of DNA methylation without additional library preparation steps. The incorporation of methyltransferases further allows the analysis of methylation footprints, thereby enabling comprehensive epigenetic profiling. Second, because LRS operates at a single-molecule resolution without the need for amplification, it can uncover molecular heterogeneity that is often overlooked by bulk short-read approaches. This PCR-free nature thus provides a more quantitative and accurate measurement of epigenetic modifications. Third, the long-read lengths inherent to LRS enable the phasing of haplotype-resolved data, enabling the distinction between maternal and paternal alleles by spanning entire haplotype blocks[4]. This long-read length capability is also valuable for accurately mapping epigenetic interactions across challenging genomic regions, such as highly repetitive regions (HRRs). Collectively, these advantages allow LRS to capture multiple epigenetic events along the same long chromatin fiber within a single-molecule read, which facilitates co-association analyses that are crucial for understanding complex genomic interactions.

New assays that use these technologies have been developed to analyze various chromatin features, including accessibility, protein–DNA interactions and 3D genome organization (Table 1). They have set the stage for comprehensive single-molecule epigenomic profiling, using long reads (Fig. 1). Simultaneously, the application of LRS technologies is rapidly expanding into investigations of the complexities of transcriptomes and the dynamics of RNA metabolism by allowing the profiling of alternative isoforms, epitranscriptomic modifications, nascent transcription and translating RNA molecules (reviewed in ref. 5), creating a new era of DNA and RNA research that is led by LRS.

In this Review, we examined the emerging applications of LRS in epigenomics research, focusing on the following three main areas:

Institute for Integrative Systems Biology, Spanish National Research Council, Paterna, Spain. e-mail: ana.conesa@csic.es

Table 1 | Overview of LRS assays for chromatin analysis

	Main enzymes/fusion proteins	Platform	Main category	Specificity	Year
Fiber-seq[55]	Hia5 (EcoGII, EcoGI, Btr192IV and Hin1523; 6mA)	PacBio, ONT	Chromatin accessibility	Available for both PacBio and ONT, supported by bioinformatics tools such as Fibertools, FIRE and FiberHMM	2020
SAMOSA[121]	EcoGII (6mA), Mnase	PacBio	Chromatin accessibility	Capture positions of nucleosomes through nucleolytic cleavage and nearby protein–DNA interactions through methylation	2020
SAMOSA-tag[122]	EcoGII (6mA), Tn5	PacBio	Chromatin accessibility	Low-input protocol using Tn5 tagmentation	2024
MeSMLR-seq[56]	CviPI (GpC)	ONT	Chromatin accessibility	The first publication on LRS chromatin accessibility.	2019
ODM-seq[59]	SssI (CpG), CviPI (GpC)	ONT	Chromatin accessibility	Protocol available for both ONT and Illumina (SRS)	2019
nanoNOMe[57]	CviPI (GpC)	ONT	Chromatin accessibility	New Nanopolish model for methylation calling	2020
SMAC-seq[54]	CviPI (GpC), SssI (CpG), EcoGII (6mA)	ONT	Chromatin accessibility	Simultaneous use of several methyltransferases	2020
Targeted LRS NOMe-seq[123]	CviPI (GpC)	ONT	Chromatin accessibility	CRISPR–Cas9-based targeted enrichment	2022
STAM-seq[58]	EcoGII (6mA)	ONT	Chromatin accessibility	Uses adaptive sampling for enrichment	2023
SAM-seq[64]	EcoGII (6mA)	ONT	Chromatin accessibility	Specifically tailored to plant genomes	2024
DiMeLo-seq[79]	pA–Hia5	PacBio, ONT	Protein–DNA interaction	First publication for LRS direct protein–DNA interaction	2022
nanoHiMe-seq[89]	pA–Hia5 with secondary antibody	ONT	Protein–DNA interaction	Extension of DiMeLo-seq with secondary antibody for enhanced specificity and methylation signal detection	2023
BIND&MODIFY[88]	pA–M.EcoGII	ONT	Protein–DNA interaction	Multi-omics analysis of epigenomics and transcriptomics	2023
LRS C-walk[92]	DpnII, T4 DNA ligase	PacBio	3D genome organization	First publication for LRS 3D genome	2016
MC-4C[93,94]	DpnII, HindIII, T4 DNA ligase, Cas9	PacBio, ONT	3D genome organization	Targeted loci profiling	2018, 2020
MC-3C[95]	DpnII, T4 DNA ligase	PacBio	3D genome organization	All-versus-all genome interaction mapping	2020
Pore-C[96]	DpnII, HindIII, NlaIII, T4 DNA ligase	ONT	3D genome organization	All-versus-all genome interaction mapping	2022
HiPore-C[97]	NlaIII, DpnII, HindIII, proteinase K Pronase, T4 DNA ligase	ONT	3D genome organization	Enhanced Pore-C protocol with reduced logging for higher resolution	2023

(1) experimental and computational approaches using LRS to characterize chromatin states—specifically DNA methylation patterns, chromatin accessibility and protein–DNA interactions—addressing traditional limitations of SRS, (2) the new technical and data analysis challenges introduced by these LRS methods, and (3) the integration of epigenomic and transcriptomic LRS data into multi-omics models, offering insights into gene regulation mechanisms previously inaccessible through SRS.

Detection of DNA methylation

Indirect DNA methylation detection in SRS

DNA methylation involves adding a methyl group to nucleotides within the DNA molecule[6]. Endogenously, this typically occurs at the fifth carbon of the cytosine ring, primarily at CpG dinucleotides in vertebrates[7]. Additionally, exogenous methylation includes modifications, such as N6-methyladenine (6mA) and GpC methylation, which are used in chromatin accessibility protocols[8]. These modifications are catalyzed by DNA methyltransferases, with S-adenosylmethionine (SAM) serving as the methyl donor.

DNA methylation has been traditionally studied using SRS, enabled by specialized library preparation methods, such as bisulfite conversion[9,10], affinity enrichment[11,12] and endonuclease cleavage[13,14], designed for DNA methylation profiling[15,16] (see ref. 17; Fig. 2a). However, these approaches have limitations. Rather than directly detecting DNA methylation through sequencing signals, these methods rely on alternative genomic DNA pretreatment, which may introduce various biases. For instance, bisulfite conversion can lead to DNA degradation and incomplete conversion rates[18]. Ultrafast BS-seq (UBS-seq)[19], along with bisulfite-free methods such as NEBNext Enzymatic Methyl-seq (EM-seq)[20] and TET-assisted pyridine borane

https://doi.org/10.1038/s41588-024-02038-5

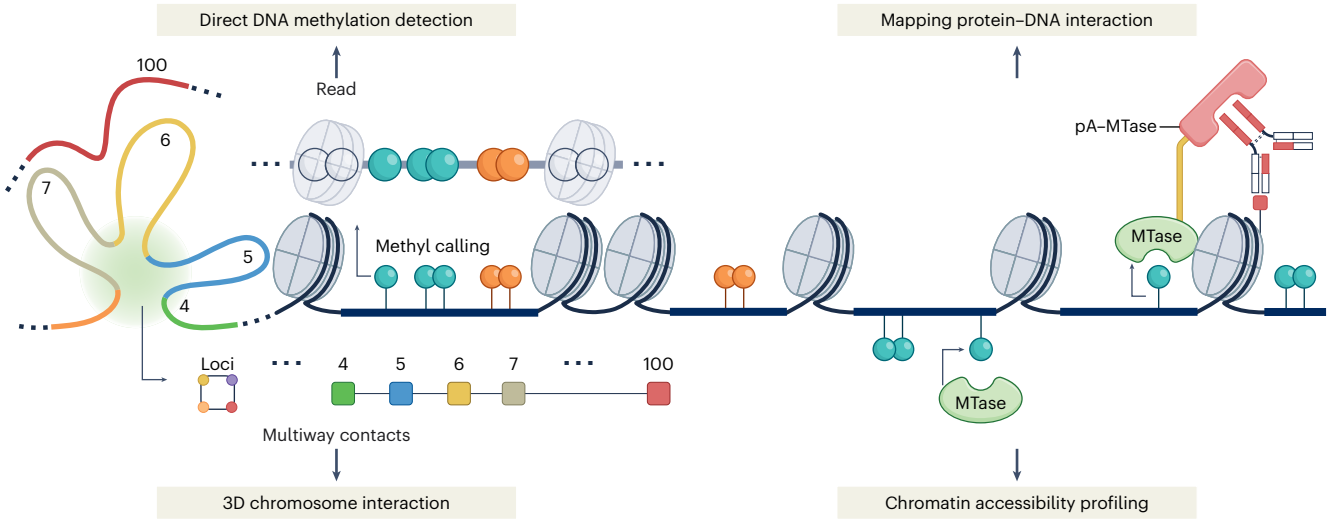


Fig. 1 | Simplified representation of single-molecule LRS strategies for epigenomic profiling and chromatin interactions. LRS technologies, such as ONT and PacBio, leverage four main strategies for epigenomic profiling. Their ability to directly detect diverse DNA methylation modifications on the reads enables mapping methylation patterns across large genomic regions. Coupling ligation of long-range interacting chromosome regions with LRS allows to resolve complex 3D chromosome topologies, capturing multiway contacts among sets of genomic loci. Treatment with methyltransferases that introduce exogenous methylation marks at open regions followed by LRS enables analysis of chromatin accessibility and nucleosome positioning. Fusion of methyltransferases to antibodies allows targeted modifications at specific chromatin sites. Sequencing these sites using long-read technology reveals the position of protein–DNA interactions, such as histone marks.

sequencing (TAPS)[21], minimize DNA damage through optimized protocols. However, UBS-seq still involves harsh chemicals, while EM-seq and TAPS may introduce conversion inefficiencies and batch variability. Affinity-enrichment-based methods are biased toward CpG-rich regions as they capture molecules that selectively bind to fragments with high CpG content[22]. Restriction-enzyme-based methods suffer from limited methylome coverage due to their reliance on specific restriction sites[23].

Moreover, these methods can only target specific modification types, for instance, 5-methylcytosine (5mC), and therefore fail to capture the full diversity and complexity of DNA modifications, such as 5-hydroxymethylcytosine (5hmC), 6mA or N4-methylcytosine (4mC). Lastly, all SRS-based approaches have inherent limitations, particularly in accurately sequencing regions with extreme GC content, mapping DNA methylation HRRs and identifying allele-specific methylation (ASM)[24–26]. These shortcomings can be addressed with LRS techniques, which can directly detect DNA methylation[17].

Direct DNA methylation detection using LRS

Both ONT and PacBio LRS technologies enable the simultaneous detection of DNA methylation during the sequencing process, thereby eliminating the need for separate library preparation steps[27–29]. The identification of methylation sites through LRS involves a systematic process comprising the following three main steps: feature extraction, read-level prediction and site-level prediction (Fig. 2b).

Feature extraction is the initial step, in which sequencing signals are analyzed for modifications introduced by methylated nucleotides. In ONT sequencing, methylation modifies the electrical current signals (known as 'squiggles') as DNA bases pass through the nanopore. These squiggle patterns are analyzed for variations that indicate methylation. In PacBio sequencing, feature extraction involves analyzing changes in polymerase activity during DNA replication by examining the interpulse distance—the time between sequencing two bases—and, to a lesser extent, pulse width[27,30], which reflects the duration of base sequencing. These parameters vary when polymerase pauses at methylated bases and indicate nucleotide methylation[27,30,31] (Fig. 2b).

Following feature extraction, read-level prediction uses computational models to interpret the extracted features and determine the presence of methylation within individual sequencing reads. This prediction is made through the following two primary approaches: (1) using statistical tests that compare the sequencing patterns to an in silico reference or to patterns derived from a nonmodified control sample[32,33] or (2) using pretrained supervised or semi-supervised models[34–36], including deep-learning and other machine-learning models[30,36–42], to accurately predict read-level methylation.

For ONT, pioneering tools such as Nanopolish[37] have been instrumental in detecting 5mC and remain widely trusted. Nanopolish was notably used in a recent deCODE genetics study for methylation analysis[43]. This study identified allele-specific methylation quantitative trait loci (ASM-QTLs) as key drivers of gene expression variability, with sequence variants influencing CpG methylation patterns in *cis*-regulatory regions. This underscores the significance of ASM-QTLs in noncoding genome regulation and their enrichment in hematological traits. Subsequent tools, such as Guppy and Megalodon, which evolved from Taiyaki (github.com/nanoporetech/taiyaki), integrated base calling and methylation detection, although early models exhibited high false-negative rates[34]. As the field progressed, separating base calling from methylation detection proved to be a more effective approach, as demonstrated by Dorado (github.com/nanoporetech/Dorado). Dorado uses Remora (github.com/nanoporetech/Remora), a methylation predictor based on a convolutional neural network architecture, to train methylation-specific models following base calling. This separation has substantially improved both the accuracy and flexibility of methylation analysis.

In PacBio sequencing, historically, read-level methylation calling involved extracting features from individual subreads to predict the modification, with these predictions then combined to generate a final circular consensus sequence (CCS) read prediction[30,31]. Modern PacBio sequencers, such as the Revio (pacb.com/revio/) and Sequel IIe systems (pacb.com/wp-content/uploads/Sequel_II_and_IIe_Data_Files.pdf), no longer provide subreads by default and the latest methylation calling methods such as Fibertools (6mA)[36] and Primrose (5mC; github.com/mattoslmp/primrose), predict methylation directly from CCS read summary features.

The final step is to determine the methylation status of each genomic site, for which the following two basic approaches exist:

direct count and model-based (Fig. 2b). In the count-based approach, the methylation status is determined by analyzing the frequency of methylation calls in the reads mapped to a particular site[30,34,35,42]. This involves comparing the number of reads identified as methylated at a genomic site with the total number of reads for that site. The model-based approach uses a more complex strategy by analyzing neighboring methylation sites[30,34,35] (github.com/PacificBiosciences/pb-CpG-tools). This approach integrates the read-level methylation probabilities of a target site with its adjacent sites, using the assumption of correlation between them to provide more confident methylation calls. Pretrained models then interpret these patterns across a sequence of sites, resulting in more refined predictions of methylation status for each site[30].

Both ONT and PacBio platforms have expanded their capabilities beyond detecting 5mC to include a variety of DNA modifications, such as 6mA and 4mC. PacBio sequencing enables the detection of these additional modifications in microbial genomes using SMRT Link microbial genome analysis tools (PacBio, pacb.com/wp-content/uploads/application-brief-measuring-dna-methylation-with-5-base-hifi-sequencing.pdf). Similarly, ONT's Dorado software supports the detection of multiple modifications, including combinations, such as 5mC with 5hmC and 4mC with 5mC, with ongoing developments aimed at addressing less common epigenetic modifications.

Although a universal model for methylation detection across species and methylation types (for example, 5mC and 6mA) would be preferable, it is likely that continuous updates to sequencing technologies, such as improvements in PacBio polymerases and ONT pores and motor proteins, will require ongoing retraining of methylation models. These updates alter the kinetic signals necessary for distinguishing methylation patterns, as seen with changes in both ONT[44] and PacBio Sequel II polymerases[36]. Moreover, detecting methylation within specific sequence contexts, such as CpG, versus nonspecific methylation, such as 6mA, may demand distinct training approaches with inherent trade-offs between precision and recall.

While initial methylation-calling methods used enzymatically modified DNA to provide known methylated regions for training[37], the newest approaches use training datasets composed of synthetically generated oligonucleotides covering all possible *k*-mers[45]. However, these datasets still need to be thoroughly benchmarked against native genomic methylation patterns to validate their accuracy with real data.

Finally, it should be noted that the benchmarking of methylation-calling tools in LRS has primarily focused on ONT[34,35], as the PacBio sequencing 5mC identification strategy initially resulted in a low signal-to-noise ratio[46,47]. However, advances in deep-learning-based methylation callers have substantially improved PacBio's performance in detecting methylation. These developments have led to a strong correlation with the traditional reference method of bisulfite sequencing, boosting the development of DNA methylation-based methods for PacBio sequencing[31,36,48], with recent comparisons demonstrating that both ONT sequencing and PacBio sequencing offer high-quality CpG methylation detection[44].

Profiling chromatin accessibility

Genome-wide mapping of nucleosome occupancy and chromatin accessibility primarily relies on cleavage-based and methyltransferase-based assays complemented by SRS. Cleavage-based methods, such as DNase I hypersensitive site sequencing (DNase-seq), assay for transposase-accessible chromatin using sequencing (ATAC-seq) and micrococcal nuclease sequencing (MNase-seq), use various enzymes to target accessible DNA regions, which are then preferentially sequenced and detected by peak calling (Fig. 3a). Methyltransferase-based assays, meanwhile, use the capacity of methyltransferase enzyme to methylate open chromatin regions but not tightly bound DNA, enabling the mapping of accessible regions through methylation detection techniques, such as bisulfite sequencing and immunoprecipitation followed by sequencing (Fig. 3b). For example, nucleosome occupancy and methylome sequencing (NOMe-seq) uses the methyltransferase M.CviPI[8,49,50], which specifically targets cytosines within GpC dinucleotides and introduces a stable 5mC modification[51], followed by bisulfite conversion and sequencing to detect open chromatin regions (reviewed in ref. 52).

In all cases, SRS-based approaches exhibit limitations due to their short-read lengths, which prevent them from capturing the coordination of accessibility between distal sites and measuring the co-occurrence of these events. This limitation makes it challenging to reveal complex *cis*-acting chromatin states. Additionally, short-read methods rely on GpC methyltransferases because bisulfite treatment can convert cytosine modifications into amplifiable base changes. This reliance presents a disadvantage compared to methods that detect 6mA methylation on long reads, which can provide a higher resolution of chromatin accessibility. Although emerging techniques involving direct deamination of cytosines regardless of sequence context may address this issue[53], SRS-based approaches still struggle to resolve chromatin accessibility patterns at segmental duplications and HRRs.

Methyltransferase-based chromatin accessibility assays with LRS

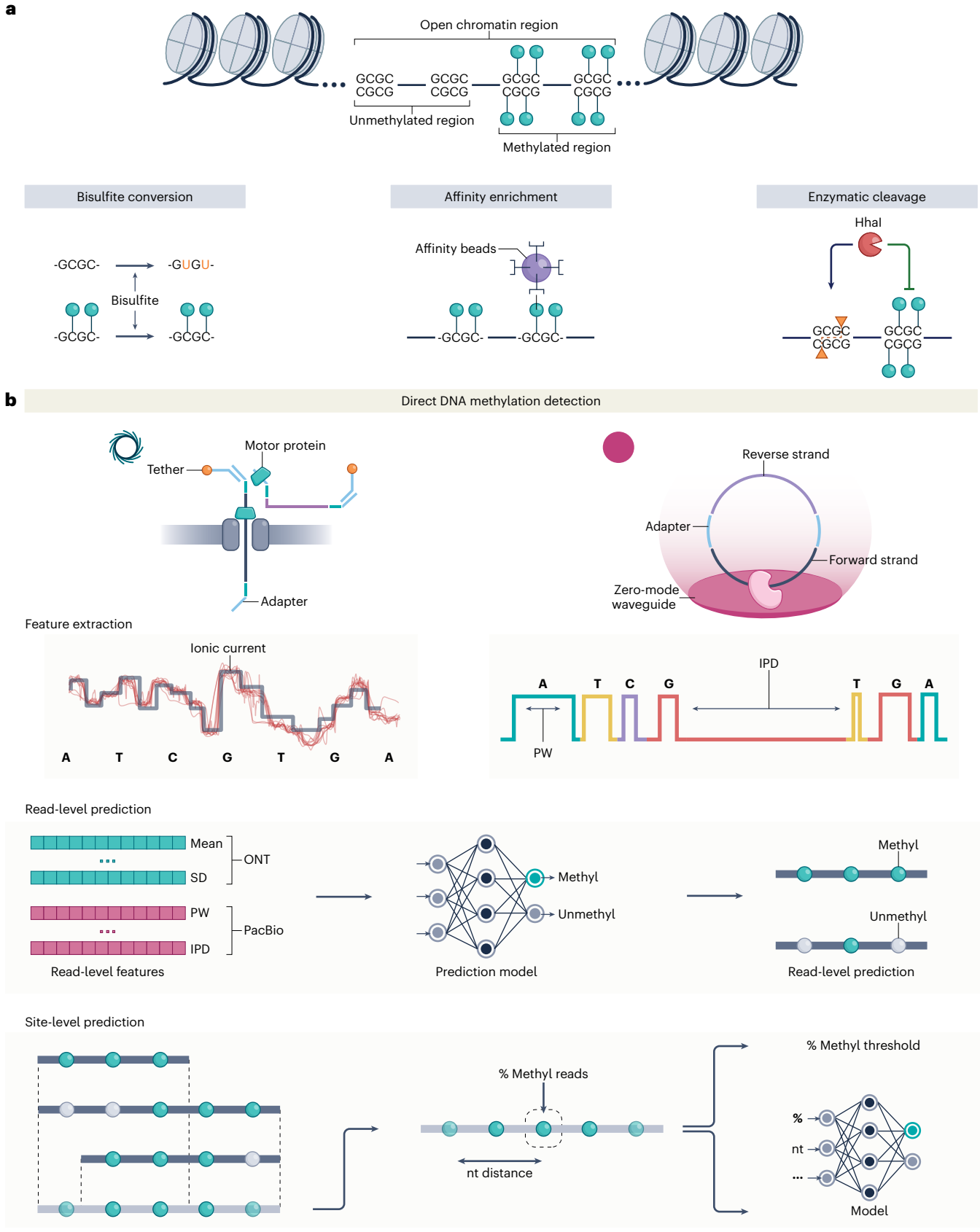
Historically, methyltransferase-based assays relied on bisulfite conversion to detect methylation marks and infer open chromatin regions. By contrast, LRS technologies enable the direct detection of methylation marks, such as 5mC and 6mA, along single DNA molecules through specialized methylation-calling processes. This direct detection eliminates the need for DNA amplification and bisulfite treatment, providing a more comprehensive and accurate assessment of chromatin accessibility across extended genomic regions[26,46] (Fig. 3b). Leveraging these advantages, various long-read-based protocols have been developed, including Fiber-seq, single-molecule adenine methylated oligonucleosome sequencing assay (SAMOSA), nanopore sequencing of nucleosome occupancy and methylome (nanoNOMe), single-molecule long-read accessible chromatin mapping sequencing assay (SMAC-seq), single-molecule targeted accessibility and methylation sequencing (STAM-seq), methyltransferase treatment followed by single-molecule long-read sequencing (MeSMLR-seq) and occupancy measurement via DNA methylation and high-throughput sequencing (ODM-seq) (Table 1). These protocols differ in the methyltransferase enzyme used, as well as the strategies for nuclei isolation and DNA

Fig. 2 | Schematic illustration of methods for detecting DNA methylation using SRS and LRS. a, SRS methods for detecting DNA methylation include bisulfite conversion, affinity enrichment and enzymatic cleavage. In bisulfite conversion, unmethylated cytosines are converted to uracil, while methylated cytosines remain unchanged, allowing their differentiation during sequencing. Affinity enrichment using methyl-binding proteins attached to beads then selectively isolates methylated DNA regions. Subsequent enzymatic cleavage with restriction enzymes (for example, HhaI) that cut at specific unmethylated sites leaves methylated regions intact. **b**, LRS methods with ONT and PacBio enable direct detection of DNA methylation by analyzing the read signals. In ONT sequencing, a motor protein controls the movement of DNA strands through a nanopore and analyzes changes in ionic current as DNA passes through the pore to identify methylated bases (top left). In PacBio sequencing, DNA polymerase synthesizes the complementary strand within a zero-mode waveguide, and methylation is detected by analyzing the interpulse duration (IPD) and pulse width (PW) (top right). These alterations in sequencing signals are used to predict methylation at the level of individual sequencing reads, known as read-level methylation detection. After mapping reads to the genome, the methylation status at each site is determined by two approaches: direct count and model based. The direct-count method calculates the proportion of methylated reads at each site, whereas the model-based approach integrates methylation probabilities of a site with neighboring sites, using assumed correlations to enhance confidence and refine methylation predictions.

https://doi.org/10.1038/s41588-024-02038-5

purification techniques, allowing researchers to select approaches best suited to their experimental requirements[54–58].

LRS methods, such as MeSMLR-seq, ODM-seq and nanoNOMe[56,57,59], use M.CviPI[8,49], the enzyme used in traditional SRS NOMe-seq. Specifically, nanoNOMe simultaneously assesses CpG DNA methylation and GpC chromatin accessibility, leveraging the native sequencing and high-throughput capabilities of the Nanopore platform to obtain PCR-free, long-range chromatin accessibility signals. Other



approaches, such as Fiber-seq, SAMOSA, SAM-seq and STAM-seq, use 6mA methyltransferase (for example, Hia5 or M.EcoGII) to mark accessible sites. Strategies based on 6mA methyltransferases typically achieve higher resolution compared with methods using CpG or GpC methyltransferases because the genomic distance between adenines is shorter than between CpG or GpCs[36,55,58].

Additionally, unlike the widespread endogenous CpG methylation, which complicates the distinction between native and experimentally introduced marks, the absence of endogenous 6mA in most eukaryotes minimizes this confounding effect, allowing clearer differentiation of exogenous methylation signals[54,60–64]. Finally, SMAC-seq integrates M.CviPI (GpC), M.SssI (CpG) and M.EcoGII (6mA) methyltransferases to achieve higher resolution and better genome coverage[54]. Spatial chromatin accessibility sequencing (SCA-seq), enhanced by proximity ligation techniques, captures both chromatin accessibility and 3D chromatin interactions using ONT at moderate coverage (36.9×)[65]. However, its reliance on the restriction enzyme DpnII inherently produces shorter and unevenly distributed chromatin fragments (~700 bp median), limiting coverage in inaccessible regions and potentially introducing biases in fragment distribution; moreover, generating 3D chromatin interactions data requires reads that scale quadratically with the number of genomic loci ($O(N^2)$), leading to substantially higher sequencing costs compared to one-dimensional assays, such as Fiber-seq and nanoNOMe[65].

Cost constraints limit LRS-based chromatin accessibility protocols to low coverage, hindering the detection of genetic alterations in a subset of cells (for example, somatic mosaic variants) and rare epigenetic states and analyzing long-range complex imprinted loci. Achieving cost-effective high coverage would enhance the resolution and reliability of epigenomic insights. However, traditional enrichment techniques, such as PCR amplification and hybridization capture, are inadequate because DNA modifications are removed during amplification (reviewed in ref. 66). Early methods used CRISPR–Cas9 ribonucleoproteins (RNPs) with guide RNA duplexes designed to introduce cuts on complementary DNA strands flanking the target regions[67,68]. However, this approach is most effective for capturing regions around 12–24 kb because larger DNA fragments are prone to shearing and degradation during extraction and handling[68], and the efficiency of ligating sequencing adapters to both ends of long fragments decreases with length, resulting in reduced coverage for larger loci[67]. Targeted Fiber-seq and targeted NOMe-seq overcome these limitations by using gentle extraction techniques that preserve high-molecular-weight DNA or by applying in-gel processing methods. These methods also involve using multiple CRISPR–Cas9 or Cas12a RNP complexes to simultaneously cleave at numerous flanking sites and optimize adapter ligation protocols to enrich DNA molecules larger than 100 kb (refs. 54,69). Alternatively, ONT's adaptive sampling offers a real-time enrichment method that selectively rejects off-target reads without requiring specialized DNA library preparations, as implemented in STAM-seq to investigate accessible, HRRs in *Arabidopsis*[58]. However, adaptive sampling requires sufficiently long DNA fragments (recommended over 5 kb) to effectively decide whether to keep or reject molecules during nanopore sequencing[70]. This limitation makes it less effective for shorter or damaged DNA, resulting in lower enrichment of such samples compared to other method.

Methods to analyze long-read chromatin accessibility data

Analyzing chromatin accessibility using LRS data leverages the single-molecule resolution to address intricate biological questions that bulk assays cannot resolve. The analysis of chromatin accessibility from LRS data typically starts with identifying methylation sites within single-molecule reads, using the base-calling algorithm compatible with the methyltransferase used in the experimental protocol to either target 6mA or 5mC[36,44]. Unlike bulk assays that average signals across many molecules, LRS allows for precisely mapping each methylation event on single DNA strands.

Following methylation identification, nucleosome footprints are detected as patterns of modified bases occurring at regular 147 bp intervals, indicative of nucleosome positioning. Once nucleosome footprints are located, these are used to identify and analyze intervening methylated regions. This single-molecule approach enables the determination of nucleosome arrangements on individual DNA molecules, providing insights into heterogeneity in chromatin states among cells, which is not possible for bulk methods (Fig. 4a). For instance, a phasing strategy has been used to align nucleosome arrays across individual reads based on the ensemble-averaged +1 nucleosome position[71]. This single-molecule alignment revealed variations in nucleosome spacing and occupancy between *Saccharomyces cerevisiae* cells, such as longer linker regions and variable nucleosome-depleted zones within gene bodies—details unattainable through bulk analyses[71].

Regions with multiple methyltransferase-induced methylation sites, typically spanning a few hundred base pairs between nucleosome footprints, are identified as methylase-sensitive patches. These MSPs are flagged as markers of accessible DNA[55,57,58] (Fig. 4b), encompassing linker DNA between histones, open chromatin regions and regions bound by transcription factors. By focusing on single-molecule data, researchers can distinguish between different accessible regions on each DNA molecule, facilitating the study of chromatin dynamics at an unprecedented resolution.

To refine analyses toward biologically relevant regions, single-molecule data allow for the filtering out of linker DNA and regions without specific transcription factor occupancy based on their precise length and methylation levels[56,72]. Statistics and machine-learning approaches can also be used to improve the accuracy of classifying single-molecule MSPs into linker DNA or open chromatin regions[57,73]. For example, the tool Fiber-seq Inferred Regulatory Elements (FIRE) identifies MSPs with features consistent with open chromatin regions using a semi-supervised machine-learning classifier that is trained using a mixed positive dataset of MSPs overlapping known DNase I hypersensitive sites and CCCTC-binding factor (CTCF) chromatin immunoprecipitation sequencing (ChIP–seq) peaks, and a negative dataset of MSPs not overlapping these sites[73] (Fig. 4b).

After refining MSPs, downstream analyses can identify protein footprints, such as RNA polymerase (Pol) or transcription factors, within accessible regions. FiberHMM uses a Hidden Markov Model (HMM) to infer states of DNA accessibility and inaccessibility, thereby pinpointing protein occupancy at single-molecule resolution[74,75] (Fig. 4b). Characterizing these footprints by their lengths or matching them to known transcription factor motifs allowed identifying Pol II and Pol III footprints spanning multikilobase (15–20 kb) regions on individual chromatin fibers in *Drosophila* S2 cells[75], revealing insights into RNA polymerase-mediated reshaping of chromatin architecture. By detecting footprints of the pre-initiation complex and promoter–proximal paused Pol II, the authors (this study) discovered that Pol II pausing leads to the destabilization and repositioning of downstream nucleosomes, facilitating transcriptional elongation by increasing chromatin accessibility. Furthermore, their single-molecule analyses revealed pervasive, distance-dependent transcriptional coupling between nearby Pol II genes, Pol III genes and transcribed enhancers. This coupling is modulated by local chromatin architecture, including topologically associating domain (TAD) boundaries and insulator proteins, which can disrupt synchronized transcriptional activity. These findings provide direct evidence of how RNA polymerases reshape chromatin architecture and coordinate transcription along individual chromatin fibers, offering new insights into the dynamic interplay between transcription machinery and chromatin structure[75].

Another interesting application made possible by LRS-based chromatin accessibility methods is co-actuation analysis. Co-actuation investigates how the accessibility of one genomic region can influence

https://doi.org/10.1038/s41588-024-02038-5

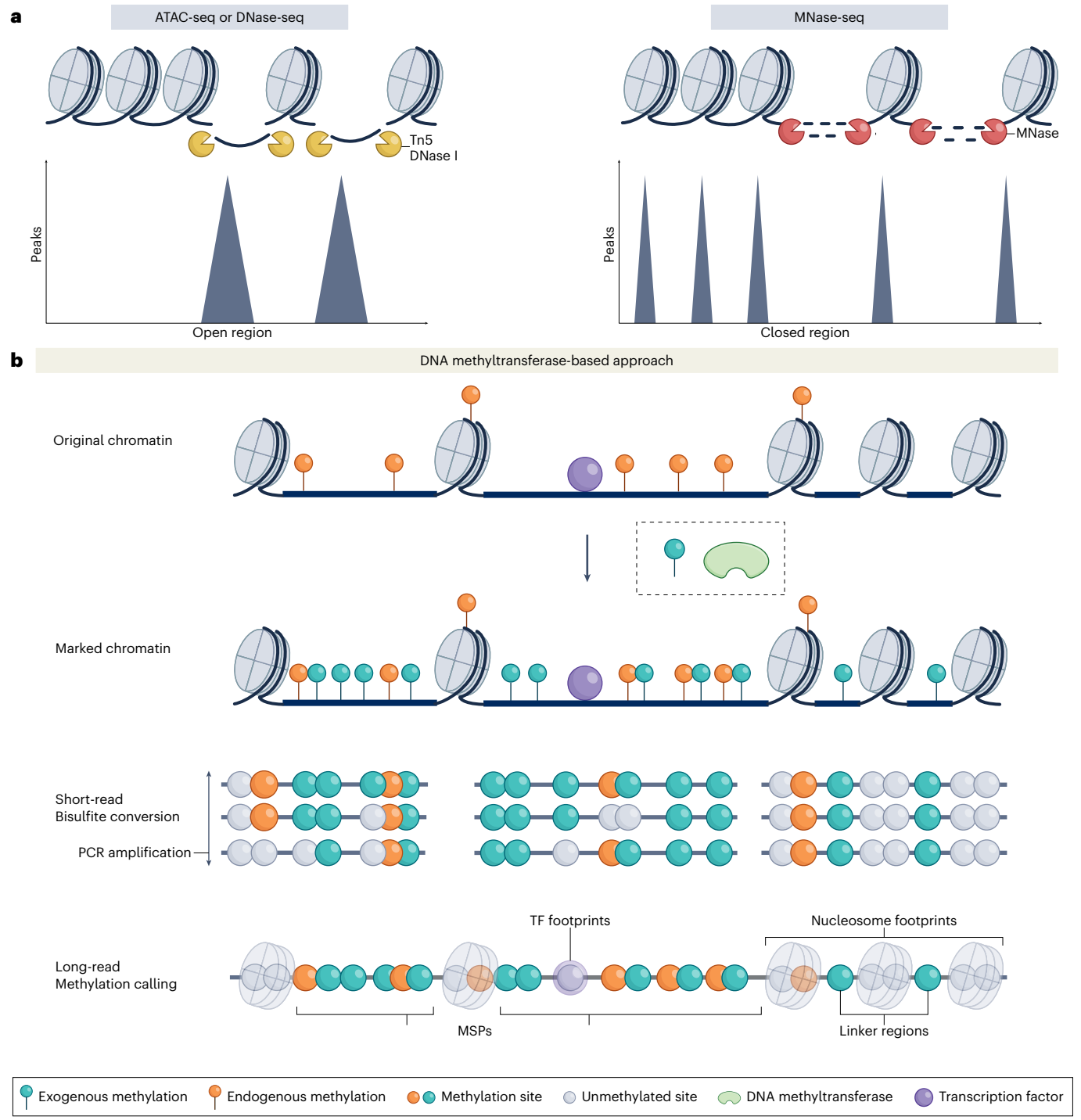


Fig. 3 | Representation of SRS and LRS methods for profiling chromatin accessibility. a, SRS methods include cleavage-based assays such as ATAC-seq or DNase-seq, which use, for instance, Tn5 transposase or DNase I to cleave accessible (open) chromatin regions, producing fragments for sequencing (left). MNase-seq uses MNase to digest linker DNA, mapping nucleosome positions and closed chromatin regions (right). **b**, In methyltransferase-based assays, such as NOMe-seq, chromatin is treated with DNA methyltransferases that methylate accessible DNA. Bisulfite sequencing is then used to detect methylation patterns, revealing open chromatin regions in SRS data. LRS methods can directly detect DNA methylation without bisulfite conversion or amplification, allowing simultaneous profiling of chromatin accessibility across long DNA molecules. MSPs, methyltransferase-sensitive patches; TF, transcription factor.

the accessibility of adjacent sections along the same chromatin, spanning several kilobases[55]. The approach involves comparing the frequency of accessibility of individual genomic regions against instances where multiple regions are accessible concurrently within chromatin fibers. By constructing a graph in which each node represents an accessible genomic region and edges depict interaction strengths, complex interactions between genomic regions can be visualized[76] (Fig. 4c). Co-actuation analysis can enhance the understanding of complex enhancer interactions, as demonstrated in the study of short tandem repeat (STR) mutations on chromosome 15q[76]. Here the STR-mutant haplotype was found to activate a thyroid-specific enhancer cluster, precisely defined by the co-actuation and spatial configuration of critical regulatory FIRE elements, which in turn substantially enhanced the expression of miR7-2 and miR-1179, two microRNAs shown to be crucial for thyroid-specific gene regulation and resistance to thyrotropin (RTSH)[76]. Finally, LRS-based chromatin accessibility data can also

be resolved for each independent allele using small structural variants, such as single-nucleotide variants, insertion–deletions, tandem repeat variants and methylation signals captured by the long reads for haplotype-phasing[73,76–78]. Building on this capability, the first quantitative genome-wide approach for haplotype-selective chromatin measurements has been developed, enabling precise mapping of chromatin accessibility across the diploid human genome at single-molecule resolution[73].

It should be noted that single-molecule approaches are inherently limited by the density of observable bases and uncertainties in assigning footprints to their cognate proteins. For instance, in Fiber-seq, observable bases are restricted to A–T base pairs, and the saturation of the methyltransferase reaction can vary across different sequence contexts, leading to footprint ambiguity. Bulk analyses address these challenges by aggregating single-molecule footprint signals, reducing the signal-to-noise ratio and combining binding data with DNA methylation information. This integrated approach provides a broader and more reliable view of transcription factor activity, as LRS methylation-calling methods can simultaneously detect endogenous m5C methylation and induced methylation marks (for example, 6mA). Consequently, a joint analysis of transcription factor binding and DNA methylation becomes possible[57,58,79] (Fig. 4d). Using this approach, CTCF-binding sites were found to be associated with lower DNA methylation and higher chromatin accessibility, which allows CTCF binding. CTCF-bound regions also show well-positioned nucleosomes, while higher methylation correlates with reduced CTCF binding and disorganized nucleosomes[57]. This suggests that DNA methylation and chromatin accessibility influence CTCF binding, which helps regulate chromatin organization and gene expression[57]. As in SRS methods, a major challenge in this approach is the degeneracy of many transcription factor motifs, which makes it difficult to identify the specific transcription factor responsible for a given footprint. Knowing the set of expressed transcription factors in the cell type under study can help narrow the analysis and improve accuracy.

Direct mapping of protein–DNA interactions

Over the past decade, the analysis of protein–DNA interactions has advanced substantially by developing various SRS-based assays tailored to this purpose. In ChIP–seq, chromatin is cross-linked, digested and probed with an antibody specific to a protein (for example, a histone or a transcription factor) or protein modification (for example, histone acetylation) to immunoprecipitate the bound DNA[80]. Subsequent SRS detects protein–DNA interactions as peaks in the genome-mapped reads[81] (Fig. 5a). Other strategies use enzyme tethering, which involves targeting the chromatin protein or modification of interest with a fusion protein and/or antibody before labeling or releasing the underlying DNA. DNA adenine methyltransferase identification (DamID) uses a fusion protein combining *Escherichia coli* DNA adenine methyltransferase (Dam) and the chromatin protein of interest to methylate the adenines near the protein's binding sites[82]. The modified sites are identified using restriction enzyme-based methylation detection and sequencing[83]. Similarly, in chromatin endogenous cleavage[84], MNase is fused to the target protein to cleave and release the underlying DNA fragments for sequencing[84]. By contrast, chromatin immunocleavage (ChIC) uses an antibody to bind to its chromatin target in fixed cells or nuclei and then adds a protein A (pA)–MNase fusion protein that tethers MNase to the chromatin-bound antibody, eliminating the need to produce different fusion proteins for each protein of interest[85]. CUT&RUN and CUT&Tag are cross-linking-free extensions of ChIC[85–87] (Fig. 5a), tethering MNase and Tn5, respectively.

Epigenetic research has progressed through SRS methods, such as bulk-level ChIP–seq and DamID, but these DNA–protein profiling techniques share limitations with SRS chromatin accessibility protocols. SRS methods cannot measure multiple protein-binding events on the same DNA molecule, limiting the ability to study combinatorial binding patterns and long-range interactions. They also cannot phase haplotype-specific protein–DNA interactions or map interactions in HRRs. Additionally, DNA methylation information is lost during PCR amplification of the fragments, hindering the simultaneous detection of protein–DNA interactions and DNA methylation. Although SRS is generally cheaper per read than LRS, the limitations in context and resolution can impact the overall cost-effectiveness and depth of insights gained.

Inspired by the short-read methods mentioned above, more recently, LRS antibody-targeted protocols have been developed to map protein–DNA interactions. Techniques such as direct methylation sequencing (DiMeLo-seq), nanohigh-resolution methylation sequencing (nanoHiMe-seq) and BIND&MODIFY use the methyltransferase activity within fusion proteins to identify antibody-targeted protein–DNA interaction sites on long, native, single DNA molecules[79,86–88] (Fig. 5b).

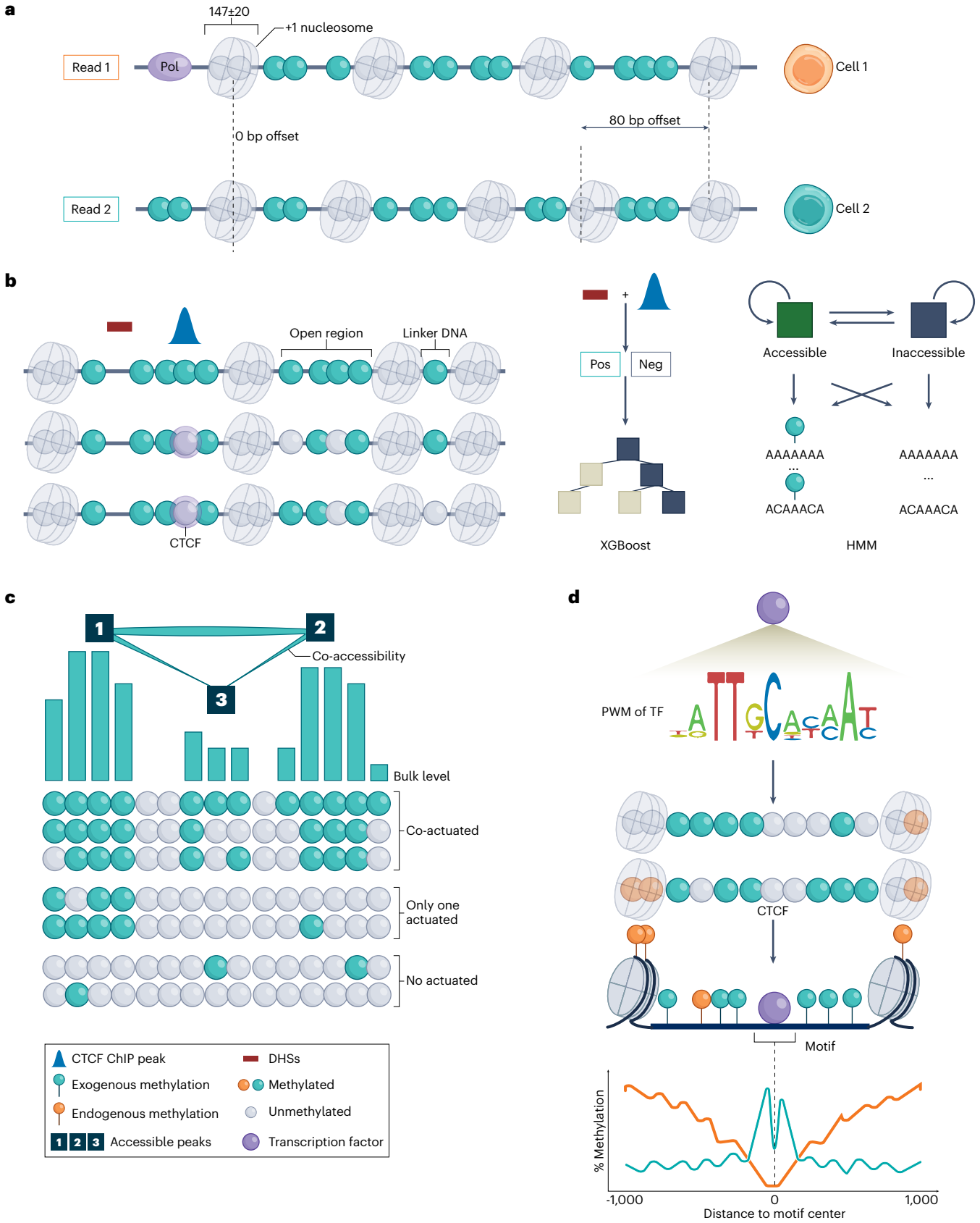
The protocol begins with nuclei permeabilization to allow the access of antibodies to target proteins or histone modifications. Magnetic beads are introduced for cell tethering and purification, minimizing the cell input required. DiMeLo-seq and nanoHiMe-seq use pA fused to the methyltransferase Hia5 (pA–Hia5)[79,89], while nanoHiMe-seq adds a secondary antibody. BIND&MODIFY uses pA fused to M.EcoGII (pA–M.EcoGII) to bind the antibody[88]. After washing away unbound components, SAM activates the fusion protein, methylating adenines near antibody target sites. The methylation signal decays exponentially from the bound site with a half-life of ~170 bp in DiMeLo-seq. Genomic DNA is then extracted and sequenced using ONT (DiMeLo-seq, nanoHiMe-seq and BIND&MODIFY) or PacBio (DiMeLo-seq). Following these LRS-based chromatin accessibility workflows, data preprocessing relies on methylation-calling methods to detect induced methylation as markers for transcription factor binding or histone modifications (Fig. 5b)

DiMeLo-seq showcases the quantitative advantages of long-read, single-molecule techniques by delivering precise and unbiased estimates of protein–DNA interaction frequencies. Unlike SRS methods, which depend on PCR amplification and produce relative quantification peaks, LRS antibody-targeted protocols sequence native, unamplified DNA molecules. This approach enables direct measurement of binding events without amplification-induced biases. For instance, DiMeLo-seq applied to the centromere dip region (CDR) of chromosome X using the AlphaHOR-RES restriction-digest-enrichment strategy achieved over 20-fold coverage in these repetitive regions[78]. The authors estimated that approximately 26 ± 5% (at least 1 in 4) of

Fig. 4 | Data analysis methods for LRS-based chromatin accessibility profiling. **a**, Identification of nucleosome footprints (regions where DNA is wrapped around histone proteins) based on patterns of inaccessible regions, approximately 147 bp in length along single DNA molecules, can reveal heterogeneity in nucleosome organization among different cells. **b**, Detection of MSPs, regions marked by methyltransferases indicative of accessible DNA between nucleosomes, is classified by machine-learning models such as extreme gradient boosting (XGBoost, a highly efficient and scalable machine-learning model based on decision trees) into either accessible region (for example, dhsS, peaks of CTCF binding) or inaccessible ones. Positive dataset (Pos)—MSPs overlapping known DNase I hypersensitive sites (DHSs) and CTCF ChIP–seq peaks. Negative dataset (Neg)—MSPs not overlapping known DHSs and CTCF ChIP–seq peaks. HMMs (statistical models that represent systems with underlying hidden states) such as FiberHMM infer DNA accessibility and protein occupancy at a single-molecule resolution. **c**, Co-actuation analysis assesses the simultaneous accessibility of adjacent genomic regions along the same chromatin fiber to identify associations among regulatory elements over kilobase distances. Interactions are visualized using graphs where nodes represent accessible regions and edges depict interaction strengths. **d**, Transcription factor footprints are identified within open chromatin regions by analyzing the lengths of footprints and matching them to known transcription factor binding motifs (for example, CTCF) using position weight matrices (PWMs).

https://doi.org/10.1038/s41588-024-02038-5

nucleosomes within the ~100-kb CDR contain the centromere-specific histone variant CENP-A; this is substantially higher than a previous ChIP–seq study, which reported an average CENP-A density of 1 in 25 nucleosomes across endogenous human centromeres, based on an assumed mean centromere size of ~1 Mb (ref. 90). By enabling absolute frequency measurements and detailed single-molecule insights, DiMeLo-seq can thus help to advance our understanding of chromatin architecture and its regulatory mechanisms.



https://doi.org/10.1038/s41588-024-02038-5

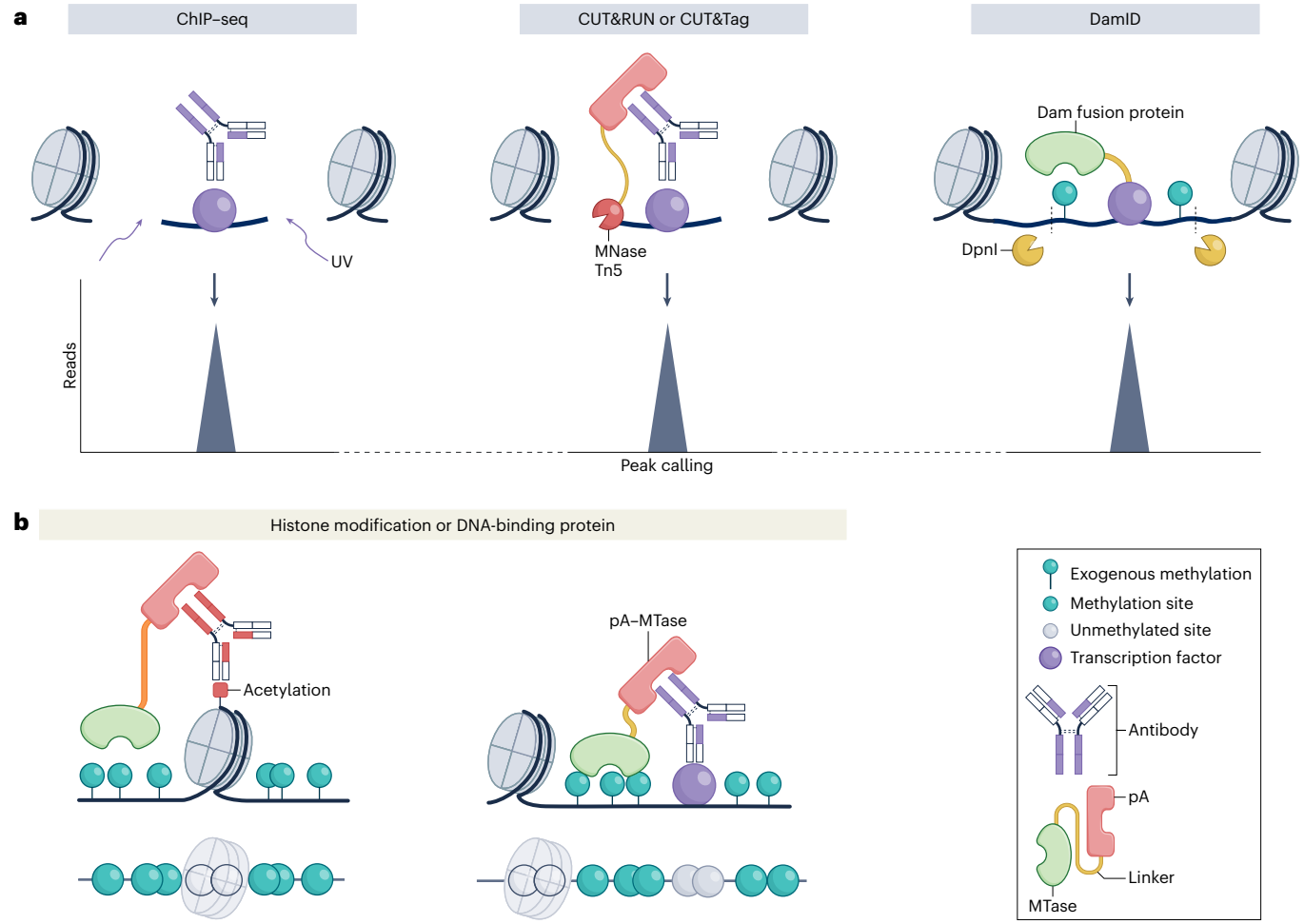


Fig. 5 | Methods for mapping protein–DNA interactions using SRS and LRS. **a**, SRS techniques are shown. In ChIP–seq, antibodies specific to proteins or histone modifications immunoprecipitate bound DNA fragments, revealing interaction sites upon sequencing. DamID uses a fusion of the protein of interest with Dam to methylate adenines near binding sites that are identified through methylation-sensitive sequencing. CUT&RUN and CUT&Tag tether MNase to antibodies against chromatin-bound proteins and cleave adjacent DNA without cross-linking, enabling the sequencing of released fragments. **b**, LRS techniques, such as DiMeLo-seq, nanoHiMe-seq and BIND&MODIFY, map protein–DNA interactions at single-molecule resolution over long genomic distances. These methods use pA fused to a methyltransferase (pA–MTase) to methylate adenines near antibody-bound targets on individual DNA molecules. The induced methylation signals are detected on long-sequencing reads, allowing direct visualization of protein–DNA interactions across extensive genomic regions, including highly repetitive sequences.

Furthermore, LRS has been used to perform a single-molecule analysis of histone modifications by calculating a histone modification score for each DNA molecule, defined as the ratio of 6mA-modified bases to total adenosines within specific genomic bins[88]. Hierarchical clustering based on gene-specific variations in H3K27me3 histone modification across associated DNA molecules categorized them into heavy (>75%), medium (25–75%) and light (<25%) epigenetic states. Comparing these states with gene expression levels revealed that genes in heavy and medium H3K27me3 states exhibit substantially lower expression. Additionally, the authors observed a correlation between H3K27me3 and CpG methylation within individual chromatin fibers, suggesting that these modifications collaborate to silence gene expression[88]. This approach provides insights into the mechanisms of epigenomic control of gene expression and addresses biological questions about the heterogeneity of epigenetic modifications at the single-molecule level.

The field needs more standardized best practices for analyzing data from these emerging direct protein–DNA interaction mapping methods. Developing benchmarks to assess the efficiency of fusion protein labeling, antibody binding efficiency and signal consistency across experimental conditions would greatly enhance data interpretation. For example, these direct mapping methods offer high specificity by generating methylation signals near antibody target sites, unlike chromatin accessibility methods such as Fiber-seq, which indirectly infer protein binding by detecting regions of unmethylated DNA. They may suffer from lower sensitivity due to variations in the binding strength of the target protein, antibody or fusion protein[79]. Therefore, computational models that account for variability in labeling efficiency are essential to enable robust data interpretation and ensure the accurate quantification of protein–DNA binding events across different genomic contexts.

3D genome organization

By incorporating LRS into the traditional Hi-C protocol[91], methods such as C-walk[92], MC-4C[93,94], MC-3C[95], Pore-C[96] and HiPore-C[97] overcome the limitations of SRS, which captures only pairwise chromatin interactions due to short-read lengths. The extended read lengths of LRS enable the detection of complex chromatin interactions involving multiple genomic loci—referred to as multiway chromatin interactions—within a single DNA molecule. Additionally, LRS facilitates the detection of epigenetic modifications through advanced methylation-calling algorithms[98]. Consequently, LRS provides a more comprehensive view of three-dimensional genome architecture.

For instance, Pore-C has been used to map multiway chromatin interactions and integrated ATAC-seq, RNA sequencing (RNA-seq) and

ChIP–seq to reveal their functional significance[98]. Pore-C identified chromatin contacts as hypergraphs linked to accessible regions by ATAC-seq. RNA-seq associated these regions with active gene expression, while ChIP–seq identified transcription factor binding sites. This combined approach revealed transcription clusters—groups of colocalized loci where accessible chromatin and transcription factor binding coordinate gene expression, highlighting both structural and regulatory aspects of multiway chromatin contacts[98].

However, the current experimental methods and computational tools are designed to capture and analyze chromatin interactions at the bulk level or, at best, at the single-cell level with limited resolution[99,100]. Additionally, 3D loop formation is short lived and dynamic[101]. Therefore, fully leveraging the single-molecule capabilities of LRS is crucial for detecting interactions between specific genomic loci on the same DNA molecule to understand the dynamic nature of chromatin architecture (reviewed in ref. 102).

LRS in transcriptomics

Long-read transcriptomic assays using ONT and PacBio platforms provide the capability to sequence full-length RNA molecules, allowing for detailed characterization of transcript isoforms and complex splicing events without assembly. ONT devices such as PromethION offer exceptionally high throughput; they generate more than 100 million reads per flow cell, which facilitates the detection of low-abundance transcripts (nanoporetech.com/products/sequence/promethion). Additionally, ONT's direct RNA-seq preserves native RNA modifications such as N6-methyladenosine (m6A), allowing epitranscriptomic studies at single-molecule resolution (reviewed in ref. 103). PacBio platforms, including the Sequel IIe and the newer Revio system, provide highly accurate HiFi reads that exceed 99% accuracy and can generate up to 100 million reads per run, making them advantageous for applications that require precise base calling and detailed isoform characterization (pacb.com/revio/). While ONT offers higher throughput and the theoretical capacity to sequence very long reads with real-time data acquisition, it has historically had higher sequencing error rates and a bias against short transcripts[104], although recent improvements have enhanced accuracy. Conversely, PacBio offers superior read accuracy and uniform read lengths, beneficial for transcriptome assembly and variant detection, albeit with lower throughput and higher cost per base; its library preparation often includes size-selection steps that may select against very long or very short transcripts.

Numerous experimental and computational methods have been developed and benchmarked for both ONT and PacBio sequencing, with adaptations underway to study transcriptional dynamics. For instance, LRS Nascent-seq sequences newly synthesized RNA molecules using long-read technology[105–110], enabling detailed analysis of transcription initiation and elongation processes, while the recent long-read Ribo-STAMP measures translation with isoform resolution[111]. We anticipate that other protocols, such as photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation[112] and high-throughput sequencing of RNA isolated by cross-linking and immunoprecipitation[113], which identify RNA-binding protein interaction sites on RNAs, will soon have LRS versions. Furthermore, both ONT and PacBio have developed protocols for LRS RNA-seq on single cells and with spatial resolution (reviewed in ref. 114).

Nevertheless, long-read transcriptomics also faces challenges. In addition to improving sequencing accuracy and throughput, which are crucial for the faithful calling of transcript variants and providing quantitative gene expression measurements, the technology must optimize library preparations and RNA capture across the wide range of transcript lengths to provide unbiased representations of transcriptomes. Moreover, there is still room for improvement in algorithmic solutions and analytical strategies to better describe the complexity of transcriptomes as revealed by LRS and to provide new insights into the functional roles of diverse transcript isoforms.

Toward long-read multi-omics applications

The advent of LRS marks the beginning of a new era of long-read-based multi-omics approaches, enabling the study of links between transcriptional regulation and epigenomic mechanisms that were challenging to address with SRS. While such relationships have been observed using SRS, the findings are constrained by the shorter read lengths and amplification-induced biases. Long-read methods, by contrast, allow the study of these biological processes at a single-molecule resolution; this facilitates the identification of specific chromatin states that promote the synthesis and modification of distinct RNA molecules, achieving a molecular resolution that is beyond the reach of SRS.

For instance, multiple SRS studies have demonstrated that transcription initiation and elongation are associated with promoter DNA methylation and specific histone marks[115–117]. Combining the analysis of promoter-wide epigenomics profiles with long-read nascent RNA will help elucidate the relationship between chromatin states that involve multiple modifications, as well as initiation and elongation rates. Additionally, coupling long-read epigenomics methods with long-read RNA-seq allows studying how histone modifications, DNA methylation and/or transcription factor binding along the chromatin fiber influence RNA splicing and isoform usage, aspects only partially resolved by short-reads. Furthermore, combining LRS-based epigenomics with direct RNA-seq could provide deeper insights into the interplay between the epigenome and the epitranscriptome[118]. However, a substantial limitation of current technologies is their inability to map different molecular modalities to the same individual molecules. This prevents us from determining the exact DNA molecule, or its epigenetic state, from which a particular RNA molecule was transcribed. Overcoming this limitation will require developing new methods capable of simultaneously capturing multiple molecular modalities at the single-molecule level or designing experimental and computational strategies to address these challenges.

Examples of LRS-based multi-omics applications are beginning to emerge. For instance, by integrating Hi-C with SRS and LRS and with short-read RNA-seq to analyze complex germline chromosomal rearrangements in individuals with congenital disorders[119], the detection of breakpoints could be improved and facilitated the reconstruction of chromosomal architecture, which revealed that large-scale rearrangements frequently involve reshuffling of TADs and lamina-associated domains, with implications for gene regulation and expression[119]. Another approach combined genome, methylome, epigenome and transcriptome data from a single LRS run to explore the molecular basis of Mendelian conditions[77]; this revealed that a patient's complex phenotype was due to a balanced translocation between chromosome X and 13 that disrupts four genes (*NBEA*, *PDK3*, *MAB21L1* and *RB1*) through distinct mechanisms, including nonsense-mediated decay, fusion transcript formation, enhancer adoption, transcriptional readthrough silencing and inappropriate chromosome X inactivation of autosomal genes and demonstrated the effectiveness of synchronized long-read multi-omics profiling in resolving complex genetic disorders. Finally, the scNanoCOOL-seq protocol uses nanopore sequencing to profile the genome, DNA methylome, chromatin accessibility and transcriptome of individual cells[120], establishing molecular chromatin transcriptome associations within single cells.

These emerging applications demonstrate the significance and potential of LRS methods for studying cellular mechanisms with a single-molecule resolution. However, these applications are in their infancy. In addition to improvements in experimental procedures to increase the signal-to-noise ratio, sequencing coverage and reduce costs, there is a strong need for analytical processes tailored to these new data types. Specifically, new statistical methods are required to identify the co-occurrence of modifications within single molecules and to model regulatory relationships across nucleic acids. The bioinformatics community faces an exciting challenge in developing benchmarks for these emerging technologies, robust signal-extraction

methodologies, new long-read simulation algorithms, as well as new statistical methods to integrate the wealth of multifaceted long-read data, ultimately delivering testable models of transcriptional regulation. With these in hand, the era of long-read multi-omics is truly approaching.

Conclusion

The rapid evolution of single-molecule LRS technologies has catalyzed a transformative shift in epigenomics, enabling the simultaneous detection of DNA methylation, chromatin accessibility and protein–DNA interactions across individual chromatin fibers. By overcoming the inherent limitations of short-read methods, LRS offers unprecedented resolution in mapping complex epigenetic landscapes and capturing long-range chromatin interactions. These advancements are now facilitating the integration of multi-omics approaches, linking chromatin states, transcriptional regulation and gene expression in a single, cohesive framework. As computational and experimental tools mature, LRS is poised to redefine our understanding of the regulatory mechanisms that govern the epigenome.

References

1. Marx, V. Method of the year: long-read sequencing. *Nat. Methods* **20**, 6–11 (2023).
2. Nurk, S. et al. The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
3. Hotaling, S., Kelley, J. L. & Frandsen, P. B. Toward a genome sequence for every animal: where are we now? *Proc. Natl Acad. Sci. USA* **118**, e2109019118 (2021).
4. Mahmoud, M. et al. Structural variant calling: the long and the short of it. *Genome Biol.* **20**, 246 (2019).
5. Oikonomopoulos, S. et al. Methodologies for transcript profiling using long-read technologies. *Front. Genet.* **11**, 606 (2020).
6. Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38 (2013).
7. Angeloni, A. & Bogdanovic, O. Sequence determinants, function, and evolution of CpG islands. *Biochem. Soc. Trans.* **49**, 1109–1119 (2021).
8. Kelly, T. K. et al. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012).
9. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
10. Meissner, A. et al. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).
11. Taiwo, O. et al. Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat. Protoc.* **7**, 617–636 (2012).
12. Cross, S. H., Charlton, J. A., Nan, X. & Bird, A. P. Purification of CpG islands using a methylated DNA binding column. *Nat. Genet.* **6**, 236–244 (1994).
13. Costello, J. F. et al. Aberrant CpG-island methylation has non-random and tumour-type–specific patterns. *Nat. Genet.* **24**, 132–138 (2000).
14. Hayashizaki, Y. et al. Restriction landmark genomic scanning method and its various applications. *Electrophoresis* **14**, 251–258 (1993).
15. Brunner, A. L. et al. Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res.* **19**, 1044–1056 (2009).
16. Maunakea, A. K. et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253–257 (2010).
17. Laird, P. W. Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.* **11**, 191–203 (2010).
18. Parrish, R., Day, J. J. & Lubin, F. D. Direct bisulfite sequencing for examination of DNA methylation patterns with gene and nucleotide resolution from brain tissues. *Curr. Protoc. Neurosci.* https://doi.org/10.1002/0471142301.ns0724s60 (2012).
19. Dai, Q. et al. Ultrafast bisulfite sequencing detection of 5-methylcytosine in DNA and RNA. *Nat. Biotechnol.* **42**, 1559–1570 (2024).
20. Vaisvila, R. et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Res.* **31**, 1280–1289 (2021).
21. Liu, Y. et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* **37**, 424–429 (2019).
22. Pastor, W. A. et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
23. Kirschner, S. A. et al. Focussing reduced representation CpG sequencing through judicious restriction enzyme choice. *Genomics* **107**, 109–119 (2016).
24. Warnecke, P. M. et al. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res.* **25**, 4422–4426 (1997).
25. Aird, D. et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, R18 (2011).
26. Lucas, M. C. & Novoa, E. M. Long-read sequencing in the era of epigenomics and epitranscriptomics. *Nat. Methods* **20**, 25–29 (2023).
27. Flusberg, B. A. et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2010).
28. Rhoads, A. & Au, K. F. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* **13**, 278–289 (2015).
29. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, 1348–1365 (2021).
30. Ni, P. et al. DNA 5-methylcytosine detection and methylation phasing using PacBio circular consensus sequencing. *Nat. Commun.* **14**, 4054 (2023).
31. Tse, O. Y. O. et al. Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proc. Natl Acad. Sci. USA* **118**, e2019768118 (2021).
32. Liu, Q., Georgieva, D. C., Egli, D. & Wang, K. NanoMod: a computational tool to detect DNA modifications using nanopore long-read sequencing data. *BMC Genomics* **20**, 78 (2019).
33. Stoiber, M. et al. De novo identification of DNA modifications enabled by genome-guided nanopore signal processing. Preprint at *bioRxiv* https://doi.org/10.1101/094672 (2017).
34. Yuen, Z. W.-S. et al. Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nat. Commun.* **12**, 3438 (2021).
35. Liu, Y. et al. DNA methylation-calling tools for Oxford nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol.* **22**, 295 (2021).
36. Jha, A. et al. DNA-m6A calling and integrated long-read epigenetic and genetic analysis with fibertools. *Genome Res.* **34**, 1976–1986 (2024).
37. Simpson, J. T. et al. Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14**, 407–410 (2017).
38. McIntyre, A. B. R. et al. Single-molecule sequencing detection of N6-methyladenine in microbial reference materials. *Nat. Commun.* **10**, 579 (2019).
39. Liu, Q. et al. Detection of DNA base modifications by deep recurrent neural network on Oxford nanopore sequencing data. *Nat. Commun.* **10**, 2449 (2019).

40. Ni, P. et al. DeepSignal: detecting DNA methylation state from nanopore sequencing reads using deep-learning. *Bioinformatics* **35**, 4586–4595 (2019).
41. Rand, A. C. et al. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* **14**, 411–413 (2017).
42. Ahsan, M. U., Gouru, A., Chan, J., Zhou, W. & Wang, K. A signal processing and deep learning framework for methylation detection using Oxford nanopore sequencing. *Nat. Commun.* **15**, 1448 (2024).
43. Stefansson, O. A. et al. The correlation between CpG methylation and gene expression is driven by sequence variants. *Nat. Genet.* **56**, 1624–1631 (2024).
44. Sigurpalsdottir, B. D. et al. A comparison of methods for detecting DNA methylation from long-read sequencing of human genomes. *Genome Biol.* **25**, 69 (2024).
45. Wang, Z. et al. Adapting nanopore sequencing basecalling models for modification detection via incremental learning and anomaly detection. *Nat. Commun.* **15**, 7148 (2024).
46. Amarasinghe, S. L. et al. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 30 (2020).
47. Gouil, Q. & Keniry, A. Latest techniques to study DNA methylation. *Essays Biochem.* **63**, 639–648 (2019).
48. Wenger, A. M. et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **37**, 1155–1162 (2019).
49. Pott, S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *eLife* **6**, e23203 (2017).
50. Krebs, A. R. et al. Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell* **67**, 411–422 (2017).
51. Xu, M., Kladde, M. P., Simpson, R. T. & Van Etten, J. L. Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC. *Nucleic Acids Res.* **26**, 3961–3966 (1998).
52. Krebs, A. R. Studying transcription factor function in the genome at molecular resolution. *Trends Genet.* **37**, 798–806 (2021).
53. He, R. et al. Human transcription factor combinations mapped by footprinting with deaminase. Preprint at *bioRxiv* https://doi.org/10.1101/2024.06.14.599019 (2024).
54. Shipony, Z. et al. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* **17**, 319–327 (2020).
55. Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S. & Stamatoyannopoulos, J. A. Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* **368**, 1449–1454 (2020).
56. Wang, Y. et al. Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res.* **29**, 1329–1342 (2019).
57. Lee, I. et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* **17**, 1191–1199 (2020).
58. Mo, W. et al. Single-molecule targeted accessibility and methylation sequencing of centromeres, telomeres and rDNAs in *Arabidopsis*. *Nat. Plants* **9**, 1439–1450 (2023).
59. Oberbeckmann, E. et al. Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Res.* **29**, 1996–2009 (2019).
60. Lister, R. et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
61. Douvlataniotis, K., Bensberg, M., Lentini, A., Gylemo, B. & Nestor, C. E. No evidence for DNA N6-methyladenine in mammals. *Sci. Adv.* **6**, eaay3335 (2020).
62. Kong, Y. et al. Critical assessment of DNA adenine methylation in eukaryotes using quantitative deconvolution. *Science* **375**, 515–522 (2022).
63. Debo, B. M., Mallory, B. J. & Stergachis, A. B. Evaluation of N6-methyldeoxyadenosine antibody-based genomic profiling in eukaryotes. *Genome Res.* **33**, 427–434 (2023).
64. Leduque, B., Edera, A., Vitte, C. & Quadrana, L. Simultaneous profiling of chromatin accessibility and DNA methylation in complete plant genomes using long-read sequencing. *Nucleic Acids Res.* **52**, 6285–6297 (2024).
65. Xie, Y. et al. Spatial chromatin accessibility sequencing resolves high-order spatial interactions of epigenomic markers. *eLife* **12**, RP87868 (2024).
66. Hook, P. W. & Timp, W. Beyond assembly: the increasing flexibility of single-molecule sequencing technology. *Nat. Rev. Genet.* **24**, 627–641 (2023).
67. Giesselmann, P. et al. Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. *Nat. Biotechnol.* **37**, 1478–1481 (2019).
68. Gilpatrick, T. et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* **38**, 433–438 (2020).
69. Bohaczuk, S. C. et al. Resolving the chromatin impact of mosaic variants with targeted Fiber-seq. Preprint at *bioRxiv* https://doi.org/10.1101/2024.07.09.602608 (2024).
70. Martin, S. et al. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol.* **23**, 11 (2022).
71. Boltengagen, M. et al. A single fiber view of the nucleosome organization in eukaryotic chromatin. *Nucleic Acids Res.* **52**, 166–185 (2024).
72. Dennis, A. F., Xu, Z. & Clark, D. J. Examining chromatin heterogeneity through PacBio long-read sequencing of M.EcoGII methylated genomes: an m6A detection efficiency and calling bias correcting pipeline. *Nucleic Acids Res.* **52**, e45 (2024).
73. Vollger, M. R. et al. A haplotype-resolved view of human gene regulation. Preprint at *bioRxiv* https://doi.org/10.1101/2024.06.14.599122 (2024).
74. Isaac, R. S. et al. Single-nucleoid architecture reveals heterogeneous packaging of mitochondrial DNA. *Nat. Struct. Mol. Biol.* **31**, 568–577 (2024).
75. Tullius, T. W. et al. RNA polymerases reshape chromatin architecture and couple transcription on individual fibers. *Mol. Cell* **84**, 3209–3222 (2024).
76. Grasberger, H. et al. STR mutations on chromosome 15q cause thyrotropin resistance by activating a primate-specific enhancer of MIR7-2/MIR1179. *Nat. Genet.* **56**, 877–888 (2024).
77. Vollger, M. R. et al. Synchronized long-read genome, methylome, epigenome, and transcriptome for resolving a Mendelian condition. Preprint at *bioRxiv* https://doi.org/10.1101/2023.09.26.559521 (2023).
78. Fu, Y. et al. MethPhaser: methylation-based long-read haplotype phasing of human genomes. *Nat. Commun.* **15**, 5327 (2024).
79. Altemose, N. et al. DiMeLo-seq: a long-read, single-molecule method for mapping protein–DNA interactions genome wide. *Nat. Methods* **19**, 711–723 (2022).
80. Landt, S. G. et al. ChIP–seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
81. Park, P. J. ChIP–seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10**, 669–680 (2009).
82. Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein–DNA interactions using DamID in mammalian cells. *Nat. Protoc.* **2**, 1467–1478 (2007).

83. Wu, F., Olson, B. G. & Yao, J. DamID-seq: Genome-wide mapping of protein–DNA interactions by high throughput sequencing of adenine-methylated DNA fragments. *J. Vis. Exp.* https://doi.org/10.3791/53620 (2016).
84. Babl, V., Stöckl, U., Tschochner, H., Milkereit, P. & Griesenbeck, J. Chromatin endogenous cleavage (ChEC) as a method to quantify protein interaction with genomic DNA in *Saccharomyces cerevisiae*. *Methods Mol. Biol.* **1334**, 219–232 (2015).
85. Schmid, M., Durussel, T. & Laemmli, U. K. ChIC and ChEC; genomic mapping of chromatin proteins. *Mol. Cell* **16**, 147–157 (2004).
86. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
87. Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
88. Weng, Z. et al. BIND&MODIFY: a long-range method for single-molecule mapping of chromatin modifications in eukaryotes. *Genome Biol.* **24**, 61 (2023).
89. Yue, X. et al. Simultaneous profiling of histone modifications and DNA methylation via nanopore sequencing. *Nat. Commun.* **13**, 7939 (2022).
90. Bodor, D. L. et al. The quantitative architecture of centromeric chromatin. *eLife* **3**, e02137 (2014).
91. Belton, J.-M. et al. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
92. Olivares-Chauvet, P. et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* **540**, 296–300 (2016).
93. Allahyar, A. et al. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018).
94. Vermeulen, C. et al. Multi-contact 4C: long-molecule sequencing of complex proximity ligation products to uncover local cooperative and competitive chromatin topologies. *Nat. Protoc.* **15**, 364–397 (2020).
95. Tavares-Cadete, F., Norouzi, D., Dekker, B., Liu, Y. & Dekker, J. Multi-contact 3C reveals that the human genome during interphase is largely not entangled. *Nat. Struct. Mol. Biol.* **27**, 1105–1114 (2020).
96. Deshpande, A. S. et al. Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. *Nat. Biotechnol.* **40**, 1488–1499 (2022).
97. Zhong, J.-Y. et al. High-throughput Pore-C reveals the single-allele topology and cell type-specificity of 3D genome folding. *Nat. Commun.* **14**, 1250 (2023).
98. Dotson, G. A. et al. Deciphering multi-way interactions in the human genome. *Nat. Commun.* **13**, 5498 (2022).
99. Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
100. Rappoport, N. et al. Single cell Hi-C identifies plastic chromosome conformations underlying the gastrulation enhancer landscape. *Nat. Commun.* **14**, 3844 (2023).
101. Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20–32 (2018).
102. Zhang, Y. et al. Computational methods for analysing multiscale 3D genome organization. *Nat. Rev. Genet.* **25**, 123–141 (2024).
103. Jain, M., Abu-Shumays, R., Olsen, H. E. & Akeson, M. Advances in nanopore direct RNA sequencing. *Nat. Methods* **19**, 1160–1164 (2022).
104. Pardo-Palacios, F. J. et al. Systematic assessment of long-read RNA-seq methods for transcript identification and quantification. *Nat. Methods* **21**, 1349–1363 (2024).
105. Herzel, L., Straube, K. & Neugebauer, K. M. Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res.* **28**, 1008–1019 (2018).
106. Long, Y., Jia, J., Mo, W., Jin, X. & Zhai, J. FLEP-seq: simultaneous detection of RNA polymerase II position, splicing status, polyadenylation site and poly(A) tail length at genome-wide scale by single-molecule nascent RNA sequencing. *Nat. Protoc.* **16**, 4355–4381 (2021).
107. Carrillo Oesterreich, F. et al. Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell* **165**, 372–381 (2016).
108. Reimer, K. A., Mimoso, C. A., Adelman, K. & Neugebauer, K. M. Co-transcriptional splicing regulates 3′ end cleavage during mammalian erythropoiesis. *Mol. Cell* **81**, 998–1012 (2021).
109. Drexler, H. L. et al. Revealing nascent RNA processing dynamics with nano-COP. *Nat. Protoc.* **16**, 1343–1375 (2021).
110. Maier, K. C., Gressel, S., Cramer, P. & Schwalb, B. Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *Genome Res.* **30**, 1332–1344 (2020).
111. Jagannatha, P. et al. Long-read Ribo-STAMP simultaneously measures transcription and translation with isoform resolution. *Genome Res.* **34**, 2012–2024 (2024).
112. Spitzer, J. et al. PAR-CLIP (photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods Enzymol.* **539**, 113–161 (2014).
113. Darnell, R. B. HITS-CLIP: panoramic views of protein–RNA regulation in living cells. *Wiley Interdiscip. Rev. RNA* **1**, 266–286 (2010).
114. Joglekar, A., Foord, C., Jarroux, J., Pollard, S. & Tilgner, H. U. From words to complete phrases: insight into single-cell isoforms using short and long reads. *Transcription* **14**, 92–104 (2023).
115. Veloso, A. et al. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* **24**, 896–905 (2014).
116. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
117. Wang, H. et al. H3K4me3 regulates RNA polymerase II promoter-proximal pause-release. *Nature* **615**, 339–348 (2023).
118. Kan, R. L., Chen, J. & Sallam, T. Crosstalk between epitranscriptomic and epigenetic mechanisms in gene regulation. *Trends Genet.* **38**, 182–193 (2022).
119. Schöpflin, R. et al. Integration of Hi-C with short and long-read genome sequencing reveals the structure of germline rearranged genomes. *Nat. Commun.* **13**, 6470 (2022).
120. Lin, J. et al. scNanoCOOL-seq: a long-read single-cell sequencing method for multi-omics profiling within individual cells. *Cell Res.* **33**, 879–882 (2023).
121. Abdulhay, N. J. et al. Massively multiplex single-molecule oligonucleosome footprinting. *eLife* **9**, e59404 (2020).
122. Nanda, A. S. et al. Direct transposition of native DNA for sensitive multimodal single-molecule sequencing. *Nat. Genet.* **56**, 1300–1309 (2024).
123. Battaglia, S. et al. Long-range phasing of dynamic, tissue-specific and allele-specific regulatory elements. *Nat. Genet.* **54**, 1504–1513 (2022).

Acknowledgements

This work has been funded by the European Union Marie Skłodowska-Curie Actions Doctoral Network project LongTREC (HORIZON-MSCA-2021-DN-01 grant agreement project 101072892). This work was partially funded by the Spanish Science Ministry, grant number PID2020-119537RB-I00.

Author contributions

T.L. and A.C. conceived the outline of the manuscript. T.L. collected literature, created figures and drafted the manuscript. A.C. supervised the work and proofread the manuscript.

Competing interests

A.C. has received in-kind funding from PacBio for library preparation and sequencing. A.C. and T.L. collaborate with Oxford Nanopore in the Marie Skłodowska-Curie Actions Doctoral Network project LongTREC.

Additional information

Correspondence and requests for materials should be addressed to Ana Conesa.

Peer review information *Nature Genetics* thanks Andrew Stergachis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.