8. Misteli, T. (2001) The concept of self-organization in cellular architecture. *J. Cell Biol.* 155, 181–185
9. Mackintosh, F. and Schmidt, C. (2010) Active cellular materials. *Curr. Opin. Cell Biol.* 22, 29–35
10. Prost, J. *et al.* (2015) Active gel physics. *Nat. Phys.* 11, 111–117
11. Qian, H. (2005) Cycle kinetics, steady-state thermodynamics and motors – a paradigm for living matter physics. *J. Phys. Condens. Matter* 17, S3783–S3794
12. Sanchez, T. *et al.* (2012) Spontaneous motion in hierarchically assembled active matter. *Nature* 491, 431–434
13. Brugues, J. and Needleman, D. (2014) Physical basis of spindle self-organization. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18496–18500
14. Mayer, M. *et al.* (2010) Anisotropies in cortical tension reveal the physical basis of polarizing cortical flows. *Nature* 467, 617–621

Special Issue: Quantitative Cell Biology

Forum

Beyond the Linear Genome: Paired-End Sequencing as a Biophysical Tool

Viviana I. Risca[1] and William J. Greenleaf[1,2,*]

Paired-end sequencing has enabled a variety of new methods for high-throughput interrogation of both genome structure and chromatin architecture. Here, we discuss how the paired-end paradigm can be used to interpret sequencing data as biophysical measurements of *in vivo* chromatin structure that report on single molecules in single cells.

Paired-End Sequencing

Pairwise correlations are perhaps the simplest and yet most powerful measurements in biology. Correlation measurements are fundamentally enabled by pairwise measurements of variables that share a fixed characteristic. Pairs of measurements can link biological elements in space (e.g., the relaxation of two proximal atoms exhibiting spin coupling), time (e.g., correlations of neuronal firing), or function (e.g., two mutations capable of compensating for each other), allowing reconstruction of biological components from macromolecular structures to complex biological signaling networks. Pairwise measurements become still more informative when deployed in high throughput to comprehensively map interactions in a biological system. Here, we explore the ways in which pairwise measurements can be made using DNA sequencing-based assays, which effectively report single-molecule information, allowing multiplex biophysical measurements in living cells.

Modern DNA sequencing technology routinely produces hundreds of millions of short reads spanning tens to hundreds of base pairs for only a few thousand dollars. From its earliest and simplest application to reading out genome sequences, DNA sequencing has evolved, through the generation of diverse assays that use short DNA fragments as a readout, into a powerful tool for cell biology and nucleic acid biophysics, enabling assays of protein–DNA interactions [1], RNA expression and splicing [1–3], and ribosome–RNA interactions [4]. These methods are complementary to established lower-throughput assays such as live-cell and immunofluorescence microscopy, which, although lower in throughput, can often access temporal dynamics that sequencing cannot and can validate observations from sequencing experiments. Although a thorough review of the many applications of sequencing to biophysical measurements is beyond the

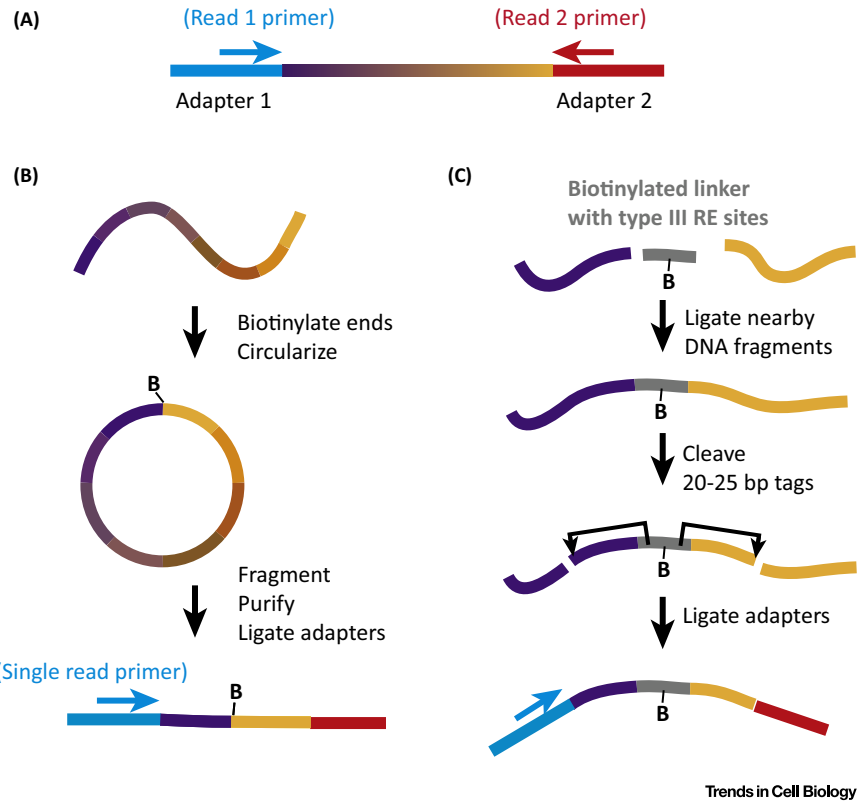

Figure 1. Types of Paired-End Libraries. (A) Simple library molecule with two adapter sequences complementary to two read primer sequences for simple paired-end sequencing. (B) Example of mate-pair library production from long (several kilobase) fragments of genomic DNA. (C) Example of paired-end tag library generation, in which a Type III restriction endonuclease (RE) with recognition sites encoded in the common linker is used to cleave 20–25 bp sequence tags from genomic DNA flanking the linker [6].

scope of this forum article, we will focus on methods that allow for correlated measurements using paired-end sequencing, a modality that is particularly promising for maximizing biophysical and cell biological insight.

In most applications, paired-end sequencing is carried out by performing two (or more) sequential rounds of sequencing-by-synthesis on each library molecule (Figure 1A), and these separate reads are identified as linked in subsequent analysis. For libraries in which the insert size of genomic DNA exceeds the length of each read (which is often true for short-read platforms), reading both ends of each insert allows mapping of the fragment onto a reference genome and determination of the insert length. If reads align discordantly (i.e., if fragment lengths exceed the known size range of the library, or if orientations are inconsistent), this information can be used to infer structural variation of the sequenced genome. Alternately, single-read methods can be applied to libraries that are circularized, creating 'mate pairs' representing inserts of several kilobases [5] (Figure 1B). Members of a library can also be cleaved and ligated consecutively as paired-end tags (PETs) [6], then read out on a single-end platform (Figure 1C).

Measurements of Molecular Contiguity

Paired-end sequencing has been extensively applied to measure the contiguity of single DNA molecules, a crucial step in *de novo* genome sequence assembly, haplotype phasing, and the detection of structural variation (Figure 2A) [5,7]. Recent extensions of these ideas have combined whole-genome amplification with paired-end sequencing to detect the emergence of chromosome rearrangements in single cells of a human embryo over a single cell cycle, enabling observation of pairs of daughter cells with reciprocal rearrangements [8]. Much longer contiguity

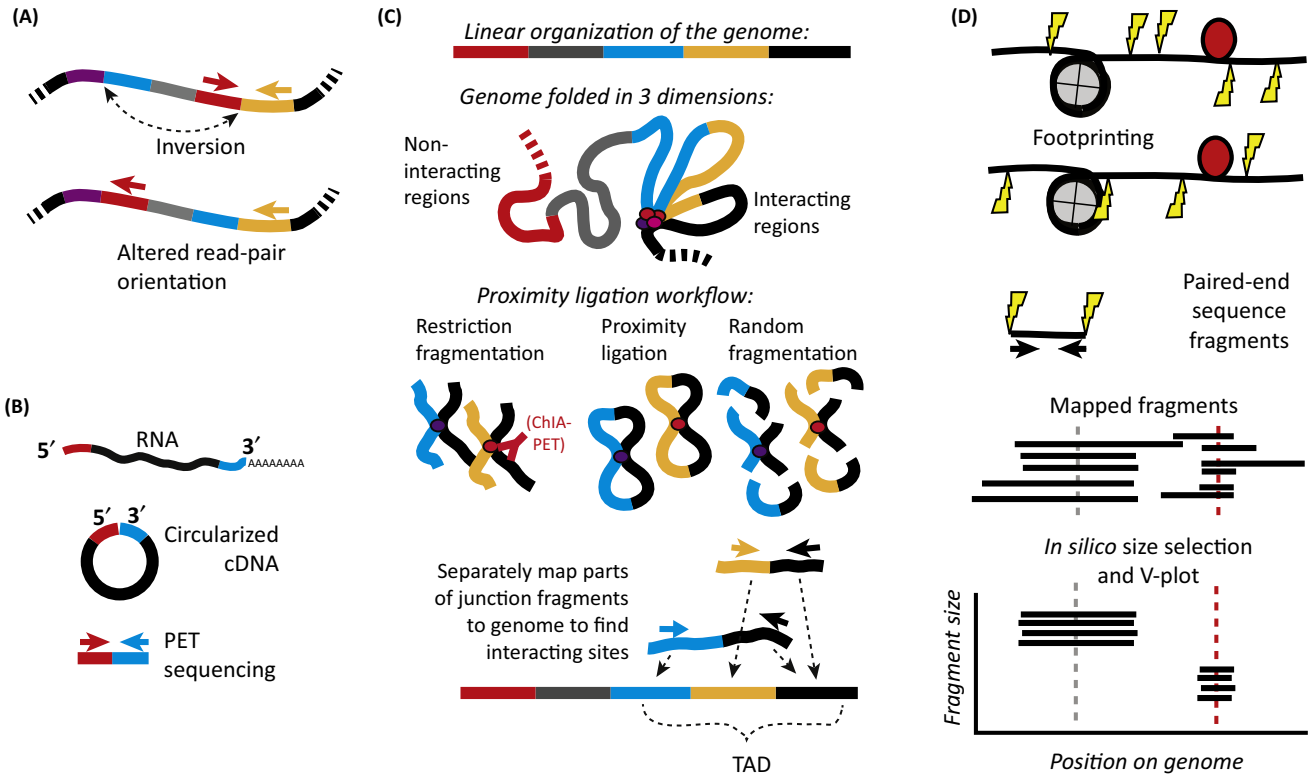

Figure 2. Extracting Biophysical and Structural Information from Mapped Paired-End Reads. (A) Paired-end or mate-paired read orientation can be used to detect structural variation in genomes, such as the inversion shown here. (B) RNA 5′ and 3′ ends can be simultaneously mapped with RNA-PET, an example of use of paired-end sequencing in transcriptomics. (C) Hi-C and ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) report on 3D contacts of loci along the linear genome (shown as colored blocks). In Hi-C [9], crosslinked and lightly permeabilized nuclei are treated with restriction endonucleases to cut the genome into short pieces spanning several hundred base pairs. The ends of spatially proximal DNA fragments held together by crosslinked proteins (red or purple circles) are then ligated to generate chimeric molecules (blue–black and yellow–black). Paired-end sequencing of these chimeric molecules yields pairs of reads that can be separately mapped to the genome, with each read pair reporting contact between a pair of loci in a single cell (e.g., blue–black, yellow–black). A region with a high density of self-contacts (blue/yellow/black) is designated a topologically associating domain (TAD). ChIA-PET [6] operates similarly, but includes immunoprecipitation (red Y) to isolate interactions mediated by a particular protein of interest (red oval). (D) Footprinting assays use cleavage of unprotected DNA by an enzyme (lightning bolt) to generate short DNA fragments. Paired-end sequencing of these fragments provides precise fragment length information that can improve signal-to-noise and permit the footprinting of multiple classes of particles with V-plots of fragment length versus genomic coordinates [1].