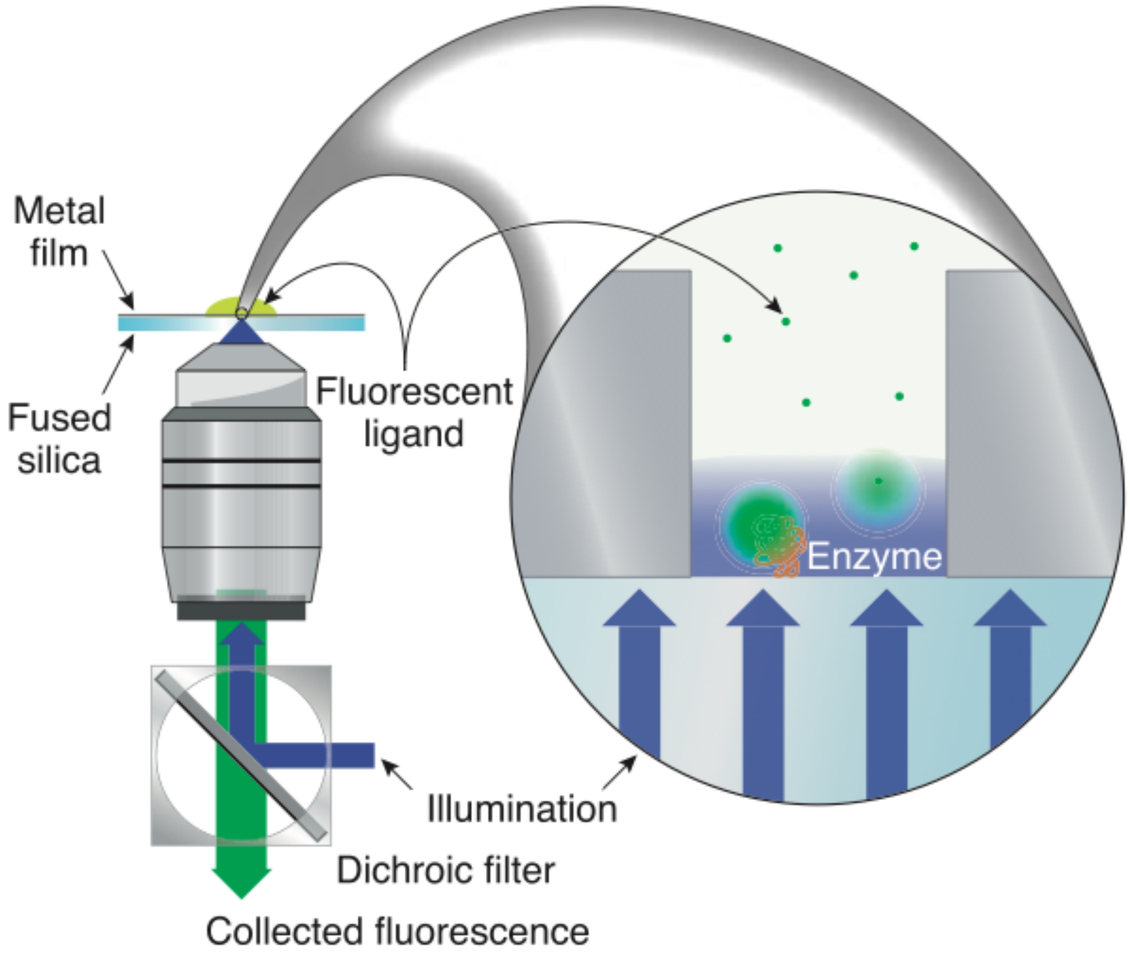
PacBio

Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations

M. J. Levene,[1] J. Korlach,[1,2] S. W. Turner,[1]* M. Foquet,[1] H. G. Craighead,[1] W. W. Webb[1]†

Optical approaches for observing the dynamics of single molecules have required pico- to nanomolar concentrations of fluorophore in order to isolate individual molecules. However, many biologically relevant processes occur at micromolar ligand concentrations, necessitating a reduction in the conventional observation volume by three orders of magnitude. We show that arrays of zero-mode waveguides consisting of subwavelength holes in a metal film provide a simple and highly parallel means for studying single-molecule dynamics at micromolar concentrations with microsecond temporal resolution. We present observations of DNA polymerase activity as an example of the effectiveness of zero-mode waveguides for performing single-molecule experiments at high concentrations.

Fig. 1. An apparatus for single-molecule enzymology using zero-mode waveguides.



Levene et al. 2003

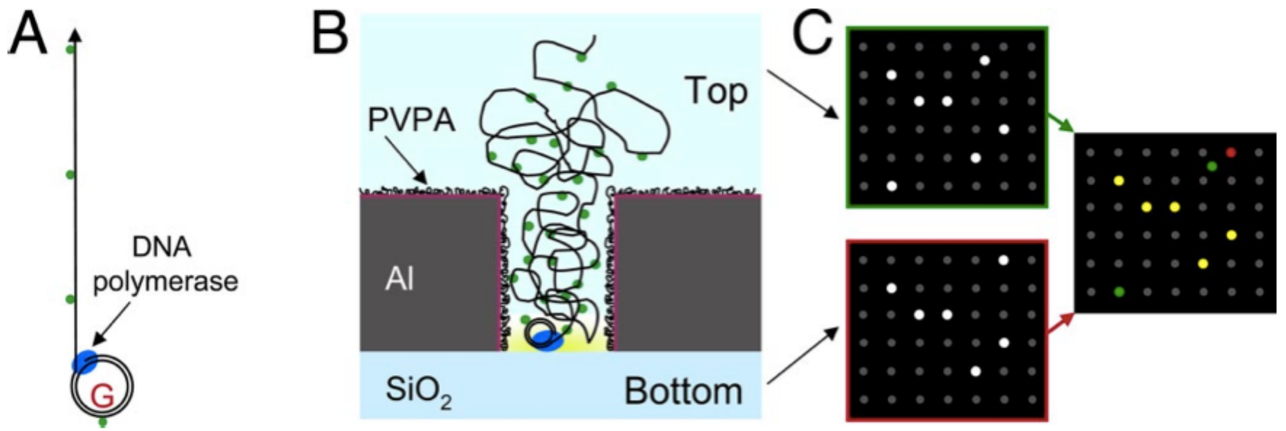


Fig. 2. Principle of observing DNA synthesis inside ZMWs. (*A*) Template design. The minicircle DNA template contained a single guanine site, allowing incorporation of a base-linked fluorescent nucleotide, Alexa Fluor 488-dCTP. Rolling circle, DNA strand displacement synthesis by ϕ29 DNA polymerase produced DNA with fluorescent labels at regular DNA length intervals (72 bases). (*B*) ZMW nanostructures were treated with PVPA, enabling selective immobilization of DNA polymerase at the bottom of ZMWs, followed by DNA extension reactions. The ZMW observation volume is highlighted in yellow. (*C*) Fluorescent DNA products were imaged from both sides of ZMW arrays. Image superposition and colocalization analysis was used to determine the bias of immobilization toward the glass floor (SiO_2, yellow and red dots) over the side wall and top surfaces (Al, green dots) and to demonstrate single molecule occupancy. The DNA length was determined by fluorescence-brightness analysis.

Korlach et al. 2008

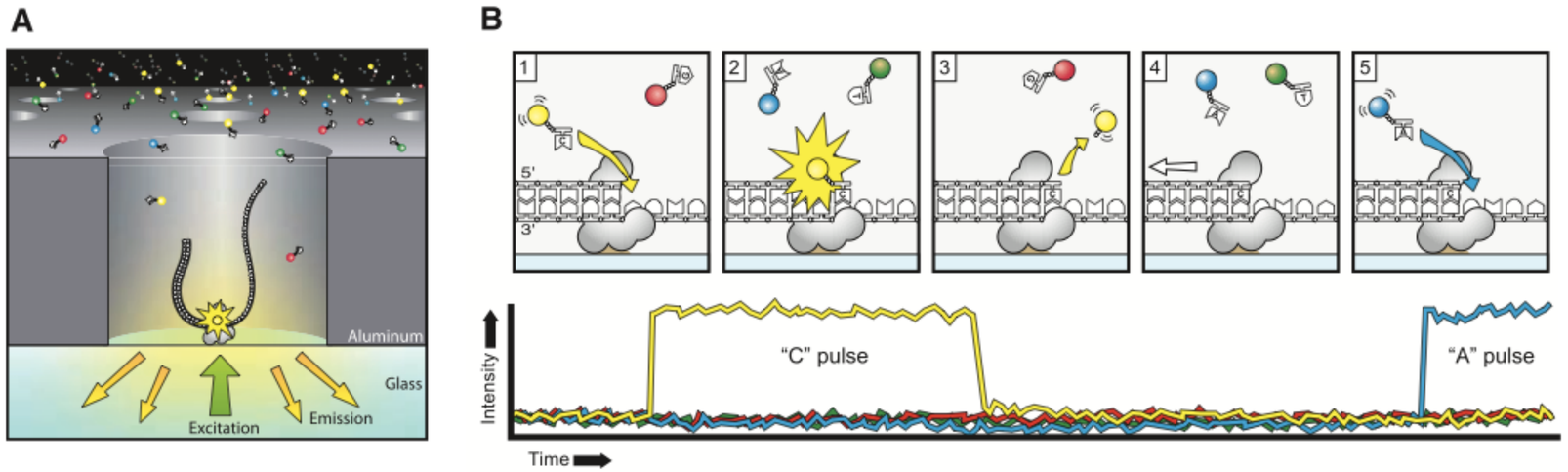


Fig. 1. Principle of single-molecule, real-time DNA sequencing. (**A**) Experimental geometry. A single molecule of DNA template-bound Φ29 DNA polymerase is immobilized at the bottom of a ZMW, which is illuminated from below by laser light. The ZMW nanostructure provides excitation confinement in the zeptoliter (10^{-21} liter) regime, enabling detection of individual phospholinked nucleotide substrates against the bulk solution background as they are incorporated into the DNA strand by the polymerase. (**B**) Schematic event sequence of the phospholinked dNTP incorporation cycle, with a corresponding expected time trace of detected fluorescence intensity from the ZMW. (1) A phospholinked nucleotide forms a cognate association with the template in the polymerase active site, (2) causing an elevation of the fluorescence output on the corresponding color channel. (3) Phosphodiester bond formation liberates the dye-linker-pyrophosphate product, which diffuses out of the ZMW, thus ending the fluorescence pulse. (4) The polymerase translocates to the next position, and (5) the next cognate nucleotide binds the active site beginning the subsequent pulse.

Eid et al. 2009

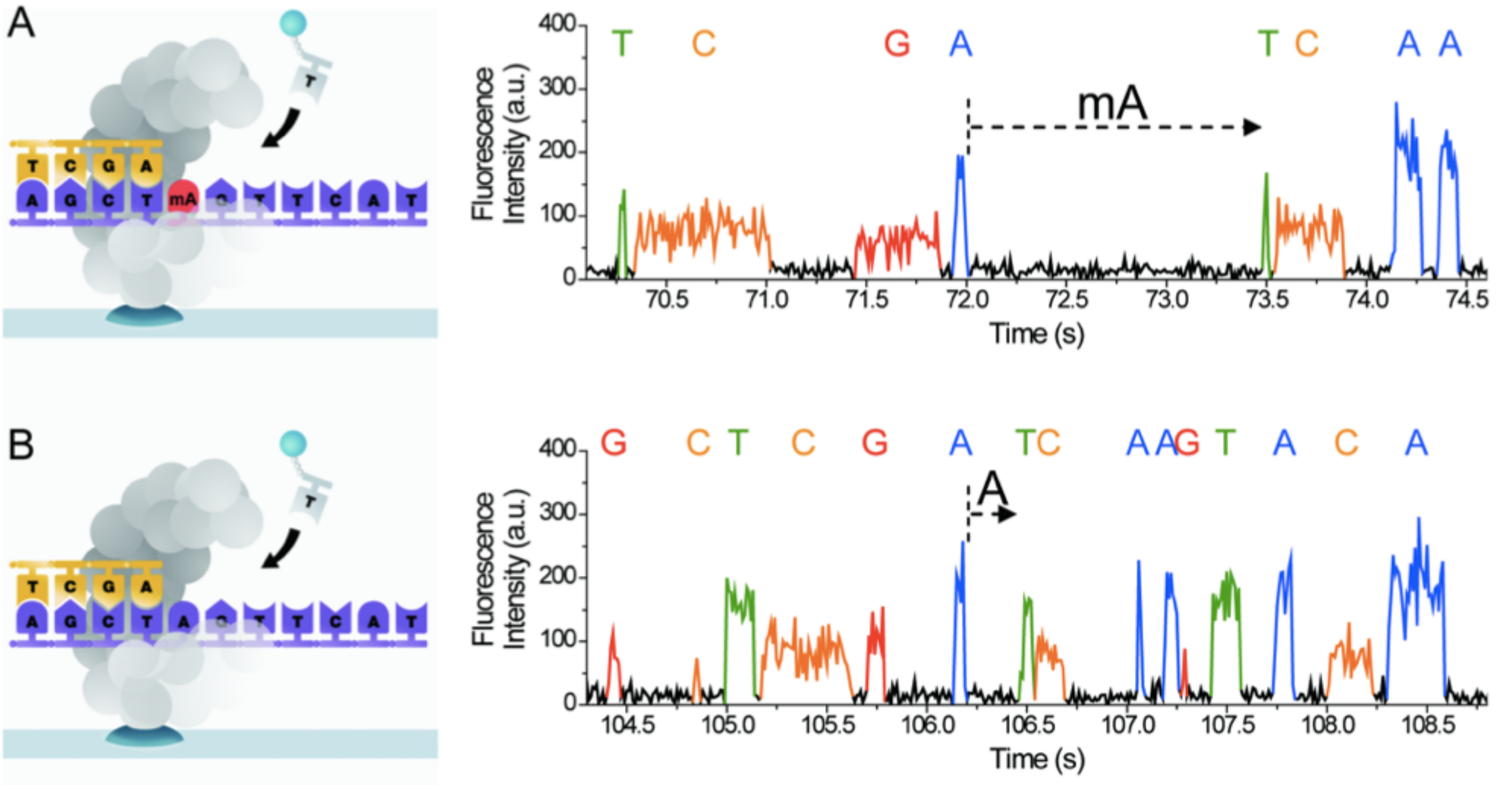
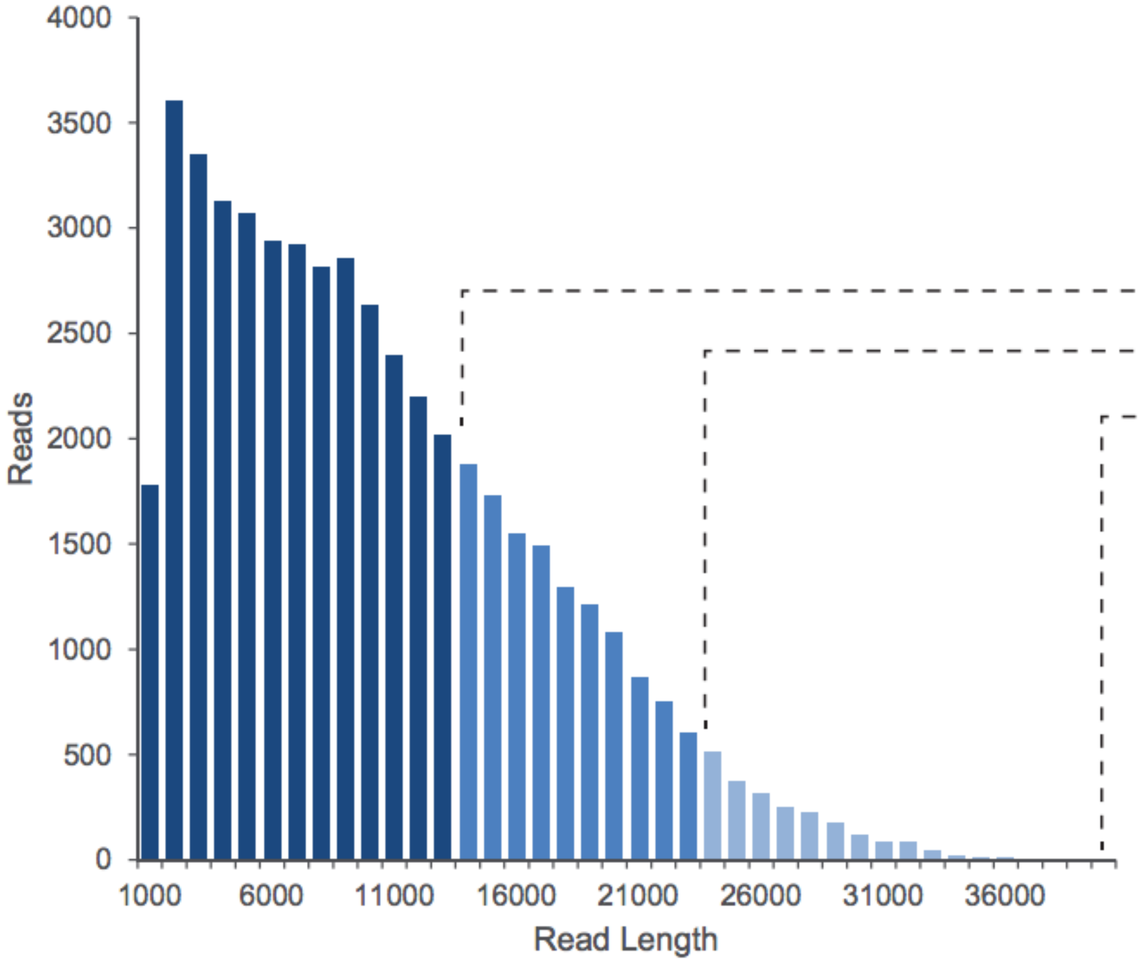


Figure 1.
Principle and corresponding example of detecting DNA methylation during SMRT sequencing. (**a**) Schematics of polymerase synthesis of DNA strands containing a methylated (top) or unmethylated (bottom) adenosine. (**b**) Typical SMRT sequencing fluorescence traces from these templates. Letters above the fluorescence trace pulses indicate the identity of the nucleotide incorporated into the growing complementary strand. The dashed arrows indicate the IPD before incorporation of the cognate T, and, for this typical example, the IPD is ~5× larger for mA in the template compared to A.

Flushberg et al. 2010

P6-C4 Chemistry



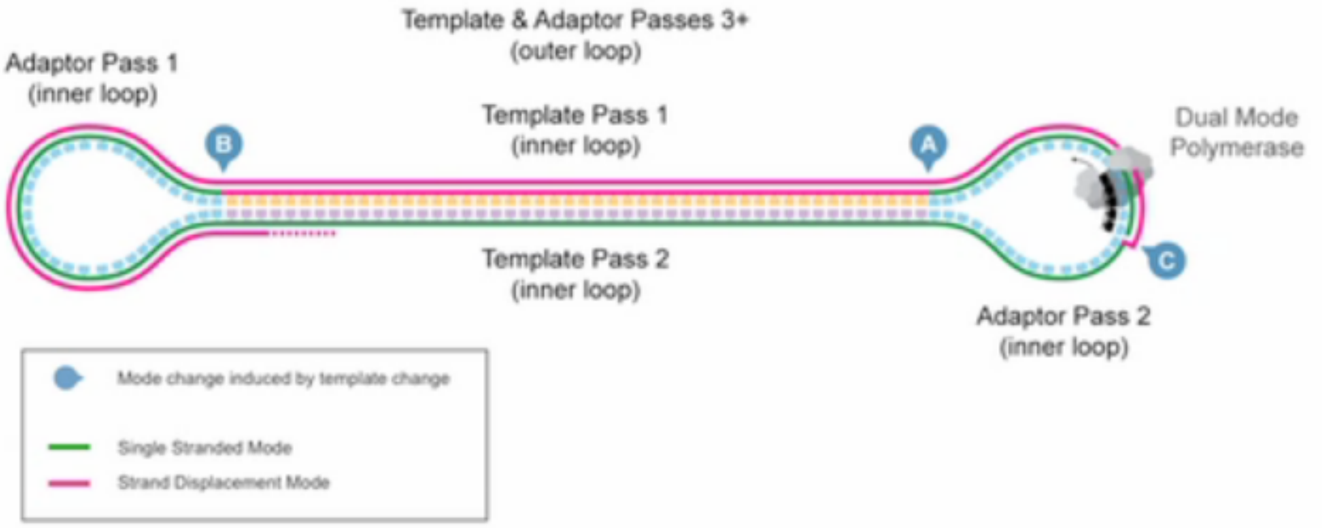
Half of data in reads: > 14 kb

Top 5% of reads: > 24 kb

Maximum read length: > 40 kb

Data per SMRT® Cell: 500 Mb – 1 Gb

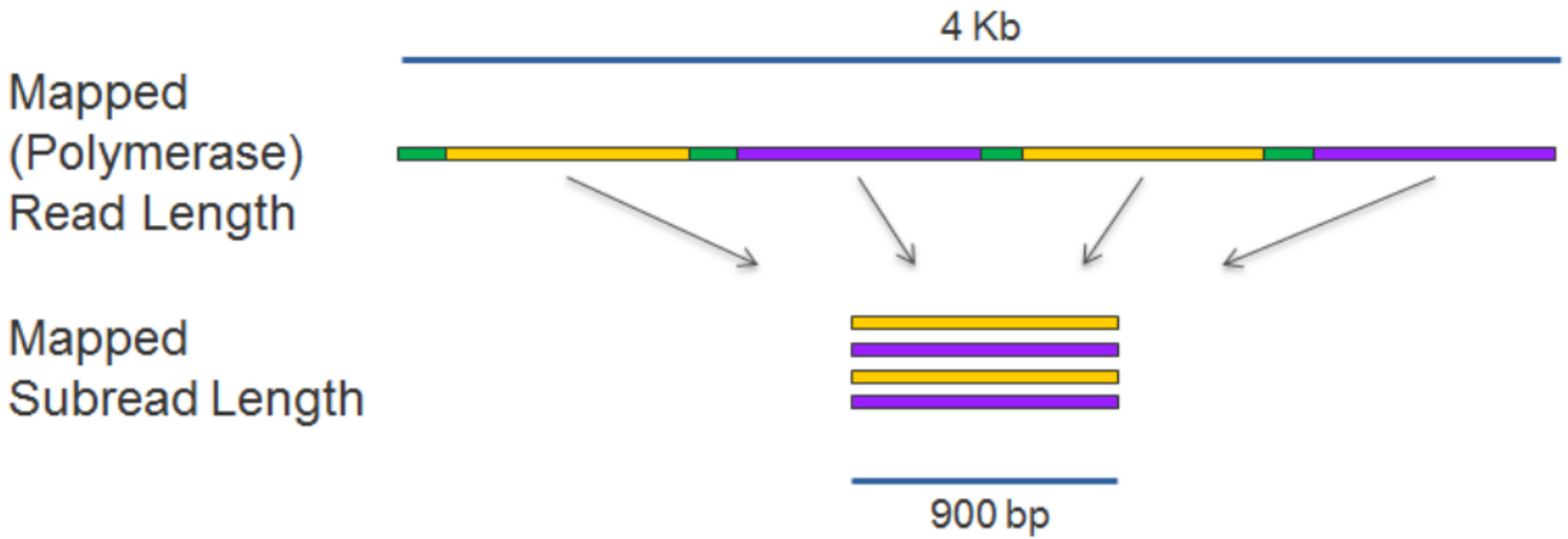
Based on data from a 20 kb size-selected *E. coli* library using a 4-hour movie.
Each SMRT Cell yields ~ 50,000 reads.



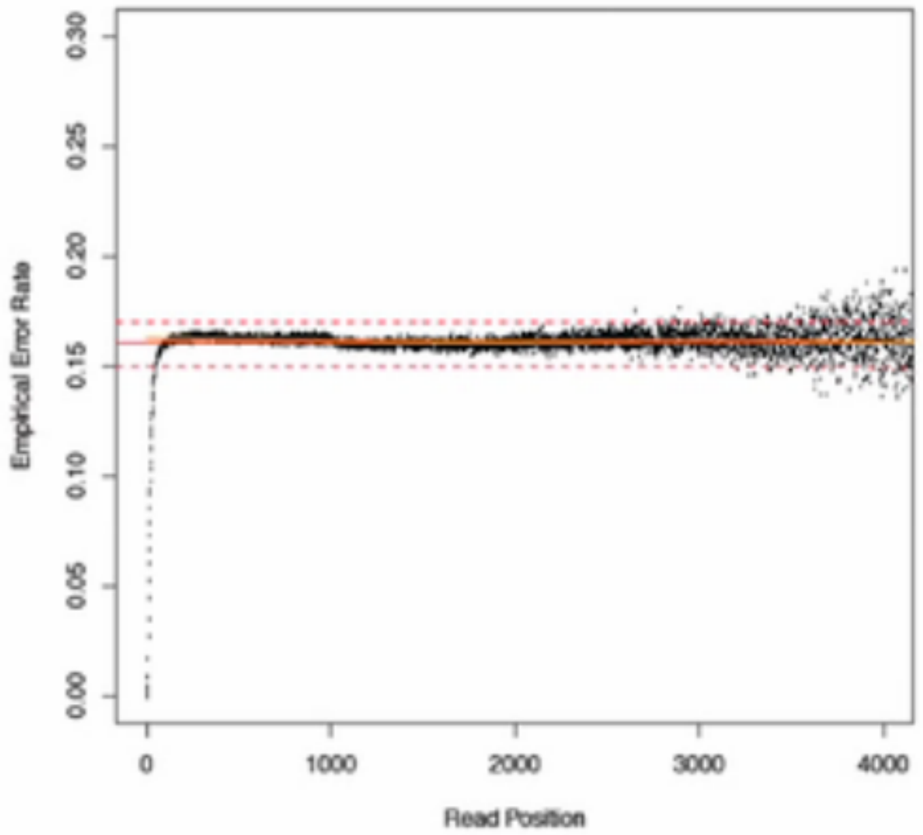
Template Pass 1	=	Strand Displacement
Adaptor Pass 1	=	Single Stranded
Template Pass 2	=	Single Stranded
Adaptor Pass 2	=	Single Stranded (some ds)
Template & Adaptor Passes 3+	=	Strand Displacement

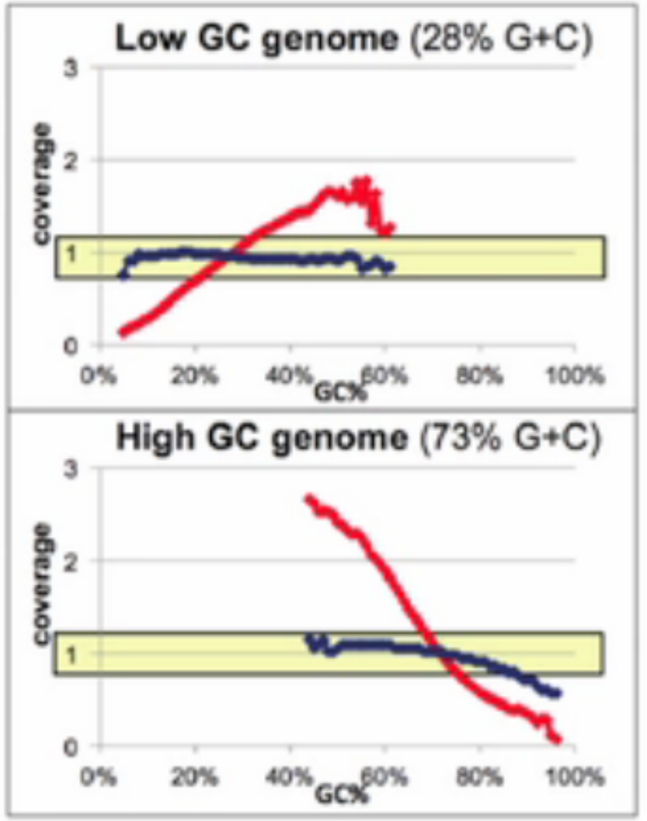
Types of runs

- CLR: Continuous Long Reads
- CCS: Circular Consensus Sequencing


4 Kb
Mapped
(Polymerase)
Read Length
Mapped
Subread Length
900 bp

Error profile and GC bias





Sergey Koren
Feng Chen, JGI

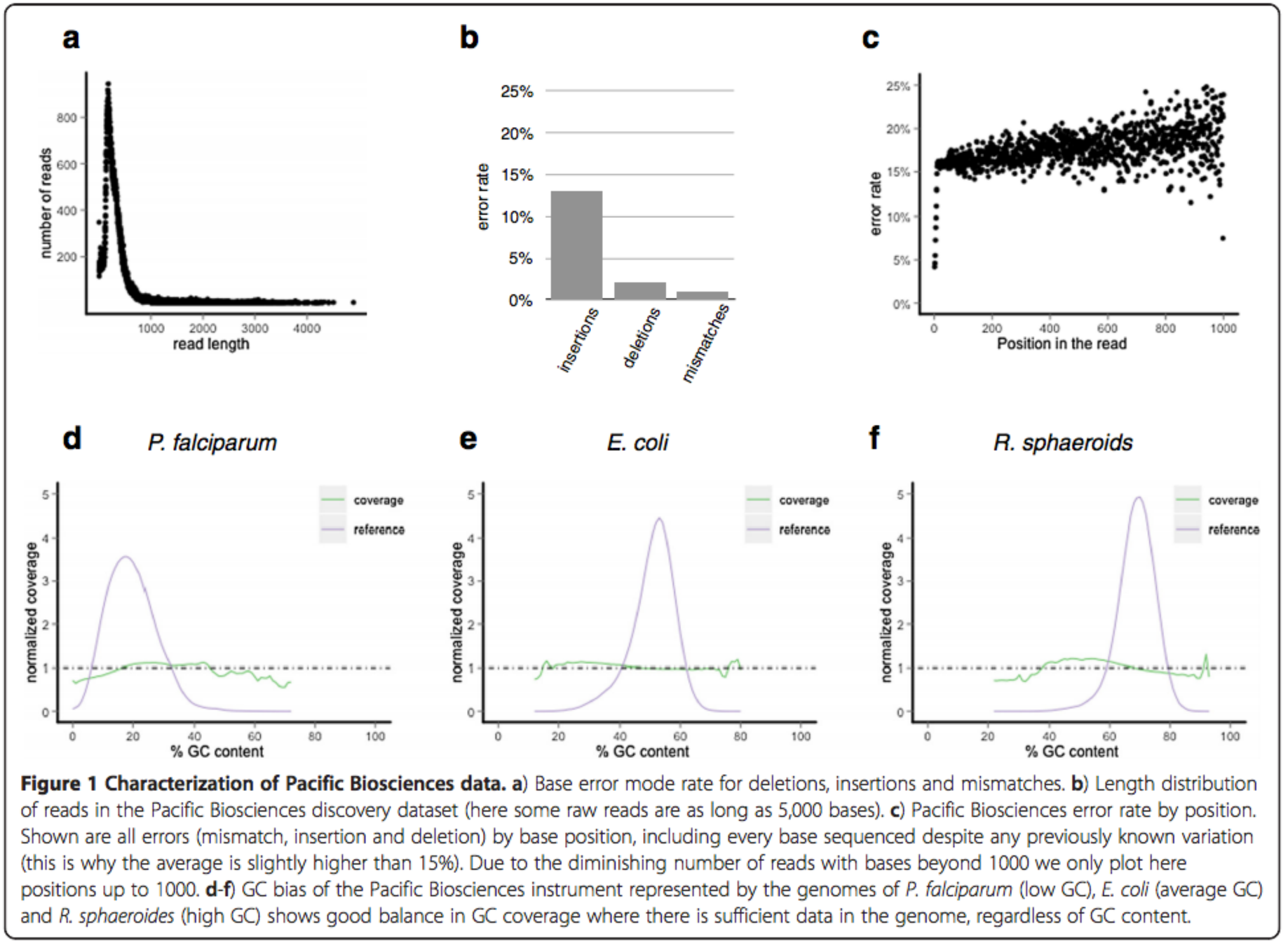


Figure 1 Characterization of Pacific Biosciences data. a) Base error mode rate for deletions, insertions and mismatches. **b**) Length distribution of reads in the Pacific Biosciences discovery dataset (here some raw reads are as long as 5,000 bases). **c**) Pacific Biosciences error rate by position. Shown are all errors (mismatch, insertion and deletion) by base position, including every base sequenced despite any previously known variation (this is why the average is slightly higher than 15%). Due to the diminishing number of reads with bases beyond 1000 we only plot here positions up to 1000. **d-f**) GC bias of the Pacific Biosciences instrument represented by the genomes of *P. falciparum* (low GC), *E. coli* (average GC) and *R. sphaeroides* (high GC) shows good balance in GC coverage where there is sufficient data in the genome, regardless of GC content.

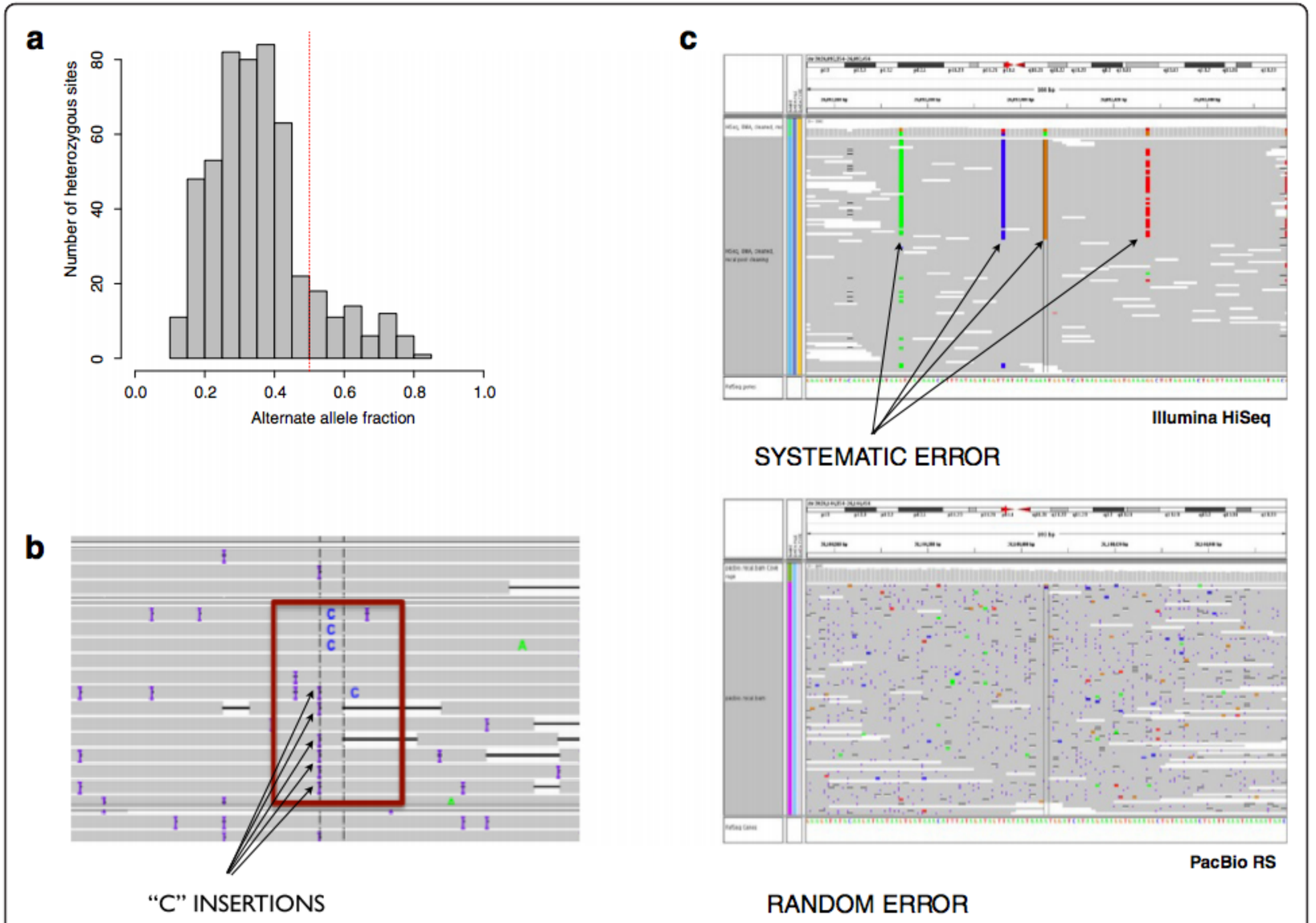


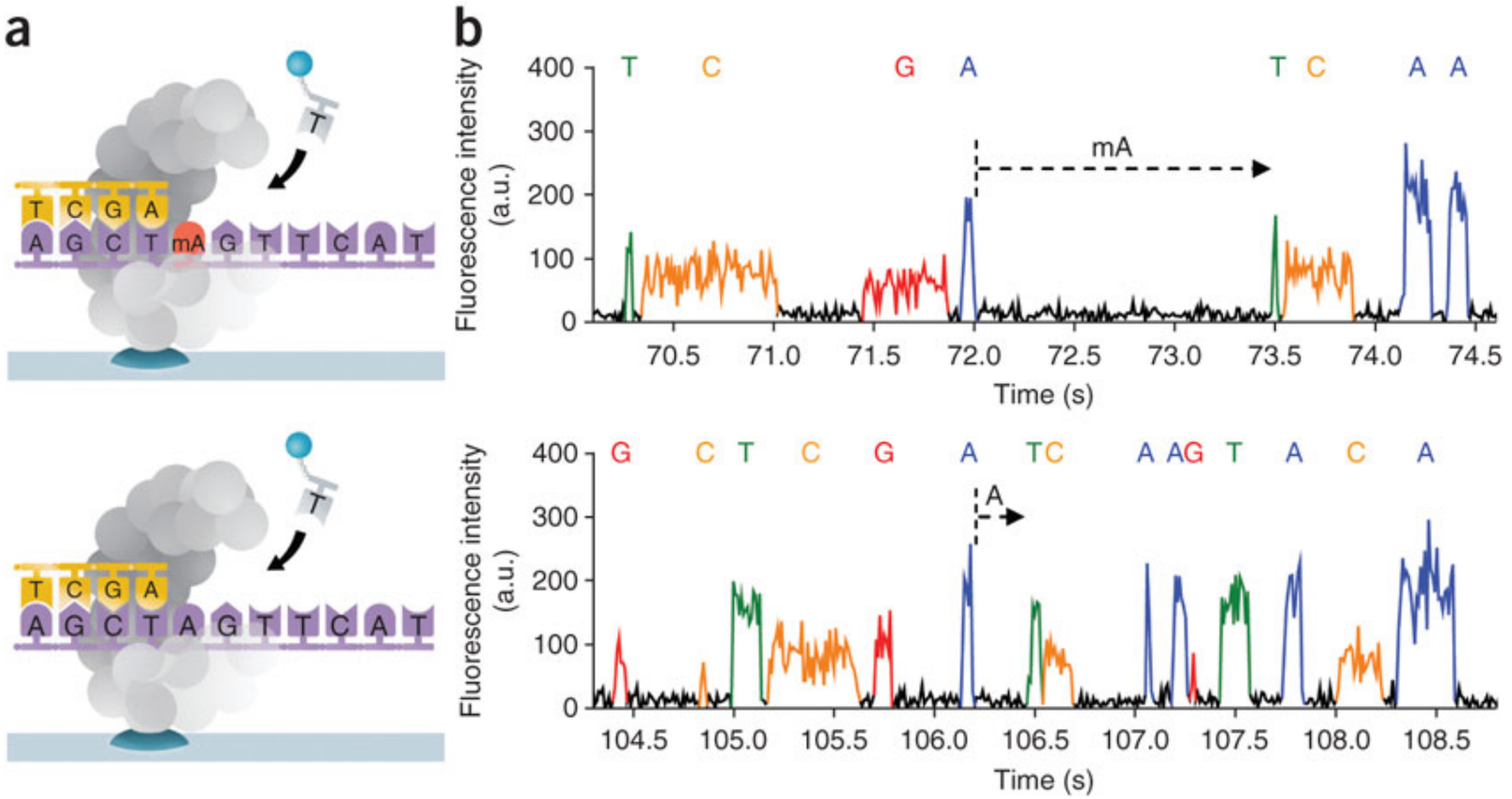
Figure 2 Error profile of Pacific Biosciences data. a) A chart showing the number of observations of the alternate allele in all heterozygous sites and how reference bias pulls the median significantly below the expected 0.5. This combination creates multiple possible alignments with the highest alignment score, allowing the aligner in some cases to hide the true alternate allele inside an insertion to maximize the alignment score at the cost of reference bias. **b**) IGV browser (http://www.broadinstitute.org/igv/) screenshot of the validation dataset showing an example of a case of aligner-created reference bias on Pacific Biosciences RS data. The true SNPs (C) are correctly called in individual reads. **c**) An IGV browser[18,19] screen snapshot of a region in the discovery dataset where Illumina HiSeq data suffers from context specific errors that makes it appear as a true heterozygous site whereas Pacific Biosciences RS data (with errors nearly random, though more frequent) clearly shows no event in this region.

Carneiro et al. 2012

Loading regimes

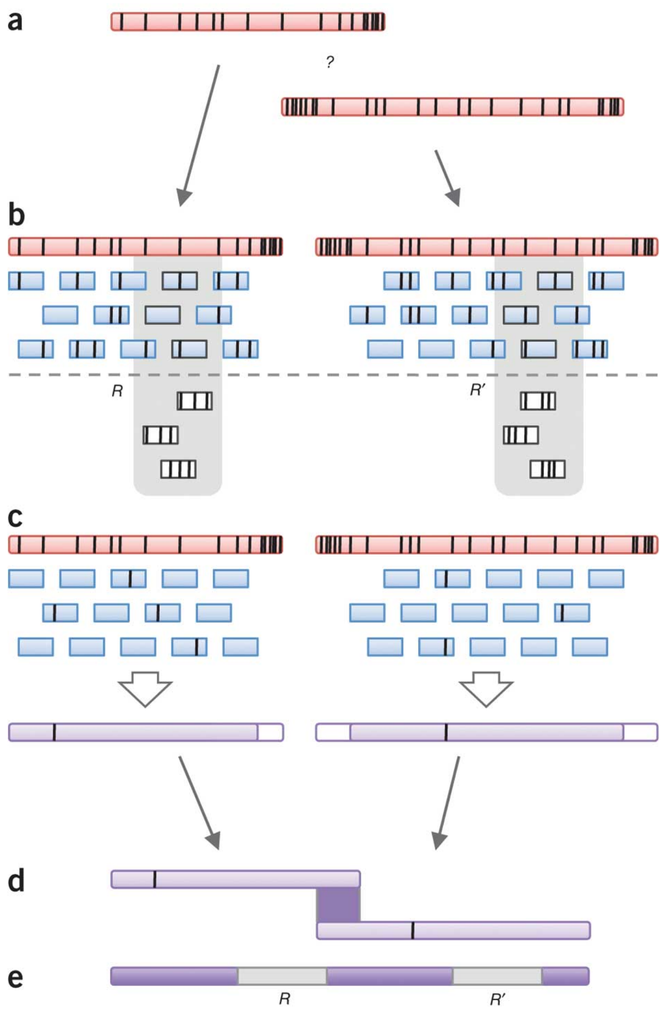
- Diffusion loading
- Magnetic bead loading

Modification detection

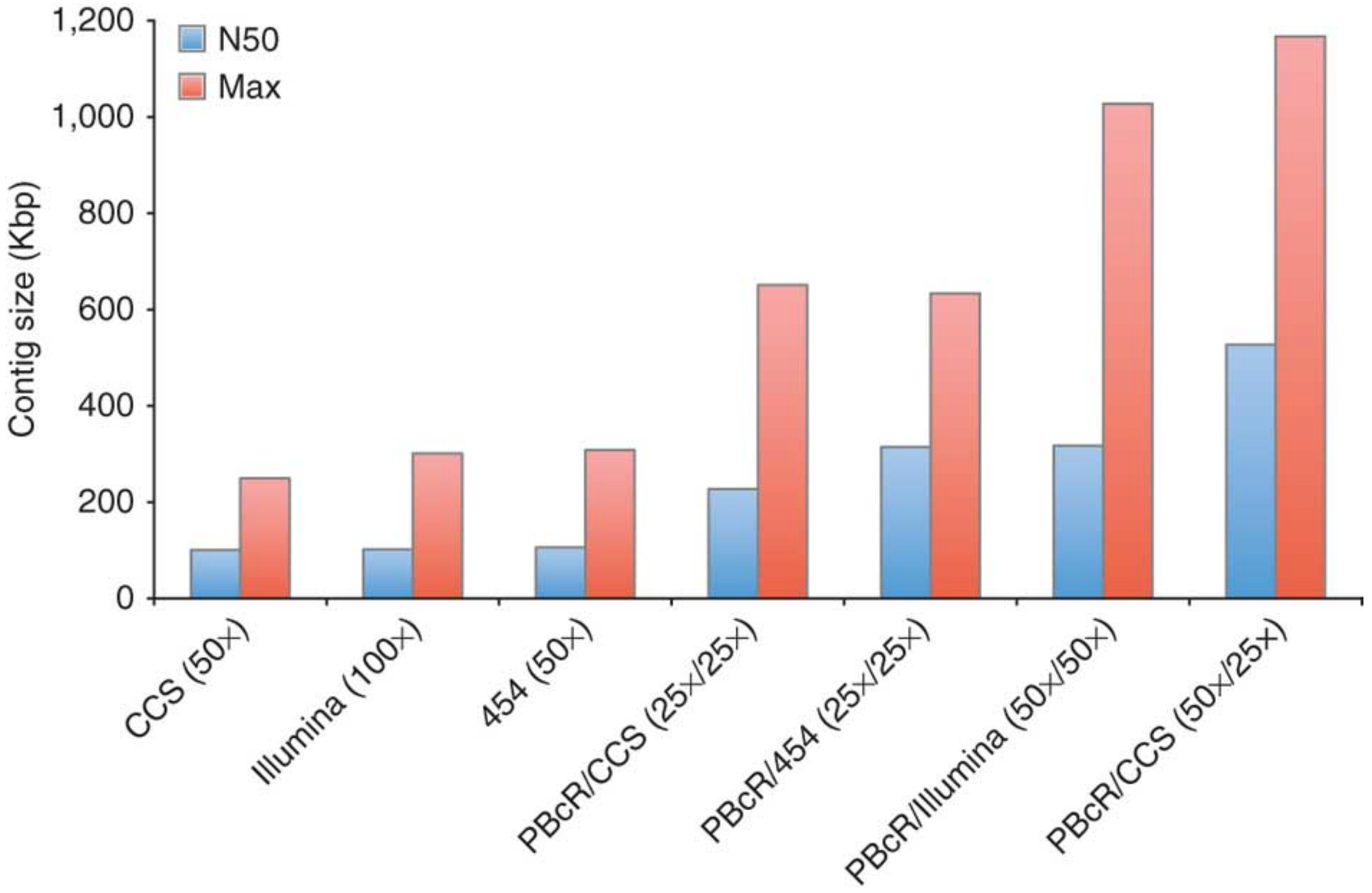


Flusberg et al. 2010

Error correction

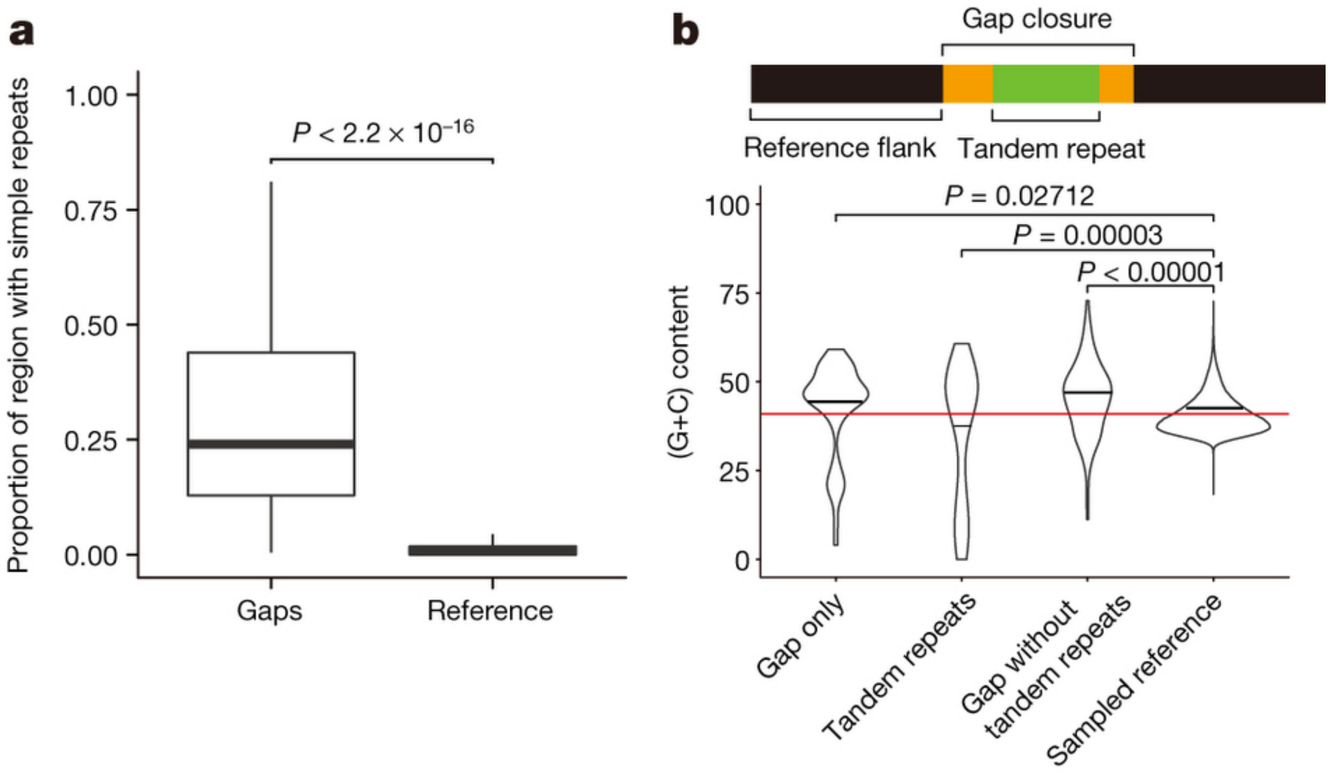


Koren et al. 2012



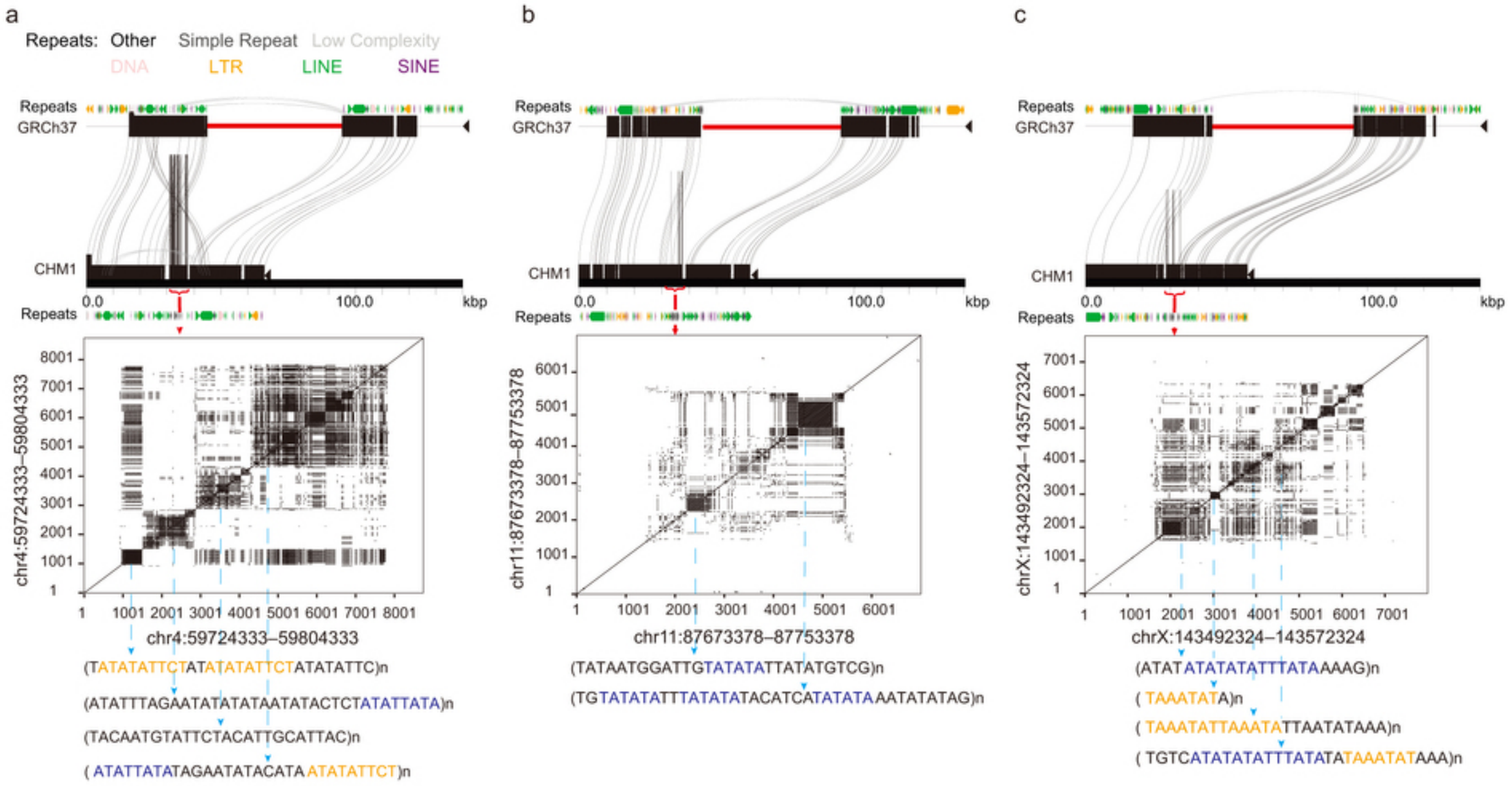
Koren et al. 2012

Closing gaps



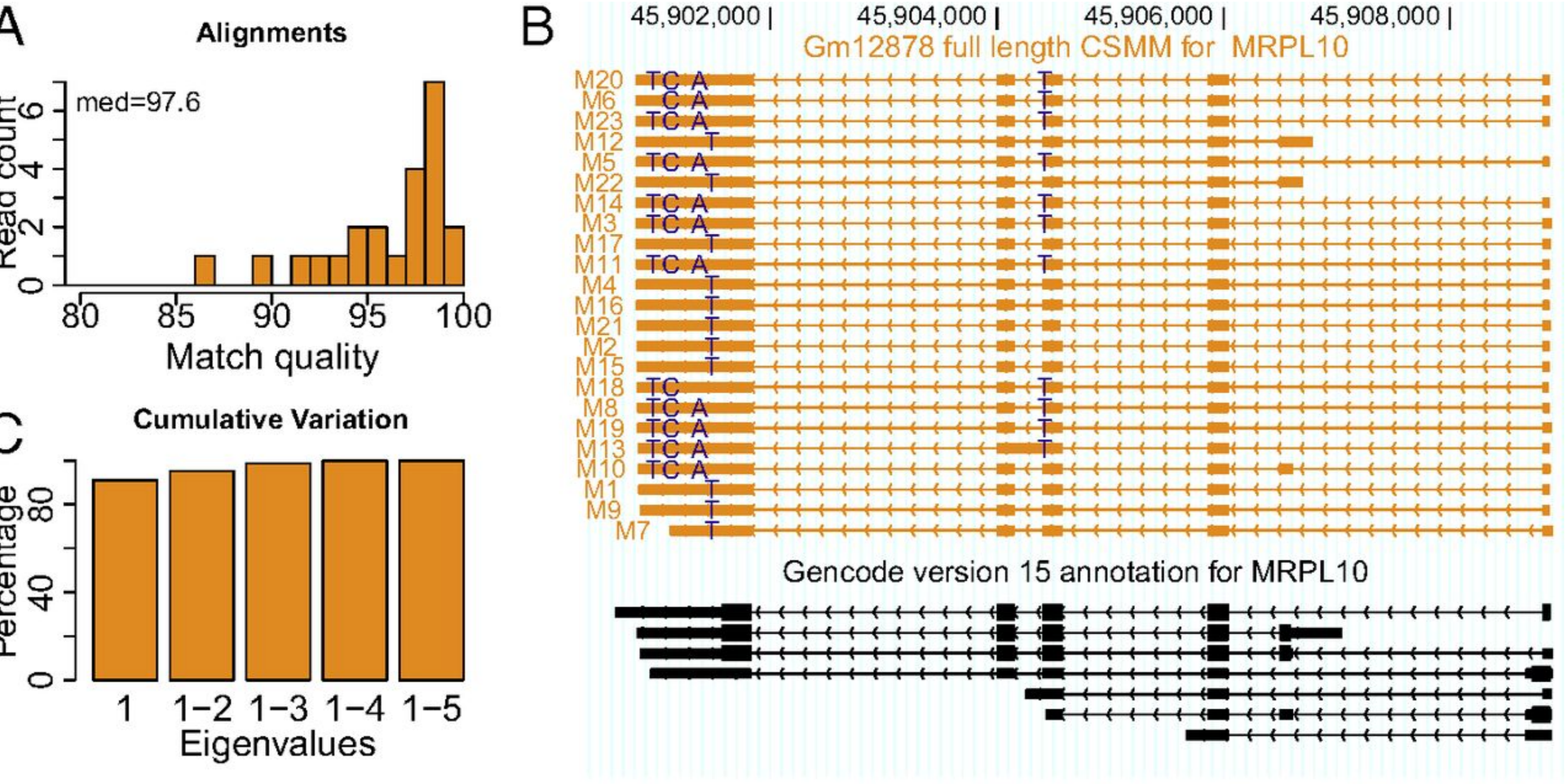
Chaisson et al. 2015

Closing gaps



Chaisson et al. 2015

Transcript sequencing



Tilgner et al. 2014