

2 - Sequencing Technologies

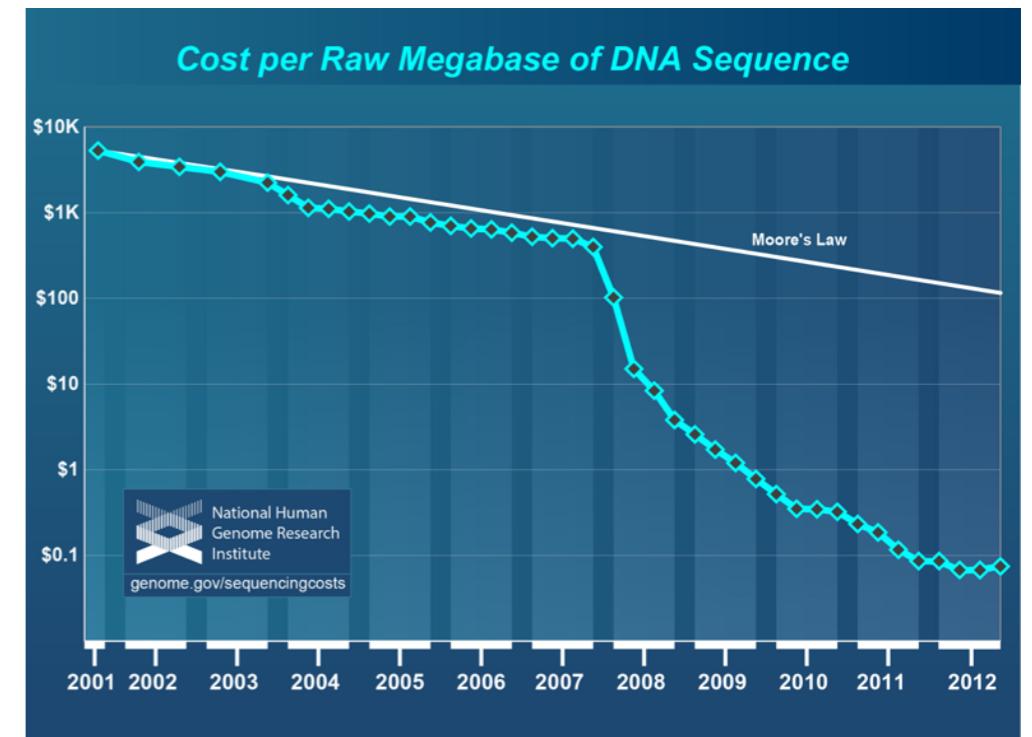
Tuesday afternoon

*Bernardo J. Clavijo
Richard Smith-Unna
Gonzalo Garcia / Jon Wright*

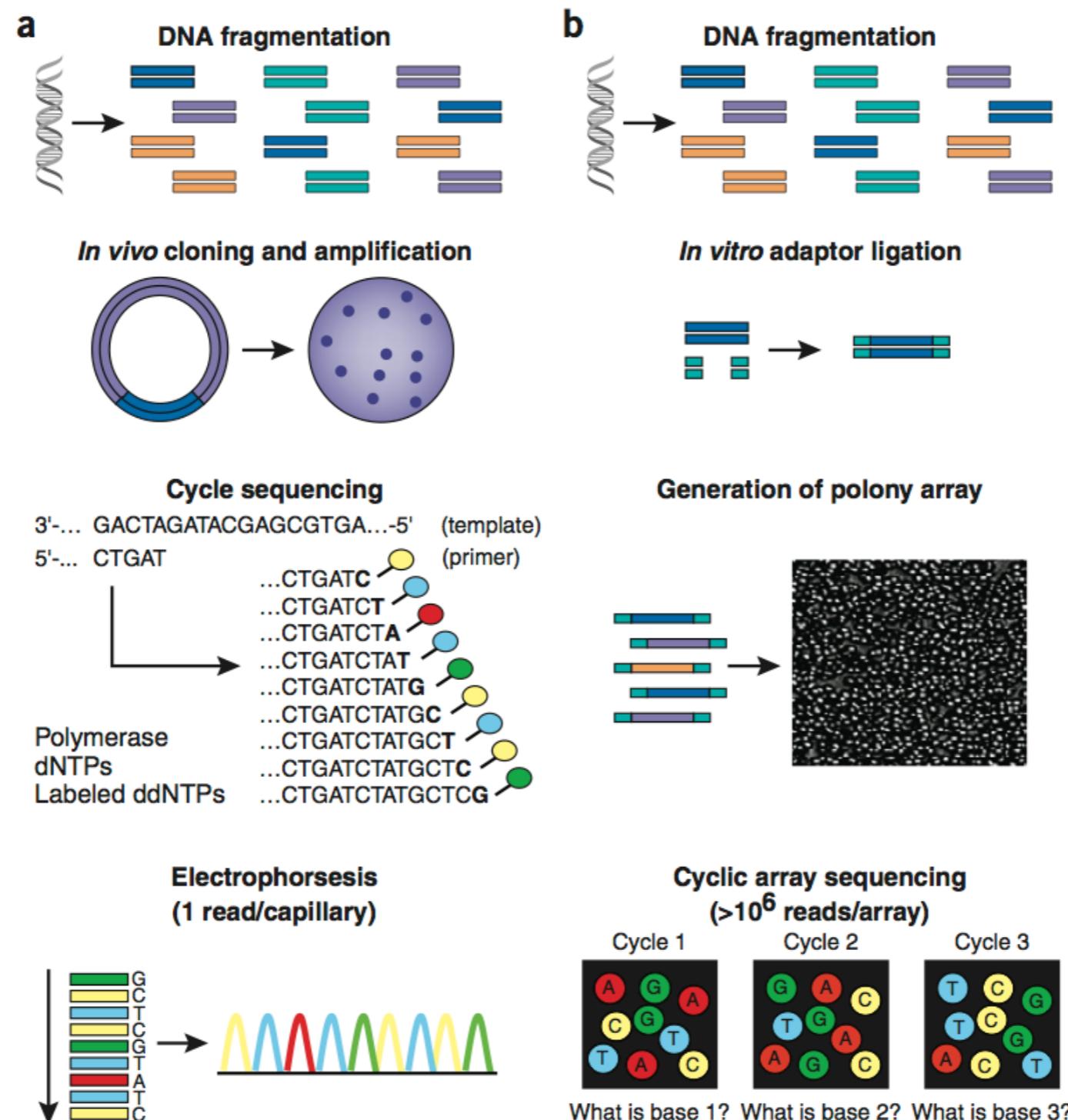


A brief history of DNA sequencing

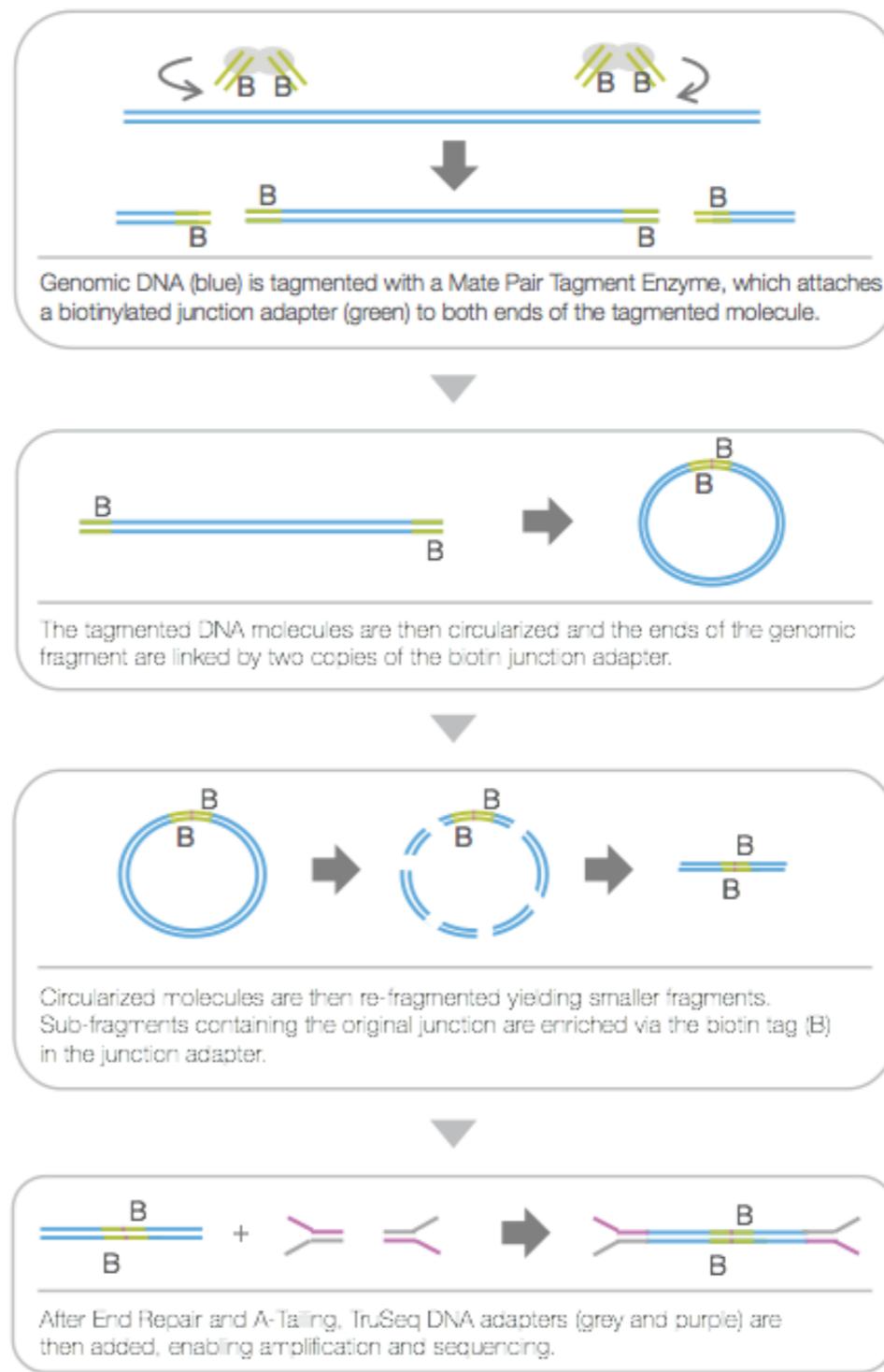
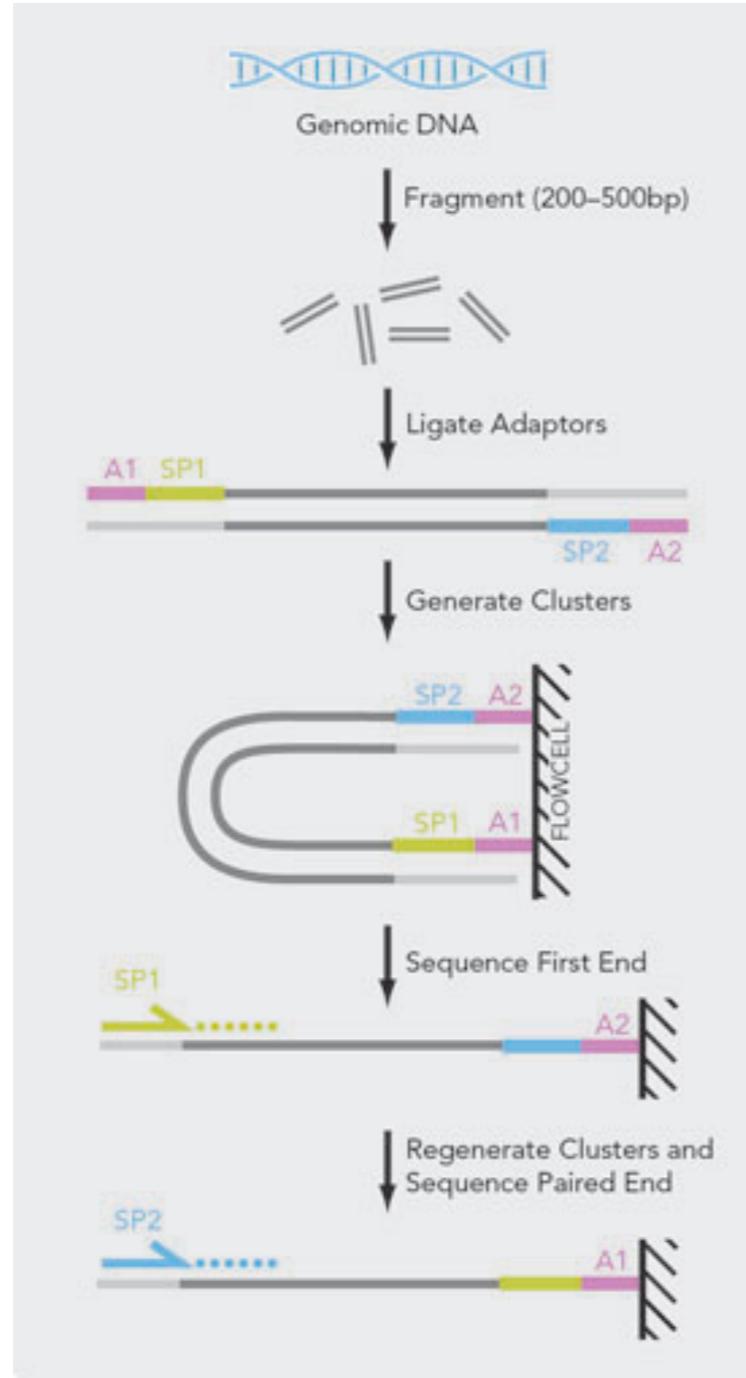
- 1953 – double helix structure, Watson & Crick
- 1977 – rapid DNA sequencing, Sanger
- 1977 – first full (5k) genome – bacteriophage Phi X
- Late 80s – first production ‘Sanger’ sequencers
- Mid 90s – DNA microarrays
- 2001 – draft human genome
- 2004 – first 454 pyrosequencing machine
- 2006 – first Solexa/Illumina sequencer
- 2011 – PacBio RS
- 2014 – Nanopore, Bionano
- 2015 – Dovetail, 10x



Next Generation Sequencing



Creating and Sequencing Paired Libraries



TGAC Sequencing Platforms



Illumina MiSeq x 3



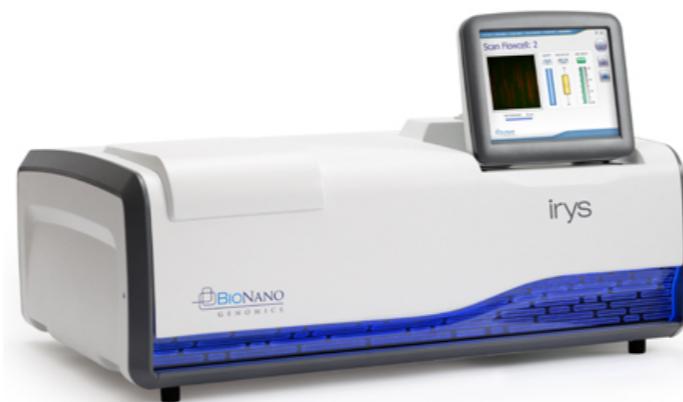
Illumina HiSeq x 3



PacBio RSII x 1



Oxford Nanopore
MinION x 1



BioNano Irys x 1



10x Genomics
Chromium x 1

Platforms compared

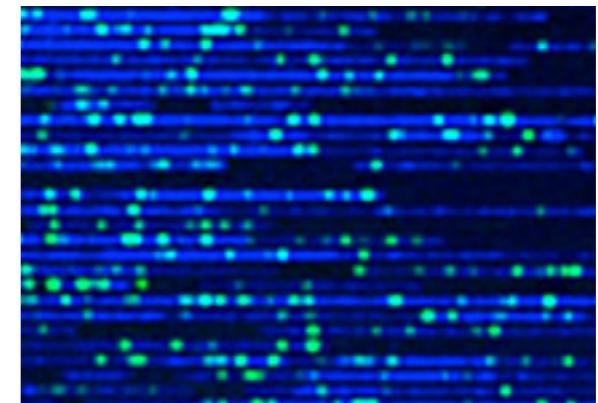
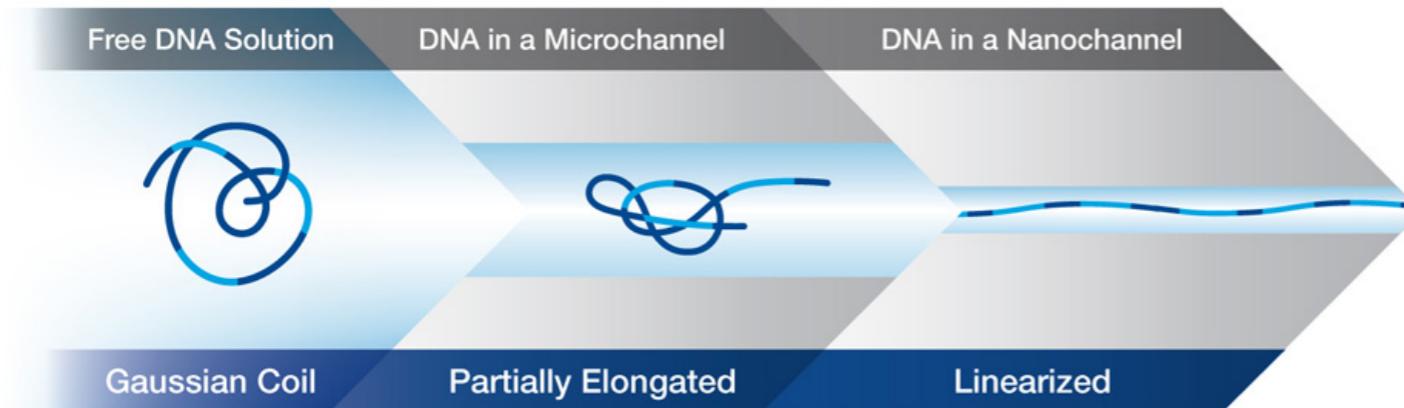
	METHOD	READ LENGTH	NUMBER OF READS	THROUGHPUT	RUN TIME	ACCURACY	APPROX. COST (£/Gb)
ILLUMINA HiSeq 2500 (High Output)	Sequencing by synthesis	Up to 125 bp PE	1.5 billion per flowcell	1 Tb	6 days	99.9%	£21
ILLUMINA HiSeq 2500 (Rapid)	Sequencing by synthesis	Up to 150 bp PE	300 million per flowcell	150 Gb	60 hours	99.9%	£36
ILLUMINA MiSeq	Sequencing by synthesis	Up to 300 bp PE	25 million per flowcell	15 Gb	55 hours	99.9%	£65
PACBIO RS II v4	Real time sequencing	Average 14 kb (up to 60 kb)	50 000 per SMRT cell	1 Gb per SMRT cell	4 hours	86%	£141
Nanopore MinION	Real time sequencing	8 to 100 kb	Up to 80kb per flow cell	500 Mb	24-48 hours	85-90%	-
BioNano Irys	Optical Mapping	300 kb	several Gb per hour	100 Gb	dependant on cycles	N/A	£10

The FASTQ file

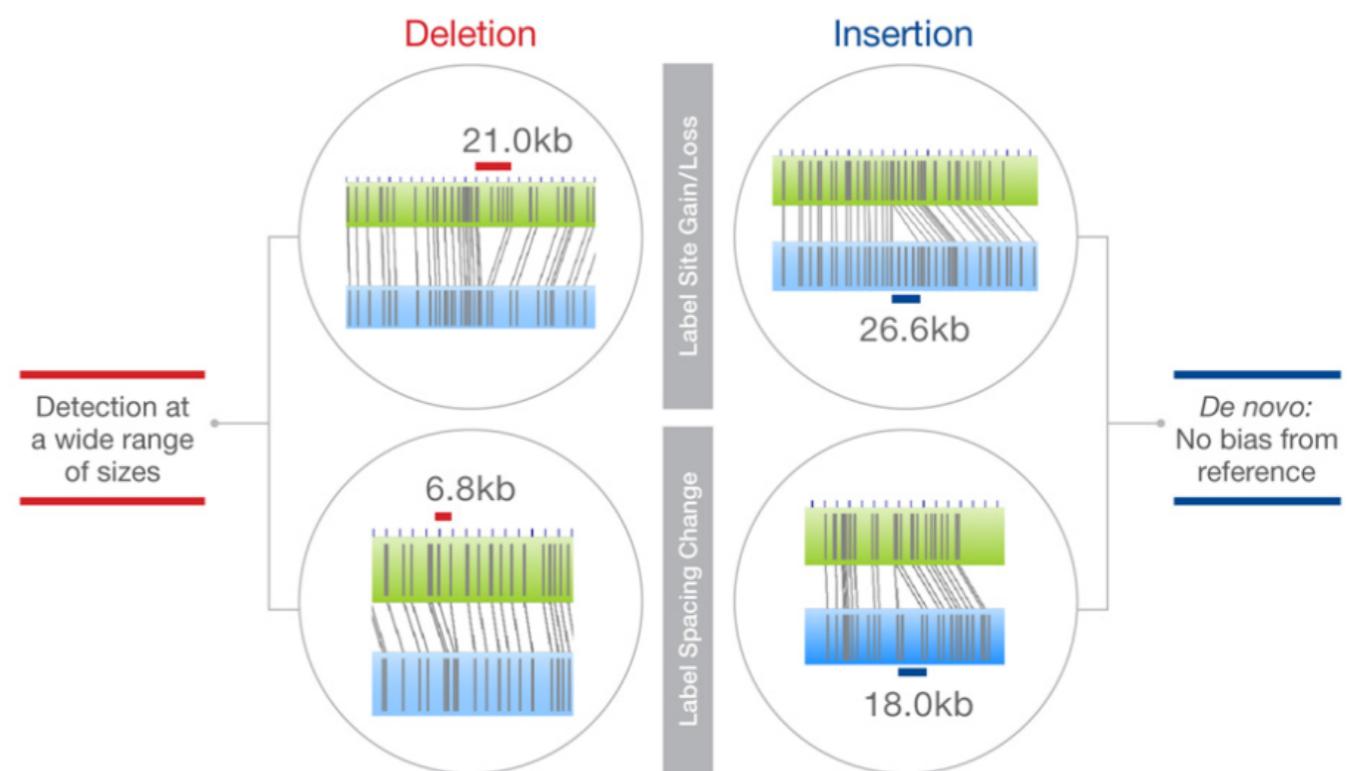
- 4 lines per read
 - Stores sequence and quality

Different Data for different Information

- Illumina paired end: a good and cheap way to get the motifs
- Long mate pairs: a hint at order and distances
- PacBio:
 - Long reads: longer, not very precise, motifs
 - Circular consensus reads: long, expensive, precise motifs
- Others include:
 - Optical maps (good positional information)
 - RNA-seq
 - Fosmid ends
 - Known deletion bins / markers / ESTs



Structural Variation Detection



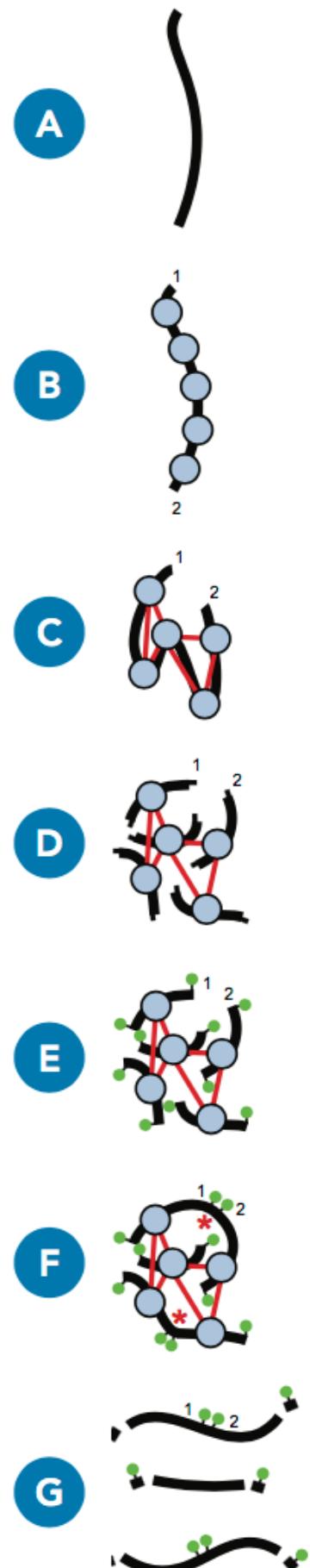


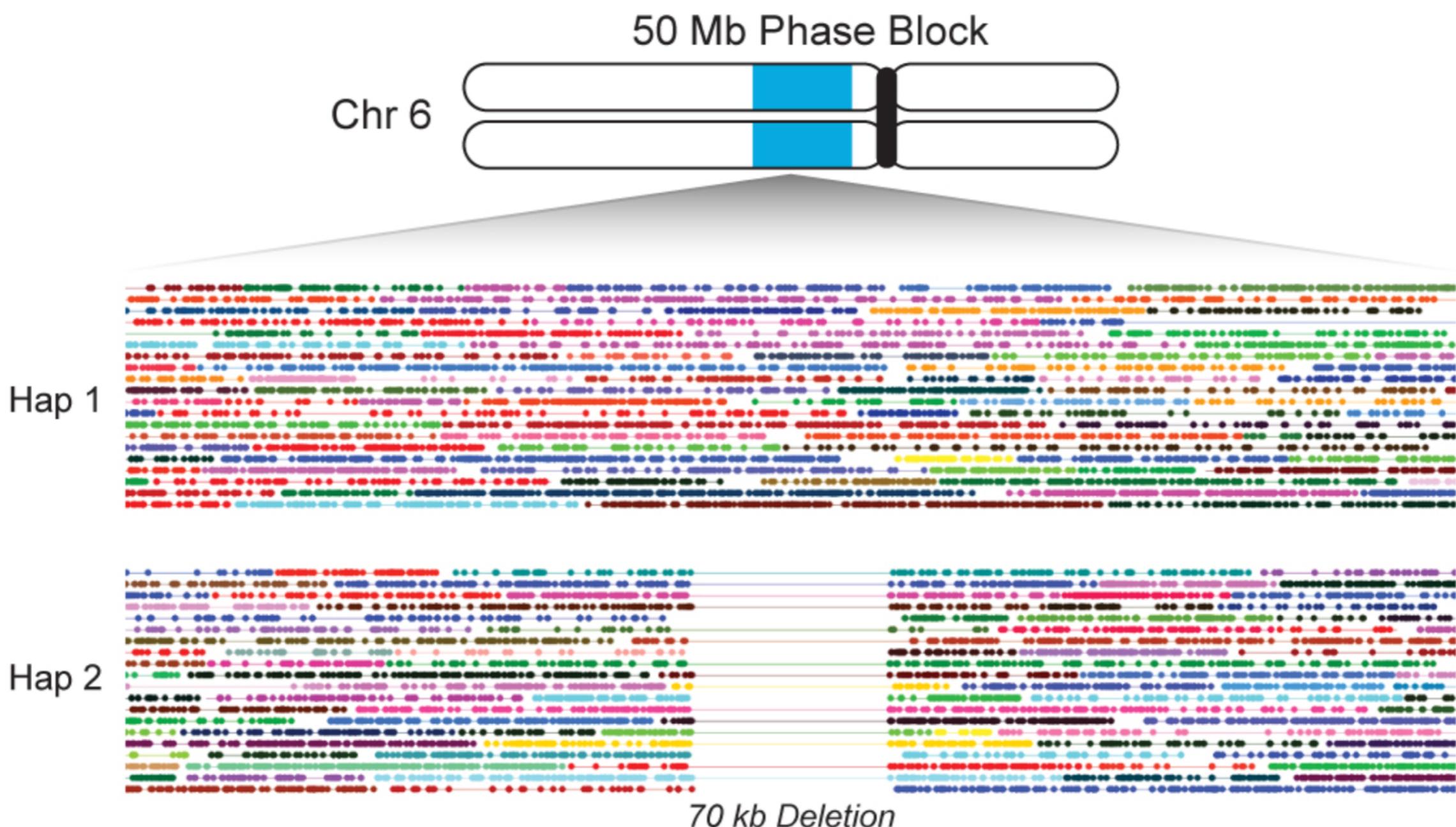
Dovetail GENOMICS

Table 1: A list of assemblies Dovetail has improved with its technologies.

Organism	Genome Size (Mbp)	Fold Physical Coverage (in 1-50 kbp bins)	Input N50 (kbp)	Final N50 (kbp)	Fold N50 Improvement
Vampire Bat	2,088	82x	5,498	13,814	3x
Cichlid	845	118x	1,208	3,395	3x
Potato	729	208x	755	5,868	8x
Butterfly	322	55x	143	3,707	26x
Pigeon	1,086	38x	70	3,739	54x
Prairie Chicken	897	111x	136	11,320	83x
Chimp	3,349	88x	72	9,969	138x
Human	3,086	39x	178	26,337	148x
Alligator	2,157	73x	81	21,540	265x

Figure 2:
An overview of the Chicago library preparation method.





Questions?

