Genome Resequencing

Rearrangements

SNPs, Indels CNVs

De novo genome Sequencing

Metagenomics

Exome Sequencing

RNA-seq Gene Expression

High Throughput Short Read Sequencing: Illumina

DNA Methylation

Splice Isoform Abundance

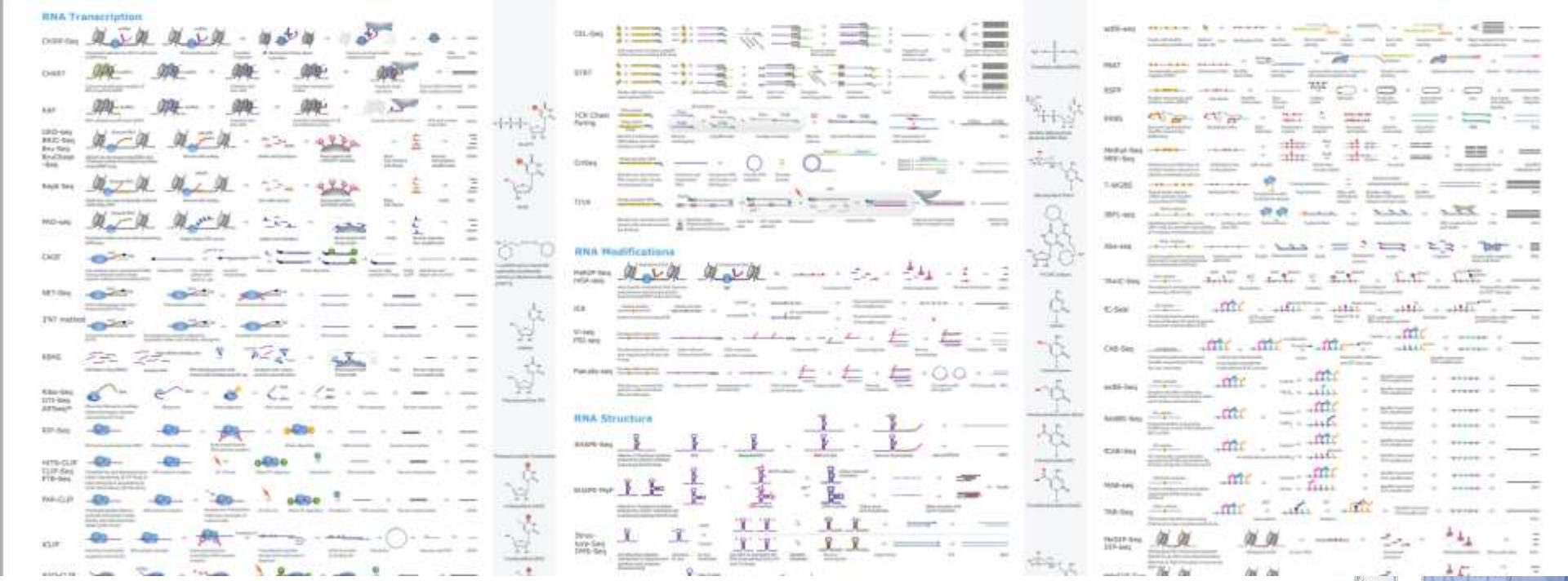
ChIP-SEQ

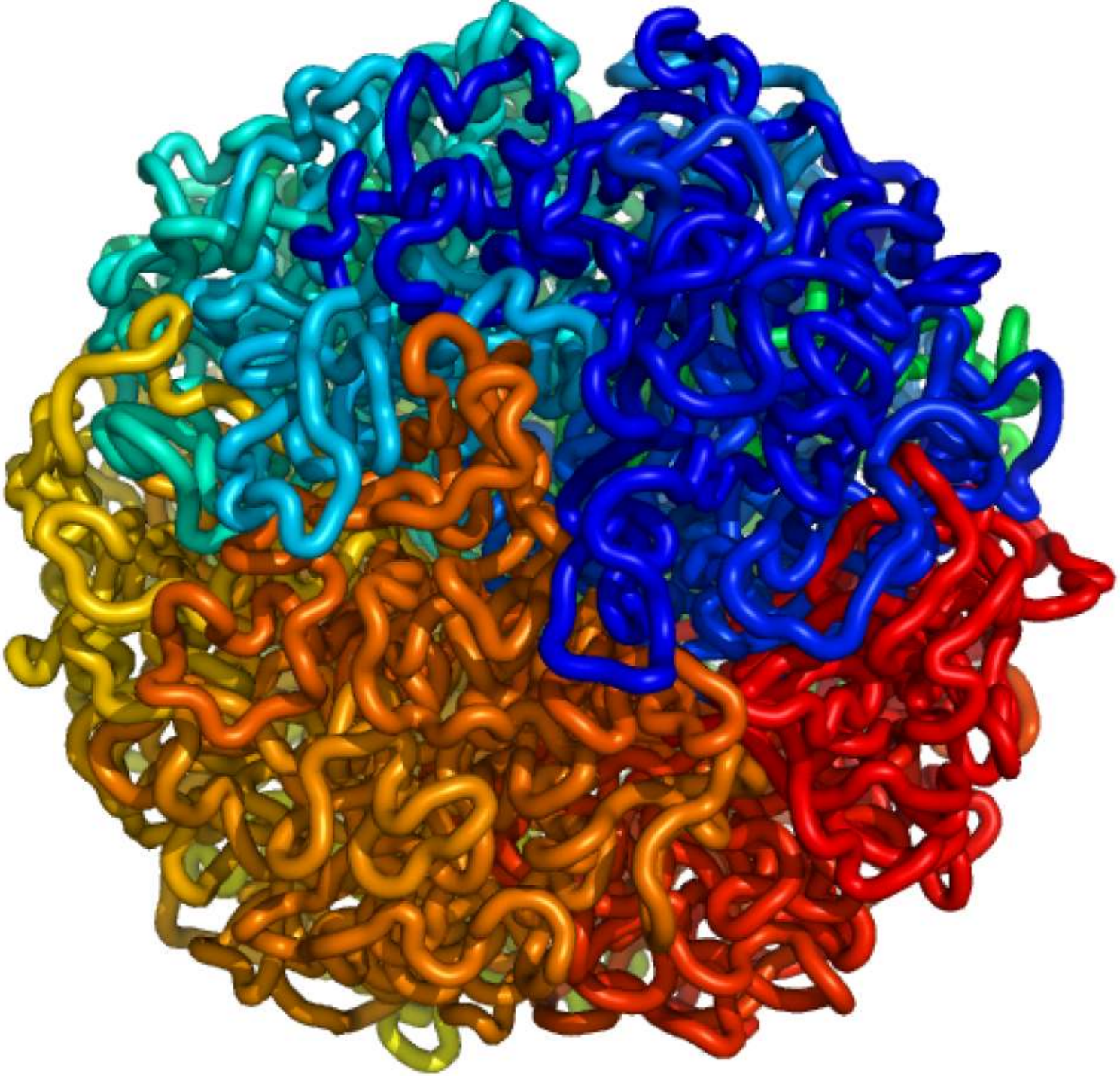
3D Organization

Genotyping

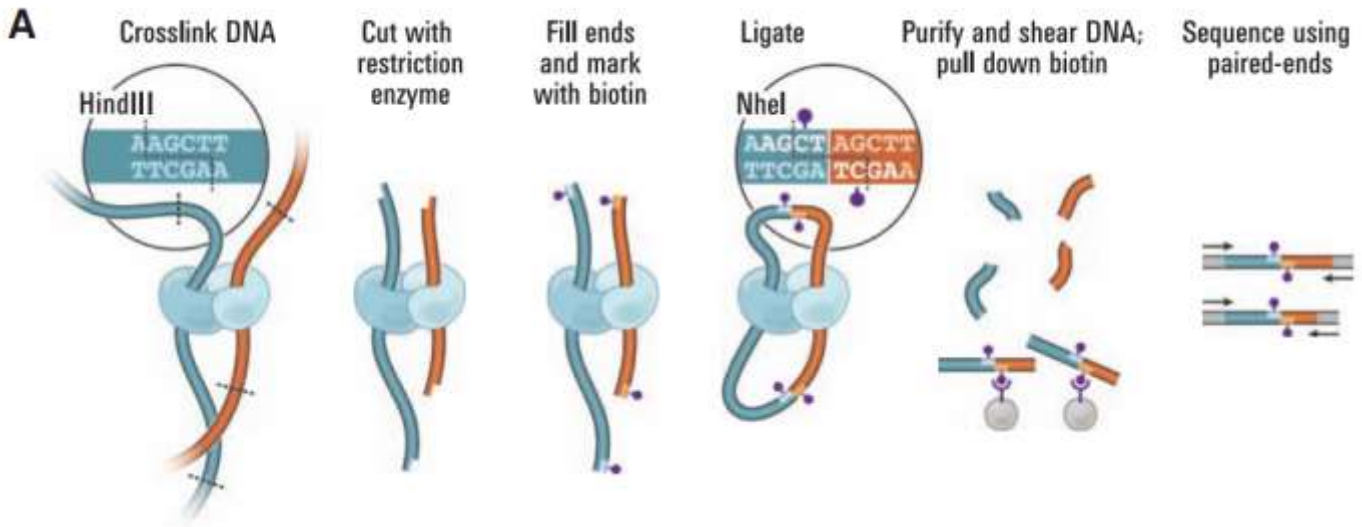
Small RNA

For all you seq...



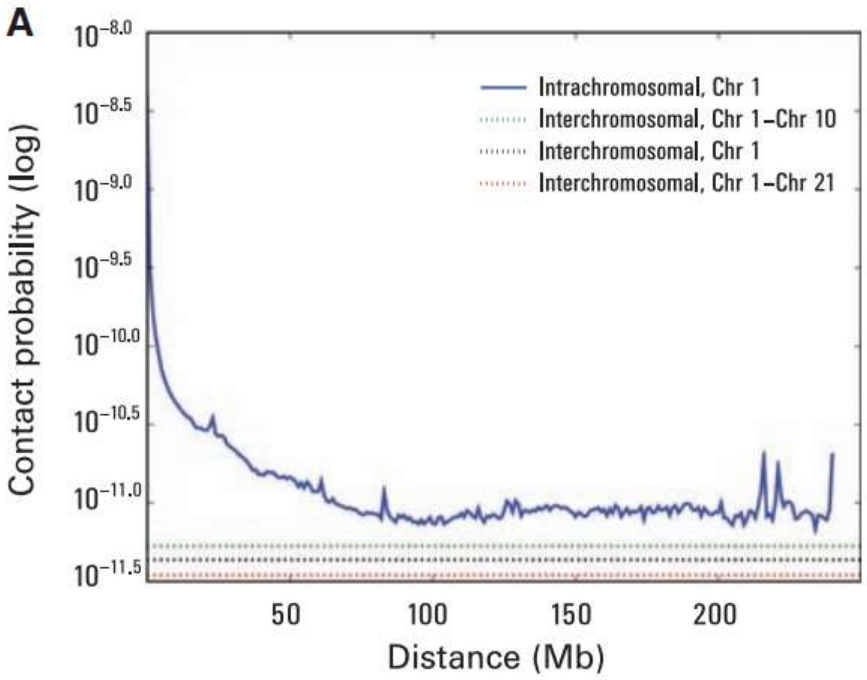


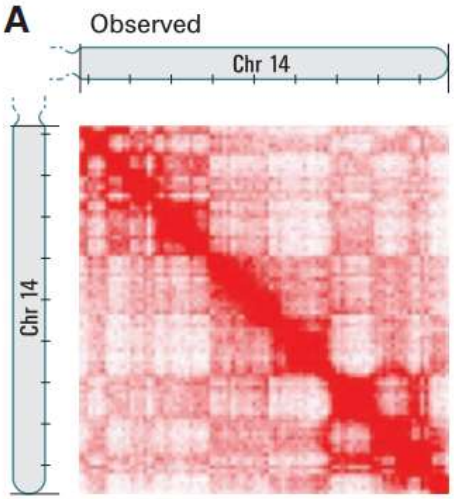
Hi-C



Lieberman-Aiden 2010

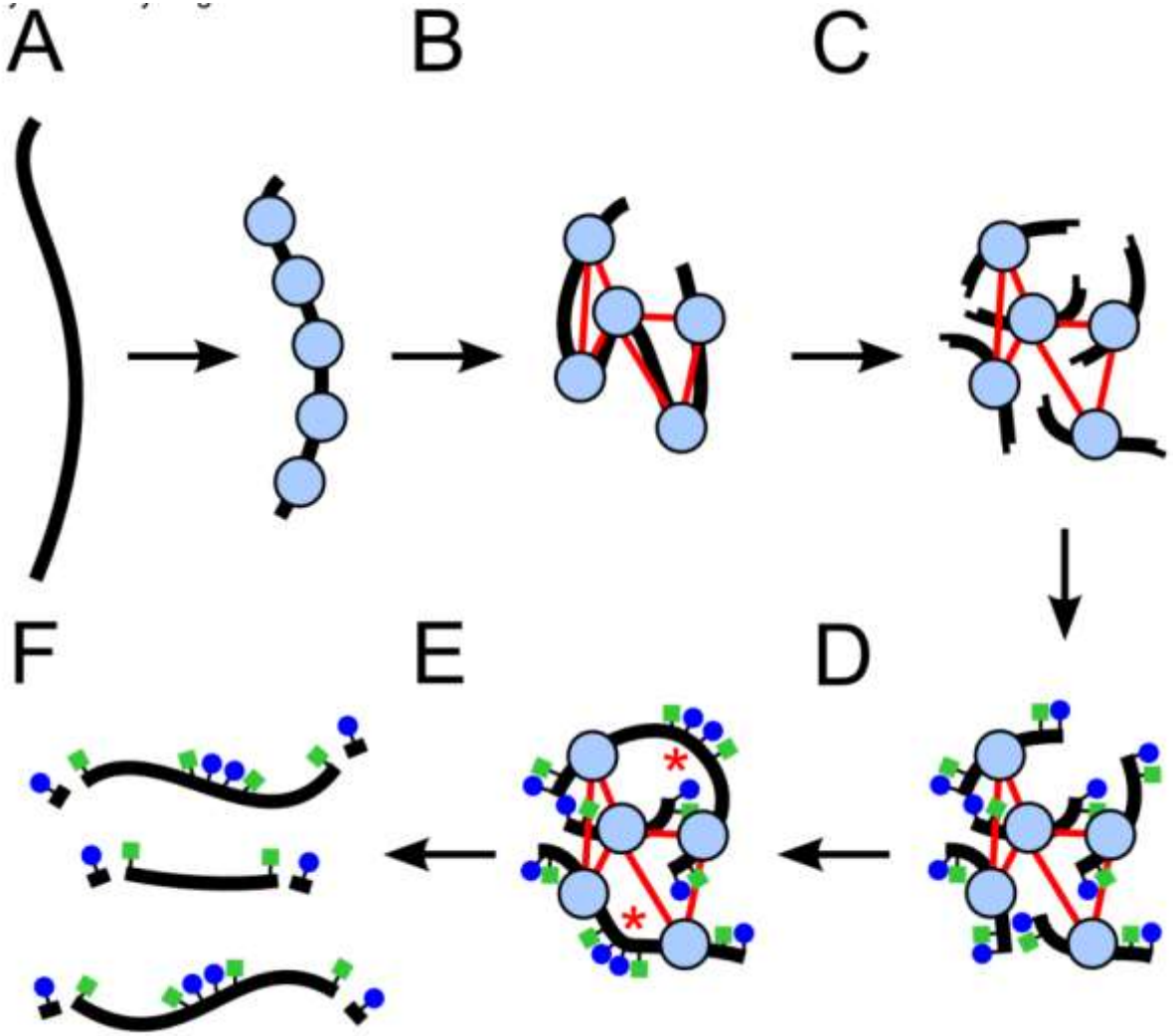
Hi-C



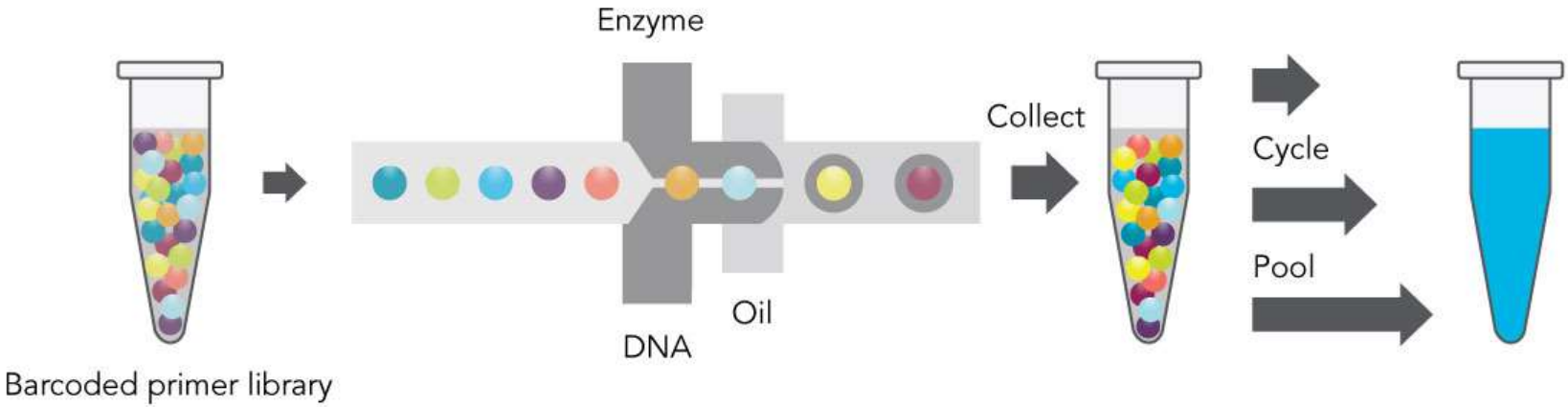


Lieberman-Aiden 2010

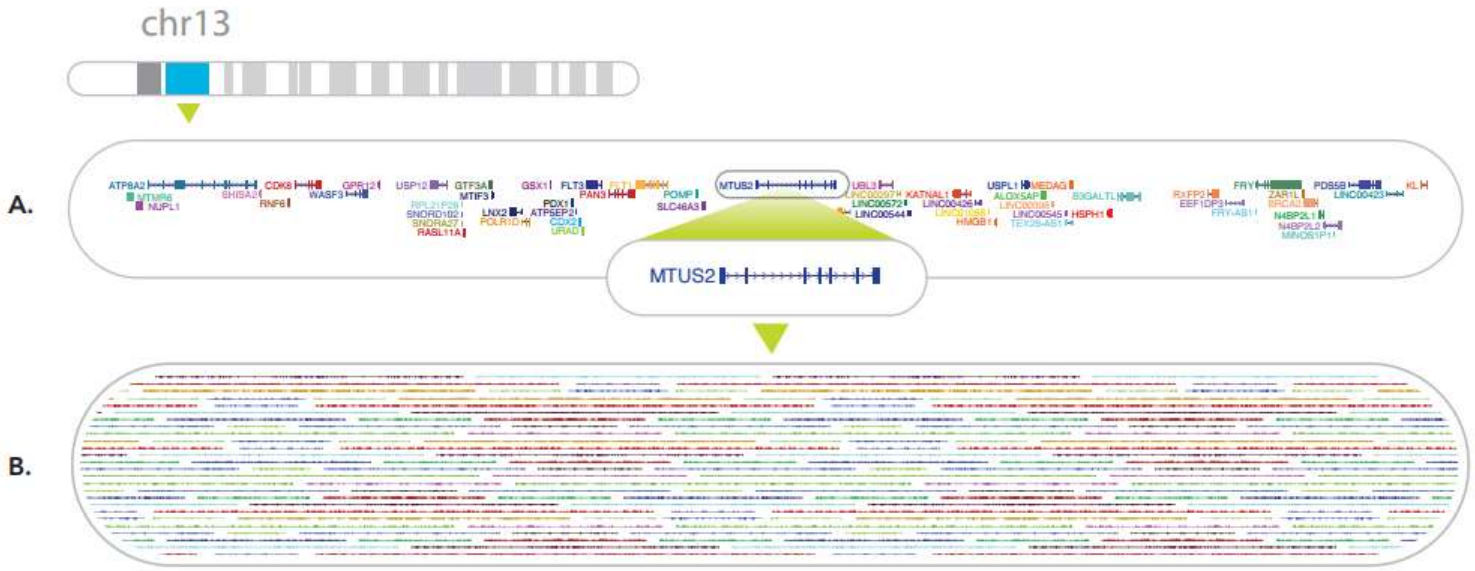
Dovetail Sequencing (Putnam 2015)



10X Genomics

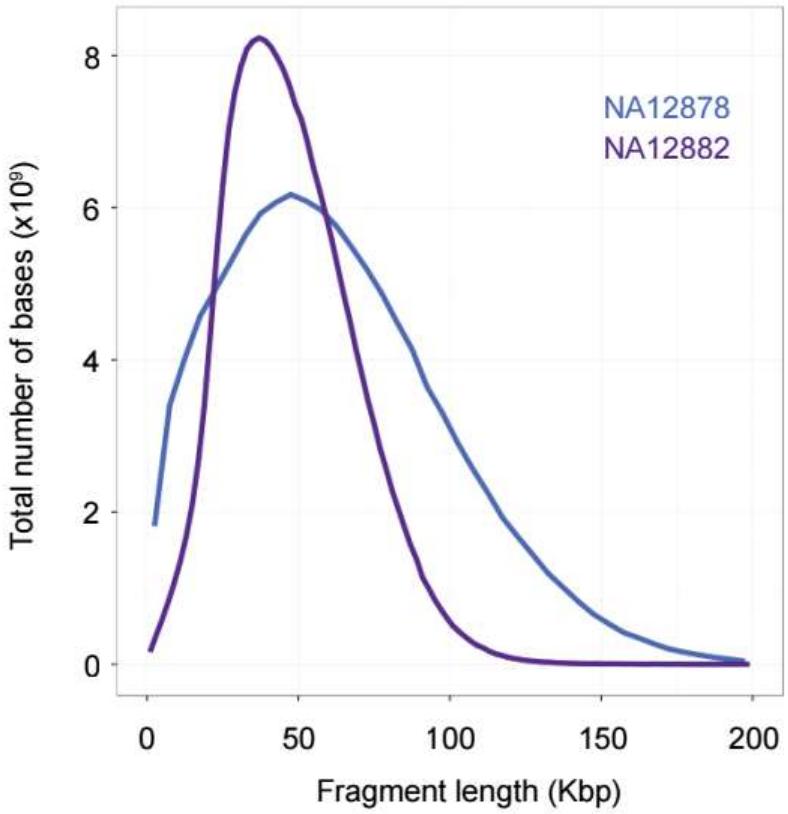

Enzyme
Collect
Cycle
Pool
Oil
DNA
Barcoded primer library

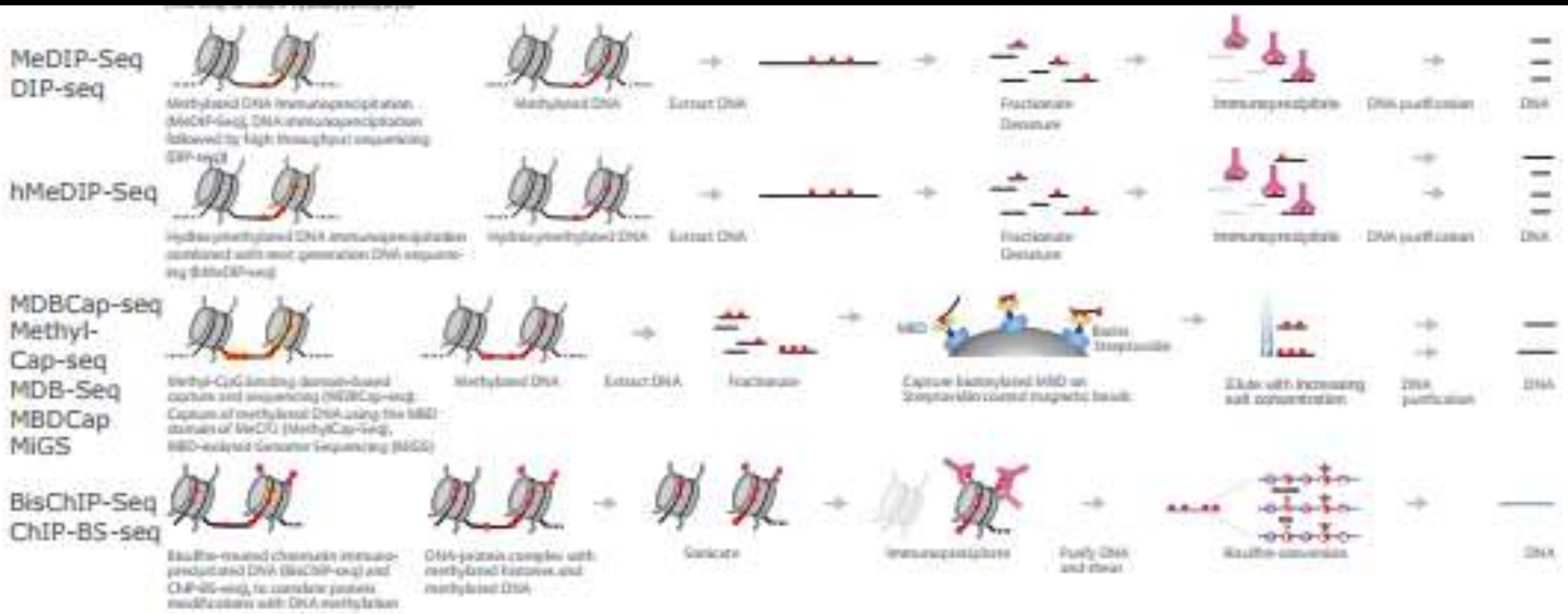
25x linked –read coverage



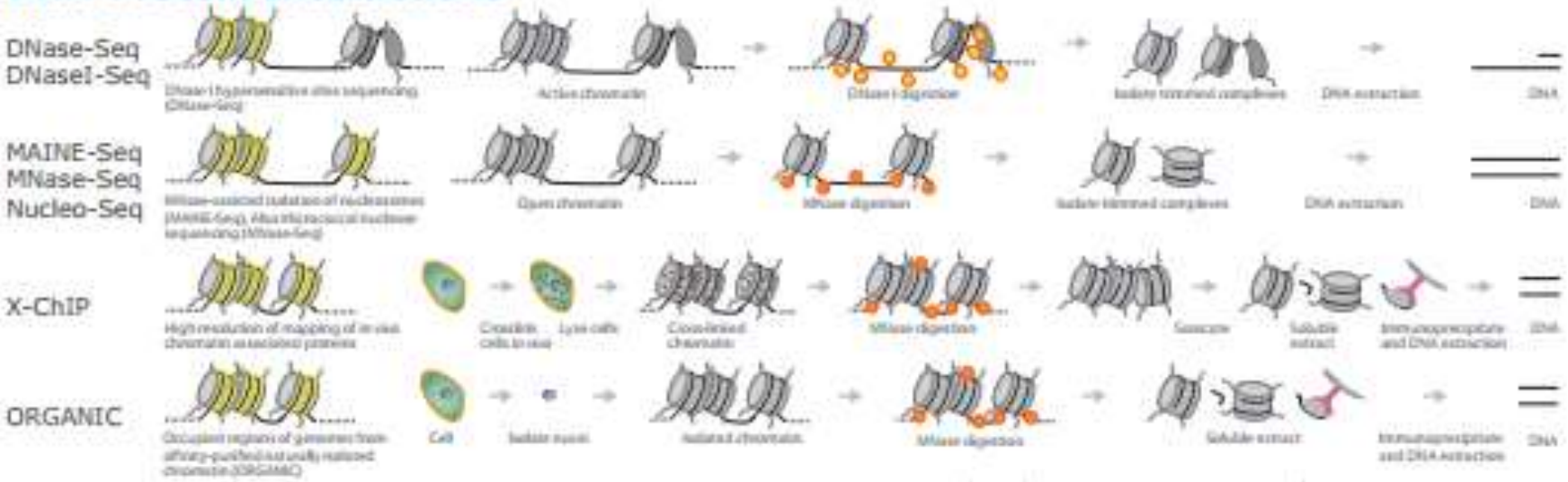
- 60 kb deletion


10X GENOMICS
Loupe · Sample 5728
File
SUMMARY
PHASING
STRUCTURAL VARIANTS
chr2+29989813-chr2+30039813,chr2+42493841-chr2+42543841
chr2
29,995,000
30,000,000
30,005,000
30,010,000
30,015,000
30,020,000
30,025,000
30,030,000
30,035,000
HAPLOTYPE 1
UNPHASED
HAPLOTYPE 2
GENES
EXONS
COVERAGE
BREAKPOINTS
ALK
SV #61
chr2
42,495,000
42,500,000
42,505,000
42,510,000
42,515,000
42,520,000
42,525,000
42,530,000
42,535,000
42,540,000
EML4
SV #61
SV #71
SV #279


NA12878
NA12882
Total number of bases (x10⁹)
0
2
4
6
8
0
50
100
150
200
Fragment length (Kbp)



DNA-Protein Interactions



Genome Resequencing

Rearrangements

SNPs, Indels CNVs

De novo genome Sequencing

Metageno mics

Exome Sequencing

RNA-seq Gene Expression

Long Read Sequencing PacBio , (Nanopore)

DNA Methylation

Splice Isoform Abundance

ChIP-SEQ

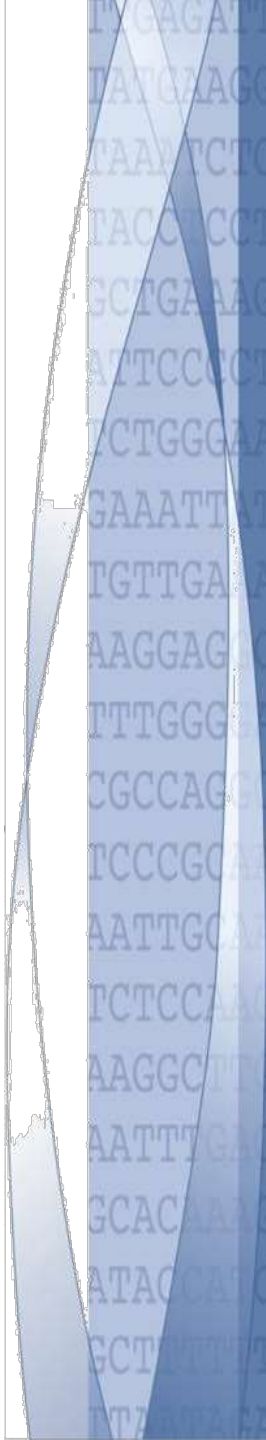
3D Organization

Genotyping

Small RNA

Illumina sequencing workflow

- Library Construction
- Cluster Formation
- Sequencing
- Data Analysis



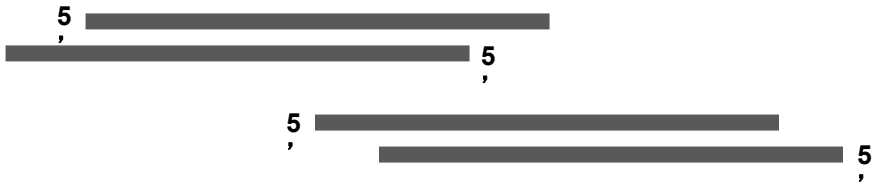
Fragmentation

- Mechanical shearing:
  - BioRuptor
  - Covaris

  DNA, RNA
- Enzymatic:
  - Fragmentase, RNAse3

  DNA, RNA
- Chemical: Mg2+, Zn2+ → **RNA**

DNA library construction



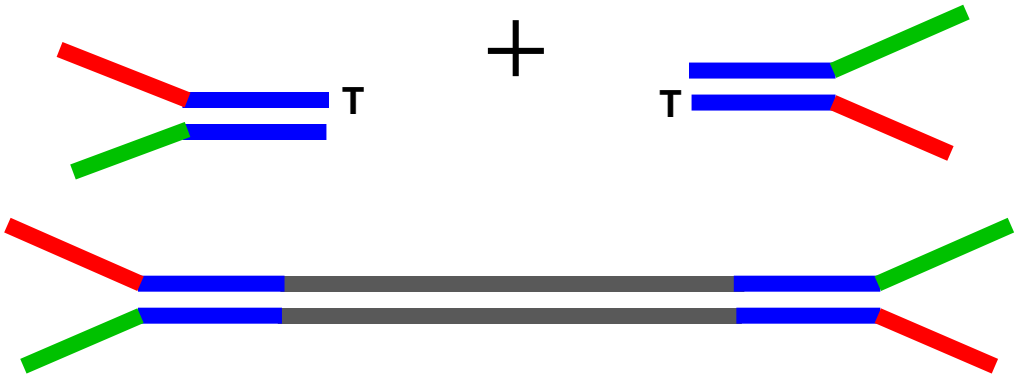
Fragmented DNA

End Repair

Blunt End Fragments

"A" Tailing

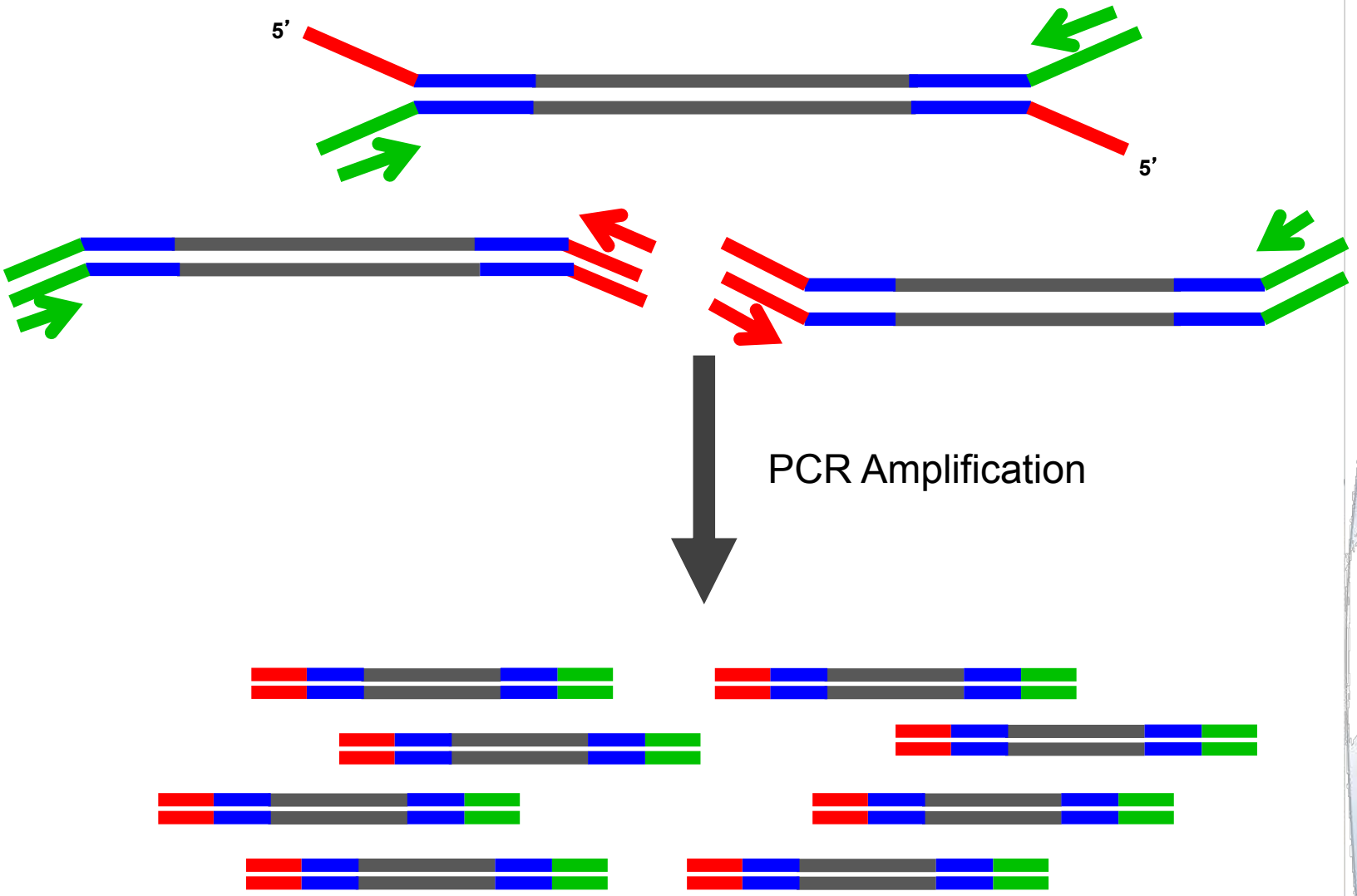
Single Overhang Fragments



Adapter Ligation

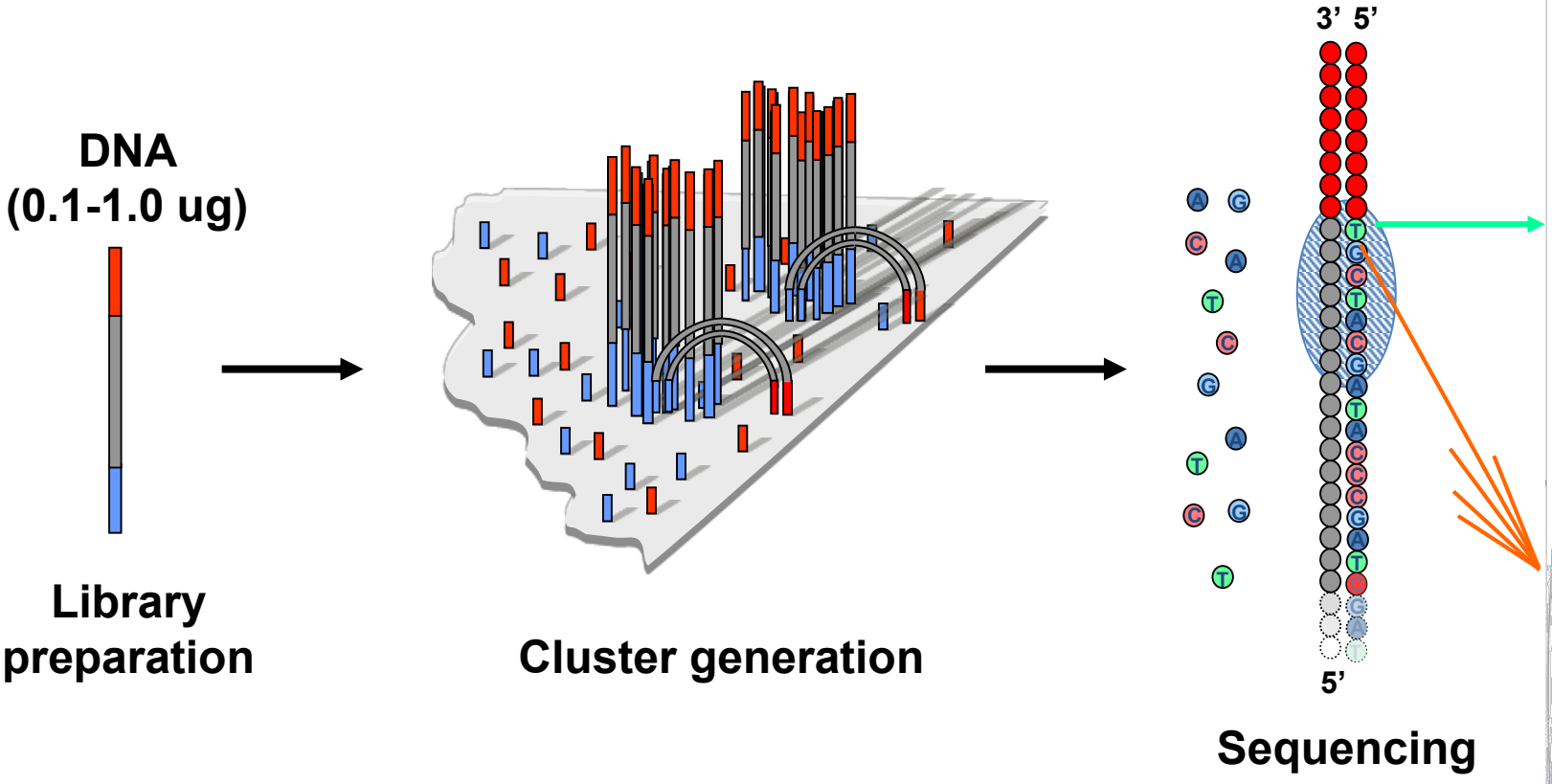
DNA Fragments with Adapter Ends

Enrichment of library fragments

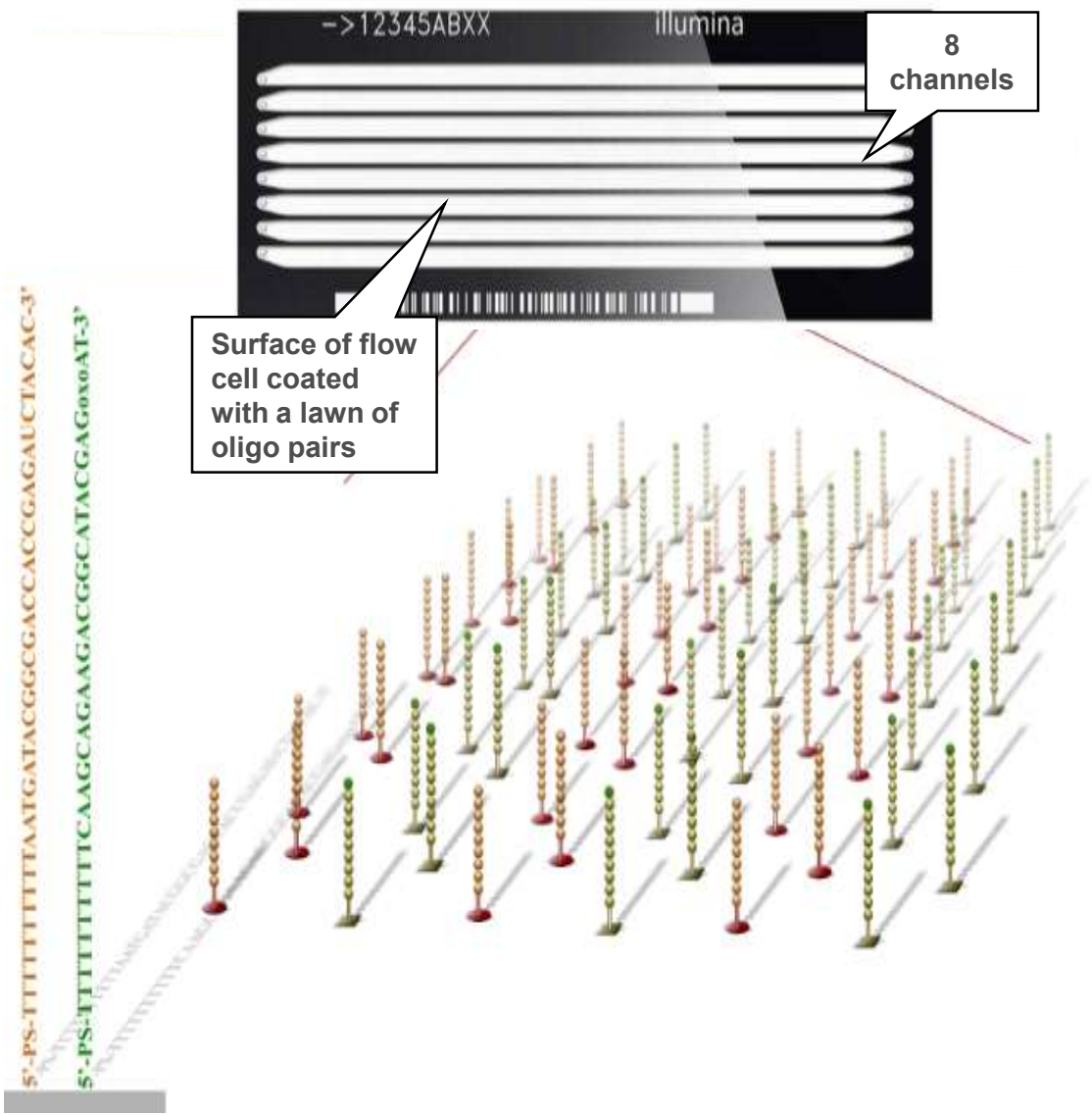


Illumina Sequencing Technology

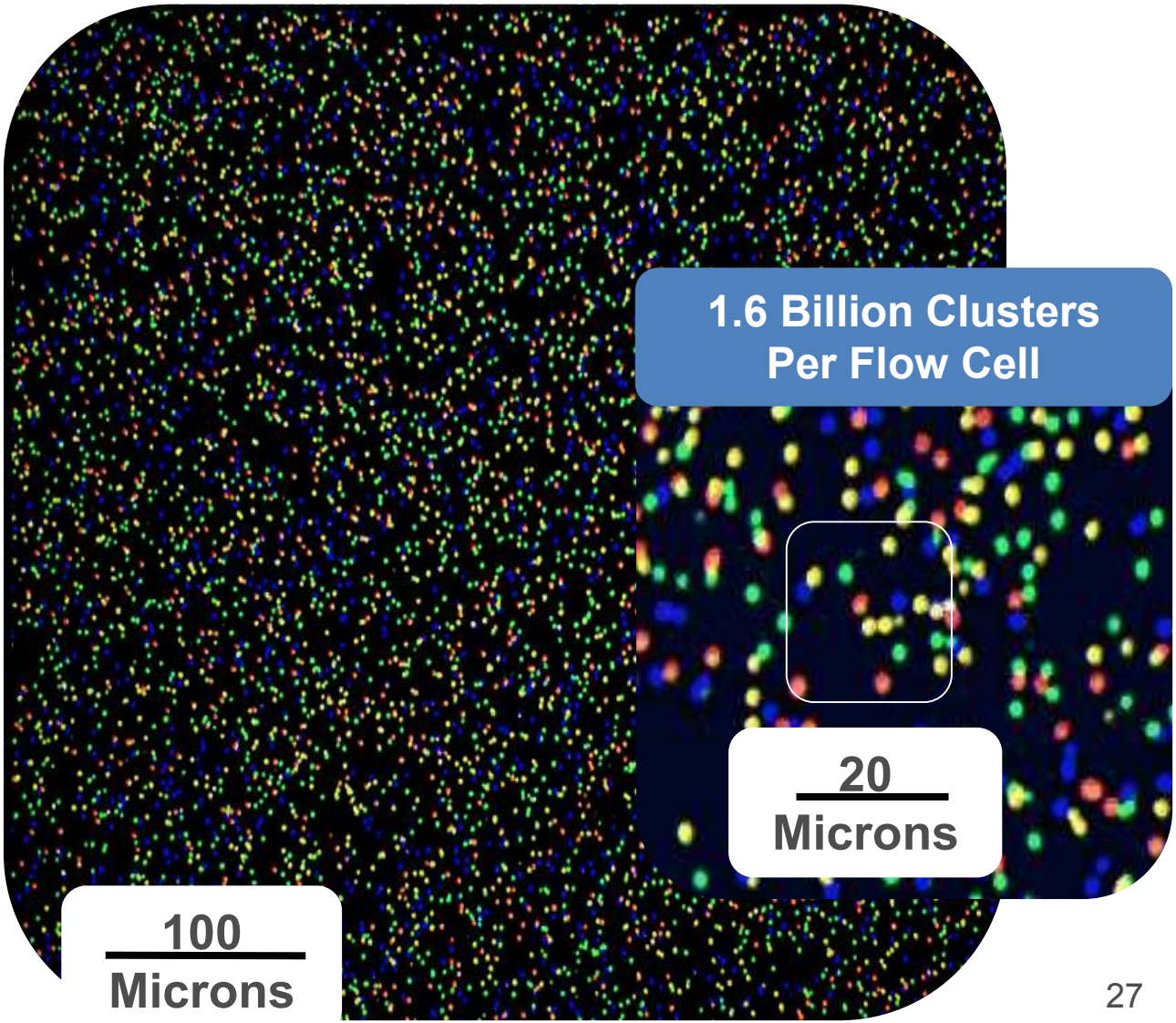
Sequencing By Synthesis (SBS) Technology



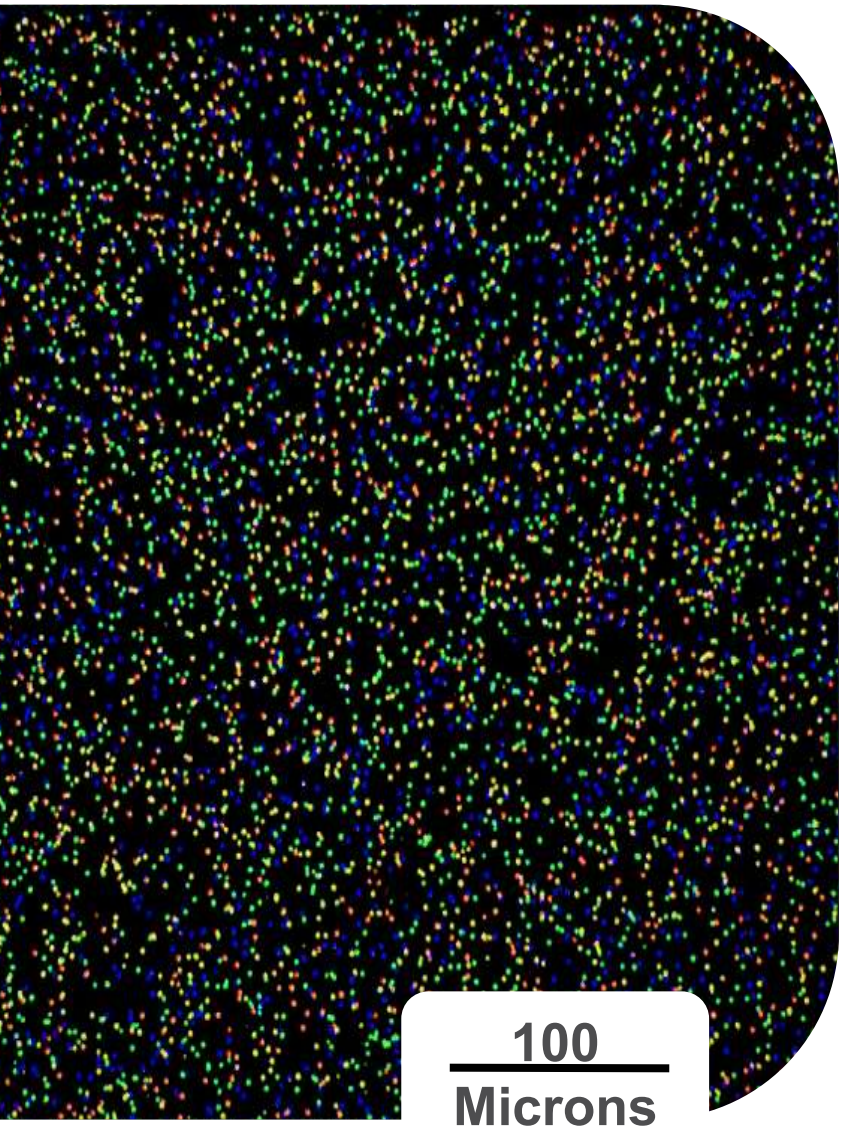
TruSeq Chemistry: Flow Cell

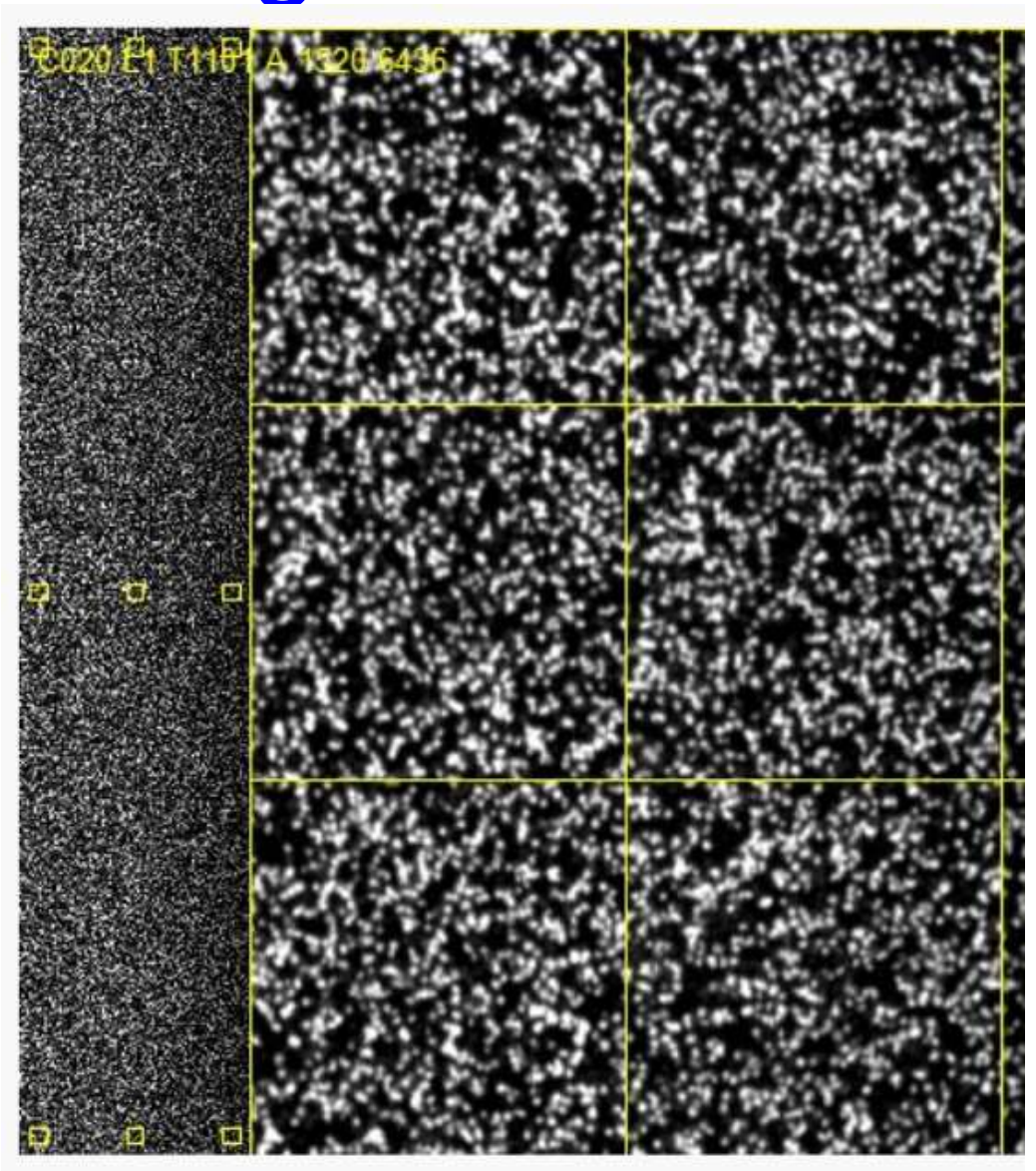


Sequencing

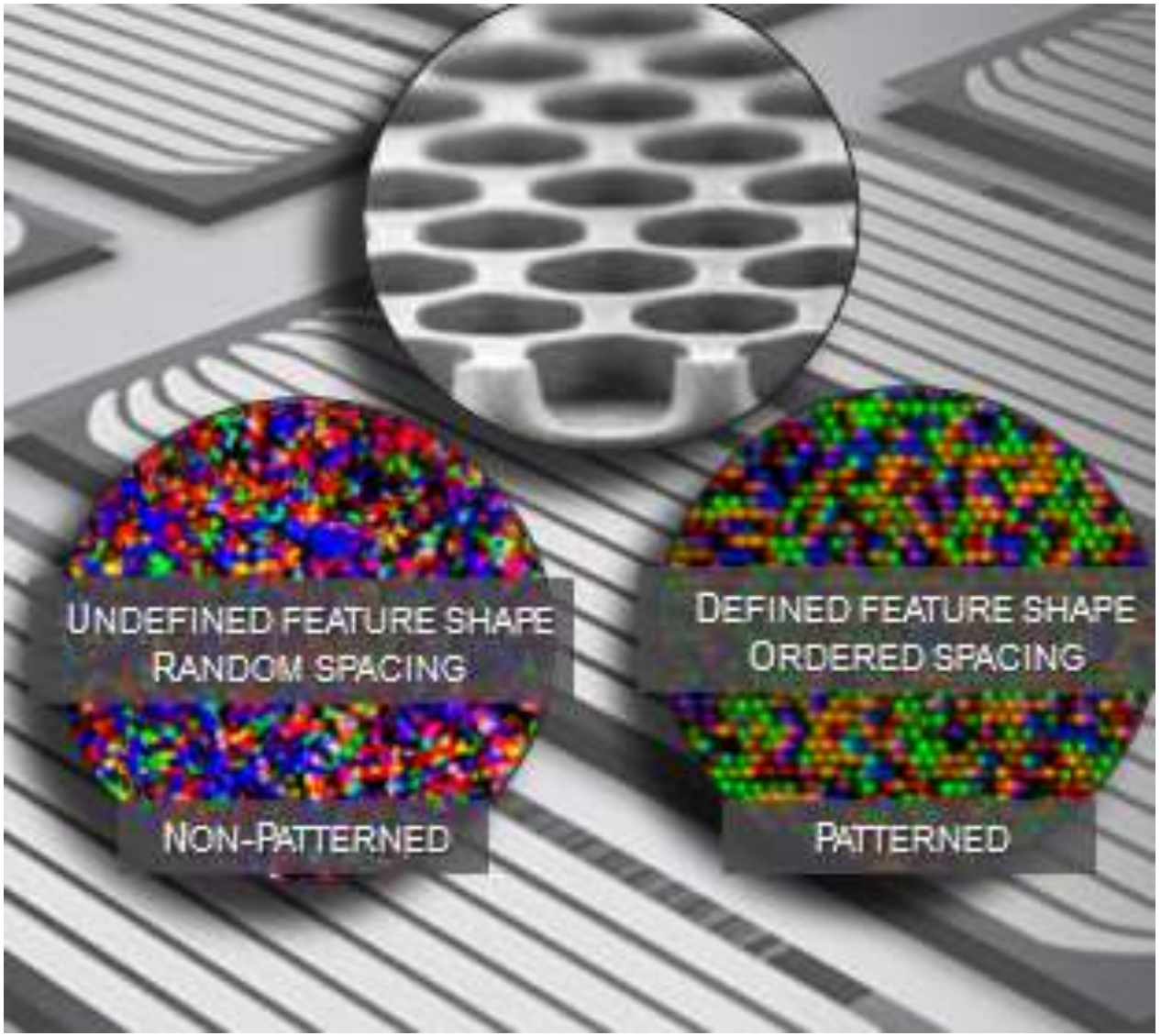


Sequencing

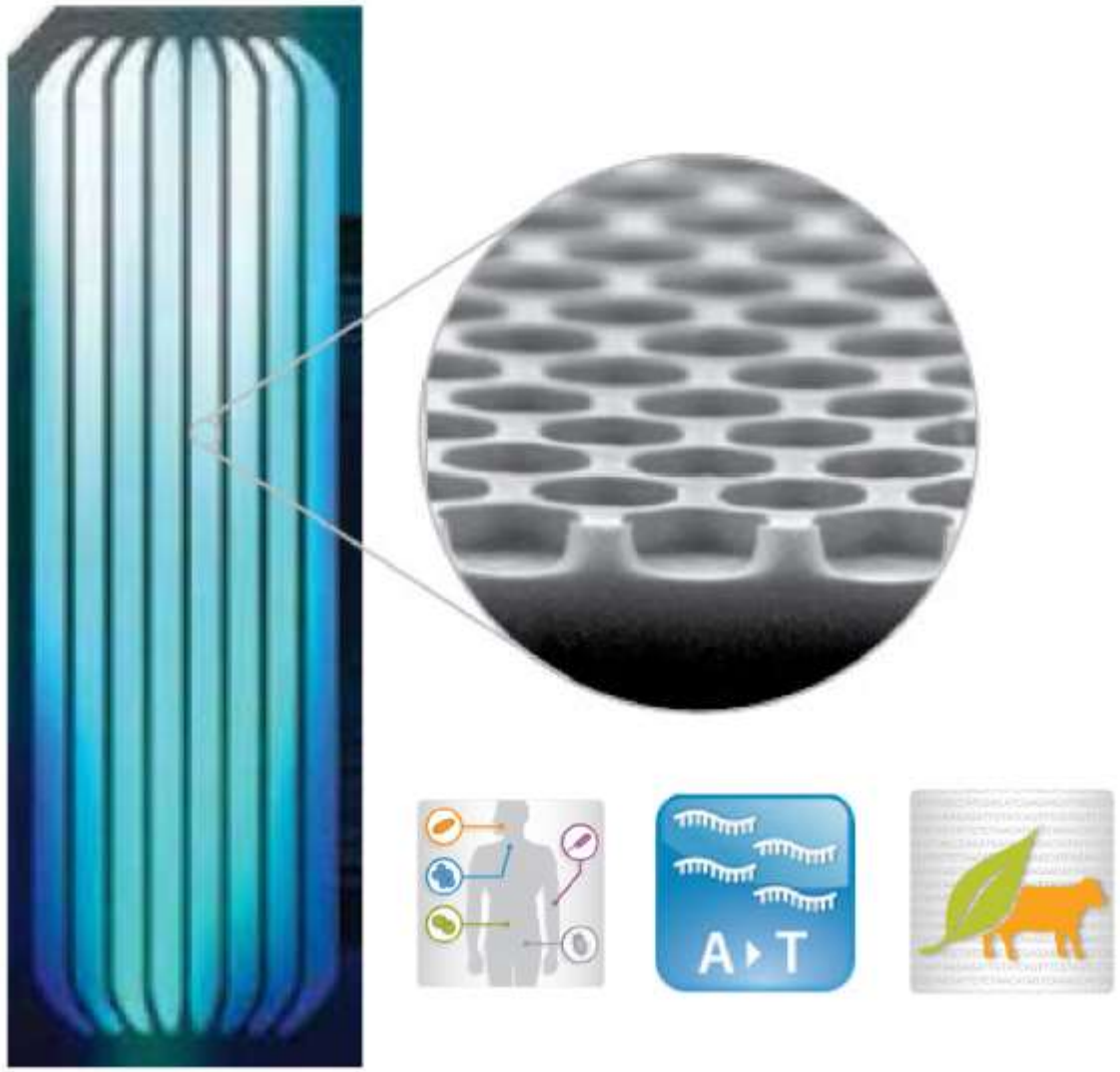




Patterned Flowcell

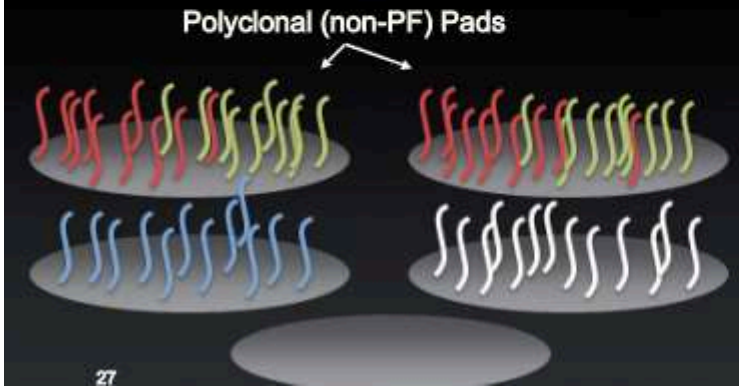


Hiseq 3000: 478 million nanowells per lane

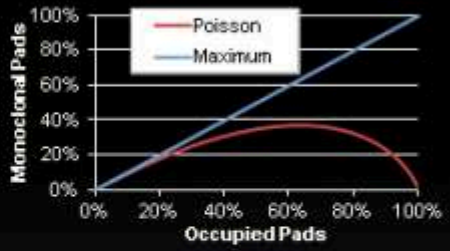


CONCEPTUAL CHALLENGE— BEATING POISSON

Amplification Phase



Maximizing Well Occupancy and Monoclonality

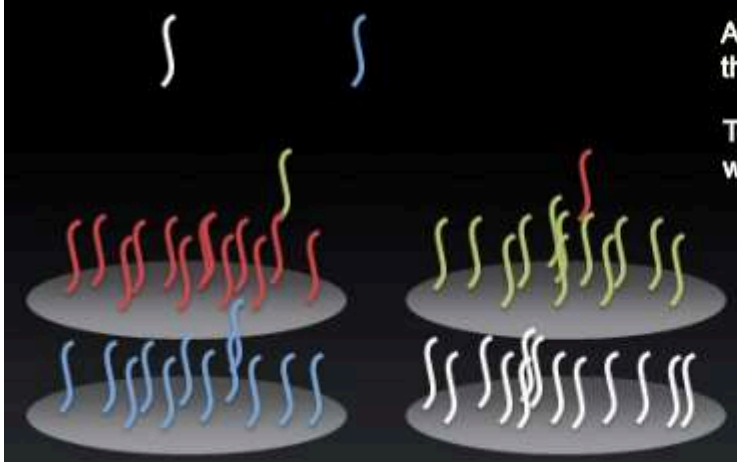


Poisson statistics limit max monoclonal occupancy < 40%

Polyclonality rises as occupancy increases

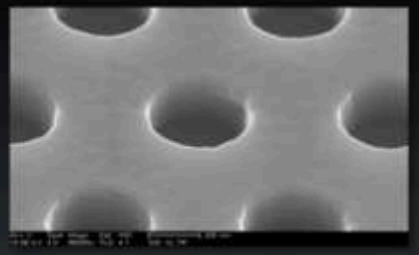
SIMULTANEOUS SEEDING AND AMPLIFICATION

Maximizing Well Occupancy and Monoclonality



Amplification occurs at rate >> faster than seeding rate

Templates excluded from occupied wells



What will go wrong ?

- cluster identification
- bubbles
- synthesis errors:

ClusterCluster
ClustsrCluster
ClusterCluster
ClusterCluster
CllsterCluster

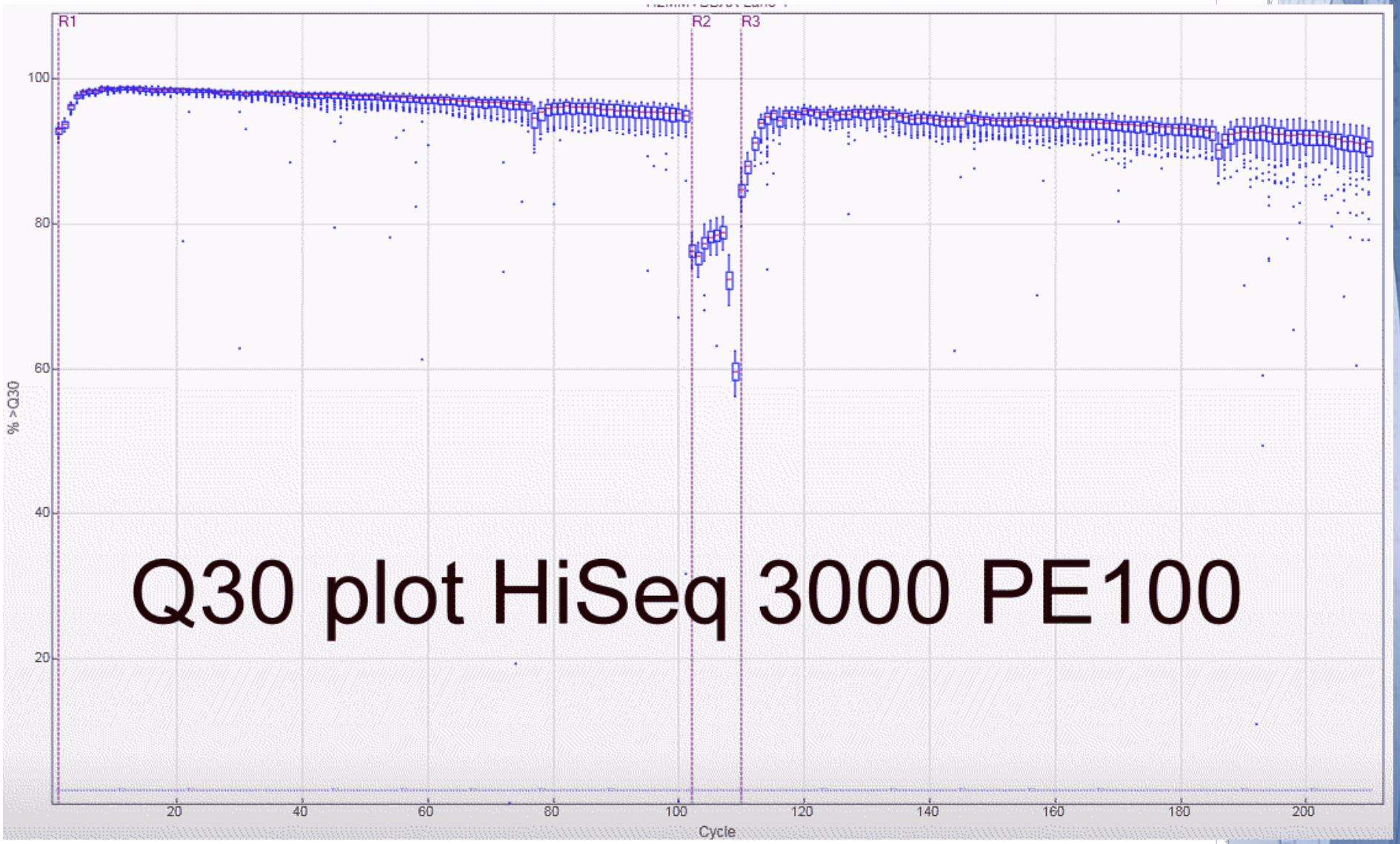
What will go wrong ?

- synthesis errors:

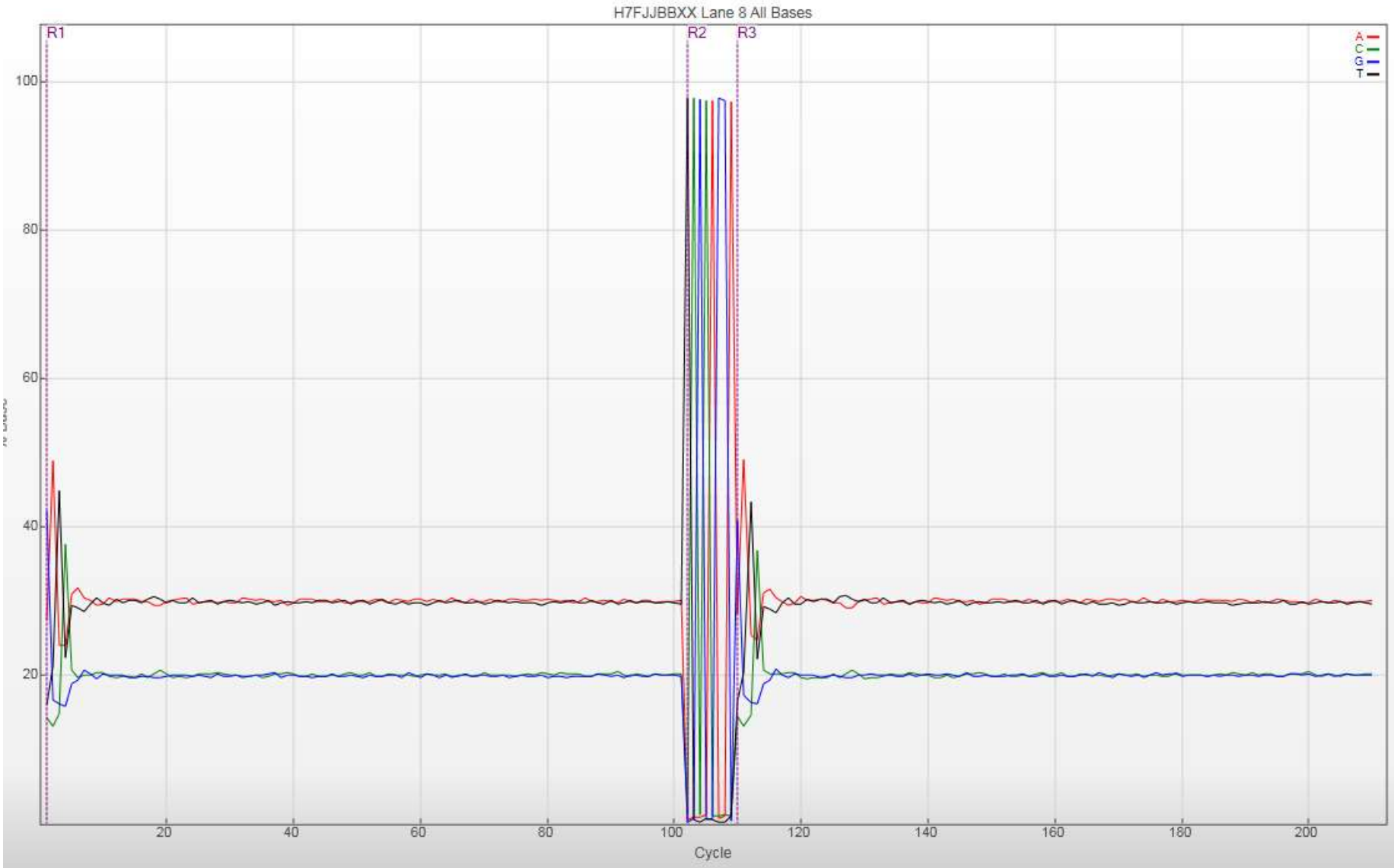
ClusterCluster	ClsterClusterC
ClustsrCluster	ClusterCluster
ClusterCluster	ClusterCluster
ClusterCluster	CllusterCluste
CllsterCluster	ClusterCluster

Phasing & Pre-Phasing problems

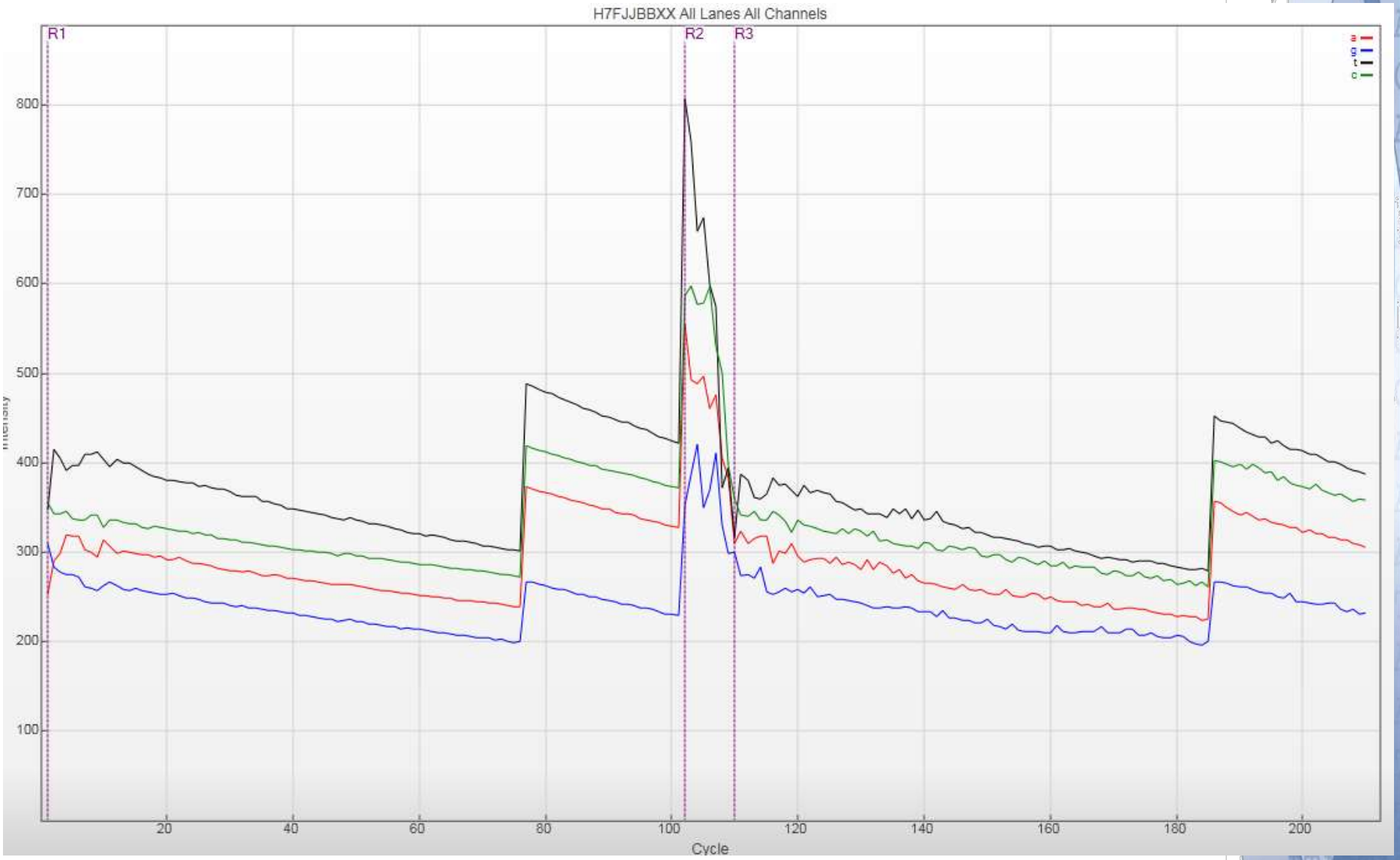
Illumina SAV viewer



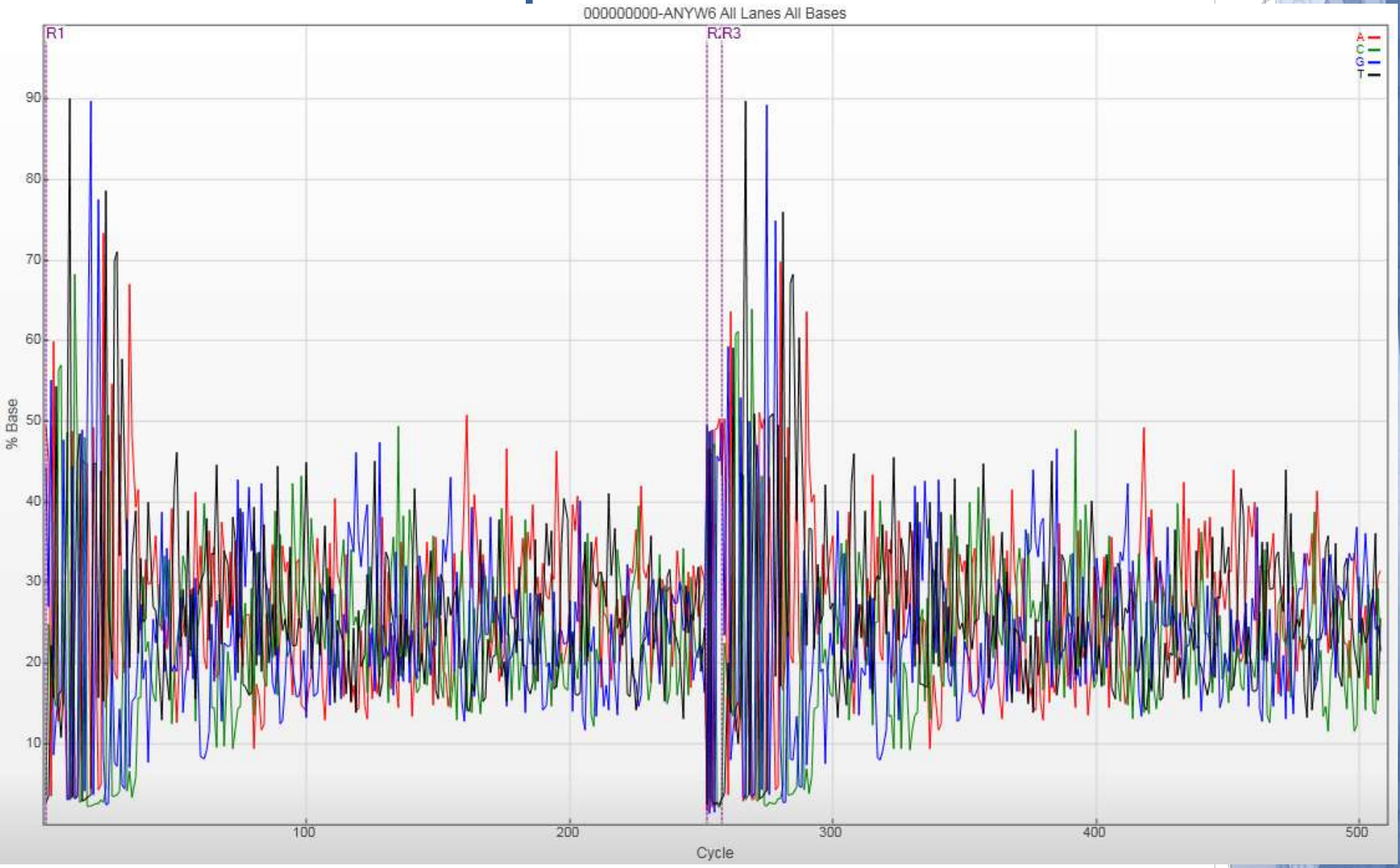
base composition



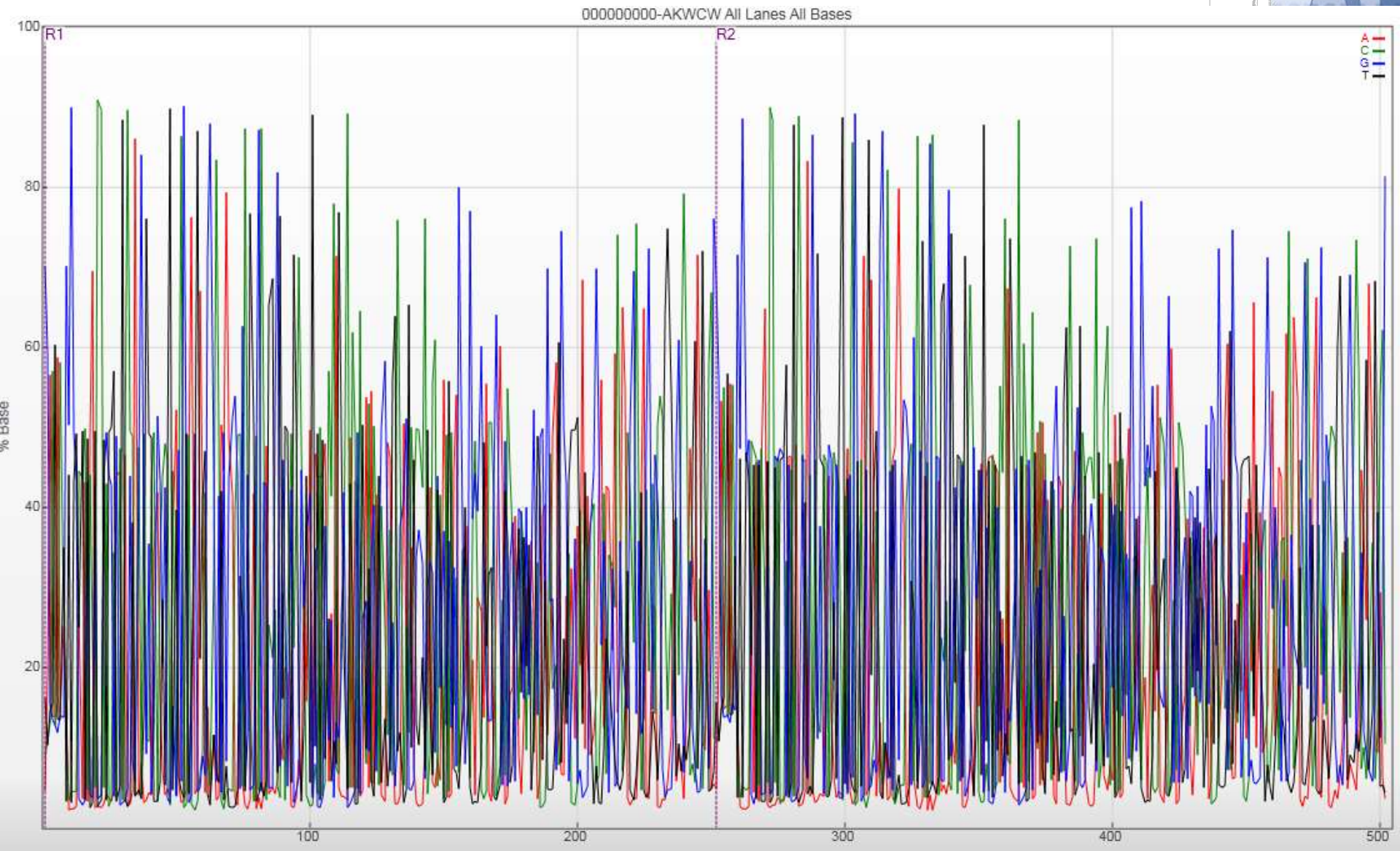
fluorescence intensity



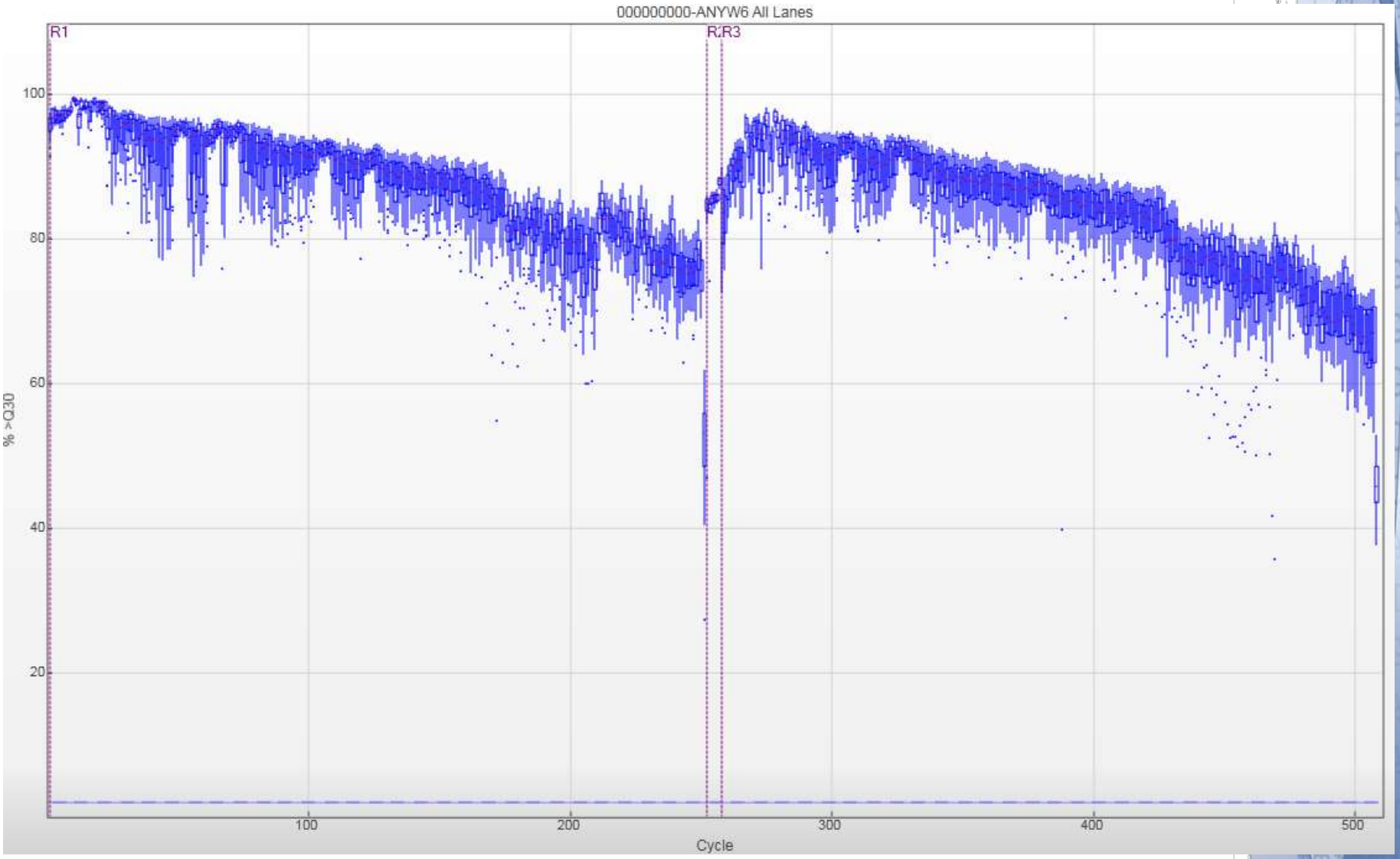
amplicon mix



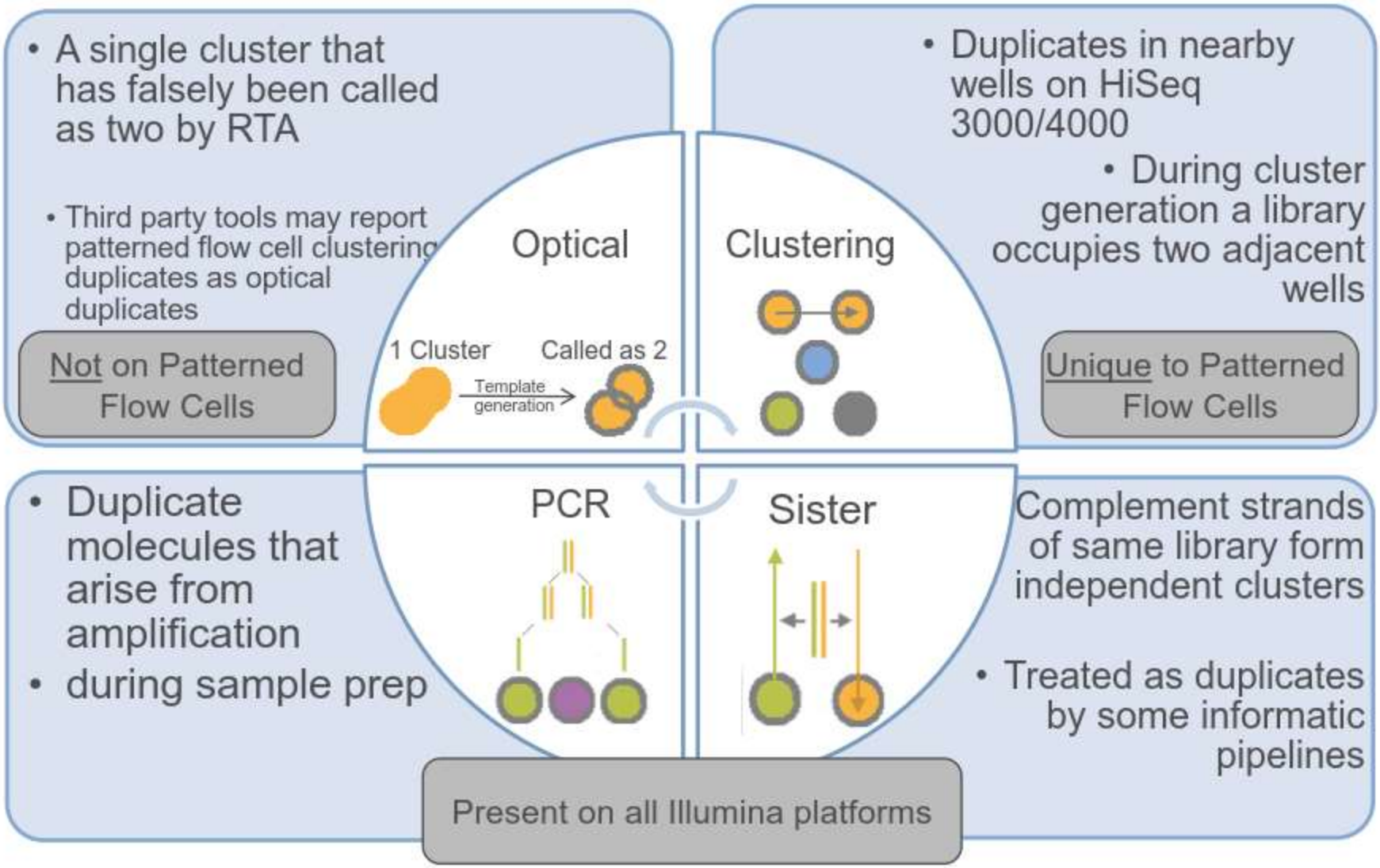
amplicon


000000000-AKWCW All Lanes All Bases
R1
R2
A
C
G
T
% Base
100
80
60
40
20
100
200
300
400
500

amplicon mix Q30

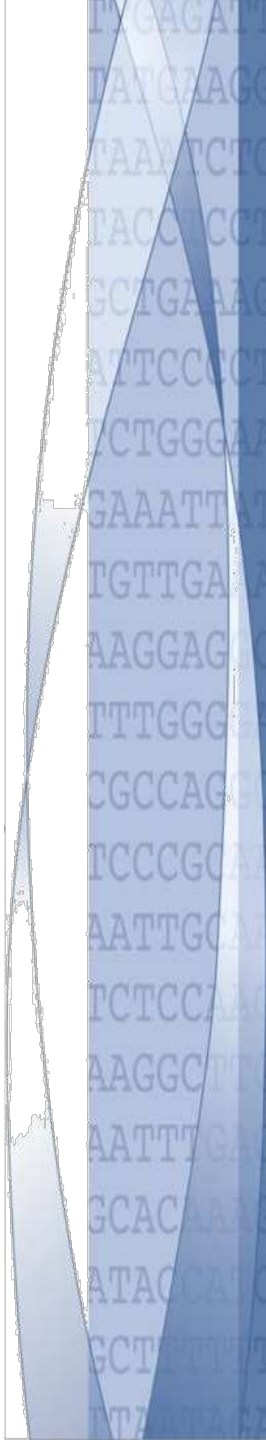


A Review of Sequencing Duplicate Types

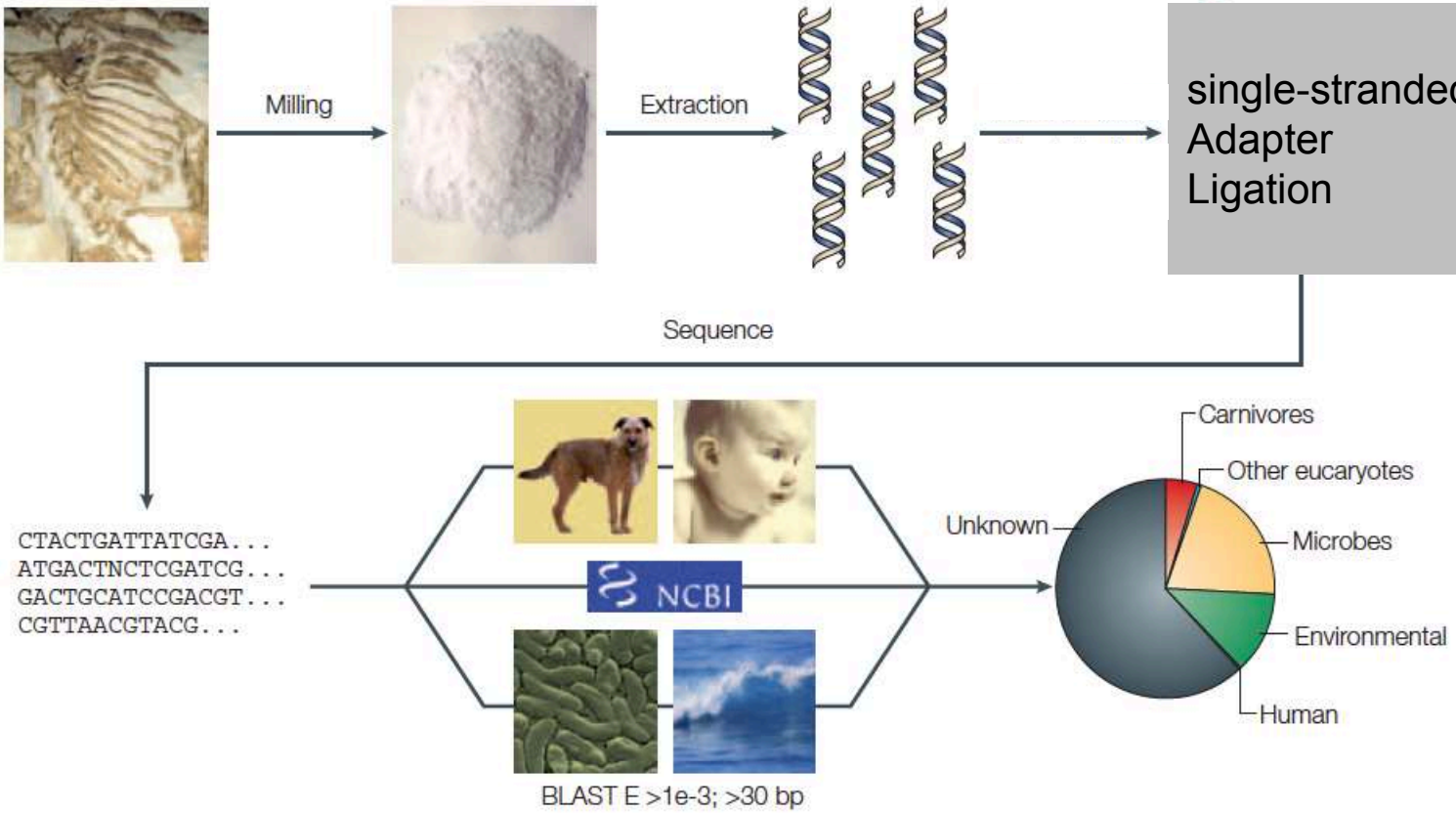


“If you can put adapters on it,

we can sequence it!”



Know your sample



No need to be scared of HTS

UC Davis Center for Plant Diversity/Herbarium

- The Herbarium archives contain over 300,000 dried specimens.

- Search for Grapevine Red Blotch-Associated Virus
- Virus traces found by PCR

Maher Al Rwahnih
UCD Plant Foundation
Plant Services

Studying historic Bean varieties from herbarium samples

- GBS (Genotyping-By-Sequencing)
- 60 year old herbarium samples

Sarah Dohle,
Gepts Lab

Quantitation & QC methods

- Intercalating dye methods (PicoGreen, Qubit, etc.):

Specific to dsDNA, accurate at low levels of DNA
Great for pooling of indexed libraries to be sequenced in one lane
Requires standard curve generation, many accurate pipetting steps

- Bioanalyzer:

Quantitation is good for rough estimate
Invaluable for library QC
High-sensitivity DNA chip allows quantitation of low DNA levels

- qPCR

Most accurate quantitation method
More labor-intensive
Must be compared to a control

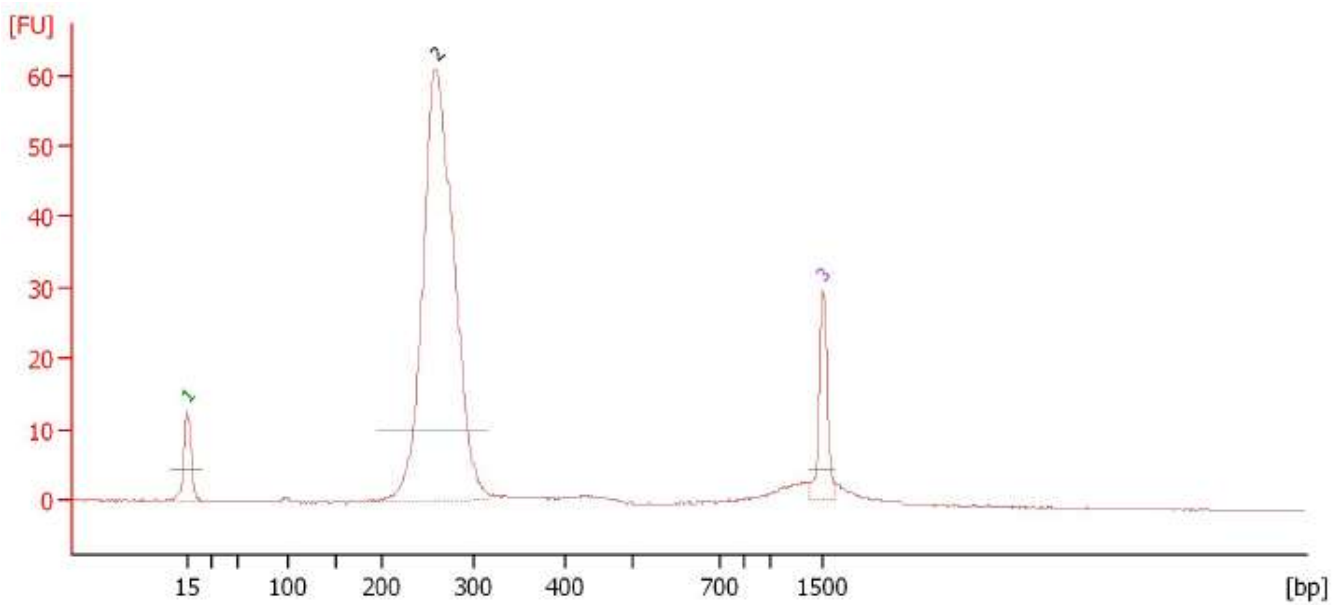
Optional: PCR-free libraries

- PCR-free library:
  - if concentration allows
  - Reduction of PCR bias against e.g. GC rich or AT rich regions, especially for metagenomic samples

OR

- Library enrichment by PCR:
  - Ideal combination: high input and low cycle number; low-bias polymerase

Library QC by Bioanalyzer

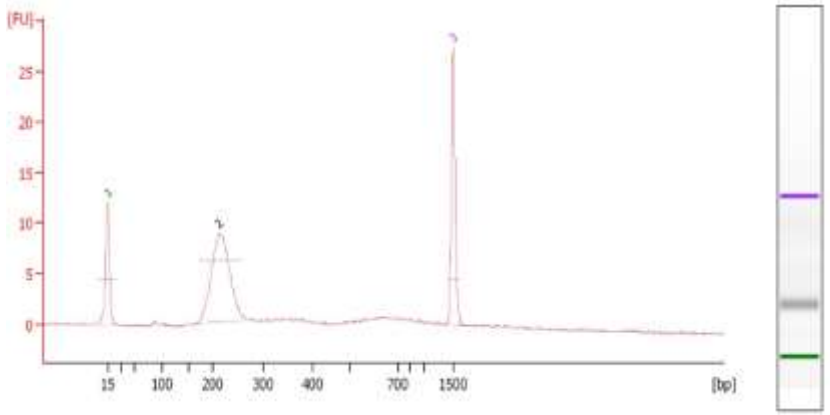


Predominant species of appropriate MW

Minimal primer dimer or adapter dimers

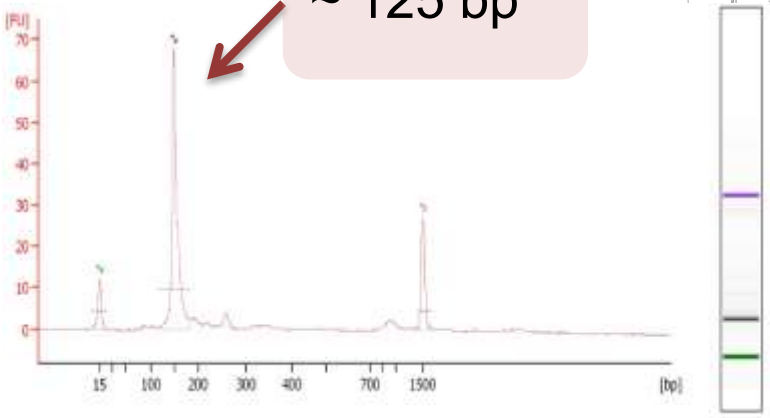
Minimal higher MW material

Library QC by Bioanalyzer

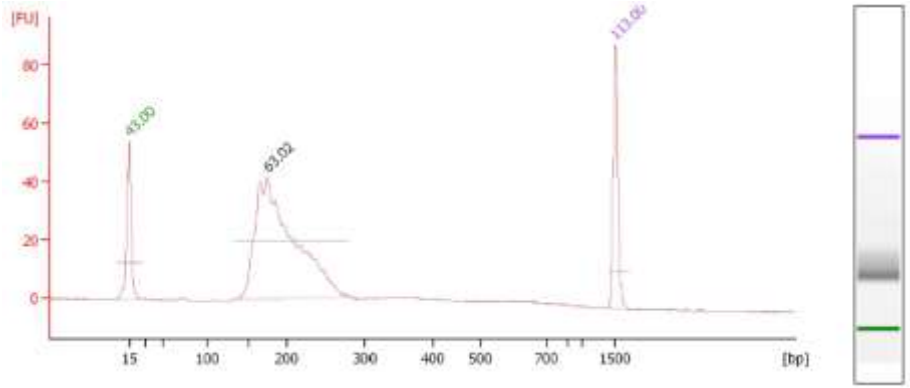


Beautiful

~ 125 bp

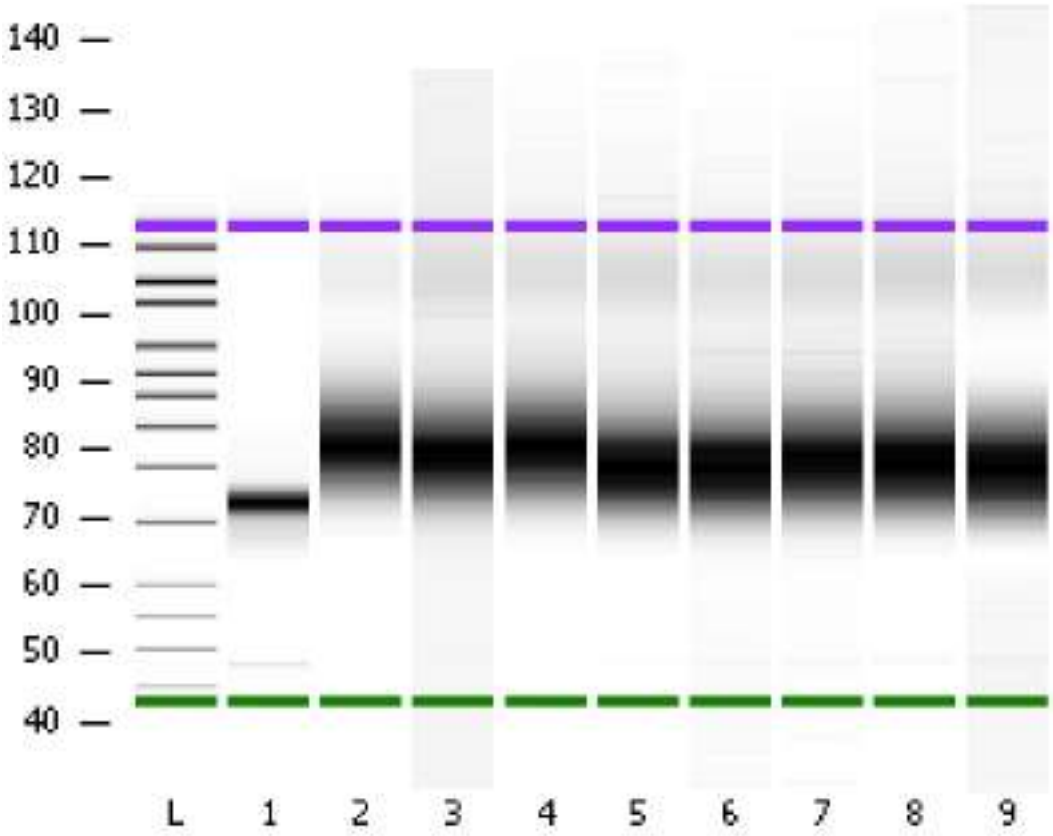


100% Adapters

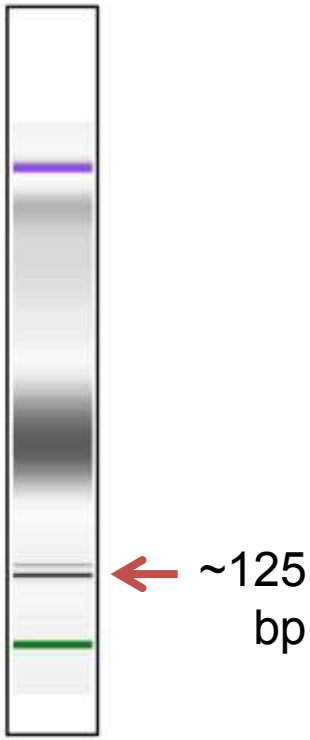


Beautiful

Library QC



Examples for successful libraries



Adapter contamination at ~125 bp

RNA-seq targeted sequencing:

- Capture-seq (Mercer et al. 2014)
- Nimblegen and Illumina

- Low quality DNA (FFPE)
- Lower read numbers 10 million reads
- Targeting lowly expressed genes.

PACIFIC BIOSCIENCES™

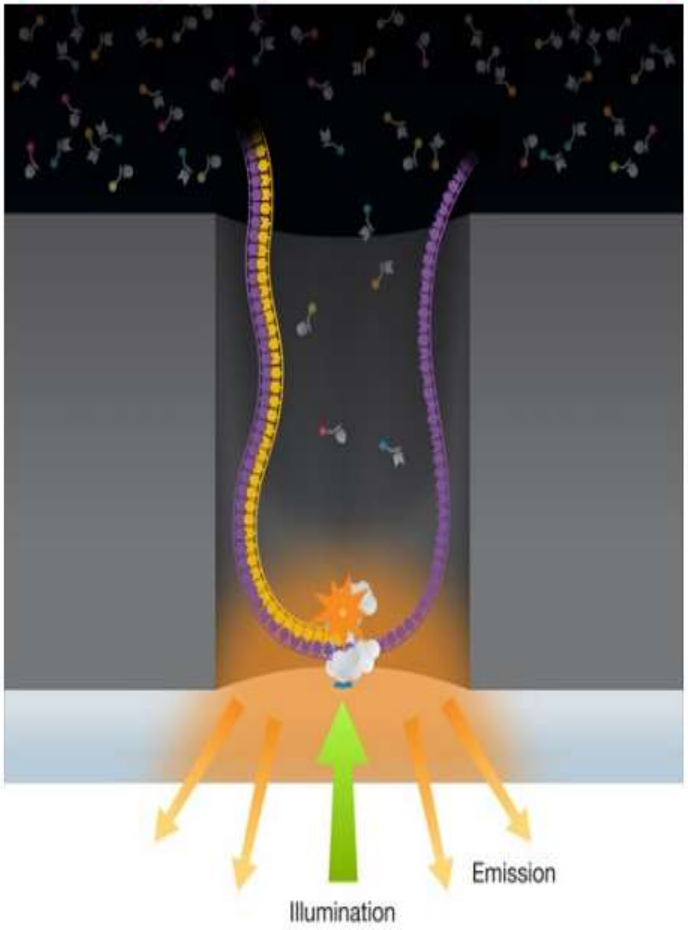
http://pacificbiosciences.com

THIRD GENERATION DNA SEQUENCING

Single Molecule Real Time (SMRT™) sequencing
Sequencing of single DNA molecule by single polymerase
Very long reads: average reads over 8 kb, up to 30 kb
High error rate (~13%).
Complementary to short accurate reads of Illumina

Third Generation Sequencing : Single Molecule Sequencing

Pacific Biosciences

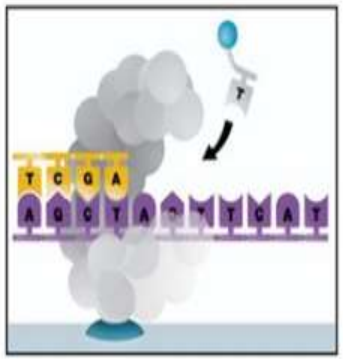


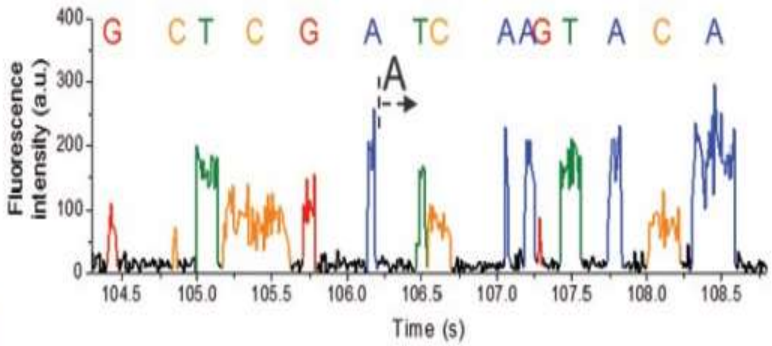
4 nucleotides with different fluorescent dye simultaneous present

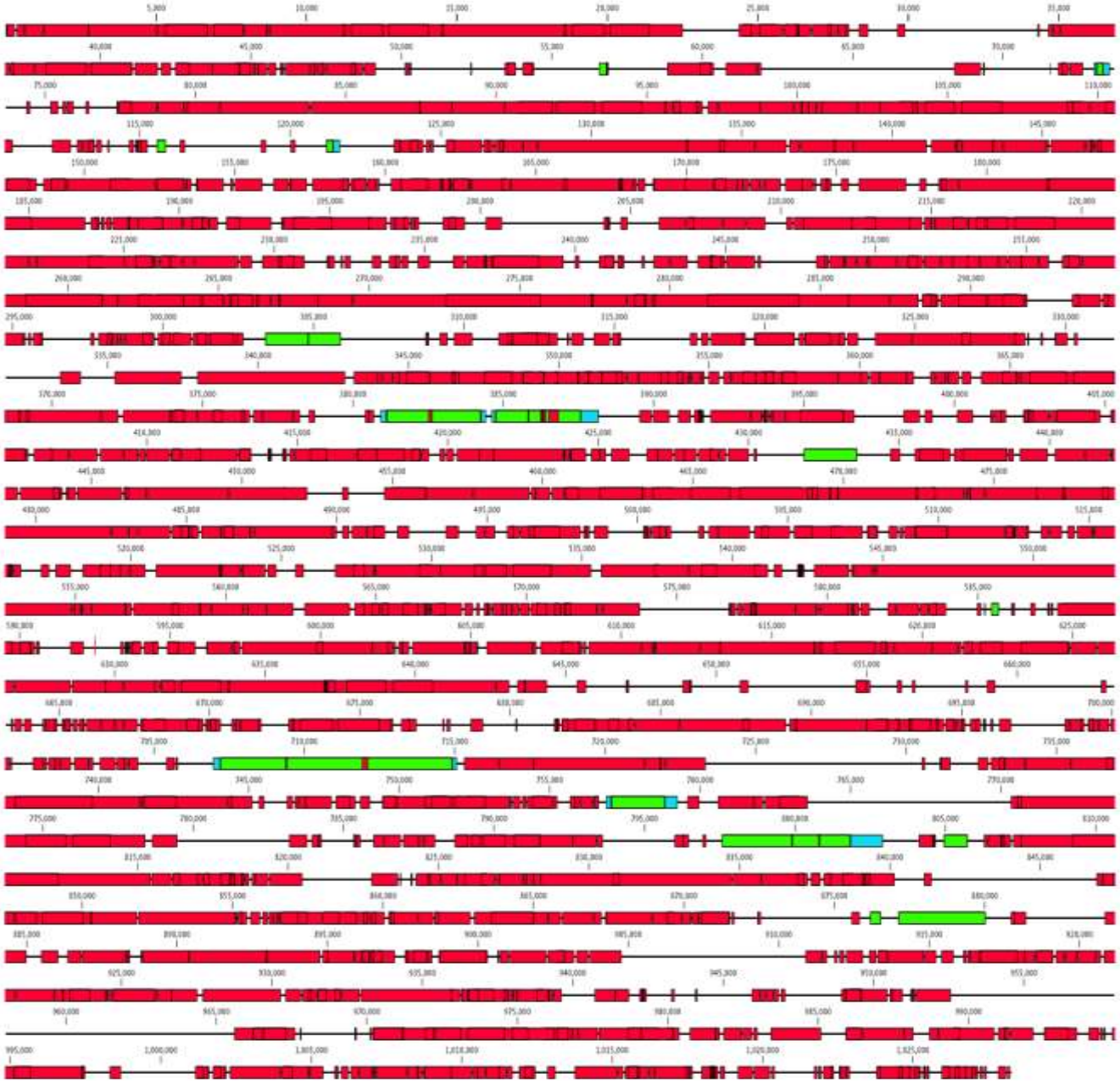
2-3 nucleotides/sec
2-3 Kb (up to 50) read length
6 TB data in 30 minutes

laser damages polymerase

70 nm aperture
“Zero Mode Waveguide”





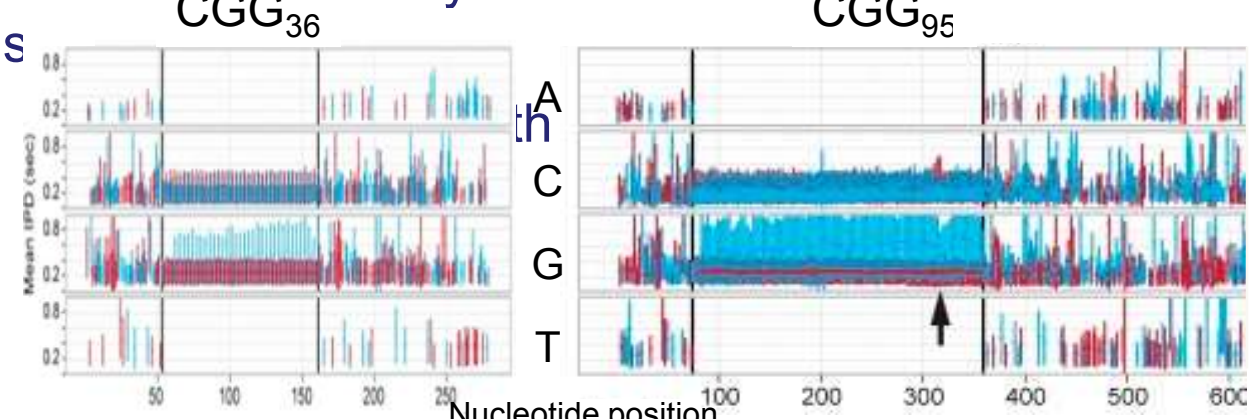


Damien Pelt

First Sequencing of CGG-repeat Alleles in Human Fragile X Syndrome using PacBio RS Sequencer

Paul Hagerman, Biochemistry and Molecular Medicine, SOM.

- Single-molecule sequencing of pure CGG array,
  - first for disease-relevant allele. Loomis *et al*. (2012) *Genome Research*.
  - applicable to many other tandem repeat disorders.
- Direct genomic DNA sequencing of methyl groups,
  - direct epigenetic sequencing (paper under review).
- Discovered 100% bias toward methylation of 20 CGG-repeat allele in female,
  – first [illegible] methylated DNA sequencing in human dis
- [illegible]th



Future's so bright

Thank you!