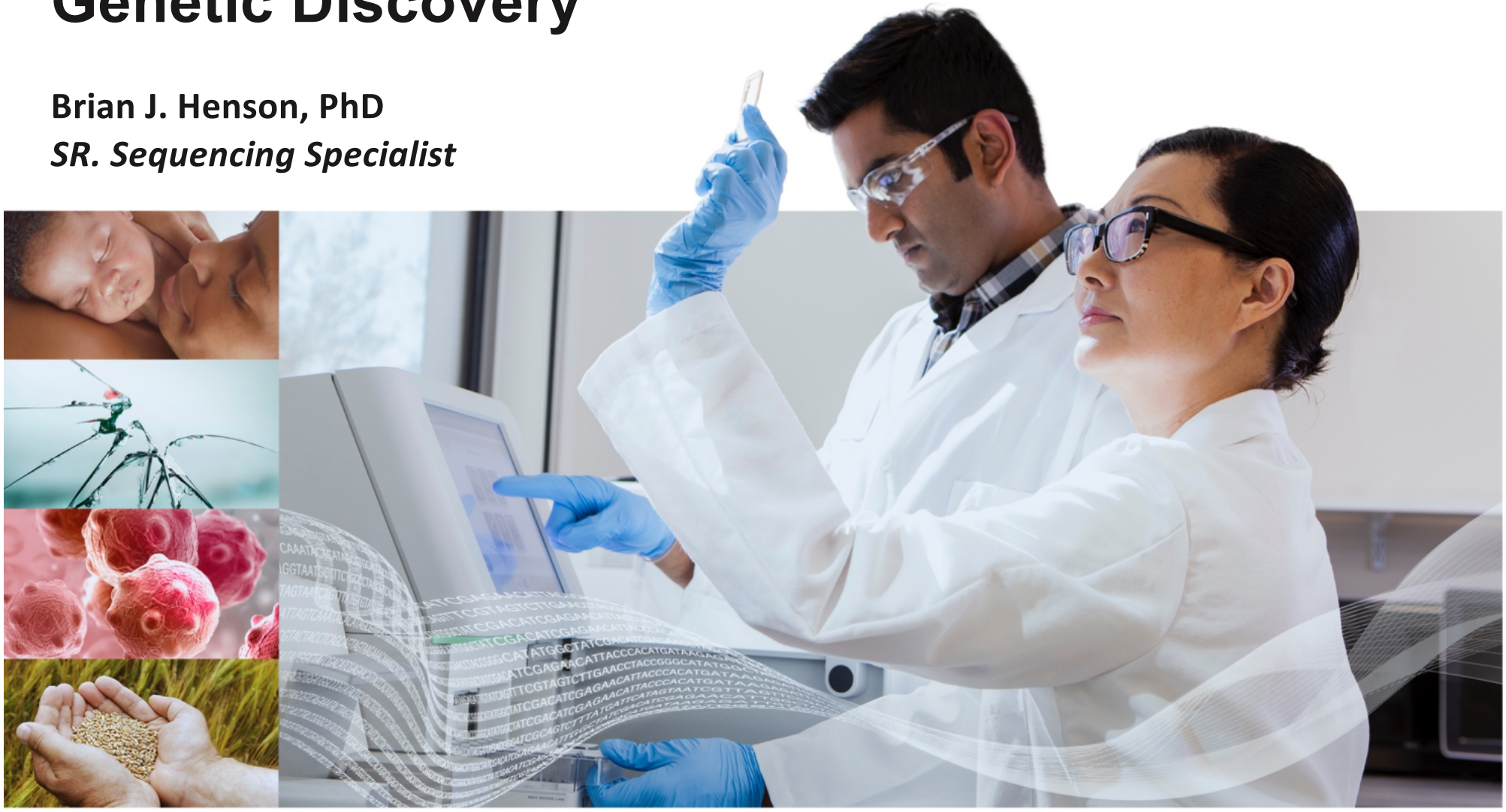
Using Illumina's NGS Technology to Empower Genetic Discovery

Brian J. Henson, PhD

SR. Sequencing Specialist



illumina®

For Research Use Only. Not for use in diagnostic procedures.

Agenda

- **Our Background and Mission**
- **Introduction to NGS**
- **Instrumentation**
- **Whole Genome Sequencing**
  - Nextera DNA Flex
  - Nextera Mate Pair
- **Data Storage and Analysis**
- **Best Practices**

Who We Serve

Innovation drives expanding market opportunities

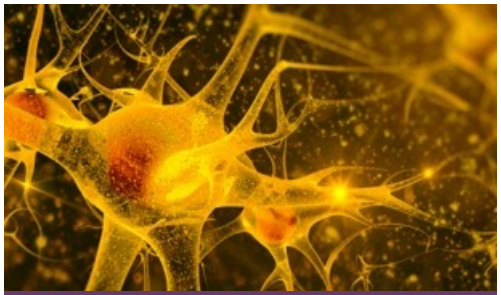
illumina®

Reproductive Health

Oncology

Population Sequencing

Research



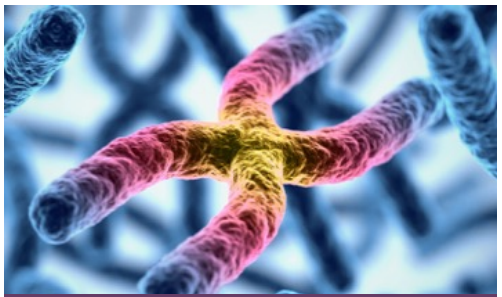
Complex Disease

Consumer

Infectious Disease

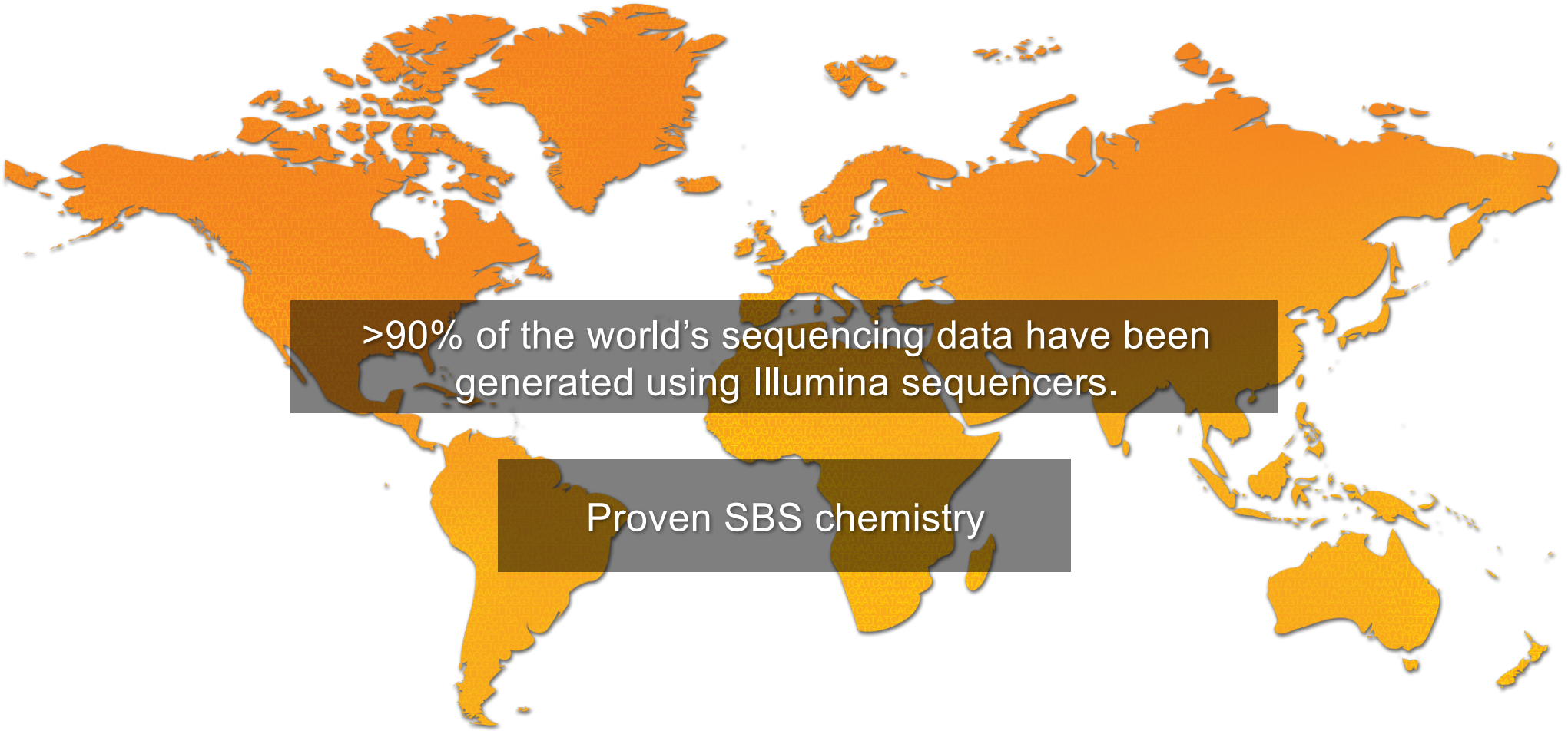
Forensics

Agriculture



Genetic Health

BioPharm


>90% of the world's sequencing data have been generated using Illumina sequencers.
Proven SBS chemistry

illumina®

Sample to Answer Integration

From library prep to downstream informatics & knowledge generation

Library Prep

Sequence

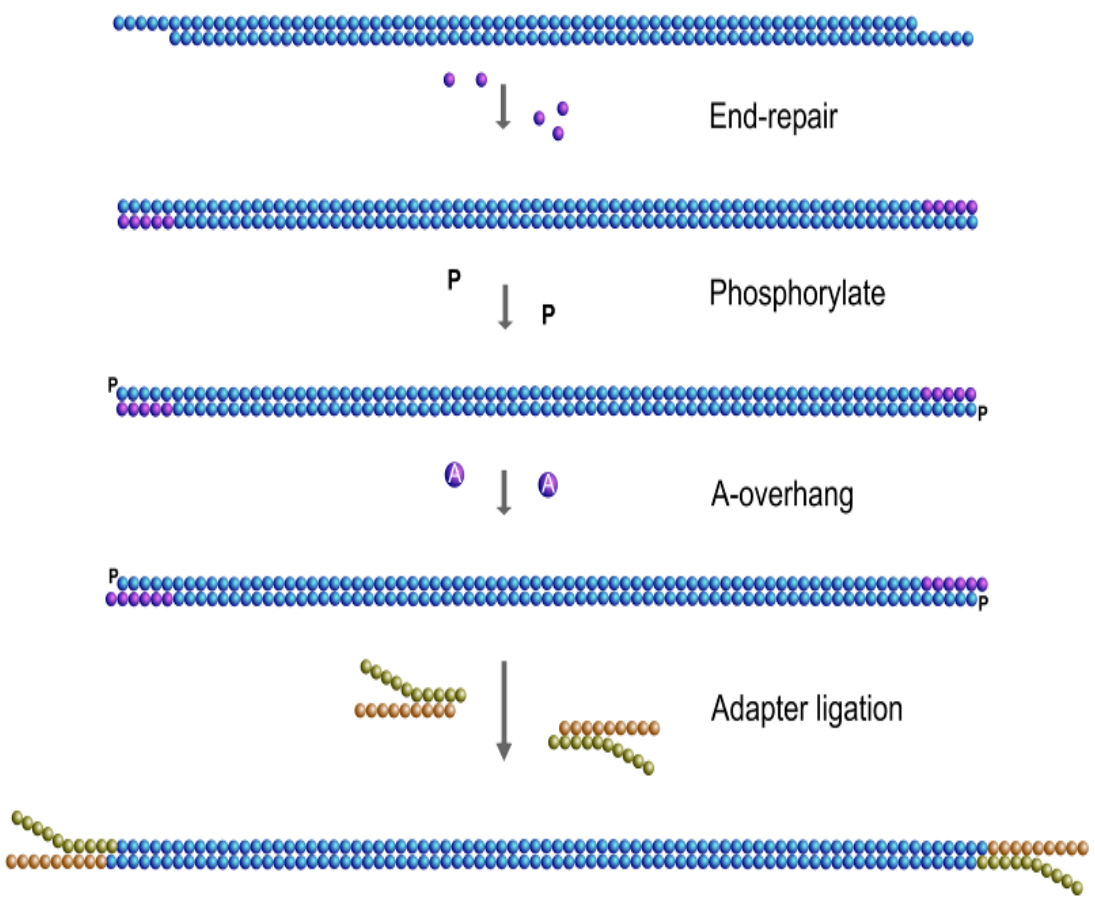
Answer

illumina®

The Next Generation Sequencing Process

illumina®

Library Preparation

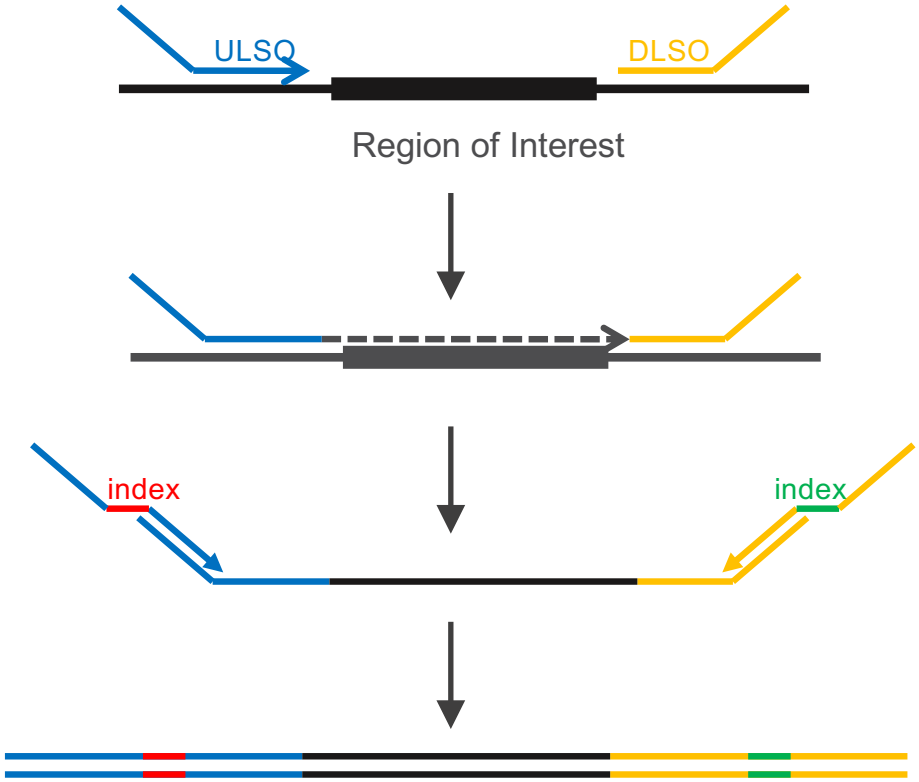


1. DNA is fragmented
2. Blunt-end fragments created
3. A-base added
4. Dual-index adapters ligated

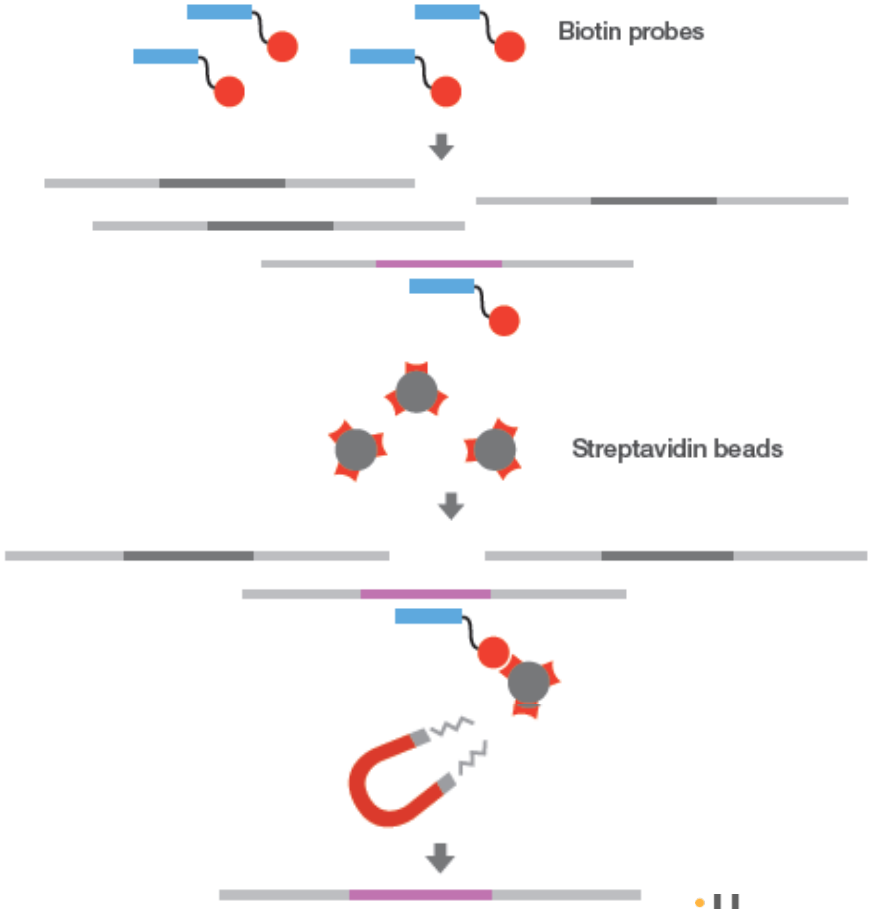
illumina®

Multiple Ways to Make a Library

- **Amplicon**

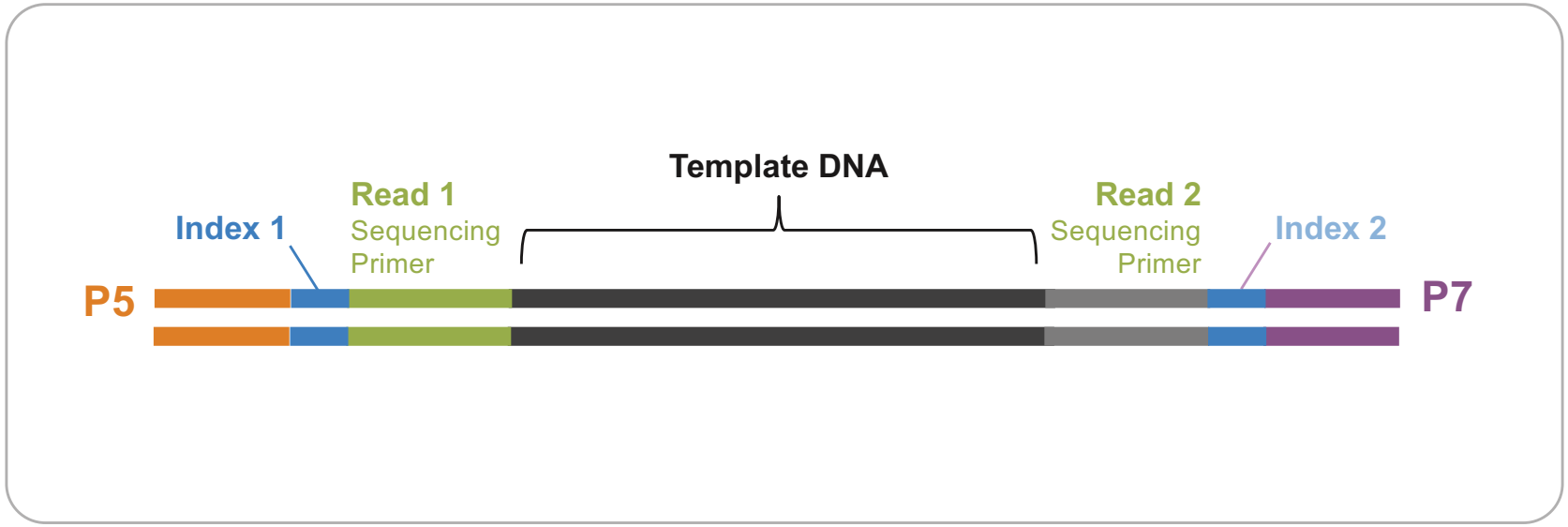


- **Enrichment**



illumina®

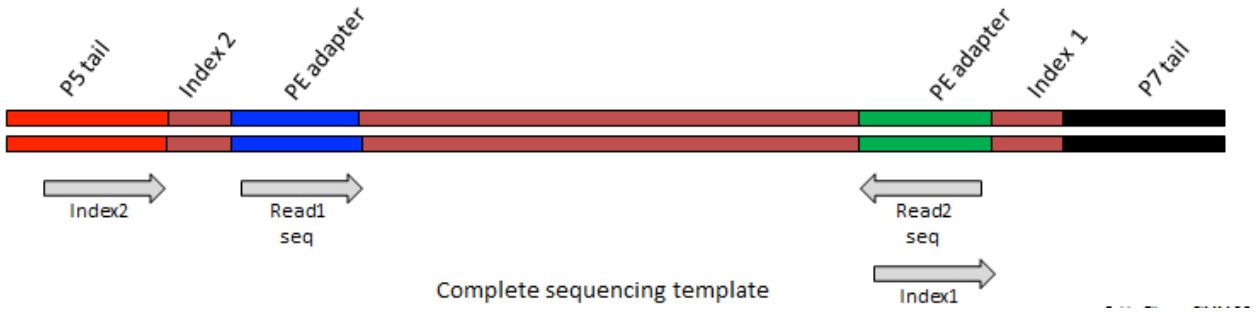
Sequence Ready Library

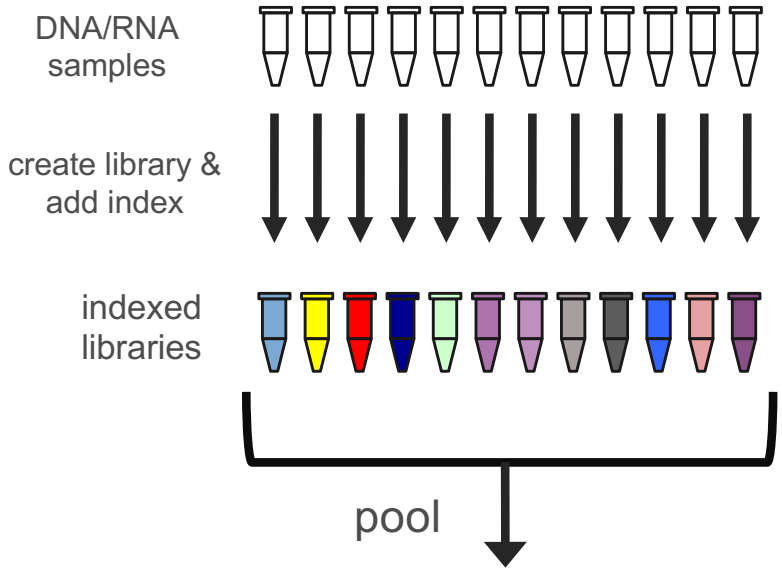


- P5/P7 bind to flow cell
- index1/2 sample specific barcodes
- Read 1/2 sequencing primers initiate sequencing

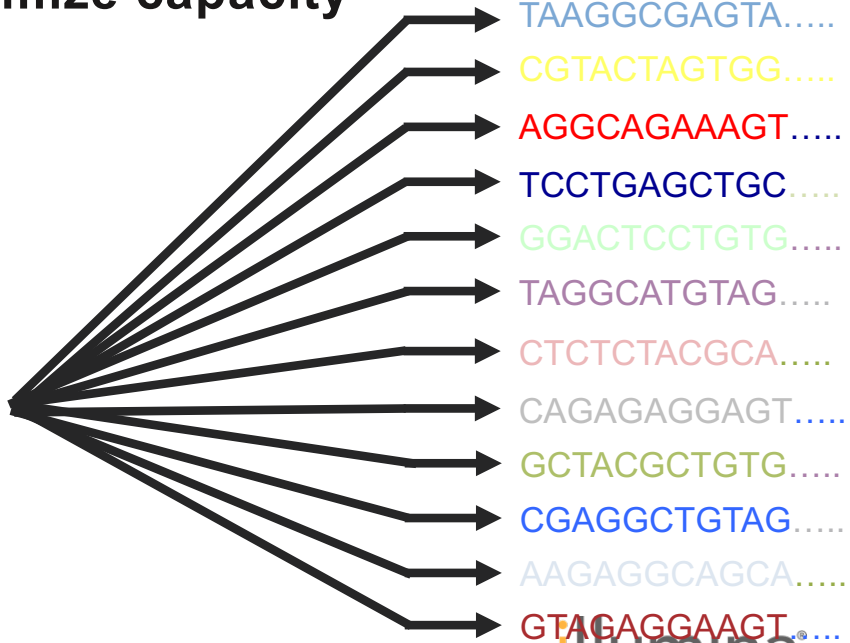
illumina®

Indexing
pooling samples





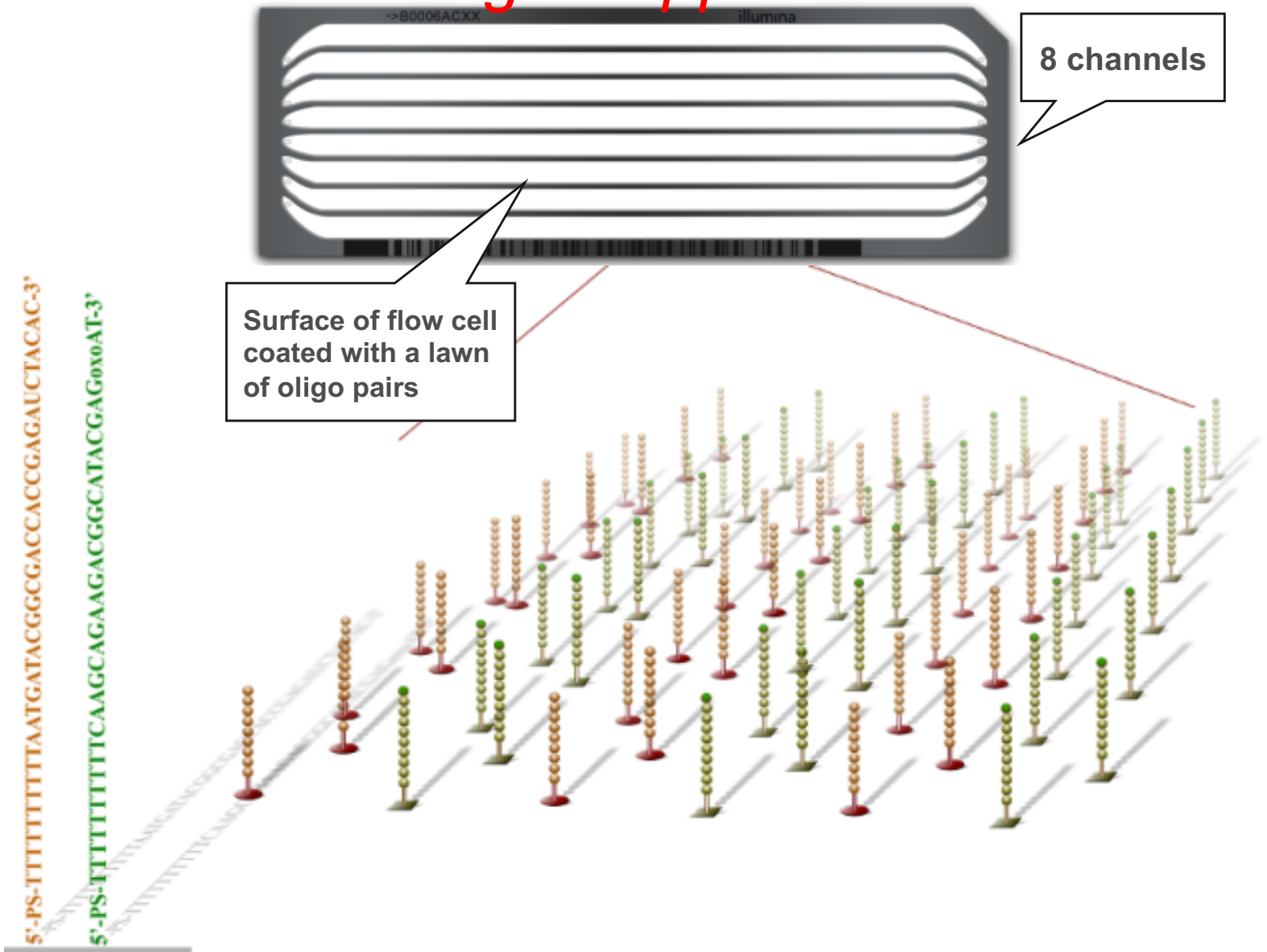
- **unique sequence "barcode"**
- **increase throughput**
- **maximize capacity**



illumina

The Flow Cell

Where the magic happens

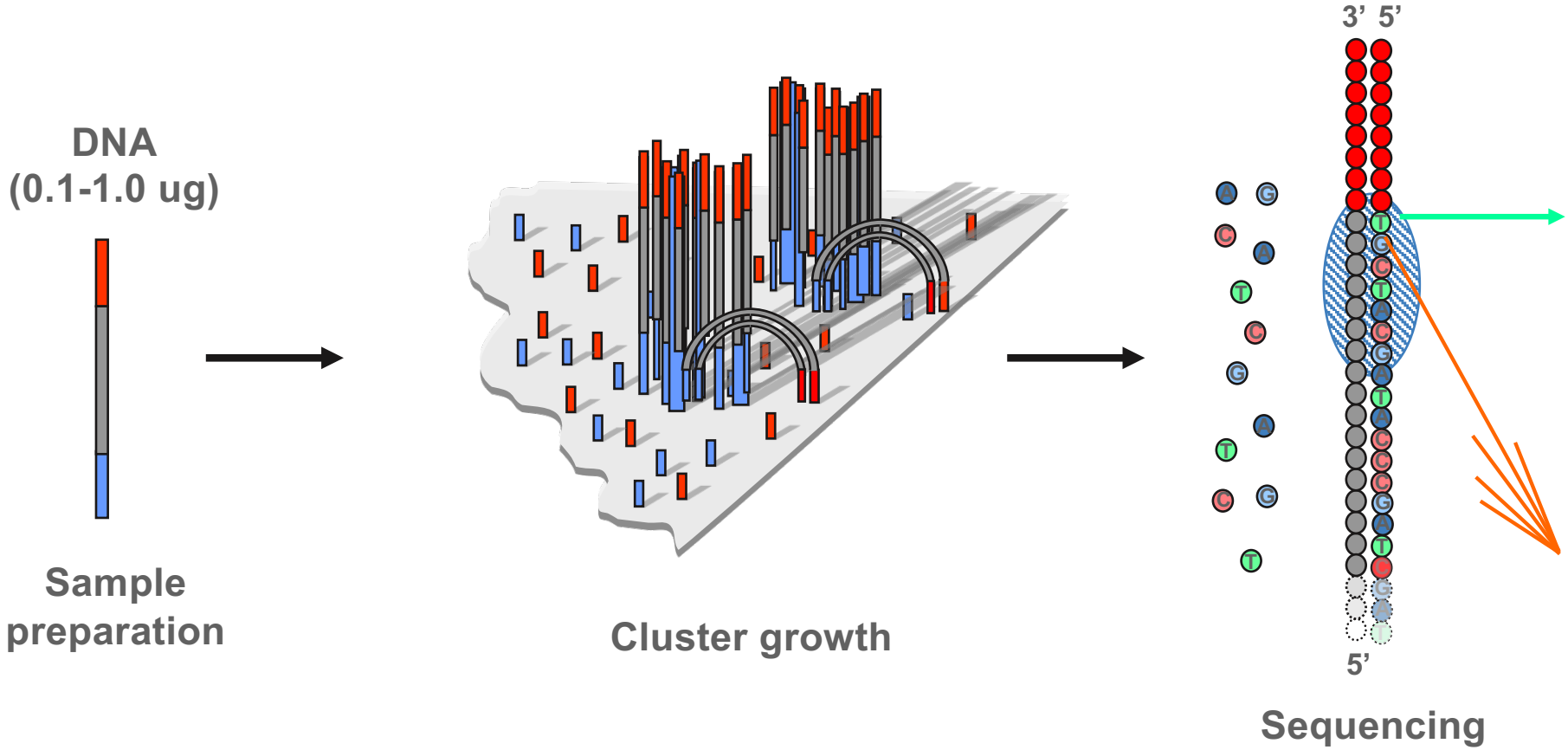


Simplified workflow

- **Clusters in a contained environment**
- **Sequencing performed in the flow cell on the clusters**

illumina®

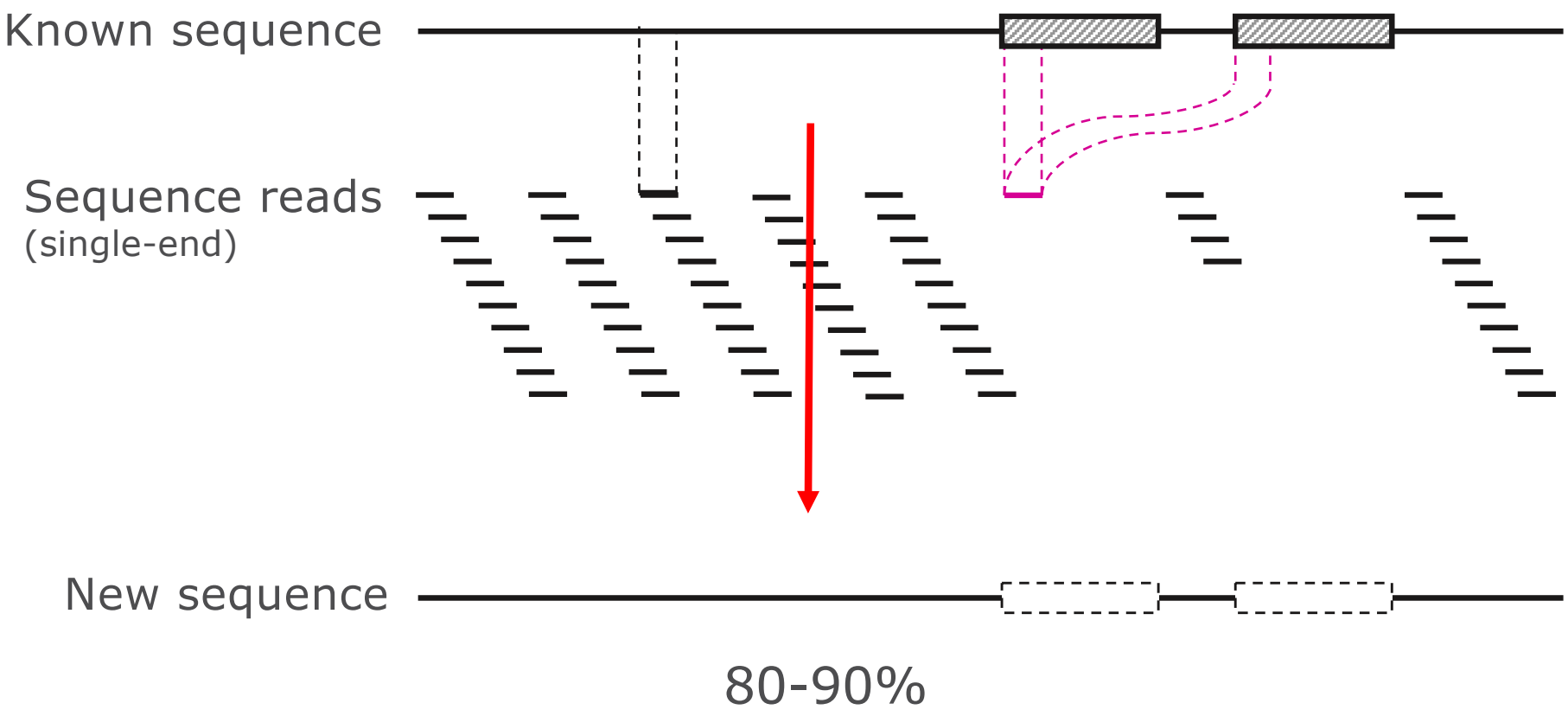
Sequencing by synthesis



illumina®

Paired-End Sequencing

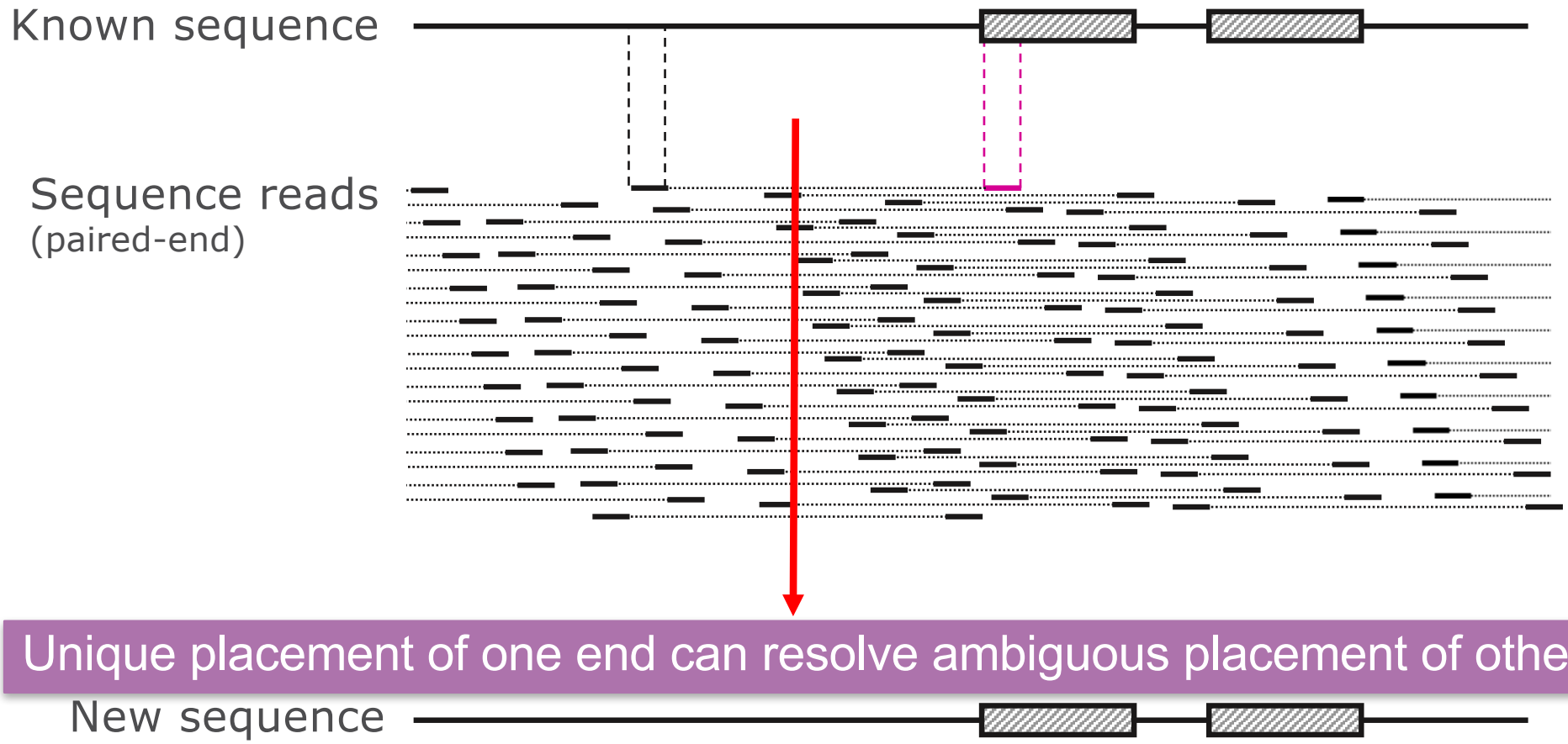
Extends the Power of the Technology



illumina®

Paired-End Sequencing

Extends the Power of the Technology

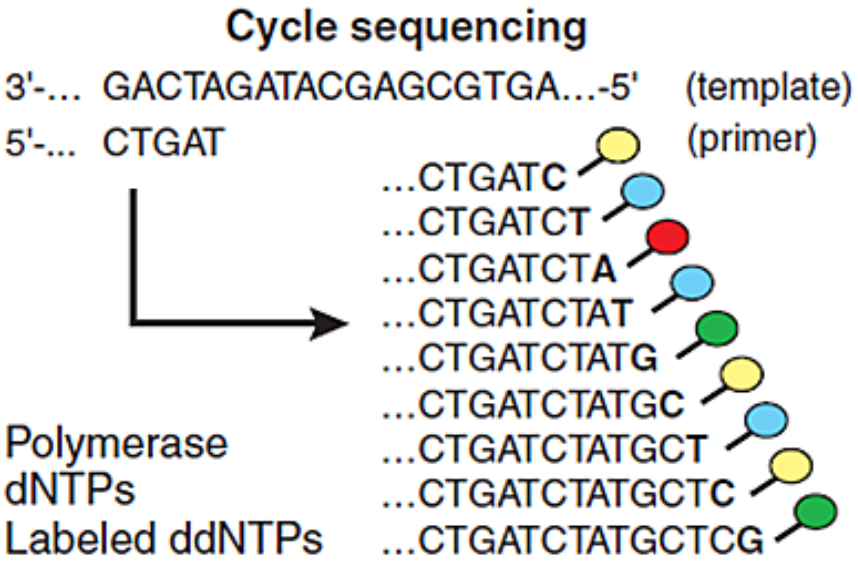


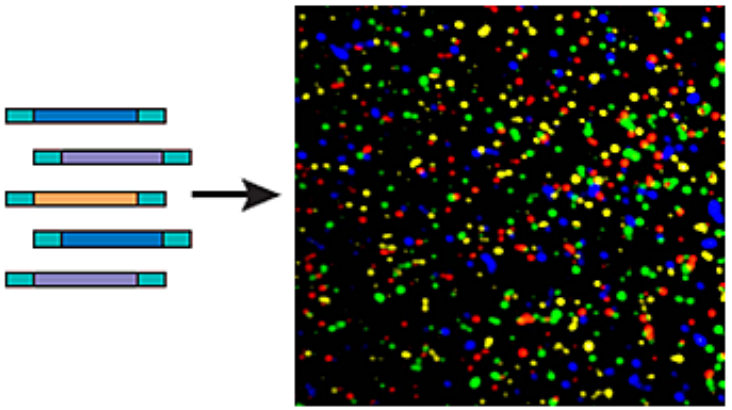
95 to >99%

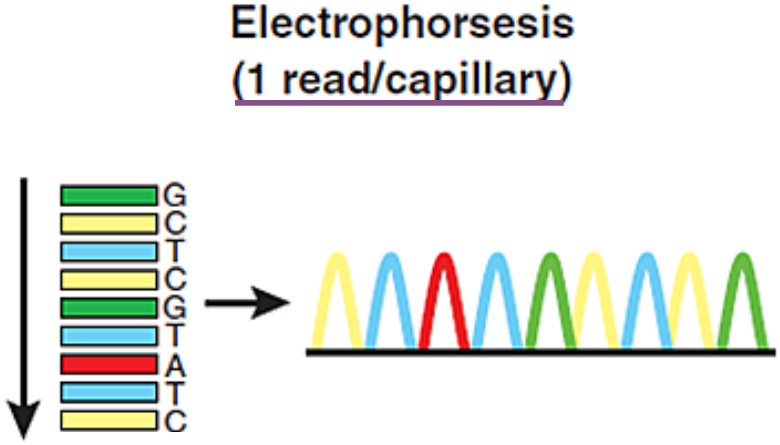
illumina®

Sanger vs. NGS

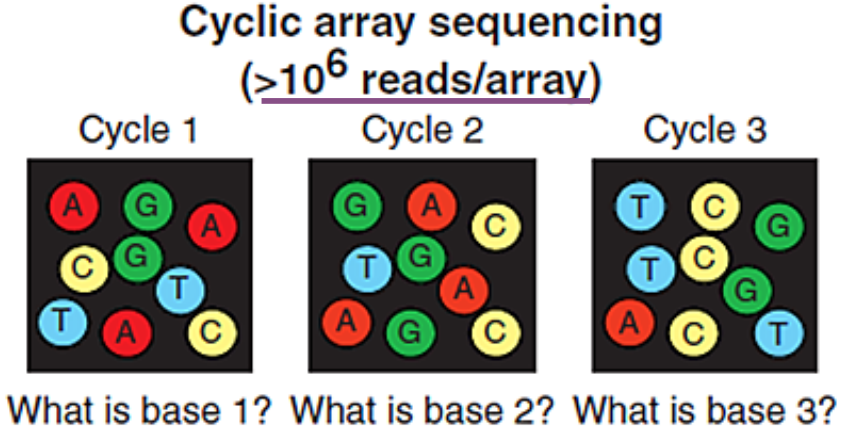
Method comparison







Sanger sequencing



NGS

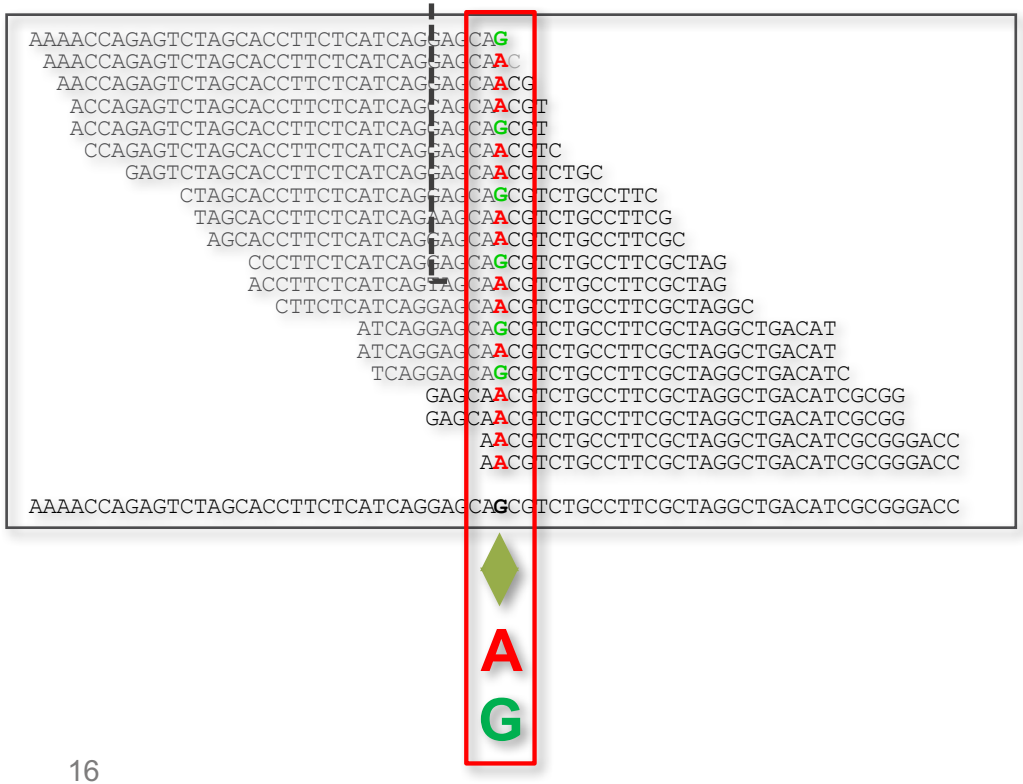
Adapted from Nat Biotechnol. 2008 Oct;26(10):1135-45

illumina®

NGS

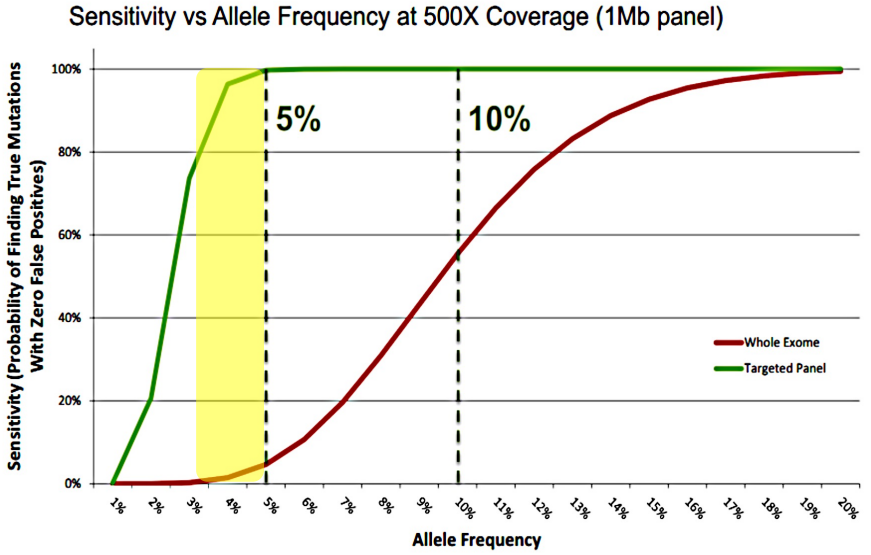
- **Limit of detection 3- 5%**

Coverage	30X	100X	1,000X	10,000X
A	28 – 29	95 – 97	950 – 970	9500 – 9700
G	1 – 2	3 – 5	30 – 50	300 – 500
Limit of Detection	~3 – 6%	3 – 5%	3 – 5%	3 – 5%



Depth of Coverage

- **The count of reads at a given base**
- **Sensitivity at desired allele**



illumina®

Sequencing Power for Every Scale

Targeted
RNA-Seq

NovaSeq 6000 – S1
500 Gb | 1.6B
2x150

NovaSeq 6000 – S4
6000 Gb | 10B
2x150

WES
WGS

WGS

HiSeq X Ten
1800 Gb | 6B
2x150

HiSeq X Five
1800 Gb | 6B
2x150

RNA-seq
Exomes
WGS
Methylation

HiSeq 4000
1500 Gb | 5B
2x150

HiSeq 3000
750 Gb | 2.5B
2x150

HiSeq 2500
1000 Gb | 4B
2x125

NextSeq
120 Gb | 400M
2x150

Prokaryote
Targeted
QC

MiSeq
1.2 Gb | 4M
2x150

MiSeq
15 Gb | 25M
2x300

MiniSeq
7.5 Gb | 25M
2x150

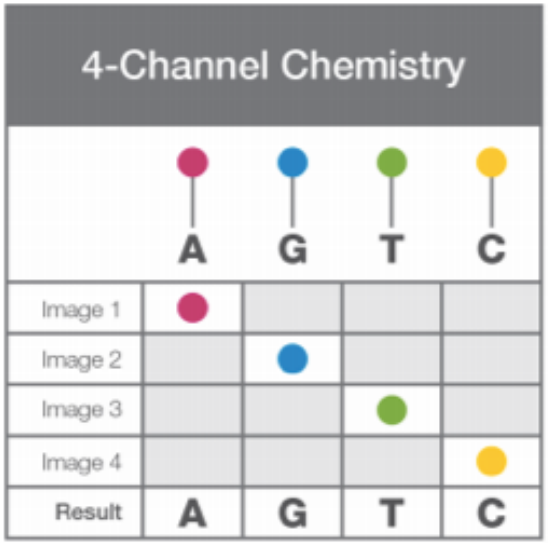
Increasing system output & price

Decreasing price per GB

illumina®

1, 2 and 4 Channel Chemistries

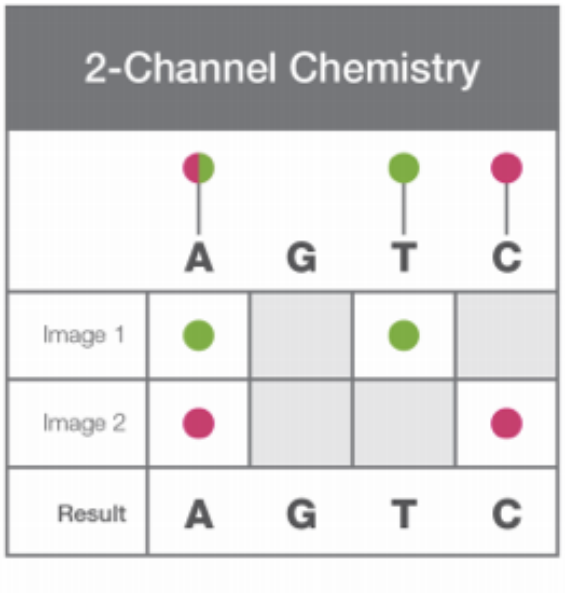
- **Channel ~= Dye**



MiSeq

HiSeq

- 4 dyes
- 4 images/cycle

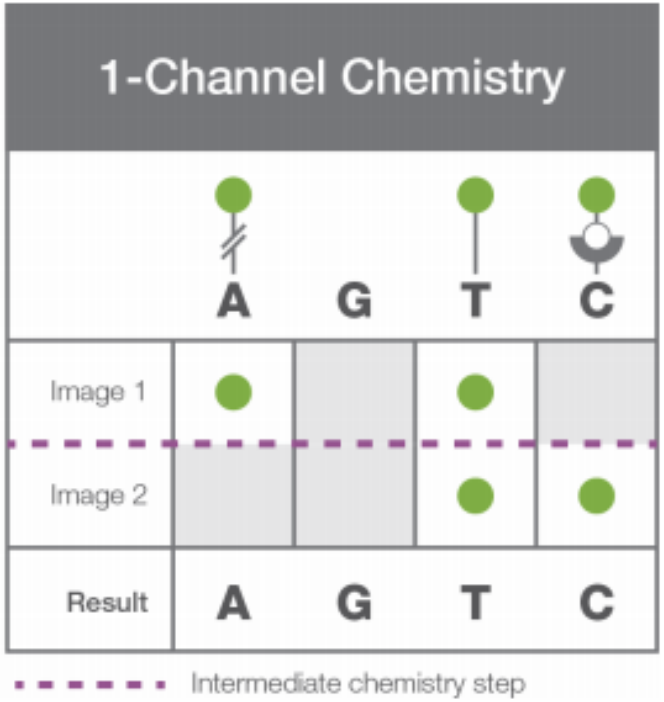


MiniSeq

NextSeq

NovaSeq

- 2 dyes
- 2 images/cycle



iSeq 100

- 1 dye
- 2 chemistry steps
- 2 imaging steps

illumina®

Whole Genome

illumina®

Whole-Genome Sequencing

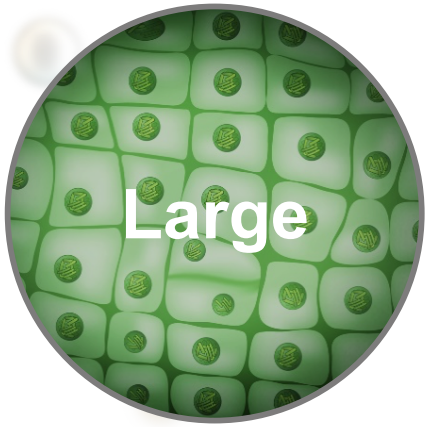
Human microbiome

Microbiology

Public health research

Metagenomics

Amplicon sequencing



Agrigenomics
(maize, wheat, bovine, etc.)

Model organisms
(fruit fly, mouse, zebrafish, etc.)

Plant / animal research

Cancer genomics research

Variant detection

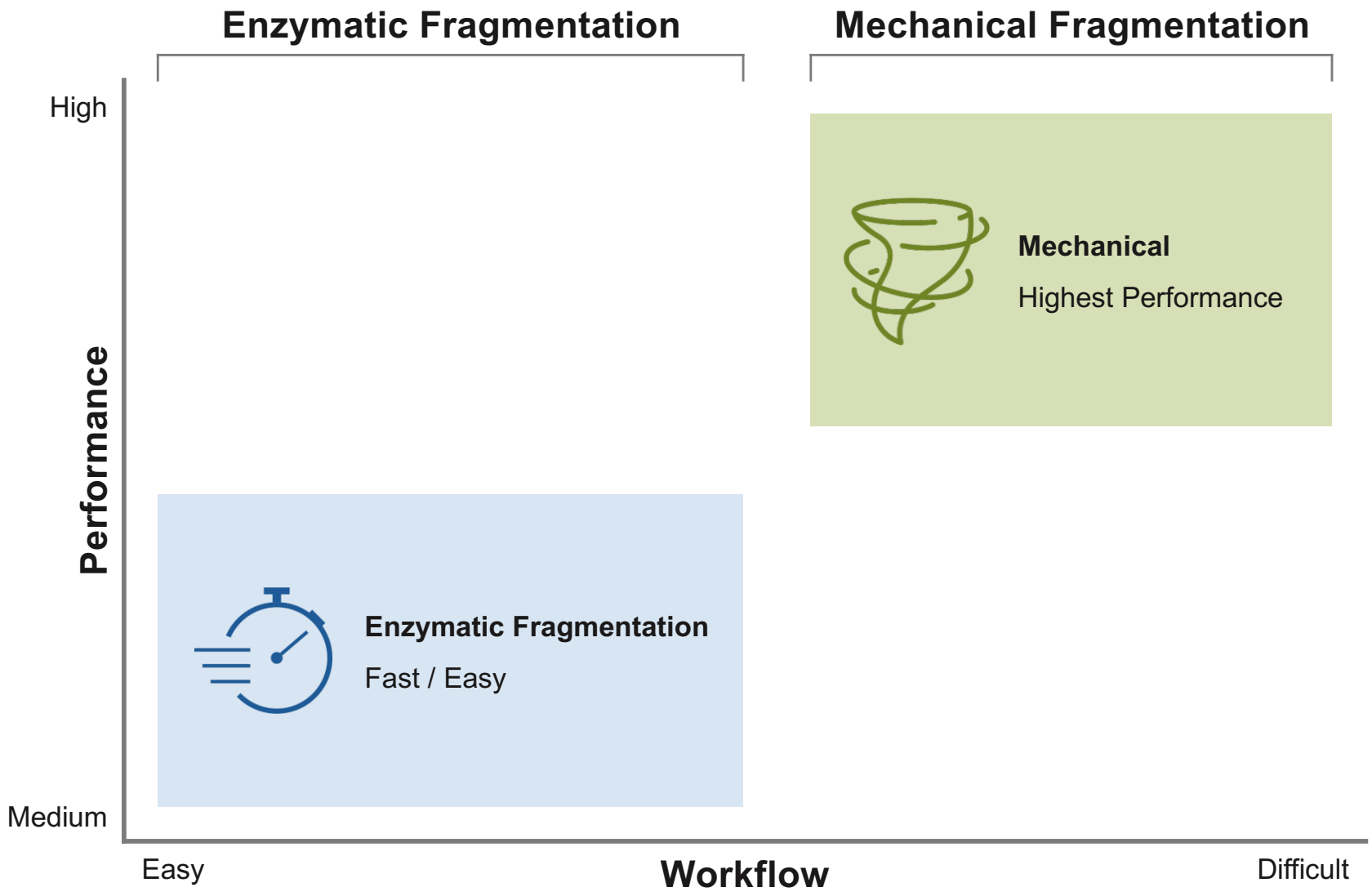
Genetic risk studies

Population genetics

illumina®

Tools for DNA Library Preparation

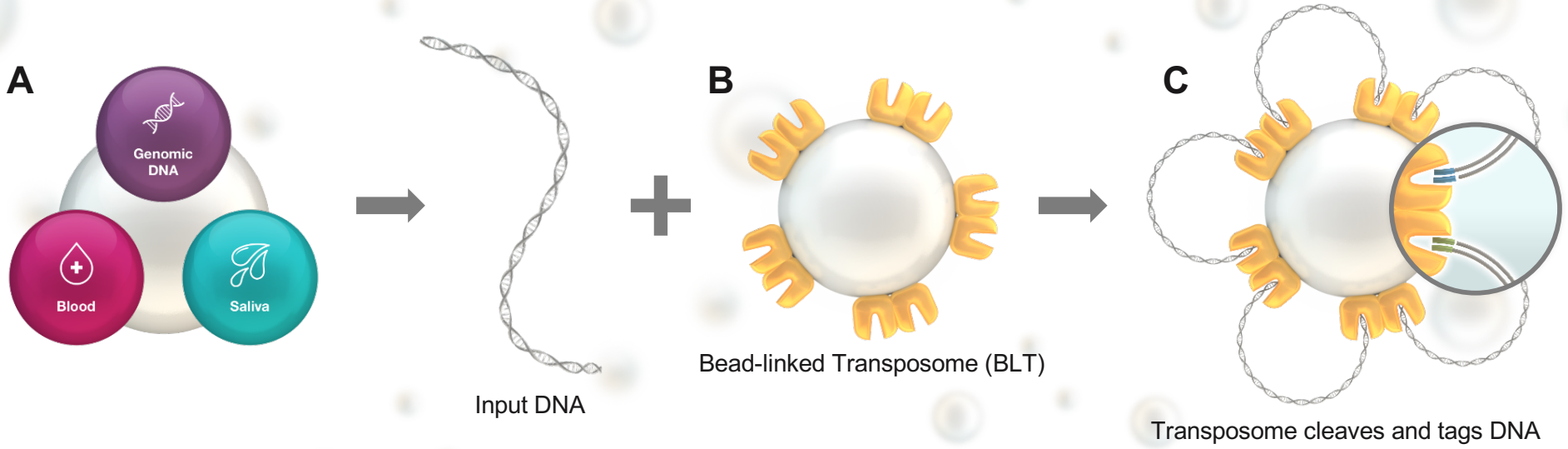
Fast or high performance



illumina®

Nextera™ DNA Flex Library Prep

Workflow overview



A **Isolate and purify DNA**
gDNA, blood, saliva, or microbe

B **Add DNA to bead-linked transposomes (BLT)**
Transposome attached to magnetic beads

C **DNA is tagmented and remains bound to the bead**

No additional tagmentation can occur after bead saturation

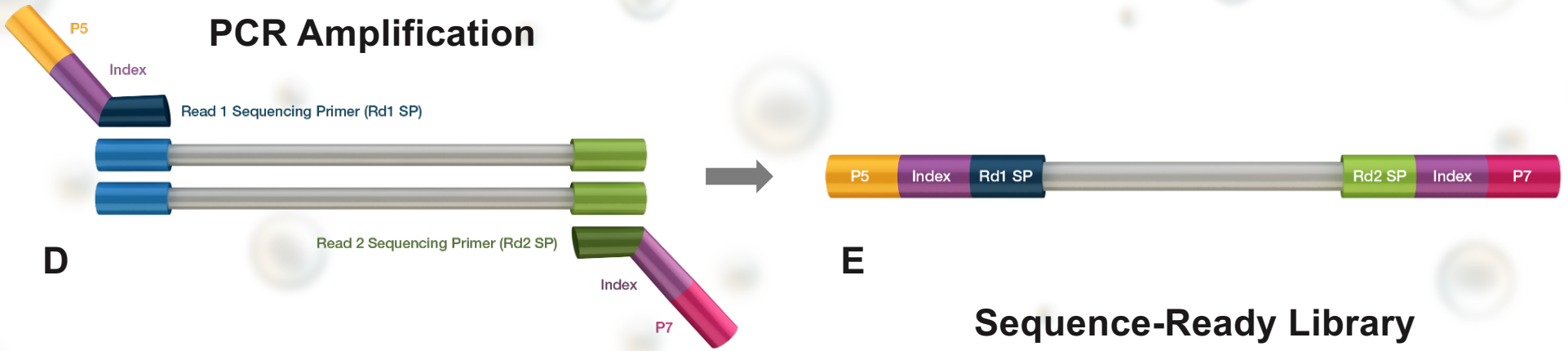
Allowing a large DNA input range (1–500ng)

Resulting in consistent insert size and normalized libraries

illumina®

Nextera™ DNA Flex Library Prep

Workflow overview



D **Index and sequencing adapter addition through PCR**

E **Normalized sequence-ready library**

Library quantification, QC, and normalization not required

(100ng–500ng)

Allowing a large range of DNA input

illumina®

Flexible sample input

gDNA, Blood, Saliva, Microbes

gDNA

User guide & kit supported process

Blood

User guide & kit supported process

Blood input requires Illumina Flex lysis reagent

Blood punch card

Demonstrated protocols available at Illumina.com

Saliva

User guide & kit supported process

Saliva input requires Oragene saliva collection kit

Microbial colony

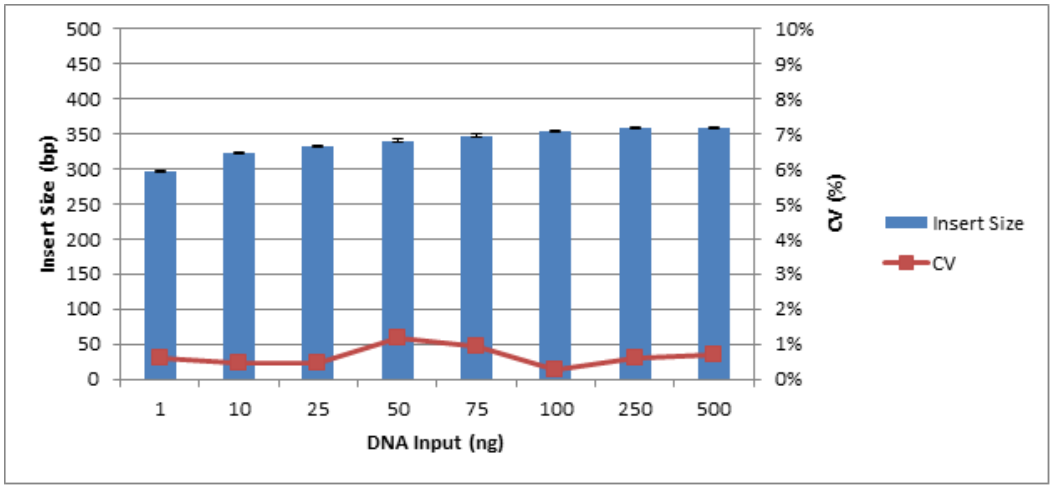
Demonstrated protocols available at Illumina.com

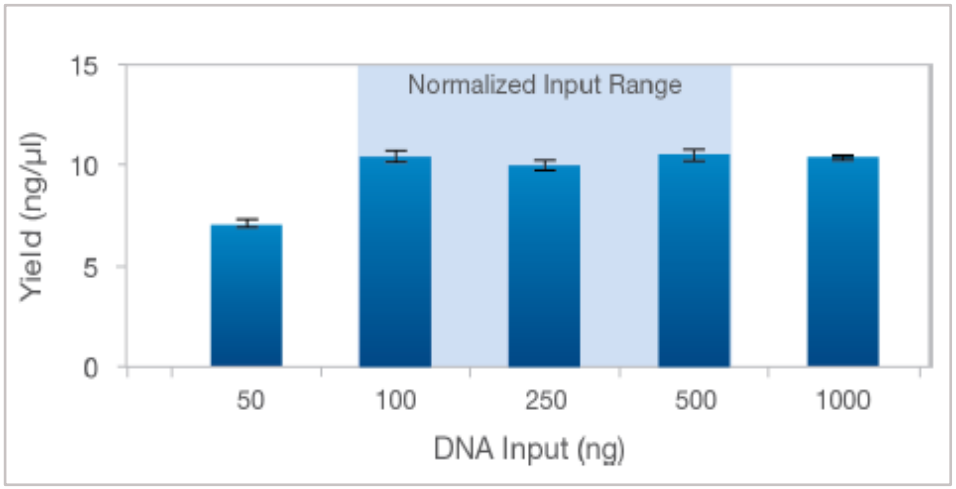
No extra DNA quantification step required post DNA extraction

illumina®

Wide DNA Input

Consistent insert sizes and Normalization





- **Consistent insert size obtained with the use of a wide DNA input range**

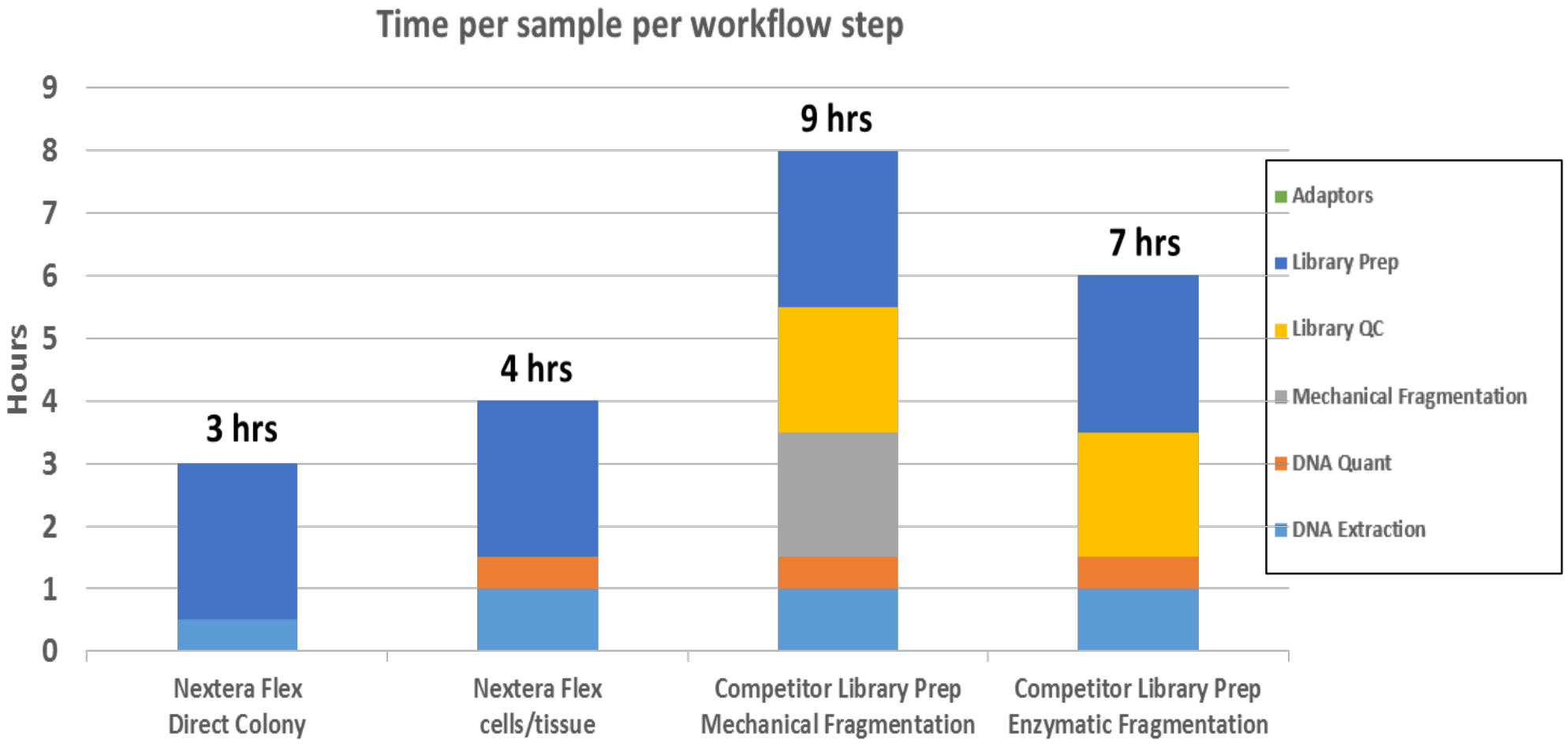
- **Normalized libraries are be obtained with:**
  - 100ng-500ng gDNA input
  - Use of the liquid blood, saliva, dried blood, or bacterial colony protocol

With Nextera DNA Flex, precise input quantification is not required to yield consistent DNA insert sizes and normalized libraries

*Data maintained in Illumina internal files 2017

illumina®

Save Time and Increase Efficiency



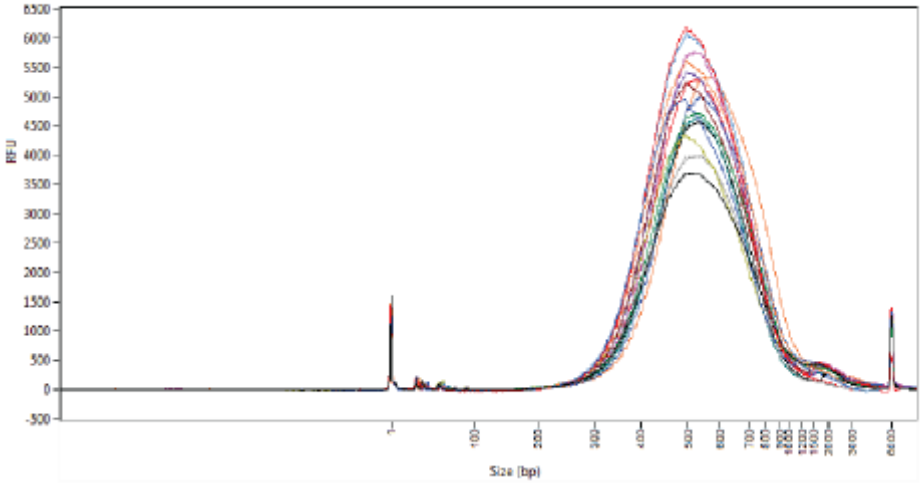
Save Time with Nextera™ DNA Flex

illumina®

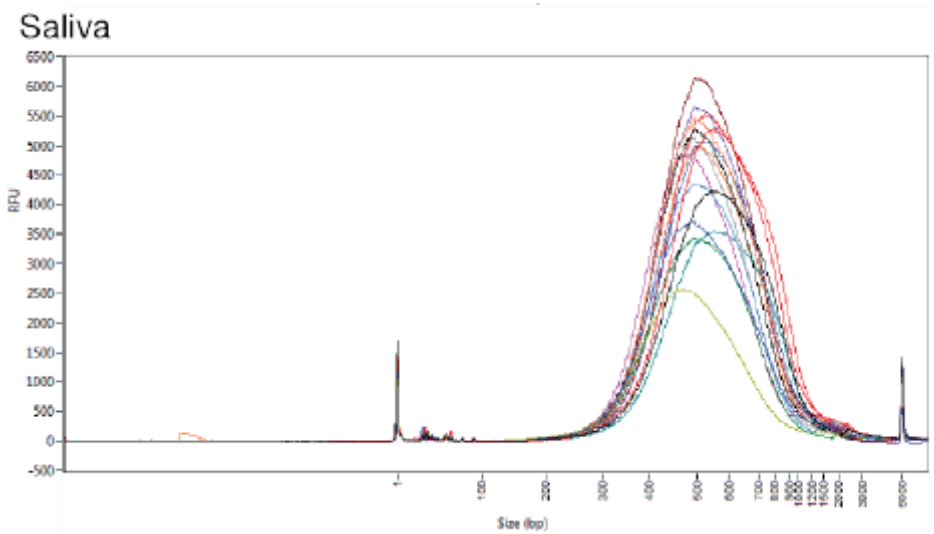
Quality Assessment of Nextera DNA Flex

Blood & Saliva libraries (fragment analyzer)

Blood



Saliva



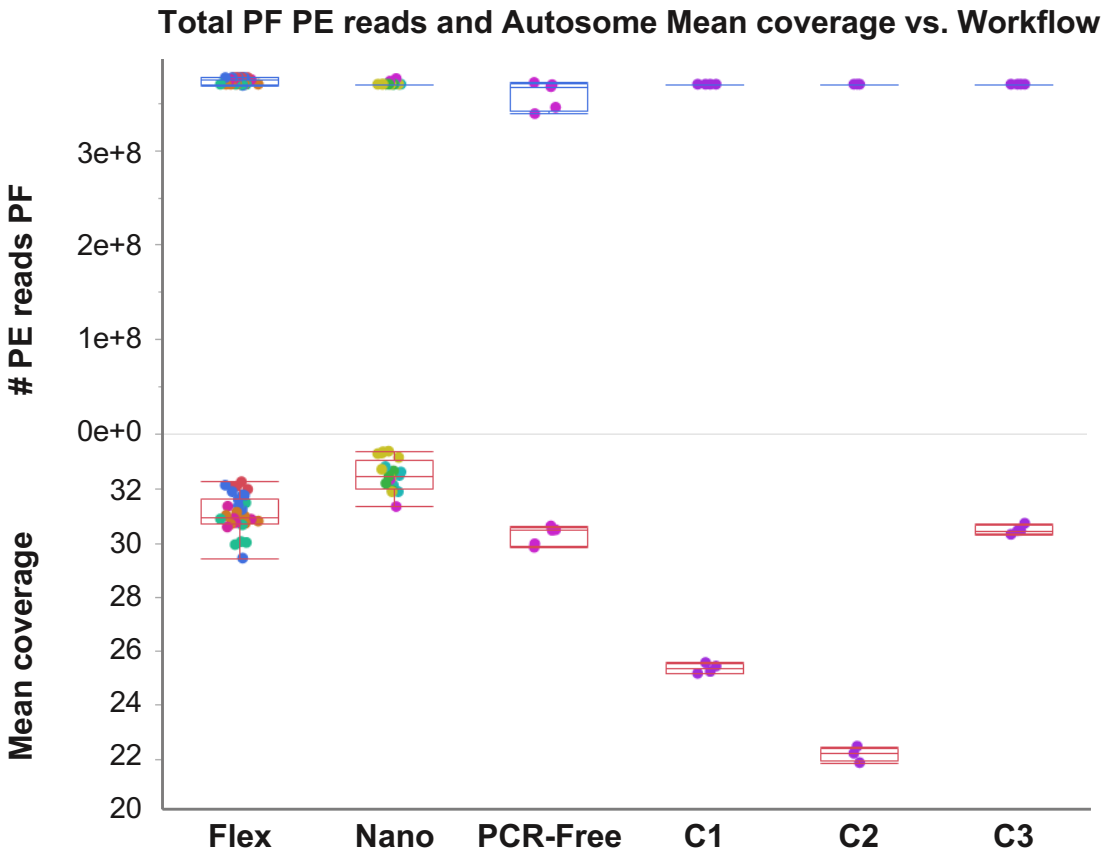
	Blood	Saliva
Average Yield (16 samples)	**9.99 ng/uL**	**8.96 ng/uL**
Fragment Size	**324**	**317**

*Data maintained in Illumina internal files 2017

- **Library yield and size consistency can be obtained with the use of blood or saliva**

illumina®

Characteristics of Genome Builds



Flex

Nextera™ DNA Flex

Nano

TruSeq™ DNA Nano

PCR-Free

TruSeq™ PCR-Free

C1

Competitor K_H

C2

Competitor K_HP

C3

Competitor N

30x coverage not attained for C1 and C2 due to short insert size

- In part, this reflects ambiguity of the workflow
- Trimmed to 2x100bp, subsampled to 550M reads (instead of 370M), re-ran analysis: conclusions unchanged

Automation

At launch protocols available

Hamilton Star

Eppendorf epMotion 5075t

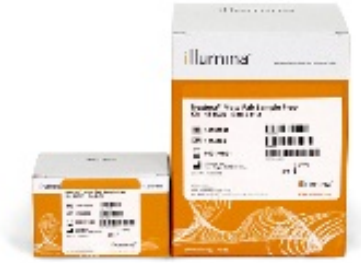
Following launch protocols available on more platforms

- Agilent
- Beckman Coulter
- Perkin Elmer
- Tecan

- **96 sample kit size is designed to be automation friendly**
  - 96 library prep kit reagents provided in increments to support smaller batches (< 96 sample runs)
  - 96 dual index kit provided in a 96 well plate format

Summary: Advantages of Nextera DNA Flex

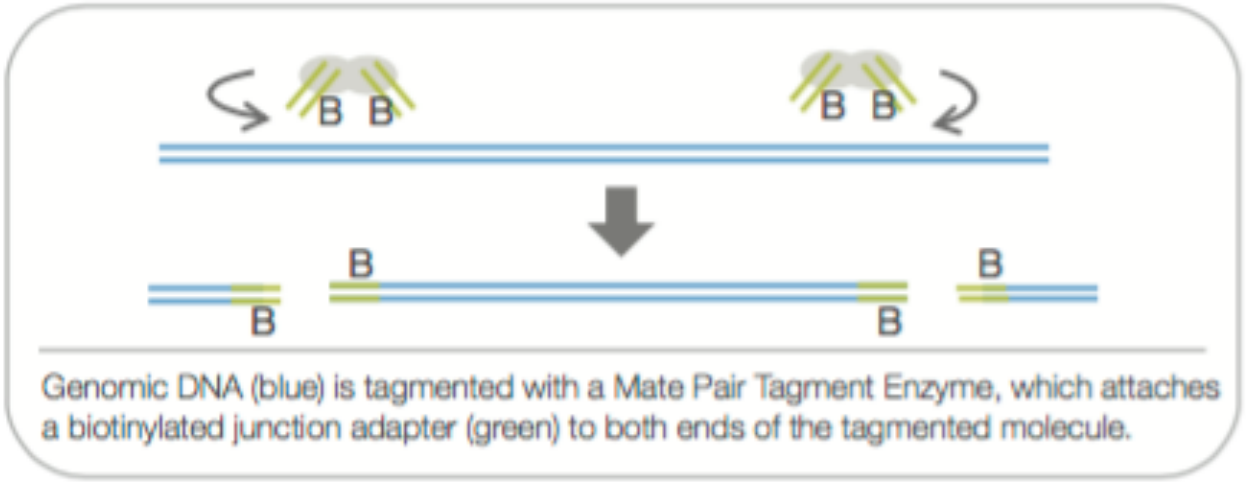
- **Full workflow is < 4 hours**
- **Wide input range**
- **Normalized libraries between 100-500 ng DNA input**
- **Consistent insert size independent of DNA input**
- **High Coverage Uniformity**
- **Can process raw sample inputs**
  - Bacterial colonies, Blood, DBS, and Saliva
- **96 indexes now, 384 by the end of year**
- **Any genome**



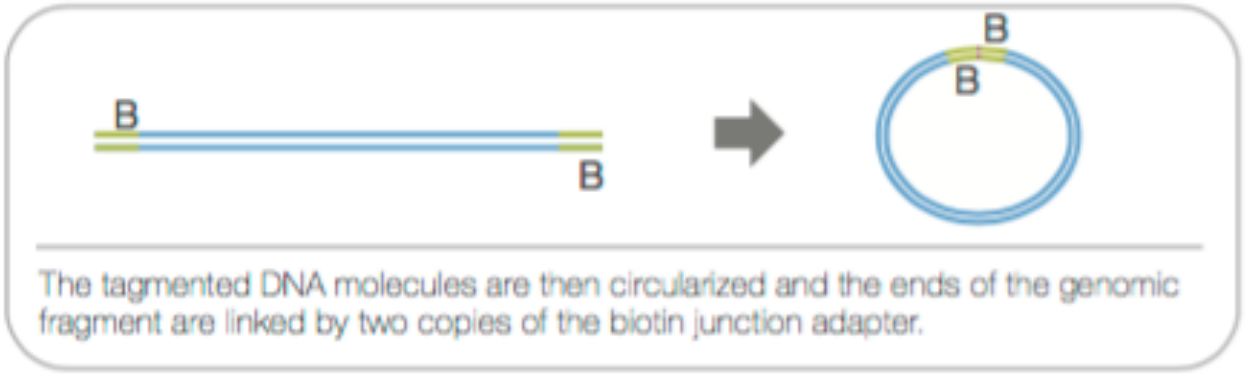
Nextera® Mate Pair Library Preparation Kit

- **An optimized library preparation method for long-insert libraries, empowering *de novo* sequencing and structural variant detection**
- **Ideal for:**
  - *de novo* sequencing
  - Genome finishing
  - Structural variants
  - Any genome
- **Low input (1ug Gel Free, 4ug Gel Plus)**
  - Gel Free smaller genomes broad range of fragment sizes, limited sample
  - Gel Plus larger more complex genomes narrow fragment size
- **1.5 to 2 day prep time**

Nextera® Mate Pair Library Preparation Kit



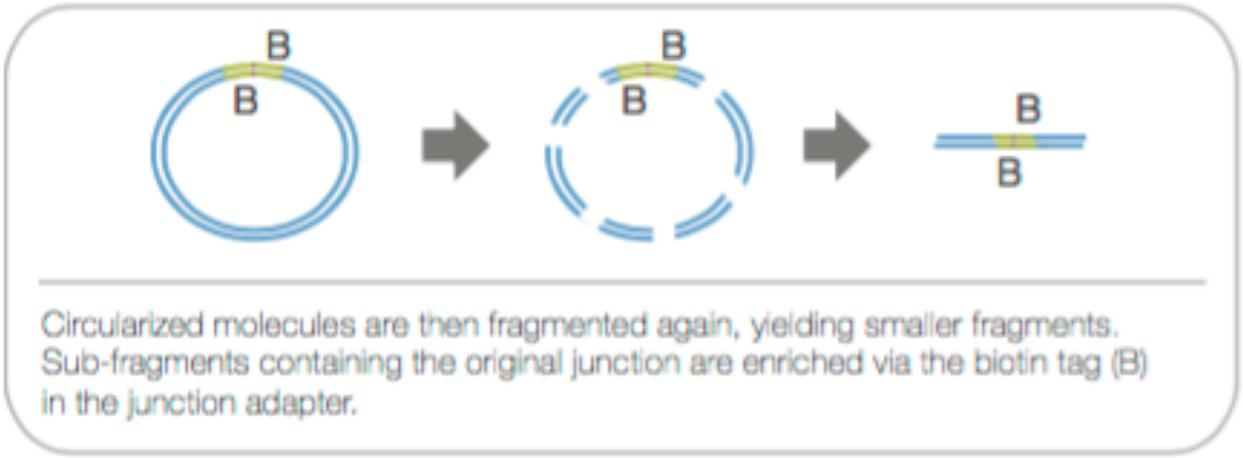
Genomic DNA (blue) is tagmented with a Mate Pair Tagment Enzyme, which attaches a biotinylated junction adapter (green) to both ends of the tagmented molecule.



The tagmented DNA molecules are then circularized and the ends of the genomic fragment are linked by two copies of the biotin junction adapter.

illumina®

Nextera® Mate Pair Library Preparation Kit

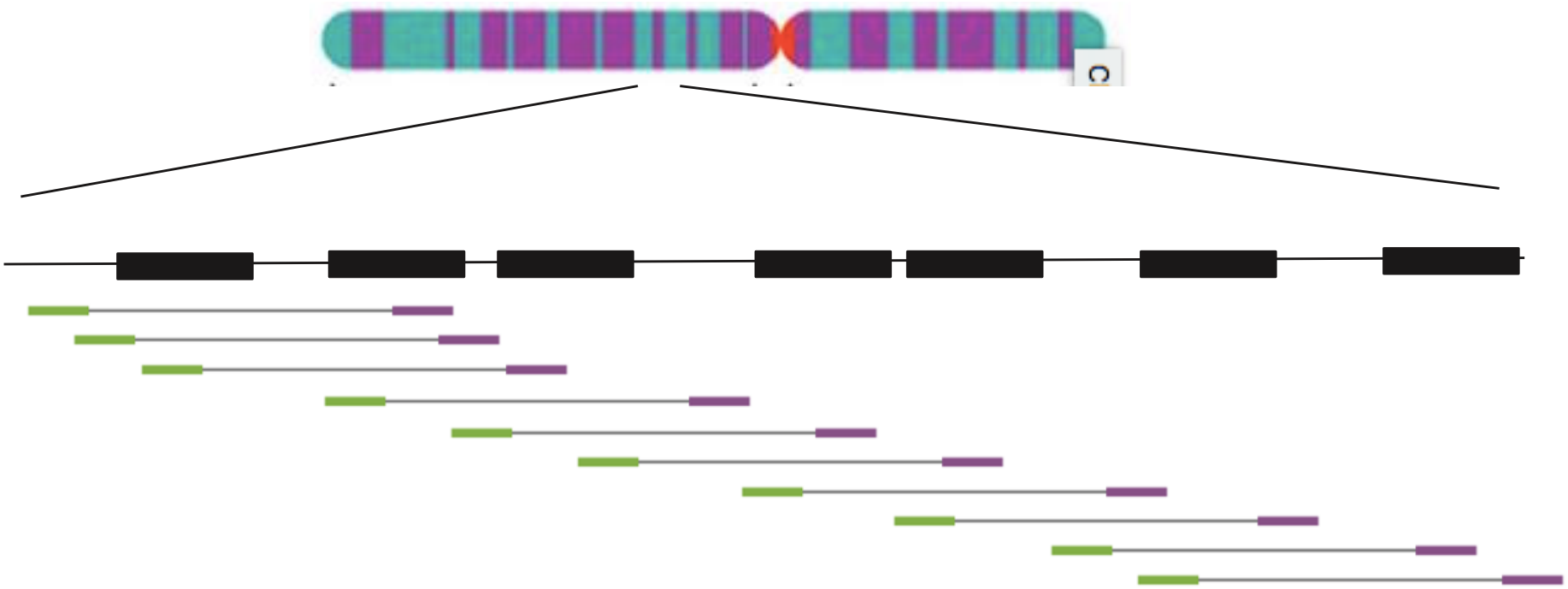


Circularized molecules are then fragmented again, yielding smaller fragments. Sub-fragments containing the original junction are enriched via the biotin tag (B) in the junction adapter.

After End Repair and A-Tailing, TruSeq DNA adapters (gray and purple) are then added, enabling amplification and sequencing.

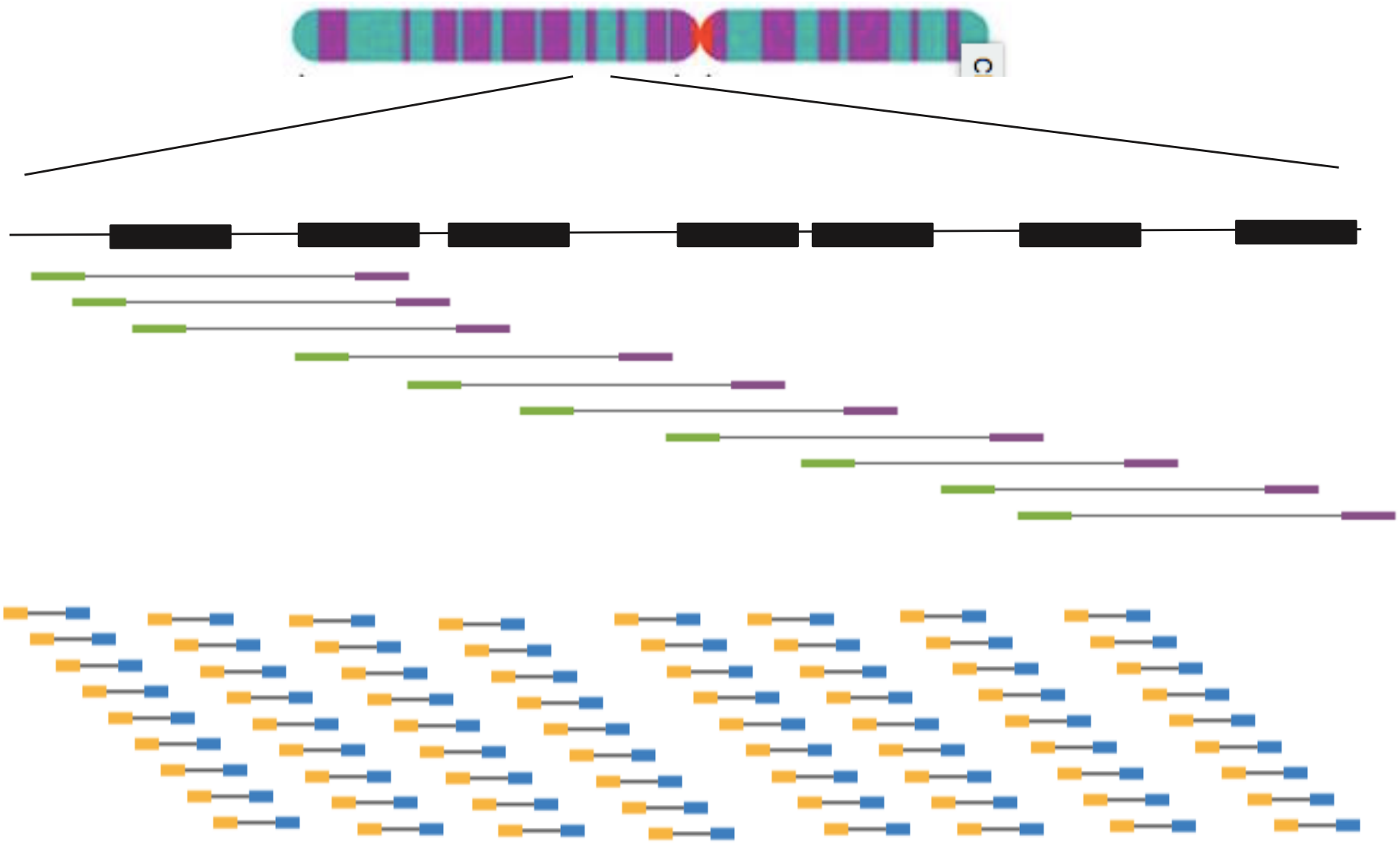
illumina®

Nextera® Mate Pair Library Preparation Kit



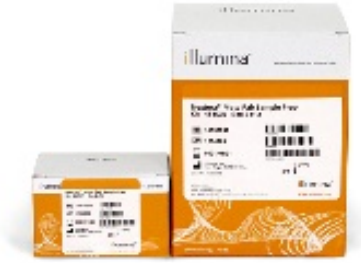
illumina®

Mate Pair Combined with Short Inserts

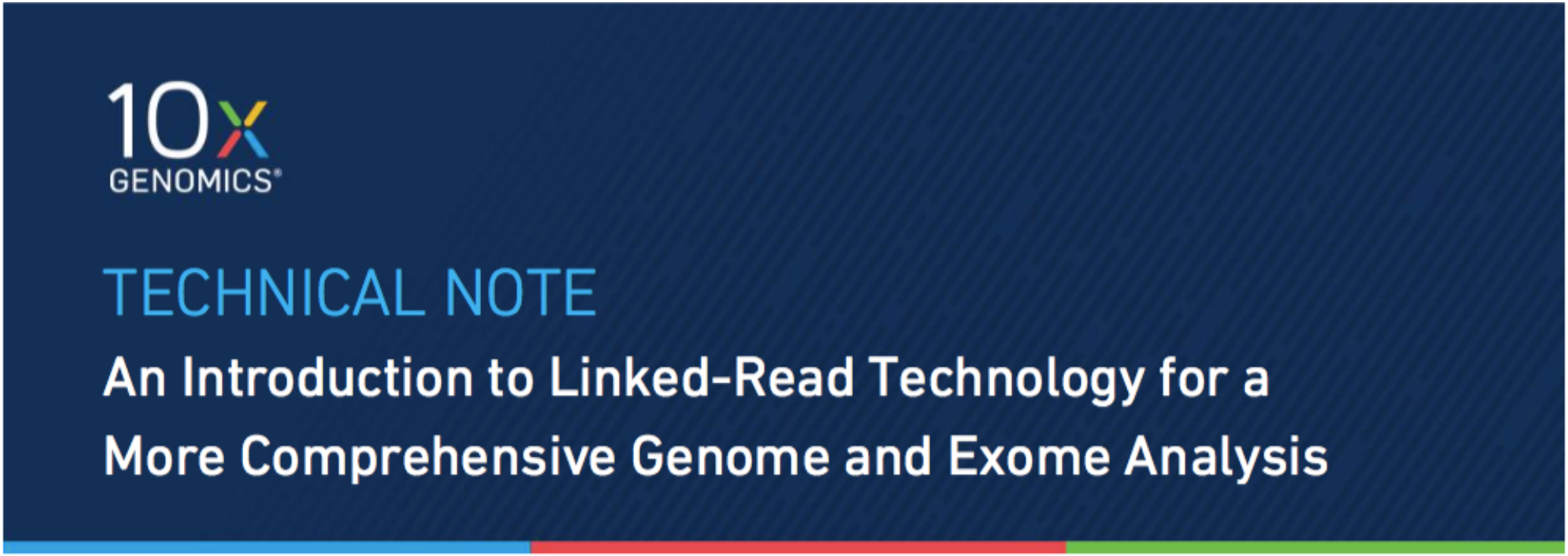


Alternative Approaches

- **Long Read Technology**
  - PacBio and Oxford Nanopore
  - Combination approach
- **Linked Reads (10X Genomics)**
- **Hi-C (chromatin conformation capture sequencing)**



Linked-Reads



illumina

Linked-Reads

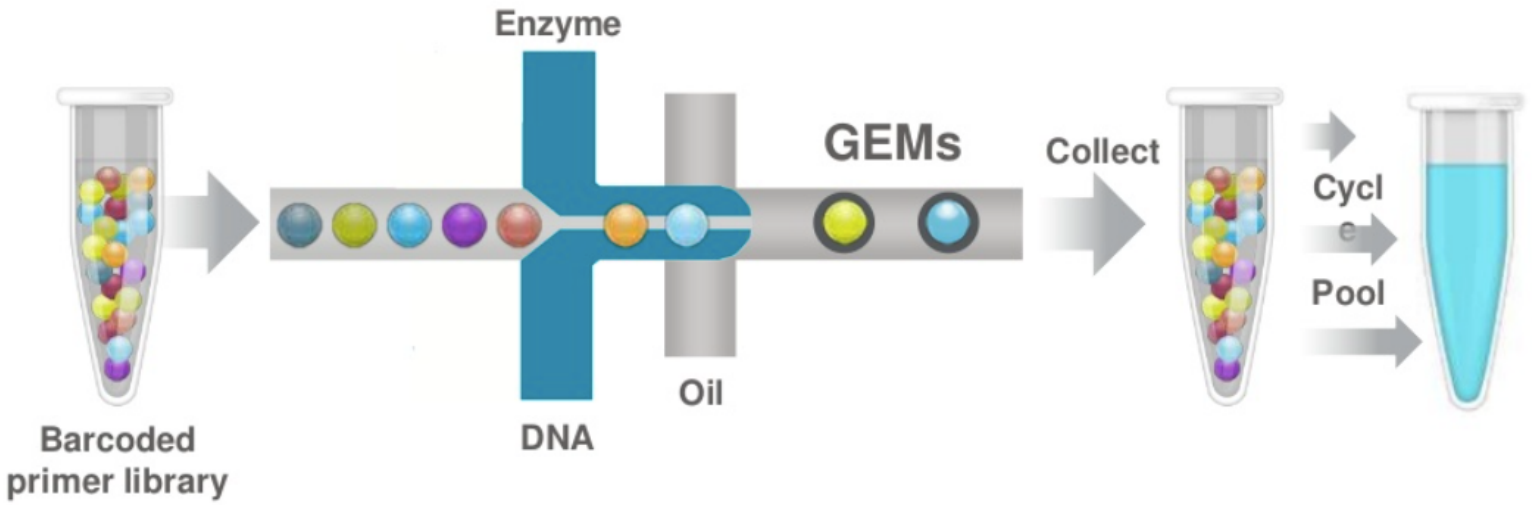


Fig 2. Chromium™ Technology mixes functionalized gel beads containing unique barcodes with enzymes and a limiting amount of genomic DNA to create >1,000,000 uniquely addressable partitions in minutes. Using a limiting dilution of molecules allows the correct mapping of reads to their corresponding molecules.

illumina®

Linked-Reads

Fig 1. Reads (short blue lines) are generated from each high-molecular weight gDNA molecule (long blue line). Reads from the same molecule will share the same barcode (shown in gold).

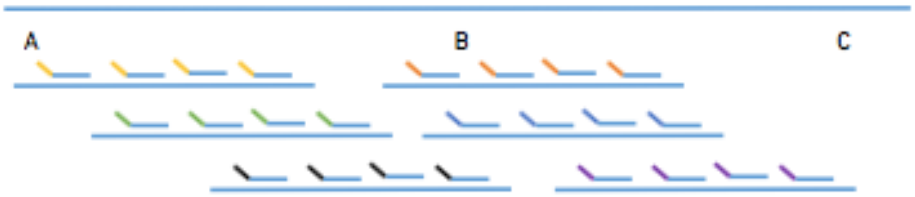


Fig 4B. Linked-Reads, with only a slight increase in standard sequencing, allow increased physical coverage that provides the power to link distant loci and reconstruct long range haplotypes. In the figure above, with the superior physical coverage of Linked-Reads, the three loci (A, B and C) can be linked.

illumina®

Linked-Reads

Long range haplotype reconstruction

Linked-Reads enable large scale haplotype reconstruction. Fig 5 shows a standard run of the NA12878 genome. Alternating colors delineate phase blocks. At standard sequencing depths, phase block lengths are determined primarily by the length of the input DNA and the diversity of the sample. For the run shown in this figure, the N50 phase block length is 4.6 Mb, and the longest phase block is 31.2 Mb. The input molecule length was 80 Kb.

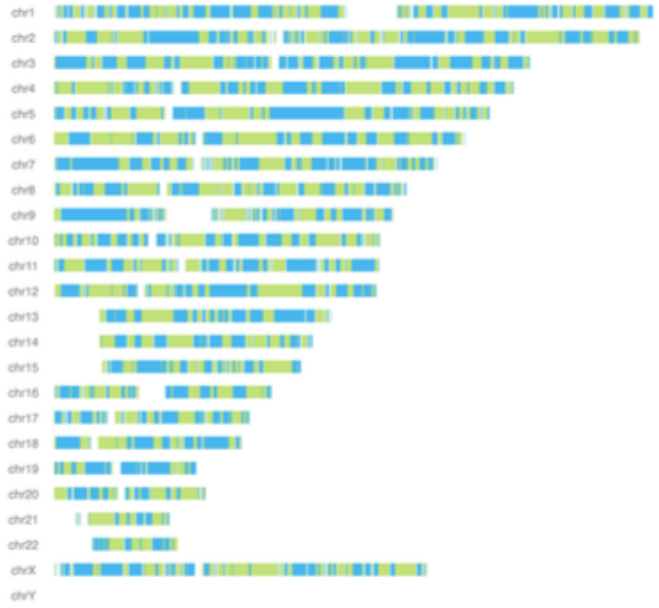


Fig 5. Standard run of the NA12878 genome. Alternating colors delineate phase blocks.

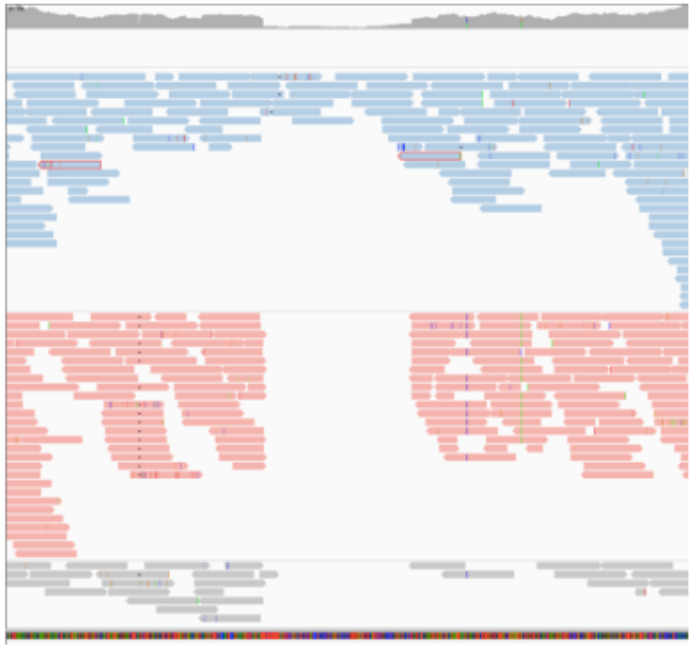
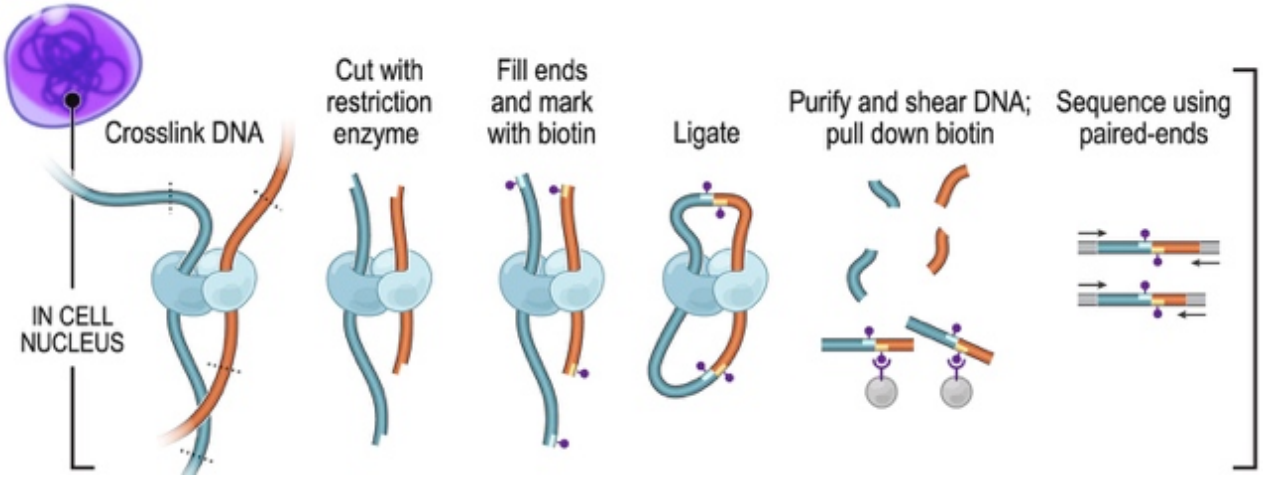


Fig 6. A 325 bp heterozygous deletion detected using Linked-Reads. Reads are partitioned into distinct Haplotypes. Haplotype1 shown in blue, haplotype 2 in pink.

illumina®

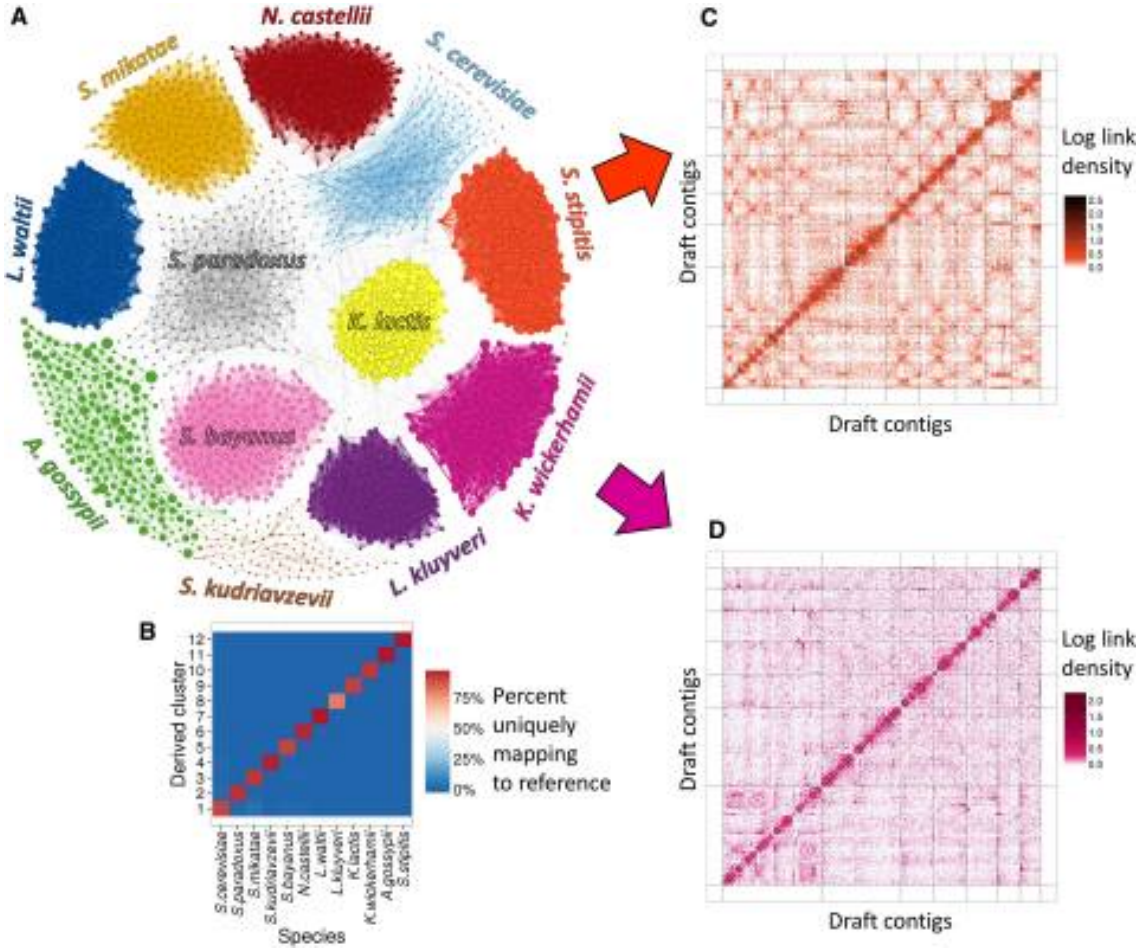
Hi-C Sequencing

- **Chromatin conformation capture sequencing**
- **Used to analyze chromatin interactions**
  - DNA/protein complexes are crosslinked
  - sample is fragmented and DNA ligated and digested
  - DNA fragments are PCR-amplified and sequenced



Hi-C Sequencing in Metagenomics

- Contact probability maps from Hi-C enable deconvolution of shotgun metagenomic assemblies
- Hi-C enables two different signals
  - Intracellularity of each pair which enables species level deconvolution
  - Correlation of Hi-C linkage with chromosomal distance, which enbales scaffolding of *de novo* assemblies



Burton, J. N., Liachko, I., Dunham, M. J. & Shendure, J. Species-Level Deconvolution of Metagenome Assemblies with Hi-C–Based Contact Probability Maps. G3: Genes|Genomes|Genetics 4, 1339–1346 (2014).

illumina®

Hi-C Sequencing in Metagenomics

- Contact probability maps from Hi-C enable deconvolution of shotgun metagenomic assemblies
- Hi-C enables two different signals
  - Intracellularity of each pair which enables species level deconvolution
  - Correlation of Hi-C linkage with chromosomal distance, which enbales scaffolding of *de novo* assemblies

Burton, J. N., Liachko, I., Dunham, M. J. & Shendure, J. Species-Level Deconvolution of Metagenome Assemblies with Hi-C–Based Contact Probability Maps. G3: Genes|Genomes|Genetics 4, 1339–1346 (2014).

Sequence Hub Cloud

Simplifying bioinformatics

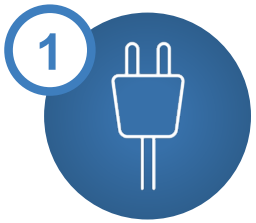
For Research Use Only. Not for use in diagnostic procedures.

illumina®

Multiple Layers of Security

- **Secure Data**
  - Data encrypted in transit, Genomic data encrypted at rest,
  - Access control, activity logging
- **Secure Employees**
  - Background checks, training on secure development
  - Training on HIPAA
- **Secure Physical Environment**
  - Built on AWS, ISO 27001 certified data centers
- **Secure Application**
  - Code reviews, penetration testing

Four Key Features



Plug and play with tight instrument integration

Easy **sharing and collaboration** worldwide

Simple **push-button analysis** with public and private analysis tools

Advanced **automation and integration**

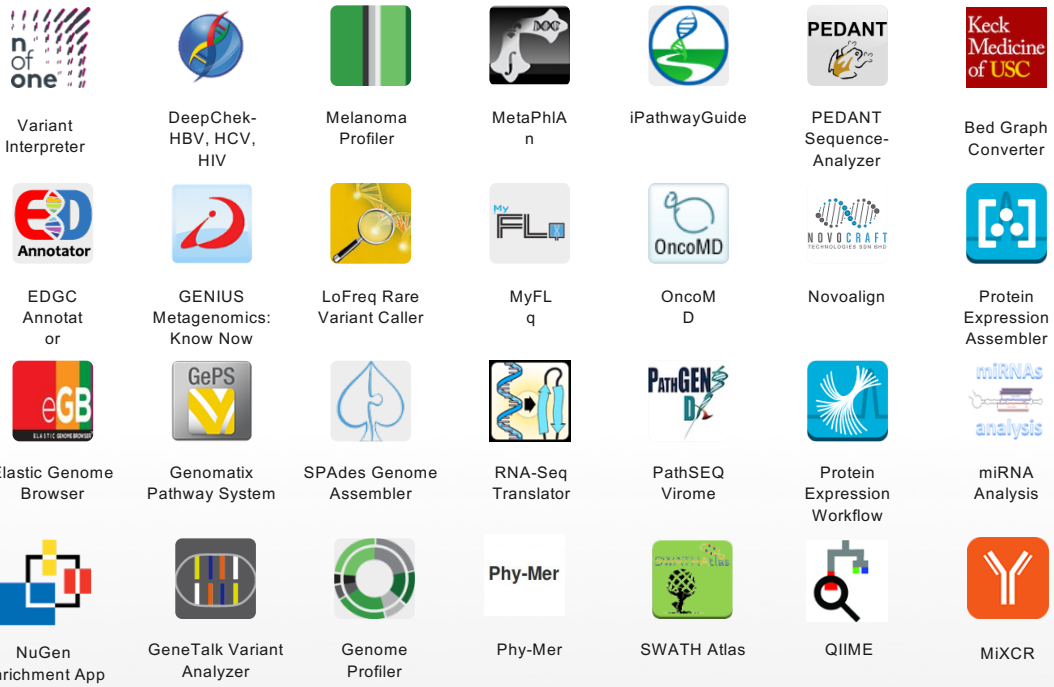
Simple **push-button analysis** with public and private analysis tools

- Over 90 published Apps supporting all of Illumina library prep kits
  - — TruSeq Amplicon, TruSeq Targeted RNA, TruSight RNA Pan-Cancer, TruSight Tumor 15

18 Sequence Hub Labs Apps

Simple **push-button analysis** with public and private analysis tools

- Over 90 published Apps supporting all of Illumina library prep kits
  - TruSeq Amplicon, TruSeq Targeted RNA, TruSight RNA Pan-Cancer, TruSight Tumor 15

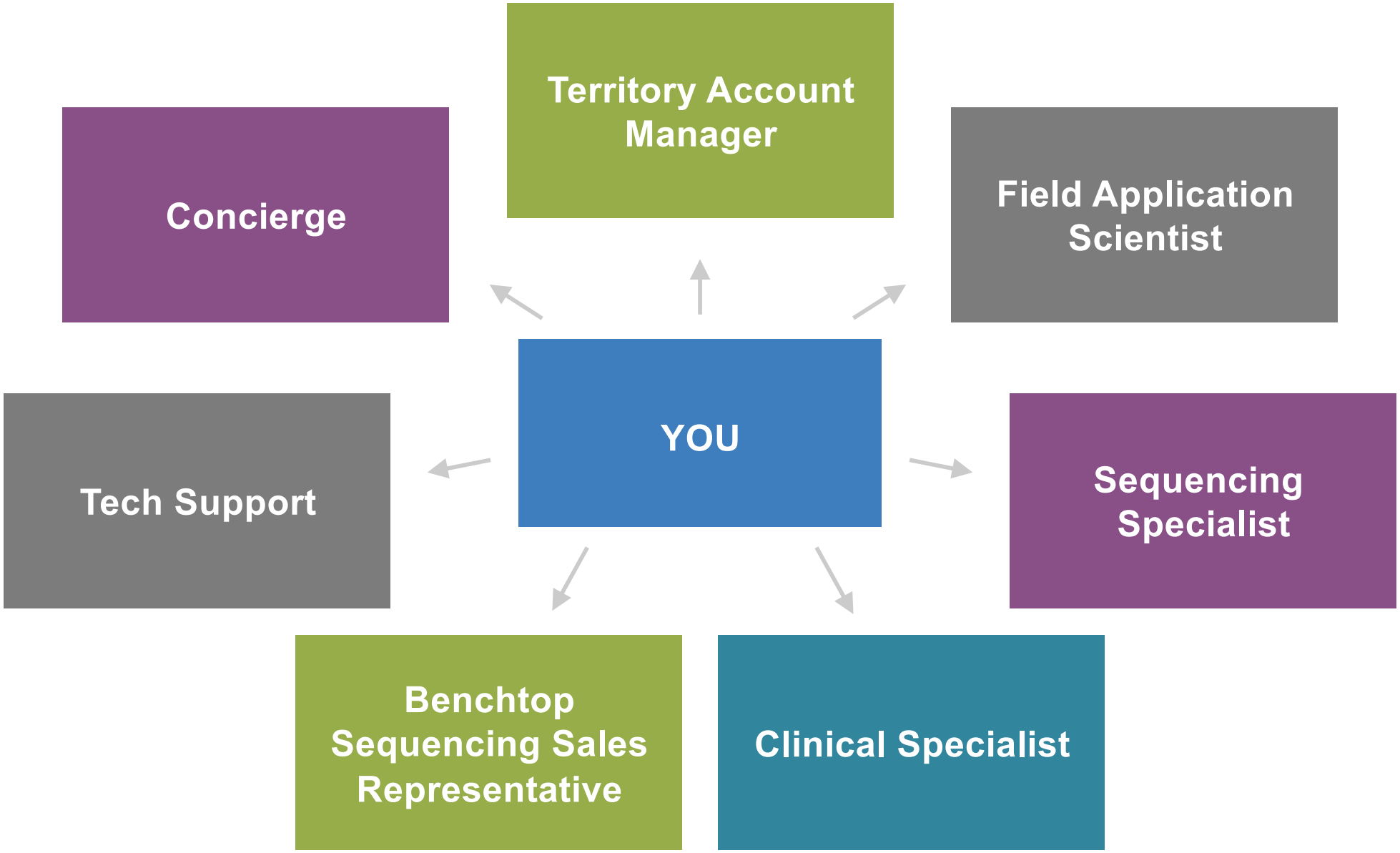


34 Third-Party Apps

Best Practices

- **DNA isolation**
  - Purity and integrity
- **Library prep**
  - Molecular profile (BioAnalyzer/Fragment Analyzer)
  - Molarity
- **Instrumentation**
  - Proper loading of libraries
  - Maintenance
- **Bioinformatics**
- **Reach out to the experts!!!**

We are Here for You!



illumina®