Illumina Sequencing Overview: Library Prep to Data Analysis

Meredith Millis
Field Applications Scientist

QB7845

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

illumina®

Comparing Technologies

q/RT-PCR

q/RT-PCR allows for the analysis of particular variants at specific locations

+ Benefits

- High sensitivity
- Capital equipment already found in most labs

- Limitations

- Low discovery power
- Low variant resolution
- Low scalability

Sanger/CE

Sanger/CE is able to interrogate a gene of interest

+ Benefits

- Cost effective for small stretches of DNA
- Well known technique

- Limitations

- Low sensitivity (down to 20%)
- Low discovery power
- Low scalability

Targeted NGS

Targeted NGS allows for simultaneous screening of several hundreds to thousands of genes

+ Benefits

- Expanded discovery power
- Maintain resolution and high sensitivity
- More data from smaller amounts of DNA and RNA
- Higher throughput with sample multiplexing

- Limitations

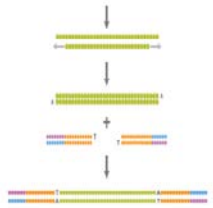
- May be less cost effective when interrogating a low number of samples

illumina

2

Illumina Sequencing Workflow

1. **Library Preparation**



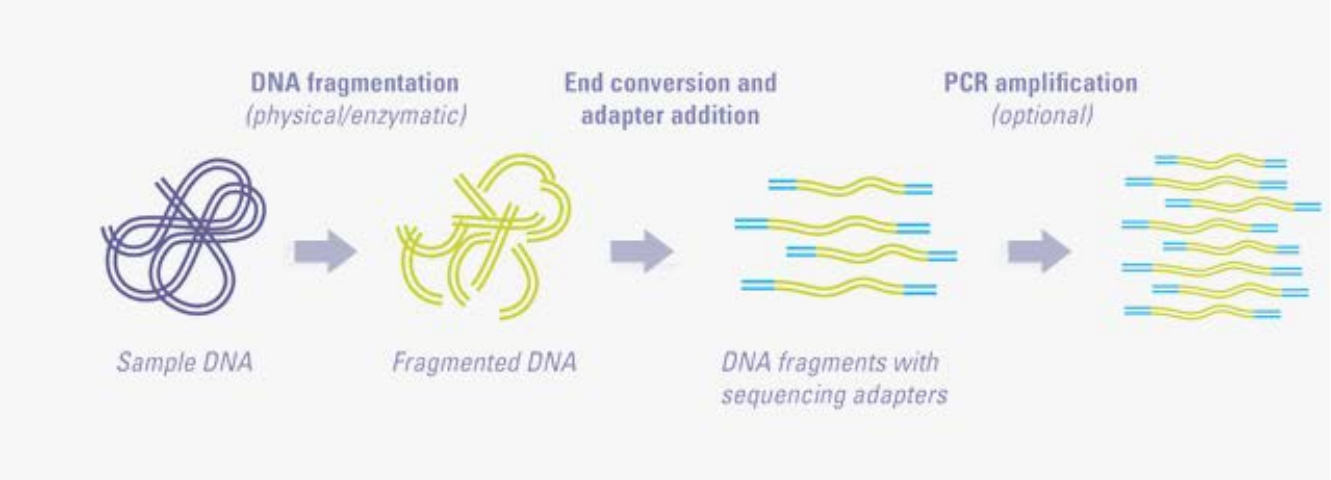
2. **Cluster Generation**

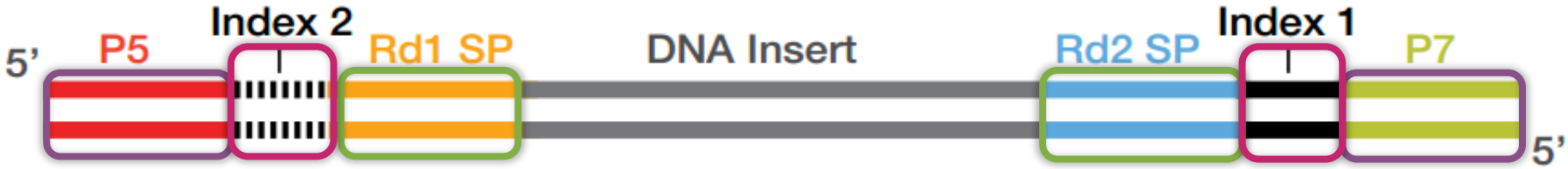
3. **Sequencing**

4. **Data Analysis**

illumina®

Library Prep is Critical for Successful Sequencing





For clustering:
Libraries must have P5 and P7 binding regions on either end of a library

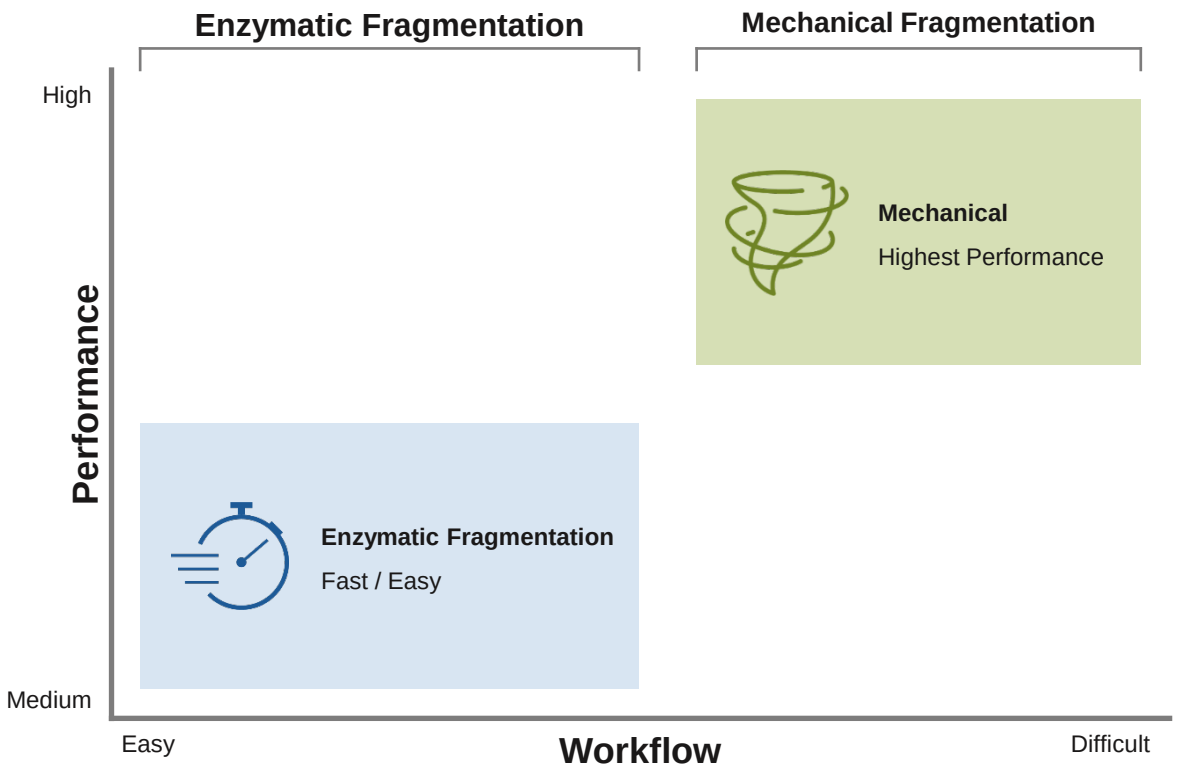
For sequencing:
Libraries must have sequencing primer binding regions

For mixing samples:
Libraries must have a unique index or barcodes sequence

illumina®

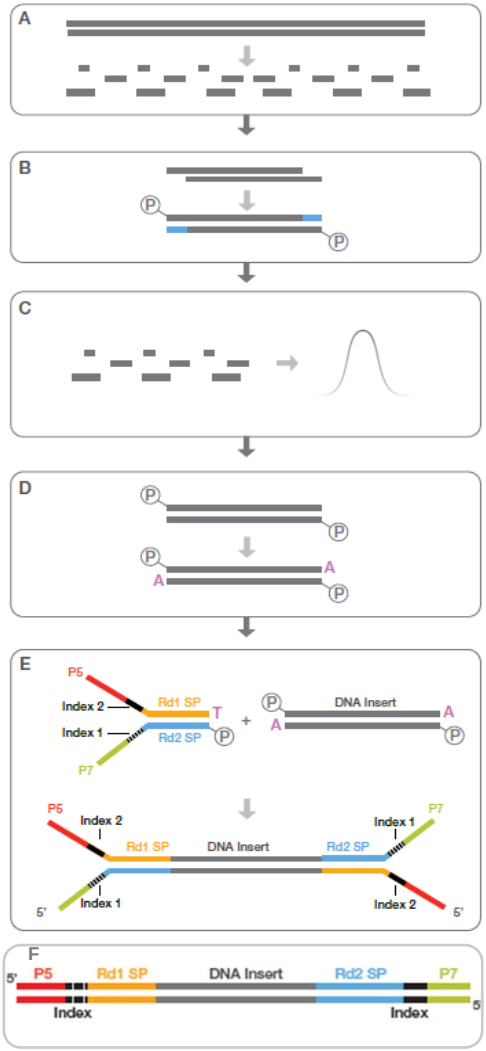
Tools for DNA Library Preparation

Fast or high performance



illumina®

Mechanical fragmentation workflow



A. Genomic DNA is fragmented

B. DNA is end-repaired and phosphorylated

C. Fragments narrowly size selected

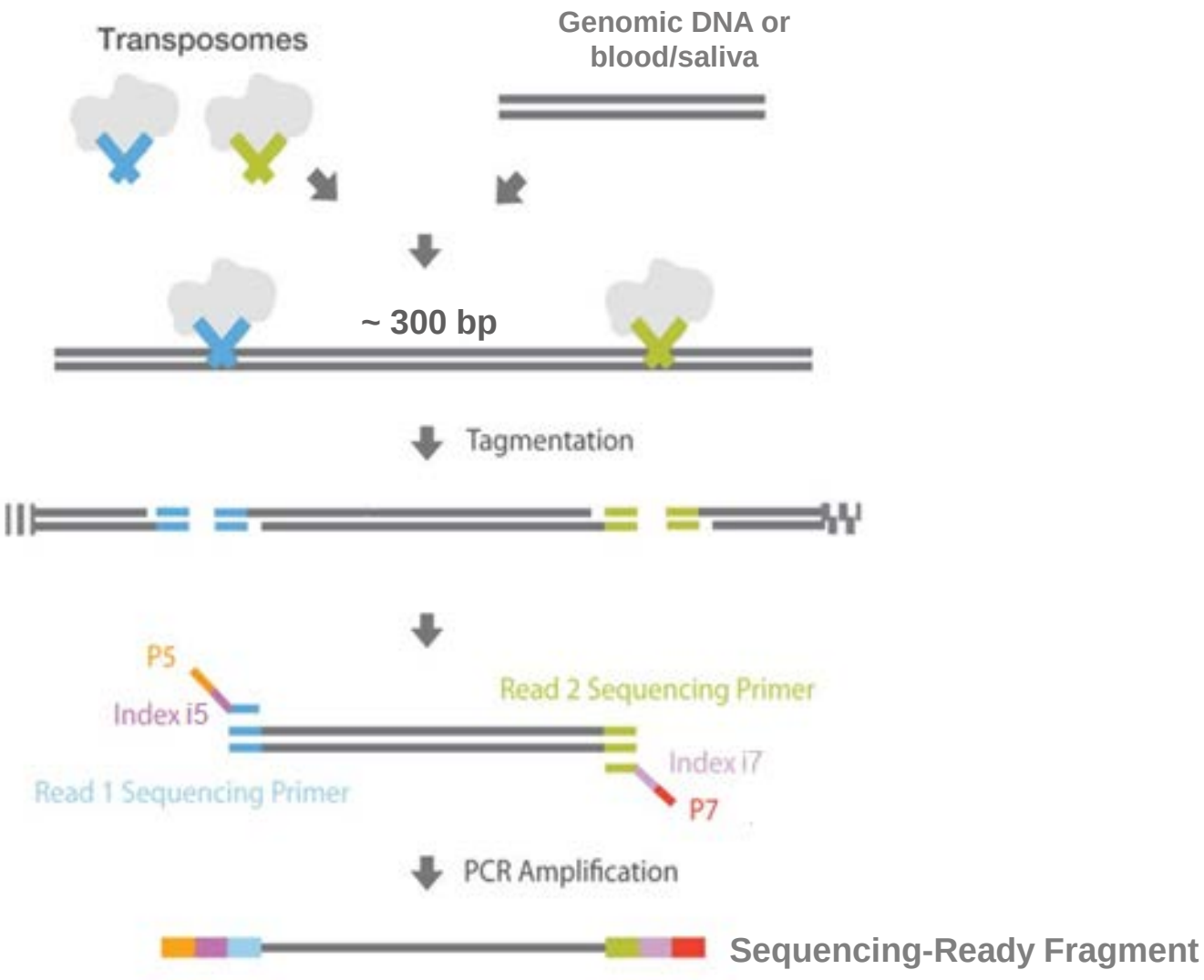
D. A-base added

E. Adapters ligated

F. (only for TruSeq Nano kit) PCR enriches complete libraries

illumina®

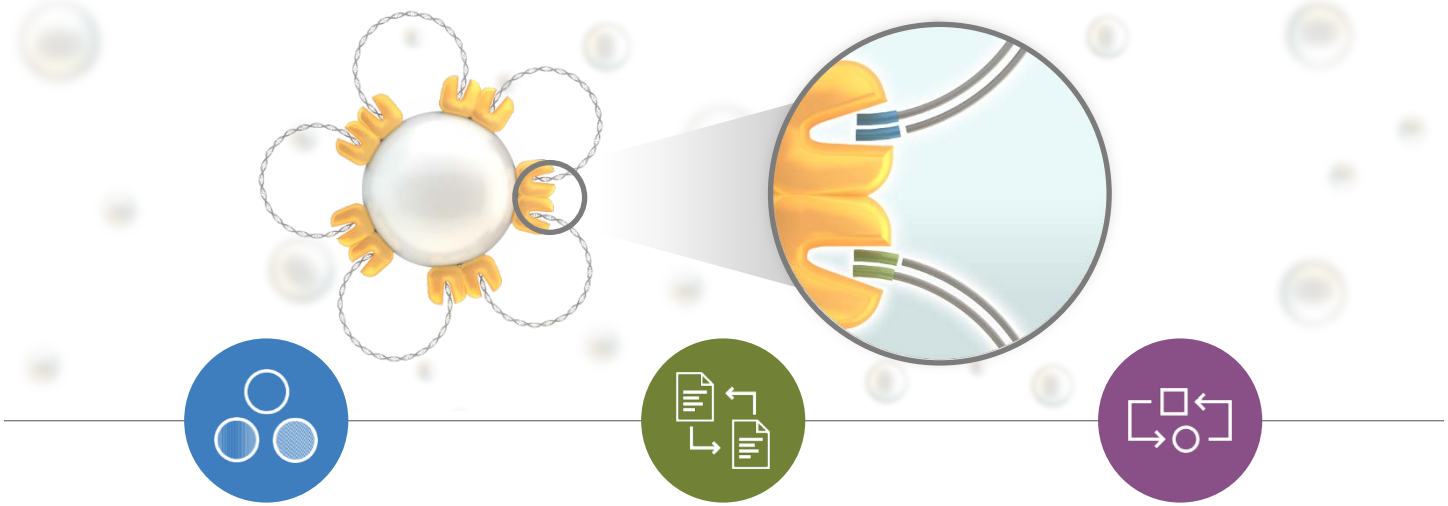
Enzymatic fragmentation workflow



illumina®

Introducing Illumina DNA Prep, (M) Tagmentation

One DNA prep, multiple solutions



Versatile

Broad DNA input range

Integrated protocol for blood, saliva, dried blood spot* & microbe colony* input

Robust and Reproducible

High uniformity of coverage

Consist insert size

Flexible Workflow

Integrated library normalization process

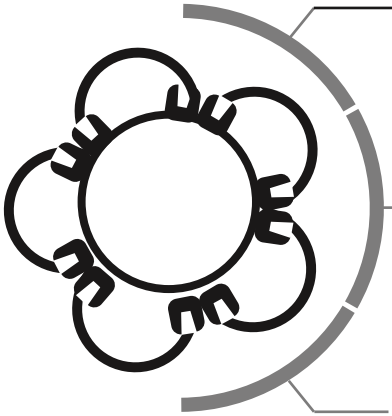
Minimal hands-on time

*Demonstrated protocol

illumina®

New Illumina Tagmentation RNA Enrichment

Bead-linked transposome extended to RNA Enrichment



Illumina DNA Prep For Enrichment DNA
RAPID, MODULAR ENRICHMENT SOLUTION

Illumina DNA Prep
ROBUST, SIMPLE DNA LIBRARY PREP

Illumina RNA Library Prep
WITH ENRICHMENT (L) TAGMENTATION

- **Significant reduction** in library prep steps allows users to focus their efforts elsewhere
- **Single hybridization** workflow allows for faster sample processing
- **Unique Dual Indices** eliminates sample “cross-talk” and supports higher output Illumina sequencing systems with **384 indices**
- **Panel modularity** allows users to build upon their standardized workflows with additional panels

illumina®

Introduction to the new Illumina RNA library prep suite

	1 | **Illumina Stranded Total RNA Prep** Ligation with Ribo-Zero Plus	2 | **Illumina Stranded mRNA Prep** Ligation	3 | **Illumina RNA Prep** Enrichment (L) Tagmentation
Detection	Whole transcriptome (coding & non-coding)	Coding transcriptome with Poly A tail	Targeted coding region
FFPE compatible	✓		✓
Turn-around time*	<8 hours	<8 hours	~9.5 hours
Minimum Input*	⇠ 10ng ⇢		
Multiplexing*	⇠ Up to 384 UDIs ⇢		
Automation		⇠ Automation Ready ⇢	
Additional Features	Includes Ribo-Zero Plus for rRNA depletion of Human, Mouse, Rat, Bacteria and Globin		Compatible with Illumina Exome

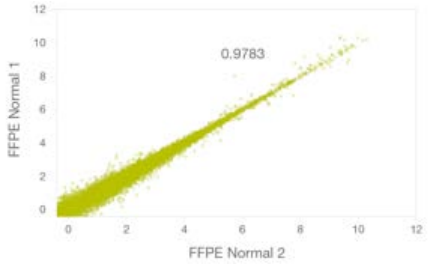
*Preliminary Specs

illumina®

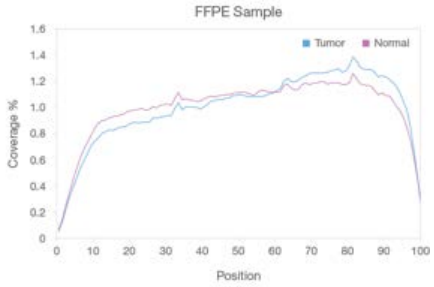
The new Illumina stranded mRNA and Total RNA ligation prep

Expanding on TruSeq™'s expectational data quality

TruSeq stranded Total RNA data quality



High data quality across replicates and runs on even the most challenging samples (FFPE)



Excellent coverage across the top 1000 expressed transcripts in both tumor and normal tissue

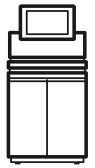
Illumina stranded mRNA and Total RNA advancements

<8-hour workflow

With less than four hours of hands-on time, users can achieve greater lab efficiency

Multi-species depletion

Single-tube rRNA depletion provides an additional level of flexibility in going from human, mouse, rat, bacteria, and epidemiology samples



Higher throughput

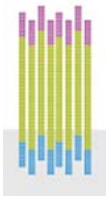
Decrease sequencing costs by loading up to 384 samples on a single NovaSeq™ 6000 S4 Flow Cell using the newly available 384 unique dual indexes

illumina®

Illumina Sequencing Workflow

1. Library Preparation

2. Cluster Generation

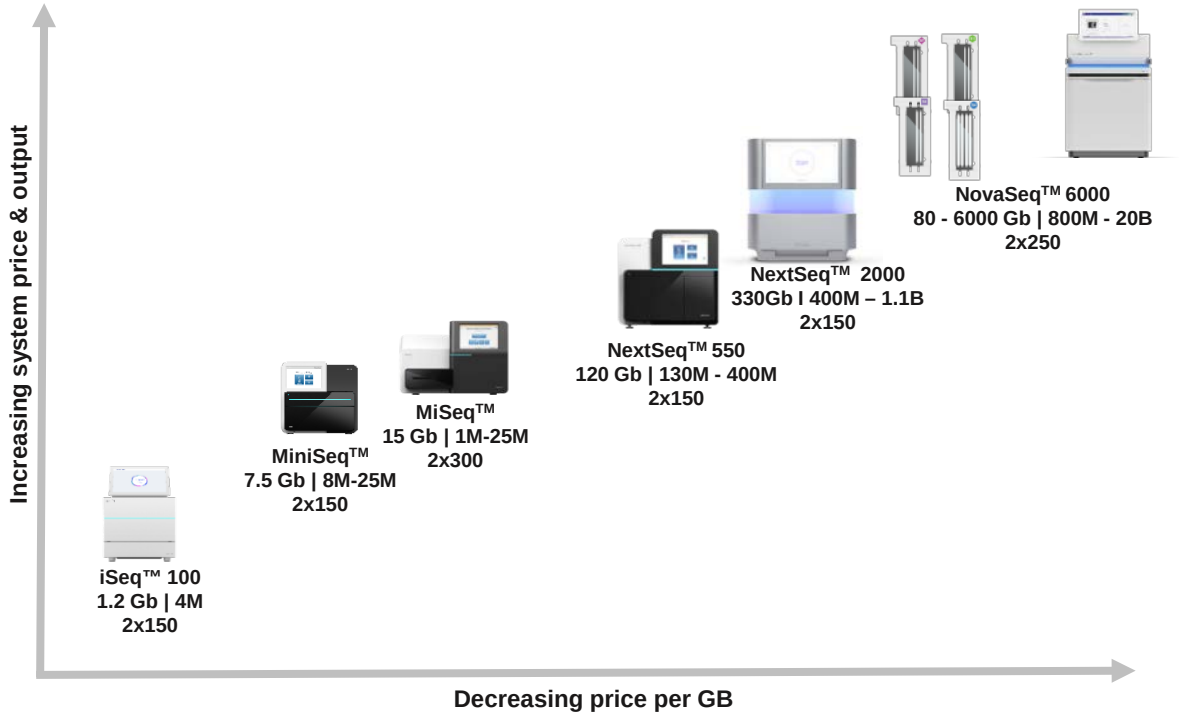


3. Sequencing

4. Data Analysis

illumina®

Illumina Sequencing Portfolio



illumina®

MiSeq System

PE300 | **75% > Q30**

0.3–15 | gigabases

1–25 million | clusters

illumina®

MiSeq Offers Scalable Sequencing

1 Million Reads → 4 Million Reads → 15 Million Reads → 25 Million Reads

MiSeq

Core Consumables Version 2 Nano	Core Consumables Version 2 Micro	Core Consumables Version 2	Core Consumables Version 3
• 500 cycles (Nano) • 300 cycles (Nano)	• 300 cycles (Micro)	• 500 cycles • 300 cycles • 150 cycles	• 600 cycles • 150 cycles

illumina®

NextSeq 550 has Tunable Output, High Data Quality, and Array Capabilities

High-Output

Up to 120 Gb

400M clusters PF

1x75 bp to 2x150 bp

Mid-Output

Up to 40 Gb

130M clusters PF

2x75 bp to 2x150 bp

MethylationEPIC and CytoSNP 850K

Clinically relevant content on arrays

SBS chemistry

90% of all NGS data are generated on an Illumina platform

High accuracy

High quality

DRAGEN™ pipelines on BaseSpace™

Industry-leading accuracy and speed

Variants include indels, small variants, and CNVs

In the cloud with BaseSpace and on premises with a server

illumina®

NextSeq 1000/2000 Configurations

NextSeq 1000

$210K I Now Available

NextSeq 1000

120GB

Max output
(100/200/300 cycles)

Field

Upgradeable

To NextSeq 2000
$150K

NextSeq 2000

$335K I Now Available

P2 Flow Cell

120GB

Max output
(100/200/300 cycles)

P3 Flow Cell

330GB

Max output
(50/100/200/300 cycles)

Further workflow improvements

Custom primers and custom recipe	DRAGEN Single Cell RNA, DRAGEN Enrichment Somatic	FastQ compression and FastQC metrics	Additional workflow features / capabilities and BSSH / control software improvements

illumina®

NovaSeq 6000

Proven Architecture

Widely published, industry leading platform built for scalability with better data economics, removing barriers to answering your biggest biological questions

Flexible Performance

Highly configurable to support the broadest range of methods at any scale – push your research further through cutting edge applications

Immense Discovery Power

Sequence deeper into the genome, expand into new applications, and run more samples to empower your studies

Streamlined Operation

Designed to increase lab efficiency with a simplified workflow and seamless user experience

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

illumina

NovaSeq System Configurations

	S Prime Flow cell	S1 Flow cell	S2 Flow cell	S4 Flow cell
Lanes	2	2	2	4
Output (based on read length)	80–400 Gb	167–500 Gb	417–1250 Gb	2000–3000 Gb
Single Reads (clusters passing filter)	0.8 B	1.6 B	4.1 B	10 B
Run Time	13–38 hours	13–25 hours	16–36 hours	36–44 hours
Max Read Length	2x250	2x150	2x150	2x150
Value†	Smaller batch size Cost effective sequencing for small projects Pilot new projects QC and optimize library concentrations Rapid turnaround Faster turnaround time for high-throughput projects‡		Run experiments to scale Enable larger batch sizes and larger cohorts Attractive per sample economics Competitive pricing per Gb and per M reads Accessibility to data-rich applications and methods Multi-modal studies, deeper sequencing	

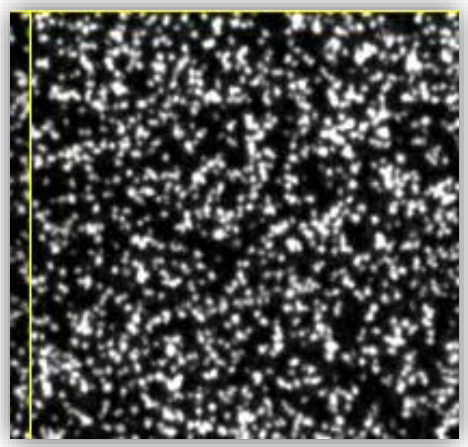
‡ Turnaround time includes cluster generation and sequencing run time
† Compared to Illumina high-throughput portfolio.

illumina®

What is a Cluster?

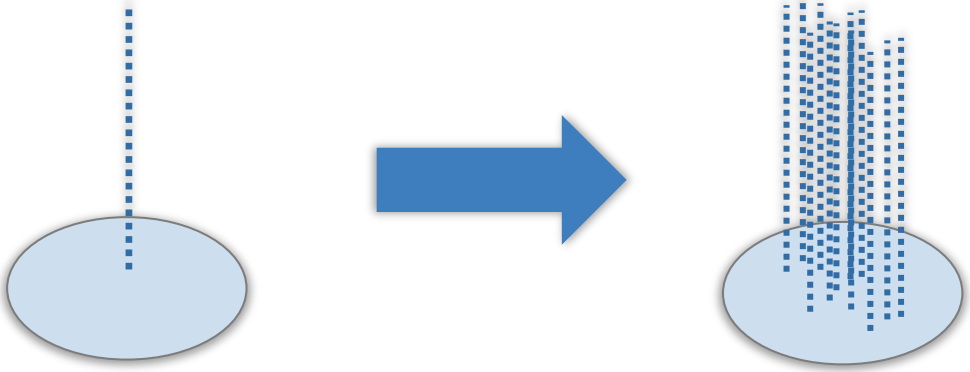
Clusters are a group of DNA strands positioned closely together

Each cluster represents thousands of copies of the same DNA strand in a 1–2 micron spot



An image of fluorescently labelled clusters on a flow cell

Single DNA Library



Amplified Clonal Cluster

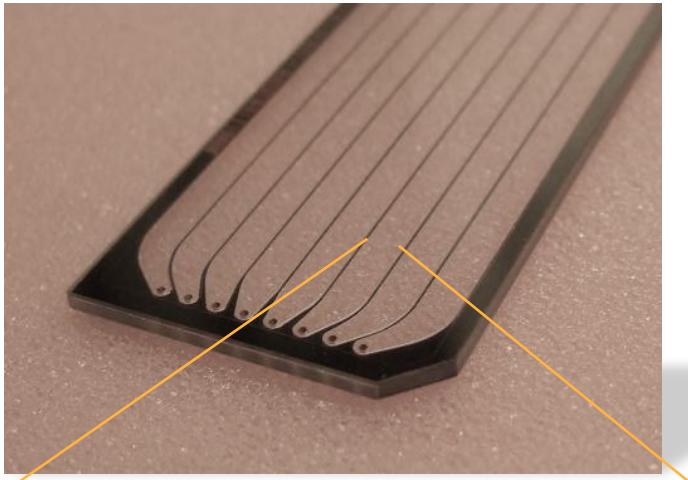
illumina®

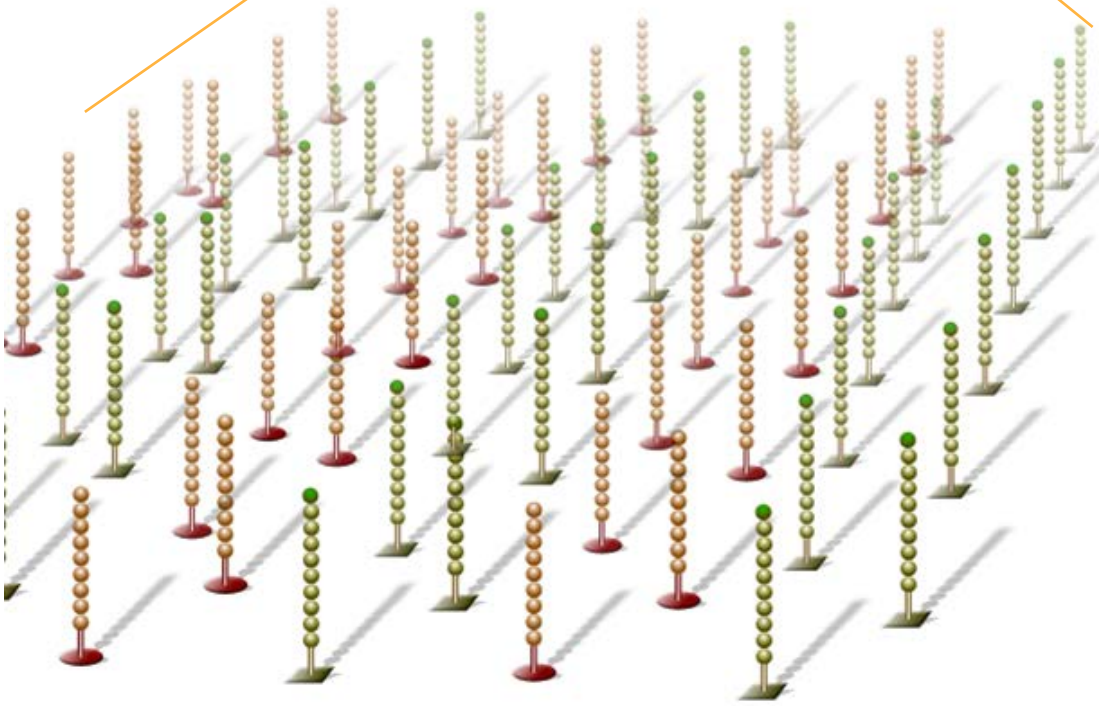
What is a flow cell?

Cluster generation occurs on a flow cell

A flow cell is a thick glass slide with channels or lanes

Each lane is coated with a lawn of oligos complementary to library adapters



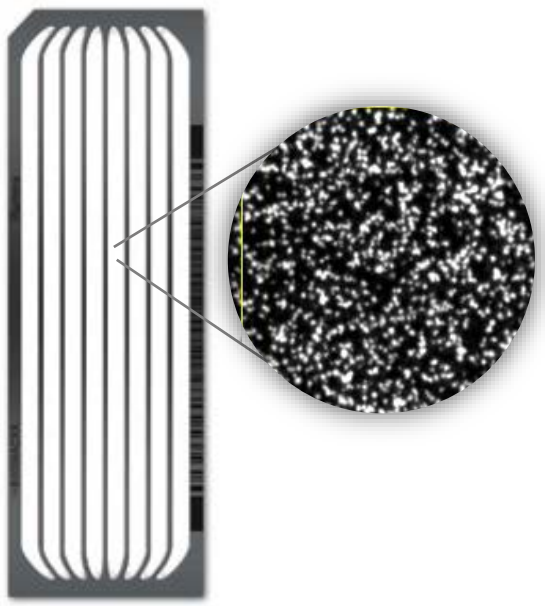


Flow cell Architecture

Random vs Patterned

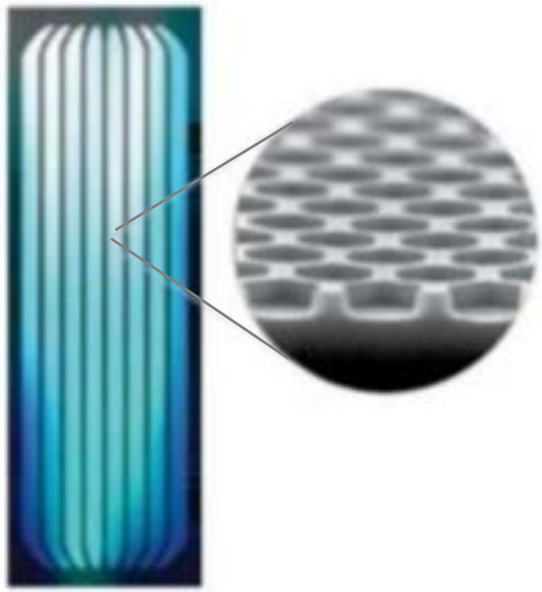
Random Flow Cell

- HiSeq™ 2500, MiSeq™, NextSeq™, MiniSeq™
- Randomly spaced clusters
- Variable Insert Sizes
- Lower Duplication Rates



Patterned Flow Cell

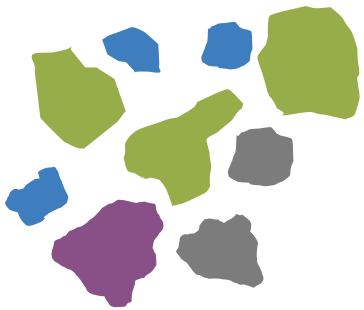
- HiSeq 3K/4K/X, NovaSeq™ 6000, iSeq™ 100
- Defined size and spacing
- Increased Cluster density
- Simplified imaging

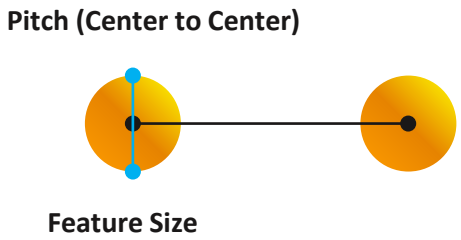


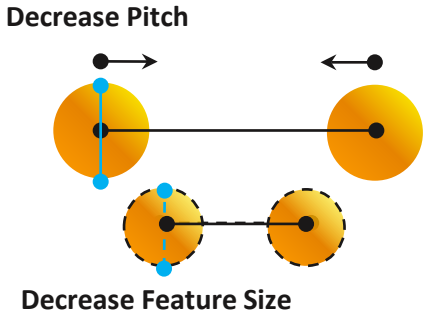
illumina®

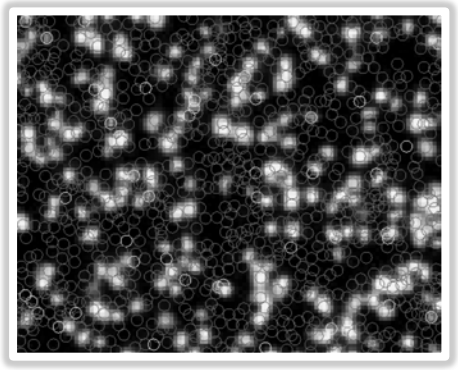
Patterned flow cells

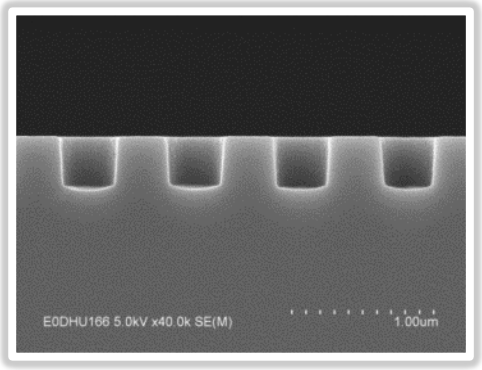
Complete control of pitch & feature size

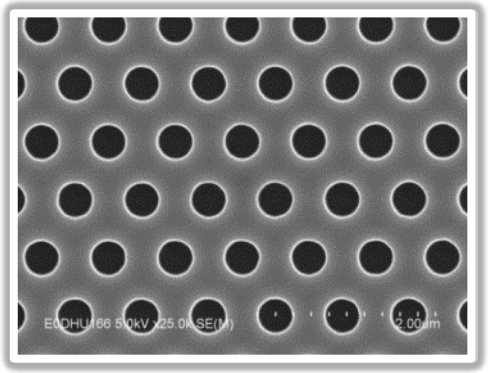








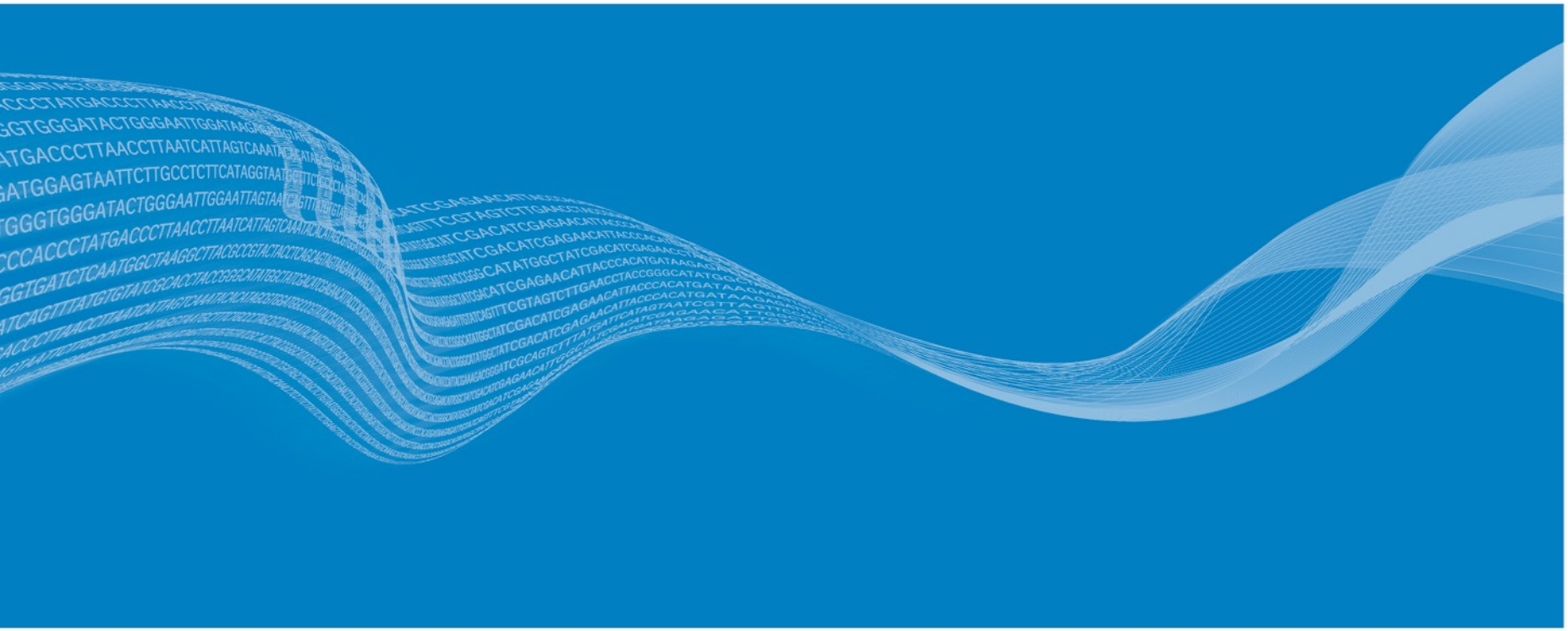




Random	Patterned
	Rigid registration reduces time by skipping template generation

illumina®

Traditional Cluster Generation

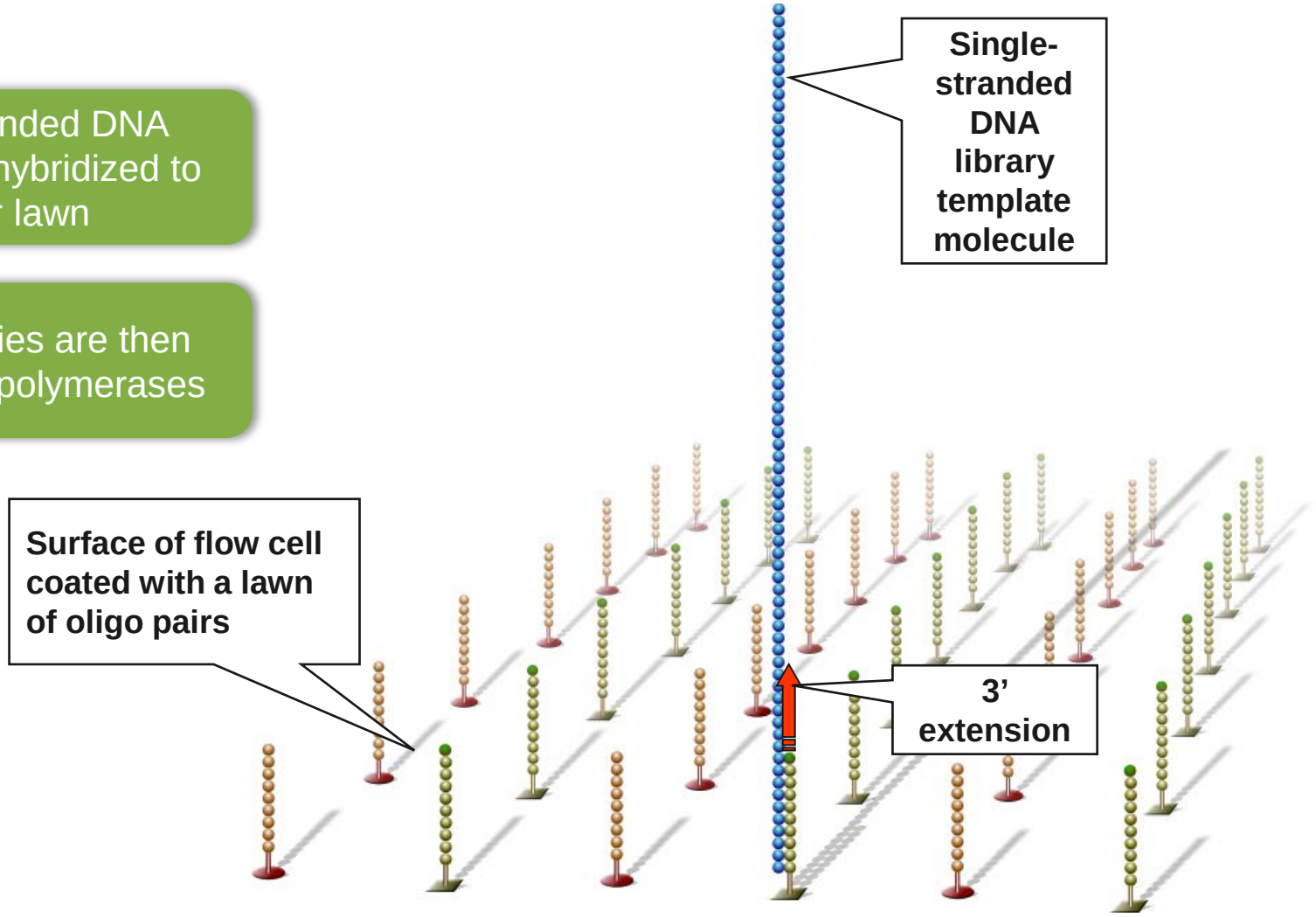


illumina®

Hybridize Fragment & Extend

Single-stranded DNA libraries are hybridized to primer lawn

Bound libraries are then extended by polymerases



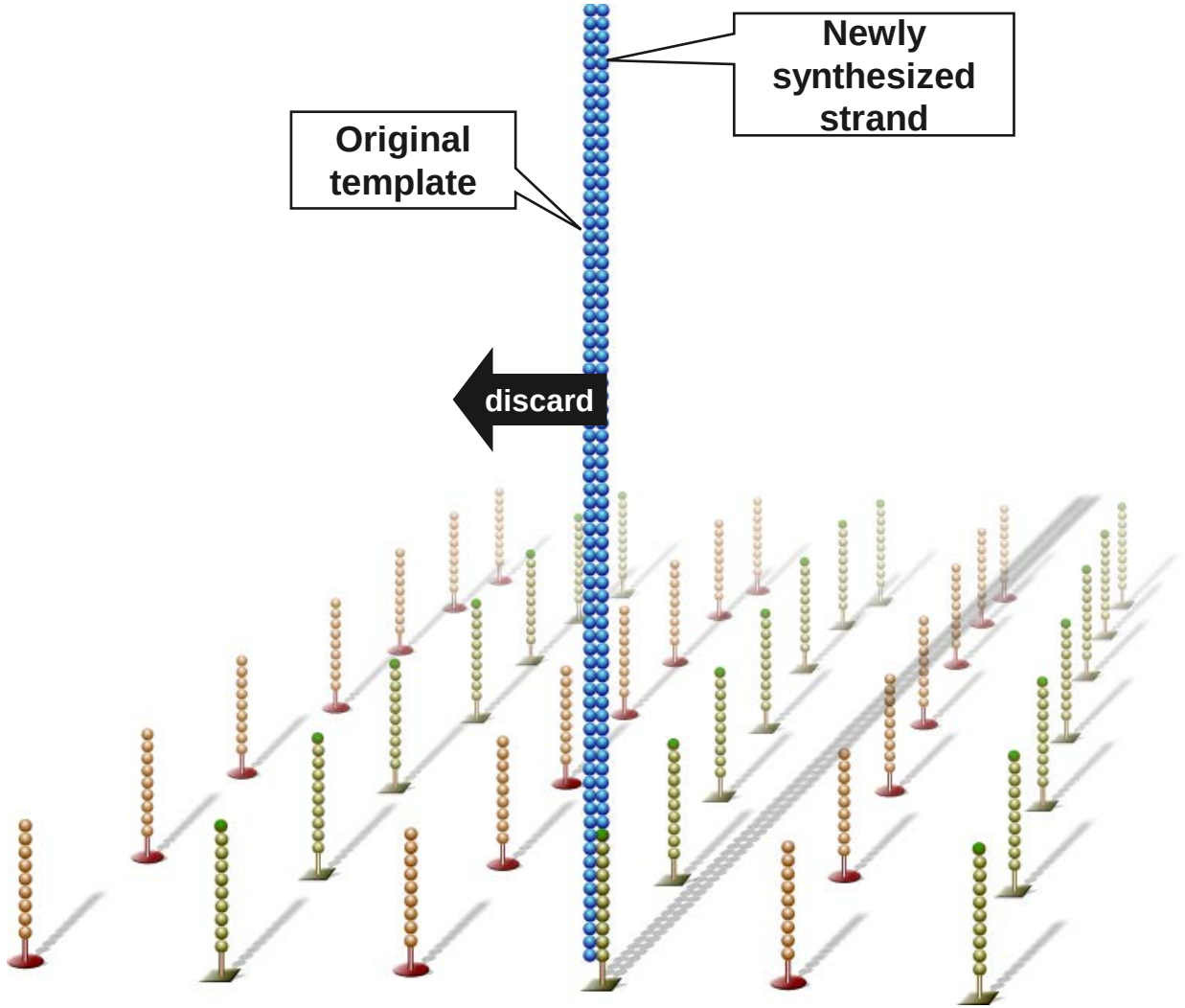
illumina®

Denature Double-Stranded DNA

Double-stranded molecule is denatured

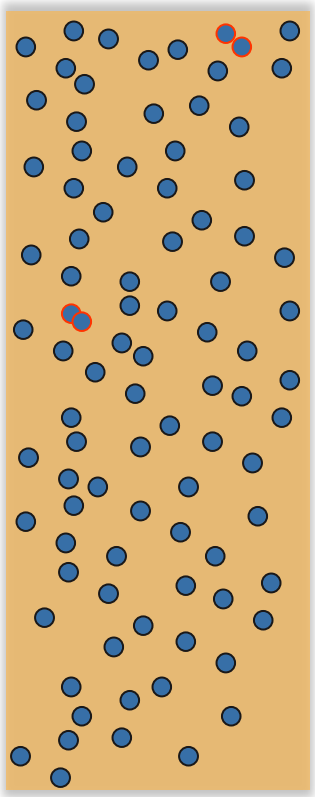
Original template washed away

Newly synthesized strand is covalently attached to flow cell surface

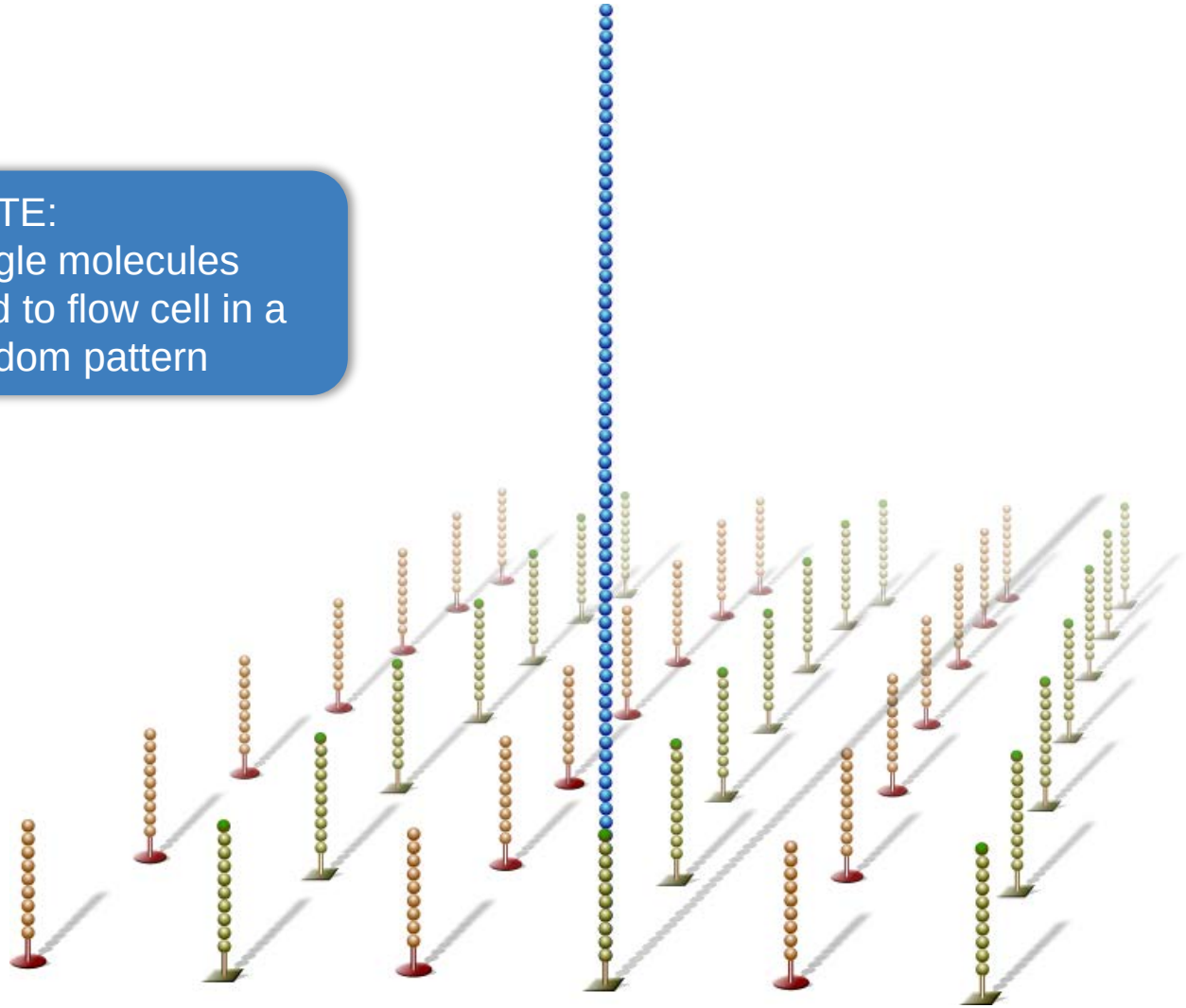


illumina®

Single-Stranded DNA



NOTE:
Single molecules bind to flow cell in a random pattern

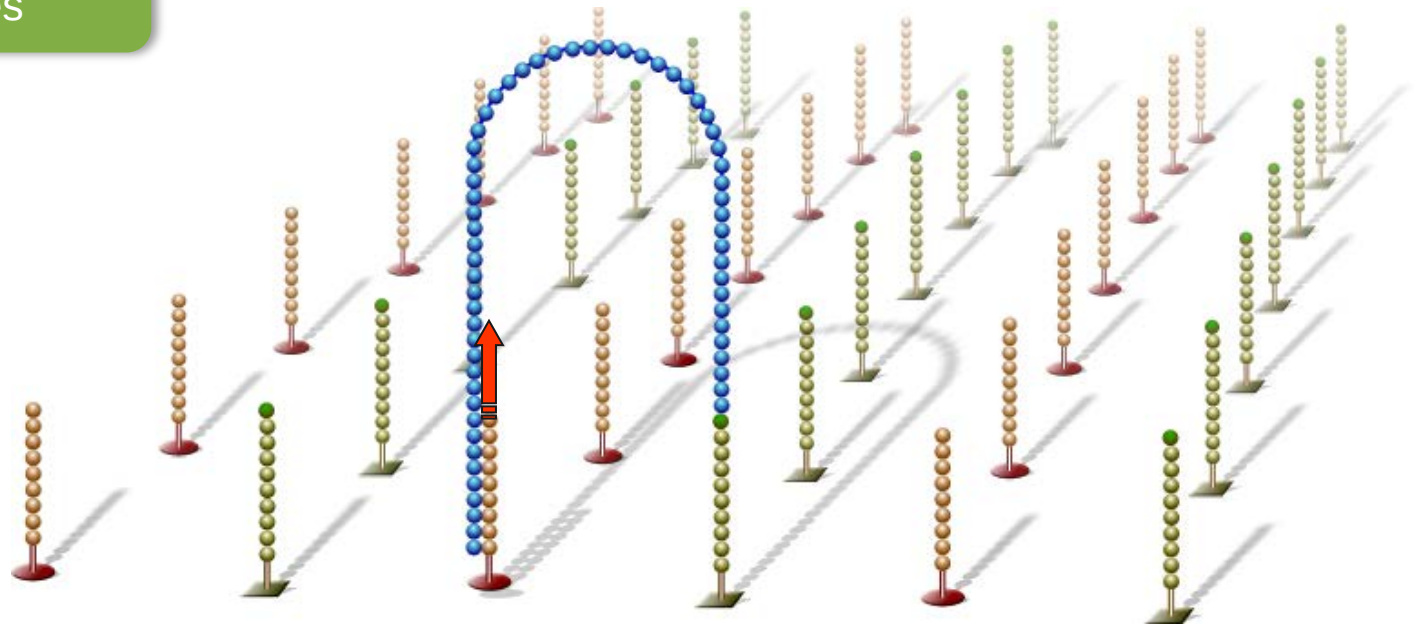


illumina®

Bridge Amplification

Single-stranded molecule flips over and forms a bridge by hybridizing to adjacent, complementary primer

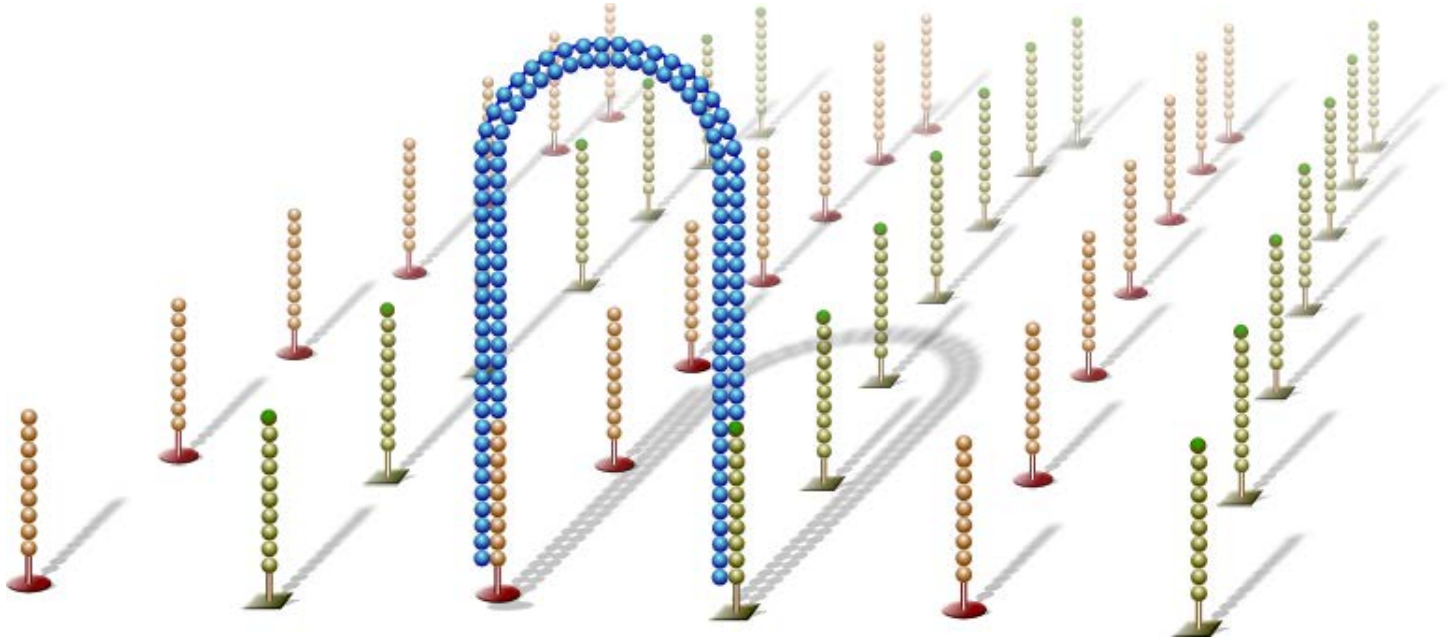
Hybridized primer extends by polymerases



illumina®

Bridge Amplification

Double-stranded bridge is formed

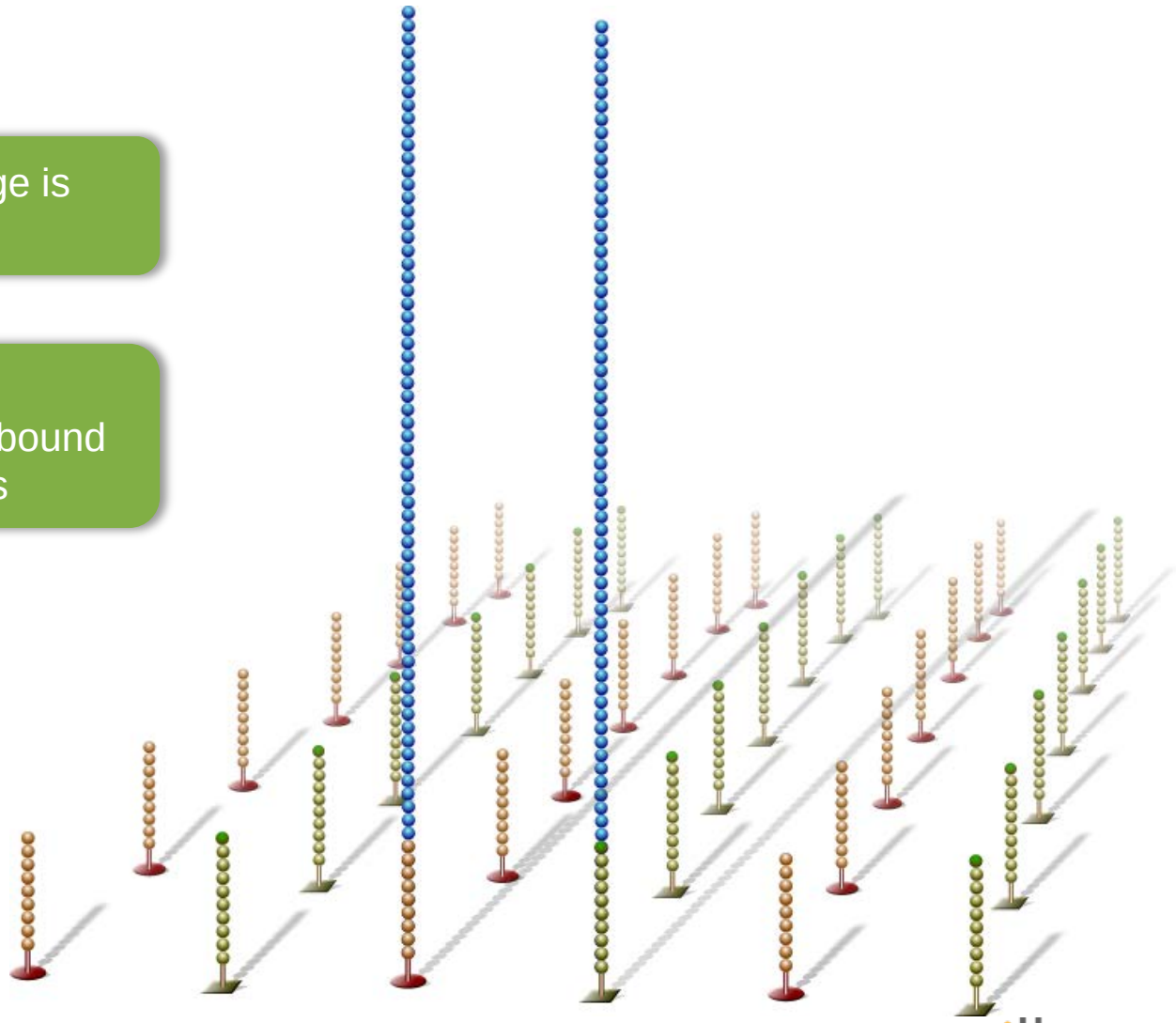


illumina®

Denature Double-Stranded Bridge

Double-stranded bridge is denatured

Result:
Two copies of covalently bound single-stranded templates

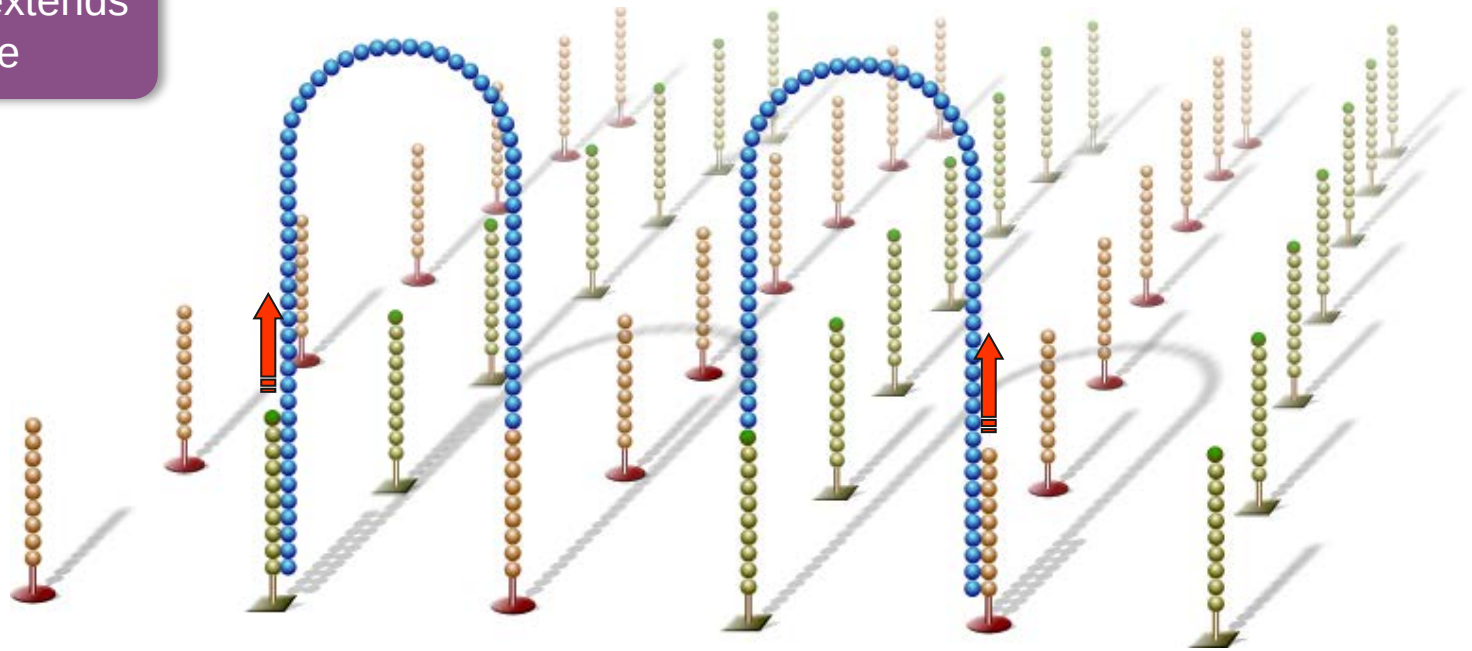


illumina®

Bridge Amplification

Single-stranded molecules flip over to hybridize to adjacent primers

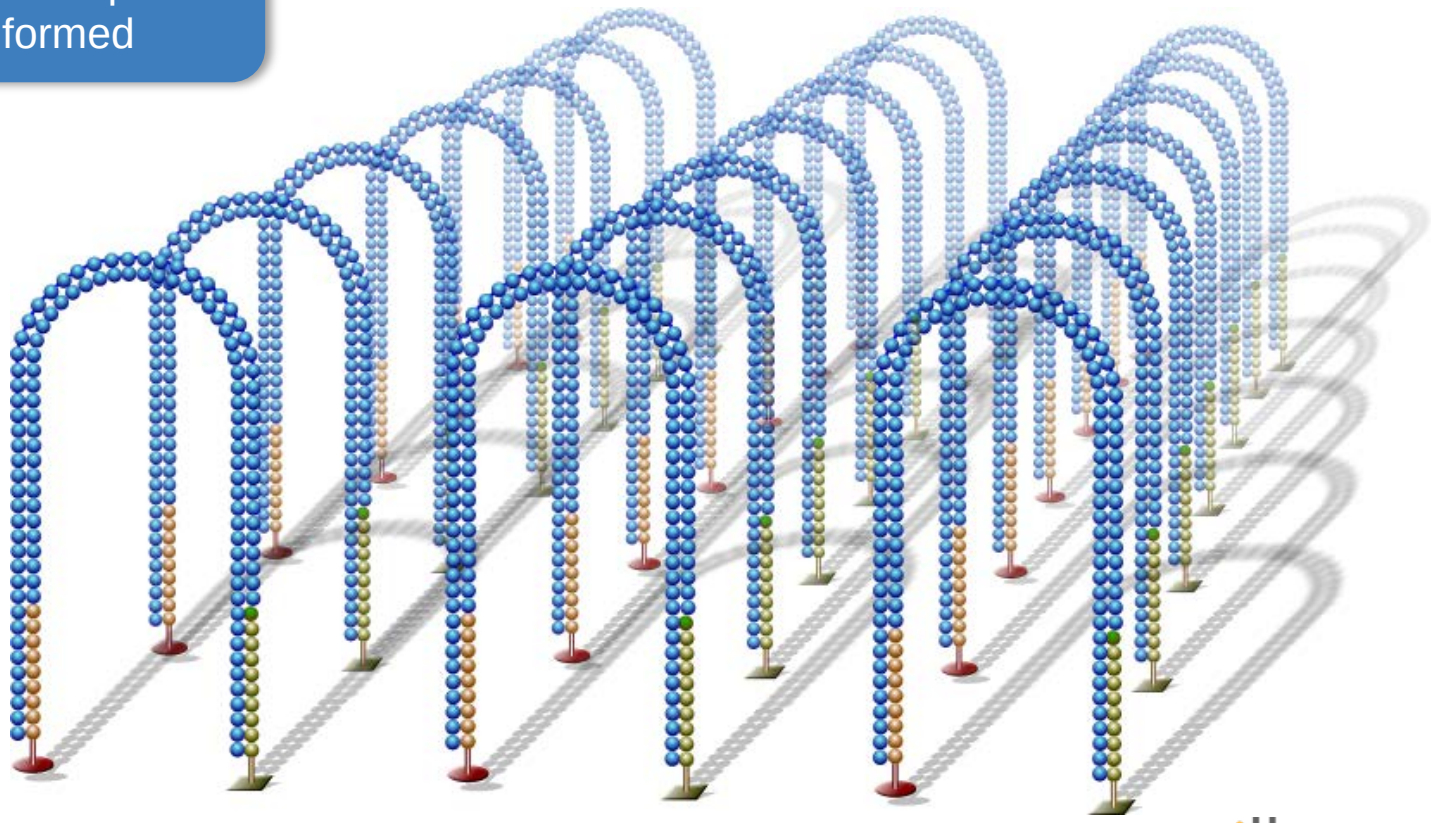
Hybridized primer extends by polymerase



illumina®

Bridge Amplification

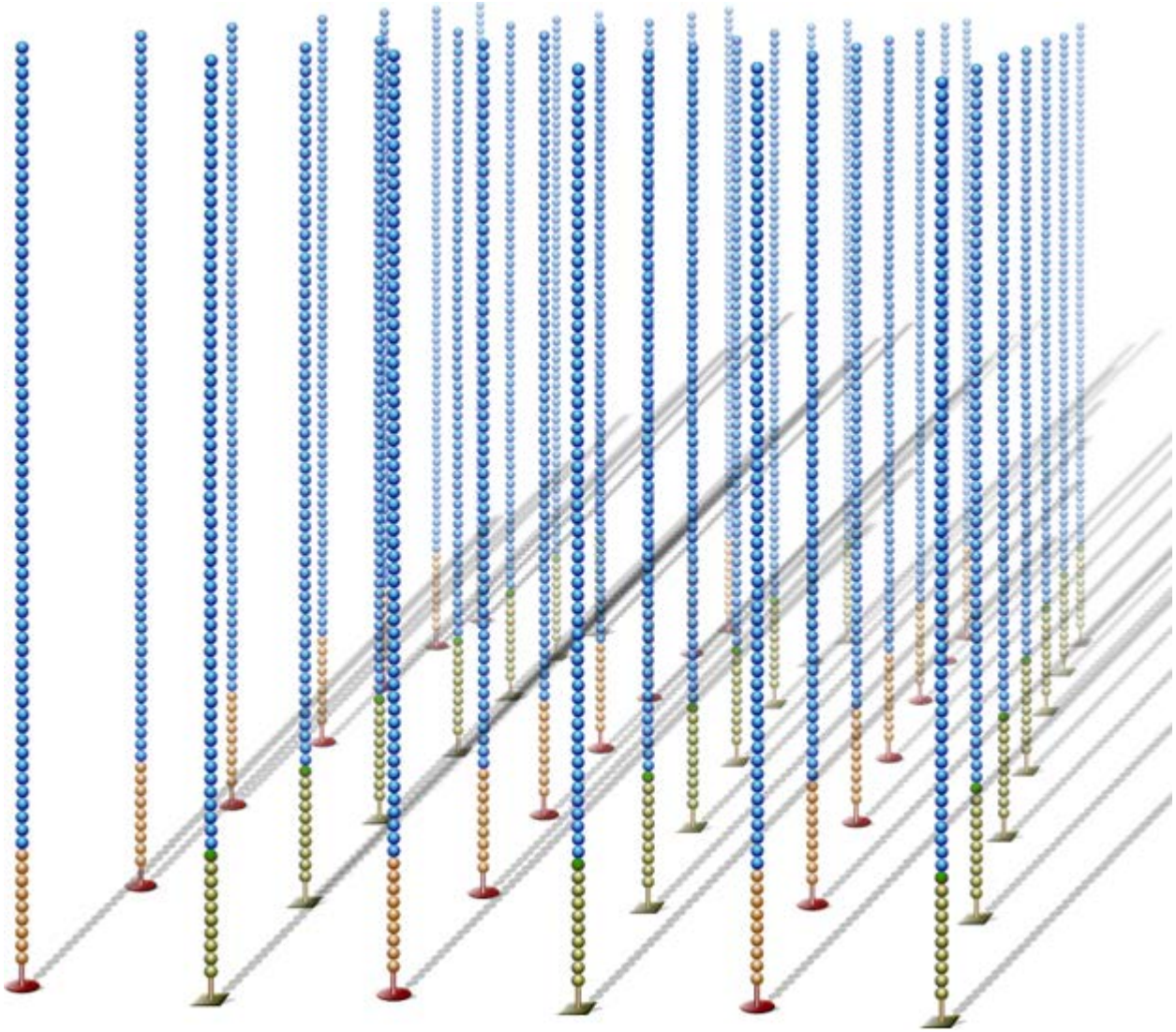
Bridge amplification cycle is repeated until multiple bridges are formed



illumina®

Linearization

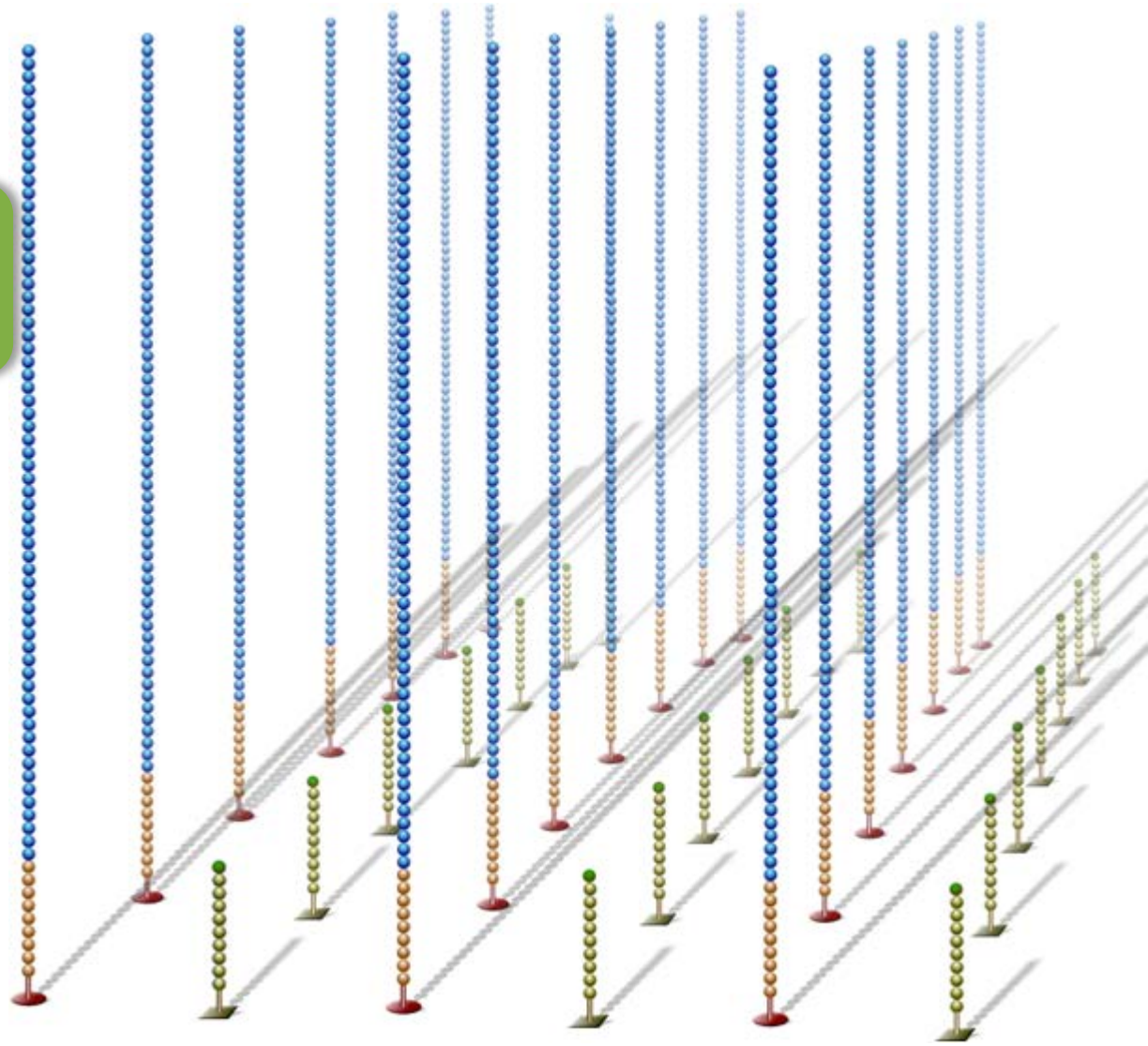
dsDNA bridges are denatured



illumina®

Reverse Strand Cleavage

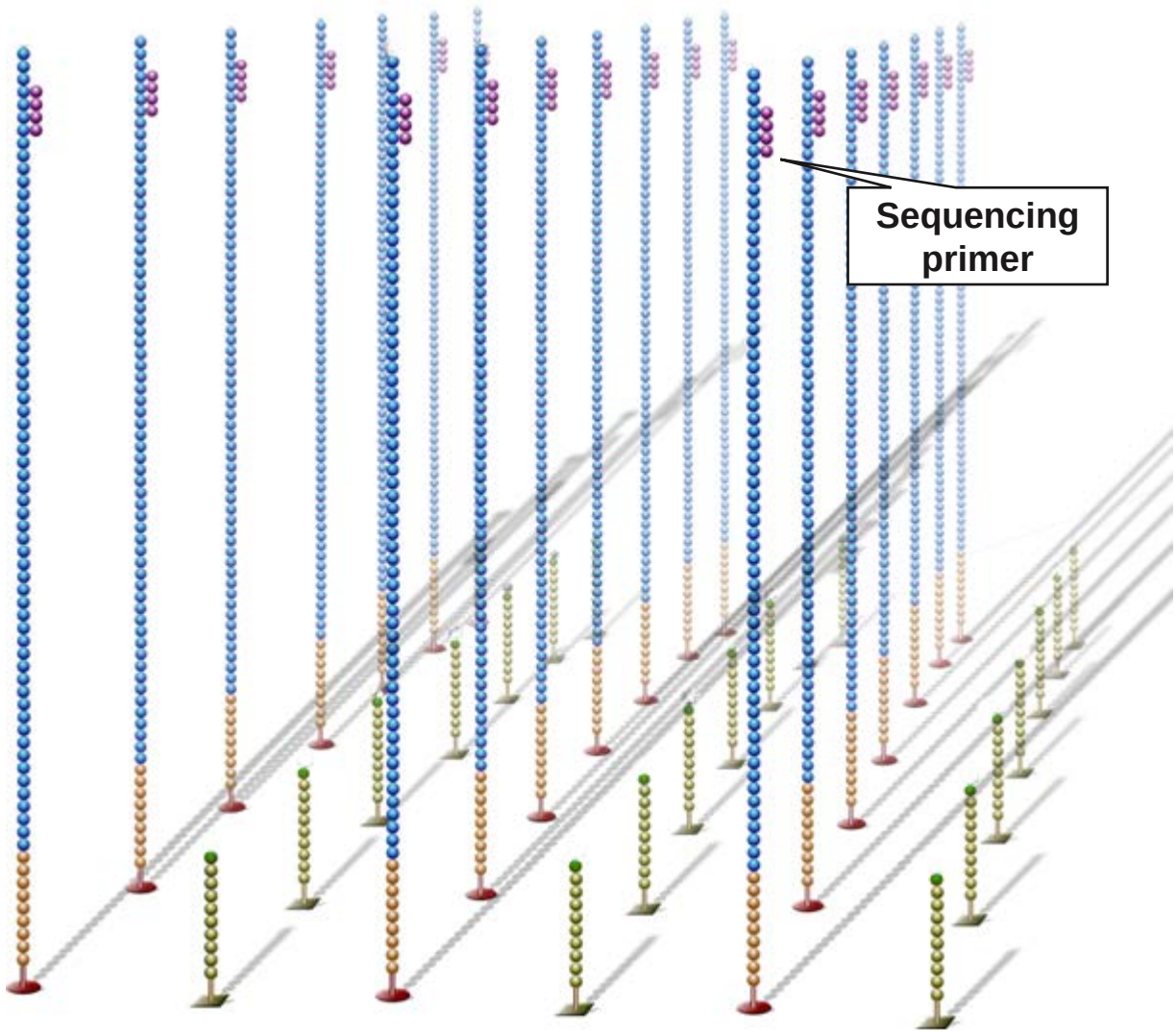
Reverse strands are cleaved and washed away, leaving a cluster with forward strands only



illumina®

Read 1 Primer Hybridization

Sequencing primer is hybridized to Read 1 sequencing primer binding site



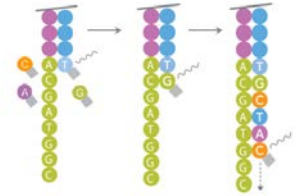
illumina®

Illumina Sequencing Workflow

1. Library Preparation

2. Cluster Generation

3. Sequencing

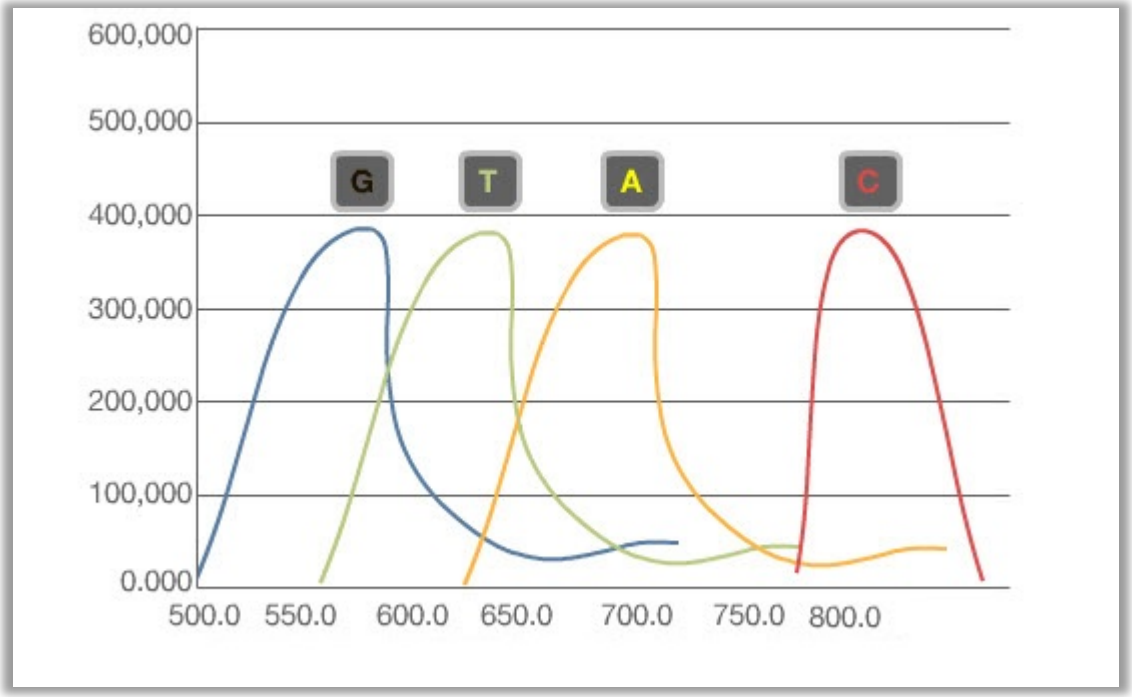


4. Data Analysis

illumina®

4-Channel SBS Chemistry:

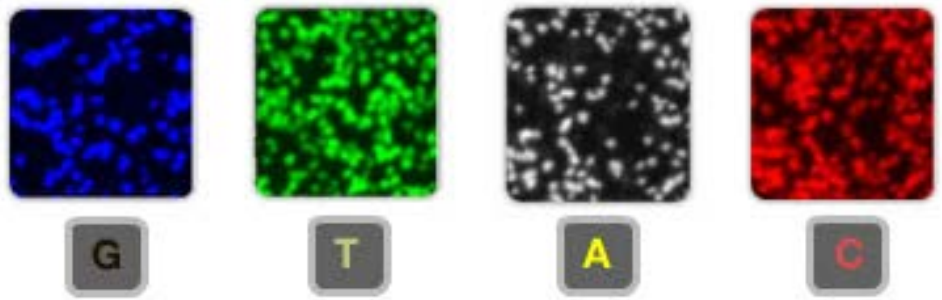
HiSeq, MiSeq



Each of the four DNA bases emits an intensity of a unique wavelength

Collects four images:

- During each cycle, each cluster appears in only one of four images



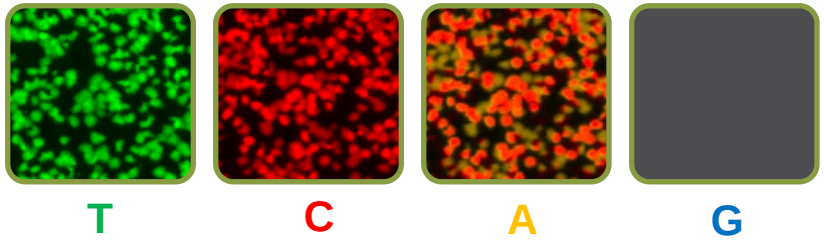
illumina®

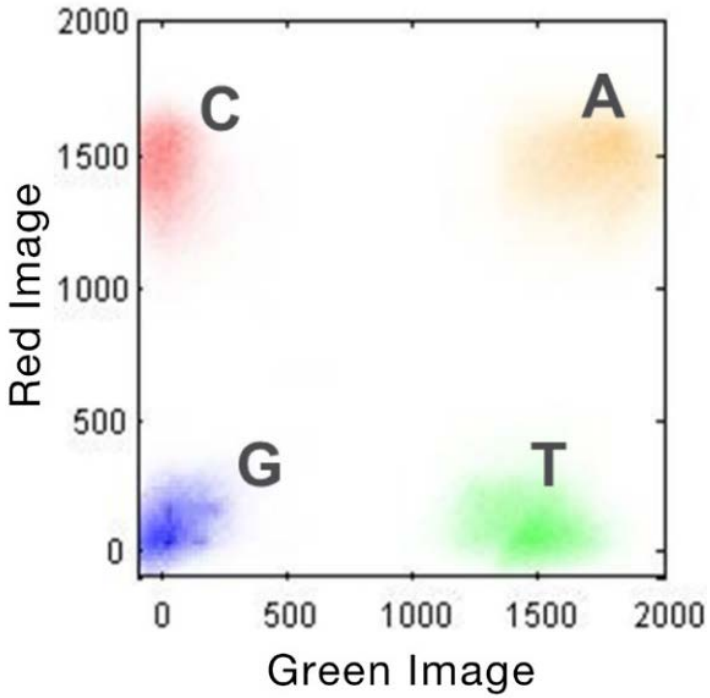
2-Channel SBS Chemistry:

NextSeq 550, MiniSeq, NovaSeq 6000

- **2-channel SBS uses two images:**
- **Clusters appearing in green only are T**
- **Clusters appearing in red only are C**
- **Clusters appearing in both images are A**
- **Clusters not present/dark are G**

- **After imaging, cluster intensities are plotted and bases called accordingly**

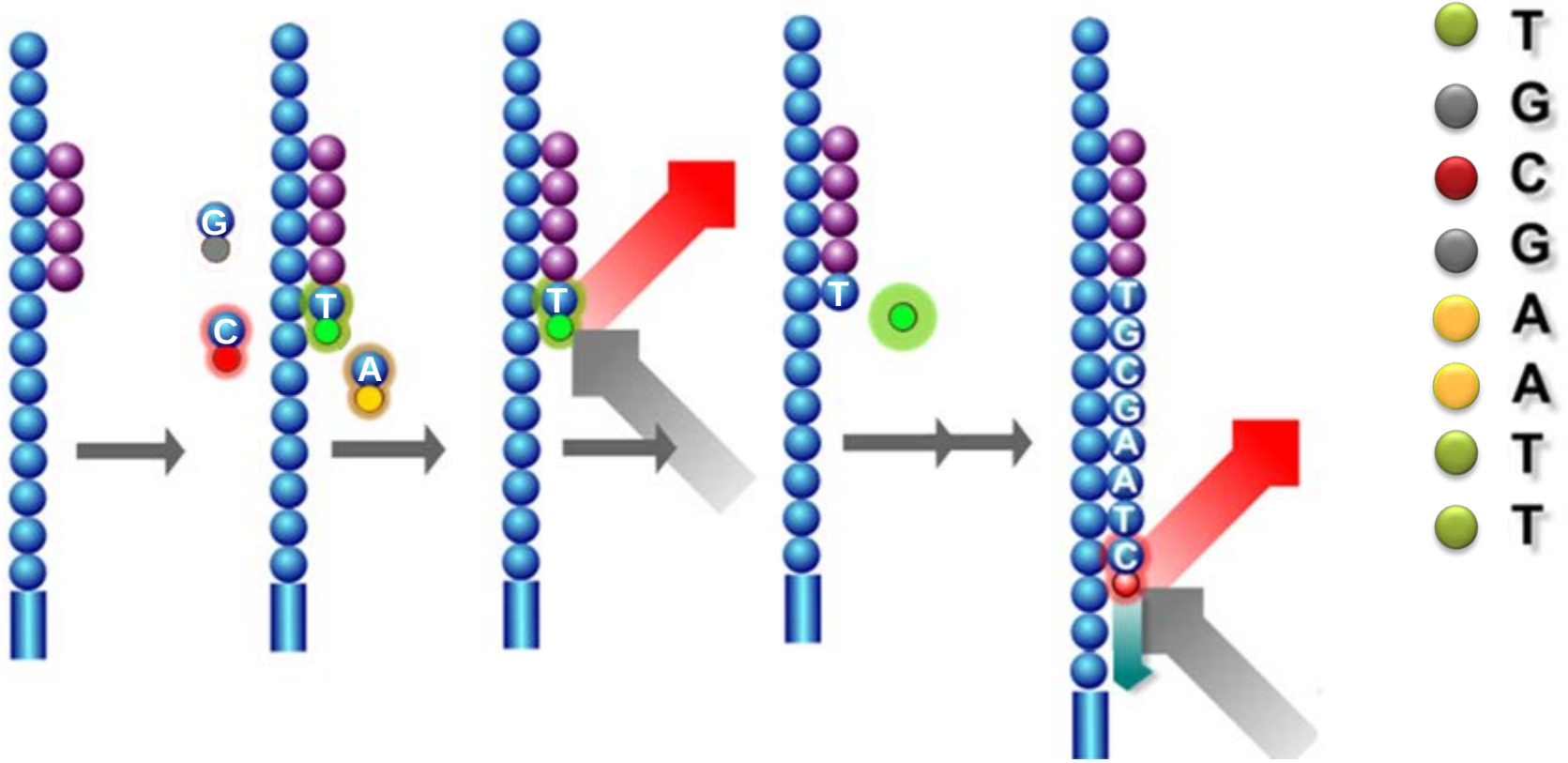




Illumina Two-Channel SBS Sequencing Technology Technote

A Closer Look At 2-Dye Chemistry

2-channel chemistry



Add 3 Fl-NTP's & nonFl-NTP + Polymerase

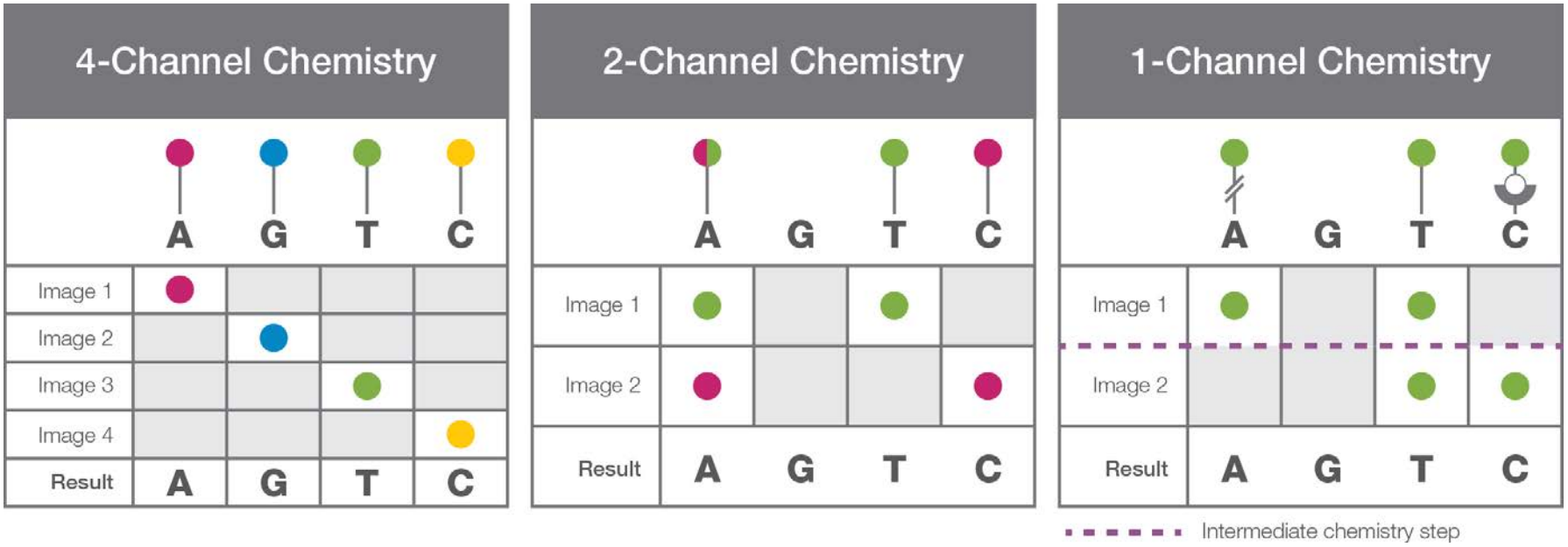
Incorporated Fl-NTPs excited then imaged (2 images/cycle)

Terminator & fluorescent dye cleaved from Fl-NTP

X # of bp in read

illumina®

Illumina Chemistry Comparison



4-channel SBS

- Bases are identified using four different fluorescent dyes, one for each base and four images per sequencing cycle

2-channel SBS

- Simplified nucleotide detection by using two fluorescent dyes and two images to determine all four base calls

1-channel SBS

- Base calling uses one fluorescent dye and two images, with chemistry step in between, to determine all four base calls

illumina®

Paired-End Sequencing

illumina®

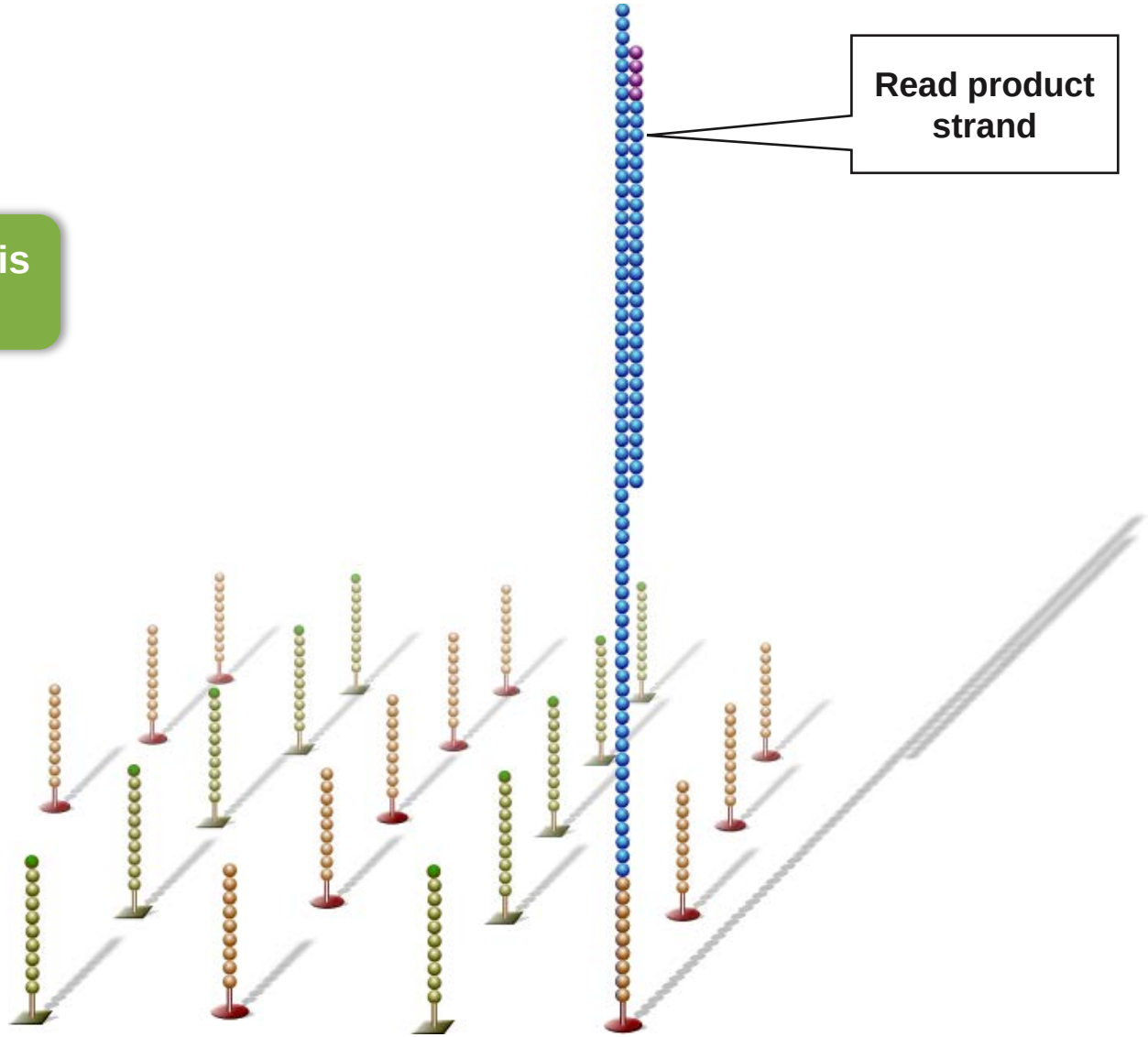
Sequencing with Paired-Ends

Reference	This is really the best way to do sequencing
Single-reads	This is
…	is really
…	really the
…	the best
…	sequencing
Paired-reads	This is (----100 characters-------) sequencing

Assembly becomes easier

Paired-End Sequencing

The read product is stripped off

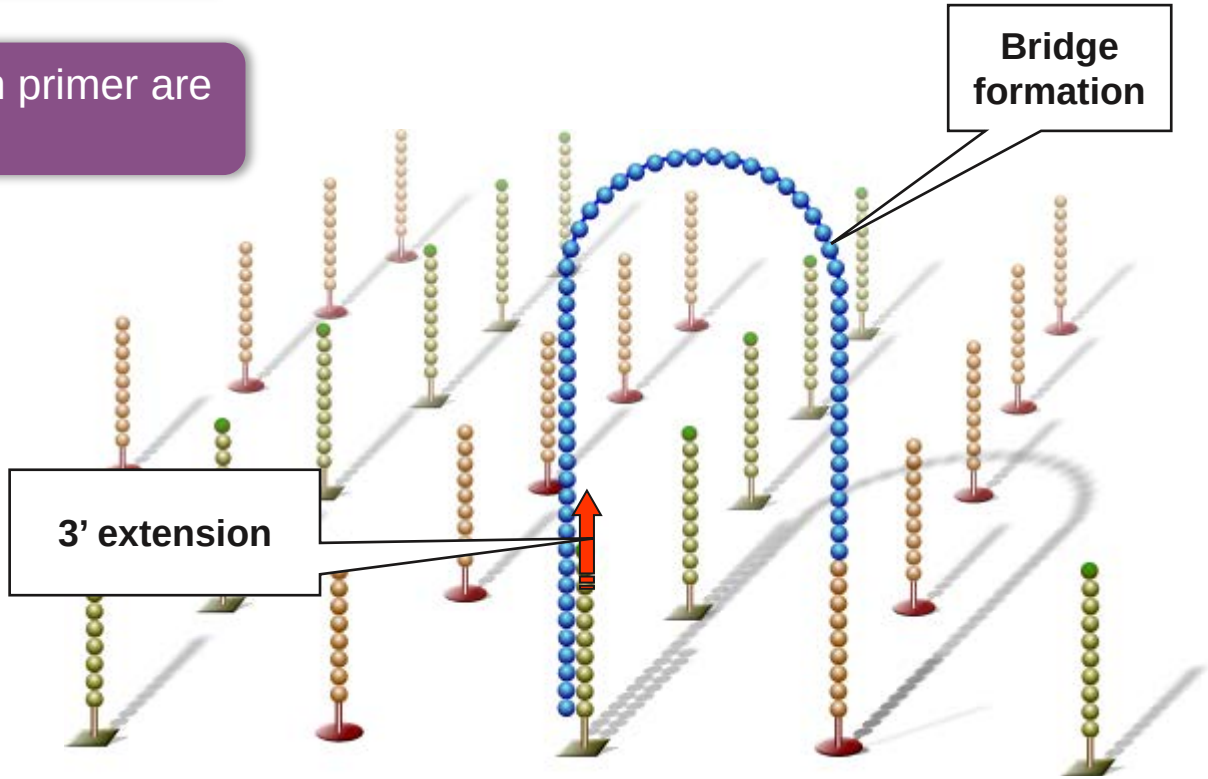


illumina®

Paired-End Sequencing

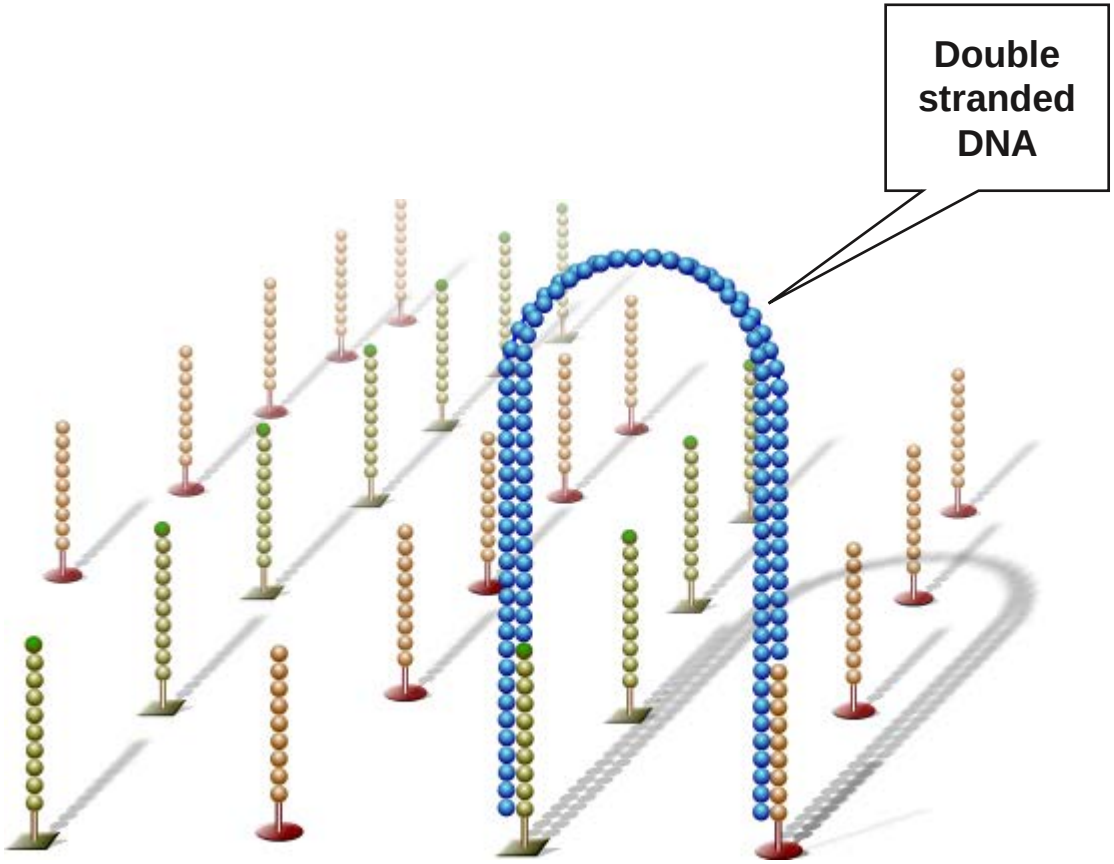
Single-stranded template loops over to form a bridge by hybridizing with a lawn primer

3’-ends of lawn primer are extended



illumina®

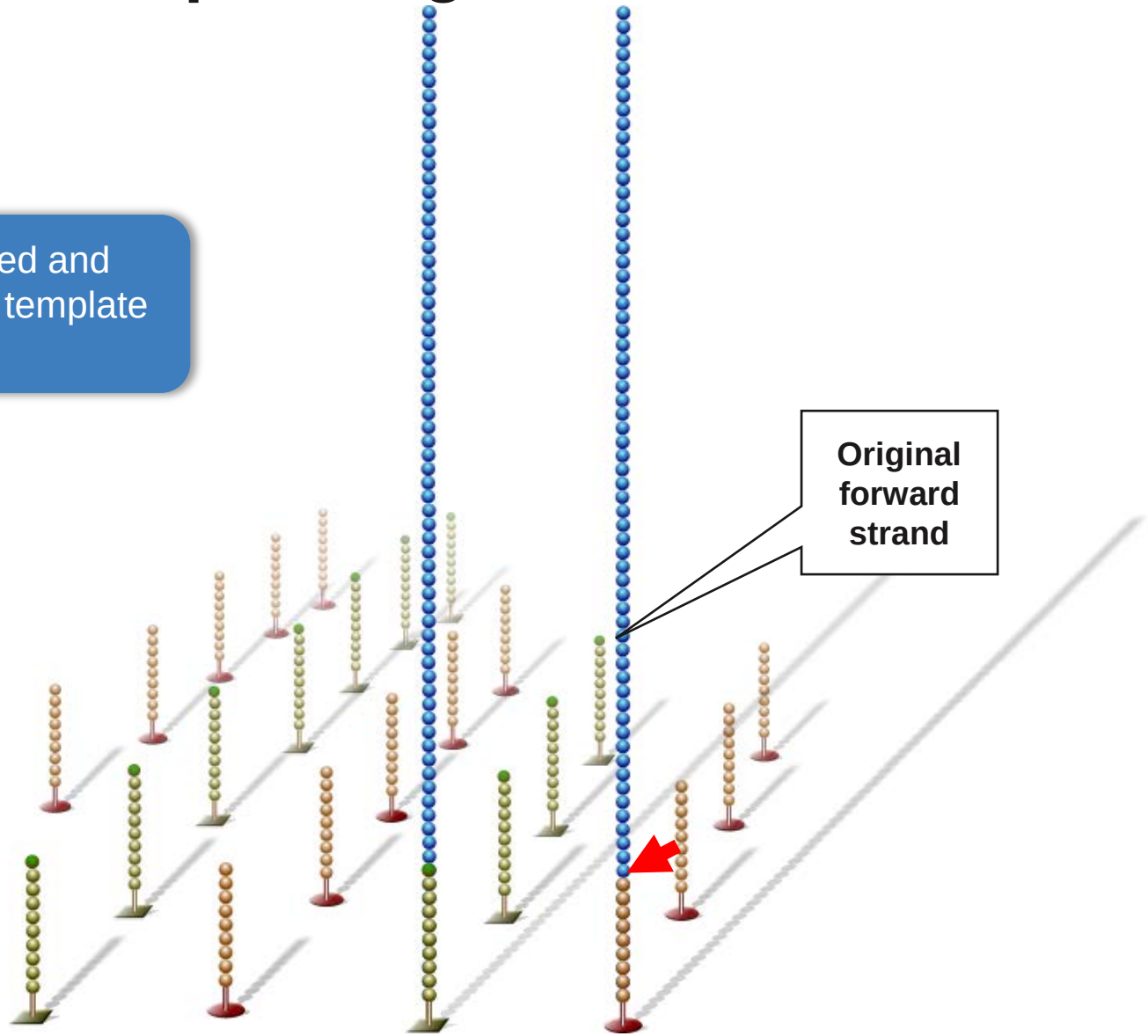
Paired-End Sequencing



illumina®

Paired-End Sequencing

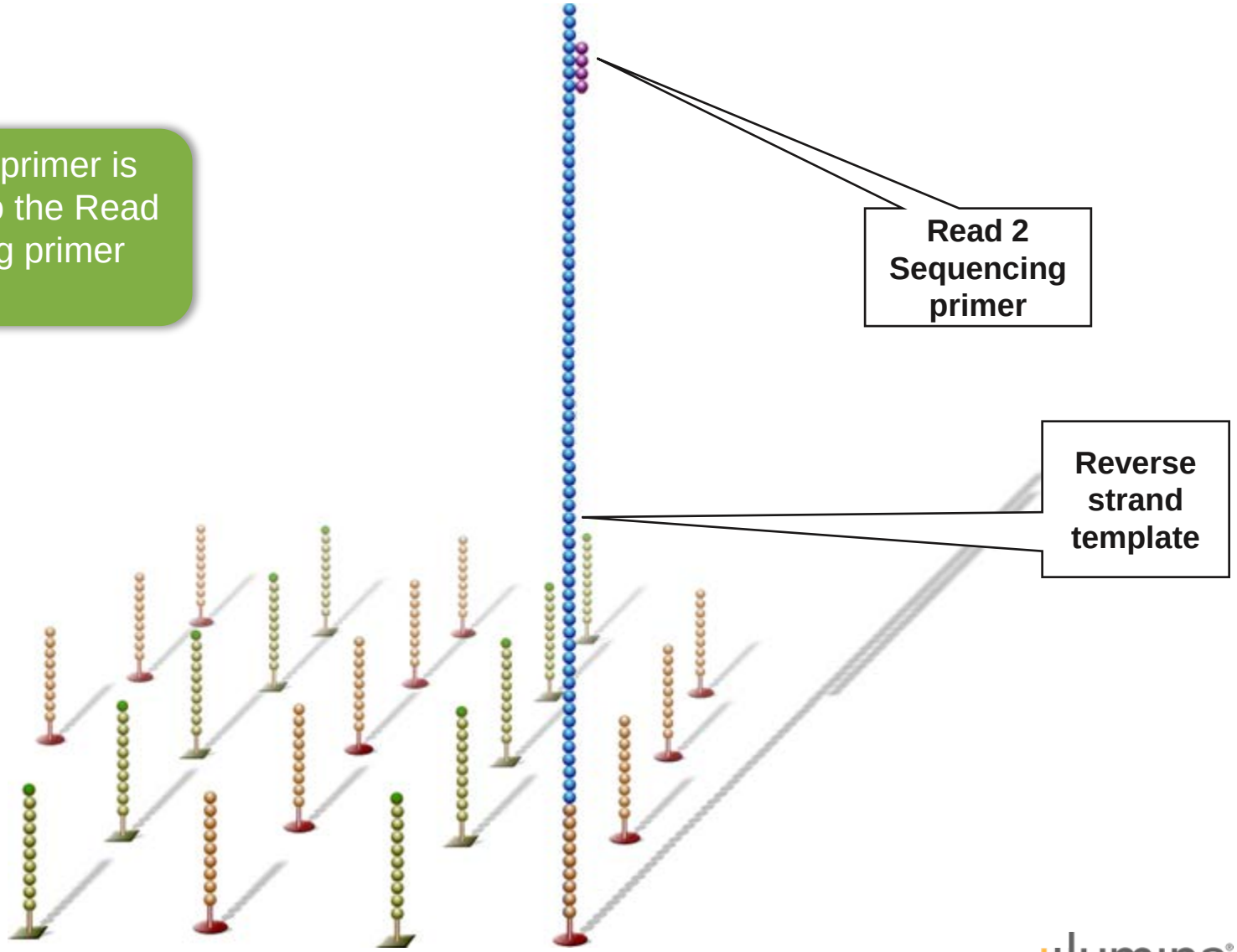
Bridges are linearized and the original forward template is cleaved



illumina®

Paired-End Sequencing

Sequencing primer is hybridized to the Read 2 sequencing primer binding site

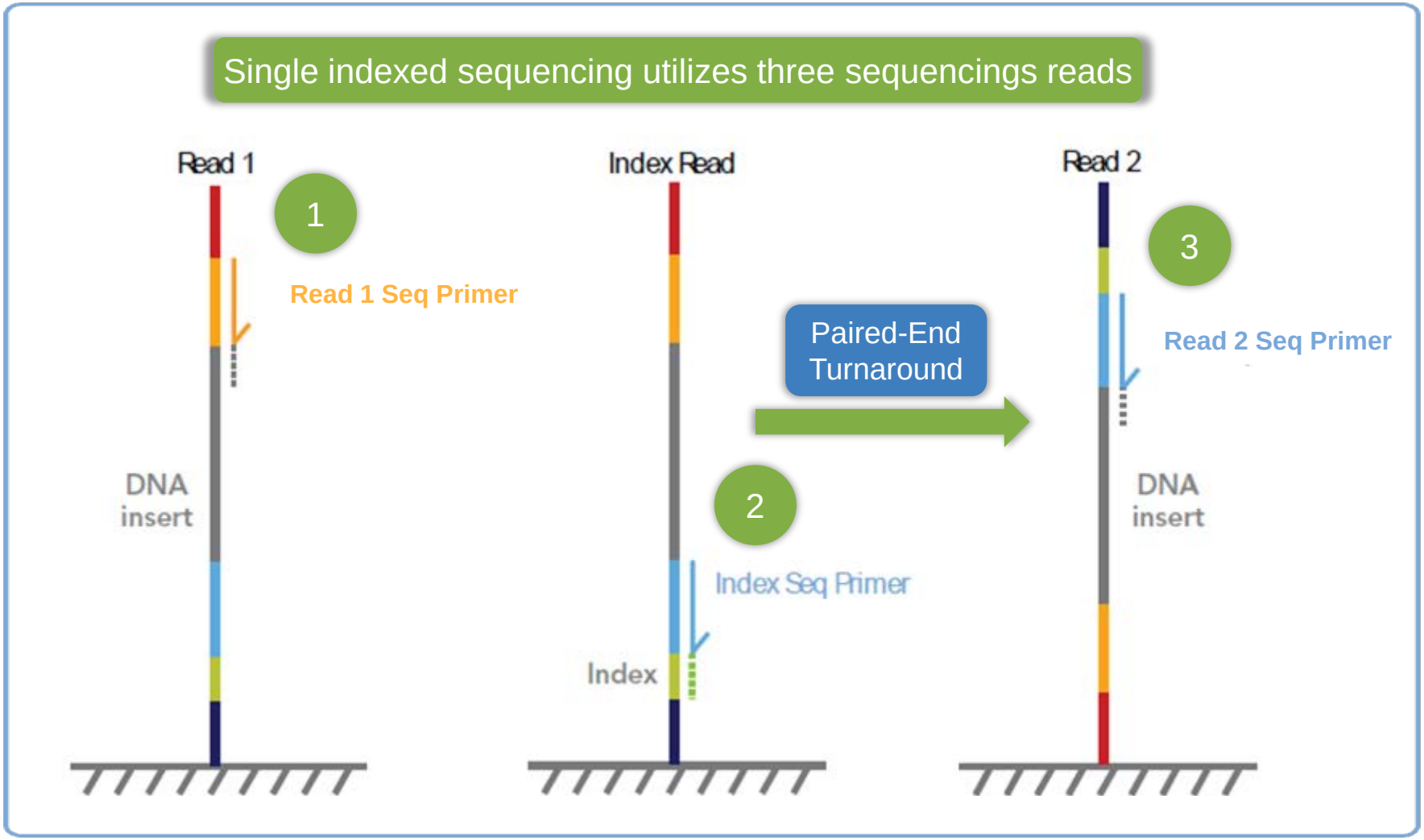


Sequencing with Index Reads

illumina®

Single Index Reads

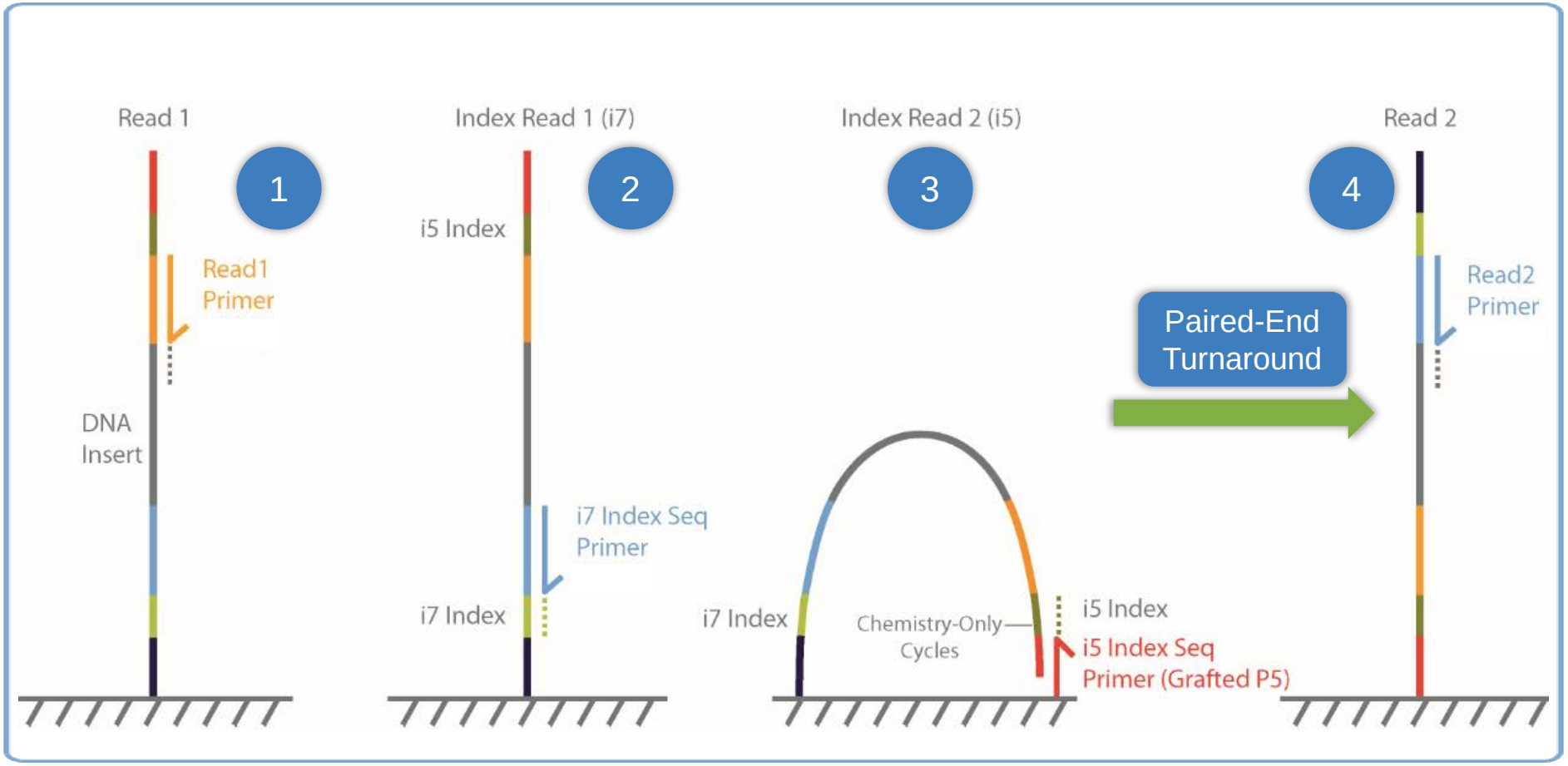
All Platforms



illumina®

Dual Index Reads

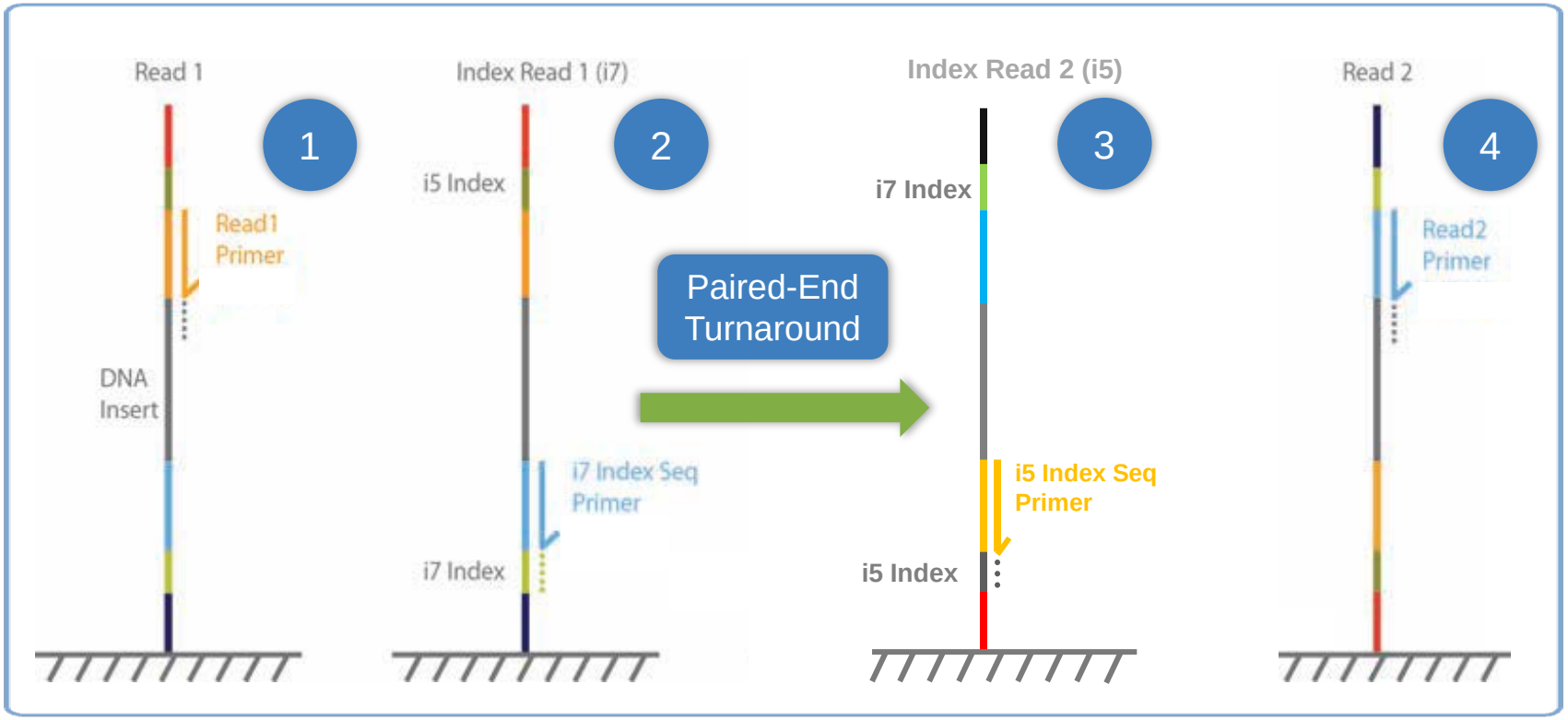
MiSeq, HiSeq 2500, NovaSeq 6000



Dual indexed sequencing utilizes four sequencing reads

Dual Index Reads

iSeq 100, MiniSeq, NextSeq, HiSeq 3000/4000



illumina®

Illumina Sequencing Workflow

1. Library Preparation

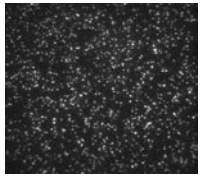
2. Cluster Generation

3. Sequencing

4. Data Analysis

illumina®

Analysis Overview

Analysis Type	Software	Outputs
Control Software		  Images, Intensities and Base Calls
Analysis Software	 	 Alignments, Variant Detection
Visualization Software	 	  Annotation, Filtering, Reports

Questions?

Meredith Millis
mmillis@illumina.com

illumina®