Illumina Sequencing Overview: Library Prep to Data Analysis

Noel Lenny, PhD
Sr. Clinical Field Applications Scientist
28-SEP-2020

QB7845

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

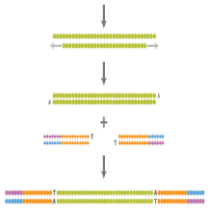
illumina®

Session Objectives

- **By the end of this training, you will be able to:**
  - List the major steps in the Illumina sequencing workflow
    - Library Preparation
    - Cluster Generation
    - Sequencing
    - Data Analysis
  - Discuss the sequencing by synthesis process
    - 4-Channel Chemistry
    - 2-Channel Chemistry
    - 1-Channel Chemistry

Illumina Sequencing Workflow

1. **Library Preparation**



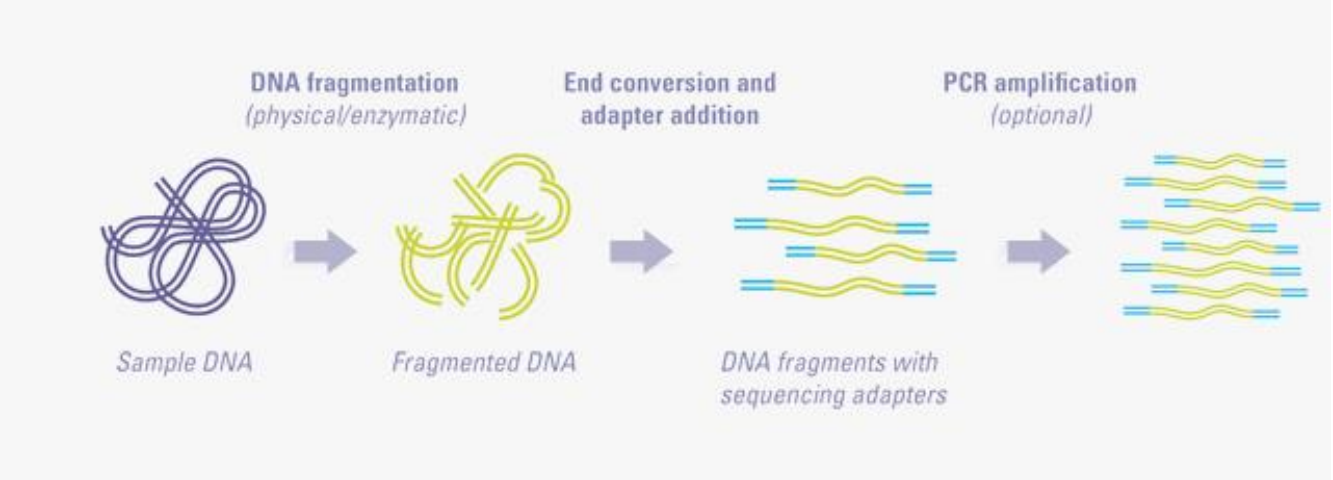
2. **Cluster Generation**

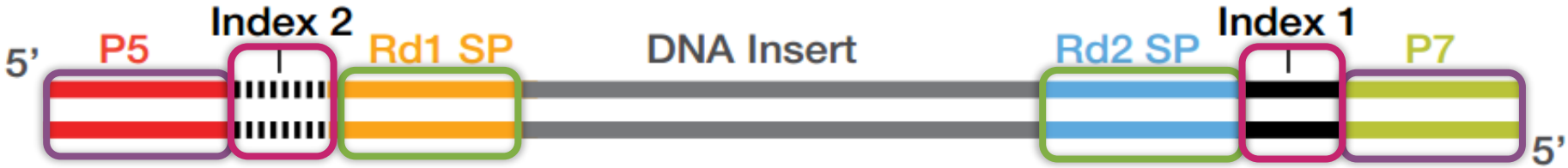
3. **Sequencing**

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

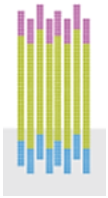
For mixing samples:
Libraries must have a unique index or barcodes sequence

illumina®

Illumina Sequencing Workflow

1. Library Preparation

2. Cluster Generation



3. Sequencing

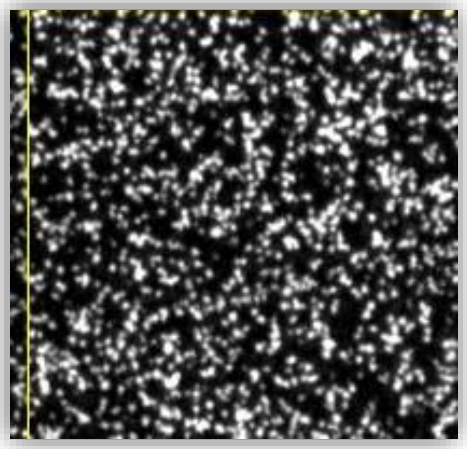
4. Data Analysis

illumina®

What is a Cluster?

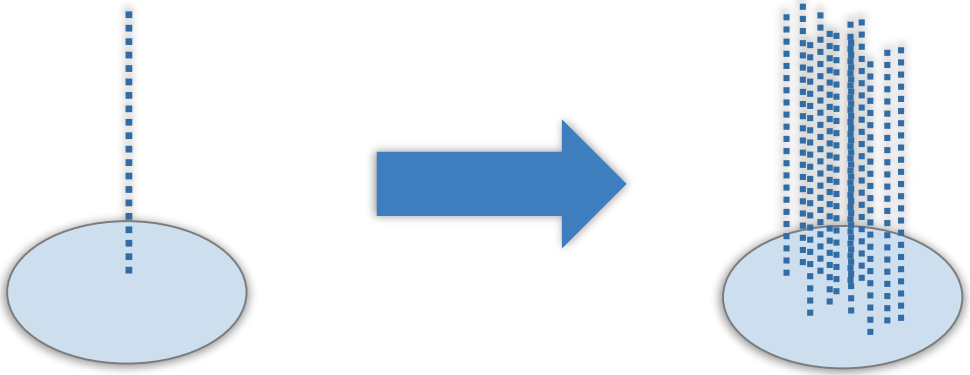
Clusters are a group of DNA strands positioned closely together

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

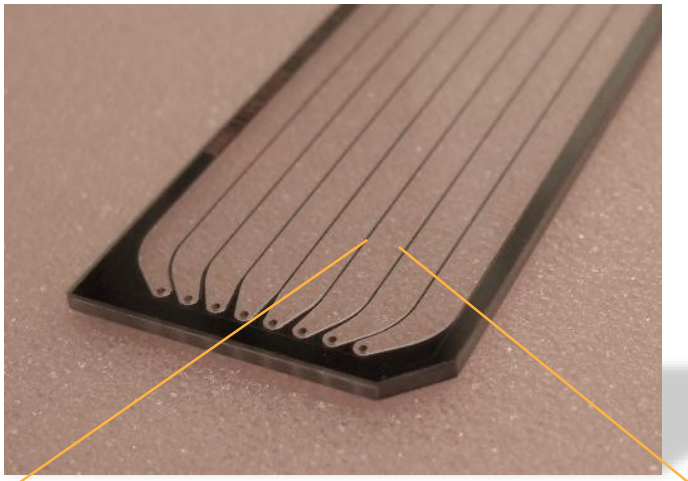
illumina®

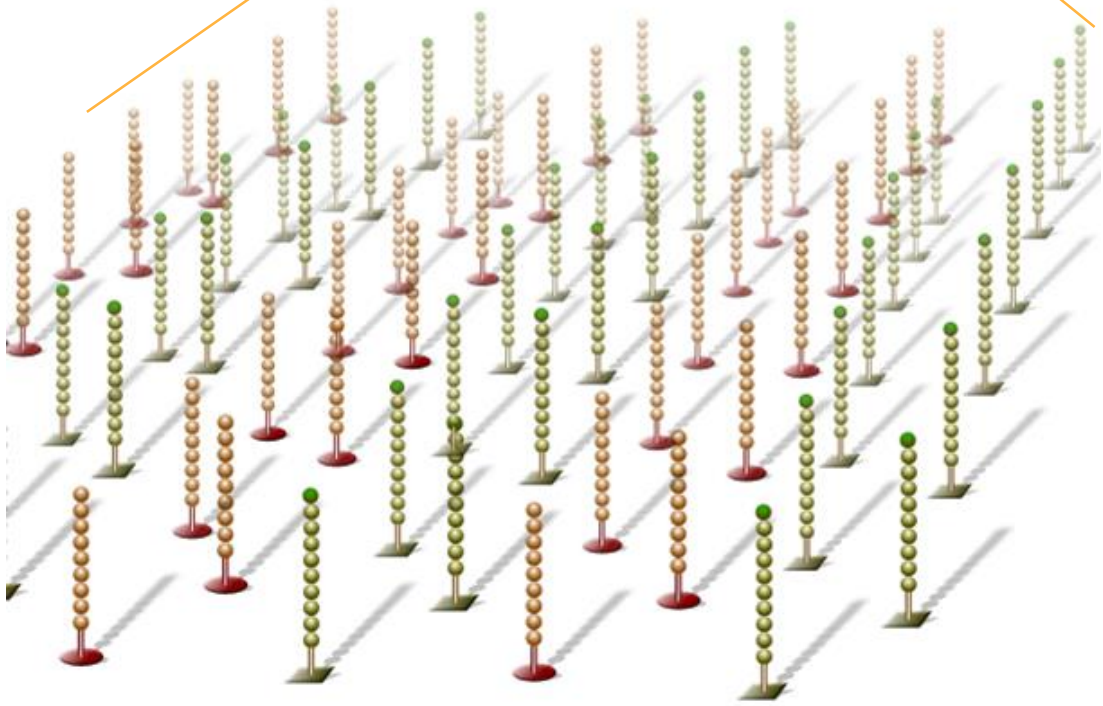
What is a flow cell?

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



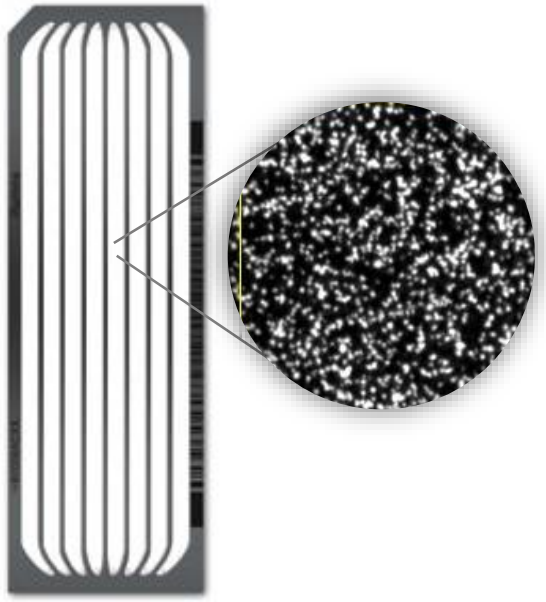


Flow cell Architecture

Random vs Patterned

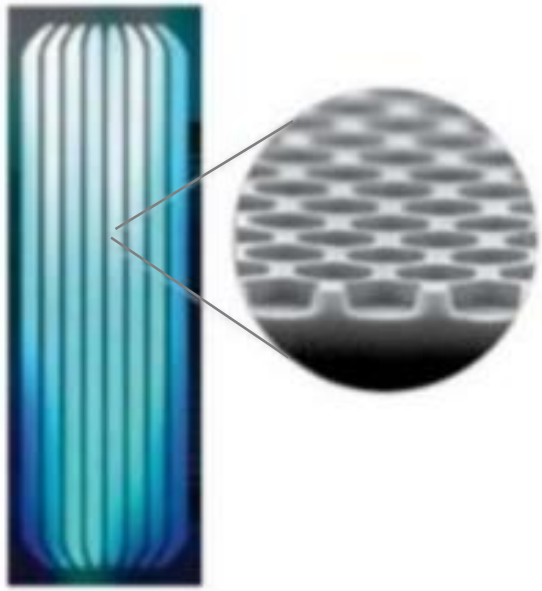
Random Flow Cell

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



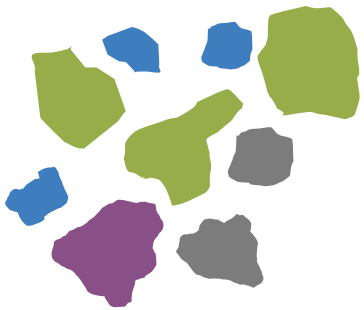
Patterned Flow Cell

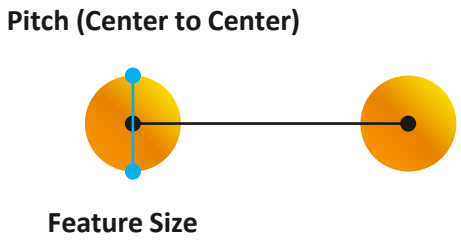
- HiSeq 3K/4K/X, NovaSeq™ 6000, iSeq™ 100
- Defined size and spacing
- Increased Cluster density
- Simplified imaging

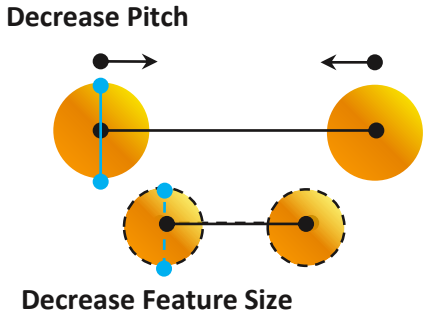


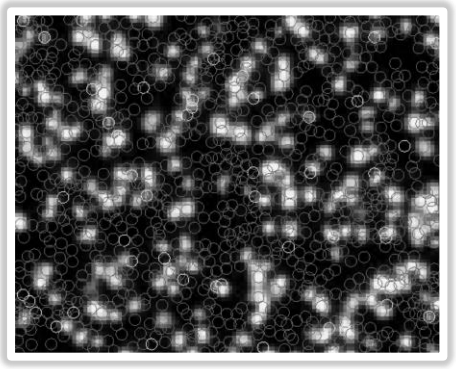
Patterned flow cells

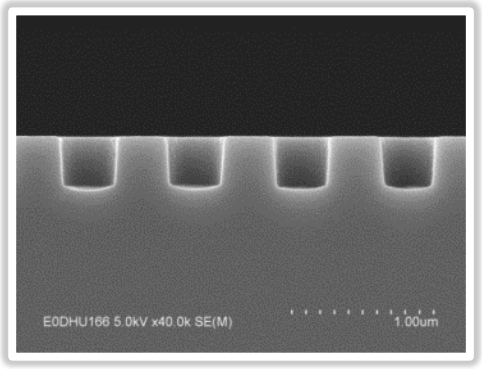
Complete control of pitch & feature size

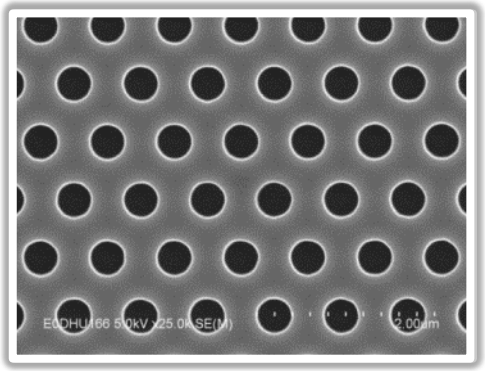








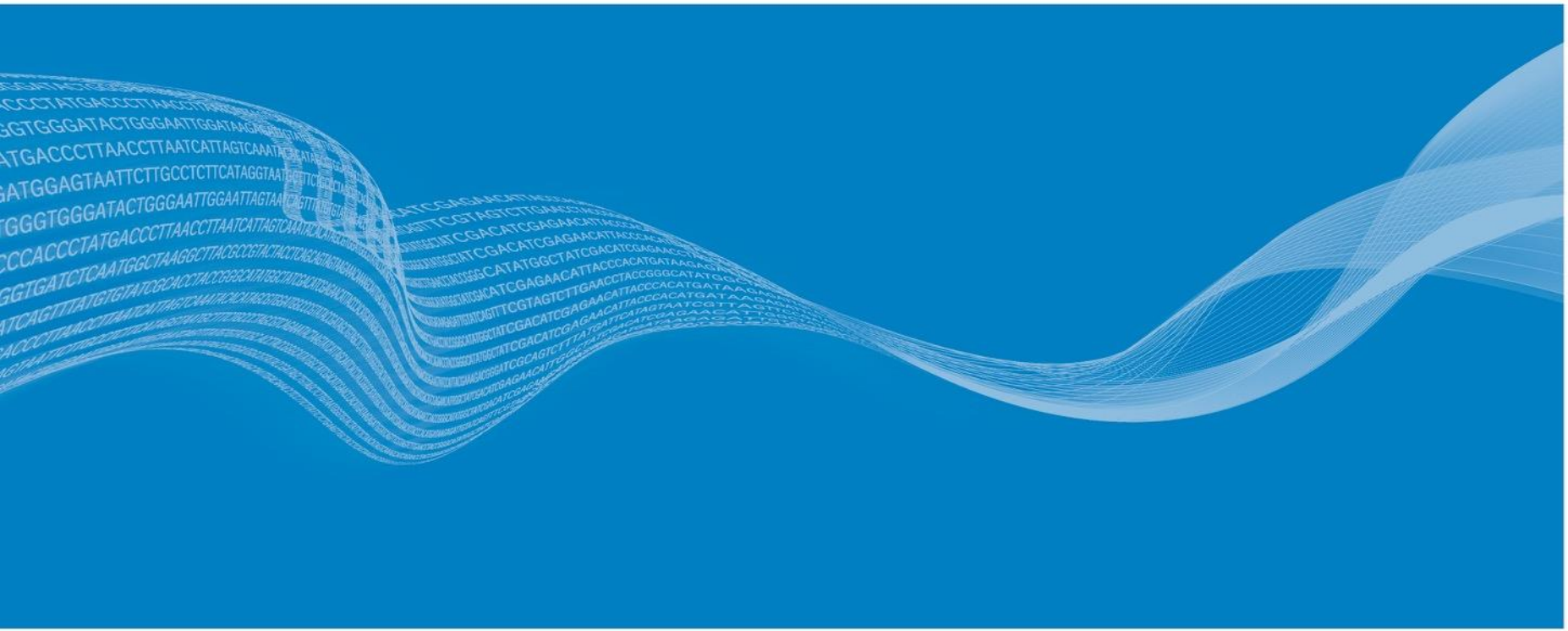




Random	Patterned
	Rigid registration reduces time by skipping template generation

illumina®

Traditional Cluster Generation

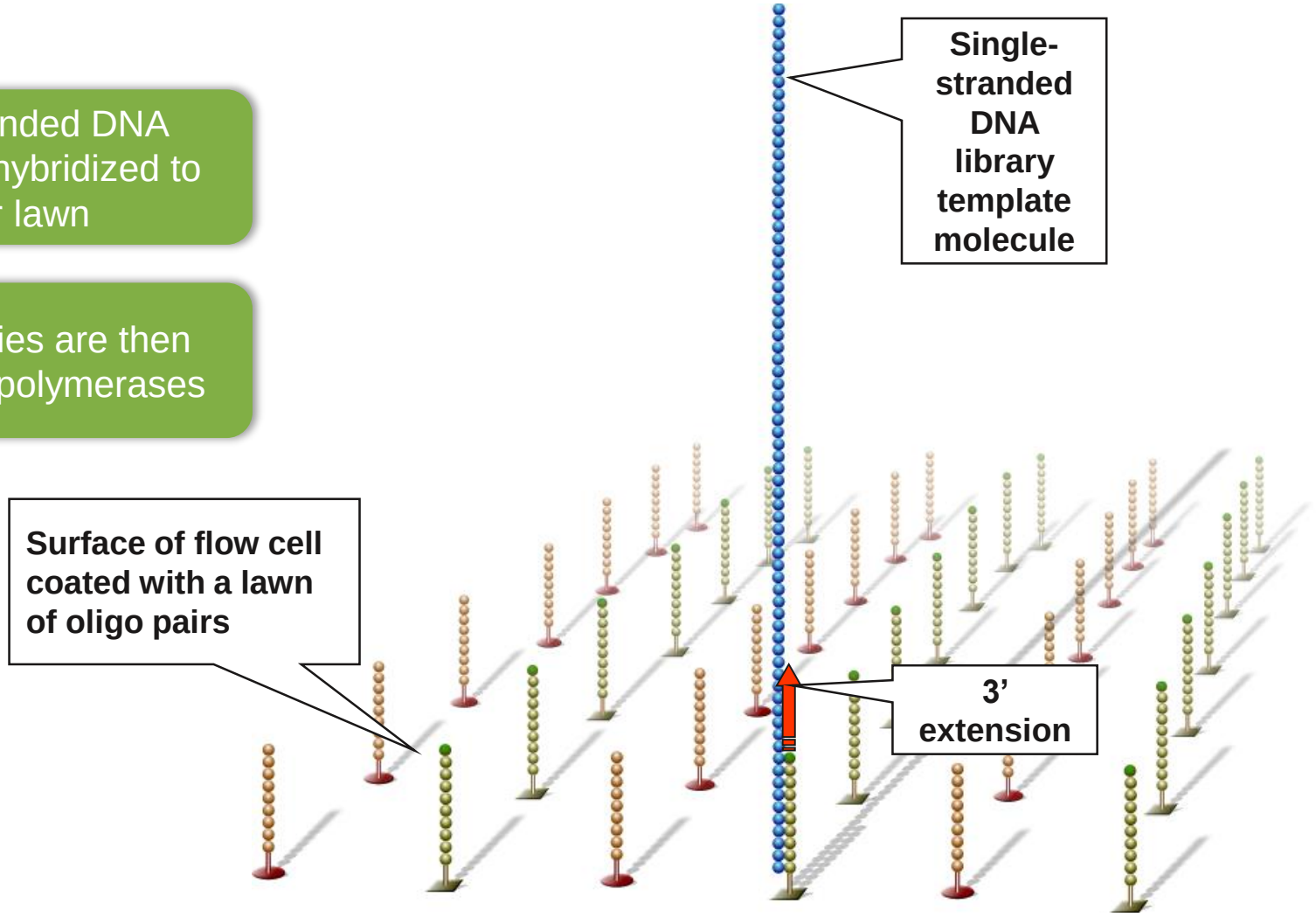


illumina®

Hybridize Fragment & Extend

Single-stranded DNA libraries are hybridized to primer lawn

Bound libraries are then extended by polymerases



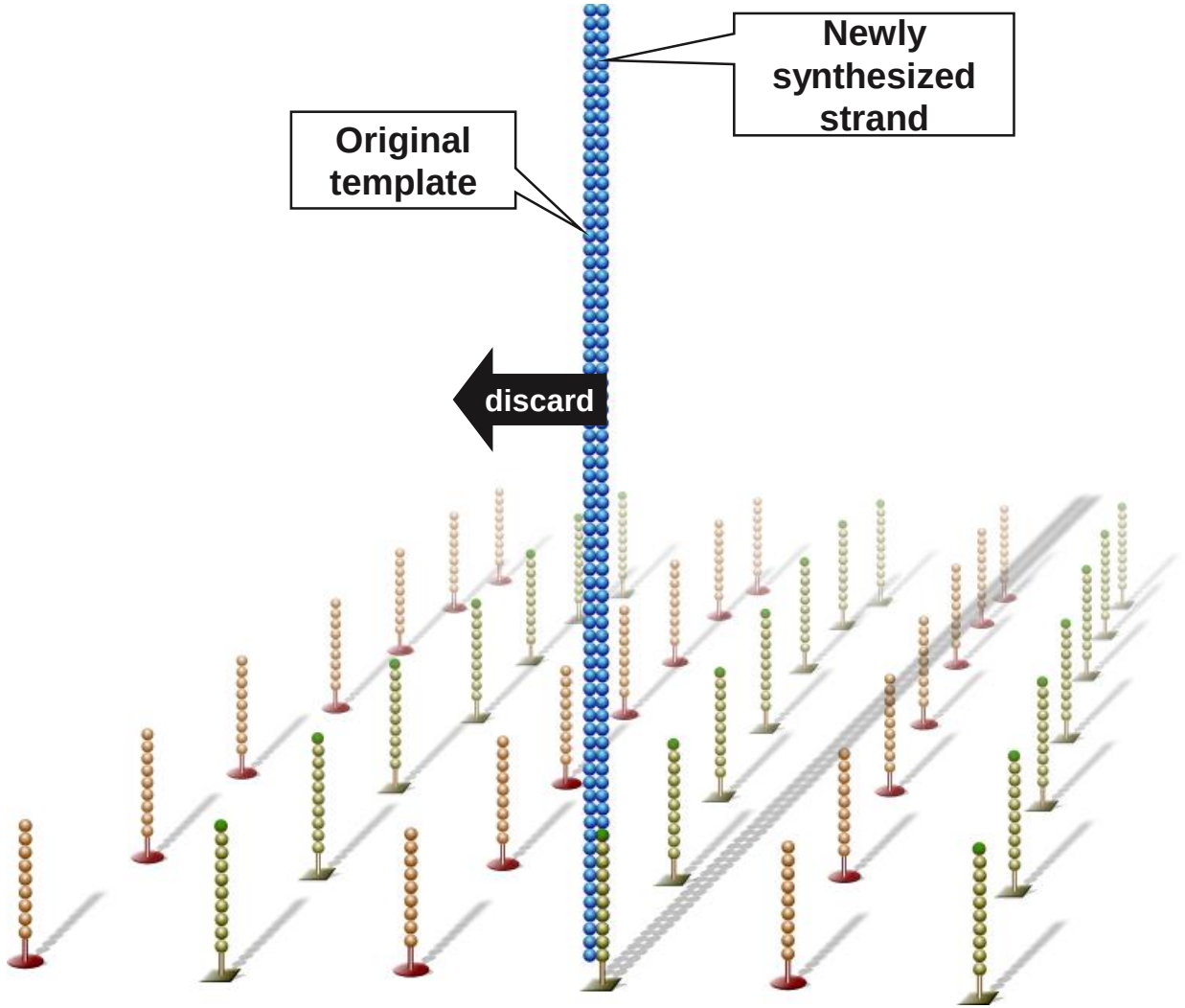
illumina®

Denature Double-Stranded DNA

Double-stranded molecule is denatured

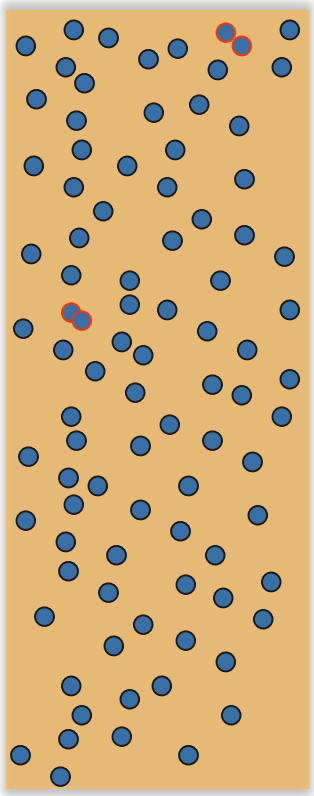
Original template washed away

Newly synthesized strand is covalently attached to flow cell surface

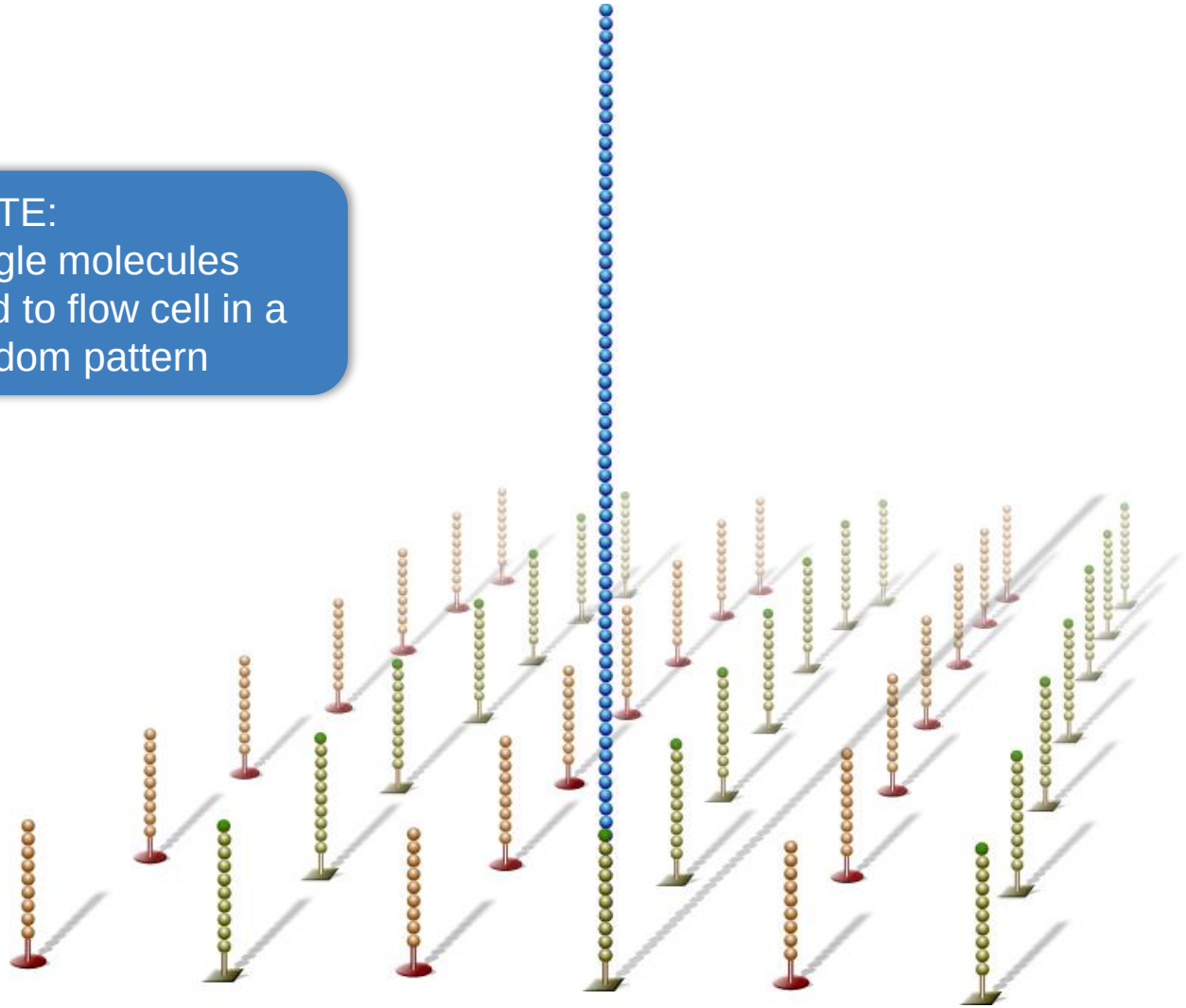


illumina®

Single-Stranded DNA



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

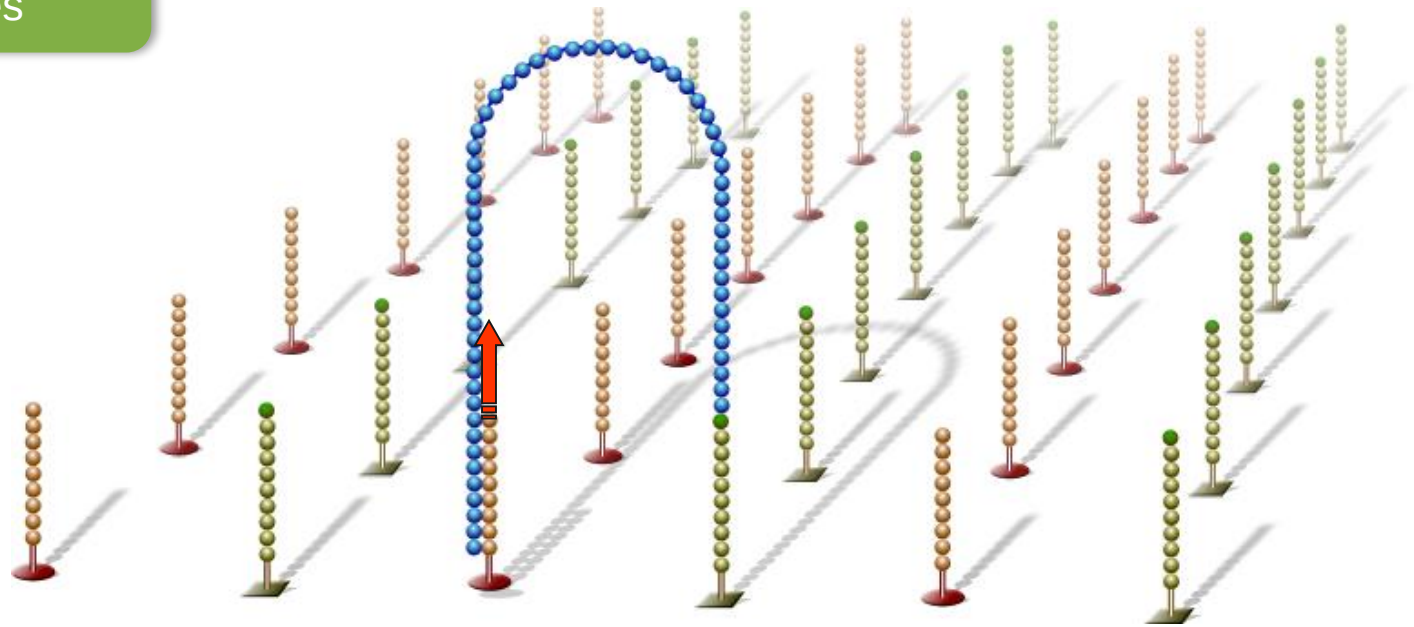


illumina®

Bridge Amplification

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

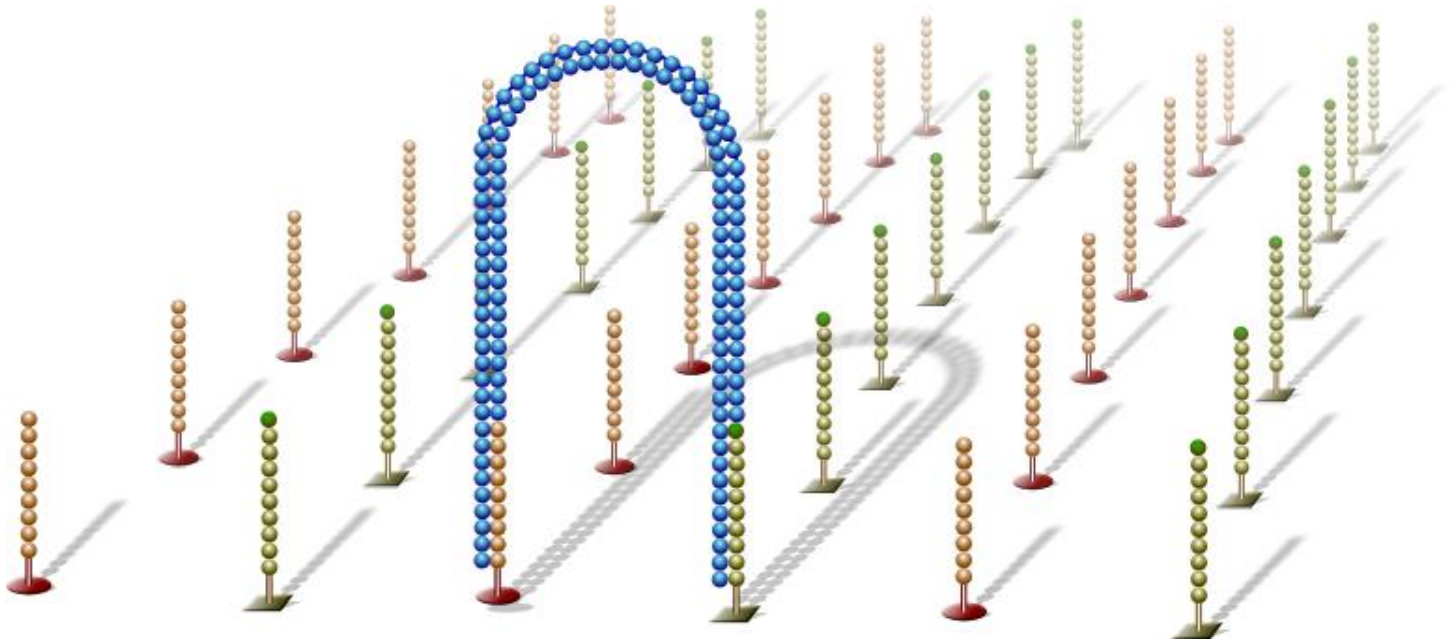
Hybridized primer extends by polymerases



illumina®

Bridge Amplification

Double-stranded bridge is formed

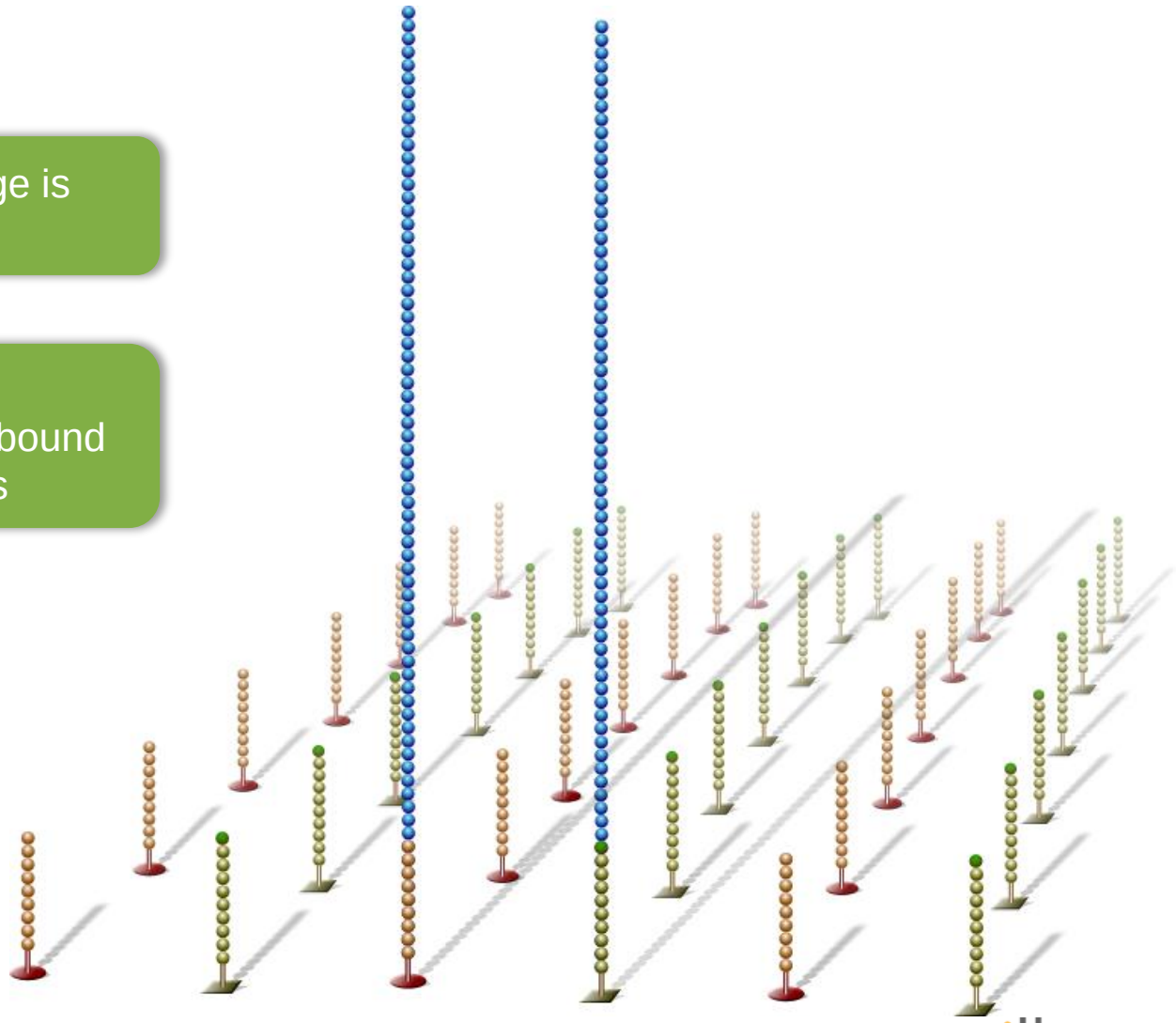


illumina®

Denature Double-Stranded Bridge

Double-stranded bridge is denatured

Result:
Two copies of covalently bound single-stranded templates

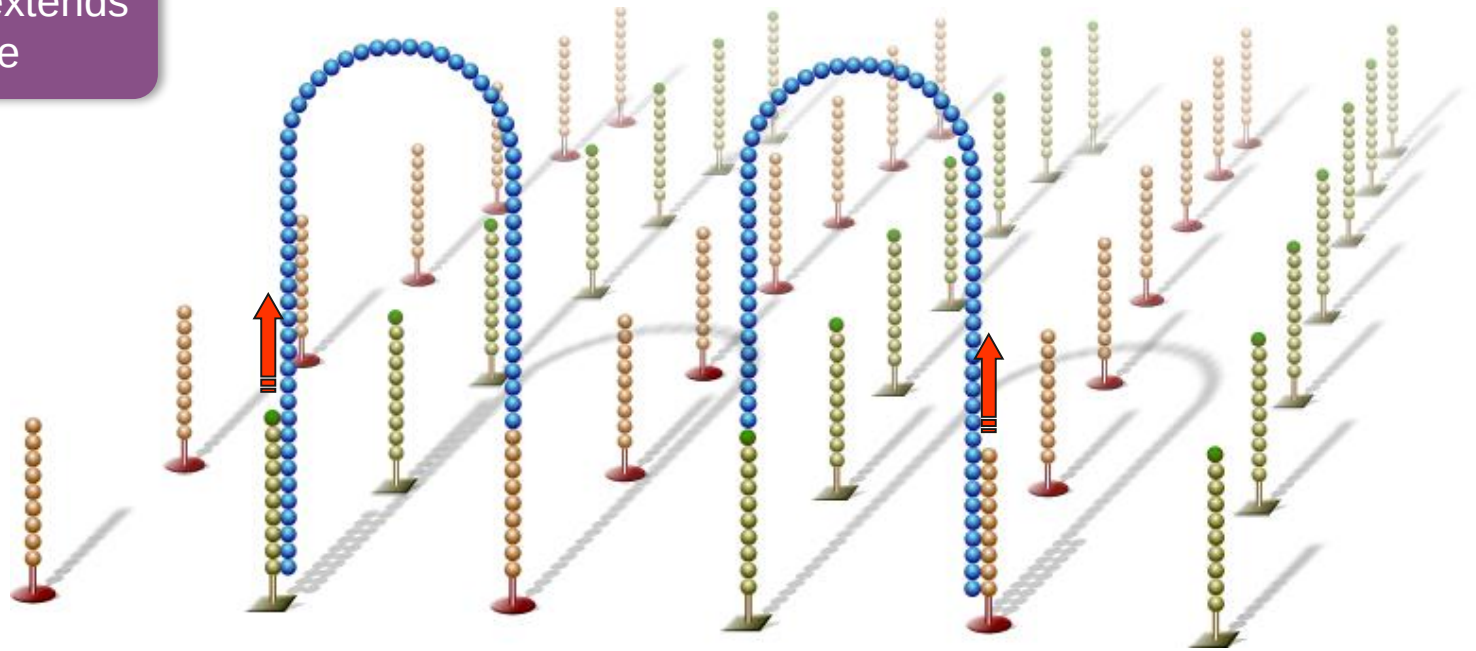


illumina®

Bridge Amplification

Single-stranded molecules flip over to hybridize to adjacent primers

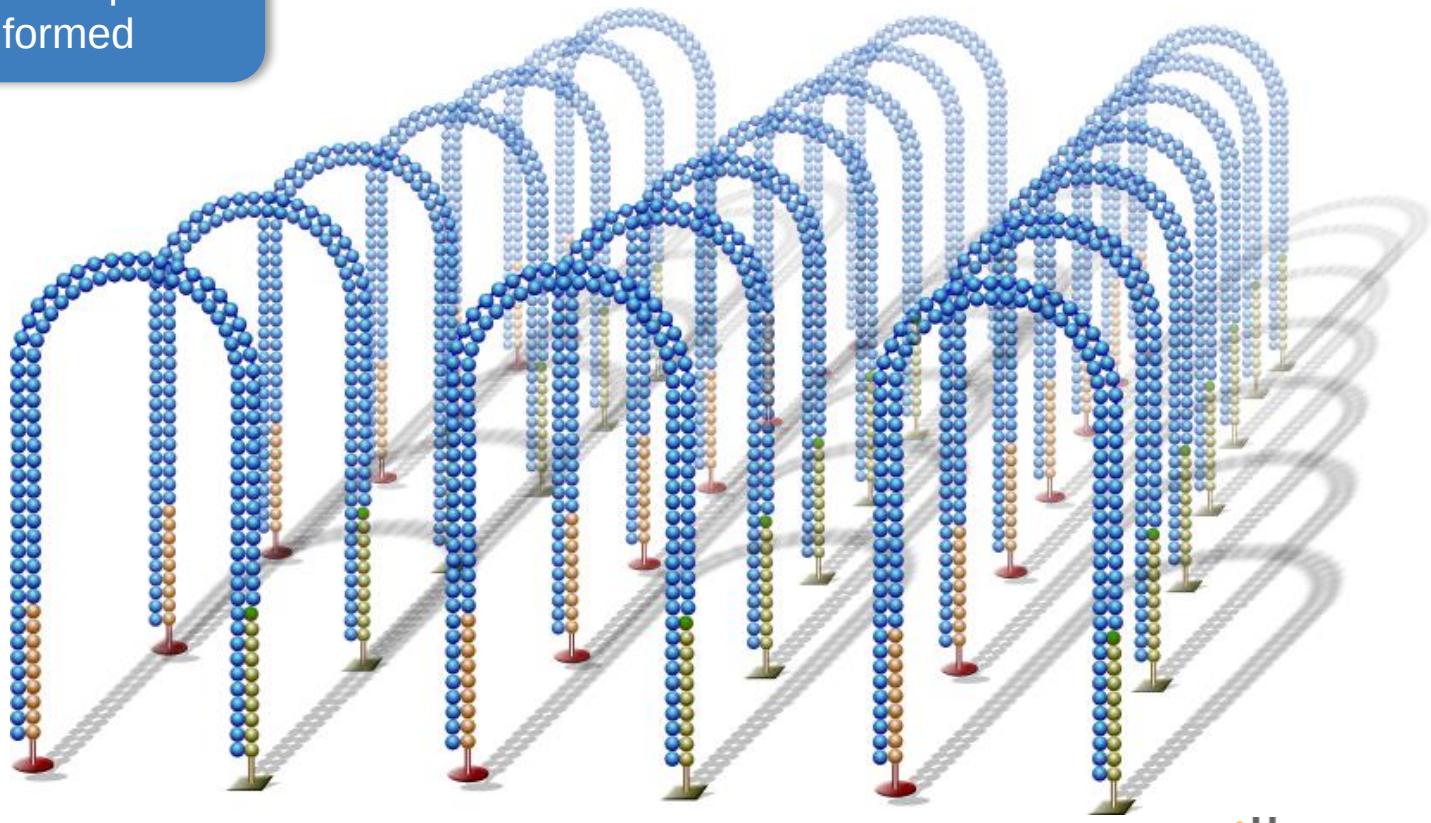
Hybridized primer extends by polymerase



illumina®

Bridge Amplification

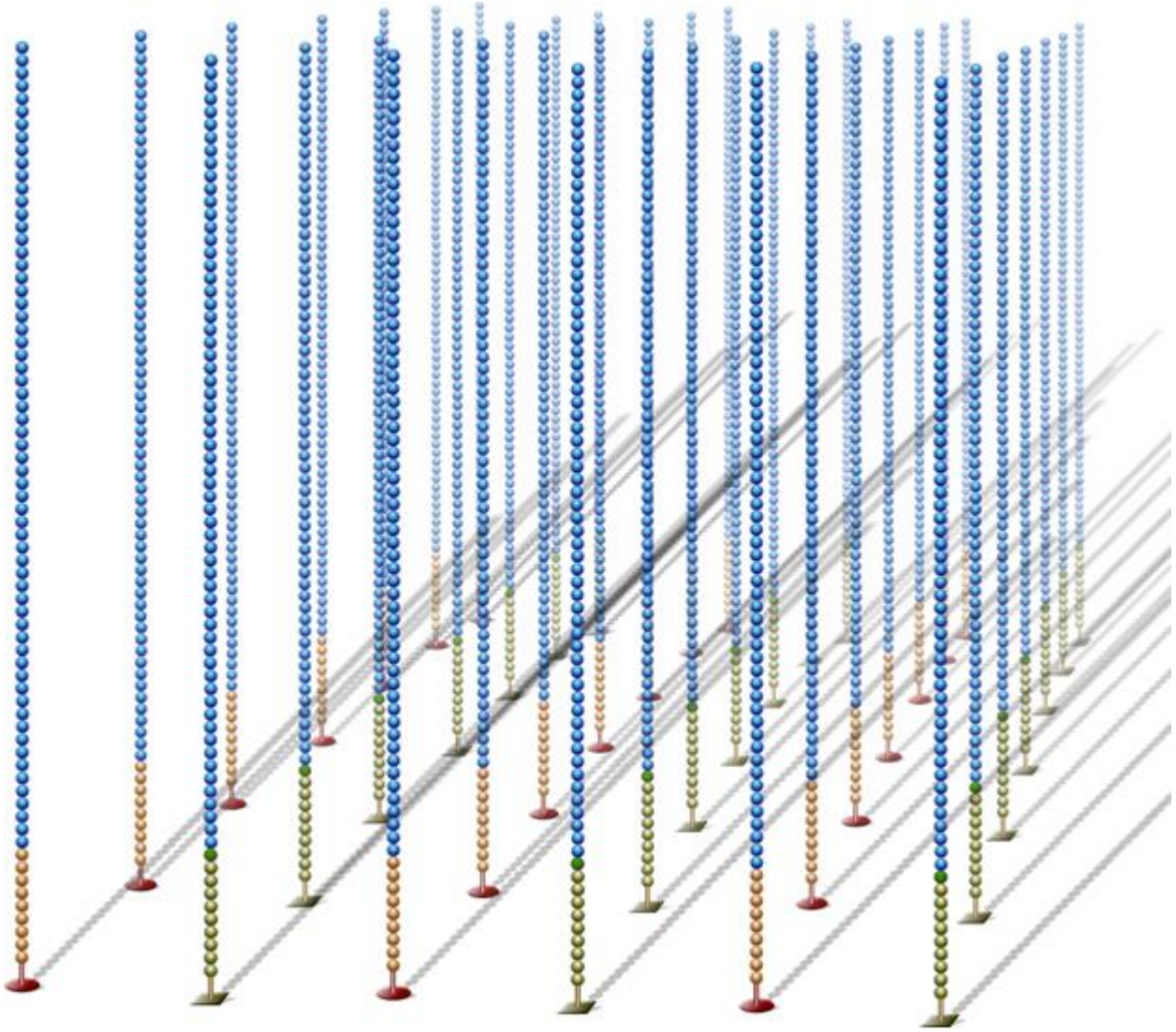
Bridge amplification cycle is repeated until multiple bridges are formed



illumina®

Linearization

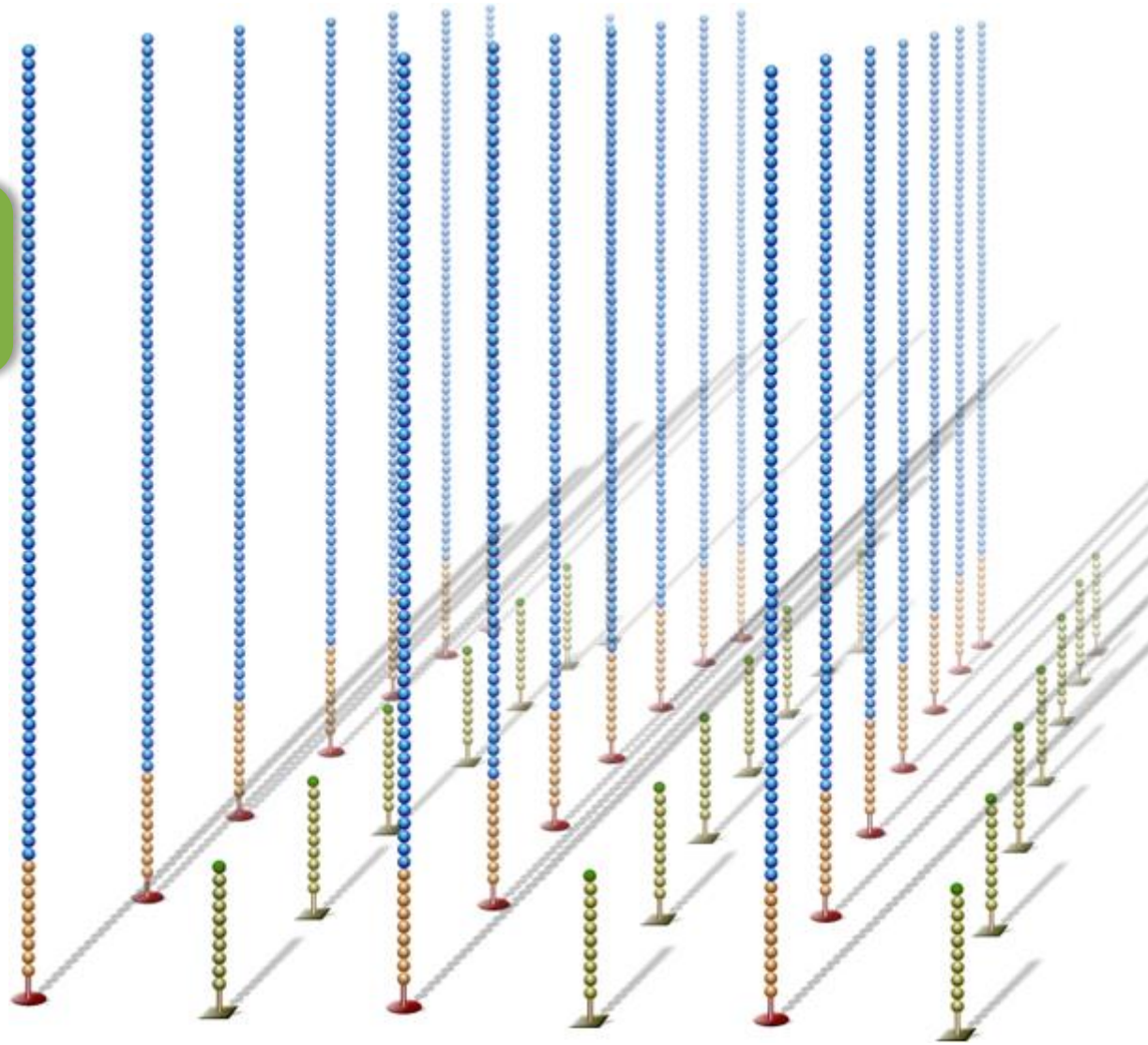
dsDNA bridges are denatured



illumina®

Reverse Strand Cleavage

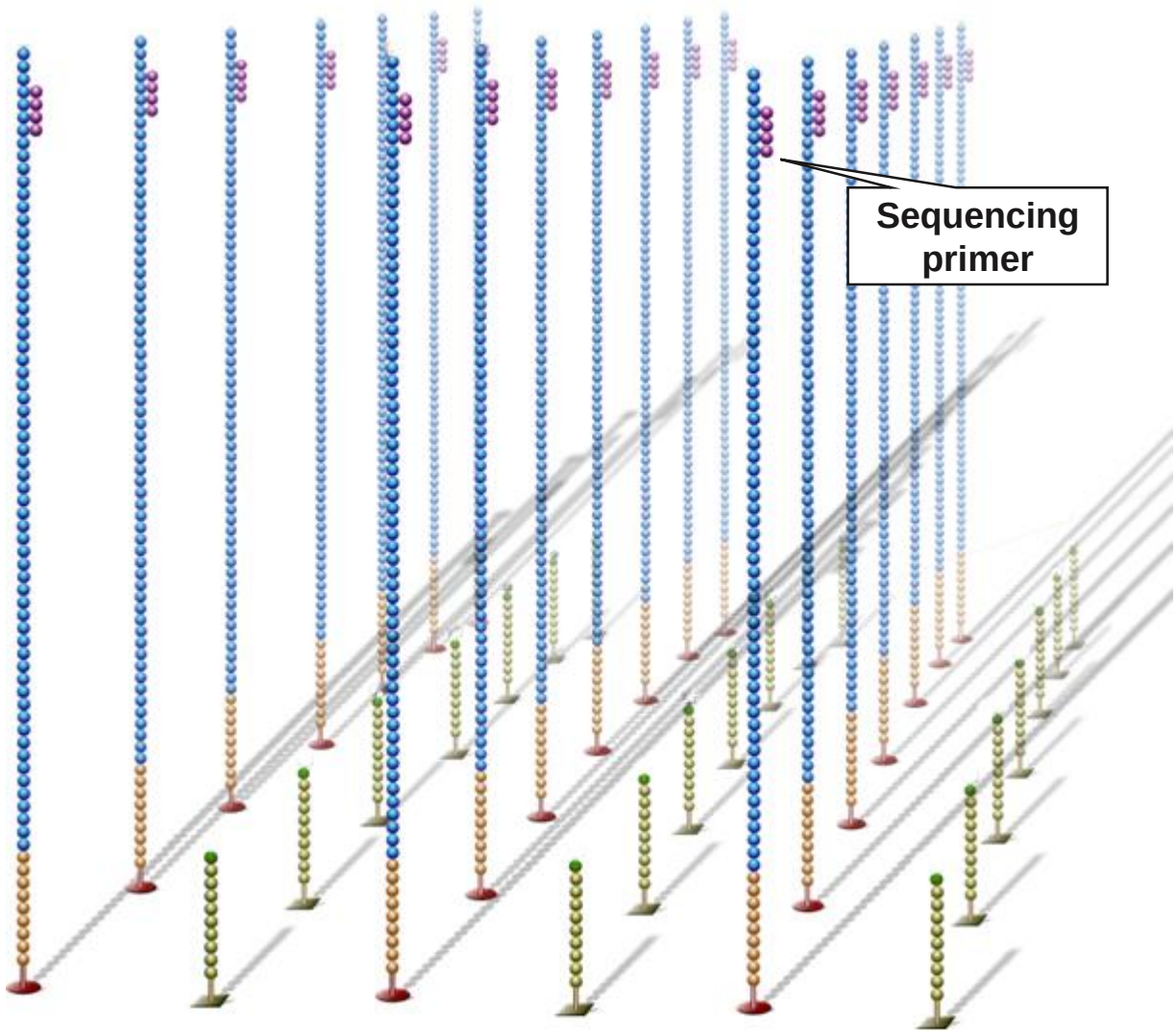
Reverse strands are cleaved and washed away, leaving a cluster with forward strands only



illumina®

Read 1 Primer Hybridization

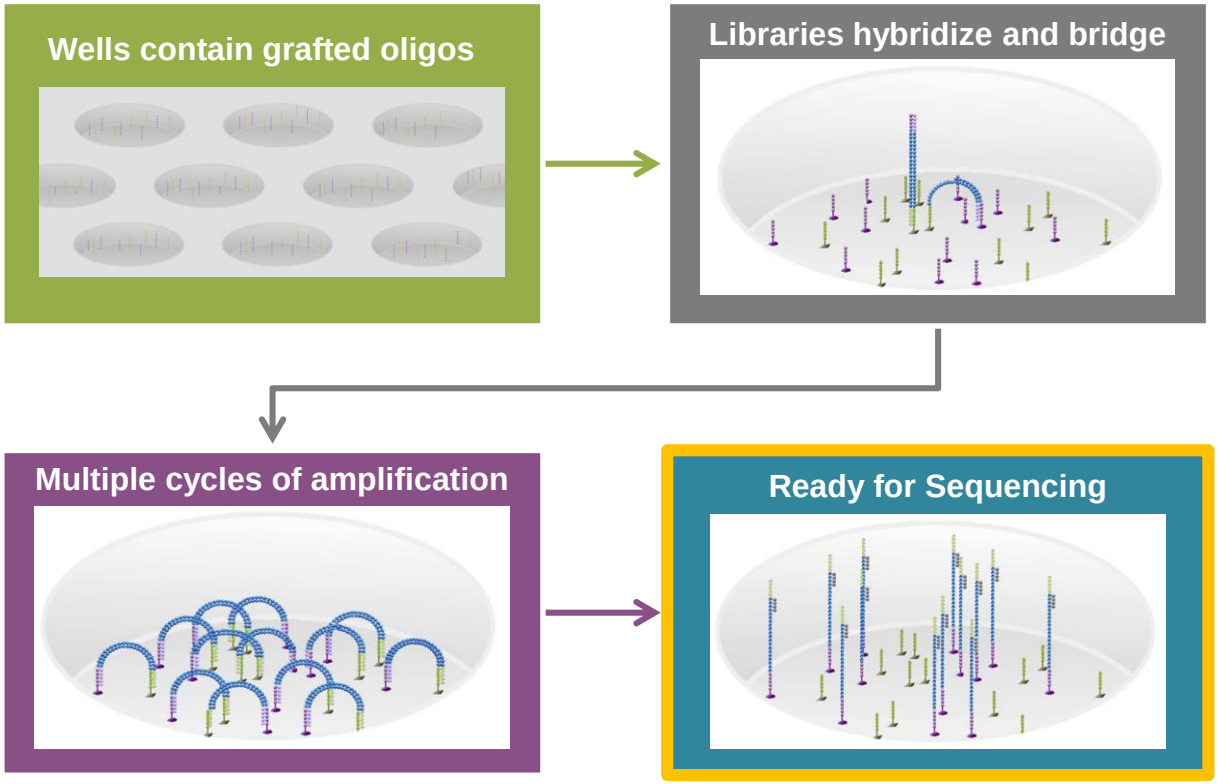
Sequencing primer is hybridized to Read 1 sequencing primer binding site



illumina®

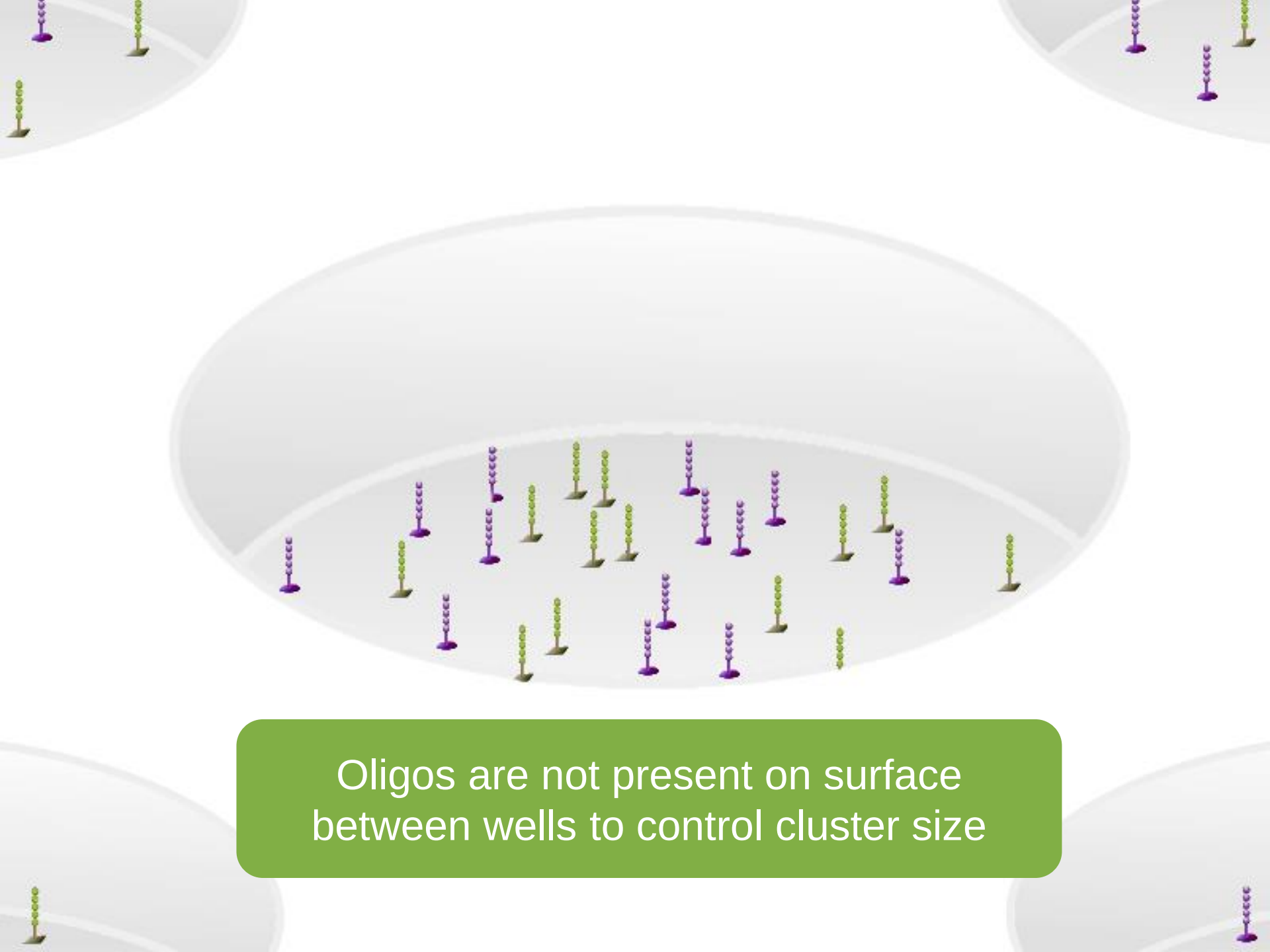
Patterned Flow Cell and ExAmp Technology

- **ExAmp technology creates clonal clusters in each well from individual library molecules**

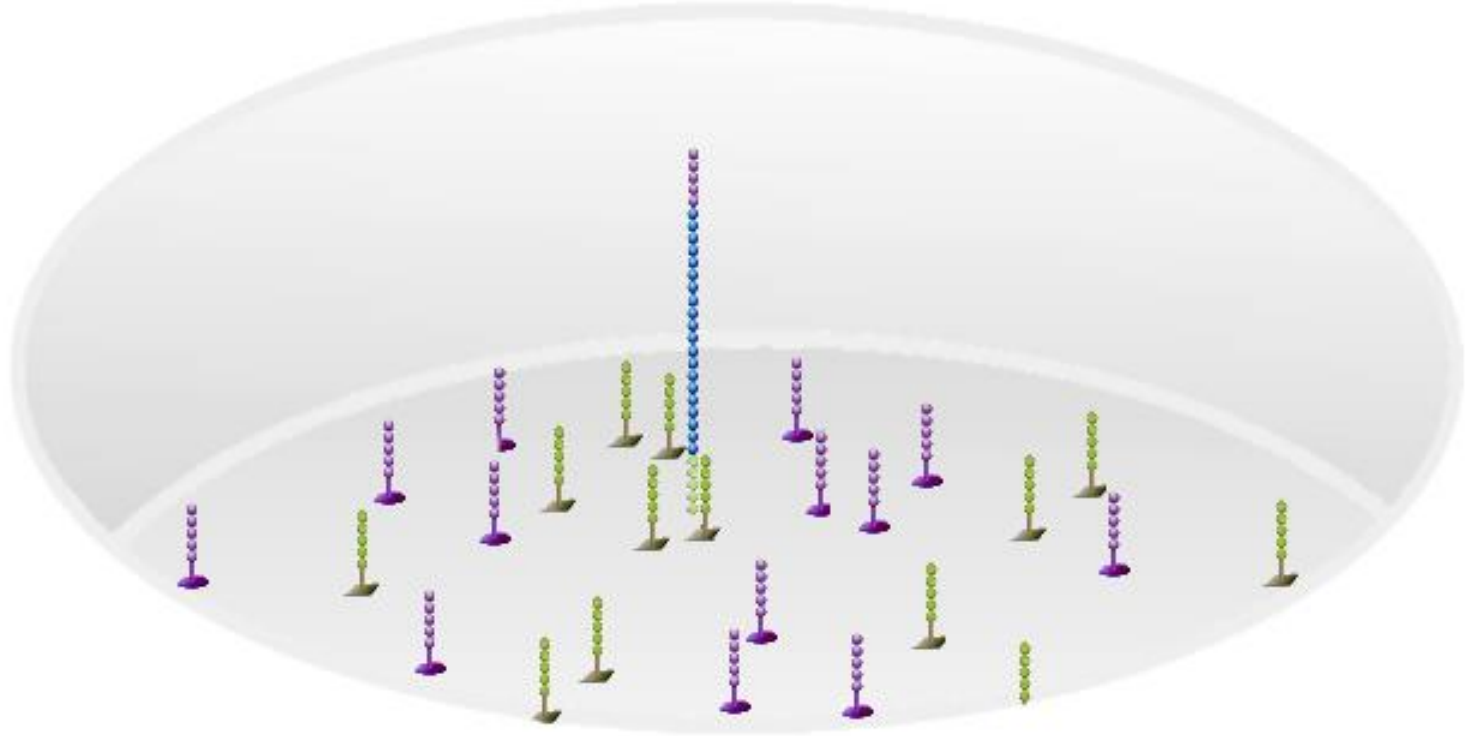


illumina®

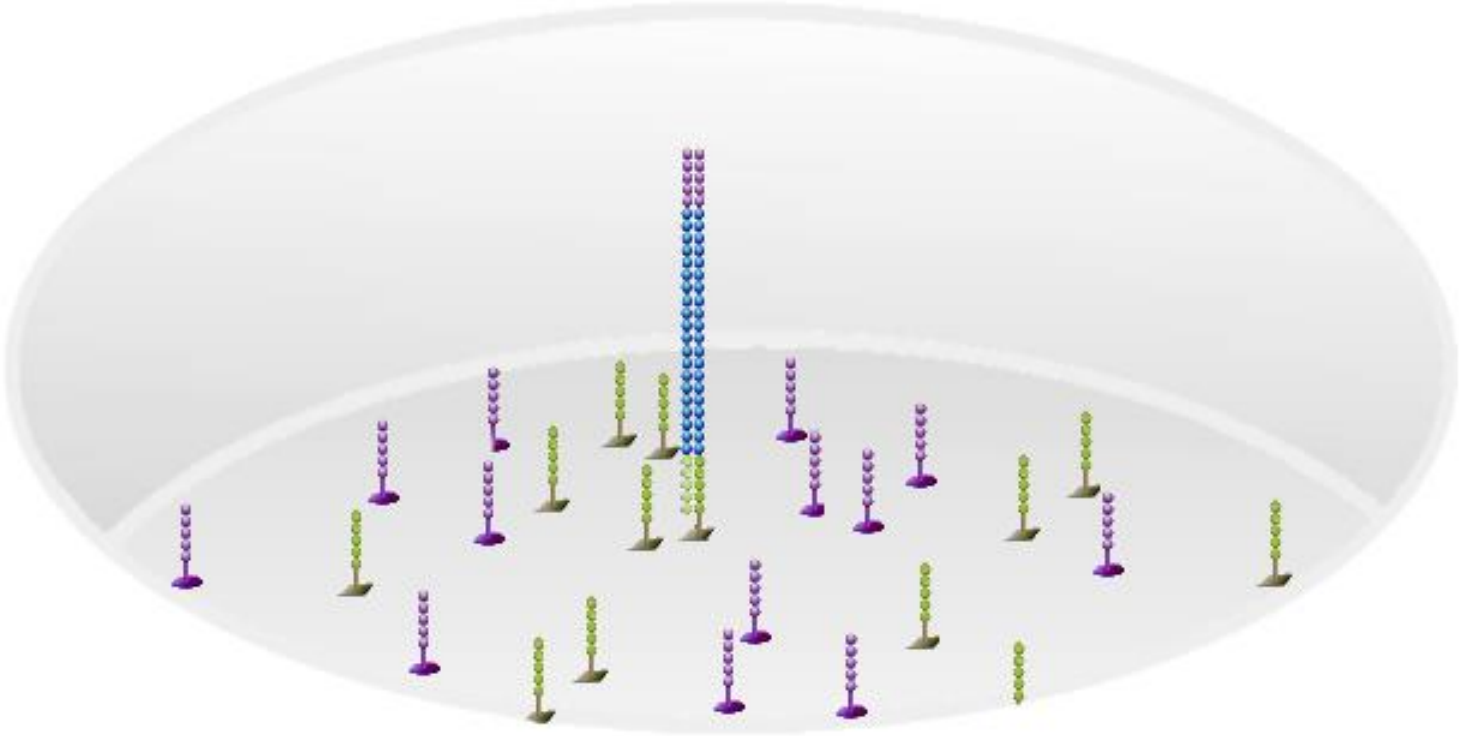
Oligos present inside patterned microwells



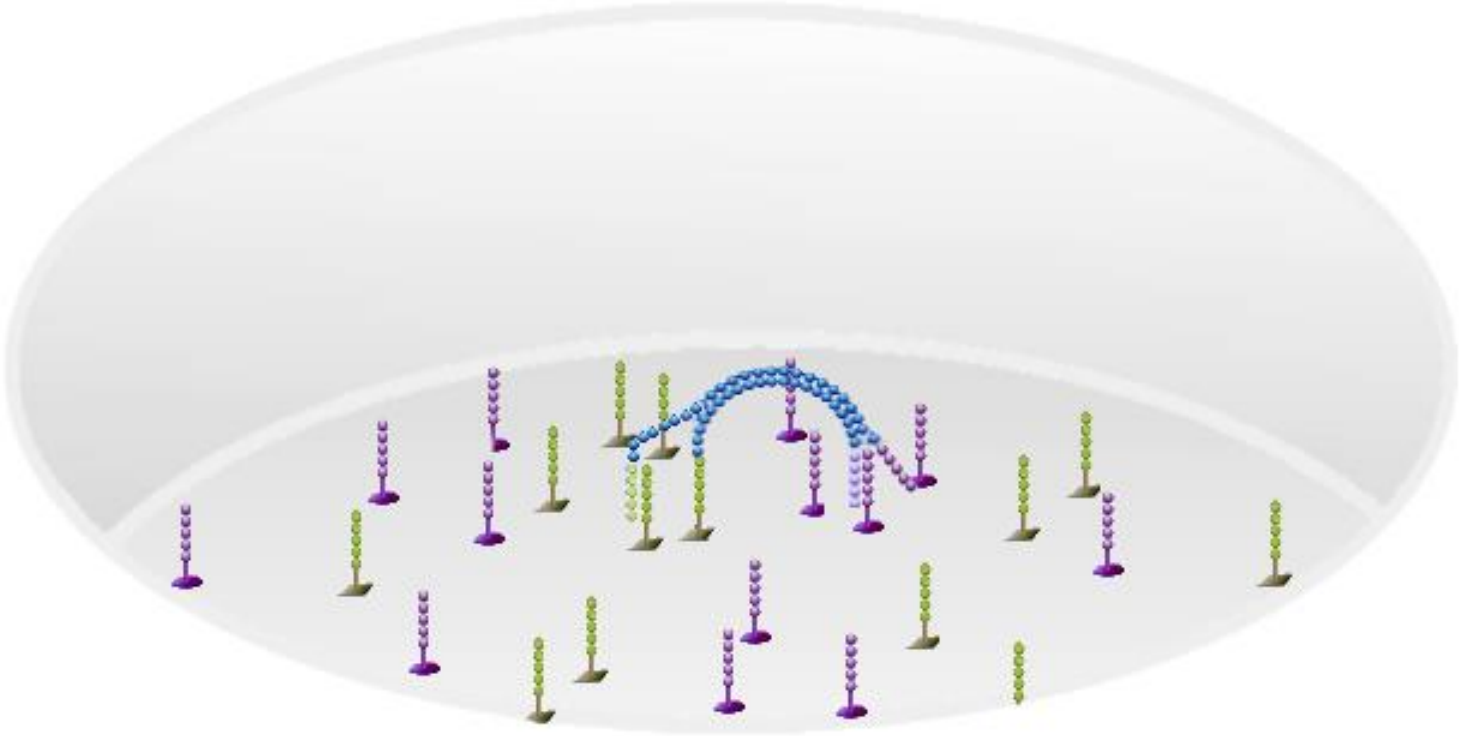
Oligos are not present on surface between wells to control cluster size



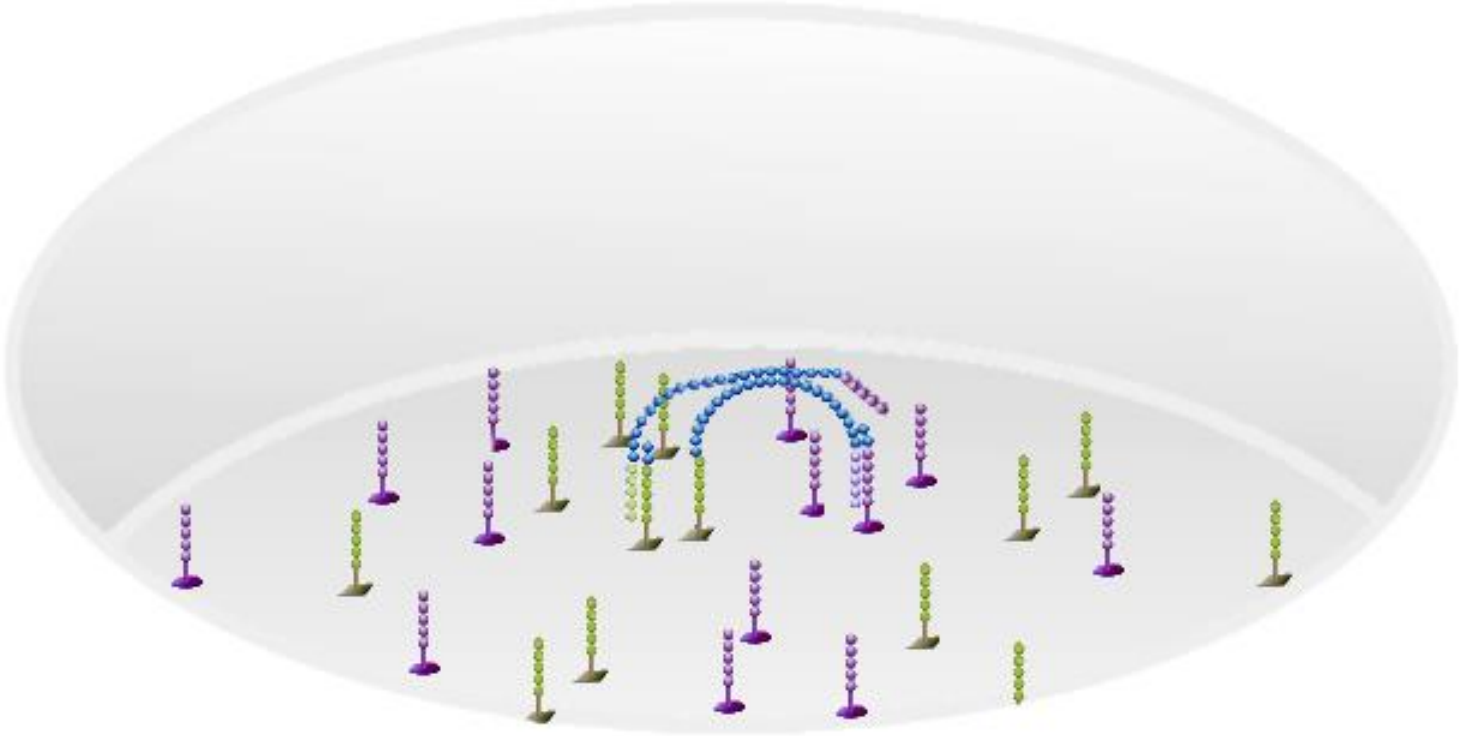
Single stranded DNA is introduced to the flow cell, which hybridizes to an oligo present in the microwell



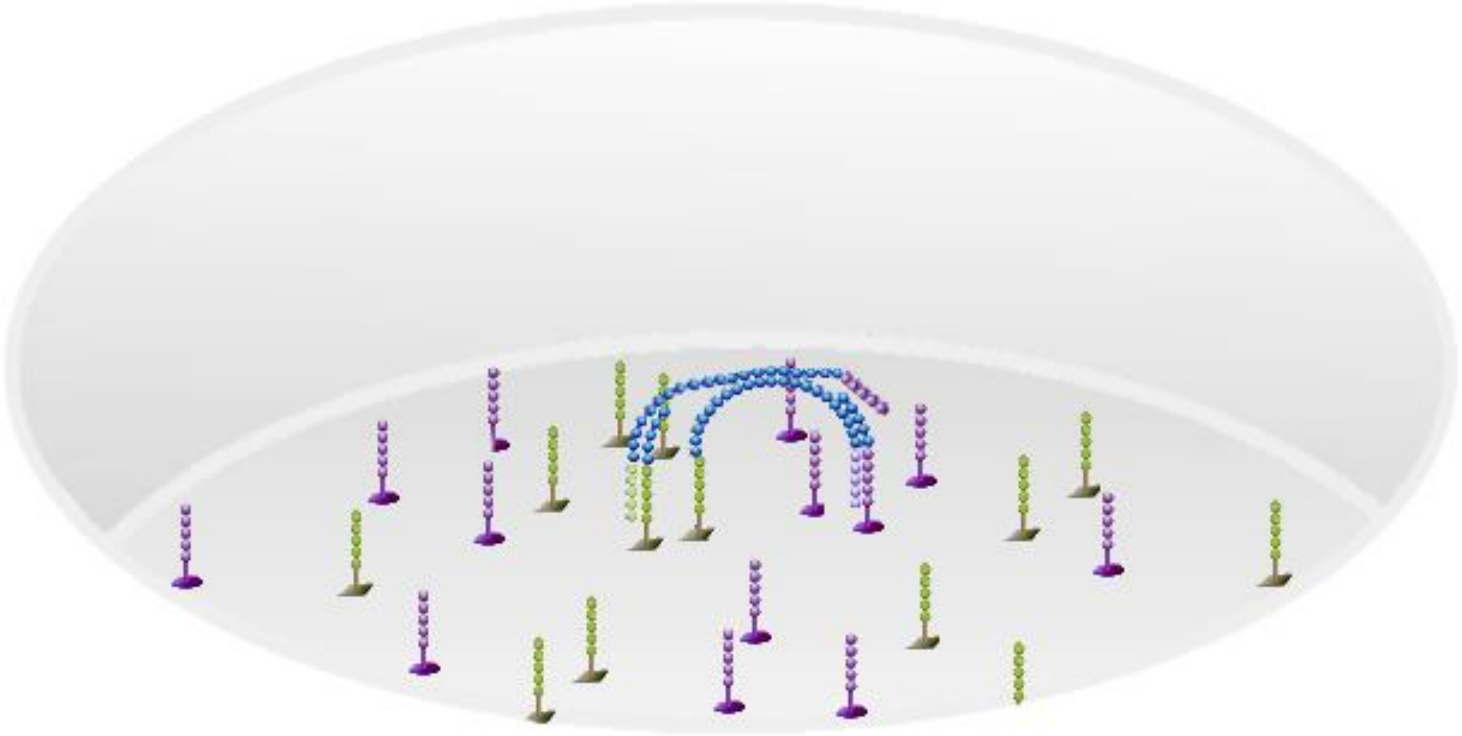
When hybridization occurs amplification immediately begins



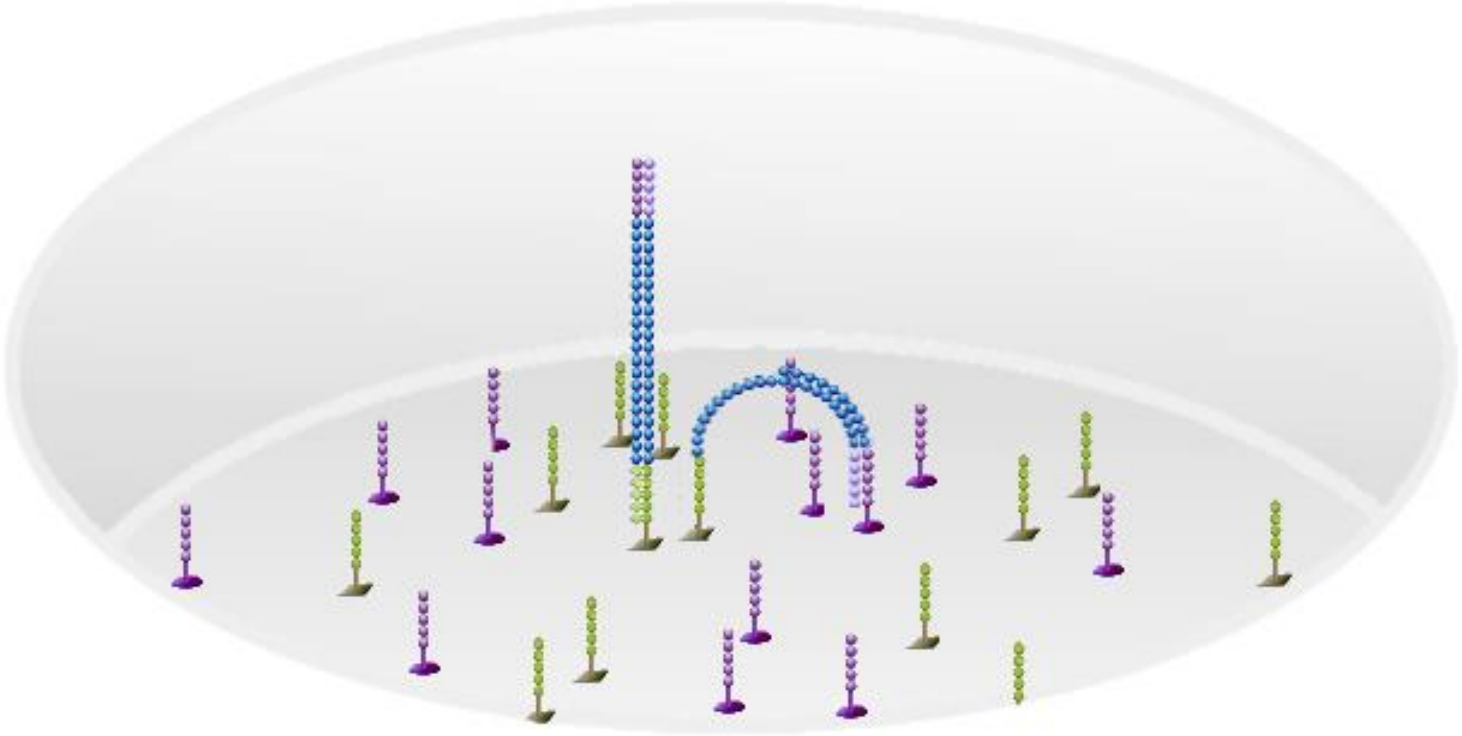
Through a proprietary process, double stranded ends of the DNA are denatured and hybridize to the flow cell



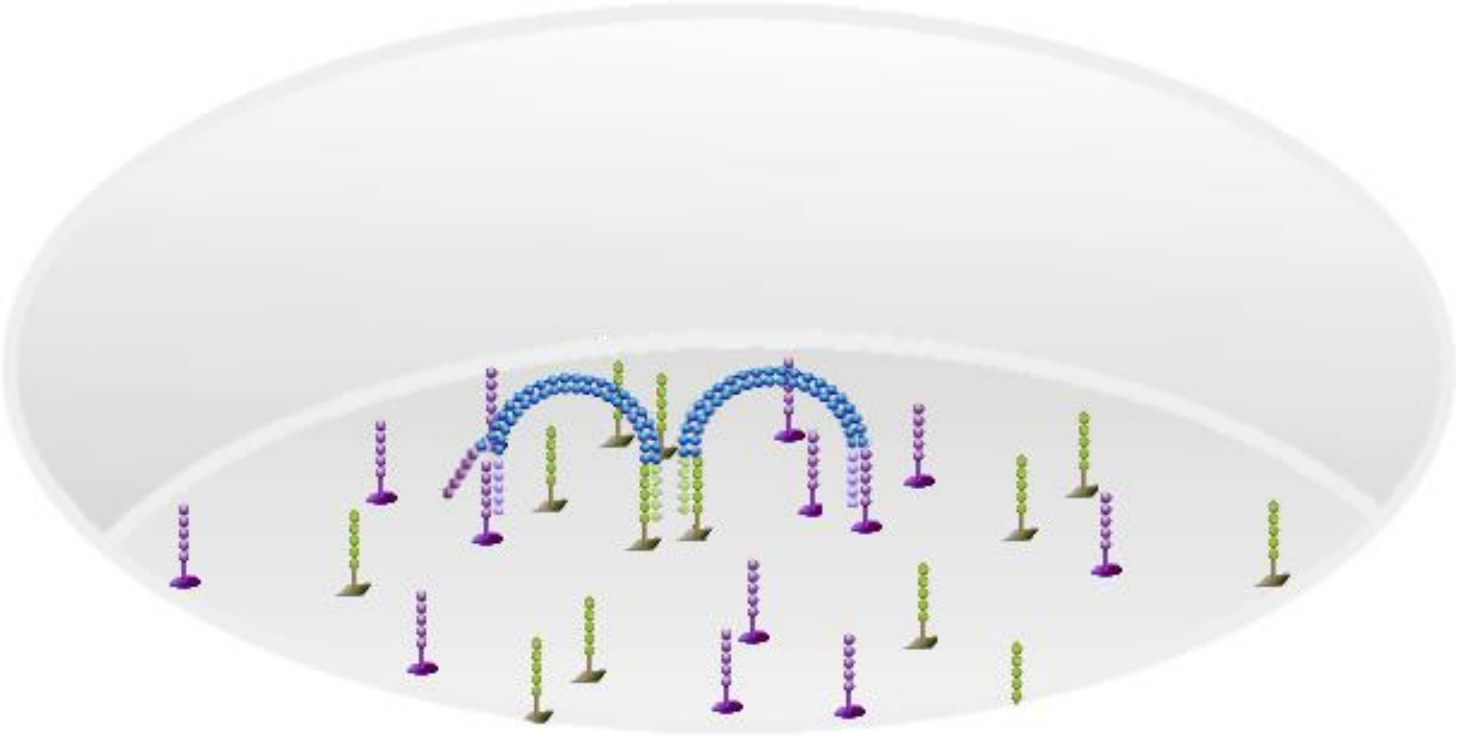
Amplification occurs on both strands simultaneously



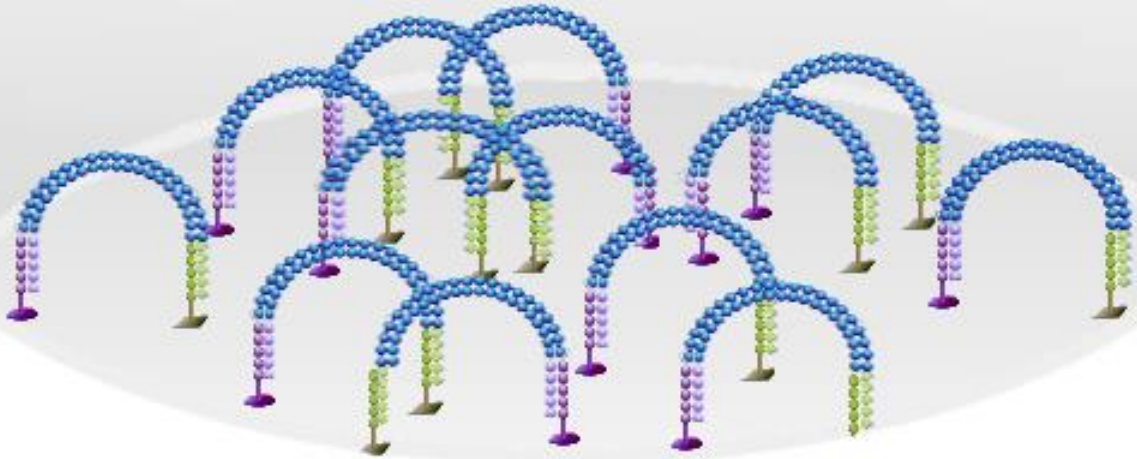
Amplification displaces double stranded DNA



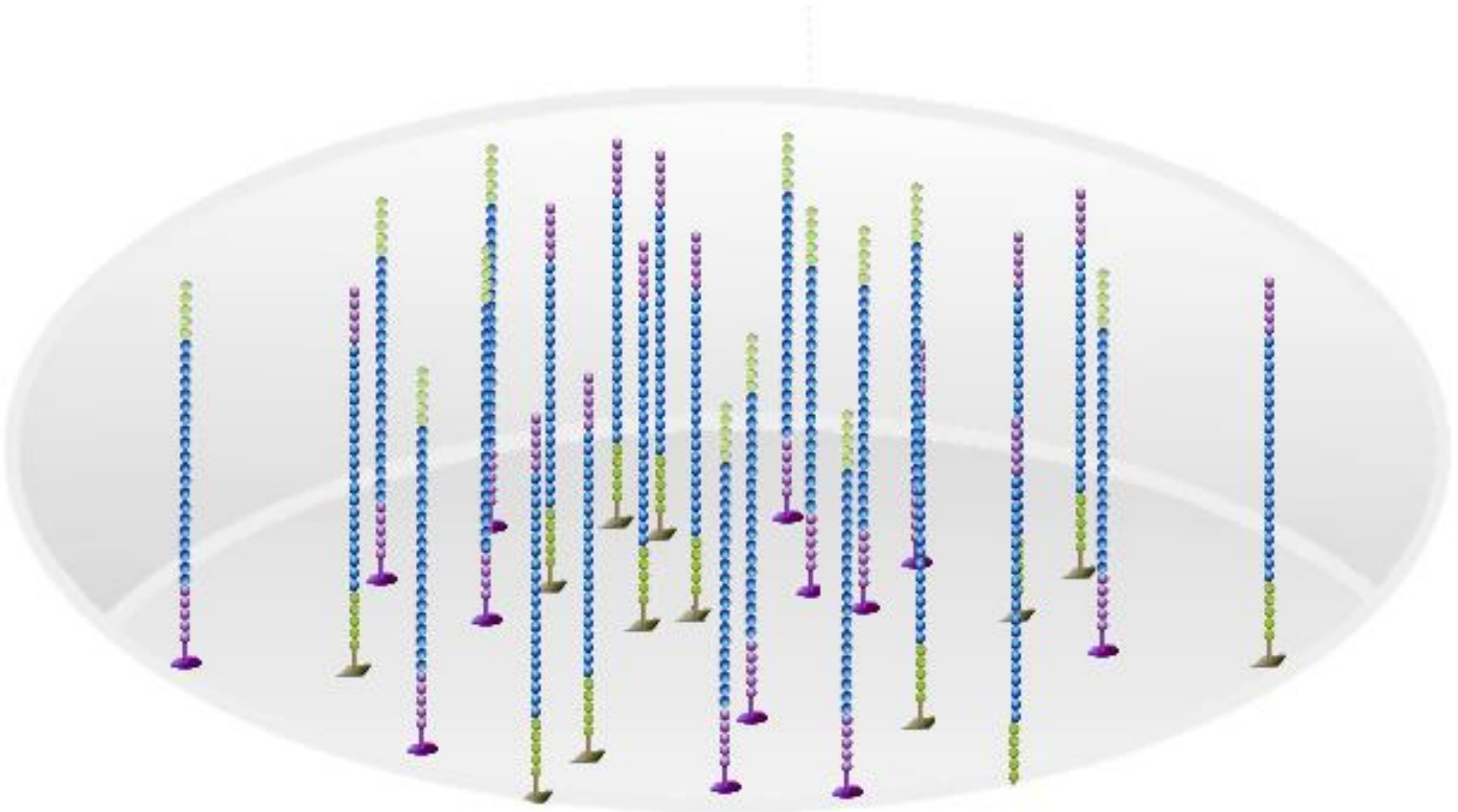
Amplification continues until fully double stranded DNA is formed



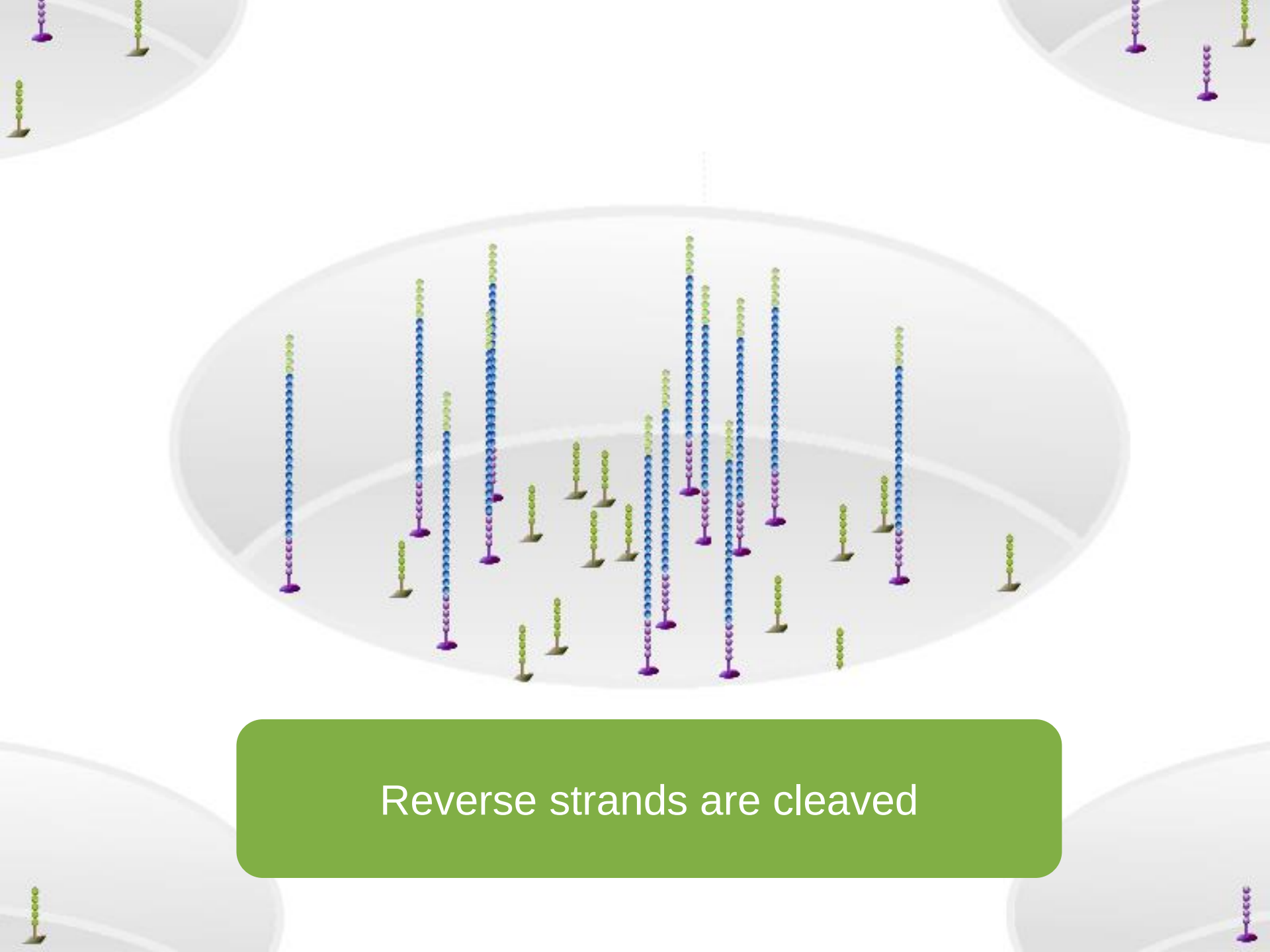
After double stranded DNA is formed, the process repeats itself



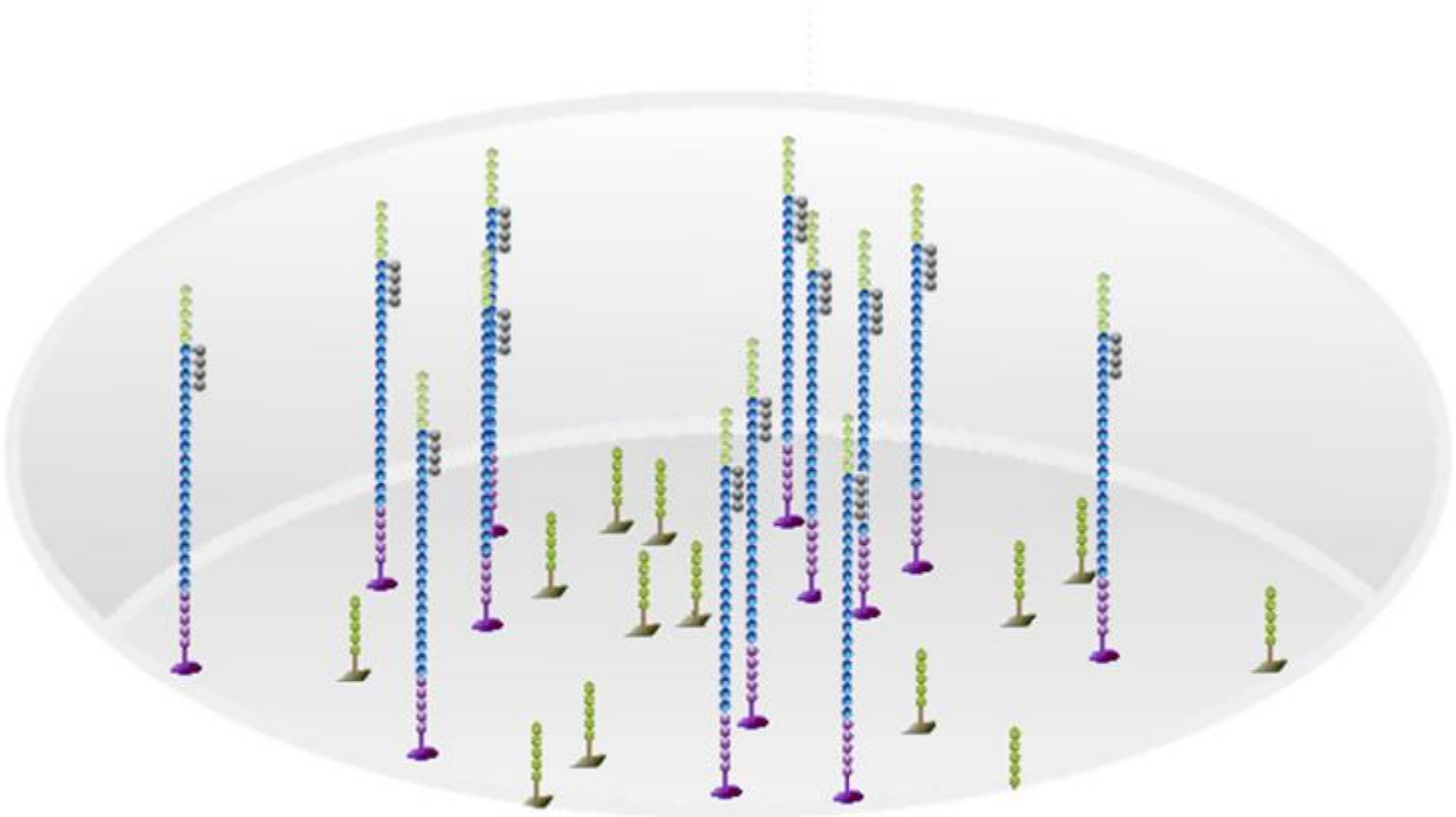
This process continues until no unused oligos are left on the flow cell surface



To prepare the strands for sequencing, linearization is performed



Reverse strands are cleaved



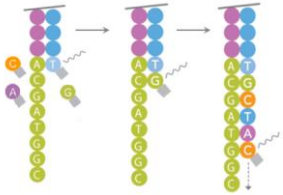
Sequencing primer is hybridized to Read 1 sequencing primer binding site

Illumina Sequencing Workflow

1. Library Preparation

2. Cluster Generation

3. Sequencing



4. Data Analysis

illumina®

Illumina Sequencing Systems

Focused Power

iSeq™ 100

MiniSeq™

MiSeq™

Flexible Power

NextSeq™ 500/550

Production Power

HiSeq™ 2500

HiSeq™ 3000/4000

Population Power

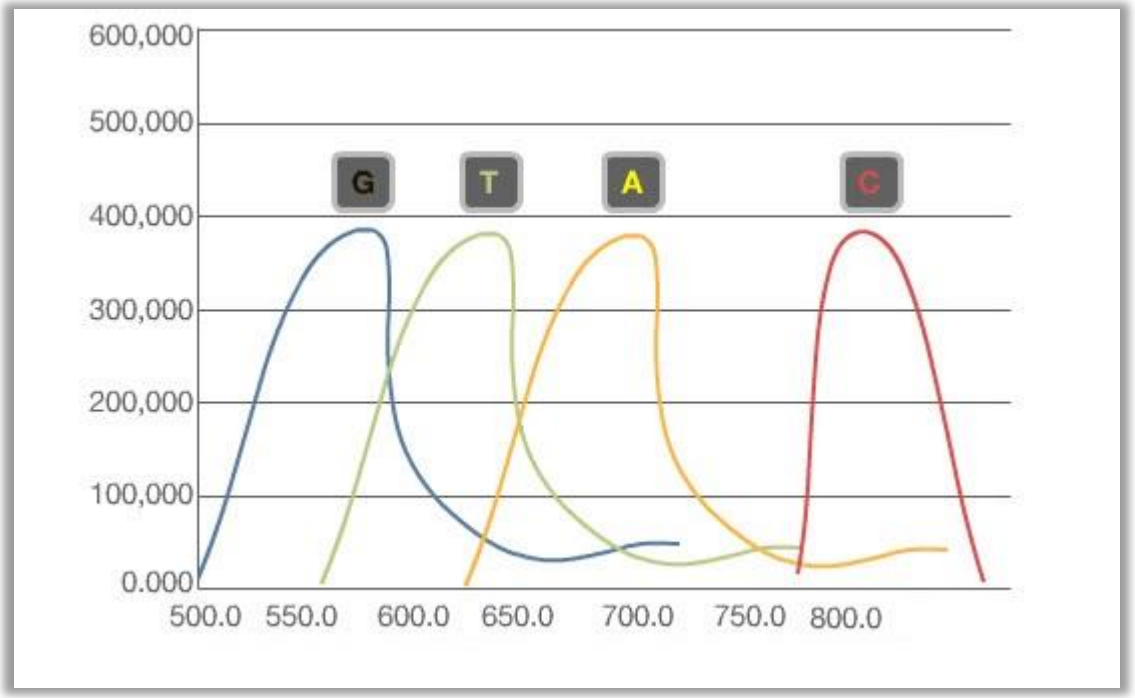
HiSeq™ X Five/Ten

NovaSeq™ 6000

illumina®

4-Channel SBS Chemistry:

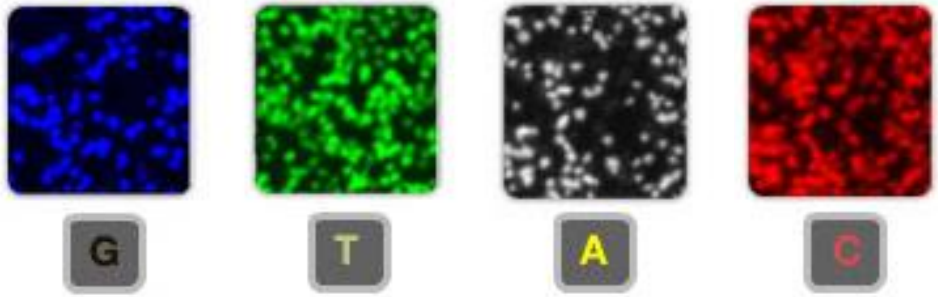
HiSeq, MiSeq



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

Collects four images:

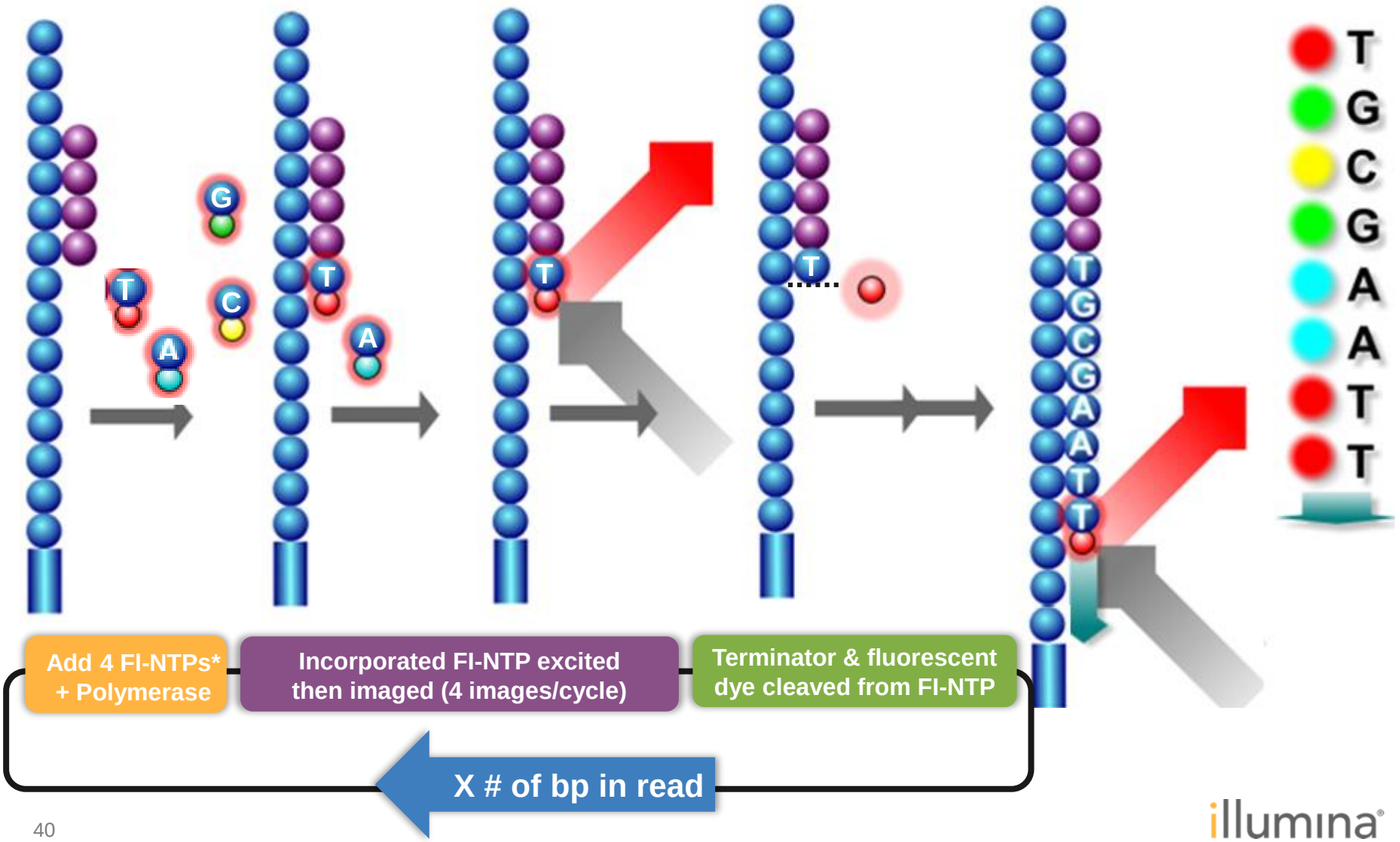
- During each cycle, each cluster appears in only one of four images



illumina®

A Closer Look At 4-Dye Chemistry

4-channel chemistry



illumina®

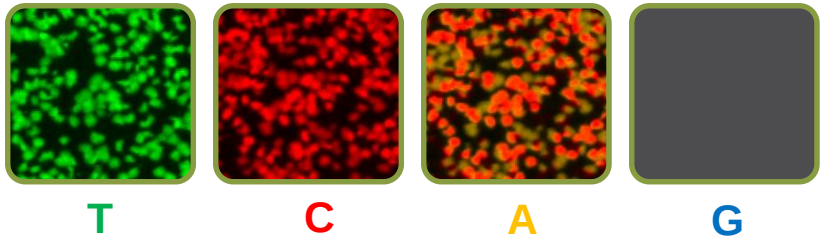
* Fluorescently labelled NTPs

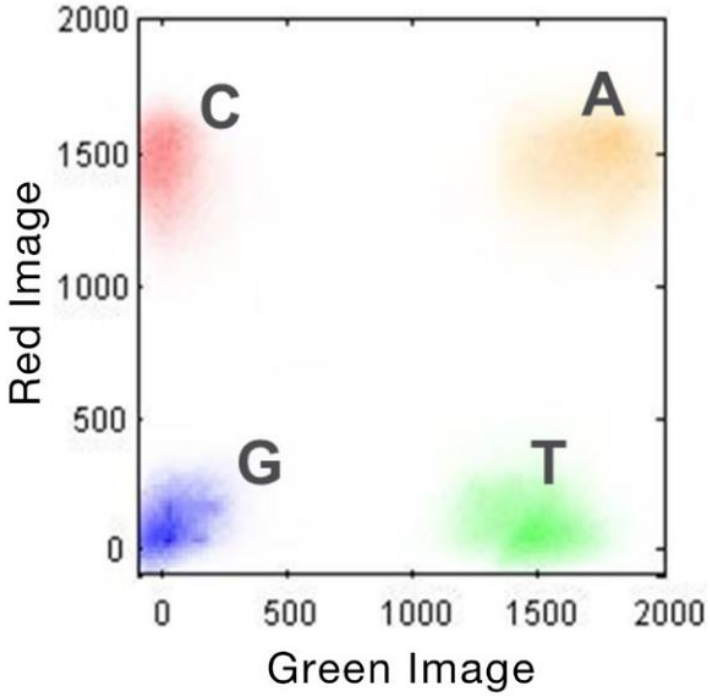
2-Channel SBS Chemistry:

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**

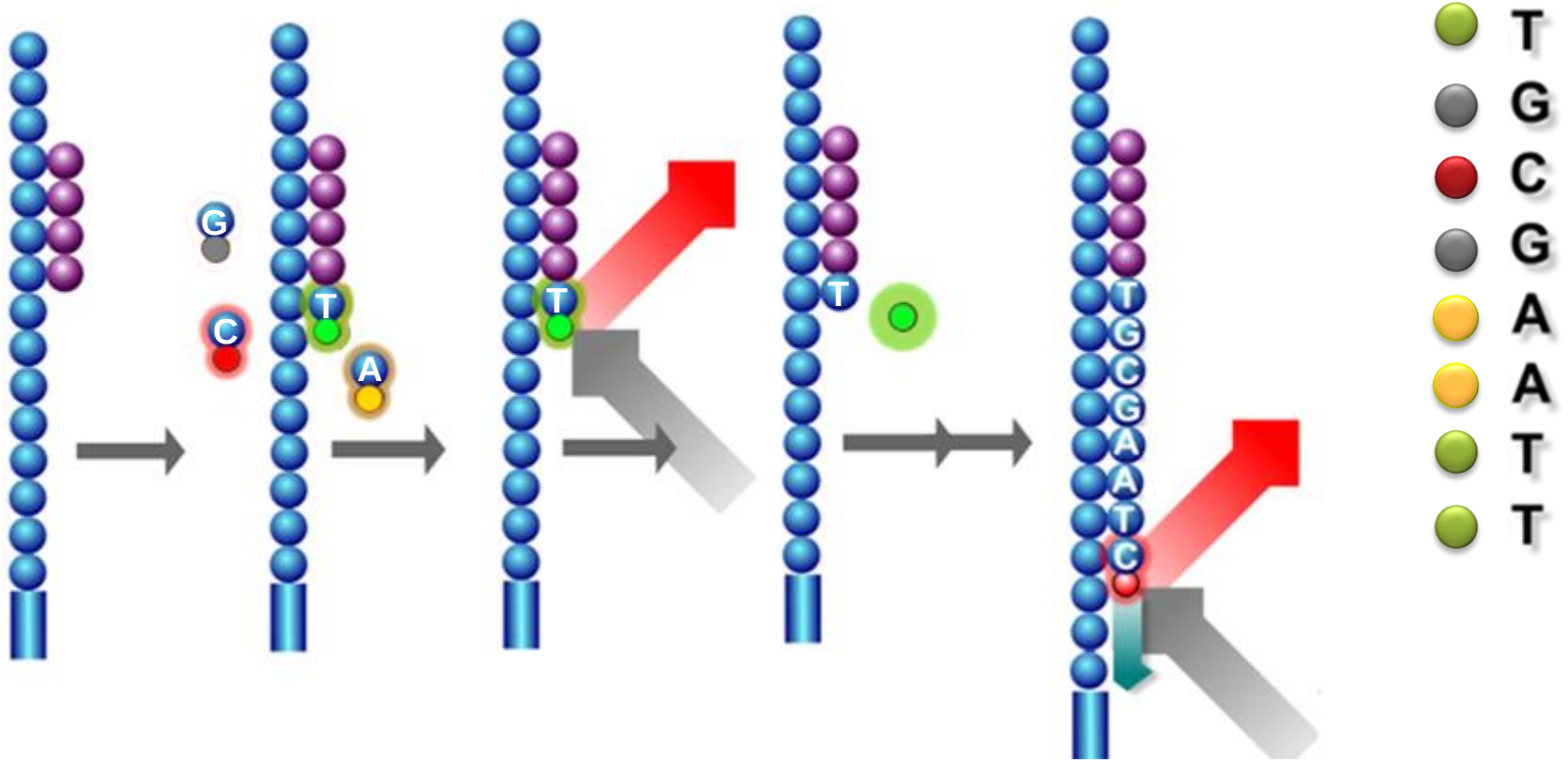




Illumina Two-Channel SBS Sequencing Technology Technote

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®

1-Channel SBS Chemistry:

iSeq 100

SBS chemistry combined with Complementary Metal-Oxide-Semiconductor (CMOS) technology

- The system uses a patterned flow cell with nanowells fabricated over a CMOS chip
- Each sequencing cycle has two chemistry steps in order to determine bases
- Two images are captured within one cycle of sequencing run

Based on the signal pattern across two images, base calls can be determined

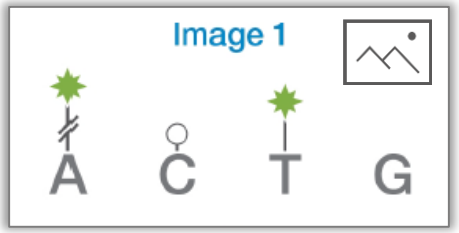
- Intensities extracted from one image and compared to a second image result in four distinct populations, each corresponding to a nucleotide.

A Closer Look At 1-Dye Chemistry

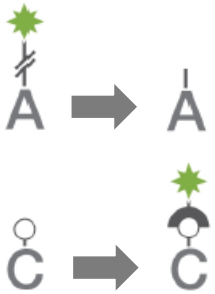
1-channel chemistry

One Sequencing Cycle

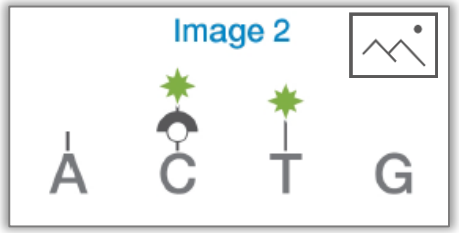
Incorporation



Imaging



Chemistry



Imaging

Incorporation → **First Image captured** → **Second Chemistry Step** → **Second Image captured**

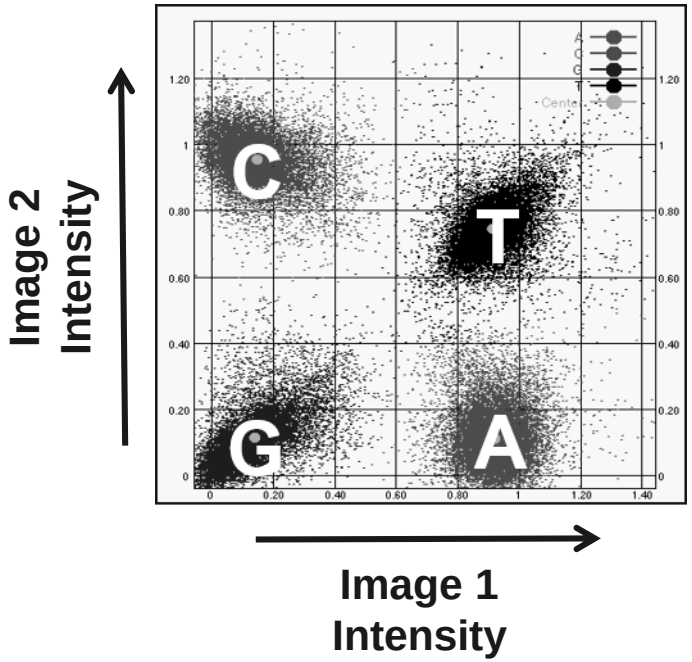
X # of bp in read

illumina®

Sequencing by synthesis with CMOS detection

1-channel chemistry

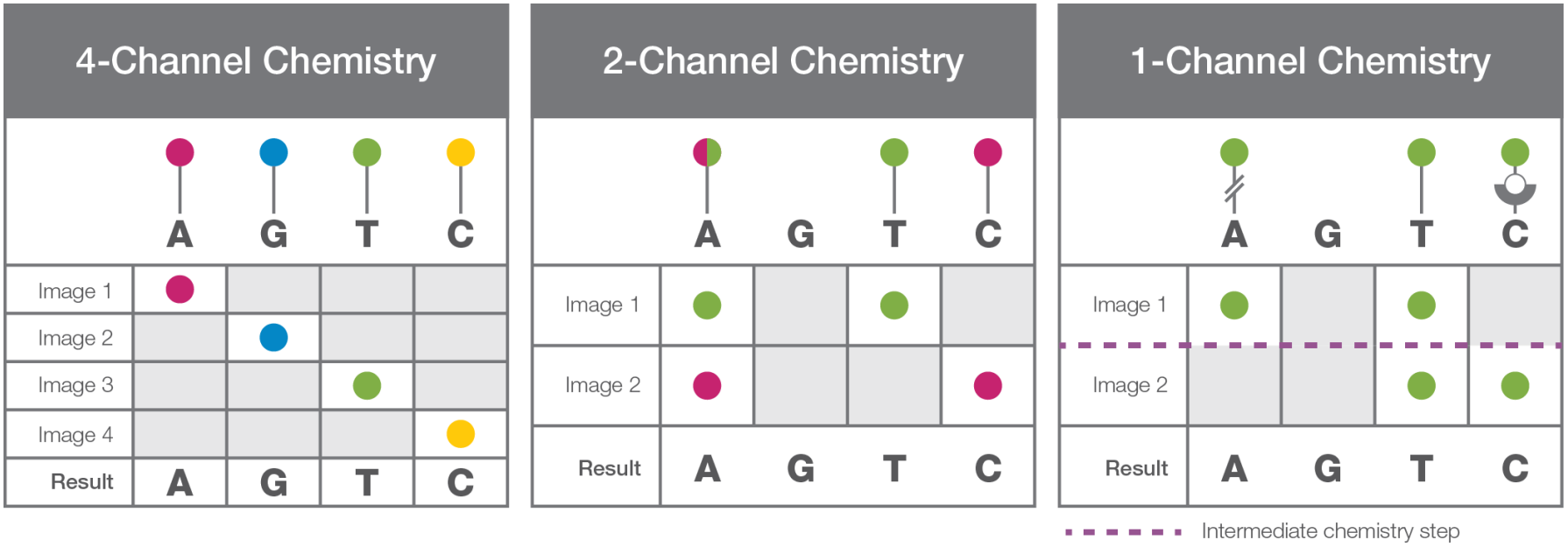
Base	Image 1	Image 2
T	ON	ON
A	ON	OFF
C	OFF	ON
G	OFF	OFF



- **The iSeq 100 System uses 1-dye sequencing, which requires one dye and two images to encode data for the four bases**

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

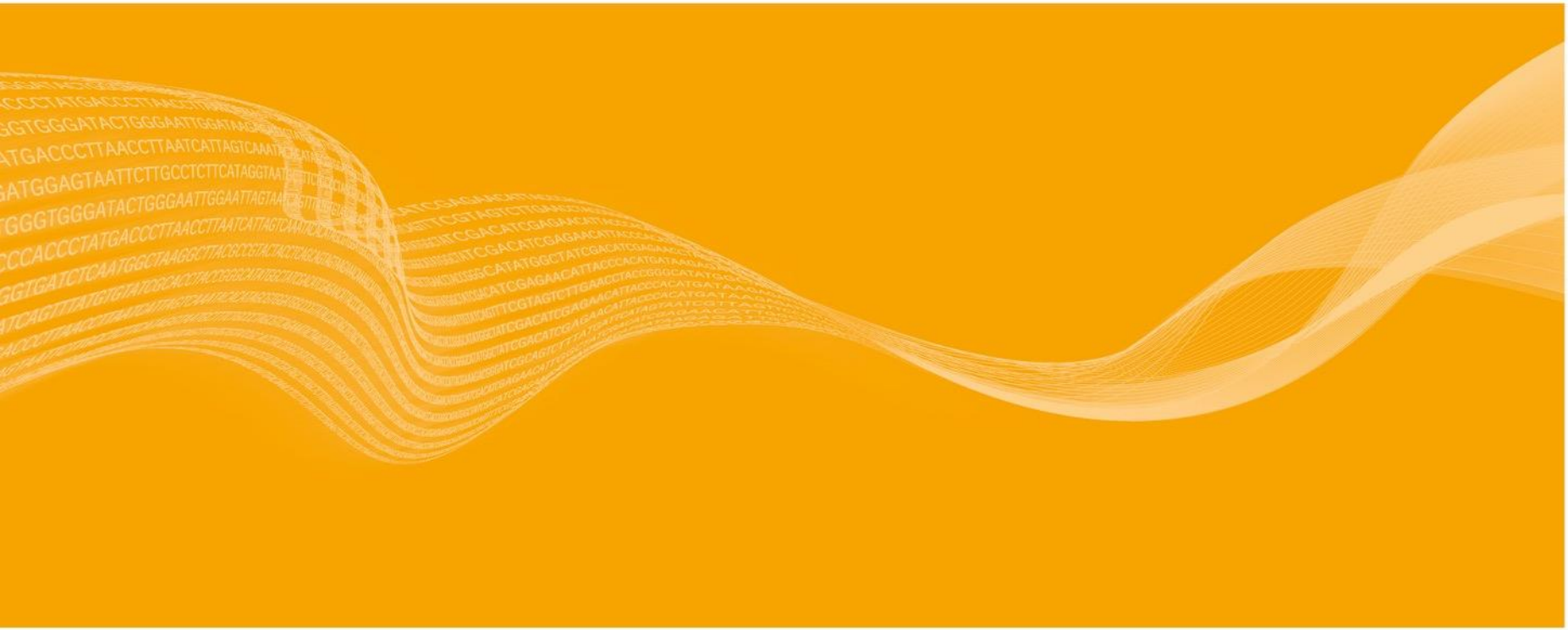
- Simplified nucleotide detection by using two fluorescent dyes and two images to determine all four base calls

1-channel SBS

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

illumina®

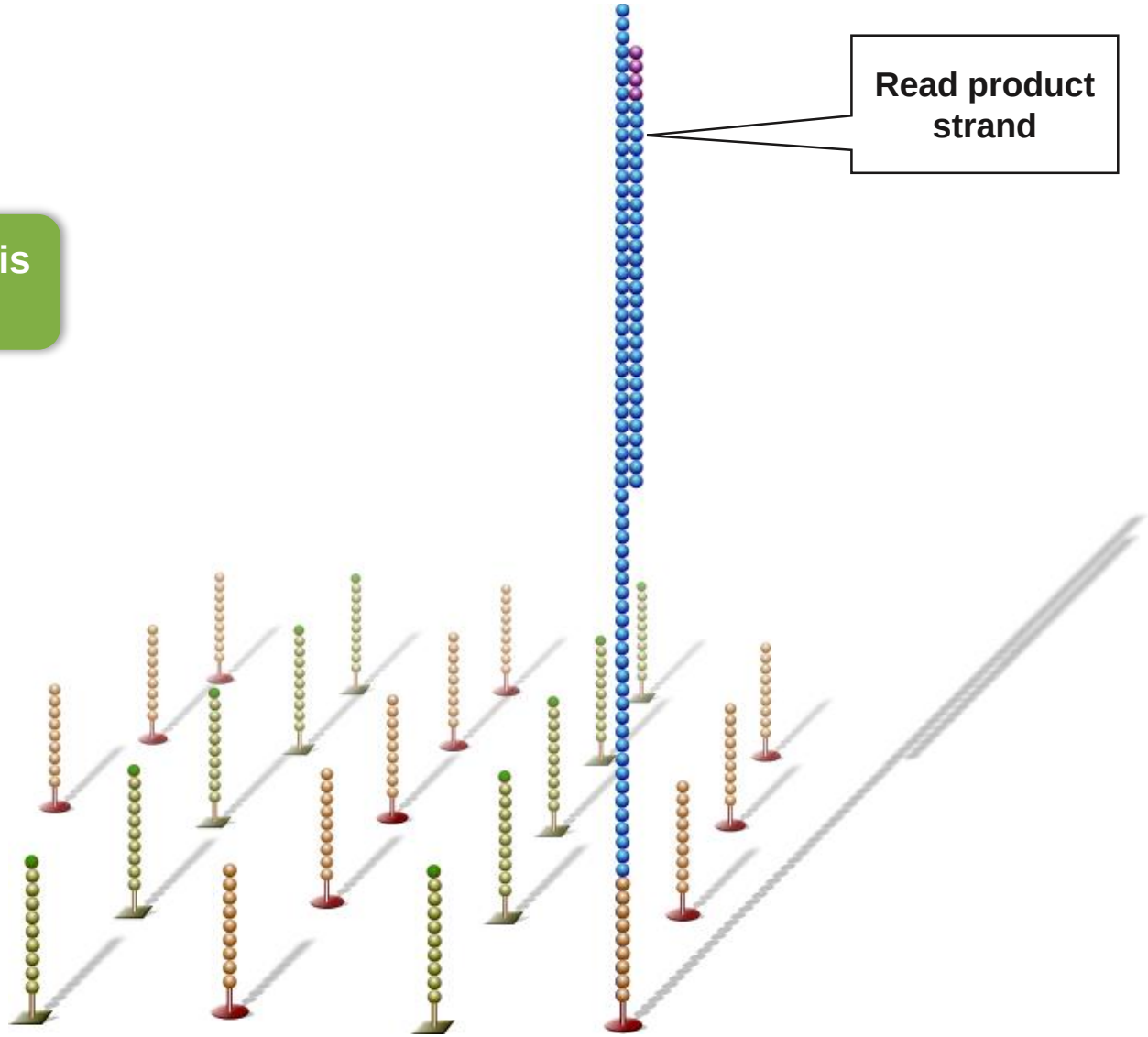
Paired-End Sequencing



illumina®

Paired-End Sequencing

The read product is stripped off

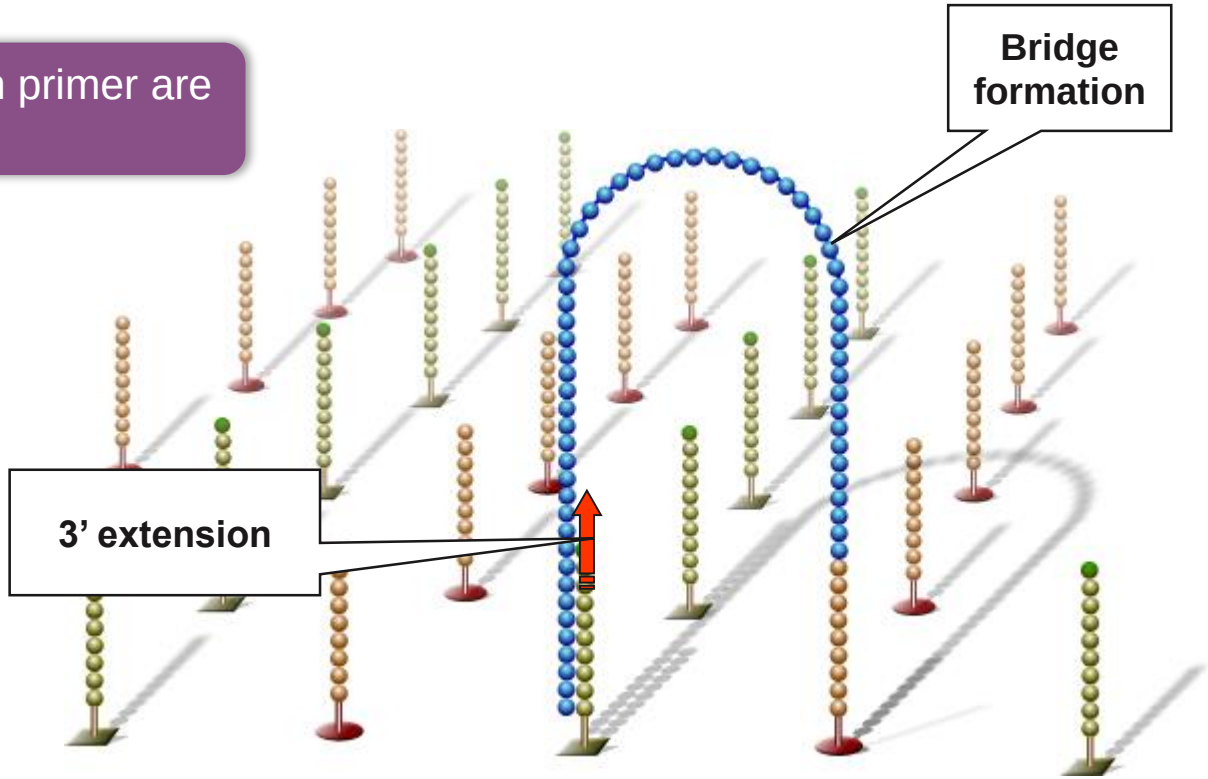


illumina®

Paired-End Sequencing

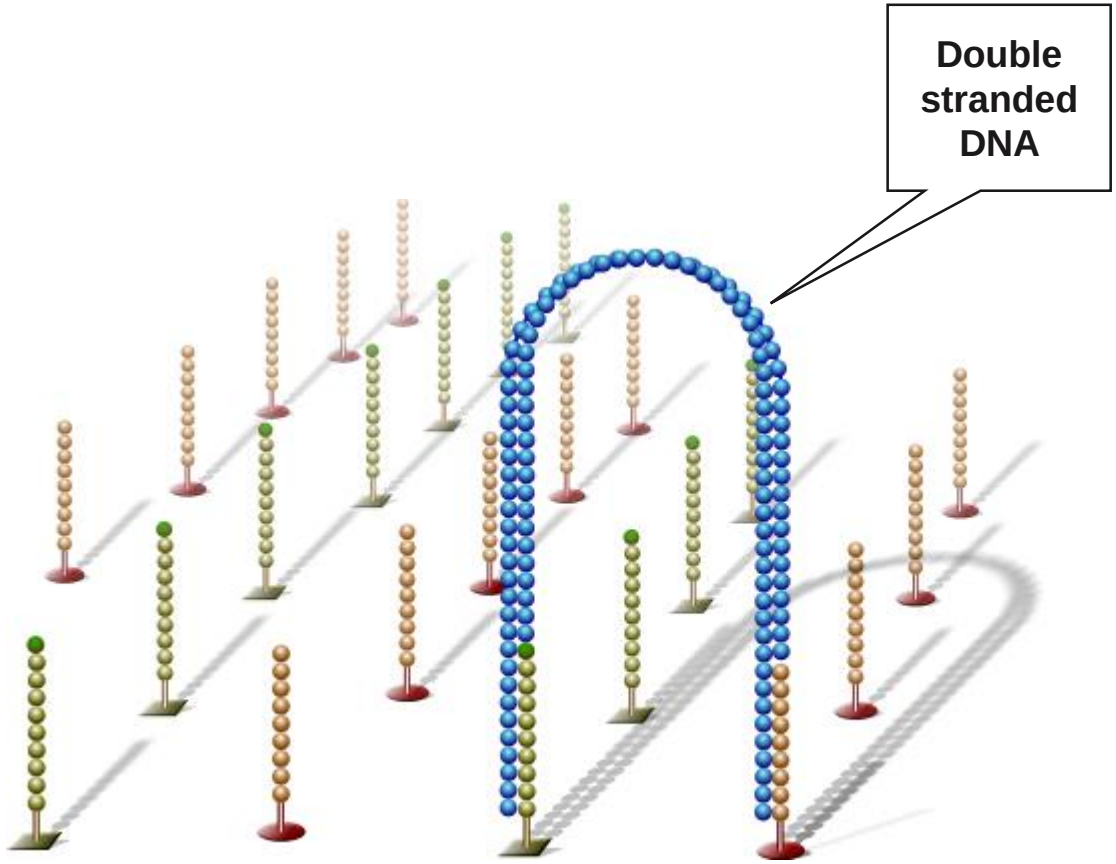
Single-stranded template loops over to form a bridge by hybridizing with a lawn primer

3’-ends of lawn primer are extended



illumina®

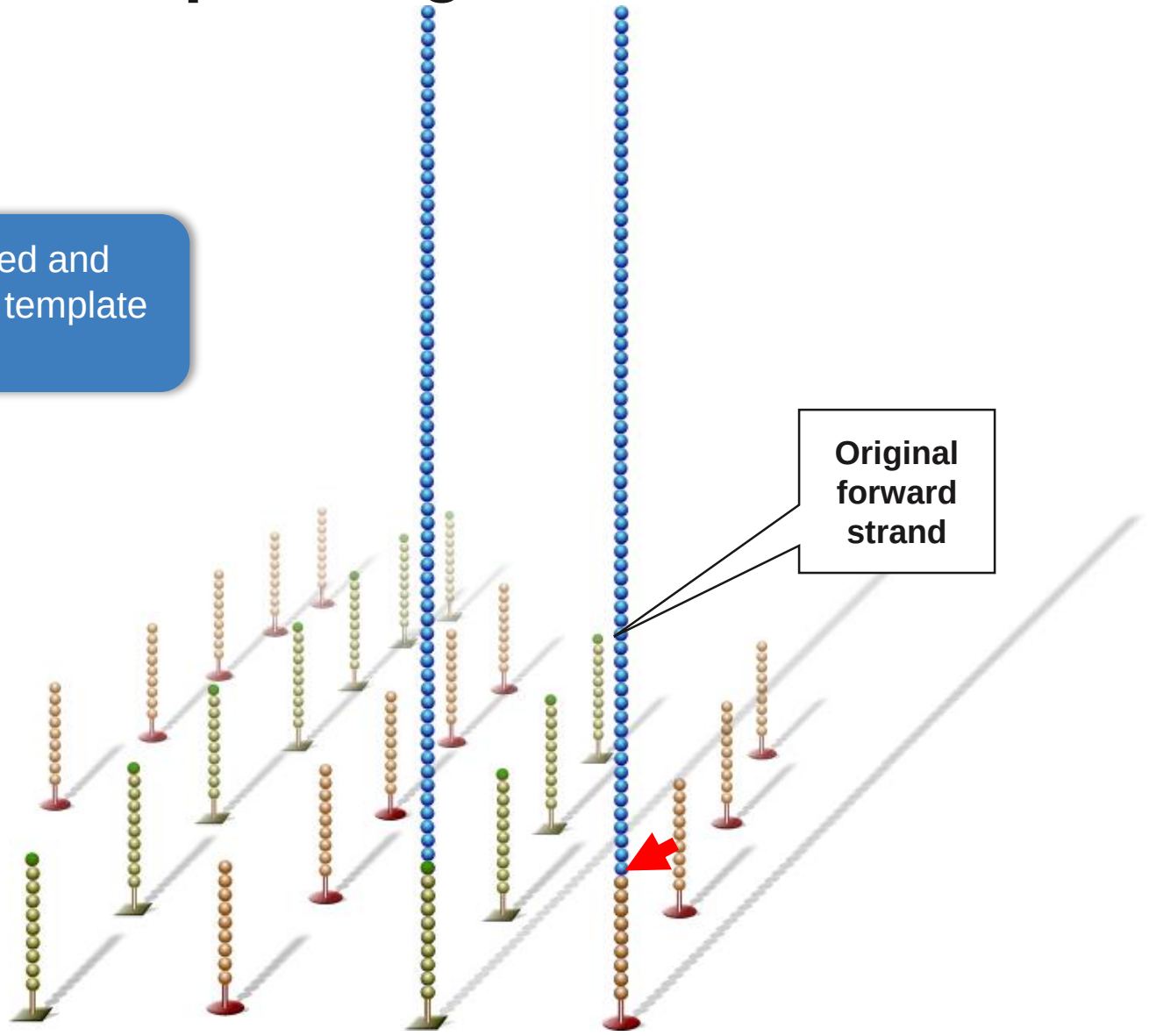
Paired-End Sequencing



illumina®

Paired-End Sequencing

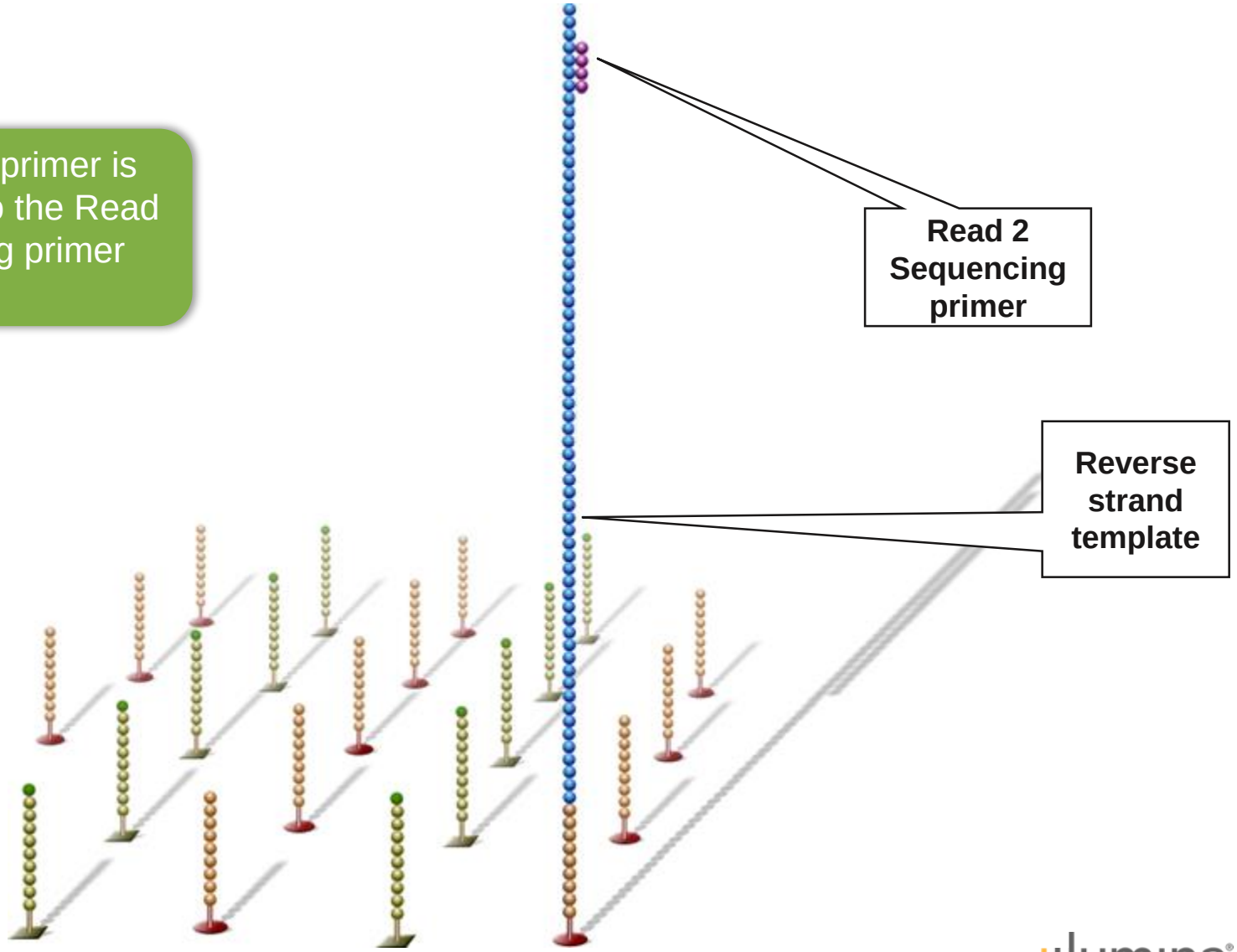
Bridges are linearized and the original forward template is cleaved



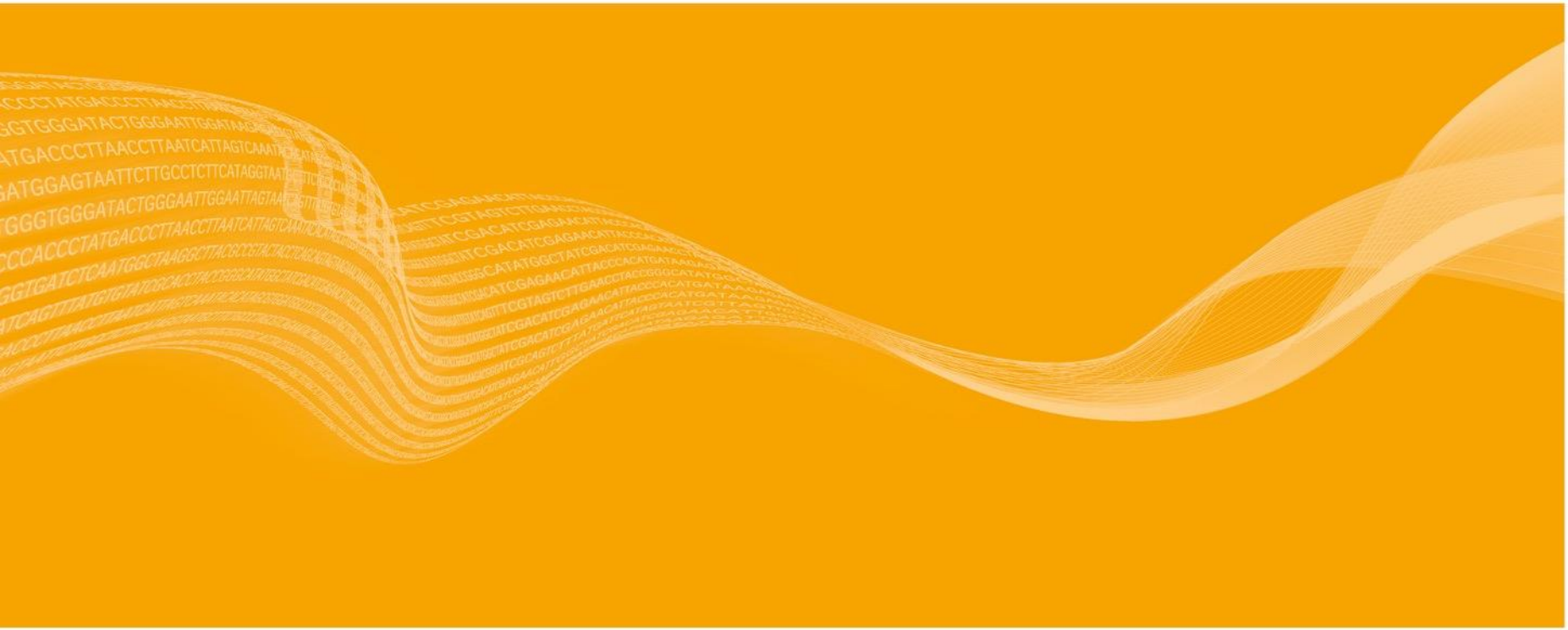
illumina®

Paired-End Sequencing

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



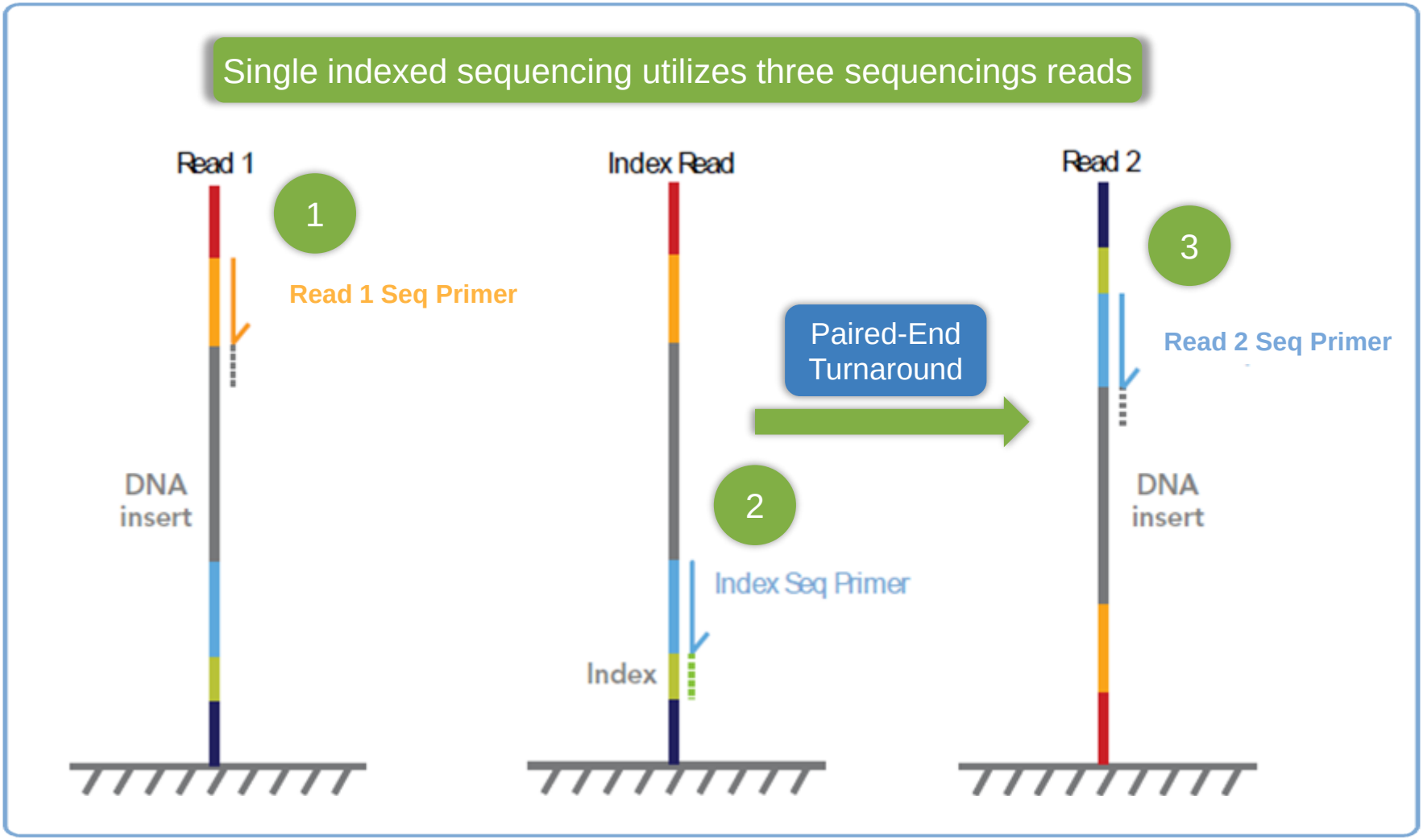
Sequencing with Index Reads



illumina®

Single Index Reads

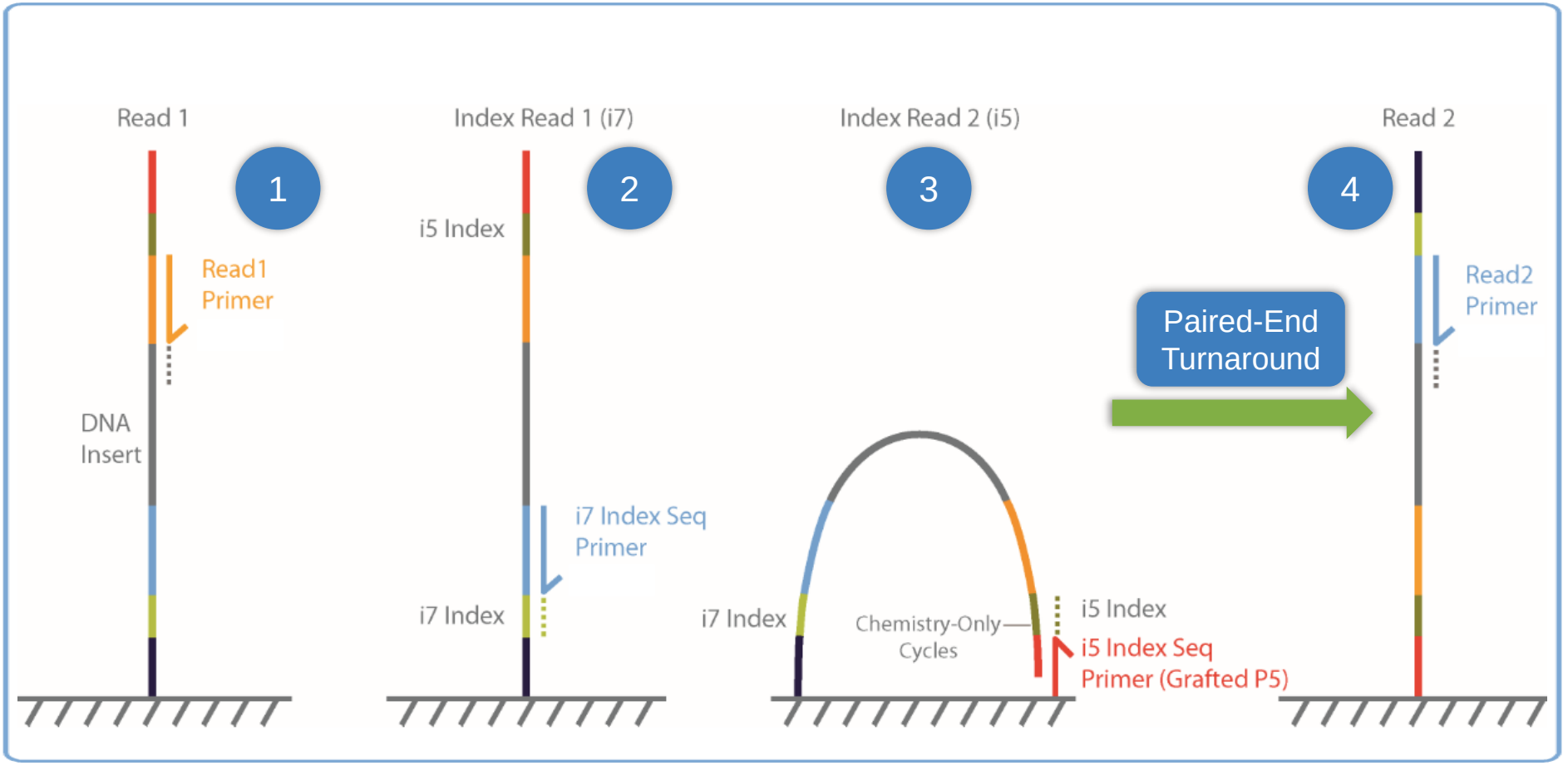
All Platforms



illumina®

Dual Index Reads

MiSeq, HiSeq 2500, NovaSeq 6000

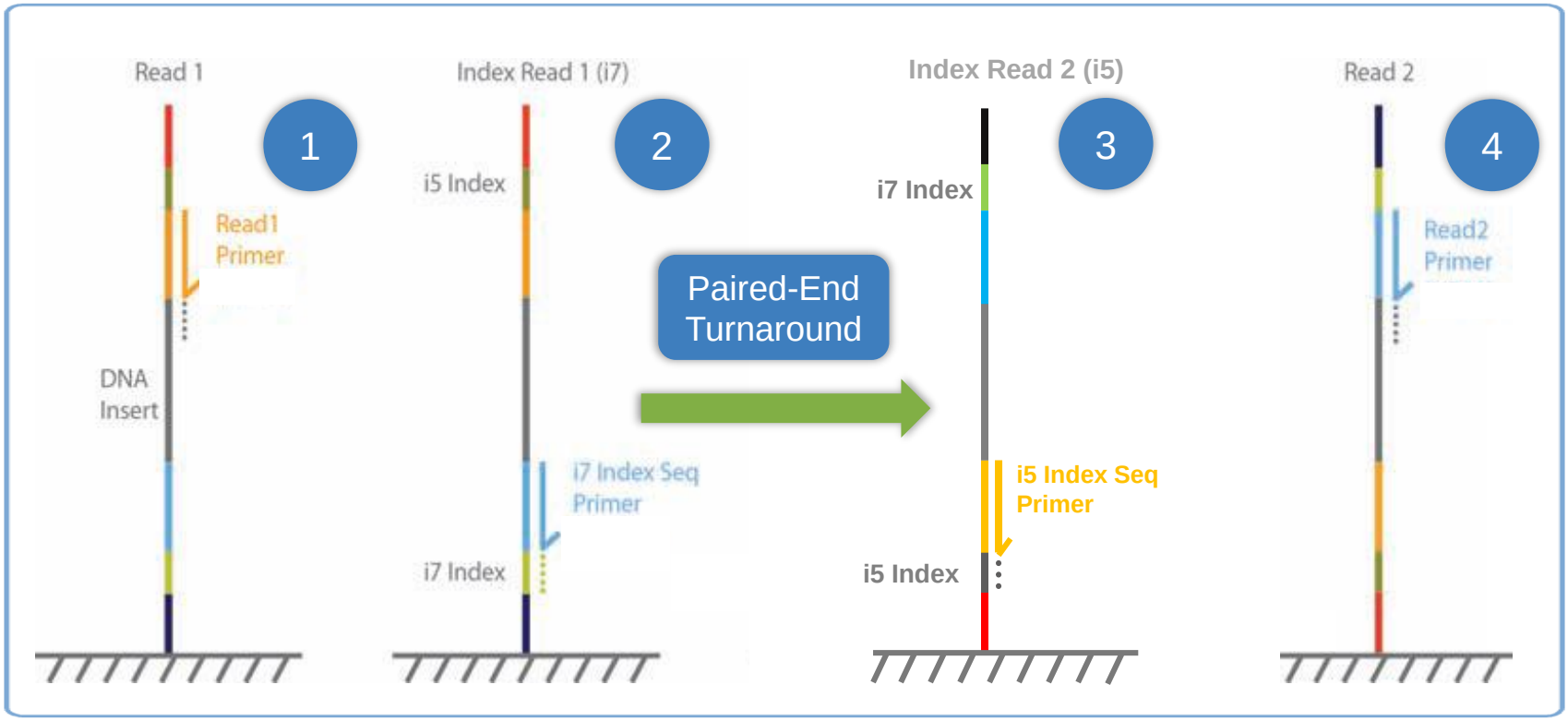


Dual indexed sequencing utilizes four sequencing reads

illumina®

Dual Index Reads

iSeq 100, MiniSeq, NextSeq, HiSeq 3000/4000



illumina®

Illumina Sequencing Workflow

1. Library Preparation

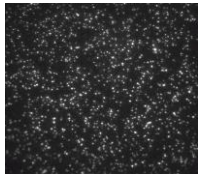
2. Cluster Generation

3. Sequencing

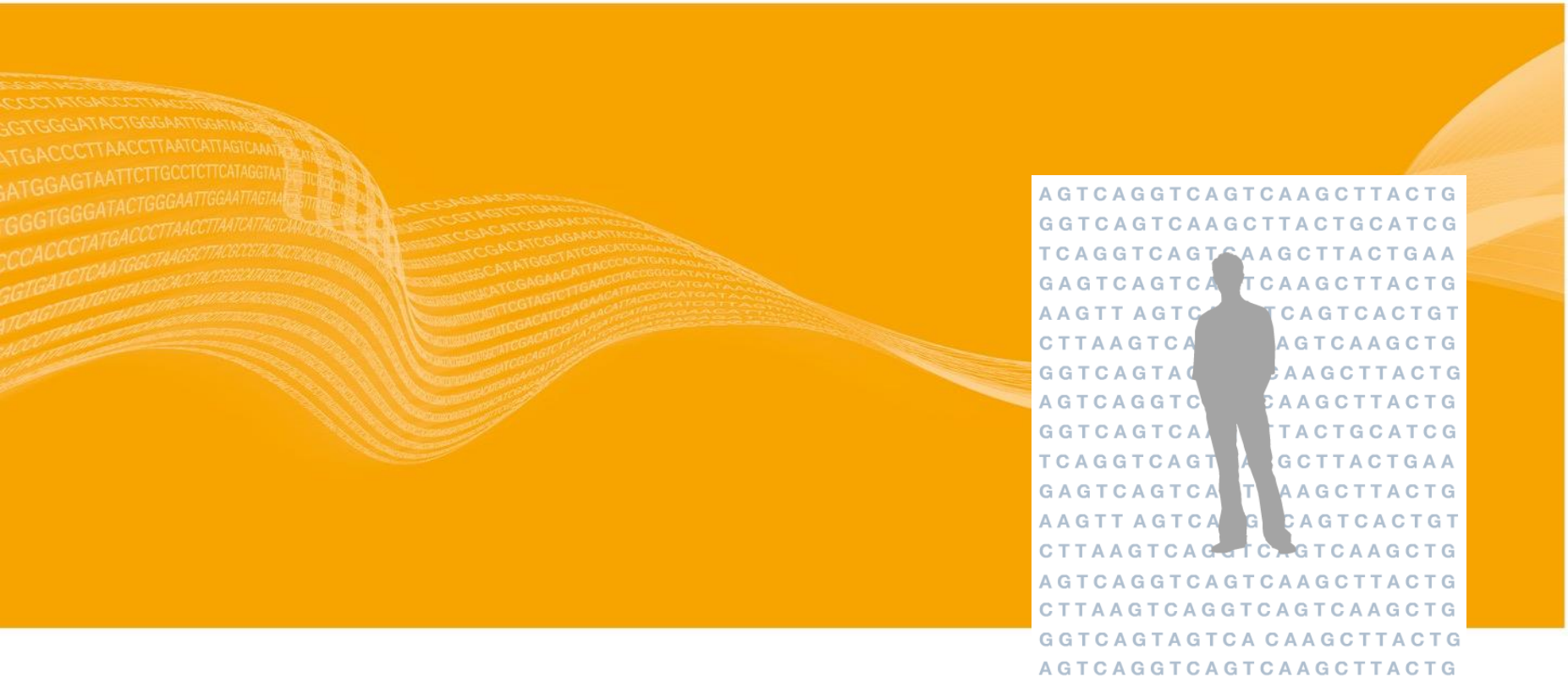
4. Data Analysis

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?



illumina®