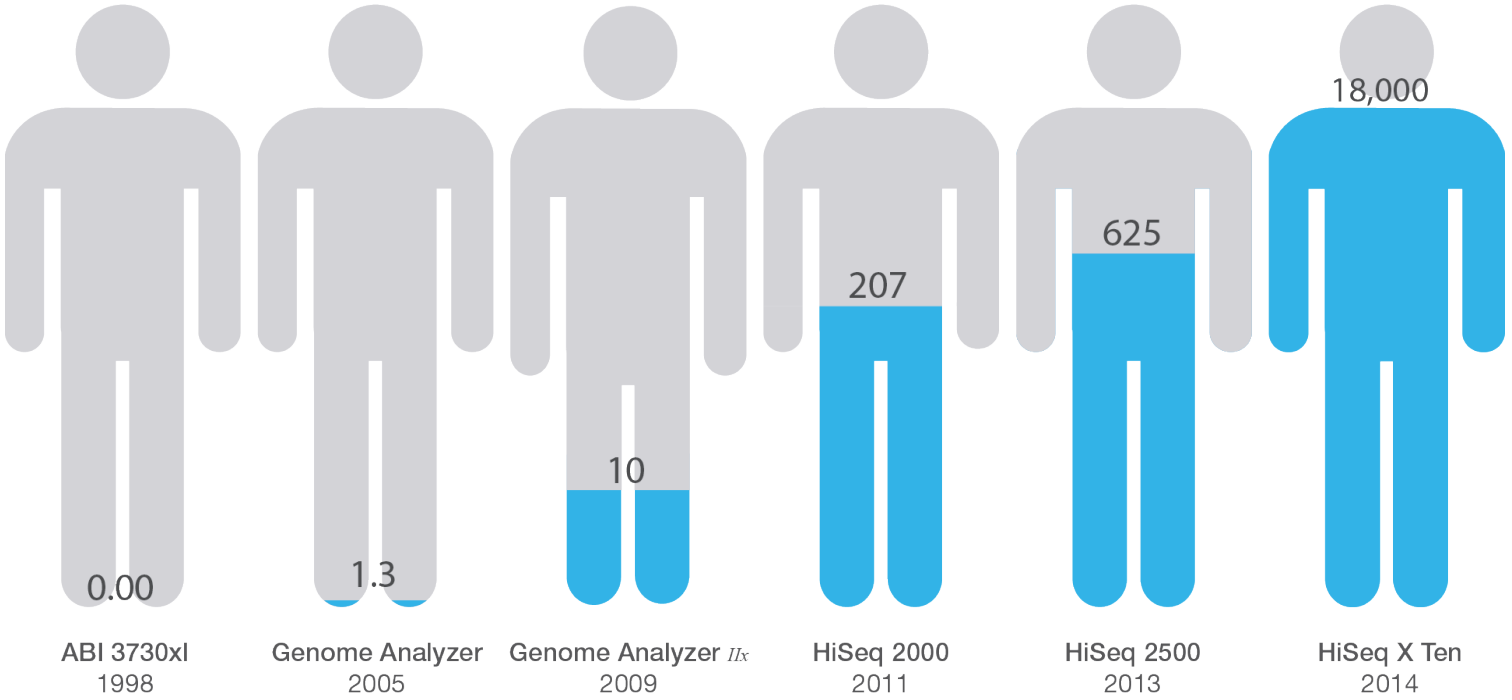


Human Genome Sequencing Over the Decades—The capacity to sequence all 3.2 billion bases of the human genome (at 30X coverage) has increased exponentially since the 1990s. In 2005, with the introduction of the Illumina Genome Analyzer System, 1.3 human genomes could be sequenced annually. Nearly 10 years later, with the Illumina HiSeq X Ten fleet of sequencing systems, the number has climbed to 18,000 human genomes a year.

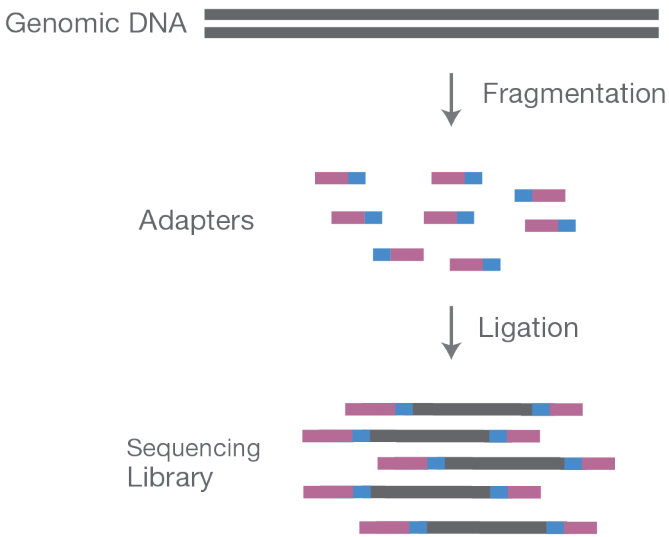
Sample preparation

Cluster Generation

Sequencing

Data Analysis

A. Library Preparation



NGS library is prepared by fragmenting a gDNA sample and ligating specialized adapters to both fragment ends.

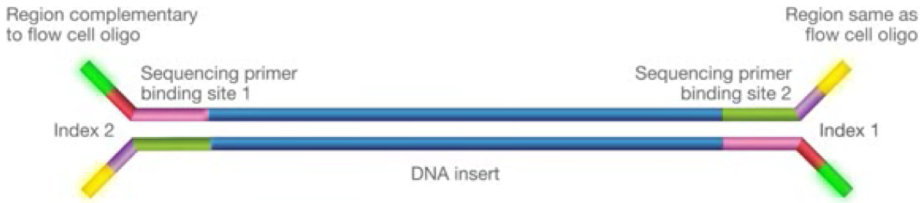
Extraction and purification of DNA by Nextera kit

Tagmentation

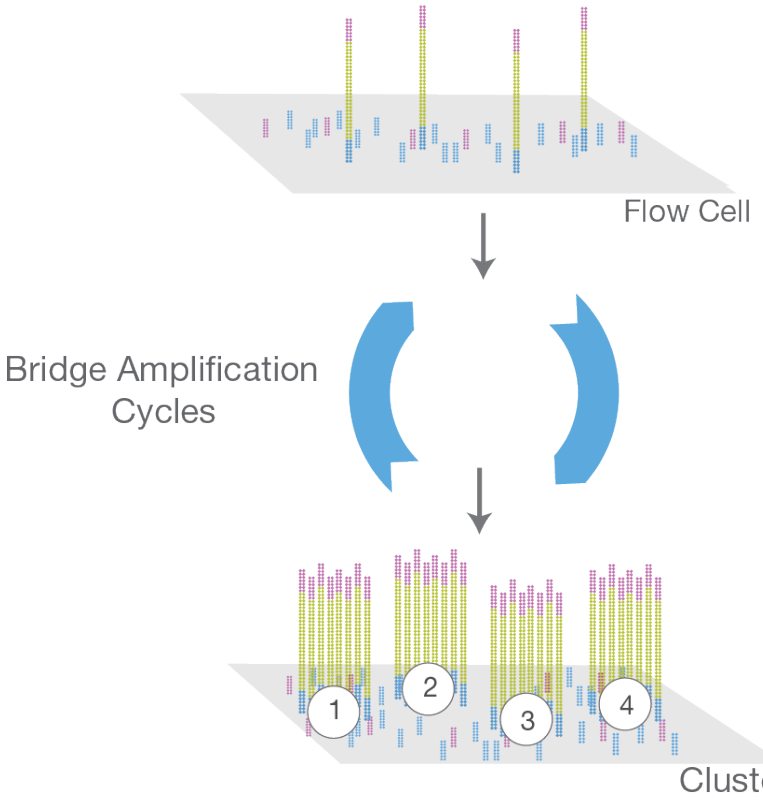
Transposome simultaneously fragment and tag the input DNA with the adapters

Reduced cycle amplification

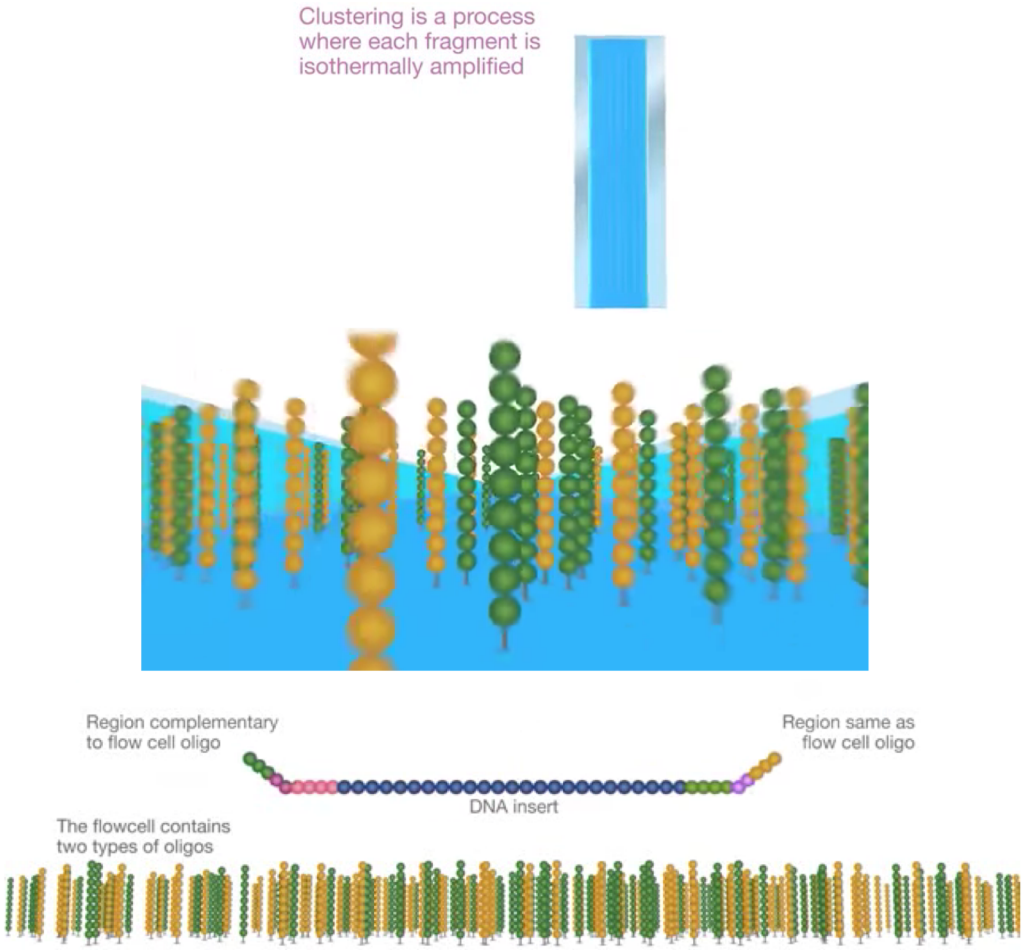
to add additional motifs:



B. Cluster Amplification

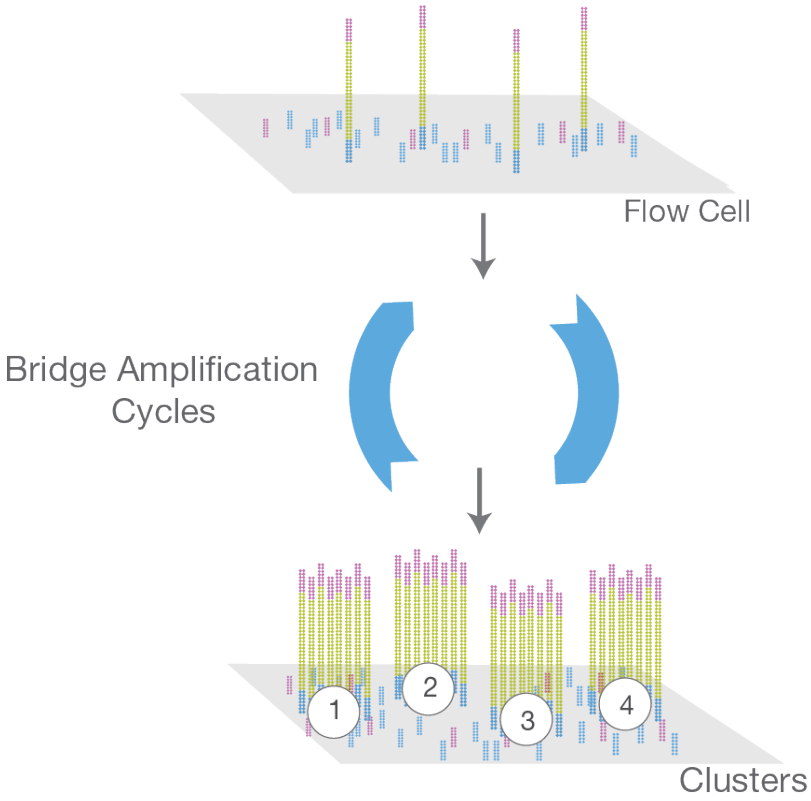


Library is loaded into a flow cell and the fragments are hybridized to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification.



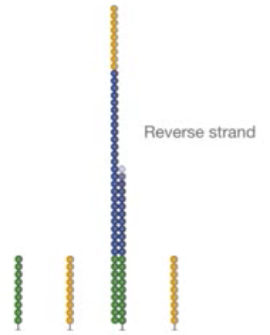
Flow cell is a glass slide divided in line. Each lane contains two types of oligo complementary to the last part of the oligo sequence in the ends of the DNA fragment

B. Cluster Amplification

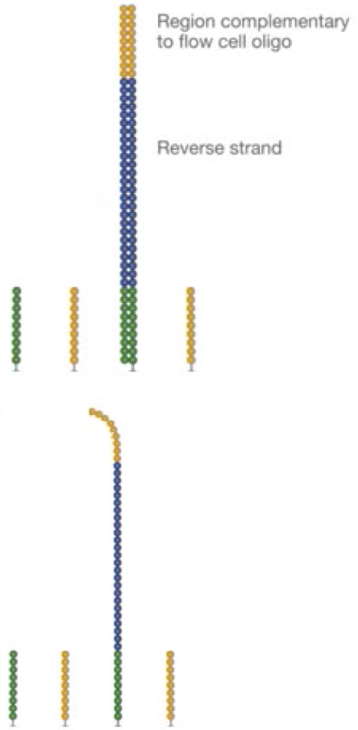


Library is loaded into a flow cell and the fragments are hybridized to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification.

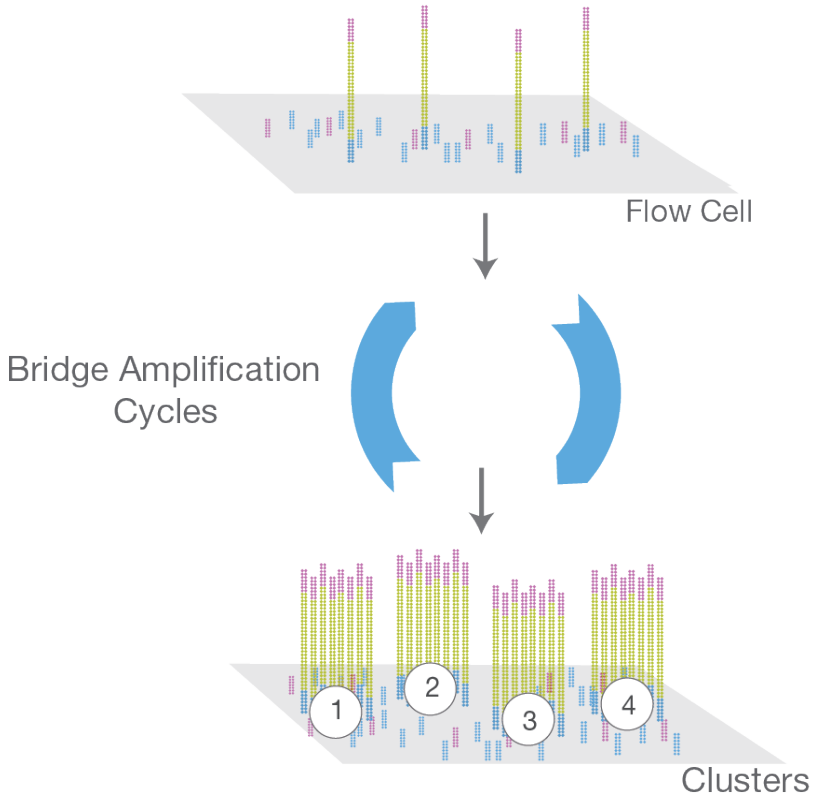
A polymerase creates a complementary fragments



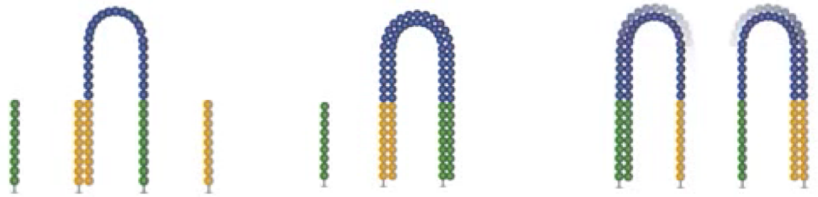
The molecules will be denature and the complementary strand will be wash away.



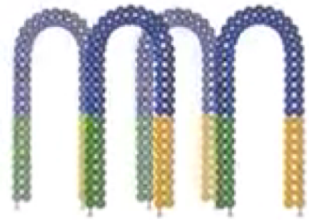
B. Cluster Amplification



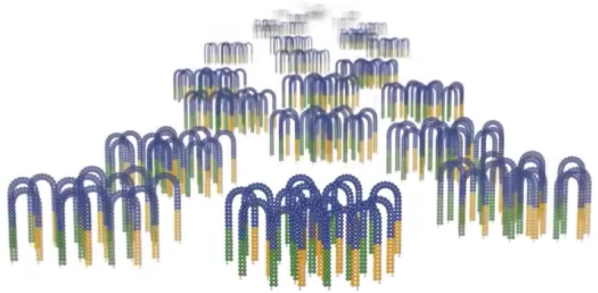
Library is loaded into a flow cell and the fragments are hybridized to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification.



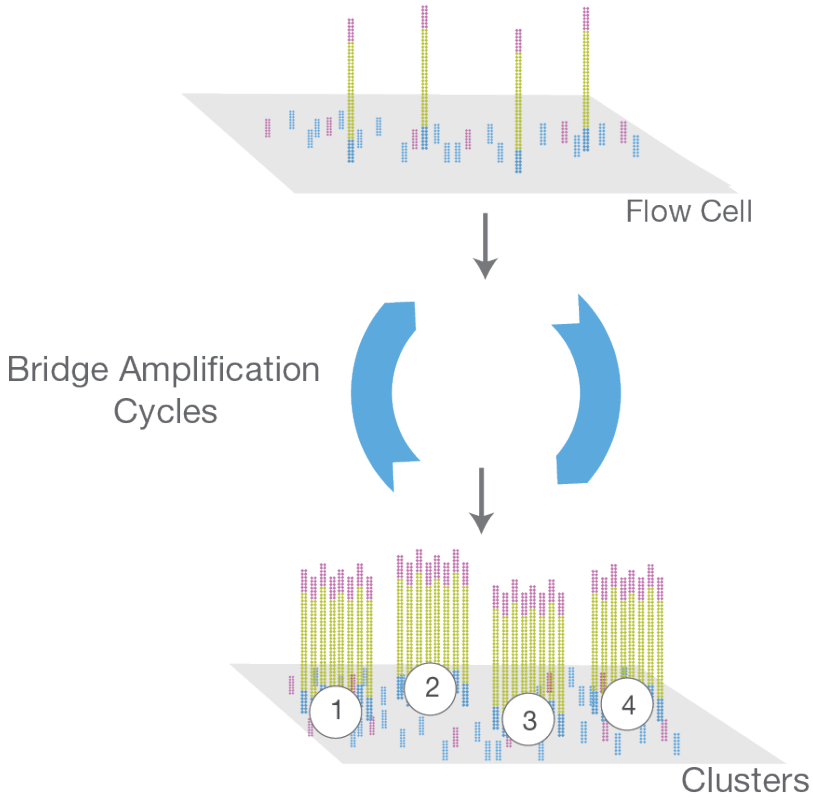
Bridge amplification



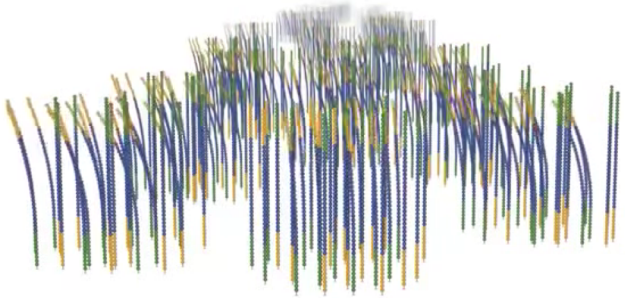
Clonal amplification of all the clusters



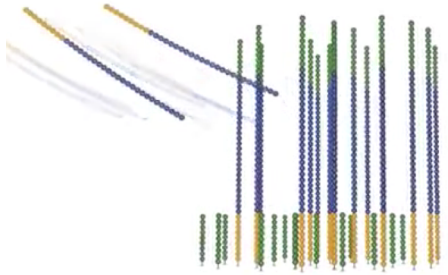
B. Cluster Amplification

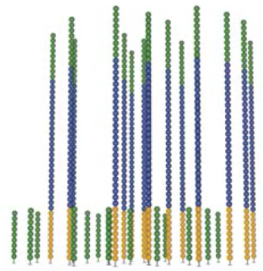


Library is loaded into a flow cell and the fragments are hybridized to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification.

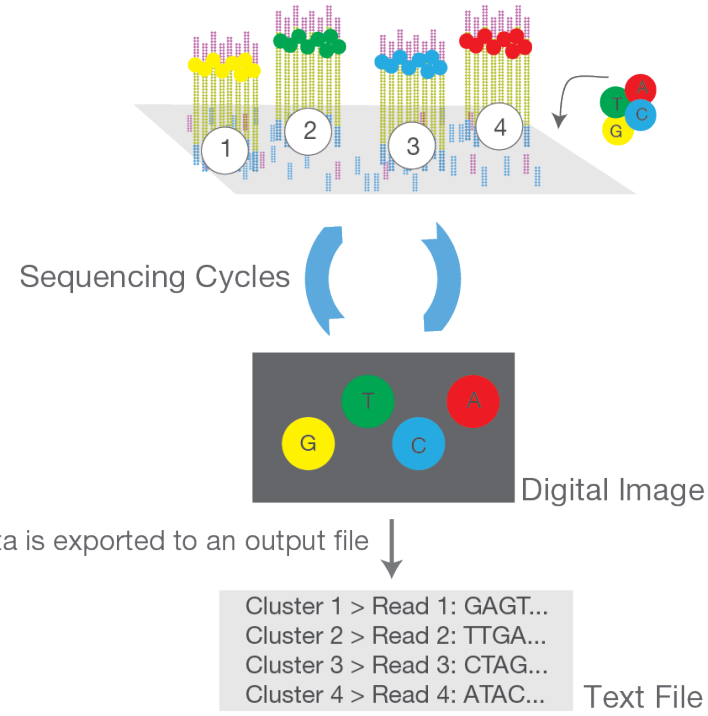


After bridge amplification the reverse strand cleved and will be wash out



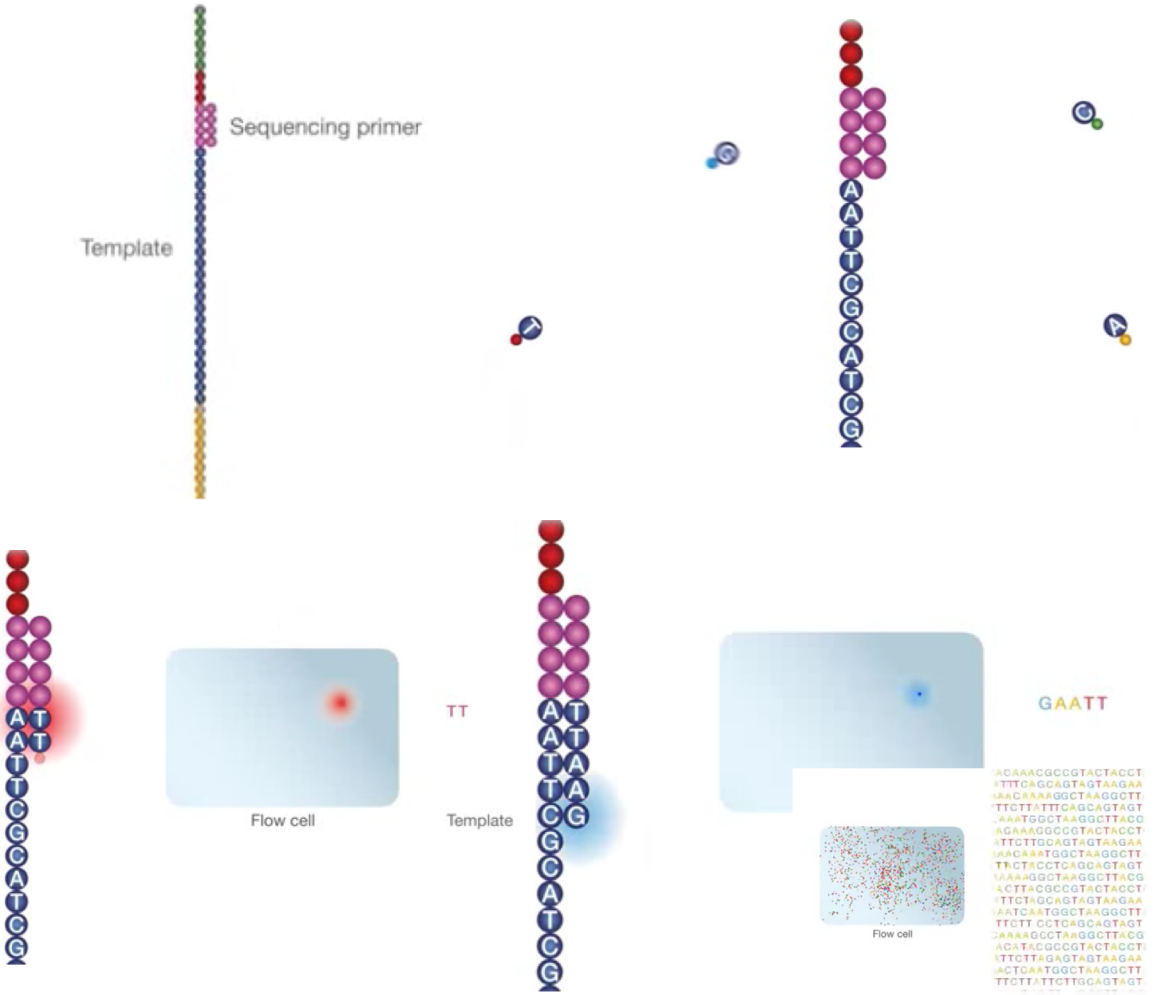


C. Sequencing



Sequencing reagents, including fluorescently labeled nucleotides, are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated “n” times to create a read length of “n” bases.

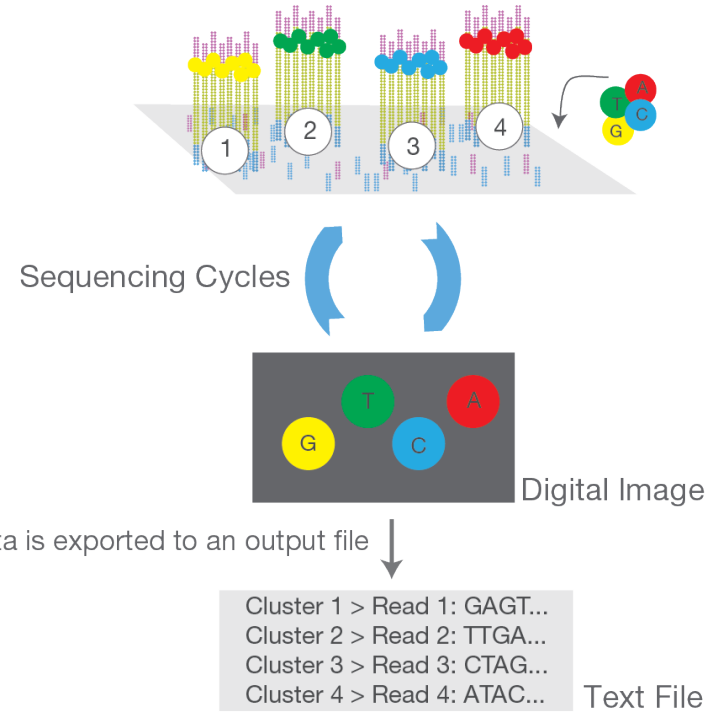
Sequence by synthesis

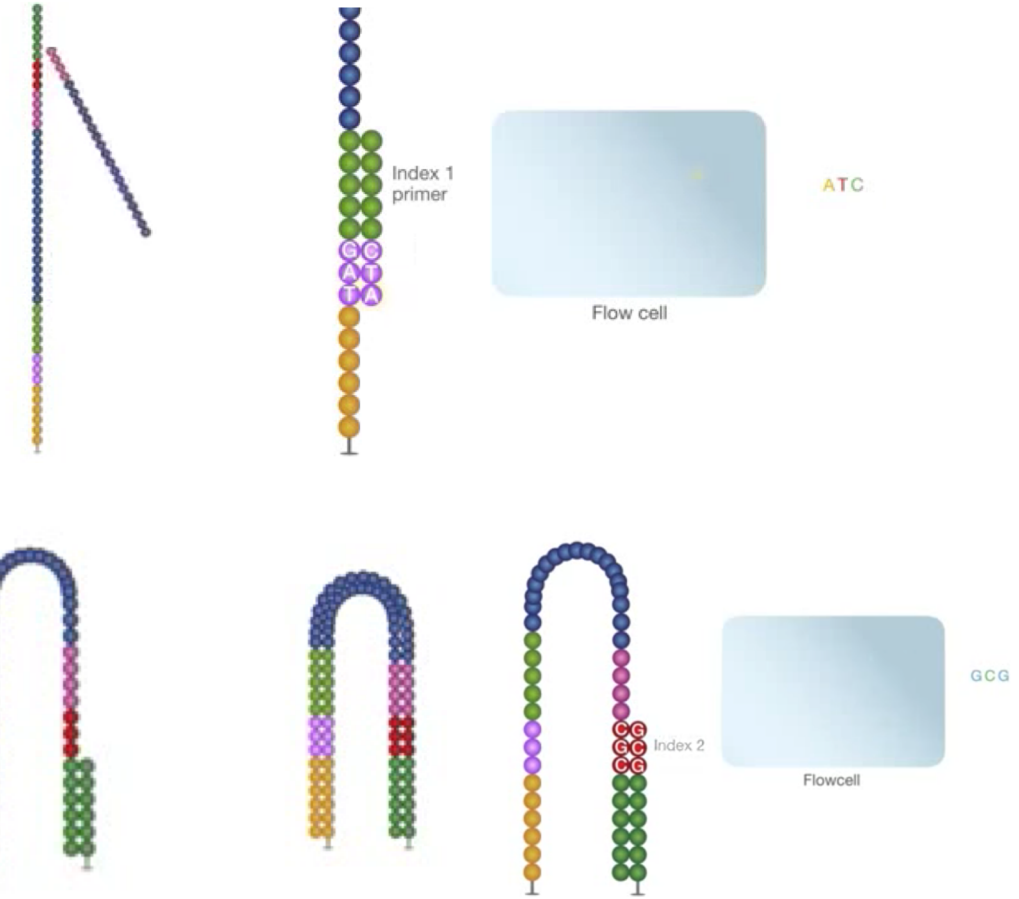


Number of cicles is linked to the lenght of the reads.

For each cluster all identical strands are read simultaneously

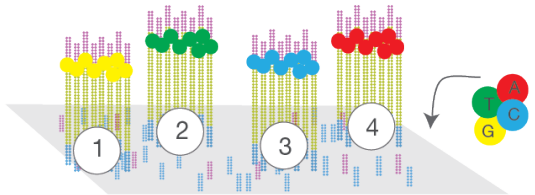
C. Sequencing





Sequencing reagents, including fluorescently labeled nucleotides, are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated “n” times to create a read length of “n” bases.

C. Sequencing



Sequencing Cycles

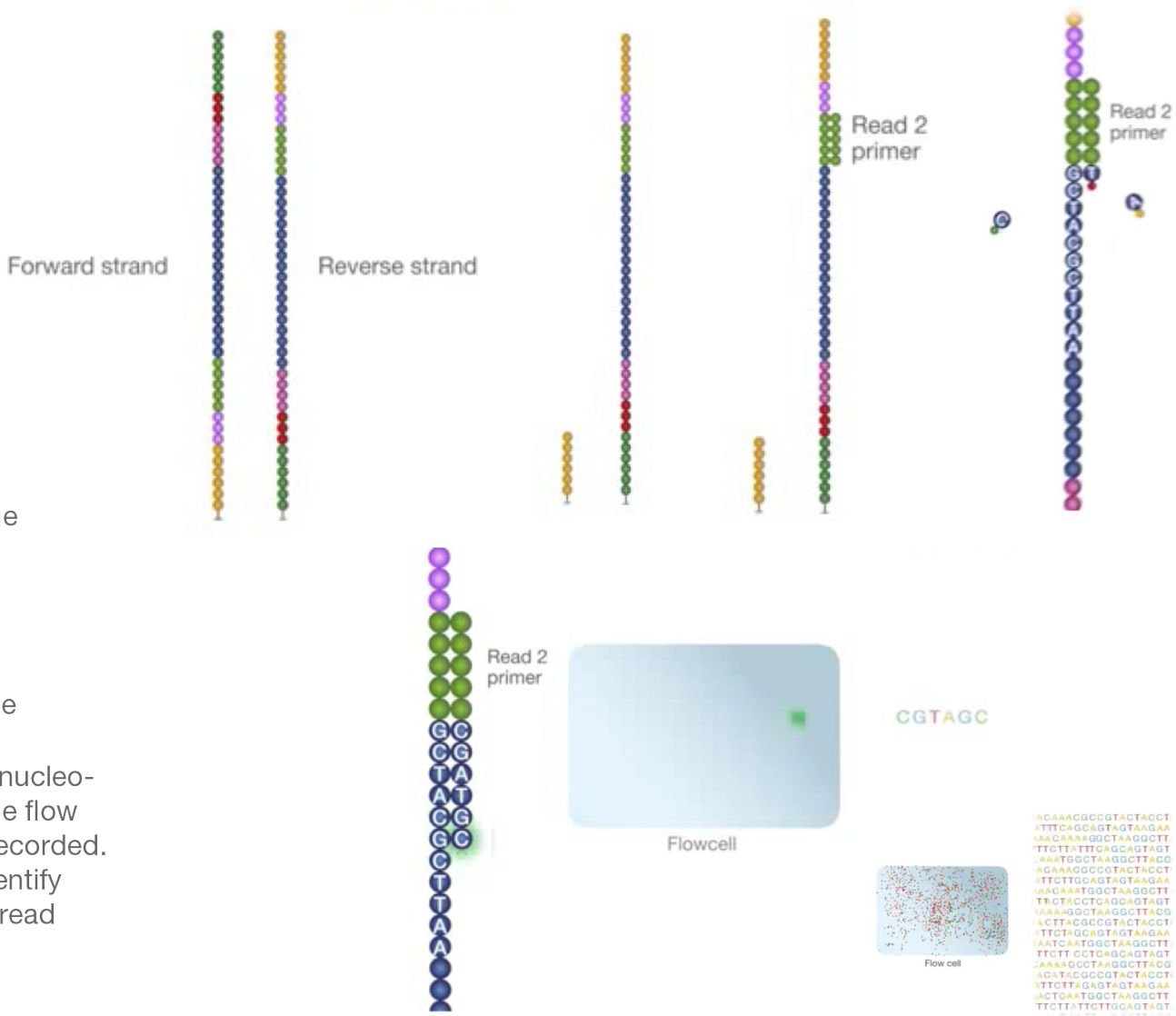
Digital Image

Data is exported to an output file

```
Cluster 1 > Read 1: GAGT...
Cluster 2 > Read 2: TTGA...
Cluster 3 > Read 3: CTAG...
Cluster 4 > Read 4: ATAC...
```

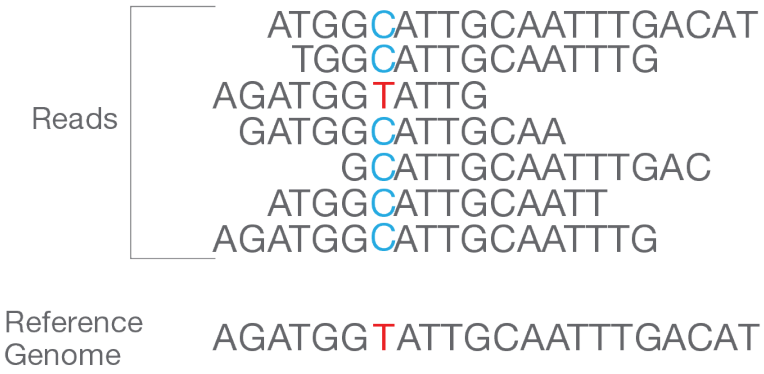
Text File

Sequencing reagents, including fluorescently labeled nucleotides, are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated “n” times to create a read length of “n” bases.



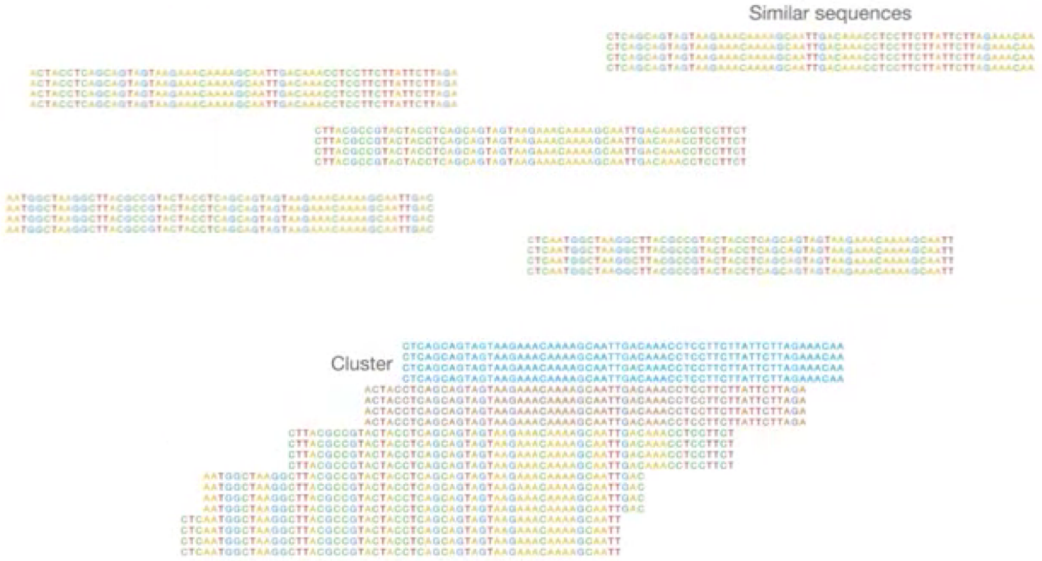
Massive parallel sequencing process

D. Alignment and Data Anaylsis



Reads are aligned to a reference sequence with bioinformatics software. After alignment, differences between the reference genome and the newly sequenced reads can be identified.

Local sequence clustering for each flow-cell cluster



Create contiguous sequences

Library preparation

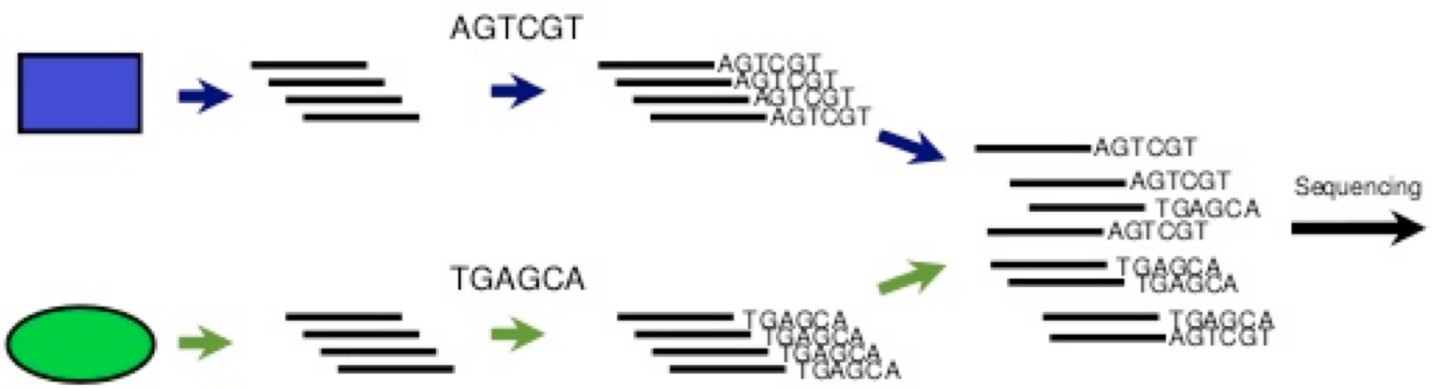
What is multiplexing?

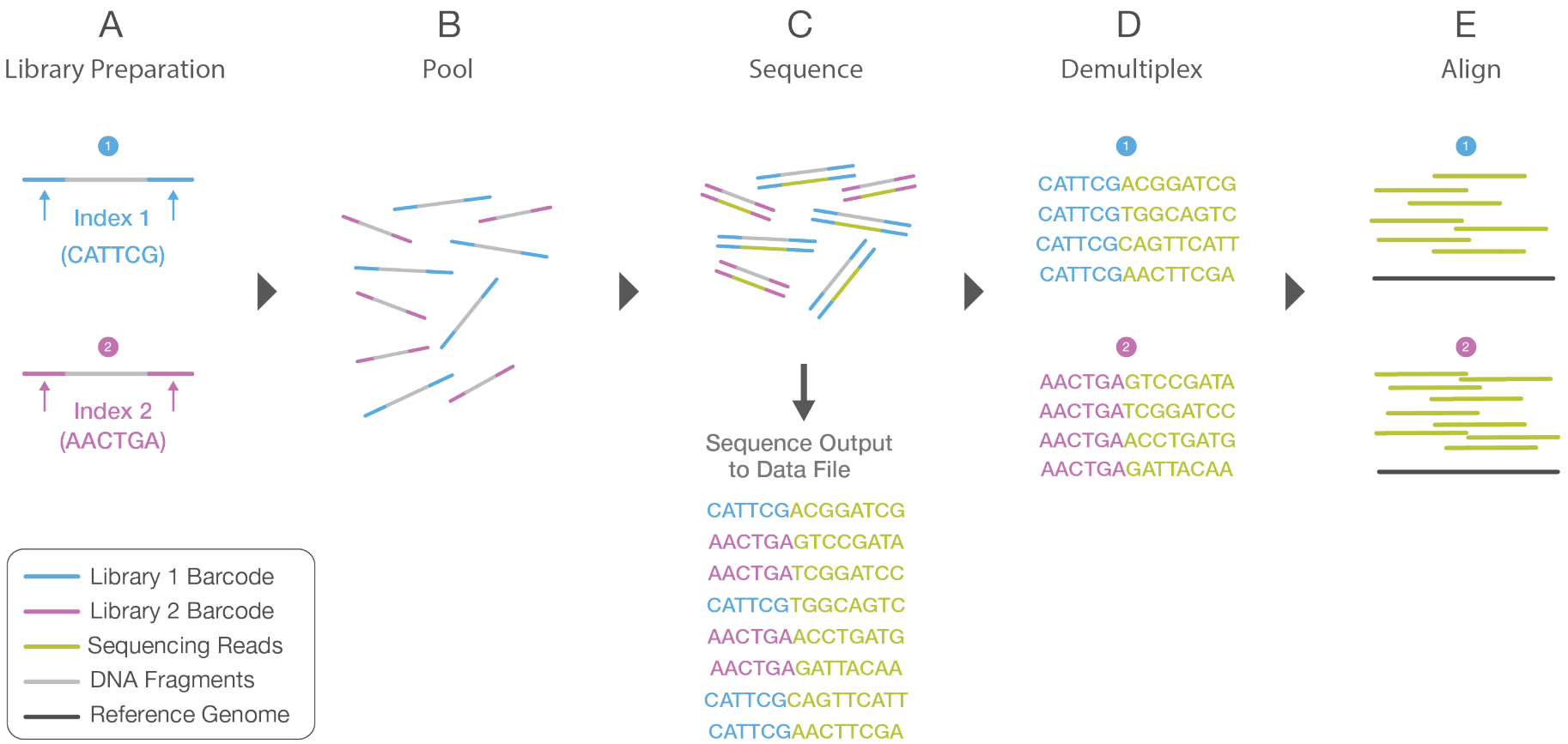
- Multiplexing: a method to analyze multiple biological samples in a single sample.
- Barcodes are unique sequence identifiers added to samples during library construction.
- Once barcodes are added, multiple libraries can be pooled together for emulsion PCR/cluster generation and sequencing.
- Sequence data is then analyzed and traced back to each source.

Library preparation

Multiplexing Libraries

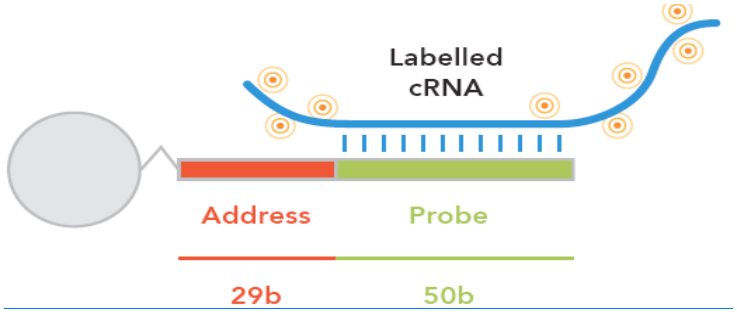
Use of different tags (4-6 nucleotides) to identify different samples in the same lane/sector.





Library Multiplexing Overview—(A) Unique index sequences are added to two different libraries during library preparation. (B) Libraries are pooled together and loaded into the same flow cell lane. (C) Libraries are sequenced together during a single instrument run. All sequences are exported to a single output file. (D) A demultiplexing algorithm sorts the reads into different files according to their indexes. (E) Each set of reads is aligned to the appropriate reference sequence.

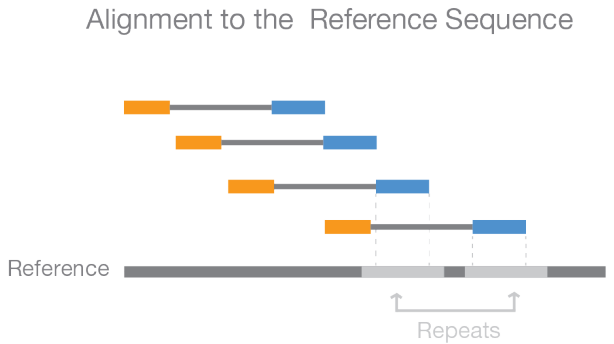
Microarray Technologies



Single end read

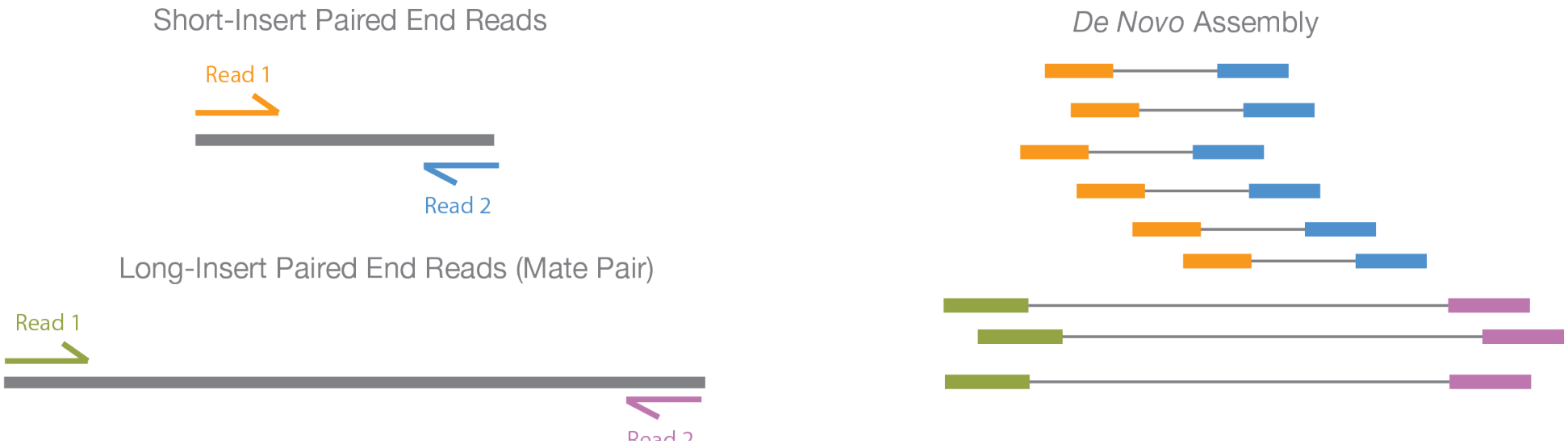
Massive parallel Technologies

Paired- end reads



De novo Sequencing

De novo sequencing refers to sequencing a novel genome where there is no reference sequence available Sequence reads are assembled as contigs and the **coverage quality** of de novo sequence data depends on the **size and continuity of the contigs** (ie, the number of gaps in the data). An important factor in generating high-quality de novo sequences is the **diversity of insert sizes included in the library**. Combining short-insert paired-end and long-insert mate pair sequences is the most powerful approach for maximal coverage across the genome. The combination of insert sizes enables detection of the widest range of structural variant types and is essential for accurately identifying more complex rearrangements.



Mate Pairs and De novo Assembly—Using a combination of short and long insert sizes with paired-end sequencing results in maximal coverage of the genome for de novo assembly.

Targeted Sequencing

With targeted sequencing, a subset of genes or regions of the genome are isolated and sequenced. Targeted sequencing allows researchers to focus time, expenses, and data analysis on specific areas of interest and enables sequencing at much higher coverage levels.