BIOINFORMATICS & HIGH-PERFORMANCE COMPUTING

UNIVERSITY of VIRGINIA | Research Computing

WHAT WE WILL DO TODAY …

- This is not a “Tool X” vs “Tool Y” workshop
- We will not go into the details of alignment algorithms

Who are you???

Bioinformatics Scientist

Identified that you need to align reads

Identified that you need to use Rivanna

What you will learn???

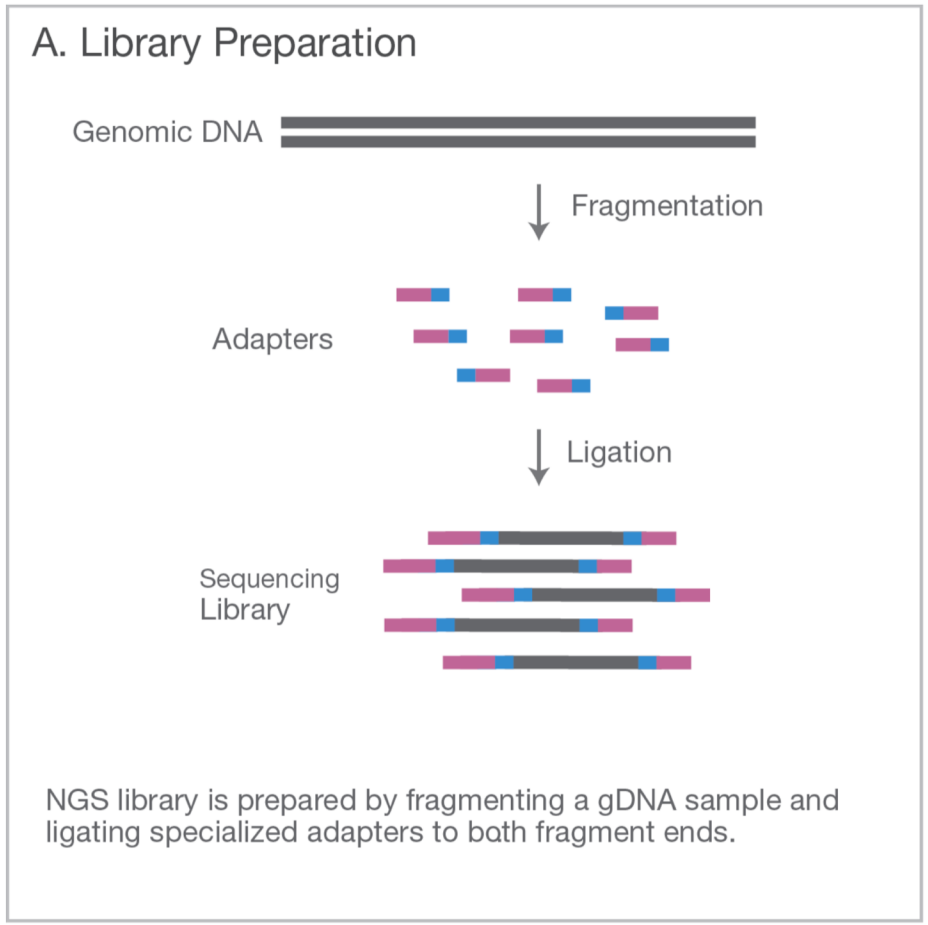
How to use bowtie2 on Rivanna

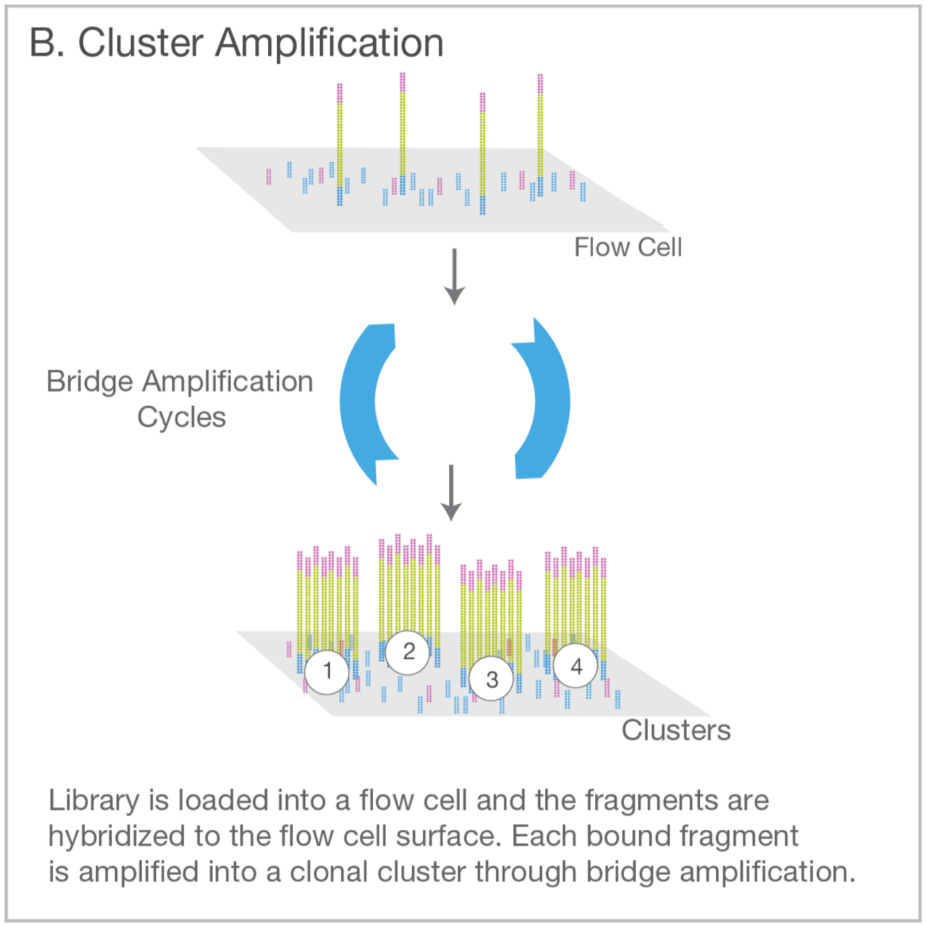
Understand SAM/BAM format

Perform downstream manipulations

BioTuple Icon

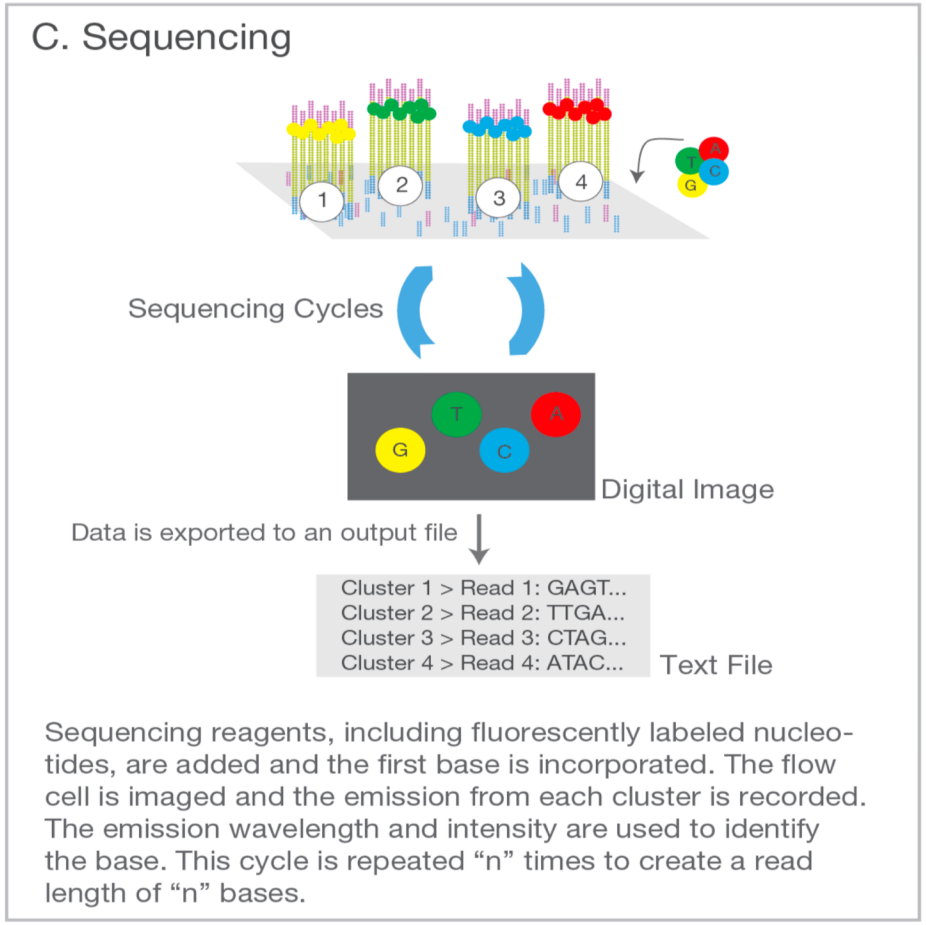
Illumina Sequencing

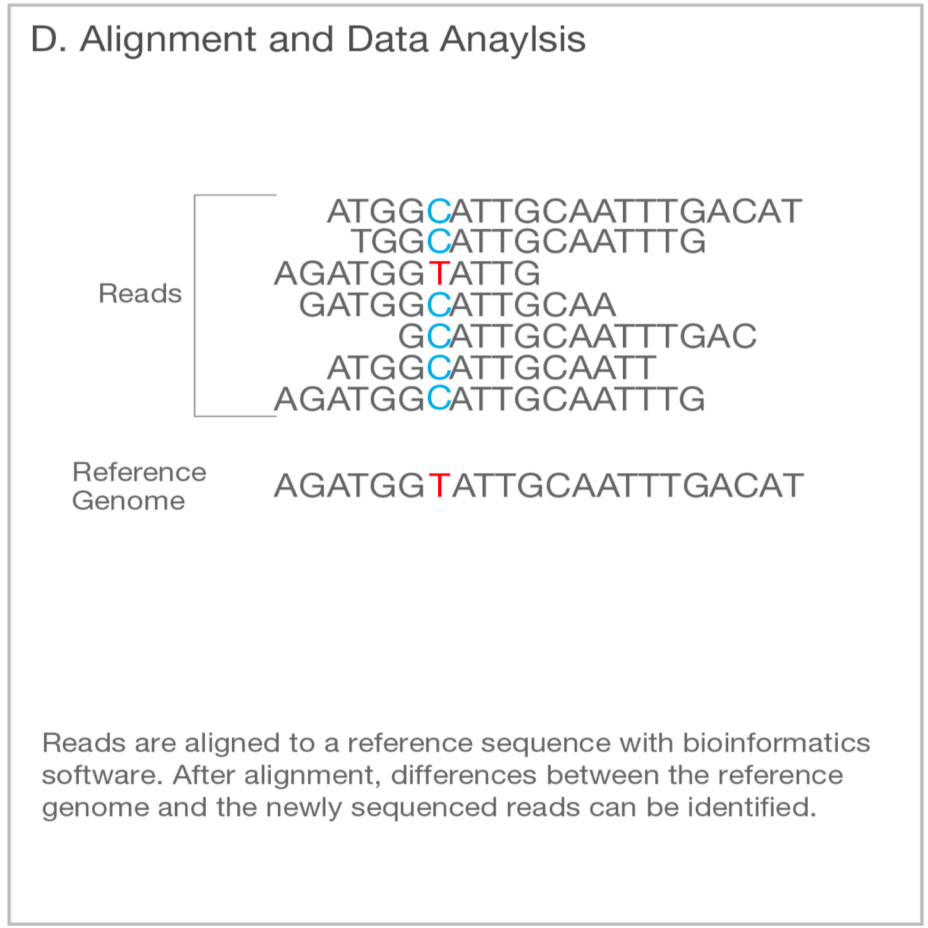




https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf

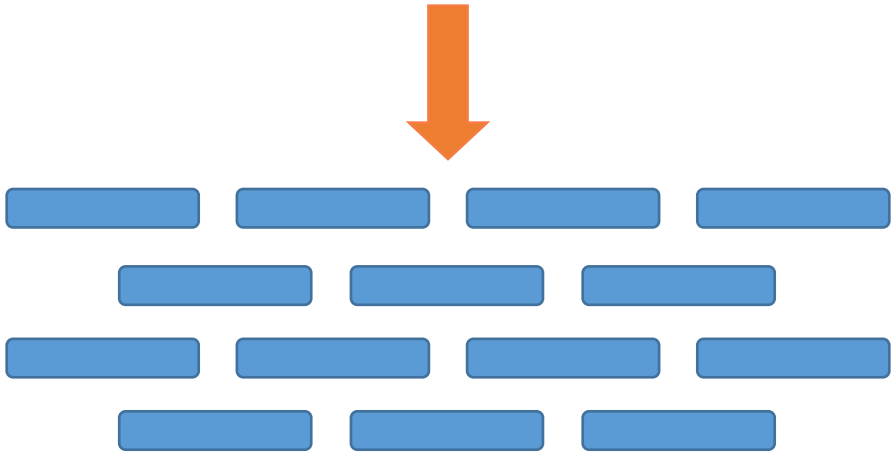
Illumina Sequencing





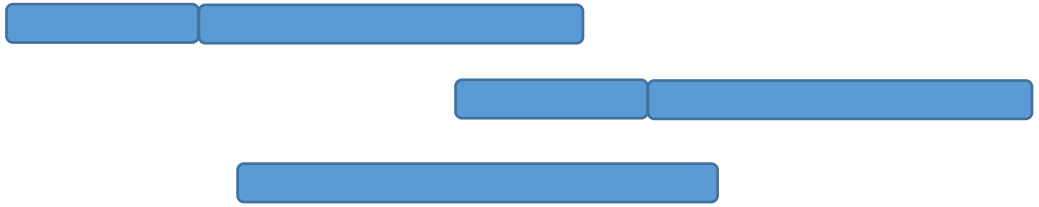
https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf

Short Read Processing



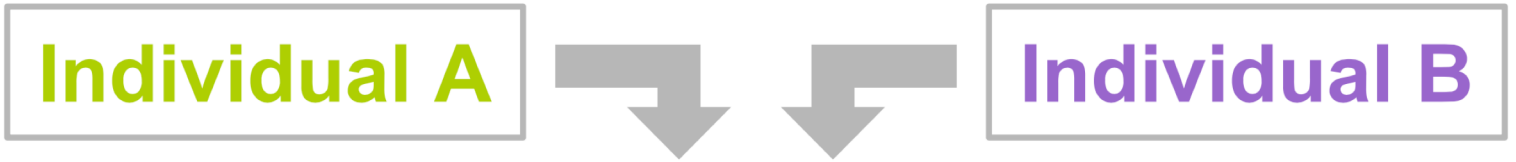
Absence of Reference Genome
de novo Assembly

Align to Reference Genome



When to Align ?

- Identify Variation in Individuals

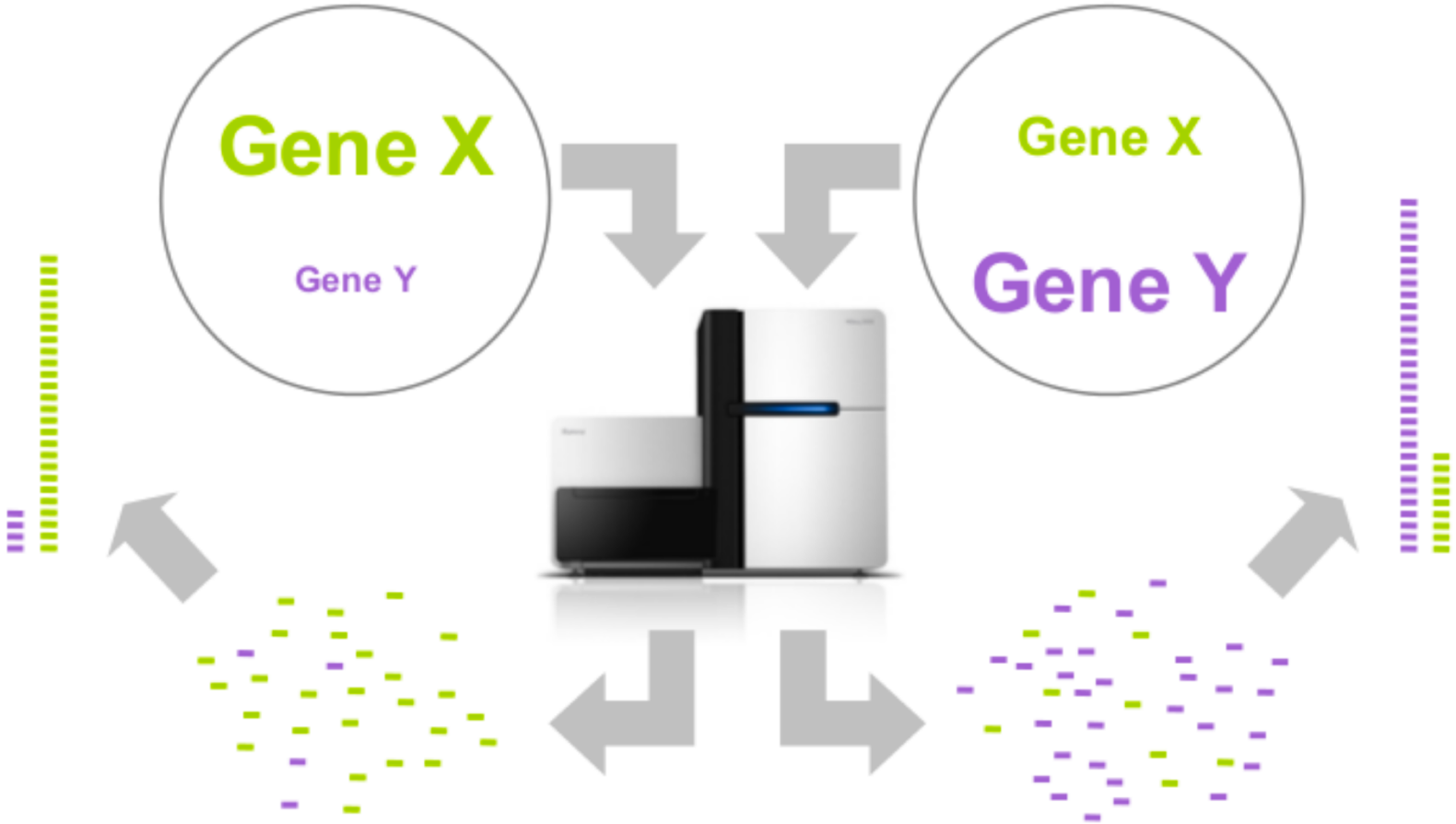


```
ATGATAGCATCGTCGGGTGTCTGCTCAATAATAGTGCCGTATCATGCTGGTGTTATAATCGCCGCATGACATGATCAATGG
                                CAATAAAAGTGCCGTATCATGCTGGTGTTACAATCGCCGCA
                                     CGTATCATGCTGGTGTTACAATCGCCGCATGACATGATCAATGG
                                TGTCTGCTCAATAAAAGTGCCGTATCATGCTGGTGTTACAATC
                           ATCGTCGGGTGTCTGCTCAATAAAAGTGCCGTATCATG--GGTGTTATAA
                                CTCAATAAGAGTGCCGTATCATG--GGTGTTATAATCGCCGCA
                                                     GTTATAATCGCCGCATGACATGATCAATGG
```

Credit: UCD Genome Center, Bioinformatics Core, Alignment 2015-06-17

When to Align ?

- Quantify Abundance



Credit: UCD Genome Center, Bioinformatics Core, Alignment 2015-06-17

Popular Aligners

- **BWA / Bowtie2**

They both use the Burrows – Wheeler Transform to index the reference genome

This massively reduces the memory footprint of long reference genomes, while allowing for rapid identification of potential origin of query sequence

- **Choice of aligner**

Documentation → can I figure out how it works?

Input features → what input can it handle?

Output → will the output be useful for downstream analysis?

Performance → do I have the computational resources to run?

- **DO NOT QUOTE ME:** In some tests, Bowtie2 was slightly faster at marginal expense of sensitivity.