

# Visualization to assess genome assembly

Armaghan Sarvar

September 28th, 2022

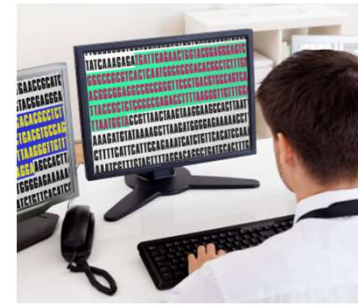
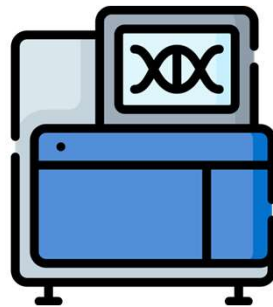


# Genome Sequencing



## ✓ Genome Sequencing:

- Determine the genetic makeup of an **organism** or **cell type**
- How specific **diseases** are formed



## ✓ Improvements:

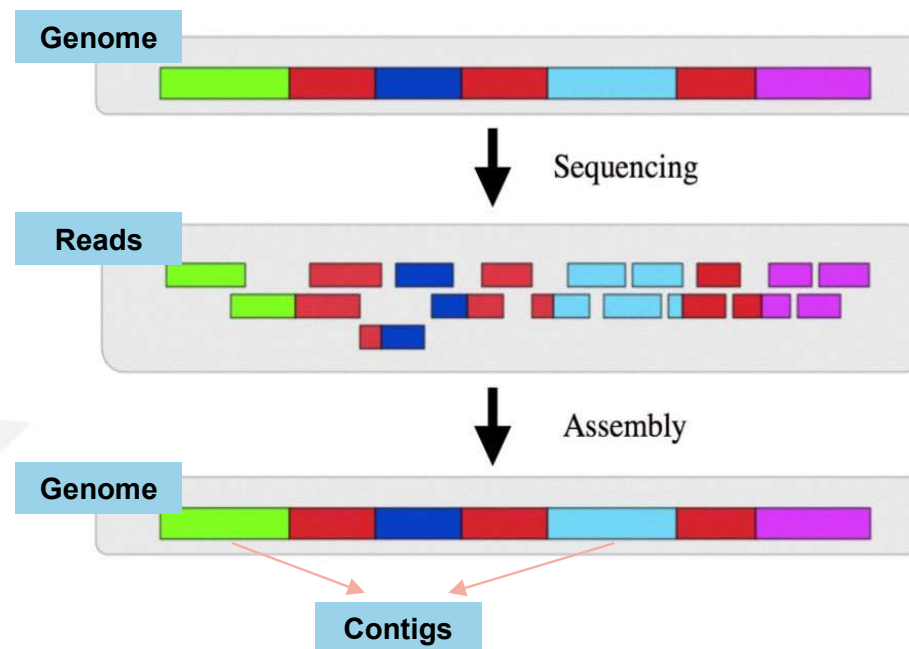


Accessible generation of sequencing **reads** from DNA data

# Genome Assembly



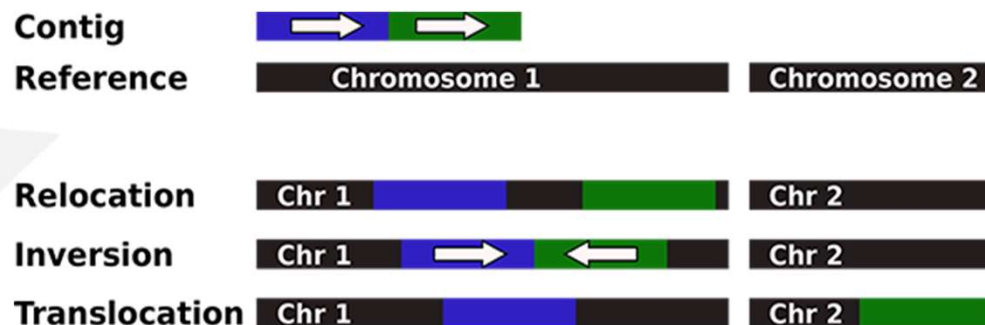
- Lengths of reads are shorter than **genomes** or even **genes**!
- ✓ Genome Assembly:
  - Concatenating nucleotide reads into the correct order



# Genome Assembly



- ✓ As a result:
  - ❓ Errors in the decisions made by assembly algorithms: [Misassemblies](#)
    - Quast [1]:
      - ✓ **Relocation**: the left and right flanking sequences align away from each other
      - ✓ **Inversion**: flanking sequences align on opposite strands
      - ✓ **Translocation**: flanking sequences align on different chromosome



# Project Ideas



- ✓ Draw assemblies according to their alignment/mapping to the reference

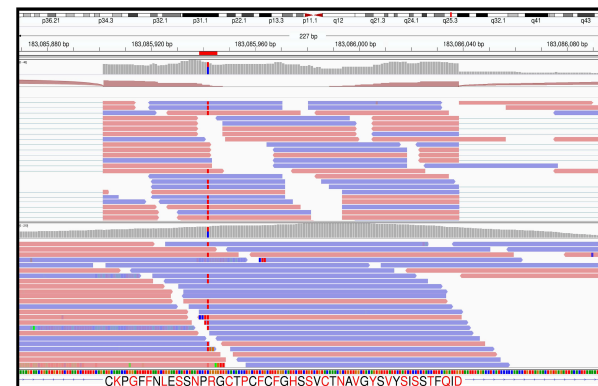
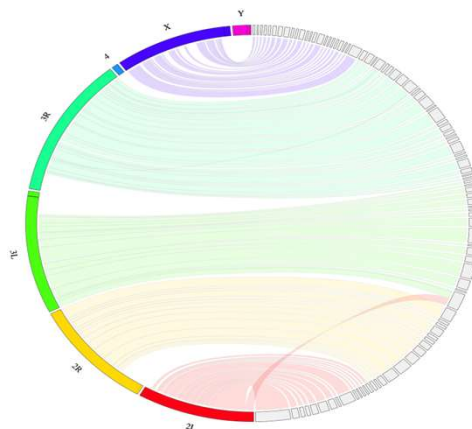
➔ Visualize positions of different types of misassemblies

- ✓ Large-scale  
(Whole Genome)

- Circos-based plots [2]

- ✓ Fine-grained  
(Specific contigs/chromosomes)

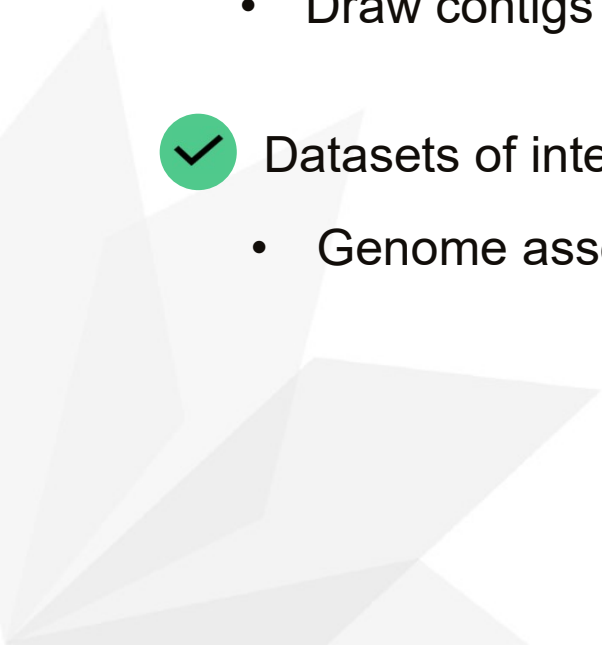
- Integrative Genomics Viewer (IGV) [3]



# Project Ideas



- Show flanking sequences incorrectly assembled
  
- Visualize assembly information such as size or contiguity of contigs
  - Draw contigs ordered based on related metrics (Nx, NGx?)
  
- Datasets of interest:
  - Genome assembly of human or other organisms from NCBI, etc.



# Reference



[1] Gurevich, Alexey, et al. "QUAST: quality assessment tool for genome assemblies." *Bioinformatics* 29.8 (2013): 1072-1075.

[2] Jackman, Shaun D., et al. "Tigmint: correcting assembly errors using linked reads from large molecules." *BMC bioinformatics* 19.1 (2018): 1-10.

[3] Thorvaldsdóttir, Helga, James T. Robinson, and Jill P. Mesirov. "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration." *Briefings in bioinformatics* 14.2 (2013): 178-192.



**Thank you**