Short Read Workshop Day 4
Trimming, Mapping, IGV

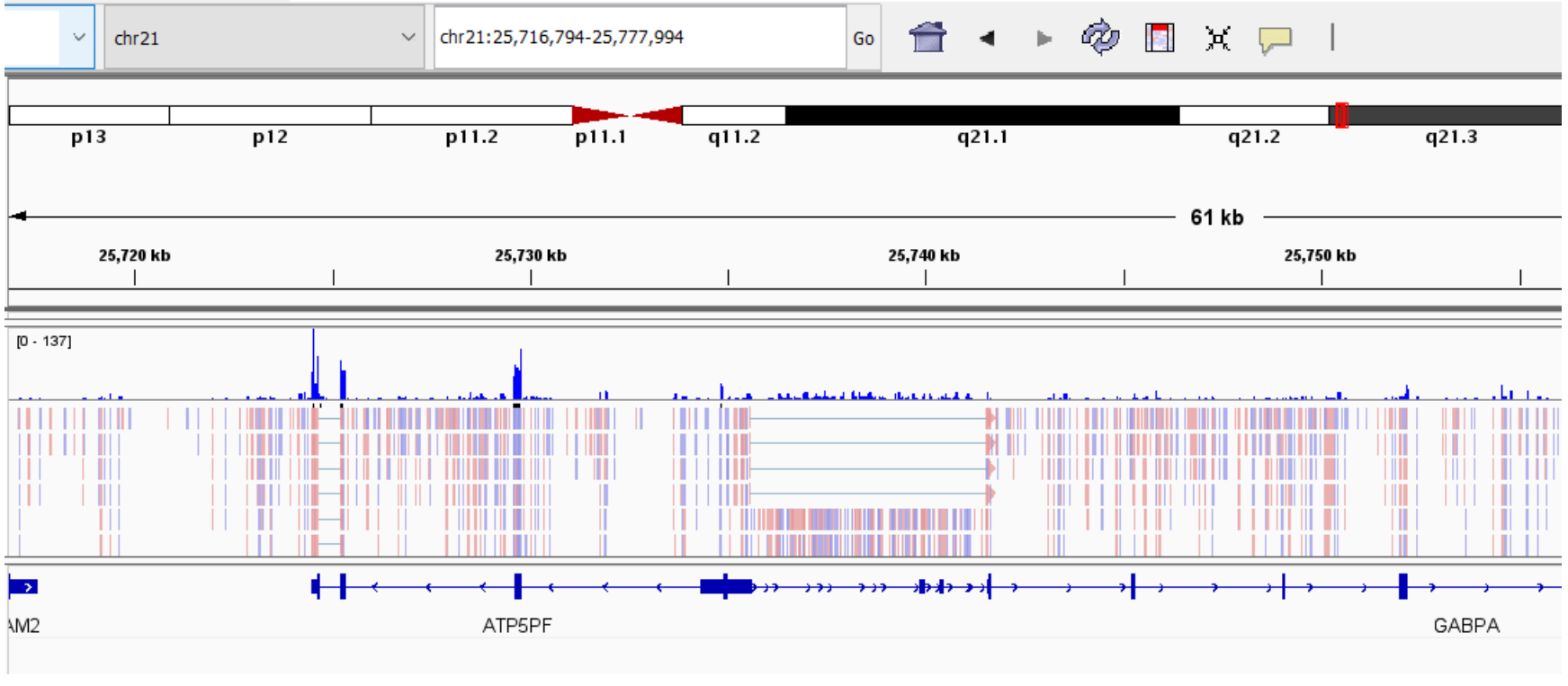
Lynn Sanford, 2022

Day 4 overview

- Trimming fastq files
- Mapping fastq files
- More about mapped file formats
- Visualizing mapped files

Goal of the Day

View sequencing data as reads aligned to a genome



Trimming/mapping recap

50 base read:
TAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGACCAGGATATAG

Trimmomatic

40 base trimmed read:
TAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGAC

HISAT2

Genome:
AGCTTCGGATCGATCGACTGACTAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGACCGCGATTACGAC
TCGAAGCCTAGCTAGCTGACTGATCCGATTGAGACATCGGGGTCCATGGTACGTATTAACTGGCGCTAATGCTG

IGV

AGCTTCGGATCGATCGACTGACTAGGCTAACTCTGTAGCCCCAGGTACCATGCATAATTGACCGCGATTACGAC
TCGAAGCCTAGCTAGCTGACTGATCCGATTGAGACATCGGGGTCCATGGTACGTATTAACTGGCGCTAATGCTG

Trimming fastq files with Trimmomatic

- Follow Trimomatic worksheet to:
- Rsync script to your /scratch/Users/<username>/ from /scratch/Shares/public/sread2022/scripts/day4/d4_trim_qc.sbatch
- Edit script to run trimmomatic

How do you trim polyA regions from both sides of reads?

- Make a new fasta file with a polyA segment, or append to the Illumina adapter file, if writeable
- >polyA

  AAAAAAAAAAAAAAAAAAAAAAAAA
- ILLUMINACLIP:<new fasta file>:2:30:10

Mapping fastq files with HISat2

- Follow Mapping/IGV worksheet to:
- Rsync script to your /scratch/Users/<username>/ from /scratch/Shares/public/sread2022/scripts/day4/d4_mapping.sbatch
- Edit script and run HISAT2
- Visualize BAM file on your local computer

Homework

Day 4 Homework – FASTQC, trimming, mapping, IGV

The assessment tomorrow will run many of the same steps as this homework. These steps are essential in ALL short read data processing.