# Tutorial 5 — BWA, SAMtools & BCFtools Part 2

MICB405 - BIOINFORMATICS - 2021W-T1 08 OCTOBER 2021 AXEL HAUDUC

### BWA MEM vs BWA ALN

#### bwa mem syntax

bwa mem ref.fasta read1.fastq read2.fastq > alignment.sam

#### bwa aln ("base BWA") syntax

bwa aln ref.fasta read1.fastq > read1.sai

bwa aln ref.fasta read2.fastq > read2.sai

bwa sampe ref.fasta read1.sai read2.sai read1.fastq read2.fastq > alignment.sam

### Recap of BWA ALN

Older, original algorithm that created SAM files

Somewhat better for read lengths below 70bp

Tailored to older sequencers that produced ≈ 36bp reads

SAI file = "suffix array index"

- Intermediate file containing Burrows-Wheeler transformed intervals of reference where each read matches
- Converted to SAM taking into account paired-end reads with secondary step "bwa sampe"

### Getting more detailed statistics on BAMs

samtools stats file.bam | grep "^SN"

samtools stats contains sets of information in lines beginning with a set of codes to denote different sections

**CHK** Checksum SN Summary numbers **FFQ** First fragment qualities LFQ Last fragment qualities GC content of first fragments **GCF** GC content of last fragments GCL ACGT content per cycle GCC ACGT content per cycle, read oriented GCT **FBC** ACGT content per cycle for first fragments only FTC ACGT raw counters for first fragments LBC ACGT content per cycle for last fragments only LTC ACGT raw counters for last fragments BCC ACGT content per cycle for BC barcode CRC ACGT content per cycle for CR barcode OXC ACGT content per cycle for OX barcode RXC ACGT content per cycle for RX barcode Quality distribution for BC barcode QTQ CYQ Quality distribution for CR barcode Quality distribution for OX barcode **BZQ** Quality distribution for RX barcode QXQ IS Insert sizes RL Read lengths Read lengths for first fragments only FRL LRL Read lengths for last fragments only ID Indel size distribution IC Indels per cycle

Coverage (depth) distribution

GC-depth

COV

GCD

### Very quick comparison of BAMs

```
diff --side-by-side \
<(samtools stats aln/bordetella.final.bam | grep "^SN") \
<(samtools stats mem/bordetella.final.bam | grep "^SN")</pre>
```

#### Process substitution:

<(COMMAND) = insert output of COMMAND as a file here

Useful when you need to pipe in 2+ items into your command

### Samtools tview

This is a commandline option to viewing read alignments

Less flexible than IGV, but easy for quick comparisons

#### samtools tview syntax

samtools tview file.bam --reference reference.fasta

### Jumping to a specific locus

samtools tview file.bam --reference reference.fasta -p chrom:position

samtools view file.bam chrom:start-end

# Filtering your BAM file

Select reads with a mapping quality above 40

samtools view file.bam -b -h -q 40 > file.filtered.bam

Keep reads w/ matching flags (4 = UNMAPPED)

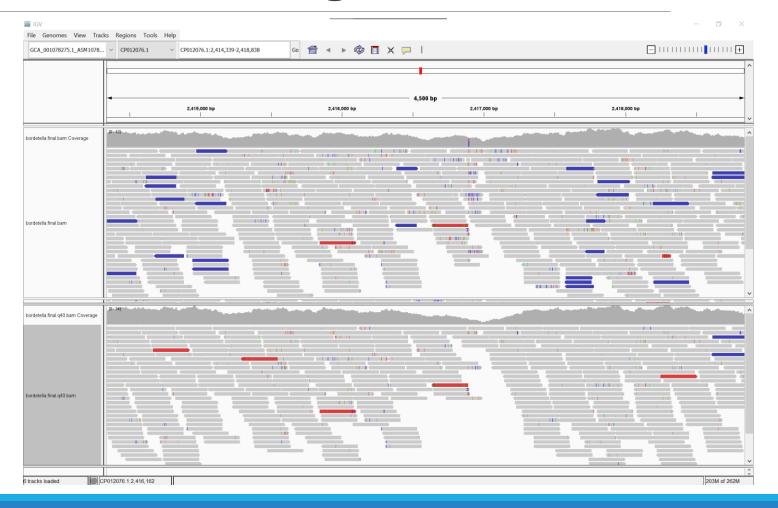
samtools view file.bam -b -h -f 4 > file.filtered.bam

Discard reads w/ matching flags

samtools view file.bam -b -h -F 4 > file.filtered.bam

# View alignments and vcfs together

scp reference, bam, bam.bai, vcf, and vcf.bai into your computer and load into IGV



08-Oct-2021 9