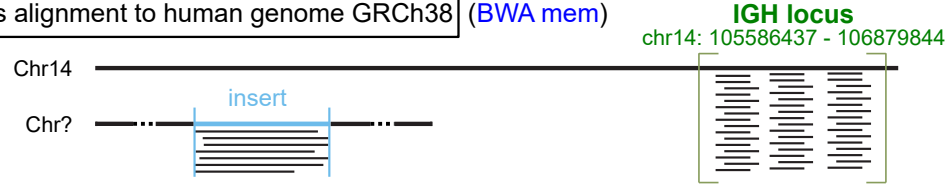


FASTQ (raw) = targeted amplicon reads (300 bp PE)

1. Quality-based trimming and adapter removal. Remove reads < 100 bp ([Trim Galore](#), [FastQC](#))

FASTQ (trimmed)

2. Reads alignment to human genome GRCh38 ([BWA mem](#))



Chr14

Chr?

insert

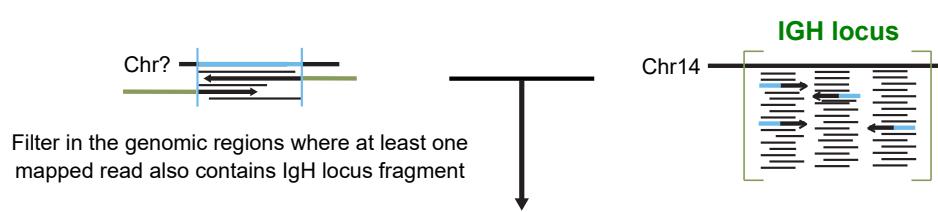
IGH locus
chr14: 105586437 - 106879844

BAM

3. Select genomic ranges where coverage ≥ 10 reads and $20 \text{ bp} \leq \text{length} < 2000 \text{ bp}$ ([bedtools](#))

4. Annotate inserts with GENCODE v29 ([Java](#), [BEDOPS](#))

5. *De novo* assembly for each potential insert ([samtools](#), [pysam](#))



Chr?

Chr14

IGH locus

Filter in the genomic regions where at least one mapped read also contains IgH locus fragment

De novo assembly ([Trinity](#))

Minimum contig length = 150 bp

FASTQ

FASTA

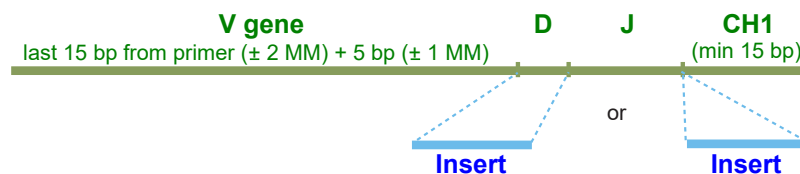
6. Validate the contig sequence

BLAST all inserts against each contig - identify insert coordinates within the contig (min length = 200 bp)
- identify "fused" inserts
- remove insert(s) from contig sequence

Identify V,D,J and CH1 genes ([IgBLAST](#), [Java](#)) and V-gene **ORF** to translate into protein

V D J CH1

Validate contig where the insert is between V gene and CH1 ([Java](#))



V gene

last 15 bp from primer ($\pm 2 \text{ MM}$) + 5 bp ($\pm 1 \text{ MM}$)

D J CH1

(min 15 bp)

Insert

or

Insert

+ Add information about putative functional domain and/or exon-exon junction
[UniProt](#), [JAGuaR](#), [BLAST](#), [Java](#)

Annotated
contigs