Alignment of NGS reads

Estudios in silico en Biomedicina

(Máster en Bioinformática, Universidad de Valencia)

David Montaner

2014-10-21

Aligners

- BWA : DNA
- Bowtie / Tophat : RNA
- ...

Alignment steps

1. Index the reference
2. Align (may be spitted up in several steps)

Other Considerations

- paired ends / single ends
- Reference genome is assumed

SAM Format

SAM : Sequence Alignment/Map format.
A TAB-delimited text format storing the alignment information.
A *header section* is optional.

SAM Mandatory fields

1. QNAME: read name
2. FLAG: bitwise FLAG
3. RNAME: reference sequence name (chromosome)
4. POS: 1-based leftmost mapping position (where the read
5. MAPQ: mapping quality
6. CIGAR: CIGAR string
7. RNEXT: name of the mate/next read (for paired ends)
8. PNEXT: Position of the mate/next read (BP 0)
9. TLEN: observed Template LENgth
10. SEQ: read sequence
11. QUAL: read quality (phred-scale +33)