Sequence Mapping/Alignment 2
BCB 511: Applied Bioinformatics

Matt Settles

February 24, 2015

Outline

Bowtie2

Bowtie2 is an many mapping application developed by Ben Langmead from John Hopkin's University.
As with all modern mappers, you need to first build a reference index, then map reads against the index.

```
bowtie2-build [options]* <reference_in> <bt2_index_base>
bowtie2 [options]* -x <bt2-idx> {-1 <m1> -2 <m2> | -U <r>} [-S <sam>]
```

Local vs Global Mode

```
Global mode forces reads ends to map, no clipping (SOFT or HARD)
For -end-to-end:
-very-fast
-fast
-sensitive
-very-sensitive
Local mode allows for SOFT and HARD clipping of ends
to allow for better alignments
For -local:
-very-fast-local
-fast-local
-sensitive-local
-very-sensitive-local
```

BWA MEM

The BWA algorithm was developed by Heng Li from the Broad
Institute.
bwa index [options]* <in.fasta>
bwa mem [options]* <idxbase> <in1.fq> [in2.fq]

samtools

http://www.htslib.org

PicardTools

http://broadinstitute.github.io/picard/

configure_Mapping.R

Like preproc_experiment, configure_Mapping.R also automates the process of mapping and some samtools processes (and optionally read extraction) and assumes the directory structure created by proproc_experiment and the presence of a samples.txt file.
the script is a part of grc/2.0
configure_Mapping.R [options]
By default, the script uses BWA MEM
When choosing Bowtie2, default is to use global mode,
- -localmode flag to change to local mode

configure_Mapping.R continued

Reference:
configure_Mapping.R will index your references for you, or use an index if one already exists.

Specify a fasta file file to use as a reference, or a mapping_targets.txt file when mapping agains multiple references. mapping_targets.txt can be a tab delimited file specifying the name and location of the mapping references. Should have 2 columns target name and target fasta location, one row for each reference you would like to map each sample to.

configure_Mapping.R reports

The script produces 3 types of reports found under the mapping results folder.

1. MappingFlagstats.txt - Tabled of results summering the flagstat report for each sample and target

2. SummarySample2Targets.txt - Report across each target, reporting which target is most like the closest (most mappable reference)

3. summary reads/proportions reports - for each target a report across samples detailing the number of reads and proportion of reads that map to each contig for each sample.