



## EyeTN workshop

March 30 - April 1 2015

### Mapping of NGS reads

During this practical session you will learn how you can map reads obtained from a sequencing run of DNA against a reference using the mapper BWA and BOWTIE2. Like with every other mapping software, the reference genome has to be indexed. Because the creation of this index on the human genome takes roughly two hours, we already calculated it. The command used to generate the index is:

```
bwa index <reference>
```

and

```
bowtie2-build <reference> <index.prefix>
```

The data for this practical session is stored in the “2\_mapping\_DNA” folder.

You do not need to run the quality analysis on this dataset as we already analyzed that.

#### 1. Mapping using *BWA MEM*

- Create a subfolder in your folder for this analysis
- Normally you would start by creating the index. As described above, the index was already created. It lies in the directory with the reference.
- Because the mapping against the human genome is very time consuming, we split the fastq file in 7 parts, so that each one of you can map one part. In the end we will merge the 7 output files to create one file containing all the mapping results.
- Discuss with your colleagues who will map which part of the reads.
- map your part of the reads against the reference using two threads. (`bwa mem`, output to `mapping.sam`)
- convert the `sam` file into a `bam` file. You need to add the reference as a header. (`samtools`)
- sort the `bam` file (`samtools`)
- create an index for the sorted `bam` file (`samtools`)

#### 2. Mapping using *BOWTIE2*

- Create a subfolder in your folder for this analysis
- Normally you would start by creating the index. As described above, the index was already created. It lies in the directory with the reference.
- Because the mapping against the human genome is very time consuming, we split the fastq file in 7 parts, so that each one of you can map one part. In the end we will merge the 7 output files to create one file containing all the mapping results.

- Discuss with your colleagues who will map which part of the reads.
- map your part of the reads against the reference using two threads. (`bowtie2`)
- convert the *sam* file into a *bam* file. You need to add the reference as a header. (`samtools`)
- sort the bam file (`samtools`)
- create an index for the sorted bam file (`samtools`)

Now we will do the merging of the *sorted bam* files. Additionally we will create an index for the merged file. Afterwards you can analyze the resulting file

### 3. Analyze the mapping

- first get an overview of the mapping quality with `samtools`
- remove pcr duplicates from the mapped bam file
- get an overview of the new bam file using `samtools`
- analyze the mapping quality (`qualimap`)
- compare the mapping results from *bwa* and from *bowtie2* in *igv*