What am i about to say?

Intro

Sample collection and DNA extraction

Sequencing

Preprocessing:

qsub ./my\_fastX\_trimmer.py 🡺 WERKT NIET OP GZ FILES 🡺 eerst unzippen dan file kopieren 🡺zulke dingen nakijken later op de job

Q33 or Phred score = error van 10^-4

Germ line variant calling

First we unzip the files because the BWA algorithm does not work with zipped files 🡺 for file in `ls |grep gz`; do gunzip $file; done

GATK:

Provides VDA(variant discovery analysis) in HTS data 🡺 provides workflow to do so

Reads are not all the same size

Picard is crazy to use

Read group adding = little identifier

BWA: is a software package for mapping low-divergent sequences against a large reference genome, such as the human genome. It consists of three algorithms: BWA-backtrack, BWA-SW and BWA-MEM. The first algorithm is designed for Illumina sequence reads up to 100bp, while the rest two for longer sequences ranged from 70bp to 1Mbp. BWA-MEM and BWA-SW share similar features such as long-read support and split alignment, but BWA-MEM, which is the latest, is generally recommended for high-quality queries as it is faster and more accurate. BWA-MEM also has better performance than BWA-backtrack for 70-100bp Illumina reads.

Depending on the questions we use the right algorithm

Using bwa mem does everything in 1 time

ID=sample, SM=sample name, LB= sample name, PL= Illumina , PU= sample name

Base quality recalibratos = tells you if there are systematic errors => detect systematic errors

From SAM to BAM

SAM tools import 🡺 full path to reference [use 1ppn] 🡺sam files names in bam veranderen = the same as same view

Then SAM sort

Then SAM index

**For every sample we run qaulimap and run multiqc again**

=> reference genome:

qualimap bamqc -bam [infile] -gff /nlustre/users/fourie/H.sapiens/intervals/trusight\_cancer\_manifest\_a.bed -outdir [outdir]

(/nlustre/users/fourie/H.sapiens/gatk\_resources\_bundle/2.8/hg19/uscs.hg19.fasta) 🡺 does not ask for a reference genome because it is normally already inbedded in the bam files

Qaulimap gene annotation files:

/nlustre/users/fourie/BIFHons/Mapping/ucsc\_hg19\_refseq.gtf

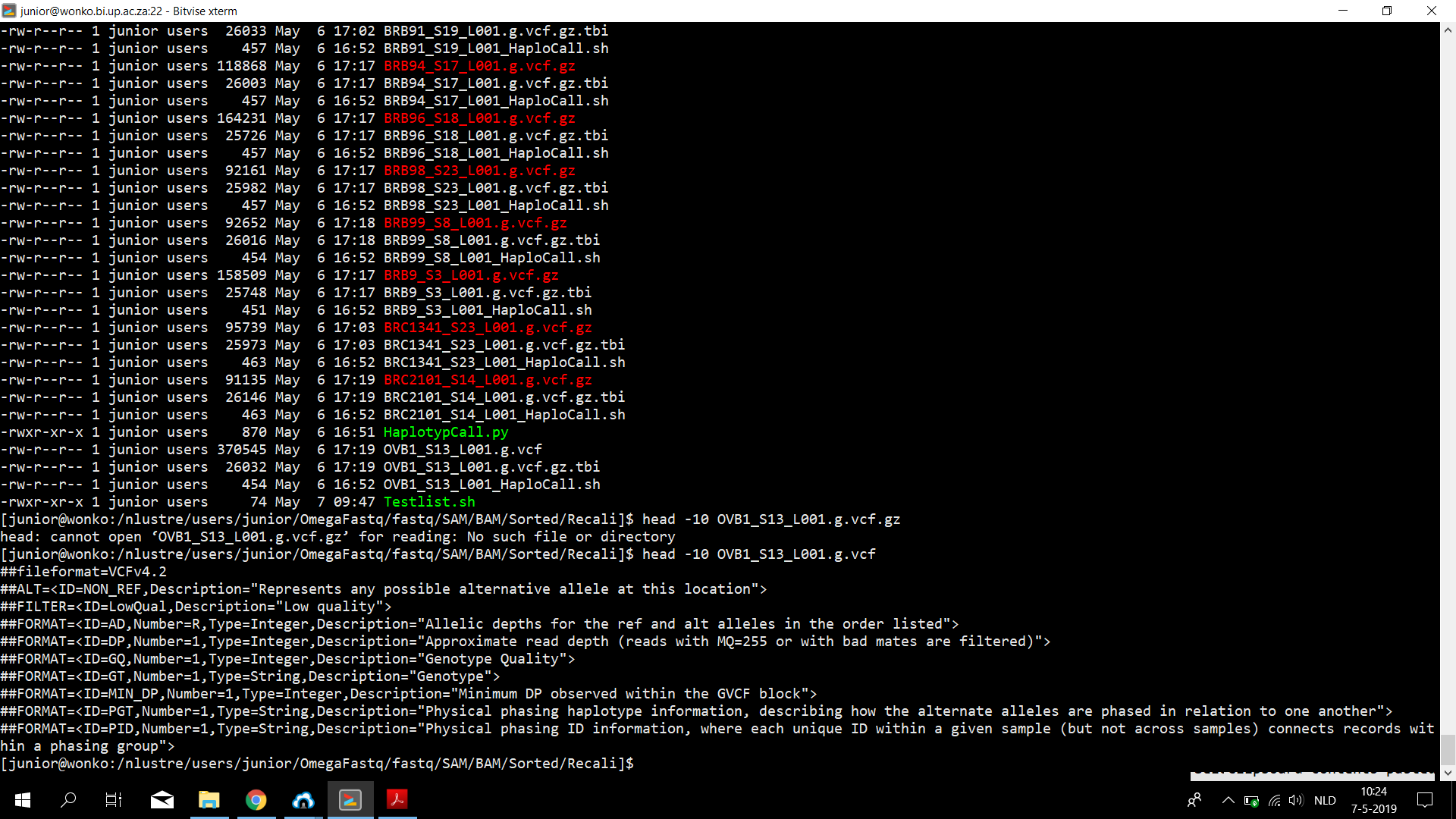
File that came with the kit

Picard:

Retrieves duplicated and adds a label to the reads to say that they are independent reads…

BQRS/ApplyBQRS or Printreads

Haplotypecaller



One file gunzipped and then head -10 $file to look at the file

Could not zip the file back so, did used previous script to create back a zipped OVB1\_S13\_L001.g.vcf.gz file



