*Emma Risson*

**ReadMe TP NGS**

**First week – Variant calling on a family**

***Installation.sh***

Script for the installation of:

* **FastQC** = quality control checks on raw sequence data coming from high throughput sequencing pipelines
* **BWA**-**MEM** = mapping sequences on a reference genome
* **SAMtools** = various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format.
* **IGV** = high-performance visualization tool for interactive exploration of large, integrated genomic datasets
* **GATK** = analysis of alignment for variant calling
* **Picard**

***Mapping.sh***

Script for the mapping of the daughter and the mother of our study, on the reference genome

***Trio-analysis.sh***

Script to take into account that we are analyzing a family, to cross information of the 3 individuals for the analysis

***Variant-calling.sh***

Script for the variant calling (“real” analysis of our study). The most important script is the loop for the 3 individuals at the end of the file.

**Second week – Population wide analysis**

***Script python***

* Extract\_pop.py: filtering by populations for the panmixia hypothesis
* Gtf\_to\_bed.py: conversion of the reference genome from gtf file to bed file
* Vcf\_analysis.py:
* Vcf\_meta\_analysis.py: statistical analysis of the variances
* Vcf\_coding\_polymorphism.py

***Installation.sh***

Everything needed for the analysis run

***Analysis.sh***

All the analysis : cleaning of the data, filtering by populations, analysis of the variances