Supervisor Meeting Minutes

* 26/05/2017
  + Met with supervisor for the first time to discuss project logistics
  + Discussed project requirements
  + Discussed preliminary report details.
  + Requested information about the file formats for the input of the pipeline.
* 30/06/2017
  + Discussed provided literature.
  + Suggested C++ as a development environment, and PLINK as the genome analysis toolset.
* 06/09/2017
  + Discussed meeting schedules for the 4th block.
  + Rehashed general project specifications and goals.
  + Discussed the use of different threading libraries for the parallelization of the pipeline.
* 20/09/2017
  + Queried the correlation between the allelic score and phenotypic load.
  + Confirmed that the beta value is equivalent to the Odds Ratio.
  + Queried correct way to change file delimiters.
    - Scott suggested using, translate and sed.
  + Scott suggested using p-values as inputs for cut-off parameters.
  + Discussed running plink from console/bash or from source, and the possibility of using a workflow language such as Nextflow.
  + Confirmed the usage of –keep and – remove plink commands.
* 29/09/2017
  + Told Scott about K-Fold cross validation functioning, and confirmed method of splitting data was sufficient.
  + Asked about performing different types of association tests. Decided that a general case was sufficient for the scope of the project.
  + Queried which aspects to parallelize, decided on PLINK and K-Fold
  + Scott suggested to test the speed of different cores
  + Discussed the final results, merging of scores and their analysis.
  + Scott requested SSH keys to provide access to the cluster.
* 06/10/2017
  + Scott suggests PICKLE for python script management.
  + Discussed how to pass files through Nextflow processes correctly.
  + Cross-validation sets can’t be taken directly from sets due to block biasing. The data must be shuffled before sets are found so that subsets are random.
  + Discussed pbs profile/job files versus running Nextflow on the cluster head node.
* 13/10/2017
  + Confirmed usage methods of the cluster.
  + Demonstrated pipeline to Scott.
  + Discussed the usage of bootstrapping for the analysis and validation of results.
* 20/10/2017
  + Queried strange behavior of number of cores on the computation time.
  + Queried K-Folds effect on reducing standard deviation, possibly the sample size.
    - Suggested investigation and discussion in report.