Αναφορά Project 11

In this project , using various tools,at first we simulate the sets of case , control datasets at SNP markers and then analyze and associate the generated genomes.

The tools we used are :

HAPGEN2 : which simulates case control datasets at SNP markers.

GTOOL : to convert HAPGEN2 files created to PLINK-compatible files

PLINK : a whole genome association analysis toolset

SNPTEST : a program for the analysis of single SNP association in genome-wide studies.

**HAPGEN2**

We chose from the hapmap offered by HAPGEN2 to simulate multiple disease SNPs on chromosome 6. With the command :

./hapgen2 -h hapmap3\_r2\_b36/hapmap3\_r2\_b36\_chr6.haps -l hapmap3\_r2\_b36/hapmap3\_r2\_b36\_chr6.legend -m hapmap3\_r2\_b36/genetic\_map\_chr6\_combined\_b36.txt -o OUTPUT -dl 2359212 1 2 2.5 -n 100 100

3 files are generated for each case and controls data and .legend , .sample files.

Those are named :

1. OUTPUT.cases.gen
2. OUTPUT.cases.haps
3. OUTPUT.cases.sample
4. OUTPUT.controls.gen
5. OUTPUT.controls.haps
6. OUTPUT.controls.sample
7. OUTPUT.legend
8. OUTPUT.summary

And finally we have our generated Datasets!

**GTOOL**

GTOOL is used to merge all case and controls files.

./gtool -M --g OUTPUT.cases.gen OUTPUT.controls.gen --s OUTPUT.cases.sample OUTPUT.controls.sample --og joined.gen --os joined.sample --threshold 0.9 --phenotype pheno

The result files are : 1) joined.gen (includes cases.gen and controls.gen) and 2) joined.sample (includes cases.sample and controls.sample).

Then we convert them to .ped and .haps files which are used by PLINK using command :

./gtool -G --g joined.gen --s joined.sample --ped joined.ped --map joined.map --threshold 0.9 --phenotype pheno

After some corrections the final .ped file is : phenojoined.ped .

**PLINK**

 PLINK is an open-source whole genome association analysis toolset. To make the association using PLINK we run the command :

./plink --noweb --ped ../phenojoined.ped --map ../joined.map --assoc --allow-no-sex --out Q

And the files generated are Q.assoc that includes the result of a 1df chi-square allelic test.

**SNPTEST**

The SNPTEST program is primarily a package to carry out tests of association at SNPs. Takes input files from hapgen2 ( .gen and .sample) and generates a file that contains the results of a score test which fully account for the uncertainty in imputed genotypes. The file is created by executing the command :

snptest\_v2.5.4-beta3\_linux\_x86\_64\_static/snptest\_v2.5.4-beta3 -data joined.gen joined.sample -frequentist 1 -method score -o SNP -pheno pheno

**Computational Requirements**

The total execution of the script is about 32:45 min ( in the worst case a.k.a. Download speed : 2,98 mbps , Upload speed : 0,73 mbps ).

The total memory , the script needs to have is 1060 MB.

**WORKFLOW**

OUTPUT.cases.gen

OUTPUT.cases.haps

OUTPUT.cases.sample

OUTPUT.controls.gen

OUTPUT.controls.haps

OUTPUT.controls.sample

OUTPUT.legend

OUTPUT.summary

HAPGEN2

joined.sample

joined.gen

GTOOL

joined.ped

joined.map

Merge

Merge

Convert

SNPTEST

PLINK

Association

Q.assoc

Association

SNP