

Genetics and Genomics in Clinical Research

An Immersion Course for Clinical Investigators at UAB

Linkage and Association analysis demonstration

April 4, 2013

By – Amit Patki

ABOUT COMPUTING HARDWARE

Where to run your Linkage/GWAS analysis? Why?

- A cluster vs. your desktop computer.
- Vetted tools with many eyes.
- Scalability.
- Dedicated SIG hardware for proteomics and genomics researchers at UAB.
- <https://docs.uabgrid.uab.edu/wiki/Cheaha>, description of the Cheaha cluster.

Basic concepts:

- Basic Unix commands
- head node
- compute nodes
- job script
- Interactive mode

LINKAGE ANALYSIS USING MERLIN

Input files

Pedigree file:

```
<contents of basic2.ped>
1 1 0 0 1 1 x 3 3 x x
1 2 0 0 2 1 x 4 4 x x
1 3 0 0 1 1 x 1 2 x x
1 4 1 2 2 1 x 4 3 x x
1 5 3 4 2 2 1.234 1 3 2 2
1 6 3 4 1 2 4.321 2 4 2 2
<end of basic2.ped>
```

Data file:

```
<contents of basic2.dat>
A some_disease
T some_trait
M some_marker
M another_marker
<end of basic2.dat>
```

Map file:

```
<contents of basic2.map>
CHROMOSOME MARKER POSITION
24 some_marker 123.4
24 another_marker 136.2
<end of basic2.map>
```

LINKAGE ANALYSIS USING MERLIN

Error detection

```
prompt> merlin -d error.dat -p error.ped -m error.map -error
```

```
Family:      2 - Founders: 2 - Descendants: 2 - Bits: 2  
MRK11 genotype for individual 3 is unlikely [0.003848]  
MRK11 genotype for individual 4 is unlikely [0.003848]  
  
Family:     73 - Founders: 2 - Descendants: 2 - Bits: 2  
MRK17 genotype for individual 3 is unlikely [0.008866]  
MRK17 genotype for individual 4 is unlikely [0.008866]  
  
Family:     81 - Founders: 2 - Descendants: 2 - Bits: 2  
MRK8 genotype for individual 3 is unlikely [0.001567]  
MRK8 genotype for individual 4 is unlikely [0.001567]  
  
Family:     94 - Founders: 2 - Descendants: 2 - Bits: 2  
MRK12 genotype for individual 3 is unlikely [0.002101]  
MRK12 genotype for individual 4 is unlikely [0.002101]
```

NON-PARAMETRIC LINKAGE

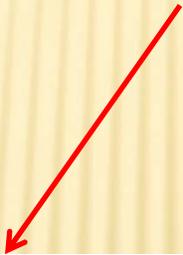
Nonparametric linkage analysis methods does not make any assumptions about the disease model.

- **Run pedstats to get pedigree statistics**
- **Run merlin for non-parametric linkage**
- **Options:**
 - General : --error, --ibd, --kinship, --information
 - Linkage : --npl, --pairs, --qtl, --deviates, --vc, --grid
 - Haplotyping : --best, --sample, --all, --founders
 - Output : --quiet, --markerNames, --pdf

PARAMETRIC LINKAGE

Parametric linkage analysis methods assumes certain disease model.

- Run merlin for parametric linkage
- **Options:** All options remain same except instead of --npl option use --model option and provide *parameter file*.



Affection	Disease Allele Frequency	Penetrances	Model Name
VERY_RARE_DISEASE	0.0001	0.0001,1.0,1.0	Rare_Dominant

PLINK

PLINK is a comprehensive GWAS tool.

It allows you to do ...

- Data management: recode SNPs, remove individuals/SNPs, merge datasets, extract SNPs, flip strands, etc.
- Data Quality Control for GWAS
- Remove bad quality data
- Run association analysis

QC USING PLINK

- Check missing data
- Check allele frequency
- Confirm self reported gender
- Check cryptic relatedness
- Check HW equilibrium
- Check for Mendelian errors (for family data)

GWAS USING PLINK

Unrelated individuals

- Case-control
 - Allelic chi-square test
 - Fisher's exact test
 - Logistic regression

- Quantitative trait
 - Allelic linear model (1df test)
 - Genotypic linear model (2df test)
 - Dominant or recessive model
 - G×E interaction for dichotomous covariate

Family data

- Dichotomous trait
 - TDT
 - DFAM: family and unrelated data together
- Quantitative trait
 - QFAM: permutation test using between and within variance