Matrix eQTL: Ultra fast eQTL analysis via large matrix operations

Andrey Shabalin

Department of Biostatistics
University of North Carolina at Chapel Hill

Main Features

- Designed for eQTL analysis of large datasets.
- Performs testing for each transcript-SNP pair.
- Ultra fast without loss of precision.
- Supports variety of linear models including
    - Least squares and ANOVA models.
    - Covariates to account for sex, population structure, surrogate variables.
    - Correlated and heteroskedastic errors.
    - Correction for multiple testing using FDR.
    - Separate FDR for cis- and trans- eQTLs.
- Matlab and R implementations.

Performance

Time of complete eQTL analysis of cystic fibrosis dataset

- 573,337 SNPs
- 22,011 transcripts
- 840 samples

Method \ # of Covariates	Zero	Ten	
Plink	8.7	146.9	days
Merlin	19.5	20.6	days
R/qtl	21.4	73.0	days
snpMatrix	2.8	5.3	days
eMap	17.8	N/A	days
FastMap	9.5	N/A	**hours**
Matrix eQTL (Matlab)	11.5	11.5	**minutes**
Matrix eQTL (Rev R)	14.4	14.4	**minutes**

Benefits

Matrix eQTL allows to:

- Perform eQTL analysis without a computing cluster.
- Study how outcome of analysis depends on
    - Model selection (Linear vs. ANOVA).
    - Inclusion of various covariates.
    - Preprocessing and normalization.
    - QC procedures (sample filtering).
- Select the right model for eQTL analysis.