Association testing with GENESIS

Stephanie Gogarten

July 19, 2018

Essential R objects

Object	Package	Description
AnnotatedDataFrame	Biobase	R data.frame with metadata describing each column
SeqVarGDSClass	SeqArray	Object providing access to a GDS file
SeqVarData	SeqVarTools	Object linking a GDS file to sample annotation
SeqVarIterator	SeqVarTools	Object to return next set of variants with each iteration
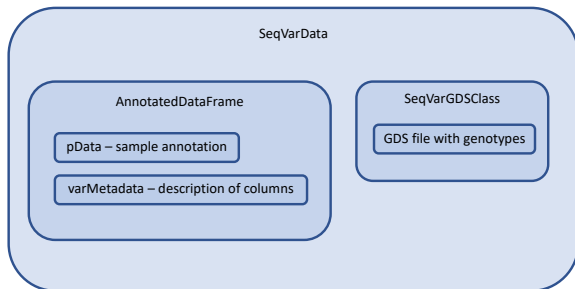
Iterator classes

Iterators extend SeqVarData, so all methods defined in SeqVarData and SeqArray can be used on them.

| Object	Description
SeqVarBlockIterator	Iterate over blocks of variants
SeqVarRangeIterator	Iterate over ranges (chromosome, start, end)
SeqVarWindowIterator	Iterate over sliding windows
SeqVarListIterator	Iterate over lists of variants

Using iterators - an example

- Store phenotype and covariates in a SeqVarData object
- Use seqSetFilter to select only rare variants
- Group variants by gene using a SeqVarListIterator
- Every call to iterateFilter will return the next gene
- Run a burden test or SKAT on all variants in this gene

GENESIS key functions

Function	Description
fitNullModel	Fit a null mixed model with fixed effects and one or more random effects (e.g., kinship, GRM)
assocTestSingle	Single-variant test using null mixed model
assocTestAggregate	Rare variant test (burden or SKAT)

Association testing flow chart
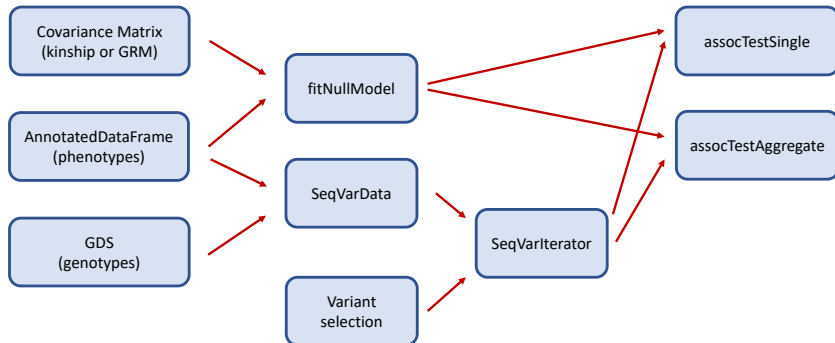
Fitting the null model

When combining samples from groups with different variances for a trait (e.g., study or ancestry group), it is recommended to allow the null model to fit heterogeneous variances by group using the parameter `group.var`.

1. Fit null mixed model including covariates (as fixed effects) and kinship/GRM (as random effect)
2. For each group separately:
   2.1 Inverse normal transform marginal residuals
   2.2 Rescale variance to match original
3. For all samples together:
   3.1 Fit null mixed model using transformed residuals as outcome
   3.2 Allow heterogeneous variance by `group.var`
   3.3 Include covariates as fixed effects
   3.4 Include kinship/GRM as random effect

The analysis pipeline implements this procedure.