

# R packages for kinship and population structure

Stephanie Gogarten

August 7, 2017

# GENESIS overview

Author: Matthew Conomos

Iterative process for separating recent familial relatedness from distant genetic ancestry

1. KING-robust<sup>1</sup> to get initial kinship estimates
2. PC-AiR<sup>2</sup> to select an informative set of unrelated samples, do PCA on unrelated, project into relatives
3. PC-Relate<sup>3</sup> to estimate kinship coefficients adjusted for population structure and admixture using PCs

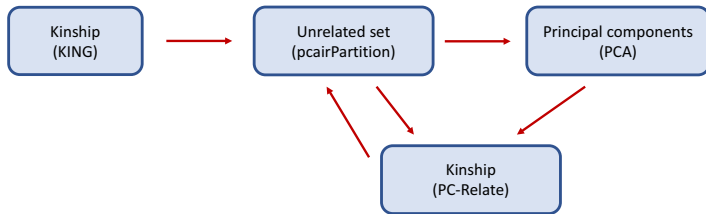
---

<sup>1</sup>Manichaikul A., Mychaleckyj J.C., Rich S.S., Daly K. Sale M., Chen W.M. (2010). *Robust relationship inference in genome-wide association studies*. *Bioinformatics*, 26(22): 2867-2873.

<sup>2</sup>Conomos, M.P., Miller, M.B., and Thornton, T.A. (2015). *Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness*. *Genetic Epidemiology* 39, 276-293.

<sup>3</sup>Conomos M.P., Reiner A.P., Weir B.S., and Thornton T.A. (2016), *Model-free estimation of recent genetic relatedness*. *American Journal of Human Genetics*, 98(1): 127-148.

## PC-Relate flow chart



## GENESIS key functions

---

Function	Description
<code>pcairPartition</code>	Select a set of unrelated samples, maximizing ancestral diversity
<code>pcair</code>	Run <code>pcairPartition</code> followed by PCA on unrelated samples and projection of relatives
<code>pcrelate</code>	Estimate kinship coefficients, adjusting for PCs
<code>pcrelateMakeGRM</code>	Return a Genetic Relatedness Matrix from <code>pcrelate</code> results
<code>pcrelateReadKinship</code>	Return a table of pairwise kinship from <code>pcrelate</code> results

---

# SNPRelate overview

Author: Xiuwen Zheng

Functions for LD pruning, Principal Component Analysis, Identity by Descent

TOPMed pipeline uses `pcairPartition` and `SNPRelate` functions for PCA in order to prune variants for LD in between selecting unrelated samples and running PCA.

## SNPRelate key functions

---

Function	Description
snpGdsLDpruning	Prune variants in linkage disequilibrium using a sliding window
snpGdsIBDKING	Kinship estimates using KING-robust
snpGdsIBDSelection	Return a table of pairwise kinship
snpGdsPCA	Principal Component Analysis
snpGdsPCASNPLoading	Calculate variant eigenvectors from PCA results
snpGdsPCASampLoading	Project samples onto eigenvectors from snpGdsPCASNPLoading

---

## Essential R objects

---

Object	Package	Description
AnnotatedDataFrame	Biobase	R data.frame with metadata describing each column
SeqVarGDSCClass	SeqArray	Object providing access to a GDS file
SeqVarData	SeqVarTools	Object linking a GDS file to sample annotation

---

