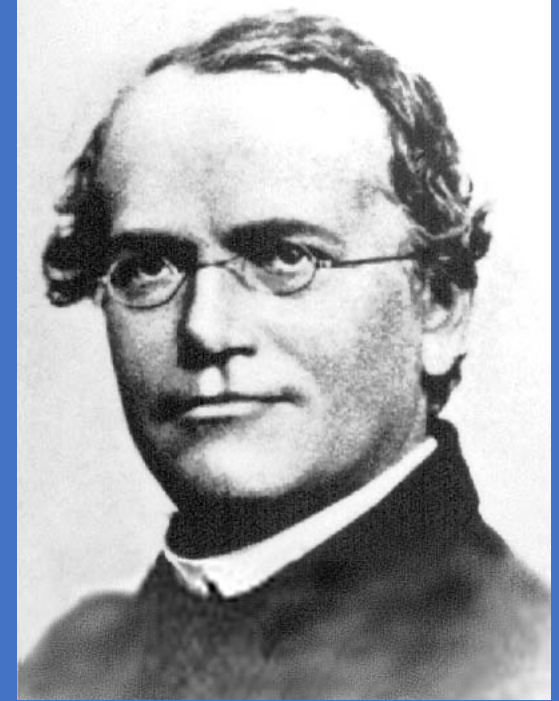


# Two stories using OpenMendel

*OpenMendel Workshop  
ASHG Annual Meeting 2020*



Jin J. Zhou PhD

Department of Epidemiology and Biostatistics

University of Arizona

# *SNPArrays.jl*

*- GRM estimation incorporating ANCESTRY*

- After diabetes diagnosis, whether patients develop different types of vascular complications is due to genetics? If so, how much is due to genetics?
- We used the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial data to answer this question.
- Estimating genetic relationship matrix is the first step.

- [MendelImpute.jl](#)
- [VCFTools.jl](#)
- [OrdinalGWAS.jl](#)
- [WiSER.jl](#)
- [MendelKinship.jl](#)
- [MendelGeneticCounseling.jl](#)
- [SnpArrays.jl](#)**
- [ADMIXTURE.jl](#)
- [Tutorials](#)
- [MendelPlots.jl](#)
- [TraitSimulation.jl](#)
- [MendelIHT.jl](#)
- [LangeSymposium-Programming...](#)
- [VarianceComponentModels.jl](#)
- [MendelGWAS.jl](#)
- [MendelBase.jl](#)
- [GeneticCounseling\\_ASHG2019](#)
- [OrdinalMultinomialModels.jl](#)
- [OpenMendel.github.io](#)
- [MendelSearch.jl](#)
- [MendelTwoPointLinkage.jl](#)
- [MendelLocationScores.jl](#)
- [MendelGeneDropping.jl](#)
- [MendelGameteCompetition.jl](#)
- [MendelEstimateFrequencies.jl](#)

# *SnpArrays.jl*

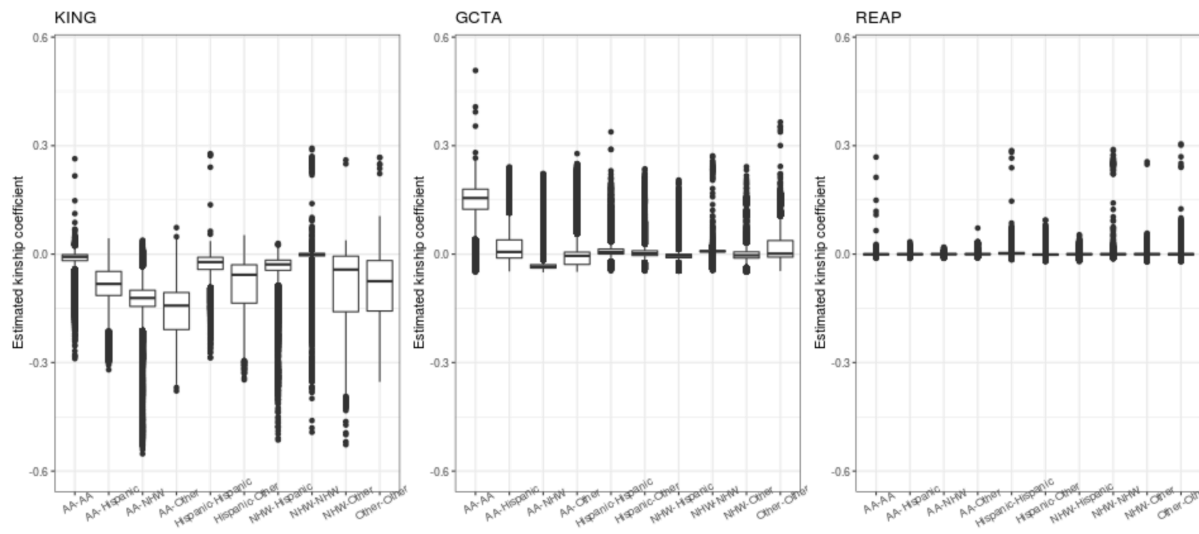
- GRM estimation incorporating ANCESTRY

REAP: <http://faculty.washington.edu/tathornt/software/REAP/index.html>

GCTA: <https://cnsgenomics.com/software/gcta/>

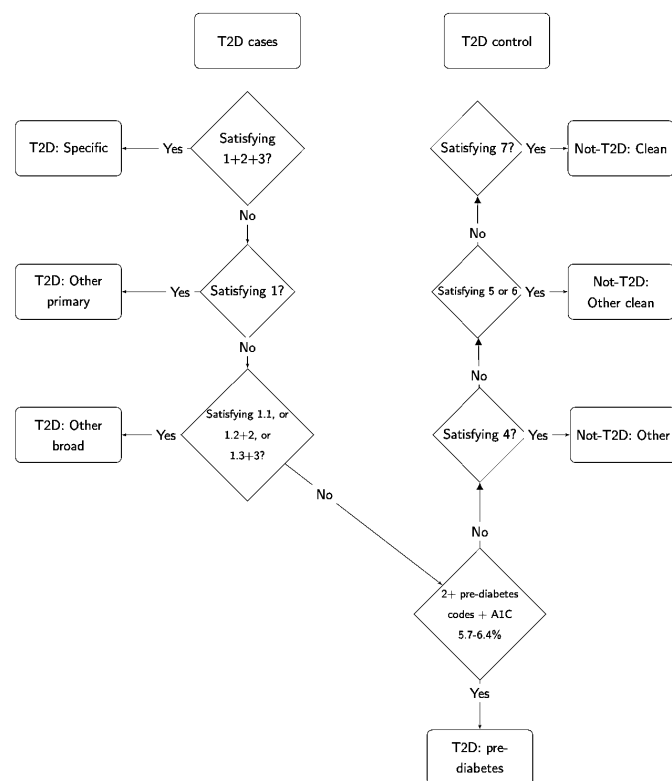
KING: <http://people.virginia.edu/~wc9c/KING/manual.html>

*SnpArrays.jl*: <https://github.com/OpenMendel/SnpArrays.jl>



- KING GRM is robust but was not positive semi-definite for down stream analysis.
- GCTA GRM removed >5000 individuals out of 6291 individuals.
- REAP took more than **6-7 hours**.
- *SnpArrays.jl* implementation took **around 6 minutes**.
- Filtering: *SnpArrays.jl* GRM 53 individuals were removed.
- **Watch Dr. Hua Zhou's *SnpArrays.jl* presentation for more details.**

# Ordinal Multi-categorical GWAS



- Disease phenotyping can be complex using data from EHR, e.g. T2D definition (left).
- It can generate multiple categories depending on the criterion used.
- There is **NO** ordinal multi-categorical GWAS tool that is available for biobank scale data.
- The current version of the US Department of Veterans Affairs' Million Veteran Program (MVP) contains genome-wide association study (GWAS) data > 800,000 veterans.
- Other large-scale biobanks include UK Biobank (> 500,000 individuals), BioVU (> 250,000), Kadoorie (> 500,000).

# OrdinalGWAS.jl

– Motivating Example: Hypertension GWAS in UKB

- Hypertension GWAS on a UK Biobank sample (~185,000 individuals and 460,000 non imputed SNPs) took **181 minutes (~6 hours)** on a standard PC
- When outcome reduces to 2 categories, *OrdinalGWAS* is case-control GWAS
- *Plink1.9* took **1999 minutes (~33 hours)** for 2 categories, i.e., case-control, analysis on this dataset
- <https://github.com/OpenMendel/OrdinalGWAS.jl>
- Watch Chris German's *Ordinal GWAS* section for details

- MendelImpute.jl
- VCFTools.jl
- OrdinalGWAS.jl**
- WiSER.jl
- MendelKinship.jl
- MendelGeneticCounseling.jl
- SnArrays.jl
- ADMIXTURE.jl
- Tutorials
- MendelPlots.jl
- TraitSimulation.jl
- MendelIHT.jl
- LangeSymposium-Programming...
- VarianceComponentModels.jl
- MendelGWAS.jl
- MendelBase.jl
- GeneticCounseling\_ASHG2019
- OrdinalMultinomialModels.jl
- OpenMendel.github.io
- MendelSearch.jl
- MendelTwoPointLinkage.jl
- MendelLocationScores.jl
- MendelGeneDropping.jl
- MendelGameteCompetition.jl
- MendelEstimateFrequencies.jl