

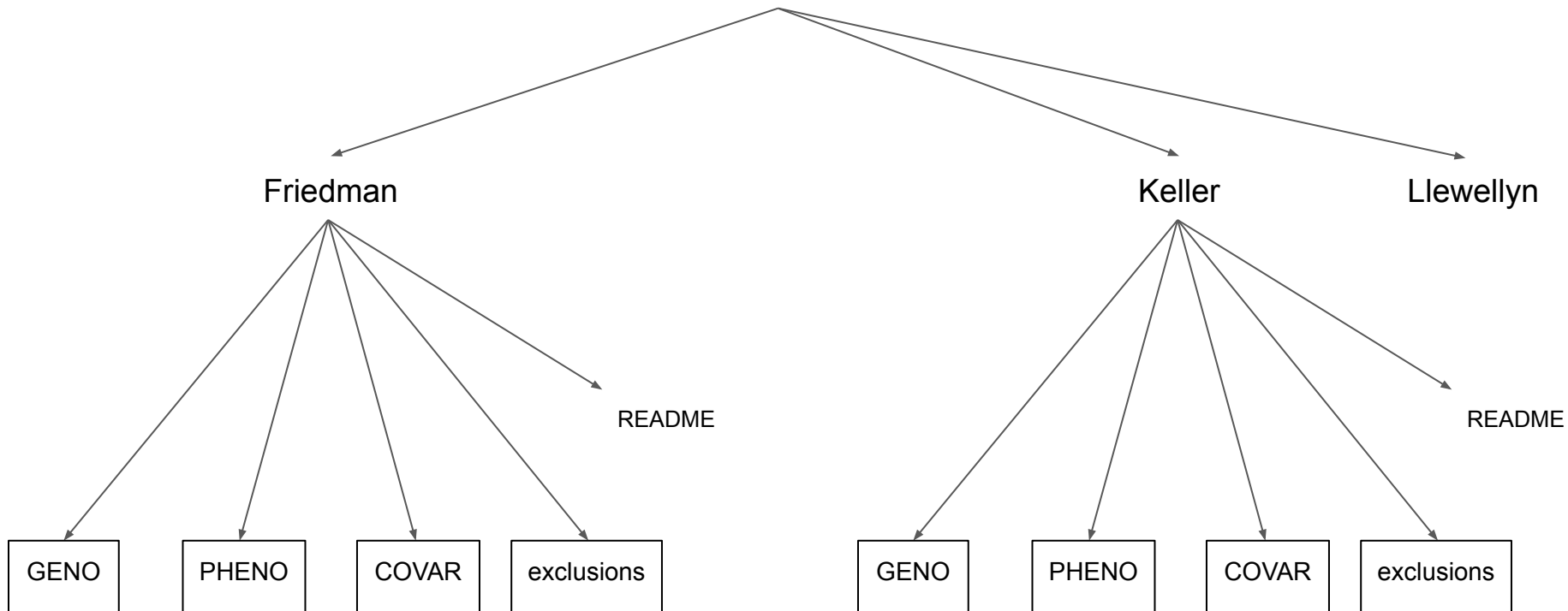
UKB data: proposed scheme

Katerina Zorina-Lichtenwalter
16 December 2020

Outline of steps

1. Present overall structure for /pl/active/IBG/UKBiobank; get input
2. Talk about general purpose scripts
3. Discuss removing v2 genotyping data
4. Create hard symbolic links in /pl/active/IBG/UKBiobank
5. Solicit scripts/general purpose files (within PI group) from others
6. Set a date for removing files from their original locations

/pl/active/IBG/UKBiobank



Genotypes

/pl/active/IBG/UKBiobank/[Friedman,Keller]

- GENO
 - v3
 - imputed
 - raw: bgen/bgi and other related files downloaded from UKB (symb. link to /pl/active/IBG/UKBiobank/Release3/EGAD00010001474 -- which will be invisible to users); sample
 - filters_QC
 - all_indiv
 - black
 - chinese
 - s_asn
 - white
 - genotyped
 - raw bed/bim downloaded from UKB (symb. link to /pl/active/IBG/UKBiobank/Release3/EGAD00010001497 -- which will be invisible to users); fam
 - filters_QC
 - all_indiv
 - black
 - chinese
 - s_asn
 - white
 - v2 (same structure as v3)

Phenotypes

/pl/active/IBG/UKBiobank/Friedman

- PHENO

- Friedman-11117
- Friedman-20707
- Friedman-20708
- Friedman-20709
- Friedman-24203
- Friedman-24204
- Friedman-24205
- Friedman-2479
- Friedman-27303
- Friedman-27304
- Friedman-27305
- Friedman-41517

/pl/active/IBG/UKBiobank/Keller

- PHENO

- derived
- raw
 - Keller-10341
 - Keller-1665

Covariates

/pl/active/IBG/UKBiobank/Friedman

- COVAR

- all_indiv
- black
- chinese
- s_asn
- white

/pl/active/IBG/UKBiobank/Keller

- COVAR

- all_indiv
- black
- chinese
- s_asn
- white

Exclusion lists

/pl/active/IBG/UKBiobank/Friedman

- Exclusions

- redactedIDs
- sex_mismatch?
- Other?

/pl/active/IBG/UKBiobank/Keller

- Exclusions

- redactedIDs
- sex_mismatch?
- Other?

READMEs

1 for /pl/active/IBG/UKBiobank/Friedman and 1 for /pl/active/IBG/UKBiobank/Keller

Description of org. scheme and list of all steps taken to move around files

General purpose scripts (parameterised) and files

QC for geno files

Relatedness filter

Genomic relationship matrix (GRM)

Analyses scripts (directory to be added in /pl/active/IBG)

GWAS

-BOLT

-SAIGE

-Regenie

-Plink?

LDSC

GCTA

GenomicSEM

??

Delete v2 geno data?

/pl/active/IBG/UKBiobank/EGAD00010001225

/pl/active/IBG/UKBiobank/EGAD00010001226

/pl/active/IBG/UKBiobank

