FinnGen & Pan-UKBB meta-analysis process

Arto Lehisto
16.12.2020

Outline

- FinnGen & Pan-UKBB phenotype matching
- Meta-analysis
- Result analysis process
- Pan-UKBB integration to FG Pheweb

Pan-UK Biobank (Pan-UKBB)

- 0.5M individuals, Genetic & phenotypic data
- 6 different populations
- 7,221 phenotypes
- Total 16,119 GWAS scans
- Openly available

https://pan.ukbb.broadinstitute.org/img/pan_ukbb_logo_v0.4-2.svg

Phenotype matching

Data:

- FinnGen endpoint definitions (R6)
  - Mapped to ICD10 manually by Aki Havulinna et al.
  - Only considers associated ICD10 codes from hospital discharge registry, cause of death registry, cancer registry
- UKBB
  - Pan-ukbb European subset (pan.ukbb.broadinstitute.org)
 - Phecode + ICD10 endpoints
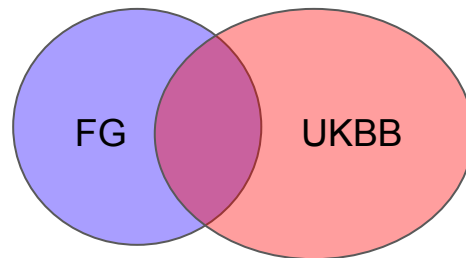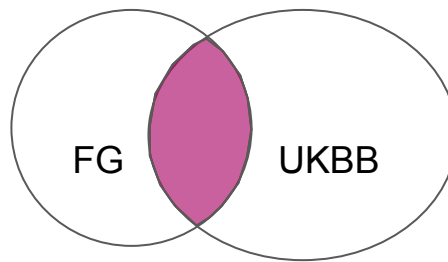
Matching:

- Match using ICD10 codes as an intermediary
- Similarity score: How much overlap is there between phenotypes when using ICD10 as an intermediary

Score = $(P1 \cap P2)/(P1 \cup P2)$

Phenotype matching results

- Approx. 25% of FinnGen phenotypes had a "complete" Pan-UKBB match
- 408 FinnGen phenotypes were excluded due to not having any associated ICD10 codes
- More information in https://github.com/FINNGEN/meta-analysis-runs

Phenotype match scores
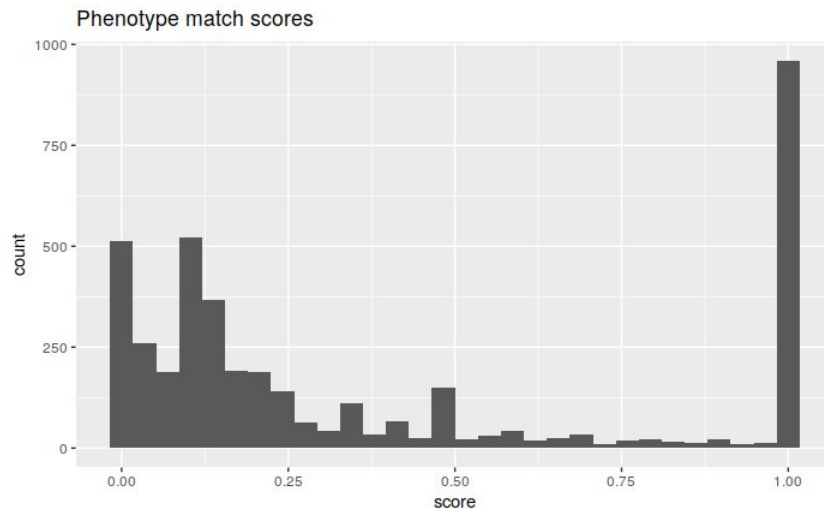
Match quality	No match	Partial match	Full match	Total
# of phenos	459	2709	956	4124

Meta-analysis

- Data preprocessing
  - UKBB variants aligned to FinnGen after UKBB liftover to hg 38.
    - Chr, pos, ref, alt matching
  - In case of no exact match (ref/alt differ), match by flipping strand and/or switch ref <--> alt.
- Inverse-variance weighted meta-analysis
  - Heterogeneity assessed with Cochran's Q
  - https://github.com/FINNGEN/META_ANALYSIS
- 860 phenotypes analysed

Meta-analysis autoreporting results

- Autoreporting ran for all 860 meta-analysed phenotypes
  - Results grouped using LD clumping with Finnish LD panel, annotated using gnomAD, compared against GWAS Catalog
- Results available in gs://r6_data_green/ukb_meta/autoreporting_results/
- Caveats:
  - Finnish LD panel does not represent both populations
  - UKBB-only variants appear as single variant groups in data, since LD can't be calculated for them

Pan-UKBB data to pheweb browser

- Pan-UKBB results to replace current Neale lab UKBB summary statistic data
- Main motivation: Replacing old linear model gwas scans with SAIGE scans
- TBD

Lead variants

chr	pos	ref	alt	locus	rsid	nearest gene	consequence	INFO	FIN enrichment	af	af cases	af controls	OR	p-value	UKBB
10	112994312	T	C	locus	rs34872471	TCF7L2	intron	0.994	0.675	0.202	0.234	0.196	1.33	1.8e-90	▲ 4.8e-6
12	4275678	T	G	locus	rs76895963	CCND2	intron	0.919	1.39	0.0311	0.0227	0.0325	0.65	1.3e-33	▼ 2.7e-3
20	44189982	G	T	locus	rs117361510	JPH2	upstream gene	0.979	2.12	0.0120	0.0175	0.0110	1.76	2.1e-27	▲ 3.5e-1
6	32742630	T	C	locus	rs115018313	HLA-DQA2	intron	0.981	2.33	0.0511	0.0631	0.0489	1.33	4.2e-27	▲ 8.2e-4
16	53765595	G	A	locus	rs9937053	FTO	intron	0.999	0.974	0.429	0.451	0.425	1.13	7.6e-26	▲ 1.9e-3
6	20675561	C	A	locus	rs35261542	CDKAL1	intron	1.00	1.25	0.329	0.348	0.326	1.13	1.7e-23	▲ 1.2e-3
11	2837210	G	A	locus	rs2237896	KCNQ1	intron	0.987	1.97	0.0756	0.0655	0.0774	0.81	4.5e-22	NA
11	92975544	C	G	locus	rs10830963	MTNR1B	intron	0.998	1.31	0.357	0.373	0.354	1.12	3.9e-21	▲ 3.3e-1
10	69554950	G	C	locus	rs537982442	NEUROG3	intergenic	0.984	22.6	0.0434	0.0514	0.0420	1.29	1.2e-20	NA
X	153634467	A	G	locus	rs5945326	DUSP9	intergenic	0.999	1.27	0.275	0.255	0.278	0.91	2.4e-18	▼ 3.5e-3

Previous | Page 1 of 7 | 10 rows | Next

Resources

- Phenotype matching
  - Writeup: https://github.com/FINNGEN/meta-analysis-runs/tree/main/PHENOMAP_R6
  - Results: https://github.com/FINNGEN/pan-ukbb-mapping
- Meta-analysis
  - Writeup: https://github.com/FINNGEN/meta-analysis-runs
  - Results:
    gs://finngen-production-library-green/finngen_R6/finngen_R6_analysis_data/ukbb_meta
- Meta-analysis autoreporting
  - Results: gs://r6_data_green/ukb_meta/autoreporting_results

Acknowledgements

UKBB meta-analysis

Juha Mehtonen
FinnGen Analysis team

FinnGen endpoint definitions:

Aki Havulinna
Tuomo Kiiskinen
FinnGen Registry team
FinnGen clinical expert groups

Phenotype matching algorithm:

Tuomo Kiiskinen

Pan-UKBB data:

Team Pan-UKBB

References

Pan-UK Biobank https://pan.ukbb.broadinstitute.org/

Questions?