



FAQ – UK BIOBANK GWAS

Q: How do I cite these results?

A: For the round 2 results [released 1st August 2018], please cite the the page <http://www.nealelab.is/uk-biobank/>

For the round 1 results [released 20th September 2017], please cite the blogpost (<http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank>).

Q: What is the data use policy of the GWAS results? Do I need permission to download and explore them?

A: No, they may be freely downloaded and used without restriction.

Q: Are there restrictions on commercial use of the results?

A: No, these results are publicly available with no restrictions on their use.

Q: How was phenotype "insert-your-favourite-phenotype-here" defined?

A: Please refer to [UK Biobank phenotype descriptions](#) and the [PHESANT paper](#) for how the raw phenotype data was then processed for analysis.

See also the phenotype summary file and the column PHESANT_transformation in particular.

Q: How can I access the UK Biobank genotype data itself?

A: See UK Biobank website for instructions on how to apply to download the raw data: <http://www.ukbiobank.ac.uk/register-apply/>

**Q: Are the manifest results updated at any time?**

A: We have released two waves of results. The first was released in September 2017 and the second in August 2018.

Q: What is the list of individuals in the data set?

A: We included in our analyses 361,194 samples (194,174 females and 167,020 males). Sample IDs are specific to project applications so can't be directly shared, but we've provided a list of the genotyping plate and well codes (i.e. matching those given by `ukb_sqc_v2.txt` from UK Biobank) that correspond to our GWAS sample in files `samples.both_sexes.tsv.bgz`, `samples.female.tsv.bgz`, and `samples.male.tsv.bgz` that are available in the results dropbox.

Q: What is the list of SNPs used?

A: See the variant annotation file, `variants.tsv.bgz`, in the results manifest.

Q: How many cases and how many controls were used in each phenotype?

A: See the phenotype summary files `phenotypes.both_sexes.tsv.bgz`, `phenotypes.female.tsv.bgz`, and `phenotypes.male.tsv.bgz`, in the results manifest.

Q: Are there descriptions of the different fields included in the annotation and results files?

A: Yes, see the [README file](#) for a description of the different fields in each file.

Q: Can I download results for all phenotypes at a specific set of SNPs?

A: Unfortunately we do not have this functionality yet but are working towards making it possible. Please check back again in the near future.

**Q: Did you include age or any other covariates in your model?**

A: In the first wave of results released in September 2017 we did not include age as a covariate; our covariates were sex and PCs 1-10. In the second wave of results released in July 2018 we include as covariates age, age², inferred_sex, age * inferred_sex, age² * inferred_sex, and PCs 1-20. In the sex-specific analyses, we include age, age², and PCs 1-20 as covariates in our model.

Q: What is there relationship between n_complete_samples, AF and AC?**Why does AF not equal AC/(2*nCompleteSamples)?**

A: There are several different allele frequency values reported in the annotation and results files. Here's how they differ:

- In the variant annotation file, variants.tsv.bgz:
 - There are AF and AC fields. These values refer to the alternate allele (the allele found in the alt field) and are calculated on all 361,194 samples included in our analysis using hardcall genotype values (0/1/2 for homRef/het/homVar).
 - There are also minor_allele and minor_AF fields in the variant annotation file. These are included because the alt allele is sometimes the major allele (AF > 0.5). When alt is the major allele, minor_allele is equal to ref and minor_AF is equal to 1.0 - AF. When alt is the minor allele, minor_allele is equal to alt and minor_AF is equal to AF.
- In each GWAS results file, *.gwas.imputed_v3.*.tsv.bgz:
 - n_complete_samples is the number of samples used in the GWAS of the particular phenotype. This may be less than 361,194 (the total number of samples in our analysis) because some phenotypes are not defined for all samples.
 - AC is the alt allele count within the n_complete_samples with defined values for the phenotype (where the alt allele is defined in the alt field of the variant annotation file). This AC is calculated on



NEALE LAB

than the AC found in the variant annotation file, even if

n_complete_samples equals 361,194.

- As in the variant annotation file, minor_allele and minor_AF fields are also included. minor_AF is calculated within the n_complete_samples for each phenotype using dosage values rather than hardcalls, so like the AC field it may be slightly different than the minor_AF field in the variant annotation file, even when n_complete_samples equals 361,194.

Q: I've read all the above FAQs and still have questions. How do I contact you?

A: Please email us at nealelab.ukb@gmail.com

TWITTER INTRANET

Powered by [Squarespace](#)