**Data Resource**

**UK Biobank GWAS results**

We developed two resource-efficient tools (called [fastGWA](file:///Users/uqljian5/Documents/gcta2_doc_github/build/index.html" \l "fastGWA) and [fastGWA-GLMM](file:///Users/uqljian5/Documents/gcta2_doc_github/build/index.html" \l "fastGWA-GLMM)) for mixed model-based association analysis in large-scale data (Jiang et al. 2019 Nat Genet and Jiang et al. 2021 Preprint). We first applied fastGWA to 2,173 traits on 456,422 array-genotyped as well as 49,960 whole-exome-sequenced individuals of European ancestry in the UK Biobank (UKB) (Jiang et al. 2019 Nat Genet). We then applied fastGWA-GLMM to 2,989 binary traits on 456,348 array-genotyped individuals of European ancestry in the UKB (Jiang et al. 2021 Preprint). In addition, we have implemented the gene-based test, ACAT-V ([Liu et al, 2019](https://www.sciencedirect.com/science/article/pii/S0002929719300023)), in GCTA. We applied ACAT-V to the GWAS summary statistics of the 2,989 UKB binary traits. See below for detailed instructions to query and download the data. One can also query or visualize the GWAS summary data using the online tool for [fastGWA](http://fastgwa.info/ukbimp) and [fastGWA-GLMM](http://fastgwa.info/ukbimpbin).

* fastGWA summary statistics using the imputed data (Jiang et al. 2019 Nat Genet): 456,422 individuals of European ancestry; 8,531,416 variants (MAF > 0.01 and missingness rate < 0.1); 2,173 traits.
  + Summary table: [UKB\_impute\_v1.1.csv](file:///Users/uqljian5/Documents/gcta2_doc_github/build/res/UKB_impute_v1.1.csv)
  + Online tool: <http://fastgwa.info/ukbimp/phenotypes>
  + Linux command to download all the summary statistics (2,173 files; 454 GB in total):

mkdir ukb && cd ukb && wget http://cnsgenomics.com/software/gcta/res/UKB\_impute\_v1.1.list && wget -i UKB\_impute\_v1.1.list

* fastGWA summary statistics using the whole-exome sequence (WES) data (Jiang et al. 2019 Nat Genet): 46,191 individuals of European ancestry; 152,327 variants (MAF > 0.01 and missingness rate < 0.1); 2,048 valid traits.
  + Summary table: [UKB\_WES\_v1.1.csv](file:///Users/uqljian5/Documents/gcta2_doc_github/build/res/UKB_WES_v1.1.csv)
  + Online tool: <http://fastgwa.info/ukbwes/phenotypes>
  + Linux command to download all the summary statistics (2,048 files; 8 GB in total):

mkdir wes && cd wes && wget http://cnsgenomics.com/software/gcta/res/UKB\_WES\_v1.1.list && wget -i UKB\_WES\_v1.1.list

* fastGWA-GLMM summary statistics using the imputed data (Jiang et al. 2021 Preprint): 456,422 individuals of European ancestry; 11,842,647 variants (MAF > 0.0001 and missingness rate < 0.1); 2,989 binary traits.
  + Summary table: [UKB\_binary\_v1.1.csv](file:///Users/uqljian5/Documents/gcta2_doc_github/build/res/UKB_binary_v1.1.csv)
  + Online tool: <http://fastgwa.info/ukbimpbin/phenotypes>
  + Linux command to download all the summary statistics (2,989 files; 1.2 TB in total):

mkdir ukb\_binary && cd ukb\_binary && wget http://cnsgenomics.com/software/gcta/res/UKB\_binary\_v1.1.list && wget -i UKB\_binary\_v1.1.list

* Gene-based test (ACAT-V) summary statistics using GWAS summary data from fastGWA-GLMM of the imputed data: 456,422 individuals of European ancestry; 26,292 genes; 2,989 binary traits.
  + Summary table: [UKB\_binary\_v1.1.csv](file:///Users/uqljian5/Documents/gcta2_doc_github/build/res/UKB_binary_v1.1.csv)
  + Summary statistics (1 file; 1.2 GB): [UKB\_ACAT\_results.v1.1.tar.gz](http://link_to_be_inserted)
  + Manhattan plots (1 file; 446 MB): [ManPlot\_UKB\_ACAT\_results.v1.1.tar.gz](http://link_to_be_inserted)

**Data format**

Columns in the summary table:

ID: the trait ID

Description: trait description

Data\_type: the type of phenotype (Continuous: quantitative trait; Ordered\_Categorical: ordered categorical trait; Binary: binary trait)

Method: LR - Linear Regression; MLM - Mixed Linear Model. Note that the program will switch to use LR for analysis if the estimated genetic variance from fastGWA is not significant (p > 0.05).

N: sample size

Ncase: number of affected individuals for a binary disease trait.

Ncontrol: number of unaffected individuals for a binary disease trait.

Gender\_specific: to indicate if it is a gender-specific trait.

URL: the link to download the summary statistics.

Format of the summary statistics from fastGWA:

CHR: chromosome

SNP: SNP ID

POS: SNP position

A1: effect allele

A2: the other allele

N: per allele sample size

AF1: the allele frequency of A1

BETA: SNP effect

SE: standard error

P: p value

Format of the summary statistics from fastGWA-GLMM:

CHR: chromosome

SNP: SNP ID

POS: SNP position

A1: effect allele

A2: the other allele

N: per allele sample size

AF1: the allele frequency of A1

T: GLMM score statistic

SE\_T: standard error of the score statistic

P\_noSPA: raw p-value

BETA: SNP effect or log(odds ratio)

SE: standard error for the estimated effect size after the SPA correction

P: p-value after the SPA correction

CONVERGE: to indicate whether the SPA correction is converged for that variant

Format of the summary statistics from ACAT-V:

chr: chromosome

gene\_name: mapped gene

start: start position of the mapped gene

stop: end position of the mapped gene

snp\_num: number of SNPs from the GWAS data mapped to the gene

p\_acat\_v: ACAT-V test p-value

Note: the names of the variants were kept the same as provided (the coordinates of the variants were based on GRCh37).

**Credits and Acknowledgements**

Zhili Zheng (online tool development and data analysis), Longda Jiang (data analysis), Jian Yang (overseeing). The online tool was developed based on the source code modified from Pheweb. We thank Alibaba Cloud - Australia and New Zealand for hosting the online tool.

**Questions and Help Requests**

If you have any question, please send an email to Jian Yang [jian.yang.qt@gmail.com](mailto:jian.yang.qt@gmail.com)

**Citation**

Jiang L, Zheng Z, Qi T, Kemper KE, Wray NR, Visscher PM, Yang J (2019) A resource-efficient tool for mixed model association analysis of large-scale data. Nature Genetics, 51: 1749–1755.

Jiang L, Zheng Z, Yang J (2021) FastGWA-GLMM: a generalized linear mixed model association tool for biobank-scale data, February 2021, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-128758/v1](https://europepmc.org/article/PPR/PPR283012)

**Update log**

* Version v1.1 (14 May 2021): added GWAS summary statistics from GCTA-fastGWA-GLMM v1.93.3 for 2,989 binary traits from the UKB, and added gene-based test summary statistics from GCTA-ACAT-V for the 2,989 UKB binary traits.
* Version v1.1 (9 Aug 2019): reran all the analyses using GCTA-fastGWA v1.92.3; removed variants with MAF < 0.01 or missingness rate > 0.1 per trait, and removed binary traits with case fraction < 0.01.
* Version v1.0 (Apr 2019): first release.