**fastGWA-GLMM**

**fastGWA-GLMM: A fast GLMM-based Genome-Wide Association tool**

fastGWA-GLMM is a resource-efficient tool for generalized linear mixed model (GLMM-) based GWAS analysis for binary traits in biobank-scale data such as the UK Biobank. It uses saddle point approximation (SPA) method account for inflation in test statistics of low-frequency variants due to case-control imbalance. Credits: [Longda Jiang](mailto:longda.jiang@uq.edu.au) (method, simulation and analysis), [Zhili Zheng](mailto:zhili.zheng@uq.edu.au) (method, software and analysis) and [Jian Yang](http://researchers.uq.edu.au/researcher/2713) (method and overseeing).

We have applied fastGWA-GLMM to 2,989 binary traits on 456,348 array-genotyped and imputed individuals in the UK Biobank. All the summary statistics are available at our data portal: <http://fastgwa.info/ukbimpbin/>, where users can query, download or visualize the summary data.

**Citation**

Jiang L, Zheng Z, Yang J (2021) FastGWA-GLMM: a generalized linear mixed model association tool for biobank-scale data, 12 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)

--make-bK-sparse 0.05 --grm-sparse  
fastGWA-GLMM adopts the same sparse GRM setting as in fastGWA. Please refer to the [fastGWA webpage](https://cnsgenomics.com/software/gcta/" \l "fastGWA) for details of these two commands.

--fastGWA-mlm-binary  
Perform a GLMM-based association analysis.

--joint-covar  
By default, fastGWA-GLMM adopts an approximation method for covariate adjustment to reduce runtime. This flag forces fastGWA-GLMM to perform the exact covariate adjustment in the association analysis, which produces slightly more accurate test statistics for all the variants but takes about twice time to run, compared to the analysis with the approximate covariate-adjustment approach (see the [preprint](https://europepmc.org/article/PPR/PPR283012) for more details).

--model-only  
To perform the variance component estimation step in fastGWA-GLMM without the association test step and save the results in \*.fastGWA.mdl.id and \*.fastGWA.mdl.bin2 files. This flag can be useful if users wish to perform the association step for each chromosome individually, or to perform association test for chromosome X based on the variance components estimated from the autosomes.

--load-model  
To load a saved model (see the --model-only flag above) to perform association tests. Note that this function only works when the sample IDs in the saved model are a subset of those in genotype data. This flag also works with all the other genotype QC flags (e.g., --maf, --extract and --geno) but is incompatible with flags to input phenotype, covariate or GRM.

**Output format**test.fastGWA (columns are chromosome, SNP, SNP position, the effect allele, the other allele, per allele sample size, frequency of A1, GLMM score statistic, standard error of the score statistic, raw p-value, effect size or log(odds ratio), standard error for the estimated effect size after the SPA correction, p-value after the SPA correction, and an indicator for whether the SPA correction is converged for the variant). Note: if the input genotype data are in BGEN v1.2 format, an additional column “INFO” will be added to the output file, which indicates the imputation INFO score of the variant.

CHR SNP POS A1 A2 N AF1 T SE\_T P\_noSPA BETA SE P CONVERGE

1 rs2531267 69569 C T 454954 0.000151664 0.407195 2.47905 0.869531 0.0662569 0.40338 0.869531 1

1 rs12238997 693731 G A 416799 0.0936579 -2.08364 56.0988 0.970372 -0.000662088 0.0178257 0.970372 1

1 rs144155419 717587 A G 445565 0.0104957 16.8041 20.4031 0.410163 0.0403668 0.0490122 0.410163 1

1 rs189787166 723329 T A 454613 0.001322 -12.1542 7.40537 0.100741 -0.221632 0.135037 0.100741 1

1 rs148120343 730087 C T 431547 0.043979 22.1704 40.2519 0.581776 0.0136837 0.0248436 0.581776 1

...

Examples

# Generate a sparse GRM from SNP data

# geno\_chrs.txt is a text file containing file paths to the SNP data of each chromosome

gcta64 --mbfile geno\_chrs.txt --make-grm --thread-num 10 --out geno\_grm

gcta64 --grm geno\_grm --make-bK-sparse 0.05 --out sp\_grm

# We may run the two steps above by one command if you do not have enough disk space to store the full dense GRM

gcta64 --mbfile geno\_chrs.txt --make-grm --sparse-cutoff 0.05 --threads 10 --out sp\_grm

# fastGWA GLMM analysis (based on the sparse GRM generated above)

gcta64 --mbfile geno\_chrs.txt --grm-sparse sp\_grm --fastGWA-mlm-binary --pheno phenotype.txt --qcovar pc.txt --covar fixed.txt --threads 10 --out geno\_assoc

# fastGWA GLMM analysis without using the approximate covariate adjustment approach

gcta64 --mbfile geno\_chrs.txt --grm-sparse sp\_grm --joint-covar --fastGWA-mlm-binary --pheno phenotype.txt --qcovar pc.txt --covar fixed.txt --threads 10 --out geno\_assoc

# To save the estimated fastGWA model parameters from an analysis for the autosomes and use them in a subsequent analysis for chrX

# chrX.idlist: a list of sample IDs used in the analysis for chromosome X (chrX)

gcta64 --mbfile geno\_chrs.txt --grm-sparse sp\_grm --fastGWA-mlm-binary --model-only --pheno phenotype.txt --qcovar pc.txt --covar fixed.txt --keep chrX.idlist --threads 10 --out geno\_assoc

# To load the saved model above to run association tests for ChrX

# chr.snplist: a list of variants on chrX to be included in this analysis

gcta64 --bfile test\_chrX --load-model geno\_assoc\_mdl.fastGWA --extract chr.snplist --geno 0.1 --out test\_chrX\_assoc --threads 10