**fastGWA-BB**

**fastGWA-BB: fastGWA-GLMM burden test for binary traits**

fastGWA-BB (fastGWA-GLMM-Burden) is an extension of the single-variant association test in fastGWA-GLMM to assess the association of a weighted rare allele count of all the tested rare variants in or around a gene with a binary trait of interest, conditioning on the sparse genetic relationship matrix. The variants weighting method follows that in SAIGE-Burden (Zhou et al. 2020 Nat Genet). Credits: [Zhili Zheng](mailto:zhili.zheng@uq.edu.au), [Longda Jiang](mailto:longda.jiang@uq.edu.au), and [Jian Yang](https://scholar.google.com/citations?user=aLuqQs8AAAAJ&hl=en).

**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)

--burden  
To perform a fastGWA-BB analysis using the estimated model components from the parameter estimation step of [fastGWA-GLMM](https://yanglab.westlake.edu.cn/software/gcta/index.html" \l "fastGWA-GLMM).

--load-model  
To load a saved model from the parameter estimation step of fastGWA-GLMM to perform the fastGWA-BB analysis (see [fastGWA-GLMM page](https://yanglab.westlake.edu.cn/software/gcta/index.html#fastGWA-GLMM) for details).

--set-list gene\_list.txt  
Input a gene list with gene start and end positions.

Input file format  
gene\_list.txt (columns are gene ID, chromosome, left- and right-side boundary of the gene region, and strand)

gene chr start end strand

DDX11L1 1 11873 14409 +

WASH7P 1 14361 29370 +

MIR6859-1 1 17368 17436 +

......

Please click the link below to download the gene list file, which is slightly different from the format used by [ACAT-V](https://yanglab.westlake.edu.cn/software/gcta/index.html#ACAT-V)

Gene list (hg19): [hg19-gene.list](https://yanglab.westlake.edu.cn/software/gcta/res/hg19-gene.list)

--max-maf 0.01  
The maximum minor allele frequency (MAF) allowed for a variant to be included in the burden test. Any variant with MAF larger than this value will be excluded.

Examples

# First, to obtain the estimated model parameters from the parameter estimation step of fastGWA-GLMM

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

# Then, to run a fastGWA-BB test for rare variants

gcta64 --burden --bfile geno\_chr1 --load-model geno\_assoc\_mdl.fastGWA --set-list hg19-gene.list --extract rare.snplist --max-maf 0.01 --maf 1e-7 --out burden\_test\_assoc\_chr1

Output file format (columns are gene ID, left- and right-side boundary of the gene region, number of qualified variants in the gene region, average MAF of the qualified variants, fastGWA-BB 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 gene)

GENE START END VAR\_N MAF\_MEAN T SE\_T P\_noSPA BETA SE P CONVERGE

SAMD11 861120 879961 51 0.0039 -4718.44 6418.35 0.462249 -0.00011453 0.000155803 0.462249 1

PLEKHN1 901876 910484 24 0.0023 -342.052 1266.68 0.787131 -0.00021318 0.000789464 0.787131 1

......