**Overview**

**About**

GCTA (Genome-wide Complex Trait Analysis) is a software package, which was initially developed to estimate the proportion of phenotypic variance explained by all genome-wide SNPs for a complex trait but has been extensively extended for many other analyses of data from genome-wide association studies (GWASs). GCTA currently supports the following analyses.

**Heritability, genetic correlation, and phenotype prediction**

* GRM: estimating genetic relationships among individuals from SNP data;
* Estimating the inbreeding coefficients of individuals in GWAS data;
* GREML: estimating the proportion of variance in a phenotype explained by all SNPs (i.e. the SNP-based heritability);
* Partitioning genetic variance into contributions from different sets of SNPs stratified by chromosome location, allele frequency, or functional annotation;
* Estimating the genetic variance attributed to the X chromosome, and testing for the effect of dosage compensation;
* GREMLd: estimating dominance variance in unrelated individuals using GWAS data;
* Bivariate GREML: estimating the genetic correlation between two traits (diseases) using GWAS data;
* Haseman-Elston regression to estimate SNP-based heritability for a trait and genetic correlation between two traits;
* sBLUP: summary-data based BLUP analysis for genomic risk prediction;

**Genome-wide association analysis**

* fastGWA: ultra-fast (mixed) linear model association analysis using a sparse GRM.
* fastGWA-GLMM: ultra-fast generalized linear mixed model-based association analysis for binary traits using a sparse GRM.
* MLMA and MLMA-LOCO: mixed linear model association analysis using a dense GRM;
* COJO: conditional & joint association analysis using GWAS summary statistics;
* mtCOJO: multi-trait-based conditional & joint association analysis using GWAS summary statistics;
* fastBAT: a gene- or set-based association test using GWAS summary statistics;
* fastGWA-BB: fastGWA-GLMM burden test;
* ACAT-V: a fast Cauchy p-value combination test for the aggregate effect of multiple rare variants.

**GWAS simulation, population genetics, and Mendelian randomisaion**

* Simulating a phenotype based on GWAS data;
* GSMR: generalised summary-data-based Mendelian randomisaion;
* PCA analysis and estimation of Fst in GWAS data;
* Estimating inbreeding coefficients of individuals from SNP data;
* Computing LD scores and searching for LD friends for a list of target SNPs;

**Latest release**[**v1.93.3beta**](https://yanglab.westlake.edu.cn/software/gcta/index.html#Download)**, click to download or view update log (1 June 2021)**

**Credits**

[Jian Yang](http://scholar.google.com.au/citations?user=aLuqQs8AAAAJ&hl=en) developed the original version of the software (with supports from [Peter Visscher](mailto:peter.visscher@uq.edu.au), [Mike Goddard](mailto:Mike.Goddard@dpi.vic.gov.au) and [Hong Lee](mailto:hong.lee@uq.edu.au)) and currently maintains the software.

[Zhili Zheng](mailto:zhili.zheng@outlook.com) programmed the fastGWA, fastGWA-GLMM and fastGWA-BB modules, rewrote the I/O and GRM modules, improved the GREML and bivariate GREML modules, extended the PCA module, and improved the SBLUP module.

[Zhihong Zhu](mailto:z.zhu1@uq.edu.au) programmed the mtCOJO and GSMR modules and improved the COJO module.

[Longda Jiang](https://yanglab.westlake.edu.cn/software/gcta/longda.jiang@ue.edu.au) and [Hailing Fang](https://yanglab.westlake.edu.cn/software/gcta/fanghailing@westlake.edu.cn) developed the ACAT-V module.

[Jian Zeng](mailto:j.zeng@imb.uq.edu.au) rewrote the GCTA-HEreg module.

[Andrew Bakshi](mailto:andrew.bakshi@gmail.com) contributed to the GCTA-fastBAT module.

[Angli Xue](https://yanglab.westlake.edu.cn/software/gcta/a.xue@imb.uq.edu.au) improved the GSMR module.

[Robert Maier](mailto:rmaier@broadinstitute.org) improved the GCTA-SBLUP module.

Contributions to the development of methods included in GCTA (e.g., GREML methods, COJO, mtCOJO, MLMA-LOCO, fastBAT, fastGWA and fastGWA-GLMM) can be found in the papers cited in the corresponding web pages.

**Questions and Help Requests**

If you have any bug reports or questions please send an email to [Jian Yang](http://scholar.google.com.au/citations?user=aLuqQs8AAAAJ&hl=en) at [jian.yang@westlake.edu.cn](mailto:jian.yang@westlake.edu.cn).

**Citations**

**GCTA Software tool:**  
Yang et al. (2011) GCTA: a tool for Genome-wide Complex Trait Analysis. Am J Hum Genet. 88(1): 76-82. [[PubMed ID: 21167468](http://www.ncbi.nlm.nih.gov/pubmed/21167468)]

**Method for estimating the variance explained by all SNPs (GREML method) with its application in human height:**  
Yang et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 42(7): 565-9. [[PubMed ID: 20562875](http://www.ncbi.nlm.nih.gov/pubmed/20562875)]

**GREML method being extended for case-control design with its application to the WTCCC data:**  
Lee et al. (2011) Estimating Missing Heritability for Disease from Genome-wide Association Studies. Am J Hum Genet. 88(3): 294-305. [[PubMed ID: 21376301](http://www.ncbi.nlm.nih.gov/pubmed?term=Estimating%20Missing%20Heritability%20for%20Disease%20from%20Genome-wide%20Association%20Studies)]

**Extension of GREML method to partition the genetic variance into individual chromosomes and genomic segments with its applications in height, BMI, vWF and QT interval:**  
Yang et al. (2011) Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet. 43(6): 519-525. [[PubMed ID: 21552263](http://www.ncbi.nlm.nih.gov/pubmed/21552263)]

**Method for conditional and joint analysis using summary statistics from GWAS with its application to the GIANT meta-analysis data for height and BMI:**  
Yang et al. (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 44(4):369-375. [[PubMed ID: 22426310](http://www.ncbi.nlm.nih.gov/pubmed/22426310)]

**Bivariate GREML method:**  
Lee et al. (2012) Estimation of pleiotropy between complex diseases using SNP-derived genomic relationships and restricted maximum likelihood. Bioinformatics. 28(19): 2540-2542. [[PubMed ID: 22843982](http://www.ncbi.nlm.nih.gov/pubmed/22843982)]

**Mixed linear model based association analysis:**  
Yang et al. (2014) Mixed model association methods: advantages and pitfalls. Nat Genet. 2014 Feb;46(2):100-6. [[Pubmed ID: 24473328](http://www.ncbi.nlm.nih.gov/pubmed/24473328)]

**GREML-LDMS method and LD-score calculation:**  
Yang et al. (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 47(10):1114-20.[[PMID: 26323059](http://www.nature.com/ng/journal/vaop/ncurrent/full/ng.3390.html)]

**Method to search for LD friends:**  
Yang et al. (2011) Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. 19(7): 807-812. [[Pubmed ID: 21407268](http://www.nature.com/ejhg/journal/v19/n7/full/ejhg201139a.html)]

**fastBAT method:**  
Bakshi et al. (2016) Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. Scientific Reports 6, 32894. [[PMID: 27604177](https://www.nature.com/articles/srep32894)]

**mtCOJO and GSMR methods:**  
Zhu et al. (2018) Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat Commun. 9, 224.[[PMID: 29335400](https://www.ncbi.nlm.nih.gov/pubmed/?term=29335400)]

**fastGWA method:**  
Jiang et al. (2019) A resource-efficient tool for mixed model association analysis of large-scale data. Nat Genet. 51, 1749–1755. [[PMID: 31768069](https://pubmed.ncbi.nlm.nih.gov/31768069/)]

**fastGWA-GLMM and fastGWA-BB methods:** Jiang et al. (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>

**Last update: 1 June 2021**