G3, 2022, **12(4),** jkac049
https://doi.org/10.1093/g3journal/jkac049
Advance Access Publication Date: 24 February 2022
Software and Data Resources

gJLS2: an R package for generalized joint location and scale analysis in X-inclusive genome-wide association studies

Wei Q. Deng [1,2,*] Lei Sun [3,4,*]

[1]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON L8P 3R2, Canada,
[2]Peter Boris Centre for Addictions Research, St. Joseph's Healthcare Hamilton, McMaster University, Hamilton, ON L8P 3R2, Canada,
[3]Department of Statistical Sciences, University of Toronto, Toronto, ON M5G 1Z5, Canada,
[4]Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5T 3M7, Canada

*Corresponding author: Department of Psychiatry and Behavioural Neurosciences, McMaster University, 100 West 5th Street, Hamilton, ON L8P 3R2, Canada. Email: dengwq@mcmaster.ca; Corresponding author: Department of Statistical Sciences, University of Toronto, Ontario Power Building, 700 University Avenue, 9th Floor, Toronto, ON M5G 1Z5, Canada. Email: sun@utstat.toronto.edu

Abstract

A joint analysis of location and scale can be a powerful tool in genome-wide association studies to uncover previously overlooked markers that influence a quantitative trait through both mean and variance, as well as to prioritize candidates for gene–environment interactions. This approach has recently been generalized to handle related samples, dosage data, and the analytically challenging X-chromosome. We disseminate the latest advances in methodology through a user-friendly R software package with added functionalities to support genome-wide analysis on individual-level or summary-level data. The implemented R package can be called from PLINK or directly in a scripting environment, to enable a streamlined genome-wide analysis for biobank-scale data. Application results on individual-level and summary-level data highlight the advantage of the joint test to discover more genome-wide signals as compared to a location or scale test alone. We hope the availability of gJLS2 software package will encourage more scale and/or joint analyses in large-scale datasets, and promote the standardized reporting of their *P*-values to be shared with the scientific community.

Keywords: gene–environment interactions; joint location and scale; PLINK; R; variance heterogeneity; X-chromosome association

Introduction

Genetic association studies examine the relationship between the genotypes of a single-nucleotide polymorphism (SNP; denoted by *G*) and a quantitative phenotype (denoted by *Y*), by testing the mean differences in *Y* according to the genotypes, otherwise known as a location test. More recently, several reports have investigated association with phenotypic variance of complex quantitative traits (Paré *et al.*, 2010; Yang *et al.* 2012; Shungin *et al.*, 2017), by testing the variance differences in *Y* across the genotype groups of a SNP, or known as a scale test, in hopes of finding biologically meaningful markers. One possible explanation of a significant scale effect is the presence of gene–gene (G×G) or gene–environment (G×E) interactions; both referred to as G×E hereinafter. Unlike a direct test of G×E interaction, a scale test can be used to indirectly infer G×E without knowledge of the interacting covariates, thus alleviates the multiple hypothesis burden of testing all possible pairwise interactions, and the assumption that all interacting environmental variables could be (accurately) measured.

A more powerful approach to prioritize biologically relevant candidates is to jointly evaluate location and scale effects (Aschard *et al.*, 2013; Cao *et al.* 2014; Soave *et al.* 2015). As compared to other existing single-marker based joint tests, a joint location-scale (JLS) association test (Soave *et al.* 2015) is easy to implement on individual-level data or in a meta-analysis, requiring only the location and scale *P*-values for each SNP, marking the first generation of the JLS tool (https://github.com/dsoave/JLS; accessed 2022 February 13). Since methods for genome-wide association studies (GWAS) of location are well-established, the main focus was on improving scale tests tailored for genetic data. Soave and Sun (2017) generalized the well-known Levene's test to handle complex data structures often observed in genetic studies, such as correlated samples and dosage data, leading to the next update, namely, generalized JLS (gJLS; https://github.com/dsoave/gJLS; accessed 2022 February 13). Following the X-inclusive trend to genome-wide analyses, robust and powerful location (Chen *et al.* 2021) and scale tests (Deng *et al.* 2019) tailored for X-chromosome are now available.

In this study, we describe a generalized joint location and scale analysis tool (gJLS2) as an update to the JLS and gJLS methodology for autosomes, with added functionalities that (1) support X-chromosome mean and variance association analyses, (2) handle imputed data as genotypic probabilities or in dosage format, (3) allow the incorporation of summary statistics for location and/or scale tests, (4) implement a flexible framework that

Received: December 02, 2021. **Accepted:** February 17, 2022

can accommodate additional covariates in both the location and scale association models, and (5) improve the computational time required for large-scale genetic data such as the UK biobank (Bycroft *et al.* 2018). We hope the availability of this unifying software package will encourage more X-inclusive, genome-wide, gJLS2 association analysis for complex continuous traits, particularly for those believed to be enriched for genetic interactions.

Materials and methods

The software can be installed in an R environment from CRAN:

```
install.packages("gJLS2")
```

or github for the most recent version:

```
#install.packages("devtools")
devtools::install_github("WeiAkaneDeng/gJLS2")
```

via the "devtools" package (Wickham *et al.* 2018). An accompanying guide that documents each of the analytical options and scenarios is available at https://weiakanedeng.github.io/gJLS2/. Any feedbacks/bugs can be reported under github's issues tab (https://github.com/WeiAkaneDeng/gJLS2/issues).

Data preparation

The gJLS2 software package requires the minimal inputs of a quantitative trait and genotype data, which can be discrete genotype counts, continuous dosage genotype values or imputed genotype probabilities. The genotype data can be supplied in PLINK format via the R plug-in option using PLINK 1.9 (Chang *et al.* 2015) or any other format that can be read in R with packages "*BGData*" and "*BEDMatrix*" (Grueneberg and de los Campos 2019). The phenotype and covariate should be supplied in the same file, and if sex is required for X-chromosome analysis, then it needs to be the first column after individual IDs.

For smaller analyses or testing purpose, it is possible to use the package in R GUI directly. However, for genome-wide analyses on larger datasets, we recommend either Rscript commands or as an R-plugin within PLINK combined with a high-performing computing cluster environment. We will demonstrate all 3 approaches using the example datasets provided.

Example datasets

The package included 2 example datasets: one comprises of simulated phenotype, covariate data and real X-chromosome genotypes from the 1000 Genomes Project (The 1000 Genomes Project Consortium 2015), denoted by "chrXdat"; and another is based on summary statistics from the Genetic Investigation of ANthropometric Traits (GIANT) consortium for body mass index (BMI) and human standing height (Yang *et al.* 2012; Yengo *et al.* 2018).

The dataset "chrXdat" consists of 5 hand-picked SNPs rs5983012 (A/G), rs986810 (C/T), rs180495 (G/A), rs5911042 (T/C), and rs4119090 (G/A) that are outside of the pseudo-autosomal region, to cover observed minor allele frequency (MAF; calculated in females and rounded to the nearest digit) of 0.1, 0.2, 0.3, 0.4, and 0.45, respectively. See Supplementary Material Section 2 for more details on the simulated dataset. The summary statistics data comprised of association *P*-values of SNPs with the *mean* of BMI and height, under the column name "gL" for location, and those with the *variance*, under the column name "gS." The example data only included chromosome 16 SNPs to keep the size of example datasets manageable.

Location association

For autosomal SNPs, a linear regression fitted using the ordinary least square (OLS) is flexible to accommodate additional covariates and the default option. To account for related samples, the generalized least square (GLS) method is used by specifying a covariance structure for error terms in smaller samples. Users can either provide the covariance matrix or specify a structure for the covariance matrix according to predefined subgroups. However, for large population studies ($n > 5{,}000$), it is recommended to run the location association analysis using the state-of-art linear mixed models, such as LMM-BOLT (Loh *et al.* 2015) or SAIGE (Zhou *et al.* 2018) and supply the results as location summary statistics for the joint analysis.

The novel contribution of gJLS2 is the addition of our recommended X-chromosome location association strategy (Chen *et al.* 2021), which has good type I error control, is robust to sex confounding, arbitrary baseline allele choice, uncertainty of X-Chromosome Inactivation (XCI), and skewed XCI. Our chosen X-chromosome association model achieves these by simultaneously including the sex information, its interaction with both the additive genetic effect and dominance effect. The default location test *P*-value is obtained by testing the null hypothesis $H_o : \beta_G = \beta_{GS} = \beta_D = 0$ under the linear model:

$$y_i \sim \beta_0 + \mathbf{C_i}\boldsymbol{\gamma} + \beta_S S_i + \beta_A G_{\text{Additive, i}} + \beta_D G_{\text{Dominance, i}} + \beta_{GS} G_{\text{Additive, i}} \times S_i, \quad (1)$$

where G_A is the additively coded genotype variable, G_D is an indicator variable for the female heterozygotes, the sex variable S is coded with males as the baseline taking value 0 and females taking value 1, and C is a vector of covariates to be adjusted for. The regression coefficients β_0, $\boldsymbol{\gamma}$, β_S are for the nongenetic covariates in the model and β_G, β_{GS}, β_D denote the regression coefficients of interest, for the additive, GxSex, and dominance effects, respectively. The function "locReg" can be called with the X-chromosome option and returns the default 3-degree of freedom (df) test *P*-value:

```
data("chrXdat")
head(chrXdat)

> locReg(GENO=chrXdat[, 7:11], SEX=chrXdat$SEX,
Y=chrXdat$PHENOTYPE, Xchr=TRUE);
  CHR         SNP       gL
1   X   rs5983012_A 0.9877674
2   X   rs4119090_G 0.9201569
3   X   rs5911042_T 0.3898029
4   X    rs986810_C 0.4619165
5   X    rs180495_G 0.8767590
```

Although the resulting 3-df test is robust to the choice of baseline allele and status of XCI, a recent report (Pazokitoroudi *et al.* 2021) has found "limited contribution of dominance heritability to

complex trait variation" and the recommended 3-df test can only be applied to discrete genotypes. Thus, an alternative 2-df test without the dominance term is recommended and is the default option for imputed SNPs.

Scale association

The scale component builds on 2 recent works that generalized to dosage genotype, related samples (Soave and Sun 2017), and the X-chromosome (Deng *et al.* 2019). Both are extensions of the generalized Levene's test via a regression framework. Besides a more flexible characterization of sample dependence structure, the generalization also allows analysis of imputed genotype data, which would otherwise be challenging with a group-based variance heterogeneity test, such as the Levene's test or the Brown-Forsythe test.

The variance test *P*-value is obtained using a 2-stage generalized Levene's test assuming additive variance effects. Briefly, the residuals d was computed under Equation (1) using the Least Absolute Deviation (LAD) algorithm, which gives the residuals in terms of each observation's distance with respect to the median (rather than the mean as in OLS) and is the default option. The absolute residuals were then fitted under the 3-df recommended model for discrete genotype:

$$d_i \sim \beta_0 + \mathbf{C_i\gamma} + \beta_S \mathrm{Sex} + \beta_A G_{\mathrm{Additive}} + \beta_D G_{\mathrm{Dominance}} + \beta_{GS} G_{\mathrm{Additive},\, i} \times S_i. \tag{2}$$

The following examples show results from LAD and OLS algorithms are very similar when the simulated phenotype is roughly symmetric:

```
> scaleReg(GENO=chrXdat[, 7:11], SEX=chrXdat$SEX,
Y=chrXdat$PHENOTYPE, Xchr=TRUE)

  CHR           SNP       gS
1  X   rs5983012_A 0.1391909
2  X   rs4119090_G 0.9828430
3  X   rs5911042_T 0.1487017
4  X    rs986810_C 0.9563390
5  X    rs180495_G 0.3476929
> scaleReg(GENO=chrXdat[, 7:11], SEX=chrXdat$SEX,
Y=chrXdat$PHENOTYPE, Xchr=TRUE, loc_alg="OLS")
  CHR         SNP         gS
1  X   rs5983012_A 0.1739062
2  X   rs4119090_G 0.9999999
3  X   rs5911042_T 0.1163023
4  X    rs986810_C 0.9581589
5  X    rs180495_G 0.3619056
```

The imputed data are analyzed either by computing the dosage value and used in place of G additively (2 df) without the dominance term, or by replacing the genotype indicators for each observation, with the corresponding group probabilities (3 df). Similar to the location association, sample relatedness is dealt with using GLS for autosomal markers at the second stage of linear regression via the correlation matrix. Additional models, such as a sex-stratified variance test, can also be specified for X-chromosome. In this case, the scale test result is given by the Fisher's method that combines female and male-specific variance test results.

```
> scaleReg(GENO=chrXdat[ , 7:11], SEX=chrXdat$SEX,
Y=chrXdat$PHENOTYPE, Xchr=TRUE, origLev=T)
  CHR     SNP       gS   LevFemale LevMale  Fisher Flagged
1 X rs5983012_A 0.1391909 0.5296098 0.08223648 0.1800391 1
2 X rs4119090_G 0.9828430 0.9143432 0.97557661 0.9939473 0
3 X rs5911042_T 0.1487017 0.1625983 0.40384172 0.2444805 0
4 X rs986810_C  0.9563390 0.5349616 0.78109151 0.7824831 1
5 X rs180495_G  0.3476929 0.9717918 0.14364539 0.4144558 1
```

Note that a "Flagged" column is appended for these results, indicating the minimum genotype count in either females/males is less than 30 (indicated by a value of 1) or not (indicated by 0). This is based on the quality control that SNPs with a minimum count below 30 should be removed to avoid inflated type I errors (Soave *et al.* 2015; Deng *et al.* 2019).

Joint location-and-scale analysis

The joint analyses can be performed automatically as part of the *gJLS2* pipeline after running location and scale tests, where the default option applies a quantile transformation to the quantitative trait such that the location and scale test can be combined without inflating the type I error rates.

```
>       gJLS2(GENO=chrXdat[       ,       7:11],
Y=chrXdat$PHENOTYPE, SEX=chrXdat$SEX, Xchr=TRUE)
  CHR         SNP         gL         gS       gJLS
1  X   rs5983012_A 0.9943538 0.1198837 0.3727472
2  X   rs4119090_G 0.8881506 0.9794193 0.9911401
3  X   rs5911042_T 0.3488576 0.1514217 0.2081701
4  X    rs986810_C 0.4898597 0.9244773 0.8116064
5  X    rs180495_G 0.8702304 0.3619588 0.67886814
```

Alternatively, summary statistics, i.e. *P*-values from location and scale tests are allowed and can be combined via Fisher's method to give the corresponding test statistic for the joint analysis:

$$W = -2[\log(gL) + \log(gS)] \sim \chi^2(4), \tag{3}$$

where gL denotes the location *P*-value and gS the scale *P*-value.

An important assumption underlying this simple method to combine evidence is the normality of the quantitative trait, which leads to gL and gS being independent under the null hypothesis. Though empirically the analyses remain valid for approximated normal distributions (Soave and Sun 2017), we strongly recommend the user to follow the default option and quantile transform the quantitative trait for location and scale association. We expect the independence between gL and gS to hold for X-chromosomal SNPs under appropriate location and scale models that account for the confounding effect of sex, XCI uncertainty and skewed XCI (Chen *et al.* 2021). The simulation study in Section 2 of the Supplementary Material was conducted to help readers gauge the effect of non-normality on the gJLS *P*-values for X-chromosome markers.

Scalability and using gJLS2 in non-GUI settings

The main contribution of the software is the ability to handle X-chromosome (location and scale) analysis and it is not our aim to compete with existing autosomal association tools that are geared toward whole-genome computations. As a result, for both location and scale association, data splitting remain our primary strategy to deal with biobank scale data. We also implemented the summary statistics option to encourage the inclusion of genome-wide results (autosome) computed using existing software as inputs for the location portion. For X-chromosome associations, we recommend using gJLS2 in non-GUI settings and provide 2 practical options for biobank-scale data.

For larger datasets, it is more convenient to run the joint analyses using the PLINK R plug-in following the a typical GWAS pipeline. The R plug-in relies on the *Rserve* package to establish communication between R and PLINK 1.9. The following script demonstrates the joint analyses for X-chromosome SNPs that included additional covariates.

```
R CMD Rserve --RS-port 8221

plink --bfile ./input/chrX_5_snp \
   --R run_gJLS2PLINK_ Xchr.R \
   --pheno ./input/Pheno.txt \
   --pheno-name pheno1 \
   --R-port 8221 \
   --covar ./input/Pheno.txt \
   --covar-name SEX, covar1, covar2, covar3 \
   --out ./output/testRun
```

Another option is to use the Rscript provided that allows additional arguments to change how frequently the results are written to the output (–write) and to increase the number of cores used (–nTasks). A core is an independent processing unit on a central processing unit (CPU). Though a modern computer, usually containing 4–8 cores, is capable of handling parallel computing, we recommend the "split-apply-combine" strategy and employing high-performance computing for large-scale analyses. Scripts for both options are available from github (https://github.com/WeiAkaneDeng/gJLS2/tree/main/inst/extdata).

```
Rscript run_gJLS2.R --bfile ./input/chrX_5_snp \
   --pfile ./input/Pheno.txt \
   --pheno pheno1 \
   --Xchr TRUE \
   --write 10 \
   --nTasks 2 \
   --covar SEX, covar1, covar2, covar3 \
   --out ./output/testRun.results.txt
```

To help assess the computational requirement at different data sizes, we sampled with replacement from the UKB X-chromosome data (restricted to 100 SNPs) to achieve sample sizes of: 1,000, 5,000, 10,000, 50,000, 100,000, 200,000, 300,000, 400,000, and repeated the joint-location-scale analysis using a single core with 10GB memory via (1) PLINK R plug-in and (2) Rscript. The reason for keeping the 100 SNPs to estimate the performance metric is because the analysis can be easily divided to chunks and combined after. Figure 1 shows the computational time and memory usage as a function of increasing sample size. These results suggest that the gJLS analysis on data with sample size up to 10,000 on less than 100 SNPs can be suitably handled in R GUI such as Rstudio, which might be the preferred option for confirmatory analyses.

For an X-chromosome wide analyses on UKB ($n = 488{,}377$, $m = 15{,}179$), the gJLS analysis took ~16 h for PLINK (using 3.5GB memory) and ~ 21 h (using 1.2GB) for Rscript. The memory efficiency of Rscript is expected as the "BEDmatrix" only maps the required portion of genotype files into memory. However, the Rscript can be parallelized, when using 4 cores with 20GB allocated memory, the wall clock run time was reduced to ~13 h (using 14.0 GB).

Results and Discussion

The *gJLS2* software supports the joint analysis on both individual-level data as well as summary statistics. To highlight the functions in our package, we present (1) an X-chromosome gJLS analysis on UK Biobank (UKB) data (Bycroft *et al.* 2018) on 4 complex traits previously studied in Deng *et al.* (2019); (2) a chromosome-wide gJLS analysis on summary statistics of location and scale from the GIANT consortium for BMI and height.

Application to UK Biobank data

The sex-stratified means and variances for these quantitative traits are summarized graphically in Supplementary Figs. 1–4. A quick visual inspection suggests that quantile normalization should be applied, which is also the default options in *gJLS2*. We restricted analyses to white British samples ($n = 276{,}694$), and for BMI, we further excluded those with diagnosed type 2 diabetes ($n = 262{,}837$). We included only bi-allelic SNPs and filtered based on MAF < 0.01, HWE *P*-value <1E-5. A further check on sex-stratified MAFs (Supplementary Fig. 5) confirmed that the remaining 15,179 X-chromosomal SNPs are of good quality.

The genotype data can be supplied in PLINK binary format via the R plug-in option using PLINK 1.9 (Chang *et al.* 2015) or any other format that can be read in R with packages "*BGData*" and "*BEDMatrix*" (Grueneberg and de los Campos 2019). The phenotype and covariate should ideally be supplied in the same file, and if not, should have a common column to match samples. For genome-wide analyses on larger datasets, we recommend the use of a high-performing computing cluster and taking advantage of multiple cores whenever available.

For the location and scale regression, we included age, genetic sex, and the first 10 genetic principal components. Since there are no additional arguments needed, such as dosage or related samples, the analysis can be done using either a PLINK R-plugin running on PLINK 1.9:

```
plink --bfile UKbiobank_ChrX_SNP_CLEANED \
   --R gJLS2PLINK_Xchr.R \
   --R-port 8221 \
   --pheno bmi_pheno.txt \
   --pheno-name BMI \
   --covar bmi_pheno.txt \
   --covar-name genetic_sex age PC1 PC2 PC3 PC4, PC5,
PC6, PC7, PC8, PC9, PC10 \
   --out UKbiobank_ChrX_SNP_bmi
```

or the Rscript option:

```
Rscript run_gJLS2.R --bfile
UKbiobank_ChrX_SNP_CLEANED \
   --pfile hip_pheno.txt \
   --pheno HIP \
   --Xchr TRUE \
   --write 100 \
   --covar age, genetic_sex, PC1, PC2, PC3, PC4, PC5,
PC6, PC7, PC8, PC9, PC10 \
   --out UKbiobank_ChrX_SNP_HIP.results.txt
```

The base scripts for PLINK (`gJLS2PLINK_Xchr.R`) and Rscript (`run_gJLS2.R`) are provided in the inst/extdata folder of the R package along with the input files. It is worth noting the main advantages of the Rscript option beyond its flexibility: (1) the gJLS2 R package supports multi-core computing via the *parallel* base package and the argument "nTasks" can be used to specify the number of cores; (2) another useful feature is the "write" option that specifies the chunk size for the results to be written while the analysis is running and thus minimizes loss in case an interruption occurred.

Since the PLINK plug-in option only supports single-thread computation, we fixed the number of cores to be 1. There is no drastic difference between the 2 options, both took ~20 h with 20G allocated memory (computing node specs 2xIntel E5-2683 v4 Broadwell @ 2.1 GHz), to complete the analysis for 276,694 unrelated European samples and 15,179 X-chromosome variants per trait. The gL, gS, and gJLS2 *P*-values are presented using a Manhattan plot, quantile–quantile plot, and a histogram (Supplementary Figs. 6–9) for each of the complex traits. We also tabulated a list of SNPs that did not pass the genome-wide significance threshold of 5E-8 for gL, but did pass for gJLS in Table 1, demonstrating the benefit of gJLS for additional genome-wide discoveries.

Application to summary statistics

The Rscript is more flexible than the PLINK plug-in solution as it can also handle analysis of summary statistics. The input file should contain at least 3 columns with headers "SNP," "gL," and "gS," while the output file has an additional column "gJLS." We re-formatted the subset of chromosome 16 summary statistics of location and scale obtained from the GIANT consortium for BMI and height as inputs.

```
Rscript run_gJLS2.R --sumfile ./input/
GIANT_BMI_chr16_gJLS_summary.txt \
   --out ./output/GIANT_BMI_Sum.chr16_results.txt

Rscript run_gJLS2.R --sumfile ./input/
GIANT_Height_chr16_gJLS_summary.txt \
   --out ./output/
GIANT_Height_Sum.chr16_results.txt
```

The gL, gS, and gJLS2 *P*-values are presented using a Manhattan plot, quantile-quantile plot, and a histogram (Supplementary Figs. 10 and 11). For chromosome 16, we gained 2 SNPs in *DNAH3* gene that passed the genome-wide significance using gJLS for BMI and an additional 23 SNPs for height (Supplementary Table 1).

Concluding remarks

The gJLS2 software package is a versatile tool for genome-wide discovery that is X-chromosome inclusive. As compared to previous versions, namely, JLS and gJLS, it has improved remarkably in terms of methodology, flexibility, computational improvements and most importantly, usability for large-scale data. Meanwhile, we expect many new features to be added with ongoing work. Naturally, the analysis of rare variants is a possible future direction for scale association test under the regression framework of a generalized Levene's test. With the available summary statistics, a systematic approach to prioritize SNPs that takes into account the joint-location, scale effects, and functional annotation is needed to produce relevant candidates for gene–environment interactions. Finally, apart from the improved signal detection, the unbiased estimation of location effects for X-chromosome and scale effects in general remain open questions, but are expected to yield improved performance of polygenic prediction.

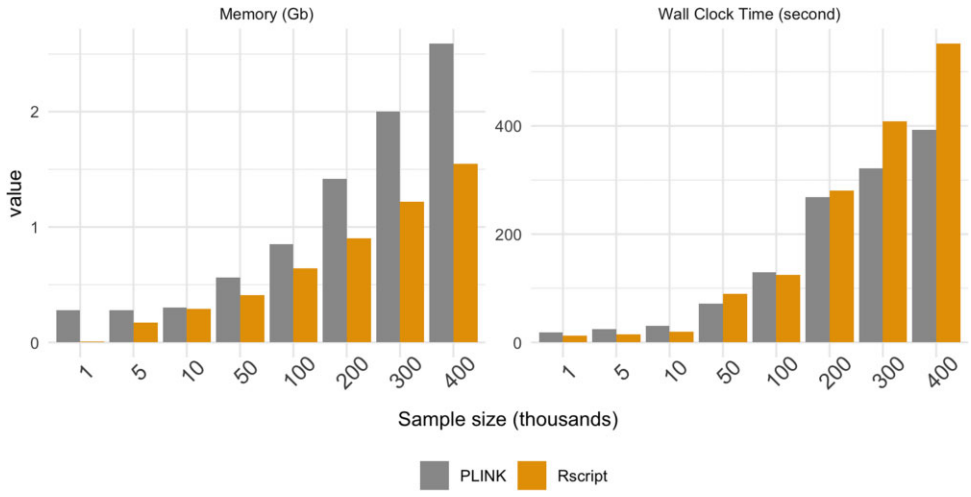

Fig. 1. Computational usage of gJLS X-chromosome association analyses. The bars represent the memory in (Gigabyte) and wall clock time (seconds) used to perform a gJLS X-chromosome association for 100 markers at various sample sizes (x-axis, per a thousand samples).

Table 1. Number of X-chromosomal SNPs gained at genome-wide significance by the joint location-scale *P*-values for 4 complex traits in UKB.

Trait	#SNPs gained	rsSNP	BP (hg19)	Gene	gL	gS	gJLS
Height	2	rs4844285	70370244	*NLGN3*	6.34E-08	0.00513854	7.45E-09
		rs5913116	79122277		5.25E-08	0.01355588	1.57E-08
Hip	2	rs2430200	117941116		4.40E-07	0.0042953	3.99E-08
		rs2495623	117898553	*IL13RA1*	9.75E-08	0.02024879	4.15E-08
Waist	1	rs5957063	117961798		8.42E-08	0.00080238	1.65E-09

Data availability

The gJLS2 package and the example datasets are available from https://github.com/WeiAkaneDeng/gJLS2, an accompanying guide for gJLS2 is available at https://weiakanedeng.github.io/gJLS2/. The simulations in this paper were based on X-chromosome genotypes from the 1000 Genomes Project (http://www.1000genomes.org), which can be freely accessed from the online data portal. The genotype and phenotype data used to demonstrate the X-chromosome analysis are available from UK Biobank (https://www.ukbiobank.ac.uk/) release in March 2018 and under project identification number 64875. Data access can be requested with information provided here: http://www.ukbiobank.ac.uk/using-the-resource/. The genome-wide summary statistics are made available from the GIANT data portal (https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files).

Supplemental material is available at *G3* online.

Acknowledgments

The authors would like to acknowledge Jiafen Gong, Sangook Kim, and Boxi Lin for testing and providing feedbacks on the initial versions of the gJLS2 software. They thank David Soave, the original creator of JLS and gJLS, for a critical reading of the study.

Funding

This research is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), the University of Toronto McLaughlin Centre, and the Canadian Institutes of Health Research (CIHR).

Conflict of interest

The authors declare that there is no conflict of interest.

Literature cited

Cao Y, Wei P, Bailey M, Kauwe JSK, Maxwell TJ. A versatile omnibus test for detecting mean and variance heterogeneity. Genet Epidemiol. 2014;38(1):51–59.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1):7.doi:10.1186/s13742-015-0047-8.

Chen B, Craiu RV, Strug LJ, Sun L. The X factor: a robust and powerful approach to X-chromosome-inclusive whole-genome association studies. Genet Epidemiol. 2021;45(7):694–709. doi:10.1002/gepi.22422.

Deng WQ, Mao S, Kalnapenkis A, Esko T, Mägi R, Paré G, Sun L. Analytical strategies to include the X-chromosome in variance heterogeneity analyses: evidence for trait-specific polygenic variance structure. Genet Epidemiol. 2019;43(7):815–830. doi:10.1002/gepi.22247.

Grueneberg A, de los Campos G. "BGData - A suite of R packages for genomic analysis with Big Data". G3 (Bethesda). 2019;9(5):1377–1383. doi:10.1534/g3.119.400018.

Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015;47(3):284–290. doi:10.1038/ng.3190.

Paré G, Cook NR, Ridker PM, Chasman DI.On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. PLoS Genet. 2010;6(6):e1000981.

Pazokitoroudi A, Chiu AM, Burch KS, Pasaniuc B, Sankararaman S. Quantifying the contribution of dominance deviation effects to complex trait variation in biobank-scale data. Am J Hum Genet. 2021;108(5):799–808. doi:10.1016/j.ajhg.2021.03.018.

Shungin D, Deng WQ, Varga TV, Luan J, Mihailov E, Metspalu A, Morris AP, Forouhi NG, Lindgren C, Magnusson PKE, *et al.* Ranking and characterization of established BMI and lipid associated loci as candidates for gene-environment interactions. PLoS Genet. 2017;13(6):e1006812.

Soave D, Corvol H, Panjwani N, Gong J, Li W, Boëlle PY, Durie PR, Paterson AD, Rommens JM, Strug LJ, *et al.* A joint location-scale test improves power to detect associated SNPs. Am J Hum Genet. 2015;97(1):125–138. doi:10.1016/j.ajhg.2015.05.015.

Soave D, Sun L. A generalized Levene's scale test for variance heterogeneity in the presence of sample correlation and group uncertainty. Biom. 2017;73(3):960–971. doi:10.1111/biom.12651.

The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526:68–74. doi:10.1038/nature15393.

Wickham H, Hester J, Chang W. devtools: tools to Make Developing R Packages Easier. 2018. https://CRAN.R-project.org/package=devtools.

Yang J, Loos RJF, Powell JE, Medland SE, Speliotes EK, Chasman DI, Rose LM, Thorleifsson G, Steinthorsdottir V, Mägi R, *et al.* FTO genotype is associated with phenotypic variability of body mass index. Nature. 2012;490(7419):267–272. doi:10.1038/nature11401.

Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM, GIANT Consortium. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum Mol Genet. 2018;27(20):3641–3649. doi:10.1093/hmg/ddy271.

Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, LeFaive J, VandeHaar P, Gagliano SA, Gifford A, *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet. 2018;50(9):1335–1341. doi:10.1038/s41588-018-0184-y.

Communicating editor R. Cantor