



OPEN ACCESS

EDITED AND REVIEWED BY
Kesheng Wang,
West Virginia University, United States

*CORRESPONDENCE
Min Zhang,
minzhang@purdue.edu
Riyan Cheng,
ric025@ucsd.edu

SPECIALTY SECTION
This article was submitted to Statistical
Genetics and Methodology,
a section of the journal
Frontiers in Genetics

RECEIVED 08 September 2022
ACCEPTED 05 October 2022
PUBLISHED 20 October 2022

CITATION
Han L, Liu H, Kang G, Zhang M and
Cheng R (2022), Editorial: Statistical
methods, computing, and resources for
genome-wide association studies,
Volume II.
Front. Genet. 13:1040022.
doi: 10.3389/fgene.2022.1040022

COPYRIGHT
© 2022 Han, Liu, Kang, Zhang and
Cheng. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Editorial: Statistical methods, computing, and resources for genome-wide association studies, Volume II

Lide Han¹, Hailan Liu², Guolian Kang³, Min Zhang^{4*} and Riyan Cheng^{5*}

¹Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, United States, ²Maize Research Institute, Sichuan Agricultural University, Chengdu, China, ³Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, United States, ⁴Department of Statistics, Purdue University, West Lafayette, IN, United States, ⁵Department of Psychiatry, University of California San Diego, San Diego, CA, United States

KEYWORDS

genome-wide association studies (GWAS), gene-based testing, longitudinal phenotypes, binary phenotypes, sparse mixture-of-experts models

Editorial on the Research Topic

Statistical methods, computing and resources for genome-wide association studies-volume II

As in the previous volume, this collection presents several research articles that address statistical methodology, computing, and resources for genome-wide association studies (GWAS). The research interest covers association studies of longitudinal data, multiple testing, computationally efficient analysis of binary phenotypes and more.

Phenotypic data is often longitudinal in plant science. Genetic analysis of longitudinal phenotypic data can treat data at different time points separately or fit a growth curve that depicts the developmental process. The latter is commonly practiced at single genetic variants while gene-based GWAS are often of interest. Li et al. propose a function-on-function regression method, a longitudinal functional data association test, which not only models the developmental process but also takes advantage of gene-based testing, and shows that their method performs well for the identification of genetic variants switching in the growth and development stage.

Proper controlling of the type I error rate is an important issue in GWAS. While the permutation test is regarded as the gold standard, its extensive computation is prohibitive in genetic studies which often include thousands of samples and millions of single nucleotide polymorphisms (SNPs). In human genetics, a broadly-accepted p -value cutoff is commonly used for statistical inference. However, the genome size and the SNP distribution often vary from study to study, and it is thus desirable to determine data-driven significance thresholds. Many computationally efficient methods have been proposed for the derivation of genome-wide thresholds, including Bonferroni

correction based on an estimated effective number of tests and Brown's method, which take the SNP dependency into account to achieve a higher power while properly controlling the type I error rate. Cinar and Viechtbauer survey several gene-based testing methods and evaluate their performance *via* simulation studies. Their work provides valuable information to those who seek computationally efficient methods for significance threshold estimation in GWAS.

Categorical phenotypes are not rare in genetic studies. Genome-wide association studies of categorical phenotypes are computationally challenging due to the incorporation of genetic relatedness matrices (GRM). The most common categorical phenotypes are binary. There have been enormous efforts to address the computational issues in GWAS with binary phenotypes. Strategies include the use of sparse GRM plus Saddle Point Approximation. With a focus on computational efficiency and the control of false positives, Gurinovich et al. empirically evaluate several known software programs such as SAIGE and fastGWA-GLMM and provide food for thought for researchers who consider software programs that perform GWAS analysis with binary phenotypes.

It is known that many diseases are heritable. Genetic data not only plays an important role in the identification of genetic factors underlying diseases but also is helpful for determining disease subtypes. Courbariaux et al. study phenotypic data that is longitudinal and arises from multiple sources, and propose a sparse mixture-of-experts model that incorporates genetic information for disease subtyping. The authors claim two advantages of their methods over others. First, their model-based clustering utilizes original data without transformation and thus facilitates the interpretation of results. Second, they address the large-scale problem due to massive genetic data by model selection. This collection also touches on two popular problems, heritability estimation and polygenic risk scores

(PRS). Heritability is often reported in GWAS. Zhang and Sun show that heritability tends to be overestimated in the presence of Hardy–Weinberg disequilibrium. Such information should be useful for those who are interested in heritability estimation. As genomic prediction, PRS concern estimation of genetic values underlying a trait of interest. Zhang et al. compare methods that utilize sex-specific PRS and share useful information with interested readers.

To sum up, this volume highlights several research interests in statistical methods, computing, and resources for GWAS which we hope contribute to the society of genetic studies.

Author contributions

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.