Check for updates

Editorial: Statistical methods for genome-wide association studies (GWAS) and transcriptome-wide association studies (TWAS) and their applications

Mengting Shao[1], Zilong Zhang[2], Huiyan Sun[3], Jingni He[4], Juexin Wang[5], Qingrun Zhang[6]* and Chen Cao[1]*

[1]School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China, [2]School of Computer Science and Technology, Hainan University, Haikou, China, [3]School of Artificial Intelligence, Jilin University, Changchun, China, [4]Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, AB, Canada, [5]Department of Biohealth Informatics, Indiana University Purdue University Indianapolis, Indianapolis, IN, United States, [6]Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada

Editorial on the Research Topic
Statistical methods for genome-wide association studies (GWAS) and transcriptome-wide association studies (TWAS) and their applications

Introduction

The Genome-Wide Association Studies (GWAS) has proven highly successful in identifying millions of risk loci associated with various diseases in the past 15 years (Klein et al., 2005). With the rapid accumulation of GWAS summary-level data, biologists now have expanded opportunities to uncover new disease-associated variants and gain insights into the mechanisms underlying complex human diseases (Michailidou et al., 2017; Sud et al., 2017; Zhang et al., 2020). While GWAS is a powerful tool, it faces challenges in pinpointing candidate disease risk genes. For instance, many disease-associated variants reside in non-coding regions, complicating the identification of their regulatory genes and underlying mechanisms due to the likely cell-type, context, and disease-specific effects of non-coding causal variants (Kossinna et al., 2022). On the other hand, GWAS finds it hard to distinguish between causal variant signals and significance signals in high Linkage Disequilibrium (LD) regions, leading to poor interpretation of GWAS signals (Christoforou et al., 2012; Cao et al., 2021a).

In response to these challenges, several post-GWAS methods have emerged, including Transcriptome-Wide Association Studies (TWAS) (Gamazon et al., 2015; Cao et al., 2021b; Cao et al., 2022; He et al., 2022), Proteome-Wide Association Study (PWAS) (Brandes et al., 2020), and Summary Data-Based Mendelian Randomization (SMR) (Zhu et al., 2016). These

methods serve as potent tools for discovering candidate disease risk genes, offering benefits such as enhanced statistical power, improved interpretability, and reduced computational costs. In recent years, numerous studies have adopted GWAS, TWAS, and SMR to delve into the intricate biological mechanisms of diseases (Baca et al., 2022).

Causal variants in multiple traits

Untangling causal signals from mere associations in GWAS presents a big challenge. Techniques such as fine mapping, Mendelian randomization, and TWAS have risen to address this challenge, facilitating the translation of GWAS findings into a functional understanding of associated traits. In this Research Topic of Frontiers in Genetics, six research articles demonstrate the efficacy of these techniques. For instance, Lu et al. utilized meticulous fine mapping to identify the rs7175517 variant as related to Body Mass Index (BMI) across diverse populations, offering fresh insights into the global obesity epidemic. In another study, Chen et al. used blood proteins as traits in GWAS, employing a two-sample Mendelian randomization analysis to identify causal proteins linked to sarcopenia-related traits. This not only identified potential therapeutic targets but also shed light on underlying genetic factors. In a separate investigation, Lu et al. focused on celiac disease, a comprehensive strategy involving TWAS and chemical-gene interaction analyses unveiled celiac disease-related genes and chemicals, providing valuable insights at both the genetics and environmental levels.

Various tools in genome-wide association studies

Another critical aspect of identifying disease-associated genes involves prioritizing trait-specific tissues, which may lead to differences in gene expression and variant regulation. To address this, Ghaffar and Nyholt developed a method called genome-wide imputed differential expression enrichment (GIDEE). GIDEE prioritizes pathogenic tissues by analyzing the enrichment of differentially expressed genes in each tissue. Additionally, the relationship between diseases plays a key role in identifying variants shared across multiple traits or diseases. To tackle this challenge, graph-GPA 2.0 (GGPA 2.0) was proposed by Deng et al. It integrates GWAS datasets of multiple diseases and utilizes functional annotations within a unified framework, successfully detecting pleiotropy between bipolar disorder and schizophrenia. Furthermore, a Visual SNP interpretation tool named SNPMap was proposed by Liu et al. to illustrate semantic relations between SNPs and traits, significance, and SNP-related information. This tool aids researchers in better understanding the link between genetic variation and disease risk.

Conclusion

Together, these articles reveal the significant potential of identifying disease susceptibility genes, understanding disease mechanisms, and discovering drug targets using GWAS and post-GWAS tools. They provide valuable knowledge resources for future medical research and clinical applications.

Author contributions

Acknowledgments

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.

References

Baca, S. C., Singler, C., Zacharia, S., Seo, J. H., Morova, T., Hach, F., et al. (2022). Genetic determinants of chromatin reveal prostate cancer risk mediated by context-dependent gene regulation. *Nat. Genet.* 54 (9), 1364–1375. doi:10.1038/s41588-022-01168-y

Brandes, N., Linial, N., and Linial, M. (2020). Pwas: proteome-wide association study-linking genes and phenotypes by functional variation in proteins. *Genome Biol.* 21 (1), 173. doi:10.1186/s13059-020-02089-x

Cao, C., Ding, B., Li, Q., Kwok, D., Wu, J., and Long, Q. (2021a). Power analysis of transcriptome-wide association studies: implications for practical protocol choice. *PLoS Genet.* 17 (2), e1009405. doi:10.1371/journal.pgen.1009405

Cao, C., Kossinna, P., Kwok, D., Li, Q., He, J., Su, L., et al. (2022). Disentangling genetic feature selection and aggregation in transcriptome-wide association studies. *Genetics* 220 (2), iyab216. doi:10.1093/genetics/iyab216

Cao, C., Kwok, D., Edie, S., Li, Q., Ding, B., Kossinna, P., et al. (2021b). kTWAS: integrating kernel machine with transcriptome-wide association studies improves statistical power and reveals novel genes. *Brief. Bioinform* 22 (4), bbaa270. doi:10.1093/bib/bbaa270

Christoforou, A., Dondrup, M., Mattingsdal, M., Mattheisen, M., Giddaluru, S., Nöthen, M. M., et al. (2012). Linkage-disequilibrium-based binning affects the interpretation of GWASs. *Am. J. Hum. Genet.* 90 (4), 727–733. doi:10.1016/j.ajhg.2012.02.025

Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* 47 (9), 1091–1098. doi:10.1038/ng.3367

He, J., Wen, W., Beeghly, A., Chen, Z., Cao, C., Shu, X. O., et al. (2022). Integrating transcription factor occupancy with transcriptome-wide association analysis identifies susceptibility genes in human cancers. *Nat. Commun.* 13 (1), 7118. doi:10.1038/s41467-022-34888-0

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* 308 (5720), 385–389. doi:10.1126/science.1109557

Kossinna, P., Cai, W. J., Lu, X. W., Shemanko, C. S., and Zhang, Q. R. (2022). Stabilized COre gene and Pathway Election uncovers pan-cancer shared pathways and a cancer-specific driver. *Sci. Adv.* 8 (51), eabo2846. doi:10.1126/sciadv.abo2846

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature* 551 (7678), 92–94. doi:10.1038/nature24284

Sud, A., Kinnersley, B., and Houlston, R. S. (2017). Genome-wide association studies of cancer: current insights and future perspectives. *Nat. Rev. Cancer* 17 (11), 692–704. doi:10.1038/nrc.2017.82

Zhang, H., Ahearn, T. U., Lecarpentier, J., Barnes, D., Beesley, J., Qi, G., et al. (2020). Genome-wide association studies identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat. Genet.* 52 (6), 572–581. doi:10.1038/s41588-020-0609-2

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48 (5), 481–487. doi:10.1038/ng.3538