

HisCoM-GGI: Software for Hierarchical Structural Component Analysis of Gene-Gene Interactions

Sungkyoung Choi¹, Sungyoung Lee², Taesung Park^{3,4*}

¹Department of Pharmacology, Yonsei University College of Medicine, Seoul 03722, Korea, ²Center for Precision Medicine, Seoul National University Hospital, Seoul 03080, Korea, ³Department of Statistics, Seoul National University, Seoul 08826, Korea, ⁴Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 08826, Korea

Gene-gene interaction (GGI) analysis is known to play an important role in explaining missing heritability. Many previous studies have already proposed software to analyze GGI, but most methods focus on a binary phenotype in a case-control design. In this study, we developed “Hierarchical structural CoMponent analysis of Gene-Gene Interactions” (HisCoM-GGI) software for GGI analysis with a continuous phenotype. The HisCoM-GGI method considers hierarchical structural relationships between genes and single nucleotide polymorphisms (SNPs), enabling both gene-level and SNP-level interaction analysis in a single model. Furthermore, this software accepts various types of genomic data and supports data management and multithreading to improve the efficiency of genome-wide association study data analysis. We expect that HisCoM-GGI software will provide advanced accessibility to researchers in genetic interaction studies and a more effective way to understand biological mechanisms of complex diseases.

Keywords: gene-gene interactions, generalized structured component analysis, genome-wide association study, ridge regression

Availability: This software and the accompanying tutorial are available at <http://statgen.snu.ac.kr/software/hiscom-ggi>.

Introduction

In the past decade, genome-wide association studies (GWASs) have successfully identified genetic variants associated with human traits and complex diseases. However, traditional GWASs have several issues and limitations that do not completely explain the heritability of complex human diseases or related traits—so-called missing heritability [1]. To address these issues, various approaches have been proposed, such as rare variant association analysis, genetic interactions, and epistasis [1]. Among possible explanations for missing heritability, there has been growing consideration of gene-gene interaction (GGI) analysis for common genetic variants in terms of statistical approaches and analysis software. Although most research models focus on the case-control study design, there are not many software programs that can analyze a continuous phenotype. Furthermore, GGI analysis has several advantages

in terms of statistical power, computing performance, and biological interpretation over single nucleotide polymorphism (SNP)–SNP interaction (SSI) analysis, but there is not much analysis software to analyze it.

In this study, we developed software for “Hierarchical Structural Component Analysis of Gene-Gene Interactions” (HisCoM-GGI) [2] that can analyze both GGI and SSI analyses for a continuous phenotype. The basic framework of the HisCoM-GGI software was “Workbench for Integrated Superfast Association study with Related Data” (WISARD) [3], which accepts various types of input formats and provides quality control or data management options that can be used to filter out individuals or genetic variants. In addition, the WISARD program supports multithreading to improve the efficiency of the analysis. By using the advantages of WISARD, HisCoM-GGI software includes all of the functions for GGI analysis on a genome-wide scale, such as input format support, data preprocessing, and multithreading.

Received December 11, 2018; Accepted December 18, 2018; Published online December 28, 2018

*Corresponding author: Tel: +82-2-880-8924, Fax: +82-2-888-6693, E-mail: tspark@stats.snu.ac.kr

Copyright © 2018 by the Korea Genome Organization

© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>).

Implementation

The workflow of the HisCoM-GGI software is shown in Fig. 1. The HisCoM-GGI method is based on the Generalized Structured Component Analysis (GSCA) method [4]. GSCA is a component-based approach to structural equation modeling. The latent variables of GSCA are defined as components and are estimated by the weighted sums of the observed variables. The HisCoM-GGI method has been proposed for genetic interaction analysis by constructing a hierarchical model that consists of SNP-level and gene-level summaries. Therefore, the HisCoM-GGI method provides the results of SSI and GGI in a single model. The HisCoM-GGI software was implemented in C++ and was developed using programs for Linux and Windows.

Input file

The HisCoM-GGI software takes three inputs: (1) a genomic dataset in PLINK or variant calling format, (2) trait information on the covariates and phenotypes, and

(3) a set file that consists of two columns for gene name and SNP ID (rsID), respectively. Furthermore, users can optionally specify a gene-gene pair list to analyze.

Output file

The HisCoM-GGI program generates the following files: (1) a '[prefix].gesca.latent.res' file that contains p-values of gene-level interactions and (2) a '[prefix].gesca.manifest.res' file that contains p-values of SNP-level interactions.

Visualization of results

We provide the circos plots for the results of GGI and SSI using R script at the following website: http://statgen.snu.ac.kr/software/hiscom-ggi/?page_id=10. The visualized results of the HisCoM-GGI software are shown in Fig. 2.

Conclusion

In this paper, we introduce the HisCoM-GGI software program, which provides a statistical test of identifying

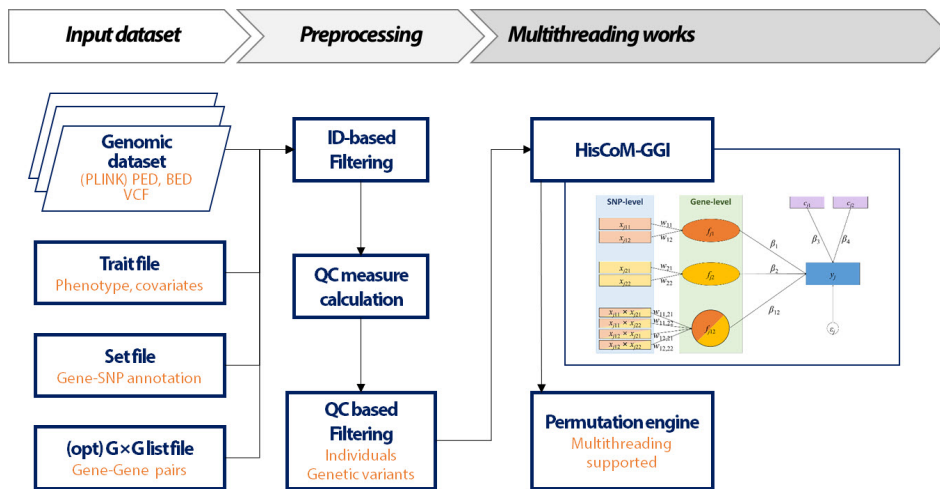
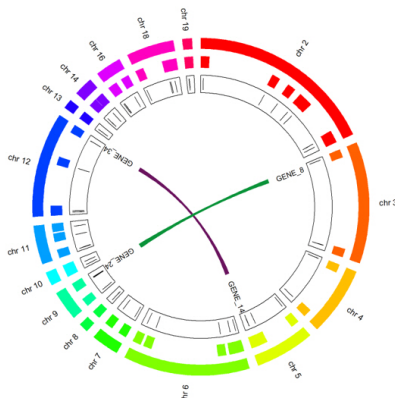


Fig. 1. Workflow of the HisCoM-GGI program. BED, binary PED; VCF, variant calling format; QC, quality check.

(A) Gene-level interactions



(B) SNP-level interactions

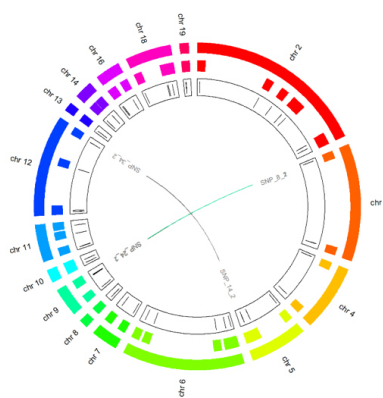


Fig. 2. Circos plots for the results of gene-level (A) and single nucleotide polymorphism (SNP)-level interactions (B).

gene-level and SNP-level interactions for continuous phenotypes. We conclude that HisCoM-GGI software may be a valuable tool for the identification of genetic interactions, allowing us to better understand biological mechanisms of complex human diseases or related traits. The software is freely available with a tutorial dataset through the website <http://statgen.snu.ac.kr/software/hiscom-ggi>.

ORCID: Sungyoung Choi: <https://orcid.org/0000-0002-4266-5911>; Sungyoung Lee: <https://orcid.org/0000-0003-3458-1440>; Taesung Park: <https://orcid.org/0000-0002-8294-590X>

Authors' contribution

Conceptualization: SC, TP
Formal analysis: SC, SL
Funding acquisition: TP
Methodology: SC, SL, TP
Writing – original draft: SC, TP
Writing – review & editing: SC, TP

Conflicts of Interest

No potential conflicts of interest relevant to this article was reported.

Acknowledgments

This work was supported by the following grants: Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (HI16C2037, HI15C2165).

References

1. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, *et al.* Finding the missing heritability of complex diseases. *Nature* 2009;461:747-753.
2. Choi S, Lee S, Kim Y, Hwang H, Park T. HisCoM-GGI: Hierarchical structural component analysis of gene-gene interactions. *J Bioinform Comput Biol* 2018;16:1840026.
3. Lee S, Choi S, Qiao D, Cho M, Silverman EK, Park T, *et al.* WISARD: workbench for integrated superfast association studies for related datasets. *BMC Med Genomics* 2018;11(Suppl 2):39.
4. Hwang H, Takane Y. Generalized structured component analysis. *Psychometrika* 2004;69:81-99.