



Specialty Grand Challenge: Systems Genetics

Rongling Wu*

Departments of Public Health Sciences and Statistics, Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA, United States

Keywords: system genetics, systems biology, network genetics, genotype-phenotype map, epistasis, omnigenic interactome network, multi-omics network

INTRODUCTION

We are currently in the midst of a genomic revolution in which an avalanche of DNA and RNA sequencing data can be produced for any organism with almost no limit. These data are fueling the dissection of quantitative traits and human diseases into their molecular machineries through the concepts of systems genetics. The past decade has witnessed the emergence of systems genetics as a discipline of genetics and its tremendous impact on genotype-phenotype mapping. Following *Cell Systems*' 2017 forum in which 16 experts were invited to share their views on the field of systems genetics (Baliga et al., 2017), the European Molecular Biology Organization (EMBO) convened a special symposium on the systems genetics of complex traits in 2019 (<https://www.embo-embl-symposia.org/symposia/2019/EES19-08/index.html>). Systems genetics studies complex traits through understanding the flow of information underlying biological processes from genotype to phenotype. This essentiality is largely above and beyond the reductionist concepts and tools used in traditional genetics.

The central theme of systems genetics is to integrate genetic and genomic data to better reveal the intrinsic complexity of biological processes underlying phenotypic variation. As an emerging discipline, the definition of systems genetics, its relevance to other fields, its standing position in modern biology and medicine, and its future development to play a more pivotal role, are dynamic, varying in the research community. Yet, there has been a well-received recognition that the heart of systems genetics is the network modeling of complex biological systems. A network is the mathematical formulation of a graph in which nodes represent individual entities and links or edges stand for the functional interconnections of different entities. In genetic networks, a "node" may present a biological entity like a SNP, gene, protein, metabolite, or even a specific disease/phenotype, whereas an "edge" may represent physical interaction, chemical induction, signal transduction, or even shared genes among different phenotypes. In this Specialty Grand Challenge, I list several key challenges for network conceptualization, application and reconstruction that are immediately faced by system genetics from a personal perspective.

OPEN ACCESS

Edited and reviewed by:

Yoram Vodovotz,
University of Pittsburgh, United States

*Correspondence:

Rongling Wu
rwu@phs.psu.edu

Specialty section:

This article was submitted to
Integrative Genetics and Genomics,
a section of the journal
Frontiers in Systems Biology

Received: 08 July 2021

Accepted: 18 August 2021

Published: 01 September 2021

Citation:

Wu R (2021) Specialty Grand
Challenge: Systems Genetics.
Front. Syst. Biol. 1:738155.
doi: 10.3389/fsysb.2021.738155

OMNIGENIC INTERACTOME NETWORKS DRIVING COMPLEX TRAITS

During the past 15 years, there have been enormous applications of genome-wide association studies (GWAS) to study the genetic architecture of complex traits and diseases. A typical approach for GWAS data analysis is to identify single DNA variants for one phenotype at a time. It turns out that each significant variant can only account for a tiny portion of genetic variation. Collectively, the portion of genetic variation explained by all significant variants detected from the whole genome is still below the heritability of a complex trait. Enormous efforts have been made to retrieve this so-called missing heritability by considering other types of variants, such as epigenetic marks, copy number variations, or rare alleles, as contributing factors, but no consensus agreement has been yet reached on where the

heritability is lost. Theoretical geneticists believe that it is impossible to retrieve this missing mystery because an extremely large number of genes of small effects are involved.

More recently, a so-called “omnigenic” model has been proposed to interpret the genetic architecture of complex traits (Boyle et al., 2017). This model states that complex phenotypes are controlled by a small number of “core” genes directly linked to phenotypic variation and a large number of “peripheral genes” that play a role through regulatory networks. This model is essentially similar in spirit to the QTL “oligogenic” model. Although some argue that the “oligogenic” model already encompasses the omnigenic extreme, the hypothesis of a few core (major) genes working along many peripheral (minor) genes has inspired widespread discussion and the potential advancement of quantitative genetics. However, the omnigenic model does not contextualize the complexity of the network to interrogate how core genes play a more important role than peripheral genes and why the existence of peripheral genes is essential given their subtle, even negligible effects.

We argue that the omnigenic model can be tested by inferring omnigenic interactome networks (OGIN) that cover a complete set of SNPs genotyped from the whole genome in a GWAS design (Wang et al., 2021). Such OGIN have many desirable properties that can fill two major gaps of systems genetics. First, the genetic effect of a SNP estimated by existing models represents the net effect of this locus. Beyond this, the OGIN partitions the net effect of each SNP into its two underlying components: the *independent effect* that arises when this SNP is assumed to be in isolation and the *dependent effect* is derived from the collective effect of interactions of other SNPs with this SNP. A core gene may not have a large independent effect, but it still can play a critical role modulating phenotypic variation because of its more links with other genes. Likewise, a peripheral gene may not necessarily be subtle, but because of its large independent effect cancelled by dependent effects in an inverse sign. Thus, by altering its interaction environments to promote or inhibit dependent effects, we can better edit and utilize a genetic locus of interest. On the other hand, even if a peripheral gene has a neglectful effect, it may still lead to an unpredicted change of phenotypic variation. This is because peripheral genes may individually wield negligible effects but they may raise the possibility of a “butterfly” effect (from chaos theory) of genetic interactions, a phenomenon of large unforeseen consequences caused by a sensitive dependence on a small initial change.

Second, core genes are defined as those with more links in the network than other genes. Beyond this definition, OGIN can classify all links of each SNP into “outgoing” and “incoming” types. An outgoing link describes an “active” process in which a SNP as a regulator actively promotes or inhibits other SNPs, whereas an incoming link is a “passive” process of a SNP receiving promotion or inhibition by other SNPs. OGIN can not only count the numbers of outgoing and incoming links for a specific SNP, but also quantify the strength of each of these links. Taken together, OGIN will outperform a traditional marginal analysis of single genes or single gene pairs, equipped with a capacity to reveal an overall picture of how the genotype is connected to the phenotype.

While OGIN can advance systems genetics, its statistical reconstruction presents a major challenge. There has been a rich body of literature on network reconstruction, but existing

approaches may not be sophisticated enough to capture favorable and unique properties of OGIN from general GWAS data. Many approaches are too specific and fragmented, hardly used to augment a generalized argument from various problem domains. For example, by reviewing over 30 network inference approaches, Marbach et al. (2012) found that no single one performs optimally across all datasets. Integration of multiple inference approaches may be robust across diverse datasets, but each approach has its own underlying mathematical rationale and assumptions, thus making the results difficult to interpret. Chen and Mar (2018) further pinpointed that existing approaches lack performance to reconstruct gene networks using heterogeneous single cell data that are becoming popular in genetic studies. Several attempts have been made to combine elements of multiple distinct disciplines to reconstruct maximally informative genetic networks that function at different levels of organization (Sun et al., 2021; Wu and Jiang 2021), producing some promising results that may advance systems genetics (Wang et al., 2021).

TRIDIMENSIONAL NETWORKS ACROSS TIME AND SPACE

Classic genetic research is aimed at analyzing the association between genotype and its remote phenotype. The mission of systems genetics is to unravel the “black box” behind the processes intermediate between genotype and phenotype by identifying biological molecules from DNA sequence variants (also epigenetic marks) to levels of transcripts, proteins and metabolites to cellular components of complex trait. Approaches have been available to line up these intermediate phenotypes or endophenotypes, some of which attempt to infer interaction networks from multi-omics data. However, these approaches generally do not attempt to chart a big picture of genotype-phenotype processes, rather focuses on some certain pathways. When modeling multi-omics data from different spaces, they do not take into account the asynchronous feature of signal transduction. For example, the expression of proteins lags behind the expression of genes, and metabolites are synthesized after perception and recognition of the signals originating from gene or protein elicitors. Thus, these approaches can only project originally existing space-dependent multilayer networks on one surface, ignoring how endophenotypes interact with each other directionally and intertwiningly across spaces.

A unified approach is needed to assemble different types of endophenotype data into a multilayer and multiplex mathematical graph. This procedure will generate a tridimensional network in which surfaces represent interaction networks of endophenotypes at a single space and vertical edges represent interaction networks of endophenotypes across different spaces. For example, when genes G1–G4 form a surface interaction network at the gene space, their transcripts T1–T4 generate a surface interaction network at the transcript space. These genes and transcripts interact across the gene and transcript spaces to form multiple channels of connectivity. By linking the gene space to transcript space to protein space to metabolite space to microbiome space finally to trait space, we

reconstruct a multiscale tridimensional network from which key nodes and key links can be identified to unveil a roadmap of genotype-phenotype relationships. This tridimensional network will be casual, signed, and weighted. It can uncover and quantify the causal relationships of endophenotypes from one space to the other or from the second to first. Statistical methods should be developed to untangle the most likely causal direction between two given spaces.

HORIZONTAL INTERACTION NETWORKS

Quantitative genetic theory has long focused on modeling how the phenotype of an organism is determined by its genes and the environment where it grows. An increasing body of evidence has unraveled that an individual's phenotype in a population is also affected by the phenotypes of other members that coexist with it. To better dissect phenotypic variation, we as a community will need to investigate not only how alleles from an individual directly affect its own phenotype, but also how alleles from the individual indirectly affect the phenotypes of members that co-exist socially with it and how alleles from different individuals affect epistatically the phenotypes of each member in a population. We define epistasis between genes from different individuals coexisting in a population as *horizontal epistasis*, as opposed to the traditional definition of *vertical epistasis* as the genetic interaction between different genes expressed in the same genome. The characterization of horizontal epistasis can help to chart a more complete atlas of genetic control for complex traits. To address this challenge, we will integrate community ecology theory and evolutionary game theory to model and quantify different types of social interactions between different individuals in a population, community or society. For example, the fetus carries DNA information from both parents and grows in the mother's uterus. Thus, the fetus and its parents form the smallest society in which each member develops and uses its optimal strategy to compete or cooperate with the other members through a coordinated social network.

Statistical methods need to be developed for reconstructing multiscale interaction networks that govern individual-individual interactions in a community. Such networks can be divided into direct genetic networks, indirect genetic networks, and horizontal epistatic networks from which we can identify key causal pathways linking genotype to phenotype and predict the growth trajectories of the intrauterine fetus in particular and community phenotypes (i.e., assembled phenotypes of multiple species) in general.

RECOVERING MAXIMALLY INFORMATIVE NETWORKS FROM STATIC DATA

The objective of network theory is to develop a tractable structure to distill relevant insight into the actions and interactions of a set of entities. Existing statistical approaches for network inference in genetics were mostly developed or modified from some aspects of network analysis in physical or social disciplines and, thereby, may not fully consider and capture biological complexities. For example, some approaches can estimate the strength of

interaction but fail to identify its direction, some can infer causality but cannot recover feedback cycles, and some can characterize all these network features but fail to determine the sign of interaction. For these reasons, many gene networks reconstructed by existing approaches are so-called nondirectional correlation networks or directed acyclic graphs. An approach that can not only fully capture all network properties, but also take into full account the biological requirements and complexities of complex trait dissection is sorely needed.

Wu and Jiang (2021) have developed a statistical model that can reconstruct bidirectional, signed, and weighted interaction networks. From a technical perspective, inferring such so-called informative networks requires temporal or perturbed data that provide an extra dimension for dynamic fitting. However, these types of data cannot be collected logistically or ethically in many studies; e.g., the genotype-tissue expression tissue (GTEx) project only can collect RNA-seq data once from dying donors. It is very challenging to collect time-series omics data from individual cells at high resolution using current sequencing techniques. In addition, transcriptional profiles of genes in cells or tissues may stochastically fluctuate over time and space, making it difficult to fit their dynamic trends using mathematical and statistical functions. Wu and Jiang (2021) developed a conceptual idea to *convert static data into their dynamic representation*, thus providing a key step to recover informative networks from static data. This methodological breakthrough facilitates the use of network tools based on static data which are much more readily collected and much more common than temporal data.

CONCLUDING REMARKS

The *Integrative Genetics and Genomics* section will provide a forum to strengthen and disseminate interdisciplinary research into systems genetics and create a synergistic environment for researchers worldwide to cross-promote and cross-pollinate the generation process of new ideas by connecting unrelated ideas brought from their own fields. We expect to publish new theories and statistical models that refresh and solidify the methodological foundation of systems genetics to better unveil life's complexities. The paradigm shift of genetic research from a reductionist thinking to holistic thinking will enable downstream researchers to predict more complete manifestations of complex traits and diseases and design rational breeding programs and therapies in translational sciences.

AUTHOR CONTRIBUTIONS

RW wrote and edited the manuscript.

ACKNOWLEDGMENTS

The author thanks many colleagues at Penn State College of Medicine for their stimulating discussions.

REFERENCES

- Baliga, N. S., Björkegren, J. L. M., Boeke, J. D., Boutros, M., Crawford, N. P. S., Dudley, A. M., et al. (2017). The State of Systems Genetics in 2017. *Cel Syst.* 4 (1), 7–15. doi:10.1016/j.cels.2017.01.005
- Boyle, E. A., Li, Y. L., and Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169, 1177–1186. doi:10.1016/j.cell.2017.05.038
- Chen, S., and Mar, J. C. (2018). Evaluating Methods of Inferring Gene Regulatory Networks Highlights Their Lack of Performance for Single Cell Gene Expression Data. *BMC Bioinf.* 19, 232. doi:10.1186/s12859-018-2217-z
- Marbach, D., Costello, J., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., et al. (2012). Wisdom of Crowds for Robust Gene Network Inference. *Nat. Methods* 9, 796–804. doi:10.1038/nmeth.2016
- Sun, L. D., Dong, A., Griffin, C., and Wu, R. L. (2021). Statistical Mechanics of Clock Gene Networks Underlying Circadian Rhythms. *Appl. Phys. Rev.* 8, 021313. doi:10.1063/5.0029993
- Wang, H. J., Ye, M. X., Fu, Y. R., Dong, A., Zhang, M. M., Feng, L., et al. (2021). Modeling Genome-wide by Environment Interactions through Omnigenic Interactome Networks. *Cell Rep* 35, 109114. doi:10.1016/j.celrep.2021.109114
- Wu, R., and Jiang, L. (2021). Recovering Dynamic Networks in Big Static Datasets. *Phys. Rep.* 912, 1–57. doi:10.1016/j.physrep.2021.01.003

Conflict of Interest: The author declares 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.

Copyright © 2021 Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.