

Predicting the Genotype-Phenotype Map of Complex Traits

Rongling Wu*

Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 17033, USA

Abstract

How to predict phenotypic development in a changing environment from the genotype of complex organisms is one of the most important and challenging questions we face in modern biology. This challenge can be addressed through establishing a framework that identifies and maps the mechanistic basis of the processes from genotype to phenotype. The central rationale of this framework is based on the genetic, developmental and regulatory dissection of phenotypic changes in response to different environments. First, a phenotype is genetically complex because of the involvement of many genes that display pervasive interactions with other genes and with environmental factors. Second, the formation of any phenotype involves a series of developmental events and biological alterations that entail cell growth, differentiation and morphogenesis. Third, DNA polymorphisms affect variation in a phenotype by perturbing transcripts, metabolites and proteins in transcriptional and regulatory networks. In this editorial, I attempt to provide a big picture of each of these three aspects on phenotypic dissection. The genotype-phenotype prediction can be enabled by integrating mathematical models for developmental processes from morphogenesis to pattern formation as well as for transcript, protein and metabolite abundance affecting high-order phenotypes through a series of biochemical steps.

Keywords: Genetic mapping; Complex trait; Genetic architecture; Regulatory network; Dynamic system

Genetic Dissection of Complex Traits

Most quantitative traits of significant importance to agriculture, biology and medicine are determined by multiple genes of unknown number, each being operational to different degrees [1]. The culmination of these genes produces a network of actions and interactions, forming a complex network of genetic architecture. This complexity can be graphically imagined by taking the elements (nodes) of the network to depict main effects of individual genes and the connections (edges) between elements as the effects of genetic interactions (also called epistasis). The concept of genetic architecture can be understood from many different perspectives, but its composite picture can be described by the following factors:

- The number of genes
- The chromosomal distribution of genes
- The main genetic effects of each gene
- The interaction between allelic effects at different genes (epistasis)
- The pleiotropic effects of genes on different traits
- The expression of alleles conditional on the physical or biological environment
- The molecular basis of allelic variation
- The regulatory or coding region of causal variants
- The parent-of-origin effects of alleles or genetic imprinting

The current theory of complex trait genetics is based on the hypothesis that genetic variants in the genetic code, such as single-nucleotide polymorphisms (SNPs), insertions or deletions (indels), and copy number variants, act in concert to determine the phenotypic value of a trait through functional alterations in the activity, expression level, stability, and splicing of the RNA and proteins they encode. Genetic mapping that attributes a phenotypic trait to its underlying quantitative trait loci (QTLs) using polymorphic markers is powerful for mapping the locations of QTLs on the genome and estimating their effects of genetic actions and interactions [2]. As a routine technique of genetic analysis, QTL mapping has been instrumental for studying the genetic architecture of complex traits [3].

Developmental Dissection of Complex Traits

Development includes a broad spectrum of processes. For example in plants, these processes include the formation of a complete embryo from a zygote, seed germination, the elaboration of a mature vegetative plant from the embryo, the formation of flowers, fruits, and seeds, and many of the plant's responses to its environment. Each of these processes is fundamental to determine the size, shape and production of all higher plants. For this reason, knowledge of the genetic basis of the variation in each process is important for understanding adaptive evolution and deriving elite domestic crop varieties. While traditional approaches for mapping QTLs with phenotypes measured at particular times fail to capture the dynamic structure and pattern of the process, two new statistical methods, called functional mapping (incorporated in a package of software FunMap [4,5]) and systems mapping, integrates biological mechanisms and dynamic processes of the trait into the genetic mapping framework through mathematical and computational models [6-11]. Functional mapping unifies the strengths of statistics, genetics, and developmental biology, thus facilitating the test of the interplay between genetic action and development.

The principle of functional mapping can be expanded to map ontogenetic QTLs that govern all developmental events in a plant's lifetime [12]. Previous work for functional mapping focused on the identification of QTLs for a particular phase of development using a mathematical model for growth trajectories during this specific phase. Thus, identified QTLs from this approach cannot be inferred to affect the landscape of ontogenetic growth and development. In plants ontogenetic QTL mapping, three major issues remain to be resolved:

*Corresponding author: Rongling Wu, Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 17033, USA, E-mail: rwu@phs.psu.edu, rwu@bifu.edu.cn

Received May 09, 2012; Accepted June 01, 2012; Published June 02, 2012

Citation: Wu R (2012) Predicting the Genotype-Phenotype Map of Complex Traits. J Biomet Biostat 3:e109. doi:10.4172/2155-6180.1000e109

Copyright: © 2012 Wu R. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Modeling change points or structural changes for a sequence of plant developmental events: Plant development includes several sequential distinct phases, seed development from a complete embryo, vegetative growth after the germination of the seed, flower growth, pollination, and seed formation. Seed development within each phase is a quantitative change over time; mathematical models can well be used to describe the pattern of development. However, the connection of any two adjacent phases through mathematical models is challenging given that they are qualitatively different from each other. New mathematical and statistical models for functional mapping are entailed to map the QTLs that regulate the transition of development from one phase to next.

Modeling epistatic interactions among genes from different regions of seed: Seed development in flowering plants is triggered by a double-fertilization process that leads to the differentiation of the embryo, endosperm, and seed coat that are the major regions of the seed and essential for seed viability and plant reproduction. Many different developmental and physiological events occur within each seed region during development that is programmed, in part, by the activity of different genes. Seed development, therefore, is the result of a mosaic of different gene expression programs occurring in parallel in different seed compartments. Statistical models have been developed for integrating genes from different regions [13,14], but new models should be developed for understanding how these genes are organized into unique regulatory circuits within the plant genome to “make a seed.”

Developing functional genome-wide association studies: To date, most QTLs in plants have been identified by linkage mapping with biparental crosses. With the advent of high-throughput SNP data in plants, genome-wide association studies (GWAS) will soon be coming of age in plants. Functional mapping has been integrated into the setting of GWAS, leading to the birth of a new statistical approach, functional genome-wide association studies or fGWAS [15]. Through fGWAS, a complete number of genes for plant traits can be identified and, more importantly, the physiological and developmental pathways in which these genes function and interact with each other can be characterized. In a platform developed for fGWAS, Bayesian lasso has been developed to select a subset of significant SNPs for developmental trajectories of traits, with the model solved using the Markov chain Monte Carlo algorithm [16,17].

Regulatory Dissection of Complex Traits

The formation of any trait can be modeled as a dynamic system in which various biological parts coordinate to determine a final phenotype through genetic regulation. The behavior and outcome of this system, i.e., trait phenotype, can be changed by altering the pathways of one or more parts. To achieve this, a profound understanding of how different parts are coordinated and organized into a whole system and what are the genetic roots of the function of these parts is crucial.

As the cost of methods for measuring mRNA, protein and other indicators continues to fall, it becomes reasonable to design experiments that capture the dynamic processes of phenotypic formation across timescales. With these data, we can reconstruct biological networks by incorporating transcriptome (the set of RNA transcripts), proteome (the set of proteins), and metabolome (the entire range of metabolites taking part in a biological process) to the functional mapping and systems mapping of final phenotypes.

The true quantitative relationship between the variation in

activity of every one of the thousands of gene-protein couples or protein-metabolite couples in a cellular system can be understood by implementing a high-dimensional system of differential equations (DEs). The DEs that model electronic networks in engineering have been successfully used to map QTLs involved in phenotypic variation [10]. The DEs have power to model several critical factors related to the regulatory network, such as the time displacement of the genetic and protein synthetic and post-translational events, their different timescales and their half-lives [18].

By integrating a system of DEs for regulatory pathways and functional mapping, we have developed a new model, called network mapping, for mapping the underlying transcriptional, proteomic, and metabolomic QTLs and interaction networks among these different types of QTLs [19]. The model has power to test what are the most important pathways that cause final phenotypes and how genes control these pathways. The regulatory network can be predicted by combining environmental and genetic perturbations through network mapping.

Outlook

Genetic analysis of complex traits has now developed to a point at which available approaches allows us to comprehend the genetic architecture of a complex trait and elucidate the rules for translating genetic variation among individuals to the phenotypic variation of the trait. Since the number of genes is usually high, the estimation of genetic effects of each gene becomes highly challenging. The following three strategies are recommended to confront this challenge:

1. Develop and use geometric series to model intrinsic changes of genetic effects over a sequence of genes on chromosomes with a fewer number of parameters. The predictive model constructed on geometric series has been confirmed by results from quantitative genetic analyses of many traits [20].
2. Derive a variable selection approach, such as lasso, to analyze all markers and their interactions at the same time [16]. Since the number of markers may be much larger than the number of samples, special regression approaches equipped by penalty will be developed. By shrinking the effects of a majority of markers to zero we obtain highly sparse estimates of marker effects.
3. Develop a new model for studying gene-environment interactions and charting the genetic basis of phenotypic plasticity for dynamic traits. A dynamic model developed [21,22] can be extended to quantify the genetic control of phenotypic plasticity over a range of discrete environments.

The future direction of genetic studies should focus on the mechanistic and process analysis of complex phenotypes by combining genetic approaches with developmental and regulatory principles underlying trait formation and progression. Such an approach enables geneticists to perform the genetic, developmental and regulatory dissection of complex traits and predict the phenotype from genotype. Computational biologists should collaborate with experimental biologists to study the genetic architecture, developmental interactions and regulatory network of complex traits. This will not only allow the conceptual models to be tested and validated by analyzing real data, but also likely glean a new insight into the genetic roots that drive the formation and development of complex phenotypes.

Acknowledgements

This work is supported by NSF/IOS-0923975, Changjiang Scholarship Award and “Thousand-person Plan” Award.

References

- Mackay TF, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 10: 565-577.
- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185-199.
- Wu R, Ma C, Casella G (2007) *Statistical Genetics of Quantitative Traits: Linkage, Maps, and QTL*. Springer-Verlag, New York.
- Ma CX, Wu R, Casella G (2004) FunMap: functional mapping of complex traits. *Bioinformatics* 20: 1808-1811.
- Tong C, Wang Z, Zhang B, Shi J, Wu R (2011) 3FunMap: Full-sib family functional mapping of dynamic traits. *Bioinformatics* 27: 2006-2008.
- Ma CX, Casella G, Wu R (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 161: 1751-1762.
- Wu R, Lin M (2006) Functional mapping - how to map and study the genetic architecture of dynamic complex traits. *Nat Rev Genet* 7: 229-237.
- Li Y, Wu R (2010) Functional mapping of growth and development. *Biol Rev Camb Philos Soc* 85: 207-216.
- He Q, Berg A, Li Y, Vallejos CE, Wu R (2010) Mapping genes for plant structure, development and evolution: functional mapping meets ontology. *Trends Genet* 26: 39-46.
- Wu R, Cao J, Huang Z, Wang Z, Gai J, et al. (2011) Systems mapping: How to improve the genetic mapping of complex traits through design principles of biological systems. *BMC Syst Biol* 5: 84.
- Fu G, Luo J, Berg A, Wang Z, Li J, et al. (2011) A dynamic model for functional mapping of biological rhythms. *J Biol Dyn* 5: 84-101.
- Zhao X, Tong C, Pang X, Wang Z, Guo Y, et al. (2012) Functional mapping of ontogeny in flowering plants. *Brief Bioinform* 13: 317-328.
- Cui Y, Wu R (2005) Statistical model for characterizing epistatic control of triploid endosperm triggered by maternal and offspring QTLs. *Genet Res* 86: 65-75.
- Cui Y, Wu R (2005) Mapping genome-genome epistasis: A high-dimensional model. *Bioinformatics* 21: 2447-2455.
- Das K, Li J, Wang Z, Tong C, Fu G, et al. (2011) A dynamic model for genome-wide association studies. *Hum Genet* 129: 629-639.
- Li J, Das K, Fu G, Li R, Wu R (2011) The Bayesian lasso for genome-wide association studies. *Bioinformatics* 27: 516-523.
- Li JH, Wang Z, Li R, Wu RL (2012) Bayesian group LASSO for nonparametric varying-coefficient models with application to functional genome-wide association studies. *Ann Appl Stat* (under revision).
- Nicholson JK, Holmes E, Lindon JC, Wilson ID (2004) The challenges of modeling mammalian biocomplexity. *Nat Biotechnol* 22: 1268-1274.
- Wang Z, Wang YQ, Wang NT, Wang JX, Wang ZH, et al. (2012) Towards a comprehensive picture of the genetic landscape of complex traits. *Brief Bioinform* [Ahead of Print].
- Li B, Wu R (1996) Genetic causes of heterosis in juvenile aspen: a quantitative comparison across intra- and interspecific hybrids. *TAG Theoretical and Applied Genetics* 93: 380-391.
- Lacaze X, Hayes PM, Korol A (2009) Genetics of phenotypic plasticity: QTL analysis in barley, *Hordeum vulgare*. *Heredity (Edinb)* 102: 163-173.
- Wang Z, Pang XM, Lv YF, Xu F, Zhou T, et al. (2012) A dynamic framework for quantifying the genetic architecture of phenotypic plasticity. *Brief Bioinform* doi: 10.1093/bib/bbs009.