

SESSION INTRODUCTION: CHARACTERIZING THE IMPORTANCE OF ENVIRONMENTAL EXPOSURES, INTERACTIONS BETWEEN THE ENVIRONMENT AND GENETIC ARCHITECTURE, AND GENETIC INTERACTIONS: NEW METHODS FOR UNDERSTANDING THE ETIOLOGY OF COMPLEX TRAITS AND DISEASE

MOLLY A. HALL

Center for Systems Genomics
Department of Biochemistry and Molecular Biology
Pennsylvania State University
University Park, PA 16802, USA
E-mail: mah546@psu.edu

SHEFALI SETIA VERMA

Center for Systems Genomics
Department of Biochemistry and Molecular Biology
Pennsylvania State University
University Park, PA 16802, USA
E-mail: szs14@psu.edu

DENNIS P. WALL

Department of Pediatrics, Division of Systems Medicine
Department of Psychiatry
Program in Biomedical Informatics
Stanford University
Stanford, CA 94305, USA
E-mail: dpwall@stanford.edu

JASON H. MOORE

Institute for Quantitative Biomedical Sciences
Department of Genetics, Geisel School of Medicine
Dartmouth College
Hanover, NH 03755
E-mail: Jason.H.Moore@dartmouth.edu

BRENDAN KEATING

Center for Applied Genomics
Children's Hospital of Philadelphia
Philadelphia, PA 19104
E-mail: bkeating@mail.med.upenn.edu

DANIEL B. CAMPBELL

Department of Psychiatry and the Behavioral Sciences
Zilkha Neurogenetic Institute
Center for Genomic Psychiatry
Neuroscience Graduate Program
Keck School of Medicine
University of Southern California
Los Angeles, CA 90089
E-mail: dbcampbe@med.usc.edu

GREGORY GIBSON

Center for Integrative Genomics
Department of Biology
Georgia Institute of Technology

Atlanta, GA 30332
E-mail: greg.gibson@biology.gatech.edu

FOLKERT W. ASSELBERGS
Department of Cardiology, Division of Heart & Lungs
University Medical Center Utrecht
The Netherlands
E-mail: F.W.Asselbergs@umcutrecht.nl

SARAH PENDERGRASS
Center for Systems Genomics
Department of Biochemistry and Molecular Biology
Pennsylvania State University
University Park, PA 16802, USA
E-mail: sap29@psu.edu

While genome-wide association studies (GWAS) [1] have identified the genetic underpinning of a number of complex traits, large portions of the heritability of common, complex diseases are still unknown [2-6]. Beyond the association between genetic variation and outcomes, the impact of environment exposure, as well as gene-gene (GxG) and gene-environment (GxE) interactions, are undoubtedly fundamental mechanisms involved in the development of complex traits. Novel methods tailored to detect these predictors have the potential to (1) reveal the impact of multiple variations in biological pathways and (2) identify genes that are only associated with a particular disease in the presence of a given environmental exposure (e.g. smoking). Such knowledge could be used to assess personal risk and to choose suitable medical interventions, based on an individual's genotype and environmental exposures. Further, a more complete picture of the genetic and environmental aspects that impact complex disease can be used to inform environmental regulations to protect vulnerable populations.

Multiple challenges exist for undertaking GxG and GxE analyses. For instance, there can be computational burden and the need to adjust for Type-I error when GxG interactions are explored via an exhaustive combinatorial search. Comprehensive tests for interactions between genome-wide single nucleotide polymorphisms (SNPs) and multiple environmental variables can also be challenging and computationally intensive, especially when using millions of sequenced or imputed variants and more than a dozen exposure measures. Further challenges are faced when exploring GxG and GxE, and the goal of this paper session was to encourage advancements in research and tool development for greater understanding of the impact of environmental exposures, GxG, and GxE on outcomes and traits.

The papers selected for the session, "Characterizing the Importance of Environmental Exposures, Interactions between the Environment and Genetic Architecture, and Genetic Interactions: New Methods for Understanding the Etiology of Complex Traits and Disease", address several current challenges faced in elucidating complex disease when taking into account environmental exposures, as well as GxG and GxE interactions.

Hu *et al.* explained their new approach for identifying GxG interactions using a human phenotype network method for GWAS data in *Genome-wide genetic interaction analysis of glaucoma using expert knowledge derived from human phenotype networks*. In this paper, the authors described their computational knowledge-based network as a filter to focus on the disease of interest. Subsequently, they employed statistical interaction tests to identify a

significant network of pairwise epistatic interactions among the prioritized SNPs to explain the complex nature of Glaucoma.

iPINBPA: An integrative network-based functional module discovery tool for genome-wide association studies by Wang *et al.* describes another novel method for handling genetic associations in a knowledge-driven manner. The authors used this method to identify and prioritize genetic associations by merging statistical data as well as biological evidence for protein interaction. The performance of this novel approach was compared to similar methods using two independent multiple sclerosis genome-wide association studies and one ImmunoChip study. *iPINBPA* demonstrated improvement over other methods in sub-network identification, and thus offers a novel approach for combining GWAS signal with knowledge of protein interactions.

In *Variable selection method for the identification of epistatic models*, Holzinger *et al.* described an approach for feature selection to detect epistatic interactions and applied the method to simulated data. This machine learning method integrates different selection parameters to identify the appropriate threshold between signal and noise and thus generate a list of variants with both interactions and main effects. These variables can further be assessed using powerful modeling tools for interpretation and prediction purposes to improve our understanding of complex human traits.

Patel *et al.* extended the environment-wide association study (EWAS) method by creating exposome correlation globes to account for inter-dependencies between exposure variables in *Development of exposome correlation globes to map out environment-wide associations*. The authors first mapped an “exposome globe”, displaying a comprehensive set of correlations that replicated across multiple datasets from the National Health and Nutrition Examination Survey (NHANES). Next, the “exposome globe” was employed to reveal interrelationships between EWAS results of two different phenotypes: type 2 diabetes and all-cause mortality. The results demonstrate the complex connections between exposures in the context of disease status, which can be applied to both EWAS and GxE interactions in future studies.

In *The global Exposome: a Bipartite Network Approach to Inferring Interactions between Environmental Exposures and Human Diseases*, Darabos *et al.* integrated environmental exposure data with human phenotypes and diseases to build a bird’s eye view of the connections between human diseases and chemical compounds. They used the CDC’s fourth national report on human exposure to environmental chemicals and obtained a total of 60 chemicals in 11 groups that are found in the environment and potentially harmful to human health. The authors also used the diseases of the NHGRI-GWAS catalog and a literature survey to compile a list of the diseases/traits that have been linked to their list of chemicals. The authors utilized this information to build a bipartite network of diseases and chemical substances and projected the network onto the disease/trait space as well as the substances space. There were many features identified in the resultant networks. The dyadicity and heterophilicity of these networks were also explored, highlighting further characteristics of these environmental exposure-human phenotype networks.

A standard approach for GxE interactions is to fit one model per GxE model and then correct for multiple testing. However, two-stage methods can be used, in which potential GxE models are first filtered, testing only specific models passing the filtering step. In *A screening-testing approach for detecting gene-environment interactions using sequential penalized and unpenalized multiple logistic regression*, Frost *et al.* introduced a two-step method using filtering first: an elastic net-penalized multiple logistic regression model to estimate a marginal

association filter statistic, or a gene enrichment correlation statistic for all genetic markers. Next, a multiple logistic regression model was used to jointly assess marginal terms and GxE interactions for all markers passing the filtering step. A likelihood-ratio test was then used to determine if the interaction terms for the evaluated models were statistically significant. The authors evaluated their method with a bladder cancer dataset, showing the statistical benefits of using their novel method for filtering then evaluating GxE models.

The ways in which an exposure variable is measured (questionnaire versus direct lab measurements) may affect GxE analyses. In *Measures of exposure impact genetic association studies: An example in vitamin K levels and VKORC1*, Crawford *et al.* presented results from a study comparing types of exposure measurements for GxE analyses. SNPs in *VKORC1*, a rate-controlling enzyme in the vitamin cycle, were tested for interaction with 2 types of vitamin K measures: survey and serum levels in the third National Health and Nutrition Examination Studies. Results suggested that *VKORC1* associations with vitamin K levels vary by whether environmental measures were survey or serum levels. The authors cautioned that the results of this case study may be relevant for other GxE interaction analyses.

Both GxG and GxE were addressed by Jeff *et al.* in *Identification of genetic modifiers within the fibrogen gene cluster for fibrogen levels in three ethnically diverse populations*. This paper provides an application of GxG and GxE assessment to medically relevant complex traits in multiple ethnic groups using multivariate linear regression. The authors assessed interactions with and without main effects and described the importance of detecting epistasis to explain missing heritability.

Restrepo *et al.* provided an additional expansion of association testing beyond nuclear DNA SNP testing by exploring associations with mitochondrial DNA and complex disease in *Mitochondrial variation and the risk of age-related macular degeneration across diverse populations*. Mitochondria are known to play an important role in health and vision sustainability. There is some evidence that mitochondrial genetic variation may contribute to age-related macular degeneration (AMD). Studies of the contribution of mitochondrial genetic variation to AMD have been limited to populations of European descent. The authors of this paper explored associations between 50 mitochondrial SNPs and AMD in non-Hispanic whites, non-Hispanic blacks, and Mexican Americans, adjusting models for well-known environmental modifiers of BMI and smoking. The authors performed analyses stratified by ancestry using individual SNPs, as well as using three mitochondrial haplogroups. A total of five SNPs were found to be associated with AMD, including three located in the mitochondrial region responsible for the initiation of transcription of the MT genome. No SNPs were found associated in the other ancestry groups, although the sample size was limited within this study.

Knowledge of the nature of environment, genetic interactions, and GxE interactions will help to determine the impact of alterations in biological pathways and identification of genes that are only found to be associated with disease in the context of the environment. This valuable information can be used to assess personal risk and to choose the most appropriate medical interventions based on unique genotype and environmental variation. The papers selected for this section offer novel insights to address the challenges in elucidating the genetic and environmental underpinnings of complex human traits.

References:

1. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106: 9362-9367.
2. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 109: 1193-1198.
3. Frazer KA, Murray SS, Schork NJ, Topol EJ (2009) Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10: 241-251.
4. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, et al. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 11: 446-450.
5. Maher B (2008) Personal genomes: The case of the missing heritability. *Nature* 456: 18-21.
6. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747-753.