A Data Driven Approach to Diagnosing and Treating Disease

Eric Schadt
Icahn School of Medicine at Mount Sinai, New York
212-659-6790, USA
eric.schadt@mssm.edu

ABSTRACT

Throughout the biomedical and life sciences research community, advanced integrative biology algorithms are employed to integrate large scale data across many different high-dimensional datatypes to construct predictive network models of disease. The causal inference approaches we employ for this purpose well complement the types of natural artificial intelligence/machine learning approaches that have become nearly standard in the life and biomedical sciences for building classifiers for a range of problems, from disease classification and subtype stratification, to the identification of responders and non-responders for a given treatment strategy. By building a causal network model that spans multiple scales (from the molecular to the cellular, to the tissue/organ, to the organism and community) we can understand the flow of information and how best to modulate that flow to improve human wellbeing, whether better diagnosing and treating disease or improving overall health(*1-4*). More specifically, we have constructed predictive network models for Alzheimer's disease, along with other common human diseases such as obesity, diabetes, heart disease, and inflammatory bowel disease, and cancer, and demonstrated a causal network common across all of these diseases(*3, 5-10*). Not only do we demonstrate that our predictive models uncover important mechanisms of disease and mechanistic connections among different diseases, but that they have led to a natural way to prioritize therapeutic points of intervention and provide optimal molecular phenotypes for high throughput screening. Our application of these models in a number of disease areas has led to the identification of novel genes that are causal for disease and that may serve as efficacious points of therapeutic intervention, as well as to personalized treatment strategies that provide a more quantitative and accurate approach to tailoring treatments to specific forms of disease.

Categories and Subject Descriptors

J.3 [**Computer Applications**]: Life and medical sciences, Biology and Genetics.

Keywords

Systems biology; multiscale biology; causal network inference; predictive network modeling; drug discovery

REFERENCES

1. A. Califano, A. J. Butte, S. Friend, T. Ideker, E. Schadt, Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nature genetics* 44, 841 (Aug, 2012).

KDD'14, August 24–27, 2014, New York, NY, USA.
ACM 978-1-4503-2956-9/14/08.
http://dx.doi.org/10.1145/2623330.2630817

2. E. Schadt, R. Chang, Genetics. A GPS for navigating DNA. *Science* 337, 1179 (Sep 7, 2012).
3. B. Zhang *et al.*, Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* 153, 707 (Apr 25, 2013).
4. B. A. Kidd, L. A. Peters, E. E. Schadt, J. T. Dudley, Unifying immunology with informatics and multiscale biology. *Nature immunology* 15, 118 (Jan 21, 2014).
5. Y. Chen *et al.*, Variations in DNA elucidate molecular networks that cause disease. *Nature* 452, 429 (Mar 27, 2008).
6. V. Emilsson *et al.*, Genetics of gene expression and its effect on disease. *Nature* 452, 423 (Mar 27, 2008).
7. L. Jostins *et al.*, Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491, 119 (Nov 1, 2012).
8. X. Yang *et al.*, Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. *Nature genetics*, (Mar 8, 2009).
9. H. Zhong *et al.*, Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS genetics* 6, e1000932 (2010).
10. H. Zhong, X. Yang, L. M. Kaplan, C. Molony, E. E. Schadt, Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *American journal of human genetics* 86, 581 (Apr 9, 2010).

Short Bibliography

Dr. Eric Schadt joined Mount Sinai Medical School as Chairman and Professor, Department of Genetics and Genomic Sciences and is the founding Director of the Icahn Institute for Genomics and Multiscale Biology in 2011. Previously, Dr. Schadt had been the Chief Scientific Officer at Pacific Biosciences, overseeing the scientific strategy for the company, including creating the vision for next-generation sequencing applications of the company's technology. Dr. Schadt is also a founding member of Sage Bionetworks, an open access genomics initiative designed to build and support databases and an accessible platform for creating innovative, dynamic models of disease. Dr. Schadt's current efforts at Mount Sinai to generate and integrate large-scale, high-dimension molecular, cellular, and clinical data to build more predictive models of disease so that we may better diagnose and treat disease, were motivated by the genomics and systems biology research he led at Merck to elucidate common human diseases and drug response using novel computational approaches applied to genetic and molecular profiling data. His research helped revolution a field in statistical genetics (the genetics of gene expression), has energized the systems biology field, and has led to a number of discoveries relating to the causes of common human diseases. Dr. Schadt received his B.S. in applied mathematics from California Polytechnic State University, his M.S. in pure mathematics from UCD, and his Ph.D. in biomathematics from UCLA.