

# Embracing the complexity of common human diseases through technology integration

Eric E. Schadt

Current and future generations of whole-genome resequencing, genotyping and gene-expression profiling technologies are expected to enable the discovery of the key disease determinants, which in turn will motivate the development of biomarkers and novel therapeutics that will have a positive impact on diseases of concern to public health. Over the past 10 years, these technologies have matured rapidly, not only providing large-scale, high-throughput and high-quality assays with signal-to-noise ratios that are enhanced beyond what most thought possible just 5 years ago, but also providing these assays at 'bargain basement' prices. The high-throughput, high-quality and low-cost combination has made possible genetic-association and gene-expression studies that have begun to uncover the complexity of common human diseases on an unprecedented scale. Monitoring transcript abundance in disease-relevant tissues has led to the identification of gene networks (patterns of expression) that define subtypes of disease and of the populations that are most likely to respond to a given therapy. One of the first studies to realize this potential identified and evaluated a 70-gene prognosis profile in breast cancer patients using microarray analysis<sup>1</sup>. A more recent study intersected patterns of expression that defined sensitivity/resistance in cancer cells with drug-associated gene-expression profiles to identify novel combination therapies for acute lymphoblastic leukaemia<sup>2</sup>. Similar levels of success have been achieved in genome-wide association studies (GWASs) that systematically test for correlations between DNA and disease-trait variations or states. The first GWAS identified variants in *HTRA1* that were robustly associated with age-related macular degeneration<sup>3,4</sup>. Many similar studies followed, including the WTCCC study that reported novel associations in seven different disease areas<sup>5</sup>, the identification of novel genes involved in the response to HIV-1 infection<sup>6</sup> and several studies of type 2 diabetes that have uncovered more susceptibility genes for this disease in the past 2 years than in all other years combined<sup>7</sup>.

However, despite much success achieved using DNA arrays for genotyping and expression profiling, uncovering the mechanisms by which genes lead to disease has been far slower. Although establishing correlations between DNA changes and disease elucidates the ultimate heritable causes of disease in the human population, such correlations do not necessarily result in a definitive identification of the causal genes, do not explain how a given gene increases disease susceptibility, and do not provide details of the broader biological context in which multiple disease-causing genes operate with respect to susceptibility.

The shortcomings of studies that focus on only a single data dimension (for example, DNA or RNA variation) can be addressed by appropriately leveraging the genetic diversity in experimental and human populations, in which multifactorial perturbations in gene networks that give rise to disease can be explicitly tracked<sup>8,9</sup>. By simultaneously measuring genotypes and transcript abundances (or other molecular phenotypes), we can correlate genetic perturbations with expression and disease-trait variation to identify components of networks that consist of molecular phenotypes that represent disease intermediates, as well as constructing causal networks that define the mechanistic underpinnings of disease. Elucidating the intermediate molecular phenotypes is crucial to forming a comprehensive view of disease, given that changes in DNA do not lead directly to disease, but instead affect the nodes of molecular networks that underlie disease phenotypes. These perturbations to the molecular networks in turn alter key pathways that result in increased disease susceptibility. This type of integrative approach has indeed already led to the identification and validation of novel genes for disease and other complex traits<sup>8,10</sup>, and has enhanced our ability to construct predictive whole-gene networks<sup>11,12</sup>.

To increase the impact that GWAS discoveries can ultimately have on human health, we must place them in the context of the biological networks that drive disease. Indeed living organisms can be defined by a hierarchy of networks. At its base lie genomics networks

that represent interactions among DNA loci that modulate gene activity and that induce synergistic effects on higher order phenotypes such as disease. Genomics networks drive molecular networks that consist of RNA, protein and metabolites, which in turn make up cellular networks. Interactions among cellular networks give rise to the complex phenotypes that define living systems, and that form the fundamental basis of tissue and whole-organism networks. DNA technologies such as whole-genome genotyping and gene-expression arrays provide the first opportunity to assess nodes in this hierarchy of networks comprehensively. As these technologies evolve (as exemplified by the emergence of rapid, high-throughput and low-cost resequencing), the extent to which we can interrogate these networks will increase. However, because common forms of disease emerge from complex network interactions (as a result of genetic and environmental perturbations to the system), experimental design and computational methods must also evolve to maximize the integration of these data, given that no single data dimension will suffice in constructing the holistic view of disease needed to have the greatest impact on human health.

Eric E. Schadt is at Rosetta Inpharmatics, LLC, Merck & Co., Inc., Seattle, Washington 98109, USA.  
e-mail: [eric\\_schadt@merck.com](mailto:eric_schadt@merck.com)

doi: 10.1038/nrg2241

- van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
- Wei, G. *et al.* Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* **10**, 331–342 (2006).
- Dewan, A. *et al.* HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science* **314**, 989–992 (2006).
- Yang, Z. *et al.* A variant of the *HTRA1* gene increases susceptibility to age-related macular degeneration. *Science* **314**, 992–993 (2006).
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
- Fellay, J. *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science* 19 July 2007 (doi:10.1126/science.1143767).
- Frayling, T. M. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nature Rev. Genet.* **8**, 657–662 (2007).
- Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genet.* **37**, 710–717 (2005).
- Lee, S. I., Pe'er, D., Dudley, A. M., Church, G. M. & Koller, D. Identifying regulatory mechanisms using individual variation reveals key role for chromatin modification. *Proc. Natl Acad. Sci. USA* **103**, 14062–14067 (2006).
- Cervino, A. C. *et al.* Integrating QTL and high-density SNP analyses in mice to identify *Insig2* as a susceptibility gene for plasma cholesterol levels. *Genomics* **86**, 505–517 (2005).
- Zhu, J. *et al.* An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet. Genome Res.* **105**, 363–374 (2004).
- Zhu, J. & Schadt, E. E. Increasing the power to detect causal associations among genes and between genes and complex traits by combining genotypic and gene expression data in segregating populations. *PLoS Comp. Biol.* **3**, e69 (2007).