

Photodisc/
Getty/NPC
 BIOINFORMATICS

Next-generation omics

The technologies used to assess the genomic, transcriptomic or metabolic profile of a cell have made impressive strides over the past few years; however, a complete picture of how a system functions will rely on approaches that combine all of these data sets. To this end, Eric Schadt and colleagues have generated a network model that simultaneously integrates several different types of data. Importantly, the network allowed new causal inferences to be made between genetic variants and cellular processes.

As their study population, the authors used the cross between a laboratory (BY) and a wild (RM) strain of *Saccharomyces cerevisiae*. They first obtained the genetic and expression variation data, as well as expression quantitative trait loci (eQTLs), that had previously been gathered on this population. In a second step, quantitative nuclear magnetic resonance (qNMR) was applied to the same population under the same conditions to measure levels of various metabolites; the genetic loci associated with variation in metabolite levels (metQTLs) were then mapped

using >2,000 SNP markers (which track almost all of the variation in this population). In the interaction network produced by combining these three data dimensions, the authors searched for overlap between the location of eQTLs and metQTLs: co-localization might reveal how an eQTL is causally linked to variation in the metabolite level through an effect on gene expression. metQTLs were identified for 16 of the 56 metabolites that could be reliably quantified in this study, and 12 of these metQTLs co-localized to 4 eQTL 'hotspots'.

Encouraged by these results, three additional data levels — namely, DNA–protein, protein–protein and protein–metabolite interactions — were extracted from public databases and incorporated into the network. This final step consisted of generating a probabilistic causal network in which the three new levels of data were used to constrain the search space occupied by the gene expression and metabolite profiles. By focusing on particular portions of the network, the authors could more fully investigate the causal molecular relationships between eQTLs and metQTLs.

This analysis clarified the connection between the eQTL hotspots and their phenotypic effects. For example, the hotspot at the *LEU2* leucine synthesis gene had been linked to alterations in the activity of the Leu3 transcription factor; this study showed that the hotspot affects Leu3 function only indirectly by controlling the levels of an intermediate product. For two of the hotspots, biological links were not previously known but were suggested by the network and were then experimentally validated.

Analytical methods that bring together two data dimensions — such as genetic and expression variation — are already in use. But by pulling together multiple dimensions in one go, this study provides a glimpse into the future of systems biology.

Tanita Casci

ORIGINAL RESEARCH PAPER Zhu J. *et al.* Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biol.* **10**, e1001301 (2012)

FURTHER READING Ghosh, S. *et al.* Software for systems biology: from tools to integrated platforms. *Nature Rev. Genet.* **12**, 821–832 (2011)