

Total information awareness for worm genetics

Sean R. Eddy

Howard Hughes Medical Institute & Department of Genetics,

Washington University School of Medicine

4444 Forest Park Blvd., Box 8510

Saint Louis, Missouri 63108 USA

eddy@genetics.wustl.edu

March 20, 2006

Genetic analysis is an art. Just by studying the effect of mutations in genes, geneticists try to deduce how organisms work. This has been compared to trying to figure out how cars work by waiting outside the factory, tying the hands of individual workers as they go in, and studying the defects of cars that come out [1]. These days, as a result of high-throughput genomics, those metaphorical factory workers - the genes - live in a veritable police state, and geneticists have friends in high places. We can tap into vast databases of gene sequences, gene expression patterns, biochemical interactions, the published literature, and more – so we know their identities, where they go and when, who they talk to, and what’s being said about them. But we face the same problem as less benevolent police states: how can we use unreliable, disparate intelligence databases to learn what they’re actually *doing*? On p. xxx of this issue, Zhong and Sternberg take a step toward total information awareness in genetic analysis [2]. They have developed an integrated database system that predicts genetic interactions in the worm *Caenorhabditis elegans*, one of the

best-studied models of how animals work.

A genetic interaction is one of the more subtle clues that geneticists use to tease apart a biological system. Two genes *A* and *B* “genetically interact” when the phenotype of the double mutant *ab* is unexpectedly not just the combination of the phenotypes of the two single mutants *a* and *b*. An interaction is called “suppression” when *ab* is more normal than expected, and “enhancement” when *ab* is more defective than expected. The “interaction” is thus a logical one between genes, not necessarily a physical one between gene products, and there are many possible mechanistic explanations for it (see Figure). By themselves, genetic interactions are usually not enough to infer how a biological process works. Rather, genetic interactions are used to identify additional genes that might play a role in a process of interest.

One classic example is from studies of the development of the fruit fly eye [3]. Single mutant phenotypes had revealed two important genes, *sevenless* and *boss*, a receptor and a signal at the top of a signalling pathway that tells one cell type (R7) to become a photoreceptor. The genes encoding the signalling pathway remained unknown, and it was thought that maybe it was an essential shared subsystem, so important elsewhere in the fly that any mutations in those genes would be lethal. An elegant enhancer screen was crafted, using a weak mutation of *sevenless* to make R7 cells uniquely sensitive to small perturbations in the signalling pathway; weak mutants in essential genes in the pathway might then produce defective R7 cells but not kill the flies. This approach worked, and seven essential genes involved in the signalling pathway were identified – which turned out to be components of one of the fundamental signalling pathways in biology, the Ras/MAP kinase cascade.

The most beautiful genetic interaction screens involve subtle, hypothesis-driven detective work. Only the simplest screens lend themselves to the assembly lines of modern high-throughput biology. Zhong and Sternberg aim to augment the geneticist, not replace him, by focusing attention on specific suspects predicted by integration of other types of high-throughput data. Recent technologies for targeted disruption of gene function (such as RNA interference) allow a geneticist to test

specific pairs of genes quickly, instead of having to mutagenize and screen the entire genome.

Informative genetic interactions are expected to involve genes that are co-expressed, that might physically interact, and that might show similar single mutant phenotypes. Zhong and Sternberg therefore integrate systematic datasets of *C. elegans* gene expression, physical interaction, and functional annotation curated from the literature. However, these worm datasets are far from complete. A crucial part of their strategy is that not only do they use *C. elegans* datasets, they also integrate comparable datasets from two other major genetic model systems, the fly *Drosophila melanogaster* and the yeast *Saccharomyces cerevisiae*. Many homologous genes function similarly in worms, flies, and yeast, so datasets from one species should be partially informative for other species. All this information is weighted and integrated by a standard statistical classifier (linear regression), trained on known examples of interacting *C. elegans* gene pairs.

Does it work? Zhong and Sternberg test two lists of predicted interactions for two genes in two different signalling pathways (*let-60/ras* in the MAP kinase signalling cascade, and *itr-1*, an inositol triphosphate receptor). Using RNA interference to create double mutants, they looked for either enhancement or suppression relative to single mutant phenotypes and confirmed twelve of the 49 novel predicted interactions they tested for *let-60/ras*, and two of six for *itr-1*. The phenotypes they score are quantitative and far from obvious, like the worm's pharynx pumping 200 times a minute instead of 180 times a minute. This points to an advantage of focusing on a short list of suspects – it would be heroic to detect such small phenotypic effects in a genome-wide mutagenesis screen.

Mind you, Zhong and Sternberg would be the first to tell you that it's not rocket science to guess that two *Caenorhabditis* genes might show a genetic interaction if they are co-expressed, they have similar mutant phenotypes, and the homologous fly genes are already known to genetically interact. The difficulty is not making predictions from the available data – the difficulty is knowing the data are available. Humans lack the time and patience to manually cross-correlate huge genomic databases across several model organisms. Integrative database analysis augments

our strong deductive ability by making up for our limited informational bandwidth.

Zhong and Sternberg have made lists of predicted genetic interactions for every gene in *C. elegans* available at their Web site. Worm geneticists will soon be perusing the lists for their favorite genes. We can expect the same computational technology to be applied to the other model genetic systems, now that Zhong and Sternberg have provided such a clear demonstration of its potential. Someday soon, a *Drosophila* geneticist will download a list of genes that might interact with her favorite wing development gene – a new and happier kind of “no-fly list” produced by database integration technology.

References

- [1] W. T. Sullivan. The salvation of Doug. <http://bio.research.ucsc.edu/people/sullivan/savedoug.html>.
- [2] W. Zhong and P. W. Sternberg. Genome-wide prediction of *C. elegans* genetic interactions. *Science*, xxx:xxx–xxx, 2006.
- [3] M. A. Simon. Signal transduction during the development of the *Drosophila* R7 photoreceptor. *Dev. Biol.*, 166:431–442, 1994.

Figure legend:

Three possible explanations for enhancement, a genetic interaction in which the double mutant *ab* shows an unexpectedly strong visible phenotype relative to either single mutation by itself. The two genes could act in redundant, parallel pathways (top); they could encode subunits of a physical complex that tolerates loss of one subunit, but falls apart upon losing both (middle); or they might have nothing special to do with each other, but the cell’s stressed responses are overwhelmed by the loss of both simultaneously (bottom). Thus, though a genetic interaction suggests that genes A and

B encode functions that are somehow related, we don't immediately know what that relationship is.

