

## GENETICS

## Guilt by phenotypic association

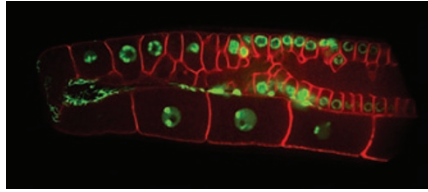
**Phenotypic screens of a complex *in vivo* structure predict function for essential genes in the worm.**

The power of a forward genetic screen is that it picks out, from all the possible genes in a genome, those that are likely to be involved in a process of interest. But following up on the specific molecular function of each hit can be laborious and, when done gene by gene, can take a very long time.

One approach to speed this up is to make predictions about gene function so that one may prioritize candidate genes to study.

Recent work illustrates that good predictions of gene function can be made in large-scale phenotypic analysis of a set of genes for which the function of some of the genes is already known. Using high-content screening of a complex organ, the syncytial worm gonad, a collaborative team at multiple locations across the United States studied mutant phenotypes of a set of genes in *Caenorhabditis elegans*.

“The idea,” explains Kris Gunsalus at New York University, one of the scientists leading the study, “is that you can increase



Confocal image of a live worm gonad with labeled plasma membrane (red) and nuclei (green). Image courtesy of Rebecca Green and Karen Oegema.

phenotypic content by looking at a complex structure.” The project was originated by her collaborator, Karen Oegema at the Ludwig Institute, who wanted to further characterize the phenotypes of the sterile genes. These 554 genes are a subset of *C. elegans* essential genes and have been difficult to study in the early embryo because mutant worms are sterile. When this work began, 166 of these genes had not been previously studied, and for over half of these, no prediction of function could be made at all.

The researchers examined the gonads of worms that had been soaked in double-stranded RNA targeting each of the 554

genes, taking three-dimensional image stacks of gonads with fluorescently labeled plasma membrane and nuclei. In what can only be called a heroic effort, they scored three worms per target gene for 94 possible defects, creating a binary (0 or 1) phenotypic profile covering all defects, for each target gene.

They then manually clustered the profiles into 102 phenotypic classes, each of which contained both previously characterized and uncharacterized genes, and predicted that the unknown genes in a given class would have molecular functions related to those of the other genes in that class—a guilt-by-association approach that is conceptually not dissimilar to genetic interaction screening. As Gunsalus puts it, “If you identify some genes that are involved in a specific function, and there are other genes that also participate in the same process or pathway or even, but not necessarily, [the same] physical complex, then those should display a similar range of phenotypes.”

Using this strategy, the researchers predicted the function of 106 of the previously

## EPIGENETICS

## PROFILING THE SIXTH BASE

**Two complementary methods facilitate genome-wide enrichment of hydroxymethylated DNA.**

In 2009, Anjana Rao and her team from Harvard Medical School characterized an enzyme responsible for producing a modified DNA base, bringing the total number of different building blocks of DNA to six, the standard four bases, methylcytosine and now hydroxymethylcytosine (hmC).

Methylcytosine has been tightly linked to transcriptionally silent chromatin. It is removed in the course of development, and a longstanding question in the field has been whether active demethylation occurs on DNA. The enzyme Rao’s team discovered, a hydroxylase of the TET family, converts methylcytosine to hmC and may be responsible for the first step in demethylation. But far from being a byproduct in an enzymatic cascade, hmC is receiving more and more attention for a role in its own right. It is abundant in embryonic stem cells, and recently a link between hematopoietic cancers and low levels of hmC was discovered.

The difficult question is, how to accurately profile the levels of hmC in a cell. In the last two years antibodies to hmC have been developed, which allow specific pull-downs, but the downside is that these antibodies are dependent on the density of hmC and do not work efficiently in genomic regions with only sparse hmC occurrence. In a recent *Nature* paper, Rao and colleagues introduced two new methods for hmC mapping (Pastor *et al.*, 2011).

Suneet Agarwal, a former graduate student in the Rao group, developed an early idea for the first approach, glucosylation, periodate oxidation, biotinylation (GLIB). “I became very interested in the idea that the hydroxyl could be exploited to covalently tag the base,” he recalls. Using T4 phage glucosyltransferases the team added glucose and, following a method developed by William Pastor, a current graduate student in the Rao lab, oxidized the sugar and then created a reactive group to which biotin was added. Pulling down the biotin-tagged hmC with streptavidin beads allowed the identification of regions bearing hmC.

Conceptually, this approach is similar to a strategy recently developed by Song *et al.*, which uses a synthetic glucose analog to label hmC (Song *et al.*, 2011); the advantage of GLIB is in the fact that no custom-made reagents are needed and that it allows for a very stringent specificity control.

In parallel to the development of GLIB, Yun Nancy Huang, a postdoc in the Rao group, worked on the second approach, cytosine 5-methylenesulphonate (CMS) pulldown. Upon reaction of hmC with sodium bisulphate, CMS is created and can be isolated with specific antibodies. “Although the CMS pulldown still has some density dependence,” says Huang, “a substantial amount of low hmC-containing fragments can still be pulled down with very low background. GLIB and CMS complement each other.”

unknown genes and experimentally validated their predictions for eight genes in five phenotypic classes. They then went on to use the manually clustered phenotypic profiles as a benchmark to develop automated clustering methods that work as well as a human scientist. For this, it was necessary to find a way to rank the functional connections, or phenotypic similarities, between genes. “Correlation-based networks necessarily give you a ton of edges because you have some correlation between everything,” points out Gunsalus. “So you have to find a way to identify which ones are meaningful in a certain biological context.”

The approach is a relatively simple one, in which a measure called the correlation specificity index (CSI) is used to weight each interaction relative to all other interactions in the network. Functional connections at a given level of specificity can thus be simultaneously examined across the network, and the CSI threshold can be adjusted to look at functional connections at different biological levels. Finally, the researchers also used this approach to integrate their newly generated phenotypic network with an already existing one made up of a partially overlapping set of essential *C. elegans* genes and generated through time-lapse imaging of early embryos.

“I was a little surprised but very pleased,” says Gunsalus, “that the CSI worked so well. We’ve been trying for some time to figure out how to level the playing field across a phenotypic network. That we were able to do that to a sufficient degree that we could combine the networks from different analyses, that was really satisfying.”

Natalie de Souza

#### RESEARCH PAPERS

Green, R.A. *et al.* A high-resolution *C. elegans* essential gene network based on phenotypic profiling of a complex tissue. *Cell* **145**, 470–482 (2011).

The team tested both approaches on mouse embryonic stem cells. They compared regions enriched for either methylcytosine or hmC and found overlap within transcribed regions. But only hmC was enriched in transcriptional start sites, 5′ untranslated regions and enhancers. Notably, hmC occurred predominantly at genes with bivalent chromatin marks, genes that are silent in embryonic stem cells but poised for activation during differentiation.

More tools to profile hmC will be invaluable for better characterization of this modification in the context of different cell types and developmental states, and this will help address the big outstanding questions in this field. The specific role of hmC during differentiation and malignant transformation still remains to be discovered. And the question of whether active demethylation actually occurs or whether hmC blocks re-methylation also still awaits answering.

Though all enrichment methods developed so far are very useful in providing a genome-wide picture of where regions enriched in hmC are, Agarwal calls it the “500-foot view”; the ‘ground-level view’ is still needed: which base is an hmC, and what are its neighbors? Such single-base resolution will require approaches that directly detect the modification on a stretch of DNA.

Nicole Rusk

#### RESEARCH PAPERS

Pastor, W.A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).

Song, C.X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5 hydroxymethylcytosine. *Nat. Biotechnol.* **29**, 68–72 (2011).

#### STRUCTURAL BIOLOGY

##### Improving molecular replacement

Molecular replacement is a standard approach for reconstructing three-dimensional protein structures from crystallography data, but this method usually fails for proteins with less than 30% sequence identity to the nearest homologous structure. DiMaio *et al.* show that the Rosetta structural modeling program, used in combination with algorithms for crystallographic structure determination, could generate high-resolution structures for several proteins that could not be solved using traditional methods.

DiMaio, F. *et al.* *Nature* **473**, 540–543 (2011).

#### BIOPHYSICS

##### Mapping cell shear stress

Krieger *et al.* describe cysteine shotgun mass spectrometry, a method to map how cysteines are exposed in proteins in cells by mechanical, thermal or drug-based stress. Exposed cysteine residues are labeled with fluorescent thiol probes; the labeled proteins can be imaged and also analyzed for site-specific cysteine modifications by shotgun mass spectrometry. The authors applied the approach to look at the cytoskeletal proteins spectrin, actin and ankyrin in red blood cells that are exposed to shear stress.

Krieger, C.C. *et al.* *Proc. Natl. Acad. Sci. USA* **108**, 8269–8274 (2011).

#### GENOMICS

##### Trinity assembles transcripts

To fully exploit the potential of high-throughput transcript sequencing (RNA-seq), analysis algorithms are essential. Grabherr *et al.* now add Trinity, a *de novo* assembler, to the toolbox. First, the Inchworm module assembles short reads into contigs, then Chrysalis clusters overlapping contigs and constructs de Bruijn graphs, and finally, Butterfly reconstructs full-length transcripts for each isoform. The authors used Trinity to analyze RNA-seq data from fission yeast, mouse and whitefly.

Grabherr, M.G. *et al.* *Nat. Biotechnol.* advance online publication (15 May 2011).

#### NEUROSCIENCE

##### Visualizing neuronal connections by MRI

Wu *et al.* report a new compound that can be used to trace neuronal connections in the *in vivo* rat brain using magnetic resonance imaging (MRI). They conjugated the anatomical tracer cholera-toxin subunit-B with gadolinium-chelate to generate a magnetic resonance-visible compound. The tracer is transported monosynaptically, whereas manganese chloride—a tracer currently used to trace neuronal connections with MRI—is transported multisynaptically, which can make the interpretation of connectivity experiments more difficult.

Wu, C.W.-H. *et al.* *Neuron* **70**, 229–243 (2011).

#### BIOPHYSICS

##### Minimum information about a simulation experiment

Minimum information guidelines are written with the aim to improve scientific reporting in a number of fields by providing clear instructions to researchers for providing sufficient information in publications to allow others to reproduce the work. Waltemath *et al.* now present the minimum information about a simulation experiment (MIASE). These guidelines describe the minimal set of information that is needed to describe a computational simulation experiment in a publication.

Waltemath, D. *et al.* *PLoS Comp. Biol.* **7**, e1001122 (2011).