nature genetics

From parts catalog to wiring diagram

Given that we know of so many levels of regulation in molecular biology, and given the possibility that a large proportion of even the expressed genomic elements remain to be identified (*e.g.*, *Cell* **116**, 499–509; 2004), surely it is premature to ask for a wiring diagram of the genome before we have the parts catalog?

Maybe not: the catalog may come with assembly hints. A productive genome annotation data set emerged from the intersection of the set of conserved noncoding elements with the set of elements binding transcription factors, data verified by experimental replication and validated by an independent technique in yeast (*Nature* **431**, 99–104; 2004). The regulatory relationships of genes and factors have been built into motifs and networks, yielding fascinating clues to their function (*Science* **298**, 799–804; 2002). Indeed, Andrew Fraser and Edward Marcotte have argued that the compete set of relationships of a protein or DNA element is itself the definition of function most suitable for exhaustive analysis of every element of the genome (*Nat. Genet.* **36**, 559–564; 2004).

But what happens if the parts resemble a puzzle with a very large number of assembly possibilities? Do we then need a genome network architecture initiative (a 'GeNetMap'?) to tackle the wiring diagram of the genome? One answer is no; this is what many are already aiming to deliver. Competitive hypothesis-and-model–driven research building on and inspired by comprehensive annotation projects such as ENCODE (http://www.genome.gov/ENCODE/) will naturally follow once we have all the parts. On the other hand, few have articulated a deliverable description of a genomic network for even a model organism. For example, how will we think of and deal with the functional topology of most of the genome that lies outside the existing regulatory frameworks of metabolic pathways and transcription networks? Ideas for this endeavor will come from collaborations with mathematicians and with chemical and electronic engineers. These investigators will need more incentive for risking interdisciplinary collaboration than the challenge of racing against the annotators. Historically, the nature of the gene, the genetic code and the DNA structure were captured by alert theorists who changed horses in time to head off the stampede of data at the nearest pass. Now, as then, collaborating biologists were essential, not only to generate data, but also to select and investigate the biological systems in which the hypothetical models were tested.

As the genome annotators work from the parts up, geneticists work toward them from function down. This complementary approach may be essential for functional analysis to indicate where parts have been missed (for instance, a transposon insertion can uncover a gene that slipped through the annotation process). At a meeting last month, fly researchers demonstrated the feasibility of a *Drosophila* ENCODE project that would annotate a whole compact genome from an organism with gene structure and multicellular complexity comparable to the human but with 5% of the genome size. We strongly support such a proposal, because this openly cooperative community of some 1,600 labs is equipped with advanced genetic tools to deliver functional insights.

Genetics also has tools for inferring architecture from functional perturbation that can be powerful if applied systematically, as a paper (on p 77) in this issue shows. Epistasis analysis is traditionally used to infer the formal order of gene action within linear portions of modules (the subnetworks of genes regulating a common process or phenotype) by phenotypic comparison of single and double gene knockouts. Roy Kishony and colleagues show that the technique can now be used systematically to group genes into a functional modules by making use the set of aggravating or buffering interactions between modules. Their proposal frees a powerful genetic technique from its current restriction to linear pathway segments and forges an important experimental tool for connecting modules into a genomic network.

Even if the metaphor of logical circuit design applies, is there a wiring diagram to uncover? Like the metabolic diagram, transcriptional networks inferred from binding site occupancy show the potential regulatory pathways available to a genome, only a subset of which may apply to a given cell under a variety of conditions. It is conceivable that different cell types differentiate to use entirely different regulatory strategies to achieve convergent results. Therefore, a summary network model of the regulatory potential of an average cell might prove misleading. Parsimony in regulatory evolution cannot be taken for granted (*Proc. Natl. Acad. USA* **100**, 13356–13361; 2003), but the common core of housekeeping genes and the systematic progression of cell types in the developmental program argue there are common regulatory principles between cell types. Even a partial wiring diagram will be useful far beyond the sum of its parts. ■