nature genetics

Systems biology unlimited

Sequencing technologies have unleashed more than enough quantitative data to test systems models of genome function, and sequence data are now driving a new systems biology. The new RNA entities uncovered may require new concepts of how genomes regulate their own expression.

There is no doubt that increased capacity for DNA sequencing has revolutionized the study of genome variation and gene expression. On page 535 Patrick Tarpey and colleagues have sequenced all the X-chromosome exons in 208 families with X-linked mental retardation. This huge experimental program raises the bar both for the use of resequencing in association studies and for the search for mendelian mutations. One rather unexpected finding of this research is that loss of function of 1% of X-chromosome genes is apparently compatible with normal life and function.

But nowhere has the flood of new sequence data altered our view of the genome more than in the RNA world. Thanks to research published in the last six months, we now have a profound appreciation of the extent of both alternative splicing (Editorial, *Nat. Genet.* **40**, 1385; 2008) and noncoding RNA (Piero Carninci, *Nature* **457**, 974–975; 2009).

Following an impressive feat of international organization, the FANTOM Consortium and the Riken Omics Science Center (p 553) report their examination of the dynamic quantitative expression changes of transcriptional start sites genome-wide in monoblasts of a myeloid leukemia cell line undergoing cell cycle arrest and monocyte differentiation. Systematically identifying transcription factor-binding motifs at these start sites, they have built and tested their transcriptional network model by monitoring the transcriptional consequences of siRNA knockdowns of a considerable number of transcription factors. A reassuring number of leukemogenic transcription factors indeed have important roles in arrest and differentiation, but it is clear that the regulatory network is more complex than hierarchic "master regulator" models would imply.

How much more complicated can it be? Evolutionary and developmental constraints suggest that the complete analysis of this transcriptional network, even if computationally intensive, cannot be impossibly complicated and need not involve modeling of all possible regulatory architectures. Transcription networks must have evolved from smaller networks comprising simple modular elements. The transcription patterns typical of undifferentiated and differentiated states have acquired the homeostatic 'attractor' property that allows a particular cell identity to resist most perturbations in gene expression. Cells differentiate relatively quickly, so they cannot afford regulatory strategies entailing cycling through long chains of global transcriptional states.

However, how can we be sure we have all the components? One problem for the future of systems genomics is that the protein-coding genome is a tiny minority of what is actually there and a tiny proportion of the regulatory capacity of the genome. Geoffrey Faulkner and colleagues (p 563) remind us that there is much well-regulated transcriptional activity from the repetitive portion of the genome, in particular the retrotransposons that show dramatic tissue-specific expression differences. Many are positioned to interact extensively with the protein-coding genome. This massive study shows the impressive regulatory potential of this genomic component but raises many more questions than this initial characterization answers.

Finally, Piero Carninci raises an intriguing possibility in his *Nature* News and Views that we may be partially looking in the wrong place for regulators. Ryan Taft and colleagues (p 572) find characteristic tiny RNAs associated with 5′ ends of many genes in humans, mice and flies that, judged by their consistent size and position, are real genomic entities rather than processed artifacts. Many overlap permissive chromatin marks associated with active promoters and could even have specific functions. Much of the regulation of promoter activity may therefore be the province of noncoding RNAs, rather than just the familiar DNA-binding protein transcription factors. ■