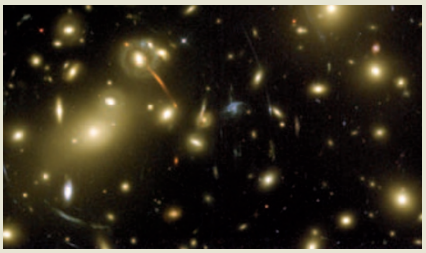
COSMOLOGY

Lenses under the lens



NASA, A. FRUCHTER & THE ERO TEAM [S. BAGGETT (STSCI), R. HOOK (ST-ECF), Z. LEVAY (STSCI)] (STSCI)

If you were to pop into a cosmology conference today, the chances are that you would see this image in at least one presentation. It is a striking snapshot of a cluster of galaxies acting as a gravitational lens: the cluster bends light from galaxies lying behind it and 'smears' the light to produce multiple images and giant arcs.

As pretty as their effects are, gravitational lenses are giving cosmologists a few headaches. For example, the observed incidence of giant arcs and their distance from the clusters' centres, which marks the size of features called Einstein rings, indicate that these clusters may have a stronger 'lensing' ability than expected in the framework of the currently accepted model of the cosmos. In a paper to appear in *Astronomy & Astrophysics*, Meneghetti *et al.* describe an analysis that advances our understanding of these systems (M. Meneghetti *et al.* Preprint at http://arXiv.org/abs/1103.0044; 2011).

The authors compared the lensing ability of a numerically simulated sample of clusters with that of a sample of well-characterized, X-ray-luminous clusters obtained by the MAssive Cluster Survey (MACS). In contrast to earlier studies, their simulations factor in elements known to affect lensing power — for example, the fact that the lenses are complex three-dimensional structures. They found that the simulated clusters produce 50% fewer arcs than do the observed MACS clusters, and that the median size of Einstein rings differs by 25% between the two samples. These are much smaller discrepancies between theory and observation than previously reported. But as the authors themselves concede, more data are needed to confirm their findings. **Ana Lopes**

promoters. Of the more than 2,000 putative promoters, 50% are already confirmed[13]. The locations of about 14,500 putative *cis*-regulatory elements were also identified. Unexpectedly, one class of active promoters does not contain the characteristic chromatin mark H3K4me3, suggesting that the genes they regulate use an alternative mode of transcriptional initiation.

Integrating the binding patterns of all transcription factors leads to hypotheses of transcription-factor partnerships, involving co-binding to regulatory elements[5,7]. But overlays of transcription-factor binding should be interpreted cautiously, particularly for factors with non-tissue-specific or partially overlapping expression: regions that are co-targeted by multiple factors are not necessarily co-bound in the same cells. Nevertheless, the complexity of some co-targeted regions is intriguing. The modENCODE researchers identified regions in the genomes of both *Drosophila*[5,7] and the nematode *Caenorhabditis elegans*[14] — the other model organism on which the project focuses — that are highly occupied by transcription factors. It remains to be determined what function, if any, such regions have in transcription.

This first phase of modENCODE has made a significant impact on refining the annotation of the *Drosophila* genome, which forms the foundation of a large body of research conducted in this organism. But where should the project go from here? First, there is the issue of completion. With the new data, the annotation of genes may be 80% complete, but the job is far from over. Despite the huge depth of coverage, almost 1,500 known genes could not be identified in any experiments[4]. Analysis of specific subpopulations of cells and tighter staging of the developmental process should greatly improve sensitivity.

Completing annotation of the 'regulatory genome' is much more challenging. Although the location of putative enhancer elements can be identified, determining which of these regions are functional, and when, is a huge task. Understanding the regulation of enhancer activity requires knowledge of which transcription factors are binding to them, in which cell types, and when. Scaling this up to the roughly 700 predicted *Drosophila* transcription factors is a monumental undertaking, but feasible given current tagging technologies[15,16].

A major drawback of the data sets is their lack of temporal and spatial resolution. Although cells in culture are extremely useful for identifying core properties of basic cellular processes, such immortalized cells, devoid of their developmental context, cannot substitute for cells within a developing embryo. On the other hand, whole-embryo studies provide merged signals from all cells in the embryo, giving no information on the tissue in which a gene, promoter or chromatin state is active. Many of the transcription factors examined are expressed across a broad range of tissues, which has the advantage of covering a wide range of *cis*-regulatory elements. But merged transcription-factor occupancy signals from multiple tissues make it very difficult to disentangle regulatory connections and thus to build reliable regulatory networks.

The general absence of functional information is perhaps the most serious limitation of the current work and a major challenge for all genomics projects. Such information is essential to understand the relevance of regulatory connections. Examining mutants was understandably beyond the scope of the present studies, but, moving forward, there is a clear need to integrate diverse types of functional data in order to make the transition from correlations to regulatory function. The thousands of *Drosophila* mutants available should provide a useful resource for this.

We can view this work[3–5] as an important chapter in a long book. The data — all freely available[17] — provide an excellent resource for identifying putative genes and regulatory elements that might be active at a particular stage of development. The sheer volume of new transcripts and putative regulatory elements, and the inherent complexity of their interactions, demonstrates how far the project has come, but also highlights the challenges that lie ahead to convert this wealth of information into regulatory networks that describe the transformation of a fertilized egg into a complex multicellular organism. To reach this goal, researchers must integrate new types of experiments that will address the function of, and connections between, genomic regions at high spatio-temporal resolution. With this in mind, we can envisage a next phase of exciting studies that will tackle these issues, and so look forward to seeing what comes next. ■

Eileen E. M. Furlong *is at the Genome Biology Unit, European Molecular Biology Laboratory, D-69117 Heidelberg, Germany.*
e-mail: furlong@embl.de

1. Adams, M. D. *et al. Science* **287,** 2185–2195 (2000).
2. Celniker, S. E. *et al. Nature* **459,** 927–930 (2009).
3. Graveley, B. R. *et al. Nature* **471,** 473–479 (2011).
4. Kharchenko, P. V. *et al. Nature* **471,** 480–485 (2011).
5. Nègre, N. *et al. Nature* **471,** 527–531 (2011).
6. Cherbas, L. *et al. Genome Res.* **21,** 301–314 (2011).
7. The modENCODE Consortium *Science* **330,** 1787–1797 (2010).
8. Filion, G. J. *et al. Cell* **143,** 212–224 (2010).
9. Papp, B. & Müller, J. *Genes Dev.* **20,** 2041–2054 (2006).
10. Kwong, C. *et al. PLoS Genet.* **4,** e1000178 (2008).
11. Oktaba, K. *et al. Dev. Cell* **15,** 877–889 (2008).
12. Enderle, D. *et al. Genome Res.* **21,** 216–226 (2011).
13. Hoskins, R. A. *et al. Genome Res.* **21,** 182–192 (2011).
14. Gerstein, M. B. *et al. Science* **330,** 1775–1787 2010).
15. Venken, K. J. T. *et al. Nature Methods* **6,** 431–434 (2009).
16. Ejsmont, R. K., Sarov, M., Winkler, S., Lipinski, K. A. & Tomancak, P. *Nature Methods* **6,** 435–437 (2009).
17. www.modencode.org

© 2011 Macmillan Publishers Limited. All rights reserved