

## GENETICS

# Revealing the Dark Matter of the Genome

Mark Blaxter

Animal embryos successfully transform the two-dimensional code of their genome into multidimensional organisms that are ready to meet the challenge of natural selection. In addition to the three dimensions of the body, animal genomes inform additional dimensions: of cells coordinating to form tissues,

by embryos, but two articles in this issue, by Gerstein *et al.* on page 1775 (1) and the modENCODE Consortium on page 1787 (2), bring this goal closer.

The genomes of multicellular animals are big and complex, but functions have been defined for only a small proportion of them. Only 1% of the human genome

Integrated data sets from two animal model organisms provide insights into the organization, structure, and function of their genomes.

expression posttranslationally, and marking of the histone proteins on which the DNA is wound with chemical tags to define regions of the genome that are active or silent.

One could analyze this complex regulatory landscape one factor or region at a time, but this would miss the big picture. The ENCODE (Encyclopedia of DNA Elements) projects are using large-scale, genome-wide assays to identify the interactions among transcription factors, ncRNA, chromatin marks, and gene expression—seeking functions for the dark genome. Initial data from the human ENCODE project (3) revealed an incredible density of regulatory marks and interactions on a small portion (1%) of the human genome. The modENCODE (model organism Encyclopedia of DNA Elements) strand of the project is using the power of model organism genomics to reveal genome-wide patterns of regulatory interactions (1, 2).

Model organisms, such as the fruit fly *Drosophila melanogaster* and the nematode *Caenorhabditis elegans*, are chosen for many reasons, including ease of cell culture and amenability to experimentation. In the era of genome science, one of their key benefits is their small genomes: 100 million bases (Mb) for *C. elegans* (5) and 180 Mb for *D. melanogaster* (6) [compared to the 3000-Mb human genome (7, 8)]—and a much larger proportion of their genomes shows signatures of evolutionary constraint. Both models have been examined with a huge armory of genetic and molecular tools, and our understanding of how their embryos develop, and how the adult organisms function, exceeds that for any other animal. Their role in modENCODE is to pilot technologies, especially those of data analysis, and to provide reference points for the emerging human ENCODE data. Fruit fly and nematode modENCODE projects have performed hundreds of experiments and produced billions of data points to permit the



**Computing the organism.** Integrated datasets across transcription, epigenome, and protein-DNA interactions describe the dynamic regulation of gene expression in the nematode and fly model organisms.

tissues functioning together as organs, and organs shaping the body's systems; and of individuals responding appropriately to the varied challenges of life and surviving to breed. Poisons in food are detoxified, pathogens are killed, parasites are eliminated, and predators avoided through the deft employment of responses encoded in the genome. It is not currently possible to compute an organism from its genome, performing the transformation so efficiently executed

is transcribed into protein-coding messenger RNA (mRNA) and non-protein-coding RNA (ncRNA), and DNA elements that control the expression of genes occupy another ~0.5%, suggesting that the remaining “dark genome” is nonfunctional padding. However, 5% of the human and other mammalian genomes are under evolutionary constraint, suggesting biological functions (3, 4). What are these functions and how are they integrated? Three interacting systems coordinate gene expression in space and time: transcription factors that bind to DNA in promoters of genes, ncRNA that modifies gene

Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. E-mail: mark.blaxter@ed.ac.uk

building of new models of gene expression regulation. These can be used to describe the idiosyncratic development and biology of each animal, but the excitement lies in the commonalities in the overall structures of the regulatory landscape they reveal.

Before modeling gene expression patterns, one first has to know what genes are present. Despite the deep annotation available for the fruit fly and nematode genomes, both projects have identified many new genes and parts of genes. In *C. elegans*, Gerstein *et al.* found evidence supporting 95% of existing protein-coding gene predictions, but 1650 new genes are now added for a total of ~22,000. The number of RNA transcripts identified is now triple the previous estimate (to about three per gene), and the ncRNA set is greater by a factor of 20. Evidence obtained by the modENCODE Consortium increases the *D. melanogaster* gene set by a similar amount, to ~17,000 distinct genes. Both reports suggest that the gene catalogs for the two model organisms may now be essentially complete, with new discoveries inherently indistinguishable from biological noise and likely to be unimportant.

The model animals have a similar complexity of patterns of chromatin marks and their correlations with gene expression. Using data for 18 different histone modifications in a *D. melanogaster* cell line, the modENCODE Consortium identified 30 chromatin states that are associated with different gene expression patterns and gene positions. *C. elegans* lacks some chromatin states found in *D. melanogaster* (such as heterochromatin—repressed DNA that makes up ~30% of the *D. melanogaster* genome), but Gerstein *et al.* found that chromatin states are similarly predictive of nematode gene expression patterns, including expression of ncRNAs.

A third shared discovery of the modENCODE teams is the very high degree of connectivity, and beguiling simplicity, in the regulatory systems (9). Regulators (transcription factors and ncRNAs) function in hierarchies with few levels, in which master regulators control many other regulators. These, in turn, feed back in a set of simple network connection motifs involving ncRNAs. In *D. melanogaster* each regulator is, on average, only two (and no more than five) links away from any other. Predictive models of gene expression based on their regulatory interactions were built and tested against observed gene expression patterns. In developing *D. melanogaster* embryos, the model predicted 62% of the expression patterns observed in isolated cell lines. Given the stochasticity of biological systems, this is a remarkable achievement.

Both Gerstein *et al.* and the modENCODE Consortium report a curious class of short (100-base) elements in the genomes called highly occupied target (HOT) regions (10, 11). HOT regions were repeatedly identified as binding many different transcription factors, but are curiously not enriched in the known DNA motifs to which these factors bind, suggesting that the interactions may be indirect. HOT regions are stable and associate with gene transcription start sites, and in *C. elegans* they are associated with genes that are universally expressed through development at high levels. In *D. melanogaster*, HOT sites are also sites of binding by proteins involved in originating DNA replication. Both studies identified novel sequence motifs that are enriched in HOT regions, but these motifs are not shared between the two species and most do not match known transcription factor binding sites, suggesting that the proteins that bind to these motifs are yet to be identified.

The Large Hadron Collider is the preeminent, long-term cooperative enterprise in the physical sciences, dedicated to gathering

data to fully parameterize the basic physical constants of the universe and understand dark matter. In the same way, the modENCODE and ENCODE deep genomics programs will, in time, deliver the power to model and predict organism function from multidimensional data, shine light on the dark genome, and hopefully allow a better understanding of the healthy human and how to treat human disease.

#### References

1. M. B. Gerstein *et al.*, *Science* **330**, 1775 (2010).
2. modENCODE Consortium *et al.*, *Science* **330**, 1787 (2010).
3. ENCODE Project Consortium, *Nature* **447**, 799 (2007).
4. L. Eory, D. L. Halligan, P. D. Keightley, *Mol. Biol. Evol.* **27**, 177 (2010).
5. *C. elegans* Sequencing Consortium, *Science* **282**, 2012 (1998).
6. M. D. Adams *et al.*, *Science* **287**, 2185 (2000).
7. J. C. Venter *et al.*, *Science* **291**, 1304 (2001).
8. International Human Genome Sequencing Consortium, *Nature* **409**, 860 (2001).
9. U. Alon, *Nat. Rev. Genet.* **8**, 450 (2007).
10. S. MacArthur *et al.*, *Genome Biol.* **10**, R80 (2009).
11. C. Moorman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12027 (2006).

Published online 22 December 2010;  
10.1126/science.1200700

#### MATERIALS SCIENCE

## Stretching Dielectric Elastomer Performance

Federico Carpi,<sup>1</sup> Siegfried Bauer,<sup>2</sup> Danilo De Rossi<sup>1</sup>

Devices using materials that deform in response to electricity are based on a phenomenon that was observed more than two centuries ago.

The idea that a solid material can deform when stimulated by electricity originated in the late-18th century with observations of ruptures in overcharged Leyden jars, the first electrical capacitors. In 1776, Italian scientist Alessandro Volta mentioned in a letter that Italian experimenter Felice Fontana had noted volume changes in the Leyden jar upon electrification (1), an observation that launched a new field of investigation—“deformable” materials affected by electricity. More than two centuries later, the concept of “electrically stretchable materials” is at the forefront of devising bioinspired robots, tactile

and haptic interfaces, and adaptive optical systems (2, 3).

This diversity of applications took a great leap 10 years ago in a landmark study by Pelrine and colleagues (4). They reported high-speed, giant-strain, electrically actuated elastomers with unprecedented electromechanical transduction performance. These materials were demonstrated for so-called dielectric elastomer actuators, deformable capacitors made of a film of a soft insulator (such as acrylic, polyurethane, or silicone elastomer), with compliant electrodes. Upon electrical charging, purely electrostatic forces caused the elastomer film to undergo substantial thickness compression and surface expansion (4). The exceptional performance of these dielectric elastomer actuators gave rise to a scientific and technological revolution in the field of

<sup>1</sup>Interdepartmental Research Centre “E. Piaggio,” School of Engineering, University of Pisa, Pisa 56100, Italy. <sup>2</sup>Department of Soft Matter Physics, Johannes Kepler University, Linz A-4040, Austria. E-mail: f.carpi@centropiaggio.unipi.it