

GENOMICS

Mysteries of heterochromatic sequences unravelled

Hitherto considered the junkyard of the genome, heterochromatin is now recognized as an important part of eukaryotic genomes, with functions that include chromosome segregation, nuclear organization and regulation of gene expression. And yet, owing to technical difficulties, all of the currently available genome sequences focus on euchromatin only. This is now set to change — two recent papers describe sequence finishing, mapping and annotation of heterochromatin in *Drosophila melanogaster*. Between them, they provide insights into the genomic organization of heterochromatic regions and pave the way for similar studies in other organisms.

In the first study, Hoskins and Carlson *et al.* re-analysed the fly whole-genome shotgun sequence (WGS3). The repetitive nature of heterochromatin hinders efficient assembly of individual reads into scaffolds. To overcome this problem, the authors selected a set of 10-kb genomic clones to fill the gaps in sequence; for higher-level assembly, they relied on BAC-based physical mapping and BAC-end sequences. The result was ~15 Mb of finished or improved heterochromatic sequence (out of 20 Mb in total), with 50% in scaffolds greater than 378 kb.

Using fluorescence *in situ* hybridization (FISH) (with single-copy probes) they created an integrated physical and cytogenetic map of the pericentromeric heterochromatin, which they used to order, orient and link scaffolds into larger contigs. Although the authors admit that new technological and computational advances are needed to study highly repetitive regions, they have shown that single-copy and middle-repetitive components of heterochromatin should be within our reach.

In the second paper, Smith and colleagues describe their computational and manual annotation of heterochromatic sequences from the same genome release. They estimate that, in the fly, heterochromatin contains ten times more repeats and transposons than euchromatin; in this respect, it resembles human euchromatin.

As well as non-protein coding genes and pseudogenes, they identify 230–254 protein-coding genes, many of which are highly conserved in other *Drosophila* species. Interestingly, DNA- and protein-binding domains are overrepresented among heterochromatic genes, prompting the authors to speculate whether these genes might in fact contribute to heterochromatin structure and function. It seems

that all nuclear genes on average have similar numbers of exons and transcripts, but heterochromatic introns are on average five times longer. Unlike in the euchromatin, intron lengths tend not to be conserved between orthologues, perhaps because their repetitive nature makes them prone to expansions and contractions. Moreover, the highly repetitive nature of gene-regulatory regions raises a possibility that regulation of heterochromatic gene expression is different from the euchromatic process.

The results, fully integrated with those for euchromatin, are now available through FlyBase and GenBank. As the authors point out, now that heterochromatin is revealing its closely guarded secrets, the similarities between it and euchromatin seem more striking than the differences.

Magdalena Skipper

ORIGINAL RESEARCH PAPERS Hoskins, R. A. & Carlson, J. W. *et al.* Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science* **316**, 1625–1629 (2007) | Smith, C. D. *et al.* The release 5.1 annotation of *Drosophila melanogaster* heterochromatin. *Science* **316**, 1586–1591 (2007)

FURTHER READING Grewal, S. I. & Jia, S. Heterochromatin revisited. *Nature Rev. Genet.* **8**, 35–46 (2007) | Matthews, K. A. *et al.* Research resources for *Drosophila*: the expanding universe. *Nature Rev. Genet.* **6**, 179–193 (2005)

WEB SITES

FlyBase: <http://flybase.bio.indiana.edu>

GenBank: <http://www.ncbi.nlm.nih.gov>

RESEARCH HIGHLIGHTS ADVISORS

MICHAEL AKAM

University of Cambridge, UK

SEAN B. CARROLL

University of Wisconsin, USA

NANCY J. COX

University of Chicago, USA

SUSAN FORSBURG University of Southern California, USA

RALPH J. GREENSPAN

The Neurosciences Institute, California, USA

YOSHIIHIDE HAYASHIZAKI

Riken Genomic Sciences Center, Japan

MARK JOBLING

University of Leicester, UK

PETER KOOPMAN

University of Queensland, Australia

LEONID KRUGLYAK

Fred Hutchinson Cancer Research Center, USA

BARBARA MEYER

University of California, Berkeley, USA

JOHN QUAKENBUSH

Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, USA

JANET ROSSANT

Mount Sinai Hospital, Toronto, Canada

MARC VIDAL Dana-Farber Cancer Institute, Boston, USA

VIRGINIA WALBOT

Stanford University, USA

DETLEF WEIGEL Max Planck Institute for Developmental Biology, Germany

PHIL ZAMORE

University of Massachusetts, USA

LEONARD I. ZON

Children's Hospital, Boston, USA