



GENOMICS

The best-laid plans of mice and men

In a coordinately published set of papers in *Nature* and other journals, the Mouse ENCODE (Encyclopedia of DNA Elements) consortium presents comprehensive data sets of functional elements in the mouse genome, as well as the first in-depth analyses of these resources. In addition to insights into the regulation of the mouse genome and its evolution, the studies reveal pivotal similarities and differences with the human genome. A list of the >1,000 generated data sets is available at [Mouse ENCODE](#), and a few of the key studies and their findings are described here.

Similar to the ENCODE project, which was developed to investigate the functional elements of the human genome, the Mouse ENCODE data represent a comprehensive catalogue of the genes and non-coding functional sequences in the mouse genome. As described in the overview by Yue *et al.*, the researchers used several high-throughput approaches across >100 mouse cell and tissue types to distinguish species-specific transcriptional and cellular regulatory programmes and those shared with humans. RNA-seq (RNA sequencing) was used to map coding regions, revealing that 46% of the mouse genome is transcribed into polyadenylated mRNAs (compared with 39% of the human genome). DNase-seq (DNase I hypersensitive sites sequencing), genomic footprinting analysis and ChIP-seq (chromatin immunoprecipitation followed by DNA sequencing) were used to map transcription factor

binding site occupancy, histone modifications and chromatin accessibility. From this information, potential *cis*-regulatory regions were identified, and regulatory function was assigned to ~12.6% of the mouse genome. In addition, genome-wide chromatin organization was determined using microarray-based techniques to generate replication-timing profiles as a proxy for mapping changes to the genomic architecture between different cell types.

Comparative analysis of human and mouse samples revealed both conserved and diverged gene expression patterns, and yielded sets of genes with expression that varies more across tissues than between species, and vice versa. Yue *et al.* posit that categorizing orthologous gene pairs into either species-specific or tissue-specific groups may enable “more informative translation of research results between mouse and human”.

Analysis of the predicted *cis*-regulatory elements in the mouse genome showed that 30–50% of candidate mouse regulatory regions do not have a human orthologue (that is, a homologous region that evolved from a common ancestral sequence). Specifically, the investigators found that 15% of predicted mouse promoters and almost 17% of candidate enhancers, which were both determined on the basis of histone modification patterns, did not have an orthologous sequence in humans.

The evolutionary principles of the candidate regulatory regions

were the main focus of the paper by Stergachis *et al.*, who report a hierarchy of conservation from poorly conserved *cis*-acting sequence elements to the preservation of *trans*-acting and network-level regulatory features. Only ~5% of individual nucleotides were conserved in *cis*-regulatory elements between mice and humans, and transcription factor binding sites showed merely 20% conservation genome-wide. Nevertheless, ~44% of individual regulatory connections between transcription factors were conserved across different mouse and human cell types and, strikingly, despite the high divergence in *cis*-regulatory sequences, the overall transcription factor regulatory network architecture was found to be nearly identical at 95%. In addition, transcription factor binding at orthologous sequences was more highly conserved at promoters than at distal regulatory elements. A third study by Cheng *et al.* further revealed that conservation of transcription factor occupancy is associated with chromatin accessibility and enhancer activity in multiple tissues, which indicates that transcription factor binding patterns show greater conservation for enhancers with pleiotropic activity.

The Mouse ENCODE data sets are likely to be a much-used resource that facilitates not only research into genome regulation but also the development of new and improved mouse models. The studies described here and their companion papers represent another crucial step towards unravelling mammalian biology and human disease pathogenesis.

Linda Koch

ORIGINAL RESEARCH PAPERS Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014) | Stergachis, A. B. *et al.* Conservation of *trans*-acting circuitry during mammalian regulatory evolution. *Nature* **515**, 365–370 (2014) | Cheng, Y. *et al.* Principles of regulatory information conservation between mouse and human. *Nature* **515**, 371–375 (2014)

WEB SITE

Mouse ENCODE: <http://mouseencode.org>

“transcription factor regulatory network architecture was found to be nearly identical at 95%”