

COMPARATIVE GENOMICS

Mammalian alignments reveal human functional elements

“These findings emphasize the large number of genomic regions that have overlapping functions in both gene regulation and coding for proteins”

A new project has carried out whole-genome sequencing to enable comparative genomics across 29 mammalian species. The result is the mapping of human functional elements in unprecedented detail.

Comparative genomics assumes that functionally important elements are constrained during evolution — that is, they are under selective pressure to conserve their sequences. Lindblad-Toh *et al.* sequenced the genomes of 20 diverse mammalian species, adding to nine previously sequenced mammalian genomes. To detect constrained regions, the human genome sequence was compared with the other 28 genomes; this provided a greater ability to identify constrained regions compared with a previous comparison of four mammalian genomes. The authors identified >3.5 million constrained human elements (compared to a few hundred thousand from the previous analysis), which constitute ~4.2% of the human genome.

Most of the newly identified constrained elements lie in intronic and intergenic regions, emphasizing that selection acts extensively beyond protein-coding regions. The authors identified potential regulatory sequence motifs based

on constrained sequences that were repeated throughout the genome. Although this approach did not identify novel types of regulatory motifs, individual elements could be more confidently detected and at a finer resolution than before.

For example, some known, longer regions of constraint could be resolved into shorter, discrete elements that correspond to individual transcription factor binding sites. The accurate detection of such sites was confirmed using existing immunoprecipitation data sets.

The authors assigned ~40,000 constrained elements as having a putative role in RNA secondary structure based on their characteristic substitution patterns during evolution (such as compensatory double mutations that maintain folding by intramolecular hybridization). The RNA secondary structures included novel hairpin families in 3'UTRs as putative regulators of mRNA stability, as well as many elements outside known protein-coding transcripts.

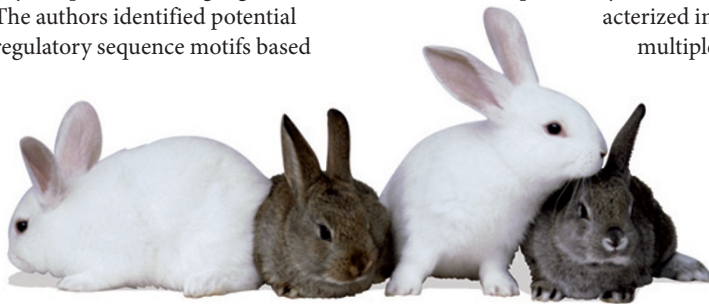
Almost one million additional constrained elements correlated with the positions of distinct chromatin states that had previously been characterized in multiple

human cell types. These elements include candidate enhancer and promoter regions that might play a part in orchestrating the local chromatin state to regulate gene expression.

In a companion paper, Lin *et al.* described in more detail an analysis of overlapping functional elements in protein-coding DNA. In addition to finding the expected avoidance of amino-acid-changing substitutions, the authors showed that more than one-quarter of genes also contained regions that avoid synonymous substitutions, revealing 10,000 synonymous constraint elements (SCEs). By examining the positions and sequence contexts of SCEs, functions were suggested for some of these sites, including roles in RNA splicing, microRNA binding and nucleosome positioning. These findings emphasize the large number of genomic regions that have overlapping functions in both gene regulation and coding for proteins, even though synonymous variants in coding regions are often assumed to be non-consequential.

These studies provide the most detailed map so far of constrained elements in the human genome. However, it is intriguing that ~40% of the elements that have been identified are yet to be annotated with putative functions. Improved functional annotation methods will be needed to uncover the roles of these uncharacterized elements.

Darren J. Burgess



ORIGINAL RESEARCH PAPERS Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 12 Oct 2011 (doi:10.1038/nature10530) | Lin, M. F. *et al.* Locating protein-coding sequences under selection for additional, overlapping functions in 29 mammalian genomes. *Genome Res.* 12 Oct 2011 (doi:10.1101/gr.108753.110)