IN BRIEF

GENOME EVOLUTION

Evidence for widespread convergent evolution around human microsatellites.

Vowles, E. & Amos, W. *PLoS Biol.* **2**, e199 (2004)

DNA-base substitutions do not occur randomly, but what forces govern the biases that we see? These authors show that, despite mutating more quickly, sequences flanking human $(AC)_n$ microsatellites are more similar to each other than expected, and that this similarity is highest around microsatellites of the same length. As $(AC)_n$ microsatellites are common, over 30% of the genome might be affected by this phenomenon, with possible implications for phylogenetic analyses and studies of mutation patterns.

GENE EXPRESSION

Tissue-specific codon usage and the expression of human genes.

Plotkin, J. B., Robins, H. & Levine, A. J. *Proc. Natl Acad. Sci. USA* **101**, 12588–12591 (2004)

Variation in synonymous codon choice is known to affect translational efficiency in several non-mammalian organisms, whereas such biases were thought to be neutral in humans. By studying the degree to which genes differ in their encoding of amino acids, Plotkin and colleagues show that six human tissues can be distinguished solely on the basis of their codon biases and that brain-specific biases are conserved between humans and mice.

MOUSE GENETICS

A focused and efficient genetic screening strategy in the mouse: identification of mutations that disrupt cortical development.

Zarbalis, K. *et al. PLoS Biol.* **2**, 1177–1187 (2004)

The use of reporter genes expressed in cellular subsets combined with mutagenesis provides a powerful tool for identifying genes involved in specific biological processes. However, this approach is not widely used in mammalian model organisms. These authors carried out chemical mutagenesis in mice carrying a β-galactosidase gene expressed in specific sets of cells in the developing forebrain and identified 13 mutations that disrupt cortical development, demonstrating the usefulness of this approach in mouse genetics.

HUMAN GENETIC DIVERSITY

Evidence for gradients of human genetic diversity within and among continents.

Serre, D. & Pääbo, S. *Genome Res.* **14**, 1679–1685 (2004)

Global patterns of human genetic diversity are a matter of debate, with some studies showing that certain patterns cluster according to continent of origin, whereas others indicate that variation changes gradually with geographical distance. Serre and Pääbo have investigated the effect of study design on such studies and found that when sampling is carried out according to geography — rather than on the basis of different populations — the results support gradual variation according to geographical distance.

GENE REGULATION

A code of transcriptional behaviour

Comparative studies have been very effective at identifying conserved *cis* sequences that might have regulatory functions; the snag, however, is that only some of those sequences will actually be bound by a regulator. Christopher Harbison, Benjamin Gordon and colleagues have now brought some much-needed clarity to this area of eukaryotic transcription: by merging data from various sources — including phylogenetic information and protein–DNA binding data — they have generated a detailed map of how yeast transcription factors interact to transcribe the genome. It's all in there: which predicted promoter elements are functional, which regulators associate with them and how, and in what way the binding associations depend on the environment.

With several genome-wide regulatory studies to its name, the yeast *Saccharomyces cerevisiae* is an excellent starting point for defining eukaryotic transcription. Starting with 203 DNA-binding regulatory proteins — probably all such proteins in the genome — the authors' first task was to find which sequences they bind to. This was done by examining the genome-wide location of DNA-bound proteins at a high level of stringency. Next, they computationally defined specific motifs that were bound at high levels of confidence by 102 of these yeast regulators by combining the regulator–DNA binding data with relevant published information and sequence comparisons among *Saccharomyces* species, and by validating previously identified regulator–DNA relationships. The information that emerges from the resulting map, which consists of 3,353 interactions and 1,296 promoter regions, is doubly useful as it incorporates genome-wide binding interactions that were carried out in different environments, such as varying cell-growth conditions.

The stringent approach with which the map was devised makes it a unique resource, but just as useful is the information that the authors were able to extract from it. For example, they found that regulator binding sites are not distributed at random, but are mostly clustered between 100 and 500 base pairs upstream of the coding region. They also defined four types of promoter based on how binding sites were organized, which in turn hints at how promoter architecture influences downstream gene transcription — for example, through combinatorial protein interactions. The impact of including the effect on the environment was felt most obviously here, as promoters could be classified according to how growth-factor status or concentration, say, affected the number and type of promoter elements that were occupied.

With this framework in place, we can begin to model the mechanisms that underlie global gene transcription, and eventually to extend the same approach to multicellular eukaryotes.

Tanita Casci

References and links

ORIGINAL RESEARCH PAPER Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).