TECHNOLOGY

On the trail of transcription factors

Using this method, they were able to mark — through Ty5 integration — the DNA sequences that were ‘visited’ by transcription factors.

Recently developed techniques such as chromatin immunoprecipitation coupled to DNA microarray analysis (ChIP–chip) or ChIPseq, in which immunoprecipitated DNA fragments are sequenced using a ‘next-generation sequencing’ technology, allow the identification of any DNA sequence that is bound by a specific protein. But is it possible to do the opposite and identify which proteins bind a certain sequence? A recent study in yeast offers a promising approach to tackle this question.

The retrotransposon Ty5 encodes an integrase that interacts with the silent information regulator 4 (Sir4) chromatin component and catalyses its insertion into the yeast genome, near Sir4-binding sites. The authors took advantage of the fact that fusing DNA-binding proteins to Sir4 directs Ty5 integration close to the binding sites of these proteins. Using this method, they were able to mark — through Ty5 integration — the DNA sequences that were ‘visited’ by

transcription factors. For each transcription factor fused to Sir4, they used a Ty5 element that was engineered to contain a short sequence that could function as a ‘barcode’ (or ‘calling card’, as the authors propose) specific for that factor.

Following cell transformation (with a plasmid carrying the fusion gene and another carrying the corresponding Ty transposon) and induction of transposition, the DNA of the cells selected for the transposition events was harvested, amplified and analysed by DNA sequencing or DNA microarray hybridization. The authors tested the method first by using two transcription factors, Gal4 and Gcn4, independently, then together with five other transcription factors, each fused to Sir4 along with its corresponding Ty5 element. Analysis of three specific promoters revealed that each had only the barcode corresponding to the transcription factor that was known to bind it, indicating that the method is

robust; moreover, it identified both known and unknown genomic targets.

Implementing this technology for all of the ~200 yeast DNA-binding proteins could allow the whole yeast transcription factor network to be dissected in a single experiment.

Francesca Pentimalli

ORIGINAL RESEARCH PAPER Wang, H., Johnston, M. & Mitra, R. D. Calling cards for DNA-binding proteins. *Genome Res.* **17**, 1202–1209 (2007)

HUMAN GENETICS

Genome-wide association for HIV

Following HIV1 infection, the levels of virus circulating in the plasma (viral load) and the time until the onset of AIDS vary considerably between individuals. A recent genome-wide association study (GWAS) has uncovered two polymorphisms that account for 14.1% of the variation in viral load, and one polymorphism that is implicated in time until the onset of disease.

The authors carefully selected 486 patients with suitable steady-state viral load for their 500k SNP study, and found three polymorphisms that associated with the phenotypes at genome-wide significance levels. The first polymorphism, explaining 9.6% of the variance in viral load, was in the major histocompatibility gene HLA complex P5 (HCP5). The variant is in strong linkage disequilibrium with an allele of HLA-B that has previously been shown to have a strong effect on disease progression. The authors note, however, that the genetic-association evidence

itself cannot distinguish between HCP5 and HLA-B as possible causal agents, and that HCP5 itself is a good candidate for also contributing to viral control. This is because HCP5 is an endogenous retroviral element with homology to the HIV pol gene that is expressed in lymphocytes, and because patients with this haplotype continue to have reduced viral load after the virus has mutated to escape being targeted by lymphocytes.

The second polymorphism is in the HLA-C gene and explains 6.5% of the variation in viral load. This variant is also significantly associated with increased HLA-C expression levels, both in the authors’ data and in previously published results, leading the authors to suggest that it is the changes in expression that cause the protective effect.

The test for association with time of disease onset provided one significant result: a cluster

of seven SNPs that include regulatory regions for the RNA polymerase subunit zinc ribbon domain-containing protein 1 (ZNRD1) and the coding region of the ring finger protein RNF39. As the identified SNPs also associate with ZNRD1 expression, the authors suggest that this is the most likely cause of the protective phenotype, possibly through an interaction with HIV transcription.

These results confirm the importance of the MHC region in HIV restriction, and they also suggest new targets for both therapy and vaccine development. The mechanisms by which HCP5 and ZNRD1 interact with HIV could be exploited therapeutically. Conversely, the role of HLA-C in the interaction between T cells and the virus could be exploited for vaccines.

Patrick Goymer

ORIGINAL RESEARCH PAPER Fellay, J. et al. A whole-genome association study of major determinants for host control of HIV-1. *Science* 19 July 2007 (doi:10.1126/science.1143767)

FURTHER READING Wang, W. Y. S. et al. Genome-wide association studies: theoretical and practical concerns. *Nature Rev. Genet.* **6**, 109–118 (2005)

WEB SITE <http://www.genome.duke.edu/centers/pg2>