

Wanted: regulatory SNPs

Thomas J. Hudson

McGill University and Genome Quebec Innovation Centre, Montreal, Quebec H3A 1A4, Canada. e-mail: tom.hudson@mcgill.ca

Reliable identification of genetic variants that affect gene regulation is still a challenge in genomics and is expected to have an important role in the molecular characterization of complex traits. A new study describes a robust functional assay that detects the presence of *cis*-acting polymorphisms that regulate gene transcription.

A key problem in human genetics is how to identify polymorphisms that confer differences in the expression of genes. In contrast to coding polymorphisms, which are relatively easy to catalog comprehensively by resequencing well defined exonic sequences across individuals¹, regulatory polymorphisms are difficult to pinpoint among the sea of polymorphisms localized in the vicinity of genes². Many diverse strategies will probably be needed to create databases of putative regulatory polymorphisms.

Current high-throughput strategies include large-scale genetic studies that compare gene expression patterns in individuals^{3,4} and *in silico* annotation of non-coding polymorphisms that are located in regions of DNA that show conservation among species or have a high probability of containing regulatory elements⁵. These approaches provide a relatively coarse and insensitive means to identify noncoding single-nucleotide polymorphisms (SNPs) that may alter gene expression.

In contrast, popular functional studies, such as transient or stable transfection⁶, have never been validated for high-throughput screening and cannot be used to assess the effects of potential regulatory SNPs (rSNPs) in their normal chromatin context. On page 469 of this issue, Julian Knight and colleagues⁷ describe a new approach for the functional association of DNA variation with expression by examining the binding of the transcriptional machinery of a gene and correlating it with DNA variation at the locus.

The target for this proof-of-concept experiment is phosphorylated RNA polymerase II (Pol II). Phosphorylation of the Pol II C-terminal domain releases the enzyme from the preinitiation complex and is a key step in initiating transcript synthesis at active promoters. The allelic differences in transcriptional activity of a specific gene caused by *cis*-acting regulatory motifs are theoretically measurable by comparing the amounts of Pol II bound to chromatin for each allelic copy of a gene. In the method of Knight *et al.*⁷, dubbed haploChIP, antibodies against Pol II are

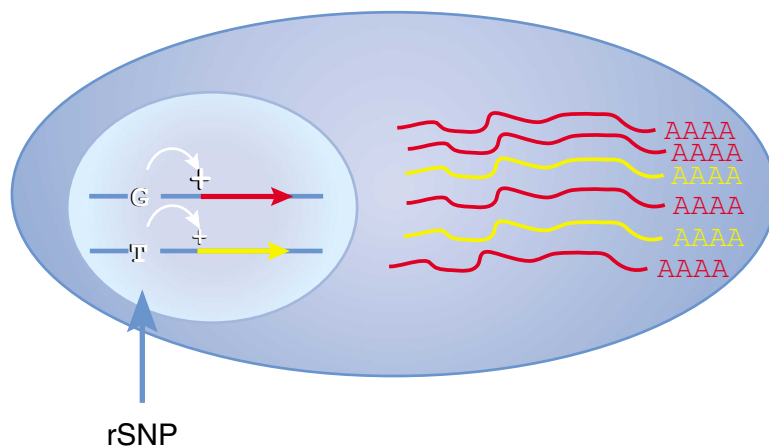
used to isolate chromatin crosslinked with Pol II. Gene-based polymorphisms can be identified using the co-immunoprecipitated DNA, and heterozygous SNPs can be assessed to see whether they maintain the expected 1:1 ratio for each allele. The theory not only works but also provides a highly sensitive means to accurately detect small deviations in allele ratios. In addition, the assay detects different alleles using dideoxy-terminated primer extension reactions followed by mass spectrometry—a strategy that could easily be adapted to create a multiplexed high-throughput assay.

A controversial rSNP

The application of the method to the investigation of a putative regulatory SNP at the gene encoding tumor necrosis factor (TNF) gave surprising results: the controversial *TNF*^{-308A} promoter polymorphism is not correlated with transcription of *TNF* but with lymphotoxin- α (encoded by *LTA*), the upstream neighbor of *TNF*. The authors show that Pol II loading is similar for both the A and the G alleles of *TNF*⁻³⁰⁸, casting doubts on the functional effect of this polymorphism. In contrast, Pol II

loading is different for specific *LTA* haplotypes that were also shown to differ in *LTA* transcriptional activity by allele-specific RT-PCR analysis: these haplotypes incorporate the *TNF*^{-308A} allele. There is no doubt that this whodunit will intrigue aspiring Hercule Poirots and that a number of labs will want to reconsider *TNF* as the susceptibility gene for asthma, malaria and other diseases. *LTA* recently became prominent as a susceptibility gene for myocardial infarction in a Japanese population⁸ and was independently shown to be transcriptionally regulated by a SNP in intron 1.

The evidence added to the *TNF/LTA* story by haploChIP leads one to wonder how many other similar occurrences there are. The high level of correlation between close polymorphisms⁹ makes it probable that a number of existing genetic associations were obtained with indirect markers in linkage disequilibrium with the functional risk variant. For example, a large common haplotype of the cytokine gene cluster that confers risk to Crohn disease¹⁰ contains many genes and many highly correlated SNPs, but the causative variant is difficult to pinpoint



Regulatory single-nucleotide polymorphism (rSNP). Two allelic variants of the same gene (shown as a red and a yellow arrow) are transcribed in different amounts as a consequence of an adjacent polymorphism. In this example, allele G, located upstream of the gene, has a higher transcript level than does allele T.

because it seems to be genetically indistinguishable from other variants embedded in this haplotype block. The advent of whole-genome association studies using the haplotype map¹¹ will probably implicate a substantial number of new disease associations with indirect markers that are in linkage disequilibrium with rSNPs.

Associations in search of a cause

Providing that Pol II (or another DNA-binding protein with similar properties) is equally useful for testing any human gene, haploChIP has the potential to allow the testing of each gene contained in a risk haplotype block for differential transcriptional activity. HaploChIP requires that suitable cells or tissues that express the gene are available and that the regulatory variant affecting the test gene is heterozygous. Another surrogate approach to *in vivo* transcriptional activity is allele-specific RT-PCR^{12,13}. This is probably easier to carry out but has an extra requirement that the RT-PCR product contain polymorphisms, which is not possible for genes (such as *TNF*) that lack exonic SNPs. By

comparing both gene transcripts in the same sample, both methods decrease the effects of environmental and other confounding factors. They also have advantages over *in vitro* techniques, such as transient transfection assays with allele-specific promoter constructs, as these studies are executed outside of their normal chromosome environment.

The haploChIP assay is useful as a surrogate marker of allelic variation in gene transcription, although it does not directly identify the *cis*-acting polymorphism or mechanism that is responsible for this variation. This is not surprising, as there are many SNPs on the haplotype associated with higher *LTA* transcriptional activity. The task is also difficult because of the extensive distances at which transcriptional control elements can be found¹⁴. Should human geneticists care about finding the true regulatory SNP when the risk gene is identified and there already are many highly correlated SNPs that tag the disease haplotype? Yes! The full knowledge of the risk variant is important, as it helps refine the knowledge of gene transcription and points

to a target for therapy. Thus, the genetic community continues to require more tools to efficiently pinpoint risk variants that affect gene expression. There is no doubt that many more inventive approaches will be developed for this. Ultimately, the manipulation of regulatory mutations that affect expression levels should be easier than repairing or modulating the effects of an abnormal protein. □

1. Cargill, M. *et al. Nat. Genet.* **22**, 231–238 (1999).
2. Sachidanandam, R. *et al. Nature* **409**, 928–933 (2001).
3. Brem, R.B., Yvert, G., Clinton, R. & Kruglyak, L. *Science* **296**, 752–755 (2002).
4. Cheung, V. *et al. Nat. Genet.* **33**, 422–425 (2003).
5. Wasserman, W.W., Palumbo, M., Thompson, W., Fickett, J.W. & Lawrence, C.E. *Nat. Genet.* **26**, 225–228 (2000).
6. Rockman, M.V. & Wray, G. *Mol. Biol. Evol.* **19**, 1991–2004 (2002).
7. Knight, J.C., Keating, B.J., Rockett, K.A. & Kwiatkowski, D.P. *Nat. Genet.* **33**, 469–475 (2003).
8. Ozaki, K. *et al. Nat. Genet.* **32**, 650–654 (2002).
9. Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J. & Lander, E.S. *Nat. Genet.* **29**, 229–232 (2001).
10. Rioux, J.D. *et al. Nat. Genet.* **29**, 223–228 (2001).
11. Gabriel, S.B. *et al. Science* **296**, 2225–2229 (2002).
12. Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B. & Kinzler, K.W. *Science* **297**, 1143 (2002).
13. Cowles, C.R., Hirschhorn, J.N., Altshuler, D. & Lander, E.S. *Nat. Genet.* **32**, 432–437 (2002).
14. Hardison, R.C., Oeltjen, J. & Miller, W. *Genome Res.* **7**, 959–966 (1997).

A clinician's plea

Judith G. Hall

Department of Pediatrics and Medical Genetics, University of British Columbia, Vancouver, British Columbia V6H 3V4, Canada. e-mail: jhall@cw.bc.ca

For the detection of human gene mutations to have practical application, clear clinical descriptions of the affected individuals (as well as those clinically affected in whom mutations are not found) should be part of the publication.

Over the last ten years, there has been an exponential increase in the number of papers describing mutations in genes responsible for human disease. Clinicians are thrilled that this new information may translate into meaningful changes in health care for affected individuals. We carefully read such papers for clinical details to help determine which individuals and which families might benefit from molecular evaluation or testing. Sometimes these details are there, but more often they are not. Thus, clinical geneticists are frustrated by an inability to determine to which condition or conditions the reported gene mutation actually refers.

The project to sequence the human genome, nearing completion this year, has been a remarkable effort. It was powered

in part by the promise that insights into human health and disease would result. The large number of excellent articles describing genes mutated in human diseases in this issue of *Nature Genetics*^{1–8} reflects the fruits (and hard work) that are to be celebrated. At the same time, however, they show how little we know about these genes, how much more work there is to be done and how urgently accurate and clear clinical descriptions are needed.

The clinical approach

A clinician can't help but wonder what was different about the 11 individuals with Ondine's curse (congenital central hypoventilation syndrome) in whom no mutation in *PHOX2B* was found, as described by Jeanne Amiel and colleagues¹

on page 459, or what additional features might have been present in the 31 individuals with anophthalmia who had no mutations in *SOX2*, as outlined by Judy Fantes and colleagues³ on page 461. Did they have other structural or functional abnormalities? What was different about the pregnancies of those individuals? Has unusual behavior developed as they age? Was their mother's health or nutrition unusual before conception or during the pregnancy?

Often we read that the 'found' gene is actually one that was previously unknown. The gene structure may allow speculation about function, and there may be related genes that suggest clues as to pathway or pathogenesis, but the actual pathogenic mechanism must be determined through the hard work of defining time-specific, tis-