



Cis-Acting Regulatory Variation in the Human Genome

Tomi Pastinen and Thomas J. Hudson

Science **306**, 647 (2004);

DOI: 10.1126/science.1101659

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of December 7, 2013):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/306/5696/647.full.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/306/5696/647.full.html#related>

This article **cites 40 articles**, 13 of which can be accessed free:

<http://www.sciencemag.org/content/306/5696/647.full.html#ref-list-1>

This article has been **cited by** 120 article(s) on the ISI Web of Science

This article has been **cited by** 55 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/306/5696/647.full.html#related-urls>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

from throughout the genome would yield a cell-specific nuclear topology, resulting in definable patterns of chromosome organization.

Recent evidence has indicated that the chromosomal organization within a nucleus is maintained upon cell division. FRAP analysis with histone-fluorescent protein fusions has revealed that chromosomes appear to remain in their relative nuclear positions (31). It is therefore possible the nucleus of a particular cell type does have its genome specifically organized for the expression of its relevant transcriptome. The spectral karyotype (SKY, which allows the simultaneous detection of all chromosomes) (32) of an interphase nucleus in Fig. 1 would therefore represent the topology that ensures the overall regulation of that cell type. Evidence supports the idea that chromosomes have specific positions within the nucleus. For example, gene-dense and gene-poor chromosomes have been shown to preferentially localize to the nuclear center and periphery, respectively (33). Additionally, a recent study focusing on a subset of chromosomes has shown that lineage-specific associations of certain chromosomes can occur (34). What remains to be established, however, is the organization of an entire genome at the level of the chromosome. Furthermore, whether a given nuclear topology changes upon cell differentiation has yet to be demonstrated.

With the idea of cell-specific nuclear topology, we have clearly moved beyond the realm in which bacteriophage λ can lead by example. The dynamic regulation of hundreds to thousands of genes requires a level of coordination unnecessary for a simple phage. Still, it is important to remember that the functions of the nucleus, such as transcription, are intertwined with its structure. If the principles of eukaryotic gene regulation find basic parallels with a less complicated example, then there is hope that these principles, aided by the merging of technologies such as SKY and 3D microscopy, may allow us to fully appreciate the dynamic organization of a genome within its nucleus.

Note added in proof: A recent study has provided evidence for the colocalization of coregulated genes on the same chromosome (35).

References and Notes

1. M. Ptashne, *A Genetic Switch* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 3, 2004).
2. S. T. Kosak, M. Groudine, *Genes Dev.* **18**, 1371 (2004).
3. A. K. Leung, A. I. Lamond, *Crit. Rev. Eukaryot. Gene Expr.* **13**, 39 (2003).
4. S. T. Kosak et al., *Science* **296**, 158 (2002).
5. C. Francastel, M. C. Walters, M. Groudine, D. I. Martin, *Cell* **99**, 259 (1999).

6. M. Bulger, M. Groudine, *Genes Dev.* **13**, 2465 (1999).
7. D. Chowdhury, R. Sen, *Immunol. Rev.* **200**, 182 (2004).
8. R. S. Mann, *Bioessays* **19**, 661 (1997).
9. M. Brand et al., *Nature Struct. Mol. Biol.* **11**, 73 (2004).
10. C. Francastel, W. Magis, M. Groudine, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12120 (2001).
11. M. Brand, M. Groudine, unpublished data.
12. H. Caron et al., *Science* **291**, 1289 (2001).
13. M. J. Lercher, A. O. Urrutia, L. D. Hurst, *Nature Genet.* **31**, 180 (2002).
14. R. H. Waterston et al., *Nature* **420**, 520 (2002).
15. B. A. Cohen, R. D. Mitra, J. D. Hughes, G. M. Church, *Nature Genet.* **26**, 183 (2000).
16. P. J. Roy, J. M. Stuart, J. Lund, S. K. Kim, *Nature* **418**, 975 (2002).
17. P. T. Spellman, G. M. Rubin, *J. Biol.* **1**, 5 (2002).
18. A. M. Boutanaev, A. I. Kalmykova, Y. Y. Shevelyov, D. I. Nurminsky, *Nature* **420**, 666 (2002).
19. L. D. Hurst, C. Pal, M. J. Lercher, *Nature Rev. Genet.* **5**, 299 (2004).
20. M. Thompson, R. A. Haeusler, P. D. Good, D. R. Engelke, *Science* **302**, 1399 (2003).
21. D. Schübeler et al., *Genes Dev.* **18**, 1263 (2004).
22. F. Greil et al., *Genes Dev.* **17**, 2825 (2003).
23. T. Misteli, *Science* **291**, 843 (2001).
24. T. Cheutin et al., *Science* **299**, 721 (2003).
25. A. Belmont, *Curr. Opin. Cell Biol.* **15**, 304 (2003).
26. E. V. Volpi et al., *J. Cell Sci.* **113**, 1565 (2000).
27. J. Wang et al., *J. Cell Biol.* **164**, 515 (2004).
28. S. Chambeyron, W. A. Bickmore, *Genes Dev.* **18**, 1119 (2004).
29. T. Ragozy, A. Telling, T. Sawado, M. Groudine, S. T. Kosak, *Chromosome Res.* **11**, 513 (2003).
30. G. Felsenfeld, M. Groudine, *Nature* **421**, 448 (2004).
31. D. Gerlich et al., *Cell* **112**, 751 (2003).
32. E. Schröck et al., *Science* **273**, 494 (1996).
33. S. Boyle et al., *Hum. Mol. Genet.* **10**, 211 (2001).
34. L. A. Parada, P. G. McQueen, T. Misteli, *Genome Biol.* **5**, R44 (2004).
35. C. S. Osborne et al., *Nature Genet.* **36**, 1065 (2004).

REVIEW

Cis-Acting Regulatory Variation in the Human Genome

Tomi Pastinen and Thomas J. Hudson*

The systematic screening of the human genome for genetic variants that affect gene regulation should advance our fundamental understanding of phenotypic diversity and lead to the identification of alleles that modify disease risk. There are several challenges in localizing regulatory polymorphisms, including the wide spectrum of cis-acting regulatory mechanisms, the inconsistent effects of regulatory variants in different tissues, and the difficulty in isolating the causal variants that are in linkage disequilibrium with many other variants. We discuss the current state of knowledge and technologies used for mapping and characterizing genetic variation controlling human gene expression.

Expression profiling and genome-wide mapping studies have shown that strong heritable factors govern differences in gene expression levels within mammalian species such as the mouse and human (1, 2). The concentration of a given mRNA allele is controlled both by

cis-acting factors (such as DNA polymorphisms and methylation) in the flanking DNA sequence of the gene and trans-acting modulators (such as transcription factors) that are themselves regulated by other genetic and environmental characteristics of the cell. Whereas heritable expression differences resulting from trans-acting mechanisms appear to be quantitatively more important, cis-acting variation may explain up to 25 to 35% of interindividual differences in gene expression. This is likely an

underestimate, as physiological feedback mechanisms can mask the impact of subtle cis-acting variants on expression levels. Evidence for the medical importance of cis-acting polymorphisms has been provided by recent positional cloning of susceptibility genes that are not associated with protein coding or splice-site polymorphisms for common diseases such as stroke and type 2 diabetes (3, 4).

Regulatory polymorphisms are DNA elements that modify the expression level of a transcript or its isoforms. Current techniques can detect expression differences as low as 1.2-fold between samples or alleles; the phenotypic consequences (if any) of such small differences are likely to depend on the function of the gene in question. Most known regulatory polymorphisms are located in gene promoter regions and function by altering gene transcription. Coding polymorphisms are also known to affect the

McGill University and Genome Quebec Innovation Centre, 740 Drive Penfield Avenue, Montreal, Quebec H3A 1A4, Canada.

*To whom correspondence should be addressed. E-mail: tom.hudson@mcgill.ca

expression of alleles, as in the case of the nonsense-mediated mRNA decay of transcripts harboring early stop codons (5). The emerging picture of regulatory sequences distributed over long distances upstream and downstream of a gene, including introns as well as the 5' and 3' untranslated regions, has led to the discovery of regulatory variants located outside promoter regions that can alter mRNA stability (6), mRNA processing efficiency (7), or mRNA isoform expression (4, 8) or induce epigenetic changes (9). However, association of allelic expression with heritable regulatory polymorphisms or epigenetic mechanisms may not be straightforward. For example, different mechanisms have been suggested to underlie allelic expression of the human *TP73* gene in noncancerous cells and tissues ranging from heritable polymorphisms (10) to tissue-specific mono-allelic expression (11).

Perils and Pitfalls in the Identification of Cis-Acting Variation

Isolating a true regulatory variant is complicated by linkage disequilibrium (LD) in the human genome. This is highlighted by studies of the lactase (*LCT*) persistence phenotype, a common monogenic trait caused by cis-acting regulatory variants. Genetic studies in the Finnish population have shown a perfect correlation between *LCT* persistence and the T allele in a single-nucleotide polymorphism (SNP) approximately 14 kb upstream of the *LCT* gene (12). Whereas these findings were supported by *in vitro* studies showing functional differences between the alleles (13), subsequent studies have identified individuals who are heterozygous for the persistence allele but show equal expression of *LCT* alleles (14). In addition, the -14-kb SNP is not associated with *LCT* persistence in some non-Caucasian populations (15); this SNP shows very high LD with other variants contained in a 1-Mb flanking region (16), suggesting that it may be in LD with the causative regulatory variant.

The modulation of gene expression caused by epigenetic mechanisms can be misconstrued to be due to regulatory polymorphisms. Classically imprinted autosomal loci display preferential expression of a single allele (monoallelic expression) that is independent of sequence variation. This

unequal expression of transcripts, which is often transmitted according to the allele's parent of origin, is usually accompanied by different patterns of cytosine methylation or posttranslational histone modifications. This phenomenon has been best studied in mice, in which nearly 50 imprinted genes localized to 15 genomic regions have been characterized (17). For some genes, only one allele of a gene is arbitrarily expressed in each cell: This process is called random monoallelic expression and is reminiscent of X-chromosome inactivation in females (18). Other naturally occurring epigenetic mechanisms, which do not follow strictly parent-of-origin or random patterns, have also been described in mammals (19). Interindividual variability in levels of allele silencing in imprinted genes can be observed in normal controls (20), and it has been suggested that diet may influence DNA

date regulatory polymorphism affects gene expression (23). Most current studies target putative promoter or upstream flanking regions; these regions are often poorly characterized and frequently do not represent the complete promoter that is active in the cell line of interest. Although experimentally validated promoters can be found in the Eukaryotic Promoter Database (www.epd.isb-sib.ch), these comprise less than 10% of human genes. The initial choice of allele-specific constructs for transfection studies can be refined by deletion experiments to delineate the most important 5' regulatory sequences. Alternatively, information from long-range regulatory sequences can be studied to use constructs containing proximal promoter regions and enhancer elements (24). More faithful reproduction of natural gene regulation can be achieved by cloning whole human

genes in bacterial artificial chromosomes (25). Such studies are technically challenging and relatively uncommon; most published studies have used relatively small promoter constructs (<1 kb) or oligonucleotide fragments studied in the context of heterotypic minimal promoters.

Transient transfection studies may be complicated by trans-acting influences on allelic expression. Studies are often performed in preexisting animal or human lines, but there are concerns about whether the observed data can be extrapolated to the human tissues of interest. Even small trans-acting differences result-

ing from other genetic variants in the host may be important, as shown by the quantitative variation of allele-specific responses in fibroblasts obtained from unrelated individuals (26).

Transient transfection assays were recently applied in a systematic, stringent survey to study proximal promoter (-0.5 kb) haplotypes from 38 human genes. Significant allele-specific expression could be reproduced in 13 out of 17 (75%) cases when independent clones were used, suggesting that >30% of proximal promoters may harbor cis-acting variants (27).

In vivo approaches. *In vivo* monitoring of allelic RNA transcripts (28) is possible in tissues or cells of individuals heterozygous for an informative marker within the locus. There are several advantages to observing relative expression of the two alleles within

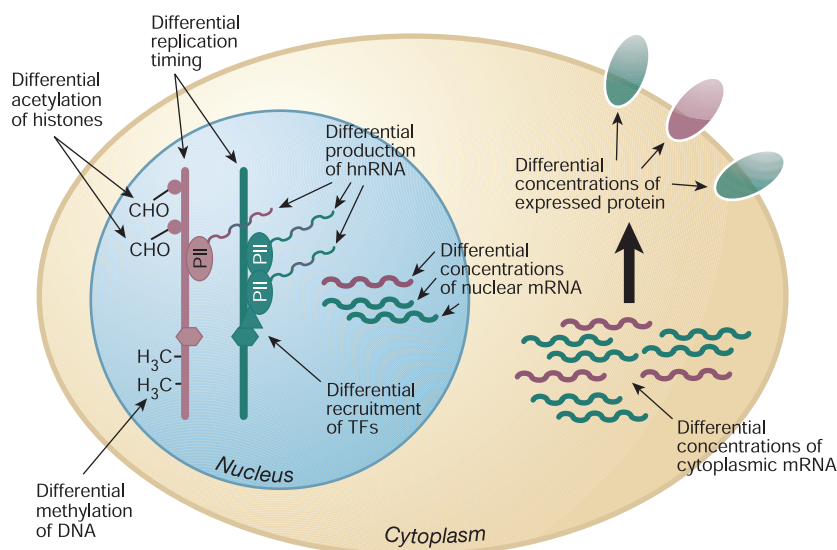


Fig. 1. Cellular phenomena associated with and measured in allele-specific expression studies. PII, RNA polymerase II.

methylation and allelic expression of epigenetically regulated loci (21). Epigenetic alterations are common in neoplastic cells, which may even be detected in peripheral blood samples as demonstrated in colon cancer patients showing increased biallelic expression of insulin-like growth factor 2 (*IGF2*) in comparison to that of healthy controls (22).

Detecting Allele-Specific Expression

Allele-specific expression of a transcript can be detected by *in vitro* and *in vivo* methods that measure the cumulative effects on a number of cellular processes (Fig. 1).

In vitro approaches. *In vitro* methods (most commonly involving transient transfection assays) monitor the transcriptional activity of a synthetic reporter construct and are best suited for testing whether a candi-

the same tissue sample: (i) Alleles are expressed in their normal environment including genomic and chromatin context; (ii) comparison of alleles is made within rather than between samples, maximizing the sensitivity of detecting cis-acting effects; (iii) the developmental and physiologic history of the tissue is unlikely to be perturbed by the presence of two low- or two high-expressing alleles; and (iv) population-based studies allow sampling of haplotype diversity within each locus.

The approach has been applied in the context of rare monogenic diseases to demonstrate underexpression of the disease allele (29). Similarly, underexpression of the wild-type allele may explain variable penetrance in dominantly inherited Mendelian traits (30). Evidence for cis-acting regulatory polymorphisms in candidate genes for complex diseases has also been sought by allelic expression measurements (8). Demonstration of parent-of-origin specific expression in tissues (or cells) (31) and monoallelic expression in cells (32) provide commonly used assays to establish imprinting and random monoallelic expression, respectively.

Direct methods of visualizing allelic expression are challenging, thus measurements are commonly performed with amplified cDNA [reverse transcription polymerase chain reaction (RT-PCR)] from tissues or cell lines of interest and require the presence of an informative polymorphism in the processed transcript. Increased informativity can be achieved with the use of nuclear premRNA [heteronuclear RNA (hnRNA)] (33). In our experience, assays performed with hnRNA have slightly higher variability and failure rates as compared with those that use mRNA, likely reflecting the lower concentration of hnRNA in total RNA preparations. Successful hnRNA assays not only increase the informativity of the allelic expression measurements but also provide evidence for transcriptional causes of allelic expression, because RNA processing differences (such as alternate splicing) can be excluded. The role of transcription in causing allele-specific expression can also be determined by the recently introduced polymerase loading assay [haplotype-specific chromatin immunoprecipitation ("Haplo-ChIP")] (34), which is based on isolating transcriptionally active DNA fragments by immunoprecipitation DNA bound to RNA polymerase II enzyme. The isolated DNA fragments can be assessed for polymorphisms in heterozygous samples to determine relative transcriptional activity of the alleles as a surrogate for relative allelic expression.

Allelic expression studies also require quantitative genotyping assays and most commonly primer extension methods have been employed for relative allelic quantita-

tion. Imbalanced allelic expression is detected when the heterozygote allele ratio in RNA (cDNA) differs from the corresponding 1:1 ratio in genomic DNA. The cut-off for calling allelic imbalance would optimally be based on the variability of cDNA heterozygote ratios in samples known to express the alleles in equal proportions; in practice, such information is not available and thresholds are commonly derived from the analysis of variability in heterozygote genomic DNA samples. The potential biases introduced by the application of genomic standards are far less serious than the artefactual allelic imbalance caused by stochastic RT-PCR amplification of one allele in low copy number targets. Stochastic effects are a particular concern in single-cell studies (32).

Recent *in vivo* screening studies of relative allelic expression in normal tissues or cell lines for hundreds of human genes suggest that 25 to 50% of genes and 5 to 25% of heterozygotes show evidence of unequally expressed alleles (10, 33, 35). The abundance of allele-specific differences in expression is notable, though variable study designs preclude consensus estimates of its prevalence in the human genome. Furthermore, the allelic expression demonstrated in informative heterozygotes has not been correlated with total expression levels across all genotypic groups; some allelic differences could be compensated if the expression of the gene were under direct negative feedback control.

Genetic Mapping Combined with Expression Profiling

Genome-wide expression profiling technology has developed to a level at which even routine clinical applications have been contemplated. This, along with improved genotyping technologies, would allow large-scale correlations of marker genotypes to gene expression levels modeled as quantitative traits (eQTLs). When genetic linkage of the eQTL coincides with the genomic location of the gene, the presence of cis-acting regulatory variants can be deduced (1, 2). Alternatively, panels of moderate sample size may have sufficient power to permit whole-genome association studies with eQTLs. The drawbacks of both approaches are that the detection of subtle cis-acting effects may require large sample sizes (i.e., thousands of RNA samples from different individuals) and that epigenetic variation is not assessed. Furthermore, without direct allele-specific expression measurements, the correlation of a marker genotype with gene expression level does not guarantee that the effect is cis-acting; a polymorphism in a trans-acting regulator in linkage disequilibrium with the marker genotype can explain the association. This may prove to be important even

with high-density mapping, because antisense transcription may be a common regulatory mechanism of human gene expression (36).

Elucidating the Causal Mechanism of an Allelic Imbalance

Heritable cis-acting effects can be demonstrated by cosegregation of unequal allelic expression with marker genotypes in pedigrees (10, 29). The lack of Mendelian inheritance of an allelic imbalance phenotype may point toward epigenetic mechanisms (33). Measurable phenomena associated with epigenetic allele-specific expression include replication asynchrony (32), differential methylation of the genomic loci (37), and allele-specific posttranslational histone modification (37). Common heritable allelic imbalance phenotypes can be mapped in unrelated individuals to establish regulatory haplotypes (34). For example, we demonstrated a strong regulatory haplotype in the human *BTN3A2* locus, which spanned at least 15 kb flanking the gene (33).

A tempting approach is to use existing bioinformatic tools to identify functional regulatory variants, but despite advances in the field (38), these computer predictions have relatively poor specificity. *In vitro* methods may also help find the functional SNP(s); however, their role is restricted by the inability of plasmid constructs to mimic the role of the natural genomic context in establishing allele-specific expression. Most of the transient transfection studies to date have been corroborated with other *in vitro* assays. For example, allele-specific DNA-protein interaction assays have been used to demonstrate that the putative regulatory polymorphism shows allele-specific differences in recruiting nuclear transcriptional activators or repressors. Similarly, indirect support for the functionality of the putative regulatory polymorphism may be sought by correlating genotypes with altered protein activity. Finally, the direct observation of cis-acting effects *in vivo*, as demonstrated for the human *LTA* gene with the use of the HaploChIP technology, provides corroboration of transcriptional events mediating allele-specific gene expression (34, 39).

Allelic Expression: Next Steps

The rapidly evolving data sets, technologies, and knowledge of regulatory variation portend the generation of genome-wide mapping and characterization of allelic variants affecting gene expression. The large-scale *in vivo* screening studies carried out to date are generally limited to descriptions of allele-specific differences in expression, leaving the underlying mechanisms largely unexplored.

A more complete picture will require genome-wide *in vivo* allelic expression analy-

ses followed by a systematic classification into probable genetic or epigenetic mechanisms with the use of family-based samples. The genetic basis of allele-specific expression phenotypes can subsequently be mapped to regulatory haplotypes. The most limiting factor for such a study is the lack of suitable panels of human tissues. In the short term, existing collections of immortalized cell lines can provide useful starting material, although these cells may express the genes of interest at low level or under abnormal transcriptional control.

Alternatively, genome-wide assessment of DNA protein interaction for transcriptionally active genes in vivo combined with allele quantitation of the protein-bound genomic fragments (34, 39) could be used to determine cis-acting polymorphisms underlying interindividual differences in response to key transcriptional regulators. The latter approach may provide a short-cut to the identification of the causative regulatory polymorphism and shift the focus to cellular processes of interest. The intersection of multidisciplinary efforts to develop tools and characterize the regulatory component of the human genome [such as the ENCODE project (40)] with genome-wide allelic expression studies and regulatory haplotype characterization will provide a

wealth of data for understanding cis-acting variation affecting the regulation of human genes and its contribution to phenotypic variance.

References and Notes

1. E. E. Schadt *et al.*, *Nature* **422**, 297 (2003).
2. M. Morley *et al.*, *Nature* **430**, 743 (2004).
3. Y. Horikawa *et al.*, *Nature Genet.* **26**, 163 (2000).
4. S. Gretarsdottir *et al.*, *Nature Genet.* **35**, 131 (2003).
5. M. C. Willing, S. P. Deschenes, R. L. Slayton, E. J. Roberts, *Am. J. Hum. Genet.* **59**, 799 (1996).
6. A. M. Carter, M. Sachchithananthan, S. Stasinopoulos, F. Maurer, R. L. Medcalf, *Thromb. Haemostasis* **87**, 846 (2002).
7. N. H. Gehring *et al.*, *Nature Genet.* **28**, 389 (2001).
8. H. Ueda *et al.*, *Nature* **423**, 506 (2003).
9. C. Ober *et al.*, *Am. J. Hum. Genet.* **72**, 1425 (2003).
10. H. Yan, W. Yuan, V. E. Velculescu, B. Vogelstein, K. W. Kinzler, *Science* **297**, 1143 (2002).
11. J. F. Hu *et al.*, *Biochim. Biophys. Acta* **1491**, 49 (2000).
12. N. S. Enattah *et al.*, *Nature Genet.* **30**, 233 (2002).
13. L. C. Olds, E. Sibley, *Hum. Mol. Genet.* **12**, 2333 (2003).
14. M. Poulter *et al.*, *Ann. Hum. Genet.* **67**, 298 (2003).
15. C. A. Mulcare *et al.*, *Am. J. Hum. Genet.* **74**, 1102 (2004).
16. T. Bersaglieri *et al.*, *Am. J. Hum. Genet.* **74**, 1111 (2004).
17. A. Ferguson-Smith, S. P. Lin, C. E. Tsai, N. Youngson, M. Tevendale, *Semin. Cell Dev. Biol.* **14**, 43 (2003).
18. R. Ohlsson, B. Tycko, C. Sapienza, *Trends Genet.* **14**, 435 (1998).
19. Z. Lin, J. Floros, *Physiol. Genomics* **11**, 235 (2002).
20. T. Sakatani *et al.*, *Biochem. Biophys. Res. Commun.* **283**, 1124 (2001).
21. D. Ingrosso *et al.*, *Lancet* **361**, 1693 (2003).
22. H. Cui *et al.*, *Science* **299**, 1753 (2003).
23. M. V. Rockman, G. A. Wray, *Mol. Biol. Evol.* **19**, 1991 (2002).
24. Y. Xu *et al.*, *J. Lipid Res.* **40**, 50 (1999).
25. R. Wade-Martins, Y. Saeki, E. A. Chiocca, *Mol. Ther.* **7**, 604 (2003).
26. J. L. Rutter *et al.*, *Cancer Res.* **58**, 5321 (1998).
27. B. Hoogendoorn *et al.*, *Hum. Mol. Genet.* **12**, 2249 (2003).
28. J. Singer-Sam, J. M. LeBon, A. Dai, A. D. Riggs, *PCR Methods Appl.* **1**, 160 (1992).
29. D. Hewett, J. Lynch, A. Child, H. Firth, B. Sykes, *Am. J. Hum. Genet.* **55**, 447 (1994).
30. S. Hutchinson *et al.*, *Hum. Mol. Genet.* **12**, 2269 (2003).
31. J. Li *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7341 (2004).
32. A. Chess, I. Simon, H. Cedar, R. Axel, *Cell* **78**, 823 (1994).
33. T. Pastinen *et al.*, *Physiol. Genomics* **16**, 184 (2004).
34. J. C. Knight, B. J. Keating, K. A. Rockett, D. P. Kwiatkowski, *Nature Genet.* **33**, 469 (2003).
35. H. S. Lo *et al.*, *Genome Res.* **13**, 1855 (2003).
36. O. Rosok, M. Sioud, *Nature Biotechnol.* **22**, 104 (2004).
37. R. Jaenisch, A. Bird, *Nature Genet.* **33** (Suppl), 245 (2003).
38. W. W. Wasserman, A. Sandelin, *Nature Rev. Genet.* **5**, 276 (2004).
39. J. C. Knight, B. J. Keating, D. P. Kwiatkowski, *Nature Genet.* **36**, 394 (2004).
40. The ENCODE Project Consortium, *Science* **306**, 647 (2004).
41. We thank E. Harmsen, J. Engert, and R. Sladek for their advice. T.J.H. and T.P. are supported by Genome Quebec and Genome Canada. T.J.H. is the recipient of a Clinician-Scientist Award in Translational Research by the Burroughs Wellcome Fund and an Investigator Award from the Canadian Institutes of Health Research.

Turn
a new
page
to...

www.sciencemag.org/books

Science
Books et al.
HOME PAGE

- ▶ the latest book reviews
- ▶ extensive review archive
- ▶ topical books received lists
- ▶ buy books online