

nature genetics

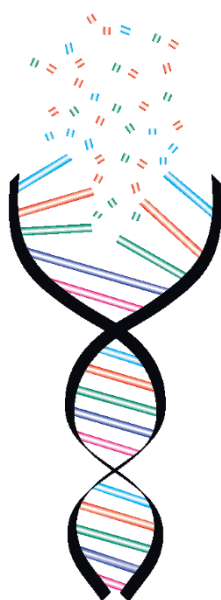
volume 27 no. 3 march 2001

Harvesting the fruits of the human genome

The papers describing the human genome sequence by the International Human Genome Sequencing Consortium¹ (IHGSC) and Celera Genomics² are epic. As such, they have inspired the recollection of other such publications, such as the two-page *Letter*³ by James Watson and Francis Crick describing the structure of DNA in 1953. The events following the description of this hypothesis-generating structure and leading up to last month's publications represent massive strides in the development of new technologies. There is cause for much celebration—not only of the final products and attendant insights, but also of the advances permitted by the continuous deposition of genome sequence by the IHGSC. One such advance is the identification of genes that influence disease susceptibility; more than 30 such genes have been cloned using the draft genome sequence from public databases. It is the combination of high-throughput technology, streamlined coordination and enormous commitment, that has given rise to the insights and tools now before us. And as to the future?

Methods of gene annotation will continue to evolve—see pages 332 (ref. 4) and 337 (ref. 5) of this issue for descriptions of two approaches. Many genes are inferred through their homology with those already known; the identity of others is suggested by mapping cDNAs onto genomic sequence or by microarray-based searches for expressed transcripts (see page 232 (ref. 6) and refs. 7,8). The sequence of the mouse genome should help. The rapid acquisition of mouse sequence will also enable detailed catalogs of changes in gene expression and genomic architecture; establishing an accessible mouse BAC library that comprehensively spans the mouse genome and can be used to generate microarray probes is therefore desirable. Its human counterpart, essential to rapid cytogenetic diagnosis, is well on the way¹⁰, as discussed by Antonarakis¹¹ (page 230).

Obtaining an accurate estimation of global gene expression is a daunting challenge, even with high-throughput *in situ* hybridization and expression analyses. The identification of regulatory sequences through comparative genomics will help. But, without a better understanding of the way in which environment directly and indirectly influences gene expression, which will involve new technologies for tracking proteins (see pages 240 (ref. 11) and 304 (ref. 12) for an example), the margin of error may be significant. An accurate understanding of gene expression will ultimately rely on a combination of data obtained using different platforms and prediction algorithms.



Nature Genetics celebrates the genome
and is delighted to make its related
commentaries freely available
(<http://genetics.nature.com/genome/>).

An impressive by-product of the sequencing endeavor is the identification and mapping of a substantive number of new single-nucleotide polymorphisms (SNPs). Those that are common to different ethnic groups are believed to have arisen during the early stages of human evolution. And so the most easily detected variation—the ‘low-hanging fruit’—is represented by a comparatively small number of very old SNPs. These are also the focus of association studies, because they are usually found in a sufficiently large fraction of patients and controls. About 1.4 million SNPs have been identified and mapped by The SNP Mapping Consortium¹³, mainly through comparison of random reads against draft sequence. On page 234, Kruglyak and Nickerson¹⁴ estimate that the collection represents up to 12% of common human allelic variation, taking into account the bias toward the selection of high-frequency SNPs. They also point out that our ability to find SNPs is poorly matched by our ability to genotype them; the lack of a robust methodology would seem to be a significant limiting factor in making the most of the tools that are currently available.

Issues concerning access are common to studies of the genome and its downstream effects. The BAC clones used for sequence analysis are, in and of themselves, reagents, and the extent to which they and other such reagents are made available will affect the rate of discovery. To ensure that advances in genome technology are put to best use, effective methods of dissemination must be established. As Little recommends (page 229, ref. 15), a concerted, international effort to establish a repository and distribution would be well placed.

Dissemination is also an activity of publishers. The debate^{16,17} about access to the sequence described by Celera Genomics² is one that is central to access to other types of data, such as the sequence identity of microarray probes, SNPs, and other genomes currently being sequenced by companies in the private sector. How to establish appropriate criteria for publication, and indeed, how to publish—so as to ensure that data central to published studies are available to the research community—are challenges that face publishers in an era of change.



1. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
2. Venter, J.C. *et al.* *Science* **10.1126/science.1058040** (2001).
3. Watson, J.D. & Crick, F.H.C. *Nature* **171**, 737–738 (1953).
4. Reboul, J. *et al.* *Nature Genet.* **27**, 332–336 (2001).
5. Gopal, S. *et al.* *Nature Genet.* **27**, 337–340 (2001).
6. Burge, C.B. *Nature Genet.* **27**, 232–234 (2001).
7. Penn, S.G., Rank, D.R., Hanzel, D.K. & Barker, D.L. *Nature Genet.* **26**, 315–318 (2001).
8. Shoemaker, D.D. *Nature* **409**, 922–927 (2001).
9. The BAC Resource Consortium **409**, 953–958 (2001).
10. Antonarakis, S.E. *Nature Genet.* **27**, 230–232 (2001).
11. Martienssen, R. *Nature Genet.* **27**, 240–241 (2001).
12. van Steensel, B. & Henikoff, S. *Nature Genet.* **27**, 304–308 (2001).
13. The International SNP Map Working Group *Nature* **409**, 928–933 (2001).
14. Kruglyak, L. & Nickerson, D.A. *Nature Genet.* **27**, 234–236 (2001).
15. Little, P. *Nature Genet.* **27**, 229–230 (2001).
16. Editorial. *Nature* **409**, 745 (2001).
17. Kennedy, D. & Jasny, B.R. *Science* **291**, 1153 (2001).