

## Human genomes, public and private

**The burgeoning commercial sector that is based on genome information poses a challenge to the norms of scientific publication. But it remains to be established that the conditions of access to published sequence data need to change.**

**T**he human genome sequence contains the genetic code that sits at the core of every one of the ten trillion cells in each human being. It profoundly influences our bodies, our behaviour and our minds; it will help the study of non-genetic influences on human development; it will unlock new insights into our origins and history as a species; and it points to new ways of combating disease. The people of many countries have invested in the Human Genome Project's determination of the sequence, and it is hard to see how that investment could have received better returns. Having released their data daily from the outset with unrestricted access, the publicly funded consortium has assembled about 92% of the sequence. *Nature* is delighted this week to publish the project's analysis, and related results, freely available to all without restriction at <http://www.nature.com> and on pages 814–958.

In so doing, *Nature* has followed a traditional model in the publishing of extensive scientific data. As indicated in our "Guide to Authors", we require the results of genome sequence analyses, as with protein structure coordinates, to be immediately available from an appropriate database without restriction. This supports an unwritten contract with our readers that what they see described is what they can use, without obstacles, whether they work in the commercial or academic sector (an increasingly blurred distinction). It supports a broader principle by which scientific results are available for searching and use with software tools. And it supports a principle enunciated by the United Nations that the human genome in particular is, in a symbolic sense, humanity's common heritage.

It is worth noting that *Nature* sees a distinction between access to essential data embodied in a paper, and access to the materials and techniques used. Making freely available all the materials used for a piece of research is sometimes impractical. All we can do is to sustain a policy of access to materials as far as possible, and to cooperate in attempts to set up new standards of access as technologies develop.

### Money makers

What, then, of companies that make a living from data discovered through their own research? There is big money at stake. The bioinformatics market alone will soon exceed \$1 billion per year. Some pharmaceutical companies are valued at hundreds of billions of dollars, much of which depends on their ability to exploit genomic information in developing blockbuster drugs. According to stock-market commentators, there are more than 100 drug-platform biotech companies chasing some \$3 billion of research and development money. Genomics-based drug companies such as Human Genome Sciences, Millennium and CuraGen are forming partnerships with big drug companies to help them grow. Companies such as Affymetrix, Celera, Gene Logic and Incyte are spending billions on software and hardware. Information-technology giants such as Motorola and IBM are increasingly bringing their expertise to bear in the supply and analysis of microarrays and other biochips.

How do information generators and curating companies make their living? Celera, a company in the vanguard of those seeking to use the human genome sequence to make money for private investors, depends on much more than the sequence to pay its way. Its sources of revenue include subscriptions, software, third-party licence revenue, consulting and diagnostics. Its total revenues are estimated at \$130 million, and its capitalization stands at over \$1 billion. Some analysts are concerned about Celera's future profitability in the face of the availability of free data from public projects. Others are confident that it will be able to add value to the basic information in a way that drug companies will pay substantial sums for. The company is gearing up to become a proteomics powerhouse and, as its president Craig Venter has said, there are too many challenges around for it to make sense for private and public scientists to duplicate one another's work.

### Data access

The private sector is essential to the improvement of human health, and its investors need financial returns. Yet many of the commercial companies, whatever their motives and products, do outstanding basic science along the way. The amount of valuable scientific information residing in research and development centres in the private sector is huge — the global R&D budget of pharmaceutical companies is estimated at \$28 billion per year.

It is in the interests of all researchers that as much of that information as possible finds its way into peer-reviewed publications. Suppose a journal were handed a wonderful and unique piece of science, but that, in the interests of the originators, some of the information had to be kept behind a wall of subscriptions and licensing arrangements. In that case the conflict of interest would be acute: on the one hand, the world is better served if the information is available, even with conditions, than if it is not available at all; on the other hand, such a situation not only breaches the traditional norms of scientific publication, but also hampers the ability of science to progress rapidly by the unrestricted reanalysis of published data.

Since we established our policy on access to genome data in January 1996, *Nature* has been able to hold the traditional line. The burden of providing proof that the line should be abandoned lies with the companies — either through rational debate or possibly by the sheer scientific significance of their output in the absence of a publicly funded equivalent. With a publicly funded project delivering data, *Nature* believes that the human genome sequence is not the place for the traditional rules to be broken. We are willing to cooperate with competing journals to maintain the model of open access as far as possible, as we have in the past. We are also willing to engage in consultation where new types of privately generated data, lacking their equivalent in the public domain or in standard public databases, require new thinking. But, in the meantime, the more that both private and public activities can stimulate each other with freely available science, the better it will be for everyone. ■