nature biotechnology

DNA confidential

As the cost of human genome sequencing plunges and large-scale genome-phenotype studies become possible, society should do more to reward those individuals who choose to disclose their data, despite the risks.

The genome sequence of Patient Zero is disclosed on p. 847 of this issue. The paper is notable not only because it provides the first description of the performance of a single-molecule platform in sequencing a human genome (~90% of it, at least), but also because Stanford professor Stephen Quake (aka Patient Zero) opted to tell the world that it was his DNA that had been sequenced. Like scientific pioneers before him, Quake is heroically self-experimenting—testing the risks in publishing identifiable personal information of the most intimate kind.

The paper thus emphasizes the urgency of the need to circumscribe the use of personal genomic information. It also signals the era of affordable full genomic determination is upon us, given one or two tweaks. So decisions on whether to control the dissemination of information, or how to do so, need to be taken very soon.

On its own, the sequence of letters in a human genome is uninformative. Its power for good arises only from associations with medical histories, behavioral characteristics, physical descriptions and environmental influences. Likewise, its capacity for ill derives only from the genome's potential for providing pointers to human qualities that serve as the basis for discrimination and defamation already prevalent in our societies—sex and sexual orientation, societally defined 'race', age, physical and mental health, aptitude or suitability for athleticism or employment, and eligibility for health, disability and life insurance. It is discrimination and its consequences in iniquitous societies, and an unduly tenacious adherence to beliefs in genetic determinism, that prompt discomfort with personal genetic information being available over the internet or in any widely available medium.

No one doubts that there are risks. But thus far discussion of the risks has dominated the debate over the use of DNA data. It is time for the debate to refocus on the benefits of data availability.

The question that should be asked is not "How can things go wrong and how can we prevent that?" [*Answer: "in a myriad of ways" and "only with a great deal of contortion"*] but "What is a necessary goal and how can we achieve it?"

The benefits from open sharing of genomic data arise, potentially, because proper intelligence and knowledge will counteract the inefficiencies of healthcare systems that still spend most of their money treating patients who are near to death. The promise of genomic medicine is preventative and predictive healthcare; and genomic medicine only works if the research community has databases combining genomic data and phenotypic information. The 'necessary good', therefore, is the rapid accumulation and organization of relevant data.

There is, however, currently a clear mismatch in the risks and the benefits of data disclosure: the potential benefits accrue to scientific research and ultimately healthcare, whereas the risks are borne now by individuals. In attaching his identity to the published sequence, Quake has signaled a belief that, ultimately, the benefits will be worth whatever risk he is taking. Others have made a similar stand, notably J. Craig Venter (*PLoS Biol.* **5**, e254, 2007), Jim Watson (*Nature* **452**, 872–876, 2008) and ten luminaries of the Personal Genome Project (PGP; http://www.personalgenomes.org/), headed by Harvard's George Church. Another 13,000 volunteers are lined up for the PGP while Venter's nonprofit institute and the 1,000 Genomes Project (http://www.1000genomes.org/) each plan to create personal human genomics databases.

However, this is just a start. Hundreds of thousands of genomes, not tens of thousands, will probably be required, along with highly structured health and lifestyle records with detailed dietary indicators (food diaries or shopping lists) to deconvolute the factors involved in human disease and wellness. That level of participation will not be achieved quickly by better education or heightened altruism on the part of individuals donating their genomes. What is really needed is an adjusted risk-benefit ratio for individuals.

Governments have thrown much legislation at reducing the risks of disclosure: the US has the Genetic Information Nondiscrimination Act (GINA): the EU has its Directive 95/46/EC on data protection; the UK amended its 2004 Human Tissue Act to make it illegal to sample someone's DNA without their consent; whereas Germany passed legislation in April specifically prohibiting anonymous paternity tests and outlawing genetic testing for predisposition to illnesses of later life. Such measures may actually reduce the risk but they also increase the perception of risk: people worry more about disclosing their genomes.

Surely it is now time to to balance the discussion of the risks to an individual with discussion of how an individual could benefit. In capitalist and competitive societies, this implies direct appeals to self-interest.

There are 'carrots' that could encourage people to put both their genome sequence and other personal data in the public domain. In the US, for instance, donation of data could serve as an additional qualification for Medicaid eligibility. Insurers might consider reduced premiums for people who offer to put their data in the communal system (or for companies that encourage their employees to do so). The federal government could provide tax breaks to companies or individuals who participate. In nations with social healthcare systems, becoming a data donor could trigger a waiving of the large employee and employer contributions to the state health insurance schemes. Or, perhaps more controversially, data donation would allow patients to rise quickly up the queues for treatment.

There will be some individuals, like Steve Quake, who will provide samples and data without an incentive; but when it comes to exploring the basis of being human and moving toward the goal of genomic medicine, society needs to do more to provide personal incentives to those who choose to disclose their data, despite the risks. After all, everybody will ultimately benefit—both those who share and those who choose not to.