In Support of Open Access for Genomic Research

by

This commentary in the Genomics Law Report's ongoing series What ELSI is New? *is contributed by* George Church, Harvard Medical School .

One of the recurring themes in this ELSI series has been the discussion of open-access vs. research-only models for genomic research (see Bobe, MacArthur, McCarty, Prainsack and Sweeney). Below I discuss the characteristics and advantages of, as well as obstacles to, an open-access data model for genomic research.

A. Self-access: Open-access and freedom of information are increasingly required by law. Medical research is increasingly holistic -- integrating a variety of (identifiable) traits and molecular signatures. Genomics is just part of this, not particularly exceptional. Multi-purpose cohorts and biobanks are displacing single trait studies. Research volunteers are increasingly expecting to see their own data and what is being doing with it. So with respect to such desired transparency, projects can be classified as ranging from 1) "no access" (HapMap, 1000 Genomes, dbGAP), to 2) limited access and no vetting exam (ClinSeq, CPMC and REVEAL), to 3) full access based on obtaining a 100% score on an exam covering risks of data sharing and re-identification (PGP).

B. Sharing: Since individuals can now easily get their medical and genomic data in digital form (outside of their actual "medical records" or any research project), and since individuals can have motivations to share these data, we can let this happen with or without scientific / non-profit / IRB guidance. If we choose the "without" route, then we will likely see Facebook / for-profit / non-IRB "DTC genomic research" proliferate. Projects that choose the "with" route, such as PGP, aim to set higher standards for how much knowledge citizens demonstrate about genetics and research before they give or receive data. The risks for both individual and society of sharing data are likely lower than many occupations (e.g. police and taxi) and possibly lower than the risks of "not sharing," but those risks still need to be communicated and appropriate guidance provided (pdf).

C. Science: Access to information can be restricted via fees, legal threats, technological censoring (e.g. GPS and encryption algorithms), and study design (eliminating useful data linkages). What has been the impact of such restrictions on science, on serendipity, collaboration, interdisciplinary research, etc. in the past? Will computer experts, (with artists, writers, etc.) create user interfaces, de-mystifying huge case-control studies in open-access systems or in closed? Will physicists and chemists make whole systems biology models if they can only see part of the data (or none of it)? Will social scientists (with ethicists and policy experts) discover alarming (or hopeful) trends in open-access systems or closed? Do we really know in advance who will contribute and who will not? Will we prioritize access based on willingness to jump through bureaucratic hoops? Is that likely to maximize the number of creative interdisciplinarians or produce the biggest out-of-the-box analytic breakthroughs? This is not about mere inconvenience, it is about a series of totally missed opportunities. The predictable positive impact of open-access is huge, and add to that impacts far beyond what we can currently predict.

D. Politics:

1) Retroactive activism: One could argue that current case-control cohorts and biobanks have enough momentum that nothing new can compete. But monuments do topple. If enough volunteers request / demand their data, then there may be pressure to give it to them, no matter what the original contract said. Any claim that the data or cells are de-identified will be untenable, since past volunteers can inexpensively provide DNA identifiers (say 100 SNPs).

2) Proactive: More importantly, going forward, larger biobanks and cohorts will likely be the most useful and new recruits may increasingly migrate to the most transparent and scientifically exciting projects.

3) Reactive: The press and the public will react to efforts that permit people to publicly share their own data; but any criticism is likely to be much less severe than the backlash following the accidental (or intentional) release of multiple volunteers without their permission. Keeping secret data about people that they cannot access will perpetuate distrust of science. In contrast, celebrating volunteers willing to become informed and share their medical information might inspire the public in a manner like astronauts in the 1960s.