In less than a decade, whole genome sequencing has moved from a big science, moon landing-style federal project to a seedbed of entrepreneurial activity. One important resulting development is the emergence of online genetic genealogy companies that enable individuals providing their own genetic samples to search extensive genetic databases for ancestral origins and more specific family connections. This is possible in large part because the Y chromosome and (in nearly all cultures) the family surname are passed from father to son; unique, highly polymorphic short tandem repeats across the Y chromosome (Y-STRs) can thus be identified and linked to family names.

Individuals have used this capacity increasingly successfully to track down family relations who—through giving up a child for adoption or confidential sperm donation—originally intended to remain anonymous. In one instance, a fifteen-year-old was able to track down his biological father using the Web and a few hundred dollars. An article by Melissa Gymrek and colleagues, published this January in Science, took things a step further: it described how the researchers used surname inferences from commercial genealogy databases and Internet searches to deduce the identity of nearly fifty research participants whose supposedly private data were stored in large, publicly available datasets. The authors projected that they could recover the surnames of approximately 12 percent of U.S. Caucasian males in this fashion.

This news comes just months after the Presidential Commission for the Study of Bioethical Issues published a report, Privacy and Progress in Whole Genome Sequencing, that expressed serious concerns about personal privacy and security in whole genome sequencing. The bioethics commission (on which we serve as chair and vice-chair) highlighted the importance of reconciling the enormous public benefits anticipated from research in this area with the potential risks to individuals’ privacy. The report also offered several specific policy proposals to help balance the potential of scientific progress with privacy and respect for persons, including the need to expand our ethical and regulatory focus to encompass both identifiability and privacy, and to establish clear policies and norms that define and enforce acceptable access to and use of genomic data.

By collecting large amounts of genomic data in databases and biorepositories, investigators are able to analyze genetic variations to better understand health and disease processes, discover innovative solutions to medical problems, and personalize therapeutic care. In order to promote widespread participation in these efforts, researchers typically attempt to ensure participants’ complete anonymity by “deidentifying” genomic data—stripping it of name, date of birth, social security number, photographs, or other traditional identifiers. It was thought that in this state, unique genetic information could not be linked to specific individuals. But clearly, the more genomic data collected, and the more refined the connections between genetic variations, disease states, and other personal characteristics, the easier it becomes to reidentify an individual and discover...
private information. Large-scale collections of genomic data therefore raise very serious privacy concerns for individuals who have contributed to them, their family members, and even their children and descendants yet unborn.

Advances in genomic research also raise important issues related to data storage and sharing since genomic databases increasingly surpass the storage abilities of individual researchers—and even of large institutions—and consequently are stored increasingly frequently in the “cloud.” While informed consent forms used in the datasets Gymrek et al. examined contained language acknowledging the potential for privacy violation, the terms of use of some databases do not explicitly prevent authorized users from working to reidentify participants. The bioethics commission noted in its report that many members of the genomic research community expressed concern about the hypothetical possibility of this kind of privacy violation; the work by Gymrek et al. makes abundantly clear that these concerns are no longer hypothetical.

As the bioethics commission noted, these concerns demand “careful and sustained ethical attention” and also a quick response from policy-makers.

The Paradigm of Identifiability

Currently, the human subjects research protections laid out in the federal regulations known as the Common Rule (45 CFR 46) are triggered by the identifiability of data. The Office for Human Research Protections guidance clarifies that the identifiability of data depends on whether a researcher may readily ascertain the identity of a subject through, for example, the use of a key. Research use of whole genome sequence data that has been stripped of traditional identifiers (such as name, geographic location smaller than a state, full face photograph, Social Security number) would therefore not be considered “identifiable.” If genomic data are not associated with information linking them to a particular person, then research with the data is not considered human subjects research. While any whole genome is uniquely identifiable (because DNA is unique to only one person), genomic data stripped of traditional identifiers are not readily identifiable because there is no key matching particular data to a particular person.3

In a response piece to Gymrek et al., Laura Rodriguez and colleagues call for a reconsideration of “whether a simplistic distinction between identifiability and nonidentifiability remains adequate as a metric for describing expectations about participant protections.” The authors point out that “some have suggested framing the risk of identifiability along a continuum rather than as an absolute.”6

It would be hard to argue that the nearly fifty people identified by Gymrek et al. were “readily identifiable.” The reidentification procedure required many steps, suggesting that, at least currently, it was a process beyond the comprehension of the average Web user. But it has been proven possible and should be addressed. The science has advanced to the point that identifiability is no longer a bright-line distinction, and our locus of concern must also evolve. The critical question is less about what makes something identifiable—since the availability of technology might make this a moot point in the near future—and much more about what keeps something private.

Consider an individual who volunteers for a research study of genomic data that are supposed to be stripped of traditional identifiers. An error occurs, resulting in the posting of the individual’s data, along with full name and date of birth, in a publicly accessible biorepository for a limited period. No researchers access the database during that time, and no one ever sees the information. Compare this to a Utah male with Northern and Western European ancestry, one of fewer than one hundred such volunteers to deposit genomic data to the 1000 Genomes Project, who is now among the names that Gymrek et al. have decoded alongside his genome. Both are identifiable, but the person whose identity is never revealed clearly has more privacy.

Privacy is secured or breached based on several key steps, starting with the individual’s decision to disclose the information, then moving to data access, and finally culminating in the use (or misuse) of those data. The ethical tools associated with protecting privacy at each of these steps must therefore consist of both informed consent and the establishment of a consistent floor of protections for access; it must include policies for maintaining security of data, as well as consistent and well-defined professional expectations for individuals who work with data, including accountability for violations.

Informed Consent, Access, and Use

Current federal and state laws protect genomic data in a discordant fashion. Many states base their protections on the collector of the genomic data—clinician, researcher, or consumer. In addition, many patients do not realize that their genomic data acquired in the clinical setting might be stripped of traditional identifiers and used for research. The bioethics commission recommended that consent processes should ascertain participant or patient preferences at the time the samples are obtained—even if future uses are planned for data stripped of traditional identifiers.8 In addition, while states should enact genomic policies that are most relevant and important to their constituents, bringing such protections to a minimum standard that addresses privacy—while still allowing individuals to share their own data—would provide just and fair protections regardless of where one happens to reside.

It is hard to control access to genomic data—we leave DNA specimens behind with every discarded coffee cup. However, just because an individual “abandons” DNA in a public place does not mean that the individual has abandoned interest in maintaining the privacy of the data that sample contains. The bioethics commission notes that unauthorized access to data is not necessarily the crux of the problem. Rather, it is the misuse of information that can be more ethically significant than unauthorized access.9 The individuals who unwittingly
participated in the Gymrek et al. study may have been warned of possible privacy breaches in the consent form. It may have simply been impossible to prevent reidentification. Yet clearly neither the individuals nor the researcher who initially collected the samples intended the data to be put to such use. Gymrek et al. acknowledge this, noting that “establishing clear policies for data sharing, educating participants about the benefits and risks of genetic study, and the legislation of proper usage of genetic information are pivotal ingredients to support the genomic endeavor.”^10

For these reasons, the bioethics commission recommended that all persons who work with whole genome sequence data should be guided by professional ethical standards related to privacy and confidentiality and not intentionally, recklessly, or negligently access or misuse data. As John Wilbanks said in testimony at a public meeting of the bioethics commission, “harm is not the act . . . of distributing data. Harm comes from actions that are taken once the data has been distributed.”^12 Professional ethical standards should guide researchers, clinicians, and others with authorized access so that they do not intentionally or inadvertently misuse genomic data.^13

To encourage public confidence in genomic research and to protect sensitive data, three foundational elements of privacy protection should be enacted:

1) Informed consent must provide individuals who submit specimens with information about the potential risks (including those related to informational privacy) and benefits of participation, and ascertain participant preferences for data use;

2) Laws, regulations, institutional policies, and information technology systems controlling access to data must be explicit, transparent, and consistent with consent and participant preference, and they should meet a common “floor” of protections across every U.S. jurisdiction; and

3) Individual researchers and clinicians, along with all those who work with genomic data, must be guided by professional ethical standards related to privacy and confidentiality and be held accountable to all state and federal laws and regulations in case of breaches of identity or confidentiality.

The principle of respect for persons compels researchers to gain informed consent before conducting any whole genome sequencing. Laws and policies should support this consent, protect individuals from unauthorized access or misuse of data, and appropriately sanction violators. Equally important, individuals who collect, handle, store, and use data must understand how critical it is to protect the privacy of the people from whom the data came, no matter how laws, regulations, policies, or information technology systems might otherwise fail them.^14

A point upon which Gymrek et al., Rodriguez et al., and the bioethics commission agree is that health care will benefit from advances in genomic technologies most rapidly when the public is willing to participate in genomic research and share their individual data.^15 Appropriate policies, however, must encourage and protect this sharing. To contribute, the public must trust that the research community will guard their data zealously and use them in the most respectful and productive manner possible. Donor consent to whole genome sequencing is critical, but warning participants of risks in a consent document is not enough: such risks must also be prevented whenever and wherever possible.

While the participants in the Gymrek et al. study were not “readily identifiable,” the data proved far more easily identifiable than expected. With rapidly evolving technology, a precise definition may be impossible. But if we move the debate from the rhetoric of identifiability to the ethical principles of public beneficence and the centrality of respecting all persons, we find that the real ethical focus must be on promoting generalizable progress while at all times respecting individual privacy.

If we move the debate from identifiability to public beneficence and respecting persons, the real ethical focus must be on promoting progress while respecting privacy.

---

5. Presidential Commission for the Study of Bioethical Issues, Privacy and Progress in Whole Genome Sequencing, 64.
Jennifer spends lots of time with dead things, dead humans actually. She works in a pathology lab. One night, she is asked to incinerate a fresh human cadaver, and she is struck that it would be a waste to throw away perfectly good meat. So, even though she is against killing and is a vegetarian for moral reasons, she cuts a chunk of flesh off the cadaver, takes it home, and has it for dinner.

Got a problem with that? Almost everyone does. But the vast majority of people cannot explain why, at least not to the level that would earn a B-plus in a freshman philosophy seminar. Instead, as Jonathan Haidt explains in *The Righteous Mind,* it is often our moral intuitions that come first, rapidly or even automatically, with ethical reasoning coming later.

Haidt’s book is one of many that have come out in recent years highlighting the relevance of psychology (and its close cousin, neuroscience) for understanding human morality. As a behavioral scientist, I have devoured many of these books. I am fascinated by human nature and love trying to understand why all of us behave and think the way we do. But as a physician ethicist, I often find myself reading these books with a parallel agenda: not just to understand human nature, but also to see whether this line of research has relevance for my work as an ethicist. My bottom line: Understanding moral psychology ought to be a required component of ethics training, not because the science will help us to differentiate between right and wrong, but because it will better equip us to teach ethics in a way that promotes moral behavior in our students.

**Ethical Reasoning as Servant to the Moral Passions**

Take a moment to imagine the following scenarios, all created by Haidt to assess people’s moral intuitions. You learn about an infertile brother and sister who decide to have consensual sex with each other. You witness a woman cutting up an American flag and using the pieces to clean toilets. Most people react viscerally to these cases and conclude that they involve immoral actions. But why? Who, if anyone, was harmed by them? Even the case of Jennifer who eats the flesh of a human corpse—was anyone hurt by her actions? If not, what is so wrong about her dietary choice?

Most laypeople have an extremely difficult time explaining why they object to these actions. To illustrate, let me paraphrase a discussion between one of Haidt’s research assistants and a person who had just read the story of a brother and sister who had consensual sex with each other:

Respondent: “I just think incest is wrong.”

Interviewer: “What’s wrong with it?”

Respondent: “Um, the whole idea . . . if the girl got pregnant . . .”

Interviewer: “But they used a condom and birth control pills.”

Respondent: “Oh, okay yeah . . . well I guess the safest sex is abstinence but, um . . . I don’t know. I just think that’s wrong. I don’t know. What did you ask me?”