Published in final edited form as:

Per Med. 2009 November ; 6(6): 691–699. doi:10.2217/pme.09.48.

Eyes wide open: the personal genome project, citizen science and veracity in informed consent

Misha Angrist

Duke University Institute for Genome Sciences & Policy, 450 Research Drive, Room B321A, Durham, NC 27708–21009, USA, Tel.: +1 919 684 2872, Fax: +1 919 613 6448

Misha Angrist: misha.angrist@duke.edu

Abstract

I am a close observer of the Personal Genome Project (PGP) and one of the original ten participants. The PGP was originally conceived as a way to test novel DNA sequencing technologies on human samples and to begin to build a database of human genomes and traits. However, its founder, Harvard geneticist George Church, was concerned about the fact that DNA is the ultimate digital identifier – individuals and many of their traits can be identified. Therefore, he believed that promising participants privacy and confidentiality would be impractical and disingenuous. Moreover, deidentification of samples would impoverish both genotypic and phenotypic data. As a result, the PGP has arguably become best known for its unprecedented approach to informed consent. All participants must pass an exam testing their knowledge of genomic science and privacy issues and agree to forgo the privacy and confidentiality of their genomic data and personal health records. Church aims to scale up to 100,000 participants. This special report discusses the impetus for the project, its early history and its potential to have a lasting impact on the treatment of human subjects in biomedical research.

Keywords

citizen science; DNA sequencing; genetic privacy; informed consent; open consent; Personal Genome Project

The idea that patients have a right to have their personal medical information kept private dates back to at least the time of Hippocrates [1]. But for research subjects in the USA, the concept of respect for privacy more recently came to the fore, especially after high-profile incidents where subjects' privacy was violated. In 1954, social scientists recorded jury deliberations surreptitiously, an act that precipitated legislation making such action criminal [2]. From 1965 to 1968, in the course of his research, sociologist Laud Humphreys secretly followed closeted homosexuals, recorded their cars' license numbers and visited them in disguise (reviewed in [3]).

In 1991, the US Department of Health and Human Services (HHS) regulations governing all federally connected research on human subjects were revised and adopted by 16 federal

© 2009 Future Medicine Ltd

For reprint orders, please contact: reprints@futuremedicine.com

Financial & competing interests disclosure

Misha Angrist is a participant in the Personal Genome Project, but has received no compensation from the PGP or any of its affiliated personnel. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

agencies, thereby becoming known as the ‘Common Rule’ [101]. The Common Rule defines a human subject as a living individual about whom an investigator conducting research obtains data or samples through intervention or interaction with the individual, or identifiable private information that could be linked back to him/her. The Common Rule also includes a description of the Institutional Review Board (IRB), whose charge at every research institution includes making sure that: ‘when appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data’. The Common Rule requires that human subjects research that uses participants’ private, identifiable information be reviewed and approved by the IRB [101].

Vulnerability of genomic data

Participant privacy and confidentiality considerations are mainstays of human subjects research involving genetics and genomics. Perhaps the most salient illustration of this can be found in the recent turbulence surrounding the NIH policy regarding genome-wide association studies (GWAS) data. In implementing its data-sharing policy in 2007, the NIH’s expectation was that unless they could offer a compelling reason not to, NIH grantees would share their human genomic data with other investigators. It was also made clear that the decision to share or not to share could have tangible consequences:

“The ability to share, and the richness of the data for maximizing the usefulness of future research, may be considered ... as part of award decisions” [102].

However, the NIH abruptly altered this policy in 2008 following the publication of a paper demonstrating that it is possible to ascertain whether a particular individual’s DNA is present in a much larger pool of samples, even if that person’s sample represents less than 1% of the pool [4]. In response to this, three institutions that serve as major repositories for pooled human DNA data – the NIH, the Wellcome Trust and the Broad Institute – removed such data from their websites [5]. A letter to *Science*, from high-ranking NIH representatives, urged the scientific community to ‘consider carefully how these data are shared and take appropriate precautions’ in order to protect participant privacy and confidentiality [6].

Personal Genome Project: open season

Even before the publication of the paper describing the reidentification of individuals from pooled DNA, Hank Greely (Stanford, CA, USA) articulated two reasons why anonymization of human genomic information might be considered a dubious practice – it is extremely difficult to do in a truly secure way and it impoverishes the data [7]. Earlier still, Harvard (MA, USA) geneticist George Church recognized the tension between genomic privacy and unfettered genomic research. In his laboratory’s 2003 proposal to create the Molecular and Genomic Imaging Center at Harvard (MA, USA) [103], Church wrote:

“The core question is: how may the gathering of increasing amounts of genetic information be made compatible with ethical and legal requirements for privacy? Anything approaching a comprehensive genotype or phenotype (including molecular phenotypes) ultimately reveals subjects’ identity in our increasingly wired world as surely as conventional identifiers like name and social security number ... This raises numerous specific questions:

- Are current informed consent practices sufficient to give human subjects adequate understanding of the potential that their identity may be discernible in large genetic datasets?

- Is enough protection afforded by allowing researchers open access to such datasets so long as they agree not to take the analytical steps that would link these data to a specific person?
- Is there a kind and level of genetic information for which it would be virtually impossible for a researcher ... to link it with a specific person?"

A number of investigators have suggested that, as things stand, the answer to each of these questions is almost certainly no. Given DNA's ubiquity, the absence of laws that restrict surreptitious DNA collection, the frequency with which computers harboring personal health information are lost/hacked, the growing availability of genealogical records on the internet and DNA's unique digital characteristics, reidentification is likely to remain a perennial risk [8–14]. In 2006, the American Society of Human Genetics issued a statement expressing concern regarding the increasing likelihood of reidentification of genetic and genomic research subjects [104].

In December 2006, the NIH sponsored a Town Hall meeting as a designated public forum to discuss the proposed policy for the sharing of GWAS data [105]. During the question and answer session, the then-National Human Genome Research Institute (NHGRI) Director (and now NIH Director) Francis Collins admitted that an openly consented cohort 'would certainly make life easier', and that going forward, 'the opportunity to do this is worth considering'. But in what I presume to be a reference to the Personal Genome Project ([PGP] and perhaps to the Watson and Venter genomes as well), Collins mentioned:

"... proposals that the first complete genome sequences ... be done very carefully from people who are less concerned about the vulnerabilities of having their information mined by others ... Those proposals are also somewhat controversial because of the sense that this might actually introduce a bias or focus specifically on people with ... a lot of resources because those are often the people who are least worried about the discrimination issue" [105].

In a subsequent commentary in *Science*, William Lowrance (Consultant in Health Research Policy and Ethics, La Grande Motte, France) and Francis Collins worried that 'novel arrangements' that returned research results to patients and participants were '... relatively uncharted territory with respect to human subjects and privacy considerations. Precedent doesn't provide sufficient guidance' [15].

In contrast with the NIH's view of privacy and confidentiality being paramount, the PGP brain trust has argued that promises of privacy and confidentiality are untenable and that researcher veracity should be the 'primary moral obligation' [16]. Thus, at least one source of tension between standard informed consent protocols and open consent, as exemplified by the PGP, might be expressed in terms of a clash of ethical values – privacy (an aspect of autonomy) versus veracity. The PGP would argue that without veracity, authentic autonomy cannot exist.

In order to champion the latter, Church developed personalgenomes.org as a 501(c)(3) charitable organization [106,107]. As part of this organization, the PGP seeks to foster and conduct research on the development of personal genomics technologies and practices that yield identifiable and improvable benefits at 'manageable' levels of risk, and that are broadly available for the benefit of the general public. The PGP has developed software in-house at Harvard Medical School (MA, USA) to manage and interpret large quantities of genomic data and associated phenotypes from PGP participants who consent to have their DNA sequences, cell lines and phenotypes made public (phenotypic data collection and dissemination remain in the rudimentary stages). The Project will also analyze participants' cells (including the construction of induced pluripotent stem cell lines), transcriptomes,

microbiomes, epigenomes, *cis*-regulomes, VDJomes (unique B- and T-cell receptors owing to genetic recombination of the V, D and J segments of immunoglobulins) and microRNAs [108]. Between 2006–2007, the first ten participants (the PGP-10) consented and donated blood to the Project (I am one of these initial ten) [109]. Recently, Church described the PGP as ‘... maybe the only project where genes and traits are brought together into one dataset that can be shared openly ...’ [110]. Participants are considered to be co-drivers of the project [George M Church, Harvard, MA, USA & Misha Angrist, Duke University Institute for Genome Sciences & Policy, NC, USA. Pers. Comm.] and as such are encouraged to make suggestions and ask questions. I completed an early version of the entrance exam and suggested changes to certain questions. I have also queried the PGP regarding their data release policy, sample availability and the rationale for specific analyses (e.g., the creation of induced pluripotent stem cells from our tissues). Three of the PGP-10 are on the PGP Board of Directors and have a direct role in its governance [Jason Bobe, Harvard, MA, USA & Misha Angrist, Duke University Institute for Genome Sciences & Policy, NC, USA. Pers. Comm.]. The PGP is funded by foundations, companies and private donors [111]. Figure 1 illustrates the current flow of information through the Project from the participants’ vantage point.

Citizen science & bottom-up genomics

What the PGP is proposing to do – collect genotype and phenotype information on 100,000 individuals and make the information public with no serious attempt at deidentification – is unprecedented. It is as if deCODE Genetics (Reykjavik, Iceland) [17], the Framingham Heart Study [18], the Estonian Genome Project [19] and/or the UK Biobank (Stockport, Cheshire, UK) [20] announced that they were going to have a perpetual ‘open house’ and all would be invited to peruse their databases. If a peruser found something of interest in one or more participants, he/she would not find it all that difficult to contact those people.

At the time of this writing (October 2009), the PGP-10 have received minimal genomic data about themselves. The public has access to our Affymetrix (CA, USA) SNP data, that is, 500,000 markers, and very preliminary exon data [109]. In October 2008, each of us had a genomic consultation with PGP medical geneticist, Joe Thakuria; however, in light of the scant available data, these meetings (I sat in on a few others) felt like a dress rehearsal.

Despite its active engagement and ongoing return of all results to participants, the PGP is hardly the first manifestation of ‘citizen science’. Volunteers have observed, collected and/or analyzed data without specialized training since, at least, the initiation of the Audubon Society’s Christmas Bird Count in 1900 [21]. More recently, citizen science has morphed into ‘community- based participatory research (CBPR)’, that is, collaborative research meant to include the participation of communities affected directly by the issue(s) under study. CBPR is characterized by the reciprocal transfer of expertise, shared decision-making and mutual ownership of the research enterprise [22]. Its growth can be readily seen in the realms of environmental and occupational health [23] – when communities are exposed to environmental agents, they often take an interest in measuring such exposures beyond those undertaken by regulatory agencies [24]. Further evidence of CBPR’s rapid development can be seen in the emergence of targeted fellowship and training programs, journals interested in publishing CBPR and new NIH-sponsored funding vehicles for CBPR (reviewed in [25]).

Historically, DNA-based research has not been a major player in citizen science or CBPR projects owing to the prohibitive costs of sequencing and genotyping. Artists and activists who have embraced DNA-related ‘biotech hobbyism’ have mainly provided social commentary [26]. However, sequencing costs have plummeted by orders of magnitude in recent years [27], making genomics-informed citizen science possible. For example, former

Church student Kay Aull has undertaken genetic testing of herself for hemochromatosis susceptibility using equipment purchased on eBay (CA, USA) and installed in her bedroom closet [28]. Indeed, any ambitious amateur molecular biologist could set up a DNA sequencing lab with a few thousand dollars worth of second-hand equipment [29]. For less hands-on types, the company 23andMe (CA, USA) offers whole-genome scans of 600,000 polymorphic DNA markers for US\$400 [30] and surveys its customers regarding their trait data [31]. Individuals can participate in the Coriell Personalized Medicine Collaborative [112] at no cost. The PGP is sequencing the whole genomes of the first ten participants and is currently evaluating the scalability and utility of generating these data for the planned 100,000 participants [113].

The PGP aims for its participants to be truly informed and not merely passive signatories of lengthy and inscrutable consent forms. As illustrated in Figure 1, the eligibility screening process includes a questionnaire regarding family circumstances and privacy preferences. Prospective participants review a study guide that discusses the potential risks of participating, outlines PGP protocols and provides a primer on basic genetics and genomics. They then take an exam that covers each of these areas: how the PGP works; knowledge of genetics; ethical principles governing human subjects research; and comfort with the idea of having one's genome and health records in the public domain [114]. Sample multiple-choice questions include:

- What are the three ethical principles espoused in the Belmont Report?
- What is the relationship among genes, DNA and chromosomes in humans?
- Huntington's disease is a genetic disorder caused by a dominant gene. Symptoms begin in adulthood and the disease is ultimately fatal. What is the ethical dilemma presented by Huntington's disease when a parent is diagnosed with the disease?

All would-be participants must score 100% on the entrance exam. If they do not, they are free to re-take it. Subsequently, in the interests of both veracity and autonomy, would-be participants are presented with the actual consent form, which emphasizes the possibility of reidentification, disclosure of nonpaternity and loss of insurance. Indeed, it also discusses more remote (some would say fanciful) scenarios precipitated by disclosure of one's genome, including learning of one's relationship to 'infamous villains', being framed for criminal activities and being cloned.

Personal Genome Project participants are not charged for their data. The project does make a nonbinding request of participants for financial support; however, the PGP insists that one's decision to pledge support has no bearing on one's chances to be selected for enrollment [106]. I believe this claim and I understand why the request for money is made – the prospect of a 'free genome sequence' may act as a coercive lure; the PGP is a nonprofit project, it does not have much in the way of government support and foundation largesse can be unpredictable. That said, this model makes me uncomfortable. In an age of direct-to-consumer genomics, I see asking for money as opening the door to the therapeutic misconception [32] and perhaps diagnostic or other sorts of misconceptions as well, no matter how thorough the exam and the consent process. One can imagine a participant saying, 'I jumped through every hoop, gave a pound of flesh and even paid you guys ... and this is all I get?' Will extremely wealthy donors expect special treatment? Finally, our country has a history of researchers exploiting patient tissues for financial gain – Henrietta Lacks and John Moore come to mind [33]. Nothing in my 3-year involvement with the PGP has led me to think that the Project will exploit its enrollees. Nevertheless, I would feel better if the PGP found ways to pay for itself that did not make any presumptions about participants' financial generosity.

Sequences & consequences

Like any start-up enterprise, the PGP has not had an entirely smooth ride. Church has been derided (unfairly, in my opinion) as a ‘celebrity genome’ [34]. To date, the NHGRI has declined to fund the project, although the National Heart, Lung and Blood Institute (MA, USA) has funded some targeted sequencing of PGP samples [115].

Technical challenges have occurred as well. One of the original goals of the PGP was to sequence each participant’s exome, that is, the complete set of exons in each genome, or approximately 1% of the human genome. The exome encompasses the entire protein-coding portion of the genome and is believed to harbor much of the functional variation [35]. In 2006, Church told me that he would rather have 100,000 exomes than 1000 complete genomes. Indeed, exomes have the potential to be powerful tools for gene discovery and analysis in humans [36]. But until recently, exon capture was not a trivial process. Moreover, costs for whole-genome sequencing have dropped significantly faster than the costs for exome capture and sequencing [Chad Nusbaum, The Broad Institute of MIT and Harvard, MA, USA & Misha Angrist, Duke University Institute for Genome Sciences & Policy, NC, USA. Pers. Comm.]. Church believes that if sequenced exomes are only a third or a fifth of the cost of sequenced whole genomes, they are probably not worth it. However, if they are ten-, or better still, 30-fold cheaper than whole genomes, then they almost certainly would be worth the cost [George M Church, Harvard, MA, USA & Misha Angrist, Duke University Institute for Genome Sciences & Policy, NC, USA. Pers. Comm.]. Exome sequencing is likely to be financially attractive for the near future, but as whole-genome sequencing costs continue to decline, it is not clear how long that situation will last.

And what of those exomes and genomes? PGP medical geneticist Joe Thakuria has the unenviable task of ‘triaging’ participants’ SNPs and sequence data. In some cases, suspicious-looking variants have been re-sequenced in a certified clinical laboratory. The PGP now uses an in-house software program (‘The Trait-o-Matic’ [developed in George Church’s laboratory, MA, USA]) to annotate participants’ genomic variants and to link out to other databases of human genetic and genomic variation [116]. Genome interpretation by the PGP (and by everyone else) remains an iterative, constantly evolving process.

Finally, the introduction of open consent, even on a small scale, has made the option of partial disclosure/redaction of genetic information more difficult to implement. All of the PGP-10’s cell lines and DNA samples are housed in the Coriell Institute for Medical Research (NJ, USA) biorepository and made available to qualified researchers at minimal cost [115]. But my decision not to make my exome data public until I was able to determine my *BRCA1* and *BRCA2* genotypes (I have a family history of breast cancer and two young daughters) led to the exclusion of my B-lymphocyte cell line from the Coriell catalog. I was subsequently tested for *BRCA* mutations by a commercial laboratory and, upon receiving a negative test, informed the PGP that I would not redact anything further and that my cells could be made available. Obviously, by making one’s cell line available to the public, any intent to redact becomes potentially meaningless – one can isolate DNA from human cells and sequence any bit of the genome one wants.

And had I tested positive for a *BRCA1* or *BRCA2* mutation? Obviously, this is where the rubber meets the road. I cannot say for certain what my wife and I would have done (we discussed it several times), but I probably would have asked to have at least that locus redacted with the understanding that anyone who was motivated enough could still secure my cells and sequence it. But even then, figuring out what my daughters inherited from me would still be a probabilistic exercise. The larger question – one put to me by PGP staff members on more than one occasion – is what happens years from now (or tomorrow) when

another deleterious mutation is found in my genome, but my genome has already been published in full? I would offer two responses. First, the probabilistic nature of our genomes – and especially what parts of them we pass to our descendants – remains. Second, like participation in any research endeavor, participation in the PGP is, finally, not a matter of legalities or policymaking, it is personal, it is an act of faith. As the Estonian philosopher Hermann Alexander Graf Keyserling wrote, ‘Faith, like courage, rests on consent to uncertainty’ [37].

It could be argued that anyone wishing to redact anything is not an ideal candidate for PGP participation anyway, and even the notion of redaction itself runs counter to the *raison d’être* of the project. However, in my view, just as people may elect not to disclose their incomes or ethnicities when they complete surveys, many research participants will want to pick and choose which parts of their genomes they share with others, if they can, assuming they have not consented to making cell lines permanently available. James Watson wanted to redact his *APOE* genotype and thereby not learn about a single allele he might carry that would predispose him to Alzheimer’s disease. Alas, even this proved to be tricky [38]. I suspect that selective disclosure – as well as a more systematic collection of phenotypic data – will have to wait until the adoption of personally controlled health records becomes more widespread. Expectations for personally-controlled health records are currently high while preparedness to use them appears to be low [39].

Conclusion & future perspective

The PGP has coalesced over the last few years as the result of the precipitous drops in the cost of DNA sequencing, the inadequacies of the current informed consent paradigm for largescale, whole-genome-based research and the vision and doggedness of George Church. I would argue that, thus far, the PGP has made more rapid progress in the ethical, legal and social realms than in the technological one. The project has multiple IRB approvals, a data safety monitoring board, and retains the services of a bioethicist, an anthropologist, a law firm and an ethics advisory board. Meanwhile, a fair amount of the early stages of the DNA sequencing has been carried out on commercial sequencing platforms outside the confines of the Church laboratory. As of October 2009, participants had received relatively scant sequence data. That said, the infrastructure is now in place to accommodate thousands of would-be participants. At personalgenomes.org [107], one can read the consent forms, study for the exam and begin the processes of eligibility screening, pre-enrollment/enrollment and ongoing participation (Figure 1).

Could something undesirable – a trait or genotype disclosure that leads to embarrassment, stigma, discrimination and/or other unforeseen problems – still happen to PGP participants? Absolutely – guaranteed assurances are a lofty but finally unattainable goal. The PGP cohort has been, and will be, screened to select individuals who can tolerate uncertainty. As such, this may introduce bias into studies, the extent of which cannot be realised in these early days. Garden-variety GWAS must contend with bias as well [40]; clearly, PGP-derived findings will demand scrutiny and independent replication.

George Church notes that for the last several years, the cost of DNA sequencing has fallen by an order of magnitude every year [41]. Clearly, such a pace is not sustainable. However, it is no longer difficult to imagine a day when whole-genome sequencing is considered as routine as newborn genetic screening and is no more expensive than a cholesterol test. When that day comes, and perhaps even long before, the PGP will have been instrumental in laying the groundwork for an alternative path in research of human subjects rooted in ‘straight talk’, and in amalgamating disparate kinds of data into a single comprehensive and open resource. Over the longer term, the hope is that the PGP will not only have been an exemplar

of ethical genomic research, but that it will have made a lasting contribution to our understanding of what it means to be human.

Acknowledgments

The author would like to thank the four anonymous referees for their helpful insights and suggestions. He is grateful to George Church, Jason Bobe and Jeantine Lunshof for generous access and for countless enlightening exchanges. He is indebted to Dan Vorhaus for astute comments and suggestions. Chad Nusbaum has been extremely helpful in discussions about DNA sequencing. The author would like to thank Amy McGuire for a conversation about consenting to uncertainty. He is grateful to his collaborators Louiqa Raschid, Ritu Agarwal, Brad Malin and Samir Khuller for their insights on privacy and citizen science.

Bibliography

Papers of special note have been highlighted as:

▪ of interest

▪▪ of considerable interest

1. Edelstein, L. The Hippocratic Oath, Text, Translation and Interpretation. Vol. VII. The Johns Hopkins Press; MD, USA: 1943. p. 64
2. Katz, J.; Capron, AM.; Glass, ES. Experimentation With Human Beings; the Authority of the Investigator, Subject, Professions, and State in the Human Experimentation Process. Vol. XLIX. Russell Sage Foundation; NY, USA: 1972. p. 1159
3. Faden, RR.; Beauchamp, TL.; King, NMP. A history and theory of informed consent. Vol. XV. Oxford University Press; NY, USA: 1986. p. 392
- 4▪. Homer N, Szelinger S, Redman M, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet.* 2008; 4:E1000167. The investigators demonstrated that individual DNA could be identified in a pool with as many as 200 other samples, thereby prompting the NIH to alter its genome-wide association study data-sharing plan. [PubMed: 18769715]
5. Couzin J. Genetic privacy. Whole-genome data not anonymous, challenging assumptions. *Science.* 2008; 321:1278. [PubMed: 18772401]
6. Zerhouni EA, Nabel EG. Protecting aggregate genomic data. *Science.* 2008; 322:44. [PubMed: 18772394]
- 7▪▪. Greely HT. The uneasy ethical and legal underpinnings of large-scale genomic biobanks. *Annu Rev Genomics Hum Genet.* 2007; 8:343–364. A devastating critique of anonymous biobanks and a clarion call for new modes of informed consent that gives research participants at least some control over, and access to, their own data. [PubMed: 17550341]
8. Greenbaum D, Du J, Gerstein M. Genomic anonymity: have we already lost it? *Am J Bioeth.* 2008; 8:71–74. [PubMed: 19003717]
9. Hull SC, Wilfond BS. What does it mean to be identifiable? *Am J Bioeth.* 2008; 8:W7–W8. [PubMed: 19003695]
- 10▪. Lin Z, Altman RB, Owen AB. Confidentiality in genome research. *Science.* 2006; 313:441–442. Points out the real-world difficulties of keeping genomic and phenotypic data secret. [PubMed: 16873628]
- 11▪. Malin B. Re-identification of familial database records. *AMIA Annu Symp Proc.* 2006:524–528. Points out the real-world difficulties of keeping genomic and phenotypic data secret. [PubMed: 17238396]
- 12▪. Malin B, Sweeney L. How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems. *J Biomed Inform.* 2004; 37:179–192. Points out the real-world difficulties of keeping genomic and phenotypic data secret. [PubMed: 15196482]
13. McGuire AL. Identifiability of DNA data: the need for consistent federal policy. *Am J Bioeth.* 2008; 8:75–76. [PubMed: 19003718]

14. McGuire AL, Fisher R, Cusenza P, et al. Confidentiality, privacy, and security of genetic and genomic test information in electronic health records: points to consider. *Genet Med*. 2008; 10:495–499. [PubMed: 18580687]
15. Lowrance WW, Collins FS. Ethics. Identifiability in genomic research. *Science*. 2007; 317:600–602. [PubMed: 17673640]
- 16•. Lunshof JE, Chadwick R, Vorhaus DB, Church GM. From genetic privacy to open consent. *Nat Rev Genet*. 2008; 9:406–411. Offers the most coherent presentation of open consent and the ethos behind the Personal Genome Project (PGP). [PubMed: 18379574]
17. Traulsen JM, Bjornsdottir I, Almarsdottir AB. ‘I’m happy if i can help’. Public views on future medicines and gene-based therapy in Iceland. *Community Genet*. 2008; 11:2–10. [PubMed: 18196912]
18. Govindaraju DR, Cupples LA, Kannel WB, et al. Genetics of the Framingham Heart Study population. *Adv Genet*. 2008; 62:33–65. [PubMed: 19010253]
19. Sutrop M, Simm K. The Estonian healthcare system and the genetic database project: from limited resources to big hopes. *Camb Q Healthc Ethics*. 2004; 13:254–262. [PubMed: 15283549]
20. An afternoon at UK Biobank. *Lancet*. 2009; 373:1146.
21. National Audubon Society. US Fish and Wildlife Service. American Birds. National Audubon Society; NY, USA: 1971.
22. Viswanathan M, Ammerman A, Eng E, et al. Community-based participatory research: assessing the evidence. *Evid Rep Technol Assess (Summ)*. 2004; 99:1–8. [PubMed: 15460504]
23. Cook WK. Integrating research and action: a systematic review of community-based participatory research to address health disparities in environmental and occupational health in the USA. *J Epidemiol Community Health*. 2008; 62(8):668–676. [PubMed: 18621950]
- 24•. Morello-Frosch, R.; Pastor, M.; Sadd, J.; Porras, C.; Prichard, M. Citizens, science, and data judo: leveraging community-based participatory research to build a regional collaborative for environmental justice in southern California. In: Israel, BA.; Eng, E.; Schulz, AJ.; Parker, EA., editors. *Methods for Conducting Community-Based Participatory Research in Public Health*. Jossey-Bass; CA, USA: 2005. p. 371-391. The authors articulate how, why and under what circumstances citizens become actively involved in scientific research (in this case, environmental) that has a bearing on their own health
25. Horowitz CR, Robinson M, Seifer S. Community-based participatory research from the margin to the mainstream: are researchers prepared? *Circulation*. 2009; 119:2633–2642. [PubMed: 19451365]
26. Riddell A. Tweaking genes in the basement. *Wired*. July 6.2006
27. Pettersson E, Lundeberg J, Ahmadian A. Generations of sequencing technologies. *Genomics*. 2009; 93:105–111. [PubMed: 18992322]
28. Johnson CY. Do-it-yourself genetic sleuthing. *The Boston Globe*. May 11.2009
29. Johnson CY. As Synthetic biology becomes affordable, amateur labs thrive. *The Boston Globe*. September 16.2008
30. Rowe A. Human genetics is now a viable hobby – 23andMe cuts its price to \$399. *Wired Science*. September 8.2008
31. Ray T. Going Wiki, 23andMe using web to recruit customers for disease-risk and ADR trials. *Pharmacogenomics Reporter*. June 4.2008
32. de Melo-Martin I, Ho A. Beyond informed consent: the therapeutic misconception and trust. *J Med Ethics*. 2008; 34:202–205. [PubMed: 18316464]
- 33•. Skloot R. Taking the least of you: most of us have tissue or blood samples on file somewhere, whether we know it or not. What we don’t typically know is what research they are being used for or how much money is being made from them. And science may want to keep things that way. *NY Times Mag*. April 16.2006 :38–45. 75, 79, 81. Examines several cases that illuminate the collision of research, tissue, intellectual property and commerce.
34. Check E. Celebrity genomes alarm researchers. *Nature*. 2007; 447:358–359. [PubMed: 17522639]
35. Ng PC, Levy S, Huang J, et al. Genetic variation in an individual human exome. *PLoS Genet*. 2008; 4:E1000160. [PubMed: 18704161]

36. Kryukov GV, Shpunt A, Stamatoyannopoulos JA, Sunyaev SR. Power of deep, all-exon resequencing for discovery of human trait genes. *Proc Natl Acad Sci USA*. 2009; 106:3871–3876. [PubMed: 19202052]
37. Keyserling, H.; Duerr, T. South American meditations on hell and heaven in the soul of man. 1. Vol. 4. Harper & Brothers; NY, USA: 1932. p. 3-420.
38. Nyholt DR, Yu CE, Visscher PM. On Jim Watson's *APOE* status: genetic information is hard to hide. *Eur J Hum Genet*. 2009; 17:147–149. [PubMed: 18941475]
39. Weitzman ER, Kaci L, Mandl KD. Acceptability of a personally controlled health record in a community-based setting. implications for policy and design. *J Med Internet Res*. 2009; 11:E14. [PubMed: 19403467]
40. Zhong H, Prentice RL. Correcting 'winner's curse' in odds ratios from genomewide association findings for major complex human diseases. *Genet Epidemiol*. 2009 (Epub ahead of print).
41. Church, GM. The \$0 genome & personalgenomes.org. Presented at: The Consumer Genetics Conference; Boston, MA, USA. 9 June 2009;

Websites

101. US Department of Health and Human Services; DHHS NIH, Office Human Research Protections. Title 45: Public Welfare, Part 46: Protection of Human Subjects (aka the Common Rule). 2005. www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm
102. National Institute of Health. NIH Genome-Wide Association Studies Data Sharing Plan. November 27. 2007 http://grants.nih.gov/grants/gwas/gwas_data_sharing_plan.pdf
103. Proposal to NIH to create the Molecular and Genomic Imaging Center. 2003. http://arep.med.harvard.edu/P50_03/
104. [Accessed 30 November 2006] American Society of Human Genetics statement on re-identification of genomic data. www.ashg.org/pages/statement_nov3006.shtml
105. National Institute of Health. Office of Extramural Research – NIH Town Hall Meeting; Bethesda, MD, USA. 14 December 2006; http://grants.nih.gov/grants/gwas/town_hall_mtg/index.htm
106. Karow, J. Personal genome project to enroll 100 more participants this summer; seeks to raise \$1.5m in donations. In *Sequence*. May 5. 2009 www.genomeweb.com/sequencing/personal-genome-project-enroll-100-more-participants-summer-seeks-raise-15m-dona?page=show
107. Personal Genome Project. www.personalgenomes.org
108. The Personal Genome Project research community. <http://openwetware.org/wiki/PGP:Studies>
109. The PGP-10. www.personalgenomes.org/pgp10.html
110. George Church on The Charlie Rose Show. June 19. 2009 www.charlierose.com/view/interview/10399
111. Karow, J. Personal Genome Project kicks off with 10 volunteers; full-scale effort to begin shortly. In *Sequence*. July 31. 2007 www.genomeweb.com/sequencing/personal-genome-project-kicks-10-volunteers-full-scale-effort-begin-shortly
112. The Coriell Personalized Medicine Collaborative™. www.coriell.org/index.php/content/view/92/257/
113. Karow, J. PGP to publish initial data sets next month as Church predicts \$1,000 genome in 2009. In *Sequence*. September 23. 2008 www.genomeweb.com/sequencing/pgp-publish-initial-data-sets-next-month-church-predicts-1000-genome-2009
114. Coriell Institute for Medical Research; search results for 'personal genome project'. <http://ccr.coriell.org/Sections/Search/Search.aspx?PgId=165&q=personal%20genome%20project>
115. Targeted 2nd generation sequencing in phenotyped Framingham and PGP populations. [http://crisp.cit.nih.gov/crisp/CRISP_LIB.getdoc?textkey=7691364&p_grant_num=5R01HL094963-02&p_query=\(pgp\)&ticket=102077676&p_audit_session_id=486760659&p_audit_score=14&p_audit_numfound=18&p_keywords=pgp](http://crisp.cit.nih.gov/crisp/CRISP_LIB.getdoc?textkey=7691364&p_grant_num=5R01HL094963-02&p_query=(pgp)&ticket=102077676&p_audit_session_id=486760659&p_audit_score=14&p_audit_numfound=18&p_keywords=pgp)

116. Download and installation – trait-o-matic GitHub.
<http://wiki.github.com/xwu/trait-o-matic/download-installation>
117. The Personal Genome Project: how it works. www.personalgenomes.org/howitworks.html

Executive summary

History

- The notion of privacy and confidentiality of health information has a long and storied history. In the USA, there were several key incidents and policy developments in the second half of the 20th Century that led to the current framework for protecting research subjects' identities, samples and data.

Vulnerability of genomic data

- Recent events – from lost and stolen laptops to the emergence of the ability to identify individual DNA samples from large pools – suggest that presumptively private human genomic research data are vulnerable to reidentification and are difficult to share in an anonymous fashion.

Personal Genome Project: open season

- The Personal Genome Project (PGP) was developed by Harvard's (MA, USA) George Church in response to falling costs in DNA sequencing, the precariousness of promises of confidentiality to human subjects in genomic research and the need for a public resource that could house and analyze large quantities of human biological and trait data without the impracticalities of deidentification.

Citizen science & bottom-up genomics

- By virtue of its entrance exam, return of all genotypic and phenotypic data to participants and ease of data-sharing, the PGP represents an unprecedented 'citizen science' initiative in the realm of genomics. However, by asking for financial contributions from participants, the PGP may be creating a perception problem for itself and misconceptions among participants.

Sequences & consequences

- The PGP continues to face challenges that are common to many, if not most, research enterprises such as obtaining funding, solving technical problems in data generation and interpretation, and developing infrastructure that will sustain the project and allow it to scale up.

Conclusion & future perspective

- Whether the open consent model is self-sustaining and widely applicable to human genomic research remains to be seen. But even if it is not, the PGP has already emerged as an intriguing and provocative alternative to the *status quo* in the interrogation of human biology.

Figure 1. The PGP: how it works

For further details see [117].

PGP: Personal Genome Project.