Genomes On Demand: User Attitudes On Commercial DNA Testing

Debjani Saha
Brook Stacy
Kiran Javkar
Sushant Patkar
Michelle L. Mazurek

University of Maryland
College Park, MD 20742, USA

dsaha@cs.umd.edu
bstacy@umd.edu
kjavkar@cs.umd.edu
spatkar@cs.umd.edu
mmazurek@cs.umd.edu

Abstract
The decreasing cost of sequencing the genome has led to the emergence of companies that sequence and analyze a user's DNA and return ancestry and health information at minimal expense. However, a market for such highly personal and sensitive information raises numerous significant privacy concerns. In this work in progress, we attempt to determine the extent to which users are aware of the depth and breadth of information that their DNA could potentially reveal, as well as evaluate users' attitudes towards commercial DNA testing in the contexts of privacy and information security. This work could help inform further avenues of research and identify areas for expanding user education.

Author Keywords
user attitudes; genetic privacy; commercial DNA testing

ACM Classification Keywords

Introduction
Commercial DNA testing has emerged as a popular service allowing consumers to learn about their genetic information. Companies such as 23andMe and Ancestry.com partially sequence customers' DNA and return results about their heritage, susceptibility to disease, and other personal information. However, such practices raise severe privacy

concerns. An individual's DNA contains a wealth of information about them and their relatives. The recent arrest in the Golden State Killer case illustrates how information gleaned from consumer genetic testing can be harnessed for public good, but also hints at means for achieving less altruistic motives. Genetic data has been used for past discrimination, resulting in the Genetic Information Nondiscrimination Act (GINA) and the Health Insurance Portability and Accountability Act (HIPAA). Despite these laws, regulations on the handling and use of genetic data are limited, with high potential for mishandling or abuse. Moreover, the full breadth of consequeces is unknown given the constant advances in DNA-related technologies and knowledge.

In this paper, we report on a pilot study aiming to evaluate user awareness of the personal information revealed through consumer DNA tests, as well as user knowledge and concern about the associated privacy risks. We conduct this evaluation through semi-structured interviews, asking users about their experiences with commercial DNA tests, the information they believe they are revealing, and their privacy concerns. This is the first study we are aware of that attempts a direct evaluation of user perceptions of genetic privacy in the context of commercial DNA testing.

Research Questions

Q1. What do users believe is revealed by their genetic information?

Q2. How concerned are users about genetic privacy?

Q3. Is their level of concern related to what they believe is revealed by their genetic information?

Q4. Does their level of concern influence their decisions about whether or not to participate in DNA testing?

Q5. How do users believe their genetic information is used by testing companies?

Related Work

The advent of technologies enabling robust sample preservation and next generation sequencing have reduced the cost of DNA sequencing and testing to the point where they can be provided as consumer services [7, 12]. This level of access comes with significant privacy concerns. Taking such a test reveals a host of potentially private data to outside parties (e.g. 23andMe). Further, closely related individuals share large portions of their genomes, enabling strong inferences about a person from their relatives' genomes.

The growing wealth of genetic information also renders entities who collect and host it susceptible to attack. Erlich details methods for reidentifying subjects in otherwise anonymized datasets [5]. Malin demonstrates it is possible to reidentify subjects from their genetic samples across multiple institutions despite measures taken to protect anonymity [8]. Moreover, datasets based on non-DNA nucleic acids such as microRNAs (not classically considered to be uniquely identifying), can be de-anonymized using membership inference methods with alarmingly high success rates of up to 90% [2, 3]. Ayday further describes genetic privacy issues from a security standpoint, then details cryptographic solutions for securely handling and sharing genomic data [1].

The legal climate surrounding collection and use of human genetic data also merits concern. Norrgard describes past cases of abuse of genetic information, but observes that they have been rare, and resulting legislation has unforeseen negative impacts on social health and research [9]. However, absence of legislation poses serious threats to individual privacy. Fendrick details the urgent need for privacy laws in genetic research, and the serious implications of their nonexistence [6]. Phillips also addresses the issue from a legal perspective, and attempts to provide consumers with information on relevant privacy risks [10, 11].

Caulfield investigates various aspects of Direct-to-Consumer (DTC) genetic testing and policy responses worldwide. Legal and ethical concerns have resulted in a ban on DTC genetic testing in Germany. Though not banned in the US, DTC companies have been criticized for false claims, engaging in unlicensed practice of medicine, and selling medical devices without appropriate regulatory oversight. Numerous organizations have either advised against the use of such unlicensed practice or recommended genetic counseling before resorting to DTC. Further, DTC companies

are not necessarily subject to HIPAA privacy regulations. There is also significant variance in privacy policies of DTC companies, with only 7 out of 32 having comprehensive consumer privacy protection policies [4].

Methods

This study was approved and conducted under the purview of the University of Maryland (UMD), College Park Institutional Review Board (IRB).

Participant Recruitment. Participants were recruited from the College Park, MD community. Flyers were posted on bulletin boards in public areas of the UMD campus, and emails were sent to UMD organizations whose members might be interested the study. Recruitment flyers and emails instructed interested individuals to fill out a screening survey (administered through Qualtrics). Eligibility criteria were that respondents be at least 18 years of age, and have taken or contemplated taking a commercial genetic test.

Interview Protocol. Eligible participants were invited to participate in the in-person interview. For this pilot study, we recruited a total of five participants, two of whom had taken a commercial genetic test (P003 and P005), and three of whom had not (P001, P002, and P004).

The interview was comprised of four sections, ordered as follows to mitigate priming effects: **Experience, Benefits and Drawbacks, Knowledge,** and **Privacy** (see sidebar, pg. 3). In particular, we discussed users' perceived benefits and drawbacks in general before delving specifically into privacy concerns. Each section began with open-ended questions (to obtain unprompted answers), followed by more targeted follow-ups to ensure important topics were addressed. For example, in the **Privacy** section, participants were first asked to describe what they believe a commercial DNA testing company does with customer data, then later asked about specific actions such a company may take (e.g. in-house research, third party sharing, etc.).

Interviews lasted 45 minutes per participant including an initial introduction, obtaining informed consent, the semi-structured interview, and a post-interview questionnaire used to collect basic demographic information (administered through Qualtrics). One author served as the primary interviewer for all interviews; the secondary interviewer rotated among other authors. The semi-structured interview was audio recorded for later transcription and analysis. Participants were compensated with $20 for their time.

Analysis. Interviews were transcribed and qualitatively analyzed using iterative open coding. MAXQDA software was used to aid this process. Two researchers developed codes independently, then met to unify the codebook and resolve disagreement, repeating this process as necessary until a final codebook was reached. After three such rounds, the researchers achieved Cohen's Kappa = 0.82, and the final codebook was applied to code the five transcripts.

Preliminary Results

Our five pilot interviews yield interesting preliminary results. Based on these results, we will update the interview protocol and conduct more interviews until saturation is reached.

What does DNA say? These initial interviews suggest that participants are aware DNA contains information about ancestry and disease markers, though no one mentioned the potential for further, yet undiscovered, insights as the role of DNA is better understood via advances in technology.

Privacy is a general concern, but not always a personal one. Although all participants were aware of general privacy concerns associated with sharing genetic data, this awareness was not necessarily reflected in their personal

Interview Outline

Experience: What was the user's experience with the tests?

Benefits and Drawbacks: What do they believe are the pros and cons of these tests?

Knowledge: What do they know about DNA and genetic information and inference?

Privacy: What are their concerns and opinions in various privacy-related contexts?

	Value	n
Age	18-29	3
	30-39	1
	50-59	1
Gender	Male	3
	Female	2
Ethnicity	White	2
	Asian	1
	Black/AA	1
	Other	1
Education	HS/GED	1
	College	1
	Bachelor's	1
	Master's	1
	Doctorate	1
Field	CS	2
	Physics	1
	BioE	1
	Food Sci	1

Table 1: Participant demographics

concerns. P001 suggested there was no cause for worry since they had nothing to hide. This sentiment was echoed by P002 and P003, who felt the benefits of these tests outweighed the concerns. P004 and P005 were personally concerned. For instance, P005 raised the possibility of being targeted by hostile groups on the basis of their genetics.

Personal privacy concerns affect participation. A participant's level of privacy concern did appear to influence their decision to take the test. Notably, P001 and P002 cited the cost of the genetic tests, not privacy concerns, as a major deterrent. In contrast, P004 explicitly cited their privacy concerns as the reason they chose not to take the test.

User autonomy. All participants agreed that the decision to take part in a commercial genetic test is the user's alone. While there was a consensus that the hereditary and possibly sensitive nature of the information in question merited prior, voluntary consultation with close relatives, the final decision to take such a test is the user's to make. However, social influence may have impacted participants' decisions, as three participants stated that friends who had taken the test made them more inclined to do so themselves.

Terms of use. There were mixed responses as to how long genetic data is stored and used. Two participants deemed it acceptable for a company like 23andMe to retain the data without explicit time limits, whereas three participants desired some restriction. Regardless of longevity, respondents unanimously agreed that the collected data should be stored in a secure environment with assurances of integrity and confidentiality. Participants were tolerant of in-house research by genetic testing companies (with the exception of P002), and accepted that the collected data may be furnished to third parties, but with certain expectations. P001 and P002 stated that customers should be told up front this sharing would happen, with P001 adding that consumers should also be given the option to opt out. Moreover, participants hoped that their data was anonymized.

Participants classified medical research for the public good as an appropriate use (in-house or third party), though it lies outside of the boundaries of the services companies advertise and provide to consumers. Marketing research by third parties such as pharmaceutical and medical device companies was viewed somewhat less unanimously, with arguments both for (personalized recommendations for treatments) and against (using data to pad profits).

Demographic influence. A participant's demographics does seem to influence their attitudes. P003 explicitly said that if they had children, they may not be so eager to release their genetic information to commercial entities. This participant was also older, so it is possible that their station in life made them less likely to see potential downsides to commercial genetic testing (they are well-established, so perhaps less concerned about the potential repercussions of mishandled data). Further, the only participant to actively choose not to take a commercial genetic test (P004) was a computer science student who likely had greater exposure to general data privacy and security concerns.

Summary

In this work in progress we present user attitudes towards commercial genetic testing in the contexts of privacy and information security. Preliminary results indicate that while users are aware of the various general privacy concerns at play, they are not necessarily reflected in an individual's personal concerns. Personal concerns do however inform their decision to take such a test, as well their attitudes towards appropriate handling and use of data collected by relevant parties. Further interviews are being conducted to validate and elaborate on initial findings.

REFERENCES

1. Erman Ayday. 2016. Cryptographic Solutions for Genomic Privacy. In *Financial Cryptography and Data Security (FC '16)*. Springer, Berlin, Heidelberg, Germany, 328–341. DOI: http://dx.doi.org/10.1007/978-3-662-53357-4_22

2. Michael Backes, Pascal Berrang, Anna Hecksteden, Mathias Humbert, Andreas Keller, and Tim Meyer. 2016a. Privacy in Epigenetics: Temporal Linkability of MicroRNA Expression Profiles. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Austin, TX, 1223–1240. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/backes_epigenetics

3. Michael Backes, Pascal Berrang, Mathias Humbert, and Praveen Manoharan. 2016b. Membership Privacy in MicroRNA-based Studies. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*. ACM, New York, NY, USA, 319–330. DOI: http://dx.doi.org/10.1145/2976749.2978355

4. Timothy Caulfield and Amy L. McGuire. 2012. Direct-to-Consumer Genetic Testing: Perceptions, Problems, and Policy Responses. *Annual Review of Medicine* 63, 1 (2012), 23–33. DOI:http://dx.doi.org/10.1146/annurev-med-062110-123753

5. Yaniv Erlich and Arvind Narayanan. 2014. Routes for Breaching and Protecting Genetic Privacy. *Nature Reviews Genetics* 15, 6 (2014), 409–421. DOI: http://dx.doi.org/10.1038/nrg3723

6. Sarah Fendrick. 2008. The Role of Privacy Law in Genetic Research. *I/S: A Journal of Law and Policy for the Information Society* 4, 3 (2008), 803–820. https://kb.osu.edu/dspace/bitstream/handle/1811/72811/1/ISJLP_V4N3_803.pdf

7. R.M. Iwasiow, A. Desbois, and H.C. Birnboim. 2011. Long-term Stability of DNA from Saliva Samples Stored in the Oragene Self-collection Kit. DNA Genotek. (2011). https://www.dnagenotek.com/US/pdf/PD-WP-005.pdf

8. Bradley Malin and Latanya Sweeney. 2004. How (Not) to Protect Genomic Data Privacy in a Distributed Network: Using Trail Re-identification to Evaluate and Design Anonymity Protection Systems. *J. of Biomedical Informatics* 37, 3 (2004), 179–192. DOI: http://dx.doi.org/10.1016/j.jbi.2004.04.005

9. Karen Norrgard. 2008. Protecting Your Genetic Identity: GINA and HIPAA. *Nature Education* 1, 1 (2008), 21. https://www.nature.com/scitable/topicpage/protecting-your-genetic-identity-gina-and-hipaa-678

10. Andelka M. Phillips. 2015. Think Before You Click: Ordering a Genetic Test Online. *The Scitech Lawyer* 11, 2 (2015), 8–11. https://www.americanbar.org/publications/scitech_lawyer/2015/winter/think_before_you_click_ordering_genetic_test_online.html

11. Andelka M. Phillips. 2016. Only a click away – DTC Genetics for Ancestry, Health, Love... and More: A View of the Business and Regulatory Landscape. *Applied & Translational Genomics* 8 (2016), 16–22. DOI:http://dx.doi.org/10.1016/j.atg.2016.01.001

12. Androniki Psifidi, Chrysostomos I. Dovas, Georgios Bramis, Thomai Lazou, Claire L. Russel, Georgios Arsenos, and Georgios Banos. 2015. Comparison of Eleven Methods for Genomic DNA Extraction Suitable for Large-Scale Whole-Genome Genotyping and Long-Term DNA Banking Using Blood Samples. *PLOS ONE* 10, 1 (2015), 1–18. DOI: http://dx.doi.org/10.1371/journal.pone.0115960