An Exploratory Ethnographic Study of Issues and Concerns with Whole Genome Sequencing

Emiliano De Cristofaro
University College London[†]
me@emilianodc.com

Abstract—Progress in Whole Genome Sequencing (WGS) will soon allow a large number of individuals to have their genome fully sequenced. This lays the foundations to improve modern healthcare, enabling a new era of personalized medicine where diagnosis and treatment is tailored to the patient's genetic makeup. It also allows individuals motivated by personal curiosity to have access to their genetic information, and use it, e.g., to trace their ancestry. However, the very same progress also amplifies a number of ethical and privacy concerns, that stem from the unprecedented sensitivity of genomic information and that are not well studied.

This paper presents an exploratory ethnographic study of users' perception of privacy and ethical issues with WGS, as well as their attitude toward different WGS programs. We report on a series of semi-structured interviews, involving 16 participants, and analyze the results both quantitatively and qualitatively. Our analysis shows that users exhibit common trust concerns and fear of discrimination, and demand to retain strict control over their genetic information. Finally, we highlight the need for further research in the area and follow-up studies that build on our initial findings.

I. INTRODUCTION

In the past half a century, research on DNA sequencing – the process of determining the precise order of nucleotides within a DNA molecule – has made tremendous progress. The first complete human genome was sequenced in 2003 within a 13-year, $3B project known as the Human Genome Project [32, 52]. Thereafter, there has been an exciting race toward faster, cheaper, and more accurate Whole Genome Sequencing (WGS) technologies. Costs have quickly plunged to $250K already in 2008 [41] and to a few thousand dollars in the 2010's [12, 13]. The long-anticipated $1,000 threshold was breached, in January 2014, by the San Diego-based company Illumina [42]. The landscape of companies and technologies in WGS is fast-evolving, nonetheless, it is not far-fetched to predict that, in 5-10 years, most individuals in developed countries will have access to WGS.

[†]Work done in part while at PARC (a Xerox Company).

The emergence of affordable WGS technologies represents an exceptional breakthrough, due to the expected medical and societal implications. Biomedical experts anticipate that advances in WGS will unlock the full potential of *personalized medicine* [20], i.e., the practice of tailoring pre-symptomatic examinations, diagnosis, and treatment to patients' genetic features. Arguably, the availability of patient's fully sequenced genome enables clinicians, doctors, and testing facilities to run a number of complex genetic tests in a matter of seconds, using specialized computational algorithms (as opposed to more expensive and slower laboratory tests). However, due to the extreme sensitivity of DNA data [11], WGS also comes at the risk of amplifying important security, privacy, and ethical concerns. These issues have been recently under the scrutiny of several researchers [3] and policy experts [40] – we overview them in Section II-B.

Motivation. The emergence of personal genomic tests as well as the anticipated availability of affordable WGS technologies, along with the associated privacy and ethical concerns, motivate the need for better understanding users' perceptions and attitudes. For instance, it is desirable to assess whether or not privacy perceptions and concerns experienced by patients correspond to what the scientific community would expect, as well as how to identify effective mechanisms to communicate potential privacy risks associated with genomic information and its disclosure. Even more importantly, much of the anticipated progress in personalized medicine, and human genome research in general, will not happen without experimental research relying on large cohorts of volunteers willing to share their genetic material.[1] Therefore, it is fundamental to understand—and address—individuals' concerns and fears (such as embarrassment or discrimination) in order to retain their trust and their willingness to participate.

A few studies (reviewed Section V) have analyzed individuals' response to learning the results of some genetic tests and potential discrimination concerns associated with them, but there is very little understanding of these concerns in the context of WGS, and the new scenarios enabled by it.

Contributions. This paper presents the results of a series of semi-structured interviews, involving 16 participants. We

[1]E.g., the US/Canada Personal Genome Project [9] and the UK 100K Genome initiative [50] rely on volunteers that agree to have their genomic data made publicly available.

aim to assess the perception of privacy and ethical issues with WGS as well as the attitude toward different sequencing programs. Results are analyzed both quantitatively and qualitatively, yielding a few interesting findings related to the issues of control and trust, as users exhibit common fear of discrimination and demand to strictly control how their genetic information is used. Our exploratory study is vital to guide follow-up (possibly larger) ethnographic studies on the topic, which we plan to conduct as part of future work.

Paper Organization. The rest of the paper is organized as follows: next section presents background information about genomics and privacy/ethical concerns, then, Section III presents our study design and participants' demographics, while Section IV analyzes study results, both quantitatively and qualitatively. After reviewing related work in Section V, the paper concludes in Section VI.

II. BACKGROUND

This section presents some background information on genomics as well as privacy threats stemming from advances in Whole Genome Sequencing (WGS).

A. Genomics Primer

We start with discussing Personalized Medicine – the practice of tailoring pre-symptomatic examinations, diagnosis, and treatment to individuals' genetic features. Genomic tests today are increasingly and more effectively used in healthcare. For instance, testing for the *tpmt* gene is already required prior to prescribing some drugs used for treating childhood leukemia [1]. Similarly, doctors prescribing Zelboraf (Roche's treatment for skin cancer) need to test the patient for the BRAFV 600E mutation. Other analogous examples include testing for mutations in the Philadelphia chromosome (in Acute Lymphoblastic Leukemia patients) or in BRCA1/BRCA2 genes (in breast and ovarian cancer patients).

Also, a few commercial companies, such as, 23andMe.com, provide customers with reports on predisposition to diseases and conditions, even though they do not yet rely on WGS. (Specifically, 23andMe.com provides customers with a low-cost (about $100) report assessing their genetic risk toward a number of diseases and conditions, by testing for almost a million specific Single-Nucleotide Polymorphisms (SNPs) [37, 47].) Also, several drugs (e.g., for cancer, HIV, or thrombosis treatment) are paired today with genetic tests needed to assess either the correct dosage or their expected effectiveness [1, 7, 38, 39].

Experts estimate that about a third of the 900 cancer drugs currently in clinical trials will come to the market with a DNA or other molecular test attached [7]. Although most predominant, cancer treatment is only one of the application fields of personalized medicine. For instance, for some cardiac patients, recovery from a common heart procedure can be complicated by a single gene responsible for drug processing, and selection of blood thinner drugs should depend on whether or not patient holds such a gene mutation [54].

In general, advances in WGS facilitate the understanding of the impact of genetic variations on the response to medications. For instance, researchers have discovered genes encoding Cytochrome P450 enzymes, which metabolize neuroleptic medications to improve drug response and reduce side-effects [10], and also genes involved in the action and metabolism of warfarin (coumadin) – a medication used as an anticoagulant [38].

The availability of whole human genomes will also facilitate a number of genetic tests that today are performed in vitro, by reducing costs and time. For instance, *ancestry and genealogical tests* allow individuals to trace their lineage by analyzing their genomic information (the scope of such tests being often quite heterogeneous). Ancestry testing is increasingly popular, e.g., to map one's own genetic heritage and/or find known ancestry. Several commercial entities (e.g., 23andMe.com and Ancestry.com) provide customers with reports on their genetic ancestry. They maintain a collection of sample genomes from individuals belonging to different ethnic groups, and compare them against their customers' genomic information to understand how they relate to known ethnic groups.

Genetic compatibility tests are used to let (potential or actual) partners assess the possibility of transmitting to their children a genetic disease with Mendelian inheritance [36]. For instance, *Beta-Thalassemia minor* causes red cells to be smaller than average, due to a mutation in the *hbb* gene. It is called *minor* when the mutation occurs only in one chromosome, while the *major* variant—that occurs when both chromosomes carry the mutation—is likely to result in premature death. Therefore, if both partners carry the *minor* form, there is a non-negligible chance that their child will carry the *major* variant.

B. WGS Privacy and Ethical Threats

While advances in genomics and sequencing technologies promise to unlock a wide range of medical and societal benefits, they also risk to amplify security, privacy, and ethical concerns [3].

The human genome not only uniquely and irrevocably identifies its owner, but also contains information about ethnic heritage, predisposition to numerous diseases and conditions, including mental disorders [8, 15, 17]. Furthermore, due to its hereditary nature, disclosing one's human genome also implies, to a certain extent, disclosing the genomes of close relatives [28].

Traditional approaches to privacy, such as de-identification or aggregation, might not be effective when applied to genomic information. [24] demonstrates the feasibility of re-identifying DNA donors from a public research database using information from popular genealogy websites. Additional results on DNA and genome re-identification include [21, 27, 33–35, 51].

As the human genome contains detailed information about susceptibility to somatic and mental conditions, ethnicity, etc., its disclosure is often associated to fears of *eugenism*, i.e., genetic discrimination, thus potentially affecting social

Age	N	%
18–24	2	12.5%
25–34	7	43.7%
35–44	3	18.7%
45–54	1	6.2%
55–64	1	6.2%
65–75	2	12.5%

Degree	N	%
College	4	25%
Master	8	50%
PhD	4	25%

Yearly Income	N	%
Less than $50K	3	18.7%
$50K-$75K	3	18.7%
More than $75K	10	62.5%

Westin Index	N	%
Unconcerned	4	25.0%
Pragmatist	7	43.7%
Fundamentalist	5	31.2%

TABLE I: Participants breakdown by age, education, (personal) yearly income, and Westin's Privacy Index [30].

dynamics as well as hiring and healthcare practices. (See [18] for a survey of self-reported genetic discrimination.)

These concerns creates the need for *informed consent* to guard against surreptitious DNA testing by requiring authorities and companies to obtain written permission from citizens before collecting, analyzing, storing or sharing their genetic information (e.g., preventing people from collecting hair or saliva samples and maliciously sequencing the victims' genome). However, on the other hand, such privacy-restrictive measures are often regarded as a potential obstacle to genomic research. Scientists typically sequence DNA from thousands of people to discover genes associated with particular diseases, thus, the informed consent restriction would mean that large genomic datasets could not be re-used to study a different disease – researchers would either need to destroy the data after each study, or track down thousands of former subjects for new authorizations [44].

As mentioned earlier, much of the anticipated progress in personalized medicine, and human genome research in general, will not happen without experimental research relying on large cohorts of volunteers willing to share their genetic material. Therefore, it is fundamental to analyze—and address—individuals' concerns and fears in order to retain their trust and their willingness to participate.

III. INTERVIEW METHODOLOGY

Our study consisted of a series of semi-structured interviews with 16 participants (8 female, 8 male), conducted in Summer 2013. It was reviewed and approved by PARC's Institutional Review Board (IRB). Participants were recruited using social networks and local mailing lists, announcing a study on the "knowledge and perception of DNA testing."

A. Participants

The 16 participants were US-based and ranged in ages from 18 to 74. We collected information about participants' personal yearly income, and assessed their Westin's Privacy Index [30], which classifies users among privacy fundamentalists, pragmatists, and unconcerned, according to their responses to a simple three-question survey. Demographics breakdown is reflected in Table I. Study participants volunteered to participate in the study, i.e., they received no monetary incentive: we decided to do so in order to recruit motivated users (who were also likely to possess some understanding of genetics).

NOTE: We deliberately decided to recruit users with college degrees and more participants in the 25-34 age range as they

constitute the representative population of early adopters of personal genomic tests and WGS, as suggested by [14, 46]. This is crucial for our study since WGS is still at an early (research) phase and is not yet available to "regular" users.

B. Experiments

Interviews lasted approximately 30 minutes (on average) and consisted of three parts: (1) first, we asked participants to provide 1-2 examples of genetic tests and describe their familiarity with genetics; (2) we interviewed participants while guiding them through a set of slides depicting a few hypothetical scenarios (this constitutes the core of our study, see details below); (3) in order to collect demographic information and assess their Westin Privacy Index [30], participants were asked to fill out a short survey.

We aimed to conduct the four experiments presented in Table II below.

Exp. A	Assessing the perception of today's genetic tests
Exp. B	Comparing the attitude toward different WGS programs
Exp. C	Assessing the perception of potential privacy/ethical issues with WGS
Exp. D	Comparing the response to medical, genomic, and personal information loss

TABLE II: Overview of study experiments and goals.

1) Experiment A: Perception of genetic tests: Participants were shown the following six examples of genetic tests. Note that the order in which the test scenarios were actually presented was randomized across multiple participants.

(A.1) Disease Predisposition (Doctor.) Scott goes to see his primary care doctor. The doctor asks Scott to run some genetic tests to assess (genetic) predisposition to certain diseases. After collecting a DNA sample (e.g., saliva), the clinic runs some tests on Scott's DNA, and the doctor tells Scott that he is predisposed to type-2 diabetes, but not to Alzheimer's.

(A.2) Genetic Compatibility. Emma and Scott are planning to have kids. Emma has Beta-Thalassemia *Minor*, a genetic disorder inherited by only one of her parents, which causes her red cells to be smaller than usual but poses no critical health threat to her. However, if Scott also has the same disorder, there is a chance that their kids will have the *Major* form, which may cause premature death. Therefore, Scott is advised to take a genetic test to make sure that he doesn't have the disorder.

(A.3) Genetic Ethnicity. Scott is a customer of a genealogy company, such as, Ancestry.com, which analyzes genetic ethnicity and reveals where one's ancestors were form. Scott needs to send a DNA sample (e.g., saliva) to the company, which, after running some tests, tells Scott that his ethnic origins are 39% Scandinavian, 17% Central European, etc.

(A.4) Disease Predisposition (Company.) Scott is a customer of a personal genomics company, such as, 23andMe.com, which provides its customers with a detailed report on their genetic chances of getting a number of diseases, like type-2 diabetes, Alzheimer's, schizophrenia, etc. For each disease, the report includes a risk percentage, the known average risk percentage, and a confidence score. Scott needs to send a DNA sample (e.g., saliva) to the company, which, after running some tests, provides Scott access to a Web interface with the report.

(A.5) Correct Drug Dosage. Scott needs to take a blood thinning drug like warfarin. His doctor, in order to better assess the right dosage for him, requests that Scott takes a genetic test. After collecting a DNA sample (e.g., saliva), the clinic runs some tests and, based on the results, the doctor prescribes the exact warfarin dosage for Scott.

(A.6) Correct Cancer Treatment. Scott is diagnosed with cancer. His doctor, prior to propose a treatment plan, requests that Scott takes a genetic test to assess whether or not he carries a certain DNA mutation. After collecting a DNA sample (e.g., saliva), the clinic runs some tests and, based on the results, the doctor suggests what the best treatment for Scott is.

After showing the above test scenarios, we presented a recap slide summarizing them. We then asked interviewees to compare those scenarios; specifically, we raised questions like *"If you were in Scott's shoes, would you do all of these tests? Which ones would you not?"* as well as more specific ones like *"If you were Scott, which tests would you feel more inclined to do? Which tests more reluctant to?"* Finally, we asked participants to actually **rank the six tests** from the one they felt most inclined to, to the one they felt most reluctant.

2) Experiment B: Attitude toward WGS: First, we presented to the participants one slide describing (with both text and pictures) Whole Genome Sequencing (WGS), i.e., the process of digitizing the complete DNA sequence of a human genome. We told participants that once a whole human genome is sequenced, all genetic tests, including those presented in Experiment A., could be performed on the sequenced data, using computer techniques. We also told them that computer techniques could be faster and cheaper than collecting a sample (and using expensive lab equipment) each time one needs a genetic test.

Afterwards, we presented (again, in random order) the following three scenarios.

(B.1) WGS with Healthcare Provider. Emma goes to see her primary care doctor, who tells her about the provider's experimental Whole Genome Sequencing program. He explains to her that, by having her genome sequenced, Emma may learn her genetic predisposition to diseases and conditions. He also tells her that the provider will keep her sequenced genome along with her medical information, so that they can use it to run genetic tests when needed for medical reasons. Also, the provider will offer a discount on the next healthcare bill.

(B.2) WGS with Personal Genomics Company. Emma decides to have her genome sequenced by using a personal genomics company, such as, 23andMe.com. The company charges her $100 to sequence her genome, give her a report about predisposition to diseases and conditions, and tell her about her genetic ancestry. Also, the company offers to store the data for her: if she wants to authorize her doctor to run some tests, she can grant him permission to contact the company.

(B.3) Data-only WGS (DVD). Emma decides to have her genome sequenced by using a sequencing laboratory, which charges her $100 to sequence her genome and send her a DVD with the data.

After being presented with the above scenarios, participants were shown a recap slide. We then invited them to compare these scenarios; specifically, we asked *"If you were Emma, would you participate in any of the three programs?* and *"Which program(s) would you feel more inclined to participate in? Which one(s) less inclined?* Finally, we asked participants to **rank** the three scenarios from the one they felt most inclined to, to the one they felt most reluctant to.

3) Experiment C: Perception of privacy/ethics issues with WGS: In this experiment, we presented (in random order) the following four hypothetical cases.

(C.1) Labor Discrimination. Emma is interviewing for a job and her potential employer requests to run a drug test and to access her genome. The employer finds out that Emma has predisposition to cancer and decides not to hire her.

(C.2) Health Insurance Discrimination. Emma wants to join a health insurance plan. The provider requests access to her medical record as well as her genome, and finds out that she has predisposition to leukemia. As a result, Emma needs to pay a higher premium.

(C.3) Sequenced Genome Stolen by Hacker. A hacker steals the file with Emma's sequenced genome and finds out her ethnicity/ancestry and that she has genetic predisposition to Huntington's disease.

(C.4) Sibling Donating Genome to Science. Emma's brother, Jack, decides to donate his genome to science (for medical research purposes). Since Emma's and Jack's genomes are 99.9% identical, Jack is actually "donating" Emma's genome as well but has not asked for her permission.

After being presented with the above scenarios, participants were shown a recap slide. We then asked to compare them; specifically, we asked: *"If you were Emma, what incident(s) would give you the most discomfort? Which one(s) the least?* Finally, we asked participants to **rank the four scenarios** from the one giving them the most discomfort, to the least.

Genetic Tests: More to less inclined	Avg	Stdev
(A.6) Correct Cancer Treatment	5.81	0.39
(A.5) Correct Drug Dosage	4.63	0.70
(A.2) Genetic Compatibility	4.06	1.25
(A.1) Disease Predisp. (Doctor)	2.63	0.99
(A.4) Disease Predisp. (Company)	2.13	0.70
(A.3) Ancestry Testing	1.75	1.09

(a) Ranking average and standard deviation for Exp A.

WGS Programs: More to less inclined	Avg	Stdev
(B.3) Data-only (DVD)	2.68	0.58
(B.1) Healthcare Provider	2.00	0.71
(B.2) Personal Genomics Company	1.31	0.46

(b) Ranking average and standard deviation for Exp B.

Incidents: More to less discomfort	Avg	Stdev
(C.1) Labor Discrimination	3.31	0.58
(C.2) Health Insurance Discrimination	3.00	0.94
(C.3) Sequenced Genome Hacked	2.56	0.93
(C.4) Sibling Donating Genome to Science	1.13	0.33

(c) Ranking average and standard deviation for Exp C.

Information loss: More to less frightened	Avg	Stdev
(D.1) Identity Theft	3.50	0.63
(D.3) Emails and Pictures	2.63	1.61
(D.4) Sequenced Genome	2.00	0.63
(D.2) Medical Records	1.88	0.48

(d) Ranking average and standard deviation for Exp D.

TABLE III: Summary of Experiments.

4) Experiment D: Response to information loss: In the last experiment, participants were asked to **rank the following four incidents** (presented in random order) from the one they found most frightening to the least.

(D.1) Identity. A hacker steals your financial information and your social security number, i.e., so-called "identity theft."

(D.2) Medical records. A hacker hacks into your clinic's databases and steals all your medical records.

(D.3) Emails and Pictures. A hacker hacks into your computer and steals your personal emails and pictures.

(D.4) Sequenced Genome. A hacker steals the file containing your sequenced genome.

IV. STUDY RESULTS

We now present the results of our study, via a quantitative and qualitative analysis.

A. Theme 1: Trust

During Experiment A, we asked participants to rank six genetic tests from the one they felt most inclined to, to the least. We report the average rankings, along with standard deviation, in Table IIIa. (Rankings are on a 1-6 scale, with 6 being the case participants felt most inclined to, 1 the least.)

(A.6), i.e., correct cancer treatment, emerged as the test to which participants felt most inclined, with average ranking 5.81 (out of 6). 13/16 participants ranked (A.6) as the top one, and all of them among the top two. Next follow (A.5), i.e., correct drug dosage, and (A.2), i.e., genetic compatibility, with, respectively, 15/16 and 13/16 of participants ranking it among the top three.

At the other extreme, (A.3), ancestry testing, emerged as the test scenario most of participants (10/16) felt the least inclined to, with an overwhelming majority (12/16) ranking it among the bottom two. However, only 2 participants mentioned that they would not want to know their ancestry, while 8 reported that they would not mind discovering it but found medical tests much more relevant. Average ranking for (A.3) was 1.75, with a relatively high standard deviation of 1.09.

(A.1), i.e., disease predisposition (doctor), and (A.4), i.e., disease predisposition (company), were ranked relatively similar, mostly in the middle of the ranking scale. The former obtained a slightly higher average, i.e., 2.63 vs 2.13, although their difference was not statistically significant.

Participants exhibited a preference toward tests they considered immediately beneficial for their health (or that of their kids), such as, (A.6), (A.5), and (A.2). In fact, the rankings for (A.2), the test with lowest average in this category, and those for (A.1), the test with highest average outside this category, are statistically significant, as per Mann-Whitney U Test ($U = 210.5$, $n_1 = n_2 = 16$, $P < 0.01$, two-tailed).

Finally, it is interesting to observe the difference between (A.1) and (A.4): recall that, in both cases, the patient discovers the genetic predisposition to certain diseases. There were 9/16 participants ranking (A.1) higher than (A.4), and 5 out of 9 justified this choice due to a lower *trust* in a commercial company. For instance, P3 mentioned that she *"would never trust a website with my genome."* The other 4 participants instead mentioned that they would prefer a doctor to explain the test results, as the effects of genetic disease predisposition were not completely clear to them. On the other hand, 7/16 participants ranked (A.4) higher than (A.1): 2 of them justified this as they do not like to go to the doctor; 2 of them preferred a broader analysis of all predispositions; and, the remaining 3 did not want their healthcare provider to know about their predisposition, owing to fears of coverage denial or higher premiums.

B. Theme 2: Control

In Experiment B, we aimed to understand and analyze participants' preference among three different options for Whole Genome Sequencing (WGS) programs. To this end, we asked participants to rank these options from the one they felt most incline to, to the least. The average rankings, along with standard deviation, are reported in Table IIIb. (Rankings are on a 1-3 scale, with 3 indicating the most favorite option, 1 the least favorite.)

The favorite option for a hypothetical WGS program was (B.3), i.e., data-only WGS (DVD), with more than two thirds

5

of the participants (12/16) ranking it at very the top (average ranking 2.68 out of 3). Whereas, half of the participants (8/16) ranked option (B.1), WGS with healthcare provider, as their second favorite. Clearly, option (B.2), WGS with personal genomics company, seems to be the least preferred (1.31 average ranking), with two thirds of the participants (11/16) ranking it at the bottom, and 16/16 ranking it among the bottom two. In fact, the difference between the rankings of (B.2) and those of (B.1) is significant as per Mann-Whitney U Test ($U = 194, n_1 = n_2 = 16, P < 0.05$, two-tailed).

We noticed that several participants, 12/16, mentioned that they *wanted to feel in control* of their genomic data. In particular, *"control"* was often used to justify the choice of ranking (B.3) as the preferred WGS option. Recall that 12/16 participants expressed preference toward (B.3), i.e., obtaining their sequenced genome on a DVD via a testing facility that would not retain data. Among these, 7 participants (6 male, 1 female) mentioned control as their *main* motivation for doing so, whereas, 5 users (all female) gave other reasons.[2]

P10 said *"I would feel in control of my data,"*, while he would not trust it with the healthcare provider, as *"they could hold it against me."* And (B.2) would be *"even worse as the company would not be subject to the same standards as a healthcare provider.* Similarly, P14 reported she *"would like to own and control my genome, or at least let my family do"* and ranked (B.1) at the bottom, mentioning a complete lack of control in giving the information to a healthcare provider. Also, P12, P15, and P16 raised control as the main motivation to prefer (B.3); nevertheless, P12 reported that he *"wouldn't mind (B.1) if they can guarantee the same level of control,"* while both P15 and P16 mentioned that it would be necessary to trust the sequencing facility not to retain the raw data.

Next, we noticed that 7/16 participants responded negatively to the idea of receiving a discount on their next health bill in exchange for letting their provider sequence their genome and give a list of disease predispositions. P6 reported that *"the discount looks fishy [...] are they going to make a deal with my genome?"* P9 said *"I don't care if I get a discount, I would prefer that they promise not to deny me coverage."* On the other hand, P5, who ranked (B.3) higher than (B.1), pointed out that she would consider changing her mind if the discount was significant (*"higher than \$1,000."*)

Additional noteworthy remarks include P4 and P8 worrying of losing the DVD with the sequenced genome and P16's fear of being served personalized ads by a personal genomics company performing the sequencing.

C. Theme 3: Discrimination

Experiment C aimed to assess perception of potential privacy and ethical issues associated with genomic tests and availability of whole genomes. Participants were asked to compare four different incidents and rank them from the one giving them the most discomfort, to the least.

---

[2]This also suggests a correlation between gender and preferring (B.3) because of increased control, with $\chi^2(1, N = 32) = 8.57$ $p < 0.01$.

The average rankings, along with standard deviation, are reflected in Table IIIc. (Rankings are on a 1-4 scale, with 4 indicating the most frightening case, and 1 the least frightening.) *Note: unlike Experiment A and Experiment B, a higher ranking in Experiment C corresponds to a negative feeling.*

(C.1), i.e., labor discrimination, and (C.2), i.e., health insurance discrimination, represented the incidents giving most discomfort to the participants, with 12/16 participants ranking either one as the top. More specifically, almost one third of the participants (10/16) ranked (C.1) and (C.2) as the two incidents giving them most discomfort, owing to related discrimination issues. Their average rankings were both above 3.00 (out of 4). (C.1) actually seems to be the most discomforting scenario as 15/16 participants placed it in the top two, compared to 11/16 for (C.2).

Then, 4/16 participants felt that (C.3), i.e., sequenced genome hacked, was the most discomforting scenario. 2 of these 4 users actually reported that (C.3) would imply (C.1) and (C.2), as a hacker could publish the genomic information or sell it to employers and healthcare providers. Whereas, the other 2 participants felt that they should be protected by the law as for (C.1) and (C.2).

(C.4), i.e., sibling donating genome to science, represented the case yielding the least discomfort, with 14/16 participants ranking it at the bottom, and 16/16 among the bottom two. The difference between rankings of (C.4) and those of (C.3), the second least discomforting case, was statistically significant, as per the Mann-Whitney U Test ($U = 238.0, n_1 = n_2 = 16, P < 0.001$, two-tailed).

Concerns related to discrimination were raised by several participants when motivating the rankings in Experiment C, but also, indirectly, when discussing their concerns with respect to Whole Genome Sequencing.

10/16 participants ranked (C.1) higher than (C.2): when asked why, 5 out of these 10 participants reported that somehow (C.2) was not too surprising. P5, P6, and P16 actually said that healthcare cost discrimination already happens today, so (C.2) would be somewhat expected. P16 also mentioned that (C.2) *"is understandable, they will be the ones to pay, and they already take into account predisposition and pre-existing conditions."*

Discomfort from (C.1) was often paired with an unfairness feeling. P7 pointed out that predisposition *"does not necessarily mean that I will get the disease,"* and P1 that predisposition *"doesn't mean that I won't be able to perform my job."*

Finally, (C.4) was the least frightening case for 14/16 participants: 9 of these said they were not concerned since the genome is donated to science, for a good cause, and 5 trusted the research lab having access to their genome, since, as pointed out by P13, *"labs are subject to ethical reviews and will not share data with anyone."*

D. Theme 4: Damage from Information Loss

The last experiment aimed to compare participants' response to incidents involving information loss. We asked them to rank four hypothetical cases from the one they considered

most frightening to the one they considered least. The average rankings, along with standard deviation, are reported in Table IIId. Rankings are on a 1-4 scale, with 4 indicating the most frightening case, and 1 the least frightening.) Again, unlike Experiment A and Experiment B, and similar to Experiment C, a higher ranking in Experiment D corresponds to a negative feeling.

(D.1) represented the most frightening case for most participants, with an average ranking of 3.50 (out of 4). Indeed, 10/16 participants ranked it as the most frightening scenario and 15/16 among the top two. Then follows (D.3), with 6/10 participants ranking it at the top. Observe that the average ranking is relatively low (2.63) with a high standard deviation (1.61); almost half of the participants (7/16) ranked it among the bottom two cases, and 5/16 at the very bottom.

(D.4) and (D.2) were ranked similar, with average ranking, resp., 2.00 and 1.88, and no one labeling them as the most frightening. We observed no statistically significant difference between the two, suggesting that participants are almost equally frightened with a hacker stealing their medical records or their sequenced genome. But, we did observe a statistically significant difference between (D.1) and (D.4), suggesting that participants are not as frightened by loss of genomic information as they are of identity theft (U = 230, $n_1 = n_2 = 16$, $P < 0.001$, two-tailed).

We observed that several participants often used the degree of damage produced by each incident involving information loss to identify the most frightening cases. Experiment D also exhibited some interesting correlations with participants' demographics. We observed that most users with incomes lower than 50K were most frightened by (D.3), i.e., their personal emails and pictures being stolen (5/6), as opposed to very few users (1/10) with higher incomes ($\chi^2(1, N = 32) = 8.60$ $p < 0.01$). Whereas, users with incomes higher than 50K were most frightened by their identity being stolen (D.1), i.e., 9/10 vs 1/6, resp., ($\chi^2(1, N = 32) = 8.60$ $p < 0.01$).

Next, we noticed that no privacy fundamentalist was most frightened by loss of personal emails and pictures (0/5), as opposed to a slight majority of pragmatists/unconcerned (6/11). Whereas, privacy fundamentalists were most frightened by identity theft (5/5) as opposed to 5/11 for non-fundamentalists. In both cases, it holds $\chi^2(1, N = 32) = 4.36$ $p < 0.05$, thus, there might be a statistically significant correlation between the Westin privacy index and being most frightened by either identity theft or loss of emails and/or pictures.

Reasons for such correlations can be explained by noticing that participants used the degree of damage as the discriminant among information loss cases. Ostensibly, high-income users have more to lose (money- and financial-wise) from identity theft – e.g., P11 mentioned *"It would cost me too much time and money."* Also, privacy fundamentalists might not be afraid of emails and pictures being stolen as they feel they take the appropriate precautions – e.g., P2 reported *"I always encrypt my sensitive emails and do not keep pictures on my laptop."*

E. Discussion

In summary, our exploratory study unveils how complex is to evaluate the perception of security, privacy, and ethical issues related to the progress in Whole Genome Sequencing (WGS), and its medical and societal consequences. We are aware that our analysis does not provide definite answers, as we interviewed a relatively small number (16) of US-based volunteers, mostly high-educated and in the 25-34 age range. Nonetheless, as suggested by [14] and [46], this seems to constitute the representative population of early adopters of personal genomic tests.

Therefore, while we expect our work to trigger further (and possibly larger) ethnographic studies in the field, we do draw some preliminary conclusions:

1) Labor and healthcare discrimination are top concerns among users, with an increase unfairness feeling associated with the former, and a lack of confidence in the protection granted by the law. Fear of discrimination was predominant in our study, even more so than privacy concerns.

2) Users consistently raised the issue of control and preferred to retain and own data with their sequenced genome. This suggests that design of genomics-related systems and applications should include mechanisms to let the users feel in control of their data.

3) Motivated by trust concerns, users prefer that doctors administered medical genetic tests, rather than specialized personal genomics companies.

4) Perception of genetic tests is strongly related to the associated perceived medical benefit. Participants were mostly inclined to genetic tests that can help fight diseases, but less to tests that can help prevent them, and neutral w.r.t. discovering their ancestry. Although not totally unexpected, this indicates the need for identifying aligned incentives and for informed consent.

5) Participants were more frightened by having their financial identity and/or personal data stolen (with a bias toward the former for high-income, privacy fundamentalist users) than medical and genomic information. However, the perception of genomic information loss varied significantly among different participants and suggested that participants were more scared of their insurance provider or their employer using genomic information against them than of a hacker having access to it.

V. RELATED WORK

The emergence of personal genomic tests and affordable Whole Genome Sequencing (WGS) motivate the need for better understanding related perception and concerns of involved users. However, much is left to be studied, since, to the best of our knowledge, there is no ethnographic study on Whole Genome Sequencing prior to our work.

Francke et al. [16] interviewed 63 customers of 23andMe.com that tested for BRCA mutations. (BRCA

gene mutations convey a high risk for breast and ovarian cancer). They analyzed customers' response among 32 mutation carriers (16 women, 16 men) and 31 non-carriers and concluded that direct access to BRCA mutation tests provided benefits to participants. None of the 25 users that had unexpectedly found out to carry the mutation reported extreme anxiety, and 4 experienced moderate, transitory anxiety, similar to the findings by Hamilton et al. [26]. Additional studies analyzing patients' response to genetic tests for disease predisposition include [45] (colon cancer), [22] (Alzheimer's disease), [4, 6] (breast cancer). Also, Andrews [2] analyzed the concept of "survivor's guilt" experienced by individuals learning not to carry a harmful mutation, while a family member does.

Brothers et al. [5] found out that, when asked about an opt-out consent process, over 90 percent of participants agreed or strongly agreed that "DNA biobank research is fine as long as people can choose not to have their DNA included." Some studies highlighted how privacy concerns are often an important obstacle to participation in large cohort studies [53]. Although 60 percent of people surveyed said they would participate in a study that involved storing data in biorepositories, 91 percent of those potential research participants would be concerned about privacy [29]. Another study showed that, while participants trusted clinicians and researchers, they were concerned that results of genetic tests could end up in the wrong hands and be used against them [19].

Lapham et al. [31] were among the first, in 1996, to analyze consumers' perspectives as to genetic discrimination and reported that people cited fear of losing insurance as a major reason to avoid genetic testing. However, discrimination by insurance companies was not a widespread reality in the 90s, as few of these cases had been filed and even fewer had been won [25]. Many states have passed laws to protect medical (and also genetic) information, such as the Health Insurance Portability and Accountability Act (HIPAA), which provides a general framework for sharing and protecting Protected Health Information. In the U.S., there also exists legislation specific against genetic discrimination – the Genetic Information Nondiscrimination Act (GINA) – which prohibits discrimination on the basis of genetic information with respect to health insurance and employment [49]. However, neither GINA or HIPAA placed any limits on health insurance rate setting.

Also, prior work analyzed the issue of labor discrimination in relationship to disease predisposition, mostly from the legal standpoint. We refer to the work by Guttmacher et al. [23] for details, along with a review of well-known rulings.

Ruiz et al. [43] compared the attitude of 279 patients from the United States and Spain who had volunteered to donate a sample for genomic research, and showed that 48% of participants would like to be informed about all individual results from future genomic studies using their donated tissue, especially those from the U.S. (71.4%) and those believing that genetic information poses special risks (69.7%). Trinidad et al. [48] explored the attitude of research participants and

possible future participants regarding Genome-Wide Association Studies (GWAS): they found out that participants expressed a variety of opinions about the acceptability of wide sharing of genetic and phenotypic information for research purposes through large, publicly accessible data repositories. Most believed that making de-identified study data available to the research community was a social good to be pursued. Privacy and confidentiality concerns were common, although not necessarily precluding participation. Also, many participants voiced reservations about sharing data with for-profit organizations.

A recent report from the Presidential Commission for the Study of Bioethical Issues [40] analyzed advances of Whole Genome Sequencing (WGS), and focused on potential privacy and ethical threats. The report listed 12 recommendations and provided a high-level effort to identify and promote policies and practices that ensure scientific research, healthcare delivery, and technological innovation are conducted in a socially and ethically responsible manner.

Thus, we conclude that, while prior work has looked at responses to learning genetic test results as well as discrimination concerns associated with genomics, no prior study has focused on the human factor in Whole Genome Sequencing.

VI. Conclusion

This paper presented the results of a series of semi-structured interviews involving 16 participants, that aimed to assess the perception of genetic tests, the attitude toward different Whole Genome Sequencing (WGS) programs, as well as their concerns with associated privacy and ethical threats. The results of our interviews were analyzed both quantitatively and qualitatively and showed that interviewees exhibited common fear of discrimination and demanded to retain strict control over their genetic information. Our exploratory analysis suggests that the issues of control, trust, and discrimination are crucial aspects that the community should take into account, not only from a policy standpoint, but also for user-centered designs of genomics-related applications that involve end-users (such as, ancestry and disease predisposition testing).

Our study highlights the need for more ethnographic studies in the field as well as the importance of informing the public with respect to privacy threats associated with genomic information disclosure and legal rights addressing discrimination issues.

As part of future work, we plan to run follow-up user studies, relying on larger populations, that can further clarify some of the issues discussed in this paper. We are also working on strategies to raise awareness of privacy, ethical, and legal issues associated with WGS, as well as on user-centered designs of privacy-respecting genomic-testing platforms.

REFERENCES

[1] A. Abbott. Special section on human genetics: With your genes? Take one of these, three times a day. *Nature*, 425(6960), 2003.

[2] L. B. Andrews. *Future perfect: confronting decisions about genetics*. Columbia University Press, 2001.

[3] E. Ayday, E. De Cristofaro, J.-P. Hubaux, and G. Tsudik. The chills and thrills of whole genome sequencing. *IEEE Computer*, 2013.

[4] J. L. Bottorff et al. Womens interest in genetic testing for breast cancer risk the influence of sociodemographics and knowledge. *Cancer Epidemiology Biomarkers & Prevention*, 11(1), 2002.

[5] K. B. Brothers, D. R. Morrison, and E. W. Clayton. Two large-scale surveys on community attitudes toward an opt-out biobank. *American Journal of Medical Genetics*, 155(12), 2011.

[6] M. Bruno et al. Awareness of breast cancer genetics and interest in predictive genetic testing: a survey of a southern Italian population. *Annals of Oncology*, 15, 2004.

[7] A. Burke. Foundation Medicine: Personalizing Cancer Drugs. MIT Technology Review, 2012. http://techre.vu/Xic52g.

[8] T. Canli. The emergence of genomic psychology. *Nature*, 8, 2007.

[9] G. M. Church. The Personal Genome Project. *Molecular Systems Biology*, 1(1), 2005.

[10] S. Cichon, M. Nöthen, M. Rietschel, and P. Propping. Pharmacogenetics of schizophrenia. *American Journal of Medical Genetics*, 97(1), 2001.

[11] F. Collins and V. McKusick. Implications of the Human Genome Project for medical science. *Journal of the American Medical Association*, 285(5), 2001.

[12] K. Darcè. Illumina drops sequencing price to $4,000. UT San Diego. http://www.utsandiego.com/news/2011/may/09/illumina-drops-human-sequencing-price-4000/.

[13] R. Drmanac et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*, 327(5961), 2010.

[14] F. M. Facio, S. Brooks, J. Loewenstein, S. Green, L. G. Biesecker, and B. B. Biesecker. Motivators for participation in a whole-genome sequencing study: implications for translational genomics research. *European Journal of Human Genetics*, 19(12), 2011.

[15] J. Fowler, J. Settle, and N. Christakis. Correlated genotypes in friendship networks. *Proceedings of the National Academy of Sciences*, 108(5), 2011.

[16] U. Francke et al. Dealing with the unexpected: consumer responses to direct-access BRCA mutation testing. *PeerJ*, 1, 2013.

[17] M. Fumagalli et al. Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *Experimental Medicine*, 206(6), 2009.

[18] L. N. Geller et al. Individual, family, and societal dimensions of genetic discrimination: a case study analysis. *Science and Engineering Ethics*, 2(1), 1996.

[19] Genetics and Public Policy Center. U.S. Public Opinion on Uses of Genetic Information and Genetic Discrimination. http://www.dnapolicy.org/resources/GINAPublic_Opinion_Genetic_Information_Discrimination.pdf.

[20] G. Ginsburg and H. Willard. Genomic and personalized medicine: foundations and applications. *Translational Research*, 154(6), 2009.

[21] J. Gitschier. Inferential genotyping of Y chromosomes in Latter-Day Saints founders and comparison to Utah samples in the HapMap project. *American Journal of Human Genetics*, 84, 2009.

[22] R. C. Green et al. Disclosure of APOE genotype for risk of Alzheimer's disease. *New England Journal of Medicine*, 361(3), 2009.

[23] A. E. Guttmacher, F. S. Collins, and E. W. Clayton. Ethical, legal, and social implications of genomic medicine. *New England Journal of Medicine*, 349(6), 2003.

[24] M. Gymrek, A. L. McGuire, D. Golan, E. Halperin, and Y. Erlich. Identifying personal genomes by surname inference. *Science*, 339(6117), 2013.

[25] M. A. Hall and S. S. Rich. Patients' fear of genetic discrimination by health insurers: the impact of legal protections. *Genetics in Medicine*, 2(4), 2000.

[26] J. G. Hamilton, M. Lobel, and A. Moyer. Emotional distress following genetic testing for hereditary breast and ovarian cancer: a meta-analytic review. *Health psychology*, 28(4), 2009.

[27] N. Homer et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genetics*, 4(8), 2008.

[28] M. Humbert, E. Ayday, J.-P. Hubaux, and A. Telenti. Addressing the Concerns of the Lacks Family: Quantification of Kin Genomic Privacy. In *CCS*, 2013.

[29] D. J. Kaufman, J. Murphy-Bollinger, J. Scott, and K. L. Hudson. Public opinion about the importance of privacy in biobank research. *American Journal of Human Genetics*, 85(5), 2009.

[30] P. Kumaraguru and L. F. Cranor. Privacy indexes: A survey of Westin's studies. http://reports-archive.adm.cs.cmu.edu/anon/isri2005/CMU-ISRI-05-138.pdf, 2005.

[31] E. V. Lapham, C. Kozma, and J. O. Weiss. Genetic discrimination: perspectives of consumers. *Science*, 274(5287), 1996.

[32] S. Levy et al. The diploid genome sequence of an individual human. *PLoS biology*, 5(10), 2007.

[33] B. Malin. Re-identification of familial database records. In *AMIA*, 2006.

[34] B. Malin and L. Sweeney. Re-identification of DNA through an automated linkage process. In *AMIA*, 2001.

[35] B. Malin and L. Sweeney. How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protec-

tion systems. *Journal of Biomedical Informatics*, 37(3), 2004.

[36] V. McKusick and S. Antonarakis. *Mendelian inheritance in man: a catalog of human genes and genetic disorders*. John Hopkins University Press, 1994.

[37] National Center for Biotechnology Information (US). Single Nucleotide Polymorphism Database. http://www.ncbi.nlm.nih.gov/projects/SNP/.

[38] J. Oldenburg, M. Watzka, S. Rost, and C. Müller. VKORC1: Molecular target of coumarins. *Journal of Thrombosis and Haemostasis*, 5(s1), 2007.

[39] A. Prat and J. Baselga. The role of hormonal therapy in the management of hormonal-receptor-positive breast cancer with co-expression of her2. *Nature Clinical Practice Oncology*, 5(9), 2008.

[40] Presidential Commission for the Study of Bioethical Issues. Privacy and Progress in Whole Genome Sequencing. http://www.bioethics.gov/cms/sites/default/files/PrivacyProgress508.pdf, 2012.

[41] M. A. Quail, I. Kozarewa, F. Smith, A. Scally, P. J. Stephens, R. Durbin, H. Swerdlow, and D. J. Turner. A large genome center's improvements to the Illumina sequencing system. *Nature methods*, 5(12), 2008.

[42] P. Rincon. Science enters $1,000 genome era. http://www.bbc.co.uk/news/science-environment-25751958, 2014.

[43] M. Ruiz-Canela, J. I. Valle-Mansilla, and D. P. Sulmasy. What research participants want to know about genetic research results: The impact of "genetic exceptionalism". *Journal of Empirical Research on Human Research Ethics*, 6(3), 2011.

[44] H. Shen. California considers DNA privacy law. *Nature*, 2012. http://www.nature.com/news/california-considers-dna-privacy-law-1.10677.

[45] K. R. Smith and R. T. Croyle. Attitudes toward genetic testing for colon cancer risk. *American Journal of Public Health*, 85(10), 1995.

[46] R. Stein. NPR News: Will Low-Cost Genome Sequencing Open 'Pandora's Box'? http://n.pr/O1ThSh, 2012.

[47] P. Stenson et al. The human gene mutation database: 2008 update. *Genome Medicine*, 1(1), 2009.

[48] S. B. Trinidad et al. Genomic research and wide data sharing: views of prospective participants. *Genetics in Medicine*, 12(8), 2010.

[49] M. Wadman. Genetics bill cruises through senate. *Nature*, 453, 2008.

[50] P. Walker. DNA of 100,000 people to be mapped for NHS. http://www.theguardian.com/science/2012/dec/10/1000000-peoples-dna-mapped, 2012.

[51] R. Wang et al. Learning your identity and disease from research papers: information leaks in Genome Wide Association Study. In *CCS*, 2009.

[52] D. Wheeler et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature*, 452(7189), 2008.

[53] S. Williams. The genetic town hall: Public opinion about research on genes, environment, and health. http://www.dnapolicy.org/images/reportpdfs/2009PCPTownHalls.pdf, 2009.

[54] S. Wood. RAPID GENE: Point-of-care genetic test singles out clopidogrel nonresponders. http://www.theheart.org/article/1379471.do, 2012.