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

Emiliano De Cristofaro
University College London
http://emilianodc.com
Genomics 101

Genomes…

Information to build/maintain an organism’s living example
At least one copy of the genome is in almost all cells
Encoded in DNA (or RNA for viruses)
  DNA: a double stranded polymer of nucleotides (A, C, G, T)
In humans, 3.2B nucleotides (in 23 chromosome pairs)

Whole Genome Sequencing (WGS)…

Determining the complete DNA sequence in a genome
WGS Progress

Some dates

1970s: DNA sequencing starts
1990: The “Human Genome Project” starts
2003: First human genome fully sequenced
2005: Personal Genome Project (PGP) starts
2012: UK announces sequencing of 100K genomes

Some numbers

$3B: Human Genome Project (2003)
$250K: Illumina (2008)
$5K: Complete Genomics (2009), Illumina (2011)
$1K: Illumina (2014)
The Good News

Affordable WGS facilitates the creation of large datasets for research purposes
   Crucial for hypothesis-driven research, e.g., GWAS

Low-cost WGS will bring genomics to the masses
   Large number of individuals will have the means to have their (fully) genome sequenced, and possibly store/retain it

Personalized medicine
   Diagnosis/treatment tailored to patient’s genetic makeup

In general, genomic tests can be done “in silico”, using specialized computation algorithms
The Bad News

The genome is a unique **identifier**

Once leaked, you cannot “revoke” it

Anonymization / de-identification useless

Gymrek et al., Identifying personal genomes by surname inference, Science, 2013

Genomic information is extremely **sensitive**

Contains ethnic heritage, predisposition to diseases and conditions (even mental), many phenotypical traits

Raises the risk of genetic discrimination – “genism”
It gets worse…

Leaking one’s genome ≈ leaking relatives’ genome

~99.9% of genomes of closely related humans identical

Basis of Gymrek’s attack

The case of Henrietta Lacks

See Humbert et al. (ACM CCS, 2013)

Sensitivity of human genomes is (almost) perpetual

Even if encrypted, can’t guarantee security of the encryption algorithm past 30-50 years

More details:

Ayday et al., Chills and Thrills of WGS, IEEE Computer
The Greater Good vs Privacy?

Advances in genomics often promoted as dependent on volunteers and data sharing.
Sharing is actually a requirement for most grants.

Sharing is an important asset for research.
Chatterjee et al. (Nature, 2013) project that several million samples may be needed for robust GWAS.

But privacy and discrimination fears may drive potential participants away?
McGuire et al. (Genetics in Medicine, 2011) finds correlation between opting out and privacy fears.
Open Questions

What do we understand about users’ perceptions and attitudes with respect to Whole Genome Sequencing?

Do privacy perceptions/concerns experienced by individuals correspond to what the scientific community would expect?

How to identify effective mechanisms to communicate risks and benefits? How to reconcile the greater good/privacy tension?

(Little understanding from prior work in context of WGS)
Methodology 1/3

Recruited 16 study volunteers in SF Bay Area

Sex: female (8), male (8)
Age: 18-24 (2), 25-34 (7), 35-44 (3), 45-54 (1), 55-64 (1), 65+(2)
Degree: College (4), Master (8), PhD (4)
Income: <$50K (3), $50K-$75K (3), >$75K (10)
Westin: Unconcerned (4), Pragmatist (7), Fundamentalist (5)

Participants skewed toward high-income/high-edu

Representative population for early WGS adopter, as per related work, e.g., Facio et al. (Nature, 2011), 2012 NPR study, …
Methodology 2/3

Participants guided through a set of slides depicting a few hypothetical scenarios

Asked to comment on and rank these scenarios

Four experiments

Exp A: Assessing perception of today’s genetic tests
Exp B: Comparing attitudes toward different WGS program
Exp C: Assessing perception of privacy/ethical issues with WGS
Exp D: Comparing the response to medical/genomic/personal information loss
Exp A – Trust

<table>
<thead>
<tr>
<th>Genetic Tests: More to less inclined</th>
<th>Avg</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A.6) Determine Cancer Treatment</td>
<td>5.81</td>
<td>0.39</td>
</tr>
<tr>
<td>(A.5) Determine Drug Dosage</td>
<td>4.63</td>
<td>0.70</td>
</tr>
<tr>
<td>(A.2) Genetic Compatibility</td>
<td>4.06</td>
<td>1.25</td>
</tr>
<tr>
<td>(A.1) Disease Predis. (Doctor)</td>
<td>2.63</td>
<td>0.99</td>
</tr>
<tr>
<td>(A.4) Disease Predis. (Company)</td>
<td>2.13</td>
<td>0.70</td>
</tr>
<tr>
<td>(A.3) Ancestry Testing</td>
<td>1.75</td>
<td>1.09</td>
</tr>
</tbody>
</table>

(A.6), (A.5), (A.2) statistically significantly higher than (A.1)
Mann-Whitney U Test (U = 210:5, n1 = n2 = 16, P < 0.01, two-tailed)

(A.1) and (A.4) close
(A.4) was ranked among the bottom because of mistrust in company
Exp B – Control

<table>
<thead>
<tr>
<th>WGS Programs: More to less inclined</th>
<th>Avg</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B.3) Data-only (DVD)</td>
<td>2.68</td>
<td>0.58</td>
</tr>
<tr>
<td>(B.1) Healthcare Provider</td>
<td>2.00</td>
<td>0.71</td>
</tr>
<tr>
<td>(B.2) Direct-to-Consumer (DTC) Company</td>
<td>1.31</td>
<td>0.46</td>
</tr>
</tbody>
</table>

(B.3) the “favorite” (12/16 ranking at the very top)

(B.2) the least “favorite” (11/16 ranking at the very bottom)

  Diff b/w (B.1) and (B.2) stat. significant (U = 194; P < 0.05, two-tailed)

12/16 participants mention they wanted to “feel in control”

  Mistrust against health provider: “use against me”, company “even worse”
  When prospecting a $1,000 discount for (B.1), even more suspicious
Exp C – Discrimination

<table>
<thead>
<tr>
<th>Incidents: More to less discomfort</th>
<th>Avg</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C.1) Labor Discrimination</td>
<td>3.31</td>
<td>0.58</td>
</tr>
<tr>
<td>(C.2) Health Insurance Discrimination</td>
<td>3.00</td>
<td>0.94</td>
</tr>
<tr>
<td>(C.3) Sequenced Genome Leaked</td>
<td>2.56</td>
<td>0.93</td>
</tr>
<tr>
<td>(C.4) Sibling Donating Genome to Science</td>
<td>1.13</td>
<td>0.33</td>
</tr>
</tbody>
</table>

(C.4) least discomforting (14/16 at the very bottom), (C.1) most discomforting (15/16 participants ranking in top two)

Some participants not surprised by (C.2)
Some participants find (C.1) extremely unjust because of environmental factors
Exp D – Harm

<table>
<thead>
<tr>
<th>Information loss: More to less frightened</th>
<th>Avg</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D.1) Identity Theft</td>
<td>3.50</td>
<td>0.63</td>
</tr>
<tr>
<td>(D.3) Emails and Pictures Leaked</td>
<td>2.63</td>
<td>1.61</td>
</tr>
<tr>
<td>(D.4) Sequenced Genome Leaked</td>
<td>2.00</td>
<td>0.63</td>
</tr>
<tr>
<td>(D.2) Medical Records Leaked</td>
<td>1.88</td>
<td>0.48</td>
</tr>
</tbody>
</table>

(D.1) and (D.4) statistically significantly different

Correlation b/w lower income and (D.3), higher income and (D.1)

\[ \chi^2 (1; N = 32) = 8.60 \ p < 0.01 \] (both cases)

Correlation b/w fundamentalists and (D.1)

\[ \chi^2 (1; N = 32) = 4.36 \ p < 0.05 \]
Summary of (Preliminary) Results

1) Users’ perception depends on related perceived medical benefit

2) Participants prefer that doctors administer/explain genetic tests, mistrust of DTC companies

3) Raising issue of control – prefer to retain & own genomics data to minimize fear of potential discrimination

4) Labor/healthcare discrimination top concerns, also due to mistrust/unawareness of legislation

5) Genetic information disclosure not well understood
Related Work

Response to learning results of genetic tests
  E.g. predisposition to cancer

Control
  Participants want the opt-out option

Genetic discrimination
  Well-known fear, not well contextualized

Informed consent issues
  Very important area of research, very hard to get the consent right
Discussion

How to use preliminary study for a larger, more focused study?

What to focus on?
How to select/recruit participants?
How to interview users on an emerging technology?

How to drive user-centered design of personal (computational) genomics?

How to raise awareness/communicate concerns and regulation from/to users?
Acknowledgments

Work done, in part, while at PARC and with PARC’s IRB approval

Thanks to Honglu Du, Darya Mohtashemi, Julien Freudiger, Greg Norcie for the helpful comments

And obviously to the study volunteers

Thanks!