

Communicating Personal Genomic Information to Non-experts: A New Frontier for Human-Computer Interaction

Orit Shaer

Wellesley College, Wellesley, MA, USA
oshaer@wellesley.edu

Oded Nov

New York University, USA
onov@nyu.edu

Lauren Westendorf

Wellesley College, Wellesley, MA, USA
lwestend@wellesley.edu

Madeleine Ball

Open Humans Foundation, Boston, MA, USA
madeleine@personalgenomes.org

now

the essence of knowledge

Boston — Delft

Foundations and Trends[®] in Human-Computer Interaction

Published, sold and distributed by:

now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:

now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

O. Shaer, O. Nov, L. Wstendorf, and M. Ball. *Communicating Personal Genomic Information to Non-experts: A New Frontier for Human-Computer Interaction*. Foundations and Trends[®] in Human-Computer Interaction, vol. 11, no. 1, pp. 1–62, 2017.

This Foundations and Trends[®] issue was typeset in L^AT_EX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-68083-254-9

© 2017 O. Shaer, O. Nov, L. Wstendorf, and M. Ball

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

**Foundations and Trends[®] in
Human-Computer Interaction**
Volume 11, Issue 1, 2017
Editorial Board

Editor-in-Chief

Desney S. Tan
Microsoft Research
United States

Editors

Ben Bederson
University of Maryland

Sheelagh Carpendale
University of Calgary

Andy Cockburn
University of Canterbury

Jon Froehlich
University of Maryland

Juan Pablo Hourcade
University of Iowa

Karrie Karahalios
*University of Illinois
at Urbana-Champaign*

Youn-Kyung Lim
*Korea Advanced Institute of Science and
Technology*

Nuria Oliver
Telefonica

Orit Shaer
Wellesley College

Kentaro Toyama
University of Michigan

Editorial Scope

Topics

Foundations and Trends[®] in Human-Computer Interaction publishes surveys and tutorials on the foundations of human-computer interaction. The scope is broad. The list of topics below is meant to illustrate some of the coverage, and is not intended to be an exhaustive list.

- History of the research community
- Design and evaluation
- Theory
- Technology
- Computer supported cooperative work
- Interdisciplinary influence
- Advanced topics and trends

Information for Librarians

Foundations and Trends[®] in Human-Computer Interaction, 2017, Volume 11, 4 issues. ISSN paper version 1551-3955. ISSN online version 1551-3963. Also available as a combined paper and online subscription.

Foundations and Trends® in Human-Computer
Interaction

Vol. 11, No. 1 (2017) 1–62

© 2017 O. Shaer, O. Nov, L. Westendorf, and M. Ball

DOI: 10.1561/1100000067



Communicating Personal Genomic Information to Non-experts: A New Frontier for Human-Computer Interaction

Orit Shaer

Wellesley College, Wellesley, MA, USA
oshaer@wellesley.edu

Oded Nov

New York University, USA
onov@nyu.edu

Lauren Westendorf

Wellesley College, Wellesley, MA, USA
lwestend@wellesley.edu

Madeleine Ball

Open Humans Foundation, Boston, MA, USA
madeleine@personalgenomes.org

Contents

1	Introduction	2
2	Personal Genetics Tutorial	6
2.1	DNA sequencing	9
3	Background	13
3.1	Historical context and new technologies	13
3.2	Online interaction with personal genomics	14
3.3	Omic technology	15
3.4	Genomic data privacy and access	17
3.5	Public genomes and follow-up research	19
3.6	Characteristics of personal genomic data	20
3.7	Uncertainty in personal genomics	20
4	Related Work	22
4.1	Personal informatics	22
4.2	Communicating uncertainty	23
4.3	Interaction with biological data sets	25
4.4	DIY genome analysis	25
5	Emerging Framework for HCI for Personal Genomics	28
5.1	Understanding users	29

5.2	Informing users	34
5.3	Empowering users	41
6	Broader Implications	48
6.1	For HCI researchers	48
6.2	For practitioners	49
6.3	For society and policymakers	49
	Acknowledgements	51
	References	52

Abstract

Recent advances in genetic testing and Internet technologies have led to a dramatic increase in the access non-experts have to their own personal genomic data. Such data are complex and sensitive, involve multiple dimensions of uncertainty, and can have substantial implications on individuals' behavior, choices, and well-being. Personal genomic data are also unique because unlike other personal data, which might change frequently, genomic data are largely stable during a person's lifetime; it is their interpretation and implications that change over time as new medical research exposes relationships between genes and health.

Future progress in genetic research and technologies is likely to further increase the availability of interactive personal genomic information to non-experts. This trend raises technological, ethical, and regulatory concerns related to how people make sense of, engage with, and rely on their personal genomic data. Such concerns are not only of paramount importance for health professionals and policymakers, but are also a pressing issue for human-computer interaction (HCI) research. HCI tools, methods, and practices can help make genomic information more accessible and understandable to non-experts. We argue that the complexity, importance, and personal relevance of this type of information makes understanding, informing, and empowering non-experts' interaction with personal genomics a key challenge that lies ahead for the HCI community.

In this article, we explore the roles HCI can play in helping non-experts contribute, understand, engage with, and share their personal genomic information. This article is also a call-to-action for those of us interested in the intersection of personal informatics and HCI, and, more broadly, in facilitating non-expert interaction with large amounts of complex, personal, and uncertain information.

1

Introduction

“A firsthand familiarity with the code of life is bound to confront us with the emotional, moral and political baggage associated with the idea of our essential nature. People have long been familiar with tests for heritable diseases, and the use of genetics to trace ancestry — the new “Roots” — is becoming familiar as well. But we are only beginning to recognize that our genome also contains information about our temperaments and abilities. Affordable genotyping may offer new kinds of answers to the question “Who am I?” — to ruminations about our ancestry, our vulnerabilities, our character and our choices in life.” [Pinker, 2009]

Recent years have seen a dramatic growth in the availability of personal genomic data to non-experts, often online and in interactive forms [Dudley and Karczewski, 2013]. The field of personal genomics is rapidly growing, as the cost of sequencing a human genome has fallen from approximately \$100 million in 2001 to \$1,000 in 2016, a rate much faster than Moore’s Law [Church, 2005] (see Figure 1.1). The Precision Medicine initiative and similar health research projects increasingly highlight the potential for genetic data to become a

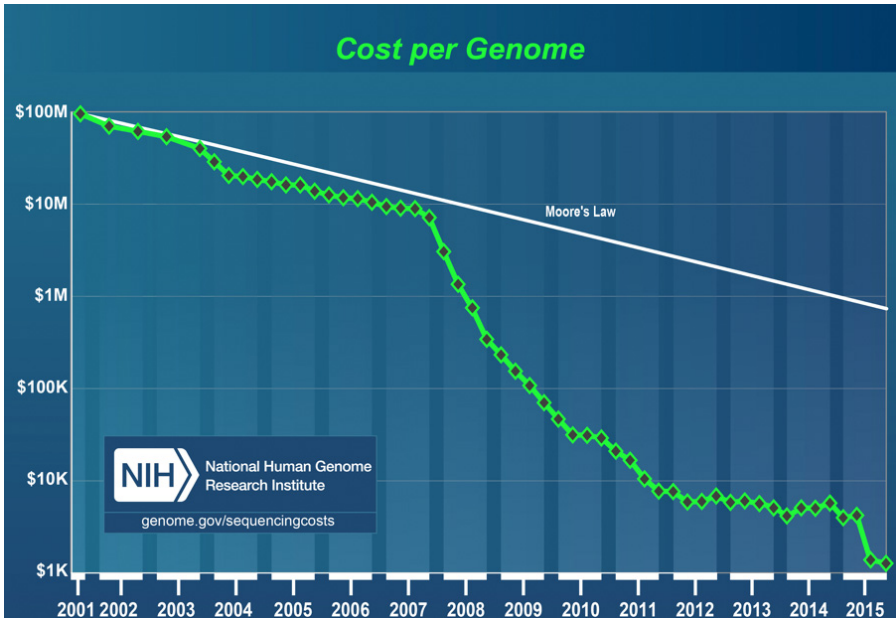


Figure 1.1: A graph demonstrating the precipitous decline in genome sequencing cost since the completion of the Human Genome Project in 2001. Whole genome sequencing is now around \$1,000 per genome, five orders of magnitude lower than the \$100 million cost of the Human Genome Project (from the National Human Genome Research Institute).

standard component of health records [NIH]. In the meantime, direct-to-consumer genetic testing (DTCGT) companies have made genomic information available to millions of individuals without the involvement of a healthcare provider through online platforms. These individuals are thereby confronted with unprecedented amount of potentially sensitive information about their genetic data, which influences their decisions, emotional states, and wellbeing [Davies, 2010].

The use of Web-based interactive technologies to deliver personal genomic information raises questions about how non-experts make sense of, engage with, and rely on their personal genomic data. Future progress in genetic research and technologies is likely to further increase the availability of interactive personal genomic information

to non-expert users. This trend raises technological, ethical, and regulatory concerns. For example, in 2013, the U.S. Food and Drug Administration (FDA) ordered 23andMe, a direct-to-consumer genetic testing company, to stop providing risk assessment reports, stating that “serious concerns are raised if test results are not adequately understood” [Pollack, 2013]. Providing consumer reports partially resumed in 2015 when 23andMe received FDA approval for a vetted set of gene variant interpretations [Pollack, 2015]. While such concerns are imperative for policymakers, it is also vital to consider these issues from the perspective of human–computer interaction (HCI) research.

Specifically, the highly personal and dynamic nature of individually-relevant personal genomic information, which is constantly updated based on new research results, raises important HCI questions, including: what are the functional requirements for supporting meaningful engagement of non-experts’ with personal genomic information? How can we design effective interaction with personal genomic information? How can we evaluate the effectiveness of techniques for interaction with personal genomic information? Can user interface design interventions impact users’ willingness to share their personal genomic data?

An additional important aspect of communicating personal genomics to non-experts is the understanding of the vast commonalities and genealogical similarities among people from different racial and ethnic backgrounds. Future HCI work on personal genomics has a critical role in making these commonalities apparent to users, just as much as the differences.

Because the technology that enables non-experts to interact with personal genomic information is new, there is little HCI research on it. Given the tremendous growth in the scale and scope of personal genomic information available to non-experts, personal genomics represents a new and fast-growing frontier where HCI research can make a significant difference in people’s lives.

This article explores the roles HCI can play in helping non-experts to understand and engage with personal genomic information. It is also a call to action for those of us interested in the intersection of personal informatics and HCI, and, more broadly, the facilitation of interaction

of non-experts with large amounts of complex and uncertain personal information.

We begin with a brief tutorial of personal genetics (Section 2), followed by a description of the historical context and new technologies of personal genomics, and a discussion of the characteristics of personal genomic data (Section 3). We then survey related work from human–computer interaction and personal informatics (Section 4). Section 5 presents a framework for HCI research focusing on understanding, informing, and empowering non-expert users of personal genomics. Finally, in Section 6 we discuss broader implications for HCI researchers, for personal genomics practitioners, for policy makers, and for society.

2

Personal Genetics Tutorial

The term personal genetics describes the information which an individual has inherited from their biological parents. An individual's physical traits — features of their biology, body, and health — are partly inherited and often are influenced by environmental and behavioral factors. At its most fundamental level, this inheritance is discrete in nature: units of inheritance are called *genes*. Some traits, like rare diseases and ABO blood types, are associated with a single gene. Many other traits (e.g. risk for diabetes, or height) are the products of many genes, and are also influenced by environmental factors like diet.

The biological basis of inherited traits is DNA — *deoxyribonucleic acid*, a linear chain of *nucleic acids* found in all cells. The sequence of these four nucleic acids, a.k.a. *bases* — adenine, cytosine, guanine, and thymine (A, C, G, and T) — is molecular information that drives almost all features of life. DNA is inherited, copied, and passed from parent to child. The collective set of DNA molecules for an individual, inherited from their mother and father, is their *genome*. Each gene is a section of a DNA molecule that contains information for creating a specific type of protein, each of which affects a specific aspect of cell biology.

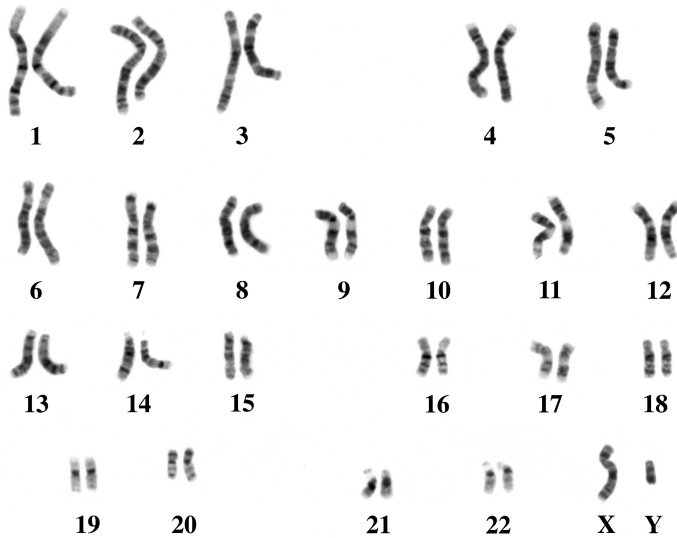


Figure 2.1: A “karyotype” of a human genetic male shows the genome’s 23 pairs of chromosomes. With the exception of the sex chromosomes in genetic males (X and Y), chromosomes exist as matched pairs: one copy is inherited from each parent. (Public domain, from NHGRI via Wikipedia.)

Proteins are the primary functioning units of life and play a variety of roles: they can be structural elements, or they may drive chemical reactions that would not otherwise occur.

Each individual human genome consists of two slightly different copies of 3 billion nucleic acids, organized into 23 pairs of chromosomes (Figure 2.1). As demonstrated in Figure 2.2, genes are distributed along chromosomes and are often interleaved with nonfunctional DNA. The functional fragments of gene DNA, accounting for 1% of the genome, are called *exons* (because this sequence is “exported” and used elsewhere in the cell).

Genetic variants account for inherited differences between individuals. However, because most of the genome is nonfunctional, most genetic variants have no biological effect. Those that do have an effect typically occur in exon regions: and they have a biological effect because they change the protein the gene produces.

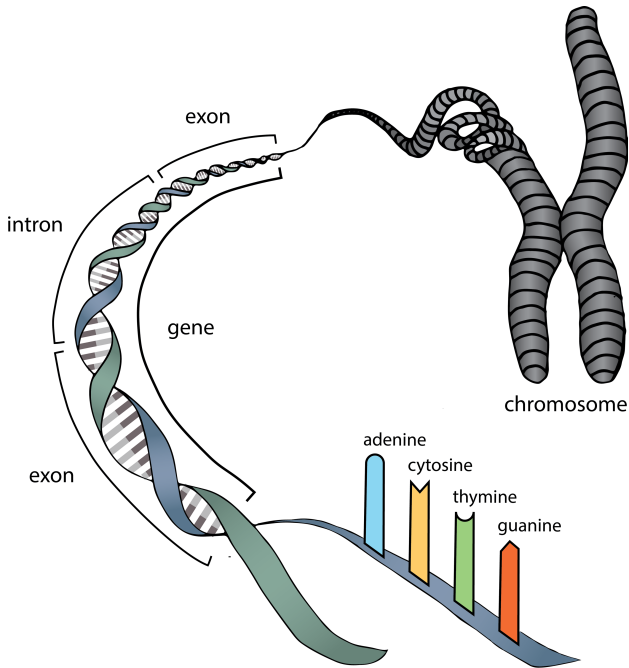


Figure 2.2: Each chromosome consists of a single, densely packed DNA strand. Genes occur as segments within the DNA strand, and genetic data is encoded within the sequence of four possible bases at each position in the DNA chain: Adenine, Cytosine, Guanine, and Thymine (A, C, G and T). Genes are often interleaved with nonfunctional DNA (intron regions). Genetic variants which change a gene's behavior typically affect exon regions — the DNA sequences used to produce a gene's protein product.

Each genome exists largely in duplicate: for most genes and DNA segments, one copy is inherited from each biological parent. When an individual carries a genetic variant, it could be present in both genes (*homozygous*, referring to the sameness of each copy) or it could be present in only one (*heterozygous*). The exception to this is the sex chromosomes: males have only one copy of each “X” and “Y” chromosome, and their variants on these chromosomes are called *hemizygous*.

The biological consequences of which involve genetic variants in a single gene can be affected by whether both copies of a gene carry

the same variant. Genetic variants that only have an effect when both copies of a gene are affected are called “recessive” (these typically have little to no effect when heterozygous, resulting in an individual having *carrier* status). While recessive variants have no immediate impact on individual health, carrier status for genetic diseases can influence reproductive decisions. In contrast, genetic variants that always have an effect are called *dominant*. Not all variants are clearly recessive or dominant: for example, a variant may have a weak effect when heterozygous, and a stronger effect when homozygous. Figure 2.3 explains the effect of genetic variants within a single gene, also referred to as monogenic effects.

2.1 DNA sequencing

There are various technologies for “reading” DNA molecules. Since information is contained within the molecular sequence of the DNA’s nucleic acids (i.e. bases), these technologies are commonly referred to as DNA sequencing. Human genome sequences are mostly identical to each other: sequence differences between two individuals account for just 0.1% of all nucleic acids. These few genetic variants account for many differences between two humans.

Although sequencing the human genome cost \$3.6 billion when it was completed in 2003 [NHGRI, 2016], improvements to DNA sequencing have since outpaced Moore’s Law: whole genome sequences can now be purchased for under \$1000 (Figure 1.1) [Keshavan, 2016, Regalado, 2016]. Two major technologies have led to vastly increased efficiency in producing genetic data: *high-throughput DNA sequencing*, and *microarray-based genotyping*. The first technology, high-throughput DNA sequencing, is comprehensive — it enables *whole genome sequencing* (see Figure 2.4) and *exome sequencing* (a lower-cost option that targets gene exon regions for around \$500 [Molteni, 2016]). The second technology, microarray-based genotyping, is less comprehensive but provides a lower-cost option: for \$100–\$200 it can be used to test a million specific locations that are known to vary between individuals [23andMe, AncestryDNA]. Because genotyping is the lowest cost

Monogenic Effects

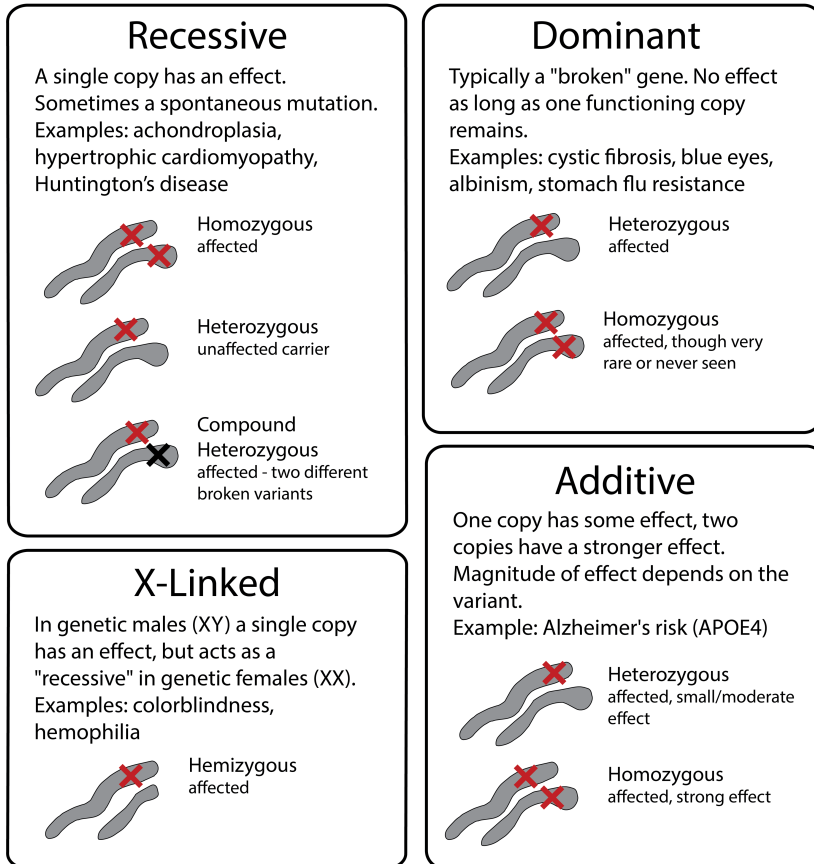


Figure 2.3: Monogenic effects involve variants in a single gene. Whether variants in a gene have an effect on traits or health depends on how many copies of the gene have variants, and on the variant's inheritance pattern: Dominant, Recessive, X-linked, or Additive.

option, it is currently used by leading direct-to-consumer genetic testing companies like [AncestryDNA](#) and [23andMe](#).

Personal genetic information gained through the process of genotyping or high-throughput DNA sequencing can be used for two seemingly disparate purposes: ancestry and biological traits. Ancestry information

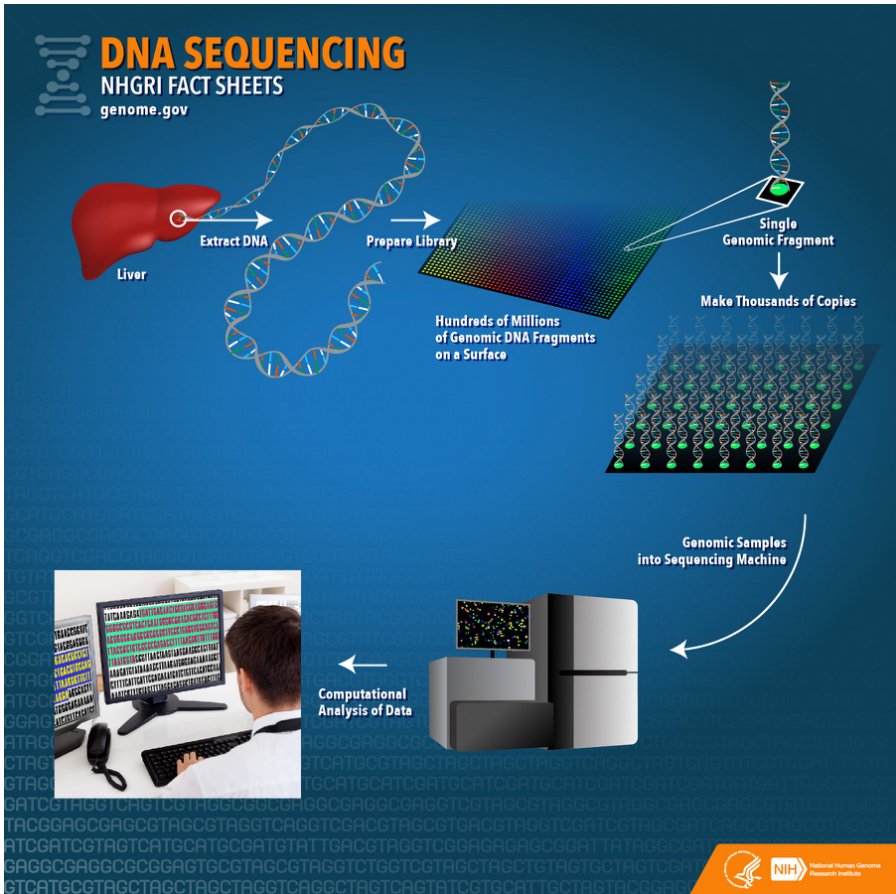


Figure 2.4: Current technology for high-throughput DNA sequencing starts with a biological sample (e.g. blood or tissue). A parallelized sequencing process generates millions of short “reads”. Reads are then computationally assembled to determine an individual’s whole genome sequence [from the National Human Genome Research Institute].

can be discovered because DNA is inherited: similarities between individual genomes can be used to discover near relatives (e.g. cousins), as well as predict more distant ancestry (i.e. geographical origins and associations with race/ethnicity). But because DNA is responsible for biological function, it can also be interpreted to understand biological

traits. This health-relevant aspect of genome information has been subject to regulatory concerns (e.g. by the United States Food and Drug Administration), because it can be used to determine clinically relevant information. As a result, while some products have interpreted customers' genetic data for both traits and ancestry purposes (e.g. [23andMe](#)), others avoid regulatory issues by only analyzing ancestry aspects (e.g., [AncestryDNA](#)).

3

Background

3.1 Historical context and new technologies

Prior to the rapid expansions in genetic technologies, genetic testing was confined to a specialized aspect of clinical practice. Testing for the diagnosis of a suspected genetic disorder (like Huntington's disease or cystic fibrosis) occurred only for affected individuals and their immediate family, in rare cases. The data and information produced was very specific, and was interpreted by a lab specialized in testing for the particular disease. The information was conveyed to the patient by trained experts called genetic counselors. These models for information management have been challenged by developments in genomic technology, which have enormously expanded both the scope of genetic data available and the number of individuals tested.

As outlined above, developments in genetic technologies have vastly lowered costs, creating a massive expansion in the **scope of data produced** by genetic testing. Analysis of this data is an area of ongoing development, as researchers and clinicians adapt and combine disease-specific and gene-specific analytical approaches. The ability to perform comprehensive analysis is also limited by current knowledge, as our understanding of the effects of gene variants continues to evolve.

Parallel to this expansion in the amount of data produced is an expansion in the **scope of individuals tested**. While genome and exome technologies have seen increasing clinical applications [[Manolio et al., 2013](#)], lowered costs for testing have also led to the growth of direct-to-consumer genetic testing (DTCGT) companies [[Turrini and Prainsack, 2016](#)]. These companies have brought genetic testing outside the clinic to millions of users (e.g. [[23andMe](#), [AncestryDNA](#)]), selling online products that enable individuals to access their own genetic data online for the purposes of expanding their understanding of personal ancestry, health, and biology.

3.2 Online interaction with personal genomics

The expansion of both the amount of data and the scope of individuals involved in genetic testing has led to radical changes both in the nature of the information returned and in the process of returning data. Automated methods and online tools have become necessary features for sharing extensive information with many individuals. In addition, the scale of the underlying data defies comprehensive interpretation, and scientific understanding of the data is constantly evolving — many potentially meaningful genetic variants are not yet known to be meaningful. What can be understood from genetic data, the potential for re-interpretation, and how to share that knowledge have provoked extensive debates [[Hughes, 2013](#), [Conley, 2014](#)].

DTCGT companies enable individuals to acquire genetic information without the involvement of a healthcare provider by sending a saliva sample to a DTCGT company, at a relatively low cost. To date, DTCGT do not offer whole genome or exome sequencing, but rather microarray-based genotyping. Results are delivered through online interactive reports. Several popular DTCGT services offer interactive online reports of ancestry information (e.g. [23andMe](#), [AncestryDNA](#), and [FamilyTreeDNA](#)). The service [23andMe](#) also provided risk assessment results for about 250 traits and conditions; however, in response to concerns from the FDA current [23andMe](#) health-related reports return

a more limited list of FDA-approved analyses [Pollack, 2013, 2015] (see Figure 3.1). Services also allow users to share and compare genetic data with each other, enabling comparison reports related to ancestry and gene inheritance (See Figures 3.2 and 3.3).

To date, all of the DTCGT services mentioned above also return raw genotyping data to users, who in turn can actively engage with their personal genomic data, for example, by learning about specific gene variants or conditions of interest. Indeed, consumers of genomic data have been observed transporting their data between services to capitalize on different features that allow them to engage more deeply with their data. For example, 23andMe users may export their data to FamilyTreeDNA for genealogy, or to various genetic data analysis tools [Bettinger, 2013]. Because this data is digital, and because its interpretation gets updated frequently based on new research findings, we anticipate increased focus on the development of online interactive report methods that perform automatic reanalysis.

3.3 Omic technology

The revolution in genetic data is paralleled by other *omic* technologies that bring similarly broad, untargeted biological data to clinical and direct-to-consumer contexts, including microbiome and metabolite profiling [Khamsi, 2014, Seetharaman, 2015, Dillet, 2015]. Microbiome profiling uses high-throughput DNA sequencing to profile microbiota that live in the human gut, as well as other tissues, potentially impacting digestion, weight, and other aspects of health. Metabolite profiling assays numerous small molecules in the blood, and can reveal abnormalities in metabolism and health. Like genome analyses, these and other omic technologies produce large amounts of data relevant to many potential conditions. It seems likely that these technologies will confront parallel challenges with management of comprehensive data, potentially limited and shifting interpretations, and the demand for tools to communicate information to non-expert individuals.

INHERITED CONDITIONS

These reports show your results for specific genetic variants that can cause certain health conditions. Many of these conditions are recessive, meaning that they only occur when you have two variants for that condition, one inherited from each parent. If you have inherited just one variant, you are said to be a "carrier". Carriers usually do not have the condition, but can pass the variant on to their children. Note that these reports cover only a subset of possible variants that may be linked to a condition. It is thus possible to have other variants not covered by these reports.

NAME	CONFIDENCE	STATUS ▾
ARSACS	★★★★	Variant Absent
Agenesis of the Corpus Callosum with Peripheral Neuropathy (ACCPN)	★★★★	Variant Absent
Alpha-1 Antitrypsin Deficiency	★★★★	Variant Absent
Autosomal Recessive Polycystic Kidney Disease	★★★★	Variant Absent
BRCA Cancer Mutations (Selected)	★★★★	Variant Absent
Beta Thalassemia	★★★★	Variant Absent
Bloom's Syndrome	★★★★	Variant Absent
Canavan Disease	★★★★	Variant Absent
Congenital Disorder of Glycosylation Type 1a (PMM2-CDG)	★★★★	Variant Absent
Connexin 26-Related Sensorineural Hearing Loss	★★★★	Variant Absent
Cystic Fibrosis	★★★★	Variant Absent
D-Bifunctional Protein Deficiency	★★★★	Variant Absent
DPD Deficiency	★★★★	Variant Absent

Figure 3.1: An example of the "Inherited Conditions" report produced by 23andMe, screening for serious genetic conditions. In this report the user is presented information for every tested gene and location. Because these tests are usually negative (no disease-causing variant found), it is common for users to have over 50 results that are entirely "Variant absent". Uncertainty in interpretation of variant effects is represented by a confidence score (number of stars in the "Confidence" column).

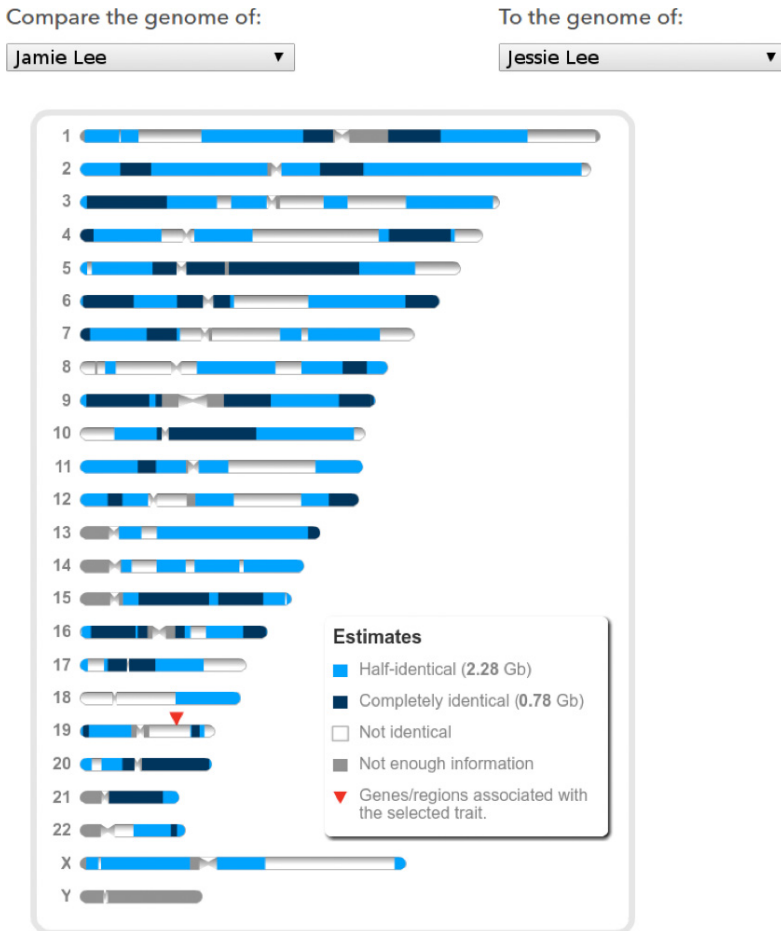


Figure 3.2: DTCGT providers typically share information about ancestry, and sometimes information is provided for specific chromosome regions. For example, this 23andMe data comparison illustrates shared DNA segments in a pair of sibling genetic data sets.

3.4 Genomic data privacy and access

Commercial companies typically provide their customers with access to raw genetic data. This raw data contains an individual's gene variant information, without interpretation, and can potentially be reanalyzed using third-party tools and resources. In contrast, research studies only

Your List of Genes

APOE

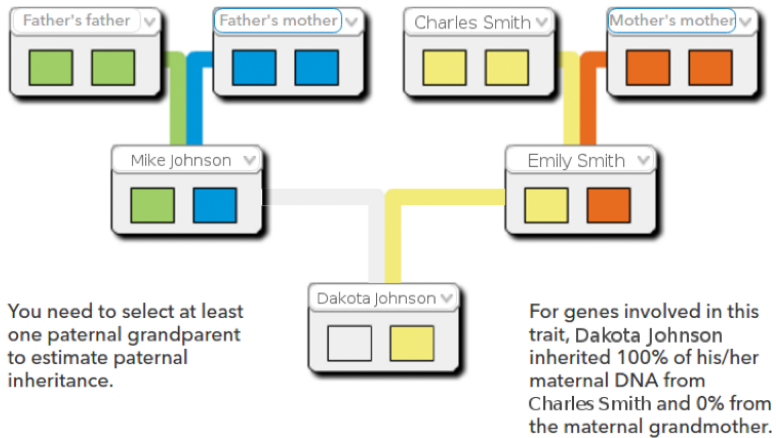


Figure 3.3: In some cases, DTCGT companies may be able to share information for more complex data comparisons. For users that have genetic data from their grandparents, 23andMe’s “family tree” tool can be used to discover which grandparents contributed genetic material for specific genes.

rarely give their research subjects access to their raw genetic data, possibly due to legal concerns regarding regulation of genetic testing [Evans, 2014, Karow, 2014].

Privacy of genetic information has been a major concern for many individuals interested in genetic data analysis. Because it contains extensive ancestry and trait information, genetic data is potentially identifiable (especially when combined with additional data), and in 2013 a re-identification study demonstrated that inference of surnames from Y-chromosome data could be used to re-identify dozens of “anonymous” public research genomes [Gymrek et al., 2013]. Many individuals are concerned about discrimination in access to insurance based on genetic data despite some legal protections (e.g. the United States Genetic Information Nondiscrimination Act of 2008) [EEOC, 2008; Sanderson et al., 2016]. Discrimination is still possible in other contexts such as employment and in some cases can be impossible to regulate. Genome data privacy has thus been a concern in both research

and commercial contexts [Nature, 2013, Phillips, 2015], and controlled-access sharing for genome data research is common (e.g. as performed by the United States Database of Genotypes and Phenotypes, dbGaP), while commercial groups often promise privacy for individual-level genetic data.

3.5 Public genomes and follow-up research

The Harvard Personal Genome Project (PGP) is a research study, initiated in 2005 by George Church of the Harvard Medical School, that seeks to improve understanding of how genomic and environmental factors influence human traits through the creation of a public dataset [Church, 2005]. PGP volunteers agree to share their genomic sequences, as well as health data, with the scientific community and the public. Today, more than 5,000 volunteers are enrolled in the project through a process of *open consent* [Lunshof et al., 2008, Ball et al., 2012, 2014] and have agreed to share their genomic information publicly. Of these, over 300 individuals have had whole genome sequencing data produced by the project and publicly shared.

Public genome data produced by the Harvard PGP is accompanied by a preliminary research report produced by the GET-Evidence system [Ball et al., 2012]. This report is given to participants prior to public data release, to better inform them regarding the data they are choosing to publicly share. Upon consent, these reports are publicly shared alongside public genome data, as is the underlying database for variant interpretations.

The Harvard PGP is distinguished from other public genome resources in its ongoing work with participants. The PGP has historically invited other research groups to work with its volunteers, enabling follow-up research with this population and providing a unique opportunity for studying HCI in genomic data. This process is now streamlined and explicitly supported within the Open Humans platform, a successor project to the PGP. In Open Humans, researchers can recruit and work with PGP participants, as well as other members with other

public and private genetic data such as microbiome, activity logs, GPS, and other diverse data sources [[Open Humans](#)].

3.6 Characteristics of personal genomic data

The data produced by genome sequencing, exome sequencing, and comprehensive genotyping spans all ~20,000 genes of the human genome. Comparing any two genomes reveals millions of genetic differences, and the database of all observed variations in all humans currently exceeds 150 million genetic variants [[dbSNP, 2016](#)]. The information that can be learned from this comprehensive data is diverse, complex, and evolving.

One of the most popular uses of this information is for ancestry analysis. By comparing individual data with population-level information, analysis of DNA segments can predict racial and ethnic origins. More recent ancestry — the discovery of cousins and near relatives — is also possible through large databases that detect shared segments of DNA between individuals. In addition to ancestry, everything about ourselves that is heritable is — in theory — contained within the data. Although many gene variants have disease associations that imply clinical utility, individuals are also motivated by curiosity and entertainment to learn about variants with no medical purpose (e.g. physical traits) [[Vayena et al., 2012](#)]. Analyzing data to produce health and trait insights is daunting: [ClinVar](#), a public database aggregating reported effects, now has over 100,000 gene variants. Aggregating and filtering results to facilitate understanding of any individual genome is challenging, and it is still unclear how to combine information from multiple genes affecting a single trait or disease.

3.7 Uncertainty in personal genomics

The relationships between genes and health effects are in many cases not well understood, and knowledge about such relationships evolves dynamically with the development of new technologies, processes, and research results. Individuals are therefore often required to continuously reconsider their genomic information against the most current research evidence. As a result, interaction with personal genomic data

is unique compared to other forms of personal data, where the dynamic element is often the data itself, which is usually sampled at intervals over time with the objective of creating an incremental feedback loop to influence an individual's behavior [Li et al., 2011]. In such traditional forms of personal informatics, uncertainty often results from the context and accuracy of data tracking. Genomic data, on the other hand, are largely stable during a person's lifetime –99.9999% accuracy is typical, and continues to improve [Peters et al., 2012]. The uncertainty in personal genomics stems from the interpretation of the data, and from the evidence of related implications for the user's health and traits (see, for example, 23andMe's "confidence" score, Figure 3.1). This form of uncertainty arises because our understanding of genetic variants changes over time: hypotheses are corrected or updated, and new medical research exposes new relationships between people's genetic makeup and their effects.

4

Related Work

4.1 Personal informatics

[Quantified Self](#), also known as personal analytics [[Regalado, 2013](#)] and personal informatics [[Li et al., 2010](#)], refers to research and practice involving communities, practices, and systems that help people collect and reflect on their personal information. The increasing availability of low-cost sensors has accelerated the practice of self-tracking as well as the rise of the Quantified Self movement [[Quantified Self](#)]. Researchers and designers have developed and studied numerous self-tracking technologies and applications for health and wellness [[Klasnja and Pratt, 2012](#), [Swan, 2009](#)]. An underlying assumption driving the growth of this field is that individual's knowledge of their data facilitates reflection, which in turn leads to self-discoveries and to lifestyle changes. While researchers of personal informatics and Quantified Self acknowledge the value of self-tracking technologies [e.g., [Bentley et al., 2013](#), [Consolvo et al., 2008](#), [Lin et al., 2006](#), [Mamykina et al., 2008](#), [Patel et al., 2012](#)], they also discovered several barriers toward the adoption and effective use of self-tracking technologies including data integration

and interpretation, unsuitable visualization and analytics tools, poor skills for analyzing data, and fragmented data scattered across multiple platforms [Bloss et al., 2010, Choe et al., 2015]. It is important to note that most of the research identifying practices and barriers in personal informatics has been conducted with expert users, *quantified-selfers*, who are early adopters, health enthusiasts, or patients.

Personal Genomics shares the main goals and assumptions of personal informatics — facilitating self-discovery based on personal information. In the popular media, personal genomics is viewed as an extension of quantified self: “It’s the ultimate quantified-self set of numbers. Forget how many steps you average in a week, or what your resting heart rate is — DNA genotyping allows us to create the complete spec-sheet for our bodies” [Lamkin, 2016]. However, personal genomics is different from other self-tracking data in being largely stable during a person’s lifetime while its interpretation, and related implications for the user’s health, are dynamic. In addition, while personal genomic information is inherently personal, it is also shared among family members, thus affecting the health and wellbeing of its owner as well as of family members and future offspring. Furthermore, personal genomic information could lead to discoveries about one’s ancestry that could impact not only one’s identity and sense of belonging but also families and communities.

Considering the complex and sensitive nature of personal genomics, its scale, and the dynamic nature of its interpretation, we view personal genomics as a new frontier for personal informatics.

4.2 Communicating uncertainty

One of the key tasks of a personal genomic report is to communicate to its reader the uncertainty that is associated with their data and its implications. An abundance of work has investigated the visualization of uncertain information in other domains.

Existing taxonomies for communicating uncertainty identify sources of uncertainty (and visual presentation techniques) [Skeels et al., 2008, Taylor and Kuyatt, 1994, Thomson et al., 2005]. Additional work

explores cognitive biases of decision-making under uncertainty and corrective visual approaches [Inbar, 2007, Tversky and Kahneman, 1974].

In particular, researchers explored public perception and communication of weather forecast uncertainty, showing that the general public understands that uncertainty is inherent in weather forecasts as well as some of the factors that increase uncertainty [Morss et al., 2008, Joslyn and Savelli, 2010, Joslyn et al., 2013]. The research shows that when communicated carefully, uncertainty in forecast leads to better outcomes such as increased precautionary action for extreme weather events [LeClerc and Joslyn, 2012, Savelli and Joslyn, 2012]. Research also indicates that communicating uncertainty in an interface or visualization of weather forecast could increase trust and help people make better decisions [Gigerenzer et al., 2005, Roulston et al., 2006, Joslyn and LeClerc, 2013]. Evidence from other application domains such as remaining driving range in electric cars [Jung et al., 2015] and weight scale [Kay et al., 2013] indicates that not including uncertainty information can decrease trust in the application.

Numerous applications tracking new types of personal and often uncertain data have explored how to present the data to encourage behavior change and reflection [Rooksby et al., 2014, Epstein et al., 2015, Choe et al., 2015, Lee et al., 2015]. A study comparing visualizations of uncertainty found that participants' judgment of these visualizations was significantly influenced by familiarity, ease of understanding, and visual appeal [Greis et al., 2016]. Another study comparing the impact of various representations of uncertainty on different activities concluded that different types of visualizations lead to different learning outcomes and suggested that an interactive display may be best for communicating uncertain information [Nadav-Greenberg et al., 2008]. Similarly, a different study shows that the amount of presented uncertainty leads to different outcomes of risk taking within a game context [Greis et al., 2016].

This growing body of work is important for researchers to draw upon when studying how to communicate uncertainty of personal genomic information.

4.3 Interaction with biological data sets

To date, little HCI research has focused on direct user engagement with personal genomic information. Instead, researchers focused on developing new ways of interacting with large-scale and complex biological data sets for use as a platform to explore novel interaction techniques, such as tangible interaction [Shaer et al., 2013, Manshaei et al., 2016]. Systems developed include a tangible interface for scientists designing new DNA molecules [Schkolne et al., 2004], a tabletop tangible interface for system biologists [Wu et al., 2011], and several tabletop interfaces for interactive visualization of biological data sets in informal and formal learning settings, such as DeepTree [Block et al., 2012] and PhyloGenie [Schneider et al., 2012]. G-nome Surfer [Shaer et al., 2011] is a tabletop interface for collaborative exploration of genomes; Tangible mtDNA is an active tangible and tabletop system for collaborative exploration of mitochondrial DNA sequencing data in breast cancer patients [Manshaei et al., 2016]. However, these systems were not designed to support non-expert users in the self-exploration of their own genomic data but rather to support researchers and learners in exploring genomic data sets.

4.4 DIY genome analysis

Individuals can potentially explore and analyze their genetic data in a variety of ways including: exploring information provided by the primary testing service; secondary analyses performed by third parties; or personal analyses drawing on various resources and databases.

However, information from primary testing services is often the only information explored. In some services this information can be quite extensive. 23andMe, for example, provides in-depth reports on each of the variants included in the health and trait analyses. These reports summarize current research, population statistics, and provide links to primary literature. Due to regulatory concerns, however, many other testing providers avoid providing information regarding health and traits associated with genetic variants. Even services that do provide this information (e.g. 23andMe) are still limited to the set of variants

Table 4.1: A selection from a growing variety of products giving consumers to access to genetic data, as well as analysis tools and resources for interpretation.

Tools, products, & resources	Type(s)	Notes	
23andMe	— Genotyping testing service	— Primary testing service	
	— Ancestry interpretation	— Exportable raw data	
	— Relative finding	— Over 1 million customers	
	— Health & trait interpretation		
AncestryDNA	— Genotyping testing service	— Primary testing service	
	— Ancestry interpretation	— Exportable raw data	
	— Relative finding	— Over 2 million customers	
ClinVar	— Variant information database	— Aggregated reports — Expert resource — Public domain	
	Enlis Genomics Personal	— Variant function analysis	— Accepts data import
	Exome Aggregation Consortium (ExAC)	— Variant information database	— Variant frequency information — Share-alike license
FamilyTreeDNA	— Genotyping testing service	— Primary testing service	
	— Ancestry interpretation	— Accepts data import	
	— Relative finding	— Exportable raw data	
Gedmatch	— Ancestry interpretation	— Accepts data import	
	— Relative finding		
Genos	— Exome sequencing	— Primary testing service — Exportable raw data	
	— Whole genome sequencing	— Primary testing service	
Illumina Understand Your Genome	— Health & trait interpretation	— Exportable raw data	
	— Ancestry interpretation	— Data analysis software	
Interpretome	— Ancestry interpretation	— Data analysis software	
	— Health & trait interpretation		

(Continued)

Table 4.1: (*Continued*)

Tools, products, & resources	Type(s)	Notes
Online Mendelian Inheritance in Man (OMIM)	— Gene information database — Variant information database	— Curated summaries of primary literature
PubMed	— Publications database	— Search primary literature
Promethease	— Health & trait interpretation	— Accepts data import
SNPedia	— Variant information database	— Crowdsourced — Non-commercial license
Veritas Genetics	— Whole genome sequencing — Health & trait interpretation	— Primary testing service — Exportable raw data (added fee)

they are reporting upon. Thus, many individuals are interested in additional analyses of raw genetic data from primary providers.

There are a variety of third-party services and tools for analyzing genetic data, which provide tools for ancestry analysis or health/trait analysis (see Table 4.1). Services like [FamilyTreeDNA](#) and [Gedmatch](#) accept third-party data for ancestry analysis, and enable relative finding within their communities. Health and ancestry analyses are also possible using third-party tools like [Enlis](#), [Interpretome](#), and [Promethease](#). These tools often provide links to resources and databases containing variant function predictions and reported effects (e.g. [ClinVar](#), [ExAC](#), [OMIM](#), [SNPedia](#), [PubMed](#)).

Table 4.1 presents a selection of primary testing providers, as well as third-party analysis services, tools, resources, and databases used for exploring personal genomic information. For each of these we summarize its main features.

The rest of this article lays the ground for future HCI research on direct engagement of non-experts with personal genomics.

5

Emerging Framework for HCI for Personal Genomics

In this section, we present a framework for HCI research for personal genomics. The framework provides a research agenda, which investigates the role of HCI methodology and interventions in personal genomics. We center our framework on three themes — Understanding, Informing, and Empowering users:

Understanding users — examining users’ motivations, needs, information practices, and concerns in engaging with and sharing their personal genomic information;

Informing users — investigating how variations in user interface design affect users’ deliberation on their consent to genetic testing, their comprehension of genomic information, as well as their intentions to share such information with others;

Empowering users — focusing on the design and development of interactive tools that empower users to engage with their personal genomic information over time in applicable and meaningful ways.

For each theme we review the recent and current research, and highlight open questions for future HCI research.

5.1 Understanding users

Since the technology that enables lay people to interact with personal genomic information directly is relatively new, many questions remain open regarding non-expert engagement with personal genomic information.

For example, what motivates non-expert individuals to acquire personal genomic information? What concerns do they have about engaging with personal genomics? What are their goals in exploring their personal genomics? How do they learn from their personal genomic information? Which tools do they use and how do users document and organize their findings? How do they validate their findings? What kind of information do they share and with whom?

Gaining insights into individuals' motivations, needs, information practices, and concerns of engaging with and sharing their personal genomic information plays a vital role in the design of new interactive tools that empower non-experts to learn from their personal genomic information. However, the importance of understanding personal genomic users goes beyond the development of new tools to informing future policy about personal genomics as well as to envisioning the future of personal informatics.

5.1.1 Current research on understanding users

Several studies have investigated the motivation of DTCGT users. In these studies, curiosity was mentioned as the participants' primary motivation for undergoing genomic testing [Goldsmith et al., 2012]. Most participants wanted to learn more about themselves, were curious about their genetic makeup, or wanted to learn about individual genetic risk factors. Participants also stated that they would use information gained from the test to take personal responsibility for their future health [McGowan et al., 2010]. Other themes included fascination with genealogy, contribution to research, and recreation [Goldsmith et al., 2012]. Studies also identified several concerns among DTCGT users, including privacy, as well as the nature of the results and their future impact [Kuznetsov et al., 2015]. For example, users may develop anxiety

surrounding the risk of particular conditions, the possibility of discrimination from employers and insurance companies, or the possession of sensitive and personal data. The Impact of Personal Genomics Study [PGen, 2016] is a current large-scale longitudinal study that surveys consumers of two U.S. DTCGT providers — 23andMe and Pathway Genomics — to determine the characteristics of consumers, the psychological, behavioral, and health impact of genetic testing, and the ethical, legal, and social issues associated with DTCGT services.

Little empirical data exists about the attitudes and motivations of people who have had their whole genome sequenced [Goldsmith et al., 2012, Linderman et al., 2016], since a relatively a small number of users around the world have had their entire genome sequenced and returned to them. Of the individuals that have access to their whole genome information, a significant subset have been volunteers of the Harvard Personal Genome Project (PGP). We established a design partnership with the Harvard PGP and have collaborated closely with its researchers and volunteers to study the information practices and needs of personal genomics users. PGP volunteers can be categorized as early adopters [Rogers, 2003] or extreme user group [Choe et al., 2014, Troshynski et al., 2008]. They tend to have advanced education and possess favorable attitudes toward science. Their motivations and information practices might not be generalizable or applicable to the broader population. However, in areas that evolve rapidly based on technical and scientific innovations, the perspectives of early adopters provide important insights, because they have used the existing technologies and have had the opportunity to identify challenges and potential solutions [Choe et al., 2014, Troshynski et al., 2008].

To better understand the needs and information practices of PGP participants, we conducted a qualitative study with 63 participants from the Harvard PGP volunteer community [Shaer and Nov, 2014]. Participants filled out an online questionnaire that consists of 10 open questions (see Shaer and Nov [2014]) regarding their motivation, goals, and information practices, in addition to demographics questions. Although study participants were motivated by a diverse set of goals, ranging from understanding traits, to identifying health risks,



Figure 5.1: Five common information tasks of non-expert personal genomic users.

to learning about their ancestry, to contributing to science, they used interactive tools to perform mostly the five common information tasks listed in Figure 5.1. The study also identified high-level needs of non-experts seeking to learn more from their personal genomic data information. Figure 5.2 presents a summary of these needs.

From an HCI perspective, these findings can serve as a basis toward the design of new interaction techniques and tools for personal genomics. To that end, it is important to gain further and more nuanced insight into how users engage with, and learn from, their annotated personal genomic reports. For that reason, we conducted a second qualitative study of personal genomics users [Shaer et al., 2015]. We interviewed and observed 36 PGP participants, who have had their whole genome sequenced, as they explored their personal genomic data using the GET-Evidence interactive genomic report [Block et al., 2012]. This study deepened our understanding of the needs and practices of personal genomic users, highlighting that individuals are predominantly

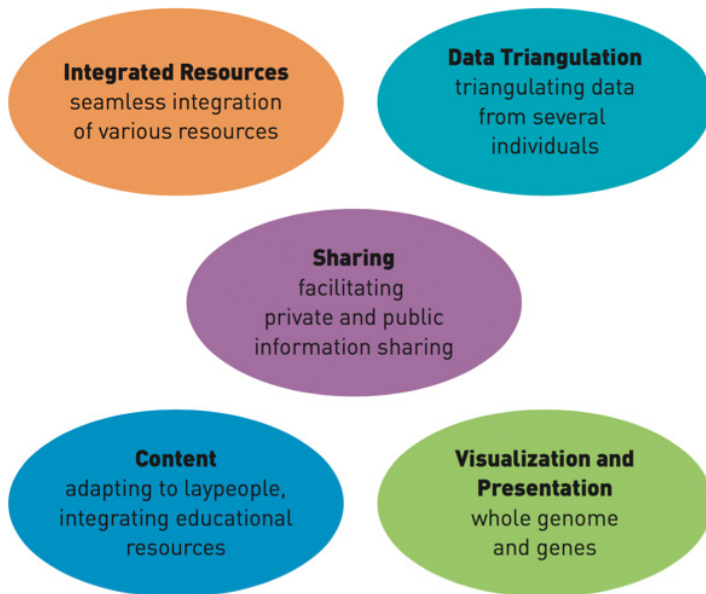


Figure 5.2: Needs of non-experts seeking to learn from their personal genomic data information.

concerned with genetic variants that are well-established, pathogenic, and have high clinical importance.

While these insights are important for the development of interactive tools and reports, important questions about user engagement with personal genomics remain open.

5.1.2 Open questions for future research

In this section, we highlight two open questions for future HCI research on understanding users:

1. *Perceiving uncertainty and discrepancies:* Although personal genomic data are largely stable during a person's lifetime, their interpretation and implications might change over time as new medical research exposes relationships between genes and health. In order to accurately communicate the inherent uncertainty of personal genomic interpretations, HCI research should investigate

how users of personal genomics perceive uncertainty and discrepancies between their personal genomic information, their experiences, and their family history. Such investigations could draw upon research the rich body of literature examining how lay individuals understand uncertainty in other domains (e.g. weather forecasting), which we discuss in Section 4.

2. *Motivation for curating and sharing information:* While personal genomic information is personal, it is also shared among families and community members. Curating and sharing findings from a personal genomic report could motivate families to seek the advice of healthcare providers, promote scientific understanding of relationships between genes and traits, enable biomedical research into rare conditions, and bring together communities of shared origins or interests. Additional research is required to understand how people curate personal genomic and related information, what factors influence people to share personal genomic information, what information they share with others, and with whom they share such information. Such research could be informed by the increasing amount of literature on family informatics [Colineau and Paris, 2011, Grimes et al., 2009, Pina et al., 2017] and on engaging patients and their family caregivers in perusal of patient-provided data and clinical data, e.g. [Grimes et al., 2009, Colineau and Paris, 2011, Chung et al., 2016, Woollen et al., 2016, Pina et al., 2017]

Qualitative HCI research methods including longitudinal semi-structured studies [Blandford et al., 2016] have an essential role in addressing these questions. However, when choosing a research method, careful attention should be directed toward considering the ethical implications of the study. Due to the sensitive nature of genetic testing, researchers should consider a range of ethics and privacy issues ranging from asking participants to share private information, to gathering sensitive information from social media and discussion groups, to presenting users with new interpretations of their personal genomic information that could drastically impact their understanding of their health risks, identities, and of their families.

5.2 Informing users

The decreasing costs of obtaining personal biological data, coupled with the proliferation of websites and digital devices for self-tracking and testing, have resulted in a reduced barrier to entry for participation in online biomedical research in general, and in genetic testing in particular, and have highlighted the importance of informed online consent processes. Taken together, these factors require us to re-evaluate the effectiveness and potential of digital, interactive consent in this new context of enrolling in biomedical research.

From an HCI perspective, interactive consent poses several challenges and opportunities to informing users: whereas individuals participating in biomedical research in general or in genetic testing in particular were formerly able to engage with a professional in a face-to-face dialogue, potential online research or DTCGT participants have fewer opportunities to ask questions and express their concerns in real time [Balestra et al., 2015]. Furthermore, presentation techniques and design interventions may influence an individual's decision to participate [Friedman et al., 2000, Das et al., 2015], raising concerns about the voluntariness of participation. In response to such concerns, federal agencies are drafting guidelines for electronic consent [FDA, 2015]. At the same time, interactivity offers unique advantages to the consent-seeking and deliberation process: through interactive consent forms, consent seekers may provide additional on-demand information based on participants' interests; through careful use of feedback and user analytics, consent seekers may also be able to gain insights into the parts of the consent forms that participants find more useful, more challenging, or requiring additional clarification. Yet another way in which web-based interactive consent forms can help the consent-seeking and deliberation process includes the use of social features, such as ratings, recommendations, and annotations. Such features could enable individuals deliberating on their consent decision to share information, evaluate different perspectives, and ultimately explore the risks and benefits of the research beyond the scope of one-on-one dialogue with a research staff member.

In addition to facilitating an informed consent process, HCI has an important role in enhancing non-expert comprehension of personal genomic reports, and in particular to help individuals make sense of the uncertainty of personal genomic data and its implications.

5.2.1 Current research on informing users

5.2.1.1 Improving online consent processes

The decision to consent to participate in medical research is mediated by two main factors: participants' comprehension of the details of the study, and their trust in the research organization [Kneipp et al., 2009]. Recent studies on consent form design focus predominantly on the impact of content structure, graphical enhancements, and multimedia on comprehension [Dresden and Levitt, 2001, Dunn et al., 2002, Murphy et al., 2007, Stiles et al., 2001].

To understand how social features can enhance online consent forms, we studied the use of social annotations in personal genomic research [Balestra et al., 2016a,b]. First, we explored the impact of social annotation, embedded in online consent forms, on individuals' informed consent beliefs and decisions. Participants were presented with an online consent form for a personal genomics study, and were randomly assigned to either a social annotation condition that exposed them to previous users' comments on-screen (Figure 5.3), or to a traditional consent form without annotation. We compared participants' perceptions about their consent decision, their trust in the organization seeking the consent, and their consent decision across conditions. We found that while consent rates did not differ across conditions, on average individuals exposed to social annotation felt that their decision was more informed, and furthermore, that the effect of the exposure to social annotation was stronger among users characterized by relatively lower levels of prior privacy-preserving behaviors.

In a second study [Balestra et al., 2016b], we focused on the influence of annotations' valence on participants' perceptions and behaviors surrounding online consent for biomedical research: participants were presented with an online consent form for a personal genomics study

PGHCI: Consent Form Task
Section 6/8

2. How can I participate?

Participation in this study involves (1) providing your trait data and genetic data to the HICFGP private online database, and (2) consenting to the use of this data for research purposes. **Your data will be used indefinitely, unless you choose to withdraw from the study.**

Trait Data: Trait data will be collected by a phone interview, which will take about 30 minutes. Upon completion of the interview, your data will be saved to the HICFGP private online database. The trait data that is collected may include, but is not limited to: date of birth, medications, vaccines, diseases, personal and familial medical history, race, ethnicity, vital signs, and lifestyle traits.


Genetic Data: You will be sent materials for an at-home saliva collection. This kit is self-administered, and typically requires you to provide 2-4 milliliters of your saliva. You will then send the saliva sample back to the laboratory in the prepaid package that is provided for you. Your genome will then be sequenced and analyzed in the lab. Your genomic data will be saved to the HICFGP private online database and a copy of the results will be sent to you.


3. Will I be compensated?

Participation in this study does not entitle you to any financial compensation. In the event that your genetic data contributes to a discovery, neither you nor your family will receive any financial benefits. However, by participating in this study, you will receive a copy of your genome data for free. This data may not be used as a clinical diagnostic, but may otherwise be used as you wish (e.g. for personal research). Purchasing your genome data through a commercial company currently costs around \$2,000.

4. How will my data be kept private?

We have taken many measures to make sure that your data is kept confidential and private within the HICFGP private online database. Furthermore, researchers who conduct statistical analyses with your full genetic data will not have access to your name, address, or user ID. On the other hand, researchers who may interview you will have access to your name and contact information; however, they will only have limited access to a small amount of your genetic information. Only non-identifying information will be published in scientific journals.

 I like that they've thought about data protection at every step of the process.
Likes: 0 Like

 Even though it says I can withdraw from the study whenever I want, your data will be used indefinitely. Kind of freaks me out, what if something happens to me? I don't want to be the next Henrietta Lacks.
Likes: 28 Like


 Leave a comment

Figure 5.3: A consent form for personal genomics research with social annotation, from [Balestra et al., 2016a,b]. Users can read comments on the right, and highlight text to leave their own comments.

that contained social annotations embedded in its margins. Individuals were randomly assigned to view the consent form with positive-, negative-, or mixed-valence comments alongside the text of the consent form. We compared participants' understanding of the material and perceptions of informedness, their trust in the organization seeking the consent, and their consent across conditions. We found that consent forms containing positive-valence annotations are likely to lead participants to feel less informed and simultaneously more trusting of the organization seeking consent. In certain cases where participants spent little time considering the content of the consent form, participants exposed to positive-valence annotations were even more likely to consent to the study.

5.2.1.2 Transforming personal genomic reports to a long-term information partnership

Traditionally, informed consent is a one-time event in which a person agrees to participating in research and is informed of what is known to the science at that point in time. However, as medical research, and information about the relationships between genetic and health information develop, digital informed consent offer opportunities for long-term information partnerships between scientists and users. In such partnerships, the person consenting can be continuously informed of new knowledge about risks and benefits emerging from research, and the ensuing new interpretations of their genetic data. Furthermore, by combining the nature of genetic information as data-static but interpretation-dynamic with design that incorporates social features and curation, genetic data can be curated by the users or relevant others (including healthcare professionals, family members, or others with exposure to the user's genetic data), such that emerging new research information is shared with the user based on their genetic profile. Such approach turns a one-time consent event to an ongoing informing mechanism which the user can enhance individually and socially. It also, however, represents potential risks of people consenting to one thing (based on existing medical knowledge at the time of consent) and being potentially exposed to information about other

things (based on advances in research). This may, in some cases, be more stressful or mentally taxing than these people had anticipated at the time of consent. Mechanisms can be designed for such cases, for example, by making the consent time-specific or with placing the control of information sharing extent with the user.

5.2.1.3 Increasing comprehension of personal genomic reports

Existing research has focused on participants' comprehension of anonymous sample personal genetic reports. [Lachance et al. \[2010\]](#) examined the informational content, literacy demands, and usability of DTCGT service websites. They found that websites vary widely, and most users would struggle to use these resources effectively. The authors suggested that future tools focus on distilling and prioritizing important information while considering readability and usability elements. Other studies have looked more specifically at users' comprehension of genomic reports. [Ostergren et al. \[2015\]](#) assessed participants' comprehension of anonymized genomic reports and found that comprehension varied widely according to demographic characteristics, numeracy and genetic knowledge, and types and format of the genetic information presented. They suggested that the presentation of genomic data be tailored to the test type and to consumer characteristics. To investigate the effect of different visualizations on consumers' understanding of personal genomic data, we conducted a comparative study, which indicated an advantage to non-zoomable visualizations, with best results (in terms of both objective comprehension and subjective preference) from using bubble graphs [[Shaer et al., 2015](#)].

In contrast to the studies, which presented users with anonymized genetic data, [Kuznetsov et al. \[2015\]](#) presented users with their own 23andMe data to understand how they make sense of and contextualize their results, critique and evaluate the underlying research, and consider the broader implications of genetic testing. They framed consumers as members of biocitizen publics in which there is an emphasis on individuals' engagement with the community and higher-order learning processes [[Airasian et al., 2001](#)], rather than merely perceiving

results and individually gathering information. The authors recommend the development of platforms for aggregating hybrid knowledge for creative reflection on professional science, and for supporting collaborations across communities.

5.2.1.4 Communicating uncertainty

The personal genomics context offers a form of uncertainty not addressed by existing taxonomies and applications. In the genomics context, unlike most personal informatics contexts, the full data set is known and is mostly stable — the source of uncertainty is the interpretation of the data, which depends on new scientific findings. Furthermore, personal genomics requires the communication of multi-dimensional uncertainty — uncertainty that emerges from the accumulation of several levels of uncertainty that the user encounters. Consider for example cases where known relationships between genetic characteristics and medical conditions convey a probabilistic message (e.g. that the carrier of a certain genetic variant has a 10% likelihood of developing a certain medical condition). On top of that uncertainty, consider cases where the data supporting this relationship between the variant and the medical condition is not well-established empirically, or that it is based on studies that have not been replicated. Such accumulation of even two sources of uncertainty is highly complicated for most non-expert users, and more research is needed in order to understand how to convey such information to non-experts in understandable ways. [Karczewski et al. \[2012\]](#) stress the importance of teaching the methods of genetic risk calculations.

5.2.2 Open issues for future research

The following are some of key issues to be addressed by future HCI research on informing users:

1. *Enhancing the online consent process with social features*: while recent research indicates that adding social annotation to online consent forms could potentially inform participants and improve their deliberation processes, many research questions remain

open. For example, what are the most effective ways to solicit useful content for social annotation? How, whether, and to what extent should the content contributed by non-expert users through social annotation tools be curated and edited before is it presented to readers of informed consent forms? If such curation takes place, who should be in charge of it? To what extent should the social intervention be based on explicit user choices — for example, should the exposure to social annotation be based on opt-in versus opt-out design defaults? What other social interventions can be considered for an interactive informed consent form? And, finally, what are the ethical concerns associated with interventions in consent forms, and what are the responsibilities of the consent seekers in this respect?

2. *Personalized presentation and user choice*: existing research suggests that the presentation of genomic information be tailored to the test type and to the characteristics of the user. However, more research is required to understand what elements should be interactive and/or customized and in what ways. What are the best practices for design interventions that are tailored, transparent, and valuable to the user? How much autonomy should the user receive to make sure they are well-informed, and to what extent should design interventions be based on explicit user choices?
3. *Communication of multidimensional uncertainty*: delivering personal genomic information to users requires the communication of multidimensional uncertainty — that is, uncertainty that emerges from the accumulation of several dimensions of uncertainty the user encounters. Communicating such accumulating uncertainty is not currently well addressed by existing taxonomies and tools, and poses a challenge to researchers and practitioners. Open questions for researchers and practitioners include: How to best construct effective representations of multidimensional uncertainty? How should different levels of visual and statistical literacy be served through interactive tools to facilitate multidimensional uncertainty communication? And finally, how to evaluate users'

comprehension of multidimensional uncertainty? These investigations could potentially draw upon the body of research on communicating uncertainty described in Section 4, as well as on the proceedings of CHI 2017 workshop on Designing for Uncertainty in HCI [Greis et al., 2017].

5.3 Empowering users

The applicability of personal genomic information to individuals, and its evolving understanding based on new scientific discoveries, evoke compelling HCI questions when considering digital tools that mediate the interaction between the individual and their data: What are the functional requirements for supporting meaningful engagement with personal genomic information? How can we design tools which allow users to explore and make sense of their personal genomic information? How can we design tools that support long-term engagement and personal research? How can we empower users to contextualize and act upon their genomic information?

Interaction with personal genomic data often begins by reviewing a personalized genetic report [Shaer et al., 2015]. Thus, the design of interactive reports for non-expert individuals that support sense-making and exploration plays a vital role in empowering people to learn from and engage with their personal genomic information. Self-exploration tools should also enable non-experts to contextualize and compare their personal genomic information with other individuals (e.g. family members), ancestry information, and family medical history [Kuznetsov et al., 2015] as well as to share information with a doctor, friends, family, or the public. Finally, since genetic research evolves with the development of novel technologies, processes, and new scientific findings, tools for non-experts should support long-term engagement, allowing people to re-interpret their information in light of new scientific findings as well as personal developments.

5.3.1 Current research on empowering users

Our studies with personal genomic users indicate that non-expert individuals mostly use the tools offered by their genetic data provider (e.g.

23andMe, PGP, AncestryDNA, 2016, fTree) [Shaer et al., 2014, Shaer and Nov, 2014]. However, tools offered by DTCGT companies provide interpretations that reflect the preferences and limitations of the genetic testing company and are mostly not customizable by the individual users. Typically, the results can only be generated and curated by the DTCGT company. Also, since DTCGT tools may use proprietary algorithms, their genome analysis is not always transparent to the user and the user may not understand how the analysis was done or be able to replicate the results with other tools [Karczewski et al., 2012]. Many of our study participants [Shaer et al., 2014, Shaer and Nov, 2014] have found the existing tools, both those offered by DTCGT providers and other online tools, to be too complicated and overwhelming in terms of the amount of information they present and their use of scientific jargon.

To address these challenges, we presented a design case study of a novel interactive tool, named GenomiX, which was developed using a user-centered design process [Shaer et al., 2016]. It is important to note that GenomiX does not provide new genome interpretations but rather draws upon the interpretation provided by the Harvard PGP's GET-Evidence report. The requirements and design goals of GenomiX draw upon findings from our previous research exploring users' motives, needs, and interaction patterns with genomic data [Shaer et al., 2015].

Drawing upon these findings, we defined the following goals for an interactive tool for exploring personal genomic information: (G1) Presenting a visual summary of personal genomic information that highlights which variants are potentially concerning and require further investigation; (G2) Communicating the level of certainty of the scientific evidence associating a particular gene variant to health conditions. Since the certainty of the evidence can change over time, the report needs to provide up-to-date evidence. (G3) Relating variants to medical conditions while conveying complex relations, which associate multiple variants with a particular condition or the same variants with multiple conditions. (G4) Allowing users to curate information about variants, giving them a basis from which to conduct further research.

GenomiX is implemented as a web tool, presenting the user with a visualization that provides an overview of their genetic variant data

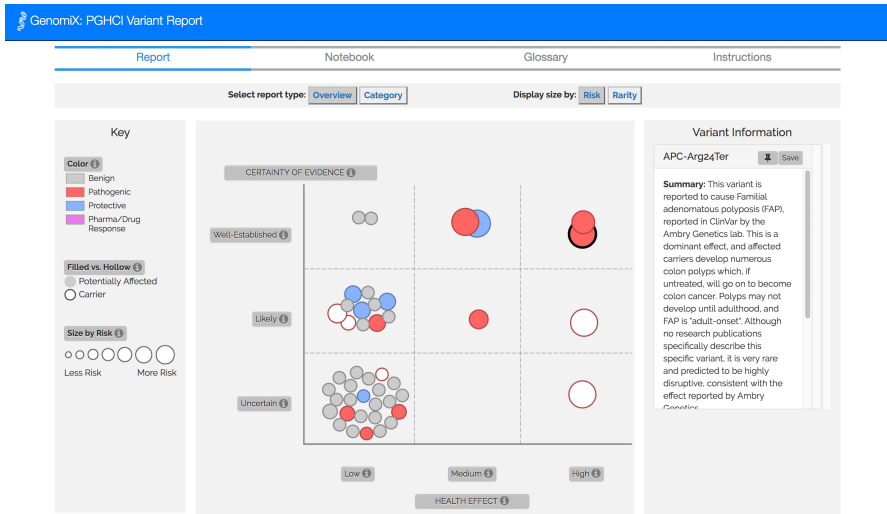


Figure 5.4: GenomiX: Gene Variant Report displaying the overview of a participant's results.

(Figure 5.4). Gene variants are represented as bubbles that are plotted between two axes: clinical importance (X axis), and certainty of evidence (Y axis); the size, color, and placement of these bubbles communicate specific information about each variant. Individuals can sort the data on the plot by either risk or rarity of the variant. Alternatively, users can view a report organized by categories where variants are sorted according to the anatomical system they might impact (see Figure 5.5). Users can save information about particular gene variants to a Notebook tab. GenomiX also offers contextual help regarding terminology as well as a glossary.

In addition to an interactive visual gene variant report, GenomiX offers Curator, a tool designed to help users curate personal genomic information, make new connections, and engage with their personal genomic information over time. Curator consists of two parts: a downloadable Google Chrome extension, which allows users to save web pages, pdf files, videos, and images; and a Notebook, which displays gene variants and web pages that a user has saved (Figure 5.6).

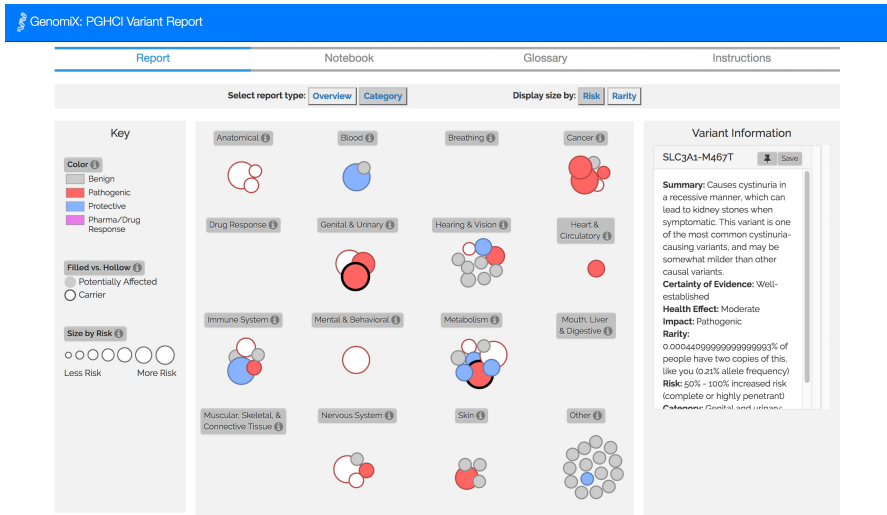


Figure 5.5: GenomiX: Gene Variant Report displaying participant’s results sorted by category.

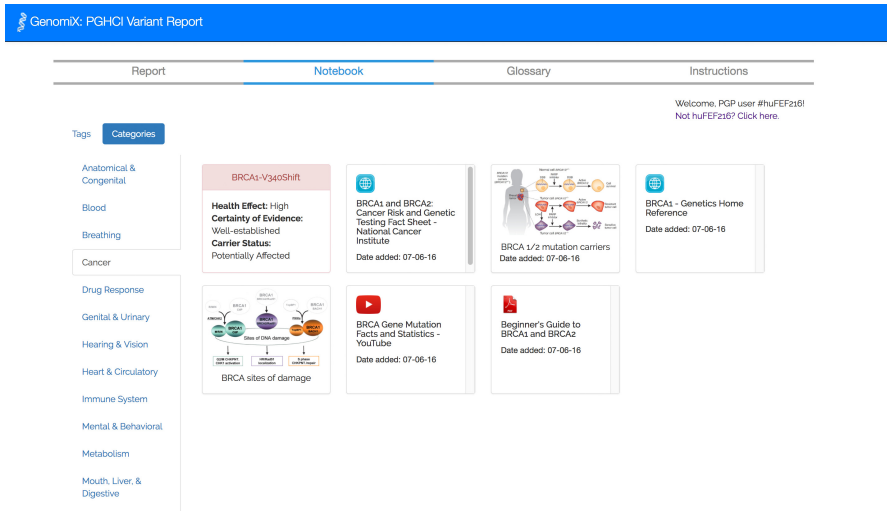


Figure 5.6: The Notebook tool, displaying gene variants and web pages that a user has saved.

To evaluate the usability and utility of GenomiX, we conducted the first study of a new genome interpretation tool to date, in which users viewed their own personal genomic data [Shaer et al., 2016]. Our findings indicate that GenomiX offered its users a number of benefits afforded by the tool's design features: visually reducing complexity improves users' ability to prioritize gene variants, communicating uncertainty helped users when interpreting genomic data and choosing which variants to focus on, and color coding helped users to identify information they overlooked in prior tool such as distinguishing protective or carrier variants.

While the design and evaluation of GenomiX offer insights into the development of future interactive personal genomics exploration tools, there are a number of key issues that should be considered in future research about tools for empowering non-experts' self-exploration of personal genomics. Future research should also draw upon lessons from a rich body of research on using genetic data to empower patients in the context of clinical care [e.g., Gilbar, 2007, Rosas-Blum et al., 2007, Berg et al., 2017]. Such lessons could be helpful in the design of novel tools for personal genomics exploration

5.3.2 Open issues for future research

Here, we highlight three key issues for future HCI research on empowering users:

1. *Comparing and relating genomes*: while existing tools present users with their own ancestry and allow relatives to view which segments of their DNA are shared (or not shared) between two individuals, additional research is needed to enable users to compare and relate genomes in order to understand trait- and health-related information. Scientific methods to understand risk in the context of genome-sharing among family members are complex and are still evolving. Contextualizing genetic testing results in respect to other family members or friends could play an important role in understanding and acting upon personal genomic information. However, many questions remain open. What criteria for comparison are meaningful and informative for

non-experts? How to visualize multiple genome comparison for non-experts in an effective manner? And how to facilitate sharing of information for comparison across individuals?

2. *Facilitating sharing*: from a societal perspective, information sharing is an important aspect of users' engagement with personal genomic information. However, sharing such information is rarely facilitated by existing tools, and is mostly limited to exploring ancestry and to discovering new biological relatives. Information sharing can be supported in a number of forms:

- (1) Sharing personal genomic information: users can share their genomic information with other users (e.g. family members), with their healthcare provider, with research projects such as Harvard's PGP, or with commercial efforts such as 23andMe. In each one of these cases, users may choose to share some parts of their information and withhold others;
- (2) Sharing meta-information with social circles: this category includes all information that is not the actual personal genomic data, but rather related to it. The recipients of the information are the user's social circles, online and offline. Examples of such meta-information sharing include: the fact the user shared their information with a research project, the user's thoughts about genomic information and the role they play and can play in modern medicine, and things the user learned as a result of engaging with their personal genomic information. The value of this type of information-sharing stems from its social impact: sharing such meta-information helps to increase awareness about the opportunities and challenges associated with sharing personal genomic information;
- (3) Sharing interpretations, insights, and other relevant information based on self and others' personal genomic information. Examples of such information sharing include new scientific research findings related to personal genomic data

(of self or others) along with explanations of how the new research findings may be relevant to the genomic information, and what other relevant information might be needed as a next step in exploring it, or pointers for others who shared their information about relevant online resources or communities (e.g. online communities of people with certain medical conditions).

In all of the cases above, in addition to understanding users' needs and concerns about sharing genomic information, it is important to investigate which design and implementation features allow for sharing while also supporting transparency regarding access, storage, and usage, and ensuring users control the level of privacy they prefer.

3. *Evaluation methods* — Personal genomic data are personal, and therefore more studies are needed in which participants are presented with their own data. Prior studies on user-facing genomics tools mostly use fictional or anonymized data; that is, genomic test results that do not belong to the participant. Thus, these studies fail to incorporate the impact and meaning of the data to the user whose data is reported. Also, existing studies tend to use novice participants who are unfamiliar with their genomic data to evaluate visualizations, thereby being unable to look at participants' evolving understanding of their genome, or how well a proposed tool or platform provides new insights. Future studies should further investigate how new tools allow users to explore and learn from their own data. In addition, tools for self-exploration of personal genomics have the potential to help users make sense of their genomic data over time as the research that links this data to health outcomes evolves, and as users' personal circumstances change. Longitudinal studies are needed to examine participants' information needs and interactions over time and to understand how the use of personal genomic tools changes and how to support users over time.

6

Broader Implications

Making interaction with personal genomic information more accessible for non-experts, has broader implications for HCI researchers, practitioners, policymakers, and society.

6.1 For HCI researchers

Personal genomics presents a unique challenge to HCI scholars and practitioners: while many contributions to HCI and personal informatics focus on novel ways to represent constantly-changing data, personal genomics represents the opposite case: data is sampled only once and is largely static over the user's lifetime, but the interpretation of the data, as well as its importance, can change dramatically over time, with new medical research revealing hitherto unknown relations between genetic characteristics on one hand and physical attributes or medical conditions on the other. This combination of static information and dynamic interpretation calls for new HCI approaches and tools that allow capturing dynamically and constantly changing new knowledge about a stable data set. Such new tools may include dynamic curation of information related to genetic data, or adaptive decision support tools and

visualization, which change based on incoming new medical research. A related issue concerns the development of tools to facilitate sharing of this unique information–interpretation combination: future tools will need to facilitate sharing with relevant others (e.g. family members, health providers, or scientists) not only personal genomic information, but also its changing interpretations, which may require the attention — and in some cases action — of others. Finally, advances in HCI research may contribute to better communication of multidimensional uncertainty by exploring the potential use of interactive, customized, and adaptive design.

6.2 For practitioners

HCI research can also contribute to the work of practitioners in areas related to personal genomics by exploring what information users value and what information they are likely to be confused by. By studying information behavior in a personal genomics context, we can improve the ways in which genetic findings are communicated to users in clinical and non-clinical settings. Products and tools based on HCI research may also underpin the development of an ecosystem of commercial and non-commercial services based on added layers of information, and on connecting people with relevant information, with relevant service providers and with each other.

6.3 For society and policymakers

Better understanding and more effective engagement of users with their personal genomic information can help increase genetic and health literacy among non-experts. Moreover, rich and effective engagement may also lead to the further growth of burgeoning genomics-based citizen science, in which members of the public contribute to professional research projects as they collectively interpret and critique relevant information [Kuznetsov et al., 2015]. The accumulation of personal genomic data and analysis contributed by members of the public can support the work of professional scientists in diverse areas such as genetics, public health, and others.

HCI work on personal genomics can also inform public discourse about the relationship between people and their personal information, and in addition, between people and entities who can benefit from access to this personal information such as researchers, health-care providers, commercial companies, and public health organizations. As the scope of personal genomics grows and its effects on the public become clearer, HCI work can inform regulators, policymakers, and standard-setting institutions as they devise policies and regulations on how personal genomics should be presented to the individuals (1) statically, (2) interactively, and (3) dynamically. Recent events in which the FDA barred and then allowed personal genomics company 23andMe to offer personal genomic health-related reports to its users [Pollack, 2015] exemplify some of the sensitivities and complexities involved in how government can regulate the exposure of individuals to their personal genomic data, and how interpretation of such data can or should be provided to non-experts.

Finally, a societal aspect of the growing availability of DTCGT services is that their customer base tends to gravitate toward white and well-educated users [Roberts and Ostergren, 2013]. It is therefore our role as researchers to increase awareness and understanding of genetic information and its implications to populations that are under-represented in this user base. It should be also noted, that at a societal level, enabling wide-spread understanding of the commonalities — and in particular, genealogical similarities — among people from different ethnic backgrounds may contribute toward reducing racial categorization and bias.

Acknowledgements

We would like to thank Christina Pollalis and Clarissa Verish for their contribution. This work was partially funded by Grants IIS-1017693 and IIS-1422706 from the National Science Foundation (NSF), Division of Information and Intelligent Systems (IIS).

References

- 23andMe. accessed 29 July 2016. <https://www.23andme.com/>, 2016.
- P. W. Airasian, K. A. Cruikshank, R. E. Mayer, P. R. Pintrich, J. Raths, and M. C. Wittrock. A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives. In L. W. Anderson and D. R. Krathwohl, editors, . Pearson, 2001.
- AncestryDNA. accessed 29 July 2016. <https://www.ancestry.com/dna/>, 2016.
- M. Balestra, O. Shaer, J. Okerlund, M. Ball, and O. Nov. The effect of exposure to social annotation on online informed consent beliefs and behavior. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 900–912, 2016a.
- M. Balestra, O. Shaer, J. Okerlund, L. Westendorf, M. Ball, and O. Nov. Social annotation valence: The impact on online informed consent beliefs and behavior. *Journal of Medical Internet Research*, 18(7), 2016b.
- M. P. Ball, J. V. Thakuria, A. W. Zaranek, T. Clegg, A. M. Rosenbaum, X. Wu, M. Angrist, J. Bhak, J. Bobe, M. J. Callow, C. Cano, M. F. Chou, W. K. Chung, S. M. Douglas, P. W. Estep, A. Gore, P. Hulick, A. Labarga, J. Lee, J. E. Lunshof, B. C. Kim, J. Kim, Z. Li, M. F. Murray, G. B. Nilsen, B. A. Peters, A. M. Raman, H. Y. Rienhoff, K. Robasky, M. T. Wheeler, W. Vandewege, D. B. Vorhaus, J. L. Yang, L. Yang, J. Aach, E. A. Ashley, R. Drmanac, S. Kim, J. B. Li, L. Peshkin, C. E. Seidman, J. Seo, K. Zhang, H. L. Rehm, and G. M. Church. A public resource facilitating clinical use of genomes. *Proceedings of the National Academy of Science of the United States of America*, 109(30):11920–11927, 2012.

- M. P. Ball, J. R. Bobe, M. F. Chou, T. Clegg, P. W. Estep, J. E. Lunshof, W. Vandewege, A. Zaranek, and G. M. Church. Harvard personal genome project: Lessons from participatory public research. *Genome Medicine*, 6(2), 2014.
- F. Bentley, K. Tollmar, P. Stephenson, L. Levy, B. Jones, S. Robertson, E. Price, R. Catrambone, and J. Wilson. Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Transactions on Computer-Human Interaction*, 20(5), 2013.
- J. S. Berg, P. B. Agrawal, D. B. Bailey Jr., A. H. Beggs, S. E. Brenner, A. M. Brower, J. A. Cakici, O. Ceyhan-Birsoy, K. Chan, F. Chen, R. J. Currier, D. Dukhovny, R. C. Green, J. Harris-Wai, I. A. Holm, B. Iglesias, G. Joseph, S. F. Kingsmore, B. A. Koenig, P. Y. Kwok, J. Lantos, S. J. Leeder, M. A. Lewis, A. L. McGuire, L. V. Milko, S. D. Mooney, R. B. Parad, S. Pereira, J. Petrikin, B. C. Powell, C. M. Powell, J. M. Puck, H. L. Rehm, N. Risch, M. Roche, J. T. Shieh, N. Veeraraghavan, M. S. Watson, L. Willig, T. W. Yu, T. Urv, and A. L. Wise. Newborn sequencing in genomic medicine and public health. *Pediatrics*, 2017.
- B. Bettinger. What else can i do with my DNA test results? *The Genetic Genealogist*, accessed 29 July 2016, <http://thegeneticgenealogist.com/2013/09/22/what-else-can-i-do-with-my-dna-test-results/>, 2013.
- A. Blandford, D. Furniss, and S. Makri. *Qualitative HCI Research: Going Behind the Scenes*. Morgan & Claypool Publishers, 2016.
- F. Block, M. S. Horn, B. C. Phillips, J. Diamond, E. Evans, and C. Shen. The deeptree exhibit: Visualizing the tree of life to facilitate informal learning. *IEEE Transactions on Visualization and Computer Graphics*, 18(2):2789–2798, 2012.
- C. S. Bloss, L. Ornowski, E. Silver, M. Cargill, V. Vanier, N. J. Schork, and E. J. Topol. Consumer perceptions of direct-to-consumer personalized genomic risk assessments. *Genetics in Medicine*, 12(9):556–566, 2010.
- E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz. Understanding quantified-selfers’ practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1143–1152, 2014.
- E. K. Choe, B. Lee, M. Kay, W. Pratt, and J. A. Kientz. Sleptight: Low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 121–132, 2015.

- C. F. Chung, K. Dew, A. Cole, J. Zia, J. Fogarty, J. A. Kientz, and S. A. Munson. Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 770–786, 2016.
- G. M. Church. The personal genome project. *Molecular Systems Biology*, 1(1), 2005.
- ClinVar. *National Center for Biotechnology Information*. accessed 29 July 2016, <http://www.ncbi.nlm.nih.gov/clinvar/>, 2016.
- N. Colineau and C. Paris. Family vs. individual profiles in a health portal: Strengths and weaknesses. In *Proceedings of Conference on Human Computer Interaction*, pages 321–330, 2011.
- J. Conley. ACMG backs down a bit. *Genomics Law Report*, accessed 29 July 2016, <http://www.genomicslawreport.com/index.php/2014/04/07/acmg-backs-down-a-bit/>, 2014.
- S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. A. Landay. Activity sensing in the wild: A field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1797–1806, 2008.
- S. Das, A. Kramer, L. Dabbish, and J. Hong. The role of social influence in security feature adoption. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1416–1426, 2015.
- K. Davies. *The \$1,000 Genome: The Revolution In Dna Sequencing and the New Era of Personalized Medicine*. New York, NY: Free Press, 2010.
- dbSNP. RELEASE: NCBI dbSNP built 148. *National Center for Biotechnology Information*, accessed 29 July 2016, http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi, 2016.
- R. Dillet. Omixy improves medical check-up by making it as easy as calling an uber. *Tech Crunch*, accessed 29 July 2016, <https://techcrunch.com/2015/11/19/omixy-improves-medical-check-up-by-making-it-as-easy-as-ordering-an-uber/>, 2015.
- G. Dresden and A. Levitt. Modifying a standard industry clinical trial consent form improves patient information retention as part of the informed consent process. *Academic Emergency Medicine*, 8(3):246–252, 2001.
- J. T. Dudley and K. J. Karczewski. *Exploring Personal Genomics*. Oxford, UK: Oxford University Press, 2013.

- L. B. Dunn, L. A. Lindamer, B. W. Palmer, S. Golshan, L. J. Schneiderman, and D. V. Jeste. Improving understanding of research consent in middle-aged and elderly patients with psychotic disorders. *American Journal of Geriatric Psychiatry*, 10(2):142–150, 2002.
- EEOC. The genetic information nondiscrimination act of 2008. US Equal Employment Opportunity Commission, 2014.
- D. A. Epstein, A. Ping, J. Fogarty, and S. A. Munson. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 731–742, 2015.
- B. J. Evans. The first amendment right to speak about the human genome. *Journal of Constitutional Law*, 16(3):549–636, 2014.
- FamilyTreeDNA. accessed 29 July 2016. <https://www.familytreedna.com/>, 2016.
- FDA. Use of electronic informed consent in clinical investigations questions and answers, guidance for industry. *US Food and Drug Administration*, accessed 29 July 2016, <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm436811.pdf&sa=D&ust=1469801013237000&usg=AFQjCNFJI8sZHaeA-Fiu6Rk7LENpHRrwKw>, 2015.
- B. Friedman, E. Felton, and L. I. Millett. Informed consent online: A conceptual model and design principles. *University of Washington Computer Science & Engineering Technical Report 00-12-2*, 2000.
- Genetic Information Nondiscrimination Act. <https://www.eeoc.gov/laws/statutes/gina.cfm>, 2008.
- G. Gigerenzer, R. Hertwig, E. Van Den Broek, B. Fasolo, and K. V. Kat-sikopoulos. A 30% chance of rain tomorrow: How does the public understand probabilistic weather forecasts? *Risk Analysis*, 25(3):623–629, 2005.
- R. Gilbar. Communicating genetic information in the family: The familial relationship as the forgotten factor. *Journal of Medical Ethics*, 33(7):390–393, 2007.
- L. Goldsmith, L. Jackson, A. O’Connor, and H. Skirton. Direct-to-consumer genomic testing: Systematic review of the literature on user perspectives. *European Journal of Human Genetics*, 20(8):811–816, 2012.
- M. Greis, P. El.Agroudy, H. Schuff, T. Machulla, and A. Schmidt. Decision-making under uncertainty: How the amount of presented uncertainty influences user behavior. *Proceedings of the Nordic Conference on Human-Computer Interaction*, 2016.

- M. Greis, J. Hullman, M. Correll, M. Kay, and O. Shaer. Designing for uncertainty in HCI: When does uncertainty help? *Extended Abstracts on Human Factors in Computing Systems*, 2017.
- A. Grimes, D. Tan, and D. Morris. Toward technologies that support family reflections on health. In *Proceedings of the ACM International Conference on Supporting Group Work*, pages 311–320, 2009.
- M. Gymrek, A. L. McGuire, D. Golan, E. Halperin, and Y. Erlich. Identifying personal genomes by surname inference. *Science*, 339(6117):321–324, 2013.
- V. Hughes. It’s time to stop obsessing about the dangers of genetic information. Slate, Date accessed 29 July 2016, URL: http://www.slate.com/articles/health_and_science/medical_examiner/2013/01/ethics_of_genetic_information_whole_genome_sequencing_is_here_and_we_need.html, 2013.
- O. Inbar. Graphical representation of statistical information in situations of judgment and decision-making. In *Proceedings of 14th European Conference on Cognitive Ergonomics*, pages 265–268, 2007.
- S. Joslyn and J. LeClerc. Decisions with uncertainty: The glass half full. *Current Directions in Psychological Science*, 22(4):308–315, 2013.
- S. Joslyn and S. Savelli. Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteorological Applications*, 17(2):180–195, 2010.
- S. Joslyn, L. Nemeč, and S. Savelli. The benefits and challenges of predictive interval forecasts and verification graphics for end-users. *Weather, Climate & Society*, 5(2):133–147, 2013.
- M. F. Jung, D. Sirkin, T. M. Gür, and M. Steinert. Displayed uncertainty improves driving experience and behavior: The case of range anxiety in an electric car. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2201–2210, 2015.
- K. J. Karczewski, R. P. Tirrell, P. Cordero, N. P. Tatonetti, J. T. Dudley, K. Salari, M. Snyder, R. B. Altman, and S. K. Kim. Interpretome: A freely available, modular, and secure personal genome interpretation engine. *Pacific Symposium on Biocomputing*, pages 339–350, 2012.
- J. Karow. First newborn sequencing study gets FDA green light while others still await approval. *GenomeWeb*, accessed 1 January 2017, 2014.
- M. Kay, D. Morris, M. C. Schraefel, and J. A. Kientz. There’s no such thing as gaining a pound: Reconsidering the bathroom scale user interface. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 401–410, 2013.

- M. Keshavan. Made in China: Direct-to-consumer whole genome sequencing for \$875. *MedCity News*, accessed 1 January 2017, <http://medcitynews.com/2016/01/made-in-china-direct-to-consumer-whole-genome-sequencing-for-895/>, 2016.
- R. Khamisi. Can gut DNA sequencing actually tell you anything about your health? *Newsweek*, accessed 29 July 2016, <http://www.newsweek.com/2014/07/25/can-gut-dna-sequencing-actually-tell-you-anything-about-your-health-259348.html>, 2014.
- P. Klasnja and W. Pratt. Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of Biomedical Informatics*, 45 (1):184–198, 2012.
- S. M. Kneipp, B. J. Lutz, and D. Means. Reasons for enrollment, the informed consent process, and trust among low-income women participating in a community-based participatory research study. *Public Health Nursing*, 26 (4):362–369, 2009.
- S. Kuznetsov, A. Kittur, and E. Paulos. Biological citizen publics: Personal genetics as a site of public engagement with science. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, pages 303–312, 2015.
- C. R. Lachance, L. A. H. Erby, B. M. Ford, V. C. Allen, and K. A. Kaphingst. Informational content, literacy demands, and usability of websites offering health-related genetic tests directly to consumers. *Genetics in Medicine*, 12(5):304–312, 2010.
- P. Lamkin. Personal genome testing: The ultimate quantified self. Wareable, accessed 29 July 2016, <http://www.wareable.com/health-and-wellbeing/personal-genome-testing-the-ultimate-quantified-self-2298>, 2016.
- J. LeClerc and S. Joslyn. Odds ratio forecasts increase precautionary action in severe weather events. *Weather, Climate & Society*, 4:263–270, 2012.
- M. K. Lee, J. Kim, J. Forlizzi, and S. Kiesler. Personalization revisited: A reflective approach helps people better personalize health services and motivates them to increase physical activity. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 743–754, 2015.
- I. Li, A. Dey, and J. Forlizzi. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 557–566, 2010.

- I. Li, A. Dey, J. Forlizzi, K. Höök, and Y. Medynskiy. Personal informatics and HCI: design, theory, and social implications. In *Extended Abstracts on Human Factors in Computing Systems*, pages 2417–2420, 2011.
- J. J. Lin, L. Mamykina, S. Lindtner, G. Delajoux, and H. B. Strub. Fish'n'steps: Encouraging physical activity with an interactive computer game. In P. Dourish and A. Friday, editors, *The 8th International Conference on Ubiquitous Computing*, pages 261–278. Springer, 2006.
- M. D. Linderman, D. E. Nielsen, and R. C. Green. Personal genome sequencing in ostensibly healthy individuals and the peopleSeq consortium. *Journal of Personalized Medicine*, 6(2):14, 2016.
- J. E. Lunshof, R. Chadwick, D. B. Vorhaus, and G. M. Church. From genetic privacy to open consent. *Nature Reviews Genetics*, 9(5):406–411, 2008.
- L. Mamykina, E. Mynatt, P. Davidson, and D. Greenblatt. MAHI: Investigation of social scaffolding for reflective thinking in diabetes management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 477–486, 2008.
- T. A. Manolio, R. L. Chisholm, B. Ozenberger, D. M. Roden, M. S. Williams, R. Wilson, D. Bick, E. P. Bottinger, M. H. Brilliant, C. Eng, K. A. Frazer, B. Korf, D. H. Ledbetter, J. R. Lupski, C. Marsh, D. Mrazek, M. F. Murray, P. H. O'Donnell, D. J. Rader, M. V. Relling, A. R. Shuldiner, D. Valle, R. Weinshilboum, E. D. Green, and G. S. Ginsburg. Implementing genomic medicine in the clinic: The future is here. *Genetics in Medicine*, 15(4):258–267, 2013.
- R. Manshaei, N. Baig, S. DeLong, S. Khayyer, B. East, and A. Mazalek. Exploring genetic mutations on mitochondrial DNA cancer data with interactive tabletop and active tangibles. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*, pages 461–464, 2016.
- M. L. McGowan, J. R. Fishman, and M. A. Lambrix. Personal genomics and individual identities: Motivations and moral imperatives of early users. *New Genetics and Society*, 29(3):261–290, 2010.
- M. Molteni. Genos will sequence your genes — and help you sell them to science. *Wired*, accessed 1 January 2017, <https://www.wired.com/2016/12/genos-will-sequence-genes-help-sell-science/>, 2016.
- R. E. Morss, J. L. Demuth, and J. K. Lazo. Communicating uncertainty in weather forecasts: A survey of the U.S. public. *Weather and Forecasting*, 23(5):974–991, 2008.

- D. A. Murphy, D. Hoffman, G. R. Seage, M. Belzer, J. Xu, S. J. Durako, and M. Geiger. Improving comprehension for HIV vaccine trial information among adolescents at risk of HIV. *AIDS Care*, 19(1):42–51, 2007.
- L. Nadav-Greenberg, S. L. Joslyn, and M. U. Taing. The effect of uncertainty visualizations on decision making in weather forecasting. *Journal of Cognitive Engineering and Decision Making*, 2(1):24–47, 2008.
- Nature. Genetic privacy. *Nature*, 493:451, 2013.
- NHGRI. The cost of sequencing a genome. *National Human Genome Research Institute*, accessed 29 July 2016, <https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/>, 2016.
- NIH. Precision medicine initiative cohort program. *National Institute of Health*, accessed 29 July 2016, <https://www.nih.gov/precision-medicine-initiative-cohort-program>, 2016.
- Open Humans. Research and data. OpenHumans, accessed 29 July 2016, <https://www.openhumans.org/research/>, 2016.
- J. E. Ostergren, M. C. Gornick, D. A. Carere, S. S. Kalie, W. R. Uhlmann, M. T. Ruffin, J. L. Mountain, R. C. Green, J. S. Roberts, and PGen Study Group. How well do customers of direct-to-consumer personal genomic testing services comprehend genetic test results? *Public Health Genomics*, 18(4):216–224, 2015.
- R. A. Patel, P. Klasnja, A. Hartzler, K. T. Unruh, and W. Pratt. Probing the benefits of real-time tracking during cancer care. In *AMIA Annual Symposium Proceedings*, pages 1340–1349, 2012.
- B. A. Peters, B. G. Kermani, A. B. Sparks, O. Alferov, P. Hong, A. Alexeev, Y. Jiang, F. Dahl, Y. T. Tang, J. Haas, K. Robasky, A. W. Zaranek, J. H. Lee, M. P. Ball, J. E. Peterson, H. Perazich, G. Yeung, J. Liu, L. Chen, M. I. Kennemer, K. Pothuraju, K. Konvicka, M. Tsoupko-Sitnikov, K. P. Pant, J. C. Ebert, G. B. Nilsen, J. Baccash, A. L. Halpern, G. M. Church, and R. Drmanac. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. *Nature*, 487(7406):190–195, 2012.
- PGen. Accessed 31 January 2016. <http://www.genomes2people.org/pgen/>, 2016.
- A. M. Phillips. Genomic privacy and direct-to-consumer genetics: Big consumer genetic data — what’s in that contract? *2015 IEEE Security and Privacy Workshops*, pages 60–64, 2015.

- L. R. Pina, S. W. Sien, T. Ward, J. C. Yip, S. A. Munson, J. Fogarty, and J. A. Kientz. From personal informatics to family informatics: Understanding family practices around health monitoring. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 2017.
- S. Pinker. My genome, my self. *New York Times*, accessed 29 July 2016, <http://www.nytimes.com/2009/01/11/magazine/11Genome-t.html>, 2009.
- A. Pollack. F.D.A. orders genetic testing firm to stop selling DNA analysis service. *New York Times*, accessed 29 July 2016, http://www.nytimes.com/2013/11/26/business/fda-demands-a-halt-to-a-dna-test-kits-marketing.html?_r=1, 2013.
- A. Pollack. 23andme will resume giving users health data. *New York Times*, accessed 29 July 2016, <http://www.nytimes.com/2015/10/21/business/23andme-will-resume-giving-users-health-data.html>, 2015.
- Quantified Self. Accessed 29 July 2016. <http://quantifiedself.com>, 2016.
- A. Regalado. Q&A: Stephen Wolfram on personal analytics. *MIT Technology Review*, 116(4), 2013.
- A. Regalado. For \$999, Veritas Genetics will put your genome on a smartphone app. *MIT Technology Review*, accessed 1 January 2017, <https://www.technologyreview.com/s/600950/for-999-veritas-genetics-will-put-your-genome-on-a-smartphone-app/>, 2016.
- J. S. Roberts and J. Ostergren. Direct-to-consumer genetic testing and personal genomics services: A review of recent empirical studies. *Current Genetic Medicine Reports*, 1(3):182–200, 2013.
- E. M. Rogers. *Diffusion of Innovations*. 5th edition, Free Press, 2003.
- J. Rooksby, M. Rost, A. Morrison, and M. Chalmers. Personal tracking as lived informatics. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, pages 1163–1172, 2014.
- E. Rosas-Blum, P. Shirsat, and M. Leiner. Communicating genetic information: A difficult challenge for future pediatricians. *BMC Medical Education*, 7(17), 2007.
- M. S. Roulston, G. E. Bolton, A. N. Kleit, and A. L. Sears-Collins. A laboratory study of the benefits of including uncertainty information in weather forecasts. *Weather and Forecasting*, 21(1):116–122, 2006.

- S. C. Sanderson, M. D. Linderman, S. A. Suckiel, G. A. Diaz, R. E. Zinberg, K. Ferryman, M. Wasserstein, A. Kasarskis, and E. E. Schadt. Motivations, concerns and preferences of personal genome sequencing research participants: Baseline findings from the healthseq project. *European Journal of Human Genetics*, 24(1):14–20, 2016.
- S. Savelli and S. Joslyn. Boater safety: Communicating weather forecast information to high stakes end users. *Weather, Climate & Society*, 4:7–19, 2012.
- S. Schkolne, H. Ishii, and P. Schroder. Immersive design of DNA molecules with a tangible interface. In *Proceedings of the Conference on Visualization*, pages 227–234, 2004.
- B. Schneider, M. Strait, L. Muller, S. Elfenbein, O. Shaer, and C. Shen. Phylogenie: Engaging students in collaborative ‘tree-thinking’ through tabletop techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3071–3080, 2012.
- D. Seetharaman. Jessica Richman seeks samples to crack the human microbiome. *Wall Street Journal*, accessed 29 July 2016, <http://www.wsj.com/articles/jessica-richman-seeks-samples-to-crack-the-human-microbiome-1445452170>, 2015.
- O. Shaer and O. Nov. HCI for personal genomics. *Interactions*, 21(5):32–37, 2014.
- O. Shaer, M. Strait, C. Valdes, T. Feng, M. Lintz, and H. Wang. Enhancing genomic learning through tabletop interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2817–2826, 2011.
- O. Shaer, A. Mazalek, B. Ullmer, and M. Konkel. From big data to insights: opportunities and challenges for TEI in genomics. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, pages 109–116, 2013.
- O. Shaer, O. Nov, A. West, and D. Eastman. Understanding information practices of interactive personal genomics users. In *Extended Abstracts on Human Factors in Computing Systems*, pages 2455–2460, 2014.
- O. Shaer, O. Nov, J. Okerlund, M. Balestra, E. Stowell, L. Ascher, J. Bi, C. Schlenker, and M. Ball. Informing the design of direct-to-consumer interactive personal genomics reports. *Journal of Medical Internet Research*, 17(6), 2015.

- O. Shaer, O. Nov, J. Okerlund, M. Balestra, E. Stowell, L. Westendorf, C. Pol-lalis, J. Davis, L. Westort, and M. Ball. Genomix: A novel interaction tool for self-exploration of personal genomic data. In *Proceedings of the 2016 Conference on Human Factors in Computing Systems*, pages 661–672, 2016.
- M. Skeels, B. Lee, G. Smith, and G. Robertson. Revealing uncertainty for information visualization. In *Proceedings of Advanced Visual Interfaces*, pages 376–379, 2008.
- P. G. Stiles, N. G. Poythress, A. Hall, D. Falkenbach, and R. Williams. Improving understanding of research consent disclosures among persons with mental illness. *Psychiatric Services*, 52(6):780–785, 2001.
- M. Swan. Emerging patient-driven health care models: An examination of health social networks, consumer personalized medicine and quantified self-tracking. *International Journal of Environmental Research and Public Health*, 6(2):492–525, 2009.
- B. N. Taylor and C. E. Kuyatt. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. NIST Technical Note 1297, 1994.
- J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel. Typology for visualizing uncertainty. In *Proceedings of the IS&T/SPIE Symposium on Electronic Imaging, Conference on Visualization and Data Analysis*, pages 146–157, 2005.
- E. Troshynski, C. Lee, and P. Dourish. Accountabilities of presence: Reframing location-based systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 487–496, 2008.
- M. Turrini and B. Prainsack. Beyond clinical utility: The multiple values of DTC genetics. *Applied & Translational Genomics*, 8(1):4–8, 2016.
- A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.
- E. Vayena, E. Gourn, J. Streuli, E. Hafen, and B. Prainsack. Experiences of early users of direct-to-consumer genomics in switzerland: An exploratory study. *Public Health Genomics*, 15(6):352–362, 2012.
- J. Woollen, J. Prey, L. Wilcox, A. Sackeim, S. Restaino, S. T. Raza, S. Bakken, S. Feiner, G. Hripcsak, and D. Vawdrey. Patient experiences using an inpatient personal health record. *Applied Clinical Informatics*, 7(2):446–460, 2016.
- A. Wu, J. B. Yim, E. Caspary, A. Mazalek, S. Chandrasekharan, and N. J. Nersessian. Kinesthetic pathways: A tabletop visualization to support discovery in systems biology. In *Proceedings of the 8th ACM Conference on Creativity and Cognition*, pages 21–30, 2011.