GenomiX: A Novel Interaction Tool for Self-Exploration of Personal Genomic Data

Orit Shaer[1], Oded Nov[2], Johanna Okerlund[1], Martina Balestra[2], Elizabeth Stowell[1], Lauren Westendorf[1], Christina Pollalis[1], Jasmine Davis[1], Liliana Westort[1], Madeleine Ball[3]

HCI Lab[1]
Wellesley College
Wellesley, MA, USA

New York University[2]
New York City, NY, USA

PersonalGenomes.org[3]
Brookline, MA, USA

ABSTRACT

The increase in the availability of personal genomic data to lay consumers using online services poses a challenge to HCI researchers: such data are complex and sensitive, involve multiple dimensions of uncertainty, and can have substantial implications for individuals' well-being. Personal genomic data are also unique because unlike other personal data, which constantly change, 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. In this paper, we present a novel tool for self exploration of personal genomic data. To evaluate the usability and utility of the tool, we conducted the first study of a genome interpretation tool to date, in which users used their own personal genomic data. We conclude by offering design implications for the development of interactive personal genomic reports.

Author Keywords

Personal genomics; personal informatics; uncertainty.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g. HCI): User Interfaces

INTRODUCTION

People's access to their own personal genomic data is rapidly increasing [8] as the cost of sequencing a human genome has fallen from approximately $100 million in 2001 to a little over $5,000 in 2015 [13], a rate much higher than Moore's Law [21]. This has rapidly accelerated the emergence of genome mapping research and commercial services, but consumers' abilities to properly interpret such data remains poor [23].

Individuals with no formal training in genetics are now able to acquire their genomic information without the involvement of a healthcare provider by sending a saliva sample to any of a number of new online direct-to-consumer genetic testing (DTCGT) services at the cost of about hundred dollars. Results are delivered via online reports and raw data are available for download. Users must then interpret a tremendous amount of data involving delicate topics like disease risk and carrier status without the guidance of a genetic counselor. The interpretation of the data may in turn influences their lifestyle decisions, emotional state, and wellbeing. To add to this complexity, the relationships between genes and disease outcomes are not currently well understood, and evolve dynamically with the development of new technologies, processes, and research results, requiring consumers to continuously reconsider their results against the most current evidence. Such confluence of trends calls for HCI researchers and practitioners to develop novel ways to help users engage with their personal genomic data.

The highly personal and dynamic nature of genomic data raises important HCI questions, including: what are the functional requirements for supporting meaningful engagement of consumers 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? Though prior studies have sought to answer these questions in the context of other types of personal informatics, we assert that interaction with personal genomic data is unique. In other forms of personal data, the dynamic element is 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 [18]. Genomic data, on the other hand, are largely stable during a person's lifetime. The certainty of the evidence, its interpretation, and related implications for the user's health, however, often change over time as new medical research exposes new relationships between people's genetic make up and their health.

In this paper, we present a design case study of a novel interactive tool, named GenomiX, aimed at empowering users as they engage with their own genomic information. We describe the conceptualization and design of GenomiX, in which we draw upon findings from previous studies examining how users perceive, interact with, and explore

CHI'16, May 07-12, 2016, San Jose, CA, USA
 ISBN 978-1-4503-3362-7/16/05…$15.00
DOI: http://dx.doi.org/10.1145/2858036.2858397

complex personal genomics data [32]. We present findings from the evaluation of GenomiX with consumers who had previously participated in personal genome mapping studies, and who had already used the industry state-of-the-art reports to view their results. This particular user group was sought out deliberately to evaluate the *ongoing* interpretation of genomic data, and to understand what *new* insights participants are able to form about their data compared with the existing tools and reports.

A novel aspect of our work is the context of consumer-facing genomic reporting. Related HCI research on personal informatics focuses on synthesizing and communicating relevant information succinctly and to highlight long-term trends in emergent data [18]. In contrast, our study focuses on the dynamic nature of the data *interpretation*. We propose methods for highlighting the most significant interpretation of the results according to the most up-to-date genomics research so that individuals can take action. We also give users the ability to save information so that they can revisit the original data and define new questions about their own health. To date, little HCI research has focused on dynamically changing interpretation in general, and on direct lay-user engagement with personal genomic information in particular.

A second novel aspect of this work lies in the visualization of uncertainty in the presentation of personal genomic results. While the majority of existing studies on visualizing uncertainty focus on its role in decision-making [12], fewer studies exist on the impact of uncertainty in personal data exploration. To the best of our knowledge, this is the first study to suggest a tangible design intervention for incorporating uncertainty into a personal genome tool.

The method of evaluating GenomiX represents a third contribution of this study. Prior studies on user-facing genomics tools use fake or anonymized data; that is, genomic test results that do not belong to the participant. In doing so, these studies fail to incorporate the impact and meaning of the data to the user. To the best of our knowledge, this is the first study to design and evaluate a visual representation of genomic data using participants' own personal data. Furthermore, these participants are already familiar with their data. Existing studies tend to use novice participants who are unfamiliar with genomic data to evaluate visualizations. Because of this sample population, most studies are unable to look at participants' *evolving* understanding of their genome, or how well the platform provides new insight.

BACKGROUND

Personal Genome Project

We established a design partnership with the Personal Genome Project (PGP) and have collaborated closely with its researchers on this and on related projects. The PGP [25] is a research study, established in 2005 out of George Church's Lab at Harvard Medical School, seeking to improve the scientific understanding of genetic and environmental contributions to human traits through the creation of a public genomic database of 100,000 volunteers [2, 3, 7]. Volunteers agree to share their genomic sequences, as well as health data, with the scientific community and the public. Today, more than 4000 volunteers are enrolled in the project through a process of "open consent" [19] to share their genomic information publicly. Those who participate in the PGP study have access to the state-of-the-art genomic report GET-Evidence [24], to navigate their results. While other DTCGT providers exist (e.g. 23andMe), they are not yet able to distribute health-related variants reports (pending FDA review [10]). The PGP GET-Evidence report is presented with emphatically non-clinical usage. The purpose of the PGP GET-Evidence report is to inform participants deciding whether to make their genomic data public. The report presents detailed information in a tabular design, including a list of gene variants reported to cause particular conditions or traits, and the frequency of each variant in the population. For each gene variant the report presents: potential impact and the certainty of that impact (e.g. well-established pathogenic, likely protective, uncertain benign); clinical importance (i.e. low, medium, or high); and a summary describing the current knowledge about a variant. It should be noted that PGP GET-Evidence reports do not interpret the user's genome; rather, they display information and require the individual's own sense making. Figure 1 provides a screenshot of the PGP GET-Evidence report.

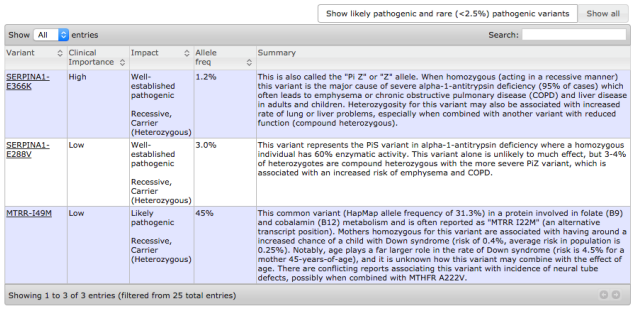

Figure 1. GET-Evidence report.

RELATED WORK

HCI For Genomics

There are a number of studies that investigate the motivation for and subjective experience of genetic testing, and of using interactive tools to understand results (e.g. [26] [10] [8]). However, these studies tend not to look at the relationship between this experience and specific design interventions. Direct lay-user engagement with personal genomic information has been relatively understudied in the HCI field. Existing research tends to focus on participants' comprehension of anonymous reports from a variety of perspectives, or on exploring novel interaction techniques for manipulating large volumes of biological data. Lachance et al. [16] examined the informational content, literacy demands, and usability of DTCGT service websites. They find that websites vary widely, and most users would struggle to use these resources effectively. The

authors suggest 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. [23] assess participants' comprehension of anonymized genomic reports and find that comprehension varies widely according to demographic characteristics, numeracy and genetic knowledge, and types and format of the genetic information presented. They suggest that the presentation of genomic data be tailored to the test type and customer characteristics.

In contrast to the studies that present users with anonymized genetic data, Kuznetsov et al. [15] present 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. Consumers are framed as members of *biocitizen publics* in which there is an emphasis on individuals' engagement with the community and higher order learning processes [1], 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. Our vision is consistent with Kuznetsov et al.'s as we work toward such a system by focusing first on the interaction with and visualization of the data.

Other studies have developed ways of interacting with large-scale and complex biological datasets and use them as a platform to explore novel interaction techniques, such as tangible interaction [31]. Systems developed include a tangible interface for designing new DNA molecules [26], and several tabletop interfaces for interactive visualization of biological datasets, such as DeepTree [4] and PhyloGenie [29]. G-nome Surfer [30] is a tabletop interface for collaborative exploration of genomes; however, it was not designed to support users in the exploration of their own genomic data.

Evaluation of tools in most studies use conventional methods in which novice participants are presented with anonymous data, and they tend to measure response precision, error rates, number of correct and incorrect responses, and measures of time to complete predefined benchmark tasks. While these are important facets of interacting with online reports, these methodologies do not capture evolving insights or the perspective of users who revisit their own data over time.

Representing Uncertainty

Abundance of work has investigated the visualization of uncertain information. Existing taxonomies for communicating uncertainty identify sources of uncertainty (and visual presentation techniques (e.g. [33], [35], [20], [34]). Additional work explores cognitive biases of decision-making under uncertainty and corrective visual approaches (e.g. [36], [14]). Numerous applications tracking new types of personal and often uncertain data have explored how to present the data to encourage behavior change and reflection [27, 6, 17, 9]. In a study comparing visualizations of uncertainty, Greis et al. [11] find that participants' judgment of these visualizations were significantly influenced by familiarity, ease of understanding, and visual appeal. Nadav-Greenberg et al. [22] compared the impact of various representations of uncertainty on different activities, concluding that different types of visualizations lead to different learning outcomes and suggest that an interactive display may be best for communicating uncertain information. However, the personal genomic context, which we investigate in this paper, offers a form of uncertainty not addressed by existing taxonomies and applications. In the genomic 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 novel technologies and new scientific findings. We seek to develop novel ways of representing this uncertainty.

GenomiX: A NEW INTERACTIVE TOOL FOR EXPLORING PERSONAL GENOMICS

GenomiX is a novel visual tool we developed that supports self-exploration of personal genomic data. GenomiX enhances learning and discovery by providing new representations and mechanisms for organizing, interacting, and curating personal genomic data.

It is important to note that GenomiX does not provide new genome interpretations but rather draws upon the interpretation provided by the PGP, which serves as the basis for the GET-Evidence report. However, by presenting a visual summary, communicating uncertainty, and allowing users to interact with their data in new ways, GenomiX empowers individuals to discover new insights from their genomic data.

Requirements and Design Goals

The requirements and design goals of GenomiX draw upon our previous research exploring users' motives, needs, and interaction patterns with genomic data [32]. In that study, we surveyed 63 participants from the Personal Genome Project interacting with their personal genomic data. User needs were synthesized and 6 specific functional requirements for future personal genomic tools were identified:

R1) Reviewing an annotated report - Participants described the difficulty of interpreting existing tabular and dense textual reports. They expressed a desire for visualizations that make the information easier to explore and understand.

R2) Integrating resources - Participants expressed a need for integrating various data resources, including annotated genomes and scientific publications.

R3) Curating information - Participants articulated a need for collecting, relating, and storing information artifacts.

R4) Making content accessible - Participants indicated a need for adapting the content and language of personal genomic reports toward consumers.

R5) Comparing genomes - Participants asked for the ability to triangulate data from several individuals in order to understand connections within families.

R6) Facilitating sharing information - Participants highlighted a need for tools that facilitate information sharing with family, friends, and genetic researchers.

In a second part of the study [32], which addressed R1-R4 as the most substantial requirements, we interviewed and observed 36 participants as they explored their personal genomic data using the GET Evidence tool. This study deepened the understanding of the needs and practices of personal genomic consumers, highlighting that users are predominantly concerned with genetic variants that are well-established, pathogenic, and have high clinical importance. Finally, the third part of that study investigated the effect of different visualizations on consumers' understanding of personal genomic data. These findings indicated an advantage to non-zoomable visualizations, with best results (in terms of both objective comprehension and subjective preference) using bubble graphs.

Drawing upon these findings, we defined new design 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.

We designed GenomiX to realize these goals. In the sections that follow, we outline the implementation, design, and functionality of the tool.

Implementation

GenomiX was developed as a web application using JavaScript with D3.js. Personal data was loaded into the visualization from a repository of genome reports hosted on our server. We generated this repository prior to users' participation in the study by scraping GET-Evidence reports of PGP participants, which are available publicly online. We created a JSON file for each person. This was done so that we would not have to scrape the online GET-Evidence reports in real time. GenomiX is also connected to a MySQL database that logs participants' actions.

Interaction Overview

When using GenomiX, the user is first prompted to input their PGP ID. The user is then presented with a visualization providing an overview of their genetic variant data (Figure 2). Gene variants are represented as bubbles that are plotted between two axes, and the size, color, and placement communicate specific information about that variant. Using controls on the left hand side of the screen, individuals can sort the data on the plot by either risk or rarity of the variant. At the top of the page, participants can click on the "categories" tab to sort the variants according to the anatomical system impacted by the variant. Users can sort these variants, like in the overview tab, according to their risk and rarity. Finally, users can click on a tab that leads to a glossary of terms.

The user therefore interacts with the tool by: 1) exploring alternative views of the information by switching back and forth between the *overview* and the *health categories* tabs; 2) selecting a variant for viewing additional information; 3) saving variants for further exploration; 4) sorting and changing the way variants are represented and organized; and 5) consulting the glossary or information buttons to learn about the terminology used.

Representation of the Variants

A key on the left side of the screen shows how graphical elements of the visualization map to the characteristics of a gene variant:

Color

The color of a bubble (a variant) represents its potential effect (*pathogenic*, *benign*, *protective*, or *pharma*). *Pathogenic* variants indicate increased risk for a disease and are therefore mapped to the color red to indicate "danger". *Protective* variants decrease the risk for a disease, and are therefore colored in blue. *Benign* variants have no health effect, and are therefore represented in neutral gray. This choice of 3-color scale is based on results from our previous studies [32]. *Pharma* variants have an effect on how one responds to certain medicines. These variants do not fit on the scale between protective, benign, and pathogenic, and are thus represented using purple.

Fill

Hollow bubbles represent variants that the user is a carrier for. These variants will not manifest in the user, but could be passed on to their children. A filled bubble represents variants that could affect the user directly.

Size

The size of a variant represents two different variables: *risk* and *rarity*, depending on the choice of the user. Users can toggle between these variables using a radio button. When size represents risk, larger bubbles indicate that an individual is at a higher risk of being affected by the associated condition. When size represents rarity, larger

circles represent variants that are rarer in the population. Users may want to pay attention to these rare variants because they are likely to be less understood, and may require more careful evaluation.

Spatial Organization

Variants are plotted according to the *certainty* of the scientific evidence that associates a variant with a particular condition or trait (well-established, likely, uncertain) and the potential *health effect* (low, medium, high). Health effect is a number calculated by PGP based on the treatability and severity associated with a variant. The plot of certainty by health effect therefore has 9 cells: well-established certainty, low health effect; well-established certainty, medium health effect; well-established certainty, high health effect, etc.

Categories

The *Category* report (see Figure 3) organizes variants according to the system it is related to (e.g. metabolism, immune system). Since variants can be associated with more than one system, multiple copies of the same variants could appear in different categories. When a particular variant copy is selected, all copies of that variant in different systems are highlighted.

Detailed Variant Information

When selecting a variant, a detailed and up-to-date summary of current knowledge about the variant is presented to the right of the main workspace. Users can save that variant and the associated summary. Saved variant information can be expanded and contracted, and persists across sessions. This feature allows users to make note of variants that they want to explore further.

Glossary And Additional Information

Each label in the report has an associated info button that provides additional information about the terminology used. Information is presented with a semi-transparent background to maintain context. We also provided a glossary that allows users to look for related terms.

EVALUATION

We evaluated GenomiX using an exploratory qualitative methodology. We drew upon insight based evaluation [28] to understand GenomiX as an interpretive and reflective tool. In particular we were interested in finding out:

1. What insights do users gain about the data from interacting with the tool? What do they learn about their data that they have not learned before?
2. What design features support users in gaining insights?
3. How can the tool be improved?

Sample

Participants were recruited from the PGP participant community, and were separately enrolled in our study. An email soliciting participation was sent to 200 qualifying individuals who had their entire genome sequenced (representing ~98% of genome sequence) and received a GET-Evidence report. Interested participants clicked on a link in the email, where they joined our study through an online informed consent form. Working with Harvard PGP allowed us access to a unique set of participants. While many genomes have been sequenced by the research community, very few individuals have access to their personal whole-genome data [5]. Harvard PGP participants have consented through a "highly informed" process [3] not only to have access to their data but also to make it publicly available through PGP. As part of this consent process, PGP participants have studied their data using GET-Evidence report and other tools. From these pre-existing public materials, our study was able to present individuals, who volunteered and gave additional consent to participate in this study, with a new interactive visualization of information they had previously received and reviewed.

Procedure

Participants were instructed to first provide demographic information and prior tool usage. They were then able to view and interact with their own data visualized using GenomiX. After exploring their data, participants completed an online questionnaire consisting of 14 questions measuring their perceptions of the new tool using a series of 5-point Likert scales (Table 2). Participants were also presented with 6 open-ended questions (Table 1) on their engagement with personal genomics services and data.

Measures and Indicators

To understand what and how users learned using GenomiX, as well as to assess their engagement and perception of this tool, we looked at various measures and indicators:

Time On Task

Study instructions did not require or limit users to a particular timeframe. Rather, we asked users to use this new tool to study their own genome. Users were free to complete the study at any point. Thus, we consider *time on task* to be a measure of engagement rather than efficiency. Time on task was measured using time stamps.

Insights

To define an insight, we draw on Saraiya et al. [28] who view insight as "an individual observation about the data by the participant, a unit of discovery" (p. 444). They group bioinformatics insights into four categories: overview (overall distribution), patterns (identification or comparison across data attributes), groups (identification of comparison of groups of entities), and details (focused information about a specific entity). In our study, we asked participants to answer three open questions (See Table 1 Q1, Q2, Q6) about learning and discovery using GenomiX, allowing users to share insights from all four categories.

Usage

To study how users used the tool, we collected information about which features were used, for how long, and in what order. Information was collected using automatic logging.

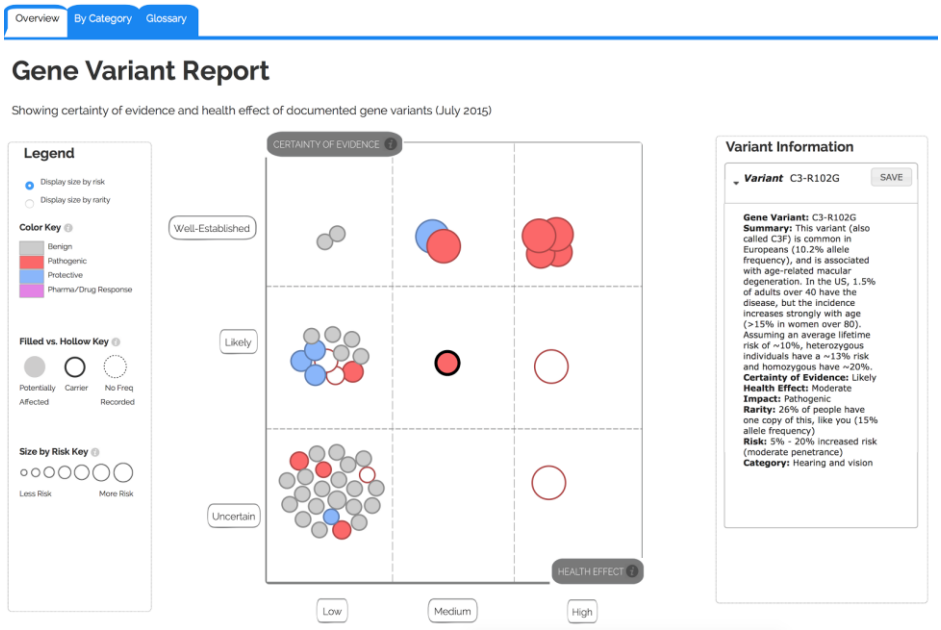

Figure 2. GenomiX: Gene Variant Report displaying the overview of a participant's results

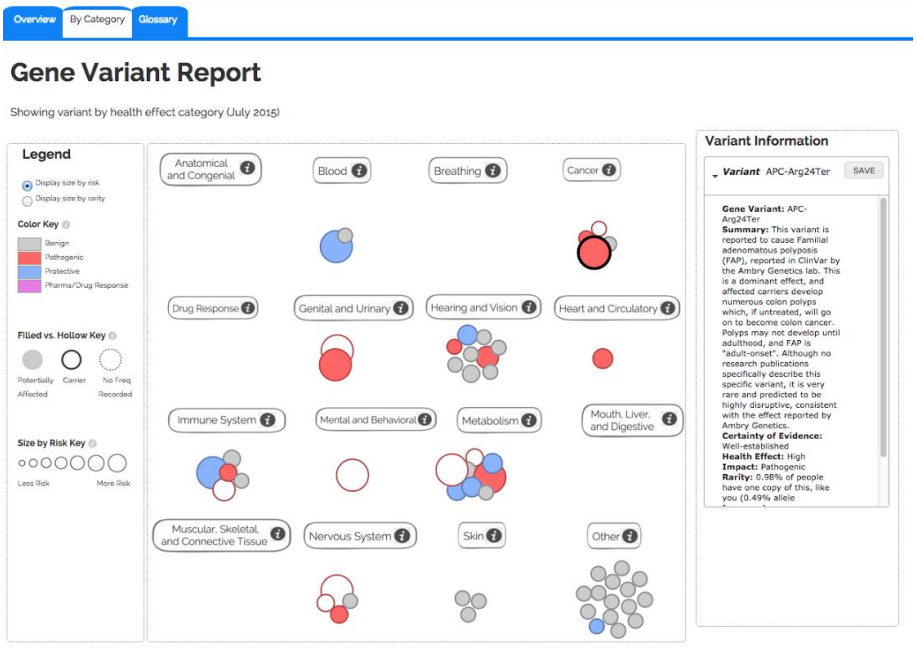

Figure 3. GenomiX: Gene Variant Report displaying participant's results sorted by category

Understandability

To assess to what extent users find the personal genomic information as presented using GenomiX understandable, we asked users to rate their agreement with a set of statements using a 5-point Likert scale (see Table 2 Q6-Q9, Q11-Q14, Q16)

Usability and Usefulness

Similarly, to assess usability and usefulness of GenomiX, we asked users to rate their agreement with a set of statements using a 5-point Likert scale (see Table 2 Q8, Q10, Q17-Q19). We also asked users which design features they found particularly useful, and, what aspects of the tools could be improved (Q3-Q4, Table 1).

Data Analysis

We analyzed the data using content analysis methods. First-level codes were developed from preliminary review by two independent coders and were then collapsed into categories based on frequency. Categories were analyzed and themes were identified. Responses to the open questions averaged

36.9 (SD=43.8) words per user. Inter-code reliability based on 30% of the data was good at 86.5%.

RESULTS

Participants

We recruited 74 participants (28 women, 36.4%), between ages 25 and 80 with average age of 51.2 years (SD=14.91) from the Harvard PGP volunteer community. 12.2% of participants (9/74) reported having some college education, 24.3% (18/74) participants had received a bachelor's degree, 27.0% (20/74) participants had received a master's degree, and 36.5% (27/74) participants had received a doctoral degree. 59.5% (44/74) of the participants studied life sciences at the college or higher level, and 40.5% (30/74) reported currently working in the life sciences. This demographic is consistent with the description of early adopters by Rogers' theory of the diffusion of innovations [26], which explains that early adopters tend to have advanced education, expert knowledge, and willingness to engage in trials of new technologies.

Previous Use Of Genome Tools

All participants had their genome sequenced previously using genetic testing services: Complete Genomics, 23andme, Microbiome, Illumina, and Family Tree DNA. 32.4% (24/74) participants reported using multiple testing services, and all participants had access to their data for over 6 months before the present study was administered. Users also reported using additional tools beyond the initial reports provided by these services to understand their data. The most commonly used tools were Google and Internet searches (10.8% (8/74)), academic paper databases such as JSTOR, Pubmed, and Medline (8.1% (6/74)), SNPedia, a wiki for information about genetic data (5.4% (4/74)), and Promethease (14.86% (11/74)), a tool that draws variant information from a number of different sources.

	Question
Q1	What insights and information about your genetics did this visualization give you?
Q2	List the gene variants you found most interesting and describe how and why you identified them as interesting.
Q3	Please use the space below to tell us which features were most helpful for understanding the report and why they were helpful.
Q4	Please use the space below to tell us how we can improve the report to make it easier to understand.
Q5	What reports or tools did you use to view and learn from your personal genomic data previous to this visualization?
Q6	Please elaborate on anything new you learned from this visual report that you didn't notice in previous reports

Table 1. Open-ended questions

	Question	M(SD)
Q6	The information in the report is presented in a clear and accessible manner.	4.37 (0.69)
Q7	The overview report is easy to understand	4.22 (0.79)
Q8	The overview report is easy to navigate	4.44 (0.76)
Q9	The categories tab is easy to understand	4.51 (0.66)
Q10	The categories tab is easy to navigate	4.53 (0.70)
Q11	I would need the help of a healthcare professional to better understand my results	2.76 (1.07)
Q12	The report gives me a firm grasp of my health and genetics	3.28 (0.94)
Q13	The visualization communicates health concerns in a clear way	4.15 (0.72)
Q14	The categories of gene variants are clear and easy to understand	4.25 (0.78)
Q15	Using this visualization I learned new insights and information about my genetics that I hadn't noticed in previous reports	3.76 (1.05)
Q16	I am able to grasp to what extent the knowledge regarding different variants is certain or uncertain	4.44 (0.72)
Q17	I found the full glossary helpful for interacting with my report	3.47 (0.97)
Q18	I found the question mark buttons helpful for interacting with my report	3.33 (0.84)
Q19	I found the ability to save variants helpful for interacting with my report	3.49 (0.94)

Table 2. Questions on perception of the new tool measured using five-point Likert scales

Usage

Participants spent an average of 30.11 minutes (SD=25.72 min) using the tool. Responses to the qualitative questions show that this relatively prolonged interaction was often indicative of high-engagement: "I think your tool is absolutely wonderful. I spent so much time with it because I found it so incredibly useful." Participants clicked on variants 716.0 times on average while examining the visualization tool. Participants changed tabs 27.4 times on average, and changed the filters (e.g. risk, rarity) 2.6 times on average per session. Finally, participants saved an average of 1.7 variants. Participants saved a total of 185 variants while using GenomiX. Almost all (98.9%, 183/185) of the variants saved by participants were saved when the "risk" filter was applied, and 70.3% (130/185) of variants saved were saved from the overview. 74.6% (138/185) of the variants saved were variants that the individual was affected by (i.e. not a carrier). The majority of saved variants, 52.4% (97/185), were pathogenic, whereas 18.4% (34/185) were protective, 17.3% (32/185) were benign, and 11.9% were pharmacogenetic. Table 3 shows how the number of variants saved break down by health effect and certainty of supporting research.

Understandability

Overall, participants found the report easy to understand. They rated the statement "I would need the help of a healthcare professional to better understand my results" with a mean score of 2.76 (SD=1.07), and the statement "The information in the report is presented in a clear and accessible manner" with a mean score of 4.37 (SD=0.69).

	Health Effect		
Certainty:	Low	Medium	High
Well-established	4.86% (9/185)	9.73% (18/185)	6.49% (12/185)
Likely	25.41% (47/185)	8.65% (16/185)	7.03% (13/185)
Uncertain	30.81% (57/185)	4.86% (9/185)	2.16% (4/185)

Table 3. Number of variants saved according to their health effect and the certainty of evidence.

Users also agreed that both the *overview* report and the *by category* report are easy to understand (see Table 2, Q7 Q9 Q13 Q14). In the words of one participant: "It made it much more clear what information is currently known about my genome and how important each bit is."

Understanding the certainty of the scientific evidence of a gene variant's effect is a crucial concept for participants when viewing their report. While some gene variants have well-established health effects, many do not and therefore should not cause undue stress or worry. Participants gave a mean rating of 4.44 (SD=0.72) to the statement "I am able to grasp to what extent the knowledge regarding different variants is certain or uncertain". In addition, 23.0% (17/74) of the users commented that they noticed the certainty of evidence of a specific gene variant, or that they understood the implications of the certainty of the evidence after using our visualization. For example, in reflecting on the insights gained using GenomiX, one participant noted that "My variants on this report tend to either have low scientific certainty or low health risk or both." Another specifies, "KCNE1-D85N was interesting but because its certainty was low, I'm not worried about it." Participants' understanding of the importance of the certainty of evidence in interpreting their data is also apparent in their rating of the statement "The report gives me a firm grasp of my health and genetics" with a mean score of 3.28 (SD=0.94). This score may indicate that participants understand that the report (due to the evolving nature of the scientific evidence) only offers limited interpretation of their personal genomic information.

New Insights Afforded by GenomiX

Despite having previously seen their results, participants rated the statement "Using this visualization I learned new insights and information about my genetics that I hadn't noticed in previous reports" with a mean score of 3.76 (SD=1.05), and 79.7% (59/74) of participants reported new insights about their genetics in the open-response questions. Of the participants who did not report new insights through our visualizations, all but one explained that they had already thoroughly reviewed their variant report before using this visualization. One user commented, "I had really poured through my gene report after I learned that I could look up specific genes so nothing brand new popped out. It would have 2 months ago." The insights garnered using GenomiX can broadly be divided into three of the four categories of insights outlined by Saraiya, et al. [28]: *details* - focused information about a particular gene variant; *overview* - overall distribution of gene variants; and *groups* - identification or comparison of groups of gene variants. Our dataset yielded no *patterns* based insights.

Focused Information About Gene Variants

48.7% (36/74) of participants noticed a specific gene variant or disease risk that they had not noticed when they had previously viewed their genome report. One participant stated, "... I was happy to find one copy of a variant that adds to longevity, and I'm a carrier of one that protects against many types of HIV. I just hadn't noticed them before, and I've looked at my report many times."

31.1% (23/74) of the participants described using the *by category* tab as effective means of identifying genes of interest, and for better understanding the impact of these genes. As one participant stated, "The variant TGIF1-P83Shift caught my eye this time, in part because it showed up as a variant in the nervous system category. I didn't notice it before, probably because I didn't understand the description until I googled what holoprosencephaly is…"
18.9% (14/74) of the participants also noted that they were able to identify and prioritize the gene variants that are most likely to affect their health more efficiently. As another participant described, "It was easy to identify what mutations may or may not be harmful - I have been trying to figure this out on my own, but have been confused."

Participants reported that they noticed and understood particular characteristics of a gene variant (such as carrier status, the certainty of the research on the gene variant, and the rarity of the gene variant), to a greater extent using GenomiX than they did using other tools. One participant commented, "Great! My ALS gene that is so bad and scary - I'm a carrier, it won't affect me. I guessed that from the previous reports but it was NOT clear at all."

Grouping and Overview

18.9% (14/74) of the participants noted grouping and overview insights. Participants indicated that sorting and re-organizing their data in meaningful ways helped them to understand particular genes in relation to other genes in their dataset, allowing them to discern high-priority information or "credible threats" (as described by one participant). Another individual referred to this process as "separat[ing] the wheat from the chaff."

Participants mentioned that they used the sorting features to visualize the data in two dimensions in a way that allowed

them to effectively and efficiently pick out genes that were located in particular regions of the visualization. In the words of one participant "[the tool] simplifies an otherwise-complicated task of sorting and weighing the gravity of an overwhelming amount of data."

Participants were particularly interested in genes that fell in the upper right-hand quadrant of the overview tab, where there is well-established evidence and high health effect. Genes in the lower left-hand quadrant, where there is low certainty and low health effect, were of less interest, "...at the moment, I have limited time to inspect my genome. So I'm interested in triaging the data points to help me focus on which data points might be useful for me to know now. This visualization helps me do that."

Using the tool, participants were also able to formulate complex questions about how gene variants in different "buckets" in the plot interact or counteract each other. For example, one user stated, "my worst thing is C3-R102G increasing risk of macular degeneration. That does stand out as interesting. I'm curious why CFH-V62I which is preventative on the same thing is in the low bucket while this is in medium on health effect." Such question formulation is a useful step in further exploring their personal genomic data, with or without the help of a health professional.

Usability and Usefulness

In general, participants found GenomiX easy to use as indicated by the mean ratings of Q8 and Q10 (see Table 2). These ratings are consistent with feedback provided in the open response questions, and comments regarding usability and usefulness. For example, one participant noted, "The Overview graph was most helpful, providing easy access to the salient features." Another commented, "I had already looked at the data, but the information is more accessible in this format." As discussed in a previous section, one of the most useful features of the overview tab seemed to be the ability to sort genes by different metrics to highlight results of interest or concern. Participants also used the categories tab as an alternative way of thinking about their results "I like toggling between category and overview, rarity and risk. It let me think about the info in different ways."

Participants reported that the mapping of particular visual elements (namely the color and size of gene variants) to qualities of the data further enhanced their experience navigating and interpreting the data: "The visuals, with different sizes and colors, made the overall picture of my health easier to view and navigate. The ability to click on the circles that appeared more important and learn more was an improvement." Indeed, 21.6% (16/74) of participants explicitly mentioned color and size coding schemes as helpful for discovering new insights. Fewer participants described using the glossary, buttons leading to more information, or the ability to save variants when discussing how they used the tool.

Suggestions for Improvement

When prompted for improvements that would make the report easier to understand, 14.9% (11/74) users commented on the content of the visualization. 6.8% (5/74) of participants wanted more variants to appear in their report, 2.7% (2/74) of participants wanted information about non-health related traits such as eye color. One user who commented on both improvements wrote, "I'm not sure if this is just a snippet of my variants, or if the dots included in the chart are the only ones with enough established data and/or known relevance to health to warrant inclusion, but I'd love to see more variants included. Also, perhaps more charts that explore non-health-related traits, like eye color, handedness, and maybe ancestry (all plotted along with their certainty of evidence)." 6.8% (5/74) of participants also commented on features they would have found useful when using the visualization. 2.7% (2/74) of participants suggested the addition of a search tool, 2.7% (2/74) of participants suggested the addition of a print feature, and 1.4% (1/74) of participants wanted the ability to toggle between more characteristics of the gene traits.

DISCUSSION AND DESIGN IMPLICATIONS

Making sense of abundant personal data that involves uncertainty poses a challenge even to expert users. In this study, we found 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. Our choice to use certainty and health effect as axes in the visualization resulted in placing variants into 'buckets'. This approach is novel within the context of personal genomics. The spatial organization of variants combined with sorting features (risk/rarity) simplified the inherent complexity of genomic data and made the relationships between gene variants explicit while allowing users to easily identify variants of interest.

The log data from participants' interactions with the tool imply that the tool allowed users to highlight variants of interest effectively: the majority of participants saved pathogenic variants or variants that would directly affect them. When reflecting on their use of the tool, and the differences between GenomiX and other tools they had used, participants pointed out that data are traditionally presented without any sort of organizing structure, making the process of discerning credible or important information labor and search-intensive, and highly reliant on existing domain knowledge. Without offering a new interpretation of the data, GenomiX provided a visual scaffold for the majority of participants, which enabled them to gain new insights from a dataset they were already familiar with, and helped them to prioritize which variants (or types of variants) they wanted to learn more about or monitor.

Furthermore, using the tool, participants were able to formulate complex questions about how gene variants in different 'buckets' relate to each other. This may show that

visual and spatial encoding reduces the complexity of the data allowing users to think more critically about their results. This, in turn, will help users to monitor the progress of new research over time, and to explore the implications for interpretation more effectively.

Uncertainty plays an important role when interpreting genomic data and choosing which variants to focus on. The personal genomic context offers a form of uncertainty not addressed by existing work on visualizing uncertain data. In this 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 novel technologies, processes, and new scientific findings. Thus, uncertainty plays an important role when interpreting genomic data and choosing which variants to focus on. In our evaluation, numerous participants mentioned the level of certainty as an important factor in interpreting their genomic results, and specifically in determining the credibility or urgency of a finding. Unlike the risk and rarity filters, the users were not offered a way to change the view of the certainty of findings, however, the emphasis given on certainty in participant responses to the open questions, demonstrates that plotting the data with certainty on one dimension and health effects on another, was effective in communicating the personal genomics unique source of uncertainty.

Personal genomic data are very personal, and therefore more studies are needed in which participants are presented with their own data. In this study we observed that users spent a fairly substantial amount of time exploring their genome. We believe that individuals were motivated to explore their results using the tool for this long because they were *their* results. Furthermore, users disclosed very personal details about their lives and the lives of their relatives as they described their reasoning processes, indicating that the exploration and evaluation of personal genomics is a fundamentally personal process informed by the individual experiences of users. This characteristic of users' engagement with GenomiX highlights the need and the value of consumer-facing genomics visualization research using individuals' own personal data rather than anonymized or fake data.

Limitations and Future Work

While the design and evaluation of GenomiX offer insights into the design of future interactive personal genomics exploration tools, there are a number of limitations to this study that should be considered in future research. First, participants were early adopters who were already familiar with personal genomics testing and reporting, and therefore could be considered "expert users". Future research should explore the use of similar tools among lay users. Furthermore, we only studied a single interaction with the tool. We believe that tools such as GenomiX 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. A longitudinal test of GenomiX usage would help us to understand participants' information needs over time and to observe how participants use the tool as their knowledge and the background research changes. Additional research is also needed to understand *how* individuals use the tool to gain the insights we observed in this study. We intend to follow with a more targeted research into users' usage patterns with GenomiX.

Implications for Design

The findings have design implications for interactive tools that enable exploring personal data with varying levels of uncertainty: first, the study suggests that there is value in studying tools that allow for exploring personal data multiple times over time as evidence about the interpretation of the data changes. Second, our study provides effective techniques to help people learn more about their data without interpreting the data for them by allowing the user to reorganize data, providing different representations of the data, and communicating the uncertainty of the data itself. By providing users tools to manipulate these features of the data, they will be able to begin to explore and make sense of their own data. Lastly, our studies show that there is a need for reflective tools for people to document and curate information based on their interpretation of their own data. We saw evidence of individuals gaining insights and making connections between data points, reflecting on the data in front of them and on what they already knew. Here lies an opportunity to design tools that support such reflection, providing users a platform to gather their growing knowledge and changing interpretation over time. In empowering the individual, we believe that such a platform enables and engages the related community of interest described by Kuznetsov et al. [15] which is necessary for the creative interpretation, debate, and action that help individuals to address shared concerns.

CONCLUSIONS

In this design case study, we introduce a tool aimed at supporting individuals who have had their genomes mapped to explore and make sense of their results. We identify personal genomics as a unique area in personal informatics in which data are largely stable but can be interpreted continuously over time. The interactive and visual features of the proposed tool helps individuals to prioritize gene variants, which will, in turn, enable them to make sense of future findings that might change the interpretation of these genetic data. Participants' responses in this study suggest that GenomiX could be a core part of a larger suite of tools where people can explore their personal genomic data.

ACKNOWLEDGMENTS

This work was partially funded by Staley Grant from Wellesley College and by grants IIS-1017693 and IIS-1422706 from the National Science Foundation.

REFERENCES

1. Peter W Airasian, Kathleen A Cruikshank, Richard E Mayer, PR Pintrich, James Raths and Merlin C Wittrock. 2001. A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives. *Anderson LW and Krathwohl DR. New York: Addison Wesley Longmann*.
2. Madeleine P Ball, Jason R Bobe, Michael F Chou, Tom Clegg, Preston W Estep, Jeantine E Lunshof, Ward Vandewege, A Zaranek and George M Church. 2014. Harvard Personal Genome Project: lessons from participatory public research. *Genome Med*, 6 (2). 10.
3. Madeleine P Ball, Joseph V Thakuria, Alexander Wait Zaranek, Tom Clegg, Abraham M Rosenbaum, Xiaodi Wu, Misha Angrist, Jong Bhak, Jason Bobe and Matthew J Callow. 2012. A public resource facilitating clinical use of genomes. *Proceedings of the National Academy of Sciences*, 109 (30). 11920-11927.
4. Florian Block, Michael S Horn, Brenda Caldwell Phillips, Judy Diamond, E Margaret Evans and Chia Shen. 2012. The deeptree exhibit: Visualizing the tree of life to facilitate informal learning. *Visualization and Computer Graphics, IEEE Transactions on*, 18 (12). 2789-2798.
5. Carole Cadwalladr. 2013. What happened when I had my genome sequenced *The Guardian*.
6. Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A. Kientz. 2015. SleepTight: 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 (UbiComp '15), 121-132.
7. George M Church. 2005. The personal genome project. *Molecular Systems Biology*, 1 (1).
8. Kevin Davies. 2010. *The $1,000 genome: the revolution in DNA sequencing and the new era of personalized medicine*. Simon and Schuster.
9. Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. 2015. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '15), 731-742.
10. FDA. 2015. Use of Electronic Informed Consent in Clinical Investigations Questions and Answers Guidance for Industry DRAFT GUIDANCE, HHS/FDA.
11. Miriam Greis, Thorsten Ohler, Niels Henze and Albrecht Schmidt. 2015. Investigating Representation Alternatives for Communicating Uncertainty to Non-experts. in *Human-Computer Interaction–INTERACT 2015*, Springer, 256-263.
12. Henning Griethe and Heidrun Schumann. 2006. Visualizing uncertainty for improved decision making. in, (SimVis.
13. National Human Genome Research Institute. 2015. Cost Per Genome.
14. Ohad Inbar. 2007. Graphical representation of statistical information in situations of judgment and decision-making. In *Proceedings of 14th European Conference on Cognitive Ergonomics* (ECCE '07), 265-268.
15. Stacey Kuznetsov, Aniket Kittur and Eric Paulos. 2015. 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*, 303-312.
16. Christina R Lachance, Lori AH Erby, Beth M Ford, Vincent C Allen and Kimberly A Kaphingst. 2010. Informational content, literacy demands, and usability of websites offering health-related genetic tests directly to consumers. *Genetics in Medicine*, 12 (5). 304-312.
17. Min Kyung Lee, Junsung Kim, Jodi Forlizzi, and Sara Kiesler. 2015. 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* (UbiComp '15), 743-754.
18. Ian Li, Anind Dey, Jodi Forlizzi, Kristina Höök and Yevgeniy Medynskiy. 2011. Personal informatics and HCI: design, theory, and social implications. in *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, 2417-2420.
19. Jeantine E Lunshof, Ruth Chadwick, Daniel B Vorhaus and George M Church. 2008. From genetic privacy to open consent. *Nature Reviews Genetics*, 9 (5). 406-411.
20. Alan M. MacEachren, Anthony Robinson, Susan Hopper, Steven Gardner, Robert Murray, Mark Gahegan, and Elisabeth Hetzler. 2005. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32 (3). 139-160.
21. Gordon E Moore. 1998. Cramming more components onto integrated circuits. *Proceedings of the IEEE*, 86 (1). 82-85.
22. Limor Nadav-Greenberg, Susan L Joslyn and Meng U Taing. 2008. The effect of uncertainty visualizations on decision making in weather forecasting. *Journal of Cognitive Engineering and Decision Making*, 2 (1). 24-47.
23. Jenny E Ostergren, Michele C Gornick, Deanna Alexis Carere, Sarah S Kalia, Wendy R Uhlmann, Mack T Ruffin, Joanna L Mountain, Robert C Green, J Scott Roberts and PGen Study Group. 2015. How Well Do

Customers of Direct-to-Consumer Personal Genomic Testing Services Comprehend Genetic Test Results? Findings from the Impact of Personal Genomics Study. *Public health genomics*, 18 (4). 216-224.

24. Personal Genome Project. GET-Evidence. Retrieved September 24, 2015 from http://evidence.pgp-hms.org/about
25. Personal Genome Project. Personal Genomes. Retrieved September 24, 2015 from http://personalgenomes.org/
26. Everett M Rogers. 2010. *Diffusion of innovations*. Simon and Schuster.
27. John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers Chalmers. 2014. Personal tracking as lived informatics. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (CHI '14). ACM, New York, NY, USA, 1163-1172.
28. Purvi Saraiya, Chris North and Karen Duca. 2005 An insight-based methodology for evaluating bioinformatics visualizations. *Visualization and Computer Graphics, IEEE Transactions on*, 11 (4). 443-456.
29. Bertrand Schneider, Megan Strait, Laurence Muller, Sarah Elfenbein, Orit Shaer and Chia Shen. 2012. Phylo-Genie: engaging students in collaborative'tree-thinking'through tabletop techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3071-3080.
30. Orit Shaer, Guy Kol, Megan Strait, Chloe Fan, Catherine Grevet and Sarah Elfenbein. 2010. G-nome surfer: a tabletop interface for collaborative exploration of genomic data. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1427-1436.
31. Orit Shaer, Ali Mazalek, Brygg Ullmer and Miriam Konkel. 2013. 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*, 109-116.
32. Orit Shaer, Oded Nov, Johanna Okerlund, Martina Balestra, Elizabeth Stowell, Laura Ascher, Joanna Bi, Claire Schlenker and Madeleine Ball. 2015. Informing the Design of Direct-to-Consumer Interactive Personal Genomics Reports. *Journal of medical Internet research*, 17 (6).
33. Meredith Skeels, Bongshin Lee, Greg Smith, and George G. Robertson. 2009. Revealing uncertainty for information visualization. *Information Visualization*, 9 (1). 70-81.
34. Barry N. Taylor and Chris E. Kuyatt. 1994. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. DIANE Publishing.
35. Judi Thomson, Elisabeth Hetzler, Alan MacEachren, Mark Gahegan, and Misha Pavel. 2005. Typology for visualizing uncertainty. In *Proceedings of the IS&T/SPIE Symposium on Electronic Imaging, Conference on Visualization and Data Analysis* (SPIE '05), 146-157.
36. Amos Tversky and Daniel Kahneman. 1974. *Science*, 185 (4157). 1124-1131.