An Environment For Long-Term Engagement with Personal Genomic Data

Johanna Okerlund
Wellesley College
106 Central St
Wellesley, MA 02481 USA
jokerlun@wellesley.edu

Orit Shaer
Wellesley College
106 Central St
Wellesley, MA 02481 USA
oshaer@wellesley.edu

Martina Balestra
New York University
5 Metrotech Center
Brooklyn, NY 11201 USA
mb5758@nyu.edu

Oded Nov
New York University
5 Metrotech Center
Brooklyn, NY 11201 USA
onov@nyu.edu

Casey Grote
Wellesley College
106 Central St
Wellesley, MA 02481 USA
caseygrote@gmail.com

Abstract
Direct to consumer genetic testing has become substantially more accessible to the average person in the last decade, providing those who seek it with a tremendous amount of important and very personal information. Genetic data is a compelling vein of personal informatics for HCI research because, unlike other varieties that sample data at regular intervals, genomic data is only collected once but its interpretation and potential application change over time based on new scientific discoveries. The perception of the data also changes over time based on users' lifestyle, environment, behavior, and an understanding of family history. Here, we discuss the properties of personal genomic data and propose functional requirements for a platform that allows users to explore raw data while tracking, organizing, and sharing their findings. We then share a preliminary design of a platform that aims to facilitate these functional requirements by providing the user means to document their findings, draw connections, and track their evolving interpretation.

Author Keywords
Personal genomics; information practices;

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g.,HCI): Miscellaneous.

Background

The cost of sequencing a human genome has fallen from approximately $100 million in 2001 to a little over $5,000 in 2014 [6], a rate much higher than Moore's Law [5]. This has rapidly accelerated the emergence of direct to consumer genetic testing. Individuals who order this testing online without doctor prescription (and who frequently have no formal genomics training) are then confronted with making sense of an unprecedented amount of raw data involving sensitive topics such as personal traits, disease risk, and carrier status. Furthermore, the relationships between genes and these topics are not currently well understood, even by expert geneticists. Genetic research is fairly young and evolves with the development of novel technologies, processes, and new scientific findings. Considering these factors, to date the FDA only permits the return of raw variant reports and prohibits the return of interpreted health reports [1]. The user must continuously re-evaluate their results against recent research findings. Ultimately, users' interpretation and perception of this data can influence their lifestyle decisions, emotional state, and well being.

The unique case of HCI for personal genomics

The quantity, sensitivity, and evolving understanding of genomic data evoke compelling HCI questions when considering the 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 does user engagement change over time as new information gets published? How can we design for effective interaction with personal genomic information? It is often the case with personal informatics that the dynamic element is the data itself, and HCI research is thus focused on how to synthesize and communicate relevant information succinctly. HCI for personal genomics data represents a unique case: genomic data is only sampled once and its *interpretation* and potential applications are dynamic, relevant research questions should instead focus on how to support personal exploration of the initial data and external resources, as well as on showcasing the interpretation as dynamic phenomena.

Functional Requirements for Interaction with Personal Genomic Data

In a previous study [8], we interviewed and observed early participants of the Personal Genome Project (an academic organization attempting to build an open repository of volunteered genomic data) interacting with their personal genomic data. User needs were synthesized and specific functional requirements for future platforms were extracted:

1. Reviewing an annotated report - Participants read and interpret an annotated overview report to understand which gene variants indicate particular traits or increased risk.
2. Searching the literature - Much of the evidence is uncertain and papers with new evidence are published often.
3. Sharing information - Users share with doctors, genetic experts, families, and friends.
4. Comparing genomes - Users compare genetic information, especially within families, to identify similarities and trends.

Figure 1. A physical workspace using a collage method to organize and develop ideas.

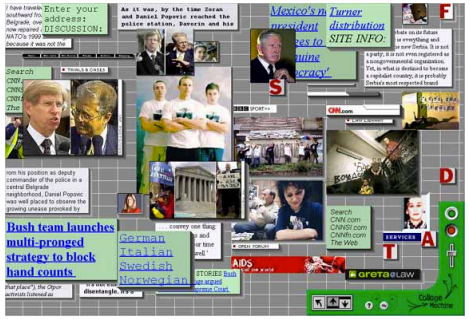

Figure 2. *CollageMachine*, a collage-like workspace for open-ended web browsing

5. Curating information – collecting, organizing, and relating information collected regarding the interpretation of individual's data.

Preliminary Platform Design

Given these functional requirements, a platform for interacting with personal genomic information should also support user exploration and active engagement over time. To that end, we looked towards the design of tools promoting active, open-ended web browsing as an analogy for this platform. In particular, we found inspiration in the "collage" model proposed and implemented by Kerne [4]. Traditional web browsing is passive because the user has few tools to document or comment on the connections between the pages they link to. Kerne asserts that a creative agent for web browsing could be collage-like, giving the user space to rearrange, organize, document, and reflect, allowing for spontaneous or intentional recombination (Figure 1, 2).

Following, we present a preliminary design of an environment for a long-term engagement with genetic data. Our design consists of two main features that form a central collage workspace: 1) a personalized and interactive genetic report; and 2) an environment for collecting and organizing published research and other information artifacts. These act as the primary materials in the collage workspace. Users may toggle between these views and can rearrange, group, draw connections, and make annotations on gene variants while collecting external resources in the workspace. In what follows we demonstrate how the collage design paradigm supports the functional requirements (outlined above).

Gene Variants Report

Interaction with the system begins by reviewing an overview personalized genetic report. The content of the report comes from comparing the user's genome to a database of gene variants that are known to be associated with medical conditions. Only gene variants that are found to be medically relevant or that are related to a particular trait are reported. To display the variants, we use a bubble chart where each gene variant is represented as a circle (Figure 3).

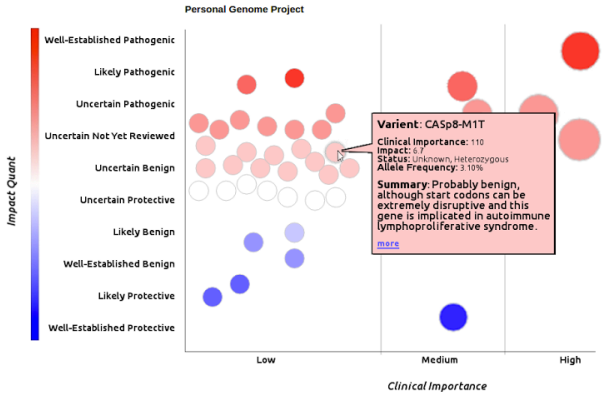

Figure 3. A sample gene variant report. Each variant is represented as a circle. The color and y-position indicate the impact. Red coding indicates pathogenic impact and blue represents protective impact. The x-position represents one of three categories of clinical importance: low, medium, and high importance.

The color, size, and placement of each circle reveal information about that variant and hovering shows its summary report. This report acts as the initial content of the collage workspace, which is accessed via a toggle button on the screen. When the user transitions from the variant report to the interactive workspace, the

axes of the graph disappear and the individual bubbles are free for the user to move around. The interactions in this workspace are described below.

Workspace Interactivity

In the workspace users can manipulate their personalized gene variant bubble chart by categorizing, grouping, and annotating gene variants (Figure 4). For example, a user may want to group variants that indicate carrier status, or that they want to ask their doctor about. A group is created by selecting the relevant sources and clicking "group". Elements of the group can also be subsequently added or removed individually. Groups can in turn become elements in larger groupings, allowing users to nest information according to their interests and intentions. Operations such as connecting, merging, and annotating can then be applied to whole groups of variants.

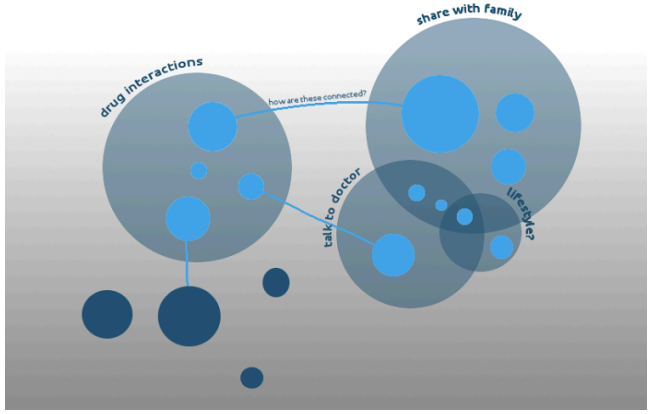

Figure 4. Rearranging, inserting, and annotating groupings of gene variants in the bubble chart.

The user can also connect gene variant representations to others (for example indicating gene interaction), as well as to information artifacts (e.g. scientific articles, or notes summarizing a meeting with an expert). Connections are a way for users to document their findings and understanding of the relationship between elements. Connections are created by drawing a line between elements and adding annotations as appropriate (Figure 5).

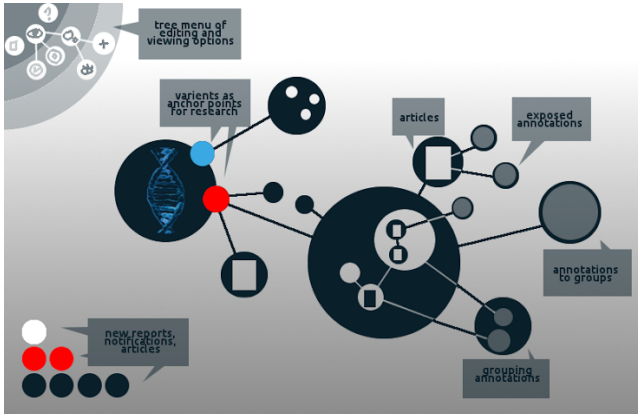

Figure 5. Web of connections between genomic data and external resources.

Open-ended Research

The workspace allows users to collect and curate information artifacts that they find while browsing the Internet outside the system (e.g scientific publications, support groups, magazine articles). The browser itself will have a button so that users can pin a resource and import it to their workspace. In the workspace, users can associate artifacts with the gene variants as well as annotating and grouping information artifacts directly (Figure 6).

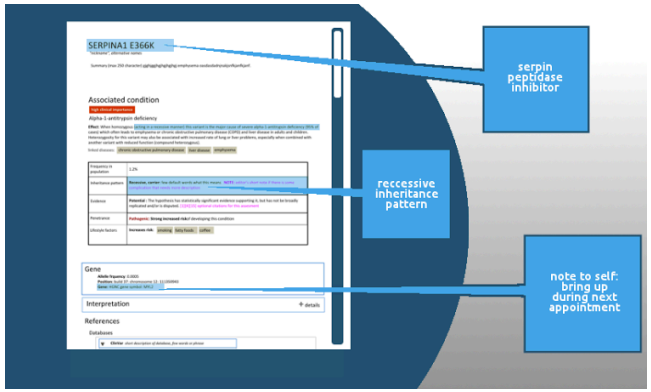

Figure 6. Annotating and highlighting imported resources

Comparing and Sharing Information

The system allows users to compare genomes, for example between family members, by importing genetic data and other shared information artifacts. The imported objects will present as a different color in the workspace, and the user can continue to group, annotate, and draw connections on both data sets.

We also envision the system as a platform for sharing information with a doctor, friends, or family. The grouping feature makes it easy to choose specific subsets of information to be shared with particular people or groups of people. For example, all external information and annotations connected to a certain gene could be hidden temporarily when the data is shared. These privacy settings are meant to provide the user with additional control over their data while promoting a culture of open research and collaboration.

Documenting Dynamic Interpretation

While the primary interactions with the system are with the dynamic web of information that users create (Figure 5), the user will have the ability to view the information on a chronological timeline accessed through a menu in the workspace. Intervals are dictated by when new information was added to the workspace (Figure 7). This view not only allows them to see how their understanding changed over time, but to keep track of the "state of the art" in relevant research and the implications such updates have on their data. For example, this feature would be helpful in understanding variants such as SNP rs9939609, a variant on the FTO gene linked to obesity [2]. Recent literature suggests that the link between the gene and obesity only exists for people born after 1942 [7]. Numerous other studies have shown that its clinical importance also depends on ethnicity and age and that obesity is highly correlated with lifestyle [3, 9]. Users concerned by how this variant may manifest for them can use this tool to not only track recent developments in research on this gene, but also to keep track of their own lifestyle changes.

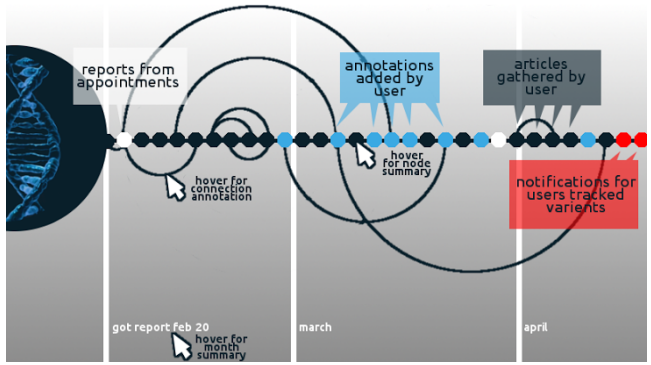

Figure 7. Information timeline view

Conclusion

In this paper, we presented personal genomics as a unique and compelling form of personal informatics to explore in an HCI setting. We also proposed collage as an interaction paradigm for interacting with this type of data over time. The evolving knowledge of genetics, the sensitivity of the interpretation, and applicability of the data to the individual invite the design of tools that allow for personal research, creative browsing, connection making, and a view that tracks interpretation over time. This is a slightly different focus from traditional personal informatics research, which has been focused on data collected over time. Our work on empowering users to conduct their own research on a complex topic could be applied to other types of personal informatics data. It is important to develop platforms and experiences that not only display or interpret data when it is collected, but also support long term engagement.

While it is important to empower users to conduct their own research, researching such complex and sensitive data may be overwhelming. Future work will focus on ways to further support users, for example through crowd sourcing with experts, other users, or an intelligent environment that proposes connections and resources. These could lead to the development of a rich and meaningful user experience that faciltates long term engagement with personal genomics.

Acknowledgements

We thank our collaborators from Harvard Personal Genome Project. This work is partially funded by NSF grant IIS-1017693.

References

[1] Retrieved from FDA: http://www.fda.gov/iceci/enforcementactions/warningletters/2013/ucm376296.html

[2] Frayling, T., Timpson, N., Weedon, M. et al. A Common Variant in the FTO Gene is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. Science 316, 889-894 (2007).

[3] Grau, K., Hansen, T., Holst, C. et al. Macronutrient-specific effect of FTO rs9939609 in response to a 10-week randomized hypo-energetic diet among obese Europeans. Int J Obes 33:1227-34 2009.

[4] Kerne, A. (2001). *CollageMachine: A Model of "Interface Ecology".* Dissertation, New York University, Computer Science, New York City.

[5] Moore, G. (1965). Cramming More Components Onto Integrated Circuits. *Electronics Magazine*.

[6] National Human Genome Research Institute. (2014, Oct. 31). *DNA Sequencing Costs: Data From The NHGRI Genome Sequencing Program (GSP)*. Retrieved Dec. 14, 2014, from genome.gov: http://www.genome.gov/sequencingcosts/

[7] Rosenquist, J., Lehrer, S., O'Malley, A. Cohort of birth modifies the association between FTO genotype and BMI. PNAS 1411893111v1-201411893 (2014).

[8] Shaer, O., Nov, O., West, A., Eastman, D. Understanding Information Practices of Interactive Personal Genomics Users. Extended Abstract, ACM CHI 2014.

[9] Vimaleswaran, K. Ängquist, L., Hansen R. et al. Association between FTO variant and change in body weight and its interaction with dietary factors: the DiOGenes study. Obesity 20: 1669-1674. 2012.