Understanding Users Information Needs and Collaborative Sensemaking of Microbiome Data

Article *in* Computer Supported Cooperative Work · August 2019

DOI: 10.1145/3274470

Orit Shaer
Wellesley College

Oded Nov
New York University

Understanding Users Information Needs and Collaborative Sensemaking of Microbiome Data

JENNIFER OTIONO[*], Wellesley College
MONSURAT OLAOSEBIKAN[*], Tufts University
ORIT SHAER, Wellesley College
ODED NOV, New York University
MAD PRICE BALL, Open Humans Foundation

Recent years are seeing a sharp increase in the availability of personal omic (e.g. genomes, microbiomes) data to non-experts through direct-to-consumer testing kits. While the scientific understanding of human -omic information is evolving, the interpretation of the data may impact well-being of users and relevant others, and therefore poses challenges and opportunities for CSCW research. We identify the information, interaction, and sense-making needs of microbiomic data users, within the broader context of social omics - the sharing and collaborative engagement with data and interpretation. Analyzing users' discussions on Reddit's r/HumanMicrobiome, we identified seven user needs for microbiome data: reviewing an annotated report, comparing microbiome data, tracking changes, receiving personalized actionable information, curating and securing information, documenting and sharing self experiments, and enhancing the communication between patients and health-care providers. We highlight the ways in which users interact with each other to collaboratively make sense of the data. We conclude with design implications, including tools for better communication with care providers, and for symptom-centered sharing and discussion.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: Online communities; Microbiome; Collaborative sensemaking

ACM Reference Format:
Jennifer Otiono, Monsurat Olaosebikan, Orit Shaer, Oded Nov, and Mad Price Ball. 2019. Understanding Users Information Needs and Collaborative Sensemaking of Microbiome Data. 1, 1 (August 2019), 21 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In recent years, there has been a sharp increase in the availability of personal 'omic' data (e.g. genomes, microbiomes) to non-experts through direct-to-consumer testing kits. "Omic" is a suffix used to denote studying components of biology in totality, commonly using next-generation DNA sequencing and other high-throughput methods. For example, personal genomics is the study of multiple genes, viral metagenomics of pathogen detection, and microbiomics of the microbiota

---

[*] Both authors contributed equally to this research.

Authors' addresses: Jennifer Otiono, jotiono@wellesley.edu, Wellesley College, Wellesley, MA, ; Monsurat Olaosebikan, Tufts University, 161 College Avenue, Medford, MA, 02155, monsurat.olaosebikan@tufts.edu; Orit Shaer, oshaer@wellesley.edu, Wellesley College, Wellesley, MA, ; Oded Nov, New York University, New York, New York, onov@nyu.edu; Mad Price Ball, Open Humans Foundation, Boston, MA,, mad@openhumans.org.

XXXX-XXXX/2019/8-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

associated with digestion, immune response, and other aspects of human health [40]. Omic data are characterized by its large scale and complexity.

As of February 2019, nearly 30 million personal genomic or microbiomics reports were provided directly to consumers by popular services such as 23andMe and uBiome [1, 63]. The widespread availability of such extensive and complex data, in need of understanding by non-experts, poses both challenges and opportunities, with substantial societal impact, for CSCW and HCI research. People with no formal training in the life sciences get access to their omic data (genomic and microbiomic) by sending a self-collected sample to a direct-to-consumer provider, and receive their results as an online report. These non-expert users then need to interpret complex data that involves sensitive information such as disease risk and potentially meaningful correlations with health and physical traits. Furthermore, the data need to be contextualized within an evolving scientific understanding regarding the meaning and appropriate interpretation of genomics and microbiome information. The interpretation of the data may impact lifestyle decisions and well-being of these users, as well as of their relevant others (e.g., family members, friends, community).

Consider, for example, a family home, where the commensal microbiota of family members can be influenced by shared lifestyle elements such as nutrition and pets, who might seek to understand how changes in their lifestyle and environment (e.g. new diet, new pet, changing seasons) might impact their microbiome and, potentially, their health. People who suffer from similar medical conditions might also seek to compare, share, and understand omic information and its implications for their well-being. However to date, there are few tools for storing, aggregating, comparing, exploring, and collaboratively making sense of such information. Recent research on personal omics studies investigated the information needs of personal genomic non-expert users [39, 68, 76] rather than microbiome data, hence little is known about the motivation and information needs of non-expert users who seek to collaboratively understand this data.

Online discussion groups on websites such as Facebook and Reddit serve as venues for exploration of knowledge sharing about omic information. However, it is not clear whether such venues are effective in facilitating meaningful informed discussions that address users' information needs. At the same time, the information seeking activities and discussions that take place on such platforms could serve as a source for identifying users' interests and information needs, and thereby inform the design of future direct-to-consumer omic data services.

In this paper, we report findings from a study of the subreddit r/HumanMicrobiome that engages users in discussions related to their microbiome and its health implications. In particular, our investigation focuses on three questions:

> **Research Question 1)** Who are the users - what are their characteristics and goals?
> **Research Question 2)** What are their information needs?
> **Research Question 3)** In what ways do users in this group interact with each other to collaboratively make sense of human microbiome information?

The CSCW and HCI communities have articulated a need to design interactive technologies that draw upon the social context of personal health and wellness data [9, 42]. This study contributes toward this goal by promoting the understanding of user motivations, information and sense-making needs as well as the social context surrounding an increasingly available new class of personal health data - personal omics. Omic data in general, and microbiomic data in particular are critical for the development and practice of proactive and personalized medicine, yet their scientific understanding is still evolving. The data are complex and sensitive and thus require a layer of curation and guidance that are not necessary for the exploration of other types of personal health informatics. This paper also contributes design considerations for interactive tools for communication and exploration of microbiomic data including tools for better communication with care providers and with others who experience similar symptoms.

We review the related work, and follow with methods and results from the study of subreddit r/HumanMicrobiome. We conclude with a discussion of the findings, design implications, and future work.

2 RELATED WORK

2.1 The Study of Human Microbiome

The past decade has seen a surge of research of human microbiota and its relevance to health, as represented by a growth of pubmed search results for 'Human Microbiome', by year, from 1974 to 2019 [52]. The Human Microbiome Project (HMP), established in 2008 and funded by the U.S. National Institute of Health, aims to generate resources for comprehensive characterization of the human microbiome and analysis of its role in human health [48]. This initiative resulted in the sequencing of over 2,200 reference strains and the publication of findings exploring the relationship between microbiome and various disease states, including inflammatory bowel disease (IBS) and type 2 diabetes[61]. At the same time, a combination of participatory research and direct-to-consumer models have launched, providing individuals with an opportunity to test and learn about their personal microbiome. The American Gut project, launched in 2012, is an academic effort that has had over 10,000 individuals contribute and take part in crowdfunded microbiome research [46]. Commercial services also offer direct-to-consumer microbiome analysis, including Viome (launched in 2016), uBiome (launched in 2012), and Thryve (launched in 2016). As of October 2018, uBiome reports having analyzed 250,000 samples, and represents the single largest entity generating microbiome data for individuals [1]. While clinical services offered by uBiome were suspended in May 2019 following legal concerns regarding billing practices, non-clinical products remain [18]. These commercial products aim to use individual microbiome data to provide personalized analyses and recommendations for improving health.

With the increasing availability of personal microbiome datasets, there is a growing need for tools that translate this data for non-experts. In 2017, researchers from Viome Inc. detailed how artificial intelligence (AI) can help individuals understand their internal biological ecosystem through the creation of a high-resolution model of their microbiome and the application of machine learning to produce an individualized wellness plan [6]. Gut Instinct integrates citizen science with online learning [55]. Helix is a commercial service which offers an array of "DNA-powered products" that allow non-experts to engage with their microbiome and genetic data through products like personalized fitness and nutrition insights, and ancestry data [27].

2.2 Social Sense-Making in Data Curation

The rise of social computing transformed data curation to a social activity [26, 33] - co-curation (or social curation). In recent years, social curation sites such as Flickr and Pinterest allowed users to create a multitude of object and data collections which can be shared and discussed in social media [2, 25, 50, 51]. In the field of Biotechnology and especially genomics research, co-curation plays an important part. As techniques of large-scale genomic analysis and functional gene annotation have progressed, the process of co-curation created paradigms for genome annotation among experts [16]. One example is the Gene Ontology (GO) which is widely used for expert annotation of molecular attributes of genes and gene products [8, 28, 64]. Other examples include the US National Center for Biotechnology Information (NCBI) and UniProt KnowledgeBase, which provide robust platforms for data sharing and knowledge dissemination. More recently, a platform coined GNPS was introduced by Nature magazine [75]. The platform allows scientists from all over the world to tap to the potential of the diverse chemistries present in natural products (NP) for biotechnology and medicine. Until recently, these knowledge bases were shared among the bio-sciences community only in the form of published papers, but the new system enables community sharing, continuous annotation of data, and co-curation of its reference libraries and data sets [24, 75].

However, the use of co-curation in health is not limited to professionals. In recent years, we see transition from personal health informatics to family health informatics [59] in addition to a rise of co-curation in health and medical contexts by patients and other non-experts. Websites like TuDiabetes, PatientsLikeMe, and Eat.ly help thousands of individuals make sense of their experiences and conditions by presenting, sharing, and commenting on health knowledge [22, 77]. These websites can elicit new concepts for health-care vocabularies, coding sets, and classifications [70], and facilitate behavior change. For example, in the context of nutrition management, Mamykina et al. (2011) have observed that the co-curation practice of collaborative tagging enhances individuals' ability to remember the nutritional values of meals [44].

Our study builds on the insights of past research on expert co-curation of genomics as well as of non-experts' health social tagging to find new and meaningful ways for collaborative curation and engagement for non-experts with personal microbiome data. While research in this area of personal and social data curation is thriving, it is difficult to deduce from current research on the sharing and curation of omic data. First, omic data such as human microbiome might be perceived as more personally sensitive and as such less prone to sharing. Second, omic data in general and microbiome data in particular is not easily delivered through short and constant un-curated updates, the same way fitness tracking data may be. Moreover, omic data is not easily understood to a non-experts and requires an additional layer of curation and interpretation.

2.3 Personal Microbiomics as Personal Informatics

We consider the self study of microbiomes, within the growing field of personal informatics, which refers to communities, practices, and systems that help people collect and reflect on their personal information [12, 41]. The increasing availability of low-cost sensors has accelerated the practice of self-tracking and the rise of the Quantified Self movement [72]. Commercial and research efforts have resulted in numerous self-tracking technologies and applications for health and wellness [38, 71] and the sharing of the data gained by self tracking [9]. Self-tracked personal data includes a wide range of activities and symptoms: fitness data [4, 20], sleep times [54], reading habits [79], food and liquid intake[34, 67], phone usage[5], different physical and emotional symptoms[15, 66] etc.

A common assumption in personal informatics is that an individual's knowledge of their data facilitates reflection that can lead to self-discoveries and to behavioral and lifestyle changes. Li et al. [41] proposed a model of how people use personal informatics tools. The model describes iterative transitions between preparation, collection, integration, reflection, and action, as well as barrier for effective use. This model had been extended by other researchers. For example, Whooley et al. [78] and Epstein et al. [17] proposed models that differentiate between stages of reflection. Other researchers characterized the barrier and challenges toward the adoption and effective use of self-tracking technologies [7, 10]. 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. Researchers have began to examine collaborations between patients and health care providers around self generated data [11, 12, 44]. Our work leverage and expand the understanding of tensions created in such collaborations surrounding new data types (personal omic in general and microbiomic in particular).

Personal microbiomics shares the main goals and assumptions of personal informatics - facilitating self-discovery based on personal information. However, similar to other personal omic areas, such as personal genomics, its interpretation and related implications for the user's health are dynamic as scientific knowledge is evolving [53]. The complex and sensitive nature of personal omic in general, and of microbiomic in particular, requires a layer of curation and guidance that are not necessary for the exploration of most other types of personal informatics [69]. In addition, while microbiome information is inherently personal, it is also shared among family members, and other community members (such as individual with similar conditions, or people who live together), thus affecting the health and well-being of its owner as well as of relevant others. We therefore consider personal microbiomics within the broader contexts of personal omic [53] and Biological Citizenship [13], which explores connections between biology and self-identity, as well as empowers individual to "take care of their own health" [13, 29, 43].

2.4 Personal Health Informatics in Online Health Communities

Social media allows communities to form in order to connect, support, and educate people who share (or care for people with) particular medical conditions. Online health communities (OHCs) serve a range of purposes including seeking advice and support, asking questions in order to make sense of information, disseminating relevant literature,and sharing personal experiences with experts and non-experts, improving understanding of symptoms and professional diagnoses, fueling professional engagement, and promoting management of chronic symptoms [30]. Studying the information seeking activities and the content of discussions taking place in such communities is an important source for identifying users' interests and information needs, and can inform the design of future tools for empowering users. There is a significant body of research studying OHCs, here we only discuss work that is highly relevant to our study.

Huh et al. [30] studied users' needs and requirements in online health communities, developing personas to illustrate the different ways people use such communities. Robillard et al. [65] studied how information about dementia is discussed and disseminated on Twitter. Park et al. [57] harnessed the Reddit platform to investigate how written communication challenges manifest in online mental health communities focusing on depression, bipolar disorder, and schizophrenia. Pappa et al. [56] investigated activity behavior and posts content on the Reddit weight management community LoseIt (r/loseit).Several studies highlighted the tensions between informational and socio-emotional needs in OHCs [49], Mayara et al. [14], and [35]. Our investigation focuses on r/HumanMicrobiome, which is different from other OHC because it centers around specific quantitative data (microbiomics) which enables a rapidly growing area of research, rather than particular medical conditions or specific goals. The human microbiome has a broad impact on human health, users are invited to join the group to learn more and to explore its implications on their own health.

3 METHODS

We follow a CSCW and HCI research tradition [23, 45] in conducting an in depth study of social media community posts to gain insight into the experiences and needs of its members. Specifically, we focused on the r/HumanMicrobiome community within reddit. Reddit is one of the top most visited websites in the US and has engagement rates higher than Twitter (as of April 2018) [31, 36]. Reddit's r/HumanMicrobiome community is defined as "*a science/evidence based sub*" dedicated to human microbiome data and its implications, where "*usage of citations to support your claims is highly encouraged and may become mandatory*" [62]. We chose to study this particular community for several reasons, including its focus on the burgeoning area of human microbiome, the active discussion it facilitates over time, and its publicly available content. Alternative communities dedicated to this topic include private Facebook groups, and forums managed by commercial testing providers. We made an ethical choice, to focus on a community where the discussion is public. We discuss further ethical considerations in the next subsection. It is also important to note that our choice for studying this particular community was guided not by aspiring to generalize the findings to other groups or to the general population, but rather, by seeking to inform the development of design recommendations for future exploration and sense making tools, based on the experiences of early adopters.

3.1 Ethical Considerations

We use content analysis methods to study social media content publicly shared on Reddit's Human Microbiome community. We intentionally chose to use only open, public data. This approach led us to focus a reddit community rather than other communities where registration, invitation or permission are required to access the community space. While the CSCW community is in early stages of discussing and forming guidelines for conducting research using public social media [73], we acknowledge the ethical concerns associated with our methods.

First, contributors to this discussion post public content to what they may perceive to be a discussion within a group. They might not be aware that the content shared could be used for academic research [19] and they did not give explicit permission for using the data for this purpose. Second, while public, some of the posts contain sensitive information about users' health conditions and symptoms which the contributor might not want to see amplified in a research paper. Third, while user names on Reddit could be independent from users' real world identity, Reddit cannot and does not guarantee anonymity. However, some users might perceive the discussion as anonymous because their user name does not give away their identity. Others choose user names that include their name. Our analysis of superusers, who are the moderators of this group (see Section 4.1), was informed by their' choice to identify themselves publicly on this community.

To mitigate these concerns we conducted a large scale study that rather than focusing on the online behavior of individuals, aims to understand the characteristics, motivations, and needs of aggregated user groups. We also approach the data from a perspective of empathizing with users, aiming to understand their information, interaction, and sense-making needs. The goal of the study is to inform the design of tools for helping and empowering users to explore and make sense of their omic data and its implications for health. We believe that the members of the r/HumanMicrobiome community will directly benefit from such tools. Finally, we slightly modified some of the quotes included in the paper as verbatim by removing potentially identifying information.

3.2 Data and Analysis

We obtained 393 posts and 3,991 comments from r/HumanMicrobiome dated between the inception of the subreddit in June 2017 to November 2018. We used SQL to extract the data from a public database of 1.7 billion Reddit posts and comments stored on Google's BigQuery. At the time of collection the data set had Reddit data from 2015 to November 2018.

To answer RQ1, who the users of r/HumanMicrobiome are and what are their characteristics and goals, we conducted an analysis of the frequency of posting by individuals. We also examined the types of posts and comments made by individuals to identify user groups based on behavioral patterns and values. Posts/comments were classified using the coding system described below.

To answer RQ2, how do users seek to use their microbiome data and what are their information needs, we used content analysis methods to analyze all posts and comments. First-level codes were developed iteratively based on literature [30, 32] and from preliminary review and discussion of the data by two independent coders. Then two coders tagged 393 posts and 3,880 comments (comments that only contained short phrases were excluded) with one or more of the following 9 codes: reference, question, knowledge sharing, data challenges, self experiments, recommendations and advice, engagement with professionals, symptom sharing, goals and hopes. We allowed for more than one code per post/comment since many contained multiple components. Comparing the most commonly associated codes we find that within posts, questions and seeking recommendations(21/393) co-occurred the most. Within comments, the code knowledge sharing often co-occurs with reference(236/3880), questions(171/3880), self-experiments(110/3880) and recommendations(139/3880). We also found symptom sharing and self-experiments to be closely coupled (195/3880). Intercoder reliability based on 100% of the data was calculated for each code separately and then averaged, yielding an average kappa score of 0.52 with 92% agreement [21, 47]. Table 1 provides a list of codes with definitions and sample quotes. Using affinity diagramming, we clustered codes into themes.

Finally, to answer RQ3, how users in this community interact with each other to collaboratively make sense of human microbiome data, we examined the types of data that people share, their questions, and knowledge sharing. We also considered the rules (set by the moderators) for interacting with the community, and examined whether and how they are enforced.

Table 1. Descriptions and examples of qualitative codes

Code	**Description**	**Sample Quote**
Reference	Links to a paper, article, blog or another website	[The Gut-Brain Axis: The Missing Link in Depression][link]
Question	Questions clarify concepts, elucidating meaning from data, and gathering insight.	*Wasn't there just a study a few weeks ago that said diet and environment matters more than genetics?*
Knowledge sharing	Consists of references, explanations, claims and conclusions, based on what the author believes is objective knowledge.	*... but to say there's "no credible evidence" that the microbiome might be able to cause mental illness is not accurate.*
Data challenges	Expressing challenges interpreting, making sense of, and utilizing human microbiome data.	*Ubiome results are very hard to interpret ...I was wondering if you guys could take a look at my results. They feel completely useless to me. [link]*
Self Experiments	Sharing experiences of self experimentation by systematically changing lifestyle factors.	*by systematically removing all the foods you listed from my diet and replacing them with organic foods only I've reduced my personal IBS symptoms by 80-90%.*
Recommendation and advice	Asking for (or receiving) general advice or specific recommendations intended to be acted upon.	*I advise you to ask yourself this: What is it you personally hope to learn?*
Engagement with professionals	Experiences or concerns regarding interaction with health care experts and professionals including doctors, health scientist, and insurers.	*As a scientist, I am very weary of forking over any kind of genetic information, including my microbiological "finger print".*

Continued on next page

Table 1 – continued from previous page

Code	Description	Sample Quote
Symptom Sharing	Sharing symptoms and medical experiences.	*After that, I started getting chronic fatigue, swollen joints, back pain, IBD*
Goals and hopes	Sharing goals and expectations of engaging with microbiome data.	*I was hoping for a miracle. I had exhausted so many years seeing Dr. after Dr. with no answers.*

4 RESULTS

4.1 RQ1) Who are the participants?

From June 2017 to November 2018 there were 613 unique contributors to r/HumanMicrobiome. Of those 613, 153 made at least 1 post and 574 made at least 1 comment. Approximately 80% (122/153) of post contributors made only 1 post. 80% (461/574) of comment contributors made between 1 and 5 comments and 20% (113/574) made more than 6 comments. Based on the presented number of subscribers to r/HumanMicrobiome in November 2018, 5K, our data suggests that about 13% of r/HumanMicrobiome subscribers are active contributors. Compared to one of the largest science based subreddit, r/science, that had 14.5 million subscribers in December 2016 with less than 1% active contributors [32], r/HumanMicrobiome shows relatively high levels of active engagement. The content of posts and comments indicate that the community is international, with users mentioning residence in North America, Asia, and Europe.

Analyzing the frequency of engagement within the community, we found that the top 1% of contributors (6 users) account for over 40% of comments and posts made in the subreddit. Following prior research on super users as a small group of users with disproportionately high number of contributions [60] or followers [74], we refer to the top 1% (6/613) contributors as super users. This group includes the moderators of the subreddit. Other active contributors wrote at least one comment. We also recognized a small (1%) of users who are experts in areas related to the microbiome. Figure 1 shows the distribution of contribution type per user group. Further content analysis allowed us to identify characteristics and motivations for three distinct user groups: super users (S), other contributors (O), and experts (E).

Super users on r/HumanMicrobiome often share from their own experiences battling with various health issues and are open to experiment with different approaches to alleviate their symptoms. The majority of their posts and comments are about sharing knowledge and references, as well as asking clarifying questions to better help other community members. They back their claims with references to scientific research and keep up with the recent scientific literature. They demand that users provide evidence for any claims they make and correct perceived misinformation posted by other users, and steer discussion in the right direction if it diverges from the goals of the subreddit. In this particular subreddit, they put extensive work into developing a wiki for the community, a guide pointing to relevant information on topics frequently asked about in the subreddit.

A small subset of contributors (less than 1%, 5/613) include scientists and experts who seek to connect with microbiome data users and practitioners. Experts include microbiologists and microbiome researchers. They reach out to the community to collect data, validate assumptions, or share knowledge. For example, one user wrote *"Don't do this at home. The dangers almost certainly outweigh the potential benefits. Even if done properly (in a controlled clinical setting), success chances of fmts depend *a lot* on the indication, on the donor and on a lot of factors we don't even begin to*

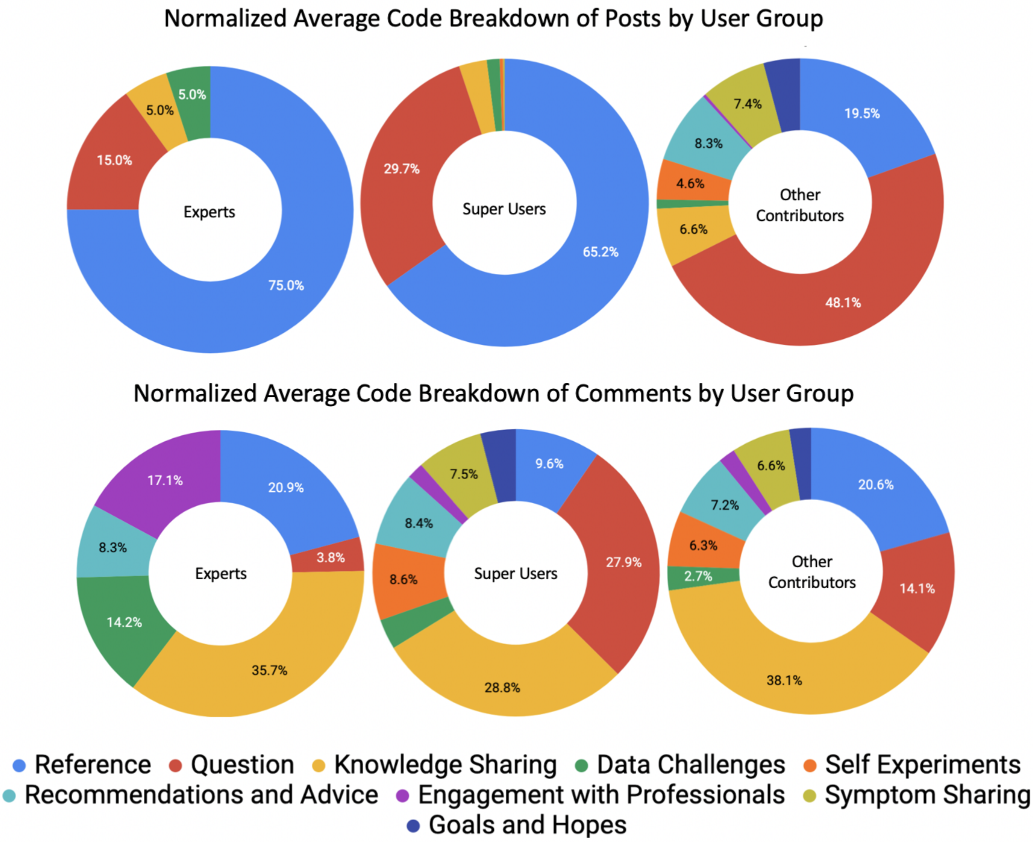

Fig. 1. Normalized average breakdown of types within posts(top) and comments(bottom) by experts, super users, and other contributors.

understand fully. Source: I'm an academic researcher working on fmt experiments and their impact on the recipient's gut flora."(E) Experts' contributions typically include sharing knowledge and references. Other participants mostly contribute posts and comments include questions, knowledge sharing, and recommendation requests. Figure 1 shows the breakdown of contribution type within posts and comments for each of these user groups.

4.1.1 User goals. We analyzed 3880 comments to find that 74 (2%) mentioned some form of concrete goal. Of these, we identified three themes for goals related to engaging with microbiome data: 1) Tracking the impact of their condition on their microbiome; 2) Improving their microbiome with hopes of alleviating reported symptoms; and 3) Learning more about the relationship between health and the microbiome. These goals are not mutually exclusive - generally, users experience health issues that they attribute to a gut microbial imbalance. Thus, they are motivated to learn more about the human microbiome through sharing personal anecdote and scientific literature that relate to their condition, in the hope of identifying treatments or lifestyle changes that will improve their quality of life. In the words of one user *"I have IBS, Microscopic Colitis and Sibo. I believe all my problems are microbiome related."*(O). Another user says *"I have MS and want to see if my gut* *microbiome has been affected. I have found that there are a few direct-to-consumer Gut Microbiome tests, but I'm not convinced they're useful. Are they?"*(O).

While specific goal setting is often seen as a core aspect of personal informatics, biomic data is more difficult to collect and interpret than other personal data types, as the collection require a third-party test provider, and the interpretation requires high-level of expertise and depends on scientific knowledge that is rapidly evolving. We speculate that due to these challenges, users' engagement with microbiomic data is more exploratory than goal driven at this stage.

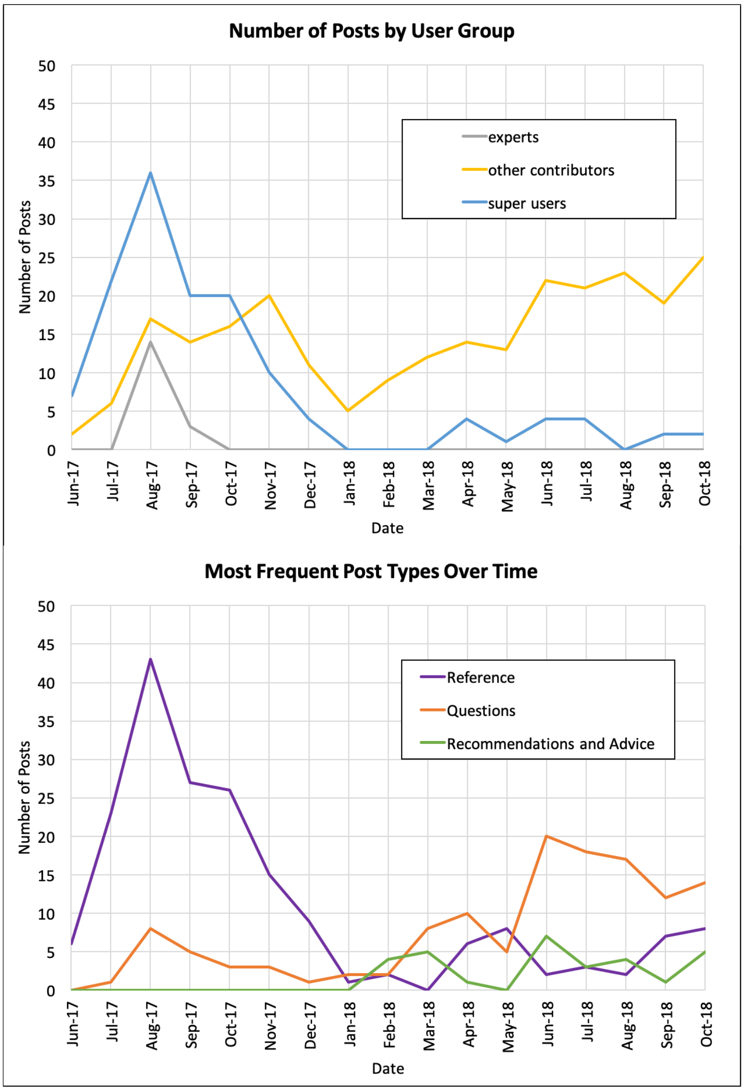

Fig. 2. a) Number of posts by user group; b) Most frequent post types over time

4.1.2 Temporal Analysis of r/HumanMicrobiome content. A temporal analysis of posts and comments showed a shift in the types of posts and made over time, reflecting the changing body of knowledge and practices of personal microbiomics. In 2017, at the start of the subreddit, posts were dominantly reference sharing of emerging research. In early 2018, posts shifted to focus more on asking questions. Also starting early 2018 we see a steady increase in posts by 'other contributors' users. Notably, the dominant provider of microbiome testing services, uBiome, was pursuing major growth in 2018 [1], and changes to the subreddit may have reflected increased sales and access to personal microbiome data during this time period. Figure 2a shows the number of posts per user group over time. Figure 2b shows the most frequent post types over time.

4.2 RQ2) What are users' information needs?

In this section we first report on observed themes, and then synthesize these themes into a list of concrete information needs.

Observed Themes

4.2.1 Quantifiable microbial data. Quantifiable microbial data which is obtained through microbiome testing, plays a central role in this community's exploration, comparison, and sense-making. While not all users obtained their own microbiome data, some of the users who did obtain their information from direct-to-consumer testing felt overwhelmed by their microbiome information. They did not know how to interpret the data or make it actionable. In particular, it is not clear to users what aspects of the data they should prioritize when seeking to learn about their microbiota. One user wrote, *"I have using Thryve. I guess my problems was I didn't know what to do with all the data."*(O) Another user wrote *"I'm not sure about how to interpret results either but I'm curious to see what other people have to say."*(O)

Users expressed a need to create a strategy to improve their microbiome, however they did not feel like they had the right tools to understand their microbiome data. In the words of one user, *"I've read most of the stuff posted but am having a hard time implement strategies to improve my own microbiome. There is just too much info to parse through it all and come up with an effective strategy to repair my own dysbiosis."*(O)

Others point out that the probiotic recommendations from companies like Ubiome are lacking personalization and leave them wondering if there is information they are missing: *"I took a Ubiome test, but I have zero clue on what I am supposed to do with the data... It seems to really push probiotics but it literally shows all of them."*(O)

Some go so far as to say the microbiome testing is of little use: *"You're right that there's very little use to these tests."*(S) Others go on to describe the limitations of testing: *"I analyze these types of data frequently, and am aware of a number of limitations. First of all, the technology is pretty noisy, so it would be hard to tell whether or not a difference is meaningful. Even if the technology was more precise, this would still be an uncontrolled experiment, and it would difficult to make comparisons to other people. Services such as uBiome aren't grouping people into proper groups (i.e. by health status, age, diet, etc.) that allow for someone to make these comparisons. Also, they don't provide enough information to make a proper statistical comparison."*(E)

Users also expressed the need to organize their microbiome information along with resources and research papers in a central location. They also wanted to keep this information private and secure: *"I am taking a look out there to find another sort of program that can work for me, until then all my resources are spread out all over the place...this is really a lot of information that we have to manage with this stuff, right? Privacy etc is important too, seems like there is no security nowadays."*(O)

4.2.2 Self Experimentation. Community members are highly proactive and methodical applying self-experimentation. About 12% of comments were from members documenting instances of self-experimentation and sharing their corresponding results with the community. This allows them to receive feedback on their methodology from other members who may share similar experiences. For example one user posted:

"...Tried FMT again with the ineffective donor to see if it's more helpful after the antibiotic, and it was helping but then I tried to "boost" the donor's stool with prebiotics ..., and this time again they were harmful and I ended up in the ER. This seems to confirm that the important microbes in FMT are the phages, not the bacteria, and thus trying to feed the new bacteria with prebiotics is misguided..."(S) The post received 10 top level detailed comments, each generating a discussion. For example, here is one one detailed response: *"I'm going to mention some suggestions below...You imply that phages are the answer. Have you also considered fungi?... You mention problems with protein and fats... what about stomach acid supplements?"*(S)

Self-experimentation often occurs following an ineffective or negative experiences with medical professionals: *"I actually saw a doctor to a clinic to get started. He felt I was a really good candidate for the transplant however because I did not have C- diff I did not qualify for clinic transplantation. He gave me all the information I needed and then I sought out my own donor ..."*(O) Although self-experimentation is highly regarded in the community as a means to improving quality of life, there is a split in the community as to the kinds of self-experimentation people are willing to support. Some people will support self-experimentation of FMT which is highly ambiguous in its results. Others are more skeptical of FMT and are more likely to stick to self-experimental procedures related to diet changes.

4.2.3 Engagement with Health Care Providers. A common theme that emerges from examining users' posts and comments about their engagement with health-care providers is frustration. Users believe that in many cases the level of care they receive is outdated and unspecialized.

21 users in 119 comments/posts expressed frustration with various health care providers when trying to inform them about their microbiome in hopes that they would be able to use that information to improve their patient's health. Users attribute their numerous (and often chronic) health issues to an unbalanced gut microbiome and find that their providers are not equipped to help them. In the words of one user: *"My post infectious IBS started after a devastating month. Ever since, my "IBS" has encompassed severe dysbiosis, SIBO, c Diff, and other pathogens. I have no idea what to do about my microbiome, including my virome. My doctor knows it's relevant but has no idea how, or how we could use info like this to help me."*(O)

Others find health care providers largely unhelpful: *".... most of the doctors I've seen are completely clueless regarding FMT and the gut microbiome."*(O) While many users in this subreddit express distrust for health professionals, it is important that health providers be kept in the loop, especially when a patient experiments with food supplements, dietary changes, and other interventions.

4.2.4 Engagement with Testing Services. Another common theme in posts and comments was related to their experiences with testing services. The dominant provider, uBiome, accounted for over 80% of mentions. Posts and comments commonly posed questions about the products: *"Does ubiome give insight re: Candida overgrowth?"*(O). Users also described these services in the context of self-experimentation, for example one user who wrote: *"I will definitely be adding kefir. I have a mission to diversify my awful microbiome. I should have a new Ubiome in a couple weeks."*(O)

Emerging Needs

4.2.5 Information needs. Considering these themes, we identified seven information needs. We map each of the following needs back to relevant observed themes:

> **Information Need 1** - Reviewing an annotated report - users expressed desire for reports that present information in a way that is easier to explore and understand. In particular, users

indicate a need for a summary that highlights most important aspects *(observed themes 4.2.1, 4.2.4).*

Information Need 2 - Comparing microbiome data - users seek to make sense of their microbiome through comparison to others. Important comparisons include: healthy individuals with similar demographic characteristics, individuals with shared medical conditions and individuals with similar diet *(observed theme 4.2.1).*

Information Need 3 - Tracking changes - users often undergo multiple testing to track changes in their own microbiome overtime, particularly before and after certain interventions e.g. probiotic or FMT (Fecal Microbiota Transplant) *(observed theme 4.2.2).*

Information Need 4 - Receiving personalized actionable information - users desire actionable recommendations (e.g. life style changes interventions) that are personalized based on their microbiome, demographic characteristics, and medical conditions *(observed themes 4.2.1, 4.2.4).*

Information Need 5 - Curating and securing information - users express a need for collecting, sharing (e.g. with family and experts), organizing, and storing various data sources including annotated reports, scientific articles, and health related data in a secure manner *(observed themes 4.2.2, 4.2.4).*

Information Need 6 - Documenting and sharing self experiments - users seek to document and share details of their self experiments in order to track progress and receive suggestions *(observed theme 4.2.2).*

Information Need 7 - Enhancing the communication between patients and health-care providers - users seek to discuss and share emerging research, new practices, and their own documentation of self experimentation with health care providers *(4.2.3).*

4.3 RQ3) How do users in this community interact with each other to collaboratively make sense of human microbiome data?

4.3.1 Learning and Knowledge Sharing. Members of this community engage in collaborative learning, sense-making, and knowledge sharing about the human microbiome and its implications for their health. The rapidly evolving scientific knowledge on the microbiome and the lack of trust in the knowledge of health care providers further motivates community members to learn from each other. In the words of one user: *"We are on the edge of science in this sub, because no-one else could help us. We must help each other here."*(O) The discussions are often centered around sense-making of gut microbiota functionality and dynamics, and of possible tools and methodologies for alleviating gut microbiome abnormalities and symptoms. 49.36% (194/393) of the posts and 21.47% (833/3880) of the comments contain a reference to an external article, website or research paper. For example, *"Study shows association between gut microbes and brain structure in people with irritable bowel syndrome [link]."*(S) Members then use comments to ask questions that guide their understanding of the microbiome, the effects of probiotics and prebiotics on the gut microbiome, and how effective fecal microbiota transplants are. For example, *"Altering your diet will no doubt alter your gut microbiota but the question is, will it be beneficial?"*(O)

About 6% (226/3880) of the comments contain a link or reference to the subreddit's main wiki that serves as a shared knowledge base. For example, Q: *""Is there any beneficial virus in our body/bowels that could be knocked out by antibiotics?"*(O), A: *"Yes. From the wiki [link]."*(S) The moderators of the subreddit regularly update the extensive wiki that lists numerous studies and information.

In general, community members are excited to learn more about their microbiome, and how it impacts the body. As one user writes, *"thanks for the info. I feel like the microbiome is severely understudied and I hope the scientific community really prioritizes this field as there is a lot we still don't know about how everything works together ..."*(O) Overall, the discussion demonstrates a high level of critical thinking. Users exchange both qualitative and quantitative information and seek to understand patterns. The moderators review posts and often correct misinformation posted by other users. They demand that users provide evidence to support their claims and encourage users to engage in an evidence-based discussion.

4.3.2 Exchanging recommendations and advice. A quarter of the users 24% (145/613) make posts and comments seeking or providing recommendations or advice for an individual's medical conditions or data. Requests are often supplemented with a link to a personal data set (e.g. microbiome results) or a detailed description of symptoms and goals. For example, *"Here's some more screen-shots, I have no idea what it all means. It seems like I have more of a lot of different bacteria than most people... am I supposed to lower that somehow? [Microbiome Graph][Metabolism1][Metabolism2] ... am I supposed to raise the low stuff and lower the high stuff in some fashion??"*(O)

Community members then offer concrete recommendations and advice, pointing the user to resources and share their own experiences. Here is a response to the request above: *"No issues! Based on just a glance of your uBiome and all the samples I've seen here on Reddit and other forums online, I'd guess that your diet isn't as rich in fiber, and richer in meat/fat."*(O) Users also reciprocate by sharing their own personal microbiome data: *"This is interesting - - do you by any chance consume yogurt or kefir on a regular basis? Here's [my uBiome] from January this year...an interesting thing is the relatively high amount of *Fusicatenibacter*."*(O)

This interaction between users highlights important social aspects of personalized medicine: users often remind others that solutions vary for different people (*"Anything to do with probiotics currently is "just try various products", because they have drastically varying person-to-person effects."*)(S), draw attention to the dearth of knowledge in the scientific literature (*"...Moreover, I can only stress again that the (vast) majority of researchers in the field would agree that (i) very little knowledge on the microbiome warrants medical action/recommendations and that (ii) FMTs in particular are not studied well enough to warrant a general recommendation. "*)(E), and advise others to temper their hopes and expectations when considering testing and new procedures (*"I think you may be glossing over the complexity of the procedure."*)(O). These findings suggest that users see their role not only as passive consumers of information but as contributors to explanation and interpretation of information.

4.3.3 Sharing Symptoms and Experiences. A quarter of the users, 25% (153/613) openly share their data, symptoms, and experiences relating to their microbiome and health. They provide detailed accounts of medical conditions they have and the symptoms they suffer. A strong motivation for sharing is the hope to alleviate or eliminate symptoms by modifying their microbiome: *"I feel like the antibiotics...have left my gut bacteria unbalanced and I need to sort things out again..."*(O)

Members often seek those similar to them for guidance in order to minimize the amount of experimentation and, in the long run, to achieve favorable results: *"I'd love to chat with you and compare symptoms, I've found very few people who suffer from these exact symptoms."*(O) Members who have successfully improved their microbiome are often found to have made specific diet and lifestyle changes. They share these successes so that others can learn and implement them, as well as failures to prevent others from making similar mistakes, and provide advice: *"Contrary to the opinions on a lot of subreddits and supplement store employees, a lot of things you can buy to put in your body actually make things worse. I wish I had learned this earlier ... Stick to changing your diet and exercise. It's the safest bet."*(O)

The open exchange of symptoms, personal information and experiences, is a key characteristic of this community. It is possible that a perception of anonymity on Reddit contributes to the sharing of sensitive personal information, however multiple users in this group share personal information using user-names that are not anonymous, and in addition, share identifiable information. It therefore seems that it is the complexity of microbiome data combined with the severity of the symptoms users are hoping to alleviate that serves as the main factor contributing to the willingness to share sensitive information. In addition, the support and help users are receiving from others in the group foster a culture of sharing.

5 DISCUSSION

Following a tradition of studying users' needs through an analysis of online health communities [30], we studied users' interactions in the r/HumanMicrobiome subreddit. As is common in many online peer production and discussion venues [30, 32], participation is highly unevenly distributed, with a small number of participants making a disproportionate number of contributions (RQ1). We identified a number of information needs of microbiome data users: reviewing an annotated report, comparing microbiome data, tracking changes, receiving personalized actionable information, curating and securing information, documenting and sharing self experiments, and enhancing the communication between patients and health-care providers. These needs overlap and complement information needs identified in the study of personal genomic data users [69], reflecting the similarities and differences in the data type and implications they have for potential actions to be taken. Specifically, the fact that microbiome data are much more sensitive to behavioral and lifestyle interventions than personal genomics was reflected in users' need for practical advice about potential interventions such as changes in nutrition. Social aspects of the data also differ: while genomic data is similar across biological family members, microbiome similarities are impacted by various factors including geographical proximity, diet, and medical interventions (RQ2).

5.1 Microbiome data as focal point for reflection on knowledge, trust and roles

Users' concerns about their interactions with health-care providers and others around microbiome data suggest a need for continuous reflection about the relationship between formal roles (experts, health-care provider, patient, knowledgeable non-expert). Such relationships are clearer in most medical domains, where the expert authority lies with the health-care professional, who generally commands trust based knowledge advantage. In the case of microbiome data, these relationships are not yet established. For example, a key finding from our study is users' mistrust in healthcare professionals' knowledge about and interpretation of microbiome data. There was a clear gap between users' expectations that their microbiome data will be a key factor in their providers' understanding of the causes and treatments of their health problems, and what users perceived as lack of knowledge or inability to draw useful insights from such data. This contrast between users' own interest in exploring and interpreting microbiome on one hand, and their perception of health-care providers is a source of tension. Prior research of other domains [32, 65] suggested a need for multi-directional interaction between experts and non-experts as a means to share and discuss emerging research. Future work in this area could potentially alleviate this tension between experts and non-experts through design interventions.

Our findings also point to a substantial gap between users' concerns, needs and interests, and the information provided by current data environments. Specifically, we found that the main motivation for users seeking and discussing microbiome data is to alleviate health symptoms (RQ1). Sharing symptoms and experiences was a common way for users to find and learn from relevant others' experiences (RQ3). Our findings, of open exchange of personally-sensitive experiences, suggest that the complexity of microbiome data, and users' sense of inadequate response from their care providers as they seek to alleviate their health concerns, lead to a relatively lax approach to personal information disclosure. This is consistent with prior research on information disclosure in health-related venues on social media [3, 37, 58].

Our findings also highlighted an important interplay between social and personal aspects of health-care discourse: users often reminded to others that ways to address biome-related problems vary for different people, they drew each other's attention to the dearth of knowledge in the scientific literature (RQ3). There is a need for personalized recommendations and ways for users to compare themselves to those who are healthy in addition to those who suffer from medical conditions similar to them (RQ3). This echos the needs of personal genomic users, who seek to compare their genomes to biological relatives [77].

5.2 Implications for Design

People's exposure to direct to consumer microbiome data is a new phenomenon, and there are no established best practices for the design of interactive tools for microbiome data exploration and communication. Our findings offer a number of insights in this direction:

5.2.1 Symptoms and prior experience as focal points for user interaction and knowledge sharing. A substantial portion of users' activity in the reddit discussions revolves around physiological symptoms and experiences users associate with their microbiome. This suggests an important distinction between upstream information exploration where the starting point is physiological outcomes (e.g. symptoms) and to which causes and treatments are sought, vs. downstream information exploration where the starting point is specific microbiome characteristics that, in turn, may lead to physiological outcomes of interest to users. Our findings suggest that while current current online reports implement a downstream approach for microbiome information exploration, future tools for non-experts should support upstream information exploration.

Users' need to share, update and collectively explore personal experience is unmet by current commercial microbiome reporting systems. Our findings suggest that future tools should acknowledge the role of users as contributors to explanation and interpretation of microbiome information, and include within-platform features that embed social interpretation and sense-making. Design interventions should revolves around personal experiences, responses to interventions (primarily changes in diet), external or internal links to relevant personal data, as well as a clustering of narratives around common experiences. Thus, allowing users to connect their data, actions, and outcomes.

5.2.2 Interaction with health-care professionals. Prior work has explored patient generated data as boundary negotiating artifacts [12]. We find multiple indications to a similar phenomenon in the case of microbiome data where the the lack of best practices for communicating and engaging with the data causes tensions between users and health-care professionals.

The contrast between users' own interest in exploring and interpreting microbiome on one hand, and their perception of health-care providers as uninformed and conservative in this respect on the other, suggests a need for designing information tools that facilitate interactions between patients and their providers. Specifically, we propose providing features that support augmenting data with references to relevant research literature. This could be helpful to patients and providers in two ways: understanding the broader complexity and gaps in current research, adjusting expectations for clinical use, and contextualizing personal experiences in respect to general knowledge.

5.3 Limitations

This study has a number of limitations. First, our data come from discussions in an online public venue, and are therefore skewed towards input from people who are comfortable with sharing their knowledge and experiences on social media. Second, scientific research on the relationship between omic data and human health is relatively new and is currently evolving. As a result, the nature of both omic data and their meaning may change rapidly, impacting users' information needs and perceptions. Third, while common in CSCW research, studying a single platform (Reddit) limits the generalizability of the findings. Finally, our analysis did not include direct observations of how people engage with their microbiome data. Future research will address this through the study of additional user populations, think-aloud observations, and analysis of system log data.

6 CONCLUSION

With the sharp increase in the availability of personal omic data to consumers, it is essential to advance our understanding of users' needs within this personal and sensitive data context. Analyzing users' discussion of their and others' experiences related to personal microbiome data, we identified information and interaction needs: reviewing an annotated report, comparing microbiome data, tracking changes, receiving personalized actionable information, curating and securing information, documenting and sharing self experiments, and enhancing the communication between patients and health-care providers. We discuss collaborative sense making of the data and offer design implications, including tools for better communication with care providers and symptom-centered sharing and discussion features.

Future research is needed to address open questions, develop and evaluate novel interactive tools. In particular, it is important to understand the social context of microbiome data - how family and community affect engagement with data, its understanding and its communication. In addition, future work could benefit from a closer examination of the role of commercial microbiome test providers in setting and reconciling users' and healthcare professionals' expectations about the value of their products and the inferences that can be made from the analyzed biome data.

Beyond the microbiome domain, the information needs we identified are relevant in other contexts in which people explore complex and personally sensitive data. For example, future research may build on the work presented here to consider design guidelines for user-focused public and personal health applications such as those available on Open Humans platform [24], or in different contexts, platforms for sharing other potentially sensitive personal information such as support groups for people with addictions. Overall, the increasing availability of sensitive personal data with social implications challenges the CSCW community to provide insights and design guidelines to ensure a strong fit between user needs and system design.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation grants IIS 1814628 and IIS 1814932. We also thank Bastian Greshake Tzovaras from Open Humans Foundation and Lenore Cowen from Tufts University for discussion of this work.

REFERENCES

[1] 2018. uBiome - The Largest Human Microbiome Dataset. Retrieved March 31, 2019 from https://ubiome.com/blog/post/ubiome-largest-human-microbiome-dataset

[2] Morgan Ames and Mor Naaman. 2007. Why we tag: motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 971–980.

[3] Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2017. Sensitive Self-disclosures, Responses, and Social Support on Instagram: the case of# depression. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. ACM, 1485–1500.

[4] Apple. [n. d.]. Apple Watch Series. https://www.apple.com/apple-watch-series-4/activity/

[5] Apple. [n. d.]. Screen Time. https://support.apple.com/en-us/HT208982

[6] Guruduth Banavar, Reza Basseda, and Matteo Rinaldi. 2017. The new era of AI will revolutionize our wellness. In *Proceedings of the 8th Balkan Conference in Informatics*. ACM, 2.

[7] Matthew J Bietz, Cinnamon S Bloss, Scout Calvert, Job G Godino, Judith Gregory, Michael P Claffey, Jerry Sheehan, and Kevin Patrick. 2015. Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association* 23, e1 (2015), e42–e48.

[8] Evelyn Camon, Michele Magrane, Daniel Barrell, Vivian Lee, Emily Dimmer, John Maslen, David Binns, Nicola Harte, Rodrigo Lopez, and Rolf Apweiler. 2004. The gene ontology annotation (goa) database: sharing knowledge in uniprot

with gene ontology. *Nucleic acids research* 32, suppl_1 (2004), D262–D266.

[9] Alan Chamberlain, Erika Poole, Sean Munson, Catalina Danis, Elizabeth Churchill, et al. 2015. Moving beyond e-health and the quantified self: the role of CSCW in collaboration, community and practice for technologically-supported proactive health and wellbeing. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*. ACM, 273–276.

[10] Eun Kyoung Choe, Bongshin Lee, et al. 2015. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE computer graphics and applications* 35, 4 (2015), 28–37.

[11] Chia-Fang Chung, Jonathan Cook, Elizabeth Bales, Jasmine Zia, and Sean A Munson. 2015. More than telemonitoring: health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. *Journal of medical Internet research* 17, 8 (2015), e203.

[12] Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine Zia, James Fogarty, Julie A Kientz, and Sean A Munson. 2016. Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 770–786.

[13] Roger Cooter. 2008. Biocitizenship. *The Lancet* 372, 9651 (2008), 1725.

[14] Mayara Costa Figueiredo, Clara Caldeira, Elizabeth Victoria Eikey, Melissa Mazmanian, and Yunan Chen. 2018. Engaging with Health Data: The Interplay Between Self-Tracking Activities and Emotions in Fertility Struggles. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 40.

[15] Kevin Doherty, Marguerite Barry, José Marcano-Belisario, Bérenger Arnaud, Cecily Morrison, Josip Car, and Gavin Doherty. 2018. A Mobile App for the Self-Report of Psychological Well-Being During Pregnancy (BrightSelf): Qualitative Design Study. *JMIR Ment Health* 5, 4 (27 Nov 2018), e10007. https://doi.org/10.2196/10007

[16] Mary E Dolan, Li Ni, Evelyn Camon, and Judith A Blake. 2005. A procedure for assessing GO annotation consistency. *Bioinformatics* 21, suppl_1 (2005), i136–i143.

[17] 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*. ACM, 731–742.

[18] Christina Farr and Angelica LaVito. 2019. Health tech start-up uBiome suspends clinical operations following FBI raid. https://www.cnbc.com/2019/05/06/ubiome-suspends-clinical-operations-after-fbi-raid.html

[19] Casey Fiesler and Nicholas Proferes. 2018. âĂIJParticipantâĂİ Perceptions of Twitter Research Ethics. *Social Media+ Society* 4, 1 (2018), 2056305118763366.

[20] fitbit. [n. d.]. fitbit. https://www.fitbit.com/home

[21] Deen G Freelon. 2010. ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science* 5, 1 (2010), 20–33.

[22] Jeana H Frost and Michael P Massagli. 2008. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one anotherâĂŹs data. *Journal of medical Internet research* 10, 3 (2008).

[23] Eva Ganglbauer, Geraldine Fitzpatrick, Özge Subasi, and Florian Güldenpfennig. 2014. Think Globally, Act Locally: A Case Study of a Free Food Sharing Community and Social Networking. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 911–921. https://doi.org/10.1145/2531602.2531664

[24] Bastian Greshake Tzovaras, Misha Angrist, Kevin Arvai, Mairi Dulaney, Vero Estrada-Galiñanes, Beau Gunderson, Tim Head, Dana Lewis, Oded Nov, Orit Shaer, Athina Tzovara, Jason Bobe, and Mad Price Ball. 2019. Open Humans: A platform for participant-centered research and personal data exploration. *GigaScience* 8, 6 (06 2019). https://doi.org/10.1093/gigascience/giz076 arXiv:http://oup.prod.sis.lan/gigascience/article-pdf/8/6/giz076/28864591/giz076.pdf

[25] Catherine Hall and Michael Zarro. 2012. Social curation on the website Pinterest. com. *proceedings of the American Society for Information Science and Technology* 49, 1 (2012), 1–9.

[26] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 203–212.

[27] Helix. [n. d.]. A New Kind of DNA Testing. https://www.helix.com/shop/

[28] David P Hill, Allan P Davis, Joel E Richardson, John P Corradi, Martin Ringwald, Janan T Eppig, and Judith A Blake. 2001. Program description: strategies for biological annotation of mammalian systems: implementing gene ontologies in mouse genome informatics. (2001).

[29] Lee Hood. 2008. A personal journey of discovery: developing technology and changing biology. *Annu. Rev. Anal. Chem.* 1 (2008), 1–43.

[30] Jina Huh, Bum Chul Kwon, Sung-Hee Kim, Sukwon Lee, Jaegul Choo, Jihoon Kim, Min-Je Choi, and Ji Soo Yi. 2016. Personas in online health communities. *Journal of biomedical informatics* 63 (2016), 212–225.

[31] Andrew Hutchinson. [n. d.]. Reddit Now Has as Many Users as Twitter, and Far Higher Engagement Rates. https://www.socialmediatoday.com/news/reddit-now-has-as-many-users-as-twitter-and-far-higher-engagement-rates/521789/

[32] Ridley Jones, Lucas Colusso, Katharina Reinecke, and Gary Hseih. 2019. r/science: Challenges and Opportunities for Online Science Communication. In *Proceedings of the 2019 CHI conference on human factors in computing systems*.

[33] Maged N Kamel Boulos and Steve Wheeler. 2007. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education 1. *Health Information & Libraries Journal* 24, 1 (2007), 2–23.

[34] Ravi Karkar, Jessica Schroeder, Daniel A. Epstein, Laura R. Pina, Jeffrey Scofield, James Fogarty, Julie A. Kientz, Sean A. Munson, Roger Vilardaga, and Jasmine Zia. 2017. TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6850–6863. https://doi.org/10.1145/3025453.3025480

[35] Elizabeth Kaziunas, Mark S Ackerman, Silvia Lindtner, and Joyce M Lee. 2017. Caring through data: Attending to the social and emotional experiences of health datafication. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2260–2272.

[36] Simon Kemp. [n. d.]. Reddit now has more active users than Twitter — and is more engaging than porn. https://thenextweb.com/contributors/2018/04/19/reddit-now-active-users-twitter-engaging-porn/

[37] Jinyoung Kim, Andrew Gambino, S Shyam Sundar, Mary Beth Rosson, Chulakorn Aritajati, Jun Ge, and Christine Fanning. 2018. Interface Cues to Promote Disclosure and Build Community: An Experimental Test of Crowd and Connectivity Cues in an Online Sexual Health Forum. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 90.

[38] Predrag Klasnja and Wanda Pratt. 2012. Healthcare in the pocket: mapping the space of mobile-phone health interventions. *Journal of biomedical informatics* 45, 1 (2012), 184–198.

[39] 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 (C&C '15)*. ACM, New York, NY, USA, 303–312. https://doi.org/10.1145/2757226.2757246

[40] Stacey Kuznetsov, Alex S Taylor, Tim Regan, Nicolas Villar, and Eric Paulos. 2012. At the seams: DIYbio and opportunities for HCI. In *Proceedings of the Designing Interactive Systems Conference*. ACM, 258–267.

[41] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 557–566.

[42] Ian Li, Yevgeniy Medynskiy, Jon Froehlich, and Jakob Larsen. 2012. Personal Informatics in Practice: Improving Quality of Life Through Data. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. ACM, New York, NY, USA, 2799–2802. https://doi.org/10.1145/2212776.2212724

[43] Deborah Lupton. 2013. The digitally engaged patient: Self-monitoring and self-care in the digital health era. *Social Theory & Health* 11, 3 (2013), 256–270.

[44] Lena Mamykina, Andrew D Miller, Catherine Grevet, Yevgeniy Medynskiy, Michael A Terry, Elizabeth D Mynatt, and Patricia R Davidson. 2011. Examining the impact of collaborative tagging on sensemaking in nutrition management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 657–666.

[45] Lena Mamykina, Drashko Nakikj, and Noemie Elhadad. 2015. Collective Sensemaking in Online Health Forums. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3217–3226. https://doi.org/10.1145/2702123.2702566

[46] Daniel McDonald, Embriette Hyde, Justine W. Debelius, James T. Morton, Antonio Gonzalez, Gail Ackermann, Alexander A. Aksenov, Bahar Behsaz, Caitriona Brennan, Yingfeng Chen, Lindsay DeRight Goldasich, Pieter C. Dorrestein, Robert R. Dunn, Ashkaan K. Fahimipour, James Gaffney, Jack A. Gilbert, Grant Gogul, Jessica L. Green, Philip Hugenholtz, Greg Humphrey, Curtis Huttenhower, Matthew A. Jackson, Stefan Janssen, Dilip V. Jeste, Lingjing Jiang, Scott T. Kelley, Dan Knights, Tomasz Kosciolek, Joshua Ladau, Jeff Leach, Clarisse Marotz, Dmitry Meleshko, Alexey V. Melnik, Jessica L. Metcalf, Hosein Mohimani, Emmanuel Montassier, Jose Navas-Molina, Tanya T. Nguyen, Shyamal Peddada, Pavel Pevzner, Katherine S. Pollard, Gholamali Rahnavard, Adam Robbins-Pianka, Naseer Sangwan, Joshua Shorenstein, Larry Smarr, Se Jin Song, Timothy Spector, Austin D. Swafford, Varykina G. Thackray, Luke R. Thompson, Anupriya Tripathi, Yoshiki Vázquez-Baeza, Alison Vrbanac, Paul Wischmeyer, Elaine Wolfe, Qiyun Zhu, , and Rob Knight. 2018. American Gut: an Open Platform for Citizen Science Microbiome Research. *mSystems* 3, 3 (2018). https://doi.org/10.1128/mSystems.00031-18 arXiv:https://msystems.asm.org/content/3/3/e00031-18.full.pdf

[47] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica* 22, 3 (2012), 276–282.

[48] Barbara A Methé, Karen E Nelson, Mihai Pop, Heather H Creasy, Michelle G Giglio, Curtis Huttenhower, Dirk Gevers, Joseph F Petrosino, Sahar Abubucker, Jonathan H Badger, et al. 2012. A framework for human microbiome research. *nature* 486, 7402 (2012), 215.

[49] Drashko Nakikj and Lena Mamykina. 2017. A park or a highway: Overcoming tensions in designing for socio-emotional and informational needs in online health communities. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1304–1319.

[50] Oded Nov, Mor Naaman, and Chen Ye. 2010. Analysis of participation in an online photo-sharing community: A multidimensional perspective. *Journal of the American Society for Information Science and Technology* 61, 3 (2010), 555–566.

[51] Oded Nov and Chen Ye. 2010. Why do people tag?: motivations for photo tagging. *Commun. ACM* 53, 7 (2010), 128–131.

[52] US National Library of Medicine. [n. d.]. PubMed. https://www.ncbi.nlm.nih.gov/pubmed

[53] Lauren Westendorf Orit Shaer, Oded Nov and Madeleine Ball. 2017. Communicating Personal Genomic Information to Non-experts: A New Frontier for Human-Computer Interaction. *Foundations and TrendsÂő in HumanâĂŞComputer Interaction* 11, 1 (2017), 1–62. https://doi.org/10.1561/1100000067

[54] Oura. [n. d.]. Oura. https://ouraring.com

[55] Vineet Pandey, Amnon Amir, Justine Debelius, Embriette R Hyde, Tomasz Kosciolek, Rob Knight, and Scott Klemmer. 2017. Gut instinct: Creating scientific theories with online learners. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. ACM, 6825–6836.

[56] Gisele Lobo Pappa, Tiago Oliveira Cunha, Paulo Viana Bicalho, Antonio Ribeiro, Ana Paula Couto Silva, Wagner Meira Jr, and Alline Maria Rezende Beleigoli. 2017. Factors associated with weight change in online weight management communities: a case study in the LoseIt Reddit community. *Journal of medical Internet research* 19, 1 (2017).

[57] Albert Park and Mike Conway. 2018. Harnessing Reddit to Understand the Written-Communication Challenges Experienced by Individuals With Mental Health Disorders: Analysis of Texts From Mental Health Communities. *Journal of medical Internet research* 20, 4 (2018).

[58] Umashanthi Pavalanathan and Munmun De Choudhury. 2015. Identity management and mental health discourse in social media. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 315–321.

[59] Laura R Pina, Sang-Wha Sien, Teresa Ward, Jason C Yip, Sean A Munson, James Fogarty, and Julie A Kientz. 2017. From personal informatics to family informatics: Understanding family practices around health monitoring. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2300–2315.

[60] Igor A Podgorny, Matthew Cannon, Chris Gielow, and Todd Goodyear. 2015. Real time detection and intervention of poorly phrased questions. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2205–2210.

[61] NIH Human Microbiome Project. [n. d.]. HMPDACC Publications. https://www.hmpdacc.org/hmp/publications.php

[62] Reddit. [n. d.]. r/HumanMicrobiome. https://www.reddit.com/r/HumanMicrobiome/

[63] Antonio Regalado. 2019. More than 26 million people have taken an at-home ancestry test. Retrieved March 31, 2019 from https://www.technologyreview.com/s/612880/more-than-26-million-people-have-taken-an-at-home-ancestry-test

[64] Seung Yon Rhee, William Beavis, Tanya Z Berardini, Guanghong Chen, David Dixon, Aisling Doyle, Margarita Garcia-Hernandez, Eva Huala, Gabriel Lander, Mary Montoya, et al. 2003. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic acids research* 31, 1 (2003), 224–228.

[65] Julie M Robillard, Thomas W Johnson, Craig Hennessey, B Lynn Beattie, and Judy Illes. 2013. Aging 2.0: health information about dementia on Twitter. *PLoS One* 8, 7 (2013), e69861.

[66] Jessica Schroeder, Chia-Fang Chung, Daniel A. Epstein, Ravi Karkar, Adele Parsons, Natalia Murinova, James Fogarty, and Sean A. Munson. 2018. Examining Self-Tracking by People with Migraine: Goals, Needs, and Opportunities in a Chronic Health Condition. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 135–148. https://doi.org/10.1145/3196709.3196738

[67] SelfNutritionData. [n. d.]. SelfNutritionData. https://nutritiondata.self.com/mynd/mytracking/welcome?returnto=/mynd/mytracking

[68] 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. *J Med Internet Res* 17, 6 (12 Jun 2015), e146. https://doi.org/10.2196/jmir.4415

[69] 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 (2015).

[70] Catherine Arnott Smith and Paul J Wicks. 2008. PatientsLikeMe: Consumer health vocabulary as a folksonomy. In *AMIA annual symposium proceedings*, Vol. 2008. American Medical Informatics Association, 682.

[71] Melanie Swan. 2009. 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 (2009), 492–525.

[72] Melanie Swan. 2013. The quantified self: Fundamental disruption in big data science and biological discovery. *Big data* 1, 2 (2013), 85–99.

[73] Jessica Vitak, Katie Shilton, and Zahra Ashktorab. 2016. Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 941–953.

[74] Gang Wang, Konark Gill, Manish Mohanlal, Haitao Zheng, and Ben Y Zhao. 2013. Wisdom in the social crowd: an analysis of quora. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 1341–1352.

[75] Mingxun Wang, Jeremy J Carver, Vanessa V Phelan, Laura M Sanchez, Neha Garg, Yao Peng, Don Duy Nguyen, Jeramie Watrous, Clifford A Kapono, Tal Luzzatto-Knaan, et al. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature biotechnology* 34, 8 (2016), 828.

[76] Lauren Westendorf, Orit Shaer, Christina Pollalis, Clarissa Verish, Oded Nov, and Mad Price Ball. 2018. Exploring Genetic Data Across Individuals: Design and Evaluation of a Novel Comparative Report Tool. *J Med Internet Res* 20, 9 (24 Sep 2018), e10297. https://doi.org/10.2196/10297

[77] Lauren Westendorf, Orit Shaer, Christina Pollalis, Clarissa Verish, Oded Nov, and Mad Price Ball. 2018. Exploring Genetic Data Across Individuals: Design and Evaluation of a Novel Comparative Report Tool. *Journal of medical Internet research* 20, 9 (2018).

[78] Mark Whooley, Bernd Ploderer, and Kathleen Gray. 2014. On the integration of self-tracking data amongst quantified self members. In *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014-Sand, Sea and Sky-Holiday HCI*. BCS, 151–160.

[79] Sanghyun Yoo, Jonatan Lemos, and Ed Finn. 2017. Nellodee 2.0: A Quantified Self Reading App for Tracking Reading Goals. In *Learning and Collaboration Technologies. Technology in Education*, Panayiotis Zaphiris and Andri Ioannou (Eds.). Springer International Publishing, Cham, 488–496.