Understanding Users Information Needs and Collaborative Sensemaking of Human Microbiome Data

ANONYMOUS AUTHOR(S)*

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 omic 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: reviewing an annotated report, comparing microbiome data, tracking changes, receiving personalized actionable information, curating information, controlling information, and documenting and sharing self experiments. 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

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 associated with digestion, immune response, and other aspects of human health [30]. 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, 48]. 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 group of roommates who share a living space where their commensal microbiota can be influenced by shared lifestyle elements such as nutrition and pets, who might seek to investigate how changes in their lifestyle and environment (e.g. diet, new pet, seasons, new furniture) might impact their microbiome and as a result, 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 studies investigated the information needs of personal genomic non-expert users [29, 51, 59] 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.
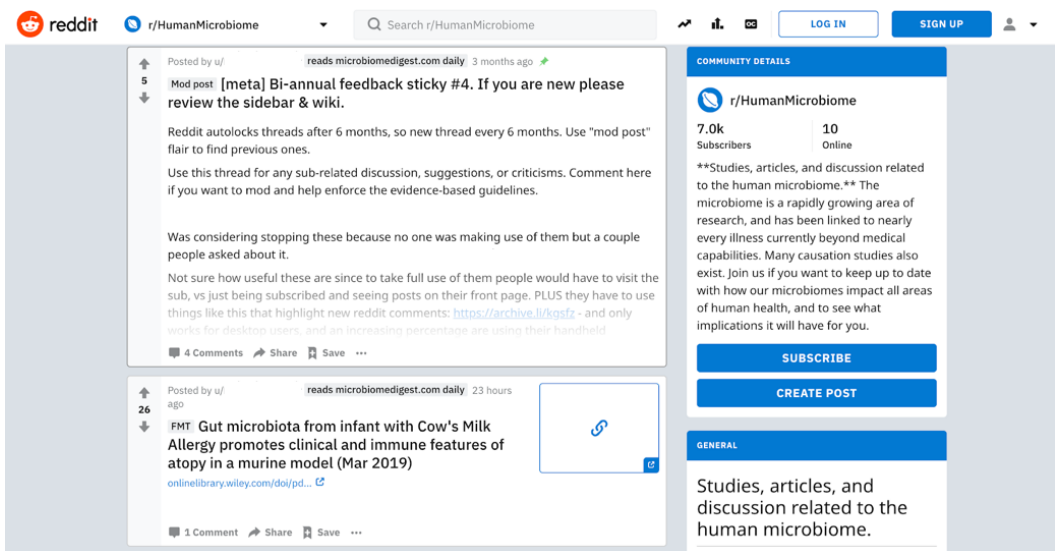
Fig. 1. the r/HumanMicrobiome subreddit page. Image was taken in March 2019. (https://www.reddit.com/r/HumanMicrobiome/).

In this paper, we report findings from a study of the subreddit r/HumanMicrobiome (see Figure 1) that engages users in discussions related to their microbiome and its health implications. In particular, our investigation focuses on three questions:

RQ1) Who are the users - what are their characteristics and motivations?
RQ2) What are their information needs?
RQ3) 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 [7, 32]. This study contributes toward this goal by promoting the understanding of user motivations, information and sensemaking 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 begin with a survey of related work followed by presenting methods and results from the study of subreddit r/HumanMicrobiome. We conclude with a discussion of the findings, implications for design, and future work.

2 RELATED WORK

To lay the foundations for our research, we examined the study of the human microbiome, as well as the areas of personal informatics, social sense-making in data curation, and online health communities.

2.1 The Study of Human Microbiome

The past decade has seen a surge of research into understanding human microbiota and its relevance to health. The Human Microbiome Project (HMP), established in 2008 and funded by the National Institute of Health (NIH), was a research initiative with the aim of generating resources that enable a comprehensive characterization of the human microbiome and analysis of its role in human health and disease [37]. This initiative resulted in the sequencing of over 2,200 reference strains from 5 major body sites: the oral cavity, nasal cavity, skin, gastrointestinal tract, and urogenital tract, of 300 healthy adults between the ages of 18 and 40. In the following years, the HMP has published findings exploring the relationship between the microbiome and various disease states, including inflammatory bowel disease (IBS) and type 2 diabetes.

While there continues to be an evolving scientific understanding of the microbiome and its relevance to health, 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 [35]. Commercial services also offer direct-to-consumer microbiome analysis, including 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]. These commercial products aim to use individual microbiome data to provide personalized analyses and recommendations for improving health.

With the increasing availability of personal and public omic datasets, there is a 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 [4]. Gut Instinct is a system that hosts online learning materials and engages learners in collaborative brainstorming about potential influences on people's microbiome with the dual objective of integrating citizen science with online learning [41]. Another approach of personal omic data is provided by Helix, a commercial service which offers

an array of "DNA-powered products" that allow non-experts to engage with their genomic data through products like personalized fitness and nutrition insights, ancestry data, and curated wine recommendations based on taste preferences [20].

2.2 Personal Omics as Personal Informatics

We consider the self study of personal omics (genomes and microbiomes), within the context of the growing field of personal informatics, which refers to communities, practices, and systems that help people collect and reflect on their personal information [9, 31]. The increasing availability of low-cost sensors has accelerated the practice of self-tracking as well as the rise of the Quantified Self movement [55]. Commercial and research efforts have resulted in numerous self-tracking technologies and applications for health and wellness [28, 54] and the sharing of the data gained by self tracking [7]. Self-tracked personal data includes a wide range of activities and symptoms including: fitness data, sleep times, reading habits, food and liquid intake, phone usage, different physical and emotional symptoms etc.

A common assumption in quantified self research and practice is that an individual's knowledge of their data facilitates reflection can lead to self-discoveries and to behavioral and lifestyle changes. Li et al. [31] 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 to differentiate between stages of reflection [14, 61] and to characterize the barrier and challenges, toward the adoption and effective use of self-tracking technologies, such as data integration, communicating uncertainty, unsuitable visualization and analytics tools, poor skills for analyzing data, and fragmented data scattered across multiple platforms [5, 8]. It is important to note that most of the research identifying practices and barriers in personal informatics has been conducted with expert users, quantified-selfers who are early adopters, health enthusiasts, or patients.

Personal omics shares the main goals and assumptions of personal informatics - facilitating self-discovery based on personal information. However, its interpretation, and related implications for the user's health, are dynamic, as scientific knowledge is evolving. Its complex and sensitive nature requires a layer of curation and guidance that are not necessary for the exploration of most other types of personal informatics. In addition, while personal omic 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 wellbeing of its owner as well as of relevant others. Personal omics should also be considered within the context of Biological Citizenship [10], which explores connections between biology and self-identity, as well as empowers individual to "take care of their own health" [10, 22, 33]. Considering the complex and sensitive nature of personal omics, its scale, its social context and health implications, as well as the dynamic nature of its interpretation, we view personal omics as a new frontier for personal informatics.

2.3 Social Sense-Making in Data Curation

Upon the constant rise in databases capacity and magnitude, the term 'data curation' has been introduced and defined as "the active and ongoing management of data through its lifecycle of interest and usefulness to scholarship, science, and education; curation activities enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time" [12]. The rise of social computing transformed and reconceived data curation as a social activity [19, 25] - co-curation (or social curation).

In recent years, social curation sites such as Flickr and Pinterest allowed users to create multitude of object and data collections which can be shared and discussed in social media [2, 18, 39, 40]. 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 [13]. One example is the Gene Ontology (GO) which is widely used for expert annotation of molecular attributes of genes and gene products [6, 21, 49]. 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 [58]. 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 [58].

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 [45] in addition to a rise of co-curation in health and medical contexts by patients and other non-experts. Websites like TuDiabetes, Patients Like Me, and Eat.ly help thousands of individuals make sense of their experiences and conditions by presenting, sharing, and commenting on health knowledge [17, 60]. These websites can elicit new concepts for healthcare vocabularies, coding sets, and classifications [53], 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 [34].

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 pervasive omic 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.

2.4 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 such as seeking advice and support, asking questions in order to makesense of information, disseminating relevant literature, sharing personal experiences with experts and non-experts, improving understanding of symptoms and professional diagnoses, fueling professional engagement, and promoting management of chronic symptoms. 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. [23] studied users' needs and requirements in online health communities, developing personas to illustrate the different ways people use such communities. Robillard et al. [50] studied how information about dementia is discussed and disseminated on Twitter. They found that a majority of tweets contain a link to news and health information sites, and that a large number of tweets discussed recent research findings. Their results highlight a need for multidirectional engagement between experts and non-expert users to discuss research advances. Park et al. [43]

harnessed the Reddit platform to investigate how written communication challenges manifest in online mental health communities focusing on depression, bipolar disorder, and schizophrenia. Their results promote the understanding of written communication challenges among individuals suffering from mental disorders. Pappa et al. [42] investigated activity behavior and posts content on the active Reddit community LoseIt (r/loseit), a weight management community. Their findings suggest that among the active users of the community, self-declaration, and engagement in discussions are associated with greater weight loss. Several studies highlighted the tensions between informational and socio-emotional needs in OHCs [38], Mayara et al. [11], and [26].

Our investigation focuses on r/HumanMicrobiome, which is different from other OHC because it centers around a specific type of quantitative data (microbiomics) which enables a rapidly growing area of research, rather than particular medical conditions or goals. The human microbiome has a broad impact on human health, users are invited to join the group to learn more and to explore implications of the human microbiome to their own health.

3 METHODS

r/HumanMicrobiome (see Figure 1) is a subreddit, a community on the Reddit social media platform, dedicated to the discussion of the human microbiome and its broad health implications. Within this community users can post new items as well as comment and vote on posts. The community is defined as "a science/evidence based sub" where "usage of citations to support your claims is highly encouraged and may become mandatory" [47]. Health or medical discussion related to microbiome is allowed but is restricted if dominating the discussion. Posts are tagged with a topic flair and can be sorted by flairs (topics).

r/HumanMicrobiome is one community in a collection of microbiome-related subreddits, where each subreddit has different aims. /r/microbiology for is a general forum for discussing the study of eukaryotes, fungi, protists, prokaryotes, viruses, and prions. It includes, among other topics, technical discussion about microbiomes. r/ScientificNutrition is dedicated to a more general discussion on diet and nutrition. /r/prebiotics contains information on feeding the human microbiome, and /r/FMTDatabase is a subreddit where users who seek Fecal Microbiota Transplants can connect with potential donors.

3.1 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 dataset had Reddit data from 2015 to November 2018.

To answer RQ1, who the users of r/HumanMicrobiome are and what are their characteristics and motivations, 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 characterized by certain 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 need, we analyzed the number of posts per topic flair and grouped flairs into high-level topics (see Table 2). We then used content analysis methods to analyze all posts and comments. First-level codes were developed iteratively based on literature [23, 24] and from preliminary review and discussion of the data by two independent coders. Then two coders tagged 393 posts and 3,880 comments 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. 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 [16, 36]. Table 1 provides a list of codes with a definition and sample quotes for each code. Using affinity diagramming, we then 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 as means for clarifying concepts, elucidating meaning from data, and gathering insight.	*I would also like to know if there is rebounding, not that I plan to run out and consume these things. I'm also curious as to what other preservatives interact negatively with the gut biome. Wasn't there just a study a few weeks ago that said diet and environment matters more than genetics?*
Knowledge sharing	An exchange which consists of explanations, claims and conclusions, based on what the author believes is objective knowledge.	*It's possible that causation goes entirely in the opposite direction (mental illness changes gut biome), but to say there's "no credible evidence" that the microbiome might be able to cause mental illness is not accurate. Correlation does not *necessarily* imply causation, but when you see correlation that is still evidence of causation, relative to not seeing correlation, unless you have further evidence against causation...*
Data challenges	Expressing challenges interpreting, making sense of, and utilizing human microbiome data.	*Hey, i know itś been mentioned many times here that UBiome results are very hard to interpret and little is known yet about the gut microbiome to make claims.. but... I was wondering if you guys could take a look at my results. They feel completely useless to me ...[link]*
	Continued on next page	

Table 1 – continued from previous page

Code	Description	Sample Quote
Self Experiments	Sharing experiences of conducting self experiments by systematically changing diet and/or other lifestyle factors.	*What's funny is 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.	*What are the specific probiotics you would recommend for my results? I advise you to ask yourself this: What is it you personally hope to learn? Will that information have a positive impact of your health and well being? Can the information you are looking for be found through alternative sources?*
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". I am a huge supporter of learning more about the microbiome, so I understand that this can seem contrarian. I feel that learning more about the impacts it has on disease is so important, but until we can be absolutely sure that information is protected and wont be penalized by insurance companies years down the road I would advise caution.*
Symptom sharing	Sharing symptoms and medical experiences.	*Amoxicillin and clarithomycin for 2 weeks. But this was right after taking penicillin for pneumonia. Both bacteria dead and gone. After that I started getting chronic fatigue, swollen joints, back pain, IBD.*
Goals and hopes	Sharing goals and expectations of engaging with microbiome data.	*To be honest I was hoping for a miracle. I had exhausted so many years of my life being sick and seeing Dr. after Dr. with no answers. I would've been happy with any improvement in my quality of life.*

3.2 Ethical Considerations

In this study we use content analysis methods to study social media content shared on Reddit's Human Microbiome community. We chose to study this particular community for several reasons, including its focus on the burgeoning area of human microbiome data and its implications, the active discussion it facilitates, and its publicly available content. While the CSCW community is in early stages of discussing and forming guidelines for conducting research using public social media [56], we acknowledge the ethical concerns associated with our methods. First, contributors to this

subreddit discussion post public content to what they may perceive to be a discussion within a group. They might not be aware that the content they share could be used for academic research [15] and they did not give explicit permission for using the data for this purpose. Second, while publicly shared, some of the posts contain sensitive information about users' health conditions and symptoms which the contribution might not want to see amplified through inclusion in a research paper. Third, while usernames on Reddit could be independent from users' real world identity, Reddit cannot and does not guarantee anonymity. However, users might perceive the discussion as anonymous.

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 sensemaking 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.

4 RESULTS

4.1 Who are the participants?

From June 2017 to November 2018 there were 153 unique post contributors and 574 unique comment contributors to r/HumanMicrobiome. Based on the presented number of subscribers to r/HumanMicrobiome in November 2018, 5K, our data suggests that less than 13% of r/HumanMicrobiome subscribers are active contributors. Analyzing 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 [46] or followers [57], we refer to the top 1% (6/613) contributors as super users. This group includes the moderators and founders 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. Figures 2 and 3 show 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, other contributors, and experts.

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 sometime correct misinformation posted by other users, and steer discussion in the right direction if it diverge 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.

Most of the other contributors on r/HumanMicrobiome join the discussion because they have health issues that they believe to be attributed to a gut microbial imbalance. They are motivated to learn more about the human microbiome in general and on scientific results that relate to their condition, in hopes of identifying discovery and treatments that will improve their condition. In the words of one user *"I have IBS, Microscopic Colitis and Sibo. I believe all my problems are microbiome*

related." 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?"* Most of their contributions are questions, knowledge sharing, and recommendation requests.

A small subset of contributors (less than 1%, 5/613) includes 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 *"I am part of an academic research group. We have had several of our students participate in the uBiome SmartGut trial. The tests showed factors contributing to UC and Crohn's about 85% of the time. The false alarm rate of this test seems quite high. We are wondering how practitioners are using this test in clinics and wonder if we simply got a bad batch of processing results."* Most of the experts' contributions include sharing knowledge and references.
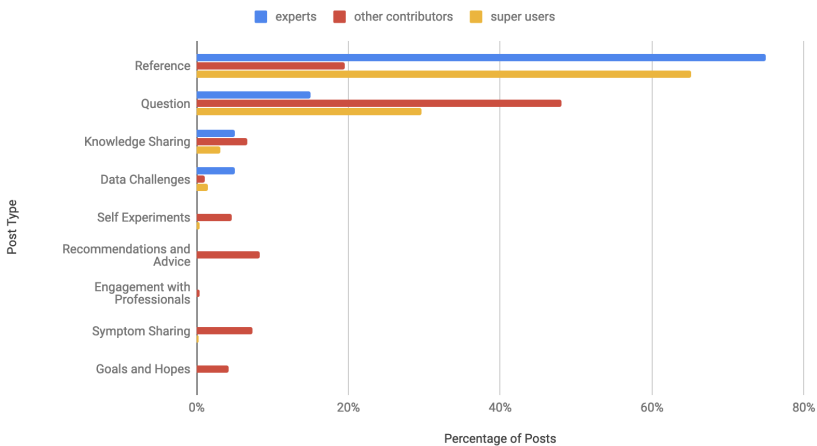

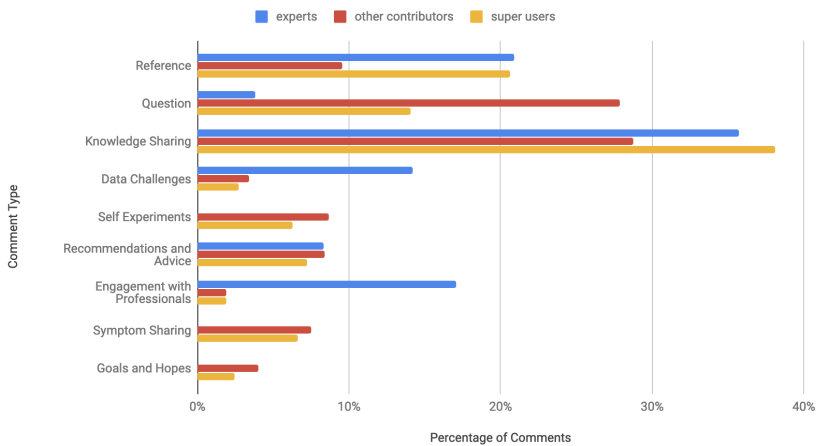Fig. 2. Normalized average breakdown of post types by experts, super users, and other contributors.

Fig. 3. Normalized average breakdown of comment types by experts, super users, and other contributors.

4.2 What topics are users interested in?

r/HumanMicrobiome posts are tagged with various flairs, indicating that users have a broad interest in the human microbiome. For example, posts under the flair testing focus on how to obtain and interpret microbiome data, posts tagged as antibiotics mostly center on recent research about the impact of antibiotics on the microbiota, and posts tagged as discussion serve as an invitation for users to share their own experiences and opinions, and are typically tagged with an additional flair such as probiotics. Table 2 shows the list of topic flairs, for each flair it lists the number of posts, and the total number of comments.

Table 2. Topic Flair Frequencies

Topic	# of Posts	# of Comments
Aging	3	17
Antibotics	14	35
Causation	4	13
Discussion	31	310
FMT	17	121
Fungus	4	17
Impact of Genetics	1	3
Peptides	5	6
Phages	9	24
Probiotics	23	106
Review	13	23
Sex Differences	1	8
Small intestine/upper G/IBS	2	8
Testing	2	36

4.3 What are users' information needs?

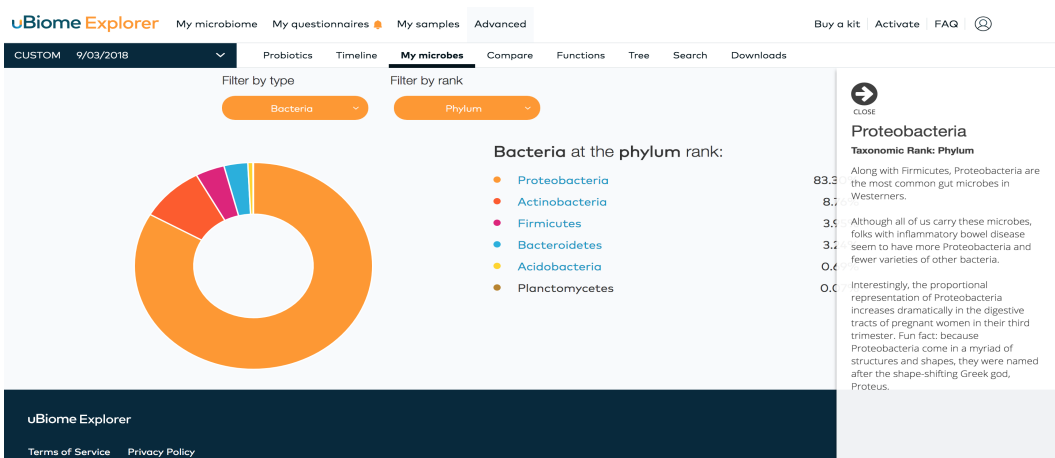Fig. 4. User's Ubiome Dashboard. The 'My microbes' tab reveals the bacterial percentage composition of a sample by phylum taxonomic classification. The drawer on the right provides background information to the bacteria. The banner at the top left labels the sample as 'CUSTOM' as opposed to the user's unique sample type input - "Kitchen Top". Image taken in March 2019. (https://explorer.ubiome.com/my-microbiome).

4.3.1   Quantifiable microbial data. Quantifiable microbial data (see Figure 4), which is obtained through microbiome testing, plays a central role in this community's exploration, comparison, and sensemaking. 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 may do it again just to compare, but not sure. I guess my problems was I didn't know what to do with all the data." 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 regarding your results."*

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...I've now completed 4 ubiome tests so I feel like I have somewhat of a baseline to go off of ... I'd be willing to pay someone that's well versed in this are to take a look at my ubiome results and help me come up with an effective strategy."*

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."*

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. Based on my results the recommendations they gave me were to eat prebiotic foods - which all harm me greatly."* Others go on to describe the limitations of testing: *"I strongly disagree about the personal testing idea. 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 (ie by health status, age, diet, etc.) that allow for someone to make these comparisons with their own data. Also, they don't provide enough information to make a proper statistical comparison."*

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. For me I definitely need everything under one roof, I'm gonna start trying out some of the other programs out there and I'll let you know if I found anything notable, cuz 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."*

Considering these themes, we identified six common information needs: reviewing an annotated report, comparing microbiome data, tracking changes, receiving personalized actionable information, curating information, and controlling information. Here we elaborate on each of these needs:

IN1) 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 and actionable aspects of the information.

IN2) Comparing microbiome data - users seek to make sense of their microbiome through comparison to others. The following comparisons seem to be most important: healthy individuals with similar demographic characteristics (e.g. age, gender, geography, ethnicity), individuals with similar medical conditions, and individuals with similar diet.

IN3) 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).

IN4) Receiving personal actionable information - users desire information that is personalized based on their microbiome, demographic characteristics, diet, and medical conditions. In particular, users look for actionable information such as concrete recommendation for diet or lifestyle changes, and interventions such as probiotics and supplements that could alleviate their symptoms.

IN5) Curating and controlling information - users expressed a need for collecting, relating, organizing, and storing various data sources including annotated reports, scientific articles, and health related data.

IN6) Controlling information - users highlighted a need for keeping their information private and secure, but at the same time, in some cases, they seek to share it with others (e.g. family, friends, and experts).

4.3.2 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:

"...One doctor finally agreed to give me flagyl and it was extremely helpful and stopped the diarrhea, but I was still having light-medium versions of the other new symptoms. 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..."

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. Hopefully there are some things you haven't tried, though I can imagine you've likely tried and read of many, many things. Those uBiome results are definitely not helpful considering how your health is and how relatively normal the sample numbers are. You imply that phages are the answer. Have you also considered fungi? I hope there's a more accessible or diverse solution for you there. You mention problems with protein and fats. I realize you've probably gone through many options, but what about stomach acid supplements? I see vinegar there, but maybe it isn't enough? Or perhaps it's too much? And what about bile salts for fats?"*

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 ..."*

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 and will often discuss the most effective means to attain an "ideal donor". Others are more skeptical of FMT and are more likely to stick to self-experimental procedures

related to diet changes. The diets most discussed are the carnivore or all-meat diet, keto, or only vegetable diet. The degree to which individual share that a diet works them is highly variable. These themes highlight an additional information need:

> IN7) Documenting and sharing self experiments - users seek to document and share details of their self experiments in order to track progress as well as to receive feedback and suggestions.

4.3.3 Engagement with Health Care Providers. A common theme that emerge from examining users' posts and comments about their engagement with healthcare 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. Basically leaves me with way more questions than I started with, but it's always good to keep learning."*

Others find health care providers largely unhelpful, in many cases when the user identified based on self research a procedure that they believe would benefit them:*".... most of the doctors I've seen are completely clueless regarding FMT and the gut microbiome. And some of the primary doctors I know who are familiar with FMT in my city won't see me because of strict requirements that the patient has c.diff, or because of insurance issues ..."*

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 health initiatives. This highlights a need to enhance the communication between patients and healthcare providers regarding emerging research and practices, as well as patient's documentation of self experimentation.

4.4 How users in this community interact with each other to collaboratively make sense of human microbiome data?

4.4.1 Learning and Knowledge Sharing. Members of this community engage in collaborative learning, sensemaking, 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."*

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. Discussion of this type revolves around users educating each other on current microbiome research to support or refute a claim related to its connection to health. For example, *"Study shows association between gut microbes and brain structure in people with irritable bowel syndrome [link]."*

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? There are a lot of diets out there that say they will improve your gut
bacteria but there is very little evidence showing that they actually lead to improvements."*

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. The moderators of the subreddit regularly update the extensive
wiki that lists numerous studies and aggregated information related to the gut microbiomes'
influence on health. Community members are often referred to the wiki in response to posts and
comments seeking information in the areas of diet, effects of antibiotic use, FMTs, microbiome
testing, prebiotics and probiotics: Q: *""Is there any beneficial virus in our body/bowels that could be
knocked out by antibiotics?"*, A: *"Yes. From the wiki [link].""*

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 ..."*

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.
For example, *"for everyone downvoting, please use rational, evidence-based arguments rather than
passive, silent votes."*

4.4.2   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 personal dataset.

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 screenshots, 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] Metabolism stuff
is weird, some stuff low some high... am I supposed to raise the low stuff and lower the high stuff in
some fashion?? Is that even possible? "*

Community members then offer concrete recommendations and advice, pointing the user to
resources and share their own experiences. Concrete recommendations range from proposing
changes in diet, to taking certain probiotics and prebiotics, to procedures like FMT. 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. *Faecalibacterium*, *Roseburia* and *Lachnospira* are butyrate-producing genera, and
they feed on resistant starches and various kinds of fiber, which is correlated with my diet. *Bilophila*
is linked with a higher meat/fat diet, and it's been non-existent in my samples."*

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. I seem to harbor a lot of the butyrate producers (*Faecalibacterium*, *Roseburia*). But an
interesting thing is the relatively high amount of *Fusicatenibacter*. Here a very [interesting article]."*

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."*),
draw attention to the dearth of knowledge in the scientific literature (*"Again, this is not how this
works. You cannot simply invert the onus of proof. The default assumption is never that a new therapy
is safe/efficient, but rather that it is not, until proven otherwise. And this proof (as provided through
rigorous, controlled clinical studies) is still lacking in the case of FMTs. Your argument suffers from a*

classical logical fallacy: argument from ignorance (that's the technical term, not implying that you are ignorant 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. "), 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. I'm not quite sure how you can both argue that the procedure is simple and should be restricted in its use."*)

4.4.3 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 I've taken recently, along with having other health problems, have left my gut bacteria unbalanced and I need to sort things out again. I'm desperately underweight and and I'm looking for anything that might help me extend my diet even a little bit so that I can regain weight and not feel so malnourished."*

Members often seek those similar to them for guidance in order to minimizes the amount of experimentations 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."* 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 and saved myself a lot of suffering. Stick to changing your diet and exercise. It's the safest bet. And be patient, it will take years /decades to learn to how fix ourselves without causing damage in the process."*

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 usernames 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 discussions in online health communities [23], we studied users' interactions in the r/HumanMicrobiome subreddit. As is common in many online peer production and discussion venues [23, 24], 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 information, and controlling information. These information needs overlap and complement needs identified in the study of personal genomic data user needs [52], reflecting the differences in the data type and implications they have for potential actions to be taken. Specifically, the fact that microbiome are more sensitive to interventions than personal genomics was reflected in users' need in practical advice about potential interventions such as changes in nutrition. Similarly, the ability to compare pre- and post-intervention data was highlighted in the

case of microbiome data, and its importance can be attributed to the differences between biome and genomic data changeability. Social aspects of the data also differed: while genomic data share similarities with biological family members, microbiome similarities are affected by geographical proximity, diet, and medical interventions (RQ2).

A recurring theme in our findings 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 potential 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 healthcare providers as uninformed and conservative in this respect on the other, is a source of tension which better systems design could potentially alleviate. For example, prior research of other domains [24, 50] suggested a need for multidirectional interaction between experts and non-expert as a means to share and discuss research progress and its implications.

The findings of our study 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 driving motivation for users seeking and discussing microbiome data was to alleviate health symptoms such as chronic fatigue, IBS and SIBO (RQ1). Sharing symptoms and experiences with probiotics was a common way for users to find others similar to them, and was used as a way to learn from relevant others' experiences (RQ3). Our findings, of open exchange of personally sensitive experiences, suggests that it is possible that the complexity of microbiome data, and users' sense of inadequate response from their care providers as they seek to alleviate their concerns, lead to a relatively lax approach to personal information disclosure. We note that this is consistent with prior research on information disclosure in health-related venues on social media [3, 27, 44].

The observed discussion underscored the social aspect of people's perceptions of personalized medicine: 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). And while users were open to experimenting with a wide variety of methods to improve their microbial imbalance, they also reported that the information and tools provided by direct-to-consumer companies have yet to reach a level of maturity where it is understandable and actionable, and offered a cautious view of the potential effectiveness about new procedures. Users sought personalized recommendations and ways to compare themselves to those that are healthy in addition to those who were similar to them (RQ3). This echos a need of personal genomic users, who seek to compare the genomes to biological relatives [60]. Symptom sharing in particular was the most common way to interact in the group and to find similarities (RQ3). However, not everyone in the community has had their microbiome sequenced so oftentimes symptoms is all they had to go by. This illustrates that there are many needs still unmet by the tools currently available for exploring microbiome data.

5.1 Implications for Design

Based on our findings, we offer a number of design recommendation. People's exposure to direct to consumer microbiome data is a new phenomenon, and therefore there is no accumulated knowledge on best practices for the design of interactive tools for microbiome data communication and exploration. Our analysis of the users' information needs offers a number of design insights in this direction:

5.1.1 Symptoms and experience as focal points for user interaction and knowledge sharing. A substantial portion of users' activity in the reddit discussions involves reporting and commenting on

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 and experiences. Our findings suggest that non experts' information needs reflect the former, whereas the design of current online reports reflects the latter.

Moreover, since much of users' conversation on microbiome data revolves around personal experiences and responses to interventions (primarily changes in diet), external or internal links to relevant personal data and narratives clustered around shared experiences is a design opportunity for embedded social interpretation and sense-making connecting user data- user action- user outcome.

5.1.2 Interaction with healthcare professionals. The contrast between users' own interest in exploring and interpreting microbiome on one hand, and their perception of healthcare providers as uninformed and conservative in this respect on the other, suggests a need for information tools that can be used to better facilitate interactions between patients and their providers. Specifically, augmenting data with references to relevant research literature would be helpful to patients in two ways: understanding the broader complexity and gaps in current research can adjust expectations for clinical use, and contextualizing personal experience to general knowledge helps patients negotiate a shared understanding of their experiences with their healthcare providers.

While anecdotal experiences and narratives of other patients which are shared on social media may be interesting to patients, they are less likely to be viewed as useful to healthcare professionals. From a design perspective, users' ability to bridge self experience, social experience and relevant research, with the option to toggle between them based on their interaction partner, could be valuable to users, and should be considered as a design alternatives for omic data reports.

5.2 Limitations

While our work contributes to the understanding of microbiome data users' needs and points to practical design implications, it has a number of limitations. First, in terms of data collection, our data comes from discussions in an online public venue, and therefore it is biased towards input from people who are comfortable with sharing their knowledge and experiences on social media. This excludes from our observed population many people who might like to engage with personal -omic information but not publicly, or otherwise those who may not be aware of such discussion groups. This is particularly relevant given the personally sensitive nature of the data and the health-related experiences they are associated with. Second, scientific research on the relationship between omic data and human health is relatively new, and major discoveries are being made constantly. As a result, the nature of both omic data and their meaning may also change rapidly, making users' information needs and perceptions subject to frequent changes and highly influenced by how new discoveries are communicated to the general public. Finally, our analysis did not include direct observations of how people engage with their microbiome data. Future research can address this through talk-aloud observations and analysis of system log data.

6   CONCLUSION

With the sharp increase in the availability of personal omic data to users, there is a growing need in understanding users' needs in this very personal and senstivie 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 information, controlling

information, and documenting and sharing self experiments. 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.

As we advance our understanding of the information and interaction needs of omic data users, future research is needed to address open questions, develop interactive tools, and evaluate them. Specifically, observing lead users who have studied their microbiome data, going through the obstacles they faced in the process and following inferences from data to action, could inform such future tools. Understanding the social context - how family and community affect the data, its understanding and its communication - would be another step in ensuring a strong fit between user needs and system design.

Beyond the omic domain, the information needs of people exploring their biome data are relevant in other contexts in which people explore complex and personally sensitive data. In particular, domains in which users' personal data is sensitive to their environment, will benefit from deeper understanding user information needs. 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 (openhumans.org), or in different contexts, platforms for sharing other potentially sensitive personal information such as support groups for people with addictions or other conditions requiring interventions and data tracking. Overall, the increasing availability of sensitive personal data with social implications is likely to keep challenging and transforming current desings of online information sharing platforms, and CSCW community is well placed to offer useful insights throughout this transformation.

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] 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.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
[10] Roger Cooter. 2008. Biocitizenship. *The Lancet* 372, 9651 (2008), 1725.
[11] 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.
[12] Melissa H Cragin, P Bryan Heidorn, Carole L Palmer, and Linda C Smith. 2007. An educational program on data curation. (2007).

[13] 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.

[14] 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.

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

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

[17] 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).

[18] 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.

[19] 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.

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

[21] 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).

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

[23] 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.

[24] 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*.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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

[30] 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.

[31] 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.

[32] 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

[33] 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.

[34] 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.

[35] 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

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

[37] 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.

[38] 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.

[39] 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.

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

[41] 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.

[42] 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).

[43] 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).

[44] 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.

[45] 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.

[46] 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.

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

[48] 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

[49] 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.

[50] 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.

[51] 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

[52] 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).

[53] 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.

[54] 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.

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

[56] 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.

[57] 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.

[58] 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.

[59] 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

[60] 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).

[61] 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.