HHS Public Access

Author manuscript

Public Health Genomics. Author manuscript; available in PMC 2019 May 08.

Published in final edited form as:

Public Health Genomics. 2018 ; 21(1-2): 93–99. doi:10.1159/000494381.

Current Social Media Conversations about Genetics and Genomics in Health: A Twitter-Based Analysis

Caitlin G. Allen[a], **Brittany Andersen**[b], **Muin J. Khoury**[c], and **Megan C. Roberts**[d]

[a]Roiiins School of Public Health, Emory University, Atlanta, GA, USA

[b]Division of Emerging Media Studies,Boston University, Boston, MA, USA

[c]Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, GA, USA

[d]The National Cancer Institute, Rockville, MD, USA

Abstract

Background—The growing availability of genomic information to the public may spur discussion about genetics and genomics on social media. Sites, including Twitter, provide a unique space for the public to access and discuss health information. The objective of this study was to better understand how social media users are sharing information about genetics and genomics in health and healthcare and what information is most commonly discussed among Twitter users.

Methods—We obtained tweets with specific genetics- and genomics-related keywords from Crimson Hexagon. We used Boolean logic to collect tweets containing chosen keywords within the timeframe of October 1, 2016, to October 1,2017. Features of the software were used to identify salient themes in conversation, conduct an emergent content analysis, and gather key demographic information.

Results—We obtained 347,196 tweets from our search. There was a monthly average volume of 28,432 tweets. The five categories of tweets included: genetic disorders/disease (45.3%), health (15.6%), genomics (8%), and genetic testing (7.3%). Top influencers in the conversation included news outlets and universities.

Conclusions—This content analysis provides insight about the types of conversation related to genomics and health. Conversations about genomics are occurring on Twitter, and they frequently emphasize rare genetic diseases and genetic disorders. These discussions tend to be driven by key influencers who primarily include news media outlets. Further understanding of the discussions related to genomics and health in social media may offer insight about topics of importance to the public.

Caitlin G. Allen, MPH, Rollins School of Public Health, Emory University, 1518 Clifton Rd, Atlanta, GA 30322 (USA), calle27@emory.edu.

Author Contributions

All authors contributed to the conception and design of the study, as well as the final approval of the content.

Statement of Ethics

This paper is not directly related to human or animal research. The authors have no ethical conflicts to disclose.

Disclosure Statement

All authors have no conflicts of interest to report. The findings and conclusions in this report are those of the authors and do not necessarily reflect the position of the Centers for Disease Control and Prevention or the National Cancer Institute.

Keywords

Genetics; Genomics; Social media; Health information; Twitter

Introduction

Genomic information is increasingly being incorporated into patient care, advertised through direct-to-consumer testing, and discussed in the news and through social media [1–5]. At the same time, social media continues to advance and alter the ways the public accesses medical and health information, shifting how people communicate about and understand their health [6–10]. Today, health information seeking is the third most popular online activity among adults with internet access with 80% of internet users having looked online for information about health topics (e.g., commentary or experience about a health topic, online videos, review of drugs or medical treatment) [8]. Historically, individuals sought information from health-specific websites, but there has been a drastic shift toward the use of social media (e.g., Facebook and Twitter) to gather and share health information [11].

The pace at which access to personalized genomic information is becoming available, combined with the public's regular use of social media to access health information, has created a new space for conversations related to genetics and genomics. Social media (specifically health-related content available on social media) may be an important contributing factor for an individual's beliefs about health conditions and subsequent health behaviors and decision-making [12, 13]. Given the speed at which genomic science has advanced and information is becoming available via social media, we sought to characterize existing conversations about genomics on Twitter, where articles and information are commonly shared in short posts called tweets [11]. The aim of this report was to provide a description of how Twitter users are sharing information about genomics and what information is most discussed among users.

Methods

To better understand user conversations related to genomics and health, we obtained Twitter data from Crimson Hexagon, a data collection and analysis toolkit that includes posts from social media networks (Boston, MA, USA). Twitter is an online micro-blogging platform that is among the most commonly accessed social media sites and used globally by millions of individuals [14]. This platform allows users to post tweets: real-time messages with up to 140 characters (during our study period), images, and videos. Individuals can communicate digital content frequently and quickly by passing information through followers, re-tweets of quotes, and replies to other users' conversations [14, 15]. Each Twitter user has an account (@) and the user can share information categorically with their followers and non-followers via hashtags (e.g., #genomics) or the user's profile. Most conversations are publicly available (90%), providing a rich source of information about people's perceptions and communication about specific topics [16].

We used search terms to collect tweets that employed Boolean logic for increased precision and were any combination of the words "genetic," "genomic," "health," "public health,"

"disease," "disorder," "condition," and "medicine." These terms could have appeared in posts or any accompanying hashtags. We collected tweets posted between October 1, 2016, and October 1, 2017. The yearlong date range offers insight about activity trends.

After having obtained the tweets from Crimson Hexagon, we used features of the software (e.g., Clusters, Topic Wheel) to identify salient themes, analyze emergent content, and pinpoint the most commonly used words in the data set. Crimson Hexagon uses an algorithm that recognizes patterns of words to interpret text and has been tested and validated by the Pew Research Center [17, 18]. The Topic Wheel algorithm analyzed the collection of tweets in the data set and identified the most frequently used words to inductively produce primary categories and subcategories of the most prevalent themes in the data set. We also used the Clusters feature, which employs an algorithm that produces a graph showing which words in a data set are most often used in conjunction with one another. The Clusters illustrated associations between words and provided a further description of key themes in conversation. Both the Topic Wheel and Clusters features of the platform allowed us to assess a random selection of 10,000 tweets for analysis. All themes and topics from Crimson Hexagon were checked by the authors for clarity and consistency.

Crimson Hexagon also described the "emotions" of the tweets in the data set. The program analyzed the tweets based on words commonly associated with positive, negative, or neutral sentiment. Beyond sentiment, the platform's algorithm computed different types of emotion, such as fear and joy. The users in the data set were further analyzed through Crimson Hexagon's user demographic features. Users' gender and age were collected from their profiles, in addition to identifying the place of origin of the posted tweet. Top influencers were also identified, based on the total number of followers and users' reach.

Results

A total of 347,179 tweets were obtained from Crimson Hexagon that met our search criteria. The average monthly volume of tweets was 28,432, with a weekly average volume of 6,549 and a daily average volume of 949 (Fig. 1). Over the course of the year, there were multiple spikes in activity; notably during the week of July 31 to August 5 there were over 30,000 tweets (conversations focused on sharing about rare genetic disorders). On a weekly basis, most of the tweets occurred on Fridays (17%). Thirteen percent of the tweets were categorized as negative and 15% as positive. The tweets categorized as having an "emotion" by Crimson Hexagon (51%) displayed joy (21%), sadness (11%), disgust (9%), and fear (8%) (Table 1; Fig. 2).

Applications of Genetic and Genomic Topics

The five primary categories of themes included: genetic disorders or disease (45.3%), health (15.6%), genomics (8%), and genetic testing (7.3%). Subcategories of the genetic disorder category included rare genetic disorders and causes of genetic disorders. The genetic disease category included heart disease, causes of genetic disease, disease risk, and rare genetic disease risk. When discussing diseases and disorders, tweets referred to rare genetic disorders and considered the causes of these disorders. The health category included discussions of genomic health, healthcare, health risk, and disease. The genomics category

included genomic medicine, genomic health, disease, data, and genetic screening. Genetic testing included disease, genetic health, bills related to genetic testing, and the FDA sequencing program. In the genetic testing category, the tweets included were related to legislation and policy around genetics (e.g., bills, FDA) (Table 2).

Influencers

The top influencers included ABC News, the Associated Press, BBC News, Cambridge University, CNN, ESPN, Forbes, The Guardian, Harvard University, and the Independent. The most popular posts associated with each influencer are listed in Table 3. Most influencers were news channels both in the USA and abroad. Among the top influencers, two non-news sources were included: Cambridge University and Harvard University. Tweets from these sources were related to research discovery. Other themes among influencers included rare diseases.

Discussion

This brief report documents the current health-related conversations about genetics and genomics occurring on Twitter. We found over 300,000 tweets related to this topic on Twitter between October 2016 and October 2017. Considering there are over 200 billion tweets per year, discussion about genetics, genomics, and health makes up only a small portion of the conversation; however, it is important to consider those who are engaged in the conversation and the type of information discussed. This analysis is an initial step in better understanding topics of interest related to genetics and genomics. It demonstrates the potential utility of using social media data to describe conversations and attitudes surrounding this topic.

The top influencers in this data set were news sources. Each of these influencers had between 2 million and 10 million followers, demonstrating the potential wide-spread reach of information they share. These accounts primarily shared information about genetic disorders deemed newsworthy, such as Ehlers-Danlos syndrome, spinal muscle atrophy, and Crohn’s disease. Because social media plays a role in shaping public opinions and conversations, the endorsement of ideas from large groups, like the influencers in this sample, tends to drive social media conversation on a topic [19–22]. Opinions on Twitter evolve rapidly, but once a public opinion stabilizes, it can be difficult to change [22]. Although this analysis is a snapshot in time, it is important to consider how certain topics and opinions about genetics are portrayed through social media. Endorsements and information shared by top influencers do not necessarily represent individuals’ views, but they do offer insight about information being shared about a topic and may have a strong effect on the conversations about a specific topic. In this instance, the conversations related to genetics and genomics were highly focused on rare conditions.

It is not surprising that a large part of content in this data set was related to rare diseases and disorders, as social media platforms can serve as a forum for individuals to share information and experiences about a topic [23]. Given the considerable lack of information available about rare diseases, individuals may turn to Twitter and other social media sites to gather information, share personal stories, and create a community [24].

Another notable theme was a tendency for users to describe how they believed genes caused certain diseases. Individuals tended to overemphasize genetics as a determinant of diseases (e.g., they shared information about the extent to which genes were related to obesity). These discussions often centered on the risk of disease and personal stories. Users also noted concerns about policies and legislation that could influence their ability to receive testing. It is important to consider how the public perceives policies related to genetics and genomics to understand the consequences of genetic testing uptake. Further, there are opportunities for reliable sources, such as health agencies, to strategically engage in and contribute to the conversation related to genomics and public health. For example, tweets that use hashtags have been associated with a higher level of engagement [25].

There are several limitations to consider. First, we only provide observational data and cannot establish associations or causalities. Second, we did not analyze the quality of the information sources or the information provided in the tweets. Other studies have identified misinformation about certain diseases being shared via social media, which influences health beliefs and behaviors [11, 26–28]. Because we did not assess the quality of the information discussed, we were not able to address any differences in spread of information based on the accuracy of the content. Thus, future studies should assess the quality of the information being shared and consider the types of information being shared related to genomics (e.g., what type of information is more widely spread). In addition, while we focused on existing health-related internet use that may influence causal beliefs about genomics and health, we did not evaluate how individual users came to use Twitter in communicating ideas. We were also unable to identify if and how this information may impact causal beliefs and subsequent health behaviors. Our search strategy was targeted with broad search terms to capture the majority of tweets about genetics, genomics, health, and disease; however, we likely underreported the number of tweets during this period related to genomics and health, as some tweets may not have used any of the specific search terms from our criteria. For example, we did not capture tweets that discussed specific diseases (e.g., leukemia) unless the tweet included the combination of search terms outlined in the Methods section. Finally, the majority of tweets were from the USA, followed by the UK. It is unclear whether these data are generalizable to other cultural settings.

Ongoing research is needed to track public discourse about genetics and other emerging health topics. In addition, future research should consider how these conversations may shape causal beliefs and be shaped by other characteristics (e.g., sociodemographic characteristics). Such research could lead to further theory-building that would inform interventions to improve the dissemination of information about genomics and health. Tracking changes in conversations over time and encouraging researchers to engage in social media to promote reliable scientific information are important steps to ensure that accurate information is available to the public about the rapidly advancing conversations related to genetics and genomics. Our study provides baseline insight about genetics- and genomics-related conversations on Twitter that reflect the current public discourse surrounding the topic. This assessment is a first step toward tracking the public discourse and could lead to better-informed, more precise efforts to intervene in the public discourse related to genomics.

Acknowledgements

There is no funding to report for this study. We would like to acknowledge Sylvia Chou for her comments on the final version of this manuscript.

References

1. Chambers DA, Feero WG, Khoury MJ. Convergence of implementation science, precision medicine, and the learning health care system: a new model for biomedical research. JAMA. 2016 5;315(18): 1941–2. [PubMed: 27163980]
2. Jolie Pitt A Angelina Jolie Pitt: diary of a surgery. The New York Times. 2015 3 24.
3. Jolie A My medical choice. The New York Times. 2013 5 14.
4. Covolo L, Rubinelli S, Ceretti E, Gelatti U. Internet-based direct-to-consumer genetic testing: a systematic review. J Med Internet Res. 2015 12;17(12):e279. [PubMed: 26677835]
5. Hudson K, Javitt G, Burke W, Byers P; American Society of Human Genetics Social Issues Committee. ASHG Statement* on direct-to-consumer genetic testing in the United States. Obstet Gynecol. 2007 12; 110(6): 1392–5. [PubMed: 18055737]
6. Lasker JN, Sogolow ED, Sharim RR. The role of an online community for people with a rare disease: content analysis of messages posted on a primary biliary cirrhosis mailinglist. J Med Internet Res. 2005 3;7(1):e10. [PubMed: 15829472]
7. Gerber BS, Eiser AR. The patient physician relationship in the Internet age: future prospects and the research agenda. J Med Internet Res. 2001 Apr-Jun;3(2):E15. [PubMed: 11720957]
8. Eysenbach G, Jadad AR. Evidence-based patient choice and consumer health informatics in the Internet age. J Med Internet Res. 2001 Apr-Jun;3(2):E19. [PubMed: 11720961]
9. Mohd Roffeei SH, Abdullah N, Basar SK. Seeking social support on Facebook for children with Autism Spectrum Disorders (ASDs). Int J Med Inform. 2015 5;84(5): 375–85. [PubMed: 25701266]
10. Lu Y, Wu Y, Liu J, Li J, Zhang P. Understanding health care social media use from different stakeholder perspectives: a content analysis of an online health community. J Med Internet Res. 2017 4;19(4):e109. [PubMed: 28389418]
11. Fox S The social life of health information [Internet]. Washington (DC): Pew Research Center 2011 5 12 [cited 2017]. Available from: http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/.
12. Waters EA, Muff J, Hamilton JG. Multifactorial beliefs about the role of genetics and behavior in common health conditions: prevalence and associations with participant characteristics and engagement in health behaviors. Genet Med. 2014 12; 16(12): 913–21. [PubMed: 24830327]
13. Glanz K, Rimer BK, Viswanath K. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco: Jossey-Bass; 2008.
14. Twitter [Internet]. Getting started with Twitter. 2016 [cited 2017]. Available from: https://Q8support.twitter.com/articles/215585.
15. Aladwani AM. Facilitators, characteristics, and impacts of Twitter use: theoretical analysis and empirical illustration. Int J Inf Manage. 2015;35(1):15–25.
16. Zimmer M, Proferes N. Privacy on Twitter, Twitter on privacy In: Jones S, editor. Twitter and Society. New York: Peter Lang; 2014 p. 55–67.
17. Hopkins DJ, King G. A method of automated nonparametric content analysis for social science. Am J Pol Sci. 2010;54(1):229–47.
18. Hitlin P Methodology: how Crimson Hexagon works [Internet]. Washington (DC): Pew Research Center 2015 4 1 Available from: http://www.journalism.org/2015/04/01/methodology-crimson-hexagon/.
19. Kamenova K, Reshef A, Caulfield T. Angelina Jolie's faulty gene: newspaper coverage of a celebrity's preventive bilateral mastectomy in Canada, the United States, and the United Kingdom. Genet Med. 2014 7;16(7):522–8. [PubMed: 24357851]
20. Marcon AR, Bieber M, Caulfield T. Representing a "revolution": how the popular press has portrayed personalized medicine. Genet Med. 2018 1;20(9):950–6. [PubMed: 29300377]

21. Mollema L, Harmsen IA, Broekhuizen E, Clijnk R, De Melker H, Paulussen T, et al. Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013. J Med Internet Res. 2015 5;17(5):e128. [PubMed: 26013683]

22. Xiong F, Liu Y. Opinion formation on social media: an empirical approach. Chaos. 2014 3;24(1): 013130. [PubMed: 24697392]

23. Pemmaraju N, Utengen A, Gupta V, Thompson MA, Lane AA. Analysis of first-year Twitter metrics of a rare disease community for blastic plasmacytoid dendritic cell neoplasm (BPDCN) on social media: #BPDCN. Curr Hematol Malig Rep. 2017 12;12(6):592–7. [PubMed: 29064025]

24. Perales MA, Drake EK, Pemmaraju N, Wood WA. Social media and the adolescent and young adult (AYA) patient with cancer. Curr Hematol Malig Rep. 2016 12;11(6):449–55. [PubMed: 26893061]

25. Bhattacharya S, Srinivasan P, Polgreen P. Engagement with health agencies on twitter. PLoS One. 2014 11;9(11):e112235. [PubMed: 25379727]

26. Dredze M, Broniatowski DA, Hilyard KM. Zika vaccine misconceptions: a social media analysis. Vaccine. 2016 6;34(30):3441–2. [PubMed: 27216759]

27. Andersen B, Hair L, Groshek J, Krishna A, Walker D. Understanding and diagnosing antimicrobial resistance on social media: a year-long overview of data and analytics. Health Commun. 2017 [Epub ahead of print]

28. Lee BY. Health misinformation abounds on Pinterest, Instagram and visual social media. Forbes. 2016 1 27.

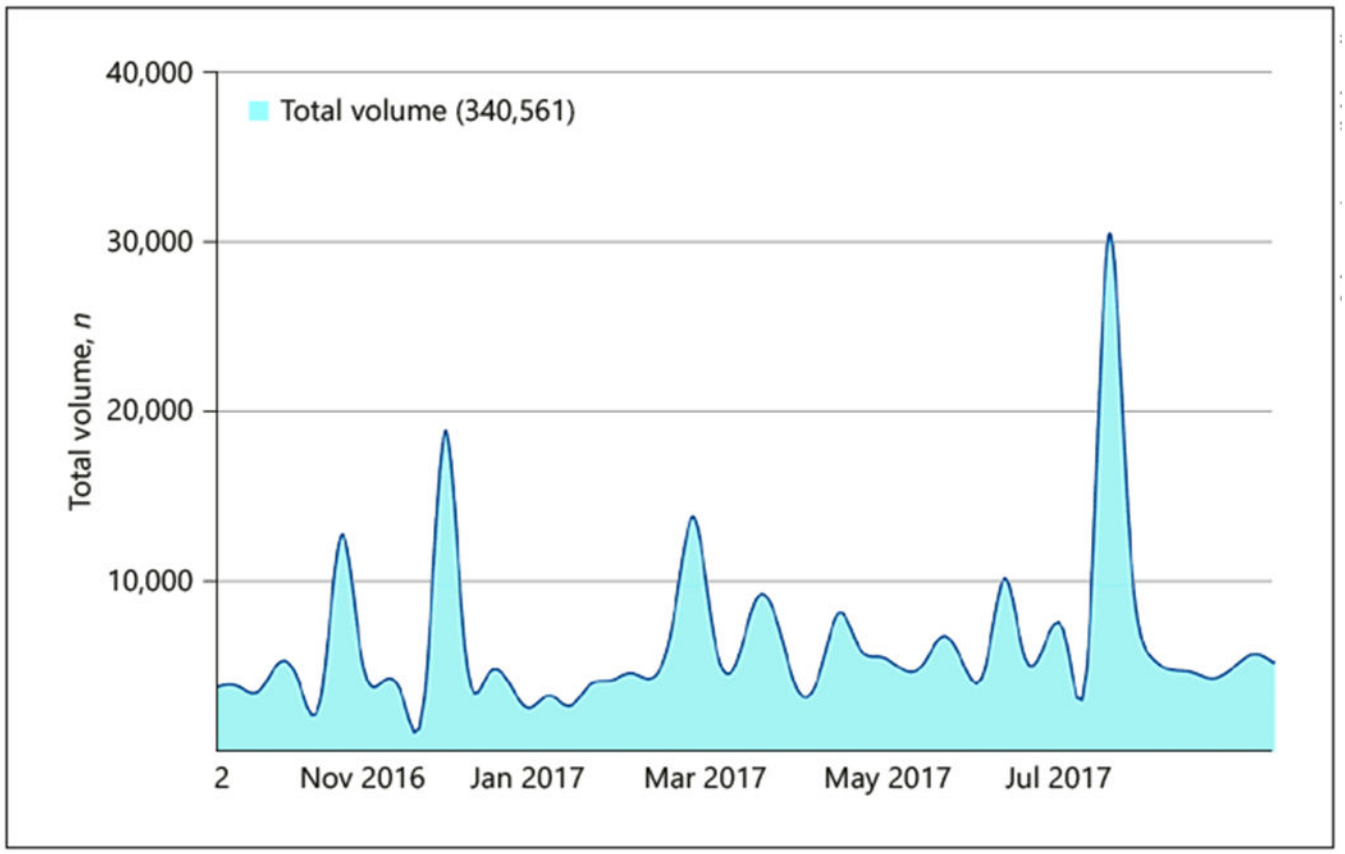

Fig. 1.
Weekly volume of genetics- and genomics-related tweets between October 1, 2016, and October 1, 2017.

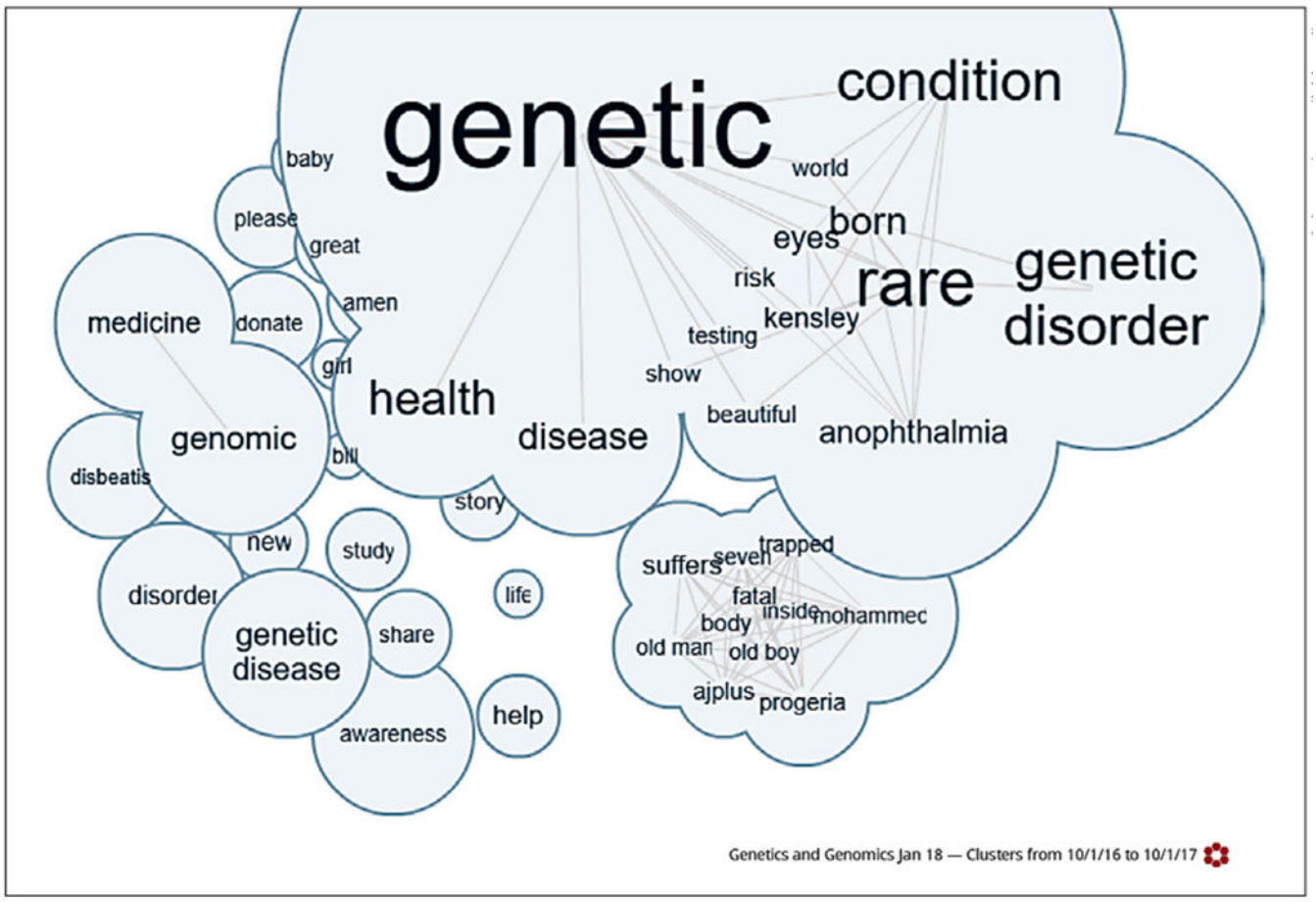

Fig. 2.

Clusters of genetic and genomic topics between October 1, 2016, and October 1, 2017.

Table 1.

Descriptive information about the tweets *(N* = 347,179)

Volume	
Weekly average volume	6,549
Daily average volume	949
Sentiment	
Neutral	248,763 (73%)
Negative	49,318 (15%)
Positive	42,480 (13%)
Emotion	
Joy	72,523 (21%)
Sadness	37,892 (11%)
Disgust	31,541 (9%)
Fear	25,430 (8%)
User gender	
Male	50%
Female	50%
Age	
<35 years	21%
≥35 years	79%
Country of origin	
USA	141,103 (59%)
UK	27,364 (12%)
India	9,316 (4%)
State of origin	
California	15,678 (16%)
New York	11,113 (11%)
Texas	8,500 (9%)
Massachusetts	5,601 (6%)

All information provided by Crimson Hexagon. User information (e.g., on age, gender, country of origin, state of origin) was available only for users that provided this information in their profiles; the total does not equal 347,179. The top posts by topic included: genetic disorder (89,210), genetic disease (78,073), genetic condition (69,533), health MS genetics (41,568), and rare genetic disorder (27,751). The most commonly used words in our sample were “genetics” (315,102), “disorder” (92,526), “disease” (90,819), “rare” (90,393), “condition” (67,820), “born” (41,903), “health” (82,053), “old” (21,855), “eyes” (27,400), and “anophthalmia” (27,000) (Fig. 2).

Table 2.

Combined content analysis: post list and Topic Wheel

Primary category	Subcategories	Examples
Genetic disorder Genetic disease (45.3%)	Rare genetic disorder (causes, awareness about specific disorder)	Inherited genetic disorder corrected in human embryos
	Causes	Explained #Saturday A Stem Cell Platform for Accelerating
	Heart disease	Genetic #Disease Research http://buff.ly/2ogV3XL
	Disease risk	Genome sequencing reveals one in five people harbor a genetic mutation for a rare disease
Health (15.6%)	Genomic health	
	Genetic code	GENOMIC HEALTH: Interpreting the Oncotype DX GPS
	Healthcare	Report #prostate #prostatecancer
	Genetic health risk	Our health is more affected by our zipcode than genetic code.
	Disease	Patients should have the same rights to our genetic information as we do to other types of health information.
Genomics (8%)	Genomic medicine	
	Genomic health	
	Disease	Is the promise of genomic medicine overhyped?
	Genomic data	@BroadInstitute's Eric Lander thinks the answer is both yes and no.
	Genomic sequencing	New! Review by Levy et al. 'Integrating cancer genomic data into electronic health records'
Genetic testing (7.3%)	Disease	
	Genetic health	This PhD student developed a genetic test so doctors can tailor treatment for children with cancer
	Bill	What You May Learn by At-Home Genetic Testing http://ow.ly/zY2B30bUJLr
	FDA	

Table 3.

Most influential authors and related posts

Author name and location	Followers, *n*	Post date (GMT)	Post text
@CNN	37,058,632	8/15/2017 18:45	25-year-old Australian bodybuilder with rare genetic disorder dies eating high-protein diet https://t.co/FTh6Hk8z9J https://t.co/PIusjtCTzU
@espn, Connecticut	30,063,434	11/3/2016 17:22	The Vasher twins were born with a genetic disorder, but also an optimistic spirit. One Michigan star took notice: https://t.co/h2MtRR5SUT https://t.co/socx4UJjiM
@Forbes, New York	13,440,729	8/18/2017 14:45	Dr. Yasmeen Abu-Fraiha is fighting genetic disease in Israel's Bedouin community https://t.co/BvSdC5w0RB https://t.co/oyN0i6zs6O
@ABC, New York	10,088,317	4/19/2017 12:44	Beauty queen uses platform to bring attention to rare genetic disorder called Ehlers-Danlos Syndrome. https://t.co/wgrfvdwBHJ https://t.co/GR0FO4wTEm
@AP	9,294,257	12/24/2016 2:31	FDA approves the first treatment for children and adults with spinal muscular atrophy, a rare genetic disorder. https://t.co/zBCtc9EgqF
@BBCNews, London	8,520,308	7/13/2017 19:38	RT @VictoriaLIVE "We just want to hug them" A father whose baby died of the same rare genetic condition as Charlie Gard speaks of his anguish for the family. https://t.co/QH2wp5faTC
@guardian, London	6,728,627	8/27/2017 7:41	Warnings over shock dementia revelations from ancestry DNA tests https://t.co/s8JnmMgWAi
@Independent, London	2,100,313	1/6/2017 16:09	Man with rare genetic condition who was made into a meme speaks out https://t.co/wHBT11yGvW
@Harvard, Massachusetts	683,079	12/21/016 16:38	RT @harvardmed Heart disease is influenced, but not determined solely by, genetic history https://t.co/4sutqMyEop https://t.co/p62GcoGITJ
@Cambridge_Uni, Cambridge	300,907	1/10/2017 10:03	Researchers have identified a series of genetic variants that affect the severity of Crohn's disease. https://t.co/AyHdjLsKGG