Published in final edited form as:

Med Ref Serv Q. 2013 ; 32(4): . doi:10.1080/02763869.2013.837726.

Using Google Blogs and Discussions to Recommend Biomedical Resources: A Case Study

Robyn B. Reed, MA, MLIS,

Biomedical Informatics and Emerging Technologies Librarian/Assistant Librarian, George T. Harrell Health Sciences Library H127, Penn State College of Medicine, Penn State Hershey, 500 University Drive, P.O. Box 850, Hershey, PA 17033-0850. (rreed4@hmc.psu.edu)

Ansuman Chattopadhyay, PhD, and

Head, Molecular Biology Information Service, Health Sciences Library System, University of Pittsburgh, 200 Scaife Hall, 3550 Terrace Street, Pittsburgh, PA 15261. (ansuman@pitt.edu)

Carrie L. Iwema, PhD, MLS, AHIP

Information Specialist in Molecular Biology, Health Sciences Library System, University of Pittsburgh, 200 Scaife Hall, 3550 Terrace Street, Pittsburgh, PA 15261. (iwema@pitt.edu)

Abstract

This case study investigated whether data gathered from discussions within the social media provide a reliable basis for a biomedical resources recommendation system. Using a search query to mine text from Google Blogs and Discussions, a ranking of biomedical resources was determined based on those most frequently mentioned. To establish quality, these results were compared to rankings by subject experts. An overall agreement between the frequency of social media discussions and subject expert recommendations was observed when identifying key bioinformatics and consumer health resources. Testing the method in more than one biomedical area implies this procedure could be employed across different subjects.

Keywords

Bioinformatics; biomedical resources; social media; subject expertise; text mining

INTRODUCTION

The rapid growth of social media in recent years has generated a wealth of information across a wide range of topics. The goal of this study was to determine the feasibility of utilizing some of these data to identify key biomedical resources when no official ranking is available. The investigation began by employing a database with which the authors have familiarity. The Molecular Biology Information Service of the Health Sciences Library System at the University of Pittsburgh maintains a searchable database of bioinformatics tools on its website. The Online Bioinformatics Resources Collection (OBRC) is an in-house developed, manually-curated tool that helps to support the efforts of biomedical researchers within the University of Pittsburgh and is freely accessible on the web.^{1, 2} This online resource <<http://www.hsls.pitt.edu/obrc>> provides researchers with annotations about and links to several thousand bioinformatics databases and tools to help them achieve specific tasks in molecular biology, e.g., PCR primer design or protein structure prediction.

The number of online bioinformatics tools and software is growing at a rapid pace, thus making manual updates to the OBRC a challenge. In 2013, the Database Issue of *Nucleic Acids Research (NAR)* introduced 88 new molecular biology resources along with updated

information on 88 existing tools. The *NAR* collection of online resources currently features 1,512 molecular biology databases.³ Presently, the OBRC contains links and annotations to more than 2,800 software tools and bioinformatics databases and will continue to grow over time.

Another challenge associated with the OBRC is the absence of a recommendation feature. Feedback from users of the system is generally positive, but the increasing number of available resources can sometimes generate a lengthy list of tools from a search, prompting users to ask which resource is best for their specific task. The OBRC currently orders search results by keyword matching rather than by popularity or recommendations. It would not be feasible to survey librarians or bioinformatics experts to build recommendations of thousands of resources on a regular basis. This study explores the creation of a system for recommending key databases within the OBRC based on discussions in the social media.

Social media and Web 2.0 technologies are widely used across many different fields, such as business and health care, for a variety of purposes. Kozinets et al. explored discussions within online blogs to understand marketing potential within the social media⁴ while Da et al. demonstrated the ability to predict stock price fluctuations following an increased frequency of Google searches on a group of stocks.⁵ Within the medical field, mining data from Twitter microblogs accurately tracked the level of H1N1 influenza in real time.⁶ Furthermore, the use of Web 2.0 technologies has been explored for the purpose of providing quality evidence-based medicine.⁷ In order to measure how National Library of Medicine (NLM) resources are being used, NLM announced a call in February 2013 for the design of a software system that mines social media including Facebook, Twitter, blogs, and other discussions. Reviewing social media entries about NLM tools would capture “sentiment and use patterns of NLM’s ‘customer’” and be used “for comparing NLM to other sources of health-related information.”⁸ While the outcome of this study remains to be seen, the use of social media by NLM is a unique approach to gaining information about resource usage.

Among the various uses of social media found today, some are designed to make recommendations or share advice on a particular topic. Oftentimes, this is achieved through crowdsourcing, or capturing the opinions of several people rather than relying on a single standard. An example of this is StockTwits®, a website for people to follow stocks, discuss financial strategies, and observe market trends in real time.⁹ SocialGuide is another website where trending television shows are displayed in real time.¹⁰ Registered users send messages via Twitter while watching television and an online trending profile is displayed indicating the most popular shows at the time.

The use of biomedical informatics has been employed for locating and recommending biomedical resources. The Biomedical Resource Ontology was used to semantically search web content to discover resources such as software, training programs, and various types of expertise.¹¹ Another study demonstrated the use of Semantic Web technologies and fuzzy linguistic modeling techniques to identify specific biomedical resources from large repositories.¹²

The system described in this pilot project is a unique and simplified approach for identifying frequently mentioned biomedical resources in the social media. The social media rankings were compared to rankings by experts to ascertain the quality of information being discussed. The results generated from this study show a high degree of agreement between the experts and the social media, suggesting this is a valid approach to creating a recommendation system for bioinformatics tools.

METHODS

In order to select bioinformatics resources for the study, five research questions were modeled after common queries that molecular biology information specialists in the Health Sciences Library at the University of Pittsburgh receive from biomedical researchers. Each question listed five possible resources that could be used to answer that question. The resources were chosen from the OBRC and listed alphabetically under each question (see Appendix 1).

Ranking of Bioinformatics Resources

Expert decisions on resources were generated by having two bioinformatics specialists independently answer the questions and rank resources in the order they would recommend them to patrons. Expert opinions were ascertained to determine if the social media results reflected professional preferences on resource selection. Knowing that domain knowledge and training between experts in this study would be varied, experts were asked for reasons for selecting resources the way they did. To maximize possibilities, they were not limited by specific criteria but rather answered using free responses.

Social media rankings of the bioinformatics resources were determined by first writing a search filter to capture discussions about the resources and to ensure that they were in a bioinformatics context. To mimic an automated search in the OBRC, the same filter was used for each resource (see Table 1). The resource was placed in quotation marks (e.g., “UCSC genome browser”) and, in order to be captured by the search, needed to contain the name of the quoted resource along with at least one word or term specific to bioinformatics or molecular biology. Each term is separated by an OR, represented in the table by a vertical line.

The sources of data for the social media values included Google Blogs and Discussions. Due to a relatively low volume of discussion in the social media on bioinformatics topics, no date restriction was placed on the data and discussions were captured “all time.” The total number of blog entries and discussions about the resources were calculated and added together. The rankings were assigned on a scale of 1 to 5 where (1) represented the highest number of discussions about a tool and (5) the least.

Ranking of Consumer Health Resources

To test the feasibility of the key resource identification process for biomedical topics other than molecular biology, consumer health websites were also examined. The same method was employed as seen with the bioinformatics samples with the following changes: two consumer health librarians rated websites by answering three questions about consumer health interests, each with five resources to order, based on how they would recommend them to patrons. (see Appendix 2) The social media rankings necessitated the creation of a new search filter. The dates of data collection were restricted from 2009 to the present.

RESULTS

The results of the bioinformatics study are displayed in Table 2. The five bioinformatics categories investigated are listed on the left and include 3-D protein structure prediction, PCR primer design, microRNA target design, multiple sequence alignment, and genome browsers. Each category had five bioinformatics tools that could be utilized to answer a research question in that area, listed in the column titled “bioinformatics tools.”

The number of Google Blogs and Discussions on each resource using the search filter is indicated in their specific columns and their total expressed as a sum of the two values. The

value in the social media ranking column was generated from these totals. Variation in the total values was observed and generated a wide range of numbers, ranging from zero to 5,323 posts. The content of a subset of the posts was examined, and the remarks suggest the posters were discussing the resources in terms of research purposes either positively or neutrally, but not negatively.

Expert Rankings Compared to Social Media Results

The expert rankings of the bioinformatics tools are included in two columns in Table 2. The experts took different approaches to creating the ranks. Both considered the usability of the resource in question, specifically examining the functionality of the tool and assessing aspects such as the quantity and quality of information retrieved as well as the ease of use of the tool. One expert also included the number of times a resource was cited in the biomedical literature as part of the determination of its rank. In each case, the experts did not consider the content of the social media when making their decisions but rather used their domain knowledge as guides.

The expert rankings initially showed a relatively high degree of variability. When comparing the results, the experts agreed 44% of the time (11 of the 25 resources, see Figure 1). When comparing the expert rankings to the social media results, only 4 of the 25 resources shared rankings between both the experts and social media. This 16% agreement value is low and initially suggested that the social media discussions did not reflect expert opinions on bioinformatics resources.

It was considered that it might be more accurate to view the data in a way that reflects how researchers typically use resources to accomplish a given task. Users of these tools may choose to employ more than one to validate results, use unique features of a given resource, or consider two or three to be equally commendable. So rather than focusing on the exact rank of a resource, it was asked what the agreement was if a tool was found in the top three categories.

Expert rankings were reevaluated based on agreement that a resource was present in the top three places. The order was not important as long as both experts ranked a resource as a one, two, or three. When calculating data in this fashion, the agreement between the experts increased to 73% (11/15 resources). The expert rankings were compared to social media results by asking the question – does at least one expert agree with the social media that a resource is in the top three places? When examined this way, the experts agreed with the social media 87% of the time (13/15 resources).

Validation of Method

To determine whether or not this ranking system is equally effective in other areas of interest or if this system only works with bioinformatics, the procedure was reassessed using three consumer health questions and a new search filter). The results of the consumer health resource rankings are in Table 3. As seen previously, there was a wide range of variability from the social media discussions. Consumer health expert rankings agreed 53% of the time (8/15 resources; Figure 1). The agreement between the experts and the social media was a mere 6.7% (1/15 resources),

The top three websites in each question were examined in a similar fashion as the bioinformatics tools. When recalculated, it demonstrated the experts agree about the top three resources 78% of the time (7/9 resources). Furthermore, when considering the top three social media resources, the experts showed a 67% agreement (6/9 resources).

DISCUSSION

Social Media Rankings

Google was employed as this study's search engine due to its advanced search features and general familiarity. However, Google uses an unknown algorithm to prioritize hits, so there is minimal control over what information is captured. Also, the number of results in a category can change regularly depending on what users post or remove. Nevertheless, it is useful to observe the varied results over time, as it is expected that the order of resources will change based on continual social media discussions. This method can be set up to automatically update the results and generate new orders based on current discussions.

The goal of this method was to create recommendations based on the most discussed biomedical resources, yet at this time it is not possible to separate positive discussions from negative ones. This challenge is also observed with citation analysis. Both Adler and Garfield noted that sometimes published articles are cited for negative criticism, such as problems with the research.^{13, 14} However, Abromo and Pendlebury believe the overall number of negative citations is small.^{15, 16}

A subset of the search results was examined to determine whether people blogged about or discussed the resources of interest for positive/neutral experiences such as data analysis or tool usage, or for negative commentary. A manual review of the first 100 Google Blog posts and 50 Google Discussion posts of the bioinformatics resources "UCSC genome browser" and "Ensembl" indicated that 100% were positive referencing of the resources. Some of the reasons noted among the discussions included using the tools for molecular biology analysis, announcement of new features, downloads, or upgrades of the resource, serving as a training guide or educational resource, or the employment of a tool in a research article. Additionally, discussions included posting inquiries on best tools for analysis or how algorithms work, as well as asking about problems encountered with the tools. Overall, these posts indicate that people are utilizing the resources and are seeking advice on how to optimize their experiences and resolve problems. However, the search algorithm should be revisited periodically to ensure consistent results.

Subject Expert Rankings

Within this analysis, subject experts in both bioinformatics and consumer health did not initially demonstrate a high rate of agreement between one another. This is a common phenomenon observed across many fields. In biomedicine, disagreement regarding how health literacy should be measured, low agreement among health professionals about the quality of health information on the Internet, and disagreement about aspects of protein folding are only a few examples of deviations in opinion.¹⁷ In the study described here, common replies for highly ranking a tool included familiarity employing a resource and knowing that the information is reliable and is updated frequently. Much of the difference in opinion among the experts was due to personal preference, such as having prior knowledge about a resource or believing that a resource has a more user-friendly interface. However, both sets of experts did demonstrate greater agreement when looking at the top three resources in their respective fields.

Overall, the consumer health rankings generated results similar to the bioinformatics study. This suggests that the online ranking system can be used to identify key resources in other biomedical topics. However, there is a discrepancy in percentages between the two areas for the top three resources identified by both the social media and experts (87% vs. 67%). While both areas showed a marked increase in agreement when considering the top three resources, the bioinformatics agreement was stronger. This could be explained by recognizing the expertise of people likely to post social media on these topics. Bioinformatics posts seem to

be generated primarily by people familiar with the resources and who understand the strengths and limitations of the resources they are using. Consumer health discussions are produced by the general population, many of whom are unfamiliar with locating health information and may not be aware of the best resources to answer their questions. Thus, it is plausible that consumer health posts may not always reflect recommendations by experts in the field.

Implementation of a Social Media Recommendation Strategy

Implementation of the social media rankings as a recommendation strategy into the existing framework of the OBRC is a tangible goal. Because of the number of resources within the database and the calculations required to determine the values, the rankings cannot be performed in real time but rather would be based on regularly scheduled searches. It is also possible to build a weighted algorithm for social media rankings. With this, a higher weight could be assigned to recent discussion and blogs, thus more accurately capturing up-to-date mentions in the social media.

One visualization option is to continue to have patron search results displayed by keyword matching but also include a tab that would re-sort the data based on social media discussions. This would allow the users to have social media recommendations created for them in real time. Another option is to have OBRC users themselves rate the resources on a Likert scale. This idea seems promising but has had limited success on another website. The Online Resource Finder for Lifesciences (OReFiL) is similar in function to the OBRC and provides users links to bioinformatics tools.¹⁸ OReFiL includes a feature where users can grade the resources, generating a social rating of the tools. Users can then display search results based on ratings. However, at the time of publishing this study, no ratings have been entered on the site, which suggests that this approach may not be the best way to show preferred usage of bioinformatics resources. The method presented in this study captures the usage of resources across a broader audience since it is not limited to users of a specific website.

Limitations

While it appears the social media ranking system described here effectively identifies top resources in bioinformatics, there are some limitations associated with the work. First, the quotation marks surrounding resources in the search filter increases the precision of the results but lowers the recall of items retrieved. This is due to being able to express a term multiple ways if it contains more than one word. For example this study only captures social media discussions where the exact phrase “UCSC genome browser” is written, and not variations of the term. While many people in the field of bioinformatics, including the experts in this study, hold the UCSC genome browser in high regard, the social media ranking placed it second behind Ensembl. One possible explanation why this occurred is simply because fewer discussions about the UCSC genome browser exist in the social media. However, it could also be due to not always retrieving hits using this search method because it was phrased in other ways. One alternate description is the “human genome browser at UCSC,” as described by Kent et al.¹⁹

Another limitation of the study is that bioinformatics topics comprise a very small part of the total social media when compared to items such as popular culture or current events. Unless the volume of posts in the areas of bioinformatics increases, the sample pool will remain small. One way to help resolve this is to add other sources of social media data. Discussions on Twitter or Facebook could be added at a later date as new search tools become available and are more amenable to this method.

In the beginning stages of this project, input from both Facebook and Twitter were considered, but after trying these sources they were removed from the final tallies. The concern with Facebook was in determining if privacy settings on personal accounts interfered with search results. Various Twitter search engines were tested in a preliminary study and difficulties were found in employing data from this resource. Half of the bioinformatics tools searched in Twitter generated zero hits, while another 20% returned non-specific Tweets. Most of the Twitter search engines did not allow for the use of the designed bioinformatics filter, so it was difficult to establish the specificity of the resource in question. Therefore, data from Facebook and Twitter were not included in this study.

Lastly, at the time of data collection for this project, Google limited queries to 32 words. This was not problematic using either of the search filters described here but may be an issue with lengthier filters. The features of various search engines, including updates to Google's options, should be considered when creating future searches.

CONCLUSION

These data from this pilot study suggest social media discussions can be used to identify key biomedical resources that generally reflect the opinions of subject experts, thus providing a useful method to identify high quality resource recommendations for information professionals who may not have strong domain knowledge in a particular subject area. Additionally, automation of the rankings would maintain the currency of the information without having to continuously survey librarians or bioinformatics experts and manually build recommendations on thousands of bioinformatics resources. In considering implementation, an explanation could be given to library patrons that the recommendations are subjective, and they will find useful any of the top three to five tools from a much larger pool of resources.

The algorithm employed is a straightforward method to gather data and can serve as a preliminary analysis tool. Over time, refinements and advances in different algorithms may allow for expansion of the query lengths, help to automatically filter negative captured comments, and allow a more detailed enhancement of a variety of parameters to improve data analysis.

Acknowledgments

The authors would like to thank Michelle Burda, MLS and Andrea Ketchum, MLS from the Health Sciences Library System at the University of Pittsburgh for serving as the consumer health experts. Also, thank you to Nancy Tannery, MLS and Rebecca Crowley, MD, MS for help at various stages of this project.

Funding for this study was supported by National Library of Medicine (NIH) grant 5-T15-LM007059-23S1.

APPENDIX 1

Bioinformatics Questions

You are a researcher in a biochemistry lab and are interested in obtaining structural data on the protein that you are studying, however no information on this topic has been reported. In order to help determine structural information about your protein, you plan to use 3-D prediction software. Please rank the following homology modeling resources from most useful (1) to least useful (5) for this particular task.

CPHmodels	http://www.cbs.dtu.dk/services/CPHmodels/
ESypred3D	http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/
GENO3D	http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html
HHPred	http://toolkit.tuebingen.mpg.de/hhpred
SWISS-MODEL	http://swissmodel.expasy.org/

You are working in a molecular biology laboratory and must design primers for upcoming PCR experiments. You know the 5 sequences that you will be amplifying from a rat liver. Please rank the following PCR primer design tools in order from the one you are most likely to choose (1) to the least likely to use (5).

IDT SciTools	http://www.idtdna.com/SciTools/SciTools.aspx
Primer3	http://primer3.sourceforge.net/
Primer Design Assistant (PDA)	http://dbb.nhri.org.tw/primer/
PrimerZ	http://genepipe.ngc.sinica.edu.tw/primerz/beginDesign.do
ProbeWiz	http://www.cbs.dtu.dk/services/DNAarray/probewiz.php

For one of your experiments, you'd like to design microRNAs to silence a specific gene. Which of the following miRNA target designers would you choose to help predict the sequences that you will need? Please rank the following resources in the order of most helpful (1) to least helpful (5).

DIANA-microT	http://diana.cslab.ece.ntua.gr/microT/
MicroInspector	http://bioinfo.uni-plovdiv.bg/microinspector/
miRGator	http://genome.ewha.ac.kr/miRGator/
siRNA target finder (Ambion)	http://www.ambion.com/techlib/misc/siRNA_finder.html
TargetScan	http://www.targetscan.org/

You are studying human gene "x" and would like to know what DNA sequence similarities exist in gene "x" in other mammalian species. In order to perform a multiple sequence alignment of these genes, please rank the resources that you would find most useful (1) to least useful (5).

Clustal W	http://www.ebi.ac.uk/clustalw/
DIALIGN	http://bibiserv.techfak.uni-bielefeld.de/dialign/
ECR Browser	http://ecrbrowser.dcode.org/
Kalign	http://www.ebi.ac.uk/Tools/msa/kalign/
TCoffee	http://www.ebi.ac.uk/Tools/msa/tcoffee/

You are studying a specific human gene in your lab and would like to examine this gene within the context of the human genome. Please rank the following genome browsers in order from most likely to use (1) to least likely to use (5).

Ensembl Genome Browser	http://useast.ensembl.org/index.html
G-Compass Genome Browser	http://www.h-invitational.jp/g-compass/
Map Viewer through NCBI	http://www.ncbi.nlm.nih.gov/mapview/
UCSC Genome Browser	http://genome.ucsc.edu/
VISTA Genome Browser	http://pipeline.lbl.gov/cgi-bin/gateway2

APPENDIX 2

Consumer Health Questions

A patron enters your library and asks you to provide her with websites that contain consumer health information. She indicates that she has type 2 diabetes and would like more information about that but is mainly looking for websites where she can locate information on a variety of conditions and diseases. Please rank the following websites below in order of most helpful (1) to least helpful (5).

Healthfinder.gov	http://www.healthfinder.gov/
MedlinePlus	http://www.nlm.nih.gov/medlineplus/
WebMD.com	http://www.webmd.com/
Wrongdiagnosis.com	http://www.wrongdiagnosis.com/
Yahoo! Health	http://health.yahoo.net/

A patron in your library tells you that his doctor has recently put him on Lipitor for his cholesterol. He indicates that he is on a few other medications and would like to know where he can find information online about the drugs he is taking. Please rank the following resources below in order of most helpful (1) to least helpful (5).

Drugs.com	http://www.drugs.com/
MedlinePlus	http://www.nlm.nih.gov/medlineplus/
PDR Health	http://www.pdrhealth.com/
Rxlist.com	http://www.rxlist.com/script/main/hp.asp
WebMD.com	http://www.webmd.com/

A patron is looking for online videos on various disease topics. Please rank the following resources below in order of most helpful (1) to least helpful (5).

Healthology	http://www.healthology.com/content/consumerhealthlibrary/
HealthScout	http://www.healthscout.com/nav/video/1/main.html
MedlinePlus	http://www.nlm.nih.gov/medlineplus/
VideoMD	http://www.videomd.com/
WebMD.com	http://www.webmd.com/

Note: The above URLs in Appendix 1&2 were active at the time of this study but may not be in the future.

REFERENCES

1. University of Pittsburgh Health Sciences Library System. [Accessed February 27, 2013] OBRC. <http://www.hsls.pitt.edu/obrc/>
2. Chen YB, Chattopadhyay A, Bergen P, Gadd C, Tannery N. The Online Bioinformatics Resources Collection at the University of Pittsburgh Health Sciences Library System--a One-Stop Gateway to Online Bioinformatics Databases and Software Tools. *Nucleic Acids Res.* 2007; 35(1):D780–D785. [PubMed: 17108360]
3. Fernández-Suárez XM, Galperin MY. The 2013 Nucleic Acids Research Database Issue and the Online Molecular Biology Database Collection. *Nucleic Acids Res.* 2013; 41(D1):D1–D7. [PubMed: 23203983]
4. Kozinets RV, de Valck K, Wojnicki AC, Wilner SJS. Networked Narratives: Understanding Word-of-Mouth Marketing in Online Communities. *J Marketing.* Mar; 2010 74(2):71–89.
5. Da Z, Engelberg J, Gao P. In Search of Attention. *J Finance.* Sep; 2011 66(5):1461–1499.
6. Signorini A, Segre AM, Polgreen PM. The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. During the Influenza A H1N1 Pandemic. *PLoS One.* May.2011 6(5):e19467. [PubMed: 21573238]
7. Metzger MJ, Flanagan AJ. Using Web 2.0 Technologies to Enhance Evidence-Based Medical Information. *J Health Commun.* Aug; 2011 16(S1):45–58. [PubMed: 21843095]
8. National Institutes of Health. [Accessed February 18, 2013] Social Media and Engagement Tool Software: Evaluating Online Resources of the National Library of Medicine by Mining Social Media. Solicitation Number: RFQ_NIHLM2013659_SRE https://www.fbo.gov/spg/HHS/NIH/OAM/RFQ_NIHLM2013659_SRE/listing.html
9. StockTwits. [Accessed February 8, 2013] About. <http://stocktwits.com/about>
10. SocialGuide. [Accessed February 8, 2013] About. <http://www.socialguide.com/about>
11. Tenenbaum JD, Whetzel PL, Anderson K, et al. The Biomedical Resource Ontology (BRO) to Enable Resource Discovery in Clinical and Translational Research. *J Biomed Inform.* 2011; 44(1): 137–145. [PubMed: 20955817]
12. Morales-del-Castillo JM, Peis E, Ruiz AA, Herrera-Viedma E. Recommending Biomedical Resources: A Fuzzy Linguistic Approach Based on Semantic Web. *Int J Intell Sys.* 2010; 25(12): 1143–1157.
13. Adler R, Ewing J, Taylor P. Citation Statistics. *Stat Sci.* 2009; 24(1):1–14.
14. Garfield E. Is Citation Analysis a Legitimate Evaluation Tool? *Scientometrics.* 1979; 1(4):359–375.
15. Abramo G, D'Angelo CA. Evaluating Research: From Informed Peer Review to Bibliometrics. *Scientometrics.* 2011; 87(3):499–514.
16. Pendlebury DA. The Use and Misuse of Journal Metrics and Other Citation Indicators. *Arch Immunol Ther Ex.* 2009; 57(1):1–11.
17. Baker DW. The Meaning and the Measure of Health Literacy. *J Gen Intern Med.* 2006; 21(8):878–883. [PubMed: 16881951]
18. [Accessed February 8, 2013] Online Resource Finder for Lifesciences. <http://orefil.dbcls.jp/>
19. Kent WJ, Sugnet CW, Furey TS, et al. The Human Genome Browser at UCSC. *Genome Res.* Jun; 2002 12(6):996–1006. [PubMed: 12045153]

FIGURE 1.
Agreement Between Experts and Social Media

TABLE 1

Social Media Search Filters

Bioinformatics	Consumer Health
[("UCSC genome browser") AND (bioinformatics genome genetics genomics computer algorithm software server database computer model protein proteomics proteome gene DNA RNA sequence alignment interactions structure modeling prediction biochemistry molecular biology computational biology)]	[("MedlinePlus") AND (consumer health health information medical information website online resource health resource personal health consumer health information drug supplement herb medication drug information prescription over the counter video online video)]
Resource in quotation marks; Vertical lines represent OR	Resource in quotation marks; Vertical lines represent OR

TABLE 2**Social Media and Expert Rankings of Bioinformatics Resources**

	Bioinformatics Tools	Blogs	Discussion	Blogs + Discussion	Social Media Rank	Expert 1 Rank	Expert 2 Rank
	SWISS-MODEL	405	102	507	1	3	1
3D Protein	CPHmodels	124	30	154	2	1	5
Prediction	HHPred	74	7	81	3	2	2
	ESypred3D	22	20	42	4	5	3
	GENO3D	8	3	11	5	4	4
	Primer3	1,130	294	1,424	1	1	1
PCR Primer	IDT SciTools	7	2	9	2	2	3
Design	Probe Wiz	5	1	6	3	5	5
	Primer Design Assistant	3	0	3	4	3	4
	PrimerZ	1	0	1	5	4	2
	Targetscan	405	75	480	1	2	2
microRNA	microInspector	34	30	64	2	4	4
Target	DIANA-microT	36	3	39	3	1	3
Design	miRGator	10	0	10	4	3	5
	siRNA target finder Ambion	6	0	6	5	5	1
	ClustalW	2,670	626	3,296	1	1	1
Multiple	DIALIGN	119	37	156	2	4	3
Sequence	Kalign	68	32	100	3	3	5
Alignment	Tcoffee	67	0	67	4	2	2
	ECR Browser	9	4	13	5	5	4
	Ensembl	4,560	763	5,323	1	3	2
Genome Browsers	UCSC Genome Browser	1,660	170	1,830	2	1	1
	NCBI Map Viewer	56	25	81	3	2	3
	VISTA Genome Browser	4	0	4	4	4	4
	G-compass Genome Browser	0	0	0	5	5	5

Data collected 1/16/13; Dates limited to all time

TABLE 3

Social Media and Expert Rankings of Consumer Health Resources

	Consumer Health Resources	Blogs	Discussion	Blogs + Discussion	Social Media Rank	Expert 1 Rank	Expert 2 Rank
	WebMD.com	1,710	9,740	11,450	1	3	2
Consumer	MedlinePlus	3,730	81	3,811	2	1	1
Health	Yahoo!Health	1,390	10	1,400	3	4	4
Information	Healthfinder.gov	380	2	382	4	2	3
	Wrongdiagnosis.com	26	4	30	5	5	5
	WebMD.com	1,710	9,740	11,450	1	4	4
Consumer	Drugs.com	5,270	99	5,369	2	5	2
Health Drug	MedlinePlus	3,730	81	3,811	3	1	1
Information	Rxlist.com	90	10	100	4	3	3
	PDR Health	1	2	3	5	2	5
	WedMD.com	1,710	9,740	11,450	1	2	3
Consumer	MedlinePlus	3,730	81	3,811	2	1	1
Health	HealthScout	10	0	10	3	4	4
Videos	Healthology	6	1	7	4	3	5
	VideoMD	3	0	3	5	5	2

[#]Data collected 1/16/13; Dates limited to 1/1/09 - present