# Crowdsourcing in biomedicine: challenges and opportunities

Ritu Khare, Benjamin M. Good, Robert Leaman, Andrew I. Su and Zhiyong Lu

Corresponding author. Zhiyong Lu, National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, Maryland 20894, USA. Tel.: 301-594-7089; Fax: 301-480-2288; E-mail: zhiyong.lu@nih.gov

## Abstract

The use of crowdsourcing to solve important but complex problems in biomedical and clinical sciences is growing and encompasses a wide variety of approaches. The crowd is diverse and includes online marketplace workers, health information seekers, science enthusiasts and domain experts. In this article, we review and highlight recent studies that use crowdsourcing to advance biomedicine. We classify these studies into two broad categories: (i) mining big data generated from a crowd (e.g. search logs) and (ii) active crowdsourcing via specific technical platforms, e.g. labor markets, wikis, scientific games and community challenges. Through describing each study in detail, we demonstrate the applicability of different methods in a variety of domains in biomedical research, including genomics, biocuration and clinical research. Furthermore, we discuss and highlight the strengths and limitations of different crowdsourcing platforms. Finally, we identify important emerging trends, opportunities and remaining challenges for future crowdsourcing research in biomedicine.

**Key words**: Amazon Mechanical Turk; big data mining; biomedicine; community challenges; crowdsourcing; games

## Introduction

Crowdsourcing is the process of getting services, information, labor or ideas by outsourcing through an open call, especially through the Internet [1]. The 'crowd' of people recruited is usually large, poorly defined and diverse. The tasks that can be crowdsourced include scientific problems that remain out of the reach of current computational algorithms, and thus require manual judgments or human expertise. Crowdsourcing approaches in biomedicine include a variety of bioinformatics research activities that are of great interest and are active research topics, e.g. text mining the search engine logs of millions of users [2, 3], seeking crowd knowledge and judgments to scale curation of biomedical databases [4, 5], designing clinical trial applications [6], solving complex problems in structural biology [7], etc. Further, the 'crowd' is expanding to explicitly include scientists and physicians (i.e. expert sourcing), and patients [8, 9]. In a previous review [10], crowdsourcing problems were broadly classified as either 'microtasks' or 'megatasks'. Microtasks are individually small tasks that typically come in big numbers where a large number of people contribute in parallel to save time and generate an aggregated solution, e.g. word sense disambiguation [11], and named entity recognition (NER) [12], which involves locating and classifying mentions in text into predefined categories, such as disease, drugs, genes [13, 14], etc. Megatasks are high difficulty tasks that are often approached through open innovation contests [15] where the key challenge is to find a few talented individuals from a large pool, e.g. challenging games such as Foldit [16].

**Ritu Khare** is an Information Scientist at the Department of Biomedical and Health Informatics of the Children's Hospital of Philadelphia. Her research focuses on electronic health record data mining and data quality.
**Benjamin M. Good** is an Assistant Professor of the Department of Molecular and Experimental Medicine at The Scripps Research Institute. His research focuses on applications of crowdsourcing in bioinformatics.
**Robert Leaman** is a Research Fellow at the National Center for Biotechnology Information, National Institutes of Health. His research focuses on information extraction and natural language processing.
**Andrew I. Su** is Associate Professor of the Department of Molecular and Experimental Medicine at The Scripps Research Institute, where he leads the Su Lab focused on crowdsourcing biology.
**Zhiyong Lu** is Earl Stadtman investigator at the National Center for Biotechnology Information (NCBI), National Institutes of Health, where he leads the biomedical text-mining group. Dr. Lu's research has been successfully applied to the NCBI's PubMed system and beyond.

The use of crowdsourcing in biomedicine is growing rapidly. Since the publication of the last review, there have been several new and highly focused workshops in this area [17]. In recent years, over 200 related publications have been added to the PubMed database. In this article, we present a review of the major recent advances in the field of crowdsourcing in biomedicine, including both new studies as well as updates of the work previously discussed in [10]. This review is organized around the classes of crowdsourcing problems. In addition to 'active crowdsourcing', which subsumes megatasking and microtasking, we introduce a new class 'mining crowd data', which involves biomedical knowledge discovery from the big data accumulated in public forums. We further categorize the works in each class based on the platforms used to conduct the crowdsourcing studies. The central part of Figure 1 presents a visual overview of the overall organization of this review. The discussion and conclusions section presents the opportunities and challenges associated with various crowdsourcing platforms, and highlights some emerging trends for future crowdsourcing research in biomedicine.

## Mining crowd data

Crowd data mining refers to the research activities of collecting and analyzing the data, produced by people through participation in public forums, for the knowledge discovery process. While in many cases, the crowd data are privately owned by the hosting organization, more sources (e.g. Twitter API, openFDA) are making their data freely available for computation, investigation and exploration. In the era of social media and big data, many recent studies address important health problems, such as identification of adverse effects of drugs and seasonality of diseases, via search logs [18, 19], Twitter [20], online patient forums [21], Food and Drug Administration (FDA) reports [22, 23] and electronic health records (EHRs) [24].

## Web search logs

Health care consumers are increasingly resorting to resources such as Web search engines [2] and public forums to seek health information or discuss health concerns such as adverse drug events. The traditional routes for reporting adverse effects of drugs on patients are slow and do not comprehensively capture all cases. Previous research has primarily focused on early detection of drug safety information from scholarly publications [25]. Recently, crowdsourcing has been explored to detect these events from wider data sources. White et al. [18] studied whether Web search patterns may give early clues to adverse effects of drug–drug interactions. This study focused on a particular case—'paroxetine and pravastatin interaction causes hyperglycemia'—reported to the FDA in 2011. White and colleagues analyzed 82 million drug/disease queries from 6 million Web searchers (based on IP addresses) in 2010. These searchers opted to share their activities through a Microsoft browser
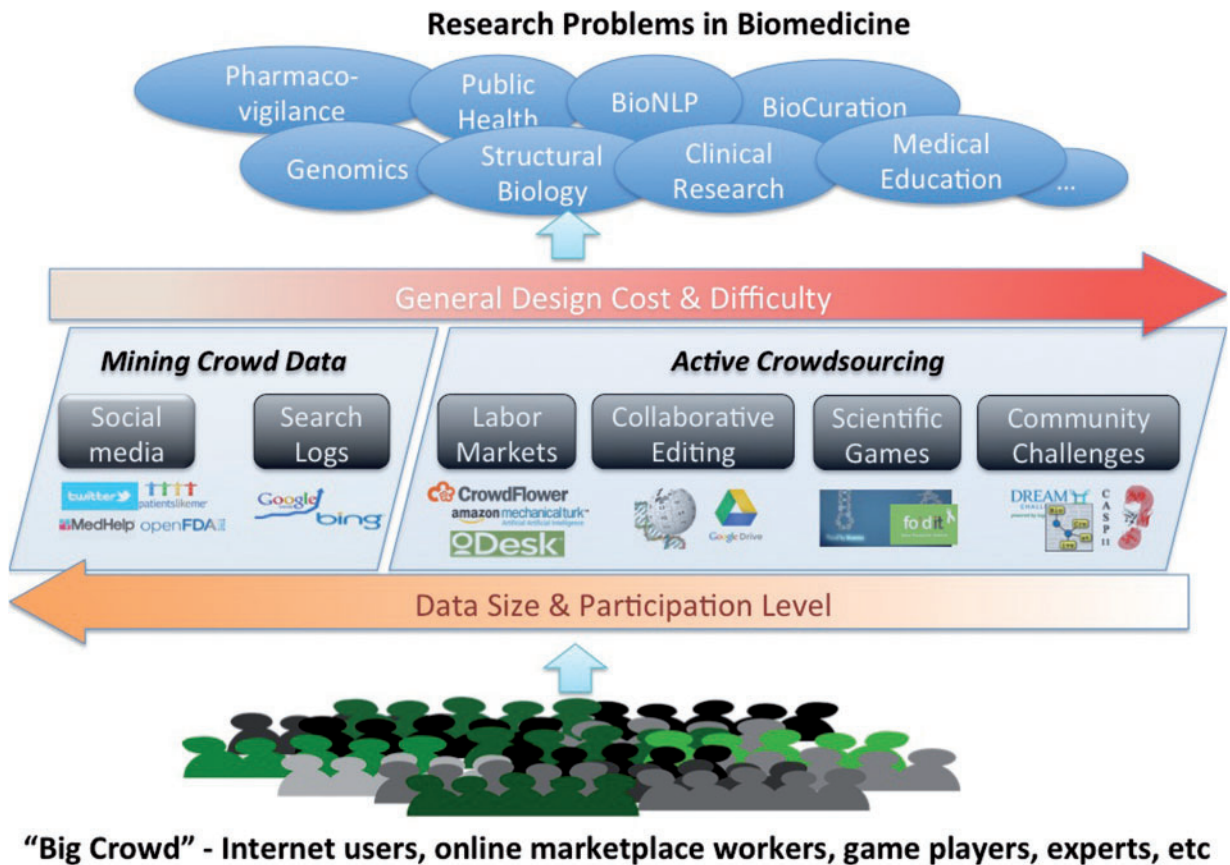


**Figure 1.** Crowdsourcing in Biomedicine: The Big Picture.

add-on application. Using disproportionality analysis, it was confirmed that the users who searched for both drugs were more likely to search for hyperglycemia-related terms than those who searched for only one of the drugs. The generalizability of this method in making predictions was also confirmed for several other established drug–drug interactions.

Ingram et al. [26] used Google Trends (http://www.google.com/trends/), an online tool that gives statistics on search term usage in Google, to study the seasonal phenomena in sleep-disordered breathing using different variations of the search terms such as 'snoring', 'sleep apnea', 'snoring children', 'sleep apnea children'. The queries from a 7-year time frame (January 2006–December 2012) were collected from two representative countries in the northern and southern hemisphere, USA and Australia, respectively. Using cosinor analysis (a regression-based peak finding technique), peak searches for all the queries were observed in the winter/early spring season in both countries, i.e. January to March for the USA, and July for Australia. The potential of using Google Trends was further empirically validated in another study [27] that compared multiple logistic regression models to forecast influenza-like-illness-related visits in an emergency department. It was found that the models performed significantly better when the Google Flu Trends data were included as a predictive variable.

## Smartphone applications and social media

According to Google, the number of worldwide smart phone users was 1.75 billion in 2014, and is expected to increase to 2 billion in 2015 [28]. People are increasingly using smartphone applications to track personal health and behaviors. In addition, the patients in this era are more willing to share health information for research purposes [29]. As an example, PatientsLikeMe (http://www.patientslikeme.com/), a patient-powered research network, has >220 000 members who share their personal health information on more than 2000 conditions [30]. These new habits are leading to a massive amount of new data ('big data') and an excellent potential knowledge resource.

In a recent study, Odgers et al. [19] studied the drug/disease query logs of patient care providers in UpToDate, a point of care medical resource. The authors investigated the hypothesis that when a provider is concerned about an adverse drug reaction, she searches for both the associated drug and disease/event concepts within a short span of time. Over 320 million queries were mined from the 2011 to 2012 search logs, and association analysis was used to quantify the strength of drug/event association. The evaluation was conducted on two data sets, one where the adverse drug reactions were long established (i.e. retrospective analysis) and another one where the event was reported after the timeline of the query logs (i.e. prospective analysis). The proposed method for distinguishing valid from spurious associations achieved a discrimination accuracy of 85 and 68% on the retrospective and prospective data sets, respectively, on an optimal surveillance period of 4–10 min. Also, >80% of the associations in the prospective data set were supported in the query logs within a surveillance period of 5 min. This study strongly indicates the feasibility and utility of mining crowd data for pharmacovigilance purposes.

In another study, Freifeld et al. [20] studied the signals of consumer-reported events from Twitter posts (November 2012–May 2013). A total of 4401 posts (related to 23 medical products) were selected using a combination of dictionary matching, human annotation, and natural language processing (NLP) techniques, and were found to be concordant with reports from the FDA Adverse Event Reporting System. Furthermore, Yang et al. [23] investigated the methods to compute the 'interestingness' of an adverse event signal mined from online communities such as MedHelp. They concluded that the metrics 'lift' and 'proportional reporting ratios' in association rule mining are highly predictive of the drug–reaction signal strength in social communities.

Turner-McGrievy et al. [31] mined data from the Eatery application, which allows users to post photographs of their meals, and receive ratings from peers on the healthiness scores of those meals (on a scale of 'fat' to 'fit'). The authors mined 450 pictures rated by 5006 raters to study how closely the crowd-sourced ratings of foods and beverages were related to the ratings by three trained raters who were knowledgeable of the Dietary Guidelines for Americans. The correlation coefficients of healthiness scores between each expert rater and peer raters were highly significant, and the peer ratings were in the expected direction as per the dietary guidelines recommendations. These results indicate the potential of using crowdsourcing to provide effective and expert-like feedback on diet quality of users.

## Active crowdsourcing

In contrast to mining crowd data, active crowdsourcing invites people to participate in solving a problem in biomedicine. Active crowdsourcing is conducted through a variety of platforms such as labor markets, casual games, hard games or community challenges.

## Labor markets (Amazon Mechanical Turk, etc.)

A labor market platform provides the technical environment to design and submit the crowdsourcing microtasks, and more importantly, acts as the human labor market so that appropriate 2workers could be recruited for the submitted tasks. Some of the most popular platforms include Amazon Mechanical Turk (MTurk) (https://www.mturk.com/mturk/welcome) and Crowdflower (http://www.crowdflower.com/). Investigators can design and customize the task interface for their experiments, using open-ended or multiple-choice questions, instructions for workers and other features such as word highlighting (e.g. Crowdflower allows building sophisticated crowdsourcing user interfaces using HTML and JavaScript). These platforms also allow configuration of the quality control mechanisms, e.g. a qualifier test to recruit high-performing workers, and interjecting control microtasks or training examples to evaluate the performance of workers. Through these platforms, one could specify the inclusion criteria for workers (e.g. demographics), payment for each microtask (usually a few cents) and the redundancy of judgments, i.e. the number of unique workers who must work on a microtask (several studies request five judgments per microtask). Once the job is finished, the study investigators aggregate the crowdsourced judgments (if applicable) using consensus building methods such as majority voting, and evaluate the consensus judgments, e.g. accuracy could be evaluated using expert curated gold standard, manual review or qualitative analysis.

In recent years, several successful studies in biomedicine have been conducted through the labor market platforms for a variety of application domains, including NLP, biocuration and literature review.

## Biomedical and clinical NLP

Biomedical NLP or text mining refers to the development of computer algorithms and tools to analyze and make sense of natural language data in the biological and medical domains (e.g. scholarly publications [32–35], EHRs [36, 37]).

In the crowdsourced NLP domain, most of the recent studies are focused on performing NER of drug or disease entities owing to their critical roles in biomedical research [2, 38]. Zhai et al. [39] used crowdsourcing for medication NER and entity linkage in eligibility criteria section of clinical trial announcement documents. The medication NER interface was developed to enable word pre-highlighting, word selection, annotation selection (name or type), tabular display of annotations (feedback) and selection of entities with discontinuous tokens. The authors also designed a correction task to edit annotations from the previous task, and a third task to link medications with attributes using pre-annotations through double annotations by experts. The authors posted 3400 microtasks drawn from 1042 clinical trial announcements on Crowdflower, and applied several quality control measures and aggregation strategies. Within 10 days of job posting, 156, 86 and 46 workers were recruited for the NER, NER correction and linking tasks, respectively. The final F-measures for medication NER and medication type NER were 90 and 76%, respectively, considered very close to the performance of the state-of-the-art text-mining tools [40–42]. This study also resulted into a very high accuracy (96%) on the linking task. In addition, no statistically significant difference was observed between the crowd-generated and expert-generated corpora, thus indicating the potential of using the crowd workers for NLP gold standard development.

MacLean and Heer [43] studied the problem of identifying medically relevant terms from patient posts in public health forums, MedHelp and CureTogether. They conducted two sets of studies, one with 30 nurse professionals recruited from oDesk (https://www.odesk.com/), a freelancing labor market, to collect three-way annotations on 1000 patient-authored sentences, and another with 50 crowd workers recruited from MTurk to collect five-way annotations on 10 000 sentences. The study was finished in 2 weeks with experts and within 17 hours of job posting with crowd workers. The crowd achieved an accuracy of 84% with expert annotations as the gold standard. It was found that the crowd-labeled data set is comparable with the expert-labeled data set in its ability to train a probabilistic model for labeling sequential tokens in a text [44].

In another study, Good et al. [45] studied the ability of crowd workers on MTurk to annotate disease mentions in PubMed abstracts. They used the expert-curated NCBI disease corpus [46, 47] as the gold standard for comparison and adapted the original annotation guidelines to suit the crowdsourcing environment. In addition to several quality control checks, users were also given precise feedback on their performance on control items with respect to the annotation guidelines. Overall, 145 workers contributed to produce annotations for 593 documents within 9 days of job posting. Each document was annotated by a minimum of 15 workers. Using voting to aggregate responses, the workers achieved an F1-measure of 87% in comparison with the gold standard.

## Biocuration

Biocuration refers to the process of extracting and curating key biological information, primarily from the literature but also other sources, into structured databases that can be examined systematically and computationally. Given the high costs of manual curation, many recent studies have proposed hybrid approaches combining text mining and crowdsourcing in curating drug indications, gene–disease associations and biomedical ontologies [48–51].

Khare et al. [4] translated the task of cataloging indications from narrative drug package inserts [52] into microtasks suitable for the average crowd worker, where the worker is asked to judge whether a highlighted disease is an indication of the given drug. The investigators posted >3000 microtasks (or drug/disease relationships) drawn from 706 drug labels on MTurk and collected 18 775 judgments from 74 workers, and achieved an aggregated accuracy of 96% on the control items.

In another study, Burger et al. [5, 53] used crowdsourcing to identify computable gene–mutation relationships from PubMed. Their microtask data set generation involved using GenNorm [54] and EMU [55] tools for identifying gene and mutation mentions respectively, and then using the crowd workers at MTurk to validate correct relationships. They simplified the task design by displaying only one gene–mutation pair at a time so that the crowd was asked to make a single yes/no judgment in a given task. Using quality control mechanisms, they posted 1354 microtasks drawn from 275 abstracts and collected five-way annotations contributed by 24 workers within 11 days. The judgment aggregation, performed using a Naïve Bayes classifier, resulted in 85% accuracy, and the active management of simulations resulted into as estimated mention-level accuracy of 90%.

Mortensen et al. [56] created microtasks on Crowdflower to curate a clinical ontology, the Systemized Nomenclature of Medicine-Clinical Terms (SNOMED-CT). The crowdsourcing interface was designed to show the hierarchical relationship between two SNOMED-CT concepts, and seek the user's judgment on the validity of the relationship. The crowdsourcing interface also showed the definitions of the two concepts, pulled from controlled vocabularies such as MeSH (Medical Subject Headings) or the National Cancer Institute Thesaurus. The crowdsourcing job consisted of 200 relationships from SNOMED-CT, where 25 separate workers evaluated each relationship. The judgments were aggregated using a Bayesian method and evaluated using a gold standard prepared by five domain experts. The average agreement of the crowd workers with the experts was approximately same as the average inter-rater agreement of the experts (0.58), indicating that the crowd is nearly indistinguishable from the experts in its ability to evaluate ontology. This study helped in identifying critical errors in 20% of the posted relationships, which were subsequently communicated to the concerning organization.

The abovementioned studies suggest that crowdsourcing in biocuration not only results in significant cost and time savings, but also leads to an accuracy comparable with that of domain experts.

## Literature review

The labor market platforms are also being used to conduct systematic literature reviews of specific research questions. Brown and Allison [57] resorted to crowdsourcing to study whether the scientific articles on nutrition obesity that are concordant with popular opinion receive more attention among peers. To study this question, they designed a pipeline of crowdsourcing tasks including, (i) preparing the corpus of relevant scientific abstracts that investigated nutrition and obesity in humans and had an unambiguous conclusion, (ii) extracting the food topic of the abstracts, (iii) providing the perceived ability of foods to

produce obesity and (iv) extracting the Google scholar citations count for the articles. As a starting point, 689 PubMed abstracts (within a 4-year range) were selected and posted to the crowdsourcing pipeline on MTurk. The crowd workers, with some expert intervention, helped prepare a corpus of 158 abstracts, extracting the corresponding foods, and more importantly providing the perceived ability of those foods in causing obesity. The workers achieved a consensus for 96% of the conclusions and 84% of the foods in the abstracts, and 99% of the workers provided complete and usable information on the perceived ability and their opinions qualitatively matched with the authors' expectations. By synthesizing all the results, it was found that there is no significant difference between intuitive and counterintuitive conclusions on nutrition obesity in the attention received by the scientific community.

## Collaborative editing using wiki-style tools

Labor markets continue to be the most popular method for crowdsourcing where the money is the predominant incentive for participation. Recently, researchers have also conducted successful studies using nonconventional crowdsourcing platforms such as Google drive and Web-based forms, inviting users to prepare educational material [9] and even contribute to research study design [30].

The integration of Wikipedia with biocuration has been proposed and much discussed in recent publications [58]. Loguercio et al. [59] developed a wiki-based online game, Dizeez, where the presumably knowledgeable players are invited to catalog valid gene–disease associations by answering multiple-choice questions. Each question is composed of one clue gene and five choices for diseases sampled from Gene Wiki [60], including one corrected link extracted from the Gene Wiki (gold standard). Within 9 months of Dizeez's release, 230 individuals played 1045 games generating 6941 unique gene–disease associations. The validity of the identified associations (measured against the gold standard) was found to be concordant with the number of votes (i.e. number of players) for that assertion. Nearly 30% of the identified associations were not present in any of the major gene–disease databases (OMIM, PharmGKB or Gene Wiki). Out of the novel associations, only six assertions were supported by four or more votes, including five assertions where the evidence could be found in recent literature and one assertion that appeared plausible but was not yet supported in any conclusive research.

Motivated by the traditional practice of classmates quizzing each other and using flashcards for exam preparation, Bow et al. [9] developed a crowdsourcing model to develop the exam preparation content for preclinical school curriculum. The investigators prepared a centralized Google drive repository where students could be invited to anonymously contribute and edit quiz questions and notes relevant to a given lecture. In addition, the investigators developed a Java-based program to automatically generate a digital flashcard from those questions. The class of 2014 Johns Hopkins University School of Medicine was invited to contribute questions for 13 months; as a result, 120 students created 16 150 questions (average 36 questions per lecture). As an outcome measure, it was found that the average exam scores of this batch were higher than those of previous batch where this model was not implemented and hence the flash card system was not available. In a student survey (23% response rate), more than half of the students referred to the flash card program right before the final examination, and >90% of the users found the flashcard system helpful in learning and retaining new material. Another unique example of collaborative editing in the clinical sciences domain is a company, called Transparency Life Sciences, that conducts clinical trials for off-label drug indications [30]. Motivated by the barriers to participation in clinical trials, this company develops its drug protocols using a crowdsourcing model. For the development of its drug application for the use of metformin in prostate cancer, 60 physicians from various specialties, and 42 patients and advocates, were invited to comment on the clinical trial design process through a secure Web form in a 6-week time frame [6]. Upon qualitative analysis (summarization and categorization) of various responses, four major and five minor changes were made to the original protocol.

## Scientific games

People are estimated to spend >3 billion hours per week playing computer games [61]. Since the original ideas of 'games with a purpose' [62], game mechanics have been used to incentivize crowd labor both in and out of the sciences. Games are an appealing form of crowdsourcing for several reasons. They do not have the inherent limits on capacity that are inherent to microtask frameworks with per-work-unit costs [63]. Done well, they tap into a variety of intrinsic incentives that can be more rewarding to players than small financial sums. Aside from accomplishing crowdsourcing tasks, games have been shown to be effective learning environments [64] and vehicles for raising awareness and interest in science [65]. Despite these powerful advantages, we see far fewer serious attempts at purposeful games because they are substantially harder and more expensive to make.

Genomics crowdsourcing is dominated by game-based projects. Phylo [66] and Ribo [61] are two multiple sequence alignment games for DNA and RNA structures, respectively. In both these games, players attempt to line up colored bricks that represent nucleotides in different sequences by sliding them back and forth along a horizontal grid. Phylo players have assisted in improving >70% of the original alignments. Ribo presents extra bricks that represent the secondary structures (base-pairing properties) of sequence elements and a different scoring function appropriate for RNA structural alignments. In a proof-of-concept experiment, 15 players solved 115 puzzles corresponding to a range of RNA alignment tasks. For the majority of cases tested, the Ribo players were able to outperform computationally generated alignments. While this was a relatively small study, the results are suggestive that this game can successfully apply human computation to this computationally challenging task.

The protein folding game 'Foldit' is the flagship of both serious games and citizen science in biology [7]. With >500 000 registered players, FoldIt is turning into a stable and important resource for the structural biology community. Solutions found by players often improve on those by state-of-the art protein folding algorithms [67]. One of the important descendants of Foldit is the game EteRNA [68], which is focused on the challenge of RNA structure design. Given a specific shape, the task of the EteRNA players is to identify the RNA sequence that would fold into this shape. In a unique twist, this game is directly coupled to real laboratory experiments. Every month, the best scoring sequences submitted by players are synthesized and their structures are determined. These experiments provide additional incentive to the players in terms of both the point structure of the game (the closer the sequence actually folds into the target structure the more points they get) and in that

their gameplay actually has a tangible impact on the real world. Furthermore, these experiments provide the players with a clear signal from which they can learn. According to its creators, EteRNA is an example of a 'massive open laboratory' [69]. The players effectively use the game to submit hypotheses that are then tested in the laboratory. Once the results are returned, the players can refine their hypotheses and repeat the cycle. The availability of such Web-accessible laboratories means that people outside of laboratory environments may now apply the traditional scientific method. The incentive structures of the game attract a diverse and engaged group of participants, but the direct connection to the laboratory is likely the most important innovation, one that is likely to be emulated in various ways in the near future. The Riedel-Kruse laboratory, for example, has focused on creating increasingly sophisticated 'biotic games', recently enabling players to directly engage with laboratory technology (e.g. microscopes) in real time [70, 71].

## Community challenges in bioinformatics

In parallel to the development of crowdsourcing initiatives that reach out to the general public to help advance science, large open challenges are now commonly used to organize the efforts of the scientific community itself. Pioneered by the Critical Assessment of protein Structure Prediction (CASP) competitions of the protein folding community (which have been running since 1994) [72], this methodology has now been applied to NLP of biomedical text [73–78], inference of gene regulatory networks [79], prediction of breast cancer prognosis [80] and others (e.g. http://dreamchallenges.org). These 'shared tasks' follow a fairly consistent recipe: a computational prediction task is posted to the community with a mechanism for submitting solutions, then at a certain point in time, the challenge is closed and the algorithms are evaluated for their performance on previously unseen, experimentally generated or expert-generated gold-standard data, and finally the results are shared with the community. Often multiple solutions submitted independently are combined to produce meta-predictors that, in some cases, outperform the best individual predictors [13, 79]. Recently, the structural biology community sought to tap into this 'wisdom of crowds' phenomenon by explicitly promoting the formation of new multi-institutional teams at the outset of a challenge called 'WeFold' [81]. A complete processing pipeline for protein structure prediction can include a variety of components. These might include for example: secondary structure prediction, homology modeling, prediction of contacts and generation and selection from a large variety of candidate solutions. In previous CASP competitions, participants had to generate complete pipelines to submit predictions. In WeFold, teams were encouraged to create folding pipelines that integrated the best subcomponents from different laboratories. The hope was to identify novel tool-chains that could outdo individual solutions. The results of the initiative were mixed. In some cases, improvements were observed, but they were not uniform. The organizers cite challenges in assembling the required cyber-infrastructure and in recruiting teams to participate. However, they remain optimistic about the approach and are actively working on the next iteration of this 'coopetition' framework.

## Discussion and conclusions

While crowdsourcing existed long before the Internet, the current face of crowdsourcing is much different—even more so in biomedicine, where the crowd includes experienced physicians, informed patients, science enthusiasts and technology enthusiasts looking for a secondary source of income through systems like MTurk [82]. In this survey, we reviewed recent advances in crowdsourcing in the biomedical sciences by organizing the works into two broad classes: (i) crowd data mining, involving the crowd generating large amounts of computable data as a by-product of some other goal, and (ii) active crowdsourcing ranging from microtasks involving identification, annotation and reviewing to megatasks involving algorithm and strategy development. Furthermore, the surveyed works are classified by the major crowdsourcing platforms that determine the success and limitations of these studies. Figure 1 illustrates the pros and cons of various platforms with respect to the degrees of difficulty in task design and crowd recruitment.

As we enter the era of big data, our ability to mine and make sense of crowd data—either by itself or combined with other experimental data—becomes increasingly important for new discovery. Search logs and social media provide access to millions of user queries and comments each day. Although access to some data remains restricted [18, 19], the general trend is to make more data open for research and discovery [27] when security and privacy issues are properly managed. Another distinctive advantage of such platforms is the ability to conduct longitudinal studies. To realize the full potential of crowd data, advanced computer technologies and expert reviews would be desired to dig gold from these very large, unstructured, heterogeneous and often noisy data [19, 23]. Mining crowd data can be powerful, but not without limitations. For instance, Google Flu Trends has been criticized for its accuracy [83].

Compared with other crowdsourcing platforms, we find that labor markets (such as MTurk) are the most used in our surveyed studies owing to their many convenient features. First, MTurk is widely accessible and the setup of an MTurk experiment is relatively straightforward to many bioinformatics researchers. Second, the large existing pool of MTurk workers and the monetary reward mechanism allow MTurk to offer a high degree of diversity and the ability to complete tasks in a short time frame. The cost of designing and running an MTurk experiment is also modest compared with hiring physicians or research scientists. However, these studies are not meant to replace but to augment the traditional expert-driven studies; e.g. in many studies, the crowd judgments have improved the gold standard rather than replacing it [45, 56]. Finally, owing to its popularity, many analytic methods have already been developed for aggregating crowdsourced judgments to obtain optimal results. Nonetheless, MTurk has its own set of limitations, not only technically (e.g. spammers or poorly performing workers) and ethically (e.g. no guarantee of payment for their work) [84], but also institutionally (e.g. difficulty in obtaining approval to conduct such research studies in federal government as it is viewed as a traditional customer survey instead of an emerging research tool). MTurk studies are also highly dependent on domain experts for quality control and result evaluation. It is also worth noting that MTurk is most suitable for simple and repetitive tasks: complex and knowledge-rich tasks such as algorithm and strategy development would not be appropriate for MTurk. In general, the success of such studies in terms of the number of participants, turnaround time and quality of results depends on the design of the task, e.g. instructions and interface.

Scientific wikis and other similar Web-based platforms (e.g. Google Drive) allow collaborative editing through the Internet. These are best suited for curating biological topics of significant importance (e.g. Gene Wiki) and are neither difficult nor expensive to create [9]. Such tools can potentially be used by anyone

for free from around the world. However, given the commonly targeted problems (e.g. specific biomedical topics), expert-level knowledge is often needed for making direct contributions through such tools. Past studies have found it to be difficult to attract sufficient volunteers, suggesting that wikis may not provide sufficient incentive to motivate the participation of experts, who are primarily researchers and scientists. As an example, although the Google Drive study [9] allowed anonymized data entry to encourage more participants, this method precluded the ability to systematically assess the impact of such collaboration on student performance. Also, given the public nature of such tools, it is difficult to measure the participation rate. Finally, it is challenging to evaluate the quality of results and immediately reward the contributors, as it often includes new information not yet in the gold standard.

Using computer games as a crowdsourcing approach for solving science problems is intriguing and promising, as it can potentially reach a large and diverse player population. Also, once the game is properly developed, no additional incentives are needed for recruiting participation, as it is inherent in game play. However, these potential benefits are commonly associated with high up-front costs and time in game development. Furthermore, highly skillful game developers may not be easy to find among typical bioinformatics researchers. Finally, the scientific games are often associated with large number of registered players but few active players.

Community challenges are highly useful at solving difficult and complex science problems through forming and strengthening a collaborative community. The competition aspect of such events strongly motivates participation and innovation. On the other hand, like game design, organizing a science challenge is non-trivial, as it often involves the preparation of gold-standard training and test data for the purpose of algorithm development and evaluation [78]. Moreover, the level of participation is limited from the crowdsourcing standpoint as expertise-level knowledge is typically required.

## Emerging trends

As a result of this survey, we also identify some new trends of crowdsourcing in biomedicine such as new crowd incentives, participatory research and development, innovation in interface design and open-source research. Although the motivation of contributors has been found to vary with time [85], past studies [10, 86] have identified several prime incentives that promote crowd participation. Besides the studies involving labor markets, all surveyed studies offered intrinsic motivation to the crowd [63], i.e. the task was worth doing for the task itself. Other prominent reasons include the interest of people in tracking their health information or looking up health information online, interest in the broader field and interest in building reputation in the scientific community through personal branding.

As the crowdsourcing research community is discovering its new role of problem solver as opposed to mere data provider [30], science is no longer confined to a closed community of scientists. Similar to the philosophy of designing with (and not just for) the user in the human computer interaction community, we are entering a new era where the role of patients is not limited to the research subjects. Some other examples include inviting clinicians to design relational databases [87] and evaluate and design medical decision support rules [88, 89], inviting the crowd to review the results of NLP informatics algorithms [90] and identify design problems in medical pictograms [91], and inviting the crowd to correct errors by the crowd [39, 43, 57, 68].

Other than the highly customized megatasking interfaces (e.g. game design), a significant effort is also being spent on the design of the microtasking interface to make the task suitable for an average crowdworker [56]. Burger et al. [5] identify one of the reasons for misjudgment by the crowd as the discrepancy between the interface instructions and the ultimate goals of the investigators. While most existing labor market platforms allow crowdsourcers to encode HTML elements (such as checkboxes, textboxes) into the user interface, recent studies have also leveraged and designed advanced features such as pre-highlighting words using text-mining tools [4, 53], real-time word highlighting [39], encoding annotation guidelines into user options [4], allowing the contributors to input evidence [59] and even incorporating some game design elements such as providing immediate feedback to contributors [45, 92].

In terms of quality control, there has been a trend of collecting higher number of judgments per microtask, e.g. 30 in [93], 25 in [56], as opposed to the most popular factor five, and using machine learning for consensus building [5, 56] and dynamic pooling of contributors [5]. Finally, the crowdsourcing research in biomedicine itself is suggesting a new direction of integrative and complementary research. Several studies are sharing their crowdsourced results [4] and the software infrastructure [39] with the larger community, thereby encouraging integration of crowd signals from multiple sources, e.g. user logs, physician logs, EHR data and FDA reports [94].

A persistent question in crowdsourcing across all platforms is to identify the small fractions of unevenly distributed contributors who would turn in the best solutions [45, 95]. Finally, the ultimate challenge is to engage the contributors and public at large, and create a connection between scientists and the community. Moreover, we agree with the earlier studies that identified three key factors determining the success of a crowdsourcing project: finding the right problem, finding the contributors with appropriate skills and aligning the motivation of contributors with the needs of the study [86].

In conclusion, we find increasing interest in mining crowd data and using crowdsourcing approaches in biomedical investigations in recent years. By reviewing these studies in detail, we found that past and current studies touch on wide application areas in biomedicine that range from NLP to biocuration, genomics and clinical research. Furthermore, through the comparison of major crowdsourcing platforms, we have highlighted the strengths and weaknesses of different approaches as a guideline for future research, and also identified emerging trends and remaining challenges such as new incentives (e.g. personal health), and innovations in task designs and quality control.

---

**Key Points**

- Recent crowdsourcing studies for biomedicine are categorized into mining crowd data and active crowdsourcing.
- The studies are summarized based on the crowdsourcing platforms, such as labor markets, scientific games, wikis and community challenges.
- Emerging themes include new incentives, participatory research and development, innovation in interface design and integrative research.

---

## Acknowledgment

## References

1. Estellés-Arolas E, González-Ladrón-de-Guevara F. Towards an integrated crowdsourcing definition. *J Inf Sci* 2012;**38**: 189–200.
2. Islamaj Dogan R, Murray GC, Neveol A, *et al*. Understanding PubMed user search behavior through log analysis. *Database* 2009;**2009**:bap018.
3. Khare R, Leaman R, Lu Z. Accessing biomedical literature in the current information landscape. *Methods Mol Biol* 2014; **1159**:11–31.
4. Khare R, Burger J, Aberdeen J, *et al*. Scaling Drug Indication Curation through Crowdsourcing. *Database* 2015; **2015**:bav016.
5. Burger J, Doughty EK, Khare R, *et al*. Hybrid curation of gene-mutation relations combining automated extraction and crowdsourcing. *Database* 2014;**2014**:bau094.
6. Leiter A, Sablinski T, Diefenbach M, *et al*. Use of crowdsourcing for cancer clinical trial development. *J Natl Cancer Inst* 2014;**106**:dju258.
7. Good BM, Su AI. Games with a scientific purpose. *Genome Biol* 2011;**12**:135.
8. McCoy AB, Wright A, Laxmisan A, *et al*. Development and evaluation of a crowdsourcing methodology for knowledge base construction: identifying relationships between clinical problems and medications. *J Am Med Inf Assoc* 2012;**19**:713–18.
9. Bow HC, Dattilo JR, Jonas AM, *et al*. A crowdsourcing model for creating preclinical medical education study tools. *Acad Med* 2013;**88**:766–70.
10. Good BM, Su AI. Crowdsourcing for bioinformatics. *Bioinformatics* 2013;**29**:1925–33.
11. Snow R, O'Connor B, Jurafsky D, *et al*. Cheap and Fast—But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks. In: *Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Honolulu, Hawaii, 2008, pp. 254–63.
12. Yetisgen-Yildiz M, Solti I, Xia F, *et al*. Preliminary experiments with Amazon's mechanical turk for annotating medical named entities. In: *NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Association for Computational Linguistics, Los Angeles, CA, 2010, pp. 180–3.
13. Lu Z, Kao HY, Wei CH, *et al*. The gene normalization task in BioCreative III. *BMC Bioinformatics* 2011;**12**(Suppl 8): S2.
14. Van Landeghem S, Bjorne J, Wei CH, *et al*. Large-scale event extraction from literature with multi-level gene normalization. *PloS One* 2013;**8**:e55814.
15. Lakhani KR, Boudreau KJ, Loh PR, *et al*. Prize-based contests can provide solutions to computational biology problems. *Nat Biotechnol* 2013;**31**:108–11.
16. Cooper S, Khatib F, Treuille A, *et al*. Predicting protein structures with a multiplayer online game. *Nature* 2010;**466**: 756–60.
17. Leaman R, Good BM, Su AI, *et al*. Session Introduction. In: *Pacific Symposium on Biocomputing*, **Vol. 20**, Hawaii, 2015, pp. 267–9.
18. White RW, Tatonetti NP, Shah NH, *et al*. Web-scale pharmacovigilance: listening to signals from the crowd. *J Am Med Inf Assoc* 2013;**20**:404–8.
19. Odgers DJ, Harpaz R, Callahan A, *et al*. Analyzing search behavior of healthcare professionals for drug safety surveillance. In: *Pacific Symposium on Biocomputing*, **Vol. 20**, Hawaii, 2015, pp. 306–17.
20. Freifeld CC, Brownstein JS, Menone CM, *et al*. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug Safety* 2014;**37**:343–50.
21. Leaman R, Wojtulewicz L, Sullivan R, *et al*. Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. In: *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, Association for Computational Linguistics, 2010, pp. 117–25.
22. Tatonetti NP, Ye PP, Daneshjou R, *et al*. Data-driven prediction of drug effects and interactions. *Sci Trans Med* 2012;**4**: 125ra131.
23. Yang CC, Yang H, Jiang L. Postmarketing drug safety surveillance using publicly available health-consumer-contributed content in social media. *ACM Trans Manag Inf Syst* 2014;**5**.
24. Ryan PB, Madigan D, Stang PE, *et al*. Medication-wide association studies. *CPT Pharmacometr Syst Pharmacol*, **2**, e76.
25. Wang W, Haerian K, Salmasian H, *et al*. A drug-adverse event extraction algorithm to support pharmacovigilance knowledge mining from PubMed citations. *AMIA* 2011;**2011**: 1464–70.
26. Ingram DG, Matthews CK, Plante DT. Seasonal trends in sleep-disordered breathing: evidence from Internet search engine query data. *Sleep Breath* 2015;**19**:79–84.
27. Araz OM, Bentley D, Muelleman RL. Using Google Flu Trends data in forecasting influenza-like–illness related ED visits in Omaha, Nebraska. *Am J Emerg Med* 2014;**32**:1016–23.
28. Bichero S. Global Smartphone Installed Base Forecast by Operating System for 88 Countries: 2007 to 2017. WWW document, http://www.strategyanalytics.com/default.aspx?mod=reportabstractviewer&a0=7834.
29. Pickard KT, Swan M. Big Desire to Share Big Health Data: A Shift in Consumer Attitudes toward Personal Health Information. In: *AAAI 2014 Spring Symposia: Big Data Becomes Personal: Knowledge into Meaning,* Association for the Advancement of Artificial Intelligence, 2014.
30. Morton CC. Innovating openly: researchers and patients turn to crowdsourcing to collaborate on clinical trials, drug discovery, and more. *IEEE Pulse* 2014;**5**:63–7.
31. Turner-McGrievy GM, Helander EE, Kaipainen K, *et al*. The use of crowdsourcing for dietary self-monitoring: crowdsourced ratings of food pictures are comparable to ratings by trained observers. *J Am Med Inf Assoc* 2014, doi: 10.1136/amiajnl-2014-002636.
32. Huang M, Neveol A, Lu Z. Recommending MeSH terms for annotating biomedical articles. *J Am Med Inf Assoc* 2011;**18**: 660–7.
33. Lu Z, Kim W, Wilbur WJ. Evaluation of Query Expansion using MeSH in PubMed. *Inf Retr Boston* 2009;**12**:69–80.
34. Lu Z, Kim W, Wilbur WJ. Evaluating relevance ranking strategies for MEDLINE retrieval. *J Am Med Inf Assoc* 2009;**16**:32–6.
35. Van Auken K, Schaeffer ML, McQuilton P, *et al*. BC4GO: a full-text corpus for the BioCreative IV GO task. *Database* 2014; **2014**:bau074.

36. Islamaj Dogan R, Neveol A, Lu Z. A context-blocks model for identifying clinical relationships in patient records. *BMC Bioinformatics* 2011;**12**(Suppl 3), S3.

37. Leaman R, Khare R, Lu Z. NCBI at 2013 ShARe/CLEF eHealth Shared Task: Disorder Normalization in Clinical Notes with DNorm. In: *CLEF 2013 Evaluation Labs and Workshop*. The CLEF Initiative, Valencia - Spain, pp. 23–26, 2013.

38. Neveol A, Islamaj Dogan R, Lu Z. Semi-automatic semantic annotation of PubMed queries: a study on quality, efficiency, satisfaction. *J Biomed Informatics*, **44**, 310–18.

39. Zhai H, Lingren T, Deleger L, *et al.* Web 2.0-based crowdsourcing for high-quality gold standard development in clinical natural language processing. *J Med Internet Res* 2013;**15**:e73.

40. Leaman R, Islamaj Dogan R, Lu Z. DNorm: disease name normalization with pairwise learning to rank. *Bioinformatics* 2013; **29**:2909–17.

41. Leaman R, Wei CH, Lu Z. tmChem: a high performance tool for chemical named entity recognition and normalization. *J Cheminform* 2015;**7**(Suppl 1):S3.

42. Baumgartner W, Lu Z, Johnson HL, *et al.* An integrated approach to concept recognition in biomedical text. In: *Proceedings of the Second BioCreative Challenge Evaluation Workshop*, 2007, Vol. 23, pp. 257–71.

43. MacLean DL, Heer J. Identifying medical terms in patient-authored text: a crowdsourcing-based approach. *J Am Med Inform Assoc* 2013;**20**:1120–7.

44. Lafferty JD, McCallum A, Pereira FCN. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: *Proceedings of ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning*, 2001, pp. 282–9.

45. Good BM, Nanis M, Wu C, *et al.* Microtask Crowdsourcing for Disease Mention Annotation in PubMed Abstracts. In: *Pacific Symposium on Biocomputing*, **Vol. 20**, Hawaii, 2015, pp. 282–93.

46. Dogan RI, Leaman R, Lu Z. NCBI disease corpus: a resource for disease name recognition and concept normalization. *J Biomed Inform* 2014;**47**:1–10.

47. Dogan RI, Lu Z. An improved corpus of disease mentions in PubMed citations. In: *Workshop on Biomedical Natural Language Processing*. Association for Computational Linguistics, Montreal, Canada, 2012, pp. 91–9.

48. Wei CH, Kao HY, Lu Z. PubTator: a web-based text mining tool for assisting biocuration. *Nucleic Acids Res* 2013;**41**:W518–22.

49. Wei CH, Harris BR, Li D, *et al.* Accelerating literature curation with text-mining tools: a case study of using PubTator to curate genes in PubMed abstracts. *Database* 2012;**2012**:bas041.

50. Arighi CN, Carterette B, Cohen KB, *et al.* An overview of the BioCreative 2012 Workshop Track III: interactive text mining task. *Database* 2013;**2013**:bas056.

51. Lu Z, Hirschman L. Biocuration workflows and text mining: overview of the BioCreative 2012 Workshop Track II. *Database* 2012;**2012**:bas043.

52. Khare R, Li J, Lu Z. LabeledIn: cataloging labeled indications for human drugs. *J Biomed Inform* 2014;**52**:448–56.

53. Burger JD, Doughty E, Bayer S, *et al.* Validating candidate gene-mutation relations in MEDLINE abstracts via crowdsourcing. *Data Integr Life Sci Lect Notes Comput Sci* 2012;**7348**: 83–91.

54. Wei CH, Kao HY, Lu Z. SR4GN: a species recognition software tool for gene normalization. *PloS One* 2012;**7**:e38460.

55. Doughty E, Kertesz-Farkas A, Bodenreider O, *et al.* Toward an automatic method for extracting cancer- and other disease-related point mutations from the biomedical literature. *Bioinformatics* 2011;**27**:408–15.

56. Mortensen JM, Minty EP, Januszyk M, *et al.* Using the wisdom of the crowds to find critical errors in biomedical ontologies: a study of SNOMED CT. *J Am Med Inform Assoc* 2014, doi: 10.1136/amiajnl-2014-002901.

57. Brown AW, Allison DB. Using crowdsourcing to evaluate published scientific literature: methods and example. *PloS One* 2014;**9**:e100647.

58. Finn RD, Gardner PP, Bateman A. Making your database available through Wikipedia: the pros and cons. *Nucleic Acids Res* 2012;**40**:D9–12.

59. Loguercio S, Good BM, Su AI. Dizeez: an online game for human gene-disease annotation. *PloS One* 2013;**8**:e71171.

60. Good BM, Howe DG, Lin SM, *et al.* Mining the Gene Wiki for functional genomic knowledge. *BMC Genomics* 2011;**12**:603.

61. Waldispühl J, Kam A, Gardner P. Crowdsourcing RNA Structural Alignments with an Online Computer Game. In: *Pacific Symposium on Biocomputing*, **Vol. 20**, Hawaii, 2015, pp. 330–41.

62. Ahn LV, Dabbish L. Designing games with a purpose. *Commun ACM* 2008;**51**:58–67.

63. Gottl F. Crowdsourcing with Gaminification. *Advances in Embedded Interactive Systems*, **Vol. 2**, Passau, Germany, 2014.

64. Stegman M. Immune Attack players perform better on a test of cellular immunology and self confidence than their classmates who play a control video game. *Faraday Discuss* 2014; **169**:403–23.

65. Perry D, Aragon C, Cruz S, *et al.* Human centered game design for bioinformatics and cyberinfrastructure learning. In: *Proceedings of the ACM Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery*. Association for Computing Machinery (ACM), 2013.

66. Kawrykow A, Roumanis G, Kam A, *et al.* Phylo: a citizen science approach for improving multiple sequence alignment. *PloS One* 2012;**7**:e31362.

67. Khatib F, DiMaio F, Foldit Contenders G, *et al.* Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat Struct Mol Biol* 2011;**18**:1175–7.

68. Lee J, Kladwang W, Lee M, *et al.* RNA design rules from a massive open laboratory. *Proc Natl Acad Sci USA* 2014;**111**:2122–7.

69. Treuille A, Das R. Scientific rigor through videogames. *Trends Biochem Sci* 2014;**39**:507–9.

70. Riedel-Kruse IH, Chung AM, Dura B, *et al.* Design, engineering and utility of biotic games. *Lab Chip* 2011;**11**:14–22.

71. Riedel-Kruse I, Blikstein P. Biotic games and cloud experimentation as novel media for biophysics education. *Bulletin of the American Physical Society* 2014;**59**.

72. Moult J, Pedersen JT, Judson R, *et al.* A large-scale experiment to assess protein structure prediction methods. *Proteins* 1995; **23**:ii–v.

73. Arighi CN, Wu CH, Cohen KB, *et al.* BioCreative-IV virtual issue. *Database* 2014;**2014**:bau039.

74. Wu CH, Arighi CN, Cohen KB, *et al.* BioCreative-2012 virtual issue. *Database* 2012;**2012**:bas049.

75. Kim JD, Pyysalo S, Ohta T, *et al.* Overview of BioNLP Shared Task 2011. In: *Proceedings of BioNLP Shared Task 2011 Workshop*. Association for Computational Linguistics, Portland, OR, 2011, pp. 1–6.

76. Hersh W, Voorhees E. TREC genomics special issue overview. *Inf Retr* 2009;**12**:1–15.

77. Uzuner O, South BR, Shen S, *et al.* 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J Am Med Inform Assoc* 2011;**18**:552–6.

78. Huang M, Lu Z. Community Challenges in Biomedical Text Mining over 10 Years: Success, Failure, and the Future. *Brief Bioinformatics* 2016;**17**(1):132–44.

79. Marbach D, Costello JC, Kuffner R, *et al*. Wisdom of crowds for robust gene network inference. *Nat Methods* 2012;**9**: 796–804.

80. Margolin AA, Bilal E, Huang E, *et al*. Systematic analysis of challenge-driven improvements in molecular prognostic models for breast cancer. *Sci Trans Med* 2013;**5**:181re181.

81. Khoury GA, Liwo A, Khatib F, *et al*. WeFold: a coopetition for protein structure prediction. *Proteins* 2014;**82**:1850–68.

82. Ross J, Irani I, Silberman MS, *et al*. Who are the Crowdworkers?: Shifting Demographics in Amazon Mechanical Turk. In: *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. ACM, Atlanta, Georgia, 2010, pp. 2863–72.

83. Butler D. When Google got flu wrong. *Nature* 2013;**494**: 155–6.

84. Fort K, Adda G, Cohen KB. Amazon Mechanical Turk: Gold Mine or Coal Mine?. In: *Computational Linguistics*, 37(2) (2011), pp. 413–20.

85. Machine D, Ophoff J. Understanding What Motivates Participation on Crowdsourcing Platforms. In: *e-Skills for Knowledge Production and Innovation Conference*, Cape Town, South Africa, 2014, pp. 191–200.

86. Parvanta C, Roth Y, Keller H. Crowdsourcing 101: a few basics to make you the leader of the pack. *Health Promot Pract* 2013; **14**:163–7.

87. Khare R, An Y, Song IY, *et al*. Can clinicians create high-quality databases: a study on a flexible electronic health record

(fEHR) system. In: *International Health Informatics Symposium*. ACM, Washington, DC, 2010, pp. 8–17.

88. Wagholikar KB, MacLaughlin KL, Kastner TM, *et al*. Formative evaluation of the accuracy of a clinical decision support system for cervical cancer screening. *J Am Med Inform Assoc* 2013; **20**:749–57.

89. Khare R, An Y, Wolf S, *et al*. Understanding the EMR error control practices among gynecologic physicians. In: *iConference 2013*, iSchools, Fort Worth, Texas, 2013, pp. 289–301.

90. Blair DR, Wang K, Nestorov S, *et al*. Quantifying the impact and extent of undocumented biomedical synonymy. *PLoS Comput Biol* 2014;**10**:e1003799.

91. Yu B, Willis M, Sun P, *et al*. Crowdsourcing participatory evaluation of medical pictograms using Amazon Mechanical Turk. *J Med Int Res* 2013;**15**:e108.

92. Nevin CR, Westfall AO, Rodriguez JM, *et al*. Gamification as a tool for enhancing graduate medical education. *Postgrad Med J* 2014;**90**:685–93.

93. Brown HR, Zeidman P, Smittenaar P, *et al*. Crowdsourcing for cognitive science–the utility of smartphones. *PloS One* 2014;**9**: e100662.

94. Harpaz R, DuMouchel W, LePendu P, *et al*. Empirical Bayes Model to Combine Signals of Adverse Drug Reactions. In: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 1339–47.

95. Neilson J. Participation inequality: lurkers vs. contributors in internet communities. *Jakob Nielsen's Alert-box,* 2006.