

# CITIZEN SCIENCE: Getting Cheap, Reliable Help from Lay Workers

By Esther Landhuis

In pre-Internet days, people sought expert advice for their purchasing decisions—consulting product ratings in magazines such as *Consumer Reports* and reading newspaper reviews of local dining spots. Now, many rely instead on feedback from ordinary folks who post to websites like Amazon or Yelp. The soaring popularity of such online reviews signals the value of the crowd: Even if any single lay opinion might seem dubious, the wisdom of the group offers a powerful source of information.

Biomedical researchers have watched these developments eagerly. “Those of us on the interface between technology and biology were thinking, ‘Hey, how can we apply crowdsourcing to problems and challenges that we care about?’” recalls bioinformatician **Andrew Su, PhD**, of Scripps Research Institute in La Jolla, California.

## Data Deluge

In recent years, the potential value of crowdsourcing has grown as biomedical researchers drown in data. Advances in online technology and DNA sequencing have plunged biomedicine into a new era in which genome-scale analyses churn out data faster than anyone can make sense of it.

So, for the last decade, Su’s research group has reached out to crowds to organize biomedical information. One of their first projects was Gene Wiki, a collection of 10,000-plus pages with information on human genes and proteins. The portal is hosted on Wikipedia, the free online encyclopedia that made a splash in 2001 and is now the world’s sixth-most-visited website with 34.5 million articles written by more than 53 million people worldwide. “None of their content is created by paid individuals. It’s all on the backs of volunteers,” Su says. “It speaks to the power of crowdsourcing.”

The idea for Gene Wiki emerged as genome-scale experiments became more powerful and feasible. It’s not uncommon now for a single experiment to produce

a list of some 500 genes expressed at different levels in cancer versus normal cells. While a researcher might know something about one or two of those genes, Su notes, “I need to quickly get up to speed on the other 498 that I’m not familiar with—to understand if they’re relevant to my system or worth further study.”

Resources like Gene Wiki require tons of biocuration—the process of combing through biomedical literature and putting its content into structured databases that can be queried for statistics and trends. The National Institutes of Health spends millions of dollars each year hiring professional scientists to do biocuration. “We hope to make that process more efficient by engaging crowds,” Su says. “The more we can get our crowd to do, the more professionals can focus on really hard problems.”

Su’s eventual goal is to build a Network of BioThings. This system would organize the torrents of data that currently flood the PubMed database at a rate of one or two new articles per minute. “Keeping up with the literature is incredibly hard,” Su says. Rather than spending weeks scouring abstracts, it would help if researchers could glean their useful tidbits by querying a knowledge base. Building it would involve surveying publications for key “bio things”—genes, proteins, mutations, diseases and drugs—and documenting relationships between them.

The screenshot shows a web interface for a PubMed abstract. On the left, there is a list of terms: "Nervous system disorder", "Genetic information", "Renal amyloidosis", and "Renal amyloidosis". On the right, there is a list of terms: "autosomal recessive disorder". A green box with the number "76" is visible. Below the lists, the title of the abstract is "MEFV-Gene analysis in armenian patients with familial Mediterranean fever: diagnostic value and unfavorable renal prognosis of the M694V homozygous genotype-genetic and therapeutic implications." The abstract text follows, starting with "Familial Mediterranean fever (FMF) is a recessively inherited disorder that is common in patients of Armenian ancestry. To date, its diagnosis, which can be made only retrospectively, is one of exclusion, based entirely on nonspecific clinical signs that result from serosal inflammation and that may lead to unnecessary surgery. Renal amyloidosis, prevented by colchicine, is the most severe complication of FMF, a disorder associated with mutations in the MEFV gene. To evaluate the diagnostic and prognostic value of MEFV-gene analysis, we investigated 90 Armenian FMF patients from 77 unrelated families that were not selected through genetic linkage analysis. Eight mutations, one of which (R128Q) is new, were found to account for 93% of the 163 independent FMF alleles, with both FMF alleles identified in 89% of the patients. In several instances, family studies provided molecular evidence for pseudodominance, heterozygosity, and incomplete penetrance of the disease phenotype. The M694V homozygous genotype was found to be associated with a higher prevalence of renal amyloidosis and FMF, compared with other genotypes (P = .0002 and P = .036, respectively). The demonstration of both the diagnostic and prognostic value of MEFV analysis and particular modes of inheritance should lead to new ways for management of FMF—including genetic counseling and therapeutic decisions in affected families." A legend at the bottom indicates "Correct annotations" and "Your annotations".

*In the training phase of the Mechanical Turk project, lay workers were given feedback on how well they did on a task, such as identifying disease names. Reprinted from BM Good, et al., *Microtask Crowdsourcing for Disease Mention Annotation in Pubmed Abstracts*. *Biocomputing* 2015: pp. 282-293.*

## Building and Debugging Databases

As a first step, Su and colleagues tested if they could crowdsource this sort of biocuration to lay people. Using Mechanical Turk—a web platform for harnessing human intelligence to do things computers can’t do well—they asked workers to highlight disease-related terms in 593 PubMed abstracts.

This job had previously been done by 12 professional biocurators working part-time for a good part of the summer, an effort that Su estimates cost tens of thousands of dollars. With crowdsourcing, 145 lay workers completed the work in nine days. Each document was scanned by 15 novices who earned six cents per abstract. The upshot: Six

“The more we can get our crowd to do, the more professionals can focus on really hard problems,” Su says.

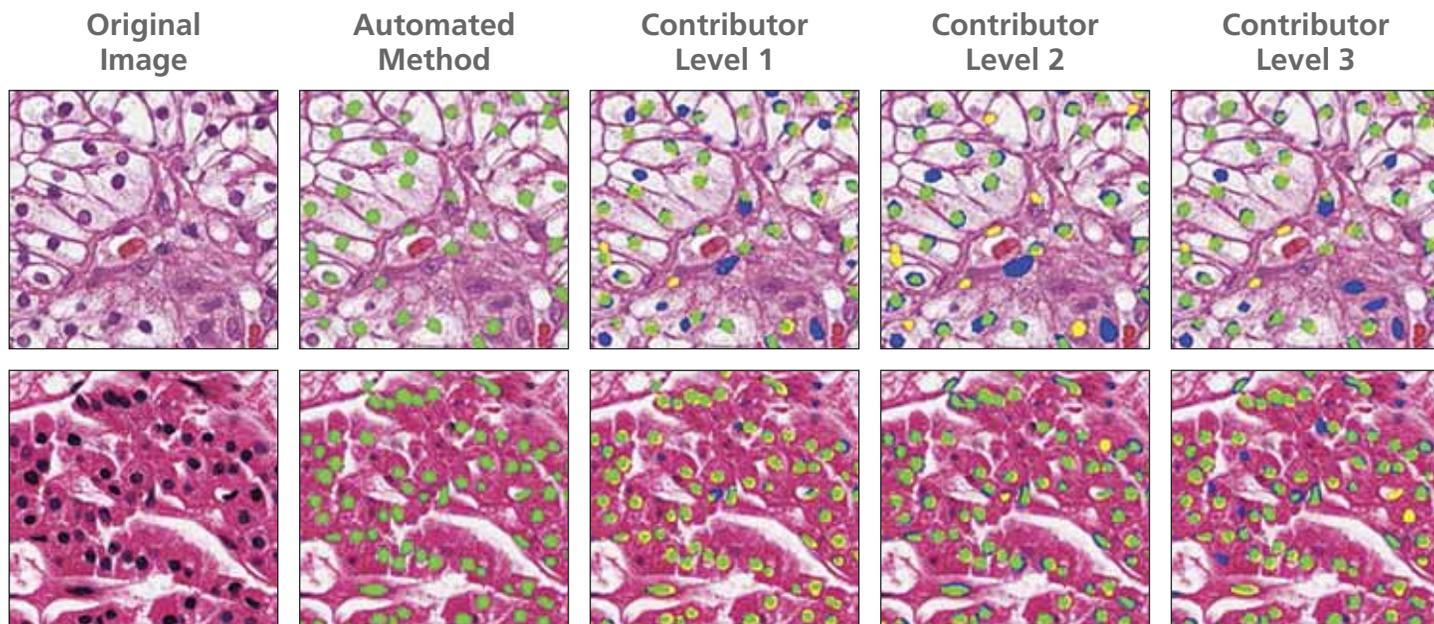
novices in aggregate did as well as, if not better than, one PhD biocurator, and at a fraction of the cost (\$631 total including time for training). Su reported these results at the Pacific Symposium on Biocomputing (PSB) held January 4–8 in Hawaii. And Su has re-

who is allergic to its key ingredient, acetaminophen, which SNOMED knows is the main ingredient of Tylenol.

Asked to verify relationships and find mistakes in a subset of SNOMED CT terms, lay workers performed “on par with experts” and

says. The challenge is getting enough annotated images to build the algorithms.

For a study he reported at the January 2015 PSB, Beck and colleagues showed lay workers a set of renal cell carcinoma images and asked them to identify and delineate



Lay workers assigned the task of determining nuclear boundaries of potentially cancerous cells did a better job than state-of-the-art automated methods. These slides show two examples (top and bottom) of nuclear segmentation using an automated method and three increasing contributor skill levels. Green region indicates a true positive region, yellow region indicates a false negative region and blue region indicates a false positive region. Reprinted from H. Irshad, et al., *Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: Evaluating experts, automated methods, and the crowd*, *Biocomputing 2015, Proceedings of the Pacific Symposium*, pp. 294-305.

cost one-fourth as much, Mortensen says. He was the lead author on a November 13 paper reporting these findings in the *Journal of the American Medical Informatics Association*.

### Many Eyes Make Light Work

Crowdsourcing has also proven useful for annotating images—a huge need in cancer research. Despite tremendous advances in molecular biology that allow researchers to probe thousands of genes and proteins within individual cells, “the single most useful tool for diagnosing cancer is a microscope. It’s the convergence of all this complex molecular data,” says Andrew Beck,

the boundaries of nuclei, which contain the cell’s DNA. The size and shape of the nucleus, as well as how dark or light it appears under a microscope, can help researchers distinguish cancer cells from normal tissue.

On the first task—identifying nuclei—automated approaches did about as well as the crowd. However, for determining nuclear boundaries, human eyes did considerably better than state-of-the-art methods, Beck says.

Though this study only focused on two small tasks, Beck says the approach could be extended “to something as complex as making diagnoses.” However, he notes, computers won’t replace human expertise.

On the first task — identifying nuclei — automated approaches did about as well as the crowd. However, for determining nuclear boundaries, human eyes did considerably better than state-of-the-art methods, Beck says.

cently shown that the same work could be done just as reliably—and free of charge to the researchers—on the Mark2Cure.org site, which allows interested people to volunteer their time to contribute to research. In May, Mark2Cure launched a biocuration campaign aimed at aiding rare disease research.

In addition to building databases, crowdsourcing can help fix them. Mark Musen, PhD, and Jonathan Mortensen of Stanford University sought crowd help to find errors in SNOMED CT, a set of clinical terms and concepts becoming more critical as hospitals switch to electronic medical records. For example, typing “Tylenol” into a SNOMED system could warn a physician to avoid prescribing this medication to a patient

MD, PhD, a molecular pathologist at Beth Israel Deaconess Medical Center in Boston.

Training computers to correlate molecular and microscopy data could help physicians tell if a tumor is benign or malignant, or predict how it might respond to treatment, Beck

When a computational method adds value, more people want to use it—which then creates more complicated results to be interpreted. “Ironically, the better the machines we have, the greater the need for human experts,” Beck says. □