The devices consist of an ipod-sized computer, glucose sensor, and insulin pump (which can be attached to the arms, legs, or stomach). A simple version (made by Medtronic) is already on the market in England: the system monitors glucose and shuts off the pump automatically when blood sugar drops too low. But Doyle’s team is going beyond such a simple feedback loop, using an advanced algorithm called “model predictive control” (which is also used in aerospace controls). “We forecast and anticipate insulin needs,” rather than simply responding to current glucose levels, Doyle says. The algorithm will even adapt to a patients’ individual patterns, such as the timing of exercise and meals, as well as to individual variation in insulin metabolism, Doyle says. “It won’t be a one-size fits all algorithm; it will be tailored and customized to the individual patient.”

Synthesizing Data on Mental Illness

A new center at the University of Chicago will explore the origins of psychiatric disease by integrating existing data from diverse disciplines and across multiple sites. The center is the newest Silvio O. Conte Center for Neurosciences Research and the first with a computational focus. It received $14 million in grants from the National Institutes of Mental Health and the Chicago Biomedical Consortium.

“We have a lot of datasets from different communities that have never been analyzed within the same model before. It’s an exciting research opportunity,” says principal investigator Andrey Rzhetsky, PhD, professor of medicine and human genetics at the University of Chicago Medical Center.

Rzhetsky will head a consortium of 15 lead investigators from seven schools that will bring together clinical data, genetic linkage data, gene pathway data, functional data on genes and proteins, drug data, and drug-gene interaction data. “The main premise of the center is to get together wonderful specialists in different disciplines; make them talk to each other; design models that span all datasets; and make predictions that can be tested experimentally.”

Rzhetsky’s team will attempt to unearth novel connections between genes, environment, and disease phenotypes, as well as between the disorders themselves. For example, Rzhetsky and colleagues have previously shown that autism, schizophrenia and bipolar disorder have considerable genetic overlap. “You can get a lot more from joint analysis of several phenotypes than from a single phenotype,” Rzhetsky says.

LEVERAGING SOCIAL MEDIA: For Biomedical Research

By Katharine Miller

It has become commonplace for people to use social media to share their healthcare stories, seek a community of individuals with the same diseases, and learn about treatment options. All this Internet activity also produces data that can be used for research.

“In the networked world, who cures cancer? We all do,” says Paul Wicks of PatientsLikeMe.

“In the networked world, who cures cancer? We all do,” says Paul Wicks, PhD, director of research and development at PatientsLikeMe, a site where people diagnosed with serious life-changing illnesses can record and share information.

For PatientsLikeMe and a number of other sites, doing biomedical research using data gathered online is part of the business plan. With names such as 23andMe, MedHelp, TUDiabetes, myMicrobes.eu, CureTogether, these sites blend community building with information gathering. They then turn to computational approaches, such as data mining and natural language processing (NLP), to analyze the information gathered.

This crowd-sourced research often reaches into realms that otherwise wouldn’t or couldn’t be studied, due to a lack of either appropriate information or financial support. Moreover, with their access to large populations of both cases and controls, these sites are rapidly producing clinical research results. That they function in a landscape of ever-changing and growing data just makes the process that much more interesting.

Doing Research That Others Can’t or Won’t

On social media healthcare sites such as PatientsLikeMe, people record and share information about their diseases. This self-re-
ported data may have some inherent biases, says Wicks, who hopes that those issues will disappear as they get to a large enough scale. But it also has some inherent strengths: Online, people talk about issues they might not raise with a physician, and they can report on and track their conditions more frequently.

To take advantage of this, PatientsLikeMe set out to “do research that’s new and novel… and not just cheaper than a survey way to capture data that no one else has the bandwidth to look at.”

**Access to Large Populations for Clinical Trials**

At PatientsLikeMe, 23andMe and MedHelp, researchers are finding that online communities offer a huge benefit to clinical research: A vast treasure trove of cases and an even vaster population of controls.

Launched in 2005, PatientsLikeMe has 115,000 users and covers about 1300 conditions. For about twenty of those conditions, PatientsLikeMe collects patient data in a structured way, requesting information on specific outcomes—the kinds of things typically used for clinical trials. “We build it so we can prepare for future research studies,” Wicks says.

For example, early on, the site created a community and several surveys for people with ALS (amyotrophic lateral sclerosis). This meant they already had lots of valuable background data when, in 2008, the community clamored for treatment with lithium. A small (16-person) study in Italy had shown that lithium could slow the progress of the disease. But PatientsLikeMe researchers were wary. “Many studies of ALS treatments kill patients faster than placebo,” Wicks says. “You want to be sure it’s not harmful.” So PatientsLikeMe immediately spent a year gathering data on off-label lithium use by 150 eager ALS patients in their community. And they matched cases to controls in a rigorous way—using an algorithm that considered data on both ALS onset and the shape of the disease progression curve, key traits that vary in significant ways among ALS patients. This was possible, Wicks says, because they had lead-in data describing the patients’ status before taking the drug. Preliminary results announced in December 2008 (just nine months after the Italian research was published) showed that lithium was not effective in slowing disease progress. Since then, this result was confirmed in randomized clinical trials. The PatientsLikeMe research was published in *Nature Biotechnology* in April 2011.

The genotyping service 23andMe does research using data they gather from people who provide not only saliva samples but also phenotype information gathered through online surveys. And the company leverages social media such as Twitter and Facebook to recruit communities of individuals with a particular disease. “Recruiting is not done through a clinical center,” says Chuong (Tom) Do, PhD, a research scientist at the company. “It’s done entirely online.”

For many communities, Do says, “we actually have the genotyping process completely sponsored, making the financial barriers to participation in the research as low as possible.” For example, using a private donation from Google founder Sergey Brin, 23andMe was able to sponsor most of the genotyping costs for PD cases in a re-

![This screenshot of the ALS tracking tool for an individual patient in the PatientsLikeMe lithium study shows how the patients entered their disease characteristics, demographics, blood levels, dosage, ALSFRS-R score (a measure of disease progression), forced vital capacity, and side effects. Reprinted from supplemental figure 1, Wicks, P, et al., Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm, Nature Biotechnology 29, 411–414 (2011).](image-url)
23andMe has the advantage of being able to use data from the population of people who pay for the service. For a less common disease like PD, Do says, the small proportion of misclassified cases mixed in with that population would have a negligible effect on the results of the study. “We just need to be sure to get enough cases,” he says. “Controls come for free. It’s actually a huge help for us.” Indeed in a recent study of PD (published in PLoS Genetics in June of 2011) involving roughly 3400 cases and 29,000 controls, they were able to identify two novel genes contributing to the risk of developing PD. Because of the control group’s size, Do says, “We could wring a lot of statistical power from our dataset.”

23andMe has also launched initiatives to study several rare disorders, namely sarcoma and myeloproliferative neoplasms. While recruitment for these conditions can be difficult and expensive in the setting of a traditional research center, 23andMe’s system allows for aggregation of individuals at low overhead to the company and without regard for geographic barriers, Do says.

With over 12 million users, MedHelp is the largest online health community. The business focuses on helping people track their diseases as well as connecting them with appropriate communities and physicians. In addition, though, they work in partnership with academics, physicians and others to extract useful knowledge from MedHelp’s accumulating data. For example, several physicians examined data on lens implant failures pulled from the eye-care forums on MedHelp (forums that were sponsored by the American Academy of Ophthalmology). The researchers found that multi-focus implants had a much higher failure rate than other types—information that was very valuable to the ophthalmology community.

Rapid Turnaround Time

 Compared to clinical research centers, those who leverage social media web sites can conduct clinical research very quickly. The PatientsLikeMe study of lithium use in ALS was completed in just twelve months—before a randomized clinical trial even began recruitment. In another example, when members of the site’s ALS community raised a question about excessive yawning, PatientsLikeMe published research on the problem in just three months. “The ability to accelerate the pace of research through social media is exciting to me,” Do says. Part of the acceleration comes from the immediate ability to amass and access large cohorts, he says. But it goes beyond that. For example, when 23andMe set out to determine whether its data was reliable enough to replicate published genome-wide association studies (GWAS), they completed the task at lightning speed compared to a typical GWAS. Indeed, it took 23andMe less than one year to replicate and present results from a PD GWAS that had taken the previous researchers almost six years from hypothesis to publication.

Flexibility

If data initially collected online is incomplete or even wrong, it is easily amended by going back to the users with revised surveys. For example, when 23andMe first attempted to replicate a GWAS for celiac disease, they did not find the expected associations. Because their survey had asked “Have you ever been diagnosed with celiac disease,” they believed their study might include some false positives. So they re-worded the question to ask: “Have you ever been diagnosed with celiac disease,” as confirmed by a biopsy of the small intestine.” And with the newly (and rapidly) acquired answers, they were able to replicate four of the six expected associations.

Dealing with changes of this kind also means re-running the GWAS. “Many times based on research results, we’ll ask new questions,” Do says. “So we end up with a very fluid dataset and the need for tools that allow us to work with the data as it constantly changes.” They often run the same GWAS studies repeatedly. “We have over 1000 that we run on a regular basis, culled from the 50 plus surveys,” Do says. That is a unique computational aspect of the work: custom-built software to conduct parallelized GWAS on the same dataset.

Today, 23andMe has more than 120,000 peoples’ genotypes in its database. “We look forward to the day when we have one million plus,” Do says. “We can only imagine the types of discoveries that will be possible with a database that size.”

---

23andMe successfully replicated previous GWAS for a number of diseases as shown here in a chart of success rate (versus total power) by disease class. Expected = number of associations they expected to replicate. Attempts = number of associations they attempted to replicate. The blue dot represents the success ratio (number of successful replications divided by number of expected replications). The black line represents the 95 percent prediction interval for the success ratio. Reprinted from Tung, J, et al., Efficient Replication of Over 180 Genetic Associations with Self-Reported Medical Data, PLoS ONE 6(8) (2011).