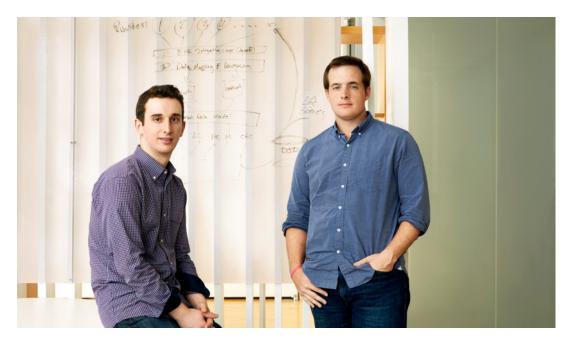# Can Big Data cure cancer?

A tale of two twenty-something computer whizzes, a mountain of money from Google, and one of the oldest, most vexing problems of all time.

You've heard the story before. A couple of whiz kids meet at an elite college, bond over their love of computers, and after a few late-night hacking sessions, build a website or an app. Before you know it, their little side project has turned into a startup, and the fresh-faced youngsters raise piles of cash from investors, decamp for SoMa or SoHo, and form a company that turns them into overnight millionaires, at least on paper.
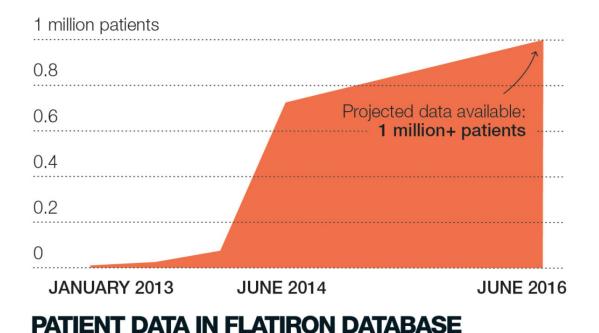
At first glance, Nat Turner and Zach Weinberg seem to fit the bill. They became fast friends after they met on the first day of their freshman year at Penn. Now the youthful duo—both are 28—run a startup in New York. But their entrepreneurial dreams are grander than most. Their budding tech company is not your typical social network, photo sharing app, or dating site. They're not building a tool for teens and twentysomethings to flirt from behind their mobile phone screens or order a late-night snack.

The company they founded two years ago, Flatiron Health, is going after a rather audacious goal: shaking up the health care world. And not just any health care problem at that. Turner and Weinberg hope to collect and analyze mountains of clinical data to make inroads into one of medicine's most complex, research-dependent, and difficult fields: cancer care. Never mind that the pair, who studied economics and entrepreneurship at the Wharton School, didn't have time for as much as a biology class.

Before you snicker, however, consider this. Flatiron is not their first company together. It's their third. They already tried the online food-ordering service for college kids, and it failed. Their second venture, Invite Media, which they started during their junior year at Penn, used big-data techniques to make digital marketing more effective. They were so successful that Google, which is both the godfather of big-data computing and the world's largest digital advertising company, scooped up Invite for more than $80 million in 2010.

Now Google has opened its checkbook again for Turner and Weinberg, investing more than $100 million in Flat- iron via Google Ventures, its venture capital unit. (In all, Flatiron has raised $138 million.) And it's not just Google that the two young ad-tech guys have

impressed. Amy Abernethy, a professor of medicine at Duke and a prominent oncologist, acknowledges that what Flatiron hopes to achieve is exceedingly difficult. Yet she was so taken by the company's thorough and systematic approach to collecting and organizing clinical data that she decided to join its ranks in July as chief medical officer. "I've bet my reputation on jumping ship to Flatiron," says Abernethy, who until recently ran the Duke Cancer Care Research Program and who has been involved in earlier "starry-eyed" efforts to use vast amounts of clinical data to improve cancer care.
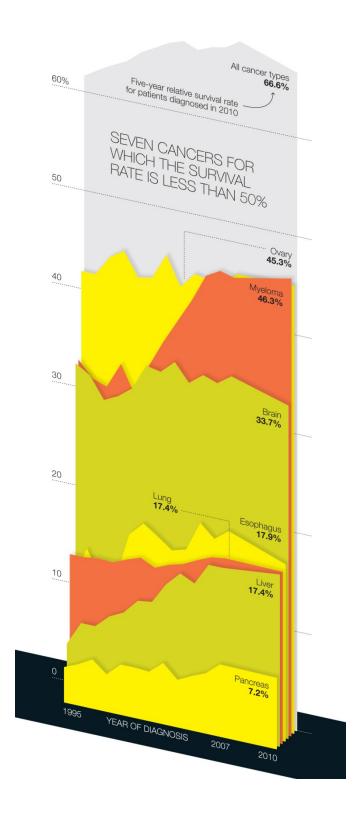

PATIENT DATA IN FLATIRON DATABASE

Such endorsements (and Googlebucks) aside, the sheer hubris of the proposition—that a pair of baby-faced techies can move the needle on cancer care much beyond what armies of research scientists and highly trained physicians have been able to achieve so far—is remarkable. Turner, who is Flatiron's CEO, is both unassuming and undaunted. "We are building a tech company that happens to be in the cancer space," he says. (He uses the word "space" a lot.)

Flatiron's thesis is the following: Currently only a small fraction of the cancer patient treatment data is being collected systematically. That collection largely takes place in randomized clinical trials, which cover about 4% of adult cancer patients (though estimates vary). By organizing and standardizing much of the information for the remaining 96% or so and then offering that data back to physicians, Flatiron thinks it can help doctors come up with better treatment options.

In theory, doctors would be able to see what therapies worked best with most patients in similar circumstances, for example, or they'd be able to evaluate their own outcomes with, say, breast-cancer treatment against those of other specialists across the nation and correct any deficiencies quickly. The data could also shine a spotlight on cost-effective therapies and, conversely, on wasteful health care spending, and it could help match more patients with suitable clinical trials, perhaps speeding up the development and approval of new medicines. Turner and Weinberg are not peddling some tech-utopian dream of ending the cancer burden through data, but they hope that the gains can be meaningful. "All we are doing is saying that if we can borrow from other industries, there is value in data," Turner says. "How much, we don't know yet." Even if the gains are small, they could affect millions. "If we can have a 5% impact across cancer types…" he says before his voice trails off. With

nearly 1.7 million Americans newly diagnosed with cancer in 2014, a 5% improvement in patient survival overall would equate to saving tens of thousands of lives this year alone (see graphic).

Five-year relative survival rate for patients diagnosed in 2010

All cancer types
66.6%

SEVEN CANCERS FOR WHICH THE SURVIVAL RATE IS LESS THAN 50%

Ovary
45.3%

Myeloma
46.3%

Brain
33.7%

Lung
17.4%

Esophagus
17.9%

Liver
17.4%

Pancreas
7.2%

YEAR OF DIAGNOSIS

1995   2007   2010

Turner, whose geophysicist father worked in oil exploration, grew up bouncing between Texas, Louisiana, the Netherlands, and Scotland. Colleagues describe him as "an old soul" with an easy smile. His light-brown hair has begun to recede, exposing a broad forehead, but he still has the youthful looks of someone who might be in grad school. On a recent morning, he's wearing a polo shirt, with a backpack slung over one shoulder and one of those rubber bracelets that are sold to raise money for various causes. Sitting in a coffee

shop across from Flatiron's former one-room "headquarters" in Tribeca (the company has since moved to more spacious digs in SoHo), Turner describes Flatiron's plans matter-of-factly. He has neither the swagger common in successful serial entrepreneurs nor the imperiousness of innovators who believe they can change the world. But he and Weinberg, who grew up on Manhattan's Upper West Side, share a firm conviction that technology can make a difference in the lives of cancer patients. "As you learn about the health care space and about oncology, you find so many problems that a good team that is humble and willing to ask the right questions can help solve," Weinberg says.

Turner first became interested in cancer in 2009. During a family vacation in North Carolina, his 7-year-old cousin, Brennan Simkins, got sick, and after a series of tests he was diagnosed with acute myeloid leukemia. It was the beginning of a years-long ordeal involving a bone-marrow transplant, a relapse, another transplant, and another recurrence of the cancer. In all, Simkins endured four grueling marrow transplants and suffered misdiagnoses as well. Now 12 years old, Simkins has been in remission since 2011.

His agonizing story of survival prompted Turner and Weinberg to start brainstorming about what they could do to help others in similar situations. As they discussed ideas for a new company, they first considered offering a second-opinion service to be delivered over the Internet. Over a period of six months, says Turner, whose rubber bracelet has Simkins's name etched in it, "we went all in on cancer." The pair—often accompanied by Krishna Yeshwant, a physician and Google Ventures partner who specializes in life sciences investments—visited some 60 cancer centers, talking with experts, going on rounds with physicians, and discussing possible business ideas.

After scores of conversations, Turner and Weinberg converged on a new idea: organizing the mountains of clinical data that were scattered in the filing systems of oncology treatment centers around the country. They proposed to collect the data—both digital and otherwise—then organize it, aggregate it, and offer it back to physicians with the goal of helping them make better decisions about how to treat patients.
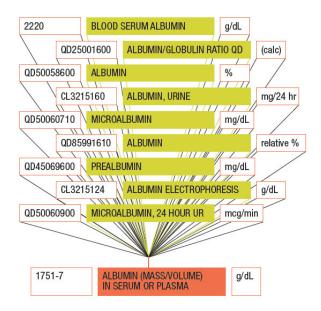
For two data geeks like Turner and Weinberg, the problem with clinical oncology data was both apparent and familiar. Despite years of effort by the medical establishment to persuade doctors and hospitals to embrace electronic medical records, or EMRs, oncology data has remained difficult to access and use. "EMR data sucks," Turner says. Data on a single patient can come from dozens of sources: internists, oncologists, radiologists, surgeons, lab and pathology reports, and more. Even when digitized, data is often in what techies call an unstructured format. Rather than being neatly organized in databases, it appears in different ways across different lab reports and records. To make matters worse, piles of data remain hidden in reports that have been written by hand and scanned, in audio recordings that no one listens to, and in low-resolution PDF files that were spit out by fax machines. Finally, a vast array of incompatible systems—and strict privacy regulations that govern personal health information—makes it yet more difficult for data to be shared across thousands of oncology practices.

The challenge, to put it bluntly, is immense.

In theory, electronic medical records were supposed to make such data aggregation and integration easy. But con- sider this: When it came to measuring the level of a single protein—albumin, commonly tested in cancer patients—a single EMR from a single cancer clinic showed results in more than 30 different formats (see graphic below). Take that challenge and multiply it by more than 100 different kinds of protein and genetic tests,

biopsies, and other diagnostic methods used in cancer care—then multiply it again across the number of separate EMR systems and cancer centers—and you begin to grasp the complexity of the problem.

A single electronic medical record (EMR) system from one cancer center showed lab results for Albumin, a protein measured in cancer patients, in over 30 ways.

Flatiron works to synthesize this to a single format by (a) creating a common data model across cancer centers and labs, (b) processing the data through matching algorithms that can identify about 90% of the terms used, (c) using a data processing engine to transcode terms in real time, and (d) flag any unmatched terms for review by a doctor or nurse. Flatiron applies the same approach to more than 100 different measurements from blood tests, genetic tests, and biopsy results commonly performed on cancer patients. And it keeps the system current as new tests are developed.

To begin tackling it, Turner and Weinberg spent more than two years building what they call a data model, a way to organize the myriad clinical information into neat categories. Doing so for all forms of cancer, they quickly realized, was far too complex. So working with a team of physician advisers, they focused on colon cancer. Using published clinical trials, they extracted more than 350 data categories on everything from demographics and location to cancer stages, biological markers of disease, and responses to therapies. Then they repeated the process for other forms of cancer.

To automate the process of extracting data from EMRs, which can be extremely labor intensive, Flatiron used various computer techniques, including matching algorithms aimed at pinpointing values in lab reports. They also fine-tuned a technique called natural-language processing that allows computers to "read" documents and extract data from them. Such systems are notoriously prone to error, so Flatiron created a hybrid human-machine learning system to catch and correct mistakes. Translation: The company hired a team of 50 nurses to enter data on 500 patients by hand, creating what Turner calls a "training set" that can be used to detect errors in data collected automatically. The discrepancies were then fed back into the system to help perfect the automated collection process. It's a dynamic system intended to improve its accuracy as it goes along. (In theory, anyway.)

Flatiron is not the first to embark on this mission. Last year the American Society of Clinical Oncology, a nonprofit professional association, announced CancerLinQ, an effort to develop a system that would tap clinical databases to help efforts to improve quality of care and speed drug discovery. Cancer Commons, a not-for-profit run by veteran computer scientist Marty Tenenbaum, is hoping to help further the standardization of oncology clinical data in a form that would be freely available to anyone. And IBM, through its Watson artificial intelligence system, is already working with cancer centers like Memorial Sloan Kettering to sift through millions of records of clinical data, text from journal articles, and reports from clinical trials to automatically provide doctors with treatment recommendations for their patients.

And earlier efforts—including a massive $500 million National Cancer Institute bioinformatics program called caBIG—either have failed miserably or have yet to yield positive results. However, Abernethy, who chaired the advisory board of CancerLinQ and has been involved in other initiatives to digitally compile clinical oncology data, says she has been impressed by Flatiron's focus on the complexity of the data. Turner and Weinberg clearly understand that "trying to solve this with just technology isn't going to work," she says. "This is the reason why I decided to work with these guys."

The Google Ventures investment not only boosted Flatiron's credibility but also gave it the means to acquire Altos Solutions, which makes an EMR service for oncology practices. The acquisition of the company, which is based in Mountain View, Calif., not far from Google, gives Flatiron a bigger installed base and closer contact with physicians. Flatiron systems are now being used in some 210 cancer centers that collectively see about 300,000 new patients every year. While most customers are community oncology clinics, large academic institutions like the Smilow Cancer Hospital at Yale-New Haven and the Abramson Cancer Center at the University of Pennsylvania are joining too. Google says it invested in part to accelerate results in a promising field. "We are trying to avoid a situation where it takes another generation for electronic medical records to become widely useful," says Bill Maris, the head of Google Ventures. "I hope we can save people a lot of heartache and suffering."

Across the country, in a nondescript office low-rise in Port Jefferson, a leafy Long Island suburb, Dr. Jeffrey Vacirca is sold on the Flatiron vision. Vacirca says the Altos EMR system, which he has been using for a few years, has helped him improve patient care, but much of its potential remains unfulfilled. "There is a lot of data out there, but no one can sort through it, and no one knows what it means," Vacirca says. "That's where I see the importance of Flatiron, of taking all that molecular data and all these treatment outcomes from these millions of patients that they are going to be able to evaluate, categorize, and see what is truly working." Dr. Vacirca calls the Flatiron system an "infrastructure for cancer care." With it, he says he will be able to detect if his medical approach to certain cancers is lagging in some category and make adjustments. He'll also be able to find out whether more patients are eligible for clinical trials. "If you can accrue patients five times faster, think how many drugs you could get through the regulatory process," he adds.

Some leading figures in the field are more skeptical of big data's promise in general when it comes to the long cancer fight. Pioneering researcher Robert Weinberg (no relation to Zach) highlighted the checkered relationship between big data and cancer in a recent essay in the journal Cell. Weinberg, a founding member of MIT's Whitehead Institute for Biomedical Research, noted that the explosion of data sets on everything from the interaction between proteins to the genetic mutations in a tumor has overwhelmed researchers' ability to interpret it. "There are people who are enthralled with bioinformatics," Weinberg told Fortune in a later interview. "The idea of aggregating data, and assuming that from that

alone, one can get insights that are qualitative and that were not previously accessible, is not obvious to me."

Weinberg adds that even if data were able to pinpoint improvements in outcomes for certain treatment protocols, the gains may not be significant enough for doctors to change their practices. "There have been a lot of brave attempts and a lot of optimistic claims," Weinberg says. "Relative to the effort that's been put into it, there's been little in take-home lessons."

John Ioannidis, a professor of medicine and health research and policy at Stanford, is more generous in his assessment—but only slightly. The ability to match patient profiles with treatments through a centralized system could help reduce the wide variability in cancer treatments across clinics and hospitals, he says. But Ioannidis doubts that major advances could result from data collected outside highly controlled clinical trials. "It's an open question as to how much we can learn from big compilations of data collected with- out experimental design," he says.

Turner and Zach Weinberg don't expect to win over skeptics quickly. But they and many in Flatiron's growing stable of customers believe their "smart data" approach will deliver better care for cancer patients. Abernethy says, among other things, it could initially help close the gap between the care available at community clinics and top academic hospitals, which often have better outcomes. Co-founder Weinberg, meanwhile, says Flatiron is in the fight against cancer for the long haul. "We're a two- year-old startup with a big, ambitious plan. We have taken a pretty good first crack at it, but ultimately this is a decades-long problem."