







# Elementary, Watson

Cancer researchers now produce more data than their human peers can keep up with. The *Jeopardy!*-playing supercomputer digests it all, and helps elite cancer centers like UNC Lineberger get treatment information to doctors quickly — wherever they are.

*by Janine Latus*



Every week, for every patient who came before the Lineberger Comprehensive Cancer Center tumor review board, Dr. Nirali Patel scoured the genetic sequencing report and the medical literature, trying to match the patient to the drug or clinical trial that might change the course of a cancer, might save a life. Standard treatments had failed these people, so she was racing against time to find something rare and significant.

Every week, Patel, a molecular pathologist, would meet with 20 or so equally overwhelmed experts in every form of cancer, each trying to absorb as much as they could from a deluge of research papers — the biomedical research world now produces about 8,000 new research papers every single day. It was time-consuming and exhausting. “If it sounds complicated, I’m not doing it justice,” said Dr. Ned Sharpless, Lineberger’s director. “If it sounds hard, it’s nearly impossible.”

In an obvious understatement, Sharpless ’88 (’93 MD) told a *60 Minutes* reporter last year, “No one has time to read 8,000 papers a day.” Almost as obviously, that puts cancer doctors in the

position of “deciding on therapy based on information that was always, in some cases, 12, 24 months out of date.”

Each patient’s genome takes up more than 100 gigabytes of data, some of it relevant and some of it background noise. Researchers are in the early stages of figuring out which genetic mutations matter and how they might respond to therapy for each individual. That burgeoning information has led to a

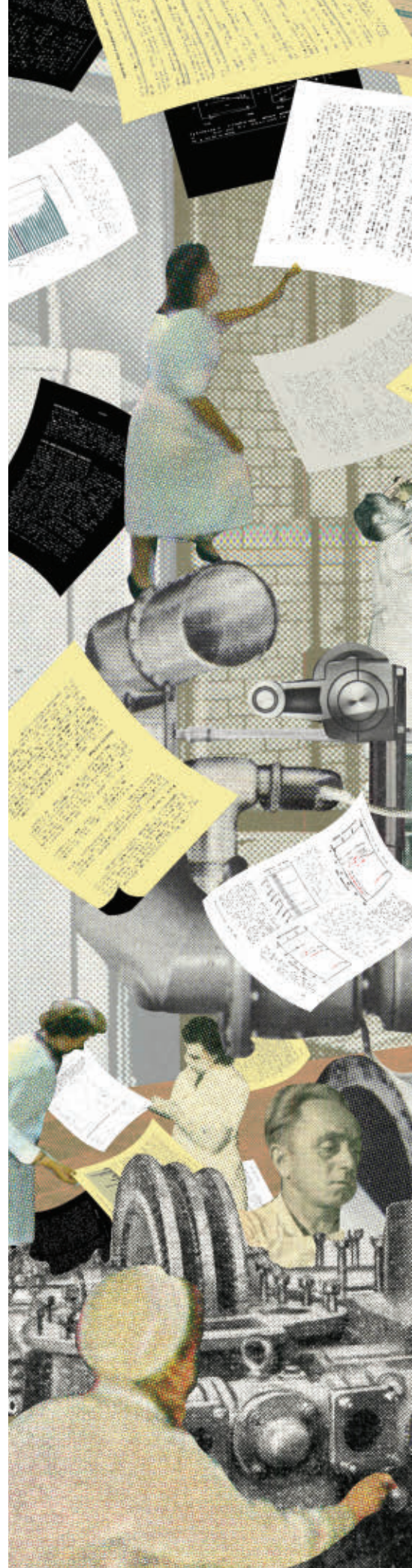
flood of studies, so Patel and other members of the tumor review board read frenetically, updating the list of important mutations as journals come out, always a few months behind because of how long it takes for studies to be published and the group to meet.

If the lung cancer doctor was out of town, his area wouldn’t be discussed that go-round. Or the tumor board members who were reading 10 articles a day on top of their other research and clinical responsibilities missed the one that was important. Add in that they had to keep track of the treatment trials constantly opening and closing at cancer centers all over the country.

Enter Watson, the artificial intelligence-wielding IBM computer that beat the best human *Jeopardy!* contestants a few years ago. Watson was able to learn words and numbers and encyclopedias’ worth of trivia as well as natural language processing and how to comb through millions of gigabytes of unstructured information and extract what was important. In

**‘If it sounds complicated, I’m not doing it justice. If it sounds hard, it’s nearly impossible. ... No one has time to read 8,000 papers a day.’**

**Dr. Ned Sharpless ’88 (’93 MD)**







other words, 8,000 articles a day — can do.

At the time, IBM had a solution searching for a problem. Watson can learn and solve problems; specific to medical research, it can combine information from journal articles, spoken words and images, absorb it all and find the patterns. It needed a big-data challenge — one that mattered. So the company approached 20 major cancer research centers, including UNC, where researchers had amassed a trove of genetic sequencing data. UNC and some of the other cancer institutes plied Watson with it.

“We now have the ability to look at DNA efficiently and cheaply,” said Dr. Neil Hayes ’96 (MD), co-director of clinical bioinformatics at Lineberger and a former member of the tumor board. “We’re the first generation who can look at it and find out how that is helpful for patients.”

### Shifting Watson’s mission

By 2015, Watson was working with 14 U.S. cancer institutes to help guide the treatment of patients, with various approaches. IBM ultimately worked with 20 cancer institutions to gather enough genomic data to train Watson. For the first year of its collaboration, UNC and IBM couldn’t find the right questions.

“UNC’s researchers wanted Watson to look at all of our genomes ... and tell us what drugs to use,” Sharpless said. “IBM had a different problem in mind.”

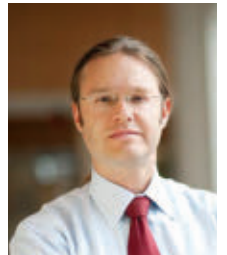
The company wanted cancer research centers to give them a couple of million data, like which drugs worked with which mutations and how patients with various mutations did on particular treatment plans. Then Watson would know which combination of mutation and treatment was ideal. Part of what frustrated then-Vice President Joe Biden during his “cancer moonshot” attempt to cure cancer, Sharpless said, was that there were data on a couple of thousand patients here and on another couple of thousand somewhere else, and the electronic records didn’t talk to each other.

IBM knew its tool was good for human-curated tasks, and the researchers knew they had a difficult task that they weren’t doing very well but was very important. That was the key. Is the data set big; does it change continuously; is it complex and unstructured?

It was Hayes who had the ah-ha moment, Sharpless said. Maybe Watson wouldn’t work for what the Lineberger researchers wanted to do, but could it keep up with all of the literature

**It was Dr. Neil Hayes ’96 (MD) who had the ah-ha moment.**

**If Watson wouldn’t work for what the Lineberger researchers wanted to do, could it keep up with all of the literature and clinical trials opening daily?**





and all of the clinical trials opening daily?

IBM heard that and recognized it as a great problem for Watson because the challenge involved reading a big data set — thousands of papers a day — and drawing conclusions that allow it to make recommendations. It would require natural language processing and deep learning capabilities, just like Watson used for *Jeopardy!*

“Watson has a great ability at this point that if we go in and teach it the language you would deal with in genetics, at a base level how a chromosome relates to a gene relates to a protein relates to amino acids, you teach it how language works in this space and what these relationships and entities are,” said Steve Harvey, vice president of IBM’s Watson Health, “and you start getting answers.”

Best of all, the problem was important, even if all Watson did was improve outcomes by a few percentage points.

“There’s no point in using Watson to make a cake taste 2 percent better,” Sharpless said. “You want problems where 2 percent change or 5 percent change is really meaningful, and I would argue that cancer diagnosis is that problem, because if you’re in that 2 percent, it’s really meaningful. We go to the mat for 2 percent better! That was the secret to our success. We found a problem for which Watson was well-suited.”

What was shocking was how quickly things changed.

### Finding what humans didn’t

“I’m used to, ‘Yeah, this idea,’ and then six months to a year later you get a sense of making progress,” Sharpless said. “IBM said, ‘We can do that,’ and then two weeks later they’d taught Watson to do that.”

In another week, it had read 25 million papers and seen tens of thousands of scans of tumors and healthy cells. It had learned what normal looks like and the many variations of abnormal, so the Lineberger researchers tested it out. They fed in the genetic sequences of more than 1,000 cancer patients and asked what the computer would recommend.

The patients’ cases had gone through two committees, so the researchers knew what the humans had found. They wanted to know whether Watson, provided the same information, would come up with anything different.

Not only did Watson find everything the humans found, but in more than 300 patients Watson found something that had eluded the humans.

“When we looked at what Watson recommended, they were legit, they were not things we should have missed,” Sharpless said. “It was not because we were bad at our job or sloppy — we committed 20 academic physicians to this committee — yet even in that idealized scenario, it still was pretty crummy.”



OLIVER WEISS



Plus, Watson can work around the clock, and it learns the papers that come out immediately rather than months later. Watson also remembers patients and will alert a doctor if a new trial or treatment option opens up.

“From that moment of, ‘Hey, Watson can do that,’ it’s fast-forward a few years and it’s the only way to do it,” Sharpless said. “Usually change takes more like a decade if not 15 years. The crank of clinical trials turns very slowly. So to go from wild-eyed experimental idea to standard of care in two years is unique in my 20 years of research.”

Which doesn’t mean that Watson is curing cancer. That’s where there has been some confusion as its reputation has spread. Some early publicity surrounding the prestigious M.D. Anderson Cancer Center at the University of Texas left the impression that Watson already was revolutionizing cancer care. More accurately, it’s being used to connect to cutting-edge trials the million or so Americans whose cancers haven’t responded to standard treatments.

“What I think about every day when I wake up is only 5 to 7 percent of people who might benefit from this type of a procedure are actually having it done,” Harvey said, “and that’s a really small number. This is one of the great things Ned and his team have done with the UNC program, and it elevates their care to a different level from what most people would have the ability to experience within the United States.”

The utility of tumor genome sequencing is still up for debate. A lot of the mutations that are found don’t seem to

have anything to do with the cancer.

On the other hand, the patients who are relying on this technology have few options, said Dr. Billy Kim, a Lineberger center member and associate professor of medicine and genetics. “So you get in this place of, what does someone with few options do?”

That’s where Watson shines, in helping direct people to clinical trials that might help prolong their lives. The decisions are still up to the physician. Watson simply presents maybe half a dozen trials, recommends which one appears to be best for this particular patient’s genome, and provides the exact research paper supporting that recommendation.

“For five years this took a lot of time per patient, but what Watson does is sort of fast-forward that research and data mining and presents everything all at one go, which helps me spend more time deciding which relevant option is useful for the patient rather than finding the data in the first place,” Patel said. “It speeds me past the icky middle step of sitting in the library and learning stuff and takes me to the higher-level information that I or Ned can rapidly

**Not only did Watson find everything the humans found, but in more than 300 patients Watson found something that had eluded the humans.**



## Southern CULTURES

Spring 2017 brings a special issue on Appalachia, and with it: Dollywood’s hillbilly identity. Country queers in Central Appalachia. Black and white Asheville through a spy camera. The soundscape of Harlan County, U.S.A. And more.

Read, research, & subscribe:  
[SouthernCultures.org](http://SouthernCultures.org)

*Photograph by Aaron Blum, from “Almost Heaven” in the Appalachia Issue (Spring 2017).*



integrate into patient decisions.”

### Spreading the wealth

So now the oncology board essentially has retired, replaced by Watson and Patel.

“Patel plus Watson is better than 20 committed physicians drinking coffee,” Sharpless said, “which frees us to do more research.”

Patel added: “Because Watson has done that groundwork, that frees me as a molecular pathologist to look at how we refine the techniques, how can we do the testing so it’s more sensitive, how we can do it more quickly.

“We no longer have to wade through the data. We can look at one picture and say, ‘Where is the clinical trial closest for this patient?’ Watson can say take drug A or drug B, and Ned can say, ‘This drug is once a week whereas this one needs to be given every day, and this patient has trouble coming to the hospital.’ Ned can figure out which is more likely to work.”

Part of the great promise of Watson is its role in disseminating research information quickly to areas many miles from an elite research center. Watson is not R2D2 or C3PO. It doesn’t have a robot body nor a robotic voice. Interacting with it is more like sending an email. You can do it from a cell phone — for instance, a cell phone from a rural community, far from a major cancer center. That means a one- or two-doctor practice in the Blue Ridge Mountains can work with the most current information.

IBM already has made it available at clinics and doctors’ offices. Now oncologists nationwide can send a portion of biopsy tissue to Watson Genomics, have it sequenced and get a report back quickly.

“The technology we helped them develop you can now buy on the internet,” Sharpless said. It’s now being packaged with sequencing machines, much like Microsoft Office comes packaged on a PC. “The thing we were doing they are now selling as a product.”

With a big hurdle cleared, UNC and Watson currently are taking a pause to determine how best to work together in the future. (No money has changed hands

between UNC and IBM. The University has given the company access to genomic data and researchers’ knowledge in exchange for access to Watson.)

Researchers now are adding in not just DNA but RNA sequences, a move that may finally unlock the secrets of rampant mutation. It’s an advance that would be nearly impossible were it not for Watson.

“Watson is comprehensive,” Kim said, “and patients want to know two things: Are we doing the right thing, and have we considered all of the options? That’s where Watson excels; it’s comprehensive so it leaves no stone unturned. That is really important for patients and their families, that a doctor and their team have thought about everything or are doing everything to help them.” ■

*Janine Latus is a freelance writer based in Chapel Hill.*

**‘We no longer have to wade through the data. We can look at one picture and say, “Where is the clinical trial closest for this patient?” Watson can say take drug A or drug B, and Ned can say, “This drug is once a week whereas this one needs to be given every day, and this patient has trouble coming to the hospital.” Ned can figure out which is more likely to work.’**

### Dr. Nirali Patel



OLIVER WEISS