Norman Sharpless of the University of North Carolina works with IBM Watson Health to analyse DNA data.

CANCER

Reshaping the cancer clinic

Big data's war on cancer is still in the early stages, but the front line is advancing.

BY CHARLIE SCHMIDT

The Cancer Genome Atlas, which catalogues cancer mutations, contains some 2.5 million gigabytes of data. This giant project, run by the US National Institutes of Health, has vastly improved our understanding of various forms of cancer — but it holds relatively little information on the clinical experience of the patients who supplied the samples.

At the other end of the cancer treatment chain, electronic health records contain a wealth of case-specific information that could be used to improve cancer care. But more often than not, such records are isolated in individual hospitals and medical practices. As a result, "most patient experiences are lost to research", says Clifford Hudis, an oncologist who specializes in breast cancer at the Memorial Sloan Kettering Cancer Center in New York.

In an effort to improve cancer treatment, Hudis and many others are now collaborating on efforts to bring together and make sense of the big data that emerge from research, patient

NATURE.COM
For more about personalized cancer care, see:
go.nature.com/dktjla

care and clinical trials. Opportunities for big data extend across most areas of medicine, but "cancer is leading the way", says Lynn Etheredge, a health-care consultant based in Chevy Chase, Maryland. But the ubiquity, variety and lethality of cancer mean that there are plenty of barriers as well as breakthroughs.

Even so, Etheredge, who in 2007 wrote an influential article for *Health Affairs* calling for "rapid learning systems" to handle big data, believes we have entered a historic period for cancer research and treatment. "We know that cancer is a genetic disease, and we have the databases and the computational power needed to analyse them," he says.

Hoping to build on early successes with personalized cancer drugs, oncologists and computer specialists are working together to harness digitized information and apply it in the clinic. These emerging ventures are competing for business and are grappling with difficult questions about privacy, data ownership and sustainable business models. "Big data is both a research tool and a proprietary commodity," Etheredge says. "It's still early days in the field and there's a lot that we need to work out."

Many organizations and approaches are bringing big data to the cancer clinic in the United States, which leads the world in some aspects of cancer treatment. Here we will consider four: a rapidly growing start-up company, a professional association's initiative, a computer giant's cognitive computing and health-care wing, and a network of academic cancer centres.

THE START-UP
Launched in 2009 by scientists at the Broad Institute in Cambridge, Massachusetts, Foundation Medicine bills insurance companies for its analytical services. Academic and community oncologists submit patients' tissue samples, and Foundation Medicine sequences them. It then screens them for genomic cancer drivers against its own growing database of molecular profiles (generated from more than 50,000 cancer patients so far) and data from other public repositories.

"The public databases aren't like Google — oncologists have no easy way to search them for genomic drivers that relate to their own patient's tumour," says Michael Pellini, chief executive of Foundation Medicine. "So we analyse the tissues and report back available therapeutic interventions, either in the form of a drug approved by the US Food and Drug Administration or a clinical trial."

Oncologists can also query Foundation Medicine's client network for advice on difficult cases. Within 72 hours, Pellini says, responses are aggregated and sent to the doctor, who can then gauge whether a particular drug or approach was effective. The company aims to make its client data more broadly available for use in clinical decision-making.

In January 2015, Swiss pharmaceutical giant Roche spent US$1 billion on a 56% stake in Foundation Medicine, the largest corporate player in this sector, expecting revenue this year of more than $85 million.

PRACTICE MAKES PERFECT

In late 2015, the American Society of Clinical Oncology (ASCO) is expecting to launch CancerLinQ, a platform designed to deliver clinical benefits by analysing aggregated electronic health records from thousands of oncology practices.

Oncologists will be able to interrogate CancerLinQ to see the effects of specific interventions, to review how their own treatment approaches stack up against established care standards, and to develop hypotheses for further study.

"Much of what we know about treating cancer comes from clinical trials that enrol just 3% of the patients diagnosed with cancer every year," says Hudis, who serves on CancerLinQ's board of governors. "With CancerLinQ, we're trying to learn from the remaining 97% who don't participate in these studies."

An initial group of 15 'vanguard practices' of varying sizes are participating in the system, which ASCO expects to contain 500,000 patient records by 2016. Researchers and clinicians will be able to query these records to compare patient outcomes by treatment. Aggregating such large amounts of data should help to reveal the effectiveness of particular drugs or approaches.

"The most important thing that CancerLinQ can do is report on outcomes, for instance, that patients who received a particular treatment lived longer, or had slower progression of their disease," says oncologist Robert Miller, medical director of ASCO's Institute for Quality. These insights will benefit patient care and come at a time, he says, when Medicare, the leading US funder of cancer treatment, is shifting from fee-for-service reimbursement to alternative payment models that reward better outcomes.

A prototype of CancerLinQ was tested in a study of 170,000 breast-cancer patients in 2013. According to Miller, unpublished data showed that the system could highlight trends in data submitted by different medical practices — for example, how they stimulate the production of red blood cells to treat anaemia after chemotherapy.

The platform extracts patient data from electronic health records, anonymizes and aggregates the data, and then integrates them with other types of information, including doctors' notes and biomarker repositories. The goal is eventually to add point-of-care decision support to aid physicians with patients whose diagnosis and treatment is problematic.

CancerLinQ currently relies on donations, but Miller says that in time it will sell effectiveness reports and data-exploration tools to make it more self-sustaining. "We are looking at a range of CancerLinQ-related products and services to help offset the operational costs of the system," says Miller.

COGNITIVE COMPUTING

Big data needs big computing, and in 2013 IBM formed a separate business unit — IBM Watson Health — to focus on commercial opportunities in cancer for its Watson cognitive computing system, which combines natural language and learning capabilities. Watson's store of biomedical knowledge includes every abstract in the PubMed database (there are currently about 25 million and counting); the US National Cancer Institute's Drug Dictionary (which has data on both approved drugs and those in clinical trials); the entire catalogue of somatic cancer mutations in the COSMIC (Catalogue of Somatic Mutations in Cancer) database, which is curated by the Wellcome Trust Sanger Institute, in Cambridge, UK; and data from many other sources.

Watson, which gained fame in 2011 by defeating human champions on the US television quiz show *Jeopardy!*, also has access to anonymized patient data. IBM Watson Health has relationships with more than a dozen medical practices, cancer centres and research organizations, says Ajay Royyuru, director of the Computational Biology Center at IBM Research in Yorktown Heights, New York.

"With CancerLinQ, we're trying to learn from the remaining 97% who don't participate in these studies."

The New York Genome Center relies on Watson to screen DNA mutations in patients enrolled in a study of glioblastoma, an often fatal brain cancer.

Physicians at the Memorial Sloan Kettering centre and at the MD Anderson Cancer Center in Houston, Texas, are training Watson to become a clinical support tool, which entails presenting the computer with anonymized and hypothetical cases. For instance, a patient's tumour might test positive for deficiencies in a gene called *STK11* that may respond to the diabetes drug metformin, Royyuru explains. But Watson might not recommend metformin because this is an off-label indication. "That would be an instance in which it could be taught to cast a wider net," Royyuru says.

Andrew Seidman, a breast-cancer specialist at the Memorial Sloan Kettering centre, adds that the use of Watson must be transparent, so that its reasoning can be easily critiqued. And Seidman cautions that Watson isn't ready for prime time yet. "I'm taking a sober view, and I say that as someone who's helping to develop the technology," he says. In particular, Watson's capacity for natural language processing remains a work in progress. For now, instead of speaking to the computer directly, clinicians have to enter the data manually.

NETWORK NEWS

One of the major challenges facing cancer research is how to match patients with targeted drugs that act on rare mutations, because enrolling enough of these patients in clinical trials is not easy. But one group of hospitals has found a way to get round the problem.

Launched in 2014 by the Moffitt Cancer Center, in Tampa, Florida, the Oncology Research Information Exchange Network (ORIEN) comprises nine academic cancer centres. Patients provide clinical data and tissue samples for analysis, and importantly agree to life-long follow-up, which allows patients to be recruited into new trials geared to their own genetic make-up. "It's a much more proactive way of doing research," says Bill Dalton, ORIEN's founding director.

Moffitt developed the protocol, which it calls "total cancer care", in 2003, and created a company — M2Gen — to handle the analyses and tissue storage in 2006. The development of ORIEN gives this protocol a national reach, with about 130,000 people enrolled so far. Member centres share clinical and molecular data, so they can collaborate on research questions.

BIG PRICE TAGS

Extracting clinical insights from big data, and using them to guide treatments, does not come cheaply, however. For example, Foundation Medicine charges nearly $6,000 to sequence and interpret the data from a single solid tumour, and more than $7,000 for a blood cancer.

But this is dwarfed by the cost of new oncology drugs, which often have price tags of more than $100,000 per treatment or per year. In July, US Medicare agreed to pay for a leukaemia drug from Amgen that will cost about $178,000 per patient.

Other countries may bargain far more aggressively with drug companies to bring down prices, or reject the drugs altogether on a cost basis, through agencies such as the UK National Institute for Health and Care Excellence.

Ideally, this big money will buy big gains in personalized treatments and cures. This is certainly the hope of the US Medicare and Medicaid officials confronted with spending more than $13 trillion on health care during the coming decade, much of it on cancer therapy. These agencies will wield enormous power over the practicalities of bringing big data into the clinic. Issues relating to data business models and costs will apply across all areas of medicine, "but cancer is forcing them to the table now", says Etheredge. ∎

Charlie Schmidt *is a freelance science writer based in Portland, Maine.*