

Comment

Cancer research needs a better map

Jesse S. Boehm, Mathew J. Garnett, David J. Adams, Hayley E. Francies, Todd R. Golub, William C. Hahn, Francesco Iorio, James M. McFarland, Leopold Parts & Francisca Vazquez

It is time to move beyond tumour sequencing data to identify vulnerabilities in cancers.

Almost 15 years ago, scientists and clinicians set out to characterize genomes of tumours from thousands of patients. The result? The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC). Nearly every targeted cancer drug approved over the past decade has drawn from the data sets generated by these efforts. This information is now also providing clues to triangulate which individuals can benefit from new types of drug, such as pembrolizumab and nivolumab, which help the immune system to fight cancer. TCGA generated more than 2.5 petabytes of data measuring mutations, gene expression and protein levels across 33 cancer types. It catalysed innovation in DNA sequencing technology and genome analysis. It ultimately collected data from some 11,000 patients – data that thousands of researchers use. This work redefined cancers on the molecular level, and painted a picture of the mutations that occur in common tumour types.

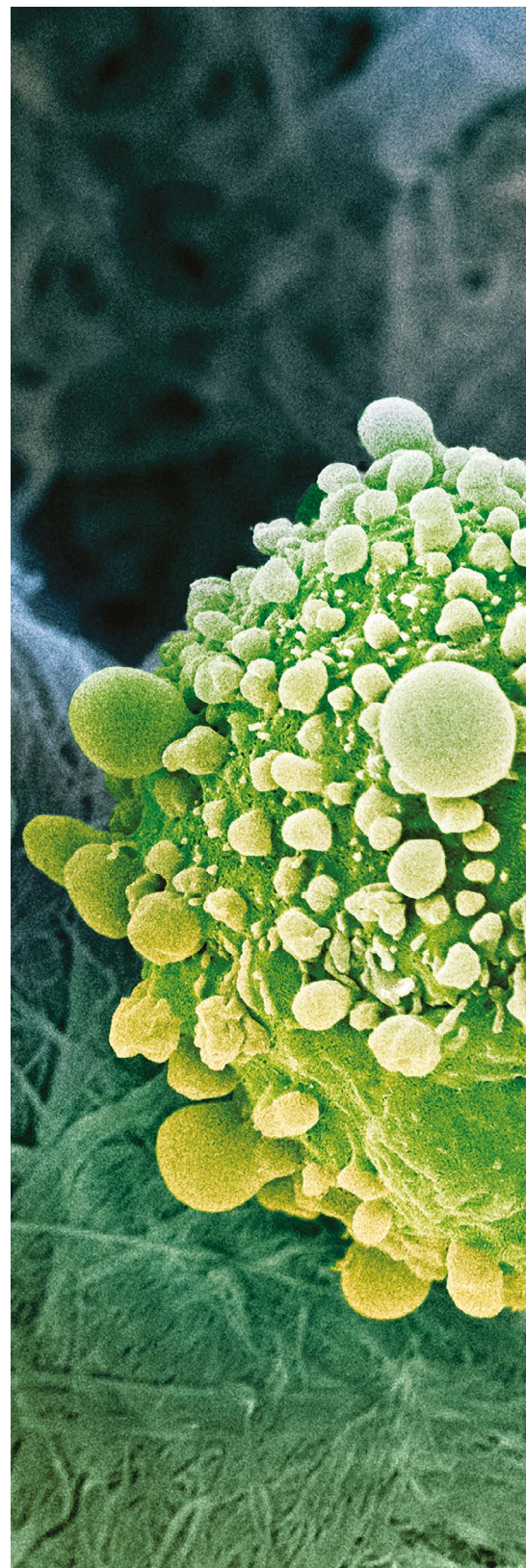
Nonetheless, less than one-quarter of people with the most common cancers benefit from precision medicine today^{1,2}. For most patients, there's a long way to go before clinicians will be able to predict drug activity and mechanism on the basis of a mutation (or other molecular marker) in an individual tumour. In 2018, an estimated 10 million people worldwide died from cancer. Too many endured invasive, toxic treatments because

physicians had no way to know what would work.

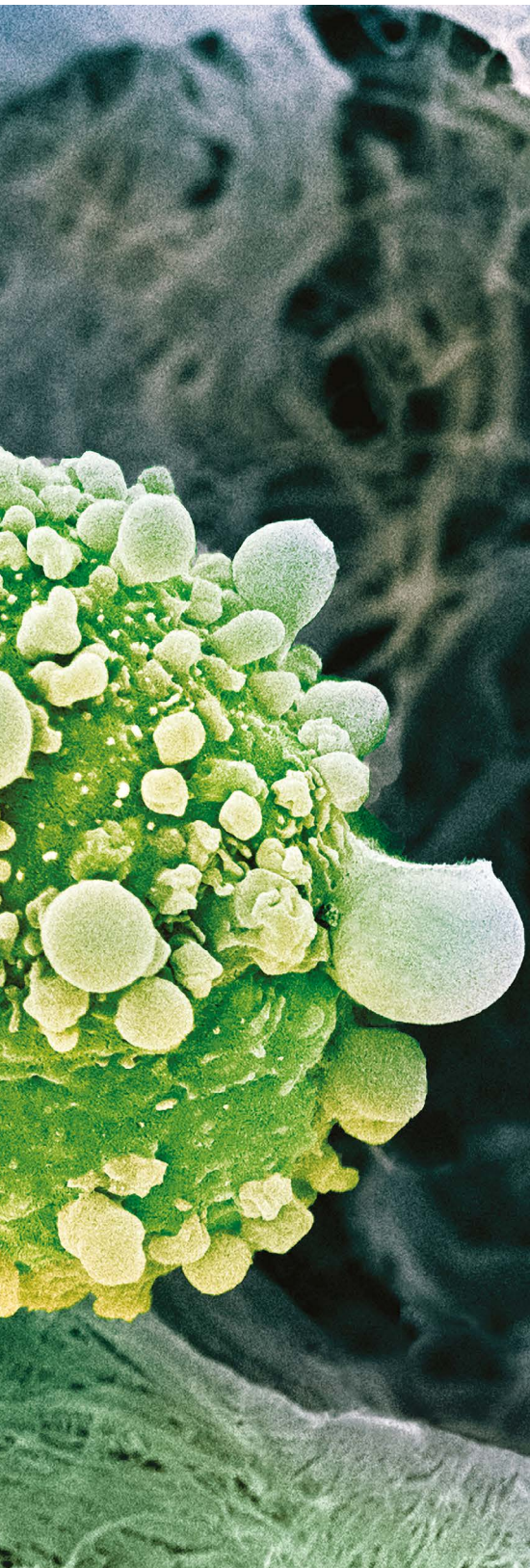
Improving the odds requires another ambitious project, one we call the Cancer Dependency Map. This would systematically map cancer vulnerabilities by perturbing genes and proteins across many cancer types as well as across clinical stages and settings. The Cancer Dependency Map would collect different data and ask different questions from sequencing projects by looking at the experimental effects of perturbations (see 'Big projects, big insights').

Our goal is audacious – some might even say naive. The aim is to evaluate every gene and drug perturbation in every possible type of cancer in laboratory experiments, and to make the data accessible to researchers and machine-learning experts worldwide. To put some ballpark numbers on this ambition, we think it will be necessary to perturb 20,000 genes and assess the activity of 10,000 drugs and drug candidates in 20,000 cancer models, and measure changes in viability, morphology, gene expression and more. Technologies from CRISPR genome editing to informatics now make this possible, given enough resources and researchers to take on the task.

We are investigators at two leading genomics-research institutes – the Wellcome Sanger Institute in Hinxton, UK, and the Broad Institute in Cambridge, Massachusetts. Our organizations have led and supported large international initiatives that produced reference genomics data sets (the Human Genome Project, 1000 Genomes, TCGA, ICGC, gnomAD, the Human Cell Atlas and more). We started a pilot project for the Cancer Dependency Map several years ago^{3–5} to assess the basic feasibility of this effort. Data from these pilot efforts are now accessed daily by 1,500 researchers across more than 100 countries. Tens of highly validated drug targets have been discovered, and clinical trials have already been launched. But the



There are few models to study rare cancers.



scale and complexity of this challenge means that we need more than the Broad and Sanger institutes alone.

Now is the time to follow in the footsteps of other large scientific projects to expand these efforts beyond our two institutes into a communal initiative coordinated and built by many hands.

More than mutations

In oncology, a ‘dependency’ is a gene, protein or other molecular feature that a tumour depends on for growth. It could also be called a vulnerability. The simplest type of dependency is an addiction⁶: a gene that spurs tumour growth when mutated, copied or overexpressed, and which becomes essential for the cells’ continued proliferation. There are therapies that can suppress such gene products directly. Relevant drugs include vemurafenib (which blocks the cell-signalling enzyme B-RAF to treat melanoma and other cancers), osimertinib (which blocks the cell receptor EGFR to treat certain lung cancers) and, more recently, sotorasib (which blocks a mutant form of the K-Ras cell-signalling

“Less than one-quarter of people with the most common cancers benefit from precision medicine today.”

protein)⁷. The addictions involving mutant genes are the type of vulnerability most readily revealed by the tumour sequencing projects.

Another type of vulnerability is more complicated, and hard to find using tumour sequence data: a protein becomes indispensable for a tumour’s growth because of alterations in other genes; this situation is known as synthetic lethality⁸. For example, some people with breast and ovarian cancers have mutations in the gene *BRCA1*. Drugs for these individuals target a protein involved in the DNA-repair mechanism that these cancers need to survive, not the *BRCA1* protein itself. Discovering and understanding more vulnerabilities, and more kinds of them, could transform cancer treatment.

The roots of the Cancer Dependency Map began with the intersection of two lines of research. One aimed to assess the effects of known drugs across hundreds of cancer lines^{5,9–12}. The other sought to disrupt nearly every gene in hundreds of cancer models^{3,4,13}.

Our estimate is that this pilot represents just a small proportion, probably less than 10%, of a ‘complete’ Cancer Dependency Map. Much remains to be done.

Mega mapping

The easiest step is to expand the range of drugs and genetic manipulations used in screens. CRISPR genome-editing tools can knock out genes, turn expression on and off and even manipulate combinations of genes. These perturbations can be performed across many tumour types and cellular contexts (such as in the presence of immune and other non-cancerous cells). Existing collections of drugs and other molecules with specific activity against known proteins can be used to inhibit or even modify proteins’ function. Classes of compounds that have other types of activity, such as promoting protein degradation, should be included in screens as our understanding of them matures.

The most challenging expansion will be in the number of cancer models (cell lines and organoids), to represent all of the diversity in cancer. Our pilot project screened about 1,000 cancer cell lines. These included few of the ‘driver genes’ that individually are found in less than 10% of tumours (but which together constitute the majority of cancer mutations). For common tumour types, populations of non-European descent are under-represented in the bulk of data collected, as are drug-resistant tumours and early stages of disease. (Projects such as TracerX show how individuals’ cancers evolve over time, but not how to treat different stages.) Rare cancers collectively represent 25% of all cancers. Most, including paediatric cancers, are under-represented in or completely absent from existing models.

The range of data that must be collected is substantial. The pilot phase of the dependency map mainly measured proxies for proliferation, so further measurements would be extremely informative. For example, identifying drug targets that promote senescence, a condition in which cells live but no longer divide, would open routes to new forms of therapeutics. Measuring gene expression in response to drugs could reveal which ones might be combined into more-effective therapies. Single-cell measurements, such as cell imaging¹⁴ or sequencing to determine RNA levels at various time points (RNA-seq), could identify differences in the responses of tumour cell populations. This information could be used to target drug-tolerant cells.

Some of the essential work for comprehensive screening is under way. Curated

STEVE GSCHMEISSNER/SPL

Comment

collections of thousands of drug-like and tool compounds have been assembled by the US National Institutes of Health (NIH), and chemical probes to new targets have been developed by the Structural Genomics Consortium. Efforts such as the Human Cancer Models Initiative are creating biobanks of new cancer cell lines, organoids and other models that are genetically and clinically annotated. Such crucial efforts will not produce a Cancer Dependency Map, however. That will take multiple, large-scale data generation and analysis centres, as well as resources and coordination.

Data standards

Data from the Cancer Dependency Map must be accessible and flow easily into analysis pipelines for academic and pharmaceutical scientists, artificial-intelligence experts and clinicians.

Data sets generated at different sites using varying technologies will be difficult or impossible to aggregate without careful advance planning. Care must be taken so that the data will be maximally useful, and can be readily integrated with databases such as the US National Cancer Institute's Genomics Data Commons, cBioPortal and the Catalogue Of Somatic Mutations In Cancer (COSMIC), as well as new dedicated data portals. For the past 20 years, various scientific communities have crafted 'minimal information' standards to enable interpretation and replication of experiments. The best known of these, MIAME (minimum information about a microarray experiment), was set up so that researchers could effectively compare gene-expression experiments in different laboratories. The Cancer Dependency Map will need something similar to support annotation, integration, reproducibility and computation.

Money and collaboration

This project is ambitious but possible. Past examples of large, successful, global projects include the ICGC and the Human Cell Atlas¹⁵, a project to categorize gene expression, lineage and the location of hundreds of cell types. The NIH and Wellcome Trust were major funders of these, and the Chan Zuckerberg Initiative also funded the Human Cell Atlas. At least half a dozen countries came together to support the Human Genome Project.

The first step forward is to assemble scientific thought leaders as well as funders to clearly define the scope and scale of data needed. The next step is to establish working groups around standards, technology, data access and cancer types. This follows the example of the Human Cell Atlas, in which research is led by independent laboratories with coordination at the level of the organ. Its working groups set data standards that enable sharing and integration.

Similar to the Human Cell Atlas, data

BIG PROJECTS, BIG INSIGHTS

The Cancer Genome Atlas revealed cancer mutations. A new project to find weak points in many cancer types using systematic experiments can suggest how to exploit these vulnerabilities for treatments.

Project	The Cancer Genome Atlas Focus on comparing DNA in healthy tissues and tumours.	The Cancer Dependency Map Focus on laboratory experiments to find tumour vulnerabilities.
Input	Tumours and healthy tissues from thousands of individuals across 33 cancer types.	Experiments to knock out genes and screen drug libraries in thousands of cancer models, representing diverse cancer types at different clinical stages and from people of various ethnicities.
Data	<ul style="list-style-type: none">• DNA sequence• Epigenetic marks• Gene expression• Protein levels and modifications.	After perturbation: <ul style="list-style-type: none">• Cell growth• Cell death• Gene expression• Cellular morphology.
Insights	Identified changes (such as mutations or deletions) specific to certain cancer types. Found genes implicated in many cancers, but did not directly explore gene functions.	How manipulating a given target in a given tumour type might affect tumour growth, including when the best targets are not in the altered genes themselves.

and technologies arising from the Cancer Dependency Map will provide huge resources for science, as will the networking and technology development spurred by the umbrella project. Just gearing up to collect measurements for the Cancer Dependency Map could drive innovation: in automation, miniaturization and in new co-culturing techniques that mimic physiological conditions.

It is true that the cell-based experiments we propose will miss many complexities of cancer biology. Still, most drugs available today arose from *in vitro* screens. As the technologies mature, we expect that experiments will extend towards new types of target, including those involving the tumour microenvironment or immune system. Early concerns about reproducibility¹⁶ seem to now be largely overcome^{12,13}. The precision of CRISPR technologies and the statistical power that comes from large-scale experiments strongly benefit reproducible experimentation.

We estimate that a thorough Cancer Dependency Map project will require an annual commitment of US\$30 million to \$50 million over a decade. (The TCGA pilot launched by the NIH cost \$100 million during its first 3 years, and continued for 12 years.) For comparison, the US National Cancer Institute's 2020 budget was \$6.4 billion, and the 2019–20 research budget for the funder Cancer Research UK was £511 million (US\$694 million). Industry, government agencies and philanthropists should contribute.

One in six deaths worldwide are due to cancer. Although more targets and therapeutic approaches are published every day, systematic investigation will enable more and faster discoveries. By testing thousands of genetic and chemical perturbations in thousands of cancer models in rigorous, coordinated experiments, the Cancer Dependency Map can reveal the cellular wiring underlying a breadth

of tumours, with data accessible to individual researchers' queries and computational approaches. In the same way that large-scale, systematic genome sequencing transformed our understanding of cancer, the Cancer Dependency Map will enable researchers to explore questions and identify routes to therapies that are currently unimaginable.

The authors

Jesse S. Boehm, Todd R. Golub, William C. Hahn, James M. McFarland and Francisca Vazquez are leaders of the Cancer Dependency Map project at the Broad Institute in Cambridge, Massachusetts, USA; T.R.G. and W.C.H. are also cancer biologists at the Dana-Farber Cancer Research Institute in Boston, Massachusetts. **Mathew J. Garnett, David J. Adams, Hayley E. Francies, Francesco Iorio and Leopold Parts** are leaders of the Cancer Dependency Map project at the Wellcome Sanger Institute in Hinxton, UK. F.I. is also a computational cancer biologist at the Human Technopole in Milan, Italy. e-mails: boehm@broadinstitute.org; mathew.garnett@sanger.ac.uk

1. Zehir, A. *et al. Nature Med.* **23**, 703–713 (2017).
2. Stockley, T. L. *et al. Genome Med.* **8**, 109 (2016).
3. Tsherniak, A. *et al. Cell* **170**, 564–576 (2017).
4. Behan, F. M. *et al. Nature* **568**, 511–516 (2019).
5. Corsello, S. M. *et al. Nature Cancer* **1**, 235–248 (2020).
6. Weinstein, I. B. & Joe, A. *Cancer Res.* **68**, 3077–3080 (2008).
7. Hong, D. S. *et al. N. Engl. J. Med.* **383**, 1207–1217 (2020).
8. Luo, J., Solimini, N. L. & Elledge, S. J. *Cell* **136**, 823–837 (2009).
9. Garnett, M. J. *et al. Nature* **483**, 570–575 (2012).
10. Iorio, F. *et al. Cell* **166**, 740–754 (2016).
11. Barretina, J. *et al. Nature* **483**, 603–607 (2012).
12. Haverty, P. M. *et al. Nature* **533**, 333–337 (2016).
13. Dempster, J. M. *et al. Nature Commun.* **10**, 5817 (2019).
14. Feldman, D. *et al. Cell* **179**, 787–799 (2019).
15. Rozenblatt-Rosen, O., Stubbington, M. J. T., Regev, A. & Teichmann, S. A. *Nature* **550**, 451–453 (2017).
16. Haibe-Kains, B. *et al. Nature* **504**, 389–393 (2013).