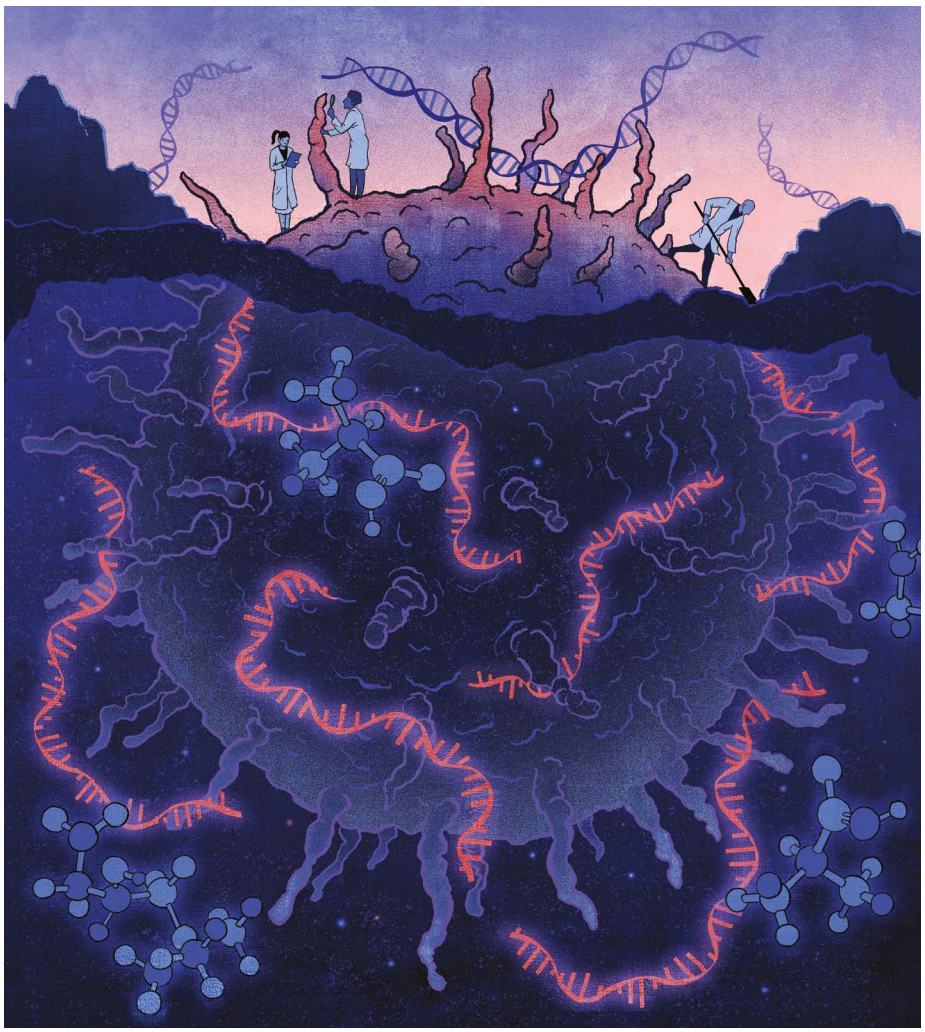
Beyond the genome

To bring precision treatments to more people, researchers are looking past the genetic blueprint to the dynamic landscape of RNA and proteins. **By Simon Makin**



ANTOINE DORE

Genomics has revolutionized cancer research. Conventional classifications of disease, in terms of which organs and tissues it affects, are being divided into subtypes defined by the specific mutations that drive the disease (see page S16). Some argue, however, that the impact on cancer care has not lived up to expectations. "Only about 5–10% of cancer patients derive any benefit from targeted therapy using genetics, and almost all of them eventually relapse," says systems biologist Andrea Califano at Columbia University in New York City. "The number that are actually cured is extremely small."

Developing a genetically targeted therapy is no easy task. It can be tricky to identify which genetic mutations are driving the cancer and which are passengers – those that are statistically linked, but that do not cause cancer. And although developers of targeted therapies focus mainly on mutations to a subset of genes called oncogenes, there is more to malignancy. "Most genetic alterations in cancer are not oncogenes, they're tumour-suppressor gene alterations," says Bert Vogelstein, a cancer researcher at Johns Hopkins University in Baltimore, Maryland. These mutations inactivate genes that usually help to guard against cancer, such as those responsible for repairing DNA damage or controlling programmed cell death. And because the proteins encoded by these genes are often not produced in the cancer cells, they are difficult to target. "If the protein isn't there, they're impossible to directly target with any drug," says Vogelstein.

In addition, cancers often relapse because tumours contain a mix of cells with different mutations. "It's clear cancer is composed of multiple clones within single tumours," says biochemist Tamar Geiger at Tel Aviv University in Israel. If even a few cells are resistant to a treatment, they "take over the tumour and you get resistance and relapse", Geiger says.

Such limitations are forcing cancer researchers to look beyond the blueprint that is the genetic code. They are, for example, exploring how epigenetic mechanisms – which modify gene function without changing the underlying code and which are influenced by developmental and environmental signals – can contribute to tumour formation. Advances in sequencing technology have allowed researchers to take snapshots of this dynamic landscape by measuring gene expression using RNA sequencing. And advances in recent years are helping researchers to study the product of genes – proteins – to construct an even fuller picture of the cellular mechanisms at work in cancer.

The Cancer Genome Atlas (TCGA) programme – launched by the US National Cancer Institute (NCI) and the National Human Genome Research Institute in 2006 – has so far characterized more than 20,000 samples spanning 33 cancer types. The types of data analysed include the complete set of epigenetic modifications, known as the epigenome; RNA transcripts, known as the transcriptome; and proteins, known as the proteome. The wealth of information provided by each 'omic' layer is helping researchers to better classify a person's cancer and to predict their response to treatment, and could lead to new lines of attack. For example, although mutations in tumour-suppressor genes are difficult to tackle directly, the events they set in motion could be targets. "If you understand how the tumour-suppressor gene works, you can figure out what's activated downstream," says Vogelstein. Analysis of networks of interaction

is revealing mechanisms that might lead to treatments that are effective against many mutations simultaneously, and protein analyses are helping to explain why some people fail to respond to certain therapies. But the most powerful studies are those that integrate multiple layers of analysis to build a complete picture of cancer biology. The hope is that these approaches will equip researchers with enough understanding of cancer's diversity and dynamism to allow them to tackle more cancers, in more people, and with treatments that do more than extend a person's life.

Links in a chain

To better understand how genes and proteins interact to produce a cancer cell, Califano is embracing transcriptomics and computational modelling. "Genetics represents the space of what could be; RNA is a snapshot of what is," he says. "It gives you a more complete picture of the regime the cell is operating in."

His idea is something he calls tumour checkpoints. Even cancers of the same genetic subtype can be highly diverse – they might have just a single mutation in common. If different combinations of mutations give rise to essentially the same disease, this suggests that they converge on a limited number of proteins. With this in mind, Califano developed an algorithm that can infer gene activity from noisy RNA data, and used it to identify proteins that channel the effects of multiple mutations. These pivotal players can be enzymes that influence transcription through epigenetic mechanisms, or transcription factors that influence gene expression more directly. "These are the proteins that run the operation room of the cancer cell," says Califano. "We call them master regulators." In an analysis of around 10,000 TGCA samples published on the preprint sever bioRxiv, Califano and his colleagues identified 407 master regulators that convey the effects of nearly all mutations implicated in the cancer samples[1]. Because master regulators are rarely mutated, genomics is not a sure-fire way to identify them.

Blocking a single master regulator could arrest aberrant cellular activity resulting from many mutations at once. "When you find the chink in the armour, the entire checkpoint collapses," Califano explains. The approach has already borne fruit. For instance, in 2015, Califano and his colleagues looked at people with breast tumours that carried mutations in the gene *HER2*, but who were resistant to the antibody drug that targets these mutations, trastuzumab (Herceptin)[2]. They found that HER2-positive cells secrete high levels of a cytokine called IL-6, which in turn activates a transcription factor known as STAT3. This process ultimately promotes the production of calprotectin – a protein complex involved in proliferation and resistance pathways. STAT3, it seems, is a master regulator that is responsible for trastuzumab resistance in breast cancer. A drug that inhibits this pathway, ruxolitinib, which is already approved for blood and bone marrow cancers, is now in a phase II trial for HER2-positive breast cancer in combination with trastuzumab.

As well as identifying master regulators in cancers, Califano and his colleagues have developed algorithms that can suggest treatments to shut down overactive master regulators and boost underactive ones. The approach is being put to the test in a clinical trial at Columbia that aims to treat 3,000 people over the next 3 years. Alongside genomic analyses, clinicians will take into account readouts from Califano's algorithms before recommending treatments. "If you have mutations shown to respond well to therapy, we should use them," says Califano. But if that isn't the case, or people relapse or fail to respond, he adds, "you really have no other option, and that's when we use RNA".

Biochemical effectors

Transcriptomics provides researchers with a more dynamic view of a cancer cell. But RNA is mainly an intermediary for biology's most fundamental players: proteins. "If we want to understand the function of a cell, the way to do it is looking at the proteins," says Geiger.

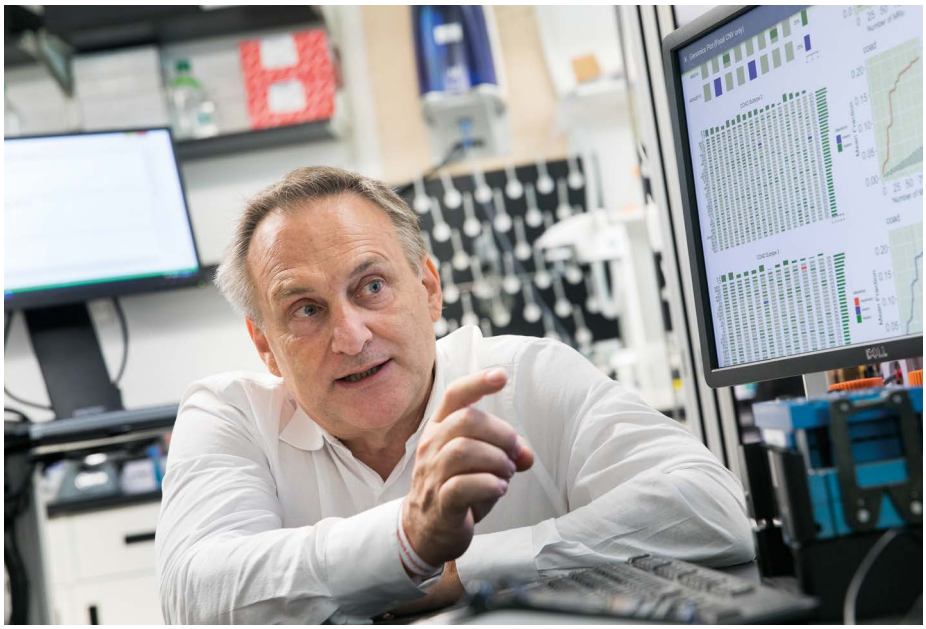
Until recently, measuring protein levels relied on techniques that required researchers to know what they were looking for before they started. But advances in mass spectrometry over the past decade have allowed for genome-wide exploration of the proteome. Geiger worked with proteomics pioneer Matthias Mann at the Max Planck Institute for Biochemistry in Martinsried, Germany, before returning to Israel to run her own lab. In 2016, the pair co-led a study demonstrating that a set of 19 proteins could distinguish between HER2-positive, oestrogen-receptor-positive and triple-negative (negative for oestrogen and progesterone receptors and excess HER2) breast-cancer subtypes[3]. And in 2018, Geiger identified a proteomic subtype of oestrogen-receptor-positive breast cancer not seen at genomic or transcriptomic levels[4]. Such protein-based groupings are often linked with different outcomes. "We see associations with survival that we don't see with RNA," Geiger says. "Clearly the protein level adds something."

One of the most promising applications of proteomics is its use in understanding and enhancing response to immunotherapy. This approach, which largely eschews genetic targeting in favour of boosting the body's natural defences against tumour cells, is revolutionizing cancer care. "When it works, it's like a cure," says biologist Karin Rodland of the Pacific Northwest National Laboratory in Richland, Washington. But it works well only for a few cancers, and even then a substantial proportion of people don't have a response. "A big question is how do you predict who's going to respond?" says Rodland. "And how do you improve the response rate for those who aren't?"

Two of the cancers that immunotherapy works best in – melanoma and lung cancer – involve tumour cells that carry more genetic mutations than other cancers. The prevailing theory is, therefore, that mutational burden determines the efficacy of immunotherapy. "If you've got lots of mutated genes, you're likely to have lots of foreign antigens," Rodland explains. "But that hasn't turned out to be very highly predictive." Instead, she and others are looking at proteins to more fully understand the biological mechanisms that determine response to immunotherapy. As part of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) – an effort launched by the US National Cancer Institute – Rodland has shown that proteomics reveals information about the extent to which cancer cells provoke an immune response.

The hypothesis that proteins are more predictive of immunotherapy response than genetic mutations has not yet been proved clinically, Rodland says, but some striking research findings have emerged. A 2019 study led by Geiger looked at the response of people with melanoma to two types of immunotherapy[5]. Her team found differences in the proteins involved in cancer-cell metabolism that predicted responses to both therapies. "We started with the aim of identifying predictive signatures to spare people treatment who aren't going to respond," says Geiger, "but saw we'd found a metabolic pathway associated with higher response." The protein differences seem to affect the presentation of antigens on cancer cells, and thus the ability of immune-system T cells to recognize the cells. "These metabolic aspects weren't seen on genomic or transcriptomic levels," says Geiger. The researchers confirmed the pathway's importance by inactivating the genes

"If we want to understand the function of a cell, the way to do it is looking at the proteins."



COLUMBIA UNIV./TIMOTHY LEE PHOTOGRAPHERS

Andrea Califano has identified 'master regulators' of cancer cells.

involved; this reduced T-cell destruction of melanoma cells. Activating this pathway might mean, therefore, that immunotherapy would work for people who would otherwise not respond. "We're testing this in melanoma now," Geiger says.

Better together

Each additional layer of information is providing fresh insights into cancer, but the full potential lies in integrating the layers. The CPTAC has carried out proteogenomic analyses that link mutational patterns to their consequences for protein output for several cancers. In a multiomic study of colorectal cancer[6], the consortium found that copy-number alterations — a type of mutation in which chunks of DNA are repeated or deleted — had substantial effects on the levels of RNA corresponding to genes in the alterations. However, this was the case only for a handful of corresponding proteins.

Researchers already knew that RNA levels do not often predict protein levels, but an important implication for cancer could be emerging. A 2018 study combining transcriptome and proteome analyses of breast-cancer tissue found levels of RNA corresponded more closely with levels of proteins in tumours than in healthy tissue — suggesting that the degree to which RNA and proteins change in lockstep might provide a crucial window into cancer biology[7].

The most highly correlated RNA–protein pairs were involved in known disease processes, and higher correlations were associated with more aggressive disease and decreased survival. RNA–protein correlations are higher when they are part of molecular pathways that cancer cells prioritize, says Sajib Chakraborty, a computational biologist at the University of Dhaka in Bangladesh. When a cancer cell needs to rewire metabolism, for example, the pathways required to make it happen come under tight regulatory control. As a result, high correlation between RNA and protein could be a sign that the cancer depends on a particular pathway. "We can identify which pathways are top priority for cancer cells," Chakraborty says. Alongside colleagues at the University of Freiburg in Germany, he is now investigating these dependencies using TCGA and CPTAC data. As well as revealing crucial cancer pathways, the team hopes to identify processes that might prevent the spread of cancer or clues to a person's likely response to treatment.

Using databases of genes affected by cancer drugs, Chakraborty is investigating how RNA–protein correlations for drug targets change over time. For instance, people given 5-fluorouracil, a common drug for colorectal cancer, often stop responding after a period of time. Chakraborty says that the team has early data suggesting that the drug's targets decline in importance as the cancer progresses. "In the early cancer stage, the target genes of 5-fluorouracil are in a tight correlation, but in the late stage their correlation becomes noisy," he says. This suggests that what makes a good treatment target might change over time.

Even proteins are not the final layer of the multiomics story. Proteins can themselves be modified after translation. These post-translational modifications take several forms, but the best understood is phosphorylation, which acts as an on–off switch for proteins. In a 2019 colorectal cancer study, the CPTAC included phosphoproteomics in its analysis for the first time[8]. It yielded immediate results, explaining an apparent paradox in the findings. Both the genomic and the proteomic analysis showed that the product of the gene *RB1* was elevated in colorectal cancer — a strange discovery, because *RB1* is involved in tumour suppression and is usually deactivated in cancer. *RB1* acts as a suppressor because its protein, Rb, inhibits a transcription factor involved in cell proliferation. Phosphorylation of Rb blocks this inhibition. "Protein abundance doesn't necessarily define protein function," says Rodland. "You need post-translational modifications to predict function."

Currently, the main limitation of proteomics is resolution. Tissue samples containing thousands of cells must be prepared for mass spectrometry instruments. Unlike RNA sequencing, the technology cannot characterize differences at the level of individual cells. This will be important for understanding the diversity in individual cancers, and targeting every cell in a tumour rather than just the most dominant type. "This complexity is critical to developing treatments that are actually going to cure patients, not just delay relapse," Geiger says.

She is optimistic that these technological barriers will be overcome in the next couple of years. Meanwhile, scientists are advancing other multiomics oncology approaches. For instance, metabolomics — the analysis of metabolic products — is already being incorporated into liquid-biopsy techniques for the early detection of cancer. However, to varying degrees, all these techniques require specialized equipment and expertise, creating a bottleneck. Bringing them into the clinic will require cheaper, more robust and reproducible techniques. "I'm hopeful that in future there will be user-friendly software, which, just by clicking, the clinician can understand what's going on in the patient's sample," says Chakraborty. "Because now, they're dependent on guys like us: computational biologists."

Simon Makin is a science writer based in London.

1. Paull, E. O. *et al.* Preprint at bioRxiv https://doi.org/10.1101/758268 (2019).
2. Rodriguez-Barrueco, R. *et al. Genes Dev.* **29**, 1631–1648 (2015).
3. Tyanova, S. *et al. Nature Commun.* **7**, 10259 (2016).
4. Yanovich, G. *et al. Cancer Res.* **78**, 6001–6010 (2018).
5. Harel, M. *et al. Cell* **179**, 236–250(2019).
6. Zhang, B. *et al. Nature* **513**, 382–387 (2014).
7. Tang, W. *et al. Genome Med.* **10**, 94 (2018).
8. Vasaiker, S. *et al. Cell* **177**, 1035–1049 (2019).