



CDC/SPL

A scientist at the US Centers for Disease Control and Prevention loads an instrument for whole-genome sequencing.

# NAVIGATING AN INCREASINGLY CROWDED SEQUENCING LANDSCAPE

With a dizzying range of technologies available, laboratories must weigh up their options to find the best platform for their projects. **By Michael Eisenstein**

**T**here's an old saying in the field of technology: "Nobody ever got fired for buying IBM" – a reference to the company's once-ubiquitous computers. Replace IBM with Illumina, a biotechnology company in San Diego, California, and the same could be said of DNA sequencing today.

Keith Robison, a computational biologist at Ginkgo Bioworks in Boston, Massachusetts, who writes about sequencing technologies on a blog called Omics! Omics!, says that for most laboratories, Illumina "is the really safe bet out there". However, IBM's days of computer market dominance are well in the past, and Illumina now faces multiple competitors who are looking to challenge – and perhaps unseat – the current giant of the sequencing marketplace.

Researchers, naturally, are paying attention. Pedro Oliveira heads the DNA-sequencing lab

at the French National Sequencing Center, also known as Genoscope, in Évry. The lab recently partnered with several big European research projects, including the European Reference Genome Atlas, which will bring in an expected workload of four genomes per week. One of Genoscope's priorities will be to increase its arsenal of Illumina instruments – but that won't be the limit of its shopping list, and Oliveira has a broad range of platforms to consider.

Some instruments use complementary approaches that generate long-sequence reads spanning thousands of nucleotides, in contrast to Illumina's 'short reads', which are typically in the 100- to 200-base range. But the past year has also seen the launch of nearly half a dozen competing short-read systems, each touting their own advantages in terms of quality, efficiency and above all, cost. "These are exciting moments that we're living in," says

Oliveira, "because this is the beginning of cheap sequencing." But the range of choices can be intimidating and confusing, given that most scientists are still waiting to see the actual data, and to assess how well these platforms match their projects.

## A safe bet

Illumina entered the sequencing market with the acquisition of a company called Solexa in 2007. Solexa's 'sequencing by synthesis' (SBS) technology exploits the same machinery that manufactures DNA in living cells. A template DNA strand is read by a DNA polymerase enzyme, which sequentially tacks on nucleotides that complement the template strand.

Each of the four DNA building blocks – A, T, G and C – is coupled to a specific fluorescent colour and a 'terminator' chemical group that pauses further DNA synthesis. Sensitive optics identify the added nucleotide from the

resulting fluorescence, after which the tag and terminator are removed and the cycle repeats. The whole process occurs in a wafer-like ‘flow cell’, in which vast numbers of DNA targets are imaged simultaneously, generating millions or even billions of short reads per run.

That approach has been staggeringly successful. By one estimate, more than 90% of the world’s sequencing data as of 2022 were generated on Illumina machines (see [go.nature.com/3abj7ng](https://go.nature.com/3abj7ng)). Dozens of would-be competitors have emerged to challenge Illumina over the years, but most have fallen by the wayside – many of them memorialized in the ‘NGS Necropolis’ (see [go.nature.com/3xwvmkt](https://go.nature.com/3xwvmkt)). Catharine Aquino, who oversees short-read sequencing at the Functional Genomics Center Zurich in Switzerland, attributes that success to technical expertise. “It’s just that the other companies were not very reliable in terms of library prep or sequencing,” she says.

Illumina’s portfolio includes both compact benchtop systems for rapid analysis of small numbers of samples, such as the iSeq Sequencing System that costs US\$20,000, and the larger high-end NovaSeq 6000, which costs nearly \$1 million but can churn out up to 6 trillion bases (6 terabases) of sequence – roughly 2,000 times the length of one of the human genome – every 2 days. Illumina’s new NovaSeq X family of production-scale sequencers use a redesigned flow cell that can accommodate a much greater density of sequencing reactions along with a retooled SBS chemistry and upgraded optics, according to Illumina’s chief technology officer, Alex Aravanis. The company reports that its new systems, which began shipping this year, can generate up to three times as much data per run as the previous generation NovaSeq 6000, lowering the cost of sequencing to just \$200 per human genome.

### An array of alternatives

Beyond human genome assembly and mutational analysis, new applications have fuelled demand for more and better short-read data at a lower cost. These include everything from epigenetics to chromosomal conformation to proteomics. Aquino estimates that 60% of her facility’s work now involves single-cell RNA-seq, a sequencing-hungry technique that profiles the gene expression of thousands or millions of individual cells. To fill that surge in demand, both start-ups and established companies have entered the ring.

One established player, MGI Tech, a Shenzhen-based spin-off company from the Chinese genomics titan BGI, offers distinctive twists on an Illumina-like SBS approach. Both MGI and Illumina use a biochemical process to generate multiple copies of every strand of template DNA on the flow-cell surface, thus boosting the fluorescent signal, but MGI’s DNBSQ platforms use a lower-cost – albeit more labour-intensive – method that converts

templates into arrays of ‘DNA nanoballs’. “The data quality is really good, and it can be much more cost-effective than Illumina,” says Ioannis Ragoussis, head of genome sciences at the McGill Genome Centre in Montreal, Canada, who has used DNBSQ instruments in his own facility.

Of the newcomers, the G4 benchtop system from Singular Genomics in San Diego is probably most like Illumina’s. But the G4 also features a flow-cell design that can make it easier to run multiple sequencing experiments simultaneously. “It’s really targeted towards these smaller, more flexible projects,” says Stephanie Pond, vice-president of emerging technologies at the Translational Genomics Research Institute in Phoenix, Arizona, which beta-tested the G4.

Ultima Genomics’ flow cell is even more distinctive. Rather than using a sealed cartridge containing complex channels to coordinate the flow of reagents, Ultima – based in Newark, California – applies sequencing reagents to the exposed surface of a spinning disc. The resulting centrifugal force distributes these materials evenly across the disc’s surface, reducing both the complexity of the

**“Cost is still one of the biggest drivers, because people only get so much money from grants.”**

flow-cell design and the amount of reagents required, and thus lowering the cost of each run. Ultima also cuts costs by using a mixture of labelled and unlabelled nucleotides rather than just the costlier labelled molecules<sup>1</sup>. In one study, early-access users at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, documented generally comparable performance to Illumina in single-cell gene-expression experiments<sup>2</sup>.

Finally, there are the chemistries developed by Element Biosciences, based in San Diego, and by Pacific Biosciences (PacBio) in Menlo Park, California, for new short-read instruments. Both rely on two-stage alternatives to the standard SBS approaches, in which fluorescently labelled nucleotides are not permanently incorporated into the newly synthesized DNA but rather bind transiently to the growing strand. Once they are imaged, they are then washed away and replaced by unlabelled nucleotides.

This results in a more natural DNA synthesis process while also allowing for careful optimization of the labelling step, and both Element and PacBio – a company already well-known for its sophisticated long-read systems – highlight the accuracy of their approaches. “We’ve been seeing extremely high-quality data,” says genomics researcher Christopher Mason at

Weill Cornell Medicine in New York City, who has used Element’s AVITI system to profile the effects of space flight on human physiology.

### Weighing up pros and cons

Sequencers fall broadly into two categories: production-scale instruments including Illumina’s NovaSeq, and smaller benchtop instruments such as Illumina’s NextSeq. For now, only Illumina and MGI operate across the full spectrum; other short-read companies target specific levels of throughput.

Production-scale instruments are massive and expensive, but such throughput is essential for many large-scale genomics or single-cell RNA-seq studies, and such instruments tend to form the backbone of core sequencing facilities. Stacey Gabriel, chief genomics officer at the Broad Institute says that almost all of the sequencing done at her centre, one of the world’s leading genomics facilities, uses such instruments. “We have 32 NovaSeqs, and we run them very hard,” she says, adding that her team will be augmenting this capacity with new NovaSeq X instruments.

Ultima also operates in this arena with its UG 100, but aims to counter the high cost of its hardware with cheaper sequencing costs. The company claims that it has the potential to deliver complete human genome sequences for \$100 – half the price of the NovaSeq X. The Broad Institute was one of the UG 100’s first users, and Gabriel says that although the technology is still maturing, she sees clear opportunities to incorporate it into their workflow for whole-genome analysis and high-throughput assays such as single-cell transcriptomics.

When it comes to purchasing decisions, equipment and reagents are only part of the calculation, and publicly announced per-genome prices don’t account for labour, maintenance and other support costs. Facilities can expect to pay 10% of an instrument’s base cost every year for service contracts, Ragoussis says, which can put even mid-range benchtop instruments out of reach for many labs. Most importantly, production-scale instruments are only more cost-effective relative to benchtop instruments when they are run at full capacity. “There are a lot of projects that just aren’t big enough, or pilot-scale projects where it’s really hard to ‘feed the beast,’” says Pond. This can also be an issue for labs dealing with multiple experiments that cannot be run simultaneously in a single flow cell.

Benchtop machines might be a better fit here, and this is the realm in which PacBio, Singular and Element currently compete. Such instruments generally cost between \$200,000 and \$400,000, and there is robust competition to deliver the most data at the lowest price-per-gigabase. “Cost is still one of the biggest drivers, because at the end of the day people only get so much money from grants,” says Mason. MGI has been using this pressure



## Work / Technology & tools

point to drive adoption of its products, Mason adds, even by offering instruments for free to some labs that are willing to spend a set amount on regular orders of reagents.

Quality is another crucial consideration, and here, too, Illumina has set a high bar. For most reads, Illumina's systems will 'call' the correct base 999 times out of 1,000 – a standard of accuracy called Q30 – and its newest-generation 'XLEAP-SBS' chemistry reportedly improves this accuracy by three-fold. PacBio claims that its new Onso instrument – which is still in beta testing – has error rates of one in 10,000 bases or lower (Q40), and Mason says his test runs with validated genomic samples have borne this out. "At the beginning of the read it's even better," he says, reporting quality nearly an order of magnitude better than Q40. Mason thinks further optimization of the computational toolbox for analysing Onso-generated data could lead to even better performance.

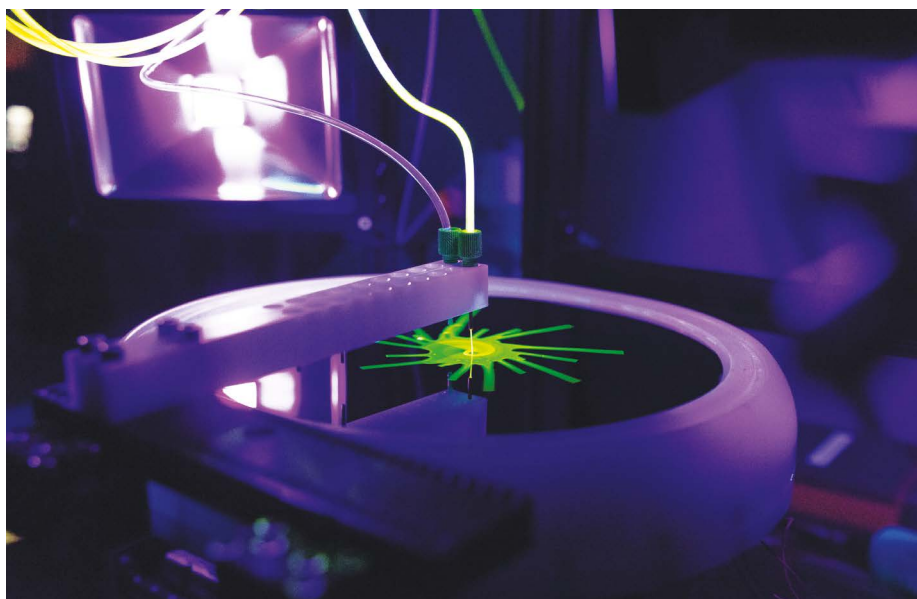
A 2022 preprint<sup>3</sup> from scientists at Element Biosciences also highlights the ability to achieve Q40 quality for most bases in a human genome sequenced with the AVITI instrument, which started shipping in June last year. The company also has a price edge over PacBio, matching Illumina's \$200 cost-per-human-genome and undercutting that of Onso by roughly seven-fold. In principle, higher quality reads reduce the amount of sequencing required for routine genomic studies and could provide a decisive advantage for applications such as the analysis of circulating tumour-derived DNA in 'liquid biopsy' assays. "There's relatively few copies in the sea of normal DNA," explains Gabriel, "so you've got to sequence very deeply."

Another consideration when choosing a sequencing platform is compatibility with existing workflows. For example, Element's workflow is largely consistent with standard Illumina processes, whereas Ultima and MGI require extra processing steps that can introduce speed bumps into existing pipelines. "It's not insurmountable – it just adds more time and labour," says Mason. Further automation might also be required to streamline the process.

Stability and reliability are also essential, because even brief downtime can disrupt lab operations. Illumina generally has an excellent reputation on this front, says Aquino. "Sometimes even before we know something is wrong, our engineer is there already," she says. "It will take all these companies a few more years to build up the support system and this array of experience."

### Going long

Not every sequencing application maps well onto short-read technologies. Therefore, companies such as PacBio and Oxford Nanopore Technologies (ONT) in the United Kingdom have worked to evolve their long-read technologies as well.



Ultima Genomics' instruments run sequencing reactions on the surface of a spinning disc.

Both companies offer systems that directly analyse individual DNA molecules spanning tens or even hundreds of thousands of nucleotides. For PacBio, this entails feeding strands of template DNA into polymerase enzymes that are tethered to a solid surface and then using sophisticated optics to detect the addition of individual labelled nucleotides as the DNA synthesis proceeds. ONT systems determine nucleotide sequences on the basis of the distinctive changes in electrical current that occur as DNA strands transit through tiny protein pores. Collectively, these systems provide insights that would be difficult or impossible to obtain with short-read systems, including large structural variations in chromosomal DNA, mRNA transcript structure and complete microbial genomes. Both systems can also directly identify and map epigenetic modifications.

PacBio offers some of the highest accuracy instruments on the market, thanks to a process called 'HiFi' in which the devices read the same segment of DNA over and over again, ironing out random errors along the way. However, they have historically been held back by high costs and low throughput. "A hundred samples in PacBio took a year, while 100 samples in Illumina took maybe two days," says Aquino. But the company's new Revio instrument, which costs \$779,000 and is scheduled to begin shipping this year, changes the equation. With the capacity to achieve 15-fold greater throughput than current-generation systems, PacBio reports that the Revio can produce a high-quality human genome for just \$1,000.

ONT offers a uniquely versatile and portable platform that can be just as easily applied to short-read applications as it can to ultra-long reads. Researchers regularly use ONT systems in the field, and Mason has even sent them to the International Space Station. "We can see applications in many remote areas," he says.

ONT also offers the lowest-cost sequencing hardware on the market, including the \$1,000 MinION, which can be run off of a standard laptop or, in newer versions, a tablet.

By contrast, ONT's high-performance PromethION can sequence up to 14 terabases in 3 days, and uses an unusual business model in which most of the upfront costs are associated with the purchase of consumables needed to run sequencing experiments. "You get an instrument where the price is associated with how many flow cells you want to use without you having to buy it," says Ragoussis, noting that this might be more appealing than spending \$300,000 or more before the lab even unpacks its first flow cell. In October last year, ONT launched its newest iteration of this platform, the portable P2Solo system, which can generate up to two human genomes per flow-cell run and allows users to get started for just over \$10,000.

In such a crowded marketplace, where change is a constant, investing in new technology requires a leap of faith. "It's very difficult to adapt every six months to a new technology – it demands a lot from the community for benchmarking, for testing and also for the bioinformatics team," says Oliveira. For now, his team is carefully weighing up the pros and cons of these emerging platforms and how they might complement or supplant his existing hardware. But competition, in general, is a good thing, driving performance up and costs down. "We are democratizing sequencing," he says.

**Michael Eisenstein** is a science writer based in Philadelphia, Pennsylvania.

1. Almog, G. et al. Preprint at bioRxiv <https://doi.org/10.1101/2022.05.29.493900> (2022).
2. Simmons, S. K. et al. *Nature Biotechnol.* **41**, 204–211 (2023).
3. Arslan, S. et al. Preprint at bioRxiv <https://doi.org/10.1101/2022.11.03.514117> (2022).