ILLUSTRATION: THE PROJECT TWINS

MOLECULAR 'BLACK BOXES' REVEAL CELLULAR SECRETS

Gene-editing tools transform cells into record-keepers, capturing hidden timelines of health and disease. **By Elie Dolgin**

Cells are poor storytellers. They live complex, eventful lives – dividing, migrating, responding to their environment. But by the time they reach the laboratory, whatever narrative they once carried is gone. Plucked from their native context and pinned beneath a microscope or cracked open to expose the genetic ledger inside, they reveal only their final state, not the path that led them there. For cell biologists, it's like glimpsing a play's finale and working backwards to deduce the plot.

For decades, researchers have strained to reconstruct cellular backstories through indirect means: static snapshots, fluorescent footprints and algorithmic inferences. But these are forensic efforts at best, post mortems of a vanished process.

A new generation of DNA-recording tools promises to change that. By outfitting cells with the molecular equivalent of flight recorders, scientists can now log, in real time, the signals that cells receive, the paths they travel and the decisions they make – all inscribed indelibly in genomic ink.

Early versions of these recording systems were crude, able to track branching lineages of cellular fate through cycles of division, but little else. With advances in gene editing and single-cell sequencing technologies, however, cells can now also record the cues they sense, the states they pass through and even the timing and intensity of signalling events and gene activity.

Researchers have used these tools to trace developmental decisions in mouse embryos, capture gene activity in animal brains and analyse gene expression in engineered bacteria as they pass through the gut. Yet these remain early demonstrations, and the research community is still focused on expanding its technical capabilities.

"The field is nascent," says Randall Platt, a genome engineer at the Swiss Federal Institute of Technology (ETH Zurich) in Basel. "There are more examples of concepts than practical use cases that have led to biological or medical breakthroughs."

Even so, momentum is building. Proof-of-concept studies are pointing towards real-world potential. And, backed by fresh investments, the pace of innovation is only expected to accelerate. "Many threads have matured," says Jay Shendure, a genome scientist at the University of Washington in Seattle. The goal now, he explains, is to "push the engineering to the next level".

Once optimized and broadly adopted, DNA-based recording tools could transform how scientists study development, disease and cellular decision-making in real time. "If we can now engineer cells to record in ways that are predictable and programmable," Shendure says, "that just opens up all kinds of frontiers."

DNA diaries

The dream of writing data into life's code dates at least to 2003, when Shendure was an MD–PhD student at Harvard Medical School (HMS) in Boston, Massachusetts. That year, he and his supervisor, geneticist George Church, filed a patent for what they called a "nucleic acid memory device": a primitive recording system in bacteria that used 'recombinase' enzymes to flip segments of DNA back and forth, like a toggle switch.

It was an ambitious vision for molecular data storage – but one constrained by the technology of the time. Even if the concept could, in principle, log gene activity, the DNA-sequencing methods then available made interpreting the records slow and impractical. "It was too early," says Church. It took another decade – and the advent of next-generation sequencing, which drastically sped up DNA reading and reduced the cost – before the first molecular recorders capable of capturing large amounts of biological information over sustained periods could be built.

Synthetic biologist Timothy Lu, then at the Massachusetts Institute of Technology (MIT) in Cambridge, led the charge, showing in 2014 how engineered bacteria could convert fleeting transcriptional signals into changes in DNA. By pairing this process with the expression of recombinase enzymes, Lu and his colleagues created permanent genetic barcodes[1]. Two years later, they extended the approach to record the identity of diverse chemical stimuli and the order in which they occur[2].

But this strategy soon hit a wall. Every extra piece of memory – every transcriptional event, signalling cue or developmental decision – demanded its own distinct enzyme that could operate independently of the others. And there were simply too few known recombinases to expand memory capacity in any substantial way.

A new technology platform was needed – and it arrived in the form of CRISPR–Cas9 gene editing.

In 2016, just a few years after CRISPR was adapted for editing human DNA, Shendure harnessed the tool to induce serial mutations at defined genomic sites, enabling his team to record lineage information across successive rounds of cellular division[3]. Known as GESTALT (genome editing of synthetic target arrays for lineage tracing), the system comprised two components: a stretch of DNA engineered with multiple CRISPR targets; and the Cas9 machinery that introduced edits at those sites (see 'Cellular recording devices').

As cells divided, cuts were repaired imperfectly, leaving unique barcodes that accumulated over time. By sequencing these evolving barcodes, researchers could reconstruct cellular family trees across tissues or even entire organisms – as Shendure and his colleagues, including Aaron McKenna, then a graduate student and now a cancer geneticist at the Dartmouth Geisel School of Medicine in Lebanon, New Hampshire, demonstrated in cultured human cells and zebrafish embryos. Others soon followed suit with CRISPR-based recorders for use in mouse and cancer models, achieving much greater memory capacity than with recombinase systems, and opening fresh avenues for studying cell dynamics in development and disease.

Armed with these molecular timekeepers, researchers revealed hidden routes for how cancer cells spread (known as metastasis) as well as insights into how cancer clones expand[4]. Developmental biologists, using similar strategies in embryonic mice, challenged long-standing views of blood formation, uncovering progenitor cell populations that make lasting contributions to the blood system[5]. "The level of detail and resolution it supports is unparalleled," Sarah Bowling, a developmental biologist at Stanford School of Medicine in California, says of the CRISPR-based lineage-tracing mice that she helped to pioneer.

But CRISPR–Cas9 also comes with trade-offs. Because it relies on double-strand breaks in DNA, it can be toxic to cells. And repeated cutting across target sites can erase the very record it is meant to preserve. As a result, although CRISPR enables richer lineage tracing than do older recombinase-powered tools, the barcodes it produces eventually saturate, cautions HMS synthetic biologist Sahand Hormoz. "If you delete enough, then everything's deleted – and you don't have any pattern you can use to reconstruct the lineage tree," he says.

Prime time

That concern has pushed the field towards more-refined editing strategies such as base editors and, more recently, prime editors, that rewrite DNA without inducing double-strand breaks. Both are built on a CRISPR–Cas framework, but whereas base editors swap one DNA letter for another at a single position, prime editors use a guided reverse transcriptase to write short custom sequences into the genome.

> **"We aim to get not just the tree topologies of which cells are related to each other, but also the timing."**

These next-generation editors are pushing the field beyond lineage recording, allowing precise logging of gene expression and signalling dynamics, and empowering researchers to record not just that an event occurred, but also what preceded and followed it. "Before, there wasn't really a good way to translate sensing what's happening in the cell to genome editing," says Junhong Choi, a synthetic biologist at the Memorial Sloan Kettering Cancer Center in New York City.

Now, by tying the activity of prime editors to stimulus-responsive promoters or transcriptional circuits, researchers can convert transient molecular events – such as bursts of gene expression or protein–protein interactions – into permanent, time-stamped DNA changes. What's more, this kind of genomic bookkeeping can be parallelized to respond to diverse cellular signals, creating simultaneous, chronological recordings of molecular activity in a type of multitrack data ledger, notes Theresa Loveless, a synthetic biologist at Rice University in Houston, Texas. "That modularity is a real advantage," she says.

Collectively, these systems – accessible through the biological materials repository Addgene – represent more than just clever uses of genome editing. They are sophisticated pieces of synthetic-biology design that integrate finely tuned logic circuits with modular genetic components: sensors to detect cellular or environmental cues; actuators to relay inputs; and writers to inscribe information into DNA. "It wraps together many components of synthetic cell manipulation into these complex machines," says Seth Shipman, a bioengineer at the Gladstone Institutes in San Francisco, California.

Still, even as these tools grow more powerful and intricate, researchers continue to find value in simpler approaches – including the kind that Shendure and Church first imagined more than two decades ago.

Last year, biophysicist Harry McNamara and his colleagues at Princeton University in New Jersey used a recombinase-based system to track the activity of key signalling molecules during early mouse-embryo development. This revealed distinct groups of cells whose molecular experiences were correlated with their eventual locations along the embryo's anterior–posterior axis[6]. It was a task for which recombinases were well suited, because the biological question required only a durable yes/no record of signalling activity, not the detailed logs that CRISPR-based tools can provide, says McNamara, who is now at Yale University in New Haven, Connecticut.

"It allowed us to really look back in time and identify these key decision points when cells decide, in a sense, where they're going to end up," he says.

Fluorescent focus

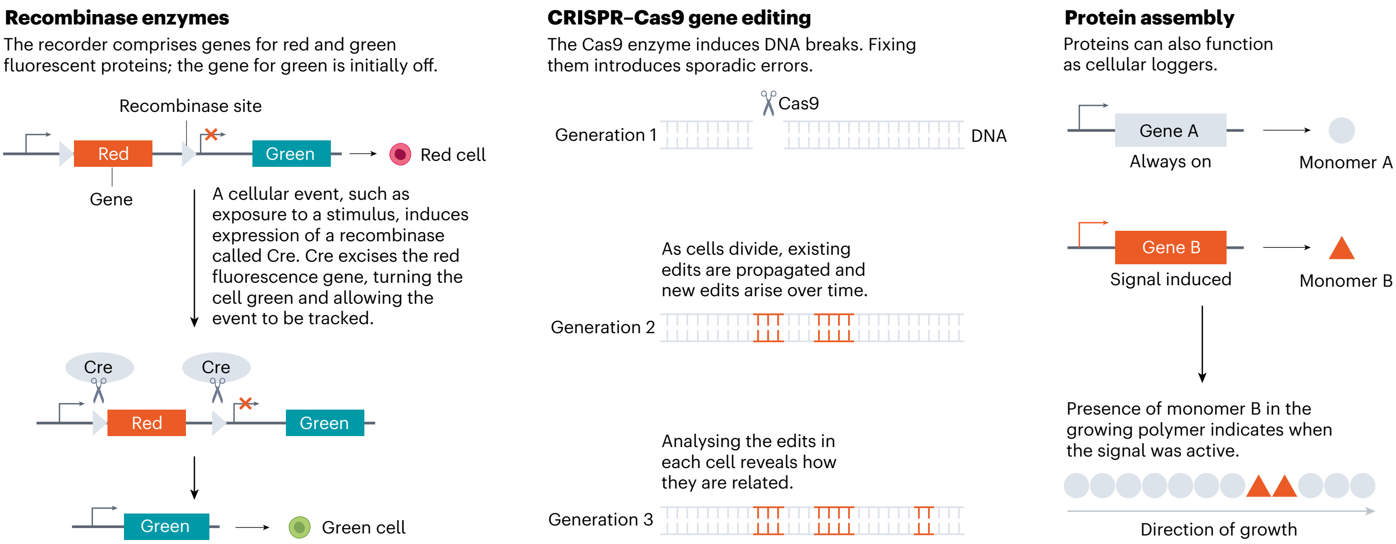
Not everyone is sold on DNA as the universal recording medium, however.

Bioengineer Changyang Linghu at the University of Michigan in Ann Arbor, for one, favours a system in which protein monomers are added to a growing super-molecular assembly in response to transcriptional events[7]. This strategy works best in non-dividing cells such as neurons, because the protein assemblies are diluted during cell division. But the approach also offers a significant advantage: by pairing fluorescence imaging with event-specific protein tags, researchers can determine not just when and how strongly a gene was active, but also where in the cell this activity occurred.

Most DNA-based recording platforms lose this spatial context, because sequencing requires cell lysis, which destroys cellular architecture. To address this, synthetic biologist Amjad Askary at the University of California, Los Angeles, developed imaging-based systems that link fluorescence signals to genetic barcodes, enabling his team to trace transcriptional histories in intact tissues[8]. Notably, however, "these are nowhere close to as easy and adaptable for labs to do as

CELLULAR RECORDING DEVICES

Researchers have devised several strategies to log the events happening inside cells. Here are three approaches.



single-cell sequencing", says Askary.

And regardless of the medium, writing and recording are only the first steps on the long road towards understanding these ticker-tape-like genetic storylines. Once data are inscribed – in DNA, proteins or other formats – scientists must then decode what they all mean.

For decades, researchers have used classical models of genetic evolution to reconstruct how species diverge and how mutations accumulate over time. But those frameworks don't map neatly onto DNA recorders, because the genetic marks that these technologies install are irreversible, not the product of random mutational drift that can revert or be overwritten over time. "This unmodifiability is actually quite an important property that influences the way we think about both what is the evolutionary process that underlies the data, but also how would algorithms behave," says Nir Yosef, a computational biologist at the Weizmann Institute of Science in Rehovot, Israel, who has built analytical frameworks tailored to the special constraints of DNA-recording systems.

The most widely used such tool is Cassiopeia, an open-source software platform that was unveiled in 2019 and is available on the code repository GitHub. The tool integrates evolutionary models with single-cell data to infer developmental lineages across time and space.

By doing so, the tool, along with later versions adapted to other editing strategies, allows researchers to piece together how cells diversify, specialize and organize during development – and how those processes go awry in diseases such as cancer.

In a mouse model of lung cancer, for example, Cassiopeia-powered lineage tracing revealed that rare, transcriptionally distinct subclones can gain a fitness advantage, take over the tumour and seed metastasis[9]. It also uncovered how these cancer populations emerge from microenvironments that promote their spread[10]. As study co-author Matt Jones, a computational biologist who trained with Yosef and is now at MIT, explains, Cassiopeia made it possible to pinpoint "these stepwise changes that are actually affecting cancer-cell progression".

Building on such lineage maps, researchers are now pushing to add a temporal dimension – covering not just which cells are related, but when key branching events occur.

"We aim to get not just the tree topologies of which cells are related to each other, but also the timing," says Tanja Stadler, a mathematical biologist at ETH Zurich in Basel. If successful, this could turn lineage tracing from a freeze-frame image into a dynamic record of developmental tempo, helping to explain why some cells bide their time whereas others race towards malignancy or differentiation.

Standing sentinel

The promise of molecular recorders extends well beyond reconstructing developmental histories and cellular decision-making, however. Researchers are also exploring how to harness these tools for real-time disease monitoring – by engineering cells that patrol the body and log signs of trouble in molecular memory.

In principle, human cells outfitted with recorders could be retrieved through biopsy to guide clinical decision-making. But, as Church notes, "that's pretty invasive". So researchers have turned to a less intrusive alternative: engineered bacteria.

In 2022, Platt and his colleagues described how they programmed *Escherichia coli* to log transcriptional changes as the bacteria transited through the intestines of mice. Recovered from faeces, these 'sentinel cells' were then sequenced to reveal which genes were activated and how the gut environment changed in different dietary and disease contexts[11].

Lu, who pioneered DNA recorders in bacteria and now leads the cancer cell-therapy company Senti Bio, based in South San Francisco, California, sees strong potential here. "The field is still craving a lot of non-invasive sensors," he says, "and leveraging microbes that are already there to read out what's going on makes a tonne of sense."

But, aside from Platt's gut-monitoring experiment, proof-of-concept clinical applications remain rare.

Shendure acknowledges the uncertainty of translational timelines and remains convinced that the pay-offs will come. "We're building a strong foundation," he says. To that end, Shendure secured more than US$80 million to launch the Seattle Hub for Synthetic Biology, which opened its doors last year and is dedicated to building versatile and scalable cell-recording systems. With continued technical progress, the curtain could at last rise on the full story cells have to tell.

"Everyone is excited," says McKenna. "But it's certainly a field that still has to prove its benefit."

Elie Dolgin is a science journalist in Somerville, Massachusetts.

1. Farzadfard, F. & Lu, T. K. *Science* **346**, 1256272 (2014).
2. Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S. & Lu, T. K. *Science* **353**, aad8559 (2016).
3. McKenna, A. *et al. Science* **353**, aaf7907 (2016).
4. Quinn, J. J. *et al. Science* **371**, eabc1944 (2021).
5. Patel, S. H. *et al. Nature* **606**, 747–753 (2022).
6. McNamara, H. M., Solley, S. C., Adamson, B., Chan, M. M. & Toettcher, J. E. *Nature Cell Biol.* **26**, 1832–1844 (2024).
7. Linghu, C. *et al. Nature Biotechnol.* **41**, 640–651 (2023).
8. Hao, K. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2024.10.11.617908 (2024).
9. Yang, D. *et al. Cell* **185**, 1905–1923 (2025).
10. Jones, M. G. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2024.10.21.619529 (2025).
11. Schmidt, F. *et al. Science* **376**, eabm6038 (2022).