nature reviews genetics

https://doi.org/10.1038/s41576-024-00788-w

Perspective

Check for updates

The lives of cells, recorded

Amjad Askary [1,15], Wei Chen [2,3,15], Junhong Choi [2,4,15], Lucia Y. Du [5,6,15], Michael B. Elowitz [6,7,8] ✉, James A. Gagnon [9] ✉, Alexander F. Schier [5,6] ✉, Sophie Seidel [10,11,15], Jay Shendure [2,6,12,13,14] ✉, Tanja Stadler [10,11] ✉ & Martin Tran [7,15]

Abstract

A paradigm for biology is emerging in which cells can be genetically programmed to write their histories into their own genomes. These records can subsequently be read, and the cellular histories reconstructed, which for each cell could include a record of its lineage relationships, extrinsic influences, internal states and physical locations, over time. DNA recording has the potential to transform the way that we study developmental and disease processes. Recent advances in genome engineering are driving the development of systems for DNA recording, and meanwhile single-cell and spatial omics technologies increasingly enable the recovery of the recorded information. Combined with advances in computational and phylogenetic inference algorithms, the DNA recording paradigm is beginning to bear fruit. In this Perspective, we explore the rationale and technical basis of DNA recording, what aspects of cellular biology might be recorded and how, and the types of discovery that we anticipate this paradigm will enable.

Sections

[1]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA, USA. [2]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [3]Department of Biochemistry, University of Washington, Seattle, WA, USA. [4]Developmental Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [5]Biozentrum, University of Basel, Basel, Switzerland. [6]Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA. [7]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [8]Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA, USA. [9]School of Biological Sciences, University of Utah, Salt Lake City, UT, USA. [10]Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland. [11]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [12]Howard Hughes Medical Institute, Seattle, WA, USA. [13]Brotman Baty Institute for Precision Medicine, University of Washington, Seattle, WA, USA. [14]Seattle Hub for Synthetic Biology, Seattle, WA, USA. [15]These authors contributed equally: Amjad Askary, Wei Chen, Junhong Choi, Lucia Y. Du, Sophie Seidel, Martin Tran. ✉e-mail: melowitz@caltech.edu; james.gagnon@utah.edu; alex.schier@unibas.ch; shendure@uw.edu; tanja.stadler@bsse.ethz.ch

Introduction

A fundamental challenge in biology is to explain biological states or processes, including normal development and disease, in terms of the events that precede them. The sequence of events taking place in individual cells can be collectively thought of as a cellular history. Cellular histories include a range of features about each cell that encapsulate past relationships and events, inform its present state, and constrain its possible futures (Fig. 1a), including:

- How a cell is related to other cells by lineage, that is, which cells are its sister, its cousins and so forth. Cellular phenotypes, functions and potentials are often inherited through, or structured by, mitotic divisions[1].
- The identity, amplitude and duration of extrinsic signals (mechanical or biochemical) that a cell, or its ancestors, received at different points in its past. Such signals provide cues for changes to cell state or cell fate[2], or responses to environmental stresses or injuries.
- The longitudinal dynamics of each cell's internal molecular state (that is, its trajectory), including the levels of transcription factors and other regulatory molecules, as well as its epigenome. These characteristics constrain and often determine a cell's functional behaviours, including its ability to change into other states[3].
- Spatial context, including the identities and states of a cell's past and present neighbours. The spatial neighbourhood of a cell informs its function within tissue and reflects its developmental history of growth, differentiation and cell movement[4].

Cellular histories have the potential to show not only how individual cells are related to one another, but also what events drove them into their observed states, in diverse biological contexts (Fig. 1b). For example, in developmental biology, cellular histories can help to delineate the developmental potentials of all progenitors, and can conclusively determine the origin(s) of both healthy and disease-associated adult cell types[5–7]. Applied to cancer progression, cellular histories reveal how clonal fitness evolves in tumours, yielding reconstructions of the emergence of invasiveness as well as the spatiotemporal dynamics of metastasis[8]. Reconstructing the intrinsic and extrinsic signalling history of individual cells can inform our understanding of cell fate determination, and of spatial and temporal patterning of tissues and organs during development[9–11]. In microbial environments, such as soil or the human body, cellular histories can reveal how microorganisms respond to different environmental cues and conditions[12].

From a technical standpoint, cellular histories are challenging to obtain. The most obvious approach is through direct visualization of cells over time. However, many model organisms and microbial environments are composed of millions to trillions of cells generated through variable patterns of cell division and traversing non-deterministic trajectories. In addition to scalability limitations, the direct observation of cells is complicated by the opaqueness of organisms, their physical movement during activities of interest, and the long timescales over which biology plays out. An alternative to visualization involves genome-wide methods for measuring cell state, such as multiplex fluorescence in situ hybridization (FISH) or single-cell RNA sequencing (scRNA-seq). For example, the widespread application of scRNA-seq has revealed a myriad of cell types and cell states in diverse organisms[13–15]. However, these approaches are typically destructive, preventing continuous monitoring of the same cells over time. Since trajectories often involve reversible or discontinuous changes, it is difficult to reliably infer a series of states traversed by individual cells from a series of snapshots[16].

Building on a rich history (Box 1), the nascent field of DNA-based recording offers a potential way of obtaining cellular histories that overcomes the limitations of current measurement paradigms. The basic premise of DNA-based recording is to engineer cells to record their histories into their genomes using DNA editors, such that a single destructive snapshot is informative with respect to not only each cell's present but also its past. A key principle that enables DNA-based recording is the generation of heritable mutations at defined target sites. Two types of system can be used to generate edits: constitutive recorders generate random edits on a set of target sites, allowing reconstruction of lineage trees[9,17,18], whereas conditional recorders generate edits at a different set of target sites at a rate proportional to the external signals, internal states or spatial contexts being sensed, allowing analysis of the past trajectory of single cells[17,19]. Many conditional editors and corresponding target sites can be included to record as many biological inputs of interest as needed. When both systems are activated, the result is a cell lineage tree 'decorated' with the relative activity of specific biological events over cell divisions or absolute time (Fig. 1c) across different biological contexts for which recording has strong potential to provide insight.

This Perspective focuses on the present and future of DNA-based recording, highlighting recent technological advances. Breakthroughs in DNA synthesis[20], assembly[21,22] and delivery[23] are facilitating the engineering of genomes to encode the necessary components of read/write systems. Site-specific recombinases[24], clustered regularly interspaced short palindrome repeats (CRISPR) systems[25–27] and other genome-editing technologies[28] are enabling the writing of information to the genome, while advances in single-cell molecular profiling, spatial transcriptomics and sequencing are enabling its recovery[29,30]. Finally, progress in big data analysis and phylogenetic inference algorithms is furthering our ability to reconstruct complex cellular histories from DNA edits[31]. Of note, although RNA-based[32–34] and protein-based[35,36] recording methods have been described recently, we focus here on recording to genomic DNA, which allows long-term storage and faithful transmission of recordings through not only cell divisions but also throughout the life of an organism. For practical guidance on how recording experiments can be designed, executed and interpreted, we refer the reader to other recent reviews[16,37–39].

Recording cellular histories

DNA recording systems integrate recent advances in three areas: (1) writing information through time as heritable DNA edits, either by constitutive or conditional editing; (2) reading recorded information from each cell at an endpoint; and (3) reconstructing the histories of cells from recovered edits (Fig. 2a).

Writing to cellular histories to DNA

Systems for writing information to the genome would ideally have several characteristics. Writing should be minimally perturbative, that is, neither the editor nor the editing activity should alter cell identity or fate. Recording loci should contain as many target sites as possible to maximize information capacity. Constitutive recorders should be active at a rate that is steady enough and high enough to generate unique edits during each cell cycle. Conditional recorders should be activated by a particular internal or external signal, and be sensitive, selective and quantitative, that is, active in proportion to the biological signals of interest that they sense (for example, enhancer activity, signal duration or protein levels), and inactive in the absence of that signal. Recorders should continue to operate in post-mitotic cells. Each target site should

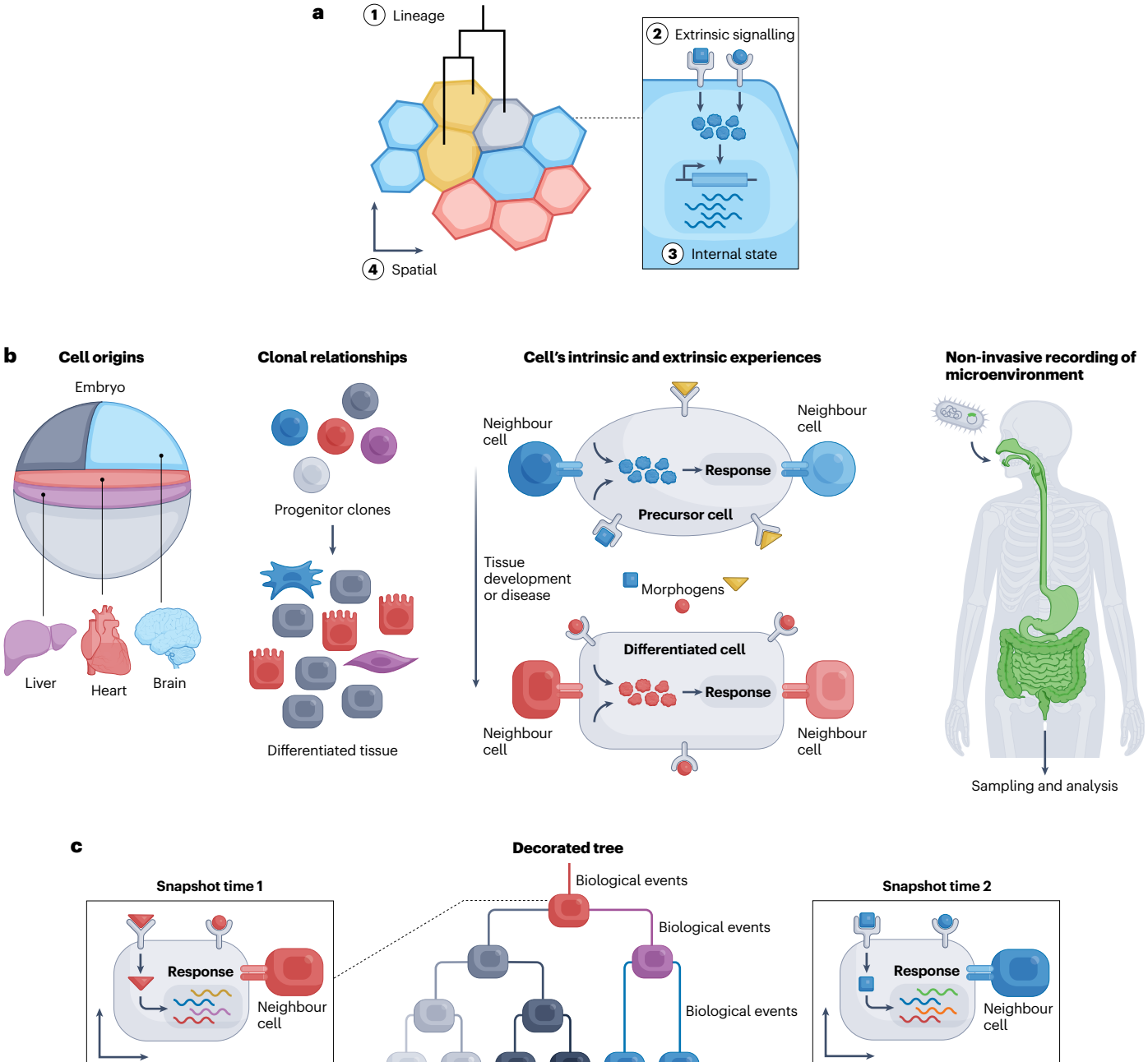


Fig. 1 | Rationale for DNA recording. a, Longitudinal recordings of cellular histories are needed to understand their dynamics, such as cell lineage (1), extrinsic signalling (2), internal state (3) and spatial relationships (4) that give rise to cell states and fates over time. **b**, Recording in diverse biological contexts may enable insights into the origin(s) of adult cell types, the evolution of clonality, the intrinsic and extrinsic signalling history associated with final cell fate, and microenvironments. **c**, A decorated tree reconstructed from DNA recordings provides multimodal cell histories (see panel **a**) over time.

be editable to many distinct and terminal 'character states' without the risk of target-site loss. Finally, the sequential order of edits should ideally be recorded to explicitly capture the temporal order of cellular events.

Several classes of editor are being explored as writers, which act by rearranging, scarring, base editing or inserting into genomic DNA. Each has its own strengths and weaknesses (Fig. 2b).

Rearranging generates edit-state complexity without DNA cleavage. Enhancer-driven site-specific recombinases (SSRs) and integrases have long been used to catalyse programmable rearrangement of DNA sequences as a form of biological recording, for example, recording chemical exposures to population of cells[40,41] or permanently marking a subpopulation of cells and its descendants

Box 1 | A historical perspective on recording cell histories

Over the past two centuries, various modes for recording cell histories have emerged that can be broadly categorized into: lineage tree reconstruction, where each branch ideally represents single cell divisions; clonal analysis, where cell fates that descended from a common progenitor are grouped together; and signal recording, where previously experienced signals provide an indelible mark within cells.

Strategy 1: Reconstructing trees of embryonic development through direct observation

The field of cell lineage tracing has deep roots that extend to the late nineteenth century, when Charles Whitman, Edwin Conklin and Edmund Wilson independently observed cells in embryos to trace the developmental origins of germ layers in various marine invertebrates[145–147]. Strikingly, they found that these embryonic cell divisions often occur in a stereotypical, invariant manner that was similar between different species. Conklin began constructing larger fate maps[148] — schematics that delineate the future fates of individual progenitor cells or regions. Decades later, John Sulston and colleagues used direct observation to map the complete lineage tree that gives rise to the cells of the hatched larvae of *Caenorhabditis elegans*[149]. These studies demonstrated how relatively simple observational methods can be used to construct detailed maps of embryonic development. However, tissues and embryos in many organisms are not transparent, are larger in scale and proceed more slowly and less deterministically in their development than in *C. elegans*, precluding the use of direct observation to reconstruct complete cell lineages.

Strategy 2: Clonal analysis through dye injections, transplantation, chimaera generation, retroviral infection and in vitro culture systems

Experimental manipulations can be used to trace the contributions of progenitor cells to descendants. For example, researchers began using dyes and radioactive tracers to label groups of progenitor cells[150–152], transplanting tissue to foreign hosts[153] and generating chimaeras through embryo aggregation[154,155]. These approaches produced fundamental discoveries about fate bias and cell migration. For example, the construction of fate maps within amphibian embryos revealed that vertebrate embryos are generally divided into three germ layers, which contain progenitors that give rise to particular sets of future organs. Later approaches permanently labelled single cells by reporter gene transfection to analyse the fates of their progeny[156–159]. In vitro culture systems were also used to track individual progenitors using time-lapse microscopy[160]. These enhanced versions of direct observation revealed that clones could be surprisingly variable in both size and descendant-cell diversity. Unlike *C. elegans* development, vertebrate development is much more plastic. However, although these approaches are powerful, portable and still widely used, they are best suited to looking at a single process or cell type at a time, and are difficult to scale to all cells in a tissue or organism.

Strategy 3: Clonal analysis in specific subsets of progenitors using genetic recombinases

Genetic recombinases began to be used widely in the 1990s to analyse groups of progenitors marked by the expression of key genes[161–163]. These experiments often involved expressing a recombinase (Cre or FLP) in a cell-type-specific or tissue-specific manner to permanently activate a reporter gene to label all descendant cells. These studies generally revealed that while some progenitor cells seem to be multipotent, giving rise to many fates, other progenitor cells can have distinct and reproducible biases in cell fate[164–166]. Nevertheless, this work was limited to analysing groups of progenitor cells, since all activated progenitors would be identically labelled, preventing high-resolution, single-cell analyses of lineage and fate. More complex recombination cassettes were later developed to label different cells with distinct colours or combinations of colours[42,43,167,168]. These allowed the simultaneous analysis of dozens of clones within a tissue, providing insight into mechanisms like clonal crypt formation in the intestine and subsets of neural progenitors in the developing brain. However, these methods were not scalable to larger tissues or organs, because limited colour diversity meant that unrelated cells could be coincidentally labelled with the same colour combination with high likelihood.

Genetic recombinases have also been adapted for recording developmental signalling pathway activation and neural activity. These approaches rely on signalling-pathway-specific promoters that express a recombinase, often gated by small-molecule activation enabling temporal control of recording, leading to permanent activation of a reporter gene. Importantly, these approaches typically record the transcriptional activation of a signalling pathway, not the expression of a signalling molecule itself. Treating embryos with a drug during a window of developmental time will permanently label all cells that activate a given signalling pathway. This approach has been used, for example, in mouse embryos to link Shh signalling to limb and digit patterning[169,170] and forebrain neurogenesis[171]. More recently, genetic recombinases were used in adult animals to label and manipulate the neural circuits that control cognition and behaviour[172,173]. Current strategies label specific subsets of neurons for manipulation based on immediate early gene expression, intracellular calcium levels[174,175] or synaptic activity[176,177]. These approaches typically also suffer from the inability to mark different cells with distinguishable labels.

Strategy 4. Reconstructing lineage trees using naturally occurring somatic mutations

In most approaches discussed thus far, labels are experimentally delivered to cells that subsequently proliferate and differentiate. These prospective approaches are limited to experimentally accessible systems and cannot be applied to the study of human embryo development. As an alternative, somatic mutations that accumulate naturally due to errors in DNA replication can be sequenced and used to reconstruct lineage trees retrospectively. As a proof of principle, researchers began to use sequencing to trace lymphoid lineages through somatic hypermutation at immunoglobulin loci[178], intestinal lineages through post-zygotic methylation patterns at CpG sites[179], and epithelial lineages through polyguanine repeat DNA sequences[180]. With new innovations in single-cell DNA and RNA sequencing in the 2010s, somatic mutations could be detected at scale and used for lineage reconstruction, including single-nucleotide variants[181–185], copy-number variants[186]

(continued from previous page)
and microsatellite repeats[187]. These approaches have been used to study human embryogenesis, revealing a surprisingly unequal contribution of early embryonic cells to adult somatic tissues[188]. However, although powerful, these approaches are only able to reconstruct lineage trees, because the transient trajectories that influence human cell-type differentiation are not recorded in DNA-replication errors.

Strategy 5: High-throughput clonal analysis by sequencing static barcodes

In 1992, Walsh and Cepko pioneered the use of diverse retrovirus libraries containing short unique DNA sequences to distinguish between individual clones within the same tissue at scale[189]. Initially, this approach was laborious because individual cells had to be excised and processed to enable PCR to detect the barcode. However, scalable single-cell sequencing in the 2010s opened up the era of single-cell clonal barcoding, allowing the simultaneous analysis of many clones containing endpoint cell states. Multiple approaches for clonal barcoding, such as diverse viral libraries, multiplexed recombination cassettes or mobile transposable elements, were first applied to the study of haematopoiesis. These studies revealed migration patterns of antigen-specific T cells[190], fate potential in lymphoid-primed multipotent progenitors[191] and hematopoietic stem cell (HSC) progenitors[43,192], and differences in HSC proliferation between irradiated transplantation settings[193] versus steady-state haematopoiesis[194]. Clonal barcoding was later applied to other systems to study fate differentiation in cancer[195] and forebrain development[196,197]. Nevertheless, the static nature of these barcodes limited the inferences of temporal dynamics. This limitation has been partially addressed by taking advantage of systems where cells can be grown in vitro. For example, Weinreb et al.[198] barcoded a pool of haematopoietic progenitors, allowed them to expand, then split them to simultaneously profile cell states immediately and at later time points after differentiation, thus linking early transcriptional states to later clonal fate compositions. Biddy et al.[199] applied lentiviral transduction successively to fibroblasts undergoing reprogramming for coarse-scale lineage tree reconstruction, in which branches represent multiple cell divisions. However, these approaches are difficult or impossible to implement in vivo, necessitating other approaches for resolving temporal dynamics.

with a fluorescent reporter (Box 1). Newer SSR-based recording systems such as intMEMOIR leverage more complex target sites, such as arrays of target sites that stochastically or iteratively rearrange to multiple character states[10,42,43]. Editing by SSRs avoids endonucleolytic cleavage of DNA strands and does not involve the endogenous DNA repair machinery[44,45]. Although the expression of an SSR can be made signal-dependent (for example, by placing it under the control of a *cis*-regulatory element that is activated by a transcription factor, drug or morphogen), the potential of SSRs to record multiple signals is limited at present by the relatively small number of well characterized, orthogonal SSRs.

Scarring efficiently generates diverse mutations at target edit sites. CRISPR–Cas9 editors can be used to induce DNA double-strand breaks (DSBs) at target sites specified by guide RNAs (gRNAs)[9,17,18,46–49]. Imperfect repair of those DSBs generates irreversible insertions and deletions (indels), termed 'scars'. This approach was taken by early attempts at DNA-based lineage recording, such as GESTALT[9,50], and is the most widely applied to date. It has been successful for in vivo recording applications thanks to its high editing rates and capacity for multiplexing with different gRNAs and CRISPR systems[9,18,46–49,51,52]. Furthermore, some in vivo scarring-based implementations place the Cas9 editor under the control of an inducible promoter, that is, single-channel conditional recording[29]. However, DSBs can be toxic to cells, and in arrays of target sites, deletion scars frequently compromise adjacent target sites[9,50,53]. Moreover, this approach does not directly capture the order in which recorded events occurred (that is, phylogenetic inference across a population of cells is required), although variant approaches based on self-targeting gRNAs can partly mitigate this issue[54,55].

Base editing allows digital editing at precise, densely encoded target sites. CRISPR base editors can generate specific point mutations at target sites[28,56,57] and have been used for DNA recording[30,58,59]. Base editors induce mutations without generating DSBs, and allow dense packing of predictably editable target sites into arrays, facilitating subsequent analysis of edit patterns. Although base editors typically generate only one character state, they can be used to produce multiple edit outcomes in two ways. First, newer base editors (such as AXBE and AYBE) allow a single initial base to be edited to all three other bases[60,61]. Second, by editing dinucleotides, rather than single-base target sites, it is possible to generate three alternative dinucleotides[62,63]. As with indel and rearranging methods, base editing in general produces unordered edits unless the sites are engineered to be edited sequentially[64,65], and can exhibit substantial off-target activity[66–69], although this can be in part addressed using newer-generation base editors with less off-target activity[70–73].

Inserting allows linearly ordered recording of multiple signal-specific symbols. Prime editors provide a way to generate temporally ordered edits. They are composed of a Cas9 nickase fused to a reverse transcriptase, and use corresponding prime editing guide RNAs (pegRNAs) to insert short sequences precisely at target sites[28,74,75]. The DNA Typewriter and peCHYRON techniques leverage prime editors to achieve a high number of potential character states for each edit (at least dozens but potentially thousands, if that many pegRNA-expressing constructs could be concurrently introduced)[76,77]. Furthermore, in both systems, each insertional edit creates a target site for a new edit. This allows edits to be concatenated sequentially, such that their temporal order is reflected in their linear order along the DNA. Furthermore, prime editors can be used to record transcriptional events by making the production of specific pegRNAs conditional on the activity of a *cis*-regulatory element, using a framework termed ENGRAM[78]. Similarly, pegRNAs can be activated through protein–protein interactions, a second kind of biologically conditional editing[79]. A drawback with prime editors is that they are currently less efficient than other editors, reducing overall edit rates and temporal resolution, although this may be addressed by newer prime editors[80].

A second class of insertional editors leverages Cas1–Cas2 systems to integrate short DNA segments (spacers) generated from reverse

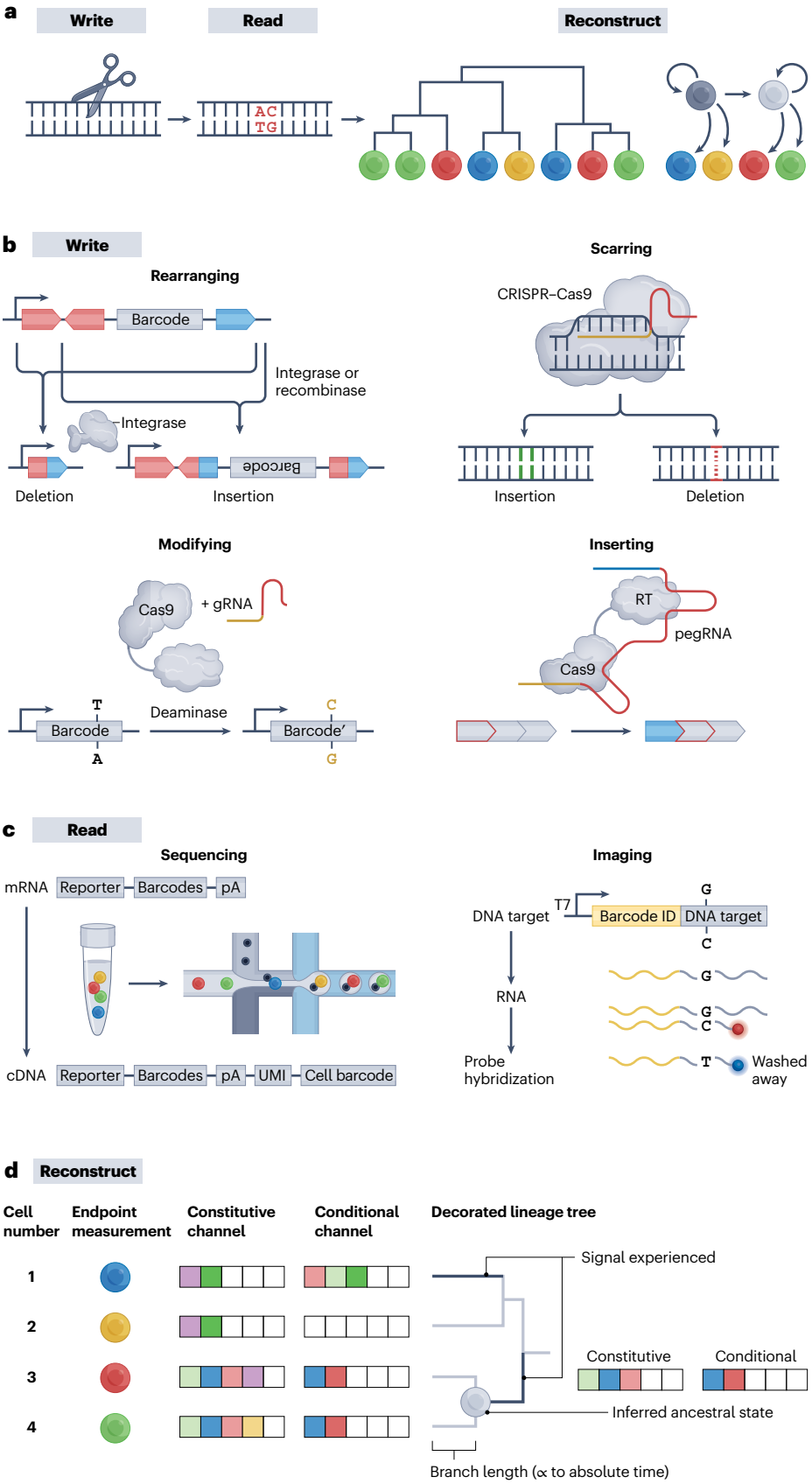


Fig. 2 | Recording and reconstructing cellular histories. a, The general DNA recording workflow consists of the following steps: writing information to encode biological information, reading recorded information and reconstructing cell histories. **b**, Writing information can be done by rearranging, scarring, modifying or inserting DNA. **c**, Reading recorded information by sequencing or imaging. For sequencing, mRNA encoding barcodes are labelled with unique molecular identifiers and cell barcodes. For imaging, primary probes that bind to the edited and unedited version of the barcode are hybridized — the probe with sequence match outcompetes the one with mismatch for binding. **d**, Reconstructing trees through DNA recordings allows estimation of the tree topology, branch timescales, ancestral node states and cell dynamics. Edits in the constitutive channel enable tree reconstruction, while edits in the conditional channel allow 'decoration' of the tree with experienced signals. cDNA, complementary DNA; CRISPR, clustered regularly interspaced short palindrome repeats; gRNA, guide RNA; pA, polyA; pegRNA, prime editing guide RNA; RT, reverse transcription; T7, T7 RNA polymerase; UMI, unique molecular identifier. Part **b** adapted with permission from ref. 10, AAAS.

transcription of intracellular RNA in CRISPR spacer arrays, providing an explicit and ordered record of a cell's gene-expression history[12,19,81–83]. With this approach, the acquisition of spacers has been shown to be both orthogonal and dose-dependent on absolute mRNA present within the cell[12,82,83]. However, these systems are currently limited to prokaryotic systems owing to reliance on accessory integration host factors.

Overall, writers that can deliver precise, information-rich and temporally ordered edits at high rates, without loss of previous recorded information and without perturbing cellular or organismal physiology, are necessary for this paradigm to reach its full potential. At present, prime editors could meet these criteria, especially if they can be optimized to achieve higher edit rates and to perform efficiently across diverse cell types. Base editors could also meet these criteria if they can be optimized to store more memory per edit site and to perform temporally ordered edits in a scalable manner. Cas1–Cas2 systems have the greatest capacity for capturing rich transcriptomic information over time but might require considerable optimization to be successfully ported to eukaryotic systems. Finally, most writers discussed so far have been constitutively expressed to record lineage, and the development of biologically conditional editors remains immature (see section 'Practical challenges for DNA-based recording' for further discussion of the challenges).

Reading DNA-recorded information

Information recorded to genomic DNA must eventually be recovered by sequencing or imaging methods. Some approaches directly recover recording information from DNA[9,18,43,58], but most approaches transcribe the single-copy genomic DNA records into RNA for capture alongside the transcriptome and/or epigenome[29,47,63,84] (Fig. 2c).

Sequencing. scRNA-seq on dissociated tissues is a convenient and scalable means of obtaining rich endpoint measurements that can also be used to recover DNA-based records. For example, target sites for editors can be embedded within the 3′ untranslated region (UTR) of expressed reporter genes, facilitating their recovery with standard scRNA-seq protocols[29,46,47], such that for each cell, historical information is recovered alongside an endpoint transcriptome. However, technical limitations of scRNA-seq, such as cellular loss, dropout and loss of spatial information, can introduce uncertainty or bias in the analysis of lineage as well as other recorded information. If scRNA-seq could enable recovery of all (or nearly all) recorded information from all (or nearly all) cells in the tissue or organism profiled, the inference of complete cellular histories would be possible. To our knowledge, the highest recovery that has been achieved while concurrently capturing DNA-based records was about 50% of cells in a monoclonal expansion of HEK293T cells in vitro[85]. An alternative to scRNA-seq would be for a rich set of recordings to be captured to a dense region and recovered by long-read sequencing, such that a single sequencing read would suffice for reading out the history of each cell.

Imaging. Imaging-based methods can also recover omic measurements while fully preserving the spatial relationships of cells. Elegant approaches for querying thousands of genes use sequential rounds of fluorescent in situ hybridization and imaging[86–88]. Adaptations of such methods can be used to amplify and query single base edits in situ to recover information from DNA-based records[30]. Such methods could be adapted to discriminate between more diverse editing outcomes. For imaging-based spatial transcriptomics methods, this requires that the set of potential character states be known in advance, such that probes can be designed to discriminate between them. However, with sequencing-based spatial transcriptomic methods, diverse editing outcomes could potentially be read out directly[89].

Reconstructing cellular histories

Lineage trees. The development of tools to reconstruct cellular histories has been led by efforts to use data from constitutive recorders to generate cell lineage trees. Lineage tree reconstruction draws on the general principle that lineage relationships can be inferred by comparing edit patterns from constitutive recorders between cells (Fig. 2d). Roughly speaking, the more similar the edit patterns are between cells, the more closely related the cells should be in the lineage tree. To enable accurate reconstruction, the properties of the recording technology have to be taken into account; for example, how character states are generated[90–92] and how the frequencies of different character states vary[9,90,91,93].

Current algorithms for reconstructing lineage trees differ in speed and accuracy. Statistical methods, such as maximum likelihood or Bayesian methods, using an appropriate model can be very accurate. However, because the number of possible tree topologies grows super-exponentially with the number of cells, reconstruction of larger lineage trees that include many cells is typically not feasible because these methods are too slow. Instead, algorithms ensuring fast tree reconstruction using heuristics are used, although they may suffer from limitations in accuracy. For example, greedy algorithms, either building the tree top-down starting with early, shared indels[94], or bottom-up by progressively merging cells with similar target sites[95], can ensure reconstruction of large trees with the risk of finding only a locally optimal tree rather than the globally best tree. Alternatively, an iterative tree-building algorithm[96] first creates a basic tree structure using a subset of cells and then iteratively adds in additional samples. However, initial choices, especially those related to the tree's backbone, may heavily bias the final tree structure. Systematic comparisons across tools with real and simulated benchmarking data are essential to evaluate the speed and accuracy of any construction algorithm[93].

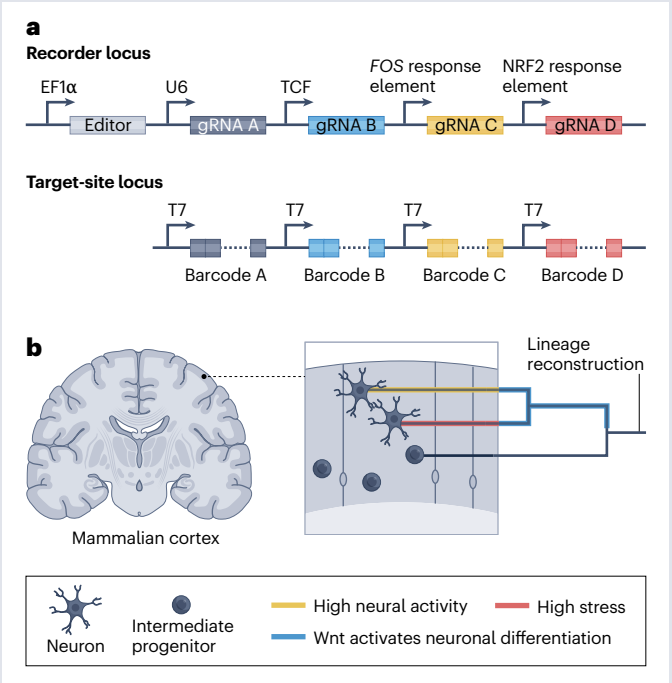
In addition to estimating an accurate tree topology, branch lengths can be calibrated to absolute time to understand the timing of various cell events such as differentiation or metastasis. Calibrating the lineage tree topologies in absolute time (such as hours)[90,95] can be achieved by modelling the edit accumulation as a function of time.

Cell population dynamics. Time-scaled lineage trees contain information about cell population dynamics. For example, if most branching events happened close to the start of the tree, the cell population probably experienced rapid early expansion and exponential growth. Alternatively, if most branching events occurred close to the end of the tree, then the cell population was probably fairly constant over time, which could be indicative of long-term maintenance of a stem cell pool[97]. The field of phylodynamics has built statistical tools to extract population dynamics from trees. In the context of single-cell biology, methods of estimating the population size and progenitor commitment times within the sampled population of cells[95] and the rates of cell division and death events in the entire cell population[90] have been proposed.

Intrinsic states, extrinsic signals and spatial context. Biological recording is possible and of interest even when it does not refer to cellular lineage information. Examples are signal event histories in a post-mitotic neuron (Box 2) or pathogen-exposed innate immune cell (Box 3). Ideally, the dynamics of multiple signals within

Box 2 | The potential of DNA-based recording in the nervous system

Understanding the complex processes that shape neural development and function is a central goal in neuroscience. By reconstructing the lineage history of individual neurons, we can gain insights into how the developmental relationship of a neuron with other neurons influences its final identity and function within the brain. Recording signalling pathway activities can further reveal the roles of specific pathways during cell-fate determination, while calcium activity and oxidative stress recordings can indicate how neurons respond to stimuli. For example, as illustrated, barcodes A, B, C and D could record lineage, Wnt signalling, calcium activity and oxidative stress, respectively (see the figure, panel **a**). In such an experiment, constitutive editing of barcode A is first used to record the lineage relationships of neurons during mouse development, by expressing an editor from an EF1α promoter and a guide RNA (gRNA) under a U6 promoter. Conditional recording is then used to record the activity of Wnt signalling. A Wnt-responsive enhancer (TCF) is coupled to the activation of an orthogonal editor that introduces edits into barcode B. Next, conditional recording is used to record neural activity in postmitotic neurons. Long-term calcium activity is recorded by coupling the immediate early gene *FOS* to another orthogonal editor that introduces edits into barcode C. Conditional recording is then used to record oxidative stress. An NRF2 (nuclear factor erythroid 2-related factor 2)-binding DNA response element, which responds to electrophiles and reactive oxygen species, is coupled to the activation of another orthogonal editor that introduces edits into barcode D. The behaviour of the mouse would also be recorded for the duration of the experiment. Finally, individual neurons are isolated and barcodes are transcribed by T7 RNA polymerase and then sequenced together with single-cell transcriptomes. Integration of data from the four barcodes and the cellular identity is used to reconstruct the history of each neuron (see the figure, panel **b**). Notably, the individual components of the proposed recorders already exist as reporter genes but would be combined here to record information into DNA.



a single cell's history, in relation to one another as well as to absolute time and regardless of whether the cell is dividing, should be recordable and recoverable if ordered editing is assumed. However, presumably because multiplex biologically conditional editors are still very new, the development of algorithms for analysis of such data remains immature. There are a few datasets available, associated with the DOMINO, CAMERA and ENGRAM methods, that may serve as entry points or inspiration for computational approaches[64,65,78]. Potential framings of the problem and how it might be addressed algorithmically are discussed in the section 'Practical challenges for DNA-based recording'.

Tree decoration. Combining the cell lineage tree with the cell's past trajectory, its extrinsic signals and its spatial context information into a cellular history involves decorating the tree with this information obtained from conditional recorders as well as endpoint measurements. Provided that the constitutive and conditional recording occurs concurrently, the tree would provide guidance on when the signals occurred. For example, if two sister cells (relationship informed by constitutive recording) share a recorded signal (obtained from conditional recording) but no other cells do, then the event recorded may have occurred in the parent of the two sister cells. Endpoint measurements can enrich such conditional recorder data by providing the state of each cell at the end of the experiment.

Rich datasets in which cell lineage, trajectory information, extrinsic signals and spatial context are concurrently recorded are not yet available. At present, decorating a tree relies merely on molecular state information of the analysed cells at the end of the experiment. Two primary methods have thus far been developed for such data: first, methods estimating transition maps between cell states from several time points, incorporating some (but by far not all) lineage information[98–100]; and second, methods that decorate the tree at ancestral nodes. For example, a maximum-parsimony approach has been applied to infer tissue location within tumour phylogenies[8].

Accounting for noise and uncertainty. Phylogenetic reconstruction and tree decoration methods include heuristic, distance-based, parsimony-based methods[18,29,47,94,96], as well as statistical maximum-likelihood and Bayesian methods[90,95]. Major challenges for reconstruction stem from the noise in the data, owing to heterogeneities in reading and writing as well as incompleteness of the recording data. Noisiness for a given cell may stem from silencing or imperfect readout, leading to 'drop out' of recorded information at some sites. Furthermore, recordings for only a subset of cells are likely to be captured and not every division of cells may be recorded (that is, sampling depth is limited). In general, methods can deal with incompleteness by imputing missing data[94,95], by explicitly modelling the

loss of information[90], or by subsampling a set of cells from the tissue[90]. The uncertainty resulting from heterogeneous and incomplete data can be considered by Bayesian phylogenetic approaches or through phylogenetic bootstrapping to obtain confidence estimates.

Heuristic methods[96] are typically fast, enabling estimation of lineage trees on millions of tips. However, the estimates may be rather uncertain or biased owing to the noise in the data and heuristics may fail to converge to global optima. By contrast, statistical methods using an appropriate model can lead to unbiased results with quantification of uncertainty. In particular, Bayesian methods naturally incorporate noise and uncertainty, but are very computationally intensive, only facilitating analysis of a few hundred cells to a couple of thousand cells. As recording becomes complex, the noise and uncertainty in the data will decrease, and non-statistical frameworks may lead to reliable decorated trees. We anticipate that expansion of recording capacity will alleviate some statistical and computational challenges.

Applications of DNA-based recording

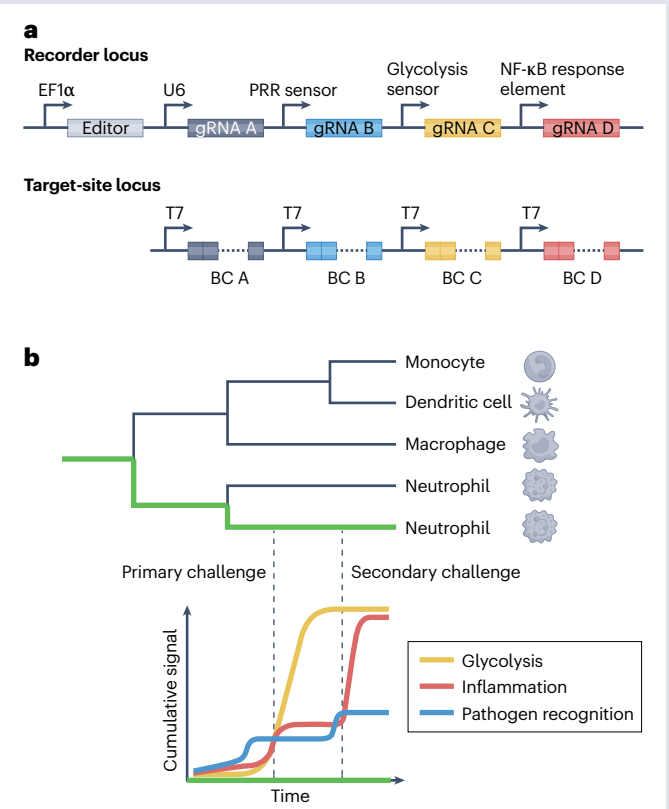
The application of DNA-based recording to generate biological insights remains in its early stages. Our selected examples illustrate the use of DNA-based recording to investigate: 1) the lineage origins of various cell types; 2) the dynamics of clonal dominance in development and disease; 3) the orchestration of cell-fate decisions by intrinsic and extrinsic signals; and 4) non-invasive bacterial monitoring of cellular behaviours and environmental conditions (Fig. 3).

Lineage origins during development and regeneration

A central goal of developmental biology is to elucidate cell lineage relationships and molecular changes during the development of a single-celled zygote into a multicellular organism composed of numerous cell types. Most studies to date have focused on clonal tracing, which defines the descendants of a cell[5–7,101,102]. For example, a major question in blood development and immunology is what kinds of progenitor give rise to diverse blood and immune cell types and how their abundance changes with age. To address this question, an inducible scarring-based barcoding system, CARLIN, was used to label mouse cells in vivo with a unique genetic barcode, which was then captured alongside the transcriptome via scRNA-seq[5]. Prior to this work, the general paradigm was that definitive haematopoietic stem cells (HSCs) derive from pre-HSCs and establish blood production in late fetal and adult mice. However, lineage tracing facilitated the identification of a new progenitor type, embryonic multipotent progenitors (eMPPs), which are also derived from pre-HSCs. Barcoding analysis revealed that eMPPs preferentially contribute to lymphoid lineages and persist lifelong, but that their output decreases with age. Conversely, adult HSCs increase productivity with age but do not compensate for the loss of lymphoid cell production

Box 3 | The potential of DNA-based recording in the immune system

Innate immune cells can remember past encounters with pathogens, enhancing their responsiveness upon re-exposure[200,201]. This phenomenon, termed trained immunity, expands our understanding of innate immune function and its importance in host defence. However, many questions remain regarding the specific immune-cell populations involved, the metabolic changes that occur, such as glycolysis, and the duration of the memory response. These questions could be addressed by a DNA recording system with four barcodes (BCs) at target sites specified by guide RNAs (gRNAs), as illustrated (see the figure, panel **a**). In this example, a DNA recorder constitutively edits barcode A to capture lineage information of the innate immune cells during their specification and differentiation. Another editor with a transcription factor coupled to pattern recognition receptors (PRRs), which detect pathogen-associated molecular patterns (PAMPs), translocates into the nucleus upon activation, where it regulates editing of barcode B[202,203]. Glycolysis, a key metabolic pathway upregulated in trained innate immunity, is monitored with a glycolytic biosensor detecting metabolites such as fructose-1,6-bisphosphate, initiating conditional editing of barcode C[204–207]. In addition, the pro-inflammatory signalling response would be recorded in barcode D under the control of NF-κB response elements[208,209]. Innate immune cells could then be sequenced to recover their cellular identities and recorded histories, providing a comprehensive view of the trained immunity process (see the figure, panel **b**). The reconstructed information would allow researchers to infer the number of pathogen challenges the cells encountered, the metabolic and pro-inflammatory response of the cells, and how these responses differ across cell types. This information would in turn provide insights into the formation of immune memory and the responses of trained innate immune cells upon pathogen re-exposure.



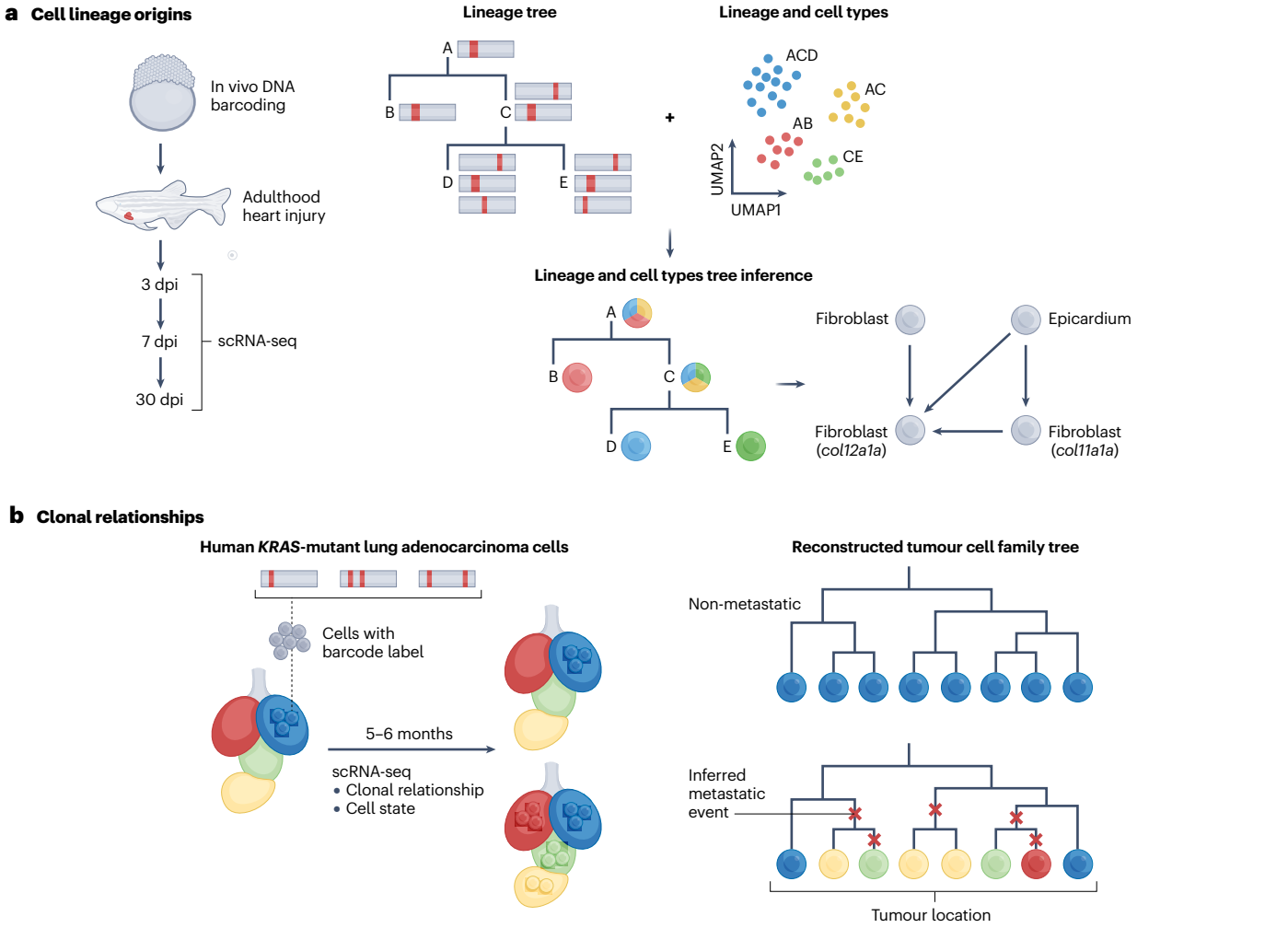

a Cell lineage origins
In vivo DNA barcoding
Adulthood heart injury
3 dpi
7 dpi
30 dpi
scRNA-seq
Lineage tree
A
B
C
D
E
Lineage and cell types
ACD
AC
AB
CE
UMAP2
UMAP1
Lineage and cell types tree inference
Fibroblast
Epicardium
Fibroblast (col12a1a)
Fibroblast (col11a1a)
b Clonal relationships
Human KRAS-mutant lung adenocarcinoma cells
Cells with barcode label
5–6 months
scRNA-seq
• Clonal relationship
• Cell state
Reconstructed tumour cell family tree
Non-metastatic
Inferred metastatic event
Tumour location
c Recording intrinsic and extrinsic experiences
Morphogen-sensitive enhancer
rtTA
Tet-responsive element
Cre
Constitutive promoter
loxP
Stop
loxP
Reporter
Recording morphogen activity in gastruloids at different time points
Early recording window (+dox)
Late recording window (+dox)
Posterior
Anterior
Signalling activity predicts cell location in final structure
d Non-invasive recording of environment
Cas1–Cas2-RT
RNA
DNA
Record-seq E. coli
Oral administration
Passage through gastrointestinal tract
• Diet
• Inflammatory state
• Microbiome
Isolate from faecal samples
RNA-seq

Fig. 3 | Biological insights from in vivo DNA recording. a, CRISPR-based DNA recording traced the origins of regenerated cardiomyocytes following heart injury in zebrafish. Single-cell RNA sequencing (scRNA-seq) at defined days post-injury (dpi) retrieved DNA recordings concurrent with cell states. These data enabled the reconstruction of lineage relationships, identification of cell types, and point to the epicardium as a likely source of *col12a1a*-expressing fibroblasts. **b**, CRISPR-based DNA recording reveals clonal relationships in cancer progression through a xenograft study involving human *KRAS*-mutant lung adenocarcinoma cells transplanted into the left lung of immunodeficient mice. Analysis of dissected and sequenced metastatic cells identifies key genes contributing to metastasis by integrating clonal relationships with spatial cell-state information. **c**, Conceptual diagram of a recombinase-based circuit reporting morphogen activity during gastruloid formation. A morphogen-responsive enhancer drives the expression of rtTA (reverse tetracycline-controlled transactivator). In the presence of doxycycline (dox), rtTA activates a Tet-responsive element, driving expression of Cre recombinase. Cre-mediated recombination at loxP sites removes a stop cassette enabling constitutive expression of a fluorescent reporter. This approach allows the visualization of cells that have experienced morphogen signalling within a defined recording window, revealing spatial patterns of signalling pathways coinciding with gastruloid symmetry breaking and elongation. **d**, Sentinel *Escherichia coli* engineered with Record-seq reverse-transcribe transient mRNA into CRISPR arrays within the gut using Cas1–Cas2-based recording. Analysing faecal samples through deep sequencing and computational methods offers insights into bacterial adaptation to gut conditions such as dietary shifts and inflammation response. CRISPR, clustered regularly interspaced short palindrome repeats; RT, reverse transcription. Part **a** adapted from ref. 7, CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/). Part **b** adapted with permission from ref. 8, AAAS. Part **d** adapted with permission from ref. 12, AAAS.

from eMPPs. These data, obtained by recording, provide a potential explanation for immune decline during ageing.

Another scarring-based method, called LINNEAUS, was used to investigate the origin and functions of key cell states in zebrafish heart regeneration[7]. A key question in regeneration is which cells contribute to the regenerated tissue. Three transient fibroblast cell types induced by heart injury were identified to have pivotal roles in tissue repair. DNA-based recording revealed that two types of fibroblast (expressing *col11a1a* or *col12a1a*) were lineage-related and originated from the heart epicardium, whereas the third type (expressing *nppc*) emerged from the endocardium (Fig. 3a). Depleting *col12a1a*-expressing fibroblasts impeded heart regeneration, underscoring their pro-regenerative function. The study showcased how DNA recording helps to reveal the distinct origin and signatures of cells with regenerative functions.

Clonal dominance in development and cancer

Embryonic progenitors can give rise to clones of drastically different sizes, thanks to a combination of intrinsic differences, extrinsic cues and stochastic processes[103,104]. Recording approaches as well as live imaging have revealed clonal dominance, where a small number of progenitors contribute disproportionately to a specific organ or cell type. For example, the DNA scarring-based method GESTALT was used to barcode embryonic progenitors to trace lineage throughout zebrafish development[9]. Remarkably, while about 20 early embryo cells seemed to give rise to the entire zebrafish blood system, as few as 5 clones explained 98% of the blood lineage in an adult animal. This study illustrates the power of DNA-based recording to reveal patterns of clonality from embryo to adult.

Clonal dominance is also evident in cancer, where evolving clones within tumours acquire specific survival, proliferation and metastatic properties. Classic studies showed intratumoural heterogeneity between primary and distant sites but did not reveal how cancer metastasis plays out within individual clones over time. Lineage tracing using DNA recorders has been recently applied to yield insight into clonal evolution and the role of specific genes in metastasis and cancer progression[8,105–108]. For example, human lung cancer cells, engineered to record their lineage by DNA scarring, were surgically implanted into the lungs of mice[8]. Phylogenetic tree reconstruction was performed for hundreds of clones as they grew and spread over months using the mutated barcodes. Reconstructing deep trees and analysing where related cells were positioned within the lung allowed quantification of metastatic capacity for each clone, which ranged from completely non-metastatic to the aggressive colonization of distant sites (Fig. 3b). This analysis also revealed a metastatic hub in lymph tissue, and complex seeding topologies disseminated clones across the body and even back to primary tumours. Gene-expression differences between clones arose early before transplantation and were mildly predictive of these divergent metastasis phenotypes, except for rare exceptions where some clones developed new potential. Candidate gene perturbations also altered invasiveness phenotypes, validating their functional role in driving metastasis. Overall, this work, as well as several other related studies[105–108], demonstrates the power of DNA recorders to illuminate metastatic progression and heterogeneity.

Intrinsic and extrinsic cellular experiences

A key challenge in developmental biology is to disentangle the effects of cell-intrinsic factors inherited through lineage and cell-extrinsic signals in the local cellular environment on cell-fate decisions. The ability to reconstruct lineage in situ could reveal the relative roles of the two kinds of cue. To this end, the image-readable, recombinase-based recording tool intMEMOIR was used to trace the emergence of fly neuron types from an embryonic progenitor pool[10]. Imaging-recombined barcodes in adult brains revealed that neurons derived from the same embryonic clone exhibited a spatially dependent similarity in terminal cell identity, with nearby cells more similar in cell type than those located further apart. By contrast, cells from different clones showed no relationship between their spatial position and terminal cell identity. These findings highlight the complex way in which lineage and extrinsic cues combine to determine cell fates in a spatially patterned tissue.

Another fundamental question in developmental biology is how symmetries are broken and axes are formed in a stereotyped fashion. Stem-cell-derived embryo models provide experimentally accessible systems to address this question. A recent study used DNA rearrangement-based recorders in the mouse gastruloid model to conditionally record signalling activated by Nodal, Bmp or Wnt morphogens[11]. The DNA rearrangements produced heritable changes in fluorescent protein expression, enabling live imaging of recorded signals during gastruloid development. This approach revealed that Nodal and Bmp exhibited spontaneous heterogeneous activity among cells, which modulated Wnt signalling. Differences in Wnt signalling led to the emergence of distinct Wnt-active and Wnt-inactive cell populations, which differed in their expression of the adhesion molecules that allow sorting into posterior and anterior domains (Fig. 3c). Interestingly, recording at different times revealed that Wnt activity was only weakly predictive of the later spatial position along the anterior–posterior

axis when recorded at 72 hours after gastruloid formation, but became strongly predictive when recorded at 96 or 120 hours after gastruloid formation. These results demonstrate that Wnt activity is a critical player in symmetry breaking, and acts as a predictive factor for future cell positions and fates along the anterior–posterior axis.

Non-invasive recording of the environment

The examples discussed above are focused on capturing lineage relationships among mammalian cells, except for the example in ref. 11, which reports recording or marking of canonical signalling pathways. Another example illustrates recording of an additional non-lineage aspect of cell states, in prokaryotes rather than eukaryotes. Specifically, bacteria were recently engineered to record their own global transcriptional responses during transit through the mouse gut[12]. A chimeric Cas1–Cas2-RT (reverse transcriptase) insertional editor was used to reverse-transcribe cellular mRNA and integrate the resulting DNA into a genomic target site. These engineered bacteria were deployed as 'sentinels', as the recorded transcripts were informative with respect to their interactions with food, host cells and other microbes during gut transit. The recordings were recovered through RNA-seq of samples collected from faeces and revealed bacterial adaptation to nutrient availability, acid stress, inflammation and microorganism–microorganism interactions (Fig. 3d). Although currently limited to *Escherichia coli*, this study and other Cas1–Cas2-based recording systems[81] open up possibilities for non-invasive analysis of complex microbiota physiology and adaptation in vivo.

Practical challenges for DNA-based recording

These initial demonstrations of recording serve as motivating prospects for what could be achievable, and eventually routine, as we develop more powerful recording systems. However, to realize their full potential, the field must address several key technical challenges to writing, reading and reconstructing DNA-based recordings.

Challenges for writing information

Programming conditional recorders. To date, most work in this field has been based on constitutive recorders. Ideally, these would be complemented by a library of conditional recorders that respond to signalling pathway activity, enhancer activity, specific transcripts, intracellular ligands, extracellular ligands, metabolic fluxes, neural activity, mechanical forces and infections, among other biological parameters. In practice, only a few studies have demonstrated conditional recorders, mostly relying on signal-dependent *cis*-regulatory elements (CREs) to drive production of a gRNA or pegRNA, or transcriptional acquisition systems in prokaryotes[11,12,19,78,80,81]. Much work remains to be done to expand and optimize the repertoire of CRE-based recorders, for example, to encompass all major signalling pathways and cell types, while transcriptional acquisition systems beg to be adapted to eukaryotic systems. Furthermore, existing signal-dependent CREs may not report linearly on pathway activation, and concatemerized synthetic reporters might behave differently from endogenous response elements. Furthermore, many biological phenomena occur post-transcriptionally, requiring alternative strategies to activate conditional recorders. For example, dual-component gRNAs can be used to create conditional recorders activated by specific protein–protein interactions[79]. Leveraging de novo designed binders or receptors, or aptamer-driven gRNA activation, could further expand the range of recordable signals. Finally, we note that these early demonstrations of conditional recorders have yet to be coupled to constitutive recorders.

Maintaining bio-orthogonality. Editors, recorders and recording media must maintain bio-orthogonality, that is, they must minimize interference with native biology. Although most cells and organisms can tolerate the extensive addition of genomic content (for example, 34 megabases (Mb) in the TcMAC21 mouse model, a transchromosomic mouse model of Down syndrome that harbours the long arm of human chromosome 21 as a mouse artificial chromosome (MAC)[109]), suggesting the potential for accommodating substantial amounts of recording 'equipment', each biological model may interact differently with any recording components used. The interference can result in low portability (for example, low efficiency of Cas1–Cas2-based editing in mammalian cells), toxicity and off-target effects (for example, Cas9-induced DSBs[50,53] or the Cre recombinase[110]), or disruption of genomic context around the integration sites of recording components. Although disruption of the genomic context can be mitigated using defined 'safe harbour' loci to integrate editors, recorders or media, addressing the former two issues requires thorough testing and validation using the biological model of interest. In the future, de novo designed binder or receptor proteins could potentially be used as sensors and/or signal transduction systems to help to maintain bio-orthogonality[111].

Timescales and capacity. Cellular histories with single-cell-cycle resolution require at least one edit per cell division. A key challenge is stabilizing the edit rate, which can diminish over time owing to factors such as silencing of editor expression and reductions in the number of unedited target sites. Ideally, editors and recorders would reside at safe harbour loci to resist silencing, and the system would possess vast amounts of information capacity or use mechanisms to stabilize the recording rate (for example, with DNA Typewriter, each successive write event both destroys and creates an editable site[76]). Recorders should ideally vary across timescales from capturing sub-cell-cycle events in seconds or minutes (for example, cell-signalling cascades) to monitoring long-term processes over hours, days, months or years (for example, cellular differentiation or disease progression). Since DNA-based recorders rely on DNA repair mechanisms, they are inherently more compatible with the longer timescales, whereas recorders based on other modalities such as RNA or proteins may be better suited to fast timescales.

Spatial information. Current spatial methods capture only a single moment in cellular interactions, lacking insight into the spatial origins of cells. In organs such as the brain and immune system, understanding past spatial proximity is crucial for subsequent development and function. For example, recording spatial proximity could shed light on migration patterns during embryogenesis or reveal anatomical locations and previous cell contact histories of immune cells[112,113]. It could also allow the spatial inference of metastasis from initial tumour to colonization of new tissues. One possible way to record the spatial trajectory of cells is to leverage synthetic signalling pathways. In this approach, 'sender' cells expressing synthetic ligands activate matching synthetic receptors in neighbouring 'receiver' cells, for example, SynNotch, Tango, MESA and BAcTrace[114–123]. SynNotch is a promising approach because it labels neighbouring cells upon contact and could potentially be coupled to CRISPR recorders. However, a challenge lies in ensuring unique labelling of each receiver cell by each sender, requiring numerous orthogonal signalling channels or intercellular genetic material transfer.

Recording across model and non-model organisms. Understanding the evolution of development and what features are conserved

or variable between species remains one of the most fundamental challenges in biology. Generating stable transgenic cells and animals with diverse recording capabilities will allow rapid characterization and comparisons of development and homeostasis across various organisms. However, complex, stably integrated recording systems may be difficult to engineer in most non-model organisms. Creating 'portable' recording systems that can be introduced into diverse animal species using viral vectors could address this limitation. These systems need to be genetically compact yet provide enough memory to record sufficient data to analyse particular developmental processes, and to maintain bio-orthogonality with respect to the native biological processes. The development of larger and more tissue-specific delivery vectors[124–126], as well as artificial chromosomes with large capacity, high mitotic stability and broad host range should help to realize this possibility[127–129].

Challenges for reading DNA records

Missing information. Many of the outstanding technical challenges for recovering DNA-recorded information are touched on above (see section 'Reading DNA-recorded information'), and essentially converge on missing information, that is, a failure to recover all information recorded in each cell (drop out), or to recover histories of all cells from a tissue or organism (sampling depth). The consequences of missing information can be substantial. For example, recovering just 1% of cells from an organism allows the generation of sparse trees that are informative of early lineage relationships, but it is uninformative for the key terminal cell divisions that drive cellular specialization. Taking the adult mouse as an example, recovering DNA records with almost 100% efficiency from nearly all of its approximately 10 billion cells is a task on a scale that cannot currently be achieved, even if the sampling depth and dropout issues were fully solved, owing to limitations inherent in scRNA-seq technology. Therefore, the field needs scRNA-seq methods to advance by several orders of magnitude, spatial transcriptomic methods to advance to a level at which entire animals can be routinely profiled at single-cell resolution, or DNA-based recording to advance to the point where the entire, longitudinal history of a cell can be captured by a single, contiguous DNA sequencing read.

Survivorship bias. Another key challenge related to recovering DNA records is survivorship bias — many if not most cells produced during an organism's life die, resulting in the loss of their records. Moreover, current approaches rely on isolating and destroying cells that contain recorders. One way these issues could be addressed is by engineering cells to export recordings over time through protective nanoparticles. For example, the COURIER system uses RNA export systems based on viral and synthetic components that efficiently package target RNAs into protective nanoparticles secreted from cells[130]. By incorporating RNA barcodes, sampling the exported RNA from culture media, and sequencing these barcodes enables longitudinal tracking of clonal population dynamics and overcomes barriers to accessing RNA from living cells in a non-destructive manner. Furthermore, if each cell produced multiple nanoparticles before division, this amplification could reduce the likelihood of cell dropouts. This approach resolved the expansion and decline of thousands of distinct cell clones over time in response to drug selection. These export tools have versatile applications, including real-time monitoring of biological phenomena for early disease detection and treatment optimization. In the future, increasing the rate at which nanoparticles are generated would help to resolve biological processes that occur at faster timescales.

Challenges for reconstructing histories

Quantifying cell dynamics. An overarching challenge is to accurately quantify variations in cell behaviour and thus to capture cellular dynamics, such as the rates of division, differentiation, apoptosis and migration. Although similar cells tend to exhibit similar dynamics, certain factors such as gene-expression variability among cells or responses to signalling molecules can have a more pronounced influence on cell dynamics than others. The phylodynamic framework should enable such quantification based on a reconstructed decorated tree[131]. As phylodynamic approaches often require in-depth adjustment even within their field of origin — epidemiology — reasonable model assumptions and approximations remain to be explored for biological recording data.

Computational scaling. The computational cost of assembling trees grows exponentially with the number of cells, necessitating more scalable computational approaches to phylodynamic inference. Some approaches developed recently for datasets generated during the COVID-19 pandemic, which contain millions of sequences, may become helpful[132–134]. First, one can use fast heuristics for tree topology estimation and, if appropriate, ignore all topology uncertainty, while using statistical tools that provide an estimation of uncertainty for more uncertain processes like timescales. Second, for phylodynamic parameter estimation, analyses with smaller numbers of cells might be informative, and analyses of smaller subtrees could be merged for overall results. Third, utilizing graphics processing units (GPUs) for calculations and large memory can further facilitate analyses.

Timings of conditional recordings. A conditional recorder captures specific features of a cell's history over time, with the intensity of recording dependent on the abundance of that feature (for example, RNA transcript or signal transduction activity levels). To interpret such data accurately, we need to establish the timing of conditional recordings, otherwise we cannot distinguish between short periods of high signal intensity versus long periods of low signal intensity, nor between the relative timing of different signals in the same system if multiple conditional recorders are present. One strategy to address this issue is to implement a constitutive time recording that runs in parallel to the conditional recording[17]. Lineage tracing through a constitutive recorder would allow conditionally recorded events to be mapped back on a branch in the lineage tree, albeit with a temporal resolution limited by the rate of cell division. However, with recording strategies that explicitly preserve the order of events in writing to DNA and furthermore write constitutive and conditional signals to a shared medium[76,78], duration versus intensity versus order of conditional signal(s) could be disentangled even between cell divisions or in post-mitotic cells.

When analysing such data, it will be necessary to consider whether a given set of cells or recordings share a common history. For sets of cells or recordings without a shared history, such as non-dividing neurons, time-series analysis approaches might be applicable[135,136]. If the cells or recordings share a common history, it will be important to account for this shared lineage to avoid biases[137]. Ideally, one would perform joint analysis of cell lineage and conditionally recorded signals, given that they may be mutually informative. Packer et al.[138] leverage *Caenorhabditis elegans* to provide a compelling early example of how this might be approached, by jointly analysing single-cell transcriptome data layered onto the invariant lineage. For such joint analyses, approaches developed for time-lapse microscopy data can also be considered[139]. However, additional complexities arise with DNA-based recording data. These include the need to reconstruct

rather than directly image the cell lineage tree, the indirect measurement of the timing of features through a separate recorder, and the fact that feature intensity is inferred from the number of recordings within a given time window.

Combining datasets. Complete molecular recording over vast spatiotemporal scales is currently not feasible. However, integrating datasets from different experiments, using data from the same biological entity but different replicates, time periods and/or molecular modalities has the potential to yield a cohesive view of the cellular dynamics governing a system of interest. Combining data from different individuals of the same species may allow us to differentiate deterministic rules from stochastic fluctuations and identify 'historical' molecular and cellular events underlying phenotypic changes due to mutations. It could also facilitate comparative developmental analyses across species, identifying both conserved and species-specific developmental programs. However, the development of methods for combining data across individuals or experiments in a coherent way remains a substantial challenge.

Infrastructure and standards. Accessible computational tools will be essential for advancing these developmental recording techniques, akin to the impact of ready-to-use computational analysis platforms for scRNA-seq or pathogen data analysis. As recording systems proliferate, and data accumulate, it will also be essential to develop public repositories and data standards, analogous to successful platforms in other areas of biology such as NextStrain (https://nextstrain.org/). These resources will enable data sharing, visualization and exploration of DNA-based recording datasets across methods and model organisms.

Glossary

Cell fate
An identity that a cell acquires during its development.

Cellular history
The complete lineage tree, trajectory, spatial context and external stimuli that direct cell fate.

Clone
In developmental biology, a set of cells descending from a single ancestor cell.

Conditional recorder
A recording system whose activity is gated by a specific stimulus.

Constitutive recorder
A recording system that is always active.

Decorated tree
A reconstructed lineage tree together with information on the trajectory, spatial context and external stimuli for the cells in the lineage tree.

DNA-based recording
Recording cellular history into genomic DNA.

Lineage tree
The sequence of ancestor–descendant relationships among cells represented as a tree structure. In a lineage tree, branching events correspond to cell divisions.

Molecular state
Information describing the molecular composition of a cell at a specific point in time, for example, its transcriptome or epigenome.

Trajectory
A chronicle of the changes in a cell's molecular state over time. In the single-cell RNA-sequencing field, this is routinely inferred from a set of the molecular states of a cell, such as in a pseudotime trajectory.

Emerging opportunities

In this section, we ask what exciting possibilities lie ahead. We focus on four domains: (1) high-capacity recording of development and homeostasis; (2) understanding the statistical nature of developmental programmes; (3) causal inference within and across individuals; and (4) engineering recorders to provide sentinel cells in humans.

Dense recording of development and homeostasis

We have yet to come even remotely close to saturating the enormous theoretical capacity for DNA-based information storage in living cells and organisms while maintaining viability. As noted above, mouse models with artificial chromosomes as large as 34 Mb are viable[109], and the upper limit on how much engineered content can be added while maintaining viability has not been seriously explored. As a thought experiment, imagine that about 1% of the mouse genome (25 Mb) were engineered to support the data-storage aspect of biological recording. At a modest density of 2 recorded bits (that is, one base) per 50 base pairs (bp), this configuration would provide about one megabit of storage per cell. Although actual developmental lineage trees tend to be asymmetric, for simplicity, if we assume a perfect binary tree of 40 cell cycles from fertilized zygote to the 10^{10} cells or so that constitute an adult mouse, only a small fraction of this storage capacity is sufficient to completely capture cell lineage relationships. For example, if ordered recording was enabled, and 8 bits were successively set at random at each cell division to distinguish daughter cells from one another ($2^8 = 256$ possibilities), only 320 bits would be required to capture a complete lineage tree from zygote to adult. In this scenario, over 99.9% of that one megabit would remain available to record aspects of biology other than cell lineage via conditional recorders (Fig. 4a).

If an additional 1% (25 Mb) or so were devoted to encoding the recorders themselves (not the target sites), at a modest density of one recorder per 10 kilobases (kb) or so, as many as 2,500 recorders could operate concurrently within each cell. A handful would be constitutive recorders, but the remainder could be conditional on signalling pathways, cell-type markers, epigenetic states, enhancer activities, transcript levels, intracellular and extracellular ligands, metabolic fluxes, neural activity, mechanical forces, infections or other aspects of cell biology. As such, it is possible to imagine a vast array of time-resolved internal recordings recoverable from individual cells. These systems could also be used to record information passing between cells. For example, systems could record cell–cell interactions or neuronal connections. The resulting datasets would realize the vision of a densely decorated lineage tree with reconstructed signal dynamics (Fig. 1c). Furthermore, ordered recording could enable the capture of quantitative dynamics in the absence of cell divisions, for example, in post-mitotic neurons or adult homeostasis, over weeks, months or years. Overall, assuming we exhaust our hypothetical DNA-based storage capacity over the course of the lifetime of an adult mouse composed of around 10^{10} cells, we could in principle capture as much as 2^{53} bits, or more than a petabyte, of information, per individual.

Statistical development

Variable development is the rule; the famous invariance of *C. elegans* lineage is an exception. The robustness with which development unfolds in most multicellular organisms is a consequence of stochastic rather than deterministic processes. However, this has been deeply explored in only a handful of cases (for example, branching morphogenesis[140]). In most contexts, the statistical analysis of development is substantially constrained by our ability to track the state of individual cells at scale.

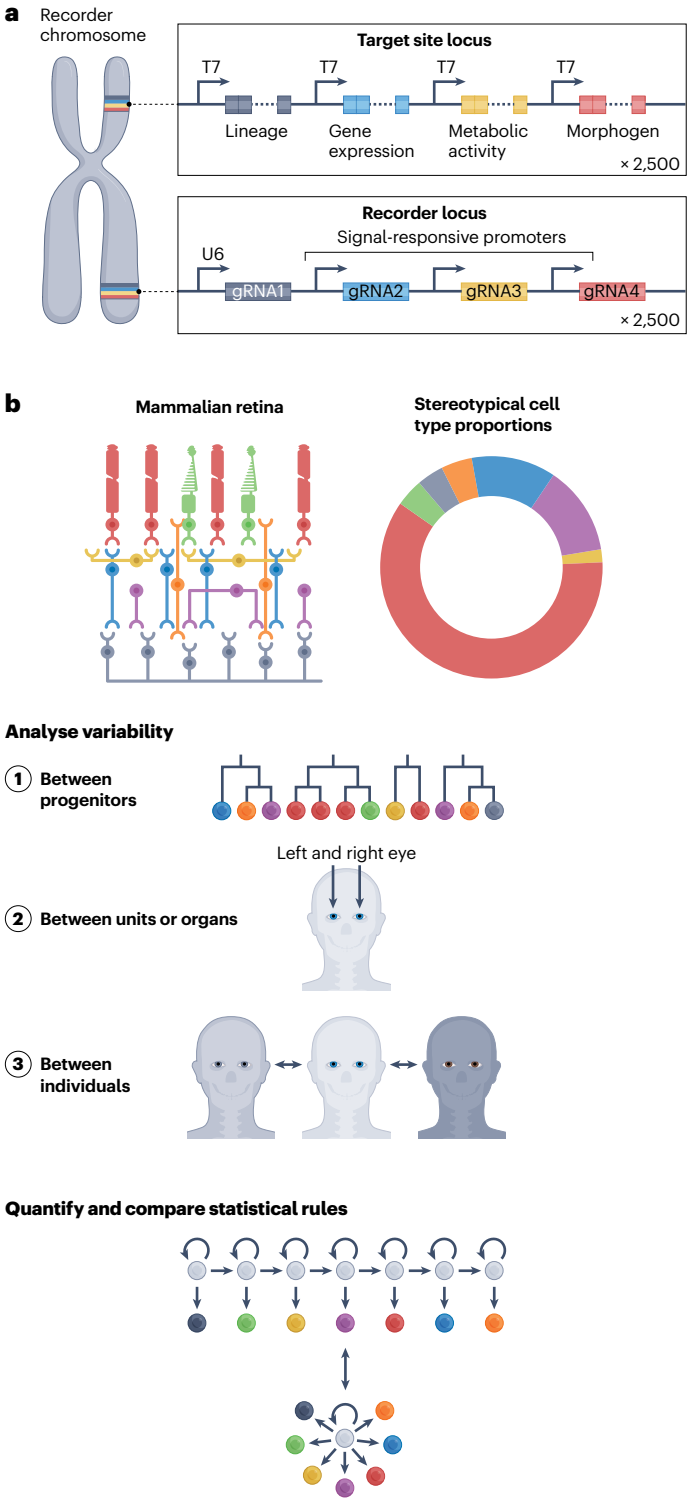


Fig. 4 | Blue-sky opportunities to track development and homeostasis. **a**, A mouse model with around 1% of its genome devoted to encoding recorders that respond to input signals and another 1% or so dedicated for sites targeted by their corresponding recorders would enable as many as 2,500 signals to be concurrently recorded in single cells. **b**, Stereotypical structures such as the mammalian retina consist of a conserved set of cell types in defined ratios, yet how this is reproducibly established during development remains unclear. Cell histories recovered through DNA recordings would help to reveal the mechanisms by which heterogeneous clones give rise to uniform structures, how these vary between homologous structures within the same individual (such as the left and right eye), and how these vary between individuals. Analysing variability at multiple scales should reveal the statistical rules that operate to generate stereotypical fate distributions and tissues. gRNA, guide RNA.

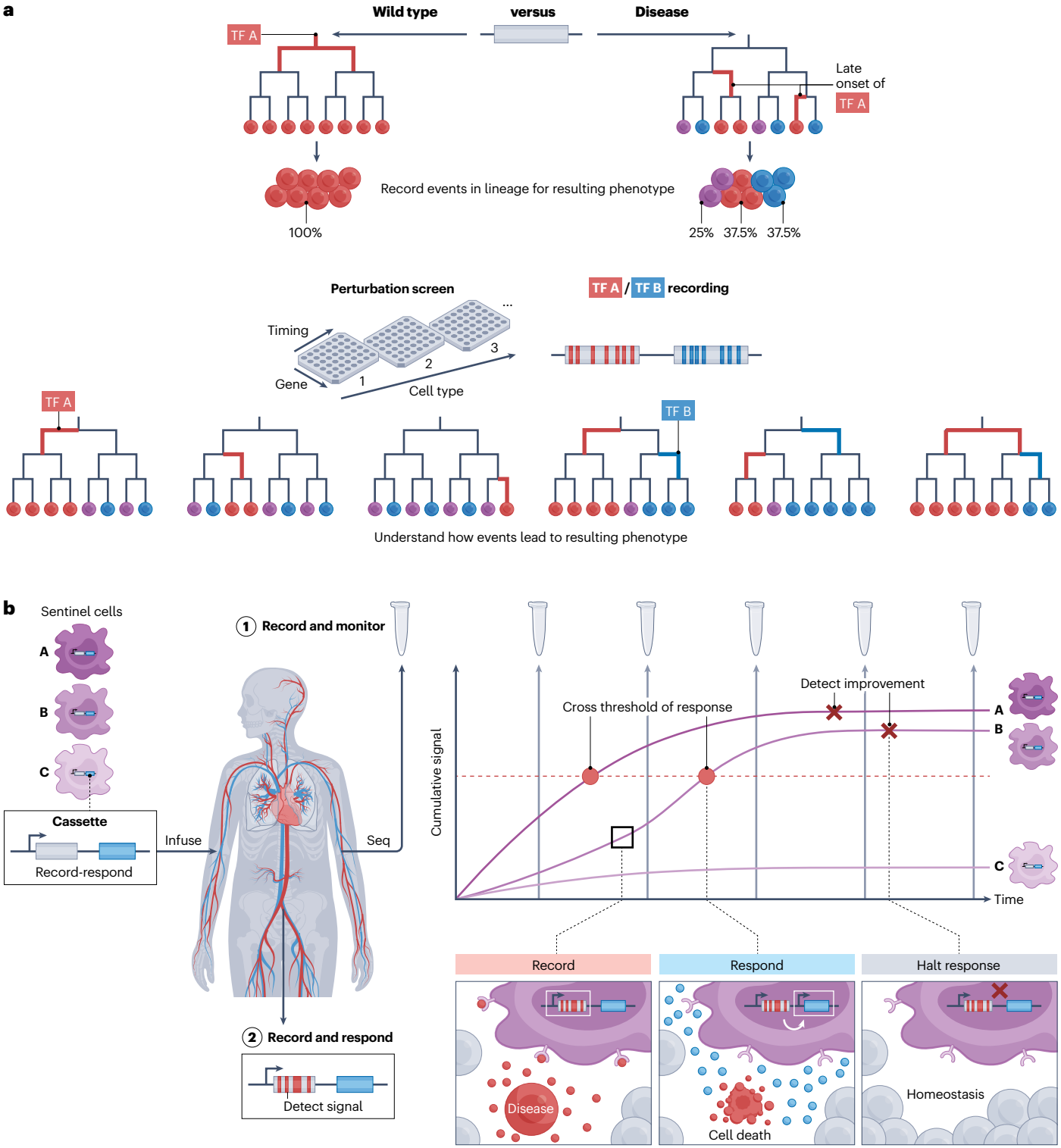
For example, in the vertebrate retina, individual progenitor cells give rise to variable clones, both in terms of size and cell-type composition. How these variable clones together create the stereotypical and reproducible structure of the retina is still an open question. Patterns in the distribution of cell fates on lineage trees can reveal the mixture of fate-restricted progenitors that in turn generate the full distribution of cell fates in the mature retina with the right proportions and spatial organization[141]. Zooming out, reconstruction of decorated lineage trees may provide fundamentally new ways of addressing some of the oldest questions in developmental biology, for example, the statistical and signal-enforced rules by which fate-biased progenitors collectively generate the cell-type distributions required for functional modules (for example, a nephron or neuronal circuit), tissues (for example, a retina) or organs (for example, a liver) (Fig. 4b). Even within a single individual, comprehensive recordings may enable us to distinguish the stochastic versus constrained aspects of such processes (for example, how many ways there are to make a nephron).

Of course, variation will manifest not only within a single individual but also between individuals, and both within genotypes as well as between phenotypes. To what extent can phenomena such as incomplete penetrance and variable expressivity be explained by the variance induced through the statistical rules underlying development? To what extent can common phenotypic variance among humans be understood as a consequence of the same statistical processes? Through the analysis of dense, decorated trees generated from many individuals, both within and across genotypes, it may be possible to extract general principles about how macroscopic phenotypic variance emerges from earlier stochastic events in the context of normal development[142,143].

Causal inference

Although we understand the genetic basis for thousands of Mendelian disorders, our understanding of the mechanisms by which mutations in particular genes give rise to specific phenotypes lags far behind. Developmental disorders are mostly pleiotropic; a multitude of system- and context-specific mechanisms can underlie the path from genotype to phenotype. Historically, the biology of developmental disorders has been studied at a limited number of timepoints, typically after a phenotype has manifested. This approach offers little insight into the early moments of causation that are buried within the developing embryo. For many disorders, combining densely decorated lineage trees from different timepoints and comparing mutant and wild-type trees could potentially pinpoint the molecular and cellular causes of downstream phenotypes (Fig. 5a). For example, shifts in statistical rules governing particular cell types at particular moments might explain how the haploinsufficiency of a gene leads to each pleiotropic aspect of a gross phenotype.

In the future, organism-scale biological recording systems could potentially be combined with synthetic biology approaches to activate user-defined genetic circuits at specific times or within specific sublineages. During development, synthetic circuits could be used



to track cell states and ectopically activate transcription factors to drive desired downstream fates or morphological characteristics. This paradigm could be used to generate novel cell types with hybrid functions: for example, pancreatic alpha or delta cells that, in addition to their normal hormone outputs, secrete insulin, or retinal cells with rod-like sensitivity but with cone-like colour detection. Alternatively, in physiology, neurons could be selectively activated according to their prior activity levels stored through recordings. This would enable analyses of how memory and learning are established in the brain, and reveal to what extent prior activity has an impact on future

Fig. 5 | Blue-sky opportunities for monitoring and altering disease progression. a, Leveraging DNA recorders to elucidate how changes in cell-type proportions arise in diseased states. Here, a synthetic circuit allows recording of transcription factor (TF) A over time and reveals how the late onset of TF A in diseased cells results in aberrant cell-type proportions. Synthetic circuits could also be used in combination with perturbation screens, in which various genes could be perturbed at different time points and/or cell types to understand how specific signalling dynamics result in particular phenotypes. **b**, Sentinel cells engineered with DNA recorders are introduced into patients, where they localize to different organs or tissues and systematically capture diverse molecular, cellular and physiological cues within their spatial environment. These sentinel cells can be programmed to trigger responses upon detection of specific signals to eliminate pathological cells that may evade the immune system. They can further be programmed to halt their responses upon detection of normal homeostasis in recovered, healthy tissues. Subsequent sequencing (seq) of sampled sentinel cells would unveil a comprehensive disease history.

responses. We can also imagine future cellular recording systems that are combined with multiplex in vivo perturbation systems to perturb multiple genes in multiple timeframes in multiple cell types, all within the context of a single animal, followed by recovery and analysis of the molecular and cellular consequences for each of these perturbations of gene, time frame or cell type.

Engineered memory sentinel cells in humans

Recording molecular and physiological states over time in individual people would provide rich datasets that enable early diagnosis, improved treatment of disease and better-informed health decisions, as well as insight into mechanisms and causes of disease. Optical and electronic devices are increasingly being deployed for non-invasive monitoring of blood glucose and some other basic physiological variables. However, most signals in the body, and the variations in those signals across tissues and organs, are inaccessible with such devices. Assuming that safety concerns could be adequately addressed, an alternative paradigm would be to engineer autologous 'memory sentinel cells' that reside within the body at different locations, and passively record a diverse range of molecular and cellular biomarkers (Fig. 5b). These cells could be engineered to allow readout of the signals through reporter systems consisting of secreted information-dense protein or nucleic acids retrieved from urine, blood or stool[130,144]. Memory sentinel cells would thus provide the possibility of surveilling all organ systems, analogous to a human-readable immune system. For example, in slow, degenerative disorders such as Huntington disease or age-related macular degeneration, sentinel cells could store and report longitudinal information, such as extracellular biomarkers and other microenvironmental parameters, alongside tissue-location information, to provide insights into the spatiotemporal mechanisms behind disease progression.

Additionally, memory sentinel cells could be extended to analyse and act upon the information they record (Fig. 5b). Such active memory sentinel cells could perform complex logical operations and enact interventions more precisely than what is possible with current approaches, such as chimeric antigen receptor (CAR) T cells. They could also potentially incorporate mechanisms to address safety concerns, for example, by incorporating engineered tumour suppressors or small-molecule-activated self-destruct circuits to serve as 'kill switches'. Recording could provide two inter-related advantages for such cells. First, it would enable the cell to store large amounts of sensing data in the genome and access those data to make decisions about whether and how to respond in a given context. This could enable targeting of pathological microenvironments defined by complex combinations of factors, ensuring that treatments are delivered only where needed. Second, by recording data over long timescales, a cell could respond not only to the present state of their environment but also to its dynamic history, identifying aspects of the physiological state that are worsening or improving. If the sentinel cells migrate or circulate, this dynamic history could integrate information over space and time. This would allow engineered cells to carry out specific functions in one tissue informed by information recorded in other tissues or organs. More generally, such cells, with complex programmable logic and access to a large memory, could execute programs whose complexity vastly exceeds what is currently envisioned in synthetic biology and cell therapy fields.

Conclusions

Here, we have considered the rationale for biological recording, the state of the art of the underlying technologies, applications of this paradigm so far, various practical challenges that are presently rate-limiting and emerging opportunities. We envision that recent technological advances in DNA-based recording, coupled with overcoming these challenges, will enable the routine recording of cellular histories. This approach will further our understanding of cell decisions over time, given their cellular ancestry and past cell trajectories, and in response to external signals and spatial context. Comparative analyses of such data will enable the deduction of general rules of cell decisions in development and disease. Looking ahead, when recordings can be coupled to responses, such technologies will offer unique opportunities within synthetic biology, in both basic research and in translation to medical applications.

Published online: 25 November 2024

References

1. Woodworth, M. B., Girskis, K. M. & Walsh, C. A. Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat. Rev. Genet.* **18**, 230–244 (2017).
2. Zon, L. I. Intrinsic and extrinsic control of haematopoietic stem-cell self-renewal. *Nature* **453**, 306–313 (2008).
3. Livesey, F. J. & Cepko, C. L. Vertebrate neural cell-fate determination: lessons from the retina. *Nat. Rev. Neurosci.* **2**, 109–118 (2001).
4. Schürch, C. M. et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* **182**, 1341–1359.e19 (2020).
5. Patel, S. H. et al. Lifelong multilineage contribution by embryonic-born blood progenitors. *Nature* **606**, 747–753 (2022).
6. Nusser, A. et al. Developmental dynamics of two bipotent thymic epithelial progenitor types. *Nature* **606**, 165–171 (2022).
7. Hu, B. et al. Origin and function of activated fibroblast states during zebrafish heart regeneration. *Nat. Genet.* **54**, 1227–1237 (2022).
8. Quinn, J. J. et al. Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. *Science* **371**, eabc1944 (2021).
9. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
10. Chow, K.-H. K. et al. Imaging cell lineage with a synthetic digital recording system. *Science* **372**, eabb3099 (2021).
11. McNamara, H. M., Solley, S. C., Adamson, B., Chan, M. M. & Toettcher, J. E. Recording morphogen signals reveals origins of gastruloid symmetry breaking. *Nat. Cell Biol.* https://doi.org/10.1038/s41556-024-01521-9 (2024).
12. Schmidt, F. et al. Noninvasive assessment of gut function using transcriptional recording sentinel cells. *Science* **376**, eabm6038 (2022).
13. Farrell, J. A. et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).
14. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
15. Li, H. et al. Fly Cell Atlas: a single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**, eabk2432 (2022).

16. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* **21**, 410–427 (2020).
17. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
18. Kalhor, R. et al. Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804 (2018).
19. Schmidt, F., Cherepkova, M. Y. & Platt, R. J. Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **562**, 380–385 (2018).
20. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507 (2014).
21. Plesa, C., Sidore, A. M., Lubock, N. B., Zhang, D. & Kosuri, S. Multiplexed gene synthesis in emulsions for exploring protein functional landscapes. *Science* **359**, 343–347 (2018).
22. Pryor, J. M. et al. Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design. *PLoS ONE* **15**, e0238592 (2020).
23. Lau, C.-H., Tin, C. & Suh, Y. CRISPR-based strategies for targeted transgene knock-in and gene correction. *Faculty Rev.* **9**, 20 (2020).
24. Merrick, C. A., Zhao, J. & Rosser, S. J. Serine integrases: advancing synthetic biology. *ACS Synth. Biol.* **7**, 299–310 (2018).
25. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
26. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
27. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
28. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
29. Raj, B. et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
30. Askary, A. et al. In situ readout of DNA barcodes and single base edits facilitated by in vitro transcription. *Nat. Biotechnol.* **38**, 66–75 (2020).
31. Bouckaert, R. et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
32. Qian, Y. et al. Programmable RNA sensing for cell monitoring and manipulation. *Nature* **610**, 713–721 (2022).
33. Kaseniit, K. E. et al. Modular, programmable RNA sensing using ADAR editing in living cells. *Nat. Biotechnol.* **41**, 482–487 (2023).
34. Jiang, K. et al. Programmable eukaryotic protein synthesis with RNA sensors by harnessing ADAR. *Nat. Biotechnol.* **41**, 698–707 (2023).
35. Lin, D. et al. Time-tagged ticker tapes for intracellular recordings. *Nat. Biotechnol.* **41**, 631–639 (2023).
36. Linghu, C. et al. Recording of cellular physiological histories along optically readable self-assembling protein chains. *Nat. Biotechnol.* **41**, 640–651 (2023).
37. VanHorn, S. & Morris, S. A. Next-generation lineage tracing and fate mapping to interrogate development. *Dev. Cell* **56**, 7–21 (2021).
38. Salvador-Martínez, I., Grillo, M., Averof, M. & Telford, M. J. Is it possible to reconstruct an accurate cell lineage using CRISPR recorders? *eLife* **8**, e40292 (2019).
39. McKenna, A. & Gagnon, J. A. Recording development with single cell dynamic lineage tracing. *Development* **146**, dev169730 (2019).
40. Farzadfard, F. & Lu, T. K. Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* **346**, 1256272 (2014).
41. Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S. & Lu, T. K. Synthetic recombinase-based state machines in living cells. *Science* **353**, aad8559 (2016).
42. Livet, J. et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
43. Pei, W. et al. Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456–460 (2017).
44. Guo, F., Gopaul, D. N. & van Duyne, G. D. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**, 40–46 (1997).
45. Stark, W. M., Boocock, M. R. & Sherratt, D. J. Catalysis by site-specific recombinases. *Trends Genet.* **8**, 432–439 (1992).
46. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018).
47. Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
48. Chan, M. M. et al. Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).
49. Bowling, S. et al. An engineered CRISPR–Cas9 mouse line for simultaneous readout of lineage histories and gene expression profiles in single cells. *Cell* **181**, 1410–1422.e27 (2020).
50. Leibowitz, M. L. et al. Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat. Genet.* **53**, 895–905 (2021).
51. Takasugi, P. R. et al. Orthogonal CRISPR–Cas tools for genome editing, inhibition, and CRISPR recording in zebrafish embryos. *Genetics* **220**, iyab196 (2022).
52. Loveless, T. B. et al. Lineage tracing and analog recording in mammalian cells by single-site DNA writing. *Nat. Chem. Biol.* **17**, 739–747 (2021).
53. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
54. Perli, S. D., Cui, C. H. & Lu, T. K. Continuous genetic recording with self-targeting CRISPR–Cas in human cells. *Science* **353**, aag0511 (2016).
55. Kalhor, R., Mali, P. & Church, G. M. Rapidly evolving homing CRISPR barcodes. *Nat. Methods* **14**, 195–200 (2017).
56. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
57. Zhang, X. et al. Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat. Biotechnol.* **38**, 856–860 (2020).
58. Liu, K. et al. Mapping single-cell-resolution cell phylogeny reveals cell population dynamics during organ development. *Nat. Methods* **18**, 1506–1514 (2021).
59. Hwang, B. et al. Lineage tracing using a Cas9–deaminase barcoding system targeting endogenous L1 elements. *Nat. Commun.* **10**, 1234 (2019).
60. Tong, H. et al. Programmable A-to-Y base editing by fusing an adenine base editor with an *N*-methylpurine DNA glycosylase. *Nat. Biotechnol.* **41**, 1080–1084 (2023).
61. Chen, L. et al. Adenine transversion editors enable precise, efficient A•T-to-C•G base editing in mammalian cells and embryos. *Nat. Biotechnol.* **42**, 638–650 (2023).
62. Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9–cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
63. Chadly, D. M. et al. Reconstructing cell histories in space with image-readable base editor recording. Preprint at *bioRxiv* https://doi.org/10.1101/2024.01.03.573434 (2024).
64. Tang, W. & Liu, D. R. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**, eaap8992 (2018).
65. Farzadfard, F. et al. Single-nucleotide-resolution computing and memory in living cells. *Mol. Cell* **75**, 769–780.e4 (2019).
66. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
67. Wienert, B. et al. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* **364**, 286–289 (2019).
68. Kim, D., Kim, D.-E., Lee, G., Cho, S.-I. & Kim, J.-S. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nat. Biotechnol.* **37**, 430–435 (2019).
69. Grünewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
70. Li, J. et al. Structure-guided engineering of adenine base editor with minimized RNA off-targeting activity. *Nat. Commun.* **12**, 2287 (2021).
71. Rees, H. A., Wilson, C., Doman, J. L. & Liu, D. R. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci. Adv.* **5**, eaax5717 (2019).
72. Li, A. et al. Cytosine base editing systems with minimized off-target effect and molecular size. *Nat. Commun.* **13**, 4531 (2022).
73. Yu, Y. et al. Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat. Commun.* **11**, 2052 (2020).
74. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
75. Chen, P. J. & Liu, D. R. Prime editing for precise and highly versatile genome manipulation. *Nat. Rev. Genet.* **24**, 161–177 (2023).
76. Choi, J. et al. A time-resolved, multi-symbol molecular recorder via sequential genome editing. *Nature* **608**, 98–107 (2022).
77. Loveless, T. B. et al. Open-ended molecular recording of sequential cellular events into DNA. Preprint at *bioRxiv* https://doi.org/10.1101/2021.11.05.467507 (2021).
78. Chen, W. et al. Symbolic recording of signalling and *cis*-regulatory element activity to DNA. *Nature* **632**, 1073–1081 (2024).
79. Choi, J., Chen, W., Liao, H., Li, X. & Shendure, J. A molecular proximity sensor based on an engineered, dual-component guide RNA. Preprint at *eLife* https://doi.org/10.7554/eLife.98110.1 (2024).
80. Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652.e29 (2021).
81. Bhattarai-Kline, S. et al. Recording gene expression order in DNA by CRISPR addition of retron barcodes. *Nature* **608**, 217–225 (2022).
82. Sheth, R. U., Yim, S. S., Wu, F. L. & Wang, H. H. Multiplex recording of cellular events over time on CRISPR biological tape. *Science* **358**, 1457–1461 (2017).
83. Tanna, T., Schmidt, F., Cherepkova, M. Y., Okoniewski, M. & Platt, R. J. Recording transcriptional histories using Record-seq. *Nat. Protoc.* **15**, 513–539 (2020).
84. Li, L. et al. A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. *Cell* **186**, 5183–5199.e22 (2023).
85. Chen, F. et al. Phylogenetic comparative analysis of single-cell transcriptomes reveals constrained accumulation of gene expression heterogeneity during clonal expansion. *Mol. Biol. Evol.* **40**, msad113 (2023).
86. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360 (2014).
87. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
88. Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235 (2019).
89. Bressan, D., Battistoni, G. & Hannon, G. J. The dawn of spatial omics. *Science* **381**, eabq4964 (2023).

90. Seidel, S. & Stadler, T. TiDeTree: a Bayesian phylogenetic framework to estimate single-cell trees and population dynamic parameters from genetic lineage tracing data. *Proc. Biol. Sci.* **289**, 20221844 (2022).
91. Feng, J. et al. Estimation of cell lineage trees by maximum-likelihood phylogenetics. *Ann. Appl. Stat.* **15**, 343–362 (2021).
92. Sashittal, P., Schmidt, H., Chan, M. & Raphael, B. J. Startle: a star homoplasy approach for CRISPR–Cas9 lineage tracing. *Cell Syst.* **14**, 1113–1121.e9 (2023).
93. Gong, W. et al. Benchmarked approaches for reconstruction of in vitro cell lineages and in silico models of *C. elegans* and *M. musculus* developmental trees. *Cell Syst.* **12**, 810–826.e4 (2021).
94. Jones, M. G. et al. Inference of single-cell phylogenies from lineage tracing data using Cassiopeia. *Genome Biol.* **21**, 92 (2020).
95. Fang, W. et al. Quantitative fate mapping: a general framework for analyzing progenitor state dynamics via retrospective lineage barcoding. *Cell* **185**, 4604–4620.e32 (2022).
96. Konno, N. et al. Deep distributed computing to reconstruct extremely large lineage trees. *Nat. Biotechnol.* **40**, 566–575 (2022).
97. Volz, E. M., Koelle, K. & Bedford, T. Viral phylodynamics. *PLoS Comput. Biol.* **9**, e1002947 (2013).
98. Hormoz, S. et al. Inferring cell-state transition dynamics from lineage trees and endpoint single-cell measurements. *Cell Syst.* **3**, 419–433.e8 (2016).
99. Wang, S. W., Herriges, M. J., Hurley, K., Kotton, D. N. & Klein, A. M. CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information. *Nat. Biotechnol.* **40**, 1066–1074 (2022).
100. Lange, M. et al. Mapping lineage-traced cells across time points with moslin. Preprint at *bioRxiv* https://doi.org/10.1101/2023.04.14.536867 (2023).
101. Iwanami, N., Petersen, M., Diekhoff, D. & Boehm, T. Clonal dynamics underlying the skewed CD4/CD8 ratio of mouse thymocytes revealed by TCR-independent barcoding. *Commun. Biol.* **5**, 911 (2022).
102. Bolondi, A. et al. Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids. *Dev. Cell* **59**, 1489–1505.e14 (2024).
103. Turner, D. L. & Cepko, C. L. A common progenitor for neurons and glia persists in rat retina late in development. *Nature* **328**, 131–136 (1987).
104. Turner, D. L., Snyder, E. Y. & Cepko, C. L. Lineage-independent determination of cell type in the embryonic mouse retina. *Neuron* **4**, 833–845 (1990).
105. Yang, D. et al. Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell* **185**, 1905–1923.e25 (2022).
106. Simeonov, K. P. et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell* **39**, 1150–1162.e9 (2021).
107. Zhang, W. et al. The bone microenvironment invigorates metastatic seeds for further dissemination. *Cell* **184**, 2471–2486.e20 (2021).
108. Islam, M. et al. Temporal recording of mammalian development and precancer. Preprint at *bioRxiv* https://doi.org/10.1101/2023.12.18.572260 (2023).
109. Kazuki, Y. et al. A non-mosaic transchromosomic mouse model of Down syndrome carrying the long arm of human chromosome 21. *eLife* **9**, e56223 (2020).
110. Silver, D. P. & Livingston, D. M. Self-excising retroviral vectors encoding the Cre recombinase overcome Cre-mediated cellular toxicity. *Mol. Cell* **8**, 233–243 (2001).
111. Quijano-Rubio, A. et al. De novo design of modular and tunable protein biosensors. *Nature* **591**, 482–487 (2021).
112. Pasqual, G. et al. Monitoring T cell–dendritic cell interactions in vivo by intercellular enzymatic labelling. *Nature* **553**, 496–500 (2018).
113. Nakandakari-Higa, S. et al. Universal recording of immune cell interactions in vivo. *Nature* **627**, 399–406 (2024).
114. Roybal, K. T. et al. Precision tumor recognition by T cells with combinatorial antigen-sensing circuits. *Cell* **164**, 770–779 (2016).
115. Roybal, K. T. et al. Engineering T cells with customized therapeutic response programs using synthetic Notch receptors. *Cell* **167**, 419–432.e16 (2016).
116. Huang, H. et al. Cell–cell contact-induced gene editing/activation in mammalian cells using a synNotch-CRISPR/Cas9 system. *Protein Cell* **11**, 299–303 (2020).
117. Zhang, S. et al. Monitoring of cell–cell communication and contact history in mammals. *Science* **378**, eabo5503 (2022).
118. Barnea, G. et al. The genetic design of signaling cascades to record receptor activation. *Proc. Natl Acad. Sci. USA* **105**, 64–69 (2008).
119. Lee, D. et al. Temporally precise labeling and control of neuromodulatory circuits in the mammalian brain. *Nat. Methods* **14**, 495–503 (2017).
120. Daringer, N. M., Dudek, R. M., Schwarz, K. A. & Leonard, J. N. Modular extracellular sensor architecture for engineering mammalian cell-based devices. *ACS Synth. Biol.* **3**, 892–902 (2014).
121. Schwarz, K. A., Daringer, N. M., Dolberg, T. B. & Leonard, J. N. Rewiring human cellular input–output using modular extracellular sensors. *Nat. Chem. Biol.* **13**, 202–209 (2016).
122. Dolberg, T. B. et al. Computation-guided optimization of split protein systems. *Nat. Chem. Biol.* **17**, 531–539 (2021).
123. Cachero, S. et al. BAcTrace, a tool for retrograde tracing of neuronal circuits in *Drosophila*. *Nat. Methods* **17**, 1254–1261 (2020).
124. Goertsen, D. et al. AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. *Nat. Neurosci.* **25**, 106–115 (2022).
125. Chan, K. Y. et al. Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nat. Neurosci.* **20**, 1172–1179 (2017).
126. Nyberg, W. A. et al. An evolved AAV variant enables efficient genetic engineering of murine T cells. *Cell* **186**, 446–460.e19 (2023).
127. Kim, J.-H. et al. Human artificial chromosome (HAC) vector with a conditional centromere for correction of genetic deficiencies in human cells. *Proc. Natl Acad. Sci. USA* **108**, 20048–20053 (2011).
128. Kazuki, Y. et al. Refined human artificial chromosome vectors for gene therapy and animal transgenesis. *Gene Ther.* **18**, 384–393 (2011).
129. Lee, N. C. O. et al. Method to assemble genomic DNA fragments or genes on human artificial chromosome with regulated kinetochore using a multi-integrase system. *ACS Synth. Biol.* **7**, 63–74 (2018).
130. Horns, F. et al. Engineering RNA export for measurement and manipulation of living cells. *Cell* **186**, 3642–3658.e32 (2023).
131. Grenfell, B. T. et al. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332 (2004).
132. De Maio, N. et al. Maximum likelihood pandemic-scale phylogenetics. *Nat. Genet.* **55**, 746–752 (2023).
133. Ye, C. et al. matOptimize: a parallel tree optimization method enables online phylogenetics for SARS-CoV-2. *Bioinformatics* **38**, 3734–3740 (2022).
134. Forster, P., Forster, L., Renfrew, C. & Forster, M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc. Natl Acad. Sci. USA* **117**, 9241–9243 (2020).
135. Chatfield, C. *The Analysis of Time Series: An Introduction* 6th edn (CRC Press, 2016).
136. Chatfield, C. *The Analysis of Time Series* (Chapman and Hall/CRC, 2003).
137. Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985).
138. Packer, J. S. et al. A lineage-resolved molecular atlas of embryogenesis at single-cell resolution. *Science* **365**, eaax1971 (2019).
139. Hilsenbeck, O. et al. Software tools for single-cell tracking and quantification of cellular and molecular properties. *Nat. Biotechnol.* **34**, 703–706 (2016).
140. Hannezo, E. et al. A unifying theory of branching morphogenesis. *Cell* **171**, 242–255.e27 (2017).
141. Tran, M., Askary, A. & Elowitz, M. B. Lineage motifs as developmental modules for control of cell type proportions. *Dev. Cell* **59**, 812–826.e3 (2024).
142. Eldar, A. et al. Partial penetrance facilitates developmental evolution in bacteria. *Nature* **460**, 510–514 (2009).
143. Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–918 (2010).
144. Lee, S. et al. Engineered serum markers for non-invasive monitoring of gene expression in the brain. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-023-02087-x (2024).
145. Whitman, C. O. A contribution to the history of the germ-layers in Clepsine. *J. Morphol.* **1**, 105–182 (1887).
146. Conklin, E. G. The embryology of crepidula, A contribution to the cell lineage and early development of some marine gasteropods. *J. Morphol.* **13**, 1–226 (1897).
147. Wilson, E. B. The cell-lineage of Nereis. A contribution to the cytogeny of the annelid body. *J. Morphol.* **6**, 361–480 (1892).
148. Conklin, E. G. *The Organization and Cell-Lineage of the Ascidian Egg* (Academy of Natural Sciences, 1905).
149. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
150. Vogt, W. Gestaltungsanalyse am Amphibienkeim mit Örtlicher Vitalfärbung: II. Teil. Gastrulation und Mesodermbildung bei Urodelen und Anuren. *Wilhelm Roux Arch. Entwickl. Mech. Org.* **120**, 384–706 (1929).
151. Weisblat, D. A., Sawyer, R. T. & Stent, G. S. Cell lineage analysis by intracellular injection of a tracer enzyme. *Science* **202**, 1295–1298 (1978).
152. Bałakier, H. & Pedersen, R. A. Allocation of cells to inner cell mass and trophectoderm lineages in preimplantation mouse embryos. *Dev. Biol.* **90**, 352–362 (1982).
153. Le Douarin, N. M. The ontogeny of the neural crest in avian embryo chimaeras. *Nature* **286**, 663–669 (1980).
154. Tarkowski, A. K. Mouse chimaeras developed from fused eggs. *Nature* **190**, 857–860 (1961).
155. Mintz, B. Genetic mosaicism in adult mice of quadriparental lineage. *Science* **148**, 1232–1233 (1965).
156. Holt, C. E., Garlick, N. & Cornel, E. Lipofection of cDNAs in the embryonic vertebrate central nervous system. *Neuron* **4**, 203–214 (1990).
157. Price, J., Turner, D. & Cepko, C. Lineage analysis in the vertebrate nervous system by retrovirus-mediated gene transfer. *Proc. Natl Acad. Sci. USA* **84**, 156–160 (1987).
158. Lemischka, I. R., Raulet, D. H. & Mulligan, R. C. Developmental potential and dynamic behavior of hematopoietic stem cells. *Cell* **45**, 917–927 (1986).
159. Holland, E. C. & Varmus, H. E. Basic fibroblast growth factor induces cell migration and proliferation after glia-specific gene transfer in mice. *Proc. Natl Acad. Sci. USA* **95**, 1218–1223 (1998).
160. Temple, S. Division and differentiation of isolated CNS blast cells in microculture. *Nature* **340**, 471–473 (1989).
161. Harrison, D. A. & Perrimon, N. Simple and efficient generation of marked clones in *Drosophila*. *Curr. Biol.* **3**, 424–433 (1993).
162. Lakso, M. et al. Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc. Natl Acad. Sci. USA* **89**, 6232–6236 (1992).
163. Orban, P. C., Chui, D. & Marth, J. D. Tissue- and site-specific DNA recombination in transgenic mice. *Proc. Natl Acad. Sci. USA* **89**, 6861–6865 (1992).
164. Nowak, J. A., Polak, L., Pasolli, H. A. & Fuchs, E. Hair follicle stem cells are specified and function in early skin morphogenesis. *Cell Stem Cell* **3**, 33–43 (2008).

165. Yang, Z., Ding, K., Pan, L., Deng, M. & Gan, L. Math5 determines the competence state of retinal ganglion cell progenitors. *Dev. Biol.* **264**, 240–254 (2003).
166. Matsuoka, T. et al. Neural crest origins of the neck and shoulder. *Nature* **436**, 347–355 (2005).
167. Liu, C. et al. Mosaic analysis with double markers reveals tumor cell of origin in glioma. *Cell* **146**, 209–221 (2011).
168. Snippert, H. J. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010).
169. Ahn, S. & Joyner, A. L. Dynamic changes in the response of cells to positive hedgehog signaling during mouse limb patterning. *Cell* **118**, 505–516 (2004).
170. Harfe, B. D. et al. Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118**, 517–528 (2004).
171. Ahn, S. & Joyner, A. L. In vivo analysis of quiescent adult neural stem cells responding to Sonic hedgehog. *Nature* **437**, 894–897 (2005).
172. Guenthner, C. J., Miyamichi, K., Yang, H. H., Heller, H. C. & Luo, L. Permanent genetic access to transiently active neurons via TRAP: targeted recombination in active populations. *Neuron* **78**, 773–784 (2013).
173. DeNardo, L. A. et al. Temporal evolution of cortical ensembles promoting remote memory retrieval. *Nat. Neurosci.* **22**, 460–469 (2019).
174. Chen, T.-W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
175. Dana, H. et al. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat. Methods* **16**, 649–657 (2019).
176. Choi, J.-H. et al. Interregional synaptic maps among engram cells underlie memory formation. *Science* **360**, 430–435 (2018).
177. Macpherson, L. J. et al. Dynamic labelling of neural connections in multiple colours by trans-synaptic fluorescence complementation. *Nat. Commun.* **6**, 10024 (2015).
178. Michael, N. et al. Effects of sequence and structure on the hypermutability of immunoglobulin genes. *Immunity* **16**, 123–134 (2002).
179. Kim, K.-M. & Shibata, D. Tracing ancestry with methylation patterns: most crypts appear distantly related in normal adult human colon. *BMC Gastroenterol.* **4**, 8 (2004).
180. Salipante, S. J. & Horwitz, M. S. Phylogenetic fate mapping. *Proc. Natl Acad. Sci. USA* **103**, 5448–5453 (2006).
181. Leung, M. L. et al. Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. *Genome Res.* **27**, 1287–1299 (2017).
182. Lodato, M. A. et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).
183. Ludwig, L. S. et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339.e22 (2019).
184. Spencer Chapman, M. et al. Lineage tracing of human development through somatic mutations. *Nature* **595**, 85–90 (2021).
185. Xu, J. et al. Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. *eLife* **8**, e45105 (2019).
186. Salehi, S. et al. Clonal fitness inferred from time-series modelling of single-cell cancer genomes. *Nature* **595**, 585–590 (2021).
187. Evrony, G. D. et al. Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59 (2015).
188. Ju, Y. S. et al. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature* **543**, 714–718 (2017).
189. Walsh, C. & Cepko, C. L. Widespread dispersion of neuronal clones across functional regions of the cerebral cortex. *Science* **255**, 434–440 (1992).
190. Schepers, K. et al. Dissecting T cell lineage relationships by cellular barcoding. *J. Exp. Med.* **205**, 2309–2318 (2008).
191. Naik, S. H. et al. Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* **496**, 229–232 (2013).
192. Pei, W. et al. Resolving fates and single-cell transcriptomes of hematopoietic stem cell clones by PolyloxExpress barcoding. *Cell Stem Cell* **27**, 383–395.e8 (2020).
193. Lu, R., Neff, N. F., Quake, S. R. & Weissman, I. L. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat. Biotechnol.* **29**, 928–933 (2011).
194. Sun, J. et al. Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327 (2014).
195. Ceresa, D. et al. Early clonal extinction in glioblastoma progression revealed by genetic barcoding. *Cancer Cell* **41**, 1466–1479.e9 (2023).
196. Bandler, R. C. et al. Single-cell delineation of lineage and genetic identity in the mouse brain. *Nature* **601**, 404–409 (2022).
197. Delgado, R. N. et al. Individual human cortical progenitors can produce excitatory and inhibitory neurons. *Nature* **601**, 397–403 (2022).
198. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, eaaw3381 (2020).
199. Biddy, B. A. et al. Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 (2018).
200. Bowdish, D. M. E., Loffredo, M. S., Mukhopadhyay, S., Mantovani, A. & Gordon, S. Macrophage receptors implicated in the 'adaptive' form of innate immunity. *Microbes Infect.* **9**, 1680–1687 (2007).
201. Netea, P. D. M. G. I. IDC key-note lecture: trained immunity: a memory for innate host defense. *J. Stem Cell Regen. Med.* **19**, 37–39 (2023).
202. Kutikhin, A. G. & Yuzhalin, A. E. *Pattern Recognition Receptors and Cancer* (Frontiers Media, 2015).
203. Quintin, J. et al. *Candida albicans* infection affords protection against reinfection via functional reprogramming of monocytes. *Cell Host Microbe* **12**, 223–232 (2012).
204. Monteiro, F. et al. Measuring glycolytic flux in single yeast cells with an orthogonal synthetic biosensor. *Mol. Syst. Biol.* **15**, e9071 (2019).
205. Koberstein, J. N. et al. Monitoring glycolytic dynamics in single cells using a fluorescent biosensor for fructose 1,6-bisphosphate. *Proc. Natl Acad. Sci. USA* **119**, e2204407119 (2022).
206. Arts, R. J. W. et al. Glutaminolysis and fumarate accumulation integrate immunometabolic and epigenetic programs in trained immunity. *Cell Metab.* **24**, 807–819 (2016).
207. Ortega, A. D. et al. A synthetic RNA-based biosensor for fructose-1,6-bisphosphate that reports glycolytic flux. *Cell Chem. Biol.* **28**, 1554–1568.e8 (2021).
208. Liu, T., Zhang, L., Joo, D. & Sun, S. C. NF-κB signaling in inflammation. *Signal Transduct. Target. Ther.* **2**, 17023 (2017).
209. Carlsen, H., Moskaug, J. Ø., Fromm, S. H. & Blomhoff, R. In vivo imaging of NF-κB activity. *J. Immunol.* **168**, 1441–1446 (2002).

Acknowledgements

The authors thank all members of the Allen Discovery Center for Cell Lineage Tracing, past and present, for valuable discussions over the course of the past six years. We also thank O. Oseth for extensive assistance in coordinating this collaborative writing project. This work was supported by the Paul G. Allen Frontiers Foundation (J.S., M.B.E. and A.F.S.), the National Eye Institute (R00EY031782 to A.A.), the National Institute of General Medical Sciences (R35GM142950 to J.A.G.), the UCLA BSCRC Transformative Technology Development Award (A.A.) and The Rose Hills Foundation (A.A.). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation programme (grant agreement number 101001077 to T.S. and 834788 to A.F.S.). M.E. and J.S. are Investigators of the Howard Hughes Medical Institute.

Author contributions

The authors contributed equally to all aspects of the article.

Competing interests

A.A., W.C., J.C., M.B.E. and J.S. have patents related to DNA-based molecular recording. J.S. is a scientific advisory board member, consultant and/or co-founder of Cajal Neuroscience, Guardant Health, Maze Therapeutics, Camp4 Therapeutics, Phase Genomics, Adaptive Biotechnologies, Scale Biosciences, Sixth Street Capital, Prime Medicine, Somite Therapeutics and Pacific Biosciences. M.B.E. is a scientific advisory board member, consultant and/or co-founder of Primordium Labs, TeraCyte, Spatial Genomics, and Asymptote Genetic Medicines. The other authors declare no competing interests.

Additional information

Peer review information *Nature Reviews Genetics* thanks Reza Kalhor, who co-reviewed with Weixiang Fang; Jan Philipp Junker; and Harris H. Wang for their contribution to the peer review of this work.