nature methods

REVIEW ARTICLE
https://doi.org/10.1038/s41592-018-0185-x

Cellular barcoding: lineage tracing, screening and beyond

Justus M. Kebschull[1,2] and Anthony M. Zador[2]*

Cellular barcoding is a technique in which individual cells are labeled with unique nucleic acid sequences, termed barcodes, so that they can be tracked through space and time. Cellular barcoding can be used to track millions of cells in parallel, and thus is an efficient approach for investigating heterogeneous populations of cells. Over the past 25 years, cellular barcoding has been used for fate mapping, lineage tracing and high-throughput screening, and has led to important insights into developmental biology and gene function. Driven by plummeting sequencing costs and the power of synthetic biology, barcoding is now expanding beyond traditional applications and into diverse fields such as neuroanatomy and the recording of cellular activity. In this review, we discuss the fundamental principles of cellular barcoding, including the underlying mathematics, and its applications in both new and established fields.

Biological systems consist of collections of heterogeneous cells with unique histories and developmental trajectories that act together to produce complex emergent phenotypes. For example, the concerted action of billions of individual neurons allows people to think, feel, remember and act. Averaging over this cellular diversity can obscure key insights. Although the advent of single-cell RNA sequencing (scRNA-seq) technology has made it possible to routinely obtain a snapshot of the transcriptome of thousands of single cells, it remains challenging to track individual cells over space and time with similar throughput.

First developed 25 years ago, cellular barcoding has emerged as an efficient strategy for tracking large numbers of cells through space, time and cell divisions. Its recent success has been fueled in part by breakthroughs in sequencing technology. Barcoding relies on the use of random, semi-random or evolving nucleic acid sequences (barcodes) as permanent or dynamic labels for individual cells. Because of the effectively unlimited number of possible barcode sequences, large populations of cells can be efficiently and cost-effectively labeled and tracked at the individual level. Exhaustive labeling of organs or even organisms is therefore conceivable and an active area of research[1–7]. Today, cellular barcoding allows the origin or history of thousands to millions of cells to be tracked over developmental[8,9] and evolutionary[10] time scales, thereby speeding up the investigation of these biological systems by many orders of magnitude. Moreover, barcode functionalization makes it possible to record cellular features such as the response to stimuli[11,12], and to map neuroanatomical features[13–15].

Despite these diverse applications, the technology underlying all uses of cellular barcoding can be discussed in the same theoretical framework, and is subject to very similar design constraints. To highlight these similarities, we first review the foundations of cellular barcoding, covering different types of barcodes, in vivo barcode generation strategies and the mathematics behind stochastic barcode assignment. We then detail classic applications of cellular barcoding to prospective lineage tracing and high-throughput screens, and finally introduce the functionalization of barcode sequences to map neural anatomy and record cellular events. Although we do not discuss the use of barcodes to tag individual DNA or RNA molecules[16–23] or samples (including use in multiplexed scRNA-seq library generation)[21,23–25], many of the same principles used for cellular barcoding are relevant to molecular barcoding. We also do not discuss the somewhat different usage of the term "barcoding" in ecology[26].

Principles and methods of cellular barcoding

Labeling cells with nucleic acid sequences. Cellular barcoding exploits the almost infinite number of unique molecules that can be generated with short sequences of nucleotides. In the simplest case, each cell is tagged with a specific sequence of a given length, such that the number of possible barcodes is equivalent to 4^N, where N is the length of the sequence (because each position can encode one of four bases). A random 10-bp barcode therefore can assume any of 4^{10} ($\sim10^6$) different sequences, and a random 30-bp barcode can assume any of 4^{30} ($\sim10^{18}$) different sequences, each of which can act as a unique label.

In addition to the use of random sequences (e.g., refs [13,27]), barcodes can be composed of semi-random nucleic acid sequences (e.g., refs [9,10]), in which some positions are constrained to one or more specific nucleotides. Barcodes can also be constructed from shuffled sequence segments (e.g., refs [6,28]), which allows for easier error correction or in vivo barcode generation, at the cost of some potential barcode diversity. Finally, barcodes can be generated via random deletions in known sequences, as is common in CRISPR–Cas9-based methods[1–4,11]. Which barcode type is chosen for any given study depends largely on the required barcode diversity and the method used to read out the barcodes.

Methods of barcode delivery. Conceptually, the easiest way to barcode a sample is to manually assign individual barcodes to cells one by one. The uniqueness of a barcode to the cell it labels is thus guaranteed, and the barcode space can be covered exhaustively—that is, every barcode will be used. One-by-one assignment has been powerful in genome-wide screens[29,30] and is still used to track a small number of conditions[31]. However, the approach is labor intensive and is limited to use with populations of cells, as it is currently very challenging to assign specific barcodes to individual cells. One-by-one labeling is therefore used only under very limited conditions.

[1]Watson School of Biological Sciences, Cold Spring Harbor, NY, USA. [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.
*e-mail: zador@cshl.edu

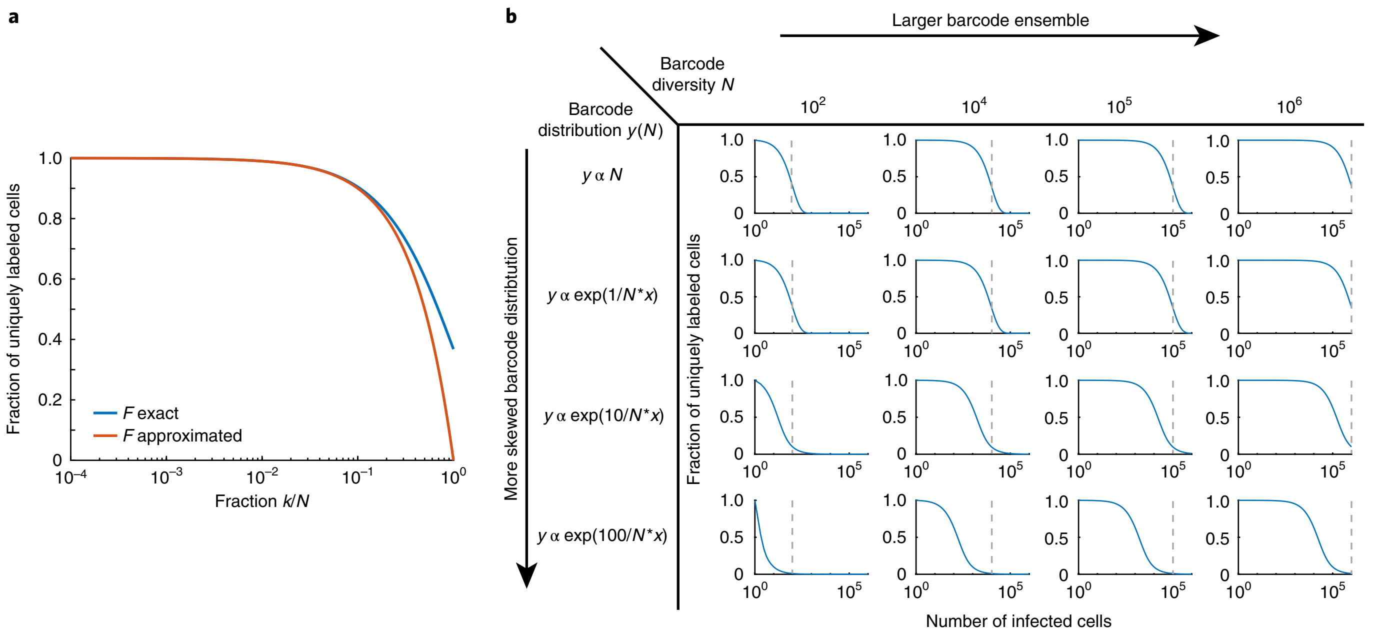


Fig. 1 | The mathematics underlying cellular barcoding. a, For a large number of uniformly distributed barcodes (N) in the ensemble and a small number of used barcodes (k), the fraction of uniquely labeled cells can be approximated as $F \approx 1 - (k/N)$. **b**, Relationships among barcode ensemble size, barcode distribution and the fraction of uniquely labeled cells. The larger the barcode ensemble is and the closer the barcode distribution is to a uniform distribution, the more cells can be labeled uniquely.

Currently, the most common, robust and efficient method to barcode individual cells relies on the production of a large pool of barcoded vectors (plasmids, viruses, etc.) in vitro. The vector pool is transfected under conditions optimized to deliver a few barcodes at most to every transfected cell, and to transfect only the desired number of target cells. The most common delivery method for such pooled barcode libraries is retroviral (including lentiviral) transfection[8,9], but other viruses such as Sindbis virus[13] and pseudo-rabies virus[32] can be used, as can nonviral delivery methods including plasmid injection[33] and electroporation[34]. Given a sufficiently large number of barcodes relative to the number of transfected cells, it is very unlikely that the same barcode will be transfected into two different cells (Box 1, Fig. 1). Every transfected cell is therefore uniquely labeled by the barcode it takes up.

In vitro barcode production is very efficient, such that vector libraries containing billions of barcodes can be relatively easily constructed and used to label millions of cells[35]. Moreover, in vitro construction allows for very compact barcode design (e.g., 30 random bases), thus facilitating readout by short-read sequencing technologies. However, applications are limited to organs, time windows and biological questions for which delivery is feasible and practical. Moreover, when barcodes are transfected under conditions ensuring a few barcodes per cell at most, Poisson statistics dictate that some cells will remain unlabeled. Thus, if the ultimate goal of exhaustive labeling of every cell in a tissue or organism is to be achieved, alternatives to Poisson-limited barcode delivery must be developed.

One way to avoid the drawbacks of experimental access and exhaustive labeling is to evolve a unique barcode within each cell from an 'ancestral' sequence—a sequence that at time zero is identical across the population. As discussed below, implementations of such in vivo barcoding rely on either the shuffling of sequence fragments or the introduction of random insertions or deletions at a specific site. So far, neither approach has been able to generate sufficient diversity to exhaustively label an adult vertebrate, but the field is progressing quickly (Table 1).

Recombinase-based in vivo barcoding. Initial approaches for in vivo barcode generation centered on the action of a DNA recombinase on an array of possible targets. The first such method was Brainbow[36,37]. In Brainbow, Cre recombinase, which can excise or flip DNA sequences flanked by specific recognition sequences, acts on an array of fluorescent protein open reading frames. Repeated Cre action leads to stochastic shuffling and collapse of the target array, generating different combinations of fluorophores in each cell that can be distinguished by imaging.

Scientists can generate higher barcode diversity in vivo by replacing fluorophores with shorter DNA sequences that can be read out by sequencing, which makes it possible to expand the array to contain more targets[6]. However, because Cre intrinsically favors excision over flipping, the target array shrinks in size over time, leading to a low final diversity that increases only linearly with the number of targets in the array[6] (Fig. 2a). Recently, the problem of array collapse was partially overcome with the *Polylox* method through limitation of Cre activity via temporary induction (thus stopping Cre action on the target array before its ultimate collapse), which allowed in vivo barcoding of hematopoietic stem cells[28]. Inducible Cre action, moreover, can restrict barcoding to a temporal window of interest. An alternative approach that avoids array collapse altogether involves the use of Rci DNA recombinase, which flips but does not excise DNA segments between recognition sites. The avoidance of excisions dramatically increases the potential barcode diversity to $2^n n!$ for n segments (Fig. 2b), and has been used successfully in bacteria[6].

A fundamental drawback of recombinase-based barcoding approaches is that target arrays tend to be long and repetitive, as dictated by the low diversity of recombinase recognition sites and their minimum spacing requirements. To achieve high barcode diversity, target arrays must contain many segments, which necessitates barcode readout by lower-throughput long-read (e.g., PacBio[6,28]) sequencing. Future improvements in long-read sequencing technologies or in situ readout of barcodes[4] might mitigate this drawback.

Table 1 | Overview of in vivo barcoding techniques and their properties

Name	Enzyme	Theoretical diversity	Demonstrated diversity per experiment	Barcode length	Readout	Reference(s)
Brainbow	Cre	—	~200	*x* insertion sites × 1,000–3,000 bp	Microscopy	36,37
Flpbow	Flp	—	—	*x* insertion sites × 1,000–3,000 bp	Microscopy	37
Polylox	Cre-ERT2	1,866,868	849	1,942 bp	PacBio	28
	Rci	176,947,200	1,723	1,472 bp	PacBio	6
Gestalt	Cas9	—	4,195	257 bp	Illumina	1
scGestalt	Cas9	—	2,213	257 bp	scRNA-seq; Illumina	54
scartrace	Cas9	—	1,572	*x* insertion sites × 700 bp *rfp* transgene	Illumina	3
Linnaeus	Cas9	—	230	*x* insertion sites × 700 bp *rfp* transgene	scRNA-seq; Illumina	40
ScarTrace	Cas9	—	—	8 insertions of H2A-gfp transgene	scRNA-seq; Illumina	3,39
mScribe	Cas9 + self-targeting gRNA	—	1,890	20–70 bp	Illumina	11
Homing barcodes	Cas9 + self-targeting gRNA	—	—	20–100 bp	Illumina; FISSEQ rolonies	2
MEMOIR	Cas9	—	~256	28 × ~1,000-bp scratchpads	FISH	4

Dashes indicate quantities that are not provided in the cited publications or are not well defined. FISSEQ, fluorescent in situ sequencing.

CRISPR-based in vivo barcoding. Recent work by a number of laboratories demonstrates the use of CRISPR–Cas9 as an alternative to DNA recombinases for barcode generation. Cas9-induced double-strand breaks in genomic DNA are often repaired by nonhomologous end joining (NHEJ)[38], an error-prone mechanism that introduces short random insertions and deletions at the cut site. These untemplated changes to the parental sequence act as a short barcode that can be used to distinguish cells. This basic idea was exploited successfully in ScarTrace[3,39,40], which uses the sequence diversity generated by a CRISPR–Cas9-mediated cut in a (potentially multicopy) transgene for lineage tracing in zebrafish. A similar approach was also demonstrated in *Caenorhabditis elegans*[41]. The GESTALT[1] and MEMOIR[4] systems increase barcode diversity by designing arrays of many perfect or mismatched CRISPR target sites (Fig. 2c).

As an alternative approach to increasing barcode diversity, mSCRIBE[11] and homing CRISPR barcodes[2] rely on an engineered guide RNA that targets its own genomic spacer sequence, instead of a target array. In a first step, the guide RNA genomic locus is cut and mutated, which produces barcode diversity. Subsequently, the mutated locus produces new guide RNA that again targets its own already mutated genomic locus. Over time, the guide RNA sequence evolves, acting as a diverse barcode sequence (Fig. 2d).

CRISPR–Cas9 approaches hold the promise of high-diversity, organism-wide, time-resolved in vivo barcode production. The initial proof-of-principle studies generated diversities too low for organism-wide barcoding, in part because of NHEJ's intrinsic bias toward the production of deletions (similar to Cre recombinase (described above)) rather than insertions, which leads to the collapse of the CRISPR barcodes over time[2]. This effect can be overcome through the use of several independently evolving barcodes per cell, which boosts the combined diversity to the product of the individual diversities[3,4,40,42]. Alternatively, an elegant approach to the production of highly diverse and compact CRISPR barcodes would be to modify NHEJ to favor insertions over deletions or to use CRISPR-directed base editors[43]. The field of CRISPR barcoding is developing rapidly, and these limitations may soon be overcome.

Methods of barcode readout. Nearly all work on cellular barcoding to date has relied on the extraction of nucleic acids followed by barcode detection or quantification in vitro. The methods used to read out barcodes have varied with the available technology, beginning with PCR amplification and sizing[8] and progressing to microarray detection[9,44], Sanger sequencing[45,46] and high-throughput sequencing[27,47]. In a paradigm shift, scRNA-seq approaches have recently been applied to dissociated cells for the simultaneous readout of a cell's barcode and transcriptome. This is an extremely powerful approach, as it combines information about cellular history or anatomy from the barcode with the independently measured high-dimensional phenotype of the cell's transcriptional state and transcriptional cell type. The combination of cellular barcoding with scRNA-seq has been exploited in genome-wide screens[48–53], lineage-tracing approaches[33,39,40,54] and neuroanatomy studies[55].

Tissue lysis and the production of single-cell suspensions, however, irrevocably destroy the 3D arrangement of cells in vivo, and with it a lot of potentially valuable information. A strategy to avoid this adapts methods developed for in situ detection of nucleic acids to the detection of barcodes. Recently, the MEMOIR method[4] was used with highly multiplexed fluorescence in situ hybridization (FISH)[56] to read out a combination of in vitro and CRISPR–Cas9-generated barcodes. Similarly, multiplexed FISH was used to register live images of bacteria to their cellular barcodes[57,58]. The detection of barcodes by FISH, however, constrains the compactness and diversity of barcodes that can be used, as hybridization probes cannot easily differentiate among a large pool of barcode sequences. We note that fluorophore-based barcoding approaches such as Brainbow have similar conceptual constraints[36,37].

In situ sequencing approaches[59,60], in which RNA is sequenced de novo in tissue, may provide an alternative strategy not subject

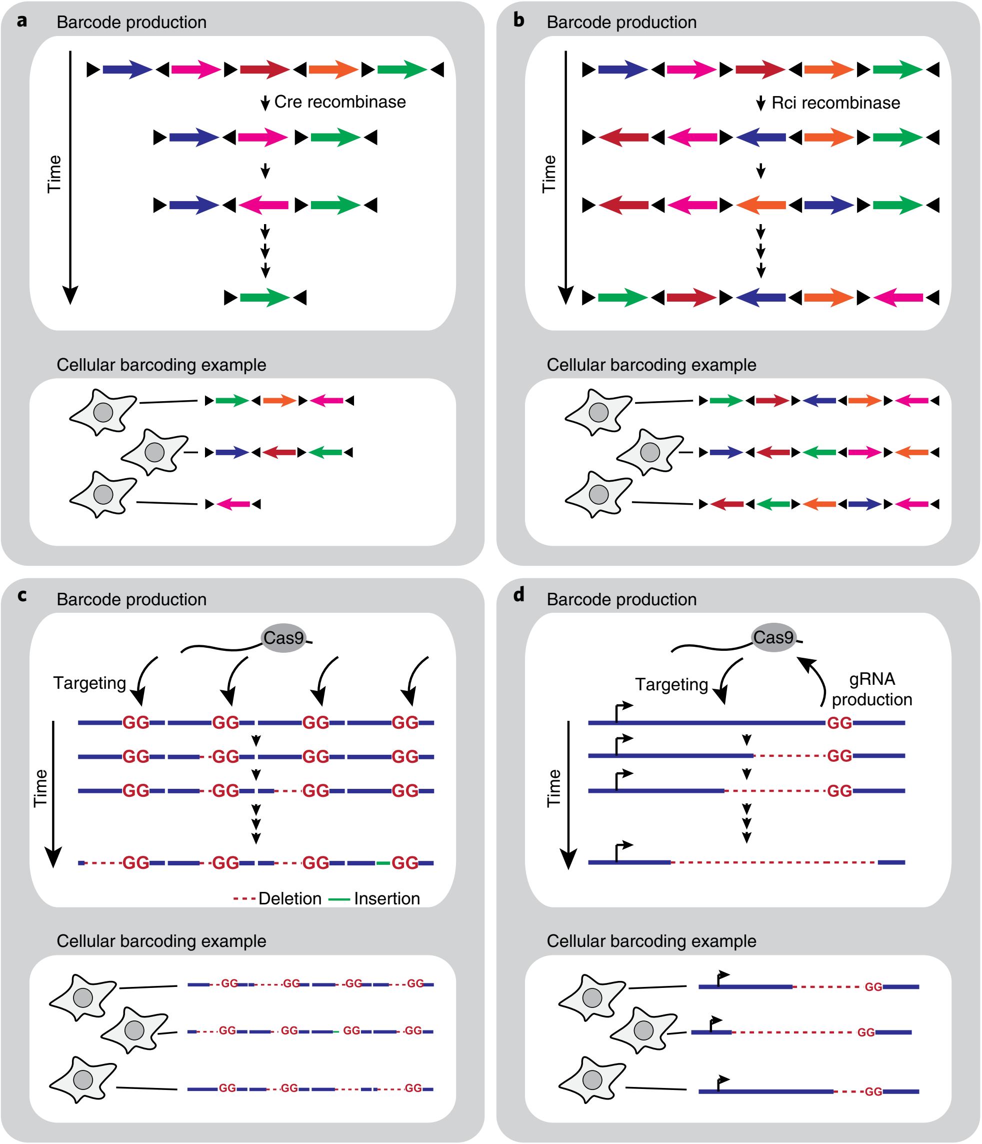


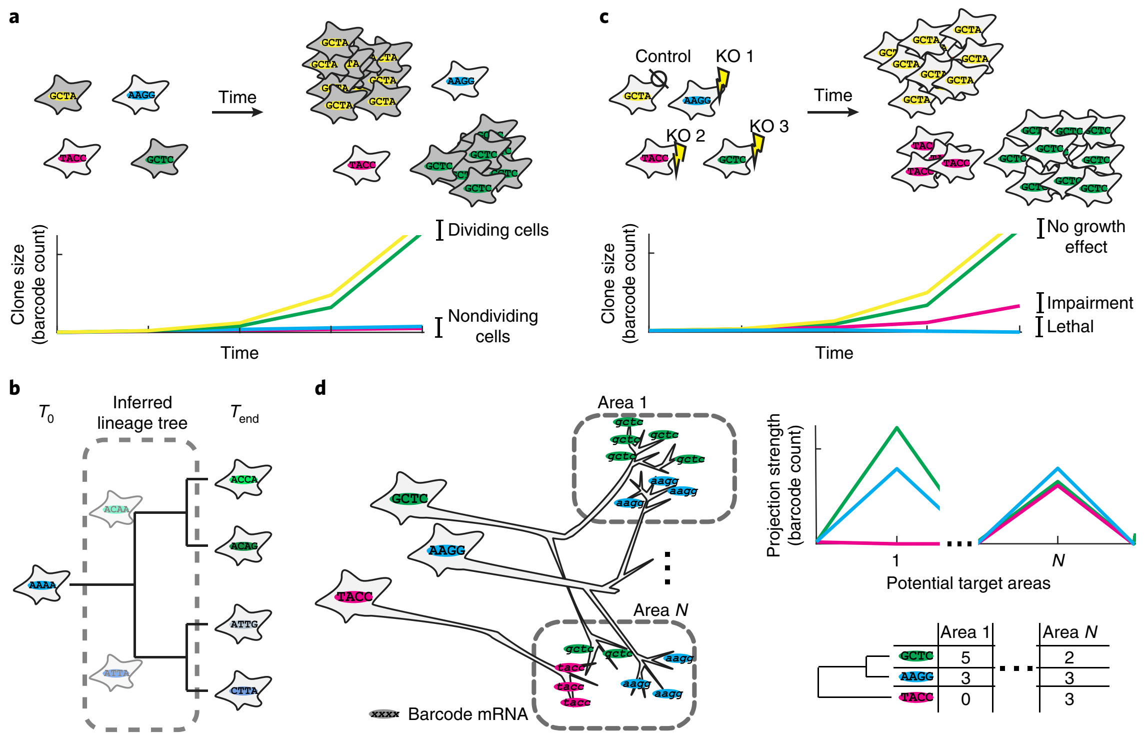
Fig. 2 | Strategies for in vivo barcode production. a, When Cre recombinase acts on an array of target sites (colored arrows) flanked by *loxP* sites (black triangles), it will excise or flip subsets of these targets, creating sequence diversity. Excessive excision will collapse the array to a single target (top), but Cre activity can be limited[28] to generate highly diverse shuffled barcodes (bottom). **b**, Rci recombinase only flips segments in an array of target sites, such that diversity increases over time while array length is maintained[6]. **c**, CRISPR–Cas9 activity will progressively introduce sequence diversity into an array of target sites over time, as a result of the insertions and deletions generated by imperfect NHEJ repair of Cas9-mediated double-strand breaks[1,4]. **d**, Alternatively, a CRISPR guide RNA (gRNA) can be engineered to target itself repeatedly, and thus build up sequence diversity at that locus[2,11].

to these constraints. Indeed, the potential for readout of homing CRISPR barcodes by targeted fluorescent in situ sequencing has been demonstrated[2]. Another approach, BaristaSeq[61], uses a combination of padlock probe hybridization[62] and gap filling followed by in situ sequencing for accurate and efficient in situ detection of cellular barcodes. Techniques such as this promise to combine the advantages of high-diversity barcode libraries with the high spatial resolution of imaging.

Every readout method is subject to errors in barcode detection. In bulk sequencing approaches, for example, these include PCR errors[63–65] such as single base substitutions[64,65], insertions/deletions[65] or template switches[65,66], and sequencing errors[67,68], as wells as errors specific to the barcoding method, such as those made by a viral polymerase during barcode transcription[69]. These errors must be taken into consideration during analysis, as they might lead to barcode misidentification. In many scenarios, however, the large range of possible sequences compared with the small number of actually used barcodes offers avenues for the correction of readout errors (e.g., ref. [70]).

Applications of cellular barcoding

Lineage tracing and fate mapping. Developmental biology provides some of the most striking examples of the value of studying cells individually rather than in bulk. Reconstruction of the precise trajectories by which individual cells arrive at their mature and differentiated states—that is, their cellular lineage—is one of the central goals of developmental biology. One powerful approach for lineage reconstruction involves labeling a particular cell, or population of cells, at one point in time and then faithfully identifying the cell's progeny by the presence of the label (Fig. 3a; also see recent reviews[71–73]). Here, we distinguish between the related concepts of lineage tracing and fate mapping. In lineage tracing, the developmental



	Area 1		Area N
GCTC	5	…	2
AAGG	3		3
TACC	0		3

Fig. 3 | Applications of cellular barcoding. a, Cellular barcoding can be applied to fate-mapping studies, for example, to count the number of dividing stem cells in a heterogeneous population of cells. **b**, Evolving barcodes allow the reconstruction of cell lineages in a single experiment by retrospective inference of cellular relationships on the basis of barcode similarity. **c**, The combination of cellular barcodes with genetic perturbations such as gene knockout (KO), shRNA-mediated inhibition and CRISPR manipulation allows for the pooled screening of a large number of different genotypes or conditions. **d**, Functionalization of cellular barcodes via expression of each cell's barcode as an mRNA that is trafficked into axonal processes allows high-throughput tracing of neuronal anatomy.

history of a cell is decoded and a tree is produced, whereas in fate mapping, cells originating from a specific progenitor are marked identically, which can obscure information about intermediate steps. Lineage tracing requires that cells in intermediate states receive distinct labels. Note that in the literature, "lineage tracing" is also used as an umbrella term encompassing all methods that interrogate cellular lineage, and as such often includes fate-mapping approaches.

Early fate-mapping experiments tracked just one or a few cells[74,75]. However, in a foundational paper published more than two decades ago, Walsh and Cepko applied cellular barcoding to map cell fates[8]. They sought to determine whether the descendants of individual neural progenitors in the developing rat neocortex stayed in a local columnar structure or dispersed across the cortex. They labeled several progenitors with randomly generated barcodes from a retroviral library and then detected the barcodes after some time. They observed that descendants of individual progenitors were spread widely throughout the adult cortex. Notably, the use of barcodes was necessary to reach this conclusion; traditional single-cell tracing approaches based on a single fluorescent marker would generally attribute widely dispersed clones to accidental labeling of multiple starter cells. Although this study and its follow-up[76] used a low-diversity barcode library (<100 sequences identified by PCR[8,76–78]), subsequent cellular barcoding libraries quickly grew to allow researchers to trace more cells in parallel[46] (Box 1). However, the throughput of single-cell-resolution fate mapping remained limited by the technology available to distinguish individual barcodes.

In 2008, Schepers et al.[9] coined the term "cellular barcoding" to describe a high-throughput fate-mapping experiment in which they overcame the barcode-detection bottleneck by using microarrays for quantification. This advance allowed them to track thousands of barcodes in parallel, and thereby to address the relationship between T cell populations after immune challenge[9]. Shortly thereafter, de novo sequencing permitted researchers to rapidly quantify barcodes of arbitrary sequence. A proof-of-concept study using Sanger sequencing[45] was quickly followed by high-throughput sequencing for barcode detection[47] and quantification[27].

Since these foundational studies, barcoding has been used extensively in fate mapping of both multicellular organisms and communities of unicellular microbes. In particular, studies of stem cell niches rely on the ability to infer the absolute number of stem cells from the number of labeled, expanded lineages (Fig. 3a). Fate mapping has been extended to the study of disease, including heterogeneity and clonality in cancer[79] and the emergence of drug resistance[35], as well as to investigate microbial evolutionary dynamics[10].

Other forms of cellular marking besides delivered barcodes have also been explored. The genomic site of a constant DNA sequence randomly inserted by retroviral infection[80] or transposon activation[81,82] acts as a heritable and unique cellular label, akin to a cellular barcode. Similarly, naturally occurring somatic mutations have been used for fate mapping and lineage reconstruction[83]. Although these approaches share similarities with cellular barcoding approaches, they are technically quite different, and have been reviewed elsewhere[71].

Reconstruction of a complete lineage tree (i.e., true lineage tracing as defined above) could theoretically be achieved by fate mapping of one cell division at a time—a truly enormous task for most multicellular organisms. The development of in vivo–generated, evolving barcodes, such as those generated by CRISPR–Cas9, now offers a potential path toward complete lineage tree reconstruction in a single experiment (Fig. 3b). A strategy common to GESTALT[1], homing barcodes[2], ScarTrace[3,39,40] and MEMOIR[4] is the use of the repeated action of Cas9 to progressively modify DNA targets. In

Box 1 | The mathematics underlying cellular barcoding

In order for individual cells in a population to be labeled uniquely, degenerate labeling (i.e., labeling of multiple cells with the same barcode) must be avoided. Ultimately, the number of available barcodes limits the number of cells that can be labeled. If the number of labeled cells exceeds the number of barcodes, some cells must share the same barcode. In practice, barcodes are usually selected randomly from the ensemble of available barcodes rather than assigned one by one to single cells (but see refs [29,30]), so the number of cells that can be labeled uniquely does not approach the number of barcodes. It is thus important to understand how the number and distribution of barcodes within an ensemble influence the number of cells that can be uniquely labeled (Fig. 1).

First, we consider the case in which every barcode sequence is equally likely to be chosen from the ensemble. Consider the labeling of k cells, each with a single barcode drawn from an ensemble of N barcodes. Every cell has a probability of $(1 - (1/N))^{k-1}$ of being uniquely labeled and $1 - (1 - (1/N))^{k-1}$ of being degenerately labeled. The expected number of degenerately labeled cells is a random variable X whose expectation $E(X)$ is given by

$$E(X) = k\left(1 - \left(1 - \frac{1}{N}\right)^{k-1}\right)$$

The fraction of uniquely labeled cells can accordingly be expressed as

$$F = 1 - \frac{E(X)}{k} = \left(1 - \frac{1}{N}\right)^{k-1}$$

which, for $N >> k$, can be simplified to (Fig. 1a)

$$F \approx 1 - \frac{k}{N}$$

The appropriate barcode diversity depends on the number of labeled cells and on experimental considerations, most notably on how sensitive the experiment is to false positives arising from degenerate labeling. For a uniform distribution, a barcode ensemble 100-fold larger than the number of labeled cells ($N/k = 100$) yields 99% unique labeling, which for many applications results in an acceptably low error rate.

Now consider the more realistic case in which not all barcodes are equally likely to be chosen. Such skewed barcode representations in the ensemble arise naturally, for example, during the production of a virus library carrying in vitro–generated barcodes[8,13,77], or because certain sequences are preferentially generated in vivo[3,4,6,11,28]. Under these circumstances it is important to determine the probability distribution of barcodes to assess the ensemble's maximal labeling capacity. Following similar reasoning as for the uniform case, but weighting the contribution of each barcode by the probability of its being chosen (p_i for barcode $i = 1 \ldots N$), we can express the expected number of degenerately labeled cells as

$$E(X) = k \sum_{i=1}^{N} p_i \left(1 - (1 - p_i)^{k-1}\right)$$

and the fraction of uniquely labeled cells F as

$$F = 1 - \sum_{i=1}^{N} p_i \left(1 - (1 - p_i)^{k-1}\right)$$

Using this formula, we can determine the maximum number of cells to label for a given ensemble of barcodes before conducting an experiment (compare also ref. [13]). We note that a uniform barcode ensemble minimizes the rate of double labeling, and that deviations from uniform labeling increase the number of double-labeled cells (Fig. 1b).

Knowing the distribution of specific barcodes within the ensemble not only allows for an estimation of the error rate due to double labeling, but also suggests a procedure for decreasing the error rate. By identifying and discarding the cells labeled with the most abundant barcodes in the ensemble, one can reduce the number of errors from degenerate labeling post hoc—at the cost of a reduced sample size. Such correction may be especially important for in vivo barcoding approaches, in which the biological processes that generate the barcodes are inherently biased[3,4,6,11,28], and was indeed recently used in the analysis of Cre-based barcoding with the *Polylox* system[28].

these methods, lineage relationships between cells at the experimental end point can be inferred from barcode similarity (either the similarity of individual barcode sequences, when single barcodes label single cells, or the similarity of sets of barcodes, when single cells contain more than one barcode).

Before complete lineage trees comparable to the famous example from *C. elegans*[84] can be reconstructed by barcoding, three important challenges need to be addressed. First, barcode diversity needs to be high enough to allow every cell present at the experimental endpoint to be uniquely labeled. Diversity that is too low either stops lineage tracing before the endpoint (if, for example, Cas9 target sites have collapsed and lost the protospacer-adjacent motif required for cutting) or severely impedes tree reconstruction, as cells in distant lineages will share the same barcode. Strategies for increasing CRISPR barcode diversity are discussed above. Second, and related, previously generated barcodes need to be protected against loss by 'overwriting' due to subsequent Cas9-mediated excision or mutation beyond recognition. Redundancy provided by multiple barcoding sites per cell or biasing of NHEJ toward insertions over deletions could mitigate this problem. Finally, barcode evolution needs to be fast enough to capture individual cell divisions that represent branch points in the reconstructed lineage tree. This can be achieved through the use of rapidly evolving barcodes, but at the cost of requiring very large potential barcode diversities. One attractive way to overcome this challenge is to synchronize barcode evolution to cell division by, for example, expressing Cas9 in a restricted phase of the cell cycle.

Barcode-derived lineage trees can be annotated using the transcriptionally determined cell types of harvested cells. Single-cell-resolution barcode and transcriptome readouts were recently conferred on GESTALT and ScarTrace/Linnaeus by scRNA-seq[39,40,54], thus providing unique multimodal insights into the correspondence of lineage relationships and adult cell types in zebrafish.

High-throughput screens. Screens for gene function have traditionally been performed one gene at a time. Genome-wide one-by-one (arrayed) screens[85], although possible, are very labor intensive and often costly. Effort and cost, however, can be greatly reduced by screening of multiple constructs at the same time. Such pooled screens are made possible by infection of each cell with only one uniquely barcoded construct. Each cell is then effectively fate-mapped (i.e., linked to a genotype) and phenotyped to reveal the cell autonomous effect of the genetic modification (Fig. 3c).

This approach was first used to generate large-scale deletion libraries in yeast, in which every strain was tagged with a different barcode sequence[29,30,86]. The knockout strains generated in this way could be pooled and grown to enable researchers to assess the fitness effects of individual deletions, thus laying the foundation for functional genomics in yeast and generating deep insights into cell biology (for a review, see ref. [87]). Since then, researchers have developed short hairpin RNA (shRNA) screens in which each shRNA construct is tagged with a known unique barcode sequence. These constructs are pooled, packaged into a retroviral or lentiviral library and delivered to a population of cells. The approach allowed the first genome-wide screens in mammalian cells[88–90]. By measuring the abundance of each barcode over time, researchers can assess the effects of each shRNA on fitness. Subsequent CRISPR knockout libraries have replaced the shRNA with a guide RNA, which itself acts as a barcode[91–93], for genome-wide screens in mammalian cells.

Barcode-enabled screening is traditionally limited to relatively simple phenotypes (e.g., viability) based on the enrichment of beneficial barcodes (Fig. 3c). Two recent proof-of-concept studies used live cell microscopy of engineered, barcoded bacteria to record and screen more complex, time-resolved phenotypes[57,58]. After fixation, the researchers read out cellular barcodes by serial FISH and then matched each cell's phenotype to its genotype by registering live cell data to the FISH images. Pooled microscopy-based screens have the potential to be very powerful in optically accessible systems. In particular, we are looking forward to applications in mammalian cells, potentially in combination with de novo barcode sequencing for increased barcode diversity and thus increased screening throughput.

Another important recent development provides rich, high-dimensional phenotypes in pooled CRISPR perturbation screens by using scRNA-seq as a readout of both cellular phenotype and guide RNA (barcode) identity[48–53].

Mapping the brain with barcodes. Beyond the traditional applications of lineage reconstruction and screening, barcodes are now being functionalized to record more than cellular identity. We recently introduced the use of cellular barcodes to rapidly and cost-effectively map neural connectivity[5]. The ultimate goal in neuroanatomy is to determine the complete wiring diagram of a brain at single-cell and single-synapse resolution. Traditional neuroanatomical methods, however, are subject to similar tradeoffs between throughput and resolution as lineage tracing prior to the advent of cellular barcoding. The choice is to either quickly map the connections of large populations of neurons through bulk tracing[94,95], or map connectivity one neuron at a time by single-neuron tracing[96] or even electron microscopy reconstructions[97].

To overcome this tradeoff, we functionalized cellular barcodes to record both cellular identity by sequence and neuroanatomical features by localization. We developed MAPseq, a method that allows the long-range projections of large numbers of individual neurons to be determined simultaneously[13]. In MAPseq, a large number of neurons are barcoded in situ by viral infection. Unlike in conventional cellular barcoding approaches, the barcodes are expressed as mRNA and trafficked into the axonal processes of each labeled neuron (Fig. 3d). Dissection of potential target brain regions followed by bulk barcode sequencing allows projections of each labeled neuron to be mapped through the quantification of barcode-labeled processes in each sequenced region. We have applied MAPseq to map projections from the locus coeruleus[13] and primary visual cortex[14] in mouse, and combined it with in situ sequencing to map projections to the auditory cortex[98], uncovering structures inaccessible at the bulk level in each case.

More recently, we demonstrated that researchers can also use barcodes to read out synaptic connectivity, by joining the cellular barcodes of connected neurons across the synapse[15].

Barcodes as molecular recording devices. Another intriguing example of barcode functionalization is the CRISPR–Cas9-based mScribe system. mScribe uses the rate of barcode divergence from the ancestral sequence to record the intensity or duration of inflammatory stimulation by placing the mutagenic Cas9 under the control of an inflammation-responsive promoter[11]. Similar ideas underlie the 'molecular ticker tape' proposed to record fast events such as neural activity in DNA in a noninvasive manner and with single-cell resolution by, for example, amplifying DNA with a polymerase whose error rate is a function of cellular Ca^{2+} concentration[99]. Transient changes in Ca^{2+} concentration are therefore permanently recorded as errors in the amplified DNA, and can be read out by sequencing.

Outlook

When it was first introduced, the use of barcodes was a means to track many cells over time. With the technological developments of the past two years, cellular barcoding is on the verge of becoming the foundation for a comprehensive, multimodal understanding of tissues and organisms with cellular resolution through time and space[100]. For the brain, for example, we envision a not too distant future in which every cell will be uniquely labeled with a barcode sequence in a typical experiment. Barcode locations will be used to map all synaptic connections between neurons in the brain (the connectome), and the barcode sequence itself will carry complete lineage information and signatures of specific, salient events in each cell's history. As all this information is stored in nucleic acid sequences, we envision that it will be read out by in situ sequencing methods, alongside each cell's transcriptome, such that barcode-based information can be integrated with transcriptomics and spatially aligned technologies.

For this vision to become reality, several hurdles still need to be overcome. First, in vivo barcoding methods currently do not produce diverse enough barcodes to uniquely label every cell in many organs, including even small mammalian brains. Moreover, the biases with which barcodes are generated in vivo are not sufficiently understood. Second, although encouraging progress has been made, the functionalization of barcodes to read out neuroanatomical features and lineage is still at an early stage in its development. Specifically, it is currently not possible to read out synaptic connectivity on the basis of barcoding at high efficiency, and lineage tracing based on barcodes is hampered by the lack of synchronization between barcode mutation and cell division. Last, in situ readout of barcodes, or the cellular transcriptome, is currently slow, inefficient or biased. More technological development is needed.

Beyond these immediate extensions and combinations of existing ideas and technologies, we expect more cellular features and cellular histories to be written into nucleic acid barcodes in the future. One might imagine a time-stamped 'interactome' of immune cells over their lifetime, high-resolution molecular ticker tapes recording neural activity or histories of gene expression, and other, stranger ideas.

Received: 13 November 2017; Accepted: 26 September 2018;
Published online: 30 October 2018

References

1. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
Development and application of CRISPR–Cas9-generated evolving barcodes for lineage tracing in zebrafish. See also refs. 2–4.
2. Kalhor, R., Mali, P. & Church, G. M. Rapidly evolving homing CRISPR barcodes. *Nat. Methods* **14**, 195–200 (2017).

3. Junker, J. P. et al. Massively parallel clonal analysis using CRISPR/Cas9 induced genetic scars. *bioRxiv* Preprint at https://www.biorxiv.org/content/early/2017/01/04/056499 (2017).
4. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
5. Zador, A. M. et al. Sequencing the connectome. *PLoS Biol.* **10**, e1001411 (2012).
6. Peikon, I. D., Gizatullina, D. I. & Zador, A. M. In vivo generation of DNA sequence diversity for cellular barcoding. *Nucleic Acids Res.* **42**, e127 (2014).
7. Frumkin, D., Wasserstrom, A., Kaplan, S., Feige, U. & Shapiro, E. Genomic variability within an organism exposes its cell lineage tree. *PLoS Comput. Biol.* **1**, e50 (2005).
8. Walsh, C. & Cepko, C. L. Widespread dispersion of neuronal clones across functional regions of the cerebral cortex. *Science* **255**, 434–440 (1992).
First use of barcodes to track cells.
9. Schepers, K. et al. Dissecting T cell lineage relationships by cellular barcoding. *J. Exp. Med.* **205**, 2309–2318 (2008).
10. Levy, S. F. et al. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181–186 (2015).
11. Perli, S. D., Cui, C. H. & Lu, T. K. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, aag0511 (2016).
Use of barcode evolution to record the duration and intensity of stimuli.
12. Chruch, G. & Shendure, J. Nucleic acid memory device. US patent application US20030228611A1 (2003).
13. Kebschull, J. M. et al. High-throughput mapping of single-neuron projections by sequencing of barcoded RNA. *Neuron* **91**, 975–987 (2016).
Use of barcodes to map axonal projections at single-cell resolution.
14. Han, Y. et al. The logic of single-cell projections from visual cortex. *Nature* **556**, 51–56 (2018).
15. Peikon, I. D. et al. Using high-throughput barcode sequencing to efficiently map connectomes. *Nucleic Acids Res.* **45**, e115 (2017).
16. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. USA* **108**, 9530–9535 (2011).
17. Shiroguchi, K., Jia, T. Z., Sims, P. A. & Xie, X. S. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. *Proc. Natl. Acad. Sci. USA* **109**, 1347–1352 (2012).
18. Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2011).
19. Fu, G. K., Hu, J., Wang, P. H. & Fodor, S. P. A. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc. Natl. Acad. Sci. USA* **108**, 9026–9031 (2011).
20. Casbon, J. A., Osborne, R. J., Brenner, S. & Lichtenstein, C. P. A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic Acids Res.* **39**, e81 (2011).
21. Miner, B. E., Stöger, R. J., Burden, A. F., Laird, C. D. & Hansen, R. S. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. *Nucleic Acids Res.* **32**, e135 (2004).
22. Brenner, S. & Macevicz, S. C. Molecular counting. WO patent application WO2007087312A3 (2007).
23. Brenner, S. Simultaneous sequencing of tagged polynucleotides. US patent US5763175A (1995).
24. Craig, D. W. et al. Identification of genetic variants using bar-coded multiplexed sequencing. *Nat. Methods* **5**, 887–893 (2008).
25. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
26. Valentini, A., Pompanon, F. & Taberlet, P. DNA barcoding for ecologists. *Trends Ecol. Evol.* **24**, 110–117 (2009).
27. Naik, S. H. et al. Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* **496**, 229–232 (2013).
28. Pei, W. et al. *Polylox* barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456–460 (2017).
29. Winzeler, E. A. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
30. Giaever, G. et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
31. Yu, C. et al. High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nat. Biotechnol.* **34**, 419–423 (2016).
32. Oyibo, H. et al. A computational framework for converting high-throughput DNA sequencing data into neural circuit connectivity. *bioRxiv* Preprint at https://www.biorxiv.org/content/early/2018/01/07/244079 (2018).
33. Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
34. Loulier, K. et al. Multiplex cell and lineage tracking with combinatorial labels. *Neuron* **81**, 505–520 (2014).
35. Bhang, H. E. et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat. Med.* **21**, 440–448 (2015).
36. Livet, J. et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
37. Cai, D., Cohen, K. B., Luo, T., Lichtman, J. W. & Sanes, J. R. Improved tools for the Brainbow toolbox. *Nat. Methods* **10**, 540–547 (2013).
38. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
39. 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).
Combination of evolving Cas9-generated barcodes and single-cell sequencing to read out both lineage and single-cell transcriptional states of individual cells. See also refs. 40,54.
40. Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
41. Schmidt, S. T., Zimmerman, S. M., Wang, J., Kim, S. K. & Quake, S. R. Quantitative analysis of synthetic cell lineage tracing using nuclease barcoding. *ACS Synth. Biol.* **6**, 936–942 (2017).
42. Kalhor, R. et al. A homing CRISPR mouse resource for barcoding and lineage tracing. *bioRxiv* Preprint at https://www.biorxiv.org/content/early/2018/03/12/280289 (2018).
43. 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).
44. van Heijst, J. W. J. et al. Recruitment of antigen-specific CD8$^+$ T cells in response to infection is markedly efficient. *Science* **325**, 1265–1269 (2009).
45. Gerrits, A. et al. Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* **115**, 2610–2618 (2010).
46. Golden, J. A., Fields-Berry, S. C. & Cepko, C. L. Construction and characterization of a highly complex retroviral library for lineage analysis. *Proc. Natl. Acad. Sci. USA* **92**, 5704–5708 (1995).
47. 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).
First use of high-throughput sequencing for reading out cellular barcodes in the hematopoietic lineage.
48. Dixit, A. et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).
49. Adamson, B. et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882 (2016).
50. Jaitin, D. A. et al. dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **167**, 1883–1896 (2016).
51. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
52. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol. Cell* **66**, 285–299 (2017).
53. Hill, A. J. et al. On the design of CRISPR-based single-cell molecular screens. *Nat. Methods* **15**, 271–274 (2018).
54. Raj, B. et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
55. Klingler, E. et al. Single-cell molecular connectomics of intracortically-projecting neurons. *bioRxiv* Preprint at https://www.biorxiv.org/content/early/2018/07/27/378760 (2018).
56. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell *in situ* RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
57. Emanuel, G., Moffitt, J. R. & Zhuang, X. High-throughput, image-based screening of pooled genetic-variant libraries. *Nat. Methods* **14**, 1159–1162 (2017).
58. Lawson, M. J. et al. *In situ* genotyping of a pooled strain library after characterizing complex phenotypes. *Mol. Syst. Biol.* **13**, 947 (2017).
59. Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
60. Ke, R. et al. *In situ* sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
61. Chen, X., Sun, Y.-C., Church, G. M., Lee, J. H. & Zador, A. M. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res.* **46**, e22 (2018).
62. Nilsson, M. et al. Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085–2088 (1994).
63. Schirmer, M. et al. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43**, e37 (2015).
64. Kebschull, J. M. & Zador, A. M. Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Res.* **43**, e143 (2015).

65. Potapov, V. & Ong, J. L. Examining sources of error in PCR by single-molecule sequencing. *PLoS One* **12**, e0169774 (2017).
66. Pääbo, S., Irwin, D. M. & Wilson, A. C. DNA damage promotes jumping between templates during enzymatic amplification. *J. Biol. Chem.* **265**, 4718–4721 (1990).
67. Schirmer, M., D'Amore, R., Ijaz, U. Z., Hall, N. & Quince, C. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics* **17**, 125 (2016).
68. Manley, L. J., Ma, D. & Levine, S. S. Monitoring error rates in Illumina sequencing. *J. Biomol. Tech.* **27**, 125–128 (2016).
69. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral mutation rates. *J. Virol.* **84**, 9733–9748 (2010).
70. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
71. 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).
72. Ma, J., Shen, Z., Yu, Y.-C. & Shi, S.-H. Neural lineage tracing in the mammalian brain. *Curr. Opin. Neurobiol.* **50**, 7–16 (2018).
73. Kretzschmar, K. & Watt, F. M. Lineage tracing. *Cell* **148**, 33–45 (2012).
74. 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).
75. Frank, E. & Sanes, J. R. Lineage of neurons and glia in chick dorsal root ganglia: analysis in vivo with a recombinant retrovirus. *Development* **111**, 895–908 (1991).
76. Walsh, C. & Cepko, C. L. Clonal dispersion in proliferative layers of developing cerebral cortex. *Nature* **362**, 632–635 (1993).
77. Kirkwood, T., Price, J. & Grove, E. The dispersion of neuronal clones across the cerebral cortex. *Science* **258**, 317–320 (1992).
78. Walsh, C., Cepko, C. L., Ryder, E. F., Church, G. M. & Tabin, C. Response. *Science* **258**, 317–320 (1992).
79. Wagenblast, E. et al. A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. *Nature* **520**, 358–362 (2015).
80. Schmidt, M. et al. Clonality analysis after retroviral-mediated gene transfer to CD34$^+$ cells from the cord blood of ADA-deficient SCID neonates. *Nat. Med* **9**, 463–468 (2003).
81. Sun, J. et al. Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327 (2014).
82. Rodriguez-Fraticelli, A. E. et al. Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216 (2018).
83. Evrony, G. D. et al. Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59 (2015).
84. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977).
85. Kamath, R. S. et al. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
86. Smith, A. M. et al. Quantitative phenotyping via deep barcode sequencing. *Genome Res.* **19**, 1836–1842 (2009).
87. Giaever, G. & Nislow, C. The yeast deletion collection: a decade of functional genomics. *Genetics* **197**, 451–465 (2014).
88. Paddison, P. J. et al. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
89. Berns, K. et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
90. Silva, J. M. et al. Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science* **319**, 617–620 (2008).
91. Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
92. Zhou, Y. et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).
93. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
94. Oh, S. W. et al. A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
95. Zingg, B. et al. Neural networks of the mouse neocortex. *Cell* **156**, 1096–1111 (2014).
96. Ghosh, S. et al. Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. *Nature* **472**, 217–220 (2011).
97. Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471**, 183–188 (2011).
98. Chen, X., Kebschull, J. M., Zhan, H., Sun, Y.-C. & Zador, A. M. High-throughput mapping of long-range neuronal projection using in situ sequencing. *bioRxiv* Preprint at https://www.biorxiv.org/content/early/2018/08/31/294637 (2018).
99. Glaser, J. I. et al. Statistical analysis of molecular signal recording. *PLoS Comput. Biol.* **9**, e1003145 (2013).
100. Marblestone, A. H. et al. Rosetta brains: a strategy for molecularly-annotated connectomics. *arXiv* Preprint at https://arxiv.org/abs/1404.5103 (2014).

Acknowledgements

We thank B. Cazakoff for comments on the manuscript. This work was supported by the US National Institutes of Health (5RO1NS073129 and 5RO1DA036913 to A.M.Z.), the Brain Research Foundation (BRF-SIA-2014-03 to A.M.Z.), IARPA (MICrONS D16PC0008 to A.M.Z.), the Simons Foundation (382793/SIMONS to A.M.Z.), a Paul Allen Distinguished Investigator Award (to A.M.Z.), the Boehringer Ingelheim Fonds (PhD fellowship to J.M.K.), and the Genentech Foundation (PhD fellowship to J.M.K.).

Competing interests

A.M.Z. is a founder of MAPneuro.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to A.M.Z.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.