nature protocols

https://doi.org/10.1038/s41596-025-01322-w

Protocol

Multichannel genomic recording of biological information with ENGRAM

Jenny F. Nathans [1,2,3,9], **Troy A. McDiarmid** [1,3,9] ✉, **Wei Chen** [3,4,5,9] ✉ & **Jay Shendure** [1,3,6,7,8] ✉

Abstract

Molecular recording is an emerging paradigm for measuring biology over time. Enhancer-mediated genomic recording of activity in multiplex (ENGRAM) is a recently described synthetic biology circuit architecture that converts the transient activity of *cis*-regulatory elements (CREs) into stable genomic records that can be retrospectively recovered via DNA sequencing. Here we provide a step-by-step protocol for conducting ENGRAM experiments and analyzing the resulting data. We also describe key design considerations for ENGRAM recorders, summarize the strengths and limitations of ENGRAM, and highlight applications, including multiplex signal recording and high-throughput CRE screening. In contrast to other systems for DNA-based recording in mammalian systems, ENGRAM relies on prime editing-mediated insertions to record the activity of a given CRE, such that it is inherently multiplexable—for example, four-base-pair insertions can represent the activities of up to 256 distinct CREs. A further contrast lies with ENGRAM's compatibility with DNA Typewriter, which facilitates the capture of signal order. For users with basic skills in molecular biology, mammalian cell culture and DNA sequencing analysis, ENGRAM experiments can typically be completed within 5–6 weeks.

Key points

- Enhancer-mediated genomic recording of activity in multiplex (ENGRAM) is a recently described synthetic biology circuit architecture that converts the transient activity of *cis*-regulatory elements into stable genomic records that can be retrospectively recovered via DNA sequencing.
- Unlike recorders based on double-strand breaks, ENGRAM utilizes a prime editor and Pol-2-driven prime editing guide RNA for recording, so it can simultaneously and sequentially record many different events as unique insertions to a shared DNA Tape.

Key reference

Chen, W. et al. *Nature* **632**, 1073–1081 (2024): https://doi.org/10.1038/s41586-024-07706-4

---

[1] Department of Genome Sciences, University of Washington, Seattle, WA, USA. [2] Medical Scientist Training Program, University of Washington, Seattle, WA, USA. [3] Seattle Hub for Synthetic Biology, Seattle, WA, USA. [4] Department of Biochemistry, University of Washington, Seattle, WA, USA. [5] Institute for Protein Design, University of Washington, Seattle, WA, USA. [6] Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. [7] Howard Hughes Medical Institute, Seattle, WA, USA. [8] Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA. [9] These authors contributed equally: Jenny F. Nathans, Troy A. McDiarmid, Wei Chen. ✉e-mail: troym13@uw.edu; wchen108@uw.edu; shendure@uw.edu

Introduction

The emerging paradigm of molecular recording seeks to monitor cellular activities and stably preserve this information within cells for retrospective recovery[1,2]. Genomic DNA is an arguably ideal medium for both encoding the machinery of molecular recorders (that is, leveraging the transcriptional and translational machinery of the cell to build recorders from within each cell with atomic precision), as well as for serving as the substrate for information storage (that is, leveraging the stability, biocompatibility and digital nature of DNA, together with high-throughput DNA sequencing technologies for the recovery of records[3]). A DNA-based molecular recorder typically comprises two key components: (1) a DNA writer: an enzyme that writes records by altering a target DNA sequence in a biologically conditional manner, and (2) a DNA Tape: an endogenous or engineered region of the genome that serves as the recording substrate for the DNA writer.

The field of DNA-based molecular recording has recently gained momentum owing to progress in genome engineering, especially clustered regularly interspaced short palindromic repeats (CRISPR)-related technologies[4–10]. Early recorders used recombinases to invert or excise specific DNA sequences[11–13]. More recent CRISPR nuclease-based recorders leverage cells' DNA repair mechanisms to generate diverse mutations at a DNA Tape after Cas9-mediated double-strand break (DSB) repair[4,5]. Alternatives include base editing[6,7,14–16], which permits precise point mutations within a defined sequence window, and prime editing (PE)[17–21], which enables users to install short insertions, deletions or point mutations at a target site. These CRISPR-based methods are arguably more effective than recombinase-based methods for stochastically creating sequence diversity, which is particularly useful for applications such as cell lineage tracing.

To move beyond recording cell lineage, biologically conditional activation of either recombinase- or CRISPR-based recorders can be achieved using tissue-specific or signal-responsive *cis*-regulatory elements (CREs) driving the expression of editing enzymes[6,7,14,16]. However, most such approaches are limited with respect to multiplexing, particularly if a different enzyme and/or DNA Tape is required for each signal. It is challenging to imagine how one would concurrently record and recover dozens to hundreds of signals in the same system with such constraints.

To record signal-specific events in cells in a manner that is compatible with extensive multiplexing, we and our colleagues recently developed enhancer-mediated genomic recording of activity in multiplex (ENGRAM)[19]—a PE-based system that converts the activity of CREs into the production of prime editing guide RNAs (pegRNAs) encoding insertional DNA barcodes (BCs) (Fig. 1a,b). These BCs, termed symbols, are then inserted into a DNA Tape by the prime editor. Each ENGRAM unit comprises an activity-specific CRE driving the production of a pegRNA encoding a CRE-specific symbol (Fig. 1a). The number of simultaneously recordable signals scales exponentially with symbol length (that is, an n-mer insertion can potentially encode the activities of 4^n distinct CREs). ENGRAM can also be combined with DNA Typewriter, a complementary technology that we concurrently developed for sequential genome editing[18]. Together, DNA Typewriter and ENGRAM enable genomic recording of the timing and order of multiple CRE activities[19].

This Protocol outlines how to design an ENGRAM experiment, including key steps for cloning ENGRAM recorders and for creating recording-compatible cell lines. In addition, we describe the retrieval of recorded information from cells, which should be straightforward for those with experience in high-throughput sequencing of PCR amplicons. We showcase ENGRAM's potential by highlighting its application to: (1) recording the activity and orders of two canonical pathways, Wnt and NFκB and (2) profiling the activity of hundreds of enhancers in cells. For experiments focused on recording the order of signals, please refer to our separate DNA Typewriter protocol[22].

Architecture of ENGRAM

In our view, a robust, inherently multichannel DNA-based molecular recording system has two key requirements. First, a unique symbol must be assignable to each signal. An alternative approach uses different enzymes, each conditional on a different signal, to edit distinct target

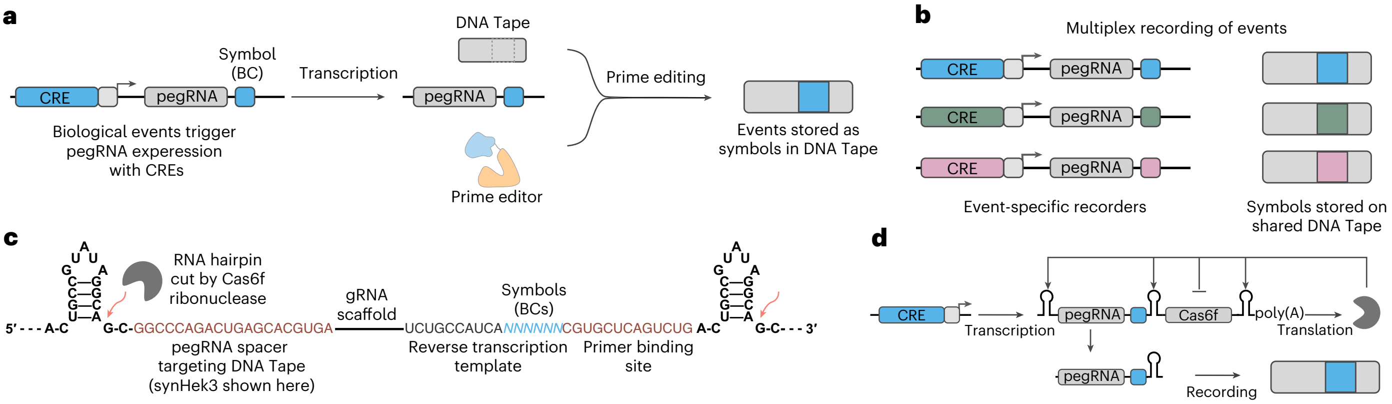

Fig. 1 | Architecture of the ENGRAM. a, A schematic of the ENGRAM. CRE activity drives the expression of a pegRNA encoding an insertional BC (that is, symbol) unique to the CRE. **b**, Multiplex or multichannel recording is achieved by constructing and concurrently introducing signal- or state-specific recorders, each encoding a unique symbol. The probability of a given symbol being written to the DNA Tape is a function of the expression level of its encoding pegRNA, which is in turn a function of CRE activity. **c**, The Pol-2-driven pegRNA is flanked by two 17-bp *cas6f* (*csy4*) hairpins and can be released by the endonuclease Cas6f (Csy4). All pegRNAs share the same spacer sequence (targeting a shared DNA Tape) and vary only in the encoded insertion (symbol). **d**, A feed-forward loop and an AND-gate circuit in 5′-ENGRAM that potentially explains the improved signal-to-noise that we observe relative to alternative architectures. Cas6f cleaves its own mRNA, resulting in the release of pegRNA and inhibition of further Cas6f production. The arrows represent activation and the bars inhibition. Part **c** is adapted from ref. 19, CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

sequences, but the number of enzymes quickly becomes limiting. The CRISPR–Cas9 system simplifies this by allowing the design of unique guide RNA (gRNA)-target pairs, with each pair representing a distinct signal. However, recording different signals to different locations makes it more challenging to recover information and also to make inferences about how these signals relate to one another in terms of order and strength. By contrast, PE enables one to write a large number of unique symbols, in the form of insertional edits, to a shared location(s). This greatly simplifies both the writing and reading of DNA-based records.

Second, the writing of a given symbol must be conditional on the activity of its associated biological signal. gRNAs, including pegRNAs, are conventionally expressed in mammalian systems under the control of constitutive Pol-3 promoters (for example, the U6 promoter), which lack inducibility in response to biological states or signals. While Pol-2 promoters permit conditional gene expression, the resulting transcripts undergo modifications such as 5′ capping and 3′ polyadenylation, which may impede PE activity. Strategies to release functional pegRNA from Pol-2 transcripts include using self-cleaving ribozymes[23], transfer RNA processing machineries[24] and ribonucleases[25]. Although these or other alternatives are probably possible, for ENGRAM's proof-of-concept, we used the ribonuclease Cas6f (also known as Csy4), which specifically recognizes and cleaves a 17-base pair (bp) *cas6f* hairpin structure (Fig. 1c). Constitutive expression of Cas6f, independent of the pegRNA, enabled recording but also resulted in high background activity. To address this issue, we designed co-expressing architectures in which the pegRNA is embedded into the untranslated regions of the transcript of Cas6f (Fig. 1d). Given that Cas6f acts as a single-turnover enzyme[26], this design creates a coherent feedforward loop with an AND gate, effectively reducing background activity while maintaining strong signal responsiveness (Fig. 1d). Although we continue to explore additional ENGRAM architectures, we currently recommend Cas6f/*cas6f* and 5′-ENGRAM (pegRNA embedded in the 5′ untranslated region) for its high signal-to-noise and simplicity in cloning.

Applications of ENGRAM

Thus far, we and our colleagues have applied ENGRAM to (1) recording the activity and orders of signaling pathways in mammalian cells and (2) concurrently recording dozens to hundreds of CRE activities. We have shown that signaling transductions that regulate transcription, such as exposure to a Wnt-pathway agonist, can be quantitatively recorded with ENGRAM (Fig. 2a,b). To record signaling events or cell states, one only has to pair each signal- or state-specific CRE

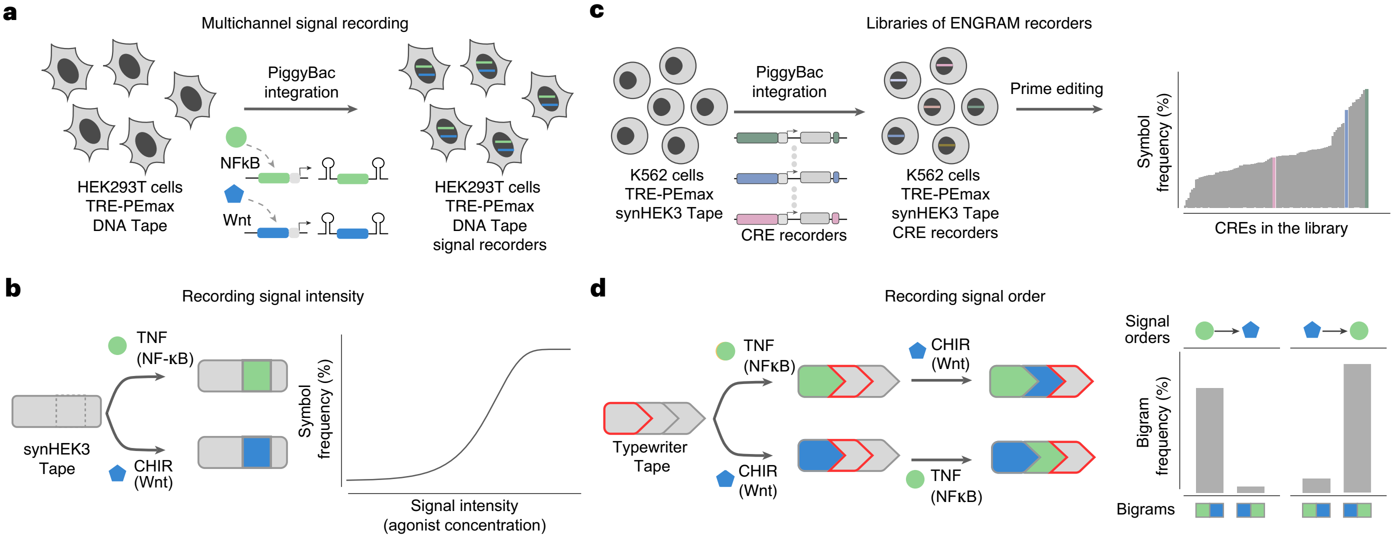

Fig. 2 | Applications of the ENGRAM. a, An illustrative example of components for multichannel signal recording with ENGRAM, including a cell line (for example, HEK293T or K562) or organism bearing an inducible writer (for example, Dox-inducible PEmax), signal-specific ENGRAM recorders (for example, NFκB and Wnt recorders shown in green and blue, respectively) and one or many copies of DNA Tape (for example, synHEK3 or DNA Typewriter). We generally introduce these components through PiggyBac-based random integration, although targeted integration is possible. **b**, An illustration showing recording signal intensity. In dose–response experiments conducted over a fixed time frame, we have found that the relative frequencies at which a signal-specific symbol in DNA Tape are observed is a nonlinear function of the concentration of the corresponding agonist. **c**, A library screening of CRE activities. A library of CREs can be associated with specific pegRNA-encoded insertions during cloning. The resulting library of ENGRAM recorders can then be integrated to target cells bearing an inducible writer (for example, Dox-inducible PEmax) and one or many copies of DNA Tape (for example, synHEK3 or DNA Typewriter). The relative activities of the many CREs can be quantified by sequencing the abundance of their corresponding symbols in DNA Tape, that is, a massively parallel recorder assay rather than an MPRA. **d**, As proof-of-concept of recording the order of signals, cells bearing pre-integrated inducible writer, DNA Typewriter Tape and two signal-specific ENGRAM recorders, were treated with the corresponding agonists in various patterns, the simplest pattern being to simply switch from one agonist to the other, or vice versa. We found that these and other ordering patterns were readily distinguishable on the basis of the bigrams of symbols observed in an ensemble of DNA Typewriter Tapes. Parts **a**,**c**,**d** are adapted from ref. 19, CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

with a pegRNA-encoded insertional symbol. This is readily achievable in a library format, as hundreds of ENGRAM recorders, each bearing a different CRE, can be concurrently cloned in *cis* with CRE-specific pegRNAs and used in a single experiment (Fig. 2c). The number of CREs whose activities can be distinguished within the context of a single system or experiment increases exponentially with the length of the inserted symbols. For example, a 4-bp symbol enables unique assignment of up to 256 (4, ref. 4) CREs, while a 5-bp symbol permits unique assignment of up to 1,024 (4, ref. 5) CREs. The presence and magnitude of a given signal can be inferred from the abundance of the corresponding symbol in DNA Tape (Fig. 2b,c). Furthermore, combining ENGRAM with DNA Typewriter[18] enables one to record the order of CRE activities by structuring the DNA Tape in a manner that is sequentially editable (Fig. 2d).To demonstrate this concept, we showed that various patterns of two signals are distinguishable from one another by the patterns of 'bigrams' that appear in DNA Typewriter Tape (DTT)[19].

In our proof-of-concept report of ENGRAM[19], we demonstrated its potential to reliably record several orthogonal signaling pathways (for example, Tet-On, NF-kB, Wnt) and its compatibility with various mammalian in vitro models, including HEK293Ts, K562s and mouse ES cell (mESC)-derived gastruloids. PE itself remains a relatively recent development in genome engineering, and we anticipate that ENGRAM will be readily adaptable to other models in which PE has been demonstrated, including other cancer cell lines, human organoids, mice[27–30] and possibly other organisms. In our hands, we have had success with ENGRAM in the cell lines described above, as well as mESCs, mESC-derived multicellular gastruloids and embryoid bodies. We and others are actively pursuing ENGRAM recording in mice in vivo and other multicellular systems in vitro, including human induced pluripotent stem cell-derived organoids. Potential challenges to such adaptations include that Cas6f toxicity may be limiting in some systems (discussed further below) and delivery and/or silencing of recording machinery over time. If encountered, such challenges can in principle be overcome with alternative gRNA release mechanisms[23,31], improved and/or multipronged delivery strategies[32] or consolidating all recording machinery in optimized configurations at silencing-resistant safe-harbor loci[33]. We also anticipate that the basic architecture can be adapted to myriad CREs that are either known to be, or can be engineered to be, biologically conditional. These include other signaling pathways, cell types, cell states and possibly even signals that do not endogenously result in transcriptional changes (for example, synNotch-triggered Pol-2 transcription of ENGRAM-released pegRNAs to record specific cell–cell or protein–protein interactions of interest[34,35]).

Comparison with other methods

Signal recording

The rapidly evolving toolkit of genetic engineering methods has enabled the permanent recording of transcriptional activity into genomic loci to label a population of cells over time. Although a substantial base of research has developed molecular recording tools in bacteria[8,9,36], here we focus exclusively on molecular recording in mammalian systems.

Molecular recording technologies that rely on the heritable gain or loss of a fluorescent marker in response to a signal-driven genome editor, for example, site-specific recombinases, are limited in ability to multiplex by the number of spectrally distinct fluorescent channels and by the lack of orthogonal enzymes that target distinct sequences[11,37,38]. Techniques utilizing DNA base editors have demonstrated event-triggered molecular recording either through direct sequencing of the edited locus or successive rounds of fluorescence in situ hybridization probe hybridization of the edited locus or loci[6,7,15]. These methods may additionally provide information about the sample’s spatial architecture if they rely on an optical readout, such as fluorescence in situ hybridization.

Unlike recorders based on DSBs, ENGRAM utilizes a nickase and exclusively relies on Pol-2-driven pegRNAs for recording, and so can be used to simultaneously and sequentially record many different events written as unique insertions to a shared DNA Tape. Base editors can also utilize Pol-2-transcribed guides, but a given recording site for base editing can only exist in one of two states, such that each signal of interest would require its own set of recording sites for multiplex signal recording. The multiplexing capacity of ENGRAM is primarily limited by PE efficiency, as dividing PE capacity between many different elements may make it more challenging to distinguish signal from noise. However, PE efficiency continues to improve with advances in the field, including optimizing the prime editor protein, modulating endogenous cellular factors that inhibit or activate PE and stabilizing the pegRNA[39–42].

Other tools to record cellular events in mammalian cells include circuits to record high and low microRNA (miRNA) expression through miRNA-mediated cleavage of a gene or endoribonuclease from a messenger RNA (mRNA) backbone. This results in the selective destabilization or stabilization of desired transcripts[43]. Intriguingly, this was utilized not only to report on miRNA levels but also to drive cell state transitions through the selective stabilization of RUNX1, a transcription factor that guides pluripotent cells toward a hematopoietic lineage. As ENGRAM similarly relies on endoribonucleases as tools to release effectors from mRNA structural elements, one can imagine analogous modifications that enable ENGRAM to encode signal-activated programs (for example, a CRE drives the production not only of a pegRNA that drives a recording event but also of a pegRNA or gRNA that mediates a cell fate-modulating genetic perturbation).

Library screening

Many of the technologies described above have been applied to, at most, the recording of two to three concurrent signals, enhancer activities or cellular events rather than high complexity libraries. By contrast, massively parallel reporter assays (MPRAs) are commonly used to measure the activity of thousands of different CREs in multiplex[44]. In the MPRA framework, CRE activity results in the transcription of RNA molecules linked to CRE-specific BCs. Recent advances in the field have developed MPRA constructs capable of distinguishing CRE strength on a single-cell level by utilizing a built-in control for construct abundance, a U6-driven circular

BC that is highly captured by 10× 3′ single-cell RNA sequencing (RNA-seq) and used to normalize a Pol-2 enhancer-driven BC on the same construct[45]. In the MPRA paradigm, the measurements of enhancer strength are transiently reported rather than permanently recorded, which may make it more challenging to interpret cellular history.

By contrast, ENGRAM records the activity of CRE and signal-response elements permanently either to endogenous genomic loci or integrated synthetic DNA Tape (for example, synHEK3 Tape or DTT). We previously demonstrated that ENGRAM is capable of recording the activity of at least hundreds of endogenous CREs and synthetic transcriptional factor-binding CREs in mammalian cells, including in a model of early mammalian development[19]. The CRE activities recorded on DNA Tape showed a strong correlation with those measured by RNA-seq. We further demonstrated that cells containing multiple synthetic DNA Tape integrations exhibit increased recording capacity, thereby reducing the number of cells needed to accurately recover CRE activities. Moreover, because ENGRAM permanently records activity in the DNA Tape, it can capture transient CRE activities that may be missed by MPRA, which only reflects activity at the time of measurement. Taken together, ENGRAM enables new possibilities for how we study the dynamics of signaling and CRE activity during complex biological processes such as embryonic development and cancer metastasis.

Recording the order of signaling events

ENGRAM pegRNAs can be engineered to target synthetic DNA Tapes that are sequentially editable, as in both the DNA Typewriter and peCHYRON systems[18,20]. In DNA Typewriter, a synthetic array of target sites is integrated into the cellular genome. Before any of these targets are edited, all but the first are truncated by 3 bp at their 5′-end—these are missing a short 3-bp sequence called the ‘key’, which is necessary for a successful PE event to occur as it completes the target sequence. When an insertion is edited into this synthetic array, the insertion is placed at the 5′-end of the next site of the synthetic array, and this insertion includes the key. The insertion of the key completes the target site of the next unit in this synthetic array, thus enabling its editing by PE machinery. We have shown that CRE-driven pegRNAs from the ENGRAM system can target this DNA Typewriter synthetic target array (for example, of a succession of monomers whose compatibility with editing is gated on the monomer 5′ to it being first edited), thus enabling us to record sequentially resolved signaling and CRE activities. The peCHYRON system differs from DNA Typewriter in that it sequentially inserts an event-specific sequence alongside the entire next DNA Tape to be edited, rather than a priori integrating the DNA Tape targets—this allows editing to continue indefinitely. While not yet demonstrated, we see no technical reason why ENGRAM-derived pegRNAs could not be used to enable Pol-2-driven sequential recording with peCHYRON DNA Tapes as well.

In addition to nucleic-acid-based mediums of sequentially resolved molecular recording, protein fibers have been used to record cellular events as well[46,47]. In these systems, both engineered and computationally designed self-assembling protein monomers containing epitope tags are expressed either constitutively (for example, through the UbC promoter) or in response to a cellular event by utilizing endogenous gene promoters or signal response elements (for example, *cFos*, *Egr1* and *Arc* promoters, or the CREB promoter). These monomers polymerize in cells into flexible, optically visible fibers that may be visualized via immunofluorescent imaging. Recent advances have resulted in the development of modular monomers capable of accommodating diverse epitopes that may all contribute to the same fiber, thus enabling multiplexed, sequentially resolved recording of cellular events over the course of days to weeks[48]. This recorded information inherently includes spatial context, as it is read optically from the cells. Compared with these methods, ENGRAM–DNA Typewriter offers several advantages. First, the DNA Tape provides a relatively compact storage medium in comparison with protein fibers that are large enough to be optically viewed. Second, all information stored to DNA Tapes across many cells with the ENGRAM–DNA Typewriter system may be read at once in the same sequencing run; to read out multiplexed signals that have recorded to the same protein fiber, multiple rounds of immunofluorescence must be performed, which becomes more time-intensive as the number of unique signals increases. Third, ENGRAM–DNA Typewriter can be readily scaled to sample millions of cells, whereas an optical readout is limited in field of view. Finally, we anticipate that ENGRAM–DNA Typewriter records can be adapted to spatial sequencing readouts as spatial sequencing tools proliferate.

Limitations of ENGRAM

Limitations of ENGRAM stem from current limitations in the field of genome and cell engineering, DNA synthesis and synthetic circuit design.

PE efficiency

The rate of recording with ENGRAM is intrinsically linked to the PE efficiency in the cellular context(s) of interest. Low PE efficiency can make it difficult to reliably detect transcriptional events, or discern signals from noise. Low PE efficiency can occur in some cell types owing to altered DNA repair mechanisms[39], nucleotide availability[49], among others, and can require optimization. Strategies to achieve high-efficiency PE in many diverse contexts have been reported[50] and are advancing rapidly. ENGRAM records reflect both signal intensity and duration, so precise determination of signal duration and intensity requires the use of molecular timestamps (that is, ENGRAM records with known absolute timepoints of induction to reconstruct timing of other records). While we do not know exactly how fast a given ENGRAM record can be installed after transcription is triggered, our current estimate is on the order of a few hours, on the basis of the general kinetics of genome editing in mammalian cells, and the fact that we could resolve the order of two signals delivered 24 h apart[19]. Thus, ENGRAM can be used to record many signals over long time periods, but it is unlikely to be well suited to decoding fast signal kinetics (for example, seconds, minutes).

Symbol length and number

The number of signals that can be simultaneously recorded, and the accuracy of recording for a given signal, is linked to the length and number of BCs that can be paired to each signal/CRE, delivered to the cells and written to the genome. Too few or too short BCs for a given library size can lead to inaccurate measurements or difficulty in demultiplexing records from independent signals (from errors in BC replication, insertion or readout). The longer the insertions that can be reliably achieved in a system, the more signals/CREs and BCs per signal/CRE (which improves measurement accuracy) can be simultaneously measured. Pooled ENGRAM screens of hundreds of signals/CREs have already been achieved in multiple systems[19], and improved PE efficiency along with more and longer BCs per signal/CRE should enable recording screens at the scale of MPRAs (hundreds of thousands of sequences)[51] in the near future.

Construct stability and design

ENGRAM, similar to many synthetic circuits, uses well-characterized parts (for example, gRNA scaffolds, Cas6f/*cas6f* hairpins, minimal promoters, among others) repeatedly in a modular fashion. This design not only enables accurate comparison of signal/CRE activity but also introduces repetition into construct designs. Repetitive constructs can pose challenges to DNA synthesis, cloning and long-term stability in cells[52–55]. Highly repetitive arrays of multiple ENGRAM constructs could lead to assembly failures, deletion or silencing of certain components. Problematic circuit designs in which individual ENGRAM cassettes are not properly distanced or insulated from one another could also cause crosstalk between enhancers[56] in neighboring ENGRAM units. Diversified parts for genome engineering and molecular recording are being developed (for example, diversified gRNA scaffolds capable of PE)[57], as are improved methods for large-scale DNA assembly[58,59], and modular cloning strategies to test diversified parts[60] that are all simplifying recorder design and assembly, which are in use in our laboratory.

Bioorthogonality

Cas6f may be cytotoxic in certain contexts[25], though this is mitigated in ENGRAM by the transient and highly destabilized nature of the ENGRAM transcript. PE and, consequently, ENGRAM does not rely on DNA DSBs the way that recorders based on Cas nucleases such as GESTALT do[17], and thus it may be less disruptive. Nonetheless, PE, similar to all genome editors and exogenous proteins, will inevitably have effects on cellular physiology[61] that should be monitored.

BOX 1

General protocol for one-pot GGA

1. Prepare the GGA reaction as follows:

	Volume (µL)	Final concentration
Plasmid	50 ng	4.17 ng/µL
Inserts	5–10 ng (5:1 molar ratio)	0.417–0.833 ng/µL
NEBridge Ligase Master Mix (3×)	4 µL	1×
BsaI-HFv2 (20 U/µL) or BsmBI-HFv2 (20 U/µL)	1 µL (20 units)	1.67 units
Nuclease-free H_2O	To 12 µL	

▲**CRITICAL STEP** The enzyme to be used in this reaction (BsaI-HFv2 or BsmBI-HFv2) will depend on the part of the cloning protocol and which specific construct is being cloned.

▲**CRITICAL STEP** The reaction volume can be adjusted from 6 µL to 20 µL depending on the complexity and amount of plasmid in the reaction. Alternatively, the Ligase Master Mix can be substituted with 400 U of T4 DNA ligase and 1× T4 ligase buffer.

2. Perform the GGA reaction as follows:

Restriction enzyme(s)	Thermocycling conditions
BsaI-HFv2, BbsI-HF	15 cycles at 37 °C for 1 min, followed by 16 °C for 1 min; 60 °C for 5 min to deactivate ligase
BsmBI-v2	15 cycles at 42 °C for 1 min, followed by 16 °C for 1 min; 60 °C for 5 min to deactivate ligase

▲**CRITICAL STEP** Consider using 30 cycles for thermocycling in complex libraries.

Silencing of recording machinery

ENGRAM requires at least two recording components, the prime editor and ENGRAM cassette, to maintain high transcription capability throughout the intended recording time period. ENGRAM circuits that use synthetic transcriptional factors to drive recording require additional components with stable expression. Silencing[33] of one or more of the recording components over time can lead to failure to record signals/events that are active at later timepoints and/or more differentiated cell states. Improved silencing-resistant constructs that enable delivery and stable expression of large payloads[62], along with improved assembly strategies (discussed above), enable consolidating recording machinery to single constructs that can all be integrated at safe harbor loci, which are less prone to silencing and simplify breeding/distribution (for example, for disease modeling).

Overview of the procedure

This Protocol provides detailed instructions for creating ENGRAM constructs and generating suitable cell lines for recording cellular events. The process involves: Procedure 1: cloning DNA Tapes with unique molecular identifiers (UMIs) (Steps 1–8) and generating a cell line with inducible PEmax expression and integrated DNA Tapes (Steps 9–21) and Procedure 2: cloning ENGRAM constructs with specific CRE and symbol pairings (Steps 1–23), generating cells with all recording components (Procedure 3, Steps 1–6; Procedure 4, Steps 1–6), recording signal intensities (Procedure 3, Steps 7–11) and orders (Procedure 3, Steps 12–17), and recording and screening CRE activities (Procedure 4, Steps 7–9). Recurrently used procedures such as Golden Gate assembly (GGA), general tissue culture and retrieval of recorded information from cells are summarized in Box 1–3, respectively. Related scripts and cloning templates can be downloaded from the ENGRAM GitHub repository (https://github.com/shendurelab/ENGRAM/tree/main/Nprotocol_scripts/ENGRAM).

ENGRAM experimental design

Several key considerations need to be addressed when designing ENGRAM experiments:

Choice of DNA writer

The efficiency of PE is critical for successful molecular recording with ENGRAM. At present, PEmax is the recommended default owing to its high efficiency.

BOX 2

General protocol for mammalian cell culture (HEK293Ts and K562s)

Procedure

1. Thaw a frozen stock vial of HEK293T or K562 cells in a 37 °C water bath until only a small chunk of ice remains.
2. With gentle pipetting, resuspend the cells in 4 mL of complete media, depending on the cell line. HEK293Ts use 90% DMEM and 10% FBS. K562s use 90% RPMI and 10% FBS.
3. Transfer this mixture of cells and media to a 15 mL Falcon tube and centrifuge for 5 min at 300*g* (relative centrifugal force) for 5 min.
4. Aspirate the supernatant from the cell pellet carefully and resuspend in 10 mL of full media.
5. Count the concentration of live cells per milliliter by using a cell counting method or instrument.
6. Aiming for 10,000–20,000 cells per square centimeter of a culture plate, dish or flask, select the size plate most appropriate for the number of cells. If the concentration is too high, dilute the cells further in the full media. We recommend a culture dish for HEK293Ts and a flask for K562s. Distribute the cells across the plate by gently shaking the cells, without disturbing them so that they spill, across both the horizontal and vertical axes.
7. Put these cells in a cell culture incubator under the following conditions: 37 °C and 5% CO_2.
8. Monitor the cells as they grow by inspecting them under a brightfield microscope.
9. When the cells have reached >70% confluence after 2–3 d post-thawing, split them onto a new plate, as described below, either for HEK293Ts (adherent) or K562s (suspension). For K562s, you may take a 1:3–1:5 aliquot of the cells in media and add that to a new flask of full K562 media. Steps 10–15 below refer to HEK293T splitting.
10. Aspirate the media and wash once with 1× PBS with a volume appropriate for the plate area, so that the PBS completely covers the cells.
11. Aspirate the PBS and add a volume of TrypLE Express Enzyme just sufficient to cover the bottom of the culture dish. For example: 1 mL of TrypLE will just cover the bottom of a 10 cm dish, and 300 µL is sufficient for one well of a 6-well plate.
12. Place in a 37 °C incubator for 5 min.
13. Inactivate the TrypLE by adding greater than four volumes of HEK293T full cell medium, as described in Step 2 of this box. Pipette the cells gently to ensure that they are fully dissociated.
14. Centrifuge this mix at 300*g* (or relative centrifugal force) for 5 min.
15. Aspirate the media and resuspend in 3–5 mL of media. Take 1 mL of this mix to split the cells 1:3–1:5, depending on confluency, and add to a new plate with a sufficient volume of full media.

To freeze the cells for long-term storage:

16. Start from K562s at Step 9 and spin them down at 300*g* for 5 min, or start from Step 14 for HEK293Ts. Both will result in cell pellets.
17. Resuspend in the cell pellet in a sufficient volume of freezing media (recommended 1 million cells per vial, with 1 mL per vial). Freezing media is 90% full media of either HEK293Ts or K562s and 10% DMSO.
18. Distribute aliquots into prelabeled cryovial tubes and freeze these in a CoolCell container at −80 °C for >8 h or overnight.
19. Transfer these frozen vials to liquid nitrogen for long-term storage.

ENGRAM architecture

Both 5′-ENGRAM and 3′-fine tuned (FT)-ENGRAM exhibit strong response to stimulation. 3′-FT-ENGRAM exhibits lower background noise and slightly higher activation, potentially improving the dynamic range of recording. However, for ease of cloning and minimized CRE–BC distance, we generally recommend 5′-ENGRAM. Throughout this protocol and in our ongoing work, we are using 5′-ENGRAM. Example cloning files for both 5′-ENGRAM and 3′-FT-ENGRAM can be downloaded from the ENGRAM GitHub repository.

Choice of cell line

While transient transfection of ENGRAM recorders and prime editor plasmids can be used for signal recording and library screening, we recommend generating a monoclonal cell line with multiple copies of DNA Tapes and stable expression of PEmax for more robust and consistent results. ENGRAM has been successfully implemented in HEK293T, K562 and mESCs[19]. This protocol uses HEK293T cells as an example for generating a stable cell line suitable for recording experiments. Specific culturing conditions for other cell lines are beyond the scope of this protocol.

Recording order of events

Determine whether recording the temporal order of events is necessary. ENGRAM can be integrated with DNA Typewriter technology to capture the sequence of signals, with a cost of reduced recording efficiency compared with targeting the synHEK3 Tape or endogenous HEK3

BOX 3

Retrieval of recording information from DNA Tape: lysis and PCR

This section details the steps for lysing cells and extracting DNA for PCR amplification (Fig. 3) of DNA Tapes.

1. Determine the number of cells using a cell counter (for example, using Countess automated cell counter).
   ▲**CRITICAL STEP** Accurate cell counting is crucial for standardizing the DNA extraction process.
2. Lyse the cells at a concentration of 1 million cells per 200 μL of lysis buffer. Gently resuspend the cell pellet in the lysis buffer by pipetting up and down several times. Ensure the pellet is completely dispersed to facilitate efficient lysis.
3. Incubate the cell lysate at 55 °C for 1 h, followed by an 80 °C enzyme inactivation step for 30 min.
   Cell lysis at high cell concentrations will initially result in a viscous solution. Once the digestion is complete, the solution should return to normal viscosity.
   ▲**CRITICAL STEP** The described lysis buffer method is also suitable for precious samples with low cell numbers, such as organoids. Estimate the cell number and adjust the volume of lysis buffer accordingly.
   ▲**CRITICAL STEP** Alternatively, gDNA can be extracted using the DNeasy Blood and Tissue Kit (Qiagen) or equivalent, following the manufacturer's instructions.
4. PCR-1: prepare a reaction to amplify the synthetic Tape using Tape_ampup_fwd and Tape_ampup_rev (Table 1), with 2× Kapa Robust DNA polymerase 2× master mix as follows:

	Volume (μL)	Final concentration
Kapa2G Robust HotStart ReadyMix	25	1×
10 μM Tape_ampup_fwd	2	0.4 μM
10 μM Tape_ampup_rev	2	0.4 μM
Cell lysis (Step 3 of this Box)	4 (equivalent 20,000 cells)	400 cells per microliter
100× SYBR green	0.5	1×
Water	32	
Total	50	

   ▲**CRITICAL STEP** Remove quantitative PCR (qPCR) samples when the amplification curve begins to rise (early in the exponential phase), typically between 18 and 22 cycles, depending on the integrated DNA Tape copy number of the selected clone. We suggest using qPCR, not standard PCR, in the first few iterations of the experiment to establish an appropriate number of cycles with which to amplify the DNA—enough that the amplification curve enters its exponential phase but is not overamplified. Once the monoclonal line is established and the required PCR cycle number is confirmed, standard PCR with a fixed number of cycles can replace qPCR. Overamplification will introduce PCR bias and require deeper sequencing to recover information. We suggest having a lower number of cycles for PCR-1 (Box 3, Step 5).
5. Run the PCR as follows:

Cycle number	Denature	Anneal	Extend
1	95 °C, 60 s		
18–22 cycles	95 °C, 10 s	65 °C, 15 s	72 °C, 30 s
23	72 °C, 60 s		

6. Confirm precise amplification of expected amplicons using TapeStation or gel electrophoresis. For TapeStation, mix 1 μL of PCR product with 3 μL of High Sensitivity D1000 Sample Buffer and run the assay by following the manufacturer's instructions. For gel electrophoresis, mix 5 μL of PCR product with 5 μL of water and 2 μL of 6× loading dye on 1% agarose gel (wt/vol) at 80 V for 30 min.
7. PCR-2: to add sequencing adaptors P5_bc_Nextera and P7_bc_Truseq (Table 1) to PCR-1 product, prepare the following PCR mix:

	Volume (μL)	Final concentration
Kapa2G Robust HotStart ReadyMix (2×)	12.5	1×
10 μM P5_bc_Nextera	1	0.4 μM
10 μM P7_bc_Truseq	1	0.4 μM
PCR-1 product	1	
Water	9.5	
Total	25	

8. Run the PCR using the following conditions:

Cycle number	Denature	Anneal	Extend
1	95 °C, 60 s		
5 cycles	95 °C, 10 s	65 °C, 15 s	72 °C, 30 s
6	72 °C, 60 s		

   ▲**CRITICAL STEP** Ensure PCR-1 is not overamplified. The overamplification of PCR-1 will lead to the incomplete adaptor attachment during PCR-2, resulting in two bands on a gel.
9. Pool 10 μL of PCR-2 product from each sample into a single tube, then take 50 μL of this pooled sample and perform AMPure XP bead cleanup according to the manufacturer's instructions.
   ▲**CRITICAL STEP** Mixing samples after PCR-2 for joint purification relies on a key assumption: that using the same amount of cell lysate for PCR-1 yields a similar concentration of PCR-1 products, which in turn leads to similar concentrations after PCR-2. An alternative method is to purify the PCR-2 products individually using AMPure XP beads (following the manufacturer's instructions), quantify them with NanoDrop spectrophotometer and then pool the purified products at equal molar ratios.

(continued from previous page)

10. Quantify the size and concentration of amplicons using TapeStation. Mix 1 μL of purified PCR product with 3 μL of High Sensitivity D1000 Sample Buffer and run the assay by following the manufacturer's instructions. The length of PCR-2 should be ~330 bp.

    ▲**CRITICAL STEP** For concentrations greater than 10 ng/μL, dilute the purified PCR product 5–10× with nuclease-free water and repeat the measurement.

    ▲**CRITICAL STEP** AMPure bead cleanup effectively removes primer dimers. However, if two bands are present, gel extraction is recommended to purify the band at ~330 bp.

    ▲**CRITICAL STEP** If TapeStation is not available, the size can be identified using agarose gel electrophoresis, and the concentration can be measured by Qubit.

11. Dilute the sample to 2 nM with nuclease-free water and load it to the Illumina Nextseq 1000/2000 system (or similar high-throughput sequencing instruments), following the manufacturer's instructions.

    ▲**CRITICAL STEP** Typically, the P1 100 cycle kit should cover the full length of the synHEK3 Tape (using single-end sequencing), and the P1 300 cycle kit should cover the full length of the 5X-DTT. While the recording efficiency of later Tapes is low, the P1 100 cycle kit with single-end sequencing should cover information stored in the first three Tapes as well.

12. Use the analysis script (ENGRAM_nprotocol.ipynb) to calculate the recording efficiency of each clone. Estimate the copy number of synTapes in each clone by counting the number of unique intBCs (integration BCs).

    ▲**CRITICAL STEP** For synHEK3, the inserted BC is extracted using pattern matching, and its frequency is quantified. For 5X-DTT, sequencing reads are first aligned to a reference sequence, and a custom Python script is used to extract information and calculate recording efficiency.

locus. We have made plasmids for both synHEK3 Tape and DTT available (see 'Plasmids' section in Materials). The two-step cloning strategy is suitable for generating ENGRAM recorders targeting either HEK3 or DTT sites. Example cloning files can be downloaded from the ENGRAM GitHub repository.

Selection of CREs

Carefully select the CREs for signal recording and screening, as well as the symbols to be associated with these CREs. Certain CREs exhibit low baseline activity and high induced activity (which are better suited to recording the presence or absence of an event), while others have high baseline activity and higher stimulated activity (which can be better for continuous recording of signaling dynamics). CREs can also vary in size by orders of magnitude (typically ~100–1,000 bp). Longer endogenous sequences defined by chromatin accessibility may better reflect endogenous CRE activity, while shorter synthetic CREs may yield more specific, tailorable responsiveness.

Selection of symbols

The choice of symbols (that is, inserted BCs) should be based on the complexity of the library. A Hamming distance of 1 is sufficient to distinguish different BCs using this protocol. We have observed variations in recording efficiency among different BCs and have curated lists of 5-bp and 8-bp insertion efficiencies[19,57]. For recording a limited number of signaling pathways, 5-bp BCs should suffice. For screening thousands of CREs, we recommend using 8-bp BCs to mitigate BC collapse.

Control selection and interpretation

If the pool of CREs has less than ~5,000 elements, we recommend two or more BCs per CRE to help control for insertion differences between BCs (that is, the impact of such differences is mitigated by the calculation of edit scores, but consistent results between edit scores derived from independent BCs will add confidence). In general, longer and more independent insertion BCs per CRE will improve recording accuracy but potentially at the cost of reduced recording efficiency. We have recorded libraries of hundreds of CREs with 5N BCs and libraries of thousands of CREs with 8N BCs[19]. We also strongly recommend spiking in 'empty' ENGRAM recorders with a minimal or no promoter driving one or several BCs as a negative control to determine the noise floor (for example, with no minimal promoter (minP), with minP alone, and/or with yeast or bacterial elements that should not be active in mammalian cells). Active CREs should drive recording levels higher than those of these negative controls[19]. The assessment of whether a given CRE is active should be performed with appropriate statistics (for example, Wilcoxon rank-sum test[63,64] with Benjamini–Hochberg adjusted *P* value[65] or DEseq2 (ref. 66)) that take into account independent BCs, technical replicates, among others. However, as a subjective ballpark, we look for activity at least two-fold greater than minP negative controls. Where needed, positive controls for ENGRAM can take the form of recorders driven by constitutive promoters (for example, cytomegalovirus) or a doxycycline (Dox)-inducible CRE[19]. However, a caution is that if positive controls are too active, they can dominate the set of recorded events and potentially compromise quantification of signal recording.

Quantifying Tape recorders with UMIs

UMIs can be used to improve quantitation of recorded information. However, as the DNA Tape is usually abundant (that is, many copies per genome equivalent) and PCR amplification is targeting the same sequence, our experience so far suggests that PCR introduces minimal amplification bias, provided that one begins with sufficient genomic DNA (gDNA) (we recommend >5,000 cells in each reaction). As attaching UMIs to gDNA requires an additional step before PCR and complicates downstream analysis, we do not typically use UMI-based approaches when reading out ENGRAM from bulk gDNA. We note, however, that UMIs are a standard component of most single-cell molecular profiling protocols, and as such, are straightforward to take advantage of when ENGRAM records are being read out by single-cell RNA-seq or when library complexity is limiting for any other reason.

Primer and oligo design

Users of this Protocol should not need to design any new primers for the experiments outlined. We have designed universal primers and oligos for cloning the DNA Tape constructs and the ENGRAM constructs and/or libraries, which will be suitable for experiments targeting both synHEK3 and Typewriter Tapes. These primers were designed for high-efficiency amplification with the Q5 high-fidelity polymerase. We have also designed primers for checking the complexity of CREs in the cloned ENGRAM library and for assessing swapping rates between CREs and their assigned BC insertions. Finally, for recovery and sequencing of the Tapes, we have designed primers to amplify the Tapes with Illumina adapters for use with the KAPA Robust polymerase. The primers and oligos described are all presented in Table 1.

Materials

Biological materials

- New England Biolabs (NEB) Stable Competent cells (NEB, cat. no. C3040H)
- NEB 10-beta Electrocompetent cells (NEB, cat. no. C3020K)
- HEK293T or desired target cell line (ATCC, cat. no. CRL-3216, RRID: CVCL_0063)
- K562 or desired target cell line (ATCC, cat. no. CCL-243, RRID: CVCL_0004)

  ▲ **CAUTION** Cell lines should be regularly screened to ensure that they are authentic and free of *Mycoplasma* contamination.

Reagents

Plasmids

- Super PiggyBac Transposase expression vector (System Bioscience, cat. no. PB210PA-1 (or equivalent))
- PiggyBac-TRE-PEmax-T2A-mCherry-PGK-rtTA-P2A-Puro (Addgene, cat no. 239974)
- PiggyBac-ENGRAM backbone (Addgene, cat no. 239972)
- PiggyBac-synHEK3-Tape (Addgene, cat no. 239975)
- PiggyBac-5x-Typewriter-Tape (Addgene, cat no. 239983)

Cloning of DNA plasmids

- Nuclease-free water (Invitrogen, cat. no. AM9937)
- Q5 High-Fidelity 2× Master Mix (NEB, cat. no. M0492L)

Table 1 | Primers and gene fragments used in this Protocol

Step	Entity ID	Sequence (5′ → 3′)	Description
Procedure 1: Step 1	BsaI_15N_fwd	GAAggtctcACACCNNNNNNNNNNNNNNNgcattgagaccCTCCAGCCCTGGCCTGGGTC	Primer with 15Ns to assign unique intBC for synTapes
Box 3	Tape_ampup_fwd	CGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNNGGCTATATCTTGTGGAAAGGACGAAACACC	Forward primer amplifying the synTape with the Nextera adaptor
Box 3: Procedure 1, Step 1	Tape_ampup_rev	CTGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGCAGGAGCTGCACATACTACCAAG	Reverse primer amplifying the synTape with the Truseq adaptor
Box 3	P5_bc_Nextera	AATGATACGGCGACCACCGAGATCTACACNNNNNNNNNNTCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	P5_bc_Nextera Illumina sequencing primer, replace NNNNNNNNNN with unique BC
Box 3	P7_bc_Truseq	CAAGCAGAAGACGGCATACGAGATNNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	P7_bc_Truseq Illumina sequencing primer, replace NNNNNNNNNN with unique BC
Procedure 2, Step 2	CRE_fwd	GTGGAAAGGACGAAACACCG	Forward primer to amplify CREs from an oligo pool
Procedure 2, Step 2	CRE_rev	GCTAAGATCGACAGCTGCCT	Reverse primer to amplify CREs from an oligo pool
Procedure 2, Step 10, 20	gblock_minP-pegRNA-scaffold	CACCGcgtctcGCACTAGAGGGTATATAATGGAAGCTCGACTTCCAGCTTGGCAATCCGGTACTGTAtcGGTactgccgtataggcagcGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCAgagacgAGGCA	Cloning minP and pegRNA scaffold (targeting HEK3 site) into Step 2 CRE–BC acceptor with BsmBI
Procedure 2, Step 10, 20	gblock_minP-pegRNA-scaffold	CACCGcgtctcGCACTAGAGGGTATATAATGGAAGCTCGACTTCCAGCTTGGCAATCCGGTACTGTAtcGGTactgccgtataggcagcGGATGATGGTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCAgagacgAGGCA	Cloning minP and pegRNA scaffold (targeting DTT site) into Step 2 CRE–BC acceptor with BsmBI
Procedure, Step 19	CRE-BC_nexteraR1_fwd	AGATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNNNNNATCTCGTATGCCGTCTTCTGCTTG	Forward primer to amplify CRE-symbol pairs from pool of plasmid
Procedure 2, Step 19	CRE-BC_truseqR2_rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTatgtcgaggtagtggtccatggtggc	Reverse primer to amplify CRE-symbol pairs from pool of plasmid

Upper and lower case in sequences were copied from the Benchling file; multiple repeats of ‘N’ are a UMI to mark unique molecules during amplification (to assist with analysis and quantification). intBC, integration BC (to mark uniquely integrated Tapes).

- KAPA2G Robust 2× Hotstart mix (Kapa, cat. no. KK5702)
- SYBR Green (Thermo Fisher, cat. no. S7563)
- Agencourt AMPure XP (Beckman Coulter, cat. no. A63882)
- Luria–Bertani (LB) medium (Thermo Fisher, cat. no. 12795027)
- LB agar powder (Thermo Fisher, cat. no. 22700025)
- Ampicillin sodium salt (Thermo Fisher, cat. no. 11593027)
- Monarch Spin Plasmid Miniprep Kit (NEB, cat. no. T1110S)
- ZymoPure Midi kit (Zymo, cat. no. D4201)
- BsaI-HF-v2 (NEB, cat. no. R3733S)
- BsmBI-HF-v2 (NEB, cat. no. R0739S)
- BbsI-HF (NEB, cat. no. R3539S)
- 10× rCutSmart buffer (NEB, cat. no. B6004S)
- 10× NEBuffer r3.1 (NEB, cat. no. B6003S)
- NEBridge Ligase Master Mix (NEB, cat. no. M1100S)
- PCR primers and gene fragments (Table 1) (Integrated DNA Technologies, single-stranded DNA oligos (for PCR primers) and gBlocks Gene Fragments (for gene fragments))
- CRE–BC oligo pools (Twist Bioscience or Integrated DNA Technologies)
- UltraPure Agarose (Thermo Fisher, cat. no. 16500-500)
- T4 DNA ligase (NEB, cat. no. M0202S)
- 10× T4 DNA ligase reaction buffer (NEB, cat. no. B0202S)
  ▲ **CRITICAL** Multiple freeze–thaw cycles will cause the adenosine triphosphate in the 10× T4 DNA ligase reaction buffer to degrade. Divide into aliquots of working volumes to minimize freeze–thaw cycles, and store as recommended by the manufacturer.

Mammalian cell culture

- Dulbecco’s modified Eagle medium (DMEM; Thermo Fisher, cat. no. 119650192)
- Roswell Park Memorial Institute medium 1640 (RPMI; Thermo Fisher, cat. no. 11875093)
- Fetal bovine serum, characterized (FBS; Cytiva, cat. no. SH30396.03)
- Penicillin–streptomycin (100× concentration) (Gibco, cat. no. 15140-122)
- Puromycin dihydrochloride (Gibco, cat. no. A1113803)
- Blasticidin S HCl (Thermo Fisher, cat. no. A1113903)
- Dulbecco’s phosphate-buffered saline 1× (PBS; Gibco, cat. no. 14190-144)
- TrypLE Express Enzyme (Thermo Fisher, cat. no. 12604013)
- Lipofectamine 3000 transfection reagent kit (Thermo Fisher, cat. no. L3000001)
- Opti-MEM reduced serum medium (Thermo Fisher, cat. no. 31985062)
- CHIR-99021(Selleck, cat. no. S2924)
- Recombinant Human TNF Protein (R&D Systems, cat. no. 210-TA-005/CF)
- Dox hyclate (Sigma, D9891)
- Proteinase K (Thermo Fisher, cat. no. EO0491)
- UltraPure 1M Tris–HCl, pH 8.0 (Thermo Fisher, cat. no. 15568025)
- Ultrapure sodium dodecyl sulfate (SDS) 10% solution (Thermo Fisher, cat. no. 15553027)
- (Optional) Dneasy blood and tissue kit (Qiagen, cat. no. 69504)
- SF Cell Line 4D-Nucleofector X Kit S (Lonza, cat no. V4XC-2032)

Sequencing library preparation

- NanoDrop spectrophotometer (Thermo Fisher, cat. no. ND-8000-GL or DNA concentration quantification tool alternatives)
- Agilent 4200 TapeStation (Agilent, cat. no. G2991BA or DNA size quantification tool alternatives)
- Illumina NextSeq 1000/2000 system (Illumina, cat. no. 20038897 or alternatives)

Equipment

- Magnetic 96-well PCR plate (Beckman Coulter, cat. no. A32782 or similar)
- Filtered pipette tips, assorted (Fisher, cat. nos. 12-111-132, 76175-406, 12-111-002 and 12-111-000)
- Serological pipettes, assorted (Eppendorf, cat. nos. 30127714, 30127722, 30127730 and 30127749)
- Standard microcentrifuge tubes (Eppendorf, cat. no. 30108418)
- Standard PCR eight-strip tubes (USA Scientific, cat. no. 1402-4700)
- Standard cryogenic vials for liquid nitrogen storage (VWR, cat. no. 89089-764)
- CoolCell (Corning, cat. no. 432000)
- Vortex mixer (VWR, cat. no. 10153-838)
- Benchtop microcentrifuge (Eppendorf, cat. no. 5405000441)
- Countess automated cell counter (Thermo Fisher, cat. no. AMQAX2000)
- BD FACS Symphony S6 cell sorter (or alternative)
- Gel electrophoresis system (PowerPac Basic Power Supply, Bio-Rad, cat. no. 1645050 and Sub-Cell GT Horizontal Electrophoresis System, Bio-Rad, cat. no. 1704401)
- Gel imaging system (GelDoc Go, Bio-Rad, cat. no. 12009077)
- Light microscope with red fluorescent protein, cyan fluorescent protein and yellow fluorescent protein excitation laser and filter (Zeiss Axio Observer 3 or similar)
- 4D-Nucleofector System (Lonza, cat. no. AAF-1003)

Reagent setup

DMEM (with 10% (vol/vol) heat-inactivated FBS)

Supplement DMEM with 10% (vol/vol) heat-inactivated FBS and 1% (vol/vol) penicillin–streptomycin. Store at 4 °C for up to 3 months.

RPMI 1640 (with 10% (vol/vol) heat-inactivated FBS)

Supplement DMEM with 10% (vol/vol) heat-inactivated FBS and 1% (vol/vol) penicillin–streptomycin. Store at 4 °C for up to 3 months.

Cell lysis buffer

To prepare 50 mL of cell lysis buffer, combine 500 μL of 1 M Tris–HCl (pH 8.0) and 250 μL of 10% SDS. This solution, without Proteinase K, can be stored at room temperature (25 °C) for up to 6 months. Immediately before cell lysis, add Proteinase K to a final concentration of 40 μg/mL by adding a 1:500 (vol/vol) dilution of a Proteinase K stock solution (Thermo Fisher, ~20 mg/mL). The final lysis buffer composition will be 10 mM Tris–HCl (pH 7.5), 0.05% SDS and 40 μg/mL Proteinase K.

Ampicillin LB agar plates

To prepare 500 mL of ampicillin LB agar plates (~20–25 plates), first prepare 500 mL of LB agar base by dissolving 12.5 g of LB agar powder in 500 mL of water and sterilizing by autoclave. Cool the sterile agar to 50–55 °C. Immediately before pouring, add ampicillin to a final concentration of 100 μg/mL by adding 500 μL of a 100 mg/mL ampicillin stock solution. The final plate composition will be 1× LB agar and 100 μg/mL ampicillin. Store at 4 °C for up to 1 month.

Procedure

Procedure 1: generating monoclonal cell lines bearing the prime editor and synthetic DNA Tape

● **TIMING** ~30 d

▲ **CRITICAL** This procedure establishes the essential groundwork for recording experiments, in which we provide detailed instructions to engineer monoclonal cell lines with foundational recording machinery (prime editor and synthetic DNA Tape) for subsequent signal and CRE activity recording using ENGRAM.

Making UMI-labeled DNA Tapes (optional)

● **TIMING** 2 d

▲ **CRITICAL** Using UMI-labeled DNA Tapes enables the accurate determination of the DNA Tape integration copy number within the genome of either a cell pool or a single clonal cell line. This method is recommended for applications requiring high-precision measurement of recording efficiency, but it is not essential for standard recording experiments. This may be useful when generating monoclonal lines with high copy-number DNA Tapes, so as to accurately select the lines with the highest copy number.

1. Prepare the PCR using PiggyBac-synHEK3-Tape plasmid as a template with BsaI_15N_fwd primer and Tape_ampup_rev primer (Table 1) using Q5 polymerase, as follows:

	Volume (μL)	Final concentration
Q5 DNA polymerase 2× master mix	25	1×
10 μM BsaI_15N_fwd	2.5	0.5 μM
10 μM Tape_ampup_rev	2.5	0.5 μM
1 ng/μL template	1	20 pg/μL
Water	19	
Sum	50	

Run the PCR using the following conditions:

Cycle number	Denature	Anneal	Extend
1	98 °C, 30 s		
15–20 cycles	98 °C, 10 s	60 °C, 15 s	72 °C, 30 s
21	72 °C, 60 s		

2. Clean and concentrate the PCR product using AMPure XP beads by following the manufacturer's instructions (or column-based DNA purification methods). Use the D1000 Tapestation to confirm the size and concentration. This product can be used for both PiggyBac-synHEK3-Tape and PiggyBac-5x-Typewriter-Tape.

   ■ **PAUSE POINT** Cleaned and concentrated PCR products can be stored at −20 °C for several months.
3. Clone the 15N PCR product from Step 2 into PiggyBac-synTape plasmid backbone using GGA with BsaI (Box 1).

   ◆ **TROUBLESHOOTING**
4. Transform 2 µL of the GGA product into 50 µL NEB 10-beta Electrocompetent *E. coli* competent cells by electroporation, following the manufacturer's instructions.

   ▲ **CRITICAL STEP** The buffer in GGA might affect the electroporation. A 2 µL GGA reaction with 50 µL competent cells is normally fine in our experience. If necessary, purify the plasmid from the GGA reaction using 0.5× AMPure XP beads by following the manufacturer's instructions, resuspend the cleaned plasmid DNA into a 7 µL volume, and proceed with the electroporation.
5. Add 948 µL Stable Outgrowth Medium and recover the transformed cells by incubating them at 30 °C for 30 min with 220 rpm shaking.

   ▲ **CRITICAL STEP** Bacteria with PiggyBac constructs should grow at 30 °C rather than 37 °C. Growth at 37 °C induces higher rates of plasmid recombination owing to the repetitive sequences such as the inverted terminal repeats. All bacterial cultures in this protocol are grown at 30 °C.
6. Take a 1 µL aliquot of the recovered cells, dilute them into 50 µL Stable Outgrowth Medium and plate them onto a selective LB agar plate containing 100 µg/mL ampicillin. Transfer the remaining outgrowth medium into a 50 mL culture tube containing LB medium supplemented with 100 µg/mL ampicillin. Culture the plate and culture tube at 30 °C for ~18 h.
7. Count the number of colony-forming units (CFUs) on the plate. A library size exceeding 10,000 unique BCs (with more than ten CFUs on the plate) is generally recommended for sufficient complexity.

   ▲ **CRITICAL STEP** A 1,000× dilution is usually a good ratio to estimate the complexity of the library. If there are too little colonies, you must perform multiple electroporations (for example, three independent times) and pool the recovered cells together before plating the 1 µL aliquot.
8. Purify the plasmid DNA from the culture in Step 6 using a Zymo Midi-Prep kit or an equivalent plasmid purification method, following the manufacturer's instructions. Ensure that the eluted DNA is of high concentration (at least 500 ng/µL) for downstream applications.

Generating a stable cell line expressing PEmax and containing the synthetic DNA Tape using the PiggyBac transposon system

● **TIMING ~7–10 d**

9. Seed 5× 105 HEK293T (or other adherent) cells into a 6-well plate a day before transfection, as described in Box 2.
10. On the day of transfection, prepare a transfection mixture containing 1,200 ng PiggyBac-TRE-PEmax-T2A-mCherry-PGK-rtTA-P2A-Puro, 800 ng PiggyBac-synHEK3-Tape and 400 ng Super PiggyBac Transposase expression plasmid, at a cargo (plasmids that need to be integrated):transposase molar ratio of 4:1.

    ▲ **CRITICAL STEP** Optionally, to generate cell lines with high copy numbers of DNA Tape, the molar ratio between the selection marker-containing PEmax plasmid and the selection-free DNA Tape plasmid can be adjusted during transfection. Reducing the molar ratio of PEmax:synTape from 1:3 to 1:10 or 1:20 and selecting cells on the basis of the puromycin selection marker can result in lines with high (~60–100) copy numbers of DNA Tapes.
11. Transfect the plasmid mixture into the target cells using Lipofectamine 3000 (or Lipofectamine 2000 for mESCs), following the manufacturer's instructions.

▲ **CRITICAL STEP** Here, we describe a protocol for adherent cell lines. This Protocol can be adopted for nonadherent cell lines such as K562 by replacing Lipofectamine 3000 with Lonza SF Cell Line 4D-Nucleofector kits to maintain transfection efficiency.

12. Change the medium the day after transfection and culture the transfected cells in their standard growth medium for an additional 48 h to enable sufficient expression of the transposase and integration of the cargo into the genome.
13. Pass the transfected cells into a T25 tissue culture flask and initiate selection of successfully integrated cells by adding puromycin to the growth medium at a final concentration of 2 µg/mL.

    ▲ **CRITICAL STEP** Integrating large genetic payloads such as PEmax using transposon systems can be relatively inefficient. Expect substantial cell death (up to 90%) during the initial selection period. Regularly monitor the cells and replenish the selection medium every 2–3 d.
14. Continue culturing the cells in puromycin-containing medium for up to 10 d, or until a population of stably resistant cells emerges.

Selection of single-cell-derived clones for robust and efficient recording, and sequencing the DNA Tape to create a whitelist of known DNA Tapes (optional but highly recommended)

▲ **CRITICAL** The selection of single-cell clones with high amounts of PE that prime edit efficiently for monoclonal line generation is optional, depending on the transfection efficiency of your cell line of use, as a higher baseline transfection efficiency will result in a polyclonal population of cells that already have high amounts of PE. However, we highly recommend following this protocol and generating monoclonal lines as this will permit precise control of the amount of PE present in each cell and may result in more consistent experimental results.

● **TIMING** ~14 d

15. Induce the expression of PEmax-T2A-mCherry by adding Dox to the cell culture medium at a final concentration of 200 ng/mL 2 d before fluorescence-activated cell sorting (FACS).
16. Use FACS to sort the top 10% mCherry-expressing cells as single cells into individual wells of two 96-well plates (one cell per well). This will provide sufficient monoclonal colonies for downstream recording experiments.

    ▲ **CRITICAL STEP** If FACS is unavailable, dilute the cells to three cells per milliliter in full culture medium and dispense 100 µL into each well of four 96-well plates. This will provide sufficient monoclonal colonies for selection.
17. Allow the sorted single cells to grow and expand in their respective wells as in Box 2, Steps 7–9, for ~1 week, or until visible colonies form. When the cells are sufficient in number to cover ~1/2 of the well, transfer the cells to a larger well by splitting them as in Box 2, Steps 9–15. Change the culture medium as needed to ensure proper cell growth and viability.
18. Select 12–24 colonies and transfer them to a new plate. Expand each clone in complete media for recording efficiency measurement and frozen stocks. For storage, cryopreserve ~1 million cells from each clone, ensuring that each vial is properly labeled and tracked, as in Box 2, Steps 16–19.
19. To assess the recording efficiency of these clones, induce the expression of PEmax by adding Dox to the cell culture medium at a final concentration of 200 ng/mL. Let the cell express for 2 d.
20. Transiently transfect each clone with a plasmid expressing a pegRNA that targets synTape (for example, U6-HEK3-pegRNA or U6-DTT-pegRNA).
21. After 48 h, collect the cells and prepare the library to analyze the recording efficiency (Box 3).

Procedure 2: cloning ENGRAM recorders targeting synHEK3 or Typewriter Tape with a two-step nested GGA

● **TIMING** 4–5 d

▲ **CRITICAL** This procedure provides detailed instructions to design, clone and prepare CRE-specific ENGRAM recorders for signal recording (Procedure 3) and CRE activity recording (Procedure 4).

1. Use the provided custom script (generate_ENGRAM_CRE_bc.py) to assign a unique BC to the CRE of interest. Order the designed CRE–BC oligos from Twist Bioscience or similar services.
   ▲ **CRITICAL STEP** Recording efficiency has been measured for 5N and 8N BC libraries. For better data interpretation, balanced recording efficiency across different BCs is critical.
   ▲ **CRITICAL STEP** Long, homologous, intervening sequences between a CRE and its paired BC can also lead to BC swapping during amplification[67] or pooled viral packaging[68,69]. BC swapping can introduce inaccuracies in recording and compress signal-to-noise. CRE–BC distance should be minimized for ENGRAM libraries (for example, as in the 5′-ENGRAM design and PiggyBac cloning scheme described with which we observe little to no swapping). For this reason, we recommend not adapting this protocol for Gibson assembly, as we have seen that Gibson assembly increases the rate of CRE–BC swapping. To clone signal-specific CREs, we recommend using specific primer pairs to amplify the oligo from the pool.
   ◆ **TROUBLESHOOTING**

Generation of individual signal-specific ENGRAM recorders

2. Amplify signal-specific CRE–BC pairs with CRE_fwd and CRE_rev (Table 1), using the following mix and qPCR conditions.

	Volume (µL)	Final concentration
Q5 DNA polymerase 2× master mix	25	1×
10 uM CRE_fwd	2.5	0.5 µM
10 uM CRE_rev	2.5	0.5 µM
1 ng/µL template	1	20 pg/µL
100× SYBR green	0.5	1×
Water	18	
Sum	50	

Cycle number	Denature	Anneal	Extend
1	98 °C, 30 s		
10–15 cycles	98 °C, 10 s	60 °C, 15 s	72 °C, 30 s
16	72 °C, 60 s		

   ▲ **CRITICAL STEP** Typically, we recommend performing qPCR so that the product can be monitored to minimize the number of PCR cycles needed. We recommend 10–15 cycles, but you can go to 20 cycles of PCR if necessary.
3. Clean and concentrate the PCR product using AMPure XP beads by following the manufacturer's instructions (or column-based DNA purification methods). Use the D1000 Tapestation to confirm the size and concentration.
4. Clone the CRE–BC PCR product into the PiggyBac-ENGRAM plasmid backbone using GGA with BsaI (Box 1).
5. Transform 1 µL of GGA product into 10 µL NEB C3040 competent cells by following the manufacturer's protocol.
6. Plate the transformation reaction onto an LB agar plate containing 100 µg/mL ampicillin and incubate overnight at 30 °C.
7. Pick four colonies for whole plasmid sequencing using Plasmidsaurus' ZeroPrep service by following their instructions; to preserve the colony, it can be suspended in 20 µL of water. Send in 10 µL of that to Plasmidsaurus' ZeroPrep service.
8. Once sequence verified, transfer the remaining 10 µL suspended colony into 5 mL LB with 100 µg/mL ampicillin. Culture the tube at 30 °C with shaking at 220 rpm for ~18 h.
9. Purify the plasmid using Monarch Spin Plasmid Miniprep Kit, following the manufacturer's instructions.
10. Clone the minP-pegRNA DNA fragment (Table 1) into the plasmid prepared from the last step using GGA with BsmBI (Box 1 and Fig. 3b).

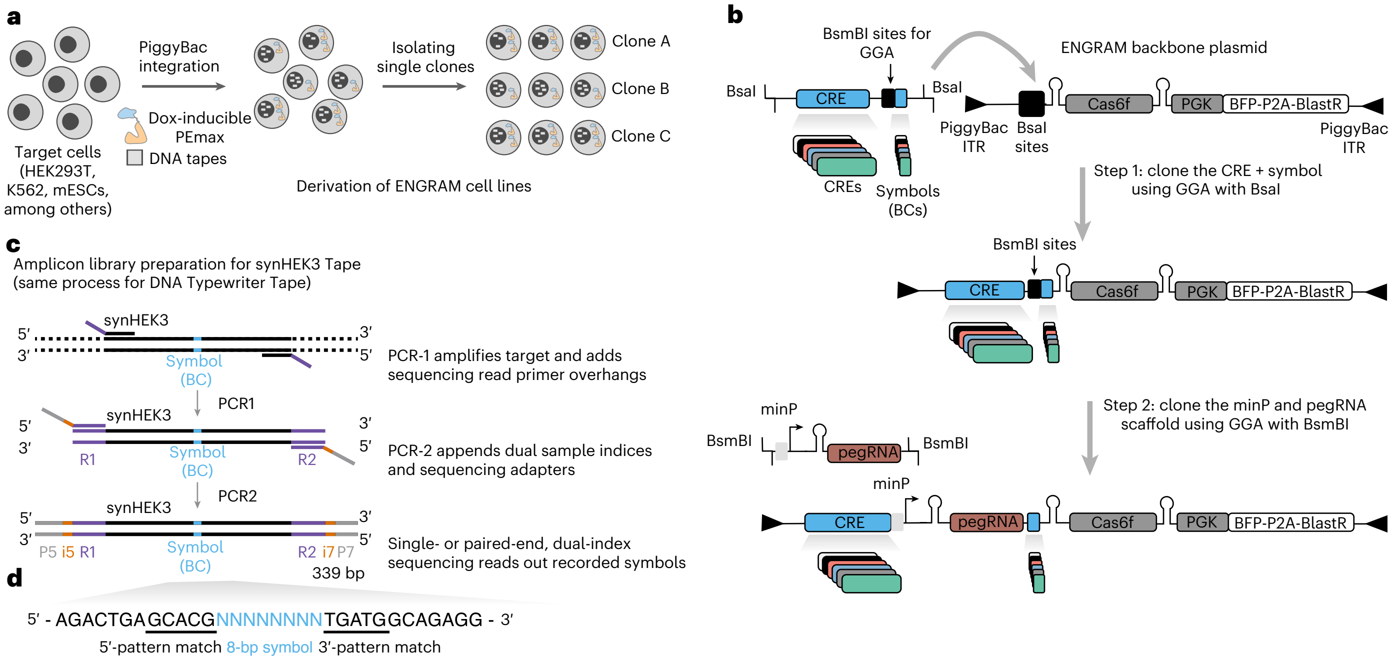

Fig. 3 | An overview of ENGRAM experimental procedures. a, A schematic of making ENGRAM cell lines. Dox-inducible PEmax and DNA Tapes are integrated into the genome via PiggyBac transposition. Deriving a monoclonal ENGRAM cell line bearing these components is recommended for robust and reliable molecular recording. **b**, A framework of two-step GGA to clone ENGRAM pegRNAs with paired symbols (BCs). In Step 1, a library of CRE–symbol pairs is cloned into the PiggyBac-ENGRAM backbone via GGA with BsaI. In Step 2, a double-stranded DNA molecule containing (1) a minP and (2) a pegRNA scaffold flanked on the 5′-end by a Cas6f hairpin is cloned into the constructs derived from Step 1 via GGA with BsmBI. **c**, A schematic of a nested PCR amplicon sequencing strategy to recover recorded information from DNA Tape. The two-step nested PCR amplicon library preparation process is essentially the same for both kinds of DNA Tapes described in this protocol (synHEK3 and DNA Typewriter; synHEK3 visualized here). The first PCR amplifies the target and adds partial sequencing primer overhangs (purple). The second PCR appends dual sample indices (orange) and sequencing adapters (gray). Single- or paired-end, dual-index high-throughput sequencing is used to demultiplex samples (index reads i5 and i7) and recover recorded information (sequencing reads R1 and/or R2). **d**, Pattern-matching functions are used to extract recorded information written as specific programmed symbols to the DNA Tape. ITR, inverted terminal repeat. Part **a** is adapted from ref. 19, CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

11. Transform 1 µL of GGA product into 10 µL NEB C3040 competent cells by following the manufacturer's protocol.
12. Plate the transformation reaction onto an LB agar plate containing 100 µg/mL of ampicillin and incubate overnight at 30 °C.
13. Pick four colonies for whole plasmid sequencing using Plasmidsaurus' ZeroPrep service by following their instructions; to preserve the colony, it can be suspended in 20 µL of water. Send in 10 µL of that to Plasmidsaurus' ZeroPrep service.

Generation of a library of CRE–ENGRAM recorders

14. Perform BsaI GGA from Step 1 as described in Steps 2–7 of Procedure 2.
15. Transform 2 µL of the GGA product into 50 µL NEB 10-beta Electrocompetent *E. coli* competent cells by electroporation, following the manufacturer's instructions. If you have a large library, repeat electroporation multiple times to achieve enough coverage.
    ▲ **CRITICAL STEP** The buffer in GGA might affect the electroporation. A 2 µL GGA reaction with 50 µL competent cells is normally fine in our experience. If necessary, purify the plasmid from the GGA reaction using 0.5× AMPure XP beads by following the manufacturer's instructions, resuspend the cleaned plasmid DNA into a volume of 7 µL and proceed with the electroporation.
16. Add 948 µL Stable Outgrowth Medium and recover the transformed cells by incubating them at 30 °C for 30 min with shaking at 220 rpm.

17. Take a 1 µL aliquot of the recovered cells, dilute them into 50 µL Stable Outgrowth Medium and plate them onto a selective LB agar plate containing 100 µg/mL ampicillin. Transfer the remaining outgrowth medium into a 50 mL culture tube containing LB medium supplemented with 100 µg/mL ampicillin. Culture the plate and culture tube at 30 °C for ~18 h.
18. Count the number of CFUs on the plate. A library size exceeding 10,000 unique BCs (with more than ten CFUs on the plate) is generally recommended for sufficient complexity.
    ▲ **CRITICAL STEP** A 1,000× dilution is usually a good ratio to estimate the complexity of the library. If there are too few colonies, you must perform multiple electroporations (for example, three independent times) and pool the recovered cells together before plating the 1 µL aliquot.
19. Purify the plasmid DNA using a Zymo Midi-Prep kit or an equivalent plasmid purification method, following the manufacturer's instructions. Ensure the eluted DNA is of high concentration (>500 ng/µL) for downstream applications.
    ▲ **CRITICAL STEP** It is recommended to quantify the complexity of the library at each step. We recommend performing PCR and next-generation sequencing on the plasmid library at each stage of cloning to estimate the coverage of the CREs and estimate the ratio of BC swapping. The only source of BC swapping that we have observed so far is the chimeric PCR product during the oligo pool amplification, which is normally ~1%. For sequencing library preparation, please refer to Box 3, Steps 4–9. In Step 4 of Box 3, use primers CRE-BC_nexteraR1_fwd and CRE-BC_truseqR2_rev (Table 1) to amplify the CRE–BC library, and continue the protocol from that PCR to assess CRE diversity and the ratio of BC swapping.
    ◆ **TROUBLESHOOTING**
20. Clone the minP-pegRNA DNA fragment (Table 1, ordered from Integrated DNA Technologies or similar DNA synthesis services) into the plasmid prepared from the last step using GGA with BsmBI (Box 1).
21. Repeat Steps 15–18 of Procedure 2 and transform 2 µL of the GGA product into 50 µL NEB 10-beta Electrocompetent *E. coli* competent cells by electroporation. Estimate the complexity of the library by counting CFUs.
    ▲ **CRITICAL STEP** Multiple electroporations are required to achieve enough coverage for a complex library.
22. Purify the plasmid DNA using a Zymo Midi-Prep kit or an equivalent plasmid purification method, following the manufacturer's instructions. Ensure that the eluted DNA is of high concentration (>500 ng/µL) for downstream applications.
23. Verify the plasmid pool by PCR as in Procedure 2, Step 19, and sequence the library using two-step PCR (Box 3, Steps 4–9).

Procedure 3: recording signaling pathway activity and order with ENGRAM

● **TIMING** 15–18 d

▲ **CRITICAL** To illustrate signal recording and intensity measurement, we use the Wnt and NFκB signaling pathways as examples. This protocol can be adapted to study other signaling pathways by utilizing appropriate signal-responsive CREs. The Protocol is compatible with both synHEK3 and Typewriter Tapes. synHEK3 Tape is recommended for quantifying signal intensity owing to its efficiency, whereas Typewriter Tape should be used for recording the temporal order of signals. In this procedure, the specific recorders and their corresponding synHEK3 or Typewriter Tape are not explicitly mentioned; users should ensure that they select the appropriate Tapes and recorders for their intended application.

Making cell lines with signal-specific ENGRAM

● **TIMING** 7–10 d

1. Seed 5× 10^5 monoclonal cells derived from Procedure 1, Steps 15–21, into a 6-well plate a day before transfection.
2. On the day of transfection, prepare a transfection mixture containing 2,000 ng signal-specific PB-ENGRAM-recorder and 400 ng PB transposase expression plasmid at a molar ratio of 5:1.

3. Transfect the plasmid mixture using Lipofectamine 3000, following the manufacturer's instructions.

   ▲ **CRITICAL STEP** Transiently transfecting the ENGRAM recorder may result in low background recording. While this background is minimal in TRE-PEmax cell lines, it can be further reduced by cotransfecting 1,000 ng of a U6 promoter-driven nontargeting pegRNA (non-targeting controls not provided; any gRNA or pegRNA should be effective).
4. Change the medium the day after transfection and culture the transfected cells in their standard growth medium for an additional 48 h to permit sufficient expression of the transposase and integration of the cargo into the genome.
5. Pass the transfected cells into a T25 tissue culture flask and initiate selection of successfully integrated cells by adding blasticin to the growth medium at a final concentration of 10 µg/mL.

   ▲ **CRITICAL STEP** PB-ENGRAM cargo is much smaller than PEmax. Most of the cells should survive and contain a few copies of the recorder. Regularly monitor the cells and change the selection medium every 2–3 d.
6. Continue culturing the cells in blasticin-containing medium for up to 10 d, or until a population of stably resistant cells emerges.

Recording the intensity of signals

● **TIMING 2 d**

7. Prepare the signal ligand medium: for Wnt and NFκB signaling pathways, create 12 serial dilutions of CHIR-99021 (0.25–16 µM) and human TNF (starting from 64 ng/mL), respectively, using a 2× stock. For CHIR-99021, include additional concentrations between 2 µM and 4 µM to cover its sensitive range.
8. Count cells and dilute with full culture medium to 4×10^5 cells/mL. Seed 250 µL of this suspension into each well of a 48-well plate.
9. Dispense 250 µL of the 2× ligand medium serial dilutions into designated wells to achieve the intended final ligand concentrations. Ensure each condition is tested in triplicate for robust recording measurement.
10. Culture the cells in ligand containing medium for 48 h.
11. Collect the cells and retrieve the recorded information by following the steps in Box 3.

    ▲ **CRITICAL STEP** Typically, a confluent well of a 48-well plate contains ~1 million cells. The protocol can be slightly modified as follows: aspirate the medium, add 200 µL of freshly prepared lysis buffer (10 mM Tris–HCl, pH 8; 0.05% SDS; 40 µg/mL proteinase) and incubate at 37 °C for 30 min. Transfer the lysate to a PCR tube and follow the protocol described in Box 3. For signal recording, one PCR reaction is sufficient to recover the information.

    ◆ **TROUBLESHOOTING**

Recording the order of signals

● **TIMING 4–6 d**

12. Prepare 2× signal ligand medium for each signal. For example, 6 µM CHIR-99021 for Wnt signaling pathway (signal A) and 20 ng/mL human TNF for NFκB signaling pathway (signal B).
13. Count the cells and dilute them with full culture medium to 4×10^5 cells/mL. Seed 250 µL of this suspension into 6 wells of a 48-well plate, representing two orders of signals (A → B, and B → A) and three replicates each.
14. Dispense 250 µL of the 2× ligand medium serial dilutions into designated wells to achieve the intended final ligand concentrations.
15. Culture the cells for 48 h for sufficient recording.
16. Passage the cells into new wells at a 1:4 ratio as described in Box 2, Steps 9–15. Culture these cells in medium containing the alternative ligand for an additional 48 h.
17. Collect the cells and retrieve the recorded information by following the steps in Box 3.

    ◆ **TROUBLESHOOTING**

Procedure 4: multiplex recording of enhancer activities with ENGRAM

● **TIMING** 15–18 d

▲ **CRITICAL** To illustrate multiplex enhancer activity measurement, we used validated enhancers in K562 cells and assigned unique BCs (symbols) to them.

Making cell lines with the CRE–ENGRAM library

● **TIMING** 4–10 d

▲ **CRITICAL** We have tested multiplex CRE recording in K562 cells. This procedure requires K562 cells with TRE-PEmax and synHEK3 DNA Tape.

1. Culture K562 cells in RPMI 1640 medium for 3 d before transfection as described in Box 2. Confirm sufficient cell quantity for the procedure.
2. To determine the required number of cells for transfection, multiply the CRE library size by 1,000. For instance, if your CRE library contains 1,000 CREs, aim to transfect 1 million cells.
3. On the day of transfection, prepare a transfection mixture containing 2,000 ng signal-specific PB-ENGRAM-recorder and 400 ng PB transposase expression plasmid at a molar ratio of 5:1 and transfect into K562 cells with Lonza SF Cell Line 4D-Nucleofector kit by following the manufacturer's instructions.

   ▲ **CRITICAL STEP** This is an example of transfecting 1 million cells. Please scale the amounts of plasmids and reagents for transfection according to the calculated cell number from the previous step. A single cuvette can transfect up to 4 million cells. Repeat transfection three times as replicates.
4. Culture the transfected cells in the standard growth medium for an additional 48 h to enable sufficient expression of the transposase and integration of the cargo into the genome.
5. Pass the transfected cells into a T25 tissue culture flask as described in Box 2 and initiate selection of successfully integrated cells by adding blasticin to the growth medium at a final concentration of 10 µg/mL.
6. Continue culturing the cells in blasticidin-containing medium for up to 10 d, or until a population of stably resistant cells emerges.

Recording CRE activities in cells

● **TIMING** 2 d

7. Induce recording by adding Dox to the cell medium at a final concentration of 200 ng/mL. Culture the cells for 48 h to enable sufficient recording of CRE activities. The optimal duration may vary depending on the cell type, CRE and experimental goals.
8. Concurrently, determine the necessary cell number to ensure adequate retrieval of recorded information. A general guideline is to aim for ~1,000 Tapes per CRE to achieve robust and statistically significant results. Consider the number of unique CREs in the library and the estimated number of DNA Tape copies per cell (determined in Procedure 1, Step 21, or in Box 3, Step 12, in each monoclonal line, for example, 10–20 Tapes per cell). For instance, with 300 CREs and an estimated ten Tapes per cell, performing one to two PCR reactions, which is equivalent to 20,000–40,000 cells, should ensure efficient information retrieval.

   ▲ **CRITICAL STEP** Determining the optimal number of reads needed for accurate measurement of CRE activity depends on several factors, including the library size and the proportion of active CREs it contains. As a general guideline, we recommend using ~1,000 Tapes per CRE and 10,000–20,000 reads per CRE. For a library containing 10,000 CREs, ~100 million reads (for example, from a single P1 kit) should provide reliable activity measurements. If the proportion of active CREs is expected to be low, the total number of reads can be reduced to ~10 million.
9. Retrieve the recorded information by following the steps in Box 3. Perform PCR-1 with the number calculated in Box 3, Step 4, with the number of cycles obtained earlier for the clone you used (Box 3).

   ◆ **TROUBLESHOOTING**

Troubleshooting

Troubleshooting advice can be found in Table 2.

Table 2 | Troubleshooting table

Step	Problem	Possible reason	Solution
Procedure 1			
Step 3	Low barcoded DNA Tape cloning efficiency	Background with the backbone plasmid is high	Increase the number of GGA cycles to 50 cycles of 37 °C for 1 min, followed by 16 °C for 1 min. This applies to any step of this protocol in which GGA is used Alternatively, precut the backbone plasmid with BsaI followed by gel purification. This ensures that no background plasmid is carried over
Procedure 2			
Step 19	Swapping between CREs and BCs	Too many PCR cycles were used to amplify the CRE–BC pairs	Use fewer PCR cycles to amplify CRE–BC pairs. Alternatively, when ordering CRE–BC pairs as a fraction of a pool, make the CRE–BC sequences a high fraction of the pool to reduce cycles required for the PCR to isolate your CRE–BC pairs from the rest of the pool
Procedure 3			
Step 17	Low editing efficiency	Low copy number of ENGRAM recorders, Tapes or PEmax in the cells	Select cells with a higher number of ENGRAM recorders by performing FACS to isolate high fluorescence cell populations. If a drug resistance marker is on the ENGRAM plasmid, use a high dose of drug to select for cells with a high ENGRAM copy number
	Low editing efficiency	Editing time-window was too short	Increase the amount of time cells have to record ENGRAM events by waiting several more days to collect, by changing the media to include fresh Dox every other day, and by adding fresh ligand as appropriate
Procedure 4			
Step 9	Low editing efficiency	Editing time window was too short	The majority of the CREs selected might be low activity or not active. Increasing the recording time helps to record these low-activity CREs

Timing

Procedure 1
Steps 1–8, making UMI-labeled DNA Tapes: 2 d
Steps 9–14, generating a stable cell line expressing PEmax and containing the synthetic DNA Tape using the piggyBac transposon system: 7–10 d
Steps 15–21, selection of single-cell-derived clones for robust and efficient recording, and sequencing the DNA Tape to create a whitelist of known DNA Tapes: 14 d
Procedure 2
Step 1, design of CRE–BC oligos for ENGRAM cloning: 1–4 h
Steps 2–13, generation of individual signal-specific ENGRAM recorders: 2–4 d
Steps 14–23, generation of a library of CRE–ENGRAM recorders: 2–4 d
Procedure 3
Steps 1–6, making cell lines with signal-specific ENGRAM: 7–10 d
Steps 7–11, recording the intensity of signals: 2 d
Steps 12–17, recording the order of signals: 4–6 d
Procedure 4
Steps 1–6, making cell lines with the CRE–ENGRAM library: 4–10 d
Steps 7–9, recording CRE activities in cells: 2 d
Boxes
Box 1, general protocol for one-pot GGA: 1–2 h
Box 2, general protocol for mammalian cell culture (HEK293Ts and K562s): 1 h
Box 3, retrieval of recording information from DNA Tape—lysis and PCR: 3–5 h

Anticipated results

Data from an ENGRAM recording experiment are counts of programmed symbols (insertional BCs) extracted from reads of the recording DNA Tape (here, sample demultiplexed high-throughput sequencing reads stored as FASTQ files). When combined with DNA Typewriter, the location of a given symbol within the sequential DNA Tape is an additional layer of information. From these counts of symbols and their sequential patterns in DNA Tape, several metrics can be quantified:

- The presence or absence of CRE/signal activity in a given sample during the recording period.
- The fold induction, or signal-to-noise ratio between stimulated and unstimulated groups for each CRE/signal recorder (for example, Fig. 2b).
- The relative timing or order of each CRE/signal activity (inferred from bigram frequencies of given symbols from sequential DNA Typewriter records, for example, Fig. 2d).
- The relative strength of CREs/signals in a pool (for example, Fig. 2c).

When combined with single-cell profiling methods, a cell BC can further be appended to sequencing reads of the DNA Tape, allowing the records to be related/combined with single cell profiles (for example, single cell transcriptomes). This allows one to quantify the activity of given CRE/signals in a precisely defined transcriptomic cell type, or to reconstruct the history of cellular events (for example, CRE/signaling activity) that produced a given cell type or state.

Key quality control metrics in an ENGRAM experiment include correlations in CRE activity (symbol counts and edit scores) between biological replicates, correlations in insertion counts between independent BCs associated with the same CRE/ENGRAM recorder, overall PE efficiency (which facilitates information recovery) and signal-to-noise ratios between stimulated and unstimulated conditions. When executing or analyzing a recording experiment, a common diversion can be to unduly focus on the increased signal (editing efficiency) rather than increased signal-to-noise (large fold-change between induced and control), the latter of which is considerably more valuable in many ENGRAM recording tasks (for example, inferring whether a CRE was ever active in a cell's past). A successful experiment will allow the user to determine which CREs/signals are active in which samples or cells, as well as their relative intensity and/or orders, with high reproducibility between biological replicates. We typically observe editing efficiencies of 20–40% with well-characterized CREs in monoclonal PEmax cell lines (though editing efficiencies for different CREs can vary widely depending on their activity) and Pearson correlations >0.9 across replicates.

ENGRAM represents a new frontier where, in theory, any event that can be coupled to Pol-2 mediated transcription can be written permanently in DNA. When combined with DNA Typewriter, additional cellular events (lineage, time, protein–protein interactions, among others) can be continuously and sequentially recorded to a universal DNA Tape. The combinatorial, single cell resolution recordings that this paradigm affords lends itself to rapid adaptation to monitor diverse signals of interest.

Data availability

The optimized ENGRAM plasmids have been deposited to Addgene with the following names and IDs: P021-mPB-ENGRAM-All-in-one-PGK-BFP-P2A-Blast (239972), P022-mPB-TRE-PEmax-T2A-mCherry-PGK-rttA-Puro (239974), P023-mPB-synHEK3-BsaI (239975) and P024-mPB-5xTape-BsaI (239983). Related raw data can be found in association with the supporting primary research articles[18,19].

Code availability

Related scripts are available via GitHub at https://github.com/shendurelab/ENGRAM/tree/main/Nprotocol_scripts.

Received: 23 June 2025; Accepted: 10 December 2025;
Published online: 11 February 2026

References

1. Askary, A. et al. The lives of cells, recorded. *Nat. Rev. Genet.* **26**, 203–222 (2025).
2. Chen, W. & Choi, J. Molecular circuits for genomic recording of cellular events. *Trends Genet.* **41**, 647–659 (2025).
3. Church, G. M., Gao, Y. & Kosuri, S. Next-generation digital information storage in DNA. *Science* **337**, 1628 (2012).
4. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
5. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
6. Tang, W. & Liu, D. R. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**, eaap8992 (2018).
7. Farzadfard, F. et al. Single-nucleotide-resolution computing and memory in living cells. *Mol. Cell* **75**, 769–780.e4 (2019).
8. 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).
9. Schmidt, F., Cherepkova, M. Y. & Platt, R. J. Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **562**, 380–385 (2018).
10. Chan, M. M. et al. Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).
11. Livet, J. et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
12. 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).
13. 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).
14. 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).
15. Hao, K. et al. Reconstructing signaling history of single cells with imaging-based molecular recording. Preprint at *bioRxiv* https://doi.org/10.1101/2024.10.11.617908 (2024).
16. Winter, E., Emiliani, F., Cook, A., Abderrahim, A. & McKenna, A. BASELINE: a CRISPR base editing platform for mammalian-scale single-cell lineage tracing. Preprint at *bioRxiv* https://doi.org/10.1101/2025.03.19.644238 (2025).
17. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
18. Choi, J. et al. A time-resolved, multi-symbol molecular recorder via sequential genome editing. *Nature* **608**, 98–107 (2022).
19. Chen, W. et al. Symbolic recording of signalling and *cis*-regulatory element activity to DNA. *Nature* **632**, 1073–1081 (2024).
20. Loveless, T. B. et al. Open-ended molecular recording of sequential cellular events into DNA. *Nat. Chem. Biol.* **21**, 512–521 (2024).
21. Koblan, L. W. et al. High-resolution spatial mapping of cell state and lineage dynamics in vivo with PEtracer. *Science* **390**, eadx3800 (2025).
22. Liao, H., Choi, J. & Shendure, J. Molecular recording using DNA Typewriter. *Nat. Protoc.* **19**, 2833–2862 (2024).
23. Gao, Y. & Zhao, Y. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J. Integr. Plant Biol.* **56**, 343–349 (2014).
24. Knapp, D. J. H. F. et al. Decoupling tRNA promoter and processing activities enables specific Pol-II Cas9 guide RNA expression. *Nat. Commun.* **10**, 1–14 (2019).
25. Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol. Cell* **54**, 698–710 (2014).
26. Haurwitz, R. E., Sternberg, S. H. & Doudna, J. A. Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *EMBO J.* **31**, 2824–2832 (2012).
27. Chardon, F. M. et al. A multiplex, prime editing framework for identifying drug resistance variants at scale. Preprint at *bioRxiv* https://doi.org/10.1101/2023.07.27.550902 (2023).
28. Schene, I. F. et al. Prime editing for functional repair in patient-derived disease models. *Nat. Commun.* **11**, 1–8 (2020).
29. Bulcaen, M. et al. Prime editing functionally corrects cystic fibrosis-causing CFTR mutations in human organoids and airway epithelial cells. *Cell Rep. Med.* **5**, 101544 (2024).
30. Davis, J. R. et al. Efficient prime editing in mouse brain, liver and heart with dual AAVs. *Nat. Biotechnol* **42**, 253–264 (2023).
31. Campa, C. C., Weisbach, N. R., Santinha, A. J., Incarnato, D. & Platt, R. J. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat. Methods* **16**, 887–893 (2019).
32. Banskota, S. et al. Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. *Cell* **185**, 250–265.e16 (2022).
33. Cabrera, A. et al. The sound of silence: transgene silencing in mammalian cell engineering. *Cell Syst.* **13**, 950–973 (2022).
34. Morsut, L. et al. Engineering customized cell sensing and response behaviors using synthetic notch receptors. *Cell* **164**, 780–791 (2016).
35. Zhu, I. et al. Modular design of synthetic receptors for programmed gene regulation in cell therapies. *Cell* **185**, 1431–1443 (2022).
36. Bhattarai-Kline, S. et al. Recording gene expression order in DNA by CRISPR addition of retron barcodes. *Nature* **608**, 217–225 (2022).
37. Liu, K., Jin, H. & Zhou, B. Genetic lineage tracing with multiple DNA recombinases: a user's guide for conducting more precise cell fate mapping studies. *J. Biol. Chem.* **295**, 6413–6424 (2020).
38. Siniscalco, A. M. et al. Barcoding Notch signaling in the developing brain. *Development* **151**, dev203102 (2024).
39. Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652.e29 (2021).
40. Yan, J. et al. Improving prime editing with an endogenous small RNA-binding protein. *Nature* **628**, 639–647 (2024).
41. Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2021).
42. Sousa, A. A. et al. Systematic optimization of prime editing for the efficient functional correction of CFTR F508del in human airway epithelial cells. *Nat. Biomed. Eng.* **9**, 7–21 (2024).
43. Wang, L. et al. Sensing and guiding cell-state transitions by using genetically encoded endoribonuclease-mediated microRNA sensors. *Nat. Biomed. Eng.* **8**, 1730–1743 (2024).
44. Klein, J. C. et al. A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nat. Methods* **17**, 1083–1091 (2020).
45. Lalanne, J.-B. et al. Multiplex profiling of developmental *cis*-regulatory elements with quantitative single-cell expression reporters. *Nat. Methods.* **21**, 983–993 (2024).
46. Zheng, L. et al. Multiplexed, scalable analog recording of gene regulation dynamics over weeks using intracellular protein tapes. Preprint at *bioRxiv* https://doi.org/10.1101/2025.05.10.653182 (2025).
47. Lin, D. et al. Time-tagged ticker tapes for intracellular recordings. *Nat. Biotechnol.* **41**, 631–639 (2023).
48. Linghu, C. et al. Recording of cellular physiological histories along optically readable self-assembling protein chains. *Nat. Biotechnol.* **41**, 640–651 (2023).
49. Liu, P. et al. Increasing intracellular dNTP levels improves prime editing efficiency. *Nat. Biotechnol.* **43**, 539–544 (2025).
50. Newby, G. A. & Liu, D. R. In vivo somatic cell base editing and prime editing. *Mol. Ther.* **29**, 3107–3124 (2021).
51. Agarwal, V. et al. Massively parallel characterization of transcriptional regulatory elements. *Nature* **639**, 411–420 (2025).
52. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat. Methods* **11**, 508–520 (2014).
53. Hughes, R. A. & Ellington, A. D. Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.* **9**, a023812 (2017).
54. Gibson, D. G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res* **37**, 6984–6990 (2009).
55. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507 (2014).
56. Coughlin, G. M. et al. Spatial genomics of AAV vectors reveals mechanism of transcriptional crosstalk that enables targeted delivery of large genetic cargo. *Nat. Biotechnol.* **44**, 133–145 (2026).
57. McDiarmid, T. A. et al. A parts list of promoters and gRNA scaffolds for mammalian genome engineering and molecular recording. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-025-02896-2 (2025).
58. Pinglay, S. et al. Synthetic regulatory reconstitution reveals principles of mammalian Hox cluster regulation. *Science* **377**, eabk2820 (2022).
59. Mitchell, L. A. et al. De novo assembly and delivery to mouse cells of a 101 kb functional human gene. *Genetics* **218**, (2021).
60. Pollak, B. et al. Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytol* **222**, 628–640 (2019).
61. Chen, P. J. & Liu, D. R. Prime editing for precise and highly versatile genome manipulation. *Nat. Rev. Genet.* **24**, 161–177 (2023).
62. Uenaka, T. et al. Prevention of transgene silencing during human pluripotent stem cell differentiation. Preprint at *bioRxiv* https://doi.org/10.1101/2025.04.07.647695 (2025).
63. Wilcoxon, F. Individual comparisons of grouped data by ranking methods. *J. Econ. Entomol* **39**, 269–270 (1946).
64. Mann, H. B. & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**, 50–60 (1947).
65. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Stat. Plan. Inference* **82**, 171–196 (1999).
66. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
67. Hegde, M., Strand, C., Hanna, R. E. & Doench, J. G. Uncoupling of sgRNAs from their associated barcodes during PCR amplification of combinatorial CRISPR screens. *PLoS ONE* **13**, e0197547 (2018).
68. Hill, A. J. et al. On the design of CRISPR-based single-cell molecular screens. *Nat. Methods* **15**, 271–274 (2018).
69. Lalanne, J.-B. et al. Extensive length and homology dependent chimerism in pool-packaged AAV libraries. Preprint at *bioRxiv* https://doi.org/10.1101/2025.01.14.632594 (2025).

Acknowledgements

We thank members of the Shendure Lab for contributions to the development of ENGRAM, and the (David) Liu Lab for developing, improving and supporting PE. This work was supported by the Brotman Baty Institute for Precision Medicine and grants from the Paul G. Allen Frontiers Group (Allen Discovery Center for Cell Lineage Tracing to J.S.), Seattle Hub for Synthetic Biology, a collaboration between the Allen Institute, Chan Zuckerberg Initiative and University of Washington (grant no. CZIF2023-008738 to J.S.) and the National Human Genome Research Institute (grant no. R01HG010632 to J.S.). W.C. is a fellow of Damon Runyon Cancer Research Foundation (DRG-2507-23). J.F.N. is supported by the UW ISCRM Fellows Program and a US National Institutes of Health F31 grant (1F31AR084852). T.A.M. was supported by a Banting Postdoctoral Fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC). J.S. is an investigator of the Howard Hughes Medical Institute.

Author contributions

This Protocol is based on a paper by W.C., J.S. and colleagues. J.F.N, T.A.M. and W.C. contributed equally to writing the manuscript and preparing the figures. All authors edited the manuscript. J.S. supervised the research and manuscript preparation. We disclose that language editing and proofreading were supported by artificial intelligence-based tools; these were not used for conceptual development or primary manuscript writing. We take full responsibility for the contents of this manuscript.

Competing interests

The University of Washington has filed a patent application partially based on this work, in which W.C. and J.S. are listed as inventors (patent application no. US-20240355418-A1, US region). J.S. is on the scientific advisory board, a consultant and/or a cofounder of Prime Medicine, Guardant Health, Camp4 Therapeutics, Phase Genomics, Adaptive Biotechnologies, Sixth Street Capital, Pacific Biosciences, Somite AI and 10x Genomics. J.N. and T.M. declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Troy A. McDiarmid, Wei Chen or Jay Shendure.

Peer review information *Nature Protocols* thanks Theresa Loveless and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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