nature protocols

https://doi.org/10.1038/s41596-024-01003-0

Protocol

Check for updates

Molecular recording using DNA Typewriter

Hanna Liao[1,2,7], Junhong Choi[1,3,7] ✉ & Jay Shendure[1,4,5,6] ✉

Abstract

Recording molecular information to genomic DNA is a powerful means of investigating topics ranging from multicellular development to cancer evolution. With molecular recording based on genome editing, events such as cell divisions and signaling pathway activity drive specific alterations in a cell's DNA, marking the genome with information about a cell's history that can be read out after the fact. Although genome editing has been used for molecular recording, capturing the temporal relationships among recorded events in mammalian cells remains challenging. The DNA Typewriter system overcomes this limitation by leveraging prime editing to facilitate sequential insertions to an engineered genomic region. DNA Typewriter includes three distinct components: DNA Tape as the 'substrate' to which edits accrue in an ordered manner, the prime editor enzyme, and prime editing guide RNAs, which program insertional edits to DNA Tape. In this protocol, we describe general design considerations for DNA Typewriter, step-by-step instructions on how to perform recording experiments by using DNA Typewriter in HEK293T cells, and example scripts for analyzing DNA Typewriter data (https://doi.org/10.6084/m9.figshare.22728758). This protocol covers two main applications of DNA Typewriter: recording sequential transfection events with programmed barcode insertions by using prime editing and recording lineage information during the expansion of a single cell to many. Compared with other methods that are compatible with mammalian cells, DNA Typewriter enables the recording of temporal information with higher recording capacities and can be completed within 4–6 weeks with basic expertise in molecular cloning, mammalian cell culturing and DNA sequencing data analysis.

Key points

- DNA Typewriter is a CRISPR genome editing-based method for recording the temporal order of molecular events by tracing the physical order of unique barcodes along a DNA Tape array.
- Compared with other existing methods, DNA Typewriter is highly multiplexable, unidirectional and sequential, capturing thousands of insertions in the precise order in which they occur, and it is active in living mammalian cells, including HEK293T, mouse embryonic stem cells and fibroblasts.

Key reference

Choi, J. et al. *Nature* **608**, 98–107 (2022): https://doi.org/10.1038/s41586-022-04922-8

[1]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [2]Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA. [3]Developmental Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [4]Howard Hughes Medical Institute, Seattle, WA, USA. [5]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. [6]Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA. [7]These authors contributed equally: Hanna Liao, Junhong Choi. ✉e-mail: choij10@mskcc.org; shendure@uw.edu

Introduction

Molecular recording allows tracking of cellular development over time by encoding molecular events into stable storage media in the cell[1–3]. Among the different molecules within the cell to which information might be recorded, genomic DNA (gDNA) is highly appealing. Various types of information can be encoded into DNA sequences by using existing genome-editing methods. The written information is maintained and reliably replicated over time throughout the process of cellular development and can be read out at a later time to reveal the history of cells.

Molecular recording has great potential to capture information about the dynamic processes of cellular development and differentiation, where the final state of the cell depends on the sequence of events that the cell has experienced in the past. In principle, molecular recording can recover molecular events that define the developmental trajectory of a cell, profiling both the final state of the cell (e.g., using single-cell RNA -sequencing (scRNA-seq), single-cell transposase-accessible chromatin or other single-cell methods compatible with sequencing specific target transcripts and genomic regions) and the history of how it got there. The ability to record temporal relationships between events is critical in reconstructing past events within each cell, especially when investigating how different orders of the same set of events may lead to different outcomes.

The field of molecular recording has advanced rapidly over the past decade, spurred by improvements in CRISPR-based genome editing methods. By using different CRISPR genome-editing methods, various applications of molecular recording, such as recording external signaling events[4,5], recording the order of internal gene expression[6,7], encoding artificial digital information in nucleic acid sequences[8,9] and even recovering past transcriptomes[10,11], have been demonstrated. However, most of these methods were demonstrated only in bacterial cells, in which various CRISPR systems were discovered. Adapting the recording devices to mammalian cells remains challenging because the cellular environment of mammalian cells is quite different from that of bacterial cells. In the case of the Cas1-Cas2 editing system, mammalian cells may lack the host factors necessary for genome editing. The generation of double-stranded DNA fragments within mammalian cells may induce different cellular responses from those in bacterial cells, such as the sensing of DNA damage. Because Cas9 activity is not dependent on bacterial host factors, editing using Cas9 and its derivatives has been overwhelmingly used as a recording method of choice to date for writing information in the mammalian genome, in which a predetermined recording locus is integrated as a substrate for the recordings. However, the ability of the Cas1-Cas2 system to write temporal information via sequential editing is missing in the Cas9-based system.

To enable the temporally ordered writing of information to the mammalian genome, we recently developed the DNA Typewriter method[12], which uses a Cas9-derivative prime editor[13,14] for sequential editing. The DNA Typewriter system further includes a designed recording locus that resembles the Cas1-Cas2 system, enforcing directional sequential editing for encoding the relative order of multiple events. Here, we describe a detailed protocol for two main applications of the DNA Typewriter system: (i) recording the order of transfection to a population of cells by using barcoded prime-editing guide RNAs (Fig. 1a) and (ii) recording single-cell lineage information (Fig. 1b and Table 1).

Development of DNA Typewriter

Molecular recording methods have progressed with advances in genome-editing tools. An advance that enabled the DNA Typewriter system is prime editing[13], which uses the engineered protein Cas9(H840A) nickase tethered to a reverse-transcriptase that is active in mammalian cells. A prime-editing guide RNA (pegRNA) complexed to the Cas9 protein specifies the insertion of a unique *k*-mer barcode, or InsertBC, at the target site as well as a 3-bp 'key' sequence that 'unlocks' the possibility of further editing. In contrast to Cas9 nuclease or base editing, the specific, desired edit can be encoded within the reverse transcriptase template of the pegRNA. Furthermore, the editing outcome is precise, resulting in either the programmed edit or an unedited sequence in the vast majority of cases.

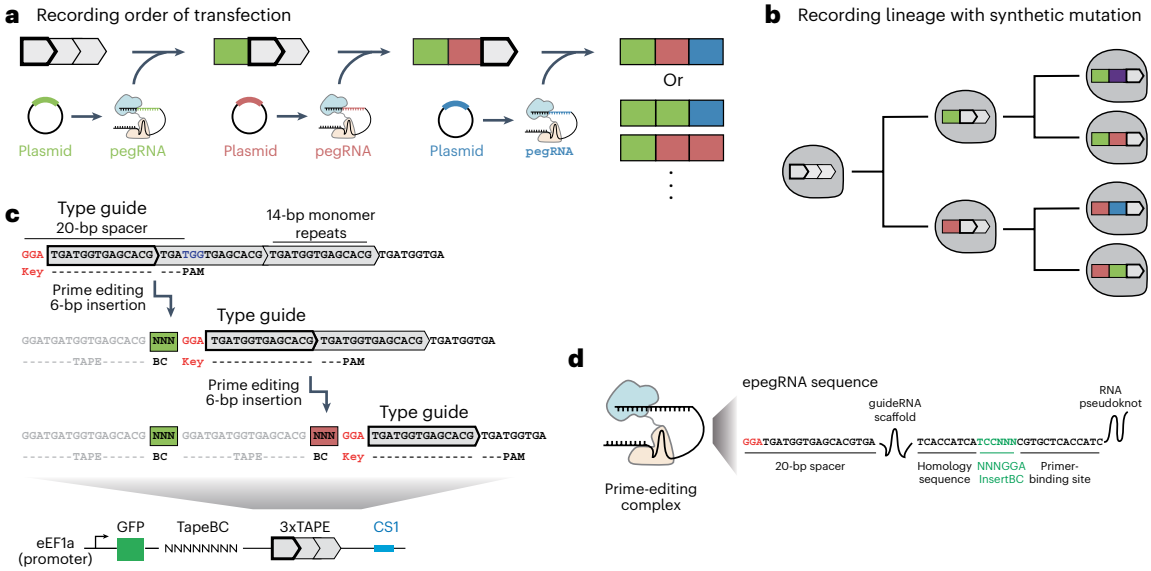

Fig. 1 | Molecular recording by using DNA Typewriter. Components and applications of DNA Typewriter system. **a**,**b**, So far, the two main applications of DNA Typewriter are recording the order of transfections (**a**) and recording cell lineage (**b**). **c**, Sequences of DNA Tape (3xTAPE). The key sequence (GGA) completes a 20-bp spacer for prime editing, which inserts a barcode (BC) and another key sequence (NNNGGA InsertBC). The order of prime-editing events is encoded on the position within the DNA Tape. **d**, The sequence of epegRNA used in DNA Typewriter. Similar to a CRISPR single-guide RNA (sgRNA), epegRNA contains a 20-bp spacer sequence with an sgRNA scaffold. The 3′ end of the sgRNA scaffold is extended with a 9-bp homology sequence, InsertBC and primer-binding sites (reverse-complementary to the target sequence). The epegRNA contains an additional RNA pseudoknot structure that prevents RNA degradation from the 3′ end. CS1, 10X Genomics Capture Sequence 1; PAM, protospacer-adjacent motif; pegRNA, prime-editing guide RNA.

Most communication systems require not only a set of symbols but also a way to write them in order. In the context of molecular recording, information about the temporal order of events must be captured in each recording. DNA Typewriter achieves this by using a recording substrate referred to as 'DNA Tape', consisting of a tandem array of partial prime-editing target sites (Fig. 1c). One unit of the DNA Tape consists of a 14-bp monomer sequence flanked by a 3-bp key (GGA) sequence at its 5′ end. The 14-bp monomer includes the TGG protospacer-adjacent motif (PAM) sequence, at positions 4–6. This array and the first six bases of the subsequent 14-bp monomer constitute a 20-bp spacer and PAM sequence. Within the DNA Tape tandem array, only the first site is an intact prime-editing target site because it includes the key sequence that completes the whole 20-bp spacer with a PAM sequence for prime editing, while the rest of the sites are truncated at their 5′ end and are therefore inactive. During the recording process, prime editing at the first target site inserts a unique event-specific barcode along with a key sequence that completes the next target site. The insertion event thus inactivates the first target site for further editing and also activates the next target site for sequential genome editing. Based on the sequential edits accumulated on the DNA Tape, the order of events within the cell can be reconstructed.

Applications of the method

To date, we have demonstrated two main applications of the DNA Typewriter system: (i) recording the order of transfections by using barcoded pegRNAs (Fig. 1a) and (ii) recording cell lineage information with recovery via scRNA-seq (Fig. 1b).

For recording the order of transfections, the plasmid to be transfected needs to carry a pegRNA-expression cassette that targets DNA Tape and inserts a variable InsertBC (Fig. 1c). The pegRNA is transcribed by using a human U6 promoter similar to that in guide RNA-expression cassettes. Based on the editing pattern of DNA Tape with InsertBCs at different locations, it is possible to infer the relative order between two InsertBCs and discriminate whether one InsertBC was introduced before another or if two InsertBCs were synchronously delivered to

the cell. For plasmids transfected concurrently, the relative ratio among transfected pegRNAs can also be inferred on the basis of the observed frequencies of each InsertBC.

In addition, DNA Typewriter can be used to track cell lineage information through the sequence of divisions required for clonal expansion (Fig. 1b). Molecular recording to track cell lineage information[15–17] has been previously used to study early development[18,19], blood cell development[20] and cancer evolution[21–23]. To apply DNA Typewriter for cell lineage tracing, all components of this system (i.e., DNA Typewriter–prime editor, DNA Tape, and pegRNAs for inserting various InsertBCs) need to be continually expressed within the cell during the clonal expansion process. Over time, InsertBCs will randomly accumulate over one or many of the multiple instances of DNA Tape integrated into the cell's genome. The combinations of InsertBCs will form a molecular barcode sequence for each cell, which will be inherited by the daughter cells after DNA replication and continue to be edited through sequential cell divisions. Finally, the edited DNA Tapes can be expressed as RNA molecules to be recovered via scRNA-seq.

We have successfully tested the DNA Typewriter system in HEK293T cells, mouse embryonic stem (ES) cells, and mouse embryonic fibroblasts (MEFs)[12]. However, prime editing has been tested in a wider variety of cell types, including human T cells[24], induced pluripotent stem cells, post-mitotic neurons[13] and cells from other model organisms such as plants[25], fruit flies[26] and *Escherichia coli*[27]. We also expect that the application of DNA Typewriter will be expanded by allowing the direct recording of cellular events. To this end, new strategies such as ENGRAM[28], a prime editing-based recording of transcription activation signals, can be used to record signaling events onto the DNA Typewriter platform. Combinations of single-cell signal-recording methods with the DNA Typewriter system could also reveal how cells develop and change over time in the context of developmental biology, cancer biology, and regenerative medicine.

Limitations

Limitations in the performance of the DNA Typewriter as a molecular recording method are linked to the limitations of prime editing. At present, the most outstanding limitation of prime editing is the rate of editing, which determines the recording rate as well as the temporal resolution of the DNA Typewriter. This limitation manifests in different forms on the basis of the application. In the context of recording the order of transfection, each round of transfection is recorded in <10% of the DNA Tapes integrated across the cell population. Thus, increasing the number of cells and performing deeper sequencing is necessary to recover the recorded information. In the context of single-cell lineage recording, a recording rate slower than one to two editing events per cell division results in a lineage tree without the resolution to reconstruct every cell division event. Although the slow editing rate could certainly hamper the application of the DNA Typewriter recording system, we expect that direct improvements in prime editing, through the engineering of prime-editing components as well as altering native DNA repair pathways[24,29–32], will address this issue.

Another limitation of the DNA Typewriter method is the delivery of the recording device to a starting cell or cells. DNA Tape needs to be stably integrated within the genome to minimize the loss of recorded DNA over time, and both the prime editor and pegRNA need to be supplied for each round of recording. In the case of recording cell lineage, the prime editor and pegRNA also need to be stably integrated to minimize perturbation of the developing system. To achieve the best recording performance, multiple copies (~20 or more) of DNA Tape would ideally be integrated into each cell, especially for single-cell lineage-recording applications[12]. The size and quantity of genetic cargo that needs to be delivered to the starting cell of interest may limit the application of DNA Typewriter to specific systems. The generation of monoclonal cell lines with multiple (ideally 20–30) integrations can also be challenging when using a conventional method for gene integration. For transposase-based integration of DNA Tapes such as using piggyBAC, a method to increase the integration number by using a small percentage of carrier plasmid during transfection with a selection marker has been reported[33].

To read out the recorded information, DNA Tape is either amplified directly from gDNA extracted from cells or transcribed into cellular RNA, which can be recovered by using scRNA-seq. Both readout methods rely on PCR amplification of the nucleic acid sequence.

Here, the most abundant forms of error are single base-pair substitutions, in which transition errors (purine-purine or pyrimidine-pyrimidine change) occur at a higher frequency than transversion errors (purine-pyrimidine change). This error could corrupt the recorded information, although one could use predetermined insertions and error-correction methods to avoid this issue. A potentially greater issue lies in the repeat structure of DNA Tape, in which the same stretch of DNA (14 bp when unedited or 17 bp when edited) is repeated multiple times for sequential editing. The tandem repeats of DNA sequences are potentially subject to expansion and contraction, both during in vivo DNA replication and during PCR amplification when preparing the sequencing library. Using long-read sequencing protocols, we have observed that larger numbers (>10) of repeats in DNA Tape are prone to array contraction[12]. In the current analysis of DNA Tape, we discard sequencing reads with unexpected expansions or contractions that may have occurred after the initial integration of DNA Tape into the genome.

Comparison with other methods

Recording the order of transfection

The use of precision genome editing such as prime editing used in DNA Typewriter (and more recently, peCHYRON[34]) enabled the recording of the order of multiple events through sequential editing. We previously demonstrated that the DNA Typewriter system can encode the relative temporal information of external stimuli such as the transfection of a DNA plasmid that expresses pegRNA targeting a DNA Tape sequence[12] (Fig. 2). Each pegRNA is capable of inserting a unique InsertBC that includes the GGA key sequence at the end (e.g., AAAGGA, AACGGA, etc.) for sequential editing. In our demonstration of encoding text messages, we pre-associated an InsertBC with a written symbol (e.g., AAAGGA representing the letter 'A', AACGGA representing 'B', etc.). Once the recording is done, the order of transfections can be inferred by sequencing the DNA Tape sequence from the bulk cell population and extrapolating the order of InsertBCs.

However, in Cas9-based recording, CRISPR–Cas9 induces double-stranded DNA breaks that are repaired via non-homologous end joining or microhomology-mediated end joining. Both processes result in a wide variety of editing outcomes, which are stochastic and highly dependent on the local DNA sequence around the cut site. Therefore, recording the order of events with Cas9 nuclease via sequential editing has been challenging to implement, because editing outcomes are hard to predict and do not occur in a unidirectional manner. However, sequential editing has been demonstrated by using other Cas-based systems, such as base editing (CAMERA[5] and DOMINO[6]) or Cas1-Cas2 combined with the retron system (Retro-Cascorder[7]).

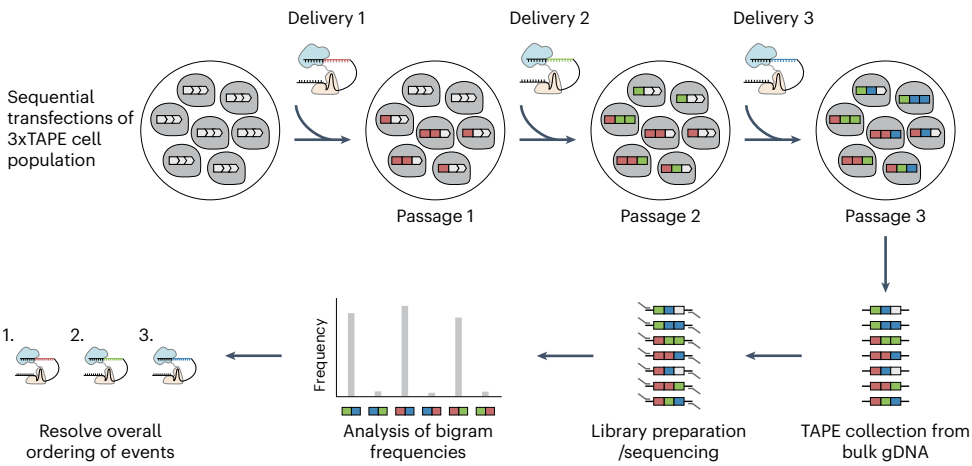

Fig. 2 | Overview of recording the order of transfections by using DNA Typewriter. As an example, a population of seven cells with 3xTAPE is propagated through three rounds of sequential transfections. Bulk gDNA collected from these seven cells after the third transfection shows different species of unidirectionally edited TAPEs. By calculating the frequencies of each bigram (i.e., pairs of adjacent editing events) from the sequencing reads, the original order of transfections can be inferred from the TAPE molecules of the entire population.

Recording single-cell lineage information

The DNA Typewriter system can be used to record and reconstruct a cell lineage tree at single-cell resolution (Fig. 3). Similar to the previous approaches based on Cas9-based lineage recorders[18,19], DNA Typewriter can be used to randomly introduce mutations in the form of insertional barcodes to DNA Tapes. These mutations are synthetic and not intended to add or alter any cellular function, but rather are used to estimate the clonal distance between cell pairs with the same set of DNA Tapes. However, we cannot rule out that the integration and expression of prime-editing components, including these arrays, may affect cellular characteristics, such as the rate of cell growth. This method is analogous to how shared natural mutation patterns across the genome could be used to estimate the phylogenetic distance between pairs of organisms after aligning their genomic sequence. Compared to other lineage reconstruction methods based on natural mutations (e.g., mutations accumulated on mitochondria or across the entire genome), lineage reconstruction using synthetic mutations has several advantages. First, the mutation rate can be tuned to achieve optimal resolution in lineage reconstruction (e.g., at a much higher rate than natural mutation). Second, mutations occur in designated genomic loci, rather than across the entire genome, to focus the sequencing efforts. Third, the genomic loci with the synthetic mutations can be transcribed into mRNA, which then can be recovered by using standard scRNA-seq platforms. These advantages come at the expense of the need to genetically introduce and maintain the recording components within the target cell, which may not be a trivial task depending on the biological system of interest.

Compared with previous Cas9-based recording methods, prime editing-based lineage recording methods such as DNA Typewriter have several notable advantages. First, prime editing generates single-stranded breaks, or 'nicks', on gDNA, which are likely to be more tolerable to mammalian cells than the toxic double-stranded DNA breaks generated by Cas9 nuclease. Second, the synthetic mutations are ordered, and insertion identity is pre-programmed by prime editing, compared to the combinations of random insertions and deletions generated during the repair of the double-stranded breaks. Because the editing outcome of prime editing is highly predictable and precise, the resulting data consist of a set of predefined mutations occurring on a set of predefined sites. Lineage-tree reconstruction methods such as the unweighted pair group method with arithmetic mean (UPGMA) can be used to infer clonal relationships more accurately compared to the Cas9-based strategy, in which editing outcomes such as large deletions can affect estimation of edit distances of the same DNA locus in different cells. Third, the sequential editing on the DNA Typewriter system

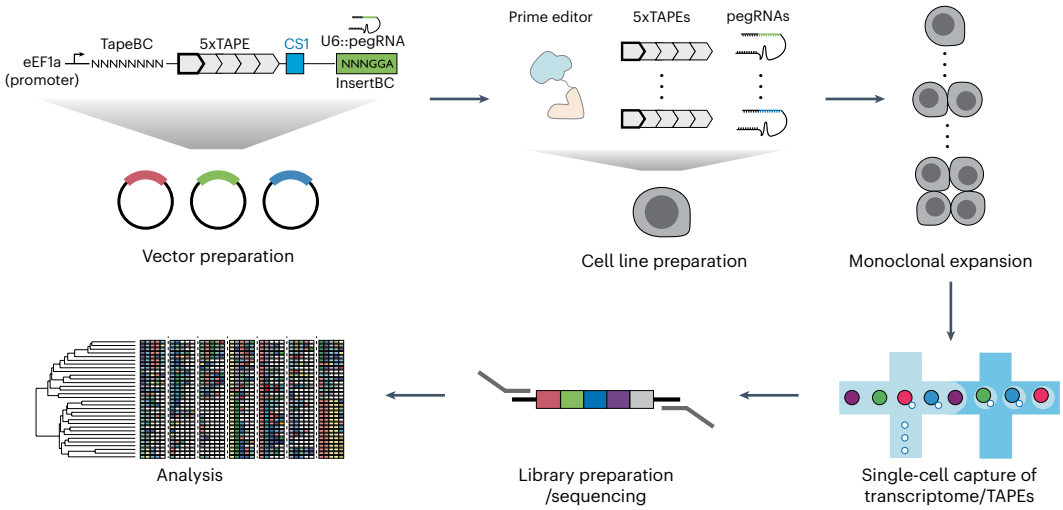

Fig. 3 | Overview of recording single-cell lineage information by using DNA Typewriter. Three DNA Typewriter components (DNA Tapes each with a unique TapeBC, pegRNAs each with a unique InsertBC and a prime editor) are stably integrated into a cell line. A single cell is isolated for monoclonal expansion, during which lineage information is recorded onto DNA Tapes by a prime editor and pegRNAs with different InsertBCs. At the end of the experiment, scRNA-seq methods can be used to capture both the transcriptome and the recorded DNA Tapes from the cells of the monoclonal culture. The clonal relationship between recovered cells can then be inferred.

allows only a portion of the potential writing sites to be available for editing at any given time, stabilizing the editing rate across a longer time period rather than focusing most of the editing early on. However, these advantages come at the expense of the need to integrate multiple pegRNA-expression cassettes to achieve a high complexity of editing. In our implementation of lineage recording using DNA Typewriter, we combined a pegRNA-expression cassette and a TapeBC-5xTAPE-expression cassette into a single lentiviral vector and aimed for a high multiplicity of infection (MOI of 20–30) to have as many writing sites (TAPEs) as possible as well as high complexity in edits that can be made (pegRNA-InsertBC).

Overview of the procedures

Here, we describe two different applications of the DNA Typewriter, which can be summarized as follows.

Procedure 1: recording the order of transfection events with barcoded pegRNAs

The steps are as follows—Steps 1–15: cloning of the DNA Typewriter components; Steps 16–19: stable integration of the recording constructs into target cells; Steps 20–25: cell culture and passaging of the cells with the DNA Typewriter recording components throughout the duration of the experiment; Steps 26–45: gDNA isolation to recover the DNA Tapes from the cell population and sequencing to infer the identity and order of events.

Procedure 2: recording cell lineage information with recovery via scRNA-seq

The steps are as follows—Steps 1–6: cloning of DNA Typewriter lineage-recording components; Steps 7–21: stable integration of the lineage-recording construct into target cells; Steps 22–33: cell culture and passaging of the cells with the DNA Typewriter recording components throughout the duration of the experiment; Steps 34–50: scRNA-seq with recovery of DNA Tapes containing lineage information along with cellular transcriptomes.

Experimental design

Designing the DNA Tape construct

Recording in the DNA Typewriter system is done at the DNA Tape locus. In the initial DNA Tape design, a 14-bp sequence (TGATGGTGAGCACG) is repeated typically five to six times, flanked by a GGA key sequence upstream and an additional downstream sequence (TGATGGTGAGC; the first 11 bp of the repeating tape sequence) to serve as a homology sequence that guides the prime-editing process on the last site[12] (Fig. 1c). We typically also include a unique molecular identifier (UMI) sequence upstream (random 8 bp referred to as 'TapeBC') that distinguishes the different instances of DNA Tape in each cell. The DNA Tape construct, which is ~200–300 bp in length, is typically embedded in the 3′ untranslated region (UTR) of a transcribed RNA cassette such as GFP for its recovery in scRNA-seq platforms. Additional capture sequences such as CS1 or CS2 from 10X Genomics can be added to boost its recovery. In most applications, the DNA Tape sequences are stably integrated into the genome by using lentiviral or transposon vectors.

A successful recording event into DNA Tape will expand each repeat sequence of 14 bp into 17 bp in length by adding the GGA key sequence, with the addition of intervening InsertBC sequences inside the tandem repeats. In our previous experience, five repeats of the Tape (5xTAPE) sequence have a relatively low frequency of losing or gaining repeats[12]. The 5xTAPE design allows the recording of up to five sequential events and fits well within the 100–200 cycles of DNA sequencing on Illumina short read sequencing platforms. The use of longer DNA Tape tends to suffer from contraction or expansion of repeats, especially during PCR amplification for the sequencing library preparation, which can corrupt the recorded information. The addition of UMIs in the form of random 8–15 bp via linear PCR amplification could help identify the change in repeat sizes from different sequencing reads. Finally, although we have not extensively tested the effect of different DNA polymerases, we expect a wide variety of DNA polymerases to work in amplifying DNA sequences containing tandem repeats used in the DNA Typewriter system. However, the PCR reaction needs to be optimized to minimize the cycle number, which may increase the chance of generating chimeric PCR products because of the highly repetitive nature of the DNA Tape sequence.

Achieving the right level of editing efficiency is key to successful recording experiments. The ideal editing efficiency per transfection would be ~10–30%, where the order of multiple events can be reliably captured within a single DNA Tape. To ensure that editing efficiency is adequate for each experiment, we recommend having both positive and negative control conditions in parallel to the recording experiment. The positive control could be the editing of a known target such as the native HEK3 locus, sharing reagents and a cell culture environment with the actual experiment as much as possible. The negative control could be a non-editing condition, which would help assess the baseline non-intentional edited product and estimate the editing rate over the course of the experiment. The editing efficiency also influences the required sequencing depth of DNA Tape to reconstruct the order of events. We have observed that for DNA Tapes with 50% editing in the first site, as few as 2,500 reads can be used to reconstruct the correct order of 16 sequential transfection events[12]. In general, we would recommend aiming for 100,000 reads per sample, in case the editing efficiency is lower than expected.

Considerations for designing pegRNAs for DNA Typewriter

The design process of pegRNAs used in DNA Typewriter is much simpler than designing a pegRNA to induce a desired genomic edit, which includes several parameters such as the spacer sequence, the length of primer-binding sequence, the length of the homology sequence, and the sequence for the editing outcome (Fig. 1d). In DNA Typewriter, we have identified a particular DNA Tape sequence (repeats of 'TAPE-1' 14 bp with GGA Key sequence) that is largely restricted to sequential editing with reasonable efficiency (5–30% per single transfection; tested in HEK293T cells, MEFs and mouse ES cells). Thus, the design parameters are limited to a single one: choosing the insertion sequence ('InsertBC'). The InsertBC must include both the event-specific barcode sequence and the shared key sequence (e.g., NNNGGA) for sequential editing. We have tested two to six random barcode sequences (inserting 5–9 bp of NNGGA to NNNNNNGGA). Although a longer InsertBC sequence can help increase the edit distance among InsertBC sequences for error-correcting barcodes, it typically decreases the efficiency of insertion. In our experience, 9-bp insertions are about half as efficient as 5- and 6-bp insertions. Given the current accuracy of DNA sequencing in base calling, NNNGGA should work well for most cases, achieving a high recording efficiency while allowing 64 different InsertBCs, or symbols, to be used within a single experiment.

The insertion efficiency of individual InsertBCs varies but is typically consistent across experiments. One could choose a set of InsertBCs with the highest editing efficiency while achieving maximal edit distance among them. The editing efficiency of each InsertBC can be used to correct the observed frequencies of the insertions, which enables an accurate estimate of relative insertion frequencies or intensity of events across different InsertBCs. It is still unclear what all the determinants of insertion efficiencies are. Anecdotally, we observe that poly-A insertion sequences tend to perform worse, possibly because they form poly-U template sequences in pegRNA prone to termination of Pol-3 during transcription[12]. Another possibility is the formation of the overall secondary and tertiary structure of the pegRNA, in which some of the InsertBC sequences can interact with the guide RNA scaffold sequence that disfavors its interaction with the Cas9 nickase portion of the prime editor[28].

Finally, any advances in designing pegRNAs can be used to increase the efficiency of pegRNAs used in DNA Typewriter. For example, recent advances in pegRNA architecture such as the addition of a 3′-end RNA pseudoknot[29] ('epegRNA') can be used to improve editing efficiency (Fig. 1d). In addition to the sequence that we have identified for DNA Typewriter, one could also search for better DNA Tape sequences with higher editing rates. In our study, we screened 48 different designs and identified two sequences for sequential editing[12], but a larger-scale screen might further increase the efficiency and/or precision of ordered recording with DNA Typewriter.

Choice of recording cell line and delivery method

To record ordered transfections of pegRNA with DNA Typewriter, two other components need to be present within the cell: the prime editor enzyme, to catalyze the recording, and DNA Tape loci as substrates for recording. Although a prime editor–expressing plasmid can be transiently delivered to cells along with each pegRNA to initiate the editing, we recommend that

the DNA Tape loci be stably integrated within the genome to minimize the loss of information over time. In addition, the large size of the prime editor plasmid (typically 10–12 kb) and of the prime editor expression cassette (6 kb encoding the prime editor with additional transcription promoter and terminator elements) may limit the delivery of the prime editor into many hard-to-transfect cell types; thus, a recorder cell line with prime editor stably expressed may be useful for increasing the percentage of cells with all three components for molecular recording.

The time interval between each pegRNA delivery event will also need to be considered. In its original implementation, DNA Typewriter was able to resolve ordered transfections in 3-day intervals, which we chose to minimize the toxicity of transfection by using liposome-based delivery of the pegRNAs to HEK293T cells. Because one needs to deliver exogenous pegRNAs into cells sequentially, the cell line of choice will need to be able to handle multiple rounds of transfection or transduction. In our case, we observed stable cell growth over 16 rounds of transfection to HEK293T cells over 48 days[12].

The choice of cell type in which to perform molecular recording will determine how to deliver the prime-editing components. Previously, we have shown the use of DNA Typewriter in HEK293T cells, mouse ES cells and MEFs[12]. Manufacturers' protocols tailored specifically to these cell types were followed for recording the order of transfection: Lipofectamine 3000 for HEK293T cells and Lonza nucleofector for mouse ES cells and MEFs. In the following protocol, we provide the DNA Typewriter procedure specific to HEK293T cells, but the details of cell line generation and nucleic acid delivery methods can be modified to suit the recording cell line of choice. We expect that most established protocols that tailor both stable integration of genetic constructs and transient transfection of nucleic acids to the cell line of choice will be suitable for molecular recording with DNA Typewriter. Lastly, in cases in which transfection or stable integration of large cargo such as the prime editor plasmid is difficult, delivery of ribonucleoprotein prime-editing complex can be considered[35], although this procedure is not covered in the following protocol.

Extracting recorded information from a DNA sequence

Information recorded to DNA Tape needs to be PCR amplified either from gDNA or complementary DNA (cDNA) generated from transcribed cellular RNA and sequenced as a short DNA amplicon. The design of the DNA amplicon should include the entire DNA Tape sequence and the TapeBC if present. In the case of using scRNA-seq platforms such as 10X Genomics, Read1 will include a 16-bp cell-identifying sequence (cell barcode, CellBC) and a 12-bp UMI sequence from the cDNA generation step. The entire DNA Tape sequence, along with any insertions, will be sequenced from Read2. To allocate sequencing cycles for DNA Tape, reading through the entire DNA Tape sequence is advantageous to check whether there has been any expansion of the repeat region during the recording. Assuming that five repeats of a writing site are fully edited with 6-bp insertions, the DNA Tape region alone will be greater than 120 bp without including flanking sequences such as the TapeBC. Therefore, a 150- or 200-cycle sequencing kit may be minimally required to sequence the entire DNA Tape region.

From the sequencing reads, reads containing errors from PCR amplification or sequencing need to be filtered out, and multiple short DNA sequences need to be extracted. Optionally, one could use paired-end sequencing reads to bidirectionally cover the entire DNA Tape and identify sequencing errors, at the expense of higher sequencing cost. In both cases, we recommend the following steps in filtering out reads: (i) use pattern matching to identify constant regions flanking either side of the DNA Tape, ensuring that the read contains the DNA Tape at the expected location; (ii) extract the TapeBC and retain an overall set of TapeBCs for analysis based on a read-count threshold; and (iii) extract the InsertBC from each position by pattern-matching the region flanking the InsertBC (e.g., for an NNNGGA InsertBC, use the flanking sequence of ACG… TGA as in *ACG*NNNGGA*TGA*). These steps would result in a table of filtered reads as the rows with the following as the columns: TapeBC, InsertBC in the first site of the DNA Tape, InsertBC in the second site of the DNA Tape and so forth. In a single-cell experiment, there would be an additional column of CellBC unique to each single cell. This table can then be used to analyze the relationships between InsertBCs within the same read to order multiple events or compare the set of InsertBCs across CellBC-TapeBC pairs to estimate clonal distances between cell pairs.

Protocol

Materials

Biological materials

- *E. coli* strains: C3040H competent cells (NEB, cat. no. C3040H)
- Mammalian cell lines: HEK293T (American Type Culture Collection, cat. no. CRL-3216; RRID: CVCL_0063) or desired target cell line
  ▲ **CAUTION** Cell lines should be regularly screened for mycoplasma contamination.

Reagents

Plasmids

- PE2 (Addgene, cat. no. 132775) or PE4max (Addgene, cat. no. 174828 for pCMV-PEmax-P2A-hMLH1dn)
- epegRNA cloning backbone (Addgene, cat. no. 174038)
- piggyBAC-rtTA (Addgene, cat. no. 126034; or equivalent)
- piggyBAC Transposase Super PiggyBac Transposase expression vector (System Bioscience, cat. no. PB210PA-1; or equivalent)
- piggyBAC-iPE2 (XLone-PE2; Addgene, cat. no. 136463; or equivalent)
- Lentiviral CRISPR droplet sequencing (CROP-seq) backbone (Addgene, cat. no. 200029)
- piggyBAC-U6-GFP-WPRE (Addgene, cat. no. 200030)

Cloning of DNA plasmids

- DNase/RNase-free distilled water (Invitrogen, cat. no. AM9937)
- Ethanol, 200 proof (Decon Labs, cat. no. 2701)
- DNA oligonucleotides for cloning and PCR (Integrated DNA Technologies custom DNA oligos, Supplementary Table 1)
- Nuclease-free water (Invitrogen, cat. no. AM9937)
- 5 M NaCl (Thermo Fisher, cat. no. AM9759)
- KAPA2G Robust 2× Hotstart mix (Kapa, cat. no. KK5702)
- SYBR Green (Thermo Fisher, cat. no. S7563)
- Agencourt AMPure XP (Beckman Coulter, cat. no. A63882)
- Ultrapure agarose (Thermo Fisher, cat. no. 16500-500)
- Ethidium bromide solution (Bio-Rad, cat. no. 1610433)
- 1 kb Plus DNA ladder (Thermo Fisher, cat. no. 10787018)
- 6× gel loading dye, purple (NEB, cat. no. B7024S)
- LB medium (e.g., Thermo Fisher, cat. no. 12795027)
- LB agar medium (e.g., Thermo Fisher, cat. no. 22700025)
- Ampicillin sodium salt (Fisher Scientific, cat. no. BP1760-25)
- Monarch gel extraction clean-up kit (NEB, cat. no. T1020S)
- Qiagen Mini kit (Qiagen, cat. no. 27106)
- ZymoPure Midi kit (Zymo, cat. no. D4201)
- 10× rCutSmart buffer (NEB, cat. no. B6004S)
- BsaI-HF-v2 (NEB, cat. no. R3733S)
- BsmBI-HF-v2 (NEB, cat. no. R0739S)
- MluI-HF (NEB, cat. no. R3198S)
- KpnI-HF (NEB, cat. no. R3142S)
- T4 polynucleotide kinase (NEB, cat. no. M0201S)
- 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 ATP in the buffer to degrade. Divide into aliquots of working volumes to minimize freeze-thaw cycles, and store as recommended by the manufacturer.
- NEBuilder HiFi DNA assembly master mix (NEB, cat. no. E2621S)

Mammalian cell culture

- DMEM (Thermo Fisher, cat. no. 119650192)
- GlutaMAX supplement (Thermo Fisher, cat. no. 35050061)
- FBS characterized (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 PBS 1× (Gibco, cat. no. 14190-144)
- Trypsin 0.25% EDTA (Thermo Fisher, cat. no. 25200056)
- Lipofectamine 3000 transfection reagent kit (Thermo Fisher, cat. no. L3000001)
- Opti-MEM reduced serum medium (Thermo Fisher, cat. no. 31985062)
- 100 mM sodium pyruvate (Thermo Fisher, cat. no. 11360070)
- ViraPower lentiviral packaging mix (Thermo Fisher, cat. no. K497500)
- Bleach solution (Fisher Scientific, cat. no. NC1796686)
- PEG-it virus precipitation solution (5×) (System Biosciences, cat. no. LV810A-1)
- ViroMag R/L (OZ Biosciences, cat. no. RL40100)
- Magnetofection magnetic plate (OZ Biosciences, cat. no. MF10000)
- Proteinase K (Thermo Fisher, cat. no. EO0491)
- 1 M Tris-HCl, pH 8.5 (VWR, cat. no. 76236-402)
- Ultrapure SDS 10% solution (Thermo Fisher, cat. no. 15553027)
- (Optional) Dneasy blood & tissue kit (Qiagen, cat. no. 69504)

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)
- 10× Genomics chromium controller (10X Genomics, cat. no. 1000204)
- Illumina NextSeq 550 system (Illumina, cat. no. SY-415-1002; or alternatives)

General equipment

- 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)
- Light microscope with GFP excitation laser and filter (Zeiss Axio Observer 3 or similar)

Software

- (Optional) CRISPResso2[36] (https://github.com/pinellolab/CRISPResso2)

Reagent setup

Single-stranded DNA oligonucleotides

Resuspend lyophilized oligos to 100 µM stock concentration in the volume of nuclease-free water or DNA elution buffer (10 mM Tris-HCl, pH 8.5) recommended by the manufacturer. Stocks can be stored at −20 °C for ≥1 year.

Complete cell lysis buffer for gDNA preparation

To 10 ml of nuclease-free water, add 100 µl of 1 M Tris-HCl pH 8.5, 50 µl of 10% ultrapure SDS solution, and 20 µl of proteinase K. Prepare complete cell lysis buffer fresh each time.

BOX 1

HEK293T cell culture and passaging

● **TIMING 1 h**

Procedure

1. Thaw a frozen stock vial of HEK293T cells in a 37 °C water bath until the ice core has just disappeared.
2. Gently resuspend the cells in 10 volumes of complete HEK293T cell medium.
3. Transfer the cell suspension to a 15-ml Falcon tube and centrifuge at 300*g* for 5 min.
4. Remove the supernatant and resuspend the cell pellet in 10 ml of HEK293T cell medium.
5. Find the concentration of live cells per milliliter by using a cell counter.
6. Calculate the volume of cell suspension needed for plating 10,000–20,000 cells per square centimeter of the culture dish or flask. Plate the cell suspension in a culture dish or flask in an appropriate volume for the container. If using a 10-cm culture dish, the total volume of the medium should be ~10–15 ml. Distribute the cells evenly across the culture dish or flask by gently shaking along the vertical axis and then the horizontal axis.
7. Place the cells in a cell culture incubator set at 37 °C and 5% CO_2.
8. Monitor cell growth by inspecting the cells under a brightfield microscope. When the cells have reached 85% confluency (after 2–3 days of culture), aspirate the culture medium and wash the cells once with 10 ml of 1× PBS, if using a 10-cm culture dish.
9. Add a volume of trypsin sufficient to cover the bottom of the culture dish (or flask) and incubate the cells in an incubator at 37 °C and 5% CO_2 for 5 min.
10. Inactivate the trypsin by adding ≥3 volumes of HEK293T cell medium supplemented with 10% (vol/vol) FBS. Pipette up and down gently to ensure full cell dissociation.
11. Centrifuge the cell suspension at 300*g* for 5 min. Aspirate the supernatant, leaving a hard cell pellet at the bottom of the tube.
12. Follow option A to propagate the culture or start a new experiment or option B to freeze the cell for long-term storage.
    (A) **Propagation of the culture: resuspend the cell pellet in 10 ml of HEK293T cell medium and plate the cell suspension in a clean culture dish for the propagation of the culture or for the start of an experiment**
    (B) **Freezing of cells for long-term storage**
    (i) Resuspend the cell pellet in 10 ml of freezing medium of 10% (vol/vol) dimethyl sulfoxide in FBS.
    (ii) Distribute aliquots of the cell suspension into cryovials and freeze the vials in a CoolCell insulator at −80 °C for ≥4 h.
    (iii) Transfer the frozen vials to liquid nitrogen for long-term storage.

Annealing buffer
To 9.8 ml of nuclease-free water, add 100 µl of 1 M Tris-HCl pH 8.0 and 100 µl of 5 M NaCl solution. The buffer can be stored at room temperature (22–25 °C) indefinitely.

HEK293T cell culturing medium
Aseptically mix 440 ml of DMEM, 5 ml of GlutaMAX supplement, 5 ml of 1× penicillin-streptomycin, and 50 ml of FBS.

Lentivirus packaging medium
Aseptically mix 469 ml of Opti-MEM I reduced serum medium, 5 ml of GlutaMAX supplement, 25 ml of FBS, and 1 ml of 100 mM sodium pyruvate.

HEK293T cell culture and passaging
Maintain HEK293T cells in culture as described in Box 1.

Procedure 1

Cloning of DNA Tape-targeting epegRNA expression cassettes

● **TIMING 3–4 days**

▲ **CRITICAL** The following cloning protocol has been adapted from the Golden Gate cloning strategy from the initial report describing prime editing[13].

1. Choose a set of InsertBCs to be cloned as epegRNAs; the chosen barcodes will be inserted into the DNA Tape via prime editing.
2. Clone epegRNA expression cassettes, by using option A (Golden Gate cloning strategy) if cloning individual cassettes, or option B (isothermal assembly also known as 'Gibson

Assembly') if cloning a library of epegRNA expression cassettes after obtaining a pool of oligonucleotides.

(A) **Generation of individual epegRNA expression cassettes by Golden Gate cloning**

● **TIMING 3–4 days**

(i) Prepare a restriction enzyme digest reaction with 5 µg of the pU6-epegRNA-GG-acceptor plasmid backbone and BsaI-HF-v2 as follows:

Component	Volume (µl)	Final concentration
pU6-epegRNA-GG-acceptor plasmid	Volume for 5 µg	–
BsaI-HF-v2 (20,000 U/ml)	5	2,500 U/ml
10× rCutSmart buffer	4	1×
Nuclease-free water	Up to 40	–

(ii) Incubate the reaction at 37 °C for ≥3 h or overnight in a thermocycler.

(iii) Prepare a 0.8% (wt/vol) agarose TAE gel supplemented with 1:10,000 (vol/vol) ethidium bromide. Mix five volumes of the restriction digest reaction with one volume of 6× purple loading dye and load the sample into the gel along with a DNA ladder in a separate lane.

(iv) Run the gel in 1× TAE buffer at 120 V cm^{-1} for 30 min.

(v) Use a clean blade to cut the relevant band from the gel and perform gel extraction by following the instructions provided in the Monarch gel extraction kit. The size of the expected digestion product is 2.2 kb.

■ **PAUSE POINT** Linearized fragments can be stored at 4 °C for several weeks or at −20 °C indefinitely.

(vi) Quantify the concentration of the extracted linearized plasmid by UV-visible spectrophotometry (Nanodrop) or an equivalent method.

(vii) Order and prepare the components for the three Golden Gate splints by following Box 2. See Supplementary Table 1 for oligo designs OLG_GG01 through OLG_GG06.

(viii) Using the linearized plasmid vector (from Step 2A(vi)) and the three splint oligos (Box 2), set up the Golden Gate assembly reaction in a PCR strip tube as follows:

Component	Volume (µl)	Final concentration
Splint 1 (1 µM)	1	0.1 µM
Splint 2 (1 µM)	1	0.1 µM
Splint 3 (1 µM)	1	0.1 µM
BsaI-HF digested pU6-epegRNA-GG-acceptor plasmid	Volume for 30 ng of digested vector	–
10× T4 ligase reaction buffer	1	1×
T4 DNA ligase (400,000 U/ml)	0.5	20,000 U/ml
BsaI-HF-v2 (20,000 U/ml)	0.2	4,000 U/ml
Nuclease-free water	Up to 10	–

(ix) Perform the Golden Gate assembly reaction by using the following thermocycler conditions:

Step	Temperature (°C)	Time
1	16	20 min
2	37	5 min
3	85	5 min
4	4	Hold

(x) (Optional) To increase the efficiency of Golden Gate assembly, follow the thermocycler protocol below:

Step	Temperature (°C)	Time	Cycling instructions
1	37	2 min	–
2	16	2 min	Go back to step 1; repeat 20–60 times

Step	Temperature (°C)	Time	Cycling instructions
3	4	Hold	–
4	65	10 min	Perform this step before transformation
5	4	Hold	–

(xi) Transform 1 µl of the assembled Golden Gate reaction into 10 µl of NEB C3040H competent cells by following the manufacturer's protocol.

(xii) Plate the entire volume of the transformation reaction onto an LB agar plate containing 50 µg/ml ampicillin and incubate overnight at 37 °C.

■ **PAUSE POINT** The transformed plate can be stored at 4 °C for 1 week.

◆ **TROUBLESHOOTING**

(xiii) Pick individual RFP-negative colonies (colonies that still carry the RFP gene will appear red by eye) and transfer them to individual 14-ml round-bottom tubes containing 5 ml of LB supplemented with 100 µg/ml ampicillin.

(xiv) Grow the cultures in a shaking incubator (200 rpm, 37 °C) overnight.

(xv) Purify plasmids from bacterial culture with the Qiagen miniprep kit and verify the insert by using Sanger sequencing.

(B) **Generation of a library of epegRNA expression cassettes by isothermal assembly**

● **TIMING 3–4 days**

(i) Perform a BsaI-HF restriction enzyme digest as described in Steps 2A(i–vii) to linearize the pU6-epegRNA-GG-acceptor plasmid for epegRNA cloning.

(ii) Order and prepare a double-stranded pool of epegRNA sequences by following the instructions in Box 2. The library will vary at the InsertBC region.

(iii) Set up the isothermal reaction as follows:

Component	Volume (µl)	Final concentration
Acceptor vector	For 30 ng of digested vector	–
Double-stranded DNA pool	For threefold molar ratio over the amount of the acceptor vector	–
NEBuilder HiFi 2× master mix	5	1×
Water	Up to 10	–

(iv) Incubate the isothermal assembly at 50 °C for 15–60 min.

(v) Transform 5 µl of the isothermal assembly reaction into 50 µl of C3040H competent cells by following the manufacturer's protocol.

(vi) Dilute the transformed cells by adding 950 µl of LB medium supplemented with 100 µg/ml ampicillin.

(vii) Plate 10 µl of the transformed cells on an LB agar plate containing ampicillin. Add the rest of the cells (~990 µl) to a 250-ml Erlenmeyer flask containing 50 ml of LB supplemented with 100 µg/ml ampicillin.

▲ **CRITICAL STEP** On the next day, the estimated number of transformation events will be 100 times the number of colonies.

(viii) Incubate the plates and the flasks at 30 °C for 24 h or 37 °C for 16–18 h. The flasks should be placed in a shaking incubator set at 200 rpm.

▲ **CRITICAL STEP** If the plasmid contains repeat elements as in the case of lentiviral or piggyBAC transposon vectors, grow C3040H cells at 30 °C to minimize recombination.

(ix) The next day, inspect the plates and count the number of individual colonies to assess the complexity of the library. Purify the plasmids from the cell cultures by using the ZymoPure Midi kit.

(x) Sequence the purified plasmid by using Sanger sequencing to verify the cloning of the plasmid pool. Alternatively, use next-generation sequencing to quantify the relative frequency of individual constructs within the pool by following Steps 30–40 and substituting gDNA with plasmid DNA.

BOX 2

Preparation of epegRNA cloning components

● **TIMING 1–4 h**

Procedure

1. Follow option A if designing oligos for Golden Gate cloning or option B for isothermal assembly.

(A) **Design and preparation of oligos for Golden Gate cloning**

(i) Using the Golden Gate splint sequences provided in Supplementary Table 1 as examples, order two oligos as the top and bottom strands of a double-stranded DNA with 5′ overhangs for each splint (splints 1–3).

▲ **CRITICAL STEP** Three splints are required: (i) guide RNA spacer, (ii) guide RNA scaffold, and (iii) 3′ extension containing the primer-binding site, insertion sequence, and homology sequence. Splint 2 (OLG_GG03 and OLG_GG04) is constant in all epegRNA designs, while splints 1 (OLG_GG01 and OLG_GG02) and 3 (OLG_GG05 and OLG_GG06) specify the sequence to target in the DNA Tape and the desired insertion outcome, respectively.

(ii) Prepare the annealing reaction of the top and bottom single-stranded DNA oligonucleotides to create double-stranded DNA with overhangs for Golden Gate assembly as follows:

Component	Volume (µl)	Final concentration (µM)
100 µM top-strand oligonucleotide	1	4
100 µM bottom-strand oligonucleotide	1	4
Annealing buffer	23	–
Total volume	25	–

(iii) Incubate the reaction in a thermocycler and anneal the oligos as follows:

Step	Condition	Time
1	95 °C	3 min
2	Ramp down to 22 °C at −0.1 °C/s	–
3	4 °C	Hold

(iv) (Optional) If oligos are purchased without the 5′ phosphorylation modification, manually phosphorylate the splint with T4 PNK enzyme in the following reaction:

Component	Volume (µl)	Final concentration
4 µM unphosphorylated splint	25	1 µM
10× T4 DNA ligase reaction buffer	10	1×
T4 PNK (10,000 U/ml)	2	200 U/ml
Nuclease-free water	63	–
Total volume	100	–

▲ **CRITICAL STEP** Only splint 2 must be phosphorylated (splints 1 and 3 can be unphosphorylated) for the Golden Gate assembly procedure. However, phosphorylation of all three splints aids the success of cloning, if needed.

(v) (Optional) Incubate the T4 PNK reaction at 37 °C for 1 h. After the phosphorylation step, the splints will be at 1 µM concentration.

(vi) (Optional) If the oligos were purchased with the 5′ phosphorylation modification, dilute the splints to 1 µM concentration by adding 75 µl of water to the 25-µl annealing reaction.

■ **PAUSE POINT** Phosphorylated splints can be stored at −20 °C for several months.

(B) **Design and preparation of an epegRNA DNA pool for isothermal assembly**

● **TIMING 3–4 days**

(i) Design a DNA pool containing epegRNA sequences targeting the same DNA Tape sequence but differing in insertions. Order this DNA pool (GF_01) as single-stranded DNA. The design in Supplementary Table 1 includes all components of the epegRNA (i.e., spacer, sgRNA scaffold, reverse transcription template, and the primer-binding site) and 15 bp of homology arms for isothermal assembly to the acceptor vector flanking the epegRNA design. The evoPreQ$_1$ knot sequence is already included in the plasmid backbone, so it is not included in the designed oligonucleotide.

(ii) Following the example in Supplementary Table 1, order a PCR primer set (OLG001 and OLG002) to amplify the pool of TAPE-targeting epegRNA fragments to make the DNA pool double stranded.

(continued from previous page)

(iii) Using the single-stranded DNA pool from Step 1B(xi) as a template and the PCR primer pair from Step 1B(xii), set up the qPCR reaction as indicated in the table below to prepare a double-stranded DNA pool of TAPE-targeting epegRNA sequences.

Component	Volume (µl)	Final concentration
DNA pool from Step 1B(xi)	1–2	–
Forward primer (100 µM)	0.1	0.4 µM
Reverse primer (100 µM)	0.1	0.4 µM
KAPA Robust 2× Hotstart mix	12.5	1×
SYBR Green (100×)	0.1	0.4×
Water	Up to 25	–

▲ **CRITICAL STEP** We recommend diluting the pool of single-stranded DNA to 1 ng/µl in water and using 1–2 µl of the diluted pool in the PCR reaction. However, a larger amount of the DNA pool can be used to minimize the number of PCR cycles required.

(iv) Perform qPCR by using the following thermocycler program:

Step	Temperature (°C)	Time	Cycling instructions
1	95	2 min	–
2	95	15 s	–
3	65	15 s	–
4	72 (with the plate read at the end)	90 s	Go back to step 2; repeat 5–8 cycles
5	72	3 min	–
6	10	Hold	–

▲ **CRITICAL STEP** The number of PCR cycles can be optimized depending on the amount of DNA template loaded, although the total number should be kept to a minimum to avoid PCR amplification biases.

(v) Purify the double-stranded DNA pool by using AMPure XP beads and following the manufacturer's instructions or by using a preferred method of DNA purification.

Cloning a pool of UMI-labeled DNA Tape expression cassettes

● **TIMING 3–4 days**

3. Design and order the DNA Tape sequence pool as a single-stranded oligonucleotide and primers to amplify the DNA Tape oligonucleotide by following the examples in Supplementary Table 1.
   ▲ **CRITICAL STEP** Each of the DNA Tape sequences in the pool contains a unique 8-bp barcode (NNNNNNNN) known as 'TapeBC', which enables the identification of unique DNA Tapes that have been integrated into the cell's genome.
4. Perform a restriction digest on the piggyBAC-U6-GFP-WPRE plasmid with the restriction enzyme KpnI-HF as follows and prepare the linearized DNA plasmid fragment to clone in the DNA Tape sequence as described in Step 2A(ii–vi).

Component	Volume (µl)	Final concentration
piggyBAC-U6-GFP-WPRE plasmid	Volume for 5 µg	–
KpnI-HF (20,000 U/ml)	5	2,500 U/ml
10× rCutSmart buffer	4	1×
Nuclease-free water	Up to 40	–

   ▲ **CRITICAL STEP** The DNA Tape sequence is placed in the 3′ UTR of GFP for later selection of cells that have stably integrated DNA Tape.
5. Set up the qPCR reaction to prepare a pool of DNA Tapes with homology arms for isothermal assembly, as described in Box 2, step 1B(iii–xv). Use the DNA pool and primers ordered in Step 3 for the qPCR reaction. Amplification must be done with a low number of cycles (fewer than five) to minimize PCR-associated issues with repetitive sequences.

▲ **CRITICAL STEP** PCR amplification might contribute to the contraction and loss of repetitive TAPE sequences. The PCR step can be avoided altogether by ordering the entire DNA Tape sequence as two separate single-stranded DNA oligos to be annealed into a double-stranded fragment as described in Box 2, step 1A(ii–iii).

6. Purify the double-stranded DNA by using AMPure XP beads and following the manufacturer's recommendations or use a preferred method of DNA purification. Quantify the concentration of this DNA pool with UV-visible spectrophotometry (NanoDrop).
7. Set up the isothermal assembly reaction as follows:

Component	Volume (µl)	Final concentration
Purified, cut vector (from Step 6)	For 30 ng of digested vector	3 ng/µl
DNA Tape pool (from Step 5)	For threefold molar ratio over the amount of the acceptor vector	–
NEBuilder HiFi 2× master mix	5	1×
Water	Up to 10	–

8. Incubate the isothermal assembly at 50 °C for 60 min in a thermocycler.
9. Transform 5 µl of the isothermal assembly reaction into 50 µl of C3040H competent cells by following the manufacturer's protocol.
10. Dilute the transformed cells by adding 950 µl of LB medium with 100 µg/ml ampicillin.
11. Plate 10 µl of the transformed cells on an LB agar plate containing ampicillin. Add the rest of the cells (~990 µl) to a 250-ml Erlenmeyer flask containing 50 ml of LB supplemented with 100 µg/ml ampicillin.
12. Incubate the plates and the flasks at 30 °C for 24 h. The flasks should be placed in a shaking incubator set at 200 rpm.
13. On the next day, inspect the plates and count the number of individual colonies to assess the complexity of the library. The estimated number of transformation events is 100 times the number of colonies. Purify the plasmids from the cell cultures by using the ZymoPure Midi kit.
14. Use Sanger sequencing to verify the cloning of the plasmid pool. Use next-generation sequencing to quantify the relative frequency of individual constructs within the pool by following Steps 30–40 and substituting gDNA with plasmid DNA.
    ◆ **TROUBLESHOOTING**

Recording the order of transfection events in HEK293T cells

● **TIMING 14–64 days**

Generating a DNA Typewriter recorder cell line

● **TIMING 10–12 days**

15. Thaw a stock vial of HEK293T cells at passage 5 or lower and culture the cells as described in Box 1, before conducting the recording experiment.
16. Generate a stable DNA Tape recorder cell line by co-transfecting cells with the DNA Tape transposon vector pool and transposase vector at a mass ratio of 8:2. Follow the manufacturer's instructions for Lipofectamine 3000 to transfect the HEK293T cells.
    ▲ **CRITICAL STEP** Different ratios between transposon and transposase can be tested to achieve high multiplicity of integration. A higher number of integration events (15–20 events) would be ideal to resolve the order between signals and allocate enough targets for editing in each cell.
17. Wait 5–7 days after the initial transfection for cells to recover and for episomal plasmid to dilute out of the transfected cell population.
18. Select the cells with stable integration of DNA Tape. The top 15% of GFP-expressing cells can be sorted via FACS as a proxy for cells with a higher number of stably integrated DNA Tapes.
    ■ **PAUSE POINT** The cells sorted for stable integration of DNA Tape can be frozen and stored in liquid nitrogen indefinitely by following the procedures described in Box 1 and used for recording experiments at a later time.
    ◆ **TROUBLESHOOTING**

Sequential transfections with the DNA Tape recorder cell line

● **TIMING 3–51 days**

19. Seed the HEK293T recorder cell line from Step 19 into a 24-well plate (WP) so that the cells reach 60–80% confluency ~24 h later at the time of transfection.
20. Following the Lipofectamine 3000 manufacturer's protocol, prepare the DNA transfection mix for transfecting cells in a 24-WP, such that a mass ratio of 1:3 epegRNA plasmid (from Step 2A(xvi) or 2B(x)) and prime editor plasmid (PE2 or PE4max) is achieved.
21. Perform co-transfection of the epegRNA plasmid and prime editor plasmid in the HEK293T recorder cell line by adding the DNA transfection mix from Step 21 dropwise to the center of the well. After returning the plate to the incubator, ensure an even distribution of the DNA transfection mix in the medium by gently shaking the plate back and forth several times along its horizontal axis, then its vertical axis.
    ▲ **CRITICAL STEP** The transfection efficiency determines the observed editing efficiency in prime editing. Transfect the cells by closely following the Lipofectamine 3000 protocol as per the manufacturer's instructions to reach the optimal transfection efficiency for HEK293T cells.
    ▲ **CRITICAL STEP** Generally, we see editing efficiency increase over the first 3 days after transfection because of an accumulation of edited targets. The percentage of edited cells and percentage of edited sites will depend on three factors: the transfection efficiency, the rate at which the transfected plasmids dilute out with cell expansion and prime-editing efficiency.
22. Within the next 6–18 h, passage 10% of the cells from each well to a new set of wells in a 24-WP. If desired, collect the remaining cells for gDNA preparation (Steps 26–29).
23. Incubate the cells at 37 °C in an incubator at 5% CO_2 and monitor their growth daily.
24. Once the cells have reached 60–80% confluency, repeat Steps 22–25 for multiple rounds of sequential transfections to record the order of transfection events. It typically takes 3 days for plated HEK293T cells to reach 60–80% confluency.
    ◆ **TROUBLESHOOTING**

Collection of the DNA Tape from cells: cell lysis and gDNA harvest

● **TIMING 2–3 h**

25. Trypsinize the monolayer of HEK293T cells (from Step 25) to lift off the cells and estimate the cell number as described in Box 1.
26. Aspirate the supernatant, leaving a hard cell pellet at the bottom of the tube.
27. Lyse the cells by resuspending the cell pellet in 1 µl of freshly prepared gDNA lysis buffer for every 2,000 cells. Pipette up and down to break up the pellet and ensure complete lysis.
    ▲ **CRITICAL STEP** Alternatively, gDNA can be purified by using commercially available kits such as DNeasy blood & tissue kit and following the manufacturer's recommendations.
    ▲ **CRITICAL STEP** Cell lysis with SDS initially makes the solution viscous at high cell concentrations. Once the protein digestion is complete, the solution returns to its original viscosity.
28. Transfer the lysis reaction into a PCR strip tube and run the following thermocycler protocol:

Step	Temperature (°C)	Time
1	50	60 min
2	85	30 min
3	4	Hold

■ **PAUSE POINT** The resulting gDNA lysis mix can now be used for subsequent library preparation. It can be stored at 4 °C for short-term storage (~1 week) or −20 °C for several months.

Preparation of the Illumina sequencing library from gDNA samples

● **TIMING 4–6 h**

29. Design and order PCR primers (OLG009 and OLG010) to selectively amplify the DNA Tape locus from gDNA (from Step 29) by following the example in Supplementary Table 1.
▲ **CRITICAL STEP** The forward PCR primer contains the Illumina Nextera handle, and the reverse PCR primer contains the Illumina TruSeq handle, as used in the original implementation of DNA Typewriter. However, we do not foresee any problems with using another combination of Nextera and/or TruSeq handles to amplify DNA Tape sequences from gDNA.
30. In a PCR tube, set up the following qPCR reaction mix:

Component	Volume (µl)	Final concentration
gDNA (from Step 29)	1.6–2	6–10 ng/µl
100 µM OLG009	0.2	0.4 µM
100 µM OLG010	0.2	0.4 µM
SYBR Green (100×)	0.2	0.4×
KAPA2G Robust 2× master mix	25	1×
Water	Up to 50	–

▲ **CRITICAL STEP** The amount of gDNA to be used may require further optimization to avoid bottlenecking the number of unique DNA Tape molecules available for sequencing in the sample; however, adding too much gDNA may inhibit the PCR reaction.

31. Load the samples from Step 31 in a thermocycler and amplify the DNA Tape loci by running the following qPCR thermocycler conditions:

Step	Temperature (°C)	Time	Cycling instructions
1	95	2 min	–
2	95	15 s	–
3	65	15 s	–
4	72 (with the plate read at the end)	90 s	Go back to step 2; repeat ~25 cycles
5	72	3 min	–
6	10	Hold	–

▲ **CRITICAL STEP** Because DNA TAPE fragments with edits are not much longer than fragments without edits, there should not be huge PCR amplification biases within the pool of edited and unedited DNA TAPE species. However, we still recommend keeping the number of PCR cycles moderate (~25), certainly less than 40 in total. Monitor the SYBR Green curves and stop the cycling reaction right before the curves reach their inflection point.

32. Purify the qPCR reactions with AMPure XP beads by following the manufacturer's instructions. Elute in 50 µl of water. We recommend a 1× AMPure XP bead clean-up for 5xTAPE.
▲ **CRITICAL STEP** Choose the appropriate AMPure XP bead ratio to size-select for the length of the amplicon containing the DNA Tape used.
33. Check for the correct size of the resulting purified product by running 1 µl on the TapeStation with a D1000 ScreenTape or by using gel electrophoresis.
34. Design and order Nextera P5 (OLG011) and TruSeq P7 (OLG012) Illumina indexing primers to prepare the final Illumina sequencing library. Follow the example in Supplementary Table 1.
35. To add P5 and P7 sequencing adapters and indices, set up the following qPCR reaction in a PCR strip tube:

Component	Volume (µl)	Final concentration
AMPure XP bead-purified PCR product from Step 33	0.5–1	–
10 µM F OLG011	1	0.4 µM
10 µM R OLG012	1	0.4 µM

Component	Volume (µl)	Final concentration
SYBR Green (100×)	0.1	0.4×
KAPA2G Robust 2× master mix	12.5	1×
Water	Up to 25	–

36. Place the tube in a thermocycler and perform the following qPCR:

Step	Temperature (°C)	Time	Cycling instructions
1	95	2 min	–
2	95	15 s	–
3	65	15 s	–
4	72 (with the plate read at the end)	90 s	Go back to step 2; repeat 7 cycles
5	72	3 min	–
6	10	Hold	–

▲ **CRITICAL STEP** 5–8 PCR cycles should be sufficient to convert most of the DNA fragments into ones with Illumina sequencing adapters.

37. Purify the qPCR reaction with AMPure XP beads by following the manufacturer's protocol and using a 0.8–0.9× bead clean-up for 5xTAPE.
▲ **CRITICAL STEP** If not using 5xTAPE, choose the appropriate AMPure XP bead ratio for the length of the DNA Tape amplicon.
38. Check for the correct size of the resulting purified product by running 1 µl on the TapeStation with a D1000 ScreenTape or by using gel electrophoresis.
39. Sequence the resulting library by using a next-generation sequencing platform (e.g., Illumina NextSeq or MiSeq instruments).
▲ **CRITICAL STEP** The loading concentration of the sequencing library should be empirically determined to minimize over-clustering on the sequencing flow cell. We typically load 1.5 pM for the NextSeq 550 system with 20–30% internal PhiX control.
▲ **CRITICAL STEP** We typically use a single read that covers the entire DNA Tape amplicon from Read1 (e.g., 148 cycles for Read1, 10 cycles for Index1 and 10 cycles for Index2 with the NextSeq 500/550 v2.5 150-cycle kit).

Analysis of sequencing data from recording the order of transfections with DNA Typewriter

● **TIMING 2–4 h**

▲ **CRITICAL** Here, we outline the steps to analyze sequencing data from DNA Typewriter experiments. Refer to the accompanying Python/R scripts for example commands for data analysis, shared through the Figshare (https://doi.org/10.6084/m9.figshare.22728758.v1)[37] platform.

40. Iterate through each of the raw sequencing reads and identify the DNA Tape sequence (e.g., 5xTAPE) and the TapeBC sequence. Use the text pattern matching package REGEX to match the two 10-bp flanking sequences around the DNA Tape and extract out the TapeBC and Tape sequence that should contain any expected insertions. Any reads that contain insertions or deletions of the TAPE 14 bp (resulting in 4xTAPE or 6xTAPE, for example) are removed from the analysis.
41. Within the DNA Tape sequencing read, extract the InsertBCs by using the text pattern matching package (e.g., REGEX) and store the identity of these InsertBCs in a table, in which the rows are a single sequencing read and the columns are the prime editing–mediated insertion events at the five DNA Tape sites.
◆ **TROUBLESHOOTING**
42. From the recording sites within the DNA Tape (e.g., Site-1 to Site-5 within 5xTAPE), count the total number of neighboring bigram sequences. A bigram consists of two InsertBCs from adjacent sites in the 5xTAPE. Count the number of times a bigram appears in Site-1 and Site-2, Site-2 and Site-3, Site-3 and Site-4 and Site-4 and Site-5. If a total of 16 possible

barcodes (InsertBCs) can be inserted into DNA Tape, 16 × 16 = 256 possible bigrams should be counted.

43. On the basis of the bigram count from Step 43, calculate the frequencies of each bigram combination. In this part of the analysis, we recommend disregarding the bigrams consisting of two of the same InsertBCs.
▲ **CRITICAL STEP** The relative order between two InsertBCs is determined by the ratio of their bigram frequencies. For example, if the bigram consisting of InsertBCs 'AACGGA' and 'TAGGGA' appears more frequently than the bigram consisting of InsertBCs 'TAGGGA' and 'AACGGA', then we can infer that the transfection of the epegRNA inserting 'AACGGA' into DNA Tape occurred before the transfection of the epegRNA inserting 'TAGGGA' into DNA Tape.
44. Perform a bubble-sorting algorithm to sort each individual transfection event on the basis of the bigram frequencies in the most parsimonious order. Briefly, the order of 16 events is initialized with their frequency in Site-1 (from highest to lowest, where the highest is probably occurred earlier), and the order of two adjacent events from earliest to latest is switched whenever we have a bigram frequency of the reverse order that is higher than the initialized order. Repeat the iteration through the list until there is no change in orders within a single iteration.

Procedure 2

Generating a lineage-recording HEK293T cell line to be used in scRNA-seq

● **TIMING 23–28 days**

Cloning the DNA Tape and epegRNA lineage-recording plasmid

● **TIMING 2–3 days**

▲ **CRITICAL** This section describes cloning a UMI-labeled pool of lineage-recording plasmids (Lenti-TapeBC-5xTAPE-pegRNA-InsertBC). Each plasmid contains a U6 promoter-driven pegRNA that programs a 6-bp insertion as well as an EF1a promoter-driven UMI-labeled DNA Tape that can be captured in scRNA-seq.

1. As described in Procedure 1, Step 2A(i–vii), set up a restriction digest reaction with the LentiCROP-PuroR-P2A-GFP-MluI-U6-BsmBI lentiviral vector and the enzyme BsmBI-v2-HF as described in the table. Incubate the reaction at 37 °C for ≥3 h or overnight. Separate the digested product by gel electrophoresis and purify the digested plasmid.

Component	Volume (µl)	Final concentration
LentiCROP-PuroR-P2A-GFP-MluI-U6-BsmBI	Volume for 5 µg	–
BsmBI-v2-HF (20,000 U/ml)	5	2,500 U/ml
10× rCutSmart buffer	4	1×
Nuclease-free water	Up to 40	–

2. Amplify pegRNA fragments (all possible NNNGGA DNA sequences) to insert into the BsmBI-digested lentiviral plasmid through isothermal assembly. Refer to Supplementary Table 1 for the pegRNA fragment oligo designs (qPCR-amplify GF003 with OLG005 and OLG006 by following Procedure 1, Step 5).
▲ **CRITICAL STEP** In this version of the lineage-recording plasmid, we use pegRNAs because epegRNAs had not been developed when we designed the protocol. The MluI cut site on the evoPreQ$_1$ pseudoknot of epegRNAs prevents the subsequent steps of the cloning scheme described in this protocol. If additional stability of the pegRNA portion is desired, one may consider adding other versions of the 3′ pseudoknot to the pegRNA that do not contain an MluI cut site[29].

3. Follow Procedure 1, Step 2B(ii–x) to ligate the DNA pool of 64 pegRNA fragments from Step 2 into the purified BsmBI-digested plasmid from Step 1. This step generates a pool of plasmids with different insertion-bearing pegRNAs, although the plasmids still lack the 5xTAPE.
4. Digest the plasmid pool from Step 3 with MluI-HF restriction enzyme to add the 5xTAPE portion by following a similar protocol to the one described in Procedure 1, Step 2A(i–vii). Incubate the reaction at 37 °C for ≥3 h or overnight. Separate the digested product by gel electrophoresis and purify the digested plasmid.

Component	Volume (µl)	Final concentration
LentiCROP-PuroR-P2A-GFP-MluI-U6-pegRNA (Step 2)	Volume for 5 µg	–
MluI-HF (20,000 U/ml)	5	2,500 U/ml
10× rCutSmart buffer	4	1×
Nuclease-free water	Up to 40	–

5. Generate UMI-labeled DNA Tapes by qPCR amplification. See Supplementary Table 1 for the DNA Tape oligo designs (qPCR-amplify GF004 with OLG007 and OLG008 by following Procedure 1, Step 5).
6. Using isothermal assembly, insert the pool of UMI-labeled DNA Tapes from Step 5 into the MluI-digested plasmid pool from Step 4, as detailed in Procedure 1, Step 2B(ii–x). This step generates the final pool of lineage-recording plasmids, in which each plasmid will contain a UMI-labeled DNA Tape and a pegRNA that specifies a 6-bp insertion (NNNGGA) into the DNA Tape.

Stable integration of a doxycycline-inducible prime editor cassette in HEK293T cells

● **TIMING 16–20 d**

7. Thaw a stock vial of HEK293T cells at passage 5 or lower and culture the cells as described in Box 1, before conducting the recording experiment.
8. 24 h before transfection, plate HEK293T cells in one well of a 24-WP such that the cells will reach 60–80% confluency in 24 h.
9. When the cells are at 60–80% confluency, transfect the piggyBAC-iPE2 plasmid for doxycycline-inducible PE2 expression, the piggyBAC-rtTA plasmid carrying the rtTA expression gene cassettes and a piggyBAC transposase (e.g., HyPBase) at a mass ratio of 4:4:2 into HEK293T cells with Lipofectamine 3000 by following the manufacturer's protocol. This transfection will generate a stable iPE2(+) HEK293T cell line with doxycycline-inducible PE2.
   ▲ **CRITICAL STEP** Prime editor cassettes can be replaced with the most recent version of PE, such as PE4max, for more efficient prime editing. However, our cloning scheme was designed before the development of more advanced, highly engineered prime editors.
10. Incubate the transfected cells at 37 °C and 5% CO_2. 24 h after transfection, remove and discard the transfection medium and replace it with drug-selection medium.
11. In a 96-WP, plate the HEK293T cells selected for stable integration of the inducible prime-editing construct (from Step 10) such that there are expected to be 0.1 cells per well.
12. Monitor the plate daily and make note of the wells that contain a population of HEK293T cells resulting from the expansion of a single cell.
13. Pick multiple wells of monoclonal iPE2(+) HEK293T cell lines and transfer the cells in each well to a new well in a 24-WP. Continue expanding each monoclonal line until there are enough cells for freezing as well as the subsequent transduction.
    ▲ **CRITICAL STEP** This HEK293T cell line will have doxycycline-inducible prime-editing activity; however, at this stage, the cell line is not yet usable for lineage recording, because it lacks stably integrated DNA Tape and epegRNAs.
    ■ **PAUSE POINT** iPE2(+) HEK293T cell lines can be frozen in a freezing medium (10% (vol/vol) dimethyl sulfoxide in FBS) and used for future applications.

14. (Optional) After a stable, monoclonal cell culture has been established, we recommend testing prime-editing activity with or without the addition of 10 ng/ul doxycycline to the culture medium to ensure inducible and non-leaky prime-editing activity in the iPE2(+) HEK293T cell line. To test the prime-editing efficiency by using the doxycycline-induced prime editor, a plasmid with an epegRNA programming an insertion at a native locus in HEK293T cells can be transfected, and the locus can be sequenced for prime-editing insertion analysis by following a similar procedure to the published protocol for prime-editing experiments in HEK293T cells[38].

Preparation of lentivirus

● **TIMING** 3 days

▲ **CAUTION** Prepare lentiviruses by following institutional precautions and practices associated with Biosafety Level 2 (BSL2).

15. Plate at least two wells of HEK293T cells in a 6-WP in lentivirus packaging medium such that the cells will be ~90% confluent in 20 h. Incubate the cells at 37 °C and 5% CO_2.
16. Once the cells reach 90% confluency, transfect lentiviral production plasmids (ViraPower packaging mix and the Lenti-TargetBC-5XTAPE-pegRNA-InsertBC plasmid from Step 4) to the HEK293T cells by following the ThermoFisher Scientific protocol 'Improve lentiviral production using Lipofectamine 3000 reagent'. Incubate the cells at 37 °C and 5% CO_2.
17. 6 h after transfection, remove and replace the medium from each well with 2 ml of fresh lentivirus packaging medium. Incubate the cells at 37 °C and 5% CO_2.
    ▲ **CAUTION** After transfection, the medium will contain infectious lentiviral particles. BSL2 practices and all institutional and regulatory guidelines must be followed for proper handling of lentivirus. Take caution to avoid contact with lentiviral particles and ensure proper disposal of biohazardous waste that has had contact with lentivirus.
18. From each well, collect the supernatant containing viral particles at two time points, 24 h and 52 h, by following the ThermoFisher Scientific protocol 'Improve lentiviral production using Lipofectamine 3000 reagent'. Store the virus at 4 °C until the second time point.
19. Combine the lentiviral supernatant from the two time points and concentrate it by using SBI PEG-*it* virus precipitation solution and following the manufacturer's instructions or by using an alternate lentivirus concentration method.
    ■ **PAUSE POINT** Virus can be frozen at −80 °C for several months to use at a later time point, but generally with reduced titer. Divide the virus into aliquots of working volumes of 50 µl to avoid multiple freeze-thaw cycles.

Transduction of iPE2(+) HEK293T monoclonal cells

● **TIMING** 2 days

▲ **CAUTION** Handling of lentivirus particles and transduction require following institutional precautions and practices associated with BSL2.

▲ **CRITICAL** In this part of the procedure, we describe the transduction protocol to generate HEK293T cells expressing all the components (i.e., prime editor, epegRNA cassettes and DNA Tape cassettes) necessary for lineage recording given the induction of prime editor transcription.

20. Seed the iPE2(+) HEK293T monoclonal cell line from Step 7 in ≥10 wells of a 96-WP in HEK293T cell culturing medium, such that the cells reach 60–80% confluency in 24 h.
    ▲ **CRITICAL STEP** Plate at different densities to ensure at least one well with 60–80% confluency the next day for lentiviral transduction.
21. Once the cells have reached 60–80% confluency, transduce iPE2(+) HEK293T cells with 5xTAPE/epegRNA lentivirus from Step 17 by using Magnetofection reagents according to the manufacturer's instructions.
    ▲ **CRITICAL STEP** We recommend aiming for a multiplicity of infection (MOI) of 20–30 to ensure enough recording sites from the DNA Tape portion of the integrated lineage-recording lentiviral construct. In addition, the set of pegRNA-expression cassettes found in each cell must include enough unique pegRNAs, because each pegRNA drives a

unique type of edit in the DNA Tape. Having enough diversity of edits to the DNA Tape will facilitate the lineage-reconstruction analysis. Magnetofection reagents may aid in reaching high MOI in the case of suboptimal viral preparation.

Recording single-cell lineage information in HEK293T cells

● **TIMING 21–28 days**

Induction of lineage recording during the expansion of a single cell

● **TIMING 21–28 days**

▲ **CRITICAL** We recommend carefully following the procedure described here and using the cell culture dilutions indicated to generate a monoclonal culture for the easiest reconstruction of lineage information (1 cell per well of a 96-WP). Lineage information can be reconstructed in a polyclonal culture, but multiple lineage trees would be generated (one per each starting cell).

22. 16–24 h after transduction (Steps 20 and21), dissociate the transduced cells by following the trypsinization procedure detailed in Box 1.
23. Count the cells with a cell counter to estimate the number of live cells per milliliter.
24. Dilute the cells at three different concentrations in 11 ml of HEK293T cell culturing medium: 2 cells/100 µl, 0.2 cells/100 µl and 0.02 cells/100 µl.
25. Add doxycycline at a concentration of 10 ng/µl to each cell dilution from Step 24 to induce the expression of prime editors and start lineage recording in the DNA Typewriter system.
26. Prepare a 96-WP per concentration condition by seeding 100 µl of each cell suspension per well, resulting in 2, 0.2 or 0.02 cells per well, respectively. Incubate the cells at 37 °C and 5% CO_2.
27. After 4–5 days, ensure that cells have adhered to the bottom of the plate. Carefully remove and discard the culture medium and replace it with fresh HEK293T cell culturing medium supplemented with 10 ng/µl doxycycline. Incubate the cells at 37 °C and 5% CO_2.
28. Throughout the next 20 days, monitor the cell cultures under a fluorescence microscope in a 96-WP and identify wells with a single colony with consistent GFP fluorescence levels throughout the colony.

    ▲ **CRITICAL STEP** Ensure that the wells chosen contain only a single colony that came from the expansion of a single cell.
29. For each well containing a single colony, dissociate the colony and plate it into a single well of a 24-WP. Incubate the cells at 37 °C and 5% CO_2 and grow them until they reach 60–80% confluency. Replace the medium in each well every other day with fresh HEK293T cell medium supplemented with 10 ug/ml doxycycline.
30. Once the cells have reached 60–80% confluency, dissociate and plate them into a single well of a 6-WP. Continue changing the medium in the wells every other day with HEK293T cell medium supplemented with doxycycline until the cells reach 80% confluency.

Choosing a monoclonal cell line for scRNA-seq profiling

● **TIMING 2–4 days**

▲ **CRITICAL** The monoclonal lines from Step 30 can be assayed for the MOI of the DNA Tape and prime-editing efficiency once they have reached 80% confluency. Those monoclonal lines showing sufficient editing of the DNA Tape will be used for scRNA-seq experiments to capture both the cellular transcriptome and molecules of TapeBC-5xTAPE with recorded lineage information.

31. Dissociate the cells from Step 30 by adding 500 µl of trypsin to each well. Incubate at 37 °C for 5 min and quench the trypsinization reaction with ≥3 volumes of pre-warmed HEK293T cell medium.
32. Collect 50% of the cell suspension from Step 31 in a 1.5-ml Eppendorf tube and centrifuge for 3 min at 500*g* at room temperature. Freeze down the cells for long-term storage in liquid nitrogen by following the freezing procedure detailed in Box 1, step 13.
33. Centrifuge the remaining 50% of the cell suspension from Step 30 at 500*g* for 3 min for gDNA extraction to prepare sequencing libraries (Procedure 1, Steps 26–40). Calculate the number of unique DNA Tape integrations in each monoclonal line as well as the percentage

of DNA Tapes that have a successful prime-editing event at the first TAPE site by following Procedure 1, Steps 41 and 42. Select the monoclonal line with ~20–30 unique DNA Tape integrations and ≥90% editing rate at the first site of the DNA Tape for the preparation of the scRNA-seq library.

Preparation of the scRNA-seq library

● **TIMING 1–4 days**

34. Once the monoclonal line for the scRNA-seq lineage-tracing experiment has been selected (Step 33), thaw the cells (Step 32) and seed them into a six-WP in HEK293T cell medium.
▲ **CRITICAL STEP** Ensure that doxycycline is not included in the culture medium so that the prime editing and further recording of lineage information are disabled.
35. Incubate the cells in 37 °C and 5% CO_2 and monitor them for adherence to the plate and for growth. Once the cells have reached 80% confluency, harvest the cells by adding 500 µl of trypsin to the well and incubating for 5 min at 37 °C.
36. Quench the trypsinization reaction by adding ≥3 volumes of HEK293T cell medium. Collect the cells in a 1.5-ml Eppendorf tube and centrifuge at 500*g* for 3 min. Remove and discard the supernatant and wash the pellet by resuspending the cells in 10 ml of HEK293T cell medium.
37. Using the cells harvested in Step 36, prepare cDNA for library construction according to the Chromium Next GEM single-cell 3′ reagent kit v3.1 with feature barcoding technology for CRISPR screening (manual part no. CG000205 Rev D) protocol from 10X Genomics Step2.3A.x.
38. To ensure the recovery of unique TapeBC-5xTAPE molecules, collect cDNA by following Step 2.3B.xiv of the 10X Genomics Chromium protocol.
39. Use the cDNA samples from both Steps 37 and 38 (TapeBC-5xTAPE cDNA molecules are present in both of the cDNA products) to set up a qPCR reaction as indicated in the first table. Run the qPCR reaction according to the amplification profile in the second table:

Component	Volume (µl)	Final concentration
cDNA from 10X Genomics Chromium protocol, Step 2.3A.x (from Step 37)	5	–
cDNA from 10X Genomics Chromium protocol, Step 2.3B.xiv (from Step 38)	5	–
OLG013 (100 µM)	0.2	0.4 µM
OLG014 (100 µM)	0.2	0.4 µM
KAPA Robust 2× Hotstart mix	25	1×
SYBR Green (100×)	0.2	0.4×
Water	Up to 50	–

Step	Temperature (°C)	Time	Cycling instructions
1	95	2 min	–
2	95	15 s	–
3	65	15 s	–
4	72 (with the plate read at the end)	90 s	Go back to step 2; repeat 15–17 cycles
5	72	3 min	–
6	10	Hold	–

▲ **CRITICAL STEP** We recommend keeping the number of PCR cycles around 15 and certainly <17 in total, to avoid the production of possible PCR byproducts such as chimeric cDNA molecules. Monitor the SYBR Green curves and stop the cycling reaction before the curves reach their inflection point.

40. Purify the pool of DNA Tapes with single-cell barcodes by using AMPure XP beads and following the manufacturer's instructions or by using a preferred method of DNA purification. Elute the purified DNA in 50 µl of water.
41. To add P5 and P7 sequencing adapters and indices to the library, set up a qPCR reaction in a PCR strip tube according to the amplification profile indicated in the second table. 5–8 PCR cycles should be sufficient to convert most of the DNA fragments into ones with Illumina sequencing adapters.

Component	Volume (µl)	Final concentration
AMPure XP bead-purified PCR product from Step 40	0.5–1	–
10 µM F OLG011	1	0.4 µM
10 µM R OLG012	1	0.4 µM
SYBR Green (100×)	0.1	0.4×
KAPA2G Robust 2× Master Mix	12.5	1×
Water	Up to 25	–

Step	Temperature (°C)	Time	Cycling instructions
1	95	2 min	–
2	95	15 s	–
3	65	15 s	–
4	72 (with the plate read at the end)	90 s	Go back to step 2; repeat 7 cycles
5	72	3 min	–
6	10	Hold	–

42. Purify the qPCR reaction with AMPure XP beads by following the manufacturer's protocol and using a 0.7–0.8× bead clean-up for 5xTAPE with 10X Genomics single-cell barcodes. If not using 5xTAPE, choose the appropriate AMPure XP bead ratio for the length of the DNA Tape amplicon.
43. Check for the correct size of the resulting purified product by running 1 µl of the product on the TapeStation with a D1000 ScreenTape or by using gel electrophoresis.
44. Sequence the resulting library by using a next-generation sequencing platform (e.g., Illumina NextSeq or MiSeq instruments).
▲ **CRITICAL STEP** The loading concentration of the sequencing library should be empirically determined to minimize overclustering on the sequencing flow cell. We typically load 1.5 pM for the NextSeq 550 system with 20–30% internal PhiX control.
▲ **CRITICAL STEP** We typically use a single read that covers the entire DNA Tape amplicon from Read2 and use Read1 for reading through the 10X Genomics CellBC and UMI (e.g., 28 cycles for Read1 for reading the CellBC and UMI added during cDNA synthesis, 10 cycles for Index1, 10 cycles for Index2 and 200 cycles for Read2 to read through the DNA Tape). We typically use the NextSeq 500/550 v2.5 300-cycle kit.

Analysis of sequencing data from single-cell lineage-recording experiments with DNA Typewriter

● **TIMING 2–4 h**

▲ **CRITICAL** Here, we outline the steps to analyze sequencing data from DNA Typewriter experiments. Refer to the accompanying Python/*R* scripts for example commands for data analysis, shared through the Figshare (https://doi.org/10.6084/m9.figshare.22728758.v1)[37] platform.

Generation of a phylogenetic tree from scRNA-seq data

● **TIMING 2–4 h**

45. Create a dataframe summarizing the sequencing data from Step 44. Each row in the table represents an individual single-cell barcoded DNA Tape read. The columns are as follows: 16-bp CellBC (from the 10X Genomics platform), 8-bp TapeBC and InsertBC sequences found in the DNA Tape (five InsertBCs in the case of 5xTAPE). Perform string processing to extract these values from each read, because these values should be located within the read in expected locations.
46. Quantify the number of unique DNA Tapes captured from each cell by grouping the table of sequences on the basis of the 16-bp CellBC. The number of unique TapeBCs per grouping indicates the number of captured DNA Tapes per cell.
47. Create an allowlist of TapeBCs for the analysis, filtering out the DNA Tapes that are not recovered in a large proportion of single cells. The cells that will be used for the lineage

reconstruction will contain every one of those allowed DNA Tapes. The allowed DNA Tapes should also contain diverse InsertBCs, indicating that lineage information was recorded at that DNA Tape locus.

▲ **CRITICAL STEP** Choosing a smaller set of allowed TapeBCs will result in fewer cells being filtered out downstream. However, the lineage information may not be sufficient to fully reconstruct the lineage tree. We recommend selecting at least eight TapeBCs with diverse InsertBCs and restricting the analysis to cells in which all eight TapeBCs are recovered.

◆ **TROUBLESHOOTING**

48. Filter for the single cells that contain all the TapeBCs in the allowlist. The rest of the cells can be removed from the analysis.
49. Create a cell-by-cell matrix of lineage distances with the following algorithm. For each cell pair within the remaining cells from Step 47, count the number of shared sequential edits across all 5xTAPEs and subtract this number from the total number of recording sites (five times the number of 5xTAPEs used).
50. The resulting lineage distances between all cell pairs can be used to generate the lineage tree by using available packages for phylogenetics (e.g., *phangorn* and *ape* packages in *R*).

Troubleshooting

Troubleshooting advice can be found in Table 1. For more detailed troubleshooting guidelines on how to design and execute prime-editing experiments in mammalian cells, see ref. 38.

Table 1 | Troubleshooting table

Step	Problem	Possible reason	Solution
Procedure 1			
2A(xii)	Low efficiency of pegRNA cloning	Background with original plasmid is high (RFP+ colonies)	Perform longer Golden Gate assembly thermocycler protocol, cycling between 16 °C (optimal for T4 DNA ligase) and 37 °C (optimal for BsaI-HF-v2) ≥50 times. The latter step should cut any re-annealed plasmids with the original insert. Taking care to fully digest the plasmid and excise only the digested fragment will help reduce background
14	Truncation or expansion of the TAPE repetitive sequence in the cloned plasmid library	Too many PCR cycles used to convert the synthesized oligomer with TAPE repeats into double-stranded DNA for insertion into the plasmid backbone	Use fewer PCR cycles by directly monitoring the amplification process with a qPCR instrument or order a gene fragment to directly insert into the plasmid backbone
		Recombination of the TAPE repetitive sequence in the plasmid during bacterial culture	Be careful not to overgrow the bacterial culture. Use recombination-deficient competent cells for transformation and grow the culture at 30 °C to minimize recombination
18	Low MOI when generating a stable TAPE cell line	No selection before isolating clones	A drug resistance gene can be incorporated into the TAPE construct as a selectable marker to identify clones with successful TAPE integration
		No sorting of high-expressing clones	Include the TAPE design as part of the 3′ UTR of a fluorescent protein. Sort the top 1–5% of the population with the highest fluorescence as a proxy for high TAPE expression
		Non-optimal transduction conditions	Concentrate virus after harvest and/or boost the infection rate with the Magnetofection system
24	The HEK293T recorder cell line reaches confluency in <3 days between the sequential transfections	Too many cells were passaged from the previous transfection event	Allow an interval of 3 days between the sequential transfections so that insertions mediated by prime editing can accumulate on the DNA Tape before the next round of transfection. Seed a lower number of cells (e.g., plate only 5% or 8% of the cells in each well into a new set of wells after the initial transfection)
41	Low efficiency of prime editing	Low transfection efficiency (i.e., low percentage of cells receiving all the components for DNA Typewriter)	Closely follow the manufacturer's protocol for transfection. Different transfection methods may need to be explored to find the optimal one
		Cells have been harvested too soon for analysis after induction of prime editing	The prime-editing efficiency with transiently transfected components usually reaches maximum editing at ~3 days in HEK293T cells. Other cell lines and methods of transgene delivery may differ in the amount of time needed
		Non-optimal pegRNA design	See refs. 31,33 for pegRNA design principles

Table 1 (continued) | Troubleshooting table

Step	Problem	Possible reason	Solution
Procedure 2			
47	Low capture rate of the DNA Tape in scRNA-seq	Relying on just oligo-dT capture of the TAPE transcript	Addition of 10× CS1 and CS2 sequences generally will lead to more capture of TAPE transcripts in 10X Genomics scRNA-seq platforms
		Cell line–specific silencing of the DNA Tape	High MOI of the DNA Tape can make up for silencing of some integrants. Waiting too long to use the cell line for experiments after selection and monoclonal isolation may coincide with silencing of TAPE expression

Timing

Procedure 1

Steps 1–2, cloning of DNA Tape–targeting epegRNA expression cassettes: 3–4 days
Steps 3–14, cloning a pool of UMI-labeled DNA Tape expression cassettes: 3–4 days
Steps 15–18, generating a DNA Typewriter recorder cell line: 10–12 days
Steps 19–24, sequential transfections with the DNA Tape recorder cell line: 3–51 days
Steps 25–28, collection of the DNA Tape from cells: cell lysis and gDNA harvest: 2–3 h
Steps 29–39, preparation of the Illumina sequencing library from gDNA samples: 4–6 h
Steps 40–44, analysis of sequencing data from recording the order of transfections with DNA Typewriter: 2–4 h

Procedure 2

Steps 1–6, cloning the DNA Tape and epegRNA lineage-recording plasmid: 2–3 days
Steps 7–14, stable integration of the doxycycline-inducible prime editor cassette in HEK293T cells: 16–20 days
Steps 15–19, preparation of the lentivirus: 3 days
Steps 20–21, transduction of the cells: 2 days
Steps 22–30, induction of lineage recording during the expansion of a single cell: 21–28 days
Steps 31–33, choosing the monoclonal cell line for scRNA-seq profiling: 2–4 days
Steps 34–44, preparation of the scRNA-seq library: 1–4 days
Steps 45–50, analysis of sequencing data from the single-cell lineage-recording experiment with DNA Typewriter: 2–4 h

Boxes

Box 1, HEK293T cell culture and passaging: 1 h
Box 2, design and order of oligo splints: 1–4 h

Anticipated results

In recording the order of transfection events by using DNA Typewriter, we anticipate that each round of editing will induce ≥5–10% editing combined across different sites of the DNA Tape. To assess the editing efficiency of the prepared reagents, a small portion of cell material can be collected after ≥1 day from the first round of recording (Procedure 1, Step 23) and subjected to the protocol described to evaluate the presence of any anticipated editing (Procedure 1, Steps 26–45). Next-generation sequencing is ideal to accurately quantify editing efficiency, although Sanger sequencing can be used to check the editing efficiency qualitatively by looking at the frequency of mixed bases at the insertion site. If new DNA Tape sequences are being tested, one could also check possible non-sequential editing within the DNA Tape, where later sites are edited while earlier sites remain unedited. If the editing efficiency seems

lower than expected, a positive control condition that targets a known native genomic locus for editing could help troubleshoot which reagent is not working as intended. After multiple rounds of recording, one could extract insertion sequences from DNA Tape sequencing reads and calculate bigram frequencies (Procedure 1, Steps 41–45) to infer the order of sequential transfection events. We anticipate that similar results can be achieved in cell culture systems that show reasonable prime-editing efficiency and are amenable to multiple rounds of transfection.

In addition, we expect the capture of DNA Tapes to integrate well with scRNA-seq platforms such as 10X Genomics, such that single-cell transcriptional profiles are simultaneously captured with lineage information. Like the experiments for recording transfection orders, achieving adequate editing efficiency is key to inferring high-resolution lineage information. In general, the editing efficiency should be high enough to record two to three edits per cell division. If, for instance, there are 20 active editing sites across 20 unique integrations of DNA Tape constructs, the rate of editing efficiency should equate to 10–15% editing per cell division. A higher number of DNA Tape integrations or a slower cell division rate could help overcome lower editing efficiency by providing more targets and time for edits to accumulate before each cell division.

In our demonstration, we generated a lineage tree for a monoclonal population of HEK293T cells that share the same ancestor cell at the start of the lineage recording. The quality of lineage recording data can be initially assessed by inspecting the resulting tree and associated edits. When lineage recording is successful, one would expect to see the same multiple edits shared in specific clades of cells. In addition, one can use bootstrapping analysis to assess the quality of each branching point within the tree. Our single-cell lineage dataset detected 19 DNA Tapes and used 16 for lineage reconstruction. During bootstrapping implemented with the APE R package[39], we resampled different DNA Tapes from the 19 detected, reconstructed a tree each time, and quantified the number of the same branching events observed in the tree reconstructed using the original set of 16 DNA Tapes. Thus, the robustness of the resulting lineage tree can be assessed by bootstrapping analysis.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Example data files have been deposited to Figshare (https://doi.org/10.6084/m9.figshare.22728758.v1)[37]. Plasmids to clone in pegRNAs and DNA Tape have been deposited to Addgene (cat. nos. 200029 for lentiviral backbone construct and 200030 for piggyBAC backbone construct).

Code availability

Along with the example data, example scripts that can be directly used to analyze example data files have been deposited to Figshare (https://doi.org/10.6084/m9.figshare.22728758.v1)[37]. For recording the order of transfection, we have included an example script (TAPE_text_sorting.ipynb) to determine the order of transfected barcodes from paired-end sequencing data. For generating a single-cell lineage tree, we run a first custom script (TAPE_10X_read2fromBAM.ipynb) to extract the single-cell barcode information along with edits in captured DNA Tape molecules. The output can then be used in the second custom script (DNATypewriter_SingleCellLineage_Rscript.ipynb) to plot the lineage tree on the basis of the shared patterns of edited DNA Tapes from single cells.

Received: 9 May 2023; Accepted: 15 March 2024;
Published online: 06 June 2024

References

1. Sheth, R. U. & Wang, H. H. DNA-based memory devices for recording cellular events. *Nat. Rev. Genet.* **19**, 718–732 (2018).
2. Farzadfard, F. & Lu, T. K. Emerging applications for DNA writers and molecular recorders. *Science* **361**, 870–875 (2018).
3. Lear, S. K. & Shipman, S. L. Molecular recording: transcriptional data collection into the genome. *Curr. Opin. Biotechnol.* **79**, 102855 (2023).
4. 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).
5. Tang, W. & Liu, D. R. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**, eaap8992 (2018).
6. Farzadfard, F. et al. Single-nucleotide-resolution computing and memory in living cells. *Mol. Cell* **75**, 769–780.e4 (2019).
7. Bhattarai-Kline, S. et al. Recording gene expression order in DNA by CRISPR addition of retron barcodes. *Nature* **608**, 217–225 (2022).
8. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016).
9. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* **547**, 345–349 (2017).
10. Schmidt, F., Cherepkova, M. Y. & Platt, R. J. Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **562**, 380–385 (2018).
11. Schmidt, F. et al. Noninvasive assessment of gut function using transcriptional recording sentinel cells. *Science* **376**, eabm6038 (2022).
12. Choi, J. et al. A time-resolved, multi-symbol molecular recorder via sequential genome editing. *Nature* **608**, 98–107 (2022).
13. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
14. Chen, P. J. & Liu, D. R. Prime editing for precise and highly versatile genome manipulation. *Nat. Rev. Genet.* **24**, 161–177 (2023).
15. Rodriguez-Fraticelli, A. & Morris, S. A. In preprints: the fast-paced field of single-cell lineage tracing. *Development* **149**, dev200877 (2022).
16. Sankaran, V. G., Weissman, J. S. & Zon, L. I. Cellular barcoding to decipher clonal dynamics in disease. *Science* **378**, eabm5874 (2022).
17. Jones, M. G., Yang, D. & Weissman, J. S. New tools for lineage tracing in cancer in vivo. *Annu. Rev. Cancer Biol.* **7**, e32888 (2023).
18. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
19. Chan, M. M. et al. Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).
20. Bowling, S. et al. An engineered CRISPR-Cas9 mouse line for simultaneous readout of lineage histories and gene expression profiles in single cells. *Cell* **181**, 1693–1694 (2020).
21. Quinn, J. J. et al. Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. *Science* **371**, eabc1944 (2021).
22. Simeonov, K. P. et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell* **39**, 1150–1162 (2021).
23. Yang, D. et al. Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell* **185**, 1905–1923.e25 (2022).
24. Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652.e29 (2021).
25. Lin, Q. et al. Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**, 582–585 (2020).
26. Bosch, J. A., Birchak, G. & Perrimon, N. Precise genome engineering in *Drosophila* using prime editing. *Proc. Natl Acad. Sci. Usa.* **118**, e2021996118 (2021).
27. Tong, Y., Jørgensen, T. S., Whitford, C. M., Weber, T. & Lee, S. Y. A versatile genetic engineering toolkit for *E. coli* based on CRISPR-prime editing. *Nat. Commun.* **12**, 5206 (2021).
28. Chen, W. et al. Multiplex genomic recording of enhancer and signal transduction activity in mammalian cells. Preprint at *bioRxiv* https://doi.org/10.1101/2021.11.05.467434 (2021).
29. Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2021).
30. Ferreira da Silva, J. et al. Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. *Nat. Commun.* **13**, 760 (2022).
31. Koeppel, J. et al. Prediction of prime editing insertion efficiencies using sequence features and DNA repair determinants. *Nat. Biotechnol.* **41**, 1446–1456 (2023).
32. Li, M. et al. Transient inhibition of p53 enhances prime editing and cytosine base-editing efficiencies in human pluripotent stem cells. *Nat. Commun.* **13**, 6354 (2022).
33. Kalhor, R. et al. Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804 (2018).
34. Loveless, T. B. et al. Molecular recording of sequential cellular events into DNA. Preprint at *bioRxiv* https://doi.org/10.1101/2021.11.05.467507 (2021).
35. Petri, K. et al. CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat. Biotechnol.* **40**, 189–193 (2022).
36. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
37. Liao, H. & Choi, J. DNA Typewriter scripts. *Figshare* https://doi.org/10.6084/m9.figshare.22728758.v1 (2023).
38. Doman, J. L., Sousa, A. A., Randolph, P. B., Chen, P. J. & Liu, D. R. Designing and executing prime editing experiments in mammalian cells. *Nat. Protoc.* **17**, 2431–2468 (2022).
39. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).

Acknowledgements
We thank R. Daza, B. Martin, H. Kim and J. Nathans for feedback on this manuscript and protocol. We thank members of Jay Shendure's laboratory for developing the DNA Typewriter platform as well as David Liu's laboratory for developing and improving the prime-editing technology. This work is supported by a grant from the Paul G. Allen Frontiers Group (Allen Discovery Center for Cell Lineage Tracing to J.S.) and the National Human Genome Research Institute (UM1HG011586 to J.S. and K99HG012973 to J.C.). H.L. is supported by the NSF Graduate Research Fellowship Program (DGE-2140004). J.C. was a Howard Hughes Medical Institute Fellow of the Damon Runyon Cancer Research Foundation (DRG-2403-20). J.S. is an investigator of the Howard Hughes Medical Institute.

Author contributions
This protocol is based on a paper by J.C. and J.S. H.L. and J.C. contributed equally and wrote the manuscript and prepared the figures. J.S. supervised the research and wrote parts of the manuscript. All authors edited the manuscript.

Competing interests
The University of Washington has filed a patent application partially based on this work, in which J.C. and J.S. are listed as inventors. J.S. is on the scientific advisory board, a consultant, and/or a co-founder of Prime Medicine, Cajal Neuroscience, Guardant Health, Maze Therapeutics, Camp4 Therapeutics, Phase Genomics, Adaptive Biotechnologies, Scale Biosciences, Sixth Street Capital and Pacific Biosciences. H.L. declares no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-024-01003-0.

Correspondence and requests for materials should be addressed to Junhong Choi or Jay Shendure.

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

Related links
Key reference using this protocol
Choi, J. et al. *Nature* **608**, 98–107 (2022): https://doi.org/10.1038/s41586-022-04922-8

nature portfolio

Corresponding author(s): Jay Shendure, Junhong Choi

Last updated by author(s): 2024-01-12

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted
  Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No data was collected for this protocols manuscript.
Data analysis	Custom Python analysis scripts are included with the manuscript as examples for users wishing to analyze data generated from following the

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Example data files are deposited to Figshare (https://doi.org/10.6084/m9.figshare.22728758.v1). Plasmids to clone in pegRNAs and DNA Tape are deposited to Addgene (Catalog #200029 for lentiviral backbone construct and #200030 for piggyBAC backbone construct).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	This study did not involve human research participants.
Population characteristics	This study did not involve human research participants.
Recruitment	This study did not involve human research participants.
Ethics oversight	This study did not involve human research participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

◉ Life sciences ○ Behavioural & social sciences ○ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. We did not do any additional experiments for this protocols paper, as we only provided step-by-step
Data exclusions	No data was excluded. We did not do any additional experiments for this protocols paper, as we only provided step-by-step instructions on how
Replication	We did not do any additional experiments for this protocols paper, as we only provided step-by-step instructions on how to perform the
Randomization	Randomization was not relevant to our study, as we only provided step-by-step instructions on how to perform the experiment and did not
Blinding	Blinding was not relevant to our study, as we only provided step-by-step instructions on how to perform the experiment and did not provide any

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing	
Data exclusions	
Non-participation	
Randomization	

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing and spatial scale	
Data exclusions	
Reproducibility	
Randomization	

Blinding

Did the study involve field work? Yes No

Field work, collection and transport

Field conditions

Location

Access & import/export

Disturbance

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
	Antibodies
	Eukaryotic cell lines
	Palaeontology and archaeology
	Animals and other organisms
	Clinical data
	Dual use research of concern

Methods

n/a	Involved in the study
	ChIP-seq
	Flow cytometry
	MRI-based neuroimaging

Antibodies

Antibodies used

Validation

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s): HEK293T sourced from ATCC

Authentication: The cell line was used as received from the vendor.

Mycoplasma contamination: Cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register)

Palaeontology and Archaeology

Specimen provenance

Specimen deposition

Dating methods

Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals

Wild animals

Reporting on sex

Field-collected samples

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration

Study protocol

Data collection

Outcomes

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
○	○	Public health
○	○	National security
○	○	Crops and/or livestock
○	○	Ecosystems
○	○	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
○	○	Demonstrate how to render a vaccine ineffective
○	○	Confer resistance to therapeutically useful antibiotics or antiviral agents
○	○	Enhance the virulence of a pathogen or render a nonpathogen virulent
○	○	Increase transmissibility of a pathogen
○	○	Alter the host range of a pathogen
○	○	Enable evasion of diagnostic/detection modalities
○	○	Enable the weaponization of a biological agent or toxin
○	○	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

Files in database submission

Genome browser session

(e.g. UCSC)

Methodology

Replicates

Sequencing depth

Antibodies

Peak calling parameters

Data quality

Software

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Instrument

Software

Cell population abundance

Gating strategy

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Design specifications

Behavioral performance measures

Acquisition

Imaging type(s)

Field strength

Sequence & imaging parameters

Area of acquisition

Diffusion MRI ○ Used ○ Not used

Preprocessing

Preprocessing software

Normalization

Normalization template

Noise and artifact removal

Volume censoring

Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis: ○ Whole brain ○ ROI-based ○ Both

Statistic type for inference (See Eklund et al. 2016)

Correction

Models & analysis

n/a | Involved in the study

- Functional and/or effective connectivity
- Graph analysis
- Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis