Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

Cell Stem Cell

CellPress
OPEN ACCESS

Review

Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants

Zhixin Kang,[1,8] Hui Chen,[2,8] Siyang Li,[3,8] Shou-Wen Wang,[3,4,5,6,*] and Bin Zhou[1,2,7,*]
[1]Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, China
[2]CAS CEMCS-CUHK Joint Laboratory, New Cornerstone Science Laboratory, Key Laboratory of Multi-Cell Systems, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai, China
[3]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou 310024, Zhejiang, China
[4]School of Life Sciences, Westlake University, Hangzhou 310024, Zhejiang, China
[5]School of Science, Westlake University, Hangzhou 310024, Zhejiang, China
[6]Center for Interdisciplinary Studies, Westlake University, Hangzhou 310024, Zhejiang, China
[7]School of Life Science and Technology, ShanghaiTech University, Shanghai, China
[8]These authors contributed equally
*Correspondence: wangshouwen@westlake.edu.cn (S.-W.W.), zhoubin@sibs.ac.cn (B.Z.)
https://doi.org/10.1016/j.stem.2026.05.001

SUMMARY

Elucidating cell fate decision-making requires linking lineage history to dynamic phenotypic states. Driven by single-cell sequencing and genome engineering, lineage tracing has evolved from observational studies into a multidimensional, high-throughput discipline. Here, we synthesize its three methodological pillars: prospective tracking via genetic markers, high-throughput mapping using synthetic barcodes, and retrospective tracing leveraging endogenous natural variants. We survey their integration with multi-omics and spatial profiling, alongside computational approaches to decode cell fates from lineage data. By detailing each approach's trade-offs, we offer a systematic guide for experimental design and highlight emerging frontiers for translating precision clonal analysis into the clinic.

INTRODUCTION

Cell fate dictates the developmental trajectory a cell follows—from its origin to its terminal identity—under physiological and pathological conditions. Mapping these fates is essential for understanding tissue development, homeostasis, and repair, as well as how aberrant lineage decisions drive disease.[1,2] Cell fate choice is often conceptualized using Waddington's landscape, where a cell rolls down a branched, contoured hill, progressively restricting its developmental potential until it settles in a terminal valley.[3,4] Translating this metaphor into a quantifiable framework requires resolving five core elements: (1) cell state—the dynamic molecular identity of a cell, (2) fate paths—the trajectories and branching histories connecting states, (3) functional outputs—the specialized roles of terminal populations, (4) fate-driving forces—the intrinsic networks and extrinsic cues that dictate lineage choices, and (5) fate manipulation—the targeted redirection of cell fate through perturbation. Together, these five elements provide a framework for assessing progress toward a complete understanding of cell fate (Figure 1A).

The quest to map these elements began around 1980 with Sulston's direct visual tracking of the complete *C. elegans* lineage.[9] To overcome the impossibility of direct observation in mammalian tissues, Cepko and colleagues introduced genetically engineered retroviruses.[10] Using *lacZ* reporters, they mapped the multipotent clonal architecture of the developing rodent brain and retina.[11] Building on these milestones, advances over the past decade have propelled the field from low-resolution tracking to systematic, high-dimensional fate mapping, enabling comprehensive decoding of the five core elements. This evolution has been driven by three methodological pillars (Figures 1B–1D).

First, prospective tracing with genetic markers remains an indispensable workhorse (Figure 1B). By using site-specific recombinases (e.g., Cre, Flp, and Dre) to activate optical reporters, defined cell populations can be labeled, tracked, and functionally interrogated *in vivo*. With further evolution through the integration of dual-recombinase logic and multicolor clonal analysis, this classic approach continues to provide foundational insights into developmental biology.

Second, to meet the demand for massive throughput and single-cell resolution, synthetic barcodes have advanced rapidly (Figure 1C). This strategy introduces diverse DNA barcodes—via viral integration or cumulative genome editing—to simultaneously track thousands to millions of clones and reconstruct phylogenetic trees. Crucially, these barcodes can be directly integrated with single-cell sequencing to enable high-resolution state-to-fate mapping.

Finally, to map lineages where transgenesis is precluded—particularly in human tissues—the field has pioneered retrospective tracing using natural variants (Figure 1D). By leveraging natively accumulated somatic mutations (e.g., single-nucleotide variants [SNVs] and mitochondrial DNA [mtDNA] variants) and stable epimutations (e.g., DNA methylation), endogenous

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

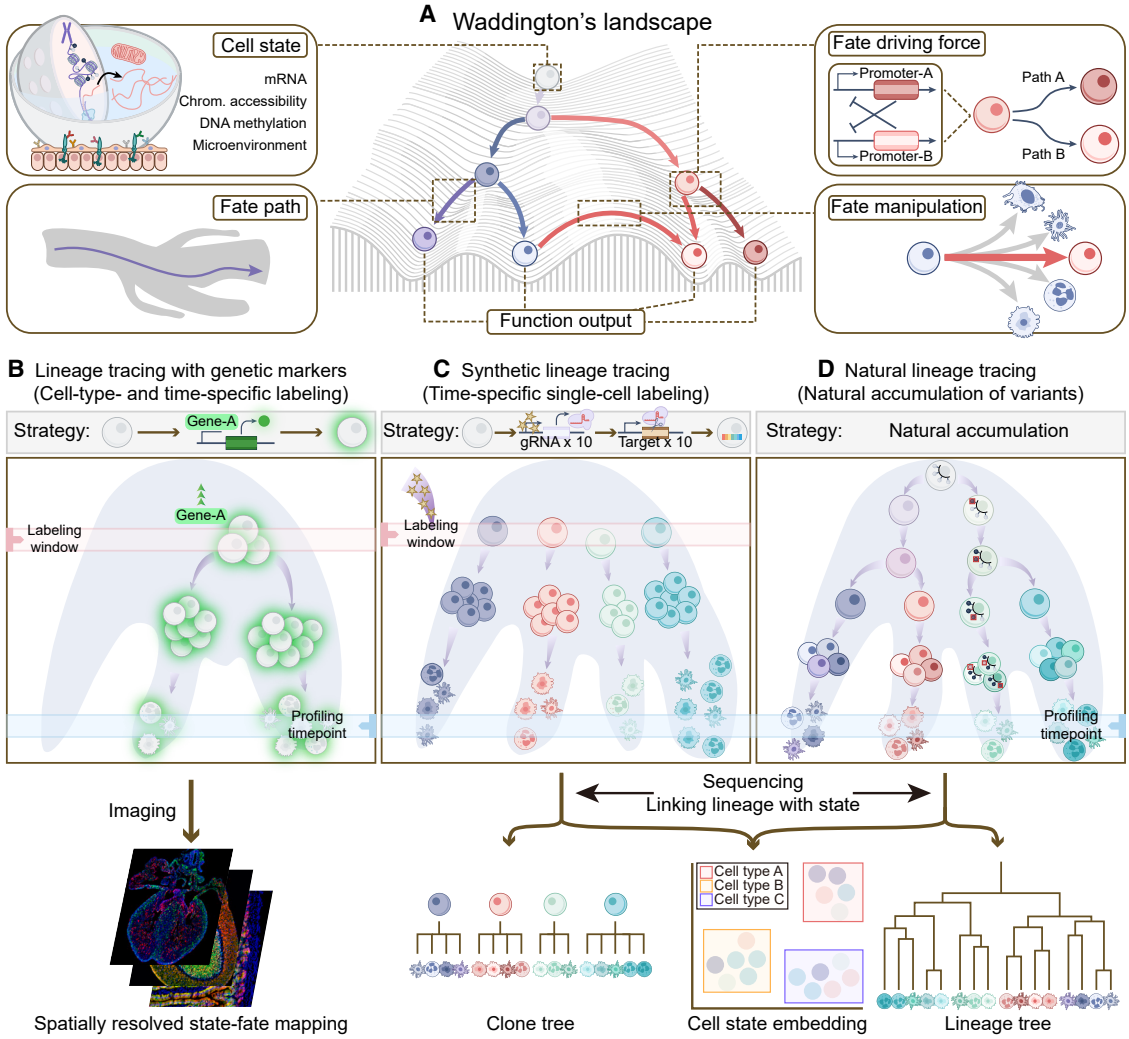

Figure 1. The three methodological pillars of lineage tracing for systematic cell fate analysis

(A) Elements of cell fate study from the Waddington landscape. Cell fate choice is represented as a ball traversing the Waddington landscape along defined fate paths. A systematic understanding requires decoding cell states, which include intrinsic molecular profiles and extrinsic niche cues, and fate paths, where cells progress along defined developmental routes. scRNA-seq enables the reconstruction of continuous cell-state manifolds to capture dynamic transitions. For higher resolution, lineage trees can be constructed by cumulative barcoding strategies that link cell division with barcode labeling. Crucially, these state manifolds may not inherently represent lineage relationships, as transcriptomic similarity does not necessarily equate to lineage proximity.[5–8] Further, fate-driving forces (e.g., transcription factors) must be identified, fate manipulation to redirect lineage outcomes performed, and the functional roles of committed fates in development and homeostasis resolved. Abbreviations are as follows: Chrom., chromatin.

(B) Lineage tracing with genetic markers. This approach uses cell-type- and time-specific drivers (e.g., Cre/Dre) to label defined populations. It is optimal for spatially resolved state-fate mapping via imaging and targeted functional perturbations.

(C) Synthetic lineage tracing. This strategy utilizes engineered barcodes (e.g., gRNA and target arrays) for time-specific single-cell labeling. By integrating single-cell sequencing, high-resolution clonal lineages can be directly linked to molecular states, enabling the systematic resolution of cell fate heterogeneity at the single-cell level. Abbreviation is as follows: gRNA, guide RNA.

(D) Natural lineage tracing. This pillar relies on the natural accumulation of endogenous variants (e.g., SNVs, mtDNA, or DNA methylation) as internal clocks. It enables the reconstruction of lineage trees where experimental manipulation is precluded, such as human tissues.

phylogenies can now be reconstructed, ushering in an era of non-invasive, multi-omic lineage tracing.

Building on previous reviews on lineage tracing that provide an excellent summary of technologies and challenges at the time,[12–17] here we provide a broad overview of recent advances across the field's three core pillars and their integration with single-cell and spatial genomics. We also examine the computational tools and challenges involved in combining lineage and cell-state data to decipher cell fate decisions. Through systematic comparison of available approaches, we offer practical guidance for selecting optimal systems and conclude by highlighting emerging opportunities and future directions.

LINEAGE TRACING WITH GENETIC MARKERS

A central goal of cell fate research is to label a defined initial state, track its dynamic path over time, and measure its functional output. Prospective tracing using genetic markers is

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

CellPress
OPEN ACCESS

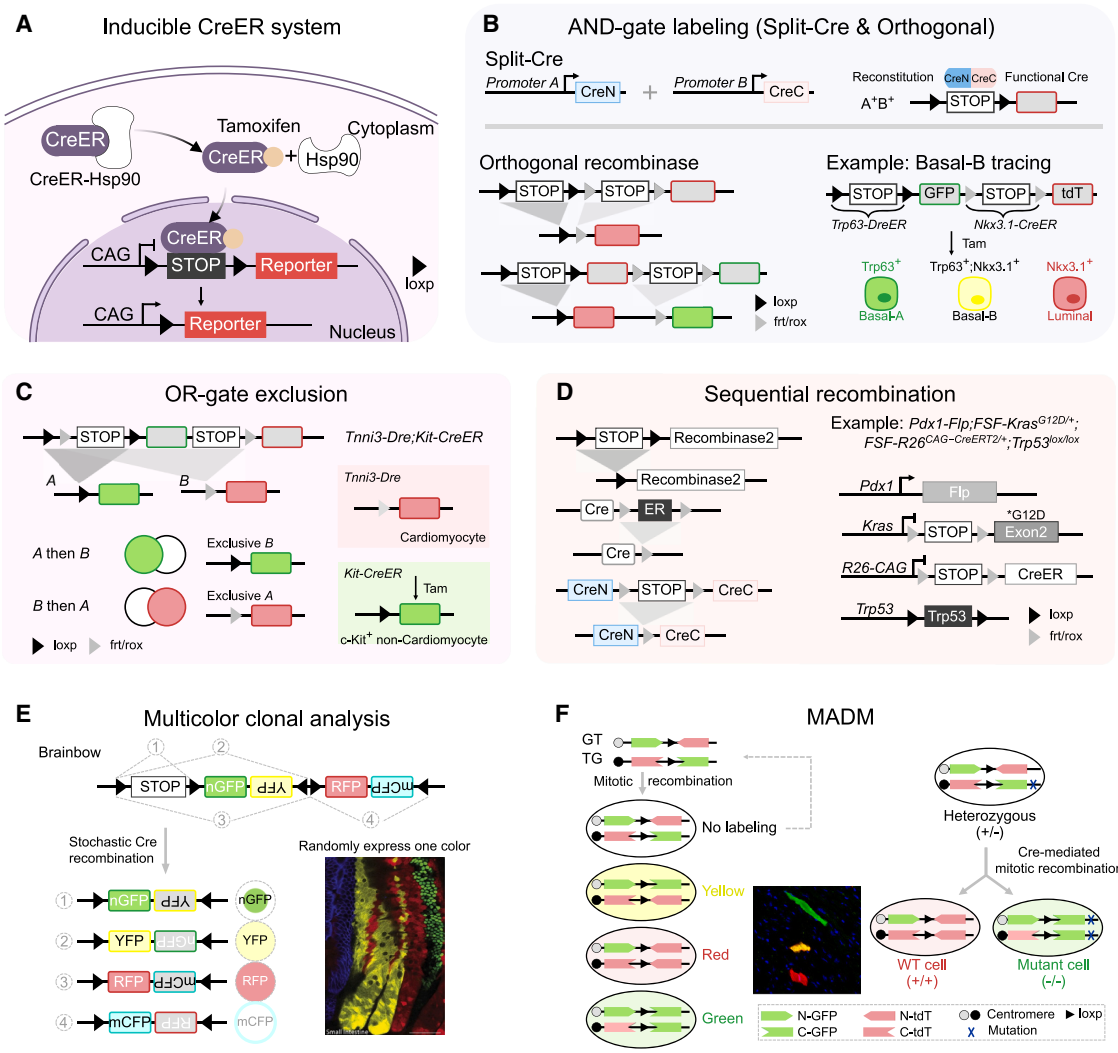

Figure 2. Genetic strategies for driver-induced lineage tracing and clonal analysis

(A) Inducible CreER system. Tamoxifen administration triggers the nuclear translocation of CreER, mediating irreversible recombination at *loxP* sites to activate a heritable reporter for temporal fate mapping. Abbreviation is as follows: Hsp90, heat shock protein 90.

(B) AND-gate labeling via split-Cre and orthogonal systems. In split-Cre architectures, functional recombinase is reconstituted only upon co-expression of N- and C-terminal fragments. Orthogonal systems (e.g., Cre/Dre) utilize two distinct recombinases to remove interleaved STOP cassettes, enabling intersectional labeling of double-positive populations such as basal-B cells. Abbreviations are as follows: Tam, tamoxifen; tdT, tdTomato.

(C) OR-gate exclusion for refined specificity. Interleaved recognition sites for two orthogonal systems (e.g., *Tnni3-Dre* and *Kit-CreER*) allow the first system to delete recognition sites for the second, ensuring reporter activation is restricted to the target non-myocytic population.

(D) Sequential recombination for stepwise control. Hierarchical control in which one recombinase activates another, such as a *Flp*-activated CreER line. This enables spatial restriction by a first driver (e.g., *Pdx1*) followed by temporal induction via tamoxifen for sophisticated multi-stage fate mapping.

(E) Stochastic multicolor reporters for clonal resolution. Systems such as Brainbow use random, irreversible recombination of incompatible sites to label individual progenitors and their descendants with distinct colors, enabling high-resolution analysis of fate heterogeneity. Abbreviations are as follows: nGFP, nuclear green fluorescent protein; YFP, yellow fluorescent protein; RFP, red fluorescent protein; mCFP, membrane-targeted cyan fluorescent protein.

(F) MADM (mosaic analysis with double markers). Cre-mediated mitotic recombination during mitosis generates differentially labeled homozygous mutant (green) and WT (red) sister cells from a heterozygous progenitor. This provides single-cell-resolution lineage tracing with an intrinsic internal control to study gene function. Abbreviations are as follows: GT, N-GFP/C-Tomato; TG, N-Tomato/C-GFP; WT, wild type.

ideally suited for this purpose, having evolved to afford increasingly precise spatiotemporal control over labeling. Crucially, because the starting population is defined by genetic logic, this approach integrates seamlessly with targeted perturbations to dissect gene or cell function. Below, we trace this progression—from classic single-recombinase systems through enhanced specificity of dual-recombinase logic to multicolor strategies for clonal analysis (Figure 2; Table 1).

Labeling cells with a single recombinase

The foundational tool for marker-based tracing is the single-recombinase system. These systems (e.g., Cre) recognize specific DNA sequences (e.g., *loxP*) to mediate genomic rearrangements, typically excising a STOP cassette to activate a fluorescent reporter. For temporal control, ligand-inducible variants such as CreER were developed.[66,67] CreER remains sequestered in the cytoplasm until an exogenous ligand (e.g., tamoxifen)

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

Table 1. Modern lineage tracing approaches

Category	Genetic marker	Synthetic barcode	Natural variant
Basic design principles	utilizes tissue- or cell-type-specific promoters to drive site-specific recombinases (e.g., Flp, Cre, and Dre), which permanently activate or perturb downstream targets to map the fate of predefined cell populations	introduces artificial, high-diversity DNA sequences into the genome—either as static identifiers or dynamically evolving edits—to uniquely tag individual cells and reconstruct complex clonal phylogenies	retrospectively reconstructs cell lineages by tracking the spontaneous, progressive accumulation of heritable genetic or epigenetic alterations that naturally occur during cell division
Representative methods	• constitutive and spatiotemporally inducible systems (e.g., Cre, Dre, CreER, and DreER)[18–20] • split recombinase (split-Cre[21–24]) • orthogonal dual-recombinase systems (Cre/Flp,[25–27] Cre/Dre[28]) • stochastic multicolor reporters (e.g., Confetti[29] and Brainbow[30]) • mosaic analysis with double markers (MADM[31])	• static viral libraries (e.g., LARRY, CellTagging, Watermelon, CaTCH, and ClonMapper)[5,32–35] • endogenous static strategies (e.g., Polylox,[36] LoxCode,[37] DRAG,[38] Sleeping Beauty,[39] and FlipJump[40]) • CRISPR-based dynamic recorders (e.g., LINNAEUS,[41] GESTALT,[42] ScarTrace,[43] DARLIN,[44] KP-Tracer,[45] macsGESTALT,[46] CREST,[6] CHYRON,[47] MARC1,[48] iTracer,[49] and eTRACER[50]) • base/prime editing (e.g., PEtracer,[51] SMALT,[52] peCHYRON,[53] DNA Typewriter,[54] and BASELINE[55])	• nuclear somatic mutations[56] • DNA methylation epimutations (e.g., MethylTree[57], EPI-Clone[58], and EVOFLUx[59]) • mtDNA heteroplasmy (e.g., MAESTER[60] and mtscATAC-seq[61])
Suggested reviews	• classic lineage tracing[16] • dual-recombinase systems[18]	• integration with single-cell omics[12] • conceptual frameworks and challenges[13] • additional reviews[14,15,62]	• epimutation[63] • mtDNA[64] • SNV lineage tracing[65]

triggers nuclear translocation. This restricts recombination—and thus heritable labeling—to a defined temporal window, such as a specific developmental phase or injury response (Figure 2A).

This precise control has been instrumental in resolving complex cell fate choices across diverse contexts. In development, tamoxifen-induced *Runx1-Mer-Cre-Mer* recombination at embryonic day 7 (E7) resolved brain macrophage ontogeny by showing that microglia derive from yolk sac progenitors, not adult hematopoietic stem cells (HSCs).[68] During homeostasis, inducible *Lgr5-CreER* showed that $Lgr5^+$ crypt base columnar cells act as actively cycling stem cells that continuously regenerate all intestinal epithelial lineages.[69] Following injury, single-recombinase tracing has uncovered remarkable cell plasticity—for example, lung Club cells dedifferentiating into basal stem cells during airway repair[70] and pancreatic α cells transdifferentiating into β cells after extreme β cell loss.[71] Together, these studies highlight how precisely timed genetic tracing decodes the adaptive cellular responses underlying tissue development, maintenance, and repair.

Resolving complex contexts: Dual recombinase and logic gates

Single-recombinase systems effectively track populations defined by one marker, but many transient or highly specific subpopulations cannot be uniquely identified by a single gene. Nonspecific or ectopic expression of the selected marker often causes these systems to capture unwanted cell types. To address this, dual-recombinase systems implementing Boolean logic gates have been developed (Figures 2B–2D). These systems use orthogonal recombinases (e.g., Cre/Flp and Cre/Dre) or split-recombinase architectures (e.g., split-Cre) to perform precise set operations on cell populations, solving key targeting challenges.

AND logic requires co-expression of two markers to activate a reporter, enabling isolation of rare double-positive subpopulations that lack a unique identifier (Figure 2B). For instance, split-Cre reconstitutes a functional enzyme only when two inactive fragments are co-expressed in the same cell. This mechanism was used to selectively label parenchymal microglia based on *Sall1* and *Cx3cr1* co-expression, distinguishing them from border-associated macrophages that share individual markers and revealing microglia's unique molecular identity in brain homeostasis.[21] However, split-Cre often shows lower recombination efficiency than intact Cre, limiting its use for comprehensive clonal recovery. To overcome this, orthogonal systems (e.g., Cre/Dre) allow independent control by distinct genes to activate reporters only in intersectional cells.[28] This approach labeled transient $Trp63^+$ $Nkx3.1^+$ "basal-B" intermediate cells in the prostate—a population masked by broad single-marker tracing—demonstrating their role as a transitional state in prostate epithelial differentiation.[72]

OR logic actively prevents reporter activation in off-target populations, eliminating background "leakiness" that frequently confounds single-recombinase tracing (Figure 2C). A prime

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

example is the DeaLT system, which uses one recombinase to preemptively excise an essential reporter component in a specific population, ensuring that a second recombinase cannot label those cells. This approach proved particularly powerful in resolving a long-standing controversy in cardiac regeneration. Single-marker *c-Kit-CreER* tracing had suggested the existence of c-Kit$^+$ cardiac stem cells capable of generating new cardiomyocytes, but results were confounded by leaky expression in mature myocytes.[73–75] By combining *Tnni3-Dre* (to excise the reporter specifically in cardiomyocytes) with inducible *c-Kit-CreER*, labeling could only occur in true non-myocyte c-Kit$^+$ cells—revealing negligible new cardiomyocyte formation from this population and disproving significant endogenous cardiomyogenesis in the adult mouse heart.[76] Similarly, in cartilage research, subtractive OR logic isolated genuine articular cartilage progenitors by eliminating reporter activation in non-skeletal Procr$^+$ cells, revealing that *Procr* marks a distinct progenitor population restricted to articular cartilage that contributes to joint homeostasis and repair.[77]

Beyond static intersections, dual-recombinase systems enable temporal labeling through sequential logic, allowing multi-step biological processes to be modeled (Figure 2D). For example, to mirror the stepwise evolution of pancreatic cancer, *Pdx1-Flp* was used to first activate *Kras*^{G12D} and establish precancerous lesions, followed by the Flp-dependent activation of CreER to delete *Trp53*. This sequential manipulation drove the transition to invasive adenocarcinoma, providing an *in vivo* platform to study molecular triggers of metastasis within an intact tumor microenvironment. Conversely, capturing transient cellular states asynchronously—without repeated drug administration—requires continuous recorders.[78] Systems such as ProTracer and αDKRC enable the cumulative tracking of cell proliferation *in vivo*.[79,80] Utilizing ProTracer for unbiased, long-term lineage tracing, recent studies demonstrated that liver homeostasis and repair are driven primarily by regional proliferation within the midzonal hepatocyte population.[79]

Dissecting heterogeneity: Clonal analysis with multicolor reporters

Standard marker-based tracing labels broad populations with a single color, masking the heterogeneity of individual cells. Although population tracing reveals collective tissue flux and ensemble behaviors, clonal analysis is indispensable for quantifying cell-to-cell heterogeneity and unmasking rare dominant clones that would otherwise be obscured in population averages. To determine whether cells within a defined population possess distinct proliferative potentials or spatial behaviors, multicolor clonal analysis is required to track the progeny of individual cells. This clonal resolution is achieved using two primary genetic strategies: stochastic multicolor reporters and mosaic analysis with double markers (MADM).

Stochastic multicolor reporters (e.g., Brainbow and Confetti) use multiple incompatible recombinase sites within a single allele.[29,30,81,82] Upon Cre induction, random recombination activates one of several fluorescent proteins, creating a unique heritable color label for a progenitor and its descendants (Figure 2E). This strategy solves the problem of distinguishing adjacent clones in dense tissues. For example, by color-coding neighboring Lgr5$^+$ stem cells with Confetti, spatial clonal competition was visually tracked within the intestinal crypt.[29] This demonstrated the principle of neutral drift, showing that random stem cell loss and replacement ultimately drive each crypt toward monoclonality.[83] Similar stochastic clonal dynamics have been observed in epidermis,[84] esophagus,[85] and spermatogenesis systems,[86] suggesting that neutral drift may represent a general principle of stem cell-driven tissue homeostasis.

Advancing beyond purely observational tracing, MADM couples color labeling directly with genetic perturbation.[31,87–89] This system uses Cre-mediated interchromosomal recombination to generate differentially colored sister cells from a single mitotic event (Figure 2F). By producing a homozygous mutant cell (e.g., labeled green) alongside a matched wild-type sister cell (e.g., labeled red), MADM provides an intrinsic internal control. This resolves a major experimental challenge: separating cell-autonomous effects from environmental influences. For instance, by comparing the *in vivo* expansion of mutant clones against their normal sister cells, oligodendrocyte precursor cells were identified as the cell-of-origin for gliomas—revealing that concurrent *p53* and *Nf1* mutations drive aberrant premalignant proliferation exclusively in this lineage.[90]

Practical considerations

Labeling specificity remains a central concern in marker-based tracing. Although widely adopted, single-recombinase systems are inherently susceptible to leaky promoter expression, off-target labeling, and residual ligand activity, as illustrated by the *Pf4-Cre* system, which inadvertently labels upstream HSCs rather than being megakaryocyte specific.[91,92] However, as is well known in the field, these limitations stem from the inherent biological and technical constraints of the tools, not from a lack of care in previous foundational studies. The deep understanding of these limitations has enabled the design of appropriate controls—such as utilizing uninduced cohorts to quantify background leakiness, conducting short-term tracing to validate the initially labeled population and performing orthogonal co-staining to confirm precise targeting—thereby ensuring the generation of reliable, foundational insights.[16,18,69,75]

Although advanced dual-recombinase logic and multicolor systems enhance targeting resolution, they introduce new practical challenges. Integrating multiple alleles necessitates complex breeding that demands substantial time and resources, particularly when driver lines are not yet established. More importantly, rigorous validation is required to ensure fate conclusions are not confounded by technical artifacts.

First, orthogonality must be verified: selected recombinases (e.g., Cre, Flp, and Dre) should not cross-react with each other's target sites—for example, Cre should not recombine rox sites nor Flp act on *loxP* sites. Second, sequential integrity must be tested. Induction protocols require tight control to prevent unintended labeling from overlapping activation windows, recombinase persistence, or residual ligand activity. This requires timed controls using single-recombinase induction at each experimental stage. Finally, background leakage must be evaluated by quantifying non-specific baseline recombination. To establish a reliable signal-to-noise ratio, "single-driver" control groups (e.g., Cre-only and Dre-only groups) should be included alongside full experimental cohorts (see Liu et al. for a comprehensive

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

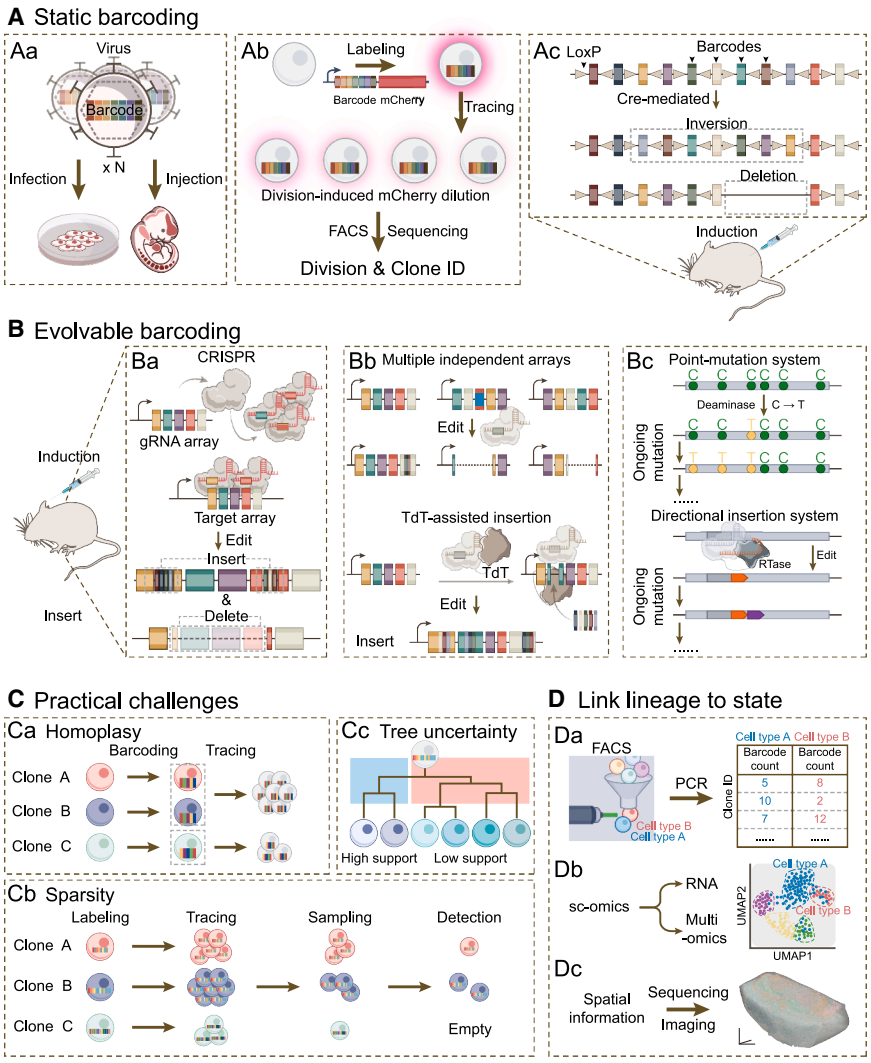

Figure 3. Strategies, practical challenges, and readout modalities of synthetic barcoding

(A) Static barcoding for clonal analysis. These methods generate high-diversity barcodes at a single time point to identify descendants of individual progenitors. Viral barcoding uses lentiviral libraries (e.g., LARRY) for *ex vivo* or *in vivo* transduction (Aa). Division-induced reporter dilution allows clonal-level monitoring of proliferative history (Ab), while Cre-mediated recombination systems (e.g., LoxCode) utilize stochastic inversion and deletion between *loxP* sites to create unique reshuffled barcodes (Ac). Abbreviations are as follows: FACS, fluorescence-activated cell sorting; mCherry, monomeric Cherry fluorescent protein; ID, identification.

(B) Evolvable barcoding for phylogenetic reconstruction. Cumulative systems record dynamic lineage histories through continuous mutational events. CRISPR-based array editing uses gRNA-directed Cas9 to create indels within a tandem array (Ba). Barcoding diversity is further enhanced by distributing multiple independent arrays across the genome or by utilizing optimized chemistries such as TdT-assisted insertion to increase information density (Bb). Next-generation non-destructive recorders (e.g., SMALT or prime editing) accumulate substitutions or precise edits without DSBs to maintain stable temporal records (Bc). Abbreviations are as follows: CRISPR, clustered regularly interspaced short palindromic repeats; TdT, terminal deoxynucleotidyl transferase; RTase, reverse transcriptase; C → T, cytosine to thymine.

(C) Practical challenges in synthetic lineage tracing that could confound data interpretation. These include homoplasy, when distinct clones are tagged with identical barcodes (Ca); sparsity, due to low recovery or sampling dropouts (Cb); and tree uncertainty, arising from insufficient or erroneous mutational signatures.

(D) Linking lineage history to molecular and spatial states. The integration of FACS and PCR enables cost-effective clonal quantification (Da). Single-cell sequencing links cell lineage to high-dimensional molecular states (RNA/multi-omics; Db), while spatial information (Dc; via sequencing or imaging) provides the architectural context for *in situ* state-to-fate mapping. Abbreviations are as follows: PCR, polymerase chain reaction; sc-omics, single-cell omics; UMAP, uniform manifold approximation and projection.

discussion on the rigorous validation protocols required for dual-recombinase systems[18]).

SYNTHETIC LINEAGE TRACING WITH ENGINEERED BARCODES

Although genetic markers excel at mapping defined populations, their limited labeling diversity limits the throughput of clonal analyses and often obscures the heterogeneity of individual cell fate decisions. To resolve how subtle variations in early states dictate divergent outcomes, massive clonal resolution is required. Synthetic barcodes achieve this by introducing highly diverse DNA barcodes—via viral integration or cumulative *in situ* genome editing. Static barcoding enables high-throughput clonal analyses, while evolvable barcoding generates deep, high-resolution lineage histories. Below, we explore both static barcoding and evolvable barcoding and discuss challenges in practical implementation (Figure 3; Table 1).

Static barcoding

Static synthetic barcoding introduces a unique, immutable DNA label into a founder cell, which is passed to all its progeny (Figure 3Aa). These barcodes are traditionally delivered *ex vivo* using complex lentiviral libraries, such as LARRY or CellTagging.[5,32] These approaches, when coupled with single-cell sequencing, excel at decoding complex differentiation trajectories. For instance, LARRY revealed that hematopoietic progenitors commit to specific lineages days before these fate biases manifest transcriptomically.[5] Furthermore, it identified the key transcription factor *Tcf15* as a crucial driver and marker of functional, long-term HSCs.[93] Similarly, CellTagging has tracked vast numbers of engineered cells, pinpointing rare, successful clonal trajectories during *in vitro* fibroblast reprogramming and identifying early molecular markers that accurately predict terminal fate outcome.[32,94]

Expanding beyond basic tracking, recent viral tools couple static barcoding with functional assays (Figure 3Ab). For

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

example, Watermelon combines DNA barcoding with label dilution to concurrently record lineage and proliferation history, identifying rare, slow-cycling "persister" clones that drive cancer relapse.[33] Meanwhile, tools such as CaTCH and ClonMapper overcome the destructive nature of single-cell sequencing by using targeted CRISPR activation to induce fluorescent reporters only in specifically barcoded cells, enabling physical isolation of live clones of interest.[34,35] This live-cell recall was used to isolate therapy-resistant clones prior to drug exposure, validating that pre-existing, intrinsically primed states drive drug resistance.

Historically, viral barcoding was restricted to *in vitro* or transplantation assays that often fail to reflect true native physiology. Although recent advances have partially circumvented this by utilizing ultrasound-guided *in utero* microinjections to deliver lentiviral libraries into embryonic niches,[95] broader *in vivo* mapping requires endogenous static barcoding strategies (Figure 3Ac). These systems use orthogonal transposases (e.g., Sleeping Beauty[39] and FlipJump[96]), stochastic recombinase-mediated rearrangement of engineered DNA cassettes (e.g., Polylox[36,97] and LoxCode[37]), or recombination within a synthetic V(D)J locus (DRAG).[38] By tracking clones in their native environment, the Sleeping Beauty system challenged previous transplantation models, demonstrating that steady-state hematopoiesis is highly polyclonal—driven by thousands of successive progenitor waves rather than a few dominant stem cells.[39] Furthermore, *in vivo* tracking showed that specific stem cell subsets are intrinsically fate-biased toward restricted lineages, such as megakaryocytes, even during normal homeostasis.[96]

Evolvable barcoding

Although static barcodes establish which endpoint cells belong to the same clone, they do not capture the continuous, sequential branching history of those cells. To reconstruct detailed phylogenetic trees, dynamic evolvable recording (or cumulative barcoding) is required, where an engineered locus continually accrues mutations over successive cell divisions (Figure 3B). Early cumulative strategies—such as GESTALT, LINNAEUS, and ScarTrace—targeted Cas9 to tandem arrays of single-guide RNA (sgRNA) recognition sites (Figure 3Ba).[41–43,98] However, these systems rely on targeted double-strand breaks (DSBs) repaired by non-homologous end joining, a process heavily biased toward deletions. This creates a severe structural bottleneck known as "inter-site dropout," where a single large deletion can erase previously encoded mutations, irreversibly destroy ancestral information, and fragment the downstream phylogenetic tree.

To overcome the limitations of finite target sites and single-locus vulnerability, distributed and self-propagating architectures were developed (Figure 3Bb). Multiplexed systems such as macsGESTALT, CREST, and KP-Tracer disperse target sites across multiple independent genomic loci to structurally prevent single large deletions from wiping out the entire cellular record.[6,45,46] Alternatively, MARC1 uses self-targeting guide RNAs (hgRNAs) that carry their own protospacer adjacent motif (PAM); this allows them to continuously edit themselves and regenerate new targets, exponentially increasing barcode capacity over time.[99] Concurrently, to alter the underlying mutational chemistry and shift the bias toward information-rich insertions, systems such as DARLIN and CHYRON fuse Cas9 to terminal deoxynucleotidyl transferase (TdT).[44,47] This polymerase drives continuous, non-templated nucleotide additions, partially bypassing the progressive target erasure inherent to standard nucleases. Combining this engineered chemistry with three distributed recording arrays, DARLIN has been established as a genetically stable mouse line with massive barcoding diversity and a high lineage recovery rate (~70%), providing a high-resolution and accurate tracking platform for mapping complex adult tissues.[44]

Although these Cas9-based improvements mitigate some barcode erasure, they still yield limited tree depth due to the inherent instability of target sites following DSBs. To circumvent this limitation, non-destructive editing strategies have emerged (Figure 3Bc). Systems using base editors (e.g., SMALT and BASELINE) accumulate stable substitution mutations without cleaving the DNA backbone.[52,55] A related strategy uses Cas9-deaminase on endogenous L1 elements to achieve edits without deletions.[100] Furthermore, prime editors (e.g., DNA Typewriter and peCHYRON) leverage pegRNAs and reverse transcriptases to insert predefined short sequences.[53,54] Both DNA Typewriter and peCHYRON physically enforce strict, sequential editing through an "edit-and-advance" mechanism. By logging the exact temporal sequence of mutations, this deterministic ordering resolves a major computational hurdle associated with stochastic Cas9 models, enabling precise reconstruction of deep and highly complex lineage trees.

The *in vivo* power of these advanced barcoding systems extends across both developmental biology and oncology. Under physiological conditions, ScarTrace systematically mapped the fate potential of early embryonic cells in zebrafish, revealing clonal relationships during hematopoiesis and in the brain and eyes.[43] Similarly, developmental phylogeny mapping in *Arabidopsis thaliana* using SMALT revealed that all shoot branch cells derive from exactly three founder cells, each belonging to one of three early-determined lineages.[101] In pathological contexts, KP-Tracer has elucidated the complex phylodynamics, plasticity, and hierarchical nature of tumor evolution in mouse lung cancer.[45] Concurrently, SMALT applied to colon cancer in mice has challenged the long-standing dogma of strictly monoclonal tumor origins, demonstrating that colon cancers actually initiate from polyclonal populations before undergoing a transition to a monoclonal state.[102] Notably, this dynamic was corroborated by Confetti-based genetic tracing in *Apc*-driven tumorigenesis, which revealed that early polyclonality enables cells to overcome initial fitness barriers before the tumor ultimately transitions to a monoclonal state.[103]

Practical challenges in synthetic lineage tracing

Despite the promise of synthetic lineage tracing, its successful application requires navigating several inherent challenges (Figure 3C). These include several technical hurdles, such as variable expression or even silencing of reporter constructs in a cell-type-specific manner, as well as genotoxicity during viral integration or genomic editing. Beyond these technical challenges, there are more generic and conceptual challenges that we will focus on below. A primary concern during the initial labeling phase is barcode homoplasy, where cells from distinct origins are inadvertently tagged with identical barcodes (Figure 3Ca).

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

This overlap can skew cell fate analyses by generating artificial "multipotent" clones. To circumvent this artifact, the effective diversity of the barcode library should be empirically assessed and must vastly outnumber the target population size at the time of labeling; specifically, achieving an error rate below 1% requires a barcode diversity at least 100 times larger than the target population.[5,36,37] Although this diversity is easily tunable in lentiviral-based libraries, it remains a rigid constraint in *in vivo* models, necessitating careful selection of a system with sufficient capacity for the biological question at hand.

The second challenge is data sparsity, which occurs when only a small fraction of cells within a clone are ultimately sampled (Figure 3Cb). This incomplete capture can create the illusion of smaller, highly coherent clones, leading to erroneous conclusions of early fate priming. Sparsity stems from two main factors: the high cost of deep single-cell sequencing and the low barcode recovery rates typical of many *in vivo* synthetic barcodes—the CARLIN mouse model, for instance, recovers only about 10% of barcodes.[104] Low barcode recovery can result from inefficient editing or poor expression of lineage barcodes. Mitigating this artifact requires a synthesis of optimized experimental models and robust computational frameworks. Experimentally, systems engineered for high lineage barcode recovery (e.g., LARRY: >70% and DARLIN: ~70%) should be prioritized.[5,44] Computationally, only biological conclusions robust to additional data subsampling should be trusted. To further mitigate the risk of misinterpretation, rigorous statistical validation can be applied by shuffling barcode data to generate a null baseline, enabling evaluation of the statistical significance of observed lineage relationships.

Finally, the robustness of the inferred lineage tree is fundamentally limited by the mutational signature of the recorder itself (Figure 3Cc). Traditional CRISPR-based approaches that generate large insertions and deletions often suffer from information loss, as subsequent cuts overwrite previous edits. This continuous erasure leads to shallow and less robust lineage trees, a structural limitation that persists even when employing advanced reconstruction algorithms such as Cassiopeia.[105] Overcoming this hurdle requires adopting richer recording chemistries that utilize non-destructive, high-information mutation signatures. For instance, the SMALT system leverages base editing to generate high-depth trees with a median statistical support value of approximately 90%.[52] Parallel to these experimental innovations, computational strategies are evolving to bridge informational gaps. Emerging tools such as LinRace and LinTIMaT attempt to construct consensus trees by integrating CRISPR barcode data directly with transcriptomic profiles.[106,107] However, these multimodal integrations must be applied cautiously, as transcriptomic similarity does not inherently dictate lineage similarity.[5–8] Ultimately, rigorous computational validation, such as bootstrap analysis, should be routinely performed to evaluate the statistical support of the inferred tree, ensuring that only highly reliable branches are used to deduce underlying biological mechanisms.

NATURAL LINEAGE TRACING WITH ENDOGENOUS VARIANTS

Both genetic markers and synthetic barcodes rely on transgenic engineering, precluding their use in human subjects. To uncover natural fate paths and functional outputs directly within native human tissues, retrospective lineage tracing is required—leveraging endogenous variants that accumulate during cell division (Figure 4; Table 1). The utility of these variants depends on their faithful mitotic inheritance. Generally, stability follows a clear hierarchy: DNA sequence > DNA methylation > histone modifications > chromatin accessibility (assay for transposase-accessible chromatin [ATAC]) and RNA expression (Figure 4A). Whereas DNA sequences are stably inherited across successive divisions, ATAC and transcriptomic profiles fluctuate dynamically in response to state changes, despite the existence of a few memory genes that persist over several divisions.[108] Highlighting this hierarchy, multi-omics tracing recently demonstrated that DNA methylation patterns of HSCs retain robust clonal memory from embryonic development through adulthood, whereas ATAC and RNA states do not.[44] Histone modifications exhibit intermediate stability,[109] but their lineage tracing potential remains to be fully developed. Consequently, somatic DNA mutations and stable epimutations stand out as the most promising candidates for accurately mapping human cellular ancestry.

Resolving histories via somatic DNA mutations

During cell division, somatic mutations naturally accumulate and are permanently inherited by daughter cells, serving as endogenous barcodes for lineage reconstruction (Figure 4B). Whereas specific genomic rearrangements can be exploited in specialized contexts—such as copy-number variations (CNVs) for mapping unstable cancer genomes or V(D)J recombination for tracking T and B cell clones—retrospective tracing in normal tissues primarily relies on SNVs.[110–116]

Although SNVs should, in principle, offer the highest fidelity for phylogenetic reconstruction, their extremely low natural mutation rate (roughly one per cell division) makes them difficult to detect against background sequencing errors. Colony-based whole-genome sequencing (WGS) remains the gold standard for overcoming this signal-to-noise barrier, as expanding isolated single cells *ex vivo* amplifies DNA prior to sequencing. By cleanly capturing rare somatic variants associated with distinct lineages, high-resolution phylogenetic trees can be reconstructed, where branch lengths directly correlate with elapsed time (Figure 4C). This approach has revealed clonal architecture during early human embryo development[112–116] and tracked lifelong clonal dynamics of native HSCs during aging.[117]

However, colony expansion is expensive, time-consuming, and restricted to culturable cells such as HSCs. Laser capture microdissection (LCM) of solid tissues offers a culture-free alternative but lacks true single-cell resolution and remains susceptible to confounding background noise. To bypass these limitations and capture state-to-fate information directly from primary solid tissues, newer chemistries such as primary template-directed amplification (PTA) were developed. PTA achieves high-fidelity, low-noise whole-genome amplification directly from an uncultured single cell.[118] Leveraging this advancement, automated platforms such as SMART-PTA improve throughput at reduced cost, and enable simultaneous transcriptomic state profiling from that same cell.[119] Although this single-cell approach remains cost-prohibitive for routine application, the ability to jointly profile somatic mutations and

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

CellPress
OPEN ACCESS

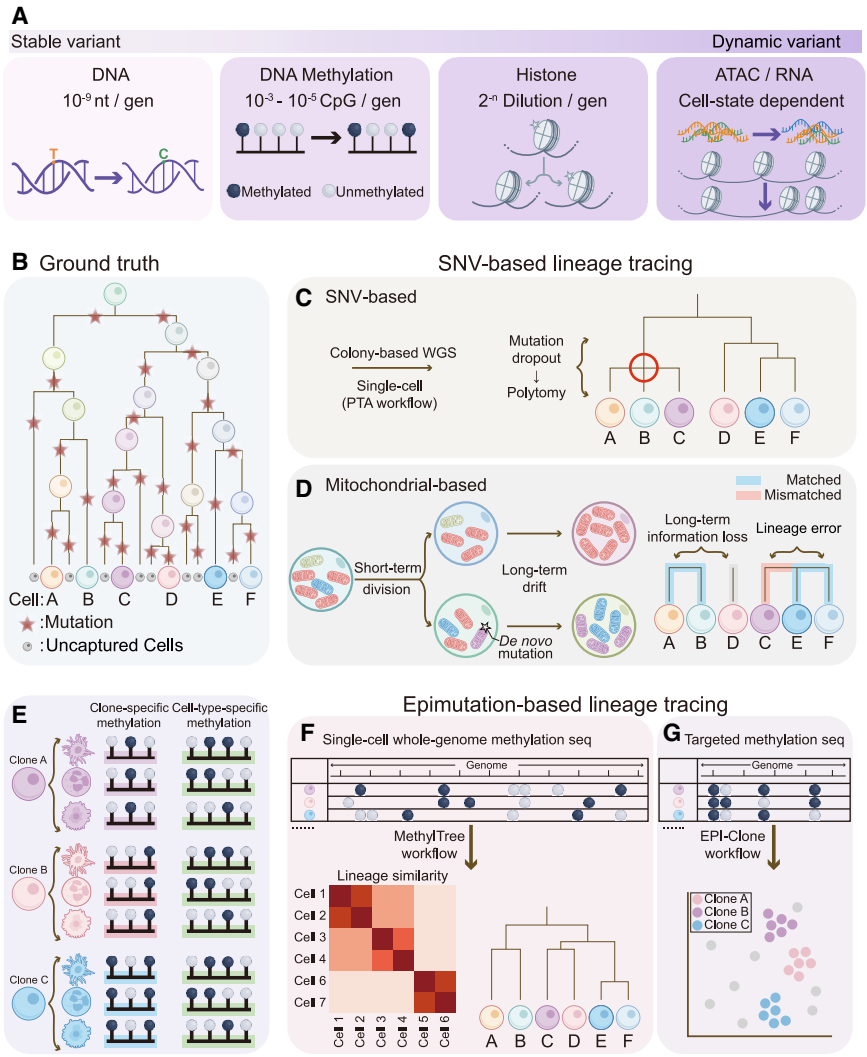

Figure 4. Natural lineage tracing using endogenous variants

(A) Stability of natural variants in the cell. Nuclear DNA is the most stable component and remains invariant during differentiation. Although most epigenetic marks are highly dynamic, DNA methylation stands out as a reliable reservoir, harboring numerous neutral or lineage-specific CpG sites that remain stable across cell divisions. Abbreviations are as follows: CpG, cytosine-phosphate-guanine site; ATAC, assay for transposase-accessible chromatin; nt, nucleotide; gen., generation.

(B) Schematic of true lineage relationships for detected cells. Stochastic mutations (stars) accumulate over successive divisions and can be measured and utilized to reconstruct the lineage tree.

(C) SNV-based lineage tracing. Somatic SNVs function as high-fidelity barcodes detected via colony-based WGS or single-cell PTA workflows. Technical hurdles such as mutation dropout can lead to phylogenetic polytomy. Abbreviations are as follows: SNV, single-nucleotide variant; WGS, whole-genome sequencing; PTA, primary template-directed amplification.

(D) Mitochondrial-based lineage tracing. Due to the asymmetric segregation of mtDNA variants and long-term genetic drift, mtDNA mutations are primarily suited for tracking short-term clonal information and are susceptible to sparse recovery and reconstruction errors. Abbreviation is as follows: mtDNA, mitochondrial DNA.

(E) Schematic of epimutation dynamics. This illustrates the distinction between clone-specific CpG methylation patterns used for tracing and cell-type-specific marks.

(F) Lineage reconstruction from single-cell whole-genome methylation. The MethylTree workflow utilizes a lineage similarity matrix to resolve ancestral relationships from genome-wide data. Abbreviation is as follows: Seq, sequencing.

(G) Identification of expanded clones with EPI-Clone. This approach utilizes targeted CpG sequencing to achieve high scalability in blood.

transcriptomic states represents a major leap forward for natural variant tracing in human tissues.

As a complementary alternative to specialized experimental approaches and deep sequencing, computational methods such as Monopogen and SpaceTracer can detect mutations from relatively shallow, high-throughput single-cell sequencing data.[120,121] Although this provides an accessible avenue for lineage and clonal analysis in humans, these algorithmic approaches are currently better suited for scenarios with high inherent mutation rates, such as cancer, rather than the low-mutation environments of normal homeostatic tissues.

Resolving histories via mitochondrial DNA mutations

To bypass the sparsity and prohibitive cost of nuclear SNVs, mtDNA mutations have recently emerged as a compelling alternative (Figure 4D).[122,123] Because the mitochondrial genome mutates at a 10- to 100-fold higher rate than the nuclear genome and is highly transcribed, these variants serve as abundant endogenous barcodes. To efficiently leverage them, targeted enrichment approaches—such as MAESTER and mtscATAC-seq—have been developed.[60,61] Crucially, these methods enable simultaneous capture of mtDNA variants and cell-state information using standard, high-throughput single-cell platforms (e.g., single-cell RNA sequencing [scRNA-seq] or scATAC-seq), making this strategy highly accessible for mapping clonal architecture in human tissues.

However, mitochondrial tracing introduces severe biological and analytical limitations due to the non-deterministic inheritance of mtDNA. Because mitochondria segregate randomly during cell division, heteroplasmic mutations are subjected to intracellular selection and neutral drift, driving variants toward either complete fixation or total loss in descendant cells. Consequently, accurate phylogenetic tree reconstruction is inherently challenging, leaving analyses susceptible to misinterpretation.[124,125] To mitigate these risks, computational frameworks such as MitoDrift simulate the drift dynamics of mtDNA mutations, enabling rigorous quantification of statistical confidence in inferred lineage trees.[126] When benchmarked against trees reconstructed from WGS-derived SNVs, MitoDrift demonstrated improved performance over traditional algorithms such as

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

neighbor joining; however, it still yields only moderate precision, even when analysis is restricted to a small subset of cells with the highest statistical support. Overall, these complex segregation dynamics and the propensity for long-term drift make mtDNA more suitable for tracking short-term clonal events, and resulting phylogenies must be interpreted with rigorous computational caution (Figure 4D).[127]

High-resolution lineage tracing via DNA methylation

Early studies estimated that DNA methylation yields an error rate of approximately 0.001 per CpG site per division,[128] suggesting that roughly 10,000 epimutations could arise across the human genome per cell cycle. This continuous accumulation constitutes a massive informational reservoir for lineage tracing (Figure 4E). Epimutation-based tracking was first explored in cancer and intestinal stem cells.[59,63,129,130] More recently, multi-omic applications of the DARLIN system have provided definitive evidence of stable clonal memory within the DNA methylome at single-cell resolution.[44] However, effectively harnessing this natural variation requires overcoming two major challenges: distinguishing neutral, history-recording CpGs from functional, cell-type-associated sites and mitigating cell-cycle noise caused by incomplete methylation maintenance immediately following division.

Two parallel methodologies have recently emerged to address these hurdles (Figures 4F and 4G). The first, MethylTree, uses standard single-cell whole-genome methylation sequencing (e.g., scBS-seq) to capture sparse, randomly sampled CpGs.[57] To navigate this extreme sparsity and reconstruct the true lineage tree, MethylTree generates cell-cell similarity matrices exclusively from CpGs detected in both cells. It then minimizes overall variance to correct for cell-cycle noise and regresses out functional, cell-type-specific signals. The resulting lineage-specific similarity matrix can then be used to reconstruct phylogenetic trees via established methods such as neighbor joining (Figure 4F). The second approach, EPI-Clone, prioritizes scalability through targeted sequencing (scTAM-seq), focusing on a predefined panel of several hundred CpG sites.[58] This targeted readout bypasses data sparsity by capturing nearly complete methylation profiles for those loci. Neutral epimutations are identified by their lack of correlation with jointly profiled cell-type labels, allowing cells to be grouped into discrete clones (Figure 4G).

Benchmarked against ground-truth synthetic barcodes, MethylTree achieves nearly 100% accuracy. Notably, epimutation-resolved clones precisely match DARLIN barcodes induced at E10. This alignment indicates that major, phylogenetically defining epimutations are established during early embryonic development, providing a stable foundational lineage tree despite continued epimutation accumulation later in life. By leveraging these deep embryonic signals, MethylTree estimated approximately 250 *de novo* HSC clones during early murine blood development. In contrast, EPI-Clone reliably detects massive, late-stage clonal expansions in aged human blood but struggles to resolve the smaller clones characteristic of younger individuals. Thus, EPI-Clone trades phylogenetic resolution for scalability, whereas MethylTree provides a reliable, tissue-agnostic approach for non-invasive lineage tracing using genome-wide CpGs (for more discussion, see Fu et al.[63]). Although the genome-wide approach is inherently more costly in terms of sequencing, there is no fundamental limitation on library throughput. Indeed, many platforms already enable profiling of the single-cell DNA methylome across tens of thousands of cells.[57,63]

Exciting progress has also been made in using bulk DNA methylation data to track clonal dynamics. Although these bulk approaches lack single-cell state information and provide lower clonal resolution, they offer a cost-effective means to quantify and monitor clonal behavior in humans. For instance, EVOFLUx tracks clonal dominance in blood cancer samples by leveraging approximately 1,000 selected epimutation CpG sites,[59] while a related method, COMET, quantifies clonal expansion during normal aging.[131] Beyond blood, the protocadherin (PCDH) gene cluster has recently been shown to harbor extensive epimutations across multiple tissues, offering a promising locus for targeted lineage tracing in humans.[63,132]

INTEGRATING LINEAGE HISTORY WITH MOLECULAR AND SPATIAL STATES

Genetic markers, synthetic barcodes, and natural variants have revolutionized ancestry reconstruction, but lineage history alone provides an incomplete picture of cell fate (Figure 1A). To decode the underlying regulatory networks, lineage readouts are now co-captured with molecular and spatial data at the single-cell level.[12,13] Moving beyond isolated phylogenetic trees, these integrated approaches establish the link between a cell's ancestry, phenotype, and physical microenvironment (Figure 3D).

Traditional bulk DNA sequencing quantifies clonal abundance, but modern synthetic tracing strategies embed barcodes within transcribed regions for simultaneous capture with the single-cell transcriptome. Moreover, understanding the regulatory mechanisms driving fate decisions also requires profiling of epigenetic states. CellTag-multi jointly profiles lineage barcodes and chromatin accessibility in single cells and computationally integrates these data with RNA modalities to enhance fate prediction.[133] Pushing further, Camellia-seq captures lineage barcodes alongside transcriptomic, chromatin accessibility, and DNA methylation profiles.[44] The discovery of strong clonal memory in DNA methylation through this platform directly motivated the design of MethylTree for endogenous lineage reconstruction. In human contexts, MethylTree enables non-invasive, multi-omics lineage tracing by jointly profiling RNA expression and DNA methylation, allowing analysis of lineage, transcriptome, and epigenome in single cells.[57] Together, these integrated readouts provide a multidimensional view of the epigenomic landscape guiding clonal evolution.

Integrating barcode information with spatial location represents another critical frontier for dissecting how cell-cell interactions and native microenvironments shape cell fate. Classic marker-based lineage tracing inherently captures spatial information at low cost and high throughput. Although labeled cells can be sorted for single-cell profiling, the lack of lineage barcode diversity fundamentally limits resolution at the single-cell level.

Recent advancements in spatial profiling—progressing from the 55-μm spots of Visium to the 500-nm resolution of Stereo-seq—have enabled spatial lineage tracing with synthetic barcodes at high resolution, albeit at high cost.[134,135] For example, iTracer applied synthetic barcodes to cerebral organoids for spatial readouts via the Visium platform.[49] Similarly, KP-Tracer

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

mapped spatial lineage dynamics in lung cancer using Slide-seq and Slide-tag,[136] while eTRACER employed Stereo-seq.[50]

Alternatively, imaging-based technologies decode barcodes alongside gene expression directly within intact tissues using multiplexed fluorescence *in situ* hybridization (FISH). Both intMEMOIR and PEtracer implement this approach to achieve high-resolution spatial lineage readouts, though they employ fundamentally distinct mutational chemistries.[51,137] intMEMOIR uses site-specific serine integrases to induce stochastic, irreversible state changes across a finite array of genomic memory elements.[137] In contrast, PEtracer leverages prime editing to progressively insert predefined nucleotide marks across dozens of independent cassettes, generating a deeper and more cumulative record of cellular ancestry.[51] Optical systems such as PolyTope instead use combinatorial fluorescent tags to track clonal architectures.[138]

Selecting a readout strategy requires balancing phylogenetic resolution, spatial context, and experimental scale. Although bulk sequencing coupled with fluorescence-activated cell sorting (FACS) lacks single-cell granularity, it enables cost-effective clonal quantification across millions of cells, mitigating the sampling bias of lower-throughput methods (Figure 3D). Conversely, single-cell sequencing provides superior resolution for reconstructing trajectories and multi-omics states but remains constrained by higher per-cell costs. A similar trade-off governs spatial lineage tracing. Sequencing-based *in situ* readouts could in principle accommodate most synthetic and even some natural lineage tracing systems while also enabling flexible multi-omics integration. However, this approach incurs higher costs and, compared with imaging-based methods, currently offers lower spatial resolution—though newer techniques such as Stereo-seq now achieve 500 nm resolution.[134,139] Advanced imaging platforms such as PEtracer achieve high-resolution mapping within intact tissues but come with lower throughput and higher costs than traditional marker-based systems. Although classic genetic markers offer superior scalability, they cannot resolve individual clonal architectures due to limited barcode diversity. Ultimately, the choice of platform is dictated by the competing needs for phylogenetic depth versus the preservation of the native physical microenvironment.

COMPUTATIONAL INTEGRATION OF LINEAGE AND STATE DYNAMICS

The complexity of high-throughput, sparse single-cell lineage tracing data necessitates sophisticated computational integration. Although scRNA-seq data alone can be utilized to infer differentiation dynamics via pseudotime analysis or RNA velocity,[140,141] these approaches are inherently limited by the absence of true temporal constraints and are shown to perform poorly in distinguishing early fate bias among progenitors.[5,8] By explicitly integrating lineage history with transcriptomic state information, the underlying differentiation processes can be captured with significantly higher fidelity.[5]

To achieve a high-resolution view of differentiation dynamics and identify early fate priming, full transition matrices between observed cell states must be established. By enforcing generic assumptions such as sparse and locally coherent transitions, CoSpar infers these matrices from both static and evolving barcodes sampled across multiple time points.[142] Although optimized for *in vitro* models permitting temporal resampling, this framework has also successfully resolved HSC fate biases using *in vivo* lineage tracing data.[44] For evolving barcodes, optimal transport (OT) algorithms, such as LineageOT and moslin,[143,144] reconcile known lineage relationships with transcriptomic distances to compute the most probable flow of cells between successive time points. However, their foundational assumption that lineage similarity dictates fate similarity must be applied cautiously, as it may fail in contexts of divergent or convergent differentiation.[13] Alternatively, PhyloVelo integrates these modalities by providing a lineage-constrained adaptation of RNA velocity, which estimates instantaneous cellular dynamics by identifying monotonically expressed genes along the lineage tree.[145]

When the identification of early fate priming is not the primary objective, analytical efforts often shift toward inferring broader differentiation hierarchies between distinct cell types. This is typically achieved by computing lineage similarity across cell types utilizing carefully designed normalization strategies, where higher similarity implies a closer developmental relationship.[142,146] Although broadly applicable across diverse lineage tracing datasets, the interpretation of these similarities requires caution; complex kinetic processes, barcode sparsity, and dropout can easily obscure the true developmental hierarchy. To resolve these ambiguities, CLADES bridges the analytical gap by simultaneously inferring all kinetic rates using temporally sampled clonal data.[147] Furthermore, several algorithms have been developed to deduce cell-type transitions directly from phylogenetic trees. For instance, PATH computes tree-based autocorrelation to map state transition kinetics in cancer models,[148] whereas Carta is designed to infer complex developmental hierarchies by balancing the structural complexity of the inferred hierarchy against discrepancies with the input data.[149]

Beyond mapping dynamics within the cell-state or cell-type space, a complementary analytical paradigm has emerged: clonal embedding. In this framework, the analytical focus shifts to the "clone space" itself. Clonal similarities are evaluated based on the aggregate state profiles of all cells within a given clone, and clones exhibiting congruent phenotypic profiles are subsequently grouped together within a low-dimensional embedding space. Data-driven approaches such as clone2vec and Clonotrace exemplify this strategy,[150,151] providing a highly effective method for organizing and exploring the diverse behaviors of thousands of distinct clones simultaneously.

Despite these recent advances, the computational integration of lineage and state information remains a formidable challenge. Resolving these analytical bottlenecks may require additional analytical paradigms that can account for challenges such as complex kinetics, heterogeneous sampling, barcode dropout, and the absence of ancestral states. As more abundant, high-fidelity lineage tracing data are generated across both model systems and human tissues, this influx of information will drive the emergence of more accurate and generalizable computational frameworks.

PRACTICAL DESIGN GUIDANCE

The three methodological pillars of lineage tracing—genetic markers, synthetic barcodes, and natural variants—each offer

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

Table 2. Summary and evaluation of modern lineage tracing approaches

Category	Genetic marker	Synthetic barcode	Natural variant
Applicable system	model organisms	model organisms; *ex vivo/in vitro*	any, including human
Labeling specificity	high (cell-type and time specific)	medium (ubiquitous or time specific)	none (spontaneous accumulation)
Lineage diversity	low (1–10)	high	high
Lineage recovery	high (minimal dropout)	variable (due to barcode dropout & homoplasy)	low/variable (depends on depth and mutation rate)
Temporal control	yes (inducible, e.g., CreERT2)	partial (delivery timing controllable; recording continuous)	no
Readout	imaging; or FACS sorting + single-cell sequencing	sequencing (destructive); *in situ* sequencing or FISH-based readout possible	sequencing of genomic DNA, mtDNA, or methylome (destructive)
Key strengths	• targeted fate mapping of defined cell populations • native spatial architecture preserved via imaging • direct functional interrogation via gene deletion (Flox) or cell ablation (DTR) • stable, heritable reporter signal; minimal silencing or dropout	• unbiased, population-scale clonal tracking without pre-defined markers • cumulative recorders enable deep temporal lineage-tree reconstruction • single sequencing run yields joint barcode and transcriptomic/epigenomic profiles	• applicable to human samples; no transgenesis required • zero perturbation; native cell states fully preserved • retrospective: full clonal history from a single biopsy, no pre-planned labeling • variants captured automatically in standard genomic/epigenomic sequencing
Key limitations	• low clonal diversity; within-population heterogeneity invisible • requires predefined driver lines; dual-recombinase strategies demand two validated lines and complex breeding • off-target labeling and background recombination from leaky Cre/Dre activity	• variable data fidelity due to barcode silencing, dropout, or homoplasy, and related artifacts • potential perturbation risks (e.g., genotoxicity, viral toxicity, or CRISPR off-target effects) • dissociation-based readout undersamples rare or slow-cycling clones • restricted to genetically manipulable systems	• no temporal control over labeling window • timing inference is challenging for mtDNA mutations and epimutations • resolution-cost tradeoff varies by variant: nuclear SNVs (high resolution, expensive); mtDNA (lower cost, lower resolution); epimutation (intermediate)

distinct trade-offs in labeling specificity, clonal resolution, and applicability (Table 2). Genetic markers provide high cell-type and temporal specificity during labeling, robust lineage recovery, and superior spatial resolution via imaging. However, they yield low lineage diversity, typically identifying only a limited number of distinct clones. Conversely, synthetic barcodes achieve massive lineage diversity and seamlessly link lineage histories to cell states through single-cell sequencing readouts. Their primary limitations include variable lineage recovery due to barcode dropout or homoplasy, alongside off-target editing risks. Crucially, both approaches require transgenesis, restricting their use to model organisms. Finally, natural variants eliminate the need for genetic engineering entirely, making them uniquely applicable to human samples. Although they capture high lineage diversity, their lineage recovery varies depending on the natural variant used and sequencing depth; furthermore, they lack temporal control and pose significant computational challenges for data interpretation. The marker-based approach excels in prospective state-fate mapping and functional interrogation, whereas synthetic barcodes and natural variants are well suited for profiling the heterogeneity of cell fate choices at single-cell resolution. Recognizing that each lineage tracing system has its niche, we provide a simplified guide to selecting the appropriate approach based on the specific biological question (Figure 5).

OPPORTUNITIES AND FUTURE DIRECTIONS

To advance the foundational tools of lineage tracing, future efforts must prioritize deeper, more precise recording capabilities. In model systems, the next generation of synthetic barcodes must enable deep, multi-generational phylogenies while strictly minimizing off-target editing and cellular toxicity. To facilitate superior state-to-fate mapping and systematically interrogate population-level heterogeneity, unprecedented spatiotemporal control over single-cell labeling is also essential. This precision can be achieved by directly integrating highly diverse synthetic barcodes with targeted driver-based systems. Translating these engineering advances into accessible, off-the-shelf model systems will be vital for democratizing deep investigation of normal development and tissue homeostasis. Conversely, achieving comparable precision in human samples requires the holistic integration

Please cite this article in press as: Kang et al., Evolving strategies for lineage tracing: Genetic markers, synthetic barcodes, and natural variants, Cell Stem Cell (2026), https://doi.org/10.1016/j.stem.2026.05.001

CellPress
OPEN ACCESS

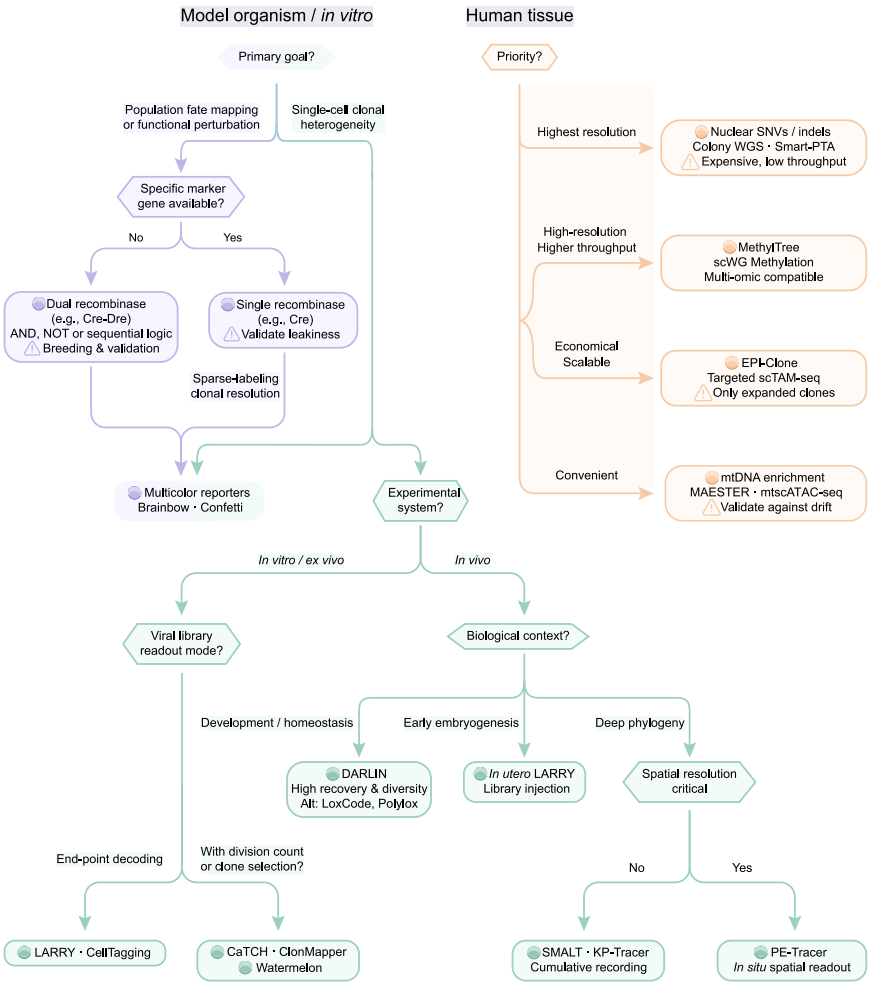

Figure 5. A flowchart for selecting appropriate lineage tracing technologies

This flowchart provides a systematic guide for choosing tracing systems based on the experimental model (model organism/*in vitro* versus human tissue), the primary biological goal, and the desired readout resolution. Recommended approaches are indicated by circular icons, while warning symbols highlight critical technical requirements, such as validating leakiness in recombinase systems, managing breeding and validation for dual drivers, or accounting for mtDNA drift.

of multiple natural variant modalities. Jointly leveraging somatic mutations, structural variations, and stable epimutations will be essential for reconstructing comprehensive, high-resolution evolutionary histories without transgenesis.

Beyond recording lineage architecture itself, future platforms must capture deeper, multidimensional resolution of cellular state. This requires moving beyond standard transcriptomics to incorporate co-profiling of the epigenome, direct mapping of physical cell-cell contacts, and logging of historical molecular signaling events. Seamlessly integrating continuous lineage trees with next-generation spatial profiling technologies will eventually illuminate how the physical microenvironment and localized niche interactions cooperatively prime cell fate decisions over time. Furthermore, coupling these multidimensional maps with clone-resolved functional perturbations will enable systematic clarification of gene function across distinct cell states within highly heterogeneous populations.

To fully harness these multi-layered datasets, computational frameworks must evolve in parallel. Advanced algorithms are urgently needed to seamlessly integrate continuous lineage phylogenies with spatial coordinates, multi-omics states, and multiplexed perturbation readouts. By mathematically synthesizing this wealth of information, next-generation computational models will pinpoint crucial regulatory networks. Ultimately, this will transition the field toward predictive modeling of cell fate, enabling forecasts of how specific genetic or pharmacological interventions will redirect a cell's developmental or regenerative trajectory.

Finally, although current tools have already unlocked unprecedented biological insights into cancer, aging, and normal development, the ultimate frontier remains routine clinical application and longitudinal disease monitoring. Achieving this requires directing high-resolution lineage analysis toward high-impact scenarios. Non-invasive tracking of somatic mutations and epimutations via liquid biopsies could enable early cancer detection and real-time monitoring of minimal residual disease. Resolving the phylogenetic architecture of circulating tumor DNA may anticipate therapy-resistant clones before clinical relapse. Similarly, tracking clonal hematopoiesis of indeterminate potential (CHIP) in aging populations offers a strategy for stratifying risk of hematologic malignancies and cardiovascular disease. Translating these approaches into the clinic requires overcoming inherent trade-offs between sequencing cost, phylogenetic precision, and cellular scalability. Developing cost-effective, targeted sequencing platforms that integrate stable epimutations with somatic mutations will position high-resolution lineage tracing as a scalable tool for precision medicine.

ACKNOWLEDGMENTS

This study was supported by the National Key Research & Development Program of China (2024YFA1803302 and 2023YFA1800700), the National Natural Science Foundation of China (82088101 and 32470700), the R&D Programs of Zhejiang Province (2024SSYS0034), the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM502), and the New Cornerstone Science Foundation. We thank Dr. Xiuxiu Liu for providing the fluorescence image in Figure 2F.

AUTHOR CONTRIBUTIONS

Z.K., H.C., and S.L. contributed equally to the literature investigation, original draft preparation, and figure generation. S.-W.W. and B.Z. conceived and designed the review and were responsible for the critical revision, commentary, and final editing of the manuscript.

DECLARATION OF INTERESTS

B.Z. is a member of the *Cell Stem Cell* Advisory Board.

REFERENCES

1. Beumer, J., and Clevers, H. (2021). Cell fate specification and differentiation in the adult mammalian intestine. Nat. Rev. Mol. Cell Biol. *22*, 39–53. https://doi.org/10.1038/s41580-020-0278-0.

2. Lytle, N.K., Barber, A.G., and Reya, T. (2018). Stem cell fate in cancer growth, progression and therapy resistance. Nat. Rev. Cancer *18*, 669–680. https://doi.org/10.1038/s41568-018-0056-x.

3. Dymecki, S.M., Ray, R.S., and Kim, J.C. (2010). Mapping cell fate and function using recombinase-based intersectional strategies. Methods Enzymol. *477*, 183–213. https://doi.org/10.1016/s0076-6879(10)77011-7.

4. Ferrell, J.E., Jr. (2012). Bistability, bifurcations, and Waddington's epigenetic landscape. Curr. Biol. *22*, R458–R466. https://doi.org/10.1016/j.cub.2012.03.045.

5. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D., and Klein, A.M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science *367*, eaaw3381. https://doi.org/10.1126/science.aaw3381.

6. Xie, L., Liu, H., You, Z., Wang, L., Li, Y., Zhang, X., Ji, X., He, H., Yuan, T., Zheng, W., et al. (2023). Comprehensive spatiotemporal mapping of single-cell lineages in developing mouse brain by CRISPR-based barcoding. Nat. Methods *20*, 1244–1255. https://doi.org/10.1038/s41592-023-01947-3.

7. Chen, W., Guillaume-Gentil, O., Rainer, P.Y., Gäbelein, C.G., Saelens, W., Gardeux, V., Klaeger, A., Dainese, R., Zachara, M., Zambelli, T., et al. (2022). Live-seq enables temporal transcriptomic recording of single cells. Nature *608*, 733–740. https://doi.org/10.1038/s41586-022-05046-9.

8. Barile, M., Imaz-Rosshandler, I., Inzani, I., Ghazanfar, S., Nichols, J., Marioni, J.C., Guibentif, C., and Göttgens, B. (2021). Coordinated changes in gene expression kinetics underlie both mouse and human erythroid maturation. Genome Biol. *22*, 197. https://doi.org/10.1186/s13059-021-02414-y.

9. Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. *100*, 64–119. https://doi.org/10.1016/0012-1606(83)90201-4.

10. Price, J., Turner, D., and Cepko, C. (1987). Lineage analysis in the vertebrate nervous system by retrovirus-mediated gene transfer. Proc. Natl. Acad. Sci. USA *84*, 156–160. https://doi.org/10.1073/pnas.84.1.156.

11. Turner, D.L., and Cepko, C.L. (1987). A common progenitor for neurons and glia persists in rat retina late in development. Nature *328*, 131–136. https://doi.org/10.1038/328131a0.

12. Kester, L., and van Oudenaarden, A. (2018). Single-Cell Transcriptomics Meets Lineage Tracing. Cell Stem Cell *23*, 166–179. https://doi.org/10.1016/j.stem.2018.04.014.

13. Wagner, D.E., and Klein, A.M. (2020). Lineage tracing meets single-cell omics: opportunities and challenges. Nat. Rev. Genet. *21*, 410–427. https://doi.org/10.1038/s41576-020-0223-2.

14. VanHorn, S., and Morris, S.A. (2021). Next-Generation Lineage Tracing and Fate Mapping to Interrogate Development. Dev. Cell *56*, 7–21. https://doi.org/10.1016/j.devcel.2020.10.021.

15. Rodriguez-Fraticelli, A.E., and Parreno, V. (2026). Charting single-cell lineages with synthetic and natural barcodes. Nat. Rev. Genet. Published online February 27, 2026. https://doi.org/10.1038/s41576-026-00943-5.

16. Kretzschmar, K., and Watt, F.M. (2012). Lineage tracing. Cell *148*, 33–45. https://doi.org/10.1016/j.cell.2012.01.002.

17. Woodworth, M.B., Girskis, K.M., and Walsh, C.A. (2017). Building a lineage from single cells: genetic techniques for cell lineage tracking. Nat. Rev. Genet. *18*, 230–244. https://doi.org/10.1038/nrg.2016.159.

18. Liu, K., Jin, H., and Zhou, B. (2020). Genetic lineage tracing with multiple DNA recombinases: A user's guide for conducting more precise cell fate mapping studies. J. Biol. Chem. *295*, 6413–6424. https://doi.org/10.1074/jbc.REV120.011631.

19. Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., and Chambon, P. (1996). Ligand-activated site-specific recombination in mice. Proc. Natl. Acad. Sci. USA *93*, 10887–10890. https://doi.org/10.1073/pnas.93.20.10887.

20. Anastassiadis, K., Fu, J., Patsch, C., Hu, S., Weidlich, S., Duerschke, K., Buchholz, F., Edenhofer, F., and Stewart, A.F. (2009). Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E. coli, mammalian cells and mice. Dis. Model. Mech. *2*, 508–515. https://doi.org/10.1242/dmm.003087.

21. Kim, J.S., Kolesnikov, M., Peled-Hajaj, S., Scheyltjens, I., Xia, Y., Trzebanski, S., Haimon, Z., Shemer, A., Lubart, A., Van Hove, H., et al. (2021). A Binary Cre Transgenic Approach Dissects Microglia and CNS Border-Associated Macrophages. Immunity *54*, 176–190.e7. https://doi.org/10.1016/j.immuni.2020.11.007.

22. Hirrlinger, J., Scheller, A., Hirrlinger, P.G., Kellert, B., Tang, W., Wehr, M.C., Goebbels, S., Reichenbach, A., Sprengel, R., Rossner, M.J., et al. (2009). Split-Cre Complementation Indicates Coincident Activity of Different Genes In Vivo. PLoS One *4*, e4286. https://doi.org/10.1371/journal.pone.0004286.

23. Wang, P., Chen, T., Sakurai, K., Han, B.X., He, Z., Feng, G., and Wang, F. (2012). Intersectional Cre Driver Lines Generated Using Split-Intein Mediated Split-Cre Reconstitution. Sci. Rep. *2*, 497. https://doi.org/10.1038/srep00497.

24. Salwig, I., Spitznagel, B., Vazquez-Armendariz, A.I., Khalooghi, K., Guenther, S., Herold, S., Szibor, M., and Braun, T. (2019). Bronchioalveolar stem cells are a main source for regeneration of distal lung epithelia in vivo. EMBO J. *38*, e102099. https://doi.org/10.15252/embj.2019102099.

25. Madisen, L., Garner, A.R., Shimaoka, D., Chuong, A.S., Klapoetke, N.C., Li, L., van der Bourg, A., Niino, Y., Egolf, L., Monetti, C., et al. (2015). Transgenic Mice for Intersectional Targeting of Neural Sensors and Effectors with High Specificity and Performance. Neuron *85*, 942–958. https://doi.org/10.1016/j.neuron.2015.02.022.

26. Fenno, L.E., Ramakrishnan, C., Kim, Y.S., Evans, K.E., Lo, M., Vesuna, S., Inoue, M., Cheung, K.Y.M., Yuen, E., Pichamoorthy, N., et al. (2020). Comprehensive Dual- and Triple-Feature Intersectional Single-Vector Delivery of Diverse Functional Payloads to Cells of Behaving Mammals. Neuron *107*, 836–853.e11. https://doi.org/10.1016/j.neuron.2020.06.003.

27. Dymecki, S.M., and Kim, J.C. (2007). Molecular Neuroanatomy's "Three Gs": A Primer. Neuron *54*, 17–34. https://doi.org/10.1016/j.neuron.2007.03.009.

28. He, L., Li, Y., Li, Y., Pu, W., Huang, X., Tian, X., Wang, Y., Zhang, H., Liu, Q., Zhang, L., et al. (2017). Enhancing the precision of genetic lineage tracing using dual recombinases. Nat. Med. *23*, 1488–1498. https://doi.org/10.1038/nm.4437.

29. Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., et al. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell *143*, 134–144. https://doi.org/10.1016/j.cell.2010.09.016.

30. Livet, J., Weissman, T.A., Kang, H., Draft, R.W., Lu, J., Bennis, R.A., Sanes, J.R., and Lichtman, J.W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature *450*, 56–62. https://doi.org/10.1038/nature06293.

31. Zong, H., Espinosa, J.S., Su, H.H., Muzumdar, M.D., and Luo, L. (2005). Mosaic analysis with double markers in mice. Cell *121*, 479–492. https://doi.org/10.1016/j.cell.2005.02.012.

32. Kong, W., Biddy, B.A., Kamimoto, K., Amrute, J.M., Butka, E.G., and Morris, S.A. (2020). CellTagging: combinatorial indexing to simultaneously map lineage and identity at single-cell resolution. Nat. Protoc. *15*, 750–772. https://doi.org/10.1038/s41596-019-0247-2.

33. Oren, Y., Tsabar, M., Cuoco, M.S., Amir-Zilberstein, L., Cabanos, H.F., Hütter, J.C., Hu, B., Thakore, P.I., Tabaka, M., Fulco, C.P., et al. (2021). Cycling cancer persister cells arise from lineages with distinct programs. Nature *596*, 576–582. https://doi.org/10.1038/s41586-021-03796-6.

34. Umkehrer, C., Holstein, F., Formenti, L., Jude, J., Froussios, K., Neumann, T., Cronin, S.M., Haas, L., Lipp, J.J., Burkard, T.R., et al. (2021). Isolating live cell clones from barcoded populations using CRISPRa-inducible reporters. Nat. Biotechnol. *39*, 174–178. https://doi.org/10.1038/s41587-020-0614-0.

35. Gutierrez, C., Al'Khafaji, A.M., Brenner, E., Johnson, K.E., Gohil, S.H., Lin, Z., Knisbacher, B.A., Durrett, R.E., Li, S., Parvin, S., et al. (2021). Multifunctional barcoding with ClonMapper enables high-resolution study of clonal dynamics during tumor evolution and treatment. Nat. Cancer *2*, 758–772. https://doi.org/10.1038/s43018-021-00222-8.

36. Pei, W., Feyerabend, T.B., Rössler, J., Wang, X., Postrach, D., Busch, K., Rode, I., Klapproth, K., Dietlein, N., Quedenau, C., et al. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. Nature *548*, 456–460. https://doi.org/10.1038/nature23653.

37. Weber, T.S., Biben, C., Miles, D.C., Glaser, S.P., Tomei, S., Lin, C.Y., Kueh, A., Pal, M., Zhang, S., Tam, P.P.L., et al. (2025). LoxCode in vivo barcoding reveals epiblast clonal fate bias to fetal organs. Cell *188*, 3882–3896.e19. https://doi.org/10.1016/j.cell.2025.04.026.

38. Urbanus, J., Cosgrove, J., Beltman, J.B., Elhanati, Y., Moral, R.A., Conrad, C., van Heijst, J.W., Tubeuf, E., Velds, A., Kok, L., et al. (2023). DRAG in situ barcoding reveals an increased number of HSPCs contributing to myelopoiesis with age. Nat. Commun. *14*, 2184. https://doi.org/10.1038/s41467-023-37167-8.

39. Sun, J., Ramos, A., Chapman, B., Johnnidis, J.B., Le, L., Ho, Y.J., Klein, A., Hofmann, O., and Camargo, F.D. (2014). Clonal dynamics of native haematopoiesis. Nature *514*, 322–327. https://doi.org/10.1038/nature13824.

40. Feng, J., Pucella, J.N., Jang, G., Alcántara-Hernández, M., Upadhaya, S., Adams, N.M., Khodadadi-Jamayran, A., Lau, C.M., Stoeckius, M., Hao, S., et al. (2022). Clonal lineage tracing reveals shared origin of conventional and plasmacytoid dendritic cells. Immunity *55*, 405–422.e11. https://doi.org/10.1016/j.immuni.2022.01.016.

41. Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N., and Junker, J.P. (2018). Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. Nat. Biotechnol. *36*, 469–473. https://doi.org/10.1038/nbt.4124.

42. Raj, B., Wagner, D.E., McKenna, A., Pandey, S., Klein, A.M., Shendure, J., Gagnon, J.A., and Schier, A.F. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. Nat. Biotechnol. *36*, 442–450. https://doi.org/10.1038/nbt.4103.

43. Alemany, A., Florescu, M., Baron, C.S., Peterson-Maduro, J., and van Oudenaarden, A. (2018). Whole-organism clone tracing using single-cell sequencing. Nature *556*, 108–112. https://doi.org/10.1038/nature25969.

44. Li, L., Bowling, S., McGeary, S.E., Yu, Q., Lemke, B., Alcedo, K., Jia, Y., Liu, X., Ferreira, M., Klein, A.M., et al. (2023). A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. Cell *186*, 5183–5199.e22. https://doi.org/10.1016/j.cell.2023.09.019.

45. Yang, D., Jones, M.G., Naranjo, S., Rideout, W.M., Min, K.H.J., Ho, R., Wu, W., Replogle, J.M., Page, J.L., Quinn, J.J., et al. (2022). Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. Cell *185*, 1905–1923.e25. https://doi.org/10.1016/j.cell.2022.04.015.

46. Simeonov, K.P., Byrns, C.N., Clark, M.L., Norgard, R.J., Martin, B., Stanger, B.Z., Shendure, J., McKenna, A., and Lengner, C.J. (2021). Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. Cancer Cell *39*, 1150–1162.e9. https://doi.org/10.1016/j.ccell.2021.05.005.

47. Loveless, T.B., Grotts, J.H., Schechter, M.W., Forouzmand, E., Carlson, C.K., Agahi, B.S., Liang, G., Ficht, M., Liu, B., Xie, X., et al. (2021). Lineage tracing and analog recording in mammalian cells by single-site DNA writing. Nat. Chem. Biol. *17*, 739–747. https://doi.org/10.1038/s41589-021-00769-8.

48. Fang, W., Bell, C.M., Sapirstein, A., Asami, S., Leeper, K., Zack, D.J., Ji, H., and Kalhor, R. (2022). Quantitative fate mapping: A general framework for analyzing progenitor state dynamics via retrospective lineage barcoding. Cell *185*, 4604–4620.e32. https://doi.org/10.1016/j.cell.2022.10.028.

49. He, Z., Maynard, A., Jain, A., Gerber, T., Petri, R., Lin, H.C., Santel, M., Ly, K., Dupré, J.S., Sidow, L., et al. (2022). Lineage recording in human cerebral organoids. Nat. Methods *19*, 90–99. https://doi.org/10.1038/s41592-021-01344-8.

50. Yang, J., Hou, L., Wang, X., Zhang, N., Bian, Y., Lu, Z., Chen, Y., Xie, D., Fang, Y., Wang, K., et al. (2025). Spatiotemporal lineage mapping of tumor immune escape with eTRACER. Prerpint at. bioRxiv. https://doi.org/10.1101/2025.08.06.668639.

51. Koblan, L.W., Yost, K.E., Zheng, P., Colgan, W.N., Jones, M.G., Yang, D., Kumar, A., Sandhu, J., Schnell, A., Sun, D., et al. (2025). High-resolution spatial mapping of cell state and lineage dynamics in vivo with PEtracer. Science *390*, eadx3800. https://doi.org/10.1126/science.adx3800.

52. Liu, K., Deng, S., Ye, C., Yao, Z., Wang, J., Gong, H., Liu, L., and He, X. (2021). Mapping single-cell-resolution cell phylogeny reveals cell population dynamics during organ development. Nat. Methods *18*, 1506–1514. https://doi.org/10.1038/s41592-021-01325-x.

53. Loveless, T.B., Carlson, C.K., Dentzel Helmy, C.A., Hu, V.J., Ross, S.K., Demelo, M.C., Murtaza, A., Liang, G., Ficht, M., Singhai, A., et al. (2025). Open-ended molecular recording of sequential cellular events into DNA. Nat. Chem. Biol. *21*, 512–521. https://doi.org/10.1038/s41589-024-01764-5.

54. Liao, H., Choi, J., and Shendure, J. (2024). Molecular recording using DNA Typewriter. Nat. Protoc. *19*, 2833–2862. https://doi.org/10.1038/s41596-024-01003-0.

55. Winter, E., Emiliani, F., Cook, A., Abderrahim, A., and McKenna, A.H. (2025). BASELINE: A CRISPR Base Editing Platform for Mammalian-Scale Single-Cell Lineage Tracing. Preprint at bioRxiv. https://doi.org/10.1101/2025.03.19.644238.

56. Lodato, M.A., Woodworth, M.B., Lee, S., Evrony, G.D., Mehta, B.K., Karger, A., Lee, S., Chittenden, T.W., D'Gama, A.M., Cai, X., et al. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. Science *350*, 94–98. https://doi.org/10.1126/science.aab1785.

57. Chen, M., Fu, R., Chen, Y., Li, L., and Wang, S.W. (2025). High-resolution, noninvasive single-cell lineage tracing in mice and humans based on DNA methylation epimutations. Nat. Methods *22*, 488–498. https://doi.org/10.1038/s41592-024-02567-1.

58. Scherer, M., Singh, I., Braun, M.M., Szu-Tu, C., Sanchez Sanchez, P., Lindenhofer, D., Jakobsen, N.A., Körber, V., Kardorff, M., Nitsch, L., et al. (2025). Clonal tracing with somatic epimutations reveals dynamics of blood ageing. Nature *643*, 478–487. https://doi.org/10.1038/s41586-025-09041-8.

59. Gabbutt, C., Duran-Ferrer, M., Grant, H.E., Mallo, D., Nadeu, F., Househam, J., Villamor, N., Müller, M., Heath, S., Raineri, E., et al. (2025). Fluctuating DNA methylation tracks cancer evolution at clinical scale. Nature *645*, 764–773. https://doi.org/10.1038/s41586-025-09374-4.

60. Miller, T.E., Lareau, C.A., Verga, J.A., DePasquale, E.A.K., Liu, V., Ssozi, D., Sandor, K., Yin, Y., Ludwig, L.S., El Farran, C.A., et al. (2022). Mitochondrial variant enrichment from high-throughput single-cell RNA sequencing resolves clonal populations. Nat. Biotechnol. *40*, 1030–1034. https://doi.org/10.1038/s41587-022-01210-8.

61. Lareau, C.A., Ludwig, L.S., Muus, C., Gohil, S.H., Zhao, T., Chiang, Z., Pelka, K., Verboon, J.M., Luo, W., Christian, E., et al. (2021). Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. Nat. Biotechnol. *39*, 451–461. https://doi.org/10.1038/s41587-020-0645-6.

62. McKenna, A., and Gagnon, J.A. (2019). Recording development with single cell dynamic lineage tracing. Development *146*, dev169730. https://doi.org/10.1242/dev.169730.

63. Fu, R., Chen, M., and Wang, S.W. (2026). DNA methylation meets lineage tracing: History, recent progress, and future directions. Quant. Biol. *14*, e70017. https://doi.org/10.1002/qub2.70017.

64. Li, S., Wang, K., Wang, X., and Hu, Z. (2026). Single-cell mitochondrial lineage tracing: Opportunities and challenges. Quant. Biol. *14*, e70018. https://doi.org/10.1002/qub2.70018.

65. Shao, D.D., Kriz, A.J., Snellings, D.A., Zhou, Z., Zhao, Y., Enyenihi, L., and Walsh, C. (2025). Advances in single-cell DNA sequencing enable insights into human somatic mosaicism. Nat. Rev. Genet. *26*, 761–774. https://doi.org/10.1038/s41576-025-00832-3.

66. Metzger, D., Clifford, J., Chiba, H., and Chambon, P. (1995). Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. Proc. Natl. Acad. Sci. USA *92*, 6991–6995. https://doi.org/10.1073/pnas.92.15.6991.

67. Indra, A.K., Warot, X., Brocard, J., Bornert, J.M., Xiao, J.H., Chambon, P., and Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases. Nucleic Acids Res. *27*, 4324–4327. https://doi.org/10.1093/nar/27.22.4324.

68. Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., Mehler, M.F., Conway, S.J., Ng, L.G., Stanley, E.R., et al. (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. Science *330*, 841–845. https://doi.org/10.1126/science.1194637.

69. Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J., et al. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003–1007. https://doi.org/10.1038/nature06196.

70. Tata, P.R., Mou, H., Pardo-Saganta, A., Zhao, R., Prabhu, M., Law, B.M., Vinarsky, V., Cho, J.L., Breton, S., Sahay, A., et al. (2013). Dedifferentiation of committed epithelial cells into stem cells in vivo. Nature *503*, 218–223. https://doi.org/10.1038/nature12777.

71. Thorel, F., Népote, V., Avril, I., Kohno, K., Desgraz, R., Chera, S., and Herrera, P.L. (2010). Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. Nature *464*, 1149–1154. https://doi.org/10.1038/nature08894.

72. Guo, W., Zhang, X., Li, L., Shao, P., Liang, C., Zhang, H., Liu, K., Wang, S., Peng, Y., Luo, J., et al. (2024). JAK/STAT signaling maintains an intermediate cell population during prostate basal cell fate determination. Nat. Genet. *56*, 2776–2789. https://doi.org/10.1038/s41588-024-01979-1.

73. Ellison, G.M., Vicinanza, C., Smith, A.J., Aquila, I., Leone, A., Waring, C.D., Henning, B.J., Stirparo, G.G., Papait, R., Scarfò, M., et al. (2013). Adult c-kit(pos) cardiac stem cells are necessary and sufficient for functional cardiac regeneration and repair. Cell *154*, 827–842. https://doi.org/10.1016/j.cell.2013.07.039.

74. Sultana, N., Zhang, L., Yan, J., Chen, J., Cai, W., Razzaque, S., Jeong, D., Sheng, W., Bu, L., Xu, M., et al. (2015). Resident c-kit(+) cells in the heart are not cardiac stem cells. Nat. Commun. *6*, 8701. https://doi.org/10.1038/ncomms9701.

75. van Berlo, J.H., Kanisicak, O., Maillet, M., Vagnozzi, R.J., Karch, J., Lin, S.C.J., Middleton, R.C., Marbán, E., and Molkentin, J.D. (2014). c-kit+ cells minimally contribute cardiomyocytes to the heart. Nature *509*, 337–341. https://doi.org/10.1038/nature13309.

76. Li, Y., He, L., Huang, X., Bhaloo, S.I., Zhao, H., Zhang, S., Pu, W., Tian, X., Li, Y., Liu, Q., et al. (2018). Genetic Lineage Tracing of Nonmyocyte Population by Dual Recombinases. Circulation *138*, 793–805. https://doi.org/10.1161/circulationaha.118.034250.

77. Zhu, Q., Yin, F., Qin, J., Shi, W., Liu, Y., Zhao, Y., Wang, J., Zhang, L., Fan, A., Cao, D., et al. (2025). Procr^{+} chondroprogenitors sense mechanical stimuli to govern articular cartilage maintenance and regeneration. Cell *188*, 5194–5211.e16. https://doi.org/10.1016/j.cell.2025.06.036.

78. Schönhuber, N., Seidler, B., Schuck, K., Veltkamp, C., Schachtler, C., Zukowska, M., Eser, S., Feyerabend, T.B., Paul, M.C., Eser, P., et al. (2014). A next-generation dual-recombinase system for time- and host-specific targeting of pancreatic cancer. Nat. Med. *20*, 1340–1347. https://doi.org/10.1038/nm.3646.

79. He, L., Pu, W., Liu, X., Zhang, Z., Han, M., Li, Y., Huang, X., Han, X., Li, Y., Liu, K., et al. (2021). Proliferation tracing reveals regional hepatocyte generation in liver homeostasis and repair. Science *371*, eabc4346. https://doi.org/10.1126/science.abc4346.

80. Bradley, L.A., Young, A., Li, H., Billcheck, H.O., and Wolf, M.J. (2021). Loss of Endogenously Cycling Adult Cardiomyocytes Worsens Myocardial Function. Circ. Res. *128*, 155–168. https://doi.org/10.1161/circresaha.120.318277.

81. Cai, D., Cohen, K.B., Luo, T., Lichtman, J.W., and Sanes, J.R. (2013). Improved tools for the Brainbow toolbox. Nat. Methods *10*, 540–547. https://doi.org/10.1038/nmeth.2450.

82. Dumas, L., Clavreul, S., Michon, F., and Loulier, K. (2022). Multicolor strategies for investigating clonal expansion and tissue plasticity. Cell. Mol. Life Sci. *79*, 141. https://doi.org/10.1007/s00018-021-04077-1.

83. Lopez-Garcia, C., Klein, A.M., Simons, B.D., and Winton, D.J. (2010). Intestinal stem cell replacement follows a pattern of neutral drift. Science *330*, 822–825. https://doi.org/10.1126/science.1196236.

84. Clayton, E., Doupé, D.P., Klein, A.M., Winton, D.J., Simons, B.D., and Jones, P.H. (2007). A single type of progenitor cell maintains normal epidermis. Nature *446*, 185–189. https://doi.org/10.1038/nature05574.

85. Doupé, D.P., Alcolea, M.P., Roshan, A., Zhang, G., Klein, A.M., Simons, B.D., and Jones, P.H. (2012). A single progenitor population switches behavior to maintain and repair esophageal epithelium. Science *337*, 1091–1093. https://doi.org/10.1126/science.1218835.

86. Klein, A.M., Nakagawa, T., Ichikawa, R., Yoshida, S., and Simons, B.D. (2010). Mouse germ line stem cells undergo rapid and stochastic turnover. Cell Stem Cell *7*, 214–224. https://doi.org/10.1016/j.stem.2010.05.017.

87. Chen, P., Wang, W., Liu, R., Lyu, J., Zhang, L., Li, B., Qiu, B., Tian, A., Jiang, W., Ying, H., et al. (2022). Olfactory sensory experience regulates gliomagenesis via neuronal IGF1. Nature *606*, 550–556. https://doi.org/10.1038/s41586-022-04719-9.

88. Shen, Z., Lin, Y., Yang, J., Jörg, D.J., Peng, Y., Zhang, X., Xu, Y., Hernandez, L., Ma, J., Simons, B.D., et al. (2021). Distinct progenitor behavior underlying neocortical gliogenesis related to tumorigenesis. Cell Rep. *34*, 108853. https://doi.org/10.1016/j.celrep.2021.108853.

89. Hippenmeyer, S., Youn, Y.H., Moon, H.M., Miyamichi, K., Zong, H., Wynshaw-Boris, A., and Luo, L. (2010). Genetic mosaic dissection of Lis1 and Ndel1 in neuronal migration. Neuron *68*, 695–709. https://doi.org/10.1016/j.neuron.2010.09.027.

90. Liu, C., Sage, J.C., Miller, M.R., Verhaak, R.G.W., Hippenmeyer, S., Vogel, H., Foreman, O., Bronson, R.T., Nishiyama, A., Luo, L., et al. (2011). Mosaic analysis with double markers reveals tumor cell of origin in glioma. Cell *146*, 209–221. https://doi.org/10.1016/j.cell.2011.06.014.

91. Calaminus, S.D.J., Guitart, A.V., Sinclair, A., Schachtner, H., Watson, S.P., Holyoake, T.L., Kranc, K.R., and Machesky, L.M. (2012). Lineage tracing of Pf4-Cre marks hematopoietic stem cells and their progeny. PLoS One *7*, e51361. https://doi.org/10.1371/journal.pone.0051361.

92. Pertuy, F., Aguilar, A., Strassel, C., Eckly, A., Freund, J.N., Duluc, I., Gachet, C., Lanza, F., and Léon, C. (2015). Broader expression of the mouse platelet factor 4-cre transgene beyond the megakaryocyte lineage. J. Thromb. Haemost. *13*, 115–125. https://doi.org/10.1111/jth.12784.

93. Rodriguez-Fraticelli, A.E., Weinreb, C., Wang, S.W., Migueles, R.P., Jankovic, M., Usart, M., Klein, A.M., Lowell, S., and Camargo, F.D. (2020). Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. Nature *583*, 585–589. https://doi.org/10.1038/s41586-020-2503-6.

94. Rivera-Gonzalez, G.C., Butka, E.G., Gonzalez, C.E., Mintz, R.L., Kleb, S.S., Josephson, V., Kong, W., Jindal, K., Kamimoto, K., Shook, B.A., et al. (2025). Comparative single-cell lineage tracing identifies distinct adipocyte precursor dynamics in skin and inguinal fat. Cell Stem Cell *32*, 1267–1284.e8. https://doi.org/10.1016/j.stem.2025.07.004.

95. de Haan, S., He, J., Corbat, A.A., Belicova, L., Ratz, M., Vinsland, E., Frisén, J., Kelley, M.W., and Andersson, E.R. (2025). Ectoderm barcoding reveals neural and cochlear compartmentalization. Science *388*, 60–68. https://doi.org/10.1126/science.adq9248.

96. Feng, J., Jang, G., Esteva, E., Adams, N.M., Jin, H., and Reizis, B. (2024). Clonal barcoding of endogenous adult hematopoietic stem cells reveals a spectrum of lineage contributions. Proc. Natl. Acad. Sci. USA *121*, e2317929121. https://doi.org/10.1073/pnas.2317929121.

97. Pei, W., Shang, F., Wang, X., Fanti, A.K., Greco, A., Busch, K., Klapproth, K., Zhang, Q., Quedenau, C., Sauer, S., et al. (2020). Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding. Cell Stem Cell *27*, 383–395.e8. https://doi.org/10.1016/j.stem.2020.07.018.

98. McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., and Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science *353*, aaf7907. https://doi.org/10.1126/science.aaf7907.

99. Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P., and Church, G.M. (2018). Developmental barcoding of whole mouse via homing CRISPR. Science *361*, eaat9804. https://doi.org/10.1126/science.aat9804.

100. Hwang, B., Lee, W., Yum, S.Y., Jeon, Y., Cho, N., Jang, G., and Bang, D. (2019). Lineage tracing using a Cas9-deaminase barcoding system targeting endogenous L1 elements. Nat. Commun. *10*, 1234. https://doi.org/10.1038/s41467-019-09203-z.

101. He, X., Xia, F.N., Liu, K., Wang, J., Liu, Z., and Li, A. (2024). Mapping zygote-to-adult developmental cell phylogeny in Arabidopsis thaliana reveals a 3-cell rule of branching. Preprint at Res. Sq. https://doi.org/10.21203/rs.3.rs-5203004/v1.

102. Lu, Z., Mo, S., Xie, D., Zhai, X., Deng, S., Zhou, K., Wang, K., Kang, X., Zhang, H., Tong, J., et al. (2024). Polyclonal-to-monoclonal transition in colorectal precancerous evolution. Nature *636*, 233–240. https://doi.org/10.1038/s41586-024-08133-1.

103. Sadien, I.D., Adler, S., Mehmed, S., Bailey, S., Sawle, A., Couturier, D.L., Eldridge, M., Adams, D.J., Kemp, R., Lourenço, F.C., et al. (2024). Polyclonality overcomes fitness barriers in Apc-driven tumorigenesis. Nature *634*, 1196–1203. https://doi.org/10.1038/s41586-024-08053-0.

104. Bowling, S., Sritharan, D., Osorio, F.G., Nguyen, M., Cheung, P., Rodriguez-Fraticelli, A., Patel, S., Yuan, W.C., Fujiwara, Y., Li, B.E., et al. (2020). An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. Cell *181*, 1410–1422.e27. https://doi.org/10.1016/j.cell.2020.04.048.

105. Jones, M.G., Khodaverdian, A., Quinn, J.J., Chan, M.M., Hussmann, J.A., Wang, R., Xu, C., Weissman, J.S., and Yosef, N. (2020). Inference of single-cell phylogenies from lineage tracing data using Cassiopeia. Genome Biol. *21*, 92. https://doi.org/10.1186/s13059-020-02000-8.

106. Pan, X., Li, H., Putta, P., and Zhang, X. (2023). LinRace: cell division history reconstruction of single cells using paired lineage barcode and gene expression data. Nat. Commun. *14*, 8388. https://doi.org/10.1038/s41467-023-44173-3.

107. Zafar, H., Lin, C., and Bar-Joseph, Z. (2020). Single-cell lineage tracing by integrating CRISPR-Cas9 mutations with transcriptomic data. Nat. Commun. *11*, 3055. https://doi.org/10.1038/s41467-020-16821-5.

108. Shaffer, S.M., Emert, B.L., Reyes Hueros, R.A., Cote, C., Harmange, G., Schaff, D.L., Sizemore, A.E., Gupte, R., Torre, E., Singh, A., et al. (2020). Memory Sequencing Reveals Heritable Single-Cell Gene Expression Programs Associated with Distinct Cellular Behaviors. Cell *182*, 947–959.e17. https://doi.org/10.1016/j.cell.2020.07.003.

109. Liu, M., Yue, Y., Chen, X., Xian, K., Dong, C., Shi, M., Xiong, H., Tian, K., Li, Y., Zhang, Q.C., et al. (2025). Genome-coverage single-cell histone modifications for embryo lineage tracing. Nature *640*, 828–839. https://doi.org/10.1038/s41586-025-08656-1.

110. Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., et al. (2011). Tumour evolution inferred by single-cell sequencing. Nature *472*, 90–94. https://doi.org/10.1038/nature09807.

111. Stubbington, M.J.T., Lönnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G., and Teichmann, S.A. (2016). T cell fate and clonality inference from single-cell transcriptomes. Nat. Methods *13*, 329–332. https://doi.org/10.1038/nmeth.3800.

112. Spencer Chapman, M., Ranzoni, A.M., Myers, B., Williams, N., Coorens, T.H.H., Mitchell, E., Butler, T., Dawson, K.J., Hooks, Y., Moore, L., et al. (2021). Lineage tracing of human development through somatic mutations. Nature *595*, 85–90. https://doi.org/10.1038/s41586-021-03548-6.

113. Bizzotto, S., Dou, Y., Ganz, J., Doan, R.N., Kwon, M., Bohrson, C.L., Kim, S.N., Bae, T., Abyzov, A., NIMH Brain Somatic Mosaicism Network, et al. (2021). Landmarks of human embryonic development inscribed in somatic mutations. Science *371*, 1249–1253. https://doi.org/10.1126/science.abe1544.

114. Park, S., Mali, N.M., Kim, R., Choi, J.W., Lee, J., Lim, J., Park, J.M., Park, J.W., Kim, D., Kim, T., et al. (2021). Clonal dynamics in early human embryogenesis inferred from somatic mutation. Nature *597*, 393–397. https://doi.org/10.1038/s41586-021-03786-8.

115. Ju, Y.S., Martincorena, I., Gerstung, M., Petljak, M., Alexandrov, L.B., Rahbari, R., Wedge, D.C., Davies, H.R., Ramakrishna, M., Fullam, A., et al. (2017). Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. Nature *543*, 714–718. https://doi.org/10.1038/nature21703.

116. Coorens, T.H.H., Moore, L., Robinson, P.S., Sanghvi, R., Christopher, J., Hewinson, J., Przybilla, M.J., Lawson, A.R.J., Spencer Chapman, M., Cagan, A., et al. (2021). Extensive phylogenies of human development inferred from somatic mutations. Nature *597*, 387–392. https://doi.org/10.1038/s41586-021-03790-y.

117. Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K.J., Mende, N., Calderbank, E.F., Jung, H., Mitchell, T., Coorens, T.H.H., Spencer, D.H., et al. (2022). Clonal dynamics of haematopoiesis across the human lifespan. Nature *606*, 343–350. https://doi.org/10.1038/s41586-022-04786-y.

118. Gonzalez-Pena, V., Natarajan, S., Xia, Y., Klein, D., Carter, R., Pang, Y., Shaner, B., Annu, K., Putnam, D., Chen, W., et al. (2021). Accurate genomic variant detection in single cells with primary template-directed amplification. Proc. Natl. Acad. Sci. USA *118*, e2024176118. https://doi.org/10.1073/pnas.2024176118.

119. Prieto, T., Yuan, D.J., Zinno, J., Hughes, C., Midler, N., Kao, S., Huuhtanen, J., Raviram, R., Fotopoulou, F., Ruthen, N., et al. (2025). Large-scale single-cell phylogenetic mapping of clonal evolution in the human aging esophagus. Preprint at bioRxiv. https://doi.org/10.1101/2025.10.11.681805.

120. Dou, J., Tan, Y., Kock, K.H., Wang, J., Cheng, X., Tan, L.M., Han, K.Y., Hon, C.C., Park, W.Y., Shin, J.W., et al. (2024). Single-nucleotide variant calling in single-cell sequencing data with Monopogen. Nat. Biotechnol. *42*, 803–812. https://doi.org/10.1038/s41587-023-01873-x.

121. Yang, Z., Yao, M., Yang, Q., Du, Y., Lu, J., Wu, X., Lin, J., Qian, Z., Hu, S., Xia, Y., et al. (2026). Detection of Somatic Point Mutations Directly from Spatial Transcriptomics Enables in vivo Spatiotemporal Lineage Tracing. Preprint at bioRxiv. https://doi.org/10.64898/2026.02.04.703493.

122. Xu, J., Nuno, K., Litzenburger, U.M., Qi, Y., Corces, M.R., Majeti, R., and Chang, H.Y. (2019). Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. eLife *8*, e45105. https://doi.org/10.7554/eLife.45105.

123. Ludwig, L.S., Lareau, C.A., Ulirsch, J.C., Christian, E., Muus, C., Li, L.H., Pelka, K., Ge, W., Oren, Y., Brack, A., et al. (2019). Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. Cell *176*, 1325–1339.e22. https://doi.org/10.1016/j.cell.2019.01.022.

124. Wang, X., Wang, K., Zhang, W., Tang, Z., Zhang, H., Cheng, Y., Zhou, D., Zhang, C., Zhong, W.Z., Ma, Q., et al. (2025). Clonal expansion dictates the efficacy of mitochondrial lineage tracing in single cells. Genome Biol. *26*, 70. https://doi.org/10.1186/s13059-025-03540-7.

125. Campbell, P., Chapman, M.S., Przybilla, M., Lawson, A., Mitchell, E., Dawson, K., Williams, N., Harvey, L., Ranzoni, A.M., Cvejic, A., et al. (2023). Mitochondrial mutation, drift and selection during human development and ageing. Preprint at Res. Sq. https://doi.org/10.21203/rs.3.rs-3083262/v1.

126. Gao, T., Weng, C., Johnson, I., Poeschla, M., Gudera, J., King, E., Rouya, C., Donovan, A., Bourke, L., Shao, Y., et al. (2026). Modeling mitochondrial inheritance enables high-precision single-cell lineage tracing in humans. Preprint at bioRxiv. https://doi.org/10.64898/2026.02.12.705660.

127. Lareau, C.A., Chapman, M.S., Penter, L., Nawy, T., Pe'er, D., and Ludwig, L.S. (2024). Artifacts in single-cell mitochondrial DNA mutation analyses misinform phylogenetic inference. Preprint at bioRxiv. https://doi.org/10.1101/2024.07.28.605517.

128. Ushijima, T., Watanabe, N., Okochi, E., Kaneda, A., Sugimura, T., and Miyamoto, K. (2003). Fidelity of the methylation pattern and its variation in the genome. Genome Res. *13*, 868–874. https://doi.org/10.1101/gr.969603.

129. Gabbutt, C., Schenck, R.O., Weisenberger, D.J., Kimberley, C., Berner, A., Househam, J., Lakatos, E., Robertson-Tessi, M., Martin, I., Patel, R., et al. (2022). Fluctuating methylation clocks for cell lineage tracing at high temporal resolution in human tissues. Nat. Biotechnol. *40*, 720–730. https://doi.org/10.1038/s41587-021-01109-w.

130. Gaiti, F., Chaligne, R., Gu, H., Brand, R.M., Kothen-Hill, S., Schulman, R.C., Grigorev, K., Risso, D., Kim, K.T., Pastore, A., et al. (2019). Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. Nature *569*, 576–580. https://doi.org/10.1038/s41586-019-1198-z.

131. Crofts, S.J.C., Grenko, C.M., Robertson, N.A., Kirschner, K., and Chandra, T. (2025). Inference of clonal hematopoiesis using the collective behaviour of DNA methylation states. Preprint at bioRxiv. https://doi.org/10.1101/2025.11.18.689122.

132. Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., Wu, X., Wen, L., Tang, F., Huang, Y., et al. (2016). Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res. *26*, 304–319. https://doi.org/10.1038/cr.2016.23.

133. Jindal, K., Adil, M.T., Yamaguchi, N., Yang, X., Wang, H.C., Kamimoto, K., Rivera-Gonzalez, G.C., and Morris, S.A. (2024). Single-cell lineage capture across genomic modalities with CellTag-multi reveals fate-specific gene regulatory changes. Nat. Biotechnol. *42*, 946–959. https://doi.org/10.1038/s41587-023-01931-4.

134. Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., et al. (2022). Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell *185*, 1777–1792.e21. https://doi.org/10.1016/j.cell.2022.04.003.

135. Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82. https://doi.org/10.1126/science.aaf2403.

136. Jones, M.G., Sun, D., Min, K.H.J., Colgan, W.N., Tian, L., Weir, J.A., Chen, V.Z., Koblan, L.W., Yost, K.E., Mathey-Andrews, N., et al. (2024). Spatiotemporal lineage tracing reveals the dynamic spatial architecture of tumor growth and metastasis. Preprint at bioRxiv. https://doi.org/10.1101/2024.10.21.619529.

137. Chow, K.K., Budde, M.W., Granados, A.A., Cabrera, M., Yoon, S., Cho, S., Huang, T.H., Koulena, N., Frieda, K.L., Cai, L., et al. (2021). Imaging cell lineage with a synthetic digital recording system. Science *372*, eabb3099. https://doi.org/10.1126/science.abb3099.

138. Postrach, D., Pritchard, C.E.J., Frank, L., van Leeuwen, T., Messal, H.A., Krimpenfort, P., van Rheenen, J., and Rodewald, H.R. (2024). Polytope: High-resolution epitope barcoding for *in vivo* spatial fate-mapping. Preprint at bioRxiv. https://doi.org/10.1101/2024.11.20.624484.

139. You, Y., Fu, Y., Li, L., Zhang, Z., Jia, S., Lu, S., Ren, W., Liu, Y., Xu, Y., Liu, X., et al. (2024). Systematic comparison of sequencing-based spatial transcriptomic methods. Nat. Methods *21*, 1743–1754. https://doi.org/10.1038/s41592-024-02325-3.

140. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.

141. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386. https://doi.org/10.1038/nbt.2859.

142. Wang, S.W., Herriges, M.J., Hurley, K., Kotton, D.N., and Klein, A.M. (2022). CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information. Nat. Biotechnol. *40*, 1066–1074. https://doi.org/10.1038/s41587-022-01209-1.

143. Forrow, A., and Schiebinger, G. (2021). LineageOT is a unified framework for lineage tracing and trajectory inference. Nat. Commun. *12*, 4940. https://doi.org/10.1038/s41467-021-25133-1.

144. Lange, M., Piran, Z., Klein, M., Spanjaard, B., Klein, D., Junker, J.P., Theis, F.J., and Nitzan, M. (2024). Mapping lineage-traced cells across time points with moslin. Genome Biol. *25*, 277. https://doi.org/10.1186/s13059-024-03422-4.

145. Wang, K., Hou, L., Wang, X., Zhai, X., Lu, Z., Zi, Z., Zhai, W., He, X., Curtis, C., Zhou, D., et al. (2024). PhyloVelo enhances transcriptomic velocity field mapping using monotonically expressed genes. Nat. Biotechnol. *42*, 778–789. https://doi.org/10.1038/s41587-023-01887-5.

146. Weinreb, C., and Klein, A.M. (2020). Lineage reconstruction from clonal correlations. Proc. Natl. Acad. Sci. USA *117*, 17041–17048. https://doi.org/10.1073/pnas.2000238117.

147. Gao, M., Barile, M., Chabra, S., Haltalli, M., Calderbank, E.F., Chao, Y., Zheng, W., Wilson, N.K., Laurenti, E., Göttgens, B., et al. (2025). CLADES: a hybrid NeuralODE-Gillespie approach for unveiling clonal cell fate and differentiation dynamics. Nat. Commun. *16*, 8174. https://doi.org/10.1038/s41467-025-63150-6.

148. Schiffman, J.S., D'Avino, A.R., Prieto, T., Pang, Y., Fan, Y., Rajagopalan, S., Potenski, C., Hara, T., Suvà, M.L., Gawad, C., et al. (2024). Defining heritability, plasticity, and transition dynamics of cellular phenotypes in somatic evolution. Nat. Genet. *56*, 2174–2184. https://doi.org/10.1038/s41588-024-01920-6.

149. Sashittal, P., Zhang, R.Y., Law, B.K., Schmidt, H., Strzalkowski, A., Bolondi, A., Chan, M.M., and Raphael, B.J. (2026). Inferring cell differentiation maps from lineage tracing data. Nat. Methods *23*, 532–541. https://doi.org/10.1038/s41592-025-02903-z.

150. Erickson, A.G., Isaev, S., Artemov, A., He, J., Semsch, B., Murtazina, A., Sun, J., Mangold, K., Chalou, A., Frisen, J., et al. (2025). Unbiased profiling of multipotency landscapes reveals spatial modulators of clonal fate biases. Preprint at bioRxiv. https://doi.org/10.1101/2024.11.15.623687.

151. Fu, Y., Mathew, D., Wang, M., Chen, X.E., Lin, K.Z., Schaff, D., Shaffer, S.M., Pardoll, D.M., Jackson, C., and Zhang, N.R. (2025). Deciphering Cell Fate and Clonal Dynamics via Integrative Single-Cell Lineage Modeling. Preprint at Res. Sq. https://doi.org/10.21203/rs.3.rs-7510946/v1.