

RESEARCH ARTICLE SUMMARY

BIOTECHNOLOGY

Developmental barcoding of whole mouse via homing CRISPR

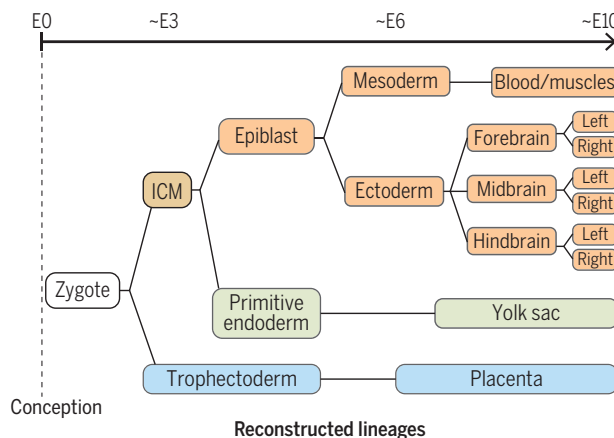
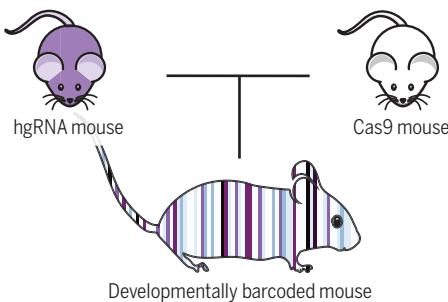
Reza Kalhor*, Kian Kalhor, Leo Mejia, Kathleen Leeper, Amanda Graveline, Prashant Mali, George M. Church*

INTRODUCTION: The remarkable development of a single cell, the zygote, into the full organism occurs through a complex series of division and differentiation events that resemble a tree, with the zygote at the base branching through lineages that end in the terminal cell types at the top. Characterizing this tree of development has long been a subject of interest, and the combination of modern genome engineering and sequencing technologies promises a powerful strategy in its service: in vivo barcoding. For in vivo barcoding, heritable random mutations are induced to accumulate during development and sequenced post hoc to reconstruct the lineage tree. Demonstrations thus far have largely focused on lower vertebrates and have used a barcoding element with a constrained window of activity for clonal tracing of individual cells or cell types. Implementation in mammalian model systems, such as the mouse, incurs unique challenges that require major enhancements.

RATIONALE: To address the complexity of mammalian development, we reasoned that multiple independent in vivo barcoding elements could be deployed in parallel to exponentially expand their recording power. Independence requires both an absence of cross-talk between the elements and an absence of interference between their mutation outcomes. A system with the potential to deliver on these requirements is homing CRISPR, a modified version of canonical CRISPR wherein the homing guide RNA (hgRNA) combines with CRISPR-Cas9 nuclease for repeated targeting of its own locus, leading to diverse mutational outcomes. Therefore, in mouse embryonic stem cells, we scattered multiple hgRNA loci with distinct spacers in the genome to serve as barcoding elements. With this arrangement, each hgRNA acts

independently as a result of its unique spacer sequence, and undesirable deletion events between multiple adjacent cut sites are less likely. Using these cells, we generated a chimeric mouse with 60 hgRNAs as the founder of the MARC1 (Mouse for Actively Recording Cells 1) line that enables barcoding and recording of cell lineages.

RESULTS: In the absence of Cas9, hgRNAs are stable and dormant; to initiate barcoding,



Developmental barcoding and lineage reconstruction in mice. Crossing the MARC1 mouse line, which carries multiple hgRNAs, with a CRISPR-Cas9 mouse line results in developmentally barcoded offspring that record lineages in their cells. These recordings were extracted and used to reconstruct lineage trees. A combination of the trees extracted from different developmentally barcoded mice is shown. ICM, inner cell mass; E0, embryonic day 0.

we crossed MARC1 mice with Cas9 knock-in mice. In the resulting offspring, hgRNAs were activated, creating diverse mutations such that an estimated 10^{23} distinct barcode combinations can be generated with only 10 hgRNAs. Furthermore, hgRNAs showed a range of activity profiles, with some mutating soon after conception while others exhibited lower activity

ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aat9804>

through most of the gestation period. This range resulted in sustained barcoding throughout gestation and recording of developmental lineages: Each cell inherits a set of unique mutations that are passed on to its daughter cells, where further unique mutations can be added. Consequently, at any stage in such developmentally barcoded mice, closely related cells have a more similar mutation profile, or barcode, than the more distant ones. These recordings remain embedded in the genomes of the cells and can be extracted by sequencing.

We used these recordings to carry out bottom-up reconstruction of the mouse lineage tree, starting with the first branches that emerged after the zygote, and observed robust reconstruction of the correct tree. We also investigated axis development in the brain by sequencing barcodes from the left and right side of the forebrain, midbrain, and hindbrain regions. We found that barcodes from the left and right sides of the same region were more closely related than those from different regions; this result suggests that in the precursor of the brain, commitment to the anterior-posterior axis is established prior to the lateral axis.

CONCLUSION: This system provides an enabling and versatile platform for in vivo barcoding and lineage tracing in a mammalian model system. It can straightforwardly create developmentally barcoded mice in which lineage information is pre-recorded in cell genomes. Combining multiple independently acting molecular recording devices greatly enhances their capacity and allows for reliable information recovery and reconstruction of deep lineage trees. ■

The list of author affiliations is available in the full article online.

*Corresponding author. Email: gchurch@genetics.med.harvard.edu (G.M.C.); kalhor@genetics.med.harvard.edu (R.K.)
Cite this article as R. Kalhor et al., *Science* **361**, eaat9804 (2018). DOI: 10.1126/science.aat9804

RESEARCH ARTICLE

BIOTECHNOLOGY

Developmental barcoding of whole mouse via homing CRISPR

Reza Kalhor^{1,2*}, Kian Kalhor³, Leo Mejia¹, Kathleen Leeper², Amanda Graveline², Prashant Mali⁴, George M. Church^{1,2*}

In vivo barcoding using nuclease-induced mutations is a powerful approach for recording biological information, including developmental lineages; however, its application in mammalian systems has been limited. We present in vivo barcoding in the mouse with multiple homing guide RNAs that each generate hundreds of mutant alleles and combine to produce an exponential diversity of barcodes. Activation upon conception and continued mutagenesis through gestation resulted in developmentally barcoded mice wherein information is recorded in lineage-specific mutations. We used these recordings for reliable post hoc reconstruction of the earliest lineages and investigation of axis development in the brain. Our results provide an enabling and versatile platform for in vivo barcoding and lineage tracing in a mammalian model system.

In sexually reproducing multicellular eukaryotes, a single totipotent zygote remarkably develops into all cells of the full organism. This development occurs through a highly orchestrated series of differentiation events that take the zygote through many lineages as it divides to create all the different cell types (1). This path resembles a tree, with the zygote at the base of the trunk branching into stems of cell lineages that eventually end in the terminal cell types at the top of the tree (2, 3). The ability to map this tree of development will have a far-ranging impact on our understanding of disease-causing developmental aberrations, our capacity to restore normal function in damaged or diseased tissues, and our capability to generate substitute tissues and organs from stem cells.

Tracing the lineage tree in non-eutelic higher eukaryotes with complex developmental pathways remains challenging. Clonal analysis, which entails cellular labeling and tracking with a distinguishable heritable marker, has been effective when evaluating a limited number of cells or lineages (4–7). Using more diverse presynthesized DNA sequences as markers, known as cellular barcoding, has allowed for analysis of larger cell numbers (7–9). What limits these approaches is the static nature of labeling that only allows analysis of a snapshot in time. Recent advances in genome engineering technologies, however, have enabled in vivo barcode generation (10, 11). In this approach, a locus is targeted for rearrangement or mutagenesis such

that a diverse set of outcomes is generated in different cells (12). As these barcodes can be generated over a sustained period of time, they drastically expand the scope of cellular barcoding strategies, promising deep and precise lineage tracing, from the single-cell to the whole-organism level (Fig. 1A) (13–15), and recording of cellular signals over time (16, 17). Multiple studies establish proof of this principle in recording and lineage tracing, with demonstrations in cultured cells (14–16, 18) and in lower vertebrates (13, 19–22). However, no demonstrations have yet been carried out in mice, a model organism more relevant to human health in many aspects such as development. The challenges associated with work in mice can account for this discrepancy. Gestation in mice takes place inside the mother's womb, rendering genetic manipulation of individual zygotes or conceptuses difficult. Additionally, the longer gestation time in mice, together with the multitude of lineages that segregate throughout its development, demands sustained generation of highly diverse barcodes with minimal unwanted overwriting events to maximize the chance for successfully recording the events of interest.

Here, we deployed multiple independent barcoding loci in parallel for robust in vivo barcoding and lineage recording in mice. We created a mouse line that carries a scattered array of 60 genomically integrated homing CRISPR guide RNA (hgRNA) loci. hgRNAs are modified versions of canonical single guide RNAs (sgRNAs) (23) that target their own loci (Fig. 1B) to create a substantially larger diversity of mutants than canonical sgRNAs (Fig. 1C) and thus act as expressed genetic barcodes (14). Crossing this hgRNA line with a Cas9 line resulted in developmentally barcoded offspring because hgRNAs stochastically accumulate mutations throughout gestation, generating unique mutations in each

lineage without deleting earlier mutations, in such a way that closely related cells have a more similar mutation profile, or barcode, than more distant ones. In developmentally barcoded mice, we extensively characterized the activity profile and mutant alleles of each hgRNA and carried out post hoc bottom-up reconstruction of the lineage tree in the early stage of development, starting with the first branches at its root and continuing through some of the germ layers. We also investigated lineage commitment with respect to the anterior-posterior and lateral axes in the brain.

Founder mouse with multiple hgRNA loci

We created a library of hgRNAs with four different transcript lengths, variable spacer sequences, and 10-base identifiers downstream of the hgRNA scaffold in a transposon backbone (Fig. 1D) (24). This library was transposed into mouse embryonic stem (mES) cells under conditions that would result in a high number of integrations per cell (Fig. 1E) (24). Transfected mES cells were injected into blastocysts, which were then implanted in surrogate females to generate chimeric mice. Of the 23 chimeric mice that resulted, eight males were more than 60% transgenic as assessed by their coat color (Fig. 1E). Five of the eight showed more than 20 total hgRNA integrations in their somatic genomes and were crossed with wild-type mice to determine the number of hgRNAs in their germlines. The chimera with the highest average number of germline hgRNAs, which were transmitted to its progeny, was selected for further studies and starting a line. We refer to this mouse as the MARCI (Mouse for Actively Recording Cells 1) founder and its progeny as the MARCI line. All results described below focus on the MARCI founder and its progeny.

Sequence, genomic position, and inheritance of hgRNA loci

By sequencing the hgRNA loci in the MARCI founder, we identified 60 different hgRNAs (Table 1 and table S1). Each hgRNA has a unique 10-base identifier and a different spacer sequence (table S1). We also sequenced the regions immediately flanking the transposed hgRNA elements (24), which allowed us to determine the genomic positions of 54 of the 60 hgRNAs (Fig. 1F, Table 1, and table S1), of which 26 are intergenic and 28 are located in an intron of a known gene (table S3) (24); none are located in an exon or are expected to disrupt the gene. We then crossed the MARCI founder with multiple females and analyzed germline transmission and the inheritance pattern of these hgRNAs in the more than 100 resulting offspring. All 60 hgRNAs were transmitted through the germline, and the offspring carrying them were fertile, had normal litter sizes, and presented no morphological abnormalities. Of these 60 hgRNAs, 55 showed a Mendelian inheritance pattern, appearing in about 50% of the offspring (Table 1 and table S1). An additional 3 of the 60, all L30 hgRNAs, were detected in fewer than 20% of the offspring,

¹Department of Genetics, Harvard Medical School, Boston, MA, USA. ²Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, USA. ³Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran. ⁴Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA. *Corresponding author. Email: gchurch@genetics.med.harvard.edu (G.M.C.); kalhor@genetics.med.harvard.edu (R.K.)

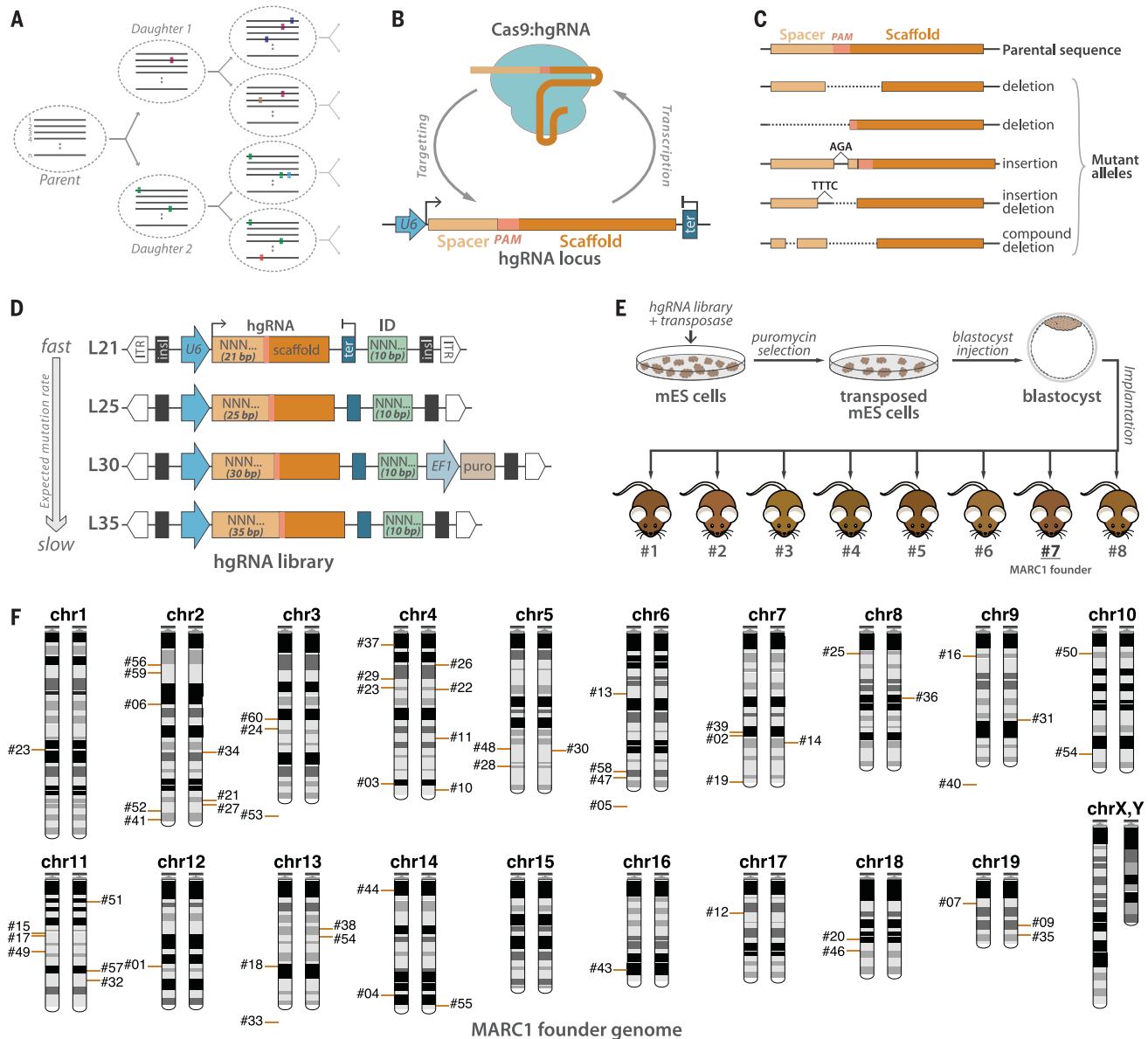


Fig. 1. In vivo barcoding with hgRNAs and strategy to generate a mouse with multiple hgRNA integrations. (A) Recording lineages using synthetically induced mutations in the genome. A number of loci (n) gradually accumulate heritable mutations as cells divide, thereby recording the lineage relationship of the cells in an array of mutational barcodes. Dashed ovals, cells; gray lines, an array of n mutating loci; colored rectangles, mutations. (B) Homing CRISPR system in which the Cas9-hgRNA complex cuts the locus encoding the hgRNA itself. As the NHEJ repair system repairs the cut (63), it introduces mutations in the hgRNA locus. (C) Example of mutations that are created in the hgRNA locus that can effectively act as barcodes. (D) Design of PiggyBac hgRNA library for creating a transgenic mouse. Four hgRNA sublibraries with 21, 25, 30, and 35 bases of distance between transcription start site (TSS) and scaffold PAM were constructed and combined. The spacer sequence

(light orange box) and the identifier sequence (green box) were composed of degenerate bases. (E) Blastocyst injection strategy for producing hgRNA mice. The hgRNA library was transposed into mES cells. Cells with a high number of transpositions were enriched using puromycin selection and injected into E3.5 mouse blastocysts to obtain chimeras. Chimera 7 was chosen as the MARC1 founder. (F) Chromosomal position of all 54 hgRNAs whose genomic position was deciphered in the MARC1 founder (red bars). Bars on the left or right copy of the chromosome indicate the hgRNAs that are linked on the same homologous copy. hgRNAs whose exact genomic position is not known but whose chromosome can be determined on the basis of linkage are shown below the chromosome. ITR, PiggyBac inverted terminal repeats; insl, insulator; U6, U6 promoter; ter, U6 terminator; ID, identifier sequence; EF1, human elongation factor-1 promoter; puro, puromycin resistance.

which we attribute to the low detection rate of L30 hgRNAs due to the performance of the polymerase chain reaction primer used for these and only these three hgRNAs (24). The remaining two hgRNAs were transmitted to almost 75% of the offspring—a result best explained by the

duplication of these hgRNAs to loci more than 50 cM away on the same chromosome or to loci on different chromosomes and confirmed by the genomic location data (table S1) (24).

We also compared the co-inheritance frequencies of MARC1 hgRNAs to those expected from

Mendelian inheritance of independently segregating loci (fig. S1A). We found no mutually exclusive cosegregating groups of hgRNAs (fig. S1A), indicating that the entire germline in the MARC1 founder was derived from only one of the injected stem cells and is thus genetically homogeneous.

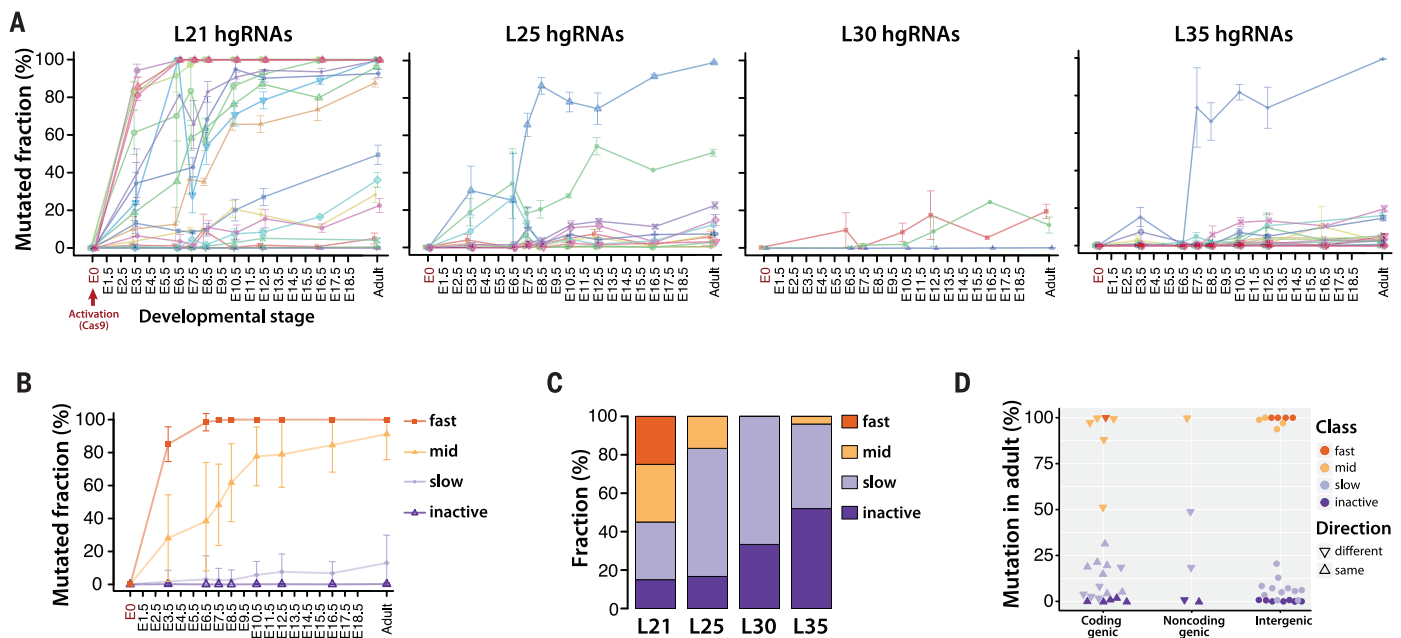


Fig. 2. Activity of MARC1 hgRNAs. (A) Activity profiles of all 60 hgRNAs in embryonic and adult progenies of the MARC1 founder crossed with Cas9 knock-in females, broken down by hgRNA length. The fraction of mutant (nonparental) spacer sequences in each hgRNA is measured. Lines connect the observed average mutation rates of one hgRNA. Means \pm SEM are shown (N is different for each value; see Table 2). See table S2 for numerical values of the plot. (B) Average activity profiles of each hgRNA class in embryonic and adult progenies of the MARC1

founder crossed with Cas9 knock-in females. Means \pm SEM are shown as representations of range of activity (N is different for each value; see Table 2). (C) Functional categorization of hgRNAs based on their activity profile in (A), broken down by length. (D) Position and transcription direction of hgRNAs with respect to all known coding and noncoding genes, annotated for their functional category. See table S3 for the genes in which hgRNAs are located; see fig. S3 for breakdown of this plot by hgRNA length.

Considering that every hgRNA detected in the somatic tissue of the MARC1 founder was also transmitted to its offspring, these results further suggest that almost all transgenic cells within this chimera were derived from one of the stem cells that were injected into its blastocyst, an observation consistent with previous studies (25, 26). The co-inheritance analysis also revealed the groups of hgRNAs that deviate from an independent segregation pattern, suggesting that they are linked on a chromosome (fig. S1B). Close examination of this linkage disequilibrium allowed us to determine which linked hgRNAs were on different homologous copies of the same chromosome or were linked on the same copy of a chromosome (Fig. 1F and fig. S1C). Combined with the genomic location information that was obtained by sequencing, this co-inheritance analysis allowed us to decipher the cytogenetic location of most hgRNAs in the MARC1 founder with a high degree of confidence (Fig. 1F).

Activity of hgRNAs

We next studied the activity of MARC1's hgRNAs upon activation with Cas9. For that, we crossed the MARC1 founder with Rosa26-Cas9 knock-in females, which constitutively express the *Streptococcus pyogenes* Cas9 protein (27). Considering that major zygotic genome activation in the mouse occurs at the two-cell stage (28), hgRNA activation is expected soon after conception. We sampled these Cas9-activated offspring at various stages

after conception to measure the fraction of mutated spacers for each hgRNA. In all, we gathered 190 samples from 102 animals in seven embryonic stages and the adult stage (Table 2). The results confirm that hgRNAs start mutating their loci soon after the introduction of Cas9 (Fig. 2A). However, the rate at which these mutations accumulate varied widely among the 60 MARC1 hgRNAs (Fig. 2A). On the basis of these activity levels, we classified hgRNAs into four categories with distinct activation profiles (Fig. 2B): (i) five “fast” hgRNAs that mutate in at least 80% of the cells in each sample by embryonic day 3.5 (E3.5) and in almost all cells by E8.5; (ii) 27 “slow” hgRNAs that mutate in only a minority of cells even in the adult stage; (iii) nine “mid” hgRNAs, intermediate between fast and slow, that accumulate mutations throughout embryonic development and are mutated in almost all cells only in later embryonic or adult stages; and (iv) 19 hgRNAs that appear to be inactive, at least with this level of Cas9 expression, mutating in fewer than 2% of sampled cells even in the adult stage (table S2). Most mutations that were detected (about 80% for fast hgRNAs) are expected to render the hgRNA nonfunctional and thus prevent further changes (fig. S2).

Transcript length clearly affects hgRNA activity: A far higher fraction of L21 hgRNAs, which have the shortest possible transcript length, were active by comparison to L25, L30, and L35 hgRNAs, which are longer by 4, 9, and 14 bases, respectively (Fig. 2, A and C). Furthermore, all fast hgRNAs

were L21 hgRNAs, whereas in longer hgRNAs the inactive proportion appeared to grow with increasing length (Fig. 2C). Beyond transcript length, we found that the variation in activity among hgRNAs with an identical length (Fig. 2A) is far more than would be expected solely on the basis of differences in their spacers (14), which suggests that genomic location may play a substantial role. Although we detected no significant difference between the activity of hgRNAs that are in intergenic regions relative to those within known genes (Wilcoxon $P > 0.1$), among hgRNAs that have landed within known coding and noncoding genes, those that transcribe in the same direction as the gene had a lower activity than those that transcribe in the opposite direction (Wilcoxon $P < 0.05$; Fig. 2D, fig. S3, and table S3). These observations suggest that hgRNA activity is affected by both genomic location and interplay with endogenous elements.

Diversity and composition of hgRNA mutants

We next analyzed the diversity produced by MARC1 hgRNAs by considering all observed mutant spacer alleles in MARC1 \times Cas9 offspring (table S4). Only a handful of mutant spacer alleles were detected for each hgRNA in each sample (Fig. 3A and fig. S4A). However, when combining mutant spacers from all offspring, on average, more than 200 distinct mutant spacers for each fast hgRNA and more than 300 for each mid hgRNA were observed (Fig. 3B and

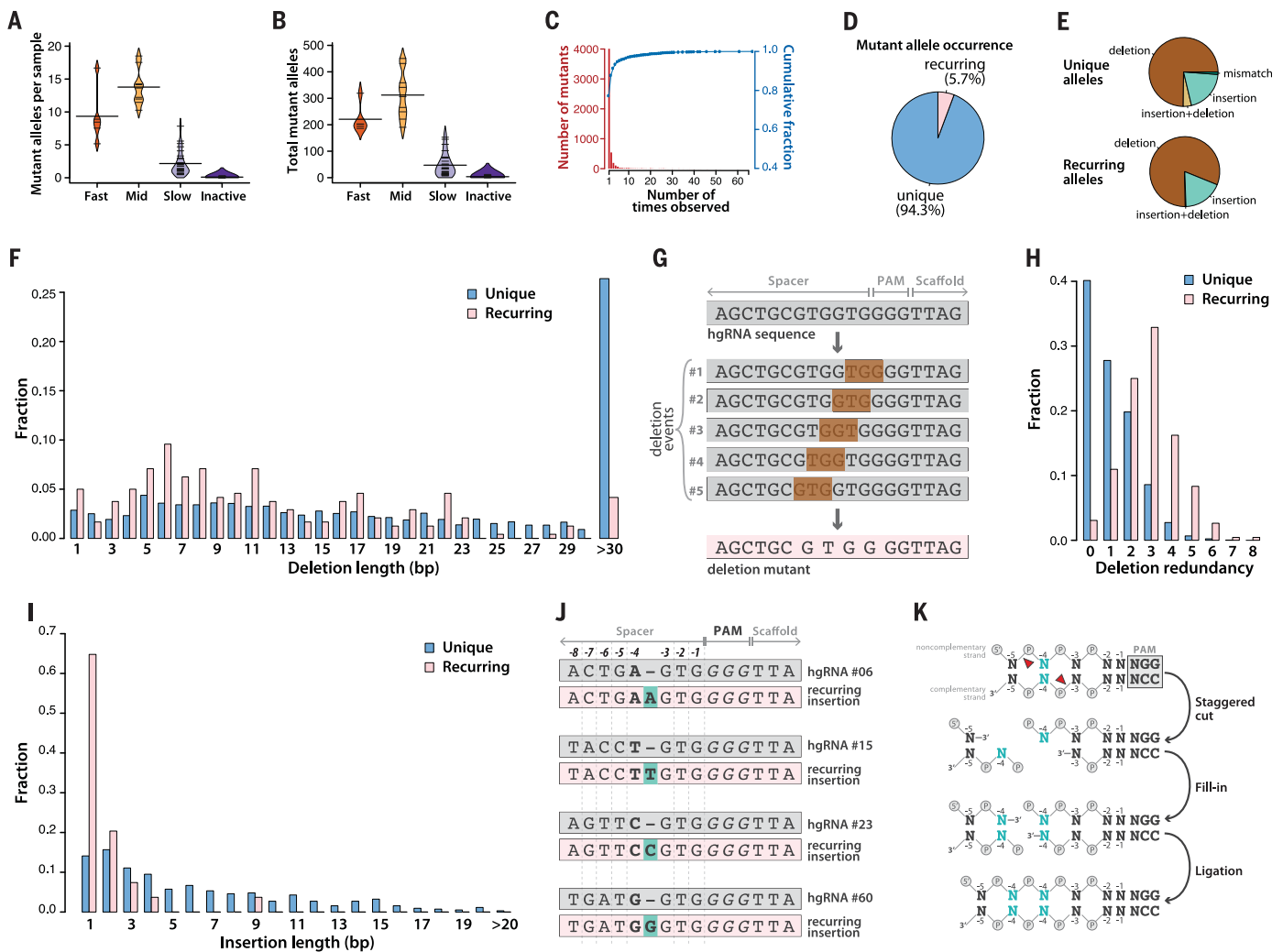


Fig. 3. Diversity of mutant hgRNA alleles in offspring of MARC1 × Cas9 cross. (A) Beanplots of the number of mutant spacer alleles observed in each mouse for each hgRNA category. Short horizontal lines mark the average for each hgRNA in the category; long horizontal lines mark the average of all the hgRNAs in the category. See fig. S4A for a separate plot for each hgRNA. (B) Beanplots of the total number of mutant spacer alleles observed for each hgRNA in all mice. See fig. S4B for a separate plot for each hgRNA. (C) Histogram (red bars) and cumulative fraction (blue connected dots) of the number of mice in which each mutant allele was observed, combined for all hgRNAs. See fig. S5 for a separate plot for each hgRNA. (D) Relative ratio of recurring mutant spacer alleles (fig. S5) (24) to the unique alleles. (E) Mutation types in unique (top) and recurring (bottom) spacer alleles. See tables S4 and S5 for the sequences and alignment of all mutants and recurring mutants, respectively, and fig. S6 for a separate plot for each hgRNA.

fig. S4B). Furthermore, about 80% of all mutant spacer alleles were unique observations in a single offspring (Fig. 3C and fig. S5), which suggests that the mutant alleles observed with our sampling level constitute only a minority of all mutant spacers possible. These results indicate that each hgRNA can produce hundreds of mutant alleles.

Notably, although most mutant spacer alleles appeared in only a single sample, about

6% recurred in multiple MARC1 × Cas9 offspring (Fig. 3, C and D, and fig. S5). To understand this phenomenon, we compared the nature of unique and recurring mutant alleles (tables S4 and S5). We observed that insertions or deletions (indels) underlie the vast majority of alleles in both unique and recurring mutations (Fig. 3E and fig. S6, A and B). The exact nature of these indel mutations, however, differs. First, short deletions of 23 base pairs (bp) or fewer are enriched in the

(F) Distribution of deletion length for unique and recurring mutant spacer alleles. Deletions larger than 30 bp have been aggregated. (G) Schematic representation of how five distinct deletion events can lead to the same mutant spacer allele. (H) Distribution of deletion redundancy—that is, the number of independent simple deletion events in the parental spacer allele that would lead to the same observed deletion mutant—for unique and recurring spacer alleles. Simple deletion is defined as deletion of a contiguous stretch of bases without creating insertions or mismatches. Redundancy of 0 represents non-simple mutant alleles, which involve insertions, mismatches, or noncontiguous deletions. (I) Distribution of insertion length for unique and recurring mutant spacer alleles. Insertions of 20 bp or longer have been aggregated. (J) Four observed examples of recurring single-base insertions, involving duplication of the –4 position, for four different hgRNAs. (K) Schematic representation of how a single-base staggered overhang generated by Cas9 can lead to duplication of the –4 position.

recurring alleles (Fig. 3F). Interestingly, these mutant alleles can be identical results of multiple different simple deletions in the parental spacer sequence (Fig. 3, G and H), which implies that this group of recurring mutations can result from distinct mutagenesis events that lead to the same sequence. Second, single-base insertions are drastically enriched among recurring insertion mutants (Fig. 3I). A closer examination of these single-base insertions revealed that many follow the same

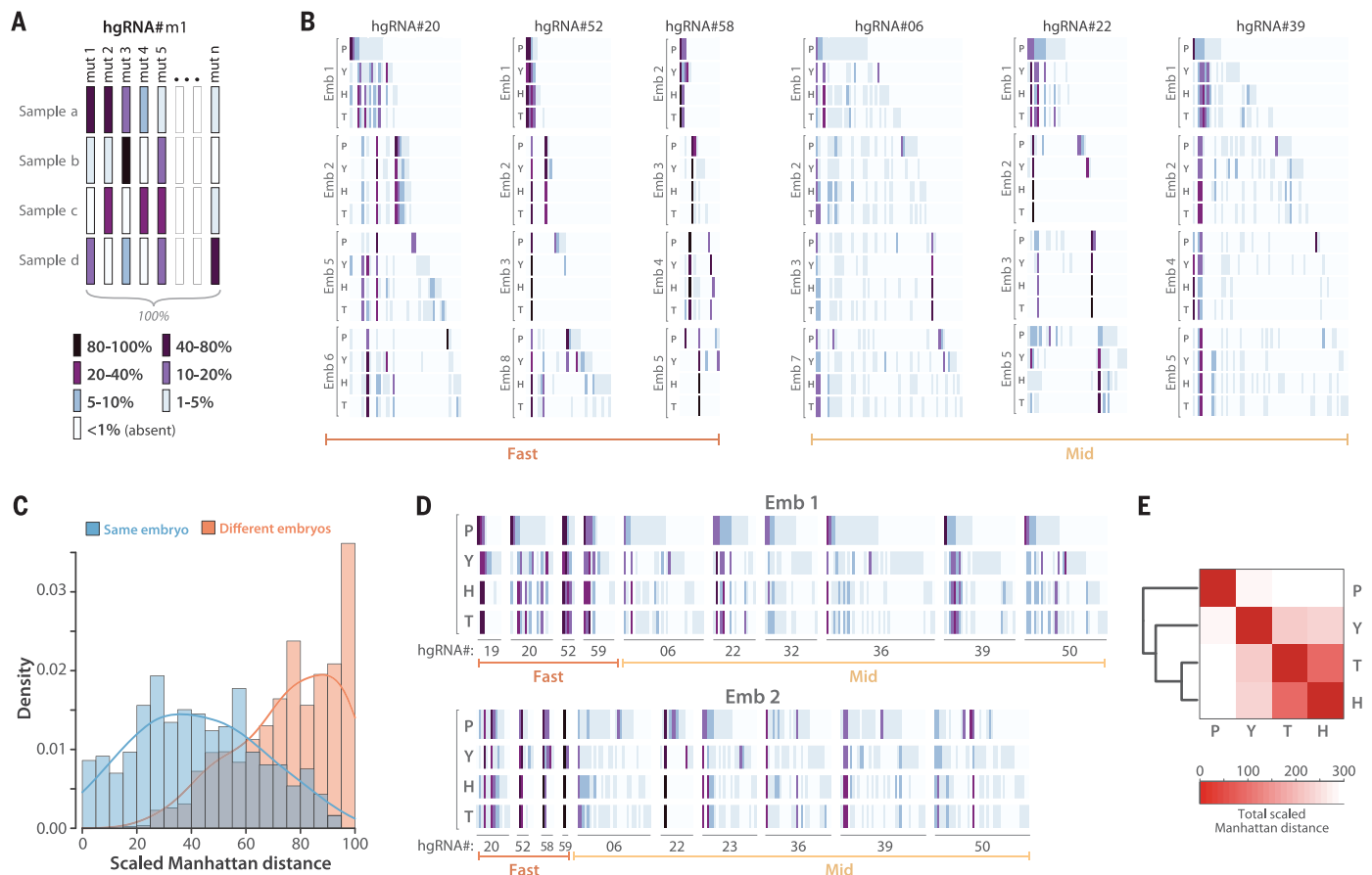


Fig. 4. In vivo barcoding in mouse embryos. (A) Barcode depiction for each hgRNA in each sample. Each column corresponds to an observed mutant spacer; each row corresponds to a sample. The color of each block represents the observed frequency of the corresponding mutant spacer in the corresponding sample. (B) In vivo-generated barcodes of three fast and three mid hgRNAs in eight embryos from a MARC1 × Cas9 cross. Four tissues were sampled from each embryo: the placenta (P), the yolk sac (Y), the head (H), and the tail (T). Embryos 1 and 2 were obtained at E16.5, whereas embryos 3 to 8 were obtained at E12.5 (Table 2). For each hgRNA, the results for a maximum of four embryos are shown. Full

barcodes for all hgRNAs are in fig. S7. The color code is as shown in (A). Only mutant alleles with a maximum abundance of more than 1% are shown. (C) Histogram of the scaled Manhattan distances (L1) between the barcodes of all possible sample pairs for each hgRNA, broken down by sample pairs belonging to the same embryo (blue) and pairs belonging to different embryos (orange). (D) The complete barcode, composed of the concatenation of all hgRNA barcodes, for embryos 1 and 2. (E) Heat map of the average Manhattan distance between the full barcodes of placenta, yolk sac, head, and tail samples in all eight embryos. For a separate map for each embryo, see fig. S8.

pattern: duplication of the base at the -4 position relative to the protospacer adjacent motif (PAM) (Fig. 3J). In fact, this type of insertion was recurring in 34 of the 41 active hgRNAs. This observation can be best explained by Cas9 cutting at the -4 position of the noncomplementary strand and at the -3 position of the complementary strand, thus creating a staggered end with a 5' overhang, which is then filled in on both ends and ligated (Fig. 3K). Therefore, our results suggest that Cas9 can produce staggered cuts, and that the nature of these cuts, together with the sequence of the target site, affects the eventual outcome of nonhomologous end joining (NHEJ) repair.

Developmental hgRNA barcodes

The results thus far indicate that MARC1 hgRNAs accumulate mutations upon activation with Cas9 nuclease after conception. We next queried whether these mutations indeed reflect developmen-

tal events. For simplicity, we focused on fast and mid hgRNAs in eight post-E12 MARC1 × Cas9 offspring for which four different tissues had been sampled (Table 2). The sampled tissues were the placenta, the yolk sac, the head, and the tail. The barcode was defined for each hgRNA in each sample as the frequency vector of the relative abundances of all observed mutant alleles (Fig. 4A). For the 32 samples under consideration (eight conceptuses with four samples each), these barcodes showed diverse and complex patterns, with each sample having a unique barcode but with varying degrees of similarity to other samples (Fig. 4B and fig. S7). To compare the hgRNA barcodes between samples, we used a scaled Manhattan distance (L1) of their frequency vectors, such that a distance of 100 would indicate a completely nonoverlapping set of mutant alleles and a distance of 0 would indicate a complete overlap of mutant alleles with identical relative frequencies (24).

Pairwise comparison of all hgRNA barcodes among all samples (Fig. 4C) showed that more than 99% of barcode pairs have a scaled Manhattan distance of more than 5, indicating unique barcoding of each sample by each hgRNA. Furthermore, barcodes from different tissues of the same embryo were more similar to each other (median distance = 41) and more distinct from different embryos (median distance = 78) (Fig. 4C), which suggests that barcodes may record information about the history of samples relative to one another.

To further evaluate this recording of sample histories, we created a “full” barcode for each sample by combining the barcodes generated by each of its hgRNAs (Fig. 4D) and compared the distance between these barcodes in the four tissues obtained from each embryo (Fig. 4E and fig. S8). The results show higher similarity between the head and tail samples, which together are the most different from placenta. The samples

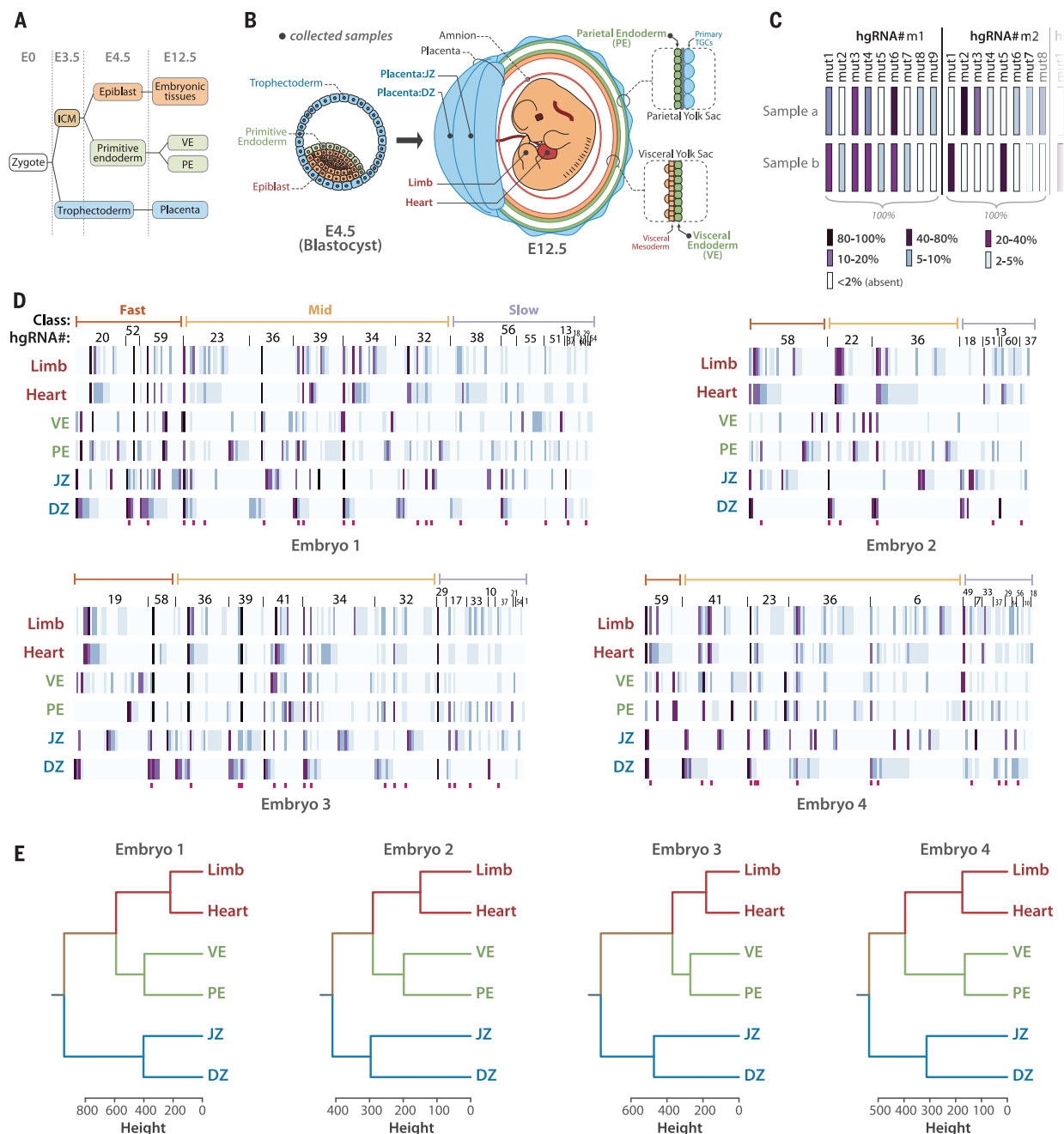


Fig. 5. Lineage derivation based on hgRNA-generated developmental barcodes. (A) Summary of the earliest lineages in mouse. (B) Schematic representation of a blastocyst and an E12.5 mouse conceptus, color-coded according to the origin of tissues in the blastocyst. Black dots show the positions and tissues of the samples obtained from E12.5 conceptuses. (C) Summary of how hgRNA barcodes were compiled for each sample. Each bar represents a mutant spacer of an hgRNA, and its color represents its abundance relative to other mutant

spacers of the same hgRNA in the same sample. (D) Full hgRNA barcodes for all samples from the four mouse embryos analyzed. The barcode is annotated in (C). Only mutant alleles with a maximum abundance of more than 2% are shown. Deep pink bars below each map mark highly recurring alleles that have been observed in more than 60% of all mice analyzed in Table 2. See table S6 for a numerical version of each barcode map. (E) Lineage tree for each embryo calculated from the barcodes in (D).

obtained here represent mixed and overlapping lineages. However, considering that the head and tail are derived from the inner cell mass (ICM) whereas the placenta is mostly derived from the trophoblast (29–31), these results suggest that hgRNA barcodes of different tissues embody their lineage histories.

First lineage tree from barcode recordings

We next assessed whether accurate lineage trees can be constructed de novo from developmentally barcoded mice. To assess this potential, we focused on the tree of the first lineages in development. The first lineage segregation events

in mammals are the differentiation of blastomeres into the trophoblast and ICM before E3.5, followed by differentiation of the ICM into the primitive endoderm and epiblast by E4.5 (Fig. 5A) (29). To reconstruct this lineage tree, we used developmentally barcoded E12.5 conceptuses and sampled two distinct tissues from each of

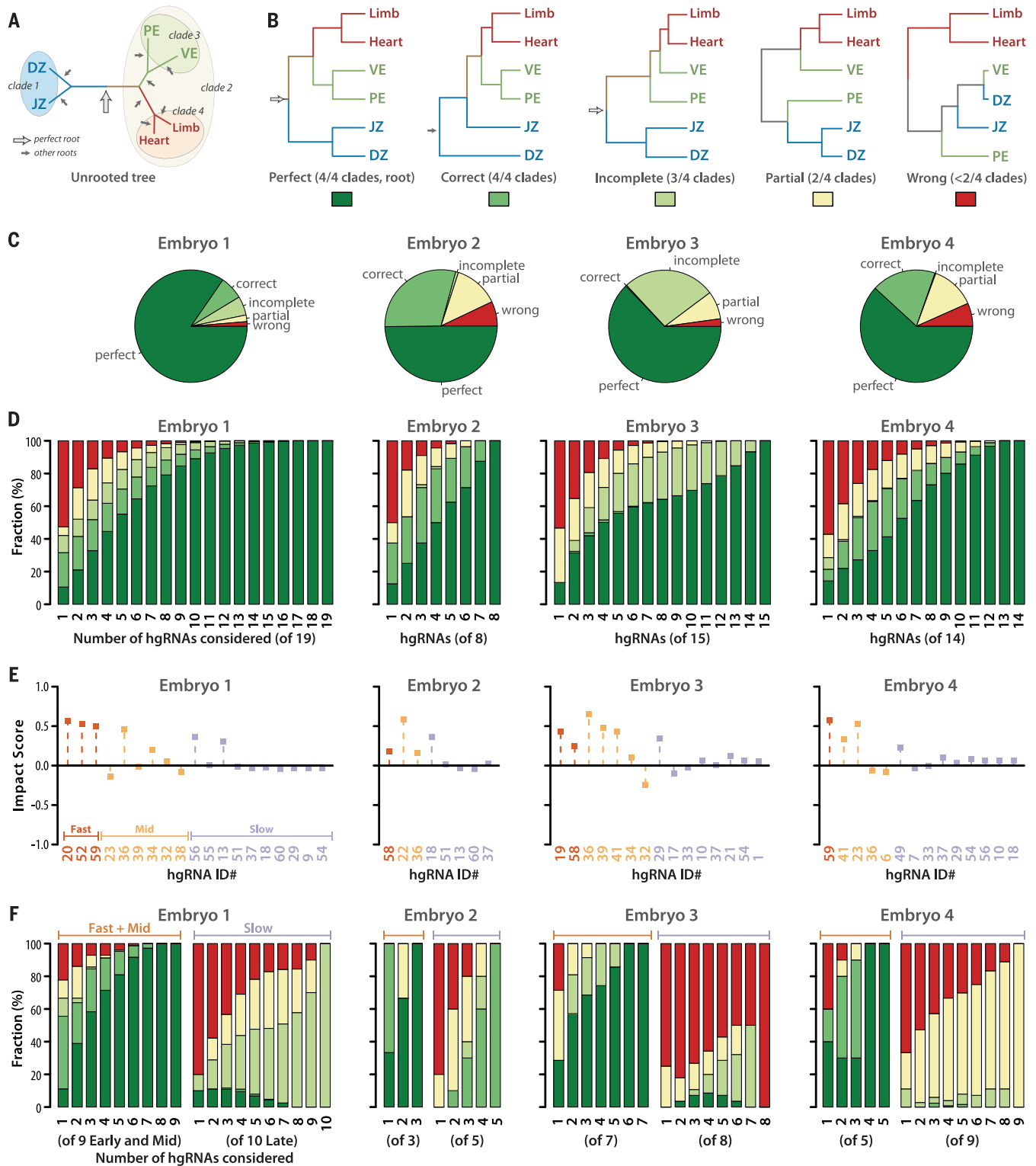


Fig. 6. Lineage tree derivation robustness and contribution of each hgRNA. (A) The correct unrooted tree topology for the earliest lineages in mouse. Arrows indicate all possible roots. The empty arrow indicates the perfect root. (B) The perfect rooted topology and an example from each of the other topology classifications. The colored boxes below each topology constitute the color key for the remaining panels of the figure. (C) For each of the four embryos analyzed, distribution of tree calculation outcomes from all possible subsets of hgRNAs ($2^n - 1$ non-null subsets for an embryo with

n hgRNAs). (D) Distribution of tree calculation outcomes when including only m of the n hgRNAs in each embryo (nC_m combinations, $1 \leq m \leq n$). Color code is as described in (B). See also figs. S9 and S10 for all combinations included and excluding each hgRNA. (E) Impact score of each hgRNA in the early lineage tree of each embryo. (F) Distribution of tree calculation outcomes when only including k of the $n_{fast} + n_{mid}$ fast and mid hgRNAs (left side of each panel, ${}^{n_{fast}+n_{mid}}C_k$ combinations) or k of the n_{slow} slow hgRNAs (right side of each panel, ${}^{n_{slow}}C_k$ combinations). Color code is as described in (B).

three lineages: the decidual zone (DZ) and the junctional zone (JZ) of the placenta, which are descendants of the trophoctoderm (30, 31); the parietal endoderm (PE) and visceral endoderm (VE) of the yolk sac, which are descendants of the primitive endoderm; and the heart and a limb bud of the embryo proper, which are descendants of the epiblast (29) (Fig. 5B). We then assembled the full barcode for each sample (Fig. 5, C and D, and table S6) and, using their Manhattan distances, clustered them to form a tree for each embryo (Fig. 5E) (24). Remarkably, despite the differences in the number and composition of hgRNAs inherited, the resulting tree perfectly matched the expected lineage in all four embryos, showing that the DZ and JZ form one clade of the tree while the other clade comprises two subclades, one with PE and VE and the other with the heart and limb bud (Fig. 5E). These results demonstrate that accurate lineage trees can be constructed from developmentally barcoded mice.

We next evaluated the robustness of lineage tree derivation from hgRNA barcodes by calculating the tree topology with only parts of the full barcodes. For a bifurcating tree with six tips (limb, heart, VE, PE, JZ, and DZ; Fig. 6A), 945 distinct rooted topologies are possible (32). Only a single one of these 945 tree topologies perfectly matches the expected lineage tree; we refer to this topology as “perfect” (Figs. 5E and 6B). Another eight topologies would be correct if unrooted—that is, if all four clades are correctly assigned but the root is misplaced because a branch other than the one connecting the (DZ, JZ) clade to the ((PE, VE), (heart, limb)) clade is the longest (Fig. 6A). We refer to these topologies as “correct” (Fig. 6B). If three, two, or fewer than two of the four clades have been assigned correctly, we consider the topologies as “incomplete,” “partial,” and “wrong,” respectively (Fig. 6B). With these distinctions, we evaluated the trees generated with all possible non-null subsets of the hgRNAs in each embryo. The results show that, depending on the embryo, 60% to 85% of all possible hgRNA subsets result in a perfect or correct topology (Fig. 6C), which compares favorably to the ~1% chance of randomly finding such topologies. With only three hgRNAs, more than 50% of all derived trees have a correct or perfect topology for each embryo (Fig. 6D). Furthermore, calculated topologies improve with increasing the number of hgRNAs (Fig. 6D). Combined, these results show that lineage tree derivation from *in vivo*-generated hgRNA barcodes is robust and that the use of a higher number of hgRNAs results in more reliable outcomes.

We then examined the contribution of each hgRNA to deriving the correct tree topology for each embryo. We defined the “impact score” of an hgRNA in each embryo’s early lineage tree as the difference between the fraction of all correct and perfect topologies in which the hgRNA was considered and the fraction of all wrong and partial topologies in which the hgRNA was considered (Fig. 6E and figs. S9 and S10) (24). As such, an impact score of +1

would indicate that whenever the hgRNA was included in tree derivation, a correct or a perfect topology was obtained, and no such topologies were obtained without that hgRNA. An impact score of -1 would indicate that when the hgRNA was included in tree derivation, only partial or wrong topologies were obtained. Values between +1 and -1 define the range between those entirely constructive or destructive outcomes, with an impact score of 0 indicating that the likelihood of obtaining a correct topology is the same with or without the hgRNA. Impact scores for hgRNAs in our four embryos show a positive average contribution by all three active hgRNA classes with slow hgRNAs, which are largely unmutated early in development (Fig. 2B), having an average impact close to 0, and mid and fast hgRNAs, which are active early in development (Fig. 2B), having increasingly positive impacts on the derivation of the correct tree (Fig. 6E). In fact, only three fast and mid hgRNAs suffice to obtain a correct or perfect topology in more than 90% of all derived trees (Fig. 6F). By contrast, exclusive use of slow hgRNAs does not recover the early lineage tree as reliably (Fig. 6F). Combined, these results suggest that active mutagenesis during a differentiation event allows it to be recorded. They also suggest that when the developmental stage in which a lineage differentiates is known, hgRNA activity profiles (Fig. 2A) can aid in choosing the appropriate hgRNAs such that correct trees can be reliably obtained with just a few hgRNAs.

Interestingly, when only slow hgRNAs were considered in tree construction for early lineages, increasing the number of hgRNAs still resulted in improved outcomes (Fig. 6F). This observation suggests that even when hgRNAs have low activity levels at the time an event is being recorded, partial recordings from multiple hgRNAs can be combined to obtain a more complete recording. As another example, four different hgRNAs from embryo 3 predict a partial tree when considered on their own, yet the perfect tree is derived when all four are considered together

(fig. S11), further supporting the integrability of hgRNA recordings.

In two of our lineage-analyzed embryo samples (Fig. 5), we noted several hgRNAs in which all ICM-derived tissue samples (PE, VE, heart, limb) were dominated by a single mutant allele, whereas the corresponding trophoctoderm-derived tissue samples (DZ, JZ) displayed a more uniform distribution of multiple mutant alleles (Fig. 7). These profiles suggest that in these embryos, these hgRNAs mutated as the trophoctoderm and ICM lineages differentiated, and that fewer blastomeres led to the ICM than to the trophoctoderm. These observations are consistent with previously reported observations (33, 34) and suggest that hgRNA mutation profiles could be used to measure both the relationship between lineages and the relative number of cells that seed lineages.

Axis development in the brain

We next used developmentally barcoded mice to address lineages above the first lineages in the tree, with a focus on the establishment of the anterior-posterior (A-P) axis versus the lateral (L-R) axis in the brain. Patterning of the nervous system and its progenitors starts in gastrulation (E6.5) when the embryo has radial symmetry (35, 36). By E8.5, both A-P and L-R axes are established in the neural tube (Fig. 8A); however, it remains unclear which axis is established first (37, 38). At a morphological level they appear concurrently (39), and previous single-cell labeling and tracing experiments carried out *ex vivo* do not adequately address the issue (40). We analyzed two developmentally barcoded mice in the adult stage. In one, we dissected the left and right cortex and cerebellum, while in the other we additionally dissected the tectum. The cortex, tectum, and cerebellum respectively originate from embryonic forebrain (prosencephalon), midbrain (mesencephalon), and hindbrain (rhombencephalon) vesicles of the neural tube (Fig. 8A). From each region, two samples of neuronal nuclei were sorted (24). We also obtained

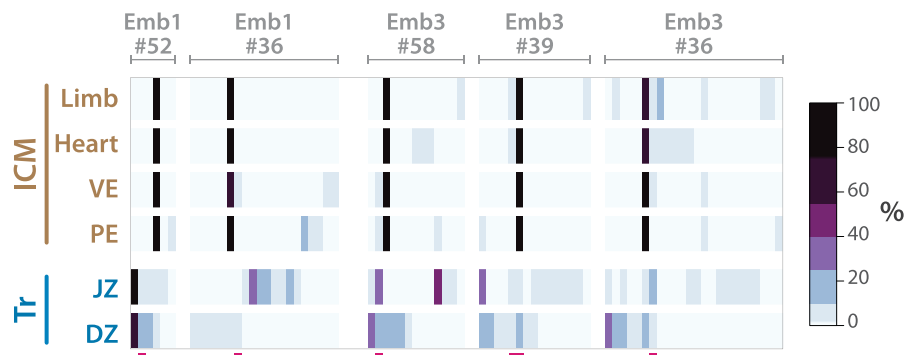


Fig. 7. Trophoctoderm and ICM barcodes show differences in their number of mutant hgRNA alleles. Five barcodes from two embryos in Fig. 5D are shown that distinguish trophoctoderm-derived and ICM-derived samples. Deep pink bars below each map mark highly recurring alleles that have been observed in more than 60% of all mice analyzed in Table 2. See table S6 for a numerical version of each barcode map. Only mutant alleles with a maximum abundance of more than 2% are shown.

samples of the blood and muscle from each mouse, both mesoderm-derived, to serve as outgroups. We then assembled the full barcode for each sample and applied clustering as before (Fig. 8, B and C, and fig. S12). In addition to segregating the mesoderm- and ectoderm-derived cells, the results clearly show that neurons from the left side of each brain region are more closely related to neurons from the right side of the same region than they are to neurons from either of the other two regions. Considering that no extensive migration of neuronal cell bodies between the regions sampled here has been reported (41), these results suggest that commitment to the A-P axis is established before commitment to the L-R axis in development of the central nervous system.

Similar to the first lineage tree analysis above (Fig. 6), we evaluated the robustness of the brain axis tree derivation as well as the contribution of each hgRNA in mouse 2 (Fig. 8, D and E) (24). We assigned topologies with all three left and right sample pairs placed closest to one another as correct, and those with two, one, or zero pairs placed as incomplete, partial, and wrong, respectively. We then calculated the distribution of tree derivation outcomes with all possible sub-

sets of active hgRNAs in mouse 2 (Fig. 8D). The results show that half of the combinations with only three hgRNAs derive a correct or partially correct topology, a ratio that only improves when including more hgRNAs. We also calculated the impact score of each hgRNA (Fig. 8E) (24). Relative to impact scores for the first lineage tree (Fig. 6E), we found smaller contributions by fast hgRNAs, which would be expected for lineages that segregate much later in development. Taken together, these results demonstrate that lineages across diverse developmental times are recorded in our developmentally barcoded mice and can be extracted.

Discussion

In this study, we created an hgRNA mouse line for in vivo barcoding and used it to generate developmentally barcoded mice in which lineage information is recorded in cell genomes and can be extracted and reconstructed. Our strategy to create the MARC1 line was designed to address challenges associated with in vivo barcoding in a mouse model. First, genetic manipulation of individual mouse embryos is more challenging than that of lower vertebrates. Therefore, a line with genomically integrated, stable, and heri-

table barcoding elements that can be activated by simply crossing with other lines is powerful, versatile, and shareable (24). Second, tracking development in mice demands that the system be capable of generating a great many barcodes with little overwriting or deletion (14). As such, we scattered hgRNAs throughout the genome instead of using a contiguous array, circumventing large deletion events that can occur with multiple adjacent cut sites (42–44) and can remove prior recordings. In fact, we estimate that less than 1% of all mutations resulted in unidentifiable alleles by removing amplification primer binding sites or all unique sequences (24). The scarcity of these unwanted deletion events led to a great success rate in analyzing barcoded mice (4/4 in Fig. 5, 2/2 in Fig. 8). Furthermore, as hgRNA loci in this scattered array accumulate mutations independently, their mutant alleles combine exponentially to create a large diversity of barcodes. Consequently, the 41 active MARC1 hgRNAs can in theory combine to create more than 10^{74} different barcodes $[\prod_{i=1}^{41} n_i]$, where n_i is the total number of observed mutant alleles for hgRNA i in this study, which is likely an underestimation (see above). Even only five fast and five mid hgRNAs can combine for roughly

Table 1. hgRNAs in the MARC1 founder male. TSS-to-PAM length, observed inheritance probability, and chromosome number for all 60 hgRNAs. See tables S1 to S3 for more details.

ID	Length (bp)	Inheritance (%)	Location	Class	ID	Length (bp)	Inheritance (%)	Location	Class
1	21	49.6	chr12	Slow	31	35	58.4	chr9	Inactive
2	35	55.2	chr7	Inactive	32	24	44	chr11	Mid
3	35	55.2	chr4	Inactive	33	21	55.2	chr13	Slow
4	21	29.6	chr14	Inactive	34	35	40	chr2	Mid
5	35	41.6	chr6	Inactive	35	25	52	chr19	Slow
6	21	54.4	chr2	Mid	36	21	45.6	chr8	Mid
7	34	43.2	chr19	Slow	37	35	51.2	chr4	Slow
8	30	16.8		Slow	38	21	38.4	chr13	Slow
9	35	51.2	chr19	Slow	39	25	54.4	chr7	Mid
10	25	42.4	chr4	Slow	40	35	45.6	chr9	Inactive
11	25	35.2	chr4	Inactive	41	21	38.4	chr2	Mid
12	35	48.8	chr17	Inactive	42	30	15.2		Slow
13	35	42.4	chr6	Slow	43	35	48.8	chr16	Slow
14	35	43.2	chr7	Inactive	44	21	51.2	chr14	Inactive
15	34	60.8	chr11	Slow	45	30	7.2		Inactive
16	35	47.2	chr9	Inactive	46	25	47.2	chr18	Slow
17	21	59.2	chr11	Slow	47	35	48.8	chr6	Inactive
18	25	50.4	chr13	Slow	48	35	34.4	chr5	Slow
19	21	39.2	chr7	Fast	49	35	58.4	chr11	Slow
20	21	49.6	chr18	Fast	50	21	44	chr10	Mid
21	35	40	chr2	Slow	51	25	36.8	chr11	Slow
22	21	42.4	chr4	Mid	52	21	53.6	chr2	Fast
23	21	80.8	chr1&4	Mid	53	35	47.2	chr3	Inactive
24	21	51.2	chr3	Inactive	54	35	73.6	chr10&13	Slow
25	25	48	chr8	Inactive	55	25	68.8	chr14	Slow
26	35	46.4	chr4	Inactive	56	21	55.2	chr2	Slow
27	21	39.2	chr2	Slow	57	34	44	chr11	Inactive
28	25	40	chr5	Slow	58	21	43.2	chr6	Fast
29	35	52.8	chr4	Slow	59	21	50.4	chr2	Fast
30	35	55.2	chr5	Inactive	60	25	51.2	chr3	Slow

10^{23} different barcodes ($200^5 \times 300^5$, where 200 and 300 are the observed average number of mutant alleles for fast and mid hgRNAs, respectively). This remarkable diversity is adequate for uniquely barcoding every one of the $\sim 10^{10}$ cells in a mouse. Furthermore, assuming a perfect binary developmental tree as a first-order approximation, this diversity is adequate for uniquely marking all the $\sim 2 \times 10^{10}$ internal and terminal nodes of the mouse developmental tree.

Close analysis of the nature of mutant alleles in hgRNA barcodes showed the interplay between target site sequence and Cas9-induced double-strand breaks that determines the possible NHEJ outcomes (Fig. 3). Specifically, short indels underlie recurring NHEJ outcomes. No-

table among these is a recurring duplication of the base at the -4 position relative to the PAM in a majority of active hgRNAs. The most likely explanation for this observation is Cas9 creating staggered cuts that produce a single-base 5' overhang (Fig. 3K), because a terminal transferase activity would not duplicate the base adjacent to the cut site, and RuvC exonuclease activity on the noncomplementary strand would not result in an insertion at all. Whether Cas9 creates blunt or staggered ends in vivo has been a subject of debate. Our observation in mice, combined with a recent report in yeast (45) and previous in vitro and in vivo evidence (44, 46-50), clarifies that Cas9 can create staggered ends as well as blunt ends, although the ratio of the two is unknown as of yet.

By crossing the MARC1 line with a line that constitutively expresses Cas9, we generated developmentally barcoded mice in which lineage information is recorded in the hgRNA barcodes. We were able to reconstruct parts of the lineage tree using these mice, with the first branches that emerge after the zygote, up to some of the germ layer, neuroectoderm, and the neural tube branches (Figs. 5 and 8). We find remarkable robustness and flexibility in these recordings (Figs. 6 and 8D). Specifically, there is overlap in recordings made by various hgRNAs, and therefore the derived lineage tree is robust to removing any part of the barcode. Furthermore, partial nonoverlapping recordings from different hgRNAs can be integrated to reconstruct a complete tree. Combined with evidence of sustained hgRNA

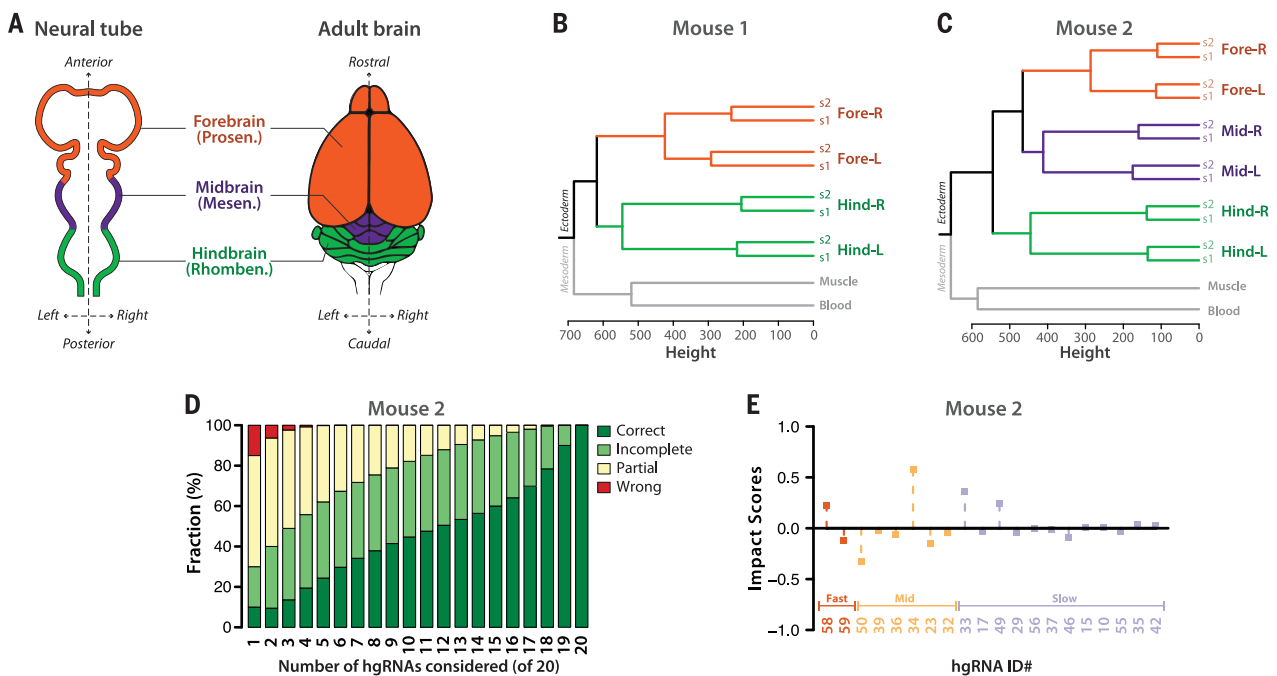


Fig. 8. The anterior-posterior axis is established before the left-right axis in the development of the brain. (A) Dorsal view of the neural tube and superior view of the adult brain in mouse. The primary brain vesicles in the neural tube and their corresponding structures in the adult brain are shown. (B and C) Calculated trees based on hgRNA barcodes in two adult mice. See fig. S12 for the full barcodes. (D) Distribution of tree calculation outcomes for mouse 2 when only including m of the n hgRNAs (nC_m combinations). Only hgRNAs with at least a 7% mutation rate in one of the samples were considered. (E) Impact score of each hgRNA in the early lineage tree of mouse 2.

Table 2. Breakdown of all mice used for hgRNA activity analysis according to developmental stage and number of samples obtained per mouse.

Samples per mouse	Stage of development										Total mice	Total samples
	E3.5	E6.5	E7.5	E8.5	E10.5	E12.5	E14.5	E15.5	E16.5	Adult		
One	11	6	3	3	1	0	0	0	0	45	69	69
Two	0	0	4	6	1	0	0	0	0	0	11	22
Three	0	0	1	0	0	0	0	0	0	0	1	3
Four	0	0	0	0	7	6	0	0	2	0	15	60
Five	0	0	0	0	0	0	0	0	0	0	0	0
Six	0	0	0	0	0	6	0	0	0	0	6	36
Total	11	6	8	9	9	12	0	0	2	45	102	190

mutagenesis throughout gestation, these results suggest that developmentally barcoded mice embody information from various stages of development in embryonic and extraembryonic tissues. Extracting such information will be a matter of the type of question being investigated and an ability to isolate cells from relevant lineages. Another interesting possibility is to create different types of barcoded mice by crossing the MARCI line with other *S. pyogenes* Cas9 lines. Among these are inducible Cas9s (51, 52), ones with different activity levels (53), tissue- or lineage-specific versions based on Cre drivers (27, 53), or base-editing Cas9s (54, 55). Such barcoded mice may enhance the capabilities of the system, overcome its shortcomings (24), or better focus its potential on specific problems.

Our results provide a platform for in vivo barcoding and lineage tracing in the mouse. Although we have focused here on the recording aspect of in vivo developmental barcoding, more effective readout strategies—in particular, those with transcriptome-coupled single-cell readouts (19, 21, 22) or with in situ readouts (56)—will be necessary. Finally, in addition to lineage-tracing applications, this platform may also be applied to recording cellular signals over time (16, 17, 57–59) and uniquely barcoding each cell in a tissue or an organism for identification purposes, such as for connectome mapping in the brain (60–62).

Methods summary

All animal procedures were approved by the Harvard University Institutional Animal Care and Use Committee (IACUC). For embryonic samples, the MARCI founder was crossed with a Cas9 knock-in female. Pregnant females were then dissected at the desired embryonic time points, designating noon of the day of vaginal plug detection as E0.5. For isolating neurons from adult barcoded female mice, brains were dissected into the regions of interest and homogenized. Nuclei were isolated from the homogenate by gradient ultracentrifugation, labeling with a NeuN antibody, and sorting the NeuN-positive fraction in flow cytometry. From all obtained samples, DNA was extracted and amplified with specific primers for hgrNA loci. The resulting amplicons were sequenced with paired ends and analyzed to identify the hgrNA itself according to the identifier sequence, and the mutant allele according to the spacer sequence. The sequencing results were processed and filtered to obtain a list of high-confidence unique spacer-identifier pairs observed in each sample and their respective abundances. For obtaining lineage trees, these lists were converted into frequency matrices and clustered hierarchically using Ward's criterion. All procedures for the experiments and data analyses are described in detail in the supplementary materials.

REFERENCES AND NOTES

- M. H. Kaufman, *The Atlas of Mouse Development* (Elsevier, 1992).
- J. E. Sulston, E. Schierenberg, J. G. White, J. N. Thomson, The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983). doi: [10.1016/0012-1606\(83\)90201-4](https://doi.org/10.1016/0012-1606(83)90201-4); pmid: 6684600
- D. L. Riddle, T. Blumenthal, B. J. Meyer, J. R. Priess, in *C. elegans II* (Cold Spring Harbor Laboratory Press, ed. 2, 1997).
- D. A. Weisblat, R. T. Sawyer, G. S. Stent, Cell lineage analysis by intracellular injection of a tracer enzyme. *Science* **202**, 1295–1298 (1978). doi: [10.1126/science.725606](https://doi.org/10.1126/science.725606); pmid: 725606
- C. Walsh, C. L. Cepko, Widespread dispersion of neuronal clones across functional regions of the cerebral cortex. *Science* **255**, 434–440 (1992). doi: [10.1126/science.1734520](https://doi.org/10.1126/science.1734520); pmid: 1734520
- S. M. Dymecki, H. Tomasiewicz, Using Flp-recombinase to characterize expansion of Wnt1-expressing neural progenitors in the mouse. *Dev. Biol.* **201**, 57–65 (1998). doi: [10.1006/dbio.1998.8971](https://doi.org/10.1006/dbio.1998.8971); pmid: 9733573
- K. Kretschmar, F. M. Watt, Lineage tracing. *Cell* **148**, 33–45 (2012). doi: [10.1016/j.cell.2012.01.002](https://doi.org/10.1016/j.cell.2012.01.002); pmid: 22265400
- S. H. Naik, T. N. Schumacher, L. Perić, Cellular barcoding: A technical appraisal. *Exp. Hematol.* **42**, 598–608 (2014). doi: [10.1016/j.exphem.2014.05.003](https://doi.org/10.1016/j.exphem.2014.05.003); pmid: 24996012
- A. Gerrits et al., Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* **115**, 2610–2618 (2010). doi: [10.1182/blood-2009-06-229757](https://doi.org/10.1182/blood-2009-06-229757); pmid: 20093403
- M. B. Woodworth, K. M. Girsakis, C. A. Walsh, Building a lineage from single cells: Genetic techniques for cell lineage tracking. *Nat. Rev. Genet.* **18**, 230–244 (2017). doi: [10.1038/nrg.2016.159](https://doi.org/10.1038/nrg.2016.159); pmid: 28111472
- J. Ma, Z. Shen, Y.-C. Yu, S.-H. Shi, Neural lineage tracing in the mammalian brain. *Curr. Opin. Neurobiol.* **50**, 7–16 (2018). doi: [10.1016/j.conb.2017.10.013](https://doi.org/10.1016/j.conb.2017.10.013); pmid: 29125960
- I. D. Peikon, D. I. Gizatullina, A. M. Zador, In vivo generation of DNA sequence diversity for cellular barcoding. *Nucleic Acids Res.* **42**, e127 (2014). doi: [10.1093/nar/gku604](https://doi.org/10.1093/nar/gku604); pmid: 25013177
- A. McKenna et al., Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016). doi: [10.1126/science.aaf7907](https://doi.org/10.1126/science.aaf7907); pmid: 27229144
- R. Kalhor, P. Mali, G. M. Church, Rapidly evolving homing CRISPR barcodes. *Nat. Methods* **14**, 195–200 (2017). doi: [10.1038/nmeth.4108](https://doi.org/10.1038/nmeth.4108); pmid: 27918539
- K. L. Frieda et al., Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017). doi: [10.1038/nature20777](https://doi.org/10.1038/nature20777); pmid: 27869821
- S. D. Perić, C. H. Cui, T. K. Lu, Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, aag0511 (2016). doi: [10.1126/science.aag0511](https://doi.org/10.1126/science.aag0511); pmid: 27540006
- R. U. Sheth, S. S. Yim, F. L. Wu, H. H. Wang, Multiplex recording of cellular events over time on CRISPR biological tape. *Science* **358**, 1457–1461 (2017). doi: [10.1126/science.aao0958](https://doi.org/10.1126/science.aao0958); pmid: 29170279
- S. T. Schmidt, S. M. Zimmerman, J. Wang, S. K. Kim, S. R. Quake, Quantitative Analysis of Synthetic Cell Lineage Tracing Using Nuclease Barcoding. *ACS Synth. Biol.* **6**, 936–942 (2017). doi: [10.1021/acssynbio.6b00309](https://doi.org/10.1021/acssynbio.6b00309); pmid: 28264564
- A. Alemany, M. Florescu, C. S. Baron, J. Peterson-Maduro, A. van Oudenaarden, Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018). doi: [10.1038/nature25969](https://doi.org/10.1038/nature25969); pmid: 29590089
- G. P. Flowers, L. D. Sanor, C. M. Crews, Lineage tracing of genome-edited alleles reveals high fidelity axolotl limb regeneration. *eLife* **6**, e25726 (2017). doi: [10.7554/eLife.25726](https://doi.org/10.7554/eLife.25726); pmid: 28917058
- B. Spanjaard et al., Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018). doi: [10.1038/nbt.4124](https://doi.org/10.1038/nbt.4124); pmid: 29644996
- B. Raj et al., Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018). doi: [10.1038/nbt.4103](https://doi.org/10.1038/nbt.4103); pmid: 29608178
- P. Mali et al., RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013). doi: [10.1126/science.1232033](https://doi.org/10.1126/science.1232033); pmid: 23287722
- See supplementary materials.
- C. L. Markert, R. M. Petters, Manufactured hexaparental mice show that adults are derived from three embryonic cells. *Science* **202**, 56–58 (1978). doi: [10.1126/science.694518](https://doi.org/10.1126/science.694518); pmid: 694518
- Z. Wang, R. Jaenisch, At most three ES cells contribute to the somatic lineages of chimeric mice and of mice produced by ES-tetraploid complementation. *Dev. Biol.* **275**, 192–201 (2004). doi: [10.1016/j.ydbio.2004.06.026](https://doi.org/10.1016/j.ydbio.2004.06.026); pmid: 15464582
- R. J. Platt et al., CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014). doi: [10.1016/j.cell.2014.09.014](https://doi.org/10.1016/j.cell.2014.09.014); pmid: 25263330
- M. T. Lee, A. R. Bonneau, A. J. Giraldez, Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* **30**, 581–613 (2014). doi: [10.1146/annurev-cellbio-100913-013027](https://doi.org/10.1146/annurev-cellbio-100913-013027); pmid: 25150012
- C. C. Lu, J. Brennan, E. J. Robertson, From fertilization to gastrulation: Axis formation in the mouse embryo. *Curr. Opin. Genet. Dev.* **11**, 384–392 (2001). doi: [10.1016/S0959-437X\(00\)00208-2](https://doi.org/10.1016/S0959-437X(00)00208-2); pmid: 11448624
- J. Rossant, J. C. Cross, in *Mouse Development: Patterning, Morphogenesis, and Organogenesis*, J. Rossant, P. P. L. Tam, Eds. (Academic Press, 2002), pp. 155–180.
- D. G. Simmons, J. C. Cross, Determinants of trophoblast lineage and cell subtype specification in the mouse placenta. *Dev. Biol.* **284**, 12–24 (2005). doi: [10.1016/j.ydbio.2005.05.010](https://doi.org/10.1016/j.ydbio.2005.05.010); pmid: 15963972
- J. Felsenstein, *Inferring Phylogenies* (Sinauer Associates, 2004).
- C. Y. Leung, M. Zhu, M. Zernicka-Goetz, Polarity in Cell-Fate Acquisition in the Early Mouse Embryo. *Curr. Top. Dev. Biol.* **120**, 203–234 (2016). doi: [10.1016/bs.ctdb.2016.04.008](https://doi.org/10.1016/bs.ctdb.2016.04.008); pmid: 27475853
- S. Anani, S. Bhat, N. Honma-Yamanaka, D. Krawchuk, Y. Yamanaka, Initiation of Hippo signaling is linked to polarity rather than to cell position in the pre-implantation mouse embryo. *Development* **141**, 2813–2824 (2014). doi: [10.1242/dev.102726](https://doi.org/10.1242/dev.102726); pmid: 24948601
- R. S. Beddington, E. J. Robertson, Axis development and early asymmetry in mammals. *Cell* **96**, 195–209 (1999). doi: [10.1016/S0092-8674\(00\)80560-7](https://doi.org/10.1016/S0092-8674(00)80560-7); pmid: 9988215
- Z. Molnár, D. J. Price, in *Kaufman's Atlas of Mouse Development Supplement*, R. Baldoock, J. Bard, D. R. Davidson, G. Morriss-Kay, Eds. (Elsevier, 2016), pp. 239–252.
- S. J. Arnold, E. J. Robertson, Making a commitment: Cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10**, 91–103 (2009). doi: [10.1038/nrm2618](https://doi.org/10.1038/nrm2618); pmid: 19129791
- C. Kiecker, A. Lumsden, The role of organizers in patterning the nervous system. *Annu. Rev. Neurosci.* **35**, 347–367 (2012). doi: [10.1146/annurev-neuro-062111-150543](https://doi.org/10.1146/annurev-neuro-062111-150543); pmid: 22462542
- C. R. Altmann, A. H. Brivanlou, Neural patterning in the vertebrate embryo. *Int. Rev. Cytol.* **203**, 447–482 (2001). doi: [10.1016/S0074-7696\(01\)03013-3](https://doi.org/10.1016/S0074-7696(01)03013-3); pmid: 11131523
- K. A. Lawson, J. J. Meneses, R. A. Pedersen, Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**, 891–911 (1991). pmid: 1821858
- R. Ayala, T. Shu, L.-H. Tsai, Trekking across the brain: The journey of neuronal migration. *Cell* **128**, 29–43 (2007). doi: [10.1016/j.cell.2006.12.021](https://doi.org/10.1016/j.cell.2006.12.021); pmid: 17218253
- H. J. Lee, E. Kim, J.-S. Kim, Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res.* **20**, 81–89 (2010). doi: [10.1101/gr.099747.109](https://doi.org/10.1101/gr.099747.109); pmid: 19952142
- M. C. Canver et al., Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289**, 21312–21324 (2014). doi: [10.1074/jbc.M114.564625](https://doi.org/10.1074/jbc.M114.564625); pmid: 24907273
- S. M. Byrne, L. Ortiz, P. Mali, J. Aach, G. M. Church, Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Res.* **43**, e21 (2015). doi: [10.1093/nar/gku1246](https://doi.org/10.1093/nar/gku1246); pmid: 25414332
- B. R. Lemos et al., CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2040–E2047 (2018). doi: [10.1073/pnas.1716855115](https://doi.org/10.1073/pnas.1716855115); pmid: 29440496
- G. Gasunas, R. Barrangou, P. Horvath, V. Siksnys, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2579–E2586 (2012). doi: [10.1073/pnas.1208507109](https://doi.org/10.1073/pnas.1208507109); pmid: 22949671
- M. Jinek et al., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012). doi: [10.1126/science.1225829](https://doi.org/10.1126/science.1225829); pmid: 22745249
- M. Jinek et al., Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014). doi: [10.1126/science.1247997](https://doi.org/10.1126/science.1247997); pmid: 24505130
- Z. Zuo, J. Liu, Cas9-catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics

- Simulations. *Sci. Rep.* **5**, 37584 (2016). doi: [10.1038/srep37584](https://doi.org/10.1038/srep37584); pmid: [27874072](https://pubmed.ncbi.nlm.nih.gov/27874072/)
50. S. Q. Tsai *et al.*, GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015). doi: [10.1038/nbt.3117](https://doi.org/10.1038/nbt.3117); pmid: [25513782](https://pubmed.ncbi.nlm.nih.gov/25513782/)
 51. L. E. Dow *et al.*, Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.* **33**, 390–394 (2015). doi: [10.1038/nbt.3155](https://doi.org/10.1038/nbt.3155); pmid: [25690852](https://pubmed.ncbi.nlm.nih.gov/25690852/)
 52. L. R. Polstein, C. A. Gersbach, A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat. Chem. Biol.* **11**, 198–200 (2015). doi: [10.1038/nchembio.1753](https://doi.org/10.1038/nchembio.1753); pmid: [25664691](https://pubmed.ncbi.nlm.nih.gov/25664691/)
 53. S.-H. Chiou *et al.*, Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. *Genes Dev.* **29**, 1576–1585 (2015). doi: [10.1101/gad.264861.115](https://doi.org/10.1101/gad.264861.115); pmid: [26178787](https://pubmed.ncbi.nlm.nih.gov/26178787/)
 54. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016). doi: [10.1038/nature17946](https://doi.org/10.1038/nature17946); pmid: [27096365](https://pubmed.ncbi.nlm.nih.gov/27096365/)
 55. N. M. Gaudelli *et al.*, Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017). doi: [10.1038/nature24644](https://doi.org/10.1038/nature24644); pmid: [29160308](https://pubmed.ncbi.nlm.nih.gov/29160308/)
 56. J. H. Lee *et al.*, Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014). doi: [10.1126/science.1250212](https://doi.org/10.1126/science.1250212); pmid: [24578530](https://pubmed.ncbi.nlm.nih.gov/24578530/)
 57. A. H. Marblestone *et al.*, Physical principles for scalable neural recording. *Front. Comput. Neurosci.* **7**, 137 (2013). doi: [10.3389/fncom.2013.00137](https://doi.org/10.3389/fncom.2013.00137); pmid: [24187539](https://pubmed.ncbi.nlm.nih.gov/24187539/)
 58. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* **547**, 345–349 (2017). doi: [10.1038/nature23017](https://doi.org/10.1038/nature23017); pmid: [28700573](https://pubmed.ncbi.nlm.nih.gov/28700573/)
 59. S. Punthambaker, R. I. Hume, Potent and long-lasting inhibition of human P2X2 receptors by copper. *Neuropharmacology* **77**, 167–176 (2014). doi: [10.1016/j.neuropharm.2013.09.001](https://doi.org/10.1016/j.neuropharm.2013.09.001); pmid: [24067922](https://pubmed.ncbi.nlm.nih.gov/24067922/)
 60. D. Cai, K. B. Cohen, T. Luo, J. W. Lichtman, J. R. Sanes, Improved tools for the Brainbow toolbox. *Nat. Methods* **10**, 540–547 (2013). doi: [10.1038/nmeth.2450](https://doi.org/10.1038/nmeth.2450)
 61. J. M. Kebschull *et al.*, High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. *Neuron* **91**, 975–987 (2016). doi: [10.1016/j.neuron.2016.07.036](https://doi.org/10.1016/j.neuron.2016.07.036); pmid: [27545715](https://pubmed.ncbi.nlm.nih.gov/27545715/)
 62. G. Church, A. Marblestone, R. Kalhor, in *The Future of the Brain*, G. Marcus, J. Freeman, Eds. (Princeton Univ. Press, 2014), pp. 50–64.
 63. M. R. Lieber, T. E. Wilson, SnapShot: Nonhomologous DNA end joining (NHEJ). *Cell* **142**, 496–496.e1 (2010). doi: [10.1016/j.cell.2010.07.035](https://doi.org/10.1016/j.cell.2010.07.035); pmid: [20691907](https://pubmed.ncbi.nlm.nih.gov/20691907/)

ACKNOWLEDGMENTS

We thank R. Jaenisch for generously sharing reagents; J. Aach for critical reading of the manuscript and helpful comments; A. Vernet for assistance with animal husbandry, G. Cuneo for assistance in FACS; L. Wu and the Genome Modification Facility for blastocyst injections as well as helpful discussions; S. Kennedy, J. Young, B. Fields, and A. Reslow for generously sharing equipment and providing technical advice; and M. Warman, C. Tabin, C. Cepko, Y. Stelzer, S. Shipman, J. Macklis, B. Khalaj, N. Davidsohn, A. Ng, and R. Kohman for helpful discussions. **Funding:** Supported by NIH grants MH103910 and HG005550

(G.M.C.), the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/Interior Business Center (DoI/IBC) contract number D16PC00008 (G.M.C.), Burroughs Wellcome Fund 1013926 (P.M.), and NIH grants R01HG009285, R01CA222826, and R01GM123313 (P.M.). **Author contributions:** R.K., P.M., and G.M.C. conceived the study; R.K. designed and carried out the experiments and analyzed the data; K.K. analyzed the genomic location data and formulated other analyses; L.M. and K.L. assisted in mouse genotyping, dissections, library preparations, and sequencing; A.G. supervised animal husbandry; R.K., K.K., and P.M. interpreted the data and wrote the manuscript with input from all other authors; and G.M.C. supervised the project. **Competing interests:** The authors declare no competing financial interest. G.M.C.'s competing financial interests can be found at v.ht/PHNc. **Data and materials availability:** All sequencing data are available in the Sequence Read Archive (SRP155997); all other data are available in supplementary materials, and other materials are available upon request.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6405/eaat9804/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 to S7
References (64–85)

26 April 2018; resubmitted 19 July 2018
Accepted 2 August 2018
Published online 9 August 2018
[10.1126/science.aat9804](https://doi.org/10.1126/science.aat9804)

Developmental barcoding of whole mouse via homing CRISPR

Reza Kalhor, Kian Kalhor, Leo Mejia, Kathleen Leeper, Amanda Graveline, Prashant Mali and George M. Church

Science **361** (6405), eaat9804.

DOI: 10.1126/science.aat9804 originally published online August 9, 2018

Lineage tracing in mouse using CRISPR

A homing guide RNA (hgRNA) that directs CRISPR-Cas9 to its own DNA locus can diversify its sequence and act as an expressed genetic barcode. Kalhor *et al.* engineered a mouse line carrying 60 independent loci of hgRNAs, thus generating a large number of unique barcodes in various embryonic and extraembryonic tissues in fully developed mice. This method demonstrates lineage tracing from the very first branches of the development tree up to organogenesis events and was used to elucidate embryonic brain patterning.

Science, this issue p. eaat9804

ARTICLE TOOLS

<http://science.sciencemag.org/content/361/6405/eaat9804>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2018/08/08/science.aat9804.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/361/6405/864.full>
<http://stm.sciencemag.org/content/scitransmed/9/372/eaah3480.full>
<http://stm.sciencemag.org/content/scitransmed/10/449/eaao3240.full>
<http://stm.sciencemag.org/content/scitransmed/9/418/eaan8081.full>
<http://stm.sciencemag.org/content/scitransmed/8/360/360ra134.full>

REFERENCES

This article cites 75 articles, 23 of which you can access for free
<http://science.sciencemag.org/content/361/6405/eaat9804#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2018 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works