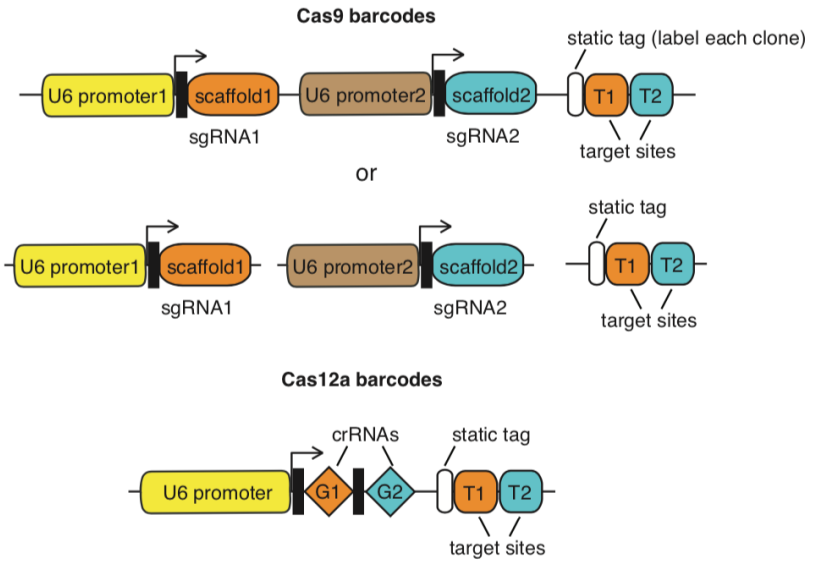
Supplemental Figures

A



B

Parameter	Cas9	Cas12a
Barcode scaffold length	~ 100nt / guideRNA (Jinek et al., 2012; Cong et al.; Mali et al., 2013)	~ 40nt / guideRNA (Zetsche et al., 2017)
Promoter Structure	1 promoter / gRNA (~300nt length) (Bowling et al.; Quinn, Jones et al.)	single promoter (crRNA array) (Zetsche et al., 2017)
Lentiviral Recombination	~50% (Hill et al., 2018)	established triple knockout efficacy (DeWeirdt et al., 2020)
Cloning Complexity	one step per sgRNA (Bowling et al.; Quinn, Jones et al.)	single-step (Zetsche et al., 2017)
Specificity (OTI)	> 2.0 (Kim et al., 2016)	0.005-0.267 (Kim et al., 2016)

Figure S1. Comparison between Cas9 and Cas12a barcode systems and features, Related to Figure 2. (A) Cas9 barcode systems contain multiplexed sgRNAs from separate U6 promoters, which can be contained in a single vector with target sites (top) or two vectors (middle). In contrast, a single vector with one U6 promoter is needed for Cas12a barcoding (bottom). (B) Table of key parameters that characterize the complexity of crRNA cloning and potential complications and considerations, like off-target effects, when performing multiplexed genome editing with either Cas9 or Cas12a.

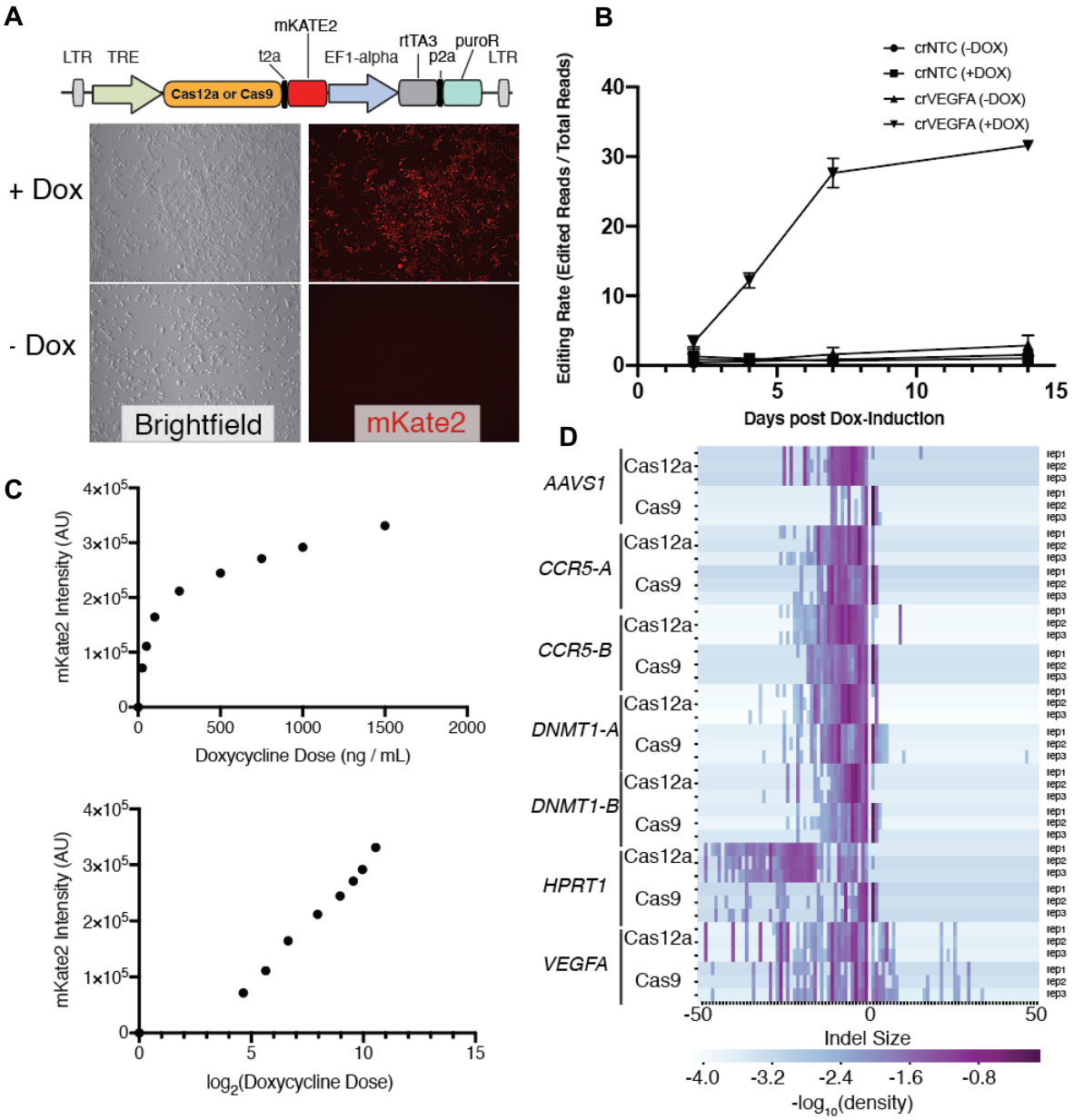

A
LTR
TRE
Cas12a or Cas9
t2a
mKATE2
EF1-alpha
rtTA3
p2a
puroR
LTR
+ Dox
- Dox
Brightfield
mKate2
B
crNTC (-DOX)
crNTC (+DOX)
crVEGFA (-DOX)
crVEGFA (+DOX)
Editing Rate (Edited Reads / Total Reads)
Days post Dox-Induction
C
mKate2 Intensity (AU)
Doxycycline Dose (ng / mL)
log2(Doxycycline Dose)
D
AAVS1
CCR5-A
CCR5-B
DNMT1-A
DNMT1-B
HPRT1
VEGFA
Cas12a
Cas9
Indel Size
-log10(density)
E
barcode 1
barcode 2
barcode 3
fraction of reads containing 2nd site editing
inter-site deletion (PAM collapse)
no deletion of PAM

Figure S2. Establishment of doxycycline-inducible Cas effector system for benchmarking editing outcome diversity, Related to Figure 2 (A) Vector schematic showing mKate2 fusion to either Cas12a or Cas9 under a TRE and a separate EF1-alpha-driven ORF expressing rtTA3-p2a-puroR (top). Micrograph showing doxycycline-dependent expression of mKate2 (bottom). (B) Doxycycline-dependent genome editing with a crRNA designed against *VEGFA* versus a non-targeting control over time. (C) Relationship between FACS-based measurement of mKate2 intensity, which is linked to Cas12a through a t2a peptide, and doxycycline dose. (bottom) Log-transform data. (D) Heatmap indicating the frequency (number of reads) of indel events with different lengths across endogenous loci that contain proximal Cas12a and Cas9 PAMs. Broader indel distributions are generally observed due to Cas12a editing rather than by Cas9. (E) Quantification of edits at the second target site in the presence or absence of an inter-site deletion (PAM collapsing) event across three barcodes tested.

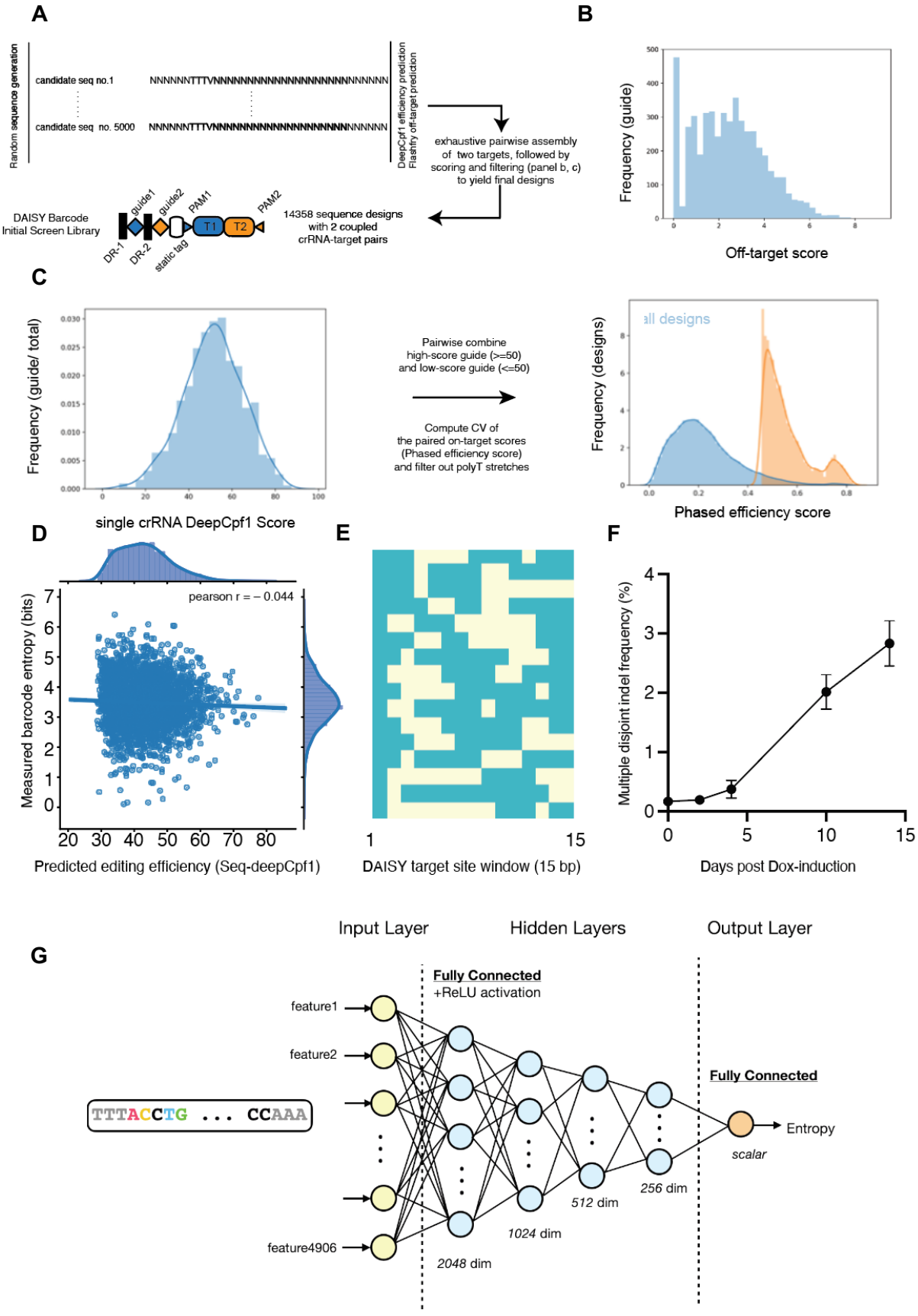

A
Random sequence generation
candidate seq no.1
NNNNNNTTTVNNNNNNNNNNNNNNNNNNNNNNNNNNNN
candidate seq no. 5000
NNNNNNTTTVNNNNNNNNNNNNNNNNNNNNNNNNNNNN
DeepCpf1 efficiency prediction
Flashfry off-target prediction
exhaustive pairwise assembly of two targets, followed by scoring and filtering (panel b, c) to yield final designs
DAISY Barcode Initial Screen Library
DR-1
guide1
DR-2
guide2
static tag
PAM1
T1
T2
PAM2
14358 sequence designs with 2 coupled crRNA-target pairs
B
Frequency (guide)
Off-target score
C
Frequency (guide/ total)
single crRNA DeepCpf1 Score
Pairwise combine high-score guide (>=50) and low-score guide (<=50)
Compute CV of the paired on-target scores (Phased efficiency score) and filter out polyT stretches
Frequency (designs)
all designs
Phased efficiency score
D
Measured barcode entropy (bits)
pearson r = – 0.044
Predicted editing efficiency (Seq-deepCpf1)
E
1
15
DAISY target site window (15 bp)
F
Multiple disjoint indel frequency (%)
Days post Dox-induction
G
Input Layer
Hidden Layers
Output Layer
Fully Connected
+ReLU activation
feature1
feature2
feature4906
TTTACCTG ... CCAAA
2048 dim
1024 dim
512 dim
256 dim
Fully Connected
Entropy
scalar

Figure S3. DAISY library design and analysis to enable machine learning optimization with a deep learning model, Related to Figure 3 (A) Schematic of DAISY library design workflow. The final DAISY cassette is shown in which two crRNAs and two target sites are assembled, generating the final cassettes which are uniquely identifiable by the design-BC for screening. (B) The 5000 randomly generated sequences were filtered based on a custom off-target score that takes into account the frequency of off-target sites and the Hamming distance between the off-target site and the guide sequence. (C) Construction of DAISYs using Seq-deepCpf1 efficiency scoring and pairwise assembly into phased designs with filtering out of polyT sequences that may inhibit PolIII transcription. (D) Correlation between measured barcode entropy and predicted editing efficiency scores (average of two target sites) using Seq-deepCpf1. **(**E) Heatmap of multiple disjoint indels within an individual target site window of a DAISY barcode. Light green blocks indicate sites of indels and dark green indicates unedited sequences. (F) The frequency of putative retargeting events (number of editing outcomes that contain multiple indels within a target site / total editing outcomes) over time. (G) Neural network architecture for the entropy prediction model. The input is 4906-dimensional features. There are 4 fully connected hidden layers with ReLU activation and decaying dimensions. The output layer is a fully connected layer that outputs a scalar entropy prediction. The parameters are trained with the Adam optimizer with weight decay.

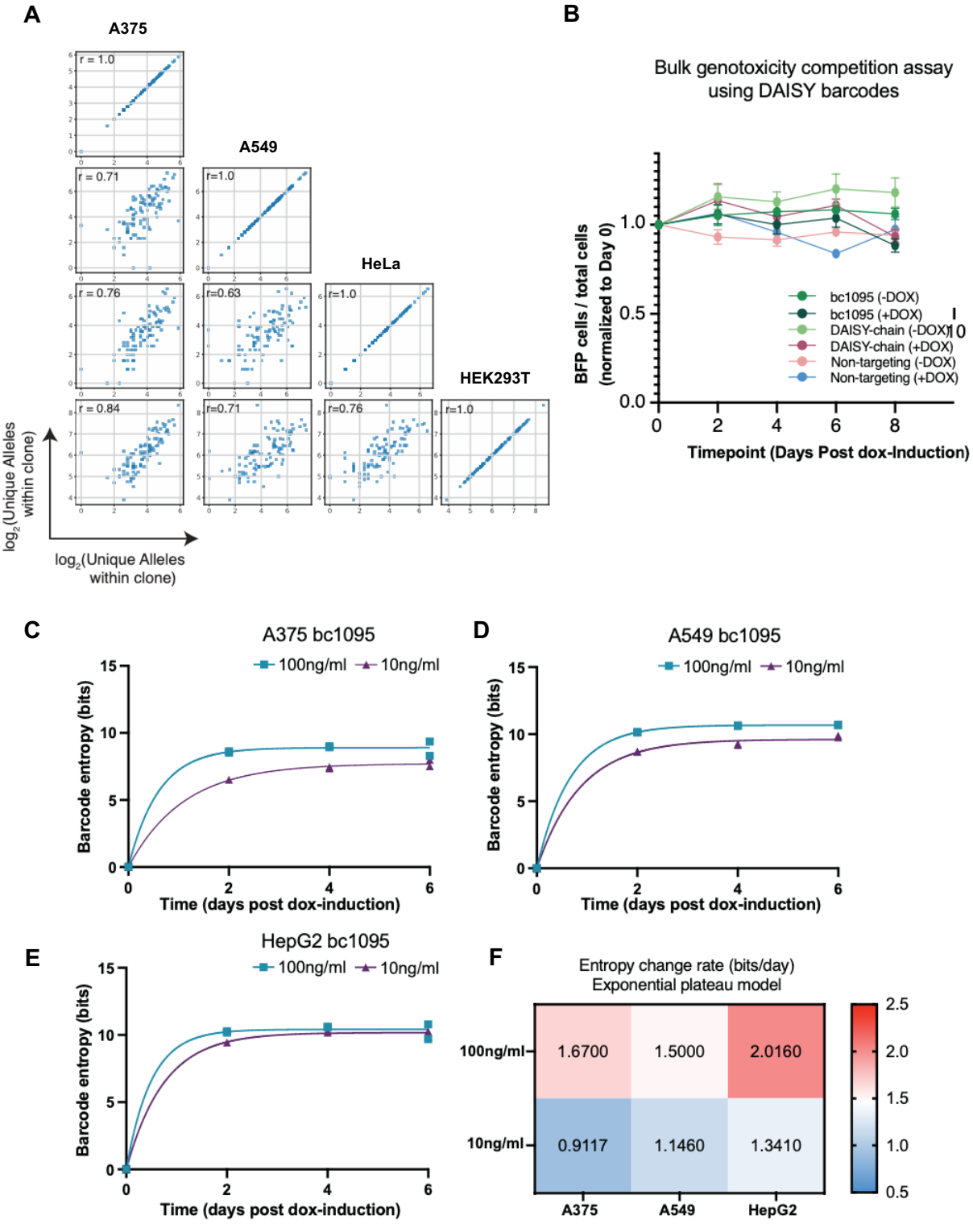

A
A375
A549
HeLa
HEK293T
r = 1.0
r = 0.71
r=1.0
r = 0.76
r=0.63
r=1.0
r = 0.84
r=0.71
r=0.76
r=1.0
log2(Unique Alleles within clone)
log2(Unique Alleles within clone)
B
Bulk genotoxicity competition assay using DAISY barcodes
BFP cells / total cells (normalized to Day 0)
1.0
0.5
0.0
0
2
4
6
8
Timepoint (Days Post dox-Induction)
bc1095 (-DOX)
bc1095 (+DOX)
DAISY-chain (-DOX)
DAISY-chain (+DOX)
Non-targeting (-DOX)
Non-targeting (+DOX)
C
A375 bc1095
100ng/ml
10ng/ml
Barcode entropy (bits)
15
10
5
0
0
2
4
6
Time (days post dox-induction)
D
A549 bc1095
100ng/ml
10ng/ml
Barcode entropy (bits)
15
10
5
0
0
2
4
6
Time (days post dox-induction)
E
HepG2 bc1095
100ng/ml
10ng/ml
Barcode entropy (bits)
15
10
5
0
0
2
4
6
Time (days post dox-induction)
F
Entropy change rate (bits/day)
Exponential plateau model
100ng/ml
10ng/ml
1.6700
1.5000
2.0160
0.9117
1.1460
1.3410
A375
A549
HepG2
2.5
2.0
1.5
1.0
0.5

Figure S4. Assessment of the portability, genotoxicity, and tunability of DAISY barcoding, Related to Figures 4 and 5 (A**)** Each dot in the dot plot represents a clonal population uniquely marked by a static tag sequence. For each cell line, the top 100 represented clones were rank-ordered by their population size estimated by the number of reads corresponding to each clone. Then the unique allele counts within each clone were plotted across cell line contexts. (B) A competition assay was performed where cells transduced with a lentiviral construct containing a DAISY barcode and BFP were mixed with parental cells that do not express the DAISY barcode. In addition, a non-targeting control was included to control for viability effects due to lentiviral transduction. The ratio of BFP-positive cells to total cells was measured overtime using FACS and used as a proxy for viability. (C-E) Change in the barcode entropy over time using 10ng/ml and 100ng/ml doxycycline in A375 (C), A549 (D) and HepG2 (E). (F) Rate kinetics of barcode entropy (based on the Exponential plateau model) across doxycycline dosages and cell lines.

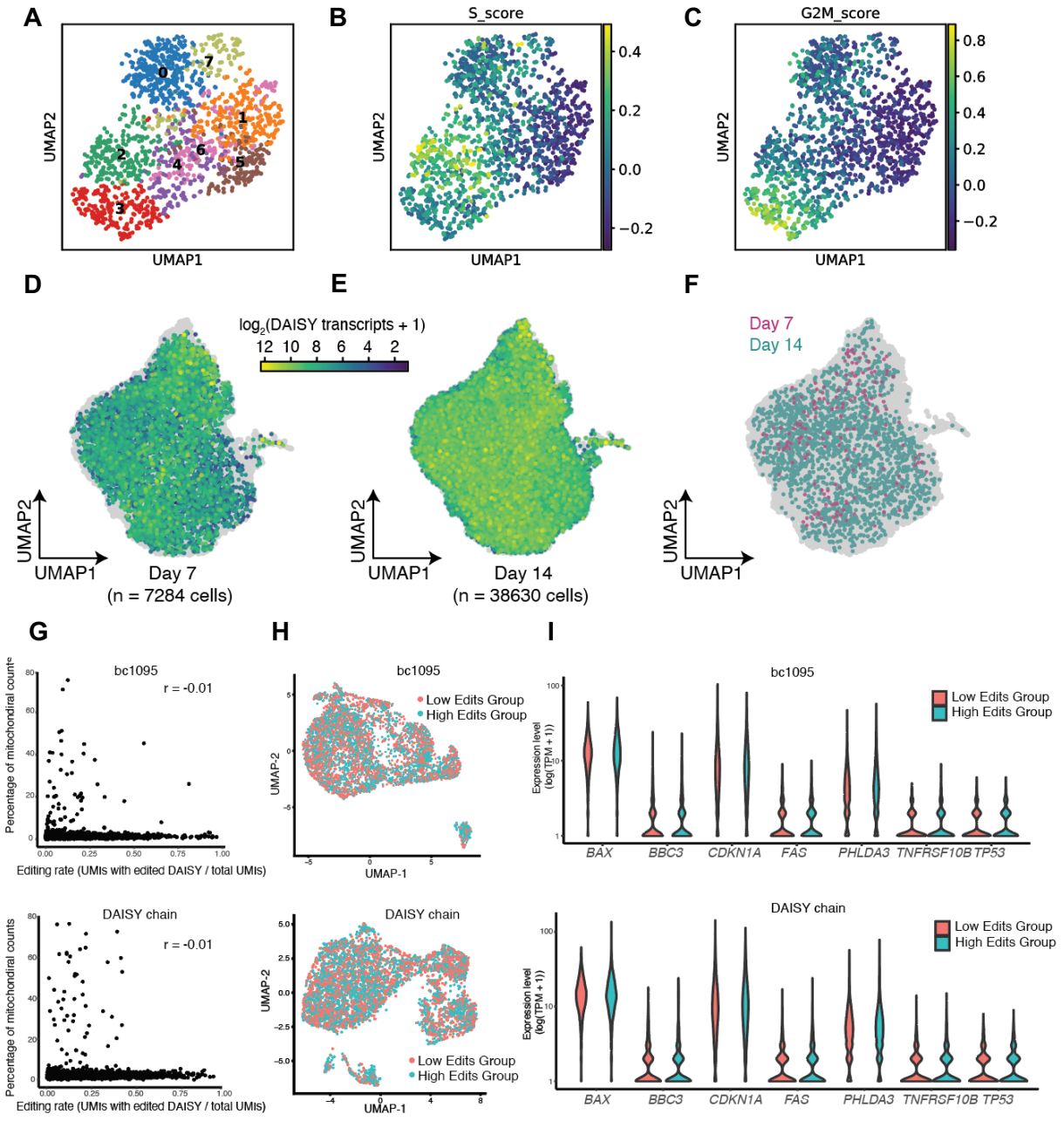


Figure S5. Analysis of transcriptional clustering and genotoxicity at single-cell resolution, Related to Figures 6 and 7 (A) UMAP embedding of gene expression states of barcoded cells from scDAISY-seq-v1. (B) Scoring of each cell by expression level of genes associated with S-phase shows enrichment of high S-scores in cluster 2. (C) Scoring by expression level of genes associated with G2M-phase shows enrichment of high G2M-scores in cluster (D) Cells collected at Day 7 colored by the expression level of the DAISY barcode and (E) at Day 14. (F) UMAP embedding shows uniform clustering of cells at Day 7 and at Day 14. (G) Scatter plot showing no correlation between fraction of mitochondrial reads within a cell (measurement of cell stress) and the editing rate within the cell (each dot is a cell). Data with bc1095 is shown in the top panel and with the DAISY chain is shown in the bottom panel. (H) UMAP plots showing uniform clustering of cells containing bc1095 (top) and the DAISY chain

(bottom) within the high (editing rate greater than or equal to 50%) and low edit groups (editing rate less than 50%). (I) Violin plots showing the expression of selected DNA damage/repair-related genes between the high and low edit groups within bc1095- (top) and DAISY chain (bottom)-expressing cells.

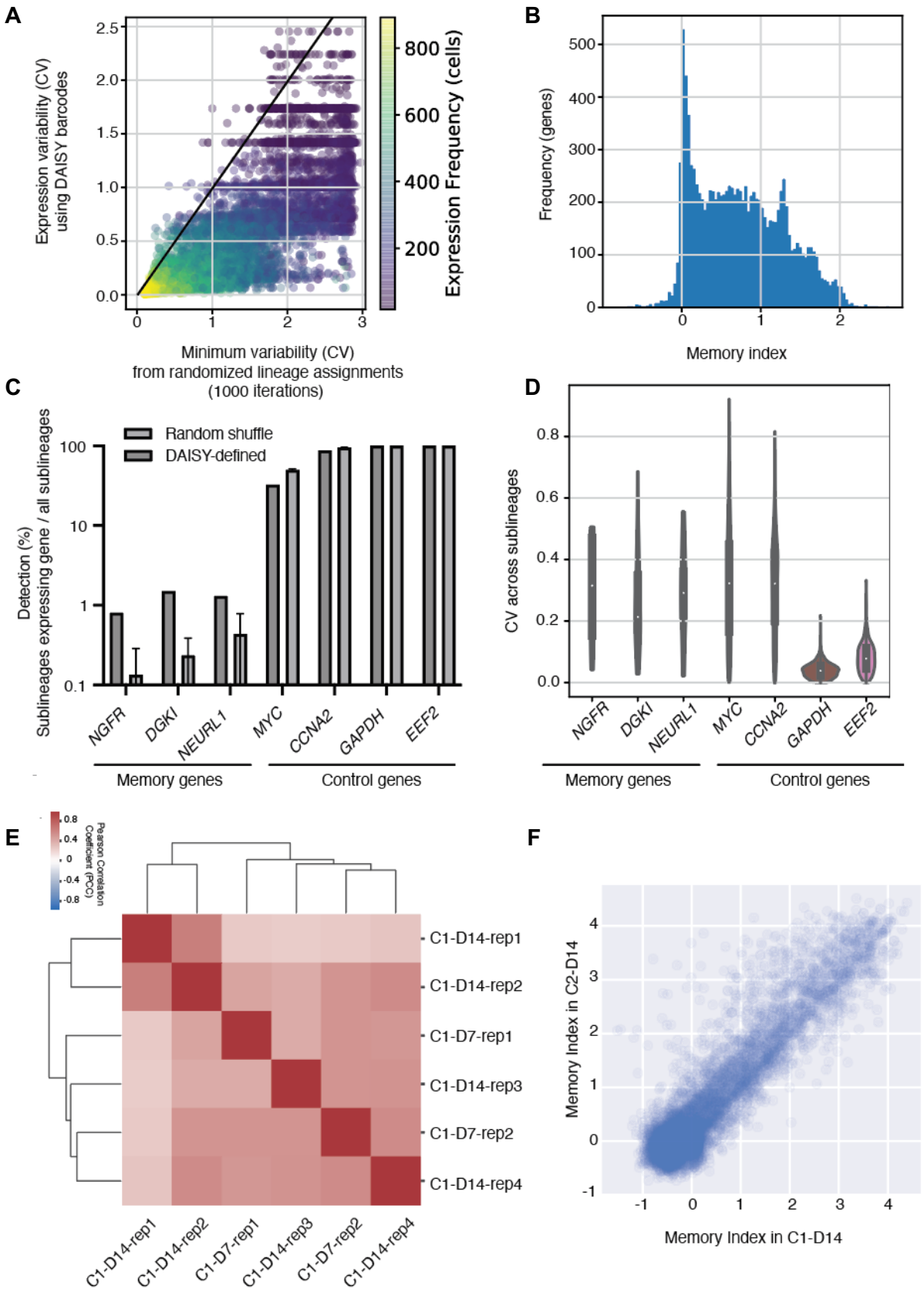

A
Expression variability (CV)
using DAISY barcodes
Minimum variability (CV)
from randomized lineage assignments
(1000 iterations)
Expression Frequency (cells)
B
Frequency (genes)
Memory index
C
Random shuffle
DAISY-defined
Detection (%)
Sublineages expressing gene / all sublineages
NGFR
DGKI
NEURL1
MYC
CCNA2
GAPDH
EEF2
Memory genes
Control genes
D
CV across sublineages
Memory genes
Control genes
E
Pearson Correlation Coefficient (PCC)
C1-D14-rep1
C1-D14-rep2
C1-D7-rep1
C1-D14-rep3
C1-D7-rep2
C1-D14-rep4
F
Memory Index in C2-D14
Memory Index in C1-D14

Figure S6. Analysis of the transcriptional memory effect within the scDAISY-seq-v2 time course experiment, Related to Figure 7 (A) The relationship between the expression variability within a DAISY-defined sublineage across all genes (dots) versus cells with randomly permuted cell labeling across 1000 iterations. The average minimum CV across all iterations is plotted. (B) The distribution of the memory index across all genes within a representative clone collected at Day 14 (n = 1712 cells). (C) Memory genes that are rarely expressed have a higher probability of being detected when DAISY barcode information is used for cell grouping as opposed to random sampling (across 1000 samples with mean sample size equal to the mean sublineage population size within the phylogenetic tree). The control genes show consistent detection rates across sampling procedures. (D) The distribution of CV expression values (minimum 3 cells expressing) across all sublineages. Housekeeping control genes (GAPDH and EEF2) show uniformly low CV across all sublineages consistent with their frequent and high expression across all cells. Memory genes show a wider distribution of CVs across sublineages. (E) Correlation matrix showing the correlation of memory indices within a clone across biological replicates and timepoints. (F) Scatter plot depicting the correlation in memory effect across two different clones collected at Day 14.

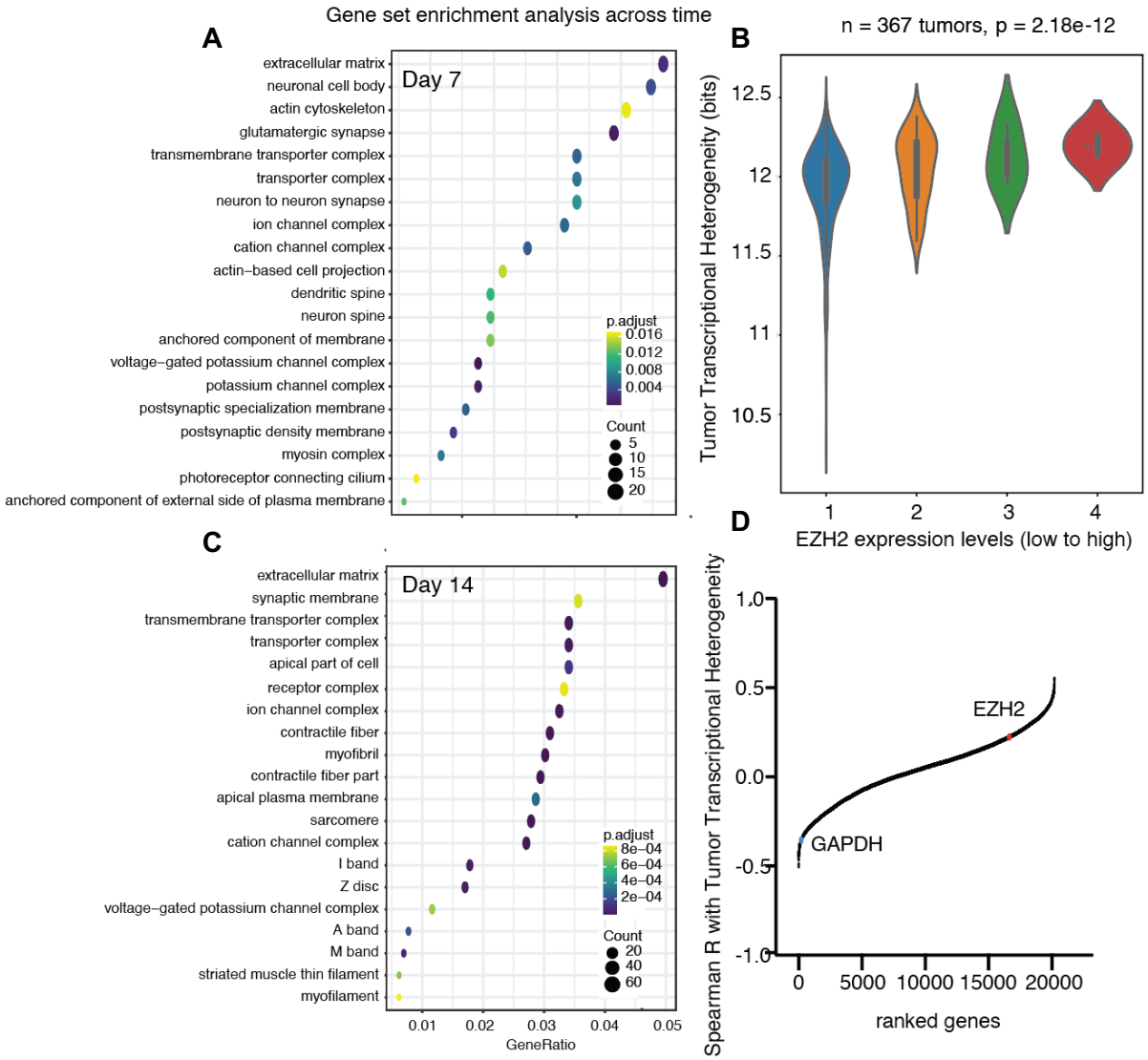


Figure S7. Gene set enrichment analysis of high memory genes across time and the relationship between *EZH2* expression and tumor transcriptional heterogeneity, Related to Figure 7 (A) Gene ontology enrichment analysis reveals that memory genes fall within pathways that may comprise a dedifferentiated cell state stably at Day 7 (B) and Day 14. (C) Relationship between the total *EZH2* expression level within tumors and the heterogeneity of gene expression across all genes (transcriptional heterogeneity) quantified as the Shannon diversity index. Tumors are binned into four groups based upon the maximum observed *EZH2* expression level. Cells expressing less than a quarter, half, three-quarters, and unity of the expression maximum are in the first, second, third, and fourth bins, respectively. (D) Distribution of spearman rank correlation values of the expression level of a gene and the transcriptional heterogeneity within a tumor across all tumors. *EZH2* is plotted with *GAPDH* as a reference.

Methods S1. Detailed Protocols related to STAR Methods

Protocol 1: DAISY Cas12a high-throughput screening

A. *In silico* design of library

1. Generate 5000 random Cas12a target sequences and filter for potential off-target effects within the human genome.
2. Pairwise assemble each randomly generated sequence together to generate all possible combinations.
3. Score the predicted editing efficiency using the Seq-DeepCpf1 tool (http://deepcrispr.info/).
4. Calculate the pairwise editing efficiency variability (CV) and keep the top 55th percentile sequences (14358 sequences) for assembly. These designs have phased efficiency and, in theory, reduce the probability of collapsing.
5. Assign each sequence a unique identifier sequence (ampID) to enable demultiplexing
6. Combine the target sites with crRNA sequences into one oligo with the below framework:

   Adapter-crRNA1-DR-crRNA2-ampID-T1-T2[rc]-Adapter
7. The script used to generate the library is provided on the Cong Lab Github (https://github.com/cong-lab/daisy_analysis)

B. Molecular construction of library

1. Resuspend synthesized oligos (Twist Biosciences) to 10 ng / uL.
2. Amplify the library using a low-cycle PCR reaction via setup below:

Oligo pool	1 uL
For Primer (10 uM)	1.25 uL
Rev Primer (10 uM)	1.25 uL
Kapa HiFi (2X)	12.5 uL
ddH20	9 uL

3. Perform PCR in a thermocycler using the following parameters:
   95^{o}C, 3 min, initial denaturation
   [98^{o}C, 20 sec
   60^{o}C, 15 sec
   72^{o}C, 15 sec]
   Repeat for a total of 12 cycles

72oC, 1 min

4. Column purify the resulting PCR product (NEB) and elute in 10 uL elution buffer
5. Assemble the amplified library into the recipient lentiviral backbone using NEBuilder HiFi DNA Assembly (NEB) using the following parameters:

Reagent	**1X**	**Notes**
pNPR028-digested (50 ng / uL)	1 uL	~ 22 fmol
Oligo PCR product	2 uL	~ 400 fmol
NEB HiFi Mix (2x)	10uL	
ddH20	7 uL	Total 20

6. Incubate the reaction at 50^{o}C for 1 hour.
7. Precipitate the assembly reaction using the following setup:

Component	**Amount per RXN (uL)**	**Final Concentration**
HiFi Assembly RXN	20	
Isopropanol	20	
GlycoBlue CoPrecipitant	0.20	0.075 ug / uL
NaCl solution (5 M)	0.40	50 mM
Total	40.6	

8. Transform bacteria (Lucigen Endura Electrocompetent Cells) using manufacturer's protocol and plate to achieve adequate coverage of library (>100 CFUs / design)
9. Scrape transformed colonies into LB and extract plasmid using a maxi-prep scale (Qiagen Plasmid Plus)

C. Transfection of HEK293FTs to make lentivirus with plasmid pool

1. One day prior to transfection, plate HEK293FT cells in DMEM (Gibco) supplemented with 10% FBS such that they achieve ~90% confluence upon transfection in a 10cm dish
2. Transfect cells using the following parameters, following the manufacturer's protocol:

jetOPTIMUS Buffer (Polyplus)	1000 uL
pCMV delta R8.2	2.5 ug

pCMV-VSV-G	950 ng
Plasmid pool	3.7 ug
jetOPTIMUS	8 uL

3. Harvest lentivirus 48 hours post-transfection and proceed to transduction

D. Transduction and tissue culture procedure (**Supplementary Protocol Figure 2**)

1. Transduce 1.5e7 A375 cells, cultured in DMEM (Gibco) supplemented with 10% FBS, (1000 cell / design coverage) with pooled lentivirus by centrifuging cell and virus mixture at 1200*g for 45 minutes at 37^{o}C
2. Select transduced cells with Blasticidin (5 ug / mL) for 48-72 hours
3. Collect 3-5e6 cells by centrifuging cells at 300*g for 5 minutes as an initial cell pellet prior to Doxycycline induction of Cas12a
4. Plate back ~1.5e7 cells and supplement media with 400 ng / mL Doxycycline
5. Every 48 hours, for 14 days, collect 3-5e6 cells and plate back remaining cells

E. Pooled PCR-based readout of editing and sequencing

1. Extract genomic DNA (gDNA) from pelleted cells using the DNeasy Blood & Tissue Kit (Qiagen) using the manufacturer's protocol
2. Perform PCR from 15 ug or ~1.5e6 genome equivalents using the following individual reaction conditions:

gDNA	1.25 ug
For Primer (10 uM)	2.5 uL
Rev Primer (10 uM)	2.5 uL
Phusion Flash MM (2X)	25 uL
ddH20	Fill to 50 uL

3. Incubate the PCR reactions using the following conditions:
   98^{o}C, 10 sec, initial denaturation
   [98^{o}C, 1 sec
   57^{o}C, 5 sec
   72^{o}C, 5 sec]
   Repeat for a total of 35 cycles

72oC, 1 min

4. Pool the first-round reactions together and column-purify (Qiagen) the pooled reactions and load into a final PCR to add Illumina sequencing adapters and sample indexes using the following parameters:

PCR1	2 uL
For Primer (10 uM)	2.5 uL
Rev Primer (10 uM)	2.5 uL
Phusion Flash MM (2X)	25 uL
ddH20	18 uL

5. Incubate the PCR reactions using the following conditions:
   98°C, 10 sec, initial denaturation
   [98°C, 1 sec
   72°C, 10 sec]
   Repeat for a total of 15 cycles
   72oC, 1 min

6. Column-purify the index PCR reactions and quantify using Qubit (Thermo Fisher Scientific)

7. Load Illumina sequencer aiming to achieve ~500X read coverage / design

F. Bioinformatic processing of data

A brief description is detailed below. The entire processing pipeline is provided on the Cong Lab
Github (https://github.com/cong-lab/daisy_analysis)

1. Merge mate pair reads using FLASh (https://ccb.jhu.edu/software/FLASH/)
2. Demultiplex each read according to its ampID
3. Align read to its reference sequence using NEEDLEALL
4. Filter out alignments that are present within the plasmid pool

Protocol 2: Integration of DAISY barcoding with single-cell RNA-seq (scDAISY-seq)

A. Preparation of lentiviral plasmid pool for single-cell recovery

1. Please refer to **Supplementary Protocol Figure 5** for a detailed description of the DAISY barcode sequence cassettes used for the single-cell RNA-seq experiments. In addition, we provide a schematic drawing of the lentiviral vector in **Supplementary Protocol Figure 3.**
2. These vectors were prepared in a similar fashion to the DAISY library in order to prepare a diverse pool of static tag sequences that serve as clonal markers. We suggest using Lucigen Endura Electrocompetent bacteria or Stbl4 ElectroMax (Invitrogen) cells to avoid recombination between repetitive sequences (like direct repeats)

3. These vectors will be deposited to Addgene as described in Table 1 listed at the end of this protocol.

B. Transfection of plasmid pool to generate DAISY-barcode lentivirus

1. Approximately 16 hours prior to transfection, plate 7.5e5 cells / well in a 6-well plate in DMEM (Gibco) supplemented with 10% FBS.
2. Transfect HEK293FT cells with the following components:

jetOPTIMUS Buffer (Polyplus)	500 uL
pCMV delta R8.2	900 ng
pCMV-VSV-G	100 ng
Plasmid pool	1 ug
jetOPTIMUS	2 uL

3. 48 hours after transfection, collect the supernatant and proceed directly to transduction or flash freeze lentivirus on dry ice and store at -80^{o}C.

C. Lentiviral delivery of DAISY barcode to mammalian cells via spin-fection

1. Mix 5e4 A375 cells with 400 uL of lentivirus and centrifuge at 1200*g for 45 minutes at 37^{o}C in DMEM supplemented with 10% FBS and 8 ug / mL Polybrene (Sigma-Aldrich)
2. Select transduced cells with Blasticidin at 5 ug / mL for 48-72 hours

D. Tissue culture procedure (using A375 cells as example)

1. Perform limited dilution of transduced cells to achieve 5-10 cells / well within a flat-bottomed 96-well plate (Sigma-Aldrich) in DMEM (Gibco) + 10% FBS + 400 ng / mL Doxycycline
2. Monitor cell outgrowth over the course of the next two weeks. Sequentially expand cells into a 48-well plate and finally a 24-well plate where the total population size will be ~2.4e5 cells upon confluence (~15 cell doublings)
3. To reduce Cas12a expression and avoid cleavage of the DAISY barcode transcript, remove Doxycycline from the media for the final 72 hours of culture.

E. Harvesting for single-cell RNA-sequencing library construction

1. Cool table-top centrifuge to 4^{o}C

2. Trypsinize cells from 24-well plate and collect into a 1.5 mL microcentrifuge tube using CMF-PBS supplemented with 2% FBS (Resusupension Buffer).

3. Pellet cells through centrifugation at 300*g for 5 minutes.

4. Resuspend cells in 200 uL of Resuspension Buffer and count cells using hemocytometer

5. Adjust cell concentration to achieve 1e3 cells / uL

6. Prepare scRNA-seq libraries according to the Chromium Next GEM Single Cell 3' v3.1 Reagent Kit (10X Genomics)

F. Targeted enrichment of barcode sequence

1. Please note that for the initial experiment, we used Capture Sequence 2 (10X Genomics) designed to hybridize to the gel beads that capture (1) the transcriptome of the cell via an oligodT and (2) the DAISY barcode of the cell via the Capture Sequence (**Supplementary Note 2**). For the timecourse DAISY chain experiment, we captured both the DAISY barcode and the transcriptome via the oligodT.

2. The targeted enrichment reaction used primers that amplified specifically off of the Nextera Read 1 (Illumina) sequence that is bound to the 10X gel bead and a DAISY barcode specific primer that included a Nextera Read 2 (Illumina) partial adapter sequence. For the polyA capture approach, we used a primer that bound specifically to Nextera Read 2 (Illumina) as it was included within the lentiviral construct.

3. Step 4.1m from the 10X protocol was used as a template for the enrichment PCR with the following setup:

Reagent	**1X**
NEB Ultra II Q5 MM	50 uL
For Primer (10 uM)	5 uL
Rev Primer (10 uM)	5 uL
cDNA (4.1m)	5 uL
ddH20	35 uL

4. The reaction was conducted with the following thermocycler parameters:

Initial Denat	98oC	30 sec
Denat	98oC	5 sec, 20 cyc
Anneal	62oC	10 sec, 20 cyc
Extension	72oC	20 sec, 20 cyc
Final Extension	72oC	2 min
Hold	4oC	

5. Column purify the PCR reaction (Qiagen) and elute in 10 uL
6. Perform a second round PCR in order to add sample indexes and Illumina adapters according to:

Reagent	**1X**
NEB Ultra II Q5 MM	25 uL
Index-Primer-For (10 uM)	2.5 uL
Index-Primer-Rev (10 uM)	2.5 uL

PCR1 (10-20 ng)	2.5 uL
ddH20	17.5 uL

7. Using the following parameters:

Initial Denat	98oC	30 sec
Denat	98oC	5 sec, 5 cyc
Anneal	65oC	10 sec, 5 cyc
Extension	72oC	20 sec, 5 cyc
Final Extension	72oC	2 min
Hold	4oC	

8. Column-purify the index PCR reactions and quantify using Qubit (Thermo Fisher Scientific)

9. Sequence the resulting libraries on a MiSeq (Illumina) using a 2x150 paired-end configuration according to sampling depth as specified in the manufacturer's protocol

G. Bioinformatic processing of sc-DAISY-seq data (**Supplementary Protocol Figure 4**)

1. Run CellRanger (10X Genomics) on target enrichment libraries. Using the Read1 (R1) trim parameter, specify Read1 to be trimmed to 28 bp (16 bp cell barcode + 12 bp UMI sequence)

2. Parse the DAISY barcode sequence from Read 2 (R2) and align to the reference barcode sequence to generate allelic information using the Needleman-Wunsch algorithm

3. Filter the alleles such that only indels that occur within a user-defined window (default: +/- 5 bp from cutsite) are included

4. Assign alleles to a lineage group based on the static tag sequence and organize into a character matrix with cells as indexes and alleles within each target site as characters

5. Use the Neighbor Joining algorithm to phylogenetically reconstruct the subclones within each lineage group

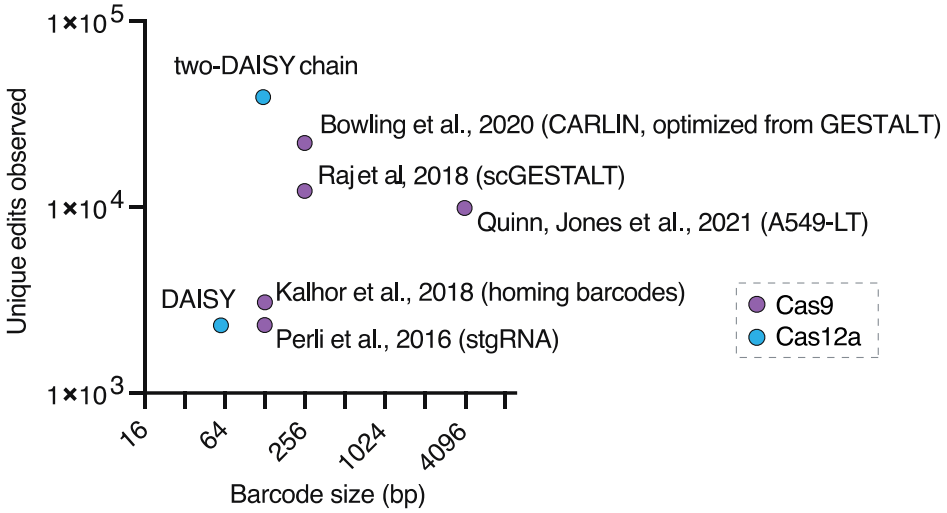
6. Scripts for the processing and for the exploratory transcriptional memory analysis are provided on the Cong Lab github (https://github.com/cong-lab/daisy_analysis/tree/main/fig6_transcriptional_memory)

Table 1. List of Plasmids (all will be available at Addgene via plasmid ID)

Plasmid ID	**Detailed Description**
DAISY Libraries	
pN026-15k	pLenti-U6-cr1-cr2-ampID-T1-T2-polyT-EF1A-BlastR-WPRE (DAISYs are from initial screen)
pN026-2k	pLenti-U6-cr1-cr2-ampID-T1-T2-polyT-EF1A-BlastR-WPRE (DAISYs are generated by CLOVER predictions)
scDAISY-seq Plasmids	
pN103-d1095	pLenti-U6-d1095-capseq2rc(10X Genomics)-polyT-EF1A-BlastR-WPRE
pN103-d859	pLenti-U6-d859-capseq2rc(10X Genomics)-polyT-EF1A-BlastR-WPRE
pN256-DAISY-chain	pLenti-5'LTR-U6-crDAISY-chain(antisense)-DAISY-chain-STATICTAG(10bp)-WPRE-GFP-TRE3GV-3'LTR

Technology Benchmarking: To contextualize DAISY barcoding technology, we performed a meta-analysis and compared the salient features (like barcode length, target sites, allelic diversity, and single-cell applicability) to existing technologies (**Supplement Protocol Figure 1**).

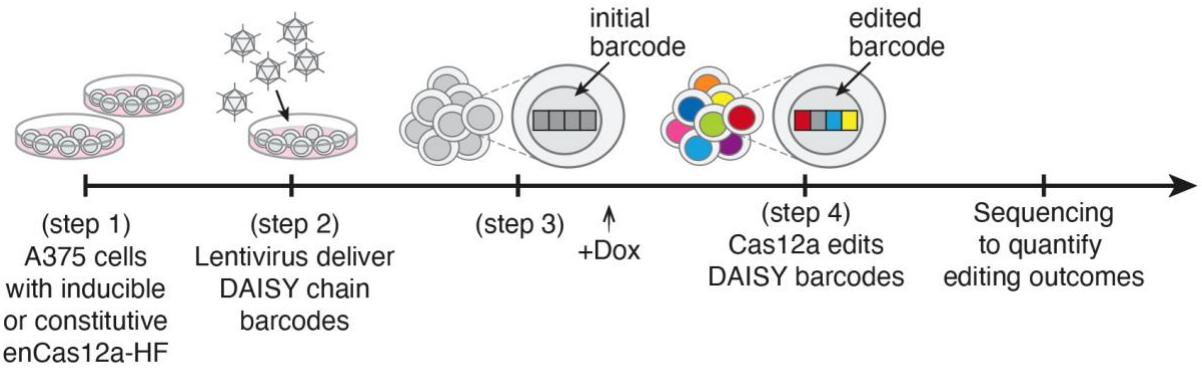
A



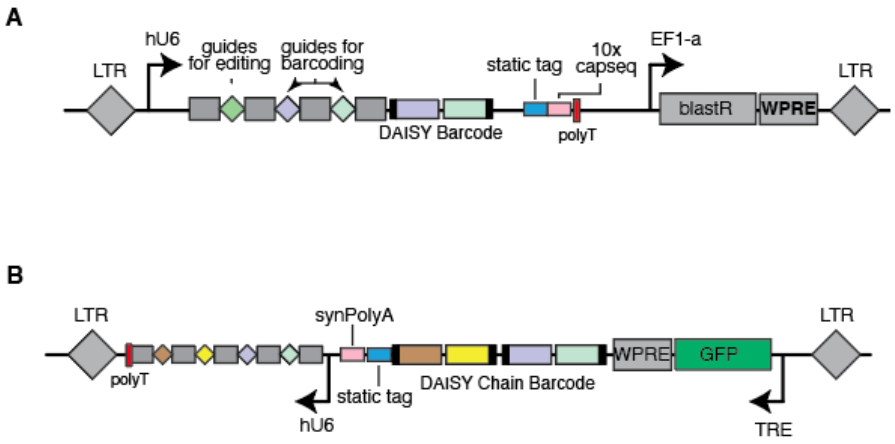
B

Parameter	Cas9 homing barcode	Cas9 array barcode	Cas12 DAISY barcode	
barcode total length	~100bp to 120bp	~270bp to 4000bp	~60bp (single DAISY)	~120bp (two-DAISY chain)
no. of target sites within full barcode	1 target site	10 target sites	2 target sites	4 target sites
unique edits observed	~3000 to 4000 (Perli, et al.; Kalhor, et al.)	~32,000 to 33,000 (Bowling et al.; Quinn, Jones et al.)	~ 3000	~ 66,000
bulk barcode entropy	5 ~ 6 bits (Kalhor, et al.)	~ 7 bits (Bowling et al.)	~ 9 bits	~ 12 bits
single-cell demonstration	No	Yes	Yes	

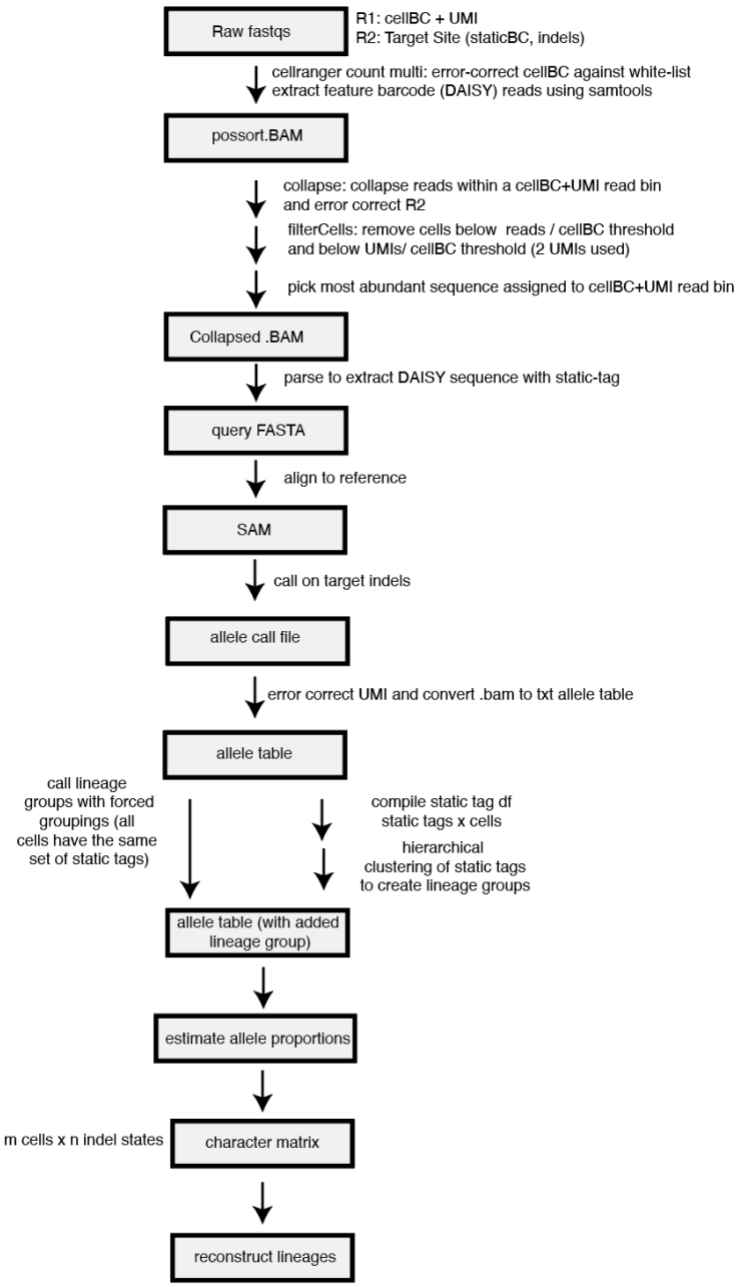
Supplementary Protocol Figure 1. Benchmarking of DAISY chain barcoding relative to existing Cas9-based approaches. (A) Summary of unique barcode edits observed and the size of barcodes across Cas9 and Cas12 barcoding systems. (B) Comparison of key features of barcodes.



Supplementary Protocol Figure 2. Experimental design using a DAISY-chain. Cells harboring constitutive enCas12a-HF or inducible enCas12a-HF were transduced at low MOI and harvested at multiple timepoints after infection to measure barcode editing and entropy.



Supplementary Protocol Figure 3. scDAISY-seq vector designs. (A) Vector design for the initial single-cell barcoding (scDAISY-seq-v1) experiment with two-target DAISY barcodes where the DAISY barcode includes an additional guide for potential genomic editing along with a static randomized tag that is captured with the 10X capture sequence. (B) Vector design for the time course DAISY-chain scDAISY-seq experiment (scDAISY-seq-v2) where the target sites are inserted into the UTR of eGFP to enable oligo-dT-based capturing on the 10X gel bead.



Supplementary Protocol Figure 4. Workflow of scDAISY-seq lineage reconstruction pipeline adapted from Cassiopeia (Jones *et al*., 2020).

Design 859

atttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagta
tttcgatttcttggctttatatatcttgtggaaaggacgaaacaccGTAATTTCTACTCTTGTAGATGTAACGCG
CTTGGGTCAAGCATCAAATTTCTACTCTAGTAGATCTACCTATTACTAGGACAAGTTAATTTCTACTGTCGTAGAT
CGGCACATTCCGAGACATTCTAAATTTCTCCTCTCGGAGATTTGCTACCTATTACTAGGACAAGTGGAGACGTTGA
GAATGTCTCGGAATGTGCCGGAAATCATAGGACACGTACACNNNNNNNNNNGCTCACCTATTAGCGGCTAAGG

hU6 promoter
Direct repeat
Spacer
Targets
Filler
Static barcode
10X Capture Sequence 2 (Reverse Complement)

Design 1095

atttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagta
tttcgatttcttggctttatatatcttgtggaaaggacgaaacaccGTAATTTCTACTCTTGTAGATGTAACGCG
CTTGGGTCAAGCATCAAATTTCTACTCTAGTAGATCATCGTATTACTAGGACAATTTAATTTCTACTGTCGTAGAT
CGGCACATCGAGAGACTTCCTAAATTTCTCCTCTCGGAGATTTGCATCGTATTACTAGGACAATTCGAGTCCTTGA
GGAAGTCTCTCGATGTGCCGGAAATCATAGGACACGTACACNNNNNNNNNNGCTCACCTATTAGCGGCTAAGG

hU6 promoter
Direct repeat
Spacer
Targets
Filler
Static barcode
10X Capture Sequence 2 (Reverse Complement)

Supplementary Protocol Figure 5. Sequences of single-cell DAISY cassettes with two highest capacity barcodes (after CLOVER optimization), showing the detailed sequence and design of the barcode cassette.