nature biotechnology

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41587-019-0331-8

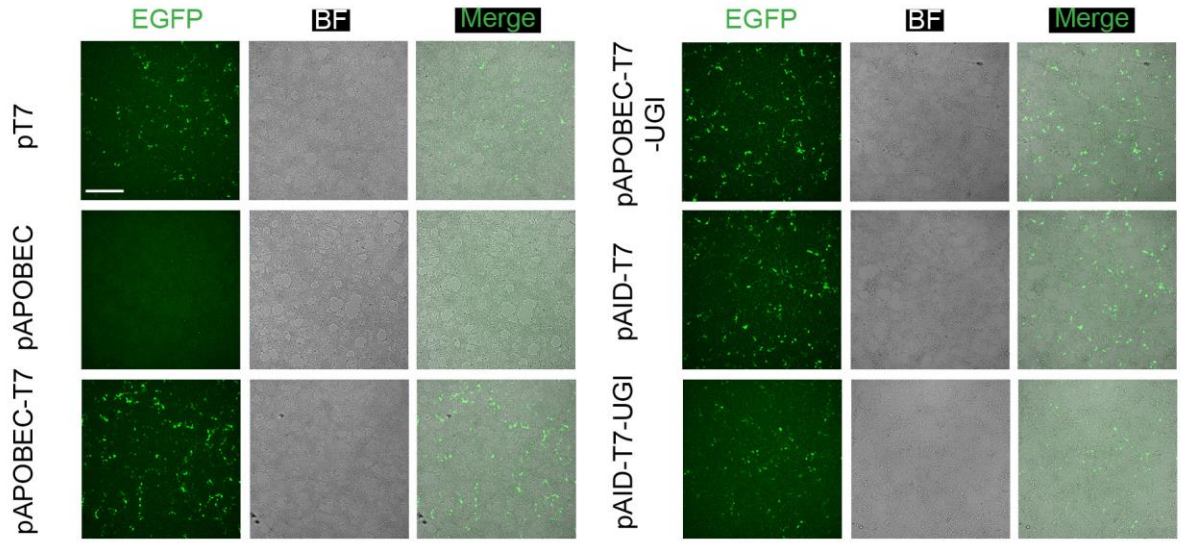
In the format provided by the authors and unedited.

Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor

Haiqi Chen[1], Sophia Liu[1,2,3,11], Samuel Padula[1,11], Daniel Lesman[1], Kettner Griswold[4,5,6,7], Allen Lin[8,9], Tongtong Zhao[1,10], Jamie L. Marshall[1] and Fei Chen[1*]

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Biophysics Program, Harvard University, Boston, MA, USA. [3]Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. [6]Department of Genetics, Harvard Medical School, Boston, MA, USA. [7]Charles Stark Draper Laboratory, Cambridge, MA, USA. [8]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [9]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. [10]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [11]These authors contributed equally: Sophia Liu, Samuel Padula. *e-mail: chenf@broadinstitute.org

Supplementary Figure 1

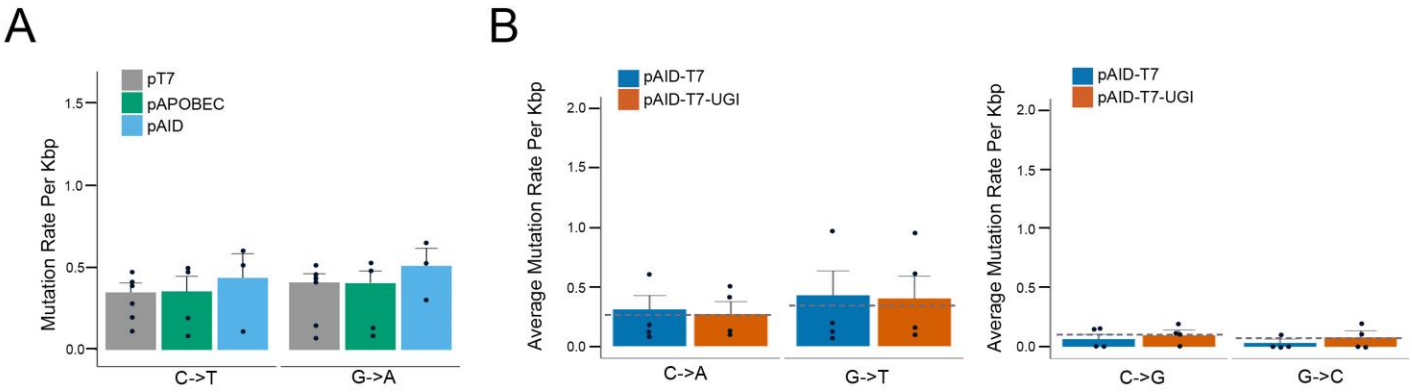


Supplementary Figure 1

Deaminase-T7 RNA polymerase (T7 RNAP) fusion maintains the transcriptional activity of T7 RNAP.

Fusing a cytidine deaminase to T7 RNAP does not significantly hinder the transcriptional activity of the T7 RNAP. Each pEditor variant was introduced into HEK293T cells together with pTarget in which the EGFP gene was solely under the control of a T7 promoter. EGFP signals were observed in cells transfected with pT7, pAPOBEC-T7, pAPOBEC-T7-UGI, pAID-T7, and pAID-T7-UGI, but not in cells transfected with pAPOEBC. The experiments were repeated 3 times with similar results. Scale bar, 200 μm, applies to all micrographs.

Supplementary Figure 2

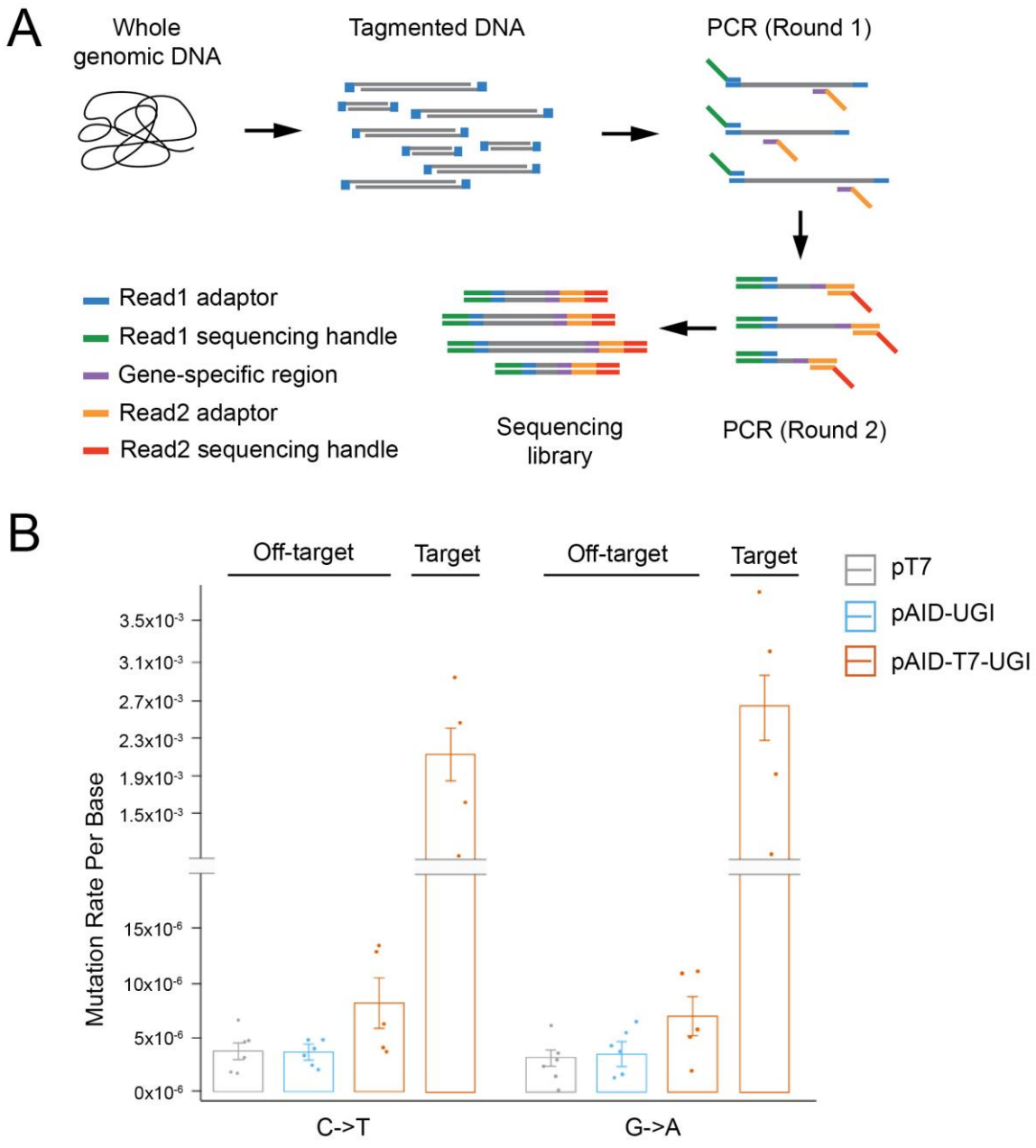


Supplementary Figure 2

TRACE demonstrates high editing specificity in human cells.

A. Overexpression of cytidine deaminases alone (pAPOBEC or pAID) in cells result in mutation rates that are not statistically different to mutation rates in the pT7 group (pT7 *vs.* pAPOBEC, two-sided *t* test, p=0.9661 in C->T, p=0.8887 in G->A; pT7 *vs.* pAID, two-sided *t* test, p=0.2511 in C->T, p=0.1903 in G->A). Bars represent mean ± SEM (N=6 biologically independent cell samples for pT7, 4 for pAPOBEC, and 3 for pAID). **B.** Mean C->A and G->T (left), C->G and G->C (right) mutation rates in pAID-T7 and pAID-T7-UGI group. Dashed line represents the mean sequencing error rate for the indicated base substitution. Bars are mean ± SEM of N=4 biologically independent cell samples for pAID-T7and pAID-T7-UGI).

Supplementary Figure 3

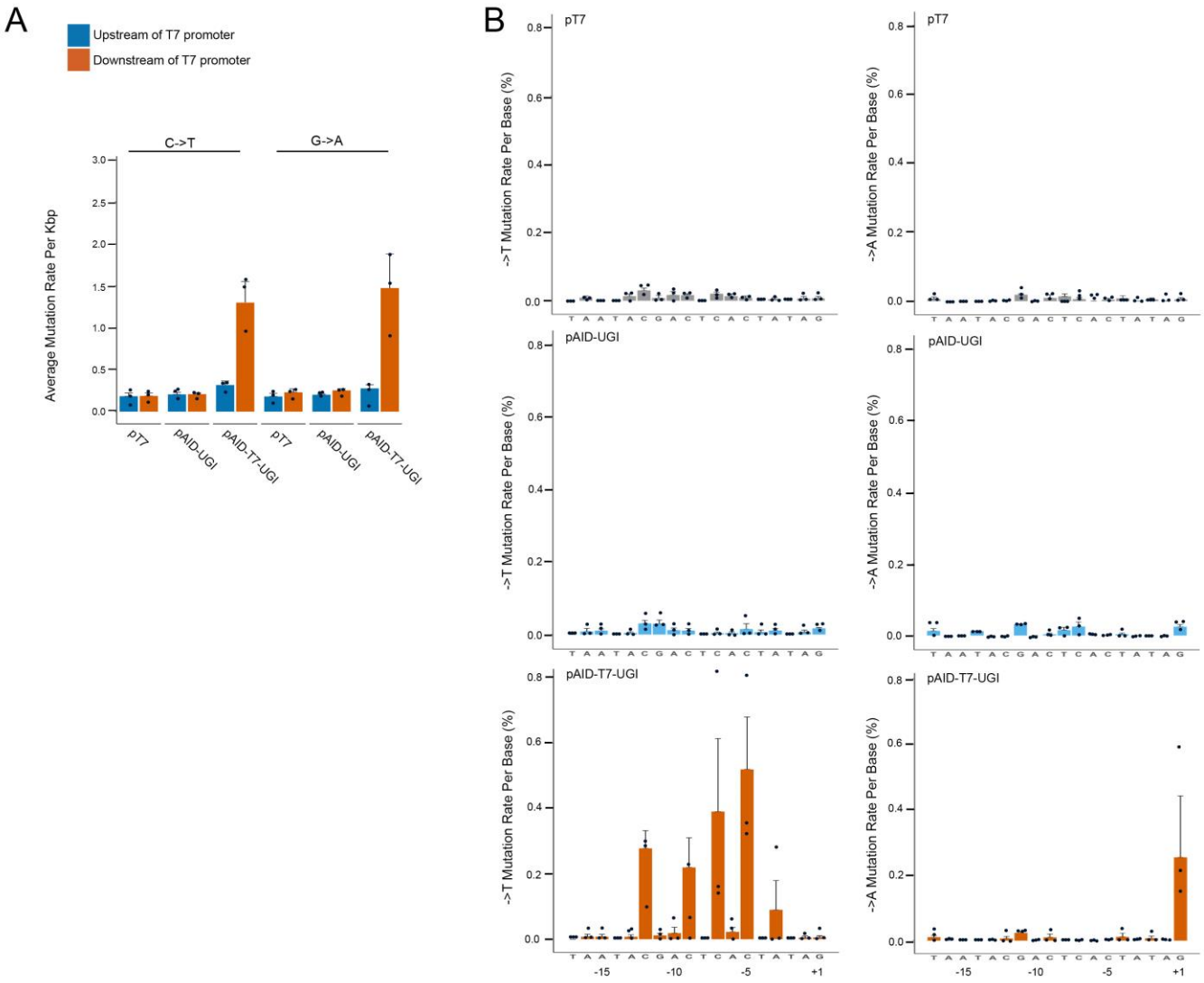


Supplementary Figure 3

TRACE induces negligible off-target effects.

A. Experimental workflow on the generation of a sequencing library for off-target analysis. **B.** Mutation rate per base in target vs. off-target regions of the genome. All bars in the figure are mean ± SEM of N = 6 biologically independent samples for off-target analysis and N=4 for on-target analysis.

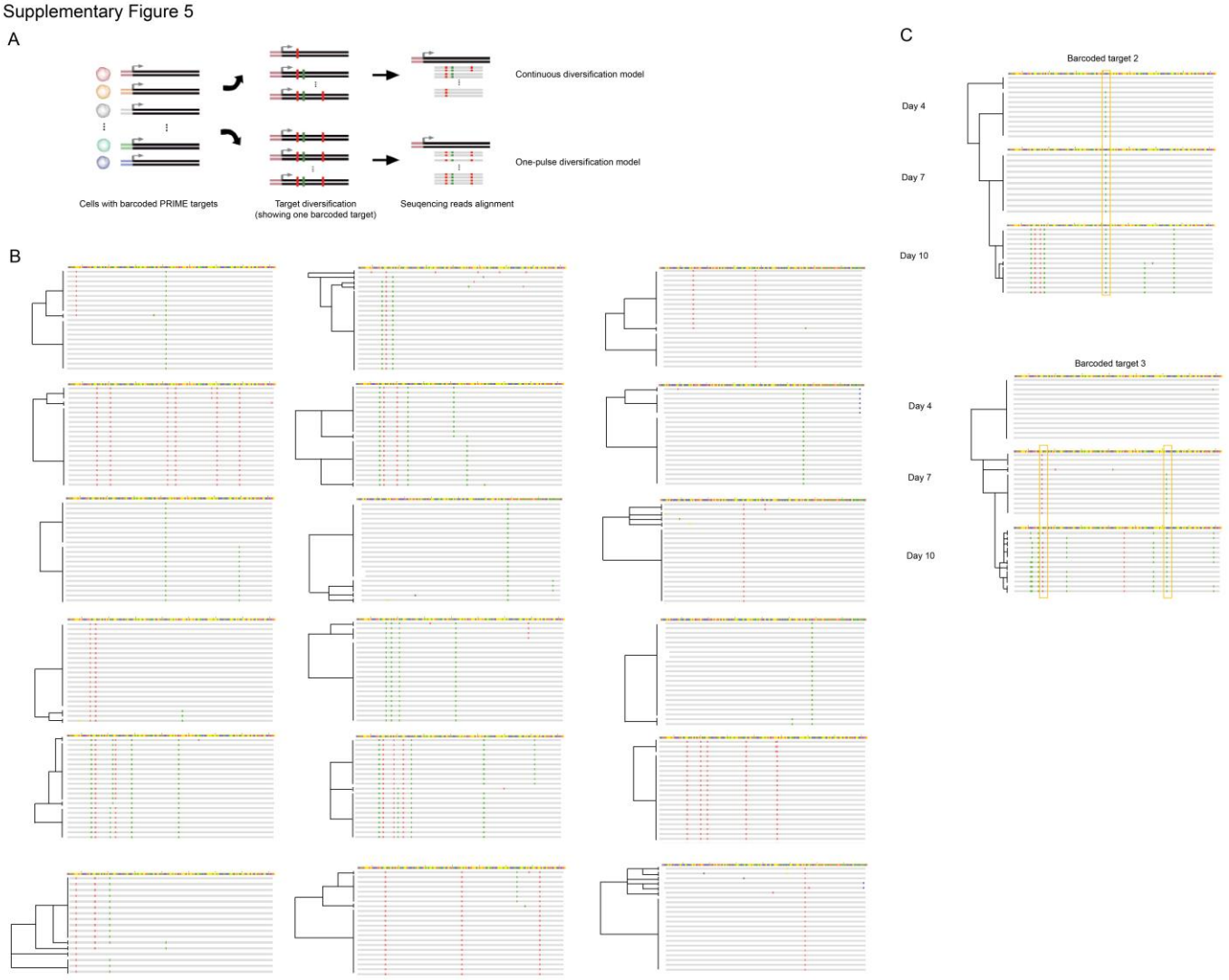
Supplementary Figure 4



Supplementary Figure 4

Mutation profiles upstream, downstream of and at the T7 promoter sequence following TRACE diversification.

A. Mean C->T and G->A mutation rates for the upstream region of the T7 promoter vs. the downstream region. Upstream vs. downstream, two-sided *t* test, p=0.0363 in C->T, p=0.0205 in G->A for pAID-T7-UGI; p=0.5403 in C->T and p=0.4251 in G->A for pT7; p=0.8983 in C->T, p=0.0861 in G->A for pAID-UGI. **B**. Mean ->T and ->A mutation rates at each base of the T7 promoter. All bars in the figure are mean ± SEM of N = 3 independent experiments.

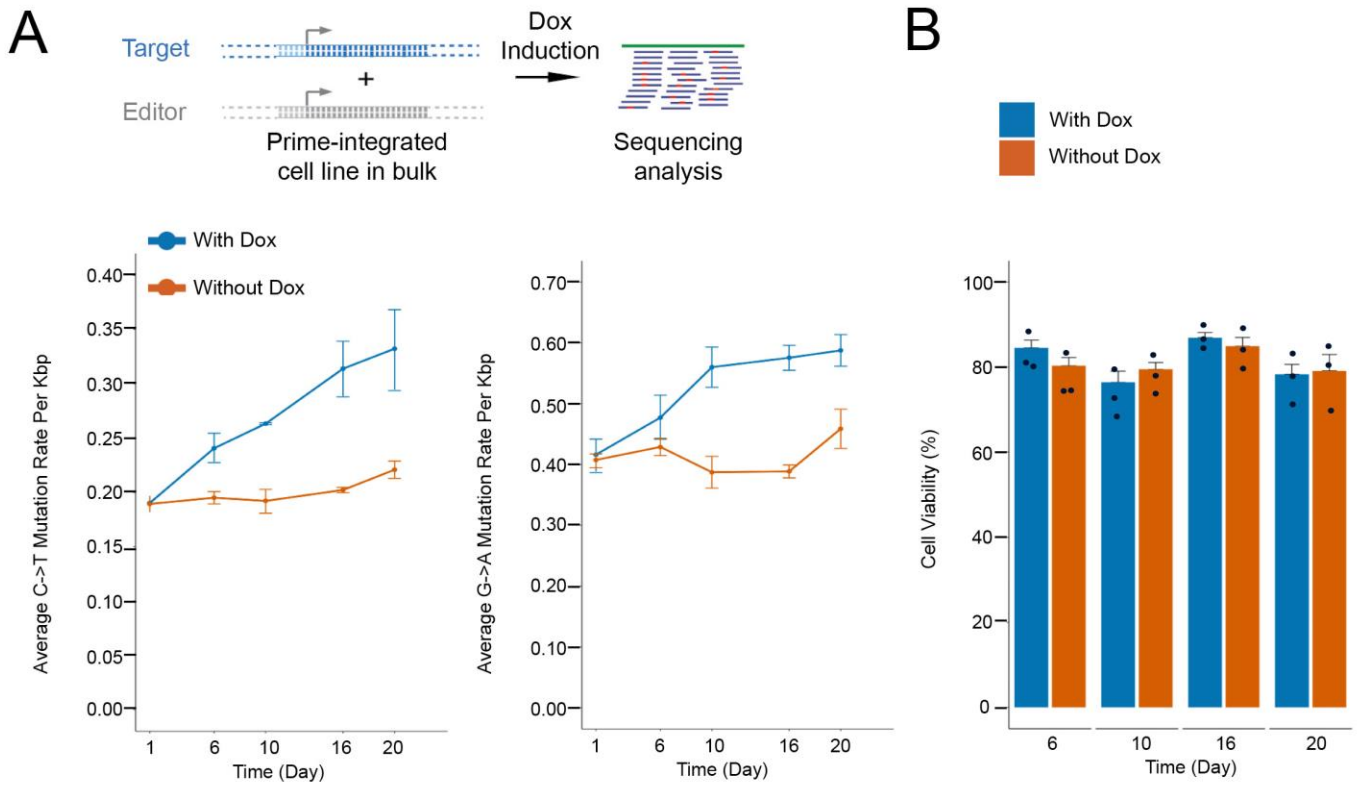


Supplementary Figure 5

TRACE induces continuous diversification on barcoded targets.

A. Schematic of two possible models of target diversification by TRACE (continuous diversification *vs*. one-pulse diversification) which could explain the results shown in **Figure 2B**. **B**. Representative sequencing reads alignments of 18 different barcoded targets. Dendrograms showing the hierarchical relationship of the sequencing reads were constructed using the mutation profile of each read and were displayed on the left of each alignment. Red: A; Green: T; Blue: C; Yellow: G. **C**. In addition to **Figure 2C**, representative sequencing reads alignment of additional 2 unique barcoded targets detected over 3 time points were shown. Dendrograms showing the hierarchical relationship of the sequencing reads across time points were constructed using the mutation profile of each read and were displayed on the left of each alignment. Shared mutations across time points were highlighted in yellow boxes. Red: A; Green: T; Blue: C; Yellow: G.

Supplementary Figure 6

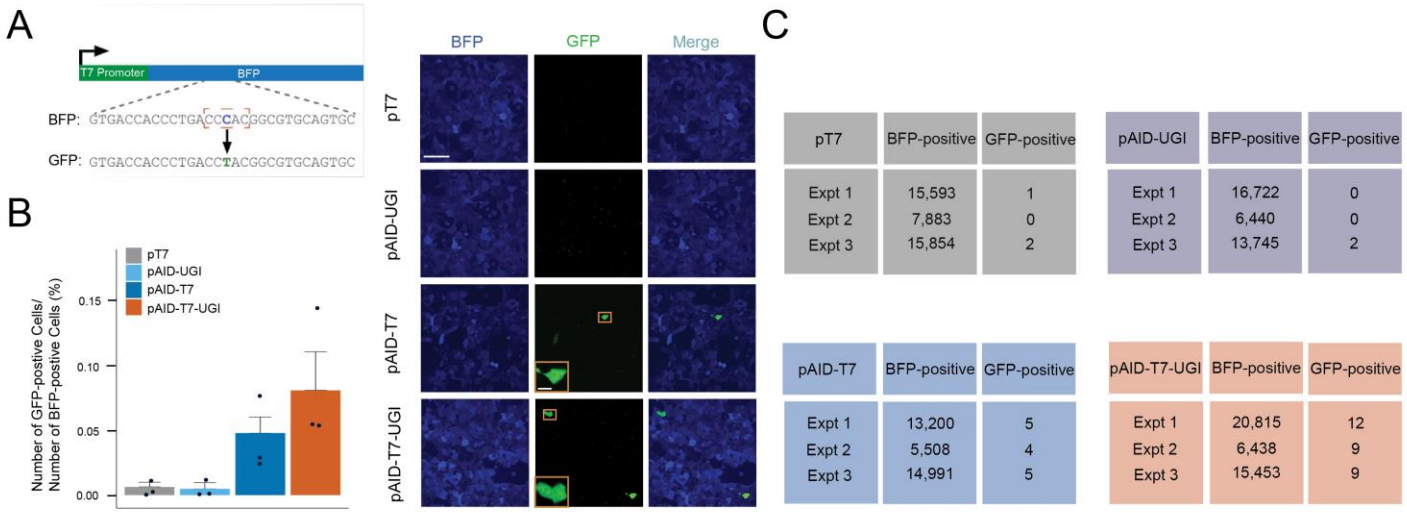


Supplementary Figure 6

Genome-integrated TRACE induces targeted somatic mutations in genomic loci.

A. Both the editor (pAID-T7) and the T7 promoter-controlled gene (T7 promoter-EBFP) were integrated into the genome of HEK293T cells. The expression of the editor protein was under the control of a doxycycline (dox)-inducible promoter (upper panel). Line plots showing C->T and G->A mutations in targeted gene loci over a period of 20 days in cells harboring the TRACE system (lower panel) starting from Dox induction. Each point is a mean ± SEM of N = 3 independent experiments. **B.** Cell viability (number of live cells/total number of cells x 100%) with and without dox induction at the indicated time points. Dox *vs*. no Dox, two-sided *t* test, $p = 0.1911$ for day 6, $p = 0.3771$ for day 10, $p = 0.4470$ for day 16, $p = 0.8675$ for day 20. All bars are mean ± SEM of N = 3 independent experiments.

Supplementary Figure 7



pT7	BFP-positive	GFP-positive
Expt 1	15,593	1
Expt 2	7,883	0
Expt 3	15,854	2

pAID-UGI	BFP-positive	GFP-positive
Expt 1	16,722	0
Expt 2	6,440	0
Expt 3	13,745	2

pAID-T7	BFP-positive	GFP-positive
Expt 1	13,200	5
Expt 2	5,508	4
Expt 3	14,991	5

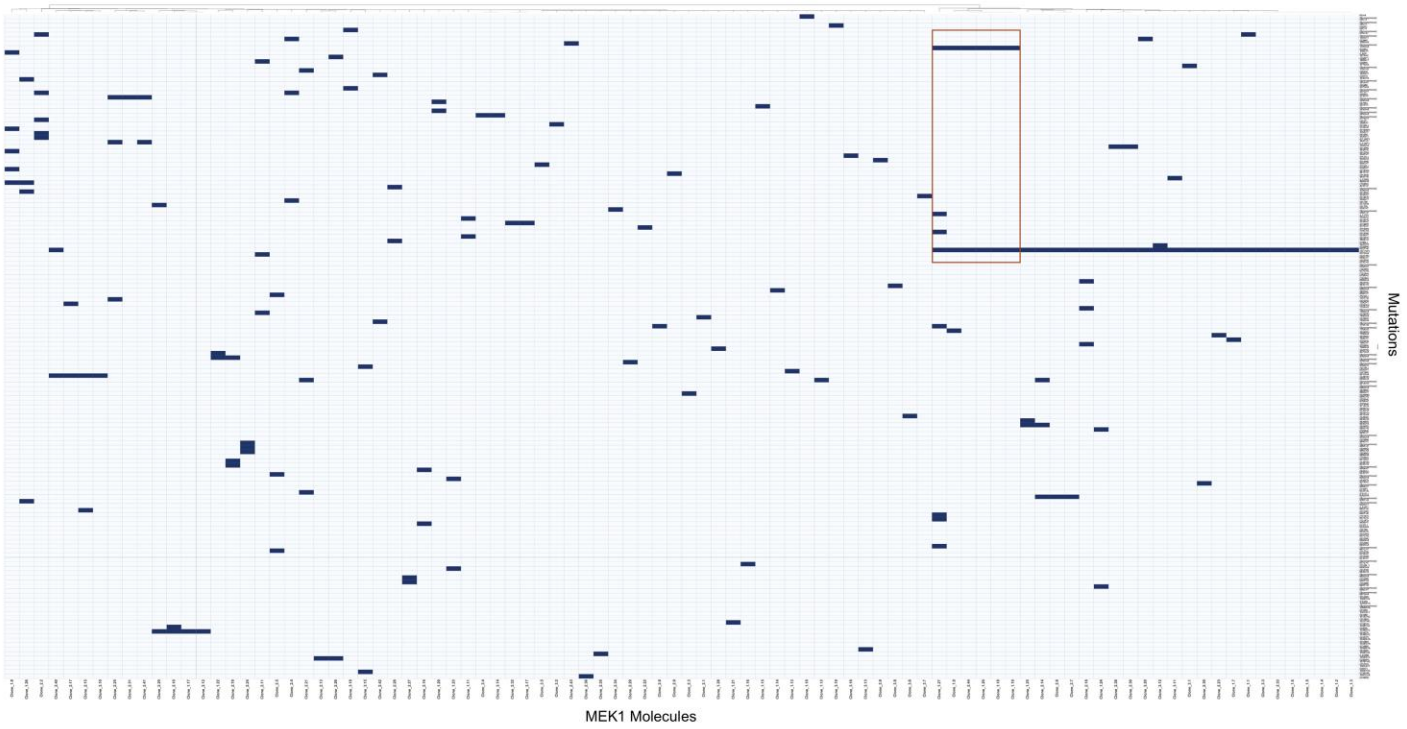
pAID-T7-UGI	BFP-positive	GFP-positive
Expt 1	20,815	12
Expt 2	6,438	9
Expt 3	15,453	9

Supplementary Figure 7

TRACE evolves BFP to GFP.

A. TRACE evolves BFP to GFP via a single H66Y amino acid substitution (left) and mutations can be visualized (right). The experiments were repeated 3 times with similar results. Scale bar, 75 μm; insets, 10 μm. **B**. Ratio of GFP-positive cells to BFP-positive cells. pAID-T7 vs. pAID-UGI, two-sided *t* test, p=0.0318; pAID-T7-UGI vs. pAID-UGI, two-sided t test, p=0.0441. Bars are the mean ± SEM of N=3 independent experiments. **C**. Summary counts for the number of BFP-positive cells and GFP-positive cells identified in microscopy images from 3 independent experiments.

Supplementary Figure 8



Supplementary Figure 8

Mutation profiles of single MEK1 molecules.

Mutation profiles of 93 single MEK1 molecules from trametinib-resistant colonies as identified by Sanger sequencing. Molecules containing both E38K and V211D mutations are highlighted with a red box.

Supplementary Table 1. Statistical analysis for Supplementary Figure 4B.

Group	Mutation	Test	*P* value
pAID-UGI vs. pAID-T7-UGI at -12C	C->T	two-sided *t* test	0.0592
pAID-UGI vs. pAID-T7-UGI at -9C	C->T	two-sided *t* test	0.3133
pAID-UGI vs. pAID-T7-UGI at -7C	C->T	two-sided *t* test	0.1603
pAID-UGI vs. pAID-T7-UGI at -5C	C->T	two-sided *t* test	0.0153
pAID-UGI vs. pAID-T7-UGI at -11G	G->A	two-sided *t* test	0.8400
pAID-UGI vs. pAID-T7-UGI at +1G	G->A	two-sided *t* test	0.2893

Supplementary Table 2. Statistical analysis for Figure 2A.

Group	Mutation	Test	*P* value
Day 5 vs. Day 1	C->T	two-sided *t* test	0.0194
Day 10 vs. Day 5	C->T	two-sided *t* test	0.0131
Day 15 vs. Day 10	C->T	two-sided *t* test	0.0093
Day 20 vs. Day 15	C->T	two-sided *t* test	0.0276
Day 5 vs. Day 1	G->A	two-sided *t* test	0.1157
Day 10 vs. Day 5	G->A	two-sided *t* test	0.3586
Day 15 vs. Day 10	G->A	two-sided *t* test	0.0040
Day 20 vs. Day 15	G->A	two-sided *t* test	0.0111

Supplementary Table 3. Statistical analysis for Supplementary Figure 6A.

Group	Mutation	Test	*P* value
Dox vs. no Dox on Day 6	C->T	two-sided *t* test	0.0093
Dox vs. no Dox on Day 10	C->T	two-sided *t* test	0.0006
Dox vs. no Dox on Day 16	C->T	two-sided *t* test	0.0037
Dox vs. no Dox on Day 20	C->T	two-sided *t* test	0.0249
Dox vs. no Dox on Day 6	G->A	two-sided *t* test	0.1355
Dox vs. no Dox on Day 10	G->A	two-sided *t* test	0.0026
Dox vs. no Dox on Day 16	G->A	two-sided *t* test	0.0002
Dox vs. no Dox on Day 20	G->A	two-sided *t* test	0.0080

Supplementary Table 4. Previous studies on the MEK1 mutations identified in the current screen.

Mutation	Information from previous studies	References
E38K	Previously identified as a recurrent missense mutation in MEK1 in patients with Langerhans cell histiocytosis (LCH).	28,29
G128D	Previously identified in LCH patients with wild type BRAF	30
V211D	Situated directly within the arylamine binding pocket of the MEK1 protein, previously demonstrated as resistant to both selumetinib and trametinib	31,32
L115R	X-ray crystal structures of the trametinib-MEK1 binding interaction suggests that this substitution prevents trametinib from making hydrophobic contacts with the MEK1 residue L115 within the αC-helix.	33

Note: apart from G->A mutations, TRACE also identified non-canonical base substitutions, such as T->A and T->G. These substitutions likely arose from previously characterized errors caused by DNA repair processes following AID activity[34-36].

References for Supplementary Table 4

28. Zeng, K. et al. BRAFV600E and MAP2K1 mutations in Langerhans cell histiocytosis occur predominantly in children. *Hematological Oncology* **35**, 845-851 (2017).
29. Brown, N.A. et al. High prevalence of somatic MAP2K1 mutations in BRAF V600E–negative Langerhans cell histiocytosis. *Blood* **124**, 1655-1658 (2014).
30. Nelson, D.S. et al. MAP2K1 and MAP3K1 mutations in Langerhans cell histiocytosis. *Genes Chromosomes & Cancer* **54**, 361-368 (2015).
31. Emery, C.M. et al. MEK1 mutations confer resistance to MEK and B-RAF inhibition. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20411-20416 (2009).
32. Oddo, D. et al. Molecular landscape of acquired resistance to targeted therapy combinations in BRAF mutant colorectal cancer. *Cancer Research* **76**, 4504-4515 (2016).
33. Roskoski Jr, R. Allosteric MEK1/2 inhibitors including cobimetanib and trametinib in the treatment of cutaneous melanomas. *Pharmacological Research* **117**, 20-31 (2017).

34. Gómez-González, B. & Aguilera, A. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8409-8414 (2007).
35. Spencer, J. & Dunn-Walters, D.K. Hypermutation at A-T Base Pairs: The A Nucleotide Replacement Spectrum Is Affected by Adjacent Nucleotides and There Is No Reverse Complementarity of Sequences Flanking Mutated A and T Nucleotides. *The Journal of Immunology* **175**, 5170-5177 (2005).
36. Thientosapol, E.S. et al. SAMHD1 enhances immunoglobulin hypermutation by promoting transversion mutation. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 4921-4926 (2018).

Supplementary Table 5. Statistical analysis (one-way ANOVA followed by Tukey test, *p*-values) for Figure 2G.

	0.001 µM	**0.03 µM**	**0.10 µM**	**0.32 µM**	**1.00 µM**	**3.16 µM**	**10.0 µM**	**31.6 µM**
WT vs. E38KV211D	0.0030	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0137
E38K vs. E38KV211D	0.0032	0.0010	0.0010	0.0010	0.0010	0.0010	0.0024	0.0179
V211D vs. E38KV211D	0.0056	0.0010	0.0012	0.0010	0.0010	0.0010	0.0023	0.0457

Supplementary Table 6. Comparison of TRACE to state-of-the-art methods for nucleotide diversification.

Method (ref)	Host species	Editing rate	Editing window
TRACE (our method)	Human	~2 kbp^{-1} per 3 days (mainly C->T & G->A)	At least 2000 bp per T7 promoter
CRISPR-X (4)	Human	~1 kbp^{-1} per ~12 days (all types of substitutions)	~100 bp per sgRNA
EvolvR (5)	*E.coli*	~0.05 kbp^{-1} per day (all types of substitutions)	~50 bp per sgRNA
OrthoRep (8)	Yeast	~0.16 kbp^{-1} per day (all types of substitutions)	At least 5000 bp