



RESEARCH ARTICLE SUMMARY

FUNCTIONAL GENOMICS

Helicase-assisted continuous editing for programmable mutagenesis of endogenous genomes

Xi Dawn Chen[†], Zeyu Chen[†], George Wythes, Yifan Zhang, Benno C. Orr, Gary Sun, Yu-Kai Chao, Andrea Navarro Torres, Ka Thao, Mounica Vallurupalli, Jing Sun, Mehdi Borji, Emre Tkacik, Haiqi Chen, Bradley E. Bernstein*, Fei Chen*

INTRODUCTION: A fundamental challenge of genomics is to chart the impact of the three billion bases in the human genome on protein function and gene regulation. Thus, a critical goal is to develop strategies for mutagenizing genomic sequences systematically and at high throughput. In particular, targeted mutagenesis of single genomic loci could emulate the natural evolution process to reveal sequence-structure relationships, gain- and loss-of-function phenotypes, and cooperative mutations. However, no method exists that can perform continuous mutagenesis at targeted regions in the endogenous genomes of mammalian cells.

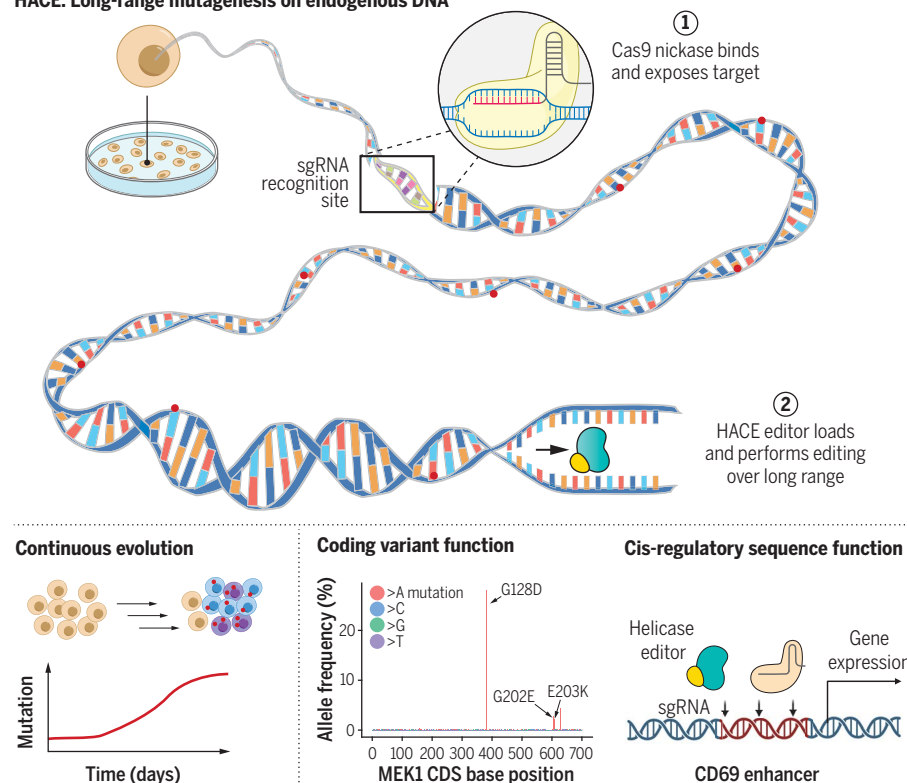
RATIONALE: We sought to develop a tool to perform targeted mutagenesis on the endogenous

mammalian genome. Looking to nature, we observed that helicases are highly processive enzymes that can traverse large genomic regions. Some helicases, including those involved in DNA damage repair, can load and start unwinding DNA at single-stranded DNA regions in the genome. We reasoned that such helicases could be used for long-range targeted mutagenesis when fused to a deaminase enzyme. The fusion construct and its interval of hypermutation could then be programmably targeted, through single-guide RNAs (sgRNAs), to specific genomic regions using a Cas9 nickase. The directional and long-range DNA-unwinding event by the recruited helicase will then generate random mutations in the region.

RESULTS: We designed a platform called helicase-assisted continuous editing (HACE), which combines long-range editing of entire loci with the sequence programmability inherent to CRISPR gene editing tools. HACE uses CRISPR-Cas9 to direct the loading of a helicase-deaminase fusion for targeted hypermutation of the downstream genomic sequence. HACE achieved locus-specific deamination across >1000 nucleotides with mutations continuously accumulating over time. We further evaluated HACE prototypes incorporating diverse helicases, Cas9 variants, and deaminases, showing that they have tunable edit rates and ranges. We also showed that HACE can be multiplexed to target multiple genomic regions with a minimal number of guide RNAs. We then applied HACE in coding and noncoding genomic contexts to functionally dissect endogenous mutations conferring drug resistance, changes in enzymatic activity, and altered cis-regulatory element function. In the coding space, we identified variants that lead to mitogen-activated protein kinase kinase 1 (MEK1)-inhibitor drug resistance and also identified variants in SF3B1, a splicing factor, that lead to alternative 3' splice-site usage. Turning to regulatory regions, we defined functional artificial variants in the enhancer regions of CD69 and pinpointed specific bases and motifs that mediate the impact of RUNX transcription factors on CD69 regulation. HACE solves two limitations faced by conventional base editing screens: the requirement of an NGG protospacer adjacent motif in the sgRNA recognition sequence and the occurrence of bystander mutations that can create artificial linkages and confound screening results. The long editing range of HACE can also uncover combinatorial effects and interactions between multiple distant mutations across a locus.

CONCLUSION: HACE makes possible the continuous, long-range, programmable diversification of endogenous mammalian genomes. We envision that HACE will substantially expand the functional genomics toolbox and enable the building of systematic sequence-function maps of both coding and noncoding genomes. Furthermore, HACE can be developed into a directed evolution system in the endogenous genome, enabling the selection of sequences for desired functions in mammalian biology. ■

HACE: Long-range mutagenesis on endogenous DNA



Development of a system to perform long-range targeted continuous mutagenesis of endogenous genomes.

The helicase-assisted continuous editing (HACE) system allows for long-range mutagenesis on endogenous DNA and also continuous evolution over multiple cell generations. We have applied HACE to identify functional variants in both the coding and noncoding genome. CDS, coding sequence; D, Asp; E, Glu; G, Gly; K, Lys.

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

*Corresponding author. Email: chenfb@broadinstitute.org (F.C.); bradley_bernstein@dfci.harvard.edu (B.E.B.)

[†]These authors contributed equally to this work.

Cite this article as X. D. Chen *et al.*, *Science* **386**, eadn5876 (2024). DOI: 10.1126/science.adn5876

READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.adn5876>

RESEARCH ARTICLE

FUNCTIONAL GENOMICS

Helicase-assisted continuous editing for programmable mutagenesis of endogenous genomes

Xi Dawn Chen^{1,2,3†}, Zeyu Chen^{1,4,5†}, George Wythes¹, Yifan Zhang¹, Benno C. Orr^{1,2}, Gary Sun^{4,5}, Yu-Kai Chao^{1,6}, Andrea Navarro Torres¹, Ka Thao⁴, Mounica Vallurupalli⁷, Jing Sun¹, Mehdi Borji¹, Emre Tkacik³, Haiqi Chen^{8,9}, Bradley E. Bernstein^{1,4,5,10*}, Fei Chen^{1,2*}

Deciphering the context-specific relationship between sequence and function is a major challenge in genomics. Existing tools for inducing locus-specific hypermutation and evolution in the native genome context are limited. Here we present a programmable platform for long-range, locus-specific hypermutation called helicase-assisted continuous editing (HACE). HACE leverages CRISPR-Cas9 to target a processive helicase-deaminase fusion that incurs mutations across large (>1000–base pair) genomic intervals. We applied HACE to identify mutations in mitogen-activated protein kinase kinase 1 (MEK1) that confer kinase inhibitor resistance, to dissect the impact of individual variants in splicing factor 3B subunit 1 (SF3B1)–dependent missplicing, and to evaluate noncoding variants in a stimulation-dependent immune enhancer of CD69. HACE provides a powerful tool for investigating coding and noncoding variants, uncovering combinatorial sequence-to-function relationships, and evolving new biological functions.

A fundamental challenge of genomics is to chart the impact of the three billion bases in the human genome on protein function and gene regulation. Therefore, a critical goal is to develop strategies for mutagenizing genomic sequences systematically and at high throughput. In particular, targeted mutagenesis of single genomic loci could emulate the natural evolution process to reveal sequence-structure relationships and gain- or loss-of-function phenotypes. By performing such mutagenesis and selection in a stepwise and/or continuous manner, this evolutionary process could be directed to generate enhanced protein functions, gene expression, or cell fitness.

However, targeted mutagenesis in the endogenous mammalian genome remains difficult for three primary reasons. First, many existing tools rely on exogenous overexpression of the

gene of interest on a plasmid or vector [e.g., deep mutational scanning (1), VEGAS (2)]. Exogenous overexpression is sensitive to gene dosage and cannot be used to evolve noncoding regions in their native chromatin contexts. Second, some tools [e.g., TRACE (3), TRIDENT (4)] require integrating exogenous sequences into the genome, which leads to experimental complexity and constrains throughput. Third, existing tools targeting the endogenous genome are either nonspecific [e.g., alkylators (5) that introduce genome-wide mutations] or confined to narrow genomic windows [e.g., CRISPR base editors (6, 7), CRISPR-X (8), or TAM (9)]. Whereas CRISPR base-editor screens have been used to investigate protein function and regulatory elements, they are limited in the base positions that can be targeted with high efficiency and can lead to artificial variant linkage owing to the base editor mutating multiple bases in the editing window (10).

Given the existing challenges in performing targeted and continuous mutagenesis in mammalian cells, we sought to develop a platform for mutagenizing endogenous loci in their native chromatin contexts. We envisioned a tool with several advantageous properties, including (i) a long mutagenesis range [>200 base pairs (bp)]; (ii) the capacity to incur multiple, potentially interacting mutations across a region of interest; (iii) a continuous and potentially tunable mutation rate for sampling variant space and exploring fitness landscape changes; and (iv) a generalizable technical framework to target genomic loci of interest individually and in combination.

Here we introduce helicase-assisted continuous editing, or HACE, which combines long-

range editing of entire loci with the advantages in sequence programmability inherent to CRISPR gene editing tools. HACE uses CRISPR-Cas9 to direct the loading of a helicase-deaminase fusion for targeted hypermutation of the downstream genomic sequence (Fig. 1, A and B). We evaluated HACE prototypes incorporating diverse helicases, Cas9 variants, and deaminases, showing that they have tunable editing rates and ranges. We then applied HACE to functionally characterize both coding and noncoding genomic elements. In the coding space, we identified variants that lead to mitogen-activated protein kinase kinase 1 (MEK1)–inhibitor drug resistance and also variants in splicing factor 3B subunit 1 (SF3B1) that lead to alternative 3' splice-site (3'ss) usage. Turning to regulatory regions, we defined functional artificial variants in the enhancer regions of CD69 and pinpointed specific bases and motifs that mediate the impact of RUNX transcription factors on CD69 regulation. HACE provides a novel strategy for targeted mutagenesis of coding genes and regulatory regions, identifying specific nucleotides that underlie genotype-phenotype associations and evolving new cellular functions.

HACE design and implementation

Helicases are highly processive enzymes that can traverse large genomic regions. Certain helicases, including those involved in DNA damage repair, can load and start unwinding DNA at single-stranded DNA (ssDNA) regions in the genome (11). We reasoned that such helicases could be used for endogenous mutagenesis. We hypothesized that a helicase and deaminase enzyme fusion could generate multiple somatic edits across an extended genomic window in the endogenous mammalian genome, similar in concept to previous targeted mutagenesis using the Cascade complex and Cas3 in *Saccharomyces cerevisiae* [CoMuTER (12)] and T7-directed mutagenesis through polymerase translocation [TRACE (3)]. The fusion and its interval of hypermutation could be targeted to specific genomic regions using a nicking variant of SpCas9 [nCas9; Asp¹⁰→Ala (D10A)].

To test this, we constructed a HACE editor (HE) protein by fusing a hyperactive mutant of activation-induced cytidine deaminase (AID*Δ) with a *Geobacillus stearothermophilus* PcrA (plasmid copy reduced A) helicase that was previously optimized for processivity (PcrA M6) (8, 13). We also appended a uracil DNA glycosylase inhibitor (UGI), which has been shown to facilitate C:G>T:A mutations (14), at the 3' end of the HE (Fig. 1C). To target the HE to the desired loci, we used the nCas9 variant and a corresponding single-guide RNA (sgRNA). We reasoned that nCas9 binding would direct the HE to load onto the exposed ssDNA, start unwinding the DNA as the helicase translocates, and generate random mutations in the process.

¹Gene Regulation Observatory, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. ²Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA. ³Systems, Synthetic, and Quantitative Biology PhD Program, Harvard University, Cambridge, MA 02138, USA. ⁴Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. ⁵Department of Cell Biology and Pathology, Harvard Medical School, Boston, MA 02115, USA. ⁶Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. ⁷Cancer Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. ⁸Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. ⁹Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. ¹⁰The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

*Corresponding author. Email: chenfb@broadinstitute.org (F.C.); bradley_bernstein@dfci.harvard.edu (B.E.B.)

†These authors contributed equally to this work.

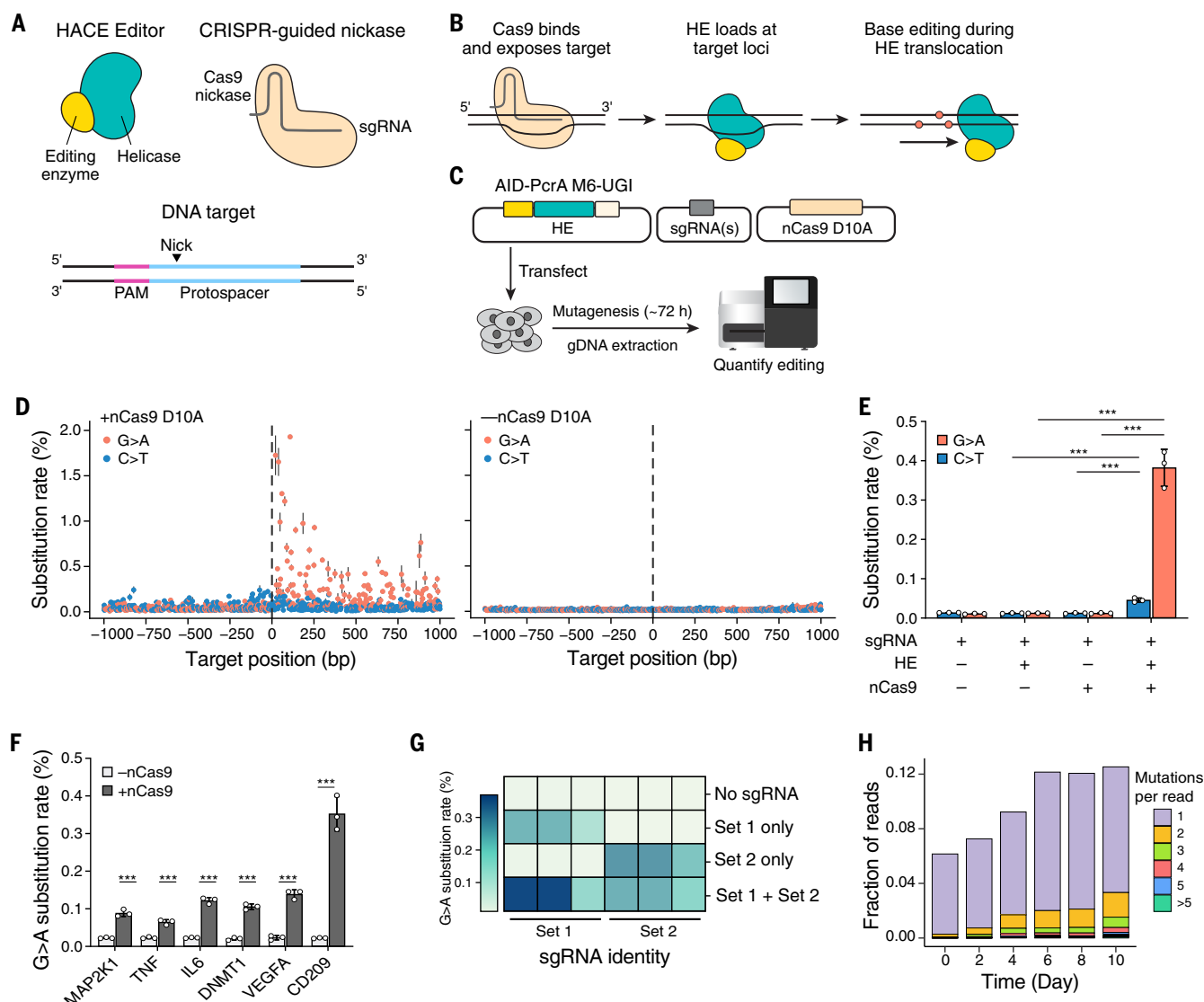


Fig. 1. Overview of the helicase-assisted continuous editing (HACE) system.

(A) Schematic of a HACE editor (HE), which is a fusion protein of helicase and base-editing enzymes (left). A CRISPR-guided nickase (Cas9 nickase) binds with a sgRNA to expose the target genomic DNA target (right). (B) Schematic model for the HACE system. A CRISPR-guided nickase targets a specific genomic position to expose the target region. The HE loads to this site. As the helicase translocates along the DNA, the base editor introduces random point mutations (orange circles). (C) Schematic of HACE experimental workflow, involving cotransfection of an HE plasmid [PcrA helicase variant (PcrA M6) fused with a hyperactive mutant of activation-induced cytidine deaminase (AID* Δ) and uracil DNA glycosylase inhibitor (UGI)], a nSpCas9(D10A) plasmid, and sgRNA plasmid(s). Editing efficiency is assessed 72 hours posttransfection through amplicon sequencing. (D) Substitution rate per base across a ~1-kb target region in the presence (left) and absence (right) of nCas9. All samples are transfected with HE and sgRNA. The vertical dashed line shows the nick site. Data are mean of experimental replicates ($n = 3$) \pm SD.

(E) Substitution rate at the target loci (*HEK3*). The average substitution rate is calculated for both C>T and G>A modes for the region downstream of the nick site, excluding the sgRNA spacer region (methods). Data are mean of experimental replicates ($n = 3$) \pm SD. Significance is determined by unpaired two-tailed t test between \pm HE or \pm nCas9 samples. *** $P < 0.001$. (F) G>A substitution rate across genomic targets using loci-specific sgRNAs in the presence or absence of nCas9. All samples are transfected with HE and sgRNA. The substitution rate is calculated for the region downstream of the nick site, excluding the sgRNA spacer region. Data are mean of experimental replicates ($n = 3$) \pm SD. Significance is determined by unpaired two-tailed t test between \pm nCas9 samples. *** $P < 0.001$. (G) Average substitution rate for multiplex sgRNA targeting. Two sets of sgRNAs (three sgRNAs per set; see table S1) are independently cotransfected with other HACE components. The substitution rate at each sgRNA target locus is depicted in the heatmap. (H) Distribution of substitutions per contiguous read over time points for *HEK3* loci.

To this end, we cotransfected separate plasmids expressing HE, nCas9, and an sgRNA targeting the *HEK3* locus into human embryonic kidney 293FT (HEK293FT) cells. Cells were collected 72 hours after transfection, and editing

rates were evaluated by amplicon sequencing. We observed editing over the targeted region in the presence of the HE, nCas9, and sgRNA (Fig. 1D). We did not observe elevated mutation levels in cells transfected with only HE or

only with nCas9 and sgRNA, suggesting that editing is driven by the HE and is guide-specific (Fig. 1E). We detected strand and directional bias, with elevated G>A substitution rates across a ~1000-bp window downstream of the nick

site. The directionality is consistent with a model in which the helicase fusion preferentially loads onto the nontarget strand of the Cas9 (i.e., the strand that is not bound by the sgRNA), which has exposed ssDNA, and then translocates in the 3' to 5' direction (fig. S1A). This directionality is consistent with known mechanisms of PcrA helicases (15, 16). The strand bias likely reflects preferential repair of mismatches on the nicked strand (14), as the deaminase installs C>T substitutions on the nontarget strand that has the exposed ssDNA.

Quantifying the editing rate downstream of the nick site, we observed an average G>A substitution rate of 0.38% per base. As expected from the directional cytidine deamination primarily on the nontarget strand, we observed a lower average C>T substitution rate of 0.046% per base (Fig. 1E; see methods). This substitution rate is significantly higher than that of cells transfected with only nCas9 or HE (unpaired *t* test, $P < 0.001$ for \pm nCas9 and \pm HE in both C>T and G>A groups). It is also substantially higher than the replication error rate of human cells (17). The editing rate upstream of the nick site was also elevated compared with cells transfected with only nCas9 or HE (fig. S1B), although the absolute editing rate was ~10-fold lower than downstream (unpaired *t* test, $P < 0.01$ for \pm nCas9 in both C>T and G>A groups). The rates of other transition and transversion mutation modes were low and comparable to background sgRNA transfection absent of Cas9 and HE, providing further support for the specificity and targeting of our fusion protein (fig. S1C).

To demonstrate that HACE can be targeted to diverse loci, we cotransfected the HE and nCas9 with different sgRNAs targeting different genomic loci (tables S1 and S2). We observed elevated substitution rates in these loci only in the presence of HE, sgRNA, and nCas9, but not in the absence of nCas9, suggesting that editing is guide- and nCas9-dependent (Fig. 1F and fig. S1D). In all locations, editing is consistently higher downstream of the sgRNA spacer than upstream (fig. S1E), confirming that editing is directionally biased. We also detected a low (<1%) indel (insertion or deletion) rate across all loci (fig. S1F). The editing rate remained stable over a 10-day period, with fluctuations within a twofold range between days 4 and 10 postediting (fig. S1G).

The modular architecture of HACE can also enable the simultaneous diversification of distinct genomic loci by coexpressing multiple sgRNAs. We coexpressed sets of three sgRNAs that target distant genomic loci. We observed elevated substitution rates in all target loci but not in untargeted control regions (Fig. 1G and fig. S1H). We also compared the editing efficiencies at target loci when a single sgRNA was used versus when multiplexed. We observed that the editing efficiencies when multiplexed

remained within a threefold range of those observed when targeted individually (fig. S1I). This suggests that HACE can couple hypermutation across distant genomic loci and thus enable multiplexing or evaluation of genetic interactions.

We also tested whether HACE could induce continuous nucleotide diversification over multiple cellular generations. We constructed an HEK293FT cell line with an integrated inducible HE. We induced HE expression after transfection of nCas9 and sgRNA and monitored the editing rate every 2 days over a 10-day period. We observed a continuous increase in substitution rates across the five loci profiled (fig. S1J), with the number of mutations per contiguous Illumina sequencing read increasing over time points (Fig. 1H and fig. S1K). These data suggest that HACE can induce continuous nucleotide diversification across 10 cell generations (~1 day per generation), making it amenable to continuous evolution applications.

HACE is flexible and programmable

We sought to optimize HACE components to modulate editing rate, range, and mutational modes. Broadly, the HACE system consists of four modular components: the helicase and editing enzyme that make up the HE and the nCas9 and sgRNA that direct the fusion to target loci (Fig. 2A). We began by designing and testing HE constructs using six different monomeric helicases and shortlisted four for systematic profiling (fig. S2A and tables S3 and S4). We also tested two variants of nCas9, one that nicks the target strand (nCas9 D10A) and one that nicks the nontarget strand (nCas9 H840A). We cloned constructs encoding different combinations of helicase and nCas9 variants, targeted the fusions to three different genomic loci (*HEK3*, *TNF*, and *IL6*), and examined substitution rate and editing range.

First, we tested HACE constructs with different nickase enzymes. We detected elevated substitution rates using either target-strand nickase (nCas9 D10A) or nontarget-strand nickase (nCas9 H840A), with all constructs showing significant editing across the three loci for at least one nickase variant (Fig. 2B). The preference for target- or nontarget-strand nickases varied depending on the locus (Fig. 2B). We then characterized the editing range of different HE constructs. For each helicase, nickase, and loci combination, we transfected the respective HE, nCas9, and sgRNA combination into HEK293FT cells, harvested genomic DNA after 3 days, and then amplified and sequenced a 1000-bp window surrounding the nick. We found that the helicases preferred to translocate in the downstream direction for both target-strand (nCas9 D10A) and nontarget-strand nickase (nCas9 H840A) (Fig. 2, C and D, and fig. S2, B and C). Although we initially reasoned

that nick generation may be required for directing the helicase, the consistent directional bias relative to sgRNA binding for both nickase variants suggested that the Cas9 binding event might be sufficient. This raised the possibility that HACE could be adapted to a catalytically inactive Cas9 (dCas9) and thus avoid generating DNA nicks. We therefore tested HACE constructs targeted with dCas9, focusing again on the *HEK3* locus. We found that HE fusions targeted with dCas9 also preferentially incurred base edits in the downstream direction (fig. S3, A to C). However, the absolute editing rates were lower with dCas9 than nCas9 D10A for three of the four HE constructs tested. Overall, our data suggest that the helicases are directed by the relative strand of sgRNA binding rather than the position of the nick.

To analyze the editing range, we calculated the local average substitution rate in 100-bp bins for each genomic locus (fig. S4A; methods). We observed a decreasing substitution rate as a function of distance from the nick for all helicases profiled with both nickase variants. BLM, NS3h, and PcrA M6 helicases all demonstrated elevated editing ($>10^{-3}$ G>A substitution rate per base) within 500 bp from the nick site. The substitution rate of PcrA M6 stabilized past 500 bp (at $\sim 10^{-3}$ G>A substitution rate per base), suggesting consistent, long-range editing up to 1000 bp from the nick site. This range is an order of magnitude longer than that achieved by previous Cas9-directed editing tools (8, 9). To determine the maximum editing range, we sequenced 10 kb downstream of the nick site at the *HEK3* loci. We observed that editing falls below the detection threshold (G>A substitution rate of 0.1%) at ~2 kb across all HE variants profiled (fig. S4B).

Next, we explored different base-editing enzymes for HACE. To identify the most efficient base editing enzyme and access diverse mutation modes, we engineered HEs fused with diverse deaminases, including (i) other cytosine deaminase enzymes that introduce C>T and G>A substitutions [rAPOBEC1 (18)], (ii) adenosine deaminase enzymes that introduce A>G and T>C substitutions [TadA-8e (19)], and (iii) an engineered dual base editor that can perform both cytosine and adenine base editing [TadDE (20)]. We tested these constructs in HEK293FT cells and quantified substitution rates by amplicon sequencing (Fig. 2, E and F). We observed that both rAPOBEC1 and AID*Δ perform well in introducing G>A base edits. We found that TadA was able to induce high rates of T>C edits. On the other hand, the dual TadDE editor only induced minor levels of G>A and T>C editing. The deaminases rAPOBEC1, AID*Δ, and TadA-8e introduced mutations across diverse genomic loci (fig. S4C), demonstrating that HACE can use different deaminase fusions to introduce diverse base edits in a programmable (i.e., sequence-directed) manner.

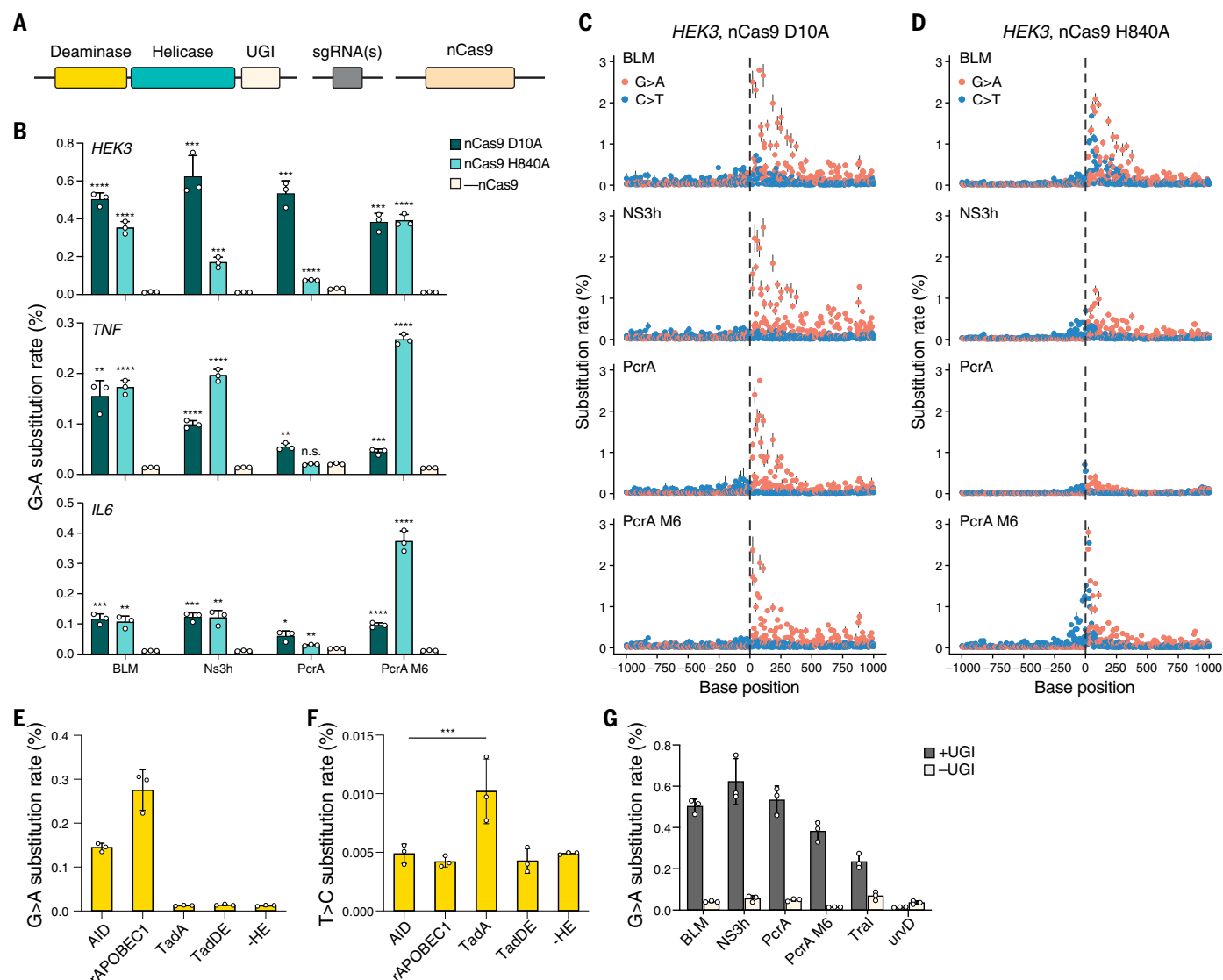


Fig. 2. The HACE system is modular and flexible. (A) Modular components of the HACE system. Each component can be independently substituted to control for editing efficiency, mode, and range. (B) G>A substitution rate at the *HEK3*, *TNF*, and *IL6* loci for HEs with different helicase variants and the nCas9 (D10A) or nCas9(H840A) nickase variant compared with -nCas9 condition. Significance is determined by unpaired two-tailed *t* test between control and nCas9 samples with multiple-testing correction. n.s., not significant. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001. (C) Substitution rate per base across a ~1-kb target region for HEs with different helicase variants and the nCas9(D10A) nickase variant. The vertical dashed line shows the nick site. Data are mean of

experimental replicates (*n* = 3) ± SD. (D) Substitution rate per base across a ~1-kb target region for HEs with different helicase variants and the nCas9 (H840A) nickase variant. The vertical dashed line shows the nick site. Data are mean of experimental replicates (*n* = 3) ± SD. (E) Average G>A substitution rate at the *HEK3* loci for HEs with different deaminase variants fused to the PcrA M6 helicase. (F) Average T>C substitution rate at the *HEK3* loci for HEs with different deaminase variants fused to the PcrA M6 helicase. Significance is determined by unpaired two-tailed *t* test between AID and TadaA groups. ****P* < 0.001. (G) Substitution rate at the *HEK3* loci for HEs with different variants with and without UGI fusion. All data are mean of experimental replicates (*n* = 3) ± SD.

Lastly, we found that the fusion of UGI significantly elevated the editing levels for AID (Fig. 2G, unpaired *t* test, *P* < 0.05 for all ±UGI groups), consistent with reports from previous cytidine-base editor studies (14). These results demonstrate that HACE editing rates can be tuned by varying the helicase and Cas9 variants and by incorporating UGI, making HACE suitable for diverse applications. Further engineering may leverage these insights to opti-

mize the editing rate and range for specific applications.

HACE is minimally perturbative in mammalian cells

Although we found the HACE constructs to be well tolerated in transfection experiments, we sought to quantify the effects of different HACE constructs on cell viability. To do so, we quantified cell viability using a luciferase-based aden-

osine triphosphate assay (CellTiter-Glo) across various helicase constructs—both with and without deaminase—along with a locus-targeting sgRNA and nCas9 (fig. S5A). We found that HEs constructed with BLM and PcrA helicases did not result in a significant decrease in cell viability (unpaired *t* test, *P* > 0.05 for each group). However, AID-NS3h-UGI did decrease in cell viability (unpaired *t* test, *P* < 0.05), possibly because of the toxicity of NS3h helicases,

which also act on RNA (21, 22). Consistently, we observed a similar decrease in cell viability for Ns3h-UGI (unpaired *t* test, $P < 0.001$). AID-PcrA M6-UGI also significantly decreased cell viability (unpaired *t* test, $P < 0.001$), whereas PcrA M6 alone did not affect cell viability (unpaired *t* test, $P = 0.118$). These results suggest that helicases used for HACE are well tolerated for cell viability.

Lastly, we explored whether HACE generated elevated substitution rates in nontargeted parts of the genome. To do so, we subjected cells expressing different HE variants or AID to whole-exome sequencing at high coverage ($\sim 1000\times$ coverage). To increase our detection power, we divided the genome into 100-kb bins, calculated the editing rate of each bin, and then compared the editing rate between HE variants and control cells (fig. S5B). In the 16,621 genomic bins, we observed that overexpression of AID alone generated the most significant

off-target bins (38 bins, Fisher's exact test). We detected 13 bins with elevated editing rates for AID-NS3h-UGI, 5 for AID-PcrA M6-UGI, 3 for AID-PcrA-UGI, and 3 for AID-BLM-UGI. We found nine bins that were elevated across multiple HE constructs, indicating common off-target sites across helicases (all sites outlined in table S5). These data suggest that there is minimal elevation of global substitution rates in the exome due to HACE and that off-target editing is likely driven by AID overexpression rather than by HE itself.

We also performed whole-exome sequencing of cells expressing HE with alternative base editors and compared against control cells (fig. S5C). We found that all base editors (raPOBE1, TadA-8e, and TadDE) generated, at most, three bins with elevated editing rates for either C>T or A>G editing modes (table S5), again suggesting minimal elevation of global substitution rates.

HACE enables the identification of MEK1 inhibitor-resistance mutations

We sought to apply HACE to perform functional mutagenesis in both coding and non-coding genome contexts. We initially screened for mutations within MEK1 (also known as MAP2K1) that promote resistance to small-molecule drug inhibition. MEK inhibitors target the MAPK-ERK (extracellular signal-regulated kinase) pathway, which is aberrantly up-regulated in one-third of all cancers (23). We used HACE to diversify exons of the *MEK1* gene in A375 cells, a melanoma line sensitive to MEK inhibition. After 3 days of mutagenesis using AID Δ -PcrA M6-UGI, we selected cells for resistance to two MEK1 inhibitors: selumetinib and trametinib (Fig. 3A). We targeted exons 2, 3, and 6, which contain previously identified mutation hotspots (24). Because the mutagenesis range of HACE is long, we only needed to design one sgRNA per exon. We placed each

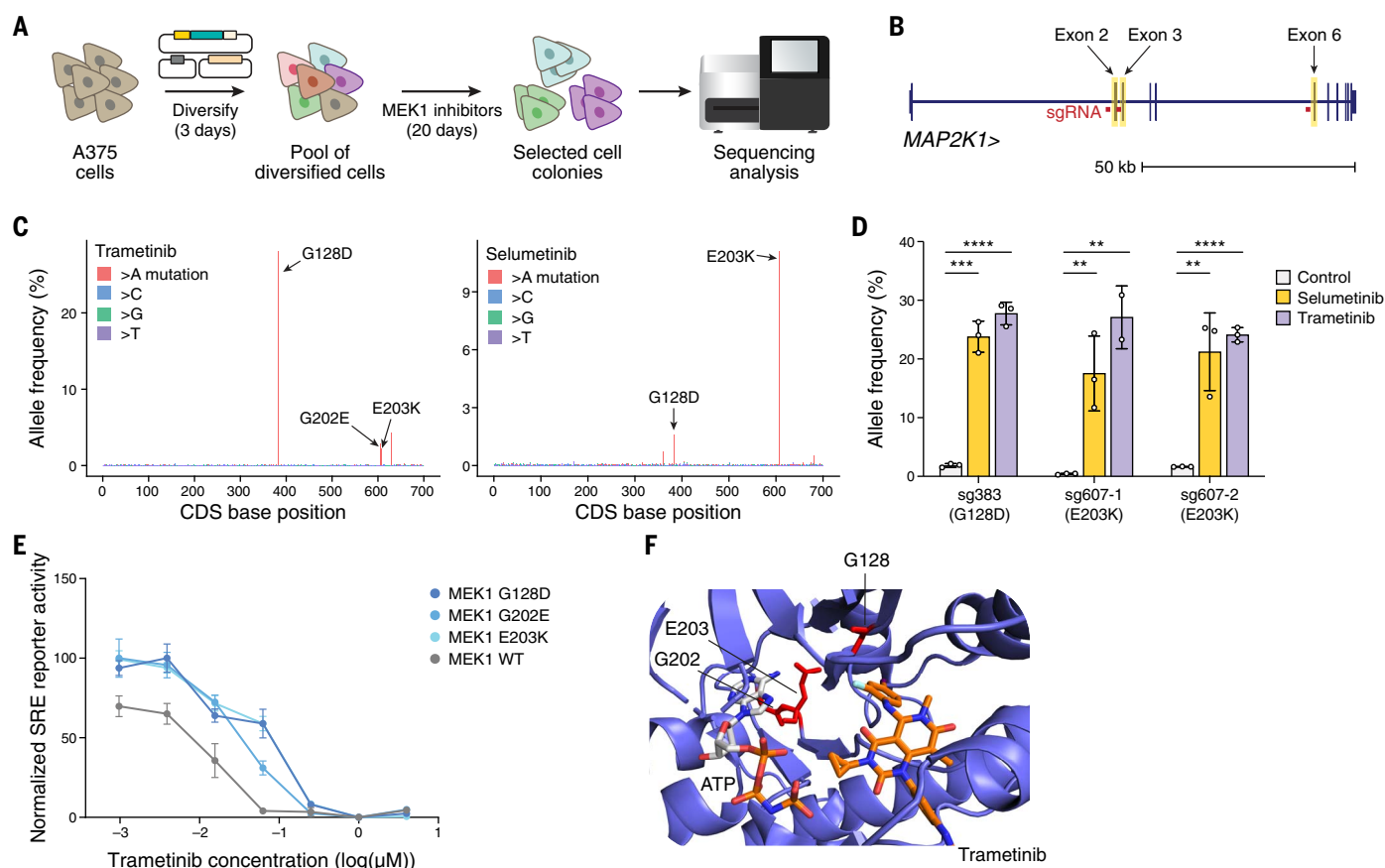


Fig. 3. HACE enables the identification of MEK1 inhibitor-resistance mutations in the endogenous genome. (A) Workflow of HACE MEK1 inhibitor-resistance screen. A375 cells are transfected with HACE and diversified for 3 days. The genomic diversified cells are selected for 20 days. Genomes of resistant clones are harvested and sequenced by amplicon sequencing.

(B) Location of sgRNAs for HACE screen. Exons 2, 3, and 6 (highlighted in yellow) are targeted for HACE diversification. Each exon-specific sgRNA (red bar) is placed ~ 100 bp upstream of the target exon. (C) Fold enrichment of MEK1 cDNA sequence in trametinib-treated (left) and selumetinib-treated (right)

samples. (D) Enrichment of mutations installed by base editing targeting G128D (sg383) and E203K (sg607-1 and sg607-2) after trametinib or selumetinib treatment. Samples are sequenced 14 days postselection by amplicon sequencing. Significance is determined by an unpaired two-tailed *t* test between control and drug-selected samples. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. (E) MAPK-ERK signaling activity as measured by luciferase SRE reporter activity for G128D, G202E, and E203K mutants (mean \pm SEM, $n = 3$ independent experiments). (F) Structure of MEK1 in complex with trametinib [Protein Data Bank (PDB) ID 7JUR]. MEK1 protein is depicted in blue, and residues that lead to drug resistance are annotated in red.

exon-specific sgRNA ~100 bp upstream of the exon, within the intronic region (Fig. 3B and table S1). We recovered drug-tolerant mutants and identified *MEK1* mutations by targeted amplicon sequencing (methods). By comparing allele frequencies between pre- and post-drug selection samples and identifying alleles that are significantly enriched after drug selection, we identified three candidate mutations that conferred resistance to trametinib (G128D, G202E, and E203K) and two candidate mutations that conferred resistance to selumetinib (G128D and E203K) (Fig. 3C, fig. S6, and table S6). Two of the mutations, G128D and E203K, conferred resistance to both selumetinib and trametinib, suggesting similar resistance mechanisms.

To validate top mutation candidates, we designed sgRNAs to introduce mutations individually into A375 cells using base editing. After 20 days of selection in the presence of drug, we evaluated allele frequencies of introduced mutations before and after selection by amplicon sequencing (Fig. 3D). We observed substantial postselection enrichment of G128D (sg383) and E203K (sg607-1 and sg607-2) with both inhibitors (sgRNAs in table S11). We could not introduce the 605G>A (G202E) mutation by base editing owing to the artificial linkage in base editing between G202E and E203K. We therefore further validated candidates that conferred resistance to trametinib using a luciferase serum response element (SRE) reporter of MAPK-ERK signaling activity and overexpression of candidate *MEK1* mutants (Fig. 3E and table S7). All three mutations individually increased trametinib resistance [median inhibitory concentration (IC_{50}) = 68.0, 46.1, and 46.1 nM for G128D, G202E, and E203K, respectively, versus 5.28 nM for wild type]. Structural analysis revealed that G128D is in the ligand-binding pocket. This mutation may function by inducing conformational changes of the binding pocket through steric interactions (Fig. 3F). On the other hand, E203K has been shown to cause constitutive *MEK1* activation and downstream ERK phosphorylation, conferring gain of function (25, 26). It is possible that the proximal G202E likely induces resistance through a mechanism similar to that of E203K. Overall, this demonstrates that HACE can identify mutations conferring drug resistance, while reducing confounding effects of artificial genetic linkage. Although previous studies have screened for *MEK1* inhibitor resistance (3, 24), our approach is a mutagenesis-based resistance screen performed in the endogenous genome context.

HACE identifies *SF3B1* variants that result in alternative 3' branch point usage

Next, we applied HACE to explore the function of individual variants in *SF3B1* for splicing regulation. RNA splicing factor mutations occur in various cancers and are especially prevalent

in hematopoietic malignancies (27–29). The most frequently mutated splicing factor, *SF3B1*, is a member of the U2 small nuclear ribonucleoprotein complex that binds to the branch point nucleotide in the precatytic spliceosome (30, 31). Pan-cancer analysis of *SF3B1* mutations has identified hotspot mutations clustered within C-terminal HEAT repeat domains 4 to 8, which display an alternative 3'ss usage signature (32, 33) (Fig. 4A). This missplicing occurs through the recognition of a different branch point sequence during 3'ss selection and results in global splicing changes associated with tumorigenesis. However, few clinical mutations have been functionally validated for their effect on splicing, and, moreover, these clinical variants only represent a small subset of the functional space of *SF3B1* variants. Distinguishing the mutations in *SF3B1* that drive missplicing could improve understanding of splicing biology and potentially guide new diagnostic and therapeutic strategies.

To screen for mutations in *SF3B1* that functionally lead to missplicing, we first sought to construct a minigene reporter that could distinguish between wild-type *SF3B1* (*SF3B1*^{WT}) and mutated *SF3B1* splicing patterns. These patterns display alternative 3'ss usage characteristic of hematopoietic malignancies. First, we compared RNA sequencing (RNA-seq) data from isogenic K562 cells containing either *SF3B1*^{WT} or mutant *SF3B1* (*SF3B1*^{K700E}, a mutation known to induce the alternative 3'ss phenotype). We shortlisted splicing events that were differential between WT and mutant cells, and constructed minigene reporters from two of the top sequences to test their ability to functionally distinguish between *SF3B1*^{WT}- and *SF3B1*^{K700E}-induced missplicing. These reporters concatenated a constant upstream exon and a green fluorescent protein (GFP) reporter, followed by the last 150 bp of the endogenous intron and its downstream exon (Fig. 4B and table S8; methods). They were designed so that correct splicing would result in a frameshift in the open reading frame of GFP, suppressing fluorescence, while missplicing would permit GFP expression. Each minigene-GFP sequence was cloned into a vector with constitutive mCherry expression, enabling quantitative assessment of mutation-dependent protein production through flow cytometry (methods).

To validate these splicing reporters, we transfected each construct into isogenic *SF3B1*^{WT} or *SF3B1*^{K700E} K562 cells and measured mutant-dependent protein expression by flow cytometry. Both reporters demonstrated mutant-dependent specificity, showing elevated GFP expression in *SF3B1*^{K700E} cells compared with *SF3B1*^{WT} cells (fig S5A). We confirmed that alternative 3'ss usage drives the reporter expression using targeted RNA-seq (fig. S7B). We selected the

reporter with the highest mutant-dependent specificity, which was derived from dihydro-lipoamide *S*-succinyltransferase (*DLST*) exon 6, for the *SF3B1* variant screen (Fig. 4C).

We used HACE to diversify exons 13 to 17 of the *SF3B1* gene in HEK293FT cells for 3 days. This region corresponds to HEAT repeat domains 4 to 8, which are enriched for known missplicing mutations. As in the *MEK1* screen, we targeted these exons with seven sgRNAs complementary to flanking intronic sequences. To widen the genetic search space, we used HACE editors that incur both C:G>T:A and A:T>G:C mutations (AID*Δ-PcrA M6-UGI and TadaA-8e-PcrA M6-UGI). We then transfected the minigene reporter into the edited pool of cells, sorted GFP[−] and GFP⁺ cells on the basis of the GFP:mCherry ratio, and performed high-throughput sequencing (Fig. 4D). Enriched mutations were identified by comparing the mutational frequencies between GFP[−] and GFP⁺ cells (Fig. 4E and table S9). We observed a high degree of replicate correlation for enriched variants between two independent biological replicates (Pearson's $\rho = 0.795$). Furthermore, nine of the most highly enriched variants in our assay (>10-fold enrichment) corresponded to clinically observed mutations (34) (fig. S7C).

To validate our hits, we used conventional base editing to introduce the top candidates into HEK293FT cells coexpressing the minigene reporter. We then quantified the fold change in the GFP:mCherry ratio compared with unedited cells. Three mutations (I617V, Y623C, and K666E) led to a significant increase in reporter fold change (unpaired *t* test, $P < 0.001$ for all base editing groups compared with control; Fig. 4F and fig. S7D). The reporter fold change values were similar to those observed for K700E, a well-validated mutation affecting *SF3B1* alternative 3'ss usage (35). We also performed targeted amplicon sequencing for each cell population to validate editing at each target base (fig. S7E). We note that two of the mutations, Y623C and K666E, have been observed in clinical datasets and are highly enriched in hematopoietic tumor samples (27, 36). K666E has been previously validated for its effect on alternative 3'ss usage (35, 37). An additional top-ranked mutation, I617V/M, was not previously observed in clinical datasets. We further validated these top candidate mutations by prime editing (Fig. 4G, fig. S7F, and table S12). Despite the low editing efficiency, we observed significantly increased splicing reporter fold changes for the mutations, relative to WT cells (unpaired *t* test, $P < 0.01$ for all prime editing groups compared with control). Overall, the editing rate for candidate mutations that affect *SF3B1* alternative 3'ss usage across validation experiments correlates well with the minigene reporter fold change (fig. S7G).

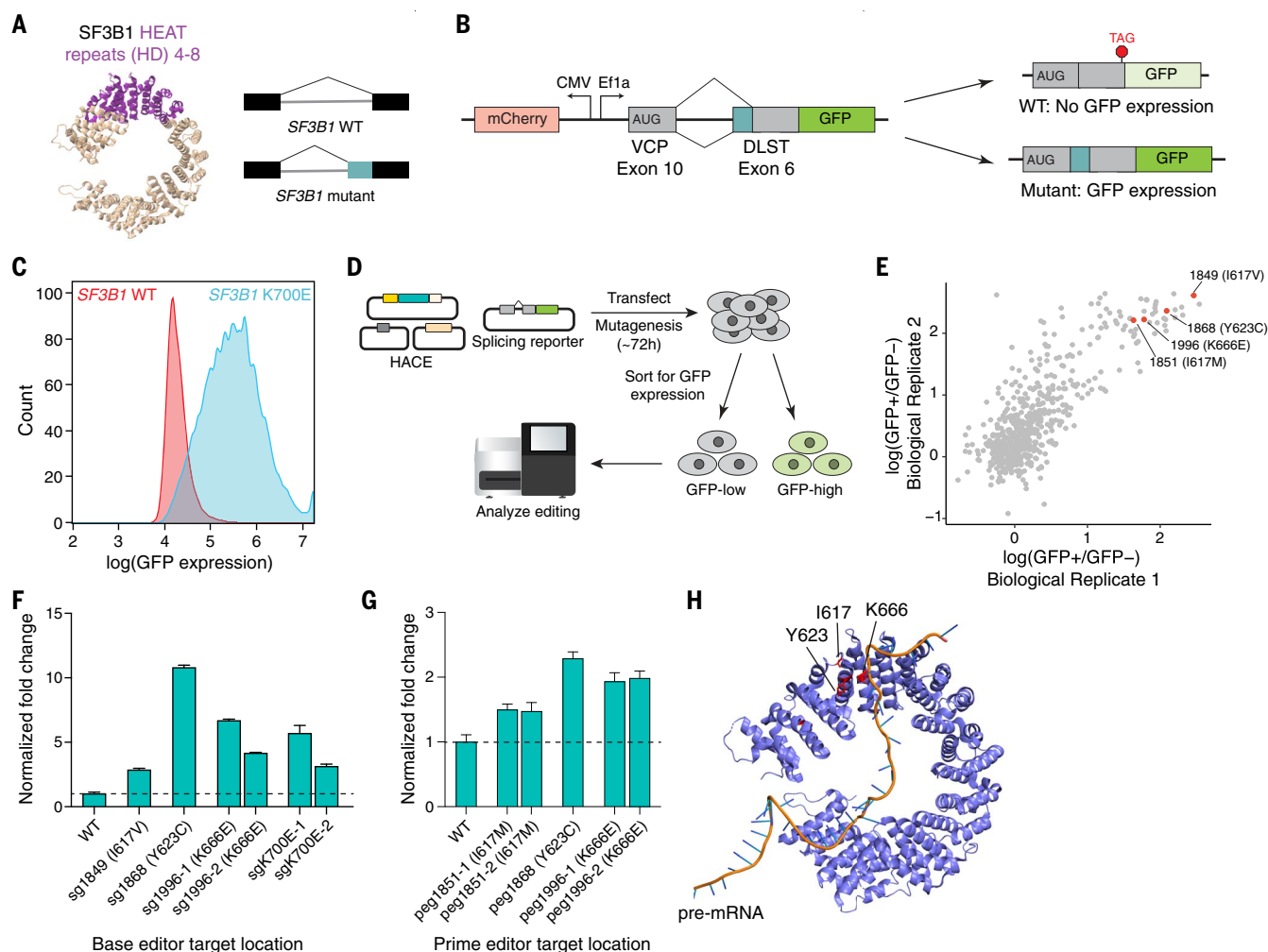


Fig. 4. Identification of variants in SF3B1 that result in alternative 3 branch point usage using HACE. (A) Structure of SF3B1 (right). HEAT repeats (HD) 4 to 8 are highlighted in purple (PDB ID 6EN4). Differential splicing patterns can result from mutations in SF3B1 (left). (B) Schematic of the SF3B1 mutant-dependent splicing reporter construct. The plasmid reporter consists of a constitutively expressed mCherry, and a minigene splicing GFP reporter (VCP exon 10 fused with DLST exon 6 with a downstream GFP). Correct splicing will not generate GFP expression, whereas SF3B1 mutant-dependent altered splicing will lead to GFP expression. (C) Histogram of GFP signal measured by flow cytometry between isogenic K562 SF3B1^{WT} and SF3B1^{K700E} cells. Cells were gated for mCherry expression. (D) Schematic of SF3B1 mutagenesis screen using HACE. HACE components and splicing reporter plasmids were cotransfected in HEK293FT cells. Mutagenesis was allowed to occur for 72 hours, and then cells were sorted for

GFP expression. The editing rate for each sorted group was assessed after genomic DNA extraction and amplicon sequencing. (E) Fold enrichment of individual bases in the SF3B1 cDNA sequence after selection across two biological replicates. Validated mutations are highlighted in red. (F) Normalized reporter activity fold change in mutations installed by base editing (mean \pm SD, $n = 3$). Significance is determined by an unpaired two-tailed t test between control and edited samples. *** $P < 0.001$ for all comparisons. The sgRNA sequences and their target bases are listed in table S11. (G) Normalized reporter activity fold change in mutations installed by prime editing (mean \pm SD, $n = 3$). Significance is determined by an unpaired two-tailed t test between control and edited samples. ** $P < 0.01$ for all comparisons. The epegRNA sequences and their target bases are listed in table S12. (H) Structure of SF3B1 (blue) in complex with pre-mRNA (orange). Validated mutations are shown in red. The structure was an overlay of PDB structures 6AHD and 5IFE.

Protein structure modeling revealed that the top-ranked mutations are all located at the edge of the HEAT repeat helices of the SF3B1 protein structure, which matches the pattern of hotspot mutations in pan-cancer analyses (Fig. 4H). These mutations are located near the pre-mRNA-binding region and have been shown to disrupt the tertiary structure of the SF3B1 protein (38). Taken together, HACE has helped us to identify clinically relevant mutations that result in SF3B1-dependent missplicing.

Mutagenesis of noncoding regulatory elements uncovers functional bases and variants

We next explored the potential of HACE to parse gene regulatory elements. Resolving the individual bases and variants that underlie the activity of enhancers remains challenging, as these elements must ideally be examined in their native chromatin contexts. Although CRISPR base editors have been used to characterize enhancers (39–41), such approaches require separate sgRNAs for each narrow tar-

get window and may be limited by the occurrence of protospacer adjacent motif (PAM) sites (~30% of locations are targetable with NGG PAM sites). Furthermore, artificial linkage between variants due to correlated mutations in an editing window limits the ability of base editors to resolve the functions of individual bases (39).

We targeted HACE to an enhancer region that regulates *CD69*, a membrane-bound lectin receptor gene that contributes to immune cell tissue residency (42–45). We designed three

sgRNAs targeting the core region of the *CD69* enhancer (table S1) and infected K562 cells with these nCas9-sgRNAs and HE (AID*Δ-PcrA M6-UGI) constructs. After 6 days, the cells were stimulated with phorbol 12-myristate 13-acetate (PMA) and ionomycin to induce *CD69* expression and then sorted on the basis of *CD69*

surface expression (methods). We assessed mutations by amplifying and sequencing the targeted region in *CD69*^{low} and *CD69*^{high} subsets (Fig. 5A and fig. S8A). The relative effect of each base edit was calculated according to its enrichment or depletion in *CD69*^{high} relative to *CD69*^{low} libraries (methods). Multiple indi-

vidual bases reduced *CD69* activation, with most of them located in motifs of immune-related transcription factors (Fig. 5B and table S10). The base enrichment pattern was highly consistent across biological replicates (Pearson's $\rho = 0.845$), confirming the robustness of our screen (Fig. 5C).

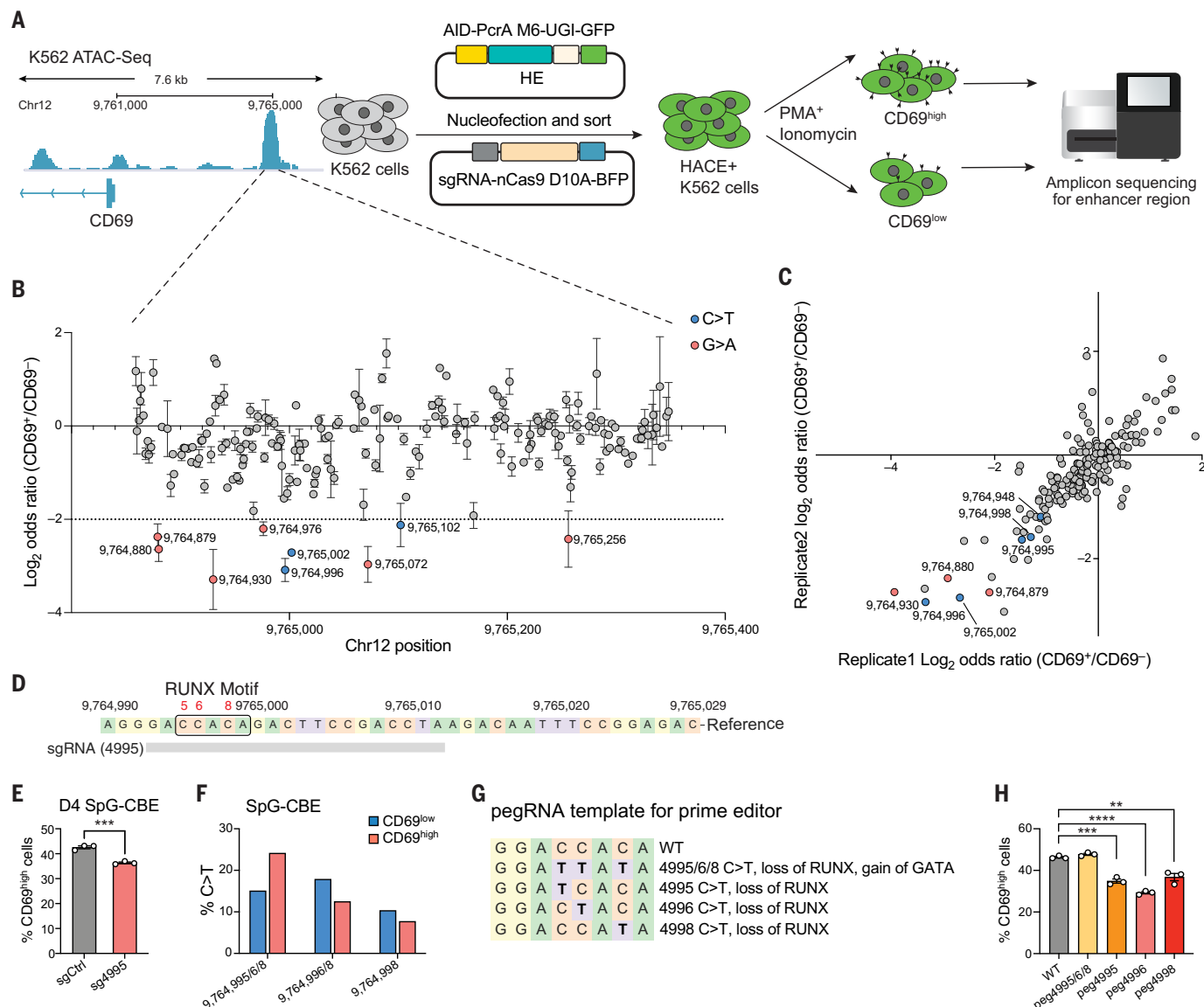


Fig. 5. Single-base tuning of cis-regulatory elements by HACE identifies transcriptional regulation of *CD69* by RUNX1/2. (A) Schematic of experimental workflow. The *CD69* enhancer region in K562 cells was identified using assay for transposase-accessible chromatin with sequencing (ATAC-seq) data and targeted by HACE sgRNAs. HACE+ K562 cells were diversified for 6 days, then stimulated with PMA and ionomycin to induce *CD69* expression. Cells are sorted into *CD69*^{high} and *CD69*^{low} populations, and the editing rate was profiled using amplicon sequencing. (B) Per-base enrichment of C>T or G>A edits in *CD69*^{high} cells relative to *CD69*^{low} cells (methods). The top most enriched C>T (blue) and G>A (red) variants in the *CD69*^{low} population are annotated. Each data point represents mean \pm SEM (*n* = 2). (C) Fold enrichment of individual bases in the *CD69* enhancer region across two biological replicates. Validated bases are highlighted in blue (C>T) or red (G>A). (D) Sequence of chr12:9764990–

9765029. RUNX motif is boxed. A sgRNA (sg4995) with NG PAM was used to target multiple cytosines in the RUNX motif. (E) Bar plot depicts the proportion of *CD69*^{high} cells in SpG-CBE-sgCtrl (gray) and SpG-CBE-sg4995 (red) after stimulation on day 4 after transfection. Significance is determined by unpaired two-tailed *t* test between groups (****P* < 0.001). Data are from three independent experiments each with three or four experimental replicates, mean \pm SEM. (F) Frequency of different incurred base edit combinations in sg4995-transfected K562 cells in *CD69*^{low} and *CD69*^{high} populations. (G) The pegRNA templates for single-base dissection around chr12:9764992–9764999. The hypothesized changes in phenotype are annotated. (H) The proportion of *CD69*^{high} poststimulation for cells edited with different pegRNAs on day 4 after transfection. Significance is determined by unpaired two-tailed *t* test between WT and edited groups. ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001.

We validated the HACE screen by evaluating mutations that were located in three immune-related transcription factor motifs using base editing. We infected K562 cells with a base editor construct and sgRNAs designed to incur the corresponding edits, activated the cells, and analyzed CD69 expression by flow cytometry. We first confirmed that a C>T transition at base Chr12:9764948 (abbreviated as the last four digits, “4948”; same nomenclature used below) suppressed CD69 induction (fig. S8B). This artificial variant, which we had identified in a prior base editor screen, disrupts a GATA transcription factor motif (39). We also validated G>A transitions at bases 4879 and 4880 that suppress CD69 induction (fig. S8C). These variants lie within a predicted IRF/ETS or IRF/STAT transcription factor motif (“GAAAGGAA”), suggesting a role for the motif and cognate factors in CD69 induction.

A particularly interesting outcome of the HACE screen was related to three adjacent C>T transitions (positions 4995, 4996, and 4998). Motif analysis suggests these variants are located within the core motif region (“CCACA”) recognized by RUNX family transcription factors, which have been implicated in CD69 regulation (Fig. 5D) (46, 47). Indeed, CD69 expression was increased by RUNX1 or RUNX2 overexpression (fig. S9A) and reduced by RUNX1 or RUNX2 knockdown in K562s (fig. S9B). These results support a role for a RUNX1/2 circuit in driving CD69 induction in K562 cells.

To validate the HACE data, we designed a sgRNA targeting the 4995 to 4998 region and introduced C>T mutations using an NG PAM cytidine base editor (Fig. 5D). We confirmed that the base edits reduced CD69 expression 4 days after editing (Fig. 5E and fig. S9C). To evaluate the impact of different combinations of variants in this window, we sorted CD69^{low} and CD69^{high} populations and performed targeted amplicon sequencing. We found that edited alleles with a single mutation at position 4998 or paired mutations at positions 4996 and 4998 were enriched in the CD69^{low} population, consistent with requirement of the RUNX motif for CD69 induction. However, edited alleles with C>T mutations at all three positions (4995, 4996, and 4998) were enriched in the CD69^{high} population (Fig. 5F and fig. S9, D and E). A likely explanation for this discordance is that the concurrent triple mutation not only disrupts the RUNX motif but also creates a GATA motif (“GATT”) at positions 4993 to 4996. GATA factors, including GATA1 and GATA2, are highly expressed in K562 cells and associated with transcriptional induction. These data suggest that de novo motif creation and GATA recruitment underlie the increased CD69 induction associated with the triple-mutant allele (Fig. 5G).

Our findings support the potential of combinatorial base editing to create gain-of-

function regulatory sequences but also highlight complications related to artificial linkage between adjacent variants. To definitively address this limitation, we used prime editing to introduce either individual or triple mutations at this locus (table S12). We again found that the respective single mutations reduced CD69 induction, thereby confirming the HACE discovery (Fig. 5H and fig. S9F). However, amplicon sequencing of sorted cells again confirmed that the CD69^{high} population is enriched for the triple-mutant allele but depleted for the respective single mutants (fig. S9, G and H). Hence, the application of HACE, followed by base and prime editing validation, revealed single nucleotide variants and combinatorial mutations that differentially affect enhancer activity by ablating or creating transcription factor motifs. Our data demonstrate the potential of HACE mutagenesis to systematically parse noncoding regulatory sequences at base resolution.

Discussion

We have developed a technology that enables programmable (sequence-directed) hypermutation of endogenous genomic loci in mammalian cells. HACE leverages CRISPR-Cas technology for targeted and multiplexable mutagenesis of the mammalian genome. It is distinctive both in its ability to distribute edits across large (>1000-bp) genomic regions and in its potential to evolve alleles over multiple cellular generations. We demonstrate HACE in forward genetic screens to identify coding mutations that confer resistance to kinase inhibitors, to identify and validate clinically relevant mutations of the SF3B1 splicing factor that induce altered 3'ss selection, and to uncover artificial variants that alter the activity of an immune enhancer.

HACE provides key advantages over existing methods for mammalian mutagenesis (table S13). First, it solves two long-standing limitations faced by conventional base editing screens: the requirement of an NGG PAM adjacent to the sgRNA recognition sequence and the occurrence of bystander mutations that can create artificial linkages and confound screening results. Instead of necessitating multiple sgRNAs to tile a target locus, the extended range of HACE confers mutations across a large genomic region using a single sgRNA, decreasing complexity and cost for pooled screens across multiple genomic targets. Second, the long editing range can potentially uncover combinatorial effects and interactions between multiple distant mutations across a locus—a capability not afforded by existing base editor methods. Although we cannot control the exact mutations introduced by HACE, the average editing rate of one to five random mutations per kilobase reduces the probability of mutation linkage compared with base edit-

ing, where the mutations primarily occur within a narrow 4- to 8-nucleotide target window (48). HACE is also amenable to continuous mutagenesis over multiple cell generations. HACE complements existing mammalian directed evolution strategies (49, 50) with an entirely different platform with the potential to chart fitness landscape changes, recapitulate evolutionary trajectories of biological phenotypes, or identify multistep synergistic mutations that confer de novo protein or gene regulatory functions within the endogenous mammalian genome.

The existing HACE system is also subject to other limitations that should be addressed through further engineering. One goal would be to increase mutation rates and range by optimizing modular components of the HACE system. For example, a wide diversity of helicases can be explored for improved efficiency, processivity, and fidelity as well as faster kinetics. Such efforts could leverage insights from the rational engineering of helicases for other applications (13, 51). Although HACE has a tunable editing rate and range, the substitution efficiency is also dependent on the targeted genomic locus. This might be a result of differences in Cas9 binding, helicase loading efficiency, target sequence, or chromatin context. Whereas we predominantly used nCas9 D10A variant for its consistent high levels of editing, further engineering with Cas9 variants to increase the editing rate can lead to superior HACE constructs that would minimize indel generation. Further analysis of HACE across genomic contexts may uncover new insights on controlling editing rate and range to achieve tunable mutagenesis. This would potentially allow for helicases with adjustable mutation ranges, which could restrict mutations to relevant genomic windows (e.g., enhancers). Additionally, HACE is currently limited to transition mutation modes owing to the use of cytidine and adenosine deaminases but might be expanded by incorporating emerging base-editing enzymes (52–54). Finally, long-range sequencing technologies could accelerate investigations of sequence coevolution in HACE mutagenized regions and thus provide new insights regarding concurrent or epistatic mutations.

Overall, HACE is a powerful tool for continuous, long-range, programmable diversification of endogenous mammalian genomes. We envision that HACE will substantially expand the functional genomics toolbox and unlock new molecular-scale insights toward building sequence-function maps of both coding and noncoding genomes.

Materials and methods

Plasmids and oligonucleotides

The guide sequences used for HACE mutagenesis (table S1) are cloned by Gibson assembly

or Golden Gate assembly. The oligos used in this study for sequencing (table S2) were purchased from Integrated DNA technologies (IDT) or Azenta/GENEWIZ. The Cas9 nickase plasmids were derived from plasmids pSpCas9 (BB)-2A-GFP (Addgene 48138) and pCMV-PEmax-P2A-GFP (Addgene 180020). The dCas9 plasmid was generated by site-directed mutagenesis. Plasmids expressing sgRNAs and prime editing guide RNAs (pegRNAs) were cloned by Gibson assembly or Golden Gate assembly. HACE editor plasmids were cloned by Gibson assembly of polymerase chain reaction (PCR) products. Individual helicases and deaminases were either subcloned from existing plasmids or synthesized by Integrated DNA Technologies after mammalian codon optimization. pEGFP-BLM was a gift from C. K.-L. Chan (Addgene plasmid #110299; <http://n2t.net/addgene:110299>; RRID:Addgene_110299). pCMV-Tag1-NS3 was a gift from X. Wang (Addgene plasmid #17645; <http://n2t.net/addgene:17645>; RRID:Addgene_17645). pET22B_SA_PcrA was a gift from M. Webb (Addgene plasmid #102999; <http://n2t.net/addgene:102999>; RRID:Addgene_102999). pSpCas9n(BB)-2A-GFP (PX461) was a gift from F. Zhang (Addgene plasmid #48140; <http://n2t.net/addgene:48140>; RRID:Addgene_48140). SpCas9 TadDE was a gift from D. Liu (Addgene plasmid #193837; <http://n2t.net/addgene:193837>; RRID:Addgene_193837). The helicases tested are summarized in table S3, and sequences of individual helicases tested are listed in table S4. All new plasmids generated during this study will be deposited on Addgene.

Mammalian cell culture

HEK293FT cells (Thermo Fisher - R70007) and A375 cells (ATCC, CRL-1619) were cultured in Dulbecco's Modified Eagle Medium with GlutaMAX (Thermo Fisher Scientific 10564011) supplemented with 10% (v/v) fetal bovine serum (FBS, Sigma-Aldrich F4135) and 1x penicillin-streptomycin (Thermo Fisher Scientific 15140122). Adherent cells were maintained at confluency below 80 to 90% at 37°C and 5% CO₂.

K562 cells (ATCC, CCL-243) were cultured in RPMI 1640 medium with GlutaMax (Thermo Fisher-61870036) supplemented with 10% (v/v) FBS and 1x penicillin-streptomycin. Suspended cells were maintained at confluency below 1.5 × 10⁶ cells/ml at 37°C and 5% CO₂. For stimulation experiments, 50 ng/ml PMA (Sigma-Aldrich, P8139) and 500 ng/ml ionomycin calcium salt from *Streptomyces conglobatus* (ionomycin, Sigma-Aldrich, I0634) are added to the cell culture media to simulate the cells for 2 to 3 hours.

Transfection of HACE plasmid and genomic DNA preparation for HACE editing assays

The day before transfection, 10,000 HEK293FT cells were seeded per well on 96-well plates

(Corning). Then, 16 to 24 hours after seeding, cells were transfected at ~70% confluency with 0.3 µl of TransIT-LT1 (Mirus Bio) according to the manufacturer's specifications. Each well was transfected with 40 ng of HACE editor plasmid and 40 ng of Cas9 variant plasmid, and 16 ng of sgRNA plasmid was delivered to each well unless otherwise specified. For control conditions, HACE editor plasmid and/or Cas9 variant plasmid were substituted with the same amount of pUC19 plasmid. Cells were cultured for 3 days after transfection. DNA was collected from transfected cells by removal of medium, resuspension in 50 µl of QuickExtract (Lucigen), and incubation at 65°C for 15 min, 68°C for 15 min, and 98°C for 10 min. After thermocycling, lysate was used directly in downstream PCR reactions as per manufacturer protocol.

High-throughput DNA sequencing of genomic DNA samples

For short amplicon (~300 bp) sequencing for HEK293FT and A375 cells, the target region was amplified from genomic DNA samples using Phusion U Hot Start PCR master mix (Thermo Fisher Scientific, F562) in a 20-µl reaction. The following program was used: 98°C for 30 s; 28 cycles of 98°C for 10 s, 65°C for 30 s, 72°C for 30 s; 72°C for 2 min, then 4°C thereafter. Barcodes and adapters for Illumina sequencing were added in a subsequent PCR amplification using Q5 High-Fidelity Hot-Start Polymerase Master Mix (2x, New England Biolabs). Amplicons were pooled and prepared for sequencing on a NextSeq (Illumina) with paired-end reads (read 1, 160 bp; index 1, 8 bp; index 2, 8 bp; read 2, 160 bp). Reads were demultiplexed and analyzed with appropriate pipelines.

For long amplicon (~2000 bp) sequencing, the targeted region was amplified using Phusion U Hot Start PCR master mix in a 20-µl reaction. The following program was used: 98°C for 30 s; 28 cycles of 98°C for 10 s, 65°C for 30 s, 72°C for 2 min; 72°C for 5 min, then 4°C thereafter. PCR products were purified using Magnetic Ampure XP beads (Beckman Coulter) using a 1:1 bead solution:DNA solution ratio to select the PCR fragments. Purified PCR products were eluted in 20 µl of water. The concentration of each sample was measured by Qubit (Thermo Fisher Scientific). The sequencing library was prepared following the Nextera XT Kit protocol (Illumina) using 1 ng of purified amplicon DNA per sample as starting material and half of the recommended amount of each kit reagent. Sequencing was performed on a NextSeq (Illumina) with paired-end reads (read 1, 100 bp; index 1, 8 bp; index 2, 8 bp; read 2, 100 bp).

Quantification of editing rate

Raw fastq reads obtained from sequencing were quality trimmed using BBduk (55) (BBMap v38.93) with the options “qtrim=r1 trimq=28

maq=25.” Next, all bases with quality scores of <28 were masked to N using seqtk v1.3 (56). The filtered reads were aligned to the reference sequence using Bowtie2 (57) (version 2.3.4.3). The pileup at each base was calculated using a custom Python script.

To calculate the substitution rate, we filtered for base positions with a sequencing coverage of at least 10,000. Bases that had a >5% substitution rate in the control condition were masked because this either indicated that it was a variant or an artifact from sequence alignment. The average G>A editing rate was calculated by extracting all positions where “G” was the reference base, then taking the average of the per-base G>A editing rate. The editing rate for other base transition and transversion modes was calculated similarly. To calculate the number of mutations per contiguous read, the aligned reads were converted to a tab-delimited format using Sam2Tsv (<https://lindenb.github.io/jvarkit/Sam2Tsv.html>). Subsequently, the number of substitutions per read was tabulated and normalized by the total number of reads per sample.

To calculate the local editing rate, the alignment was centered such that the nick site is centered at base position 0. For every base position, the local G>A editing rate was calculated by extracting all the “G” bases within a 100-bp window (50 bp upstream and 50 bp downstream) and then taking the average of all per-base G>A editing rates. Code for quantifying the edit rate is available at <https://github.com/chen-dawn/hace>. Indel rates were quantified using CRISPResso2 (58).

Long-term monitoring of mutation accumulation

A tetracycline-inducible promoter-controlled HE variant (AID-PcrA M6-UGI) was stably integrated into the HEK293FT cell genome. The sgRNA-nCas9 plasmid were transfected into the cells and doxycycline (Dox, an analog of tetracycline) was added into the culture medium (10 ng/ml) to activate the HE expression. The expression of the HE [blue fluorescent protein (BFP)] and sgRNA-nCas9 (GFP) were visually monitored daily. A second transfection of the sgRNA-nCas9 plasmid was performed on day 5. Genomic DNA (gDNA) samples from day 0 (before diversification) and days 2, 4, 6, 8, and 10 were collected where the target loci was sequenced and analyzed as described above.

Cell viability measurements

Cell viability was measured using the CellTiter-Glo Luminescent Cell Viability Assay (Promega) following the manufacturer's protocols. Briefly, HEK293FT cells were seeded at a density of 10,000 cells per 100 µl per well in a 96-well plate in biological triplicates. The next day, cells were transfected with respective HACE plasmids according to the above protocol. Cell viability was measured 72 hours after transfection.

Luminescence readings were performed using a SpectraMax M5 (Molecular Devices) plate reader.

Whole-exome sequencing and off-target analysis

The day before transfection, 50,000 HEK293FT cells were seeded in a 24-well plate. The next day, individual HE constructs were transfected together with a sgRNA targeting the *MAP2K1* locus. Genomic DNA was extracted from cells 3 days after transfection using the Zymo Quick-DNA Miniprep Kit (Cat D3024). Amplicon sequencing was performed at the *MAP2K1* loci to confirm that there is HACE-dependent editing at the target loci in each condition. The whole-genome DNA sequencing library was prepared using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (Cat E7805S). Exome sequences were enriched using the xGen Exome Hybridization Panel (IDT 10005152) following the manufacturer's protocols. Exome libraries were sequenced on a NovaSeq X (Illumina) with paired-end reads (read 1, 150 bp; index 1, 8 bp; index 2, 8 bp; read 2, 150 bp) at a minimum of 100 million reads per sample.

The sequencing output was demultiplexed using bcl2fastq, and the paired-end reads were aligned to the reference genome hg38 using HISAT2 v2.2.1 (59). Aligned reads from each replicate were subsampled using reformat.sh (BBMap v38.93), and 100 million aggregated reads per replicate for each condition were used for further analysis. The HEK293FT-specific single-nucleotide polymorphisms (SNPs) were determined following the GATK4 variant calling workflow for germline short variant discovery (<https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels>) on WT HEK293FT exome libraries (>50× coverage). In brief, the aligned reads were deduplicated using Picard v2.27.5. HaplotypeCaller (GATK4) was used for calling variants, and known variants in dbSNP version 138 were used during base-quality recalibration. The chromosomal coordinates where SNPs were detected were excluded from subsequent analysis.

To quantify the per-base editing rate of the exome, the base pileup at each base was calculated using samtools mpileup (v1.15.1), followed by post processing using mpileup2readcounts (<https://github.com/IARCBioinfo/mpileup2readcounts>). Bases with <50 total read depth were excluded from subsequent analysis. The genome was binned into 100-kb bins using bcftools v1.15.1. The off-target C>T editing rates for each genomic bin were obtained using a custom R script by counting the number of C and T bases in each bin. Fisher's exact test was used to quantify significant changes in editing for each bin relative to cells transfected with only nCas9, using the false discovery rate correction to adjust for multiple hypothesis test-

ing. Significant off-target sites are listed in table S5.

MEK1 inhibitor-resistance screen

A375 cells were diversified for 3 days by transfection of HE variant AID-PcrA M6-UGI, nCas9 D10A, and sgRNAs targeting exons 2, 3, and 6 of the *MEK1* gene using TransIT-2020 (Mirus Bio). Approximately 5 million cells in a 15-cm dish were placed under selection with either 100 nM selumetinib or 5 nM trametinib for 20 days. A portion of preselection cells were harvested as a control. Cells were passaged every 3 days to ensure that they were maintained at <70% confluency. After selection, cells were harvested, and genomic DNA was extracted using QuickExtract (Lucigen). The *MEK1* exons were amplified with exon-specific primers (table S2) using Phusion U Hot Start Master Mix. Concurrently, we also harvested RNA from selected cells using the Qiagen RNeasy Mini Plus Kit (Cat 74134). The cDNA was generated by reverse transcription using Maxima H Minus Reverse Transcriptase (Thermo Fisher). Sequencing libraries for cDNA were generated using the modified Nextera XT Kit protocol described in the "High-throughput DNA sequencing of genomic DNA samples" section above. All libraries were sequenced on a NextSeq (Illumina) with paired-end reads (read 1, 160 bp; index 1, 8 bp; index 2, 8 bp; read 2, 160 bp). The substitution rate (allele frequency) for each base of the *MEK1* sequence was calculated for both pre- and postselection samples. Significant mutations were identified by comparing the base counts between pre- and postselection samples using a Fisher's exact test (table S6). We compared the substitution rate between RNA and DNA samples and found that they had a high correlation.

SRE reporter assay

pEF1a-*MEK1* wild type, pEF1a-*MEK1*G128D, pEF1a-*MEK1*G202E, and pEF1a-*MEK1*E203K were generated using Gibson assembly. Sequences of *MEK1*-derived constructs are available in table S7. The SRE reporter assay was performed using the SRE reporter kit (BPS Biosciences) according to the manufacturer's protocols. In brief, ~10,000 HEK293FT cells in 100 µl of growth medium were seeded in 96-well white opaque assay plates. The cells were transfected with 60 ng of reporter plasmid and 40 ng of respective *MEK1* plasmids. The culture medium was replaced 6 hours after transfection with 50 µl of trametinib-containing medium with 0.5% FBS. After 12 hours, the cells were washed and incubated with 50 µl of 0.5% FBS-containing culture medium supplemented with recombinant human epidermal growth factor protein (Life Technologies) at a final concentration of 10 ng/ml. After 6 hours of incubation, the reporter activity was assayed using a dual luciferase (Firefly-Renilla) assay system (BPS Bioscience) according to the manufacturer's instructions using a SpectraMax M5 (Molecular Devices) plate reader. The ratio between Firefly luminescence and Renilla luminescence intensity was calculated for each well after background subtraction.

The minigene reporter to probe SF3B1 function was constructed by Gibson assembly of a synthetic minigene sequence (synthesized by Twist Biosciences) into a custom bicistronic mCherry/GFP reporter plasmid. To construct the minigene reporter, we fused the VCP exon 10 sequence and 150 bp of its immediate downstream intron with DLST exon 6 and 97 bp of its immediate upstream intron. We also appended an "ATG" start codon at the beginning of the sequence. The open reading frame was adjusted such that correct splicing in WT cells will result in premature termination before the GFP. In contrast, the alternative 3' splice site usage in *SF3B1* mutant cells will result in full-length GFP expression. The minigene reporter sequences are annotated in table S8.

Design of SF3B1 splicing minigene reporter

The minigene reporter to probe SF3B1 function was constructed by Gibson assembly of a synthetic minigene sequence (synthesized by Twist Biosciences) into a custom bicistronic mCherry/GFP reporter plasmid. To construct the minigene reporter, we fused the VCP exon 10 sequence and 150 bp of its immediate downstream intron with DLST exon 6 and 97 bp of its immediate upstream intron. We also appended an "ATG" start codon at the beginning of the sequence. The open reading frame was adjusted such that correct splicing in WT cells will result in premature termination before the GFP. In contrast, the alternative 3' splice site usage in *SF3B1* mutant cells will result in full-length GFP expression. The minigene reporter sequences are annotated in table S8.

SF3B1 missplicing screen

HEK293FT cells were diversified for 3 days by cotransfection of HE variants AID-PcrA M6-UGI and TadA-PcrA M6-UGI, nCas9 D10A, sgRNAs targeting *SF3B1* exons 13 to 17, and splicing minigene reporter. Cells transfected with only minigene reporter were used as undiversified control. The experiment was performed in triplicate, with ~10 million cells transfected per replicate. After diversification, cells were prepared for flow sorting by washing and resuspending in 1× phosphate-buffered saline (PBS) with 2% bovine serum albumin. Cells were sorted using a SONY MA900 sorter, where mCherry-positive cells were sorted into a GFP⁻ and GFP⁺ bin. At least 1 million cells were collected for each cell population. After flow sorting, the RNA of the cells was extracted using the Qiagen RNeasy Mini Plus Kit. The cDNA was generated by reverse transcription using Maxima H Minus Reverse Transcriptase. Sequencing libraries for cDNA were generated using the modified Nextera XT Kit protocol and sequenced on a NextSeq. Fold enrichment was calculated by dividing the substitution rate in GFP⁺ by that of GFP⁻ samples. The significant mutations were identified using a Fisher's exact test and are shown in table S9. The clinical mutations that are observed in *SF3B1* were retrieved from COSMIC (34). A mutation was considered high frequency if there were at least three observations in the dataset.

CD69 enhancer tiling for functional bases

K562 cells were nucleofected with 2.5 µg of HE and 2.5 µg of nCas9 and sgRNA plasmids

using the SF Cell Line 4D-Nucleofector X Kit L (Lonza V4XC-2024), following the manufacturer's protocol. Each plasmid contained a fluorescent protein reporter (sgRNA mCherry, nCas9 GFP, HE BFP). Approximately 1.5×10^6 to 2×10^6 cells were used per nucleofection reaction. After 24 hours, cells were sorted using either SONY SH800 or BD Aria flow cytometry sorter to isolate cells expressing all plasmid components.

On day 7 after nucleofection, cells were stimulated with PMA/ionomycin for 2 to 3 hours. The cells were stained with the antibody cocktail in the staining buffer of a 1:1 mix of PBS and Brilliant Stain Buffer (BD 566349) at room temperature for 20 min or at 4°C for 30 min. The following antibodies and dyes from BioLegend were used: Brilliant Violet 510 anti-human CD69 Antibody (310936), APC anti-human CD69 Antibody (310910), and Zombie NIR Fixable Viability Kit (423106). Cells were washed once in PBS with 1% FBS and then resuspended in the same buffer to prepare for flow sorting. Subsequently, the top 40% of cells showing high CD69 expression (CD69^{high}) and the bottom 20% with low CD69 expression (CD69^{low}) were sorted using the SONY SH800 flow cytometer. A minimum of 100,000 cells were collected per tube. Genomic DNA was then isolated from these cells either by using the QIAGEN DNA micro isolation kit (Cat #56304) or by lysis buffer [0.5% Triton X-100, 0.1 AU/ml QIAGEN Protease (Cat 19157) in H₂O]. The lysis process involved incubation at 56°C for 20 min and at 72°C for 20 min at 600 rpm on a thermo shaker. Amplicon PCR for the genomic DNA was processed using the KAPA HiFi HotStart ReadyMix PCR Kit (Roche, KR0370). The following program was used: 95°C for 5 min; 30 cycles of 95°C for 30 s, 60°C for 30 s, 72°C for 30 s; 72°C for 5 min; 4°C forever. The amplicon libraries were sequenced on a NextSeq.

To identify enriched bases, the %C>T or %G>A of each group were first calculated for both CD69^{high} and CD69^{low} groups (%high or %low, respectively). Then the log₂ odds ratio of CD69^{high} versus CD69^{low} was calculated as $\log_2 \text{OR} = \log_2 \{ [\% \text{high} / (1 - \% \text{high})] / [\% \text{low} / (1 - \% \text{low})] \}$. The correlation of technical replicates was plotted using GraphPad Prism 10.0. The top hits are recorded in table S10.

Base editing validation

The following base editors are used in this study: pRDA_478 (Addgene 179096), pRDA_479 (Addgene 179099), pCAG-CBE4max-SpG-P2A-EGFP (Addgene: RTW4552/139998), pCAG-CBE4max-SpRY-P2A-EGFP (Addgene: RTW5133/139999), pCMV-T7-ABE8.20m-nSpCas9-NG-P2A-EGFP (Addgene: KAC1164/185919), and pCMV-T7-ABE8.20m-nSpRY-P2A-EGFP (Addgene: KAC1335/185917). The validation sgRNAs are listed in table S11. The sgRNA sequences were cloned into pCMV-BFP-U6-sgRNA (Addgene: 196725, gift from B. Schmierer) or directly into

pRDA_478 or pRDA_479. Mutations for sgRNAs and bystander editing rates were quantified using CRISPResso2 (58).

To validate *MEK1* variants, 67 ng of cytidine base editor and 33 ng sgRNA plasmids were transfected into A375 cells per well in a 96-well format. A nontargeting sgRNA was used as a control. After diversification for 3 days, the cells were selected with either 100 nM selumetinib or 5 nM trametinib for 14 days. The pre- and postselection substitution rates were analyzed by amplicon sequencing. All experiments were conducted in triplicates.

For validation of *SF3B1* variants, HEK293FT cells in a 96-well format were transfected with 67 ng of the base editor-sgRNA plasmid and 33 ng of the minigene splicing reporter per well. A nontargeting sgRNA was used as a control. Cells were diversified for 3 days, then the GFP:mCherry ratio in each well was quantified by confocal microscopy using a custom cell segmentation and quantification pipeline. Briefly, individual cells were segmented by watershed segmentation using the mCherry channel. For each segmented cell, the total pixel area and mean intensity of the pixels were computed for GFP (488 nm) and mCherry (561 nm) channels to obtain a “pseudo-flow cytometry” dataset. The fluorescence background for each channel was subtracted from all conditions in that channel, and aggregated values for each condition were divided by area to obtain average fluorescence intensity. Standard deviation was computed by comparing average values in three technical transfection replicates. Editing at each sgRNA was quantified by amplicon sequencing of genomic DNA samples from each well. All experiments were conducted in triplicates.

For validation of *CD69* variants, 2 µg of the base editor plasmid and 2 µg of the sgRNA plasmid were nucleofected into 1.5×10^6 K562 cells using the SF Cell Line 4D-Nucleofector X Kit L according to the manufacturer's protocol. At 24 hours after nucleofection, the cells coexpressing base editor and sgRNA were sorted on the basis of reporter expression. On day 4 after nucleofection, cells were stimulated with PMA/ionomycin for 2 to 3 hours, and the top 40% of CD69 high expression cells and bottom 20% of CD69 low expression cells were sorted using a Sony SH800 flow cytometer, collecting at least 10,000 cells per tube. Genomic DNA was isolated from the sorted cells using the QIAGEN DNA Micro Kit (Cat# 56304) and prepared for amplicon sequencing using the protocol described in the “High-throughput DNA sequencing of genomic DNA samples” section. The substitution rate at each locus was quantified using CRISPResso2.

Prime editing validation

The following prime editor plasmids are used in this study: pCMV-PEmax-P2A-hMLH1dn

(Addgene: 174828), pCMV-PEmax-P2A-GFP (Addgene: 180020), and pEF1a-hMLH1dn (Addgene: 174824). Desired pegRNA and nickase sgRNA sequences were designed using PrimeDesign (60). The engineered pegRNA (epegRNA) overhang was designed using pegLIT (67). Sequences of pegRNAs are shown in table S12.

For prime editing validation of *SF3B1* variants, HEK293FT cells in a 96-well format were transfected with 150 ng of PEmax, 50 ng of epegRNA, 25 ng of nicking sgRNA, and 50 ng of minigene splicing reporter using 0.5 µl of TransIT-LT1 per well. Cells were diversified for 3 days, then the GFP:mCherry ratio in each well was quantified by confocal microscopy as described above. Editing at each sgRNA was quantified by amplicon sequencing of genomic DNA samples from each well using CRISPResso2.

For prime editing validations for *CD69* enhancer variants in K562 cells, 2 µg of the prime editor plasmid, 1 µg of hMLH1dn plasmid, 1 µg of epegRNA plasmid, and 0.5 µg of nickase sgRNA plasmid were nucleofected in 1.5×10^6 cells using SF Cell Line 4D-Nucleofector X Kit L according to the manufacturer's protocols. After 24 hours, the cells that were positive for both prime editor and epegRNA were sorted according to the GFP and mCherry reporters and cultured in regular complete RPMI media. A second round of nucleofection and sorting was performed 4 days after transfection to increase prime editing efficiency. On day 5 after the second nucleofection, CD69 expression levels were quantified by flow cytometry. Genomic DNA was harvested from CD69^{high} (top 40%) and CD69^{low} (bottom 20%) cells, and the editing efficiency was quantified by performing amplicon sequencing using the protocols described above. The substitution rate at each locus was quantified using CRISPResso2.

REFERENCES AND NOTES

1. D. M. Fowler, S. Fields, Deep mutational scanning: A new style of protein science. *Nat. Methods* **11**, 801–807 (2014). doi: [10.1038/nmeth.3027](https://doi.org/10.1038/nmeth.3027); pmid: [25075907](https://pubmed.ncbi.nlm.nih.gov/25075907/)
2. J. G. English et al., VEGAS as a platform for facile directed evolution in mammalian cells. *Cell* **178**, 748–761.e17 (2019). doi: [10.1016/j.cell.2019.05.051](https://doi.org/10.1016/j.cell.2019.05.051); pmid: [31280962](https://pubmed.ncbi.nlm.nih.gov/31280962/)
3. H. Chen et al., Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. *Nat. Biotechnol.* **38**, 165–168 (2020). doi: [10.1038/s41587-019-0331-8](https://doi.org/10.1038/s41587-019-0331-8); pmid: [31844291](https://pubmed.ncbi.nlm.nih.gov/31844291/)
4. A. Cravens, O. K. Jamil, D. Kong, J. T. Sockolovsky, C. D. Smolke, Polymerase-guided base editing enables in vivo mutagenesis and rapid protein engineering. *Nat. Commun.* **12**, 1579 (2021). doi: [10.1038/s41467-021-21876-z](https://doi.org/10.1038/s41467-021-21876-z); pmid: [33707425](https://pubmed.ncbi.nlm.nih.gov/33707425/)
5. A. Acevedo-Arozena et al., ENU mutagenesis, a way forward to understand gene function. *Annu. Rev. Genomics Hum. Genet.* **9**, 49–69 (2008). doi: [10.1146/annurev.genom.9.081307.164224](https://doi.org/10.1146/annurev.genom.9.081307.164224); pmid: [18949851](https://pubmed.ncbi.nlm.nih.gov/18949851/)
6. R. Cuella-Martin et al., Functional interrogation of DNA damage response variants with base editing screens. *Cell* **184**, 1081–1097.e19 (2021). doi: [10.1016/j.cell.2021.01.041](https://doi.org/10.1016/j.cell.2021.01.041); pmid: [33606978](https://pubmed.ncbi.nlm.nih.gov/33606978/)
7. R. E. Hanna et al., Massively parallel assessment of human variants with base editor screens. *Cell* **184**, 1064–1080.e20 (2021). doi: [10.1016/j.cell.2021.01.012](https://doi.org/10.1016/j.cell.2021.01.012); pmid: [33606977](https://pubmed.ncbi.nlm.nih.gov/33606977/)
8. G. T. Hess et al., Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016). doi: [10.1038/nmeth.4038](https://doi.org/10.1038/nmeth.4038); pmid: [27798611](https://pubmed.ncbi.nlm.nih.gov/27798611/)

9. Y. Ma *et al.*, Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016). doi: [10.1038/nmeth.4027](https://doi.org/10.1038/nmeth.4027); pmid: [27723754](https://pubmed.ncbi.nlm.nih.gov/27723754/)
10. N. Z. Lue, B. B. Liaw, Base editor screens for in situ mutational scanning at scale. *Mol. Cell* **83**, 2167–2187 (2023). doi: [10.1016/j.molcel.2023.06.009](https://doi.org/10.1016/j.molcel.2023.06.009); pmid: [37390819](https://pubmed.ncbi.nlm.nih.gov/37390819/)
11. R. M. Brosh Jr., DNA helicases involved in DNA repair and their roles in cancer. *Nat. Rev. Cancer* **13**, 542–558 (2013). doi: [10.1038/nrc3560](https://doi.org/10.1038/nrc3560); pmid: [23842644](https://pubmed.ncbi.nlm.nih.gov/23842644/)
12. A. Zimmermann *et al.*, A Cas3-base editing tool for targetable in vivo mutagenesis. *Nat. Commun.* **14**, 3389 (2023). doi: [10.1038/s41467-023-39087-z](https://doi.org/10.1038/s41467-023-39087-z); pmid: [37296137](https://pubmed.ncbi.nlm.nih.gov/37296137/)
13. M. Gavrilov *et al.*, Engineered helicase replaces thermocycler in DNA amplification while retaining desired PCR characteristics. *Nat. Commun.* **13**, 6312 (2022). doi: [10.1038/s41467-022-34076-0](https://doi.org/10.1038/s41467-022-34076-0); pmid: [36274095](https://pubmed.ncbi.nlm.nih.gov/36274095/)
14. 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/)
15. M. S. Dillingham, P. Soultanas, P. Wiley, M. R. Webb, D. B. Wigley, Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8381–8387 (2001). doi: [10.1073/pnas.131009598](https://doi.org/10.1073/pnas.131009598); pmid: [11459979](https://pubmed.ncbi.nlm.nih.gov/11459979/)
16. J. Park *et al.*, PcrA helicase dismantles RecA filaments by reeling in DNA in uniform steps. *Cell* **142**, 544–555 (2010). doi: [10.1016/j.cell.2010.07.016](https://doi.org/10.1016/j.cell.2010.07.016); pmid: [20723756](https://pubmed.ncbi.nlm.nih.gov/20723756/)
17. T. A. Kunkel, Considering the cancer consequences of altered DNA polymerase function. *Cancer Cell* **3**, 105–110 (2003). doi: [10.1016/S1535-6108\(03\)00027-8](https://doi.org/10.1016/S1535-6108(03)00027-8); pmid: [12620405](https://pubmed.ncbi.nlm.nih.gov/12620405/)
18. L. W. Koblan *et al.*, Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018). doi: [10.1038/nbt.4172](https://doi.org/10.1038/nbt.4172); pmid: [29813047](https://pubmed.ncbi.nlm.nih.gov/29813047/)
19. M. F. Richter *et al.*, Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020). doi: [10.1038/s41587-020-0453-z](https://doi.org/10.1038/s41587-020-0453-z); pmid: [32433547](https://pubmed.ncbi.nlm.nih.gov/32433547/)
20. M. E. Neugebauer *et al.*, Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. *Nat. Biotechnol.* **41**, 673–685 (2023). doi: [10.1038/s41587-022-01533-6](https://doi.org/10.1038/s41587-022-01533-6); pmid: [36357719](https://pubmed.ncbi.nlm.nih.gov/36357719/)
21. M. Gu, C. M. Rice, Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 521–528 (2010). doi: [10.1073/pnas.0913380107](https://doi.org/10.1073/pnas.0913380107); pmid: [2080715](https://pubmed.ncbi.nlm.nih.gov/2080715/)
22. J. L. Kim *et al.*, Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: The crystal structure provides insights into the mode of unwinding. *Structure* **6**, 89–100 (1998). doi: [10.1016/S0969-2126\(98\)00010-0](https://doi.org/10.1016/S0969-2126(98)00010-0); pmid: [9493270](https://pubmed.ncbi.nlm.nih.gov/9493270/)
23. G. S. Martin, Cell signaling and cancer. *Cancer Cell* **4**, 167–174 (2003). doi: [10.1016/S1535-6108\(03\)00216-2](https://doi.org/10.1016/S1535-6108(03)00216-2); pmid: [14522250](https://pubmed.ncbi.nlm.nih.gov/14522250/)
24. C. M. Emery *et al.*, MEK1 mutations confer resistance to MEK and B-RAF inhibition. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20411–20416 (2009). doi: [10.1073/pnas.0905833106](https://doi.org/10.1073/pnas.0905833106); pmid: [19915144](https://pubmed.ncbi.nlm.nih.gov/19915144/)
25. S. I. Nikolaev *et al.*, Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nat. Genet.* **44**, 133–139 (2011). doi: [10.1038/ng.1026](https://doi.org/10.1038/ng.1026); pmid: [22197931](https://pubmed.ncbi.nlm.nih.gov/22197931/)
26. Y. Gao *et al.*, Allele-specific mechanisms of activation of MEK1 mutants determine their properties. *Cancer Discov.* **8**, 648–661 (2018). doi: [10.1158/2159-8290.CD-17-1452](https://doi.org/10.1158/2159-8290.CD-17-1452); pmid: [29483135](https://pubmed.ncbi.nlm.nih.gov/29483135/)
27. K. Yoshida *et al.*, Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011). doi: [10.1038/nature10496](https://doi.org/10.1038/nature10496); pmid: [21909114](https://pubmed.ncbi.nlm.nih.gov/21909114/)
28. E. Papaemmanuil *et al.*, Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.* **365**, 1384–1395 (2011). doi: [10.1056/NEJMoa1103283](https://doi.org/10.1056/NEJMoa1103283); pmid: [21995386](https://pubmed.ncbi.nlm.nih.gov/21995386/)
29. V. Quesada *et al.*, Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 47–52 (2011). doi: [10.1038/ng.1032](https://doi.org/10.1038/ng.1032); pmid: [22158541](https://pubmed.ncbi.nlm.nih.gov/22158541/)
30. K. Yoshida, S. Ogawa, Splicing factor mutations and cancer. *Wiley Interdiscip. Rev. RNA* **5**, 445–459 (2014). doi: [10.1002/wrna.1222](https://doi.org/10.1002/wrna.1222); pmid: [24523246](https://pubmed.ncbi.nlm.nih.gov/24523246/)
31. C. Yan, R. Wan, R. Bai, G. Huang, Y. Shi, Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* **353**, 904–911 (2016). doi: [10.1126/science.aag0291](https://doi.org/10.1126/science.aag0291); pmid: [27445306](https://pubmed.ncbi.nlm.nih.gov/27445306/)
32. R. B. Darman *et al.*, Cancer-associated SF3B1 hotspot mutations induce cryptic 3' splice site selection through use of a different branch point. *Cell Rep.* **13**, 1033–1045 (2015). doi: [10.1016/j.celrep.2015.09.053](https://doi.org/10.1016/j.celrep.2015.09.053); pmid: [26565915](https://pubmed.ncbi.nlm.nih.gov/26565915/)
33. M. Seiler *et al.*, Somatic mutational landscape of splicing factor genes and their functional consequences across 33 cancer types. *Cell Rep.* **23**, 282–296.e4 (2018). doi: [10.1016/j.celrep.2018.01.088](https://doi.org/10.1016/j.celrep.2018.01.088); pmid: [29617667](https://pubmed.ncbi.nlm.nih.gov/29617667/)
34. J. G. Tate *et al.*, COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019). doi: [10.1093/nar/gky1015](https://doi.org/10.1093/nar/gky1015); pmid: [30371878](https://pubmed.ncbi.nlm.nih.gov/30371878/)
35. D. Inoue *et al.*, Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature* **574**, 432–436 (2019). doi: [10.1038/s41586-019-1646-9](https://doi.org/10.1038/s41586-019-1646-9); pmid: [31597964](https://pubmed.ncbi.nlm.nih.gov/31597964/)
36. H. Makishima *et al.*, Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood* **119**, 3203–3210 (2012). doi: [10.1182/blood-2011-12-399774](https://doi.org/10.1182/blood-2011-12-399774); pmid: [22323480](https://pubmed.ncbi.nlm.nih.gov/22323480/)
37. K. North *et al.*, Synthetic introns enable splicing factor mutation-dependent targeting of cancer cells. *Nat. Biotechnol.* **40**, 1103–1113 (2022). doi: [10.1038/s41587-022-01224-2](https://doi.org/10.1038/s41587-022-01224-2); pmid: [35241838](https://pubmed.ncbi.nlm.nih.gov/35241838/)
38. C. Cretu *et al.*, Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol. Cell* **64**, 307–319 (2016). doi: [10.1016/j.molcel.2016.08.036](https://doi.org/10.1016/j.molcel.2016.08.036); pmid: [27206643](https://pubmed.ncbi.nlm.nih.gov/27206643/)
39. Z. Chen *et al.*, Integrative dissection of gene regulatory elements at base resolution. *Cell Genomics* **3**, 100318 (2023). doi: [10.1016/j.xgen.2023.100318](https://doi.org/10.1016/j.xgen.2023.100318); pmid: [37388913](https://pubmed.ncbi.nlm.nih.gov/37388913/)
40. J. D. Martin-Rufino *et al.*, Massively parallel base editing to map variant effects in human hematopoiesis. *Cell* **186**, 2456–2474.e24 (2023). doi: [10.1016/j.cell.2023.03.035](https://doi.org/10.1016/j.cell.2023.03.035); pmid: [37137305](https://pubmed.ncbi.nlm.nih.gov/37137305/)
41. J. A. Morris *et al.*, Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. *Science* **380**, eadh7699 (2023). doi: [10.1126/science.adh7699](https://doi.org/10.1126/science.adh7699); pmid: [37141313](https://pubmed.ncbi.nlm.nih.gov/37141313/)
42. D. Cibrián, F. Sánchez-Madrid, CD69: From activation marker to metabolic gatekeeper. *Eur. J. Immunol.* **47**, 946–953 (2017). doi: [10.1002/eji.201646837](https://doi.org/10.1002/eji.201646837); pmid: [28475283](https://pubmed.ncbi.nlm.nih.gov/28475283/)
43. D. Sancho, M. Gómez, F. Sánchez-Madrid, CD69 is an immunoregulatory molecule induced following activation. *Trends Immunol.* **26**, 136–140 (2005). doi: [10.1016/j.it.2004.12.006](https://doi.org/10.1016/j.it.2004.12.006); pmid: [15745855](https://pubmed.ncbi.nlm.nih.gov/15745855/)
44. T. Sathaliyawala *et al.*, Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets. *Immunity* **38**, 187–197 (2013). doi: [10.1016/j.jimmuni.2012.09.020](https://doi.org/10.1016/j.jimmuni.2012.09.020); pmid: [23260195](https://pubmed.ncbi.nlm.nih.gov/23260195/)
45. J. M. Schenkel, D. Masopust, Tissue-resident memory T cells. *Immunity* **41**, 886–897 (2014). doi: [10.1016/j.jimmuni.2014.12.007](https://doi.org/10.1016/j.jimmuni.2014.12.007); pmid: [25526304](https://pubmed.ncbi.nlm.nih.gov/25526304/)
46. T. Laguna *et al.*, New insights on the transcriptional regulation of CD69 gene through a potent enhancer located in the conserved non-coding sequence 2. *Mol. Immunol.* **66**, 171–179 (2015). doi: [10.1016/j.molimm.2015.02.031](https://doi.org/10.1016/j.molimm.2015.02.031); pmid: [25801305](https://pubmed.ncbi.nlm.nih.gov/25801305/)
47. S. Wahlen *et al.*, The transcription factor RUNX2 drives the generation of human NK cells and promotes tissue residency. *eLife* **11**, e80320 (2022). doi: [10.7554/eLife.80320](https://doi.org/10.7554/eLife.80320); pmid: [35793229](https://pubmed.ncbi.nlm.nih.gov/35793229/)
48. H. A. Rees, D. R. Liu, Base editing: Precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018). doi: [10.1038/s41576-018-0059-1](https://doi.org/10.1038/s41576-018-0059-1); pmid: [30323312](https://pubmed.ncbi.nlm.nih.gov/30323312/)
49. S. J. Hendel, M. D. Shoulders, Directed evolution in mammalian cells. *Nat. Methods* **18**, 346–357 (2021). doi: [10.1038/s41592-021-01090-x](https://doi.org/10.1038/s41592-021-01090-x); pmid: [33828274](https://pubmed.ncbi.nlm.nih.gov/33828274/)
50. R. S. Molina *et al.*, In vivo hypermutation and continuous evolution. *Nat. Rev. Methods Primers* **2**, 36 (2022). doi: [10.1038/s43586-022-00119-5](https://doi.org/10.1038/s43586-022-00119-5); pmid: [37073402](https://pubmed.ncbi.nlm.nih.gov/37073402/)
51. S. Arslan, R. Khafizov, C. D. Thomas, Y. R. Chemla, T. Ha, Engineering of a superhelicase through conformational control. *Science* **348**, 344–347 (2015). doi: [10.1126/science.aaa0445](https://doi.org/10.1126/science.aaa0445); pmid: [25883358](https://pubmed.ncbi.nlm.nih.gov/25883358/)
52. D. Zhao *et al.*, Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat. Biotechnol.* **39**, 35–40 (2021). doi: [10.1038/s41587-020-0592-2](https://doi.org/10.1038/s41587-020-0592-2); pmid: [32690970](https://pubmed.ncbi.nlm.nih.gov/32690970/)
53. I. C. Kurt *et al.*, CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **39**, 41–46 (2021). doi: [10.1038/s41587-020-0609-x](https://doi.org/10.1038/s41587-020-0609-x); pmid: [32690971](https://pubmed.ncbi.nlm.nih.gov/32690971/)
54. H. Tong *et al.*, Programmable A-to-Y base editing by fusing an adenine base editor with an N-methylpurine DNA glycosylase. *Nat. Biotechnol.* **41**, 1080–1084 (2023). doi: [10.1038/s41587-022-01595-6](https://doi.org/10.1038/s41587-022-01595-6); pmid: [36624150](https://pubmed.ncbi.nlm.nih.gov/36624150/)
55. B. Bushnell, “BBMap: A fast, accurate, splice-aware aligner,” LBNL-7065E (Lawrence Berkeley National Lab, 2014); <https://www.osti.gov/biblio/1241166>
56. H. Li, seqtk: Toolkit for processing sequences in FASTA/Q formats, GitHub (2023); <https://github.com/hb3/seqtk>
57. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923); pmid: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
58. K. Clement *et al.*, CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019). doi: [10.1038/s41587-019-0032-3](https://doi.org/10.1038/s41587-019-0032-3); pmid: [30809026](https://pubmed.ncbi.nlm.nih.gov/30809026/)
59. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019). doi: [10.1038/s41587-019-0201-4](https://doi.org/10.1038/s41587-019-0201-4); pmid: [31375807](https://pubmed.ncbi.nlm.nih.gov/31375807/)
60. J. Y. Hsu *et al.*, PrimeDesign software for rapid and simplified design of prime editing guide RNAs. *Nat. Commun.* **12**, 1034 (2021). doi: [10.1038/s41467-021-21337-7](https://doi.org/10.1038/s41467-021-21337-7); pmid: [33589617](https://pubmed.ncbi.nlm.nih.gov/33589617/)
61. J. W. Nelson *et al.*, Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2022). doi: [10.1038/s41587-021-01039-7](https://doi.org/10.1038/s41587-021-01039-7); pmid: [34608327](https://pubmed.ncbi.nlm.nih.gov/34608327/)
62. D. Chen, Helicase-assisted continuous editing for programmable mutagenesis of endogenous genomes, Zenodo (2024); <https://doi.org/10.5281/zenodo.11436318>

ACKNOWLEDGMENTS

We thank the members of the Chen and Bernstein labs for helpful discussions. We also thank A. Krishnan, L. Guan, and A. Guan for their support and helpful discussions. We thank L. Gaffney at the Broad Institute for help with graphic design and figure editing. We also thank A. Patentreger, A. Masse, and N. Pirete at the Broad Institute Flow Cytometry Core Facility for their experiment support. **Funding:** X.D.C. is supported by an American Heart Association Predoctoral Fellowship (23PRE1011742). Z.C. is supported by NCI-CA-234842. M.V. is the David M. Livingston, MD, Physician-Scientist of the Damon Runyon Cancer Research Foundation, and an Edward P. Evans Foundation EvansMDS Young Investigator. This work was supported by funds from the NCI/NIH Director's Fund (DP1CA216873 to B.E.B.), the Gene Regulation Observatory, and the Variant-to-Function Initiative at the Broad Institute. F.C. acknowledges support from an NIH Early Independence Award (DP5, IDP5OD024583), the NHGRI (R01, R01HG010647), a Burroughs Wellcome Fund CASI Award, the Searle Scholars Foundation, the Harvard Stem Cell Institute, and the Merkin Institute. This research was supported by the New York Stem Cell Foundation (NYSCF). F.C. is a NYSCF – Robertson Investigator. **Author contributions:** F.C., G.W., and X.D.C. conceived of the study. X.D.C., Z.C., B.C.O., Y.Z., G.W., G.S., K.T., Y.-K.C., A.N.T., M.B., and E.T. designed and performed the experiments in this study and analyzed the data. M.V. and J.S. helped with experiments related to SF3B1. H.C. helped with the experiment design for MEK1 studies. B.E.B. provided additional supervision. X.D.C., Z.C., B.E.B., and F.C. wrote the manuscript, with input from all authors. **Competing interests:** A patent application has been filed related to this work. B.E.B. declares outside interests in Fulcrum Therapeutics, HiFiBio, Arsenal Biosciences, Design Pharmaceuticals, Cell Signaling Technologies, and Chroma Medicine. F.C. is a founder of Curio Bioscience and Doppler Bio. **Data and materials availability:** Sequencing data have been deposited at the Sequence Read Archive database under accession no. PRJNA1111669. Expression plasmids will be made available on Addgene under a UBMTA (Uniform Biological Material Transfer Agreement). The code for this paper is available in Zenodo (62). **License information:** Copyright © 2024 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adn5876

Figs. S1 to S9
Tables S1 to S13
References (63–69)
MDAR Reproducibility Checklist

Submitted 17 December 2023; accepted 14 August 2024
10.1126/science.adn5876