Comprehensive mapping of key regulatory networks that drive oncogene expression

Lin, L.; Holmes, B.; Shen, M.W.; Kammeron, D.; Geijsen, N.; Gifford, D.K.; Sherwood, R.I.

Cell Reports

CellPress
OPEN ACCESS

Article

Comprehensive Mapping of Key Regulatory Networks that Drive Oncogene Expression

Lin Lin,[1,2] Benjamin Holmes,[3] Max W. Shen,[3,4] Darnell Kammeron,[2] Niels Geijsen,[2,5,*] David K. Gifford,[3,6,7,8,*] and Richard I. Sherwood[1,2,9,*]

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
[2]Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences, Utrecht 3584 CT, the Netherlands
[3]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5]Department of Anatomy and Embryology, Leiden University Medical Center, Leiden 2300 RC, the Netherlands
[6]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[7]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[9]Lead Contact
*Correspondence: n.geijsen@lumc.nl (N.G.), gifford@mit.edu (D.K.G.), rsherwood@rics.bwh.harvard.edu (R.I.S.)
https://doi.org/10.1016/j.celrep.2020.108426

SUMMARY

Gene expression is controlled by the collective binding of transcription factors to *cis*-regulatory regions. Deciphering gene-centered regulatory networks is vital to understanding and controlling gene misexpression in human disease; however, systematic approaches to uncovering regulatory networks have been lacking. Here we present high-throughput interrogation of gene-centered activation networks (HIGAN), a pipeline that employs a suite of multifaceted genomic approaches to connect upstream signaling inputs, *trans*-acting TFs, and *cis*-regulatory elements. We apply HIGAN to understand the aberrant activation of the cytidine deaminase APOBEC3B, an intrinsic source of cancer hypermutation. We reveal that nuclear factor κB (NF-κB) and AP-1 pathways are the most salient *trans*-acting inputs, with minor roles for other inflammatory pathways. We identify a *cis*-regulatory architecture dominated by a major intronic enhancer that requires coordinated NF-κB and AP-1 activity with secondary inputs from distal regulatory regions. Our data demonstrate how integration of *cis* and *trans* genomic screening platforms provides a paradigm for building gene-centered regulatory networks.

INTRODUCTION

Signaling pathways act through *trans*-acting transcription factors (TFs) at *cis*-acting regulatory elements to alter gene expression in response to internal and environmental cues. Understanding gene-regulatory mechanisms has tremendous implications in both fundamental biology and the identification of druggable targets for therapeutic intervention. However, methods such as chromatin immunoprecipitation sequencing (ChIP-seq) and yeast one-hybrid assay remain low-throughput and fail to draw a causal link between the physical TF-DNA binding and its functional impact on modulating gene expression. To complement these observational approaches, we have used CRISPR-Cas9 technology to interrogate the function of both the coding and regulatory genome *in situ* through perturbing TFs and *cis*-regulatory elements, offering scalable approaches to decipher gene-regulatory mechanisms. Previous efforts have mainly focused on identifying in *cis* non-coding elements without connecting them to regulatory TFs (Canver et al., 2015; Diao et al., 2017; Fulco et al., 2016; Gasperini et al., 2017; Klann et al., 2017; Korkmaz et al., 2016; Rajagopal et al., 2016; Sanjana et al., 2016) or to study in *trans* effects of disrupting TF expression on global gene expression without relating them to *cis*-regulatory sequences (Adamson et al., 2016; Dixit et al., 2016; Jaitin et al., 2016). None of these previous approaches have provided a causal structure that could connect upstream signaling pathways to their downstream TFs to specific *cis*-regulatory sequences mediating gene expression. Here we present high-throughput interrogation of gene-centered activation networks (HIGAN), a pipeline that allows the systematic, high-resolution mapping of *cis*- and *trans*-acting regulators and their functional relevance in controlling gene expression. HIGAN involves a CRISPRi-mediated tiling *cis*-regulatory region screen (Figure 1A) in parallel with a *trans*-regulatory TF screen (Figure 1B), followed by mutant tiled STARR-seq dissection of candidate regions (Figure 1C). With HIGAN, enhancer regions and signaling pathways are prioritized for further evaluation and manipulation (Figure 1D). As a proof-of-concept demonstration of this technology to obtain mechanistic insights into gene regulation, we applied HIGAN to study the cancer-specific upregulation of the cytidine deaminase APOBEC3B.

APOBEC cytidine deaminases, originally identified as host antiviral genes, target single-stranded polynucleotide substrate

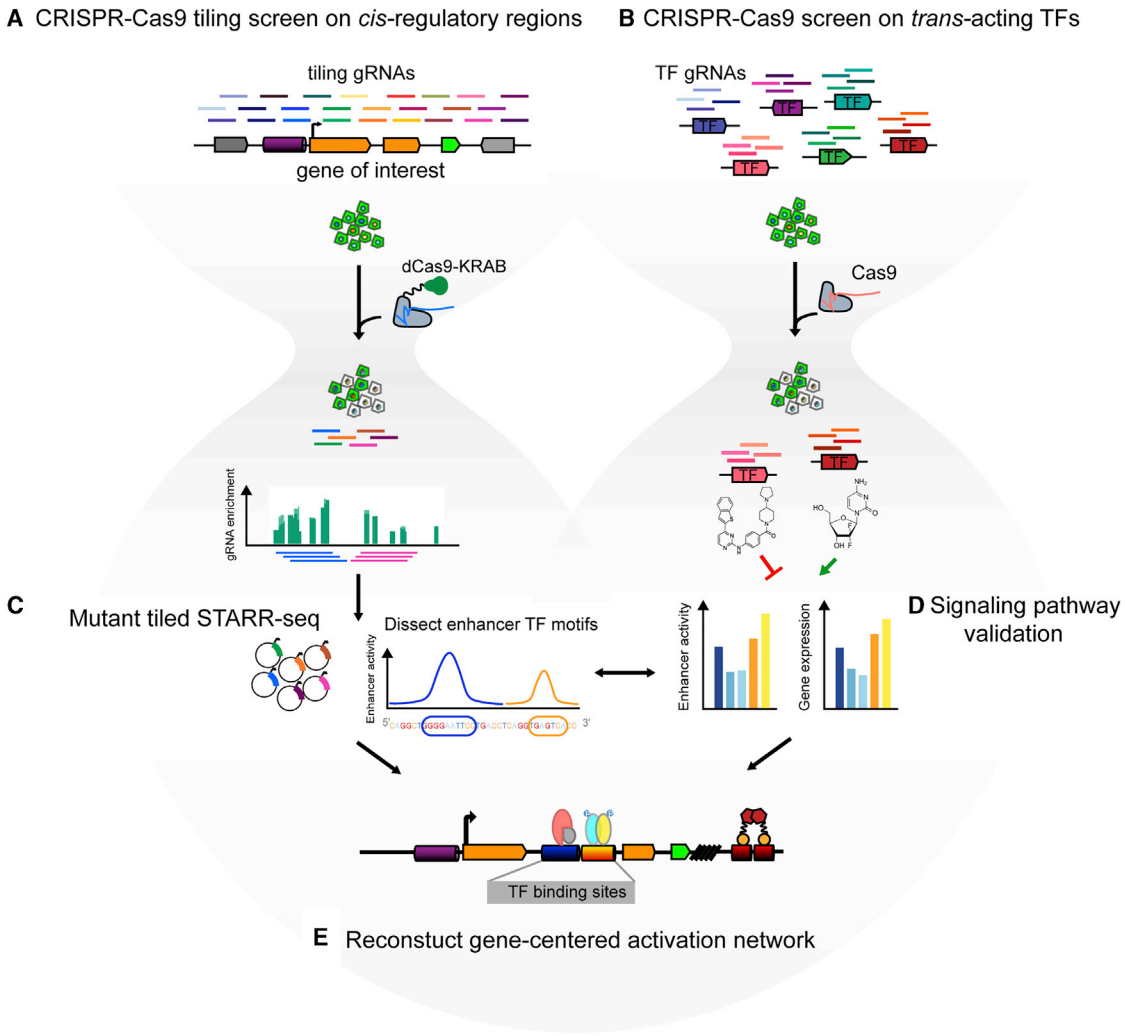


Figure 1. Workflow of High-Throughput Interrogation of Gene-Centered Activation Networks (HIGAN) Pipeline

This pipeline consists of: (A) Tiled CRISPRi screening of *cis*-regulatory region, (B) TF-wide CRISPR knockout screen on *trans*-regulatory modulators, (C) tiling STARR-seq assay with mutant tiles designed from identified regulatory regions, (D) signaling validation through small molecule perturbation, (E) HIGAN output converged to reconstruct a gene-centered activation network.

of hosts and pathogens and play an important role in the innate immune system (Stavrou and Ross, 2015). More recently, it was shown that the aberrant overexpression of APOBEC3B has been correlated with increased mutation burden in various cancer types and also with poor prognosis and drug resistance (Alexandrov et al., 2013; De Bruin et al., 2014; Law et al., 2016; Leonard et al., 2013; Sieuwerts et al., 2017; Supek and Lehner, 2017; Taylor et al., 2013; Wang et al., 2018). In contrast, a recent study has demonstrated that forced overexpression of APOBEC3B in tumor cells generates novel neoantigens through APOBEC3B-mediated mutations that sensitize cancer cells to checkpoint blockage immunotherapy (Driscoll et al., 2020). Although there are thus compelling arguments for the therapeutic utility of modulating APOBEC3B expression, how APOBEC3B expression is regulated in cancer cells remains incompletely understood. It has recently been shown that, in breast cancer cell lines, DNA replication stress correlated with APOBEC3 induction, which can be attenuated by inhibiting phosphatidylinositol 3-kinase (PI3K) and mitogen-activated protein kinase (MAPK) signaling, suggesting that these pathways are involved in APOBEC3 activation (Kanu et al., 2016). Recent studies have also revealed an important role for both classical and alternative nuclear factor κB (NF-κB) signaling pathways in regulating APOBEC3B (Leonard et al., 2015; Maruyama et al., 2016; Tilborghs et al., 2017). However, owing to the fact that NF-κB members are ubiquitously expressed, these previous studies did not pinpoint the specific coactivators and the *cis*-regulatory loci that trigger dysregulated APOBEC3B expression in cancer cells. Thus, it is as yet unclear how NF-κB cooperates with other pathways to induce APOBEC3B and why only a subset of NF-κB-induced cancers upregulates APOBEC3B expression.

To decode the transcriptional dysregulation of APOBEC3B in cancer cells, we applied HIGAN in two cancer cell lines. We identified 13 intronic and distal regions required for APOBEC3B

expression and mapped the network of TFs that functionally regulated APOBEC3B expression involved in a number of inflammatory signaling pathways. Furthermore, HIGAN systematically dissected *cis*-regulatory modules and mapped these TF candidates at high resolution to the functional binding elements. In a final step, we showed that different signaling pathways orchestrated APOBEC3B gene activation through validation by drug treatments. Collectively, we demonstrate that HIGAN is a robust, generalizable screening pipeline, to interrogate gene-centric regulatory networks.

RESULTS

CRISPR-Cas9 Tiling Screen Provides Candidate Regulatory Regions for APOBEC3B Gene Expression

To investigate *cis*-regulatory inputs to APOBEC3B gene expression in its native locus, we designed a library of 12,270 gRNAs tiling a 600-kb *cis*-regulatory region surrounding the APOBEC3 gene family locus with a median inter-gRNA distance of 256 bp, together with 125 non-targeting gRNAs as negative controls (Figure 2A; STAR Methods) (Doench et al., 2016). Next, we cloned the gRNA library into a Tol2 transposon-based vector containing a hygromycin expression cassette allowing genome integration, gRNA expression, and population selection (STAR Methods) (Shen et al., 2018). To facilitate a sensitive readout of APOBEC3B gene expression, we generated EGFP knockin reporters of the APOBEC3B locus linked by the self-cleaving P2A peptide in human U2OS osteosarcoma and DLD-1 colorectal adenocarcinoma cells. These two cell lines have a relatively high expression of APOBEC3B, allowing robust quantification of the APOBEC3B-GFP reporter (STAR Methods) (Figures S1A and S1B). Next, we established a high-throughput protocol adapted from previous tiled *cis*-regulatory CRISPR screening efforts (Fulco et al., 2019; Rajagopal et al., 2016; Sanjana et al., 2016; Sher et al., 2019) to perform pooled screening with a transposon-based delivery strategy (STAR Methods) (Figure 2B). We stably integrated the in *cis* gRNA library into cancer cells, titrating the library in an empirically calculated ratio with an empty plasmid to ensure that most cells integrate a single gRNA, followed by dCas9-KRAB (CRISPRi) targeting through a second round of Tol2 transposon-based delivery with blasticidin selection. Two weeks after dCas9-KRAB inhibition, we detected fractional loss of EGFP in the bulk population, and then we performed fluorescence-activated cell sorting (FACS) to enrich the EGFP-negative (EGFP$^-$) population (Figure 2C). Next-generation DNA sequencing (NGS) was performed on genomic DNA from the bulk population before FACS and the enriched EGFP$^-$ population after FACS, along with the bulk population prior to Cas9 targeting (two independent pre-Cas9 biological replicates and two independent post-Cas9 biological replicates for each pre-Cas9 pool).

We collected an average of ~1.6 million NGS reads per library, recovering 80%–100% of gRNAs with over 10 reads in all unsorted replicates (Figure S1C) with Pearson's correlation coefficient (r) of normalized read counts for each gRNA between biological replicates and cell lines ranging from 0.21 to 0.3 (Figure S1D). To identify gRNAs enriched in EGFP$^-$ cells, we performed gRNA differential enrichment analysis using the MAGeCK pipeline (STAR Methods) (Li et al., 2014) (Table S7). As expected, gRNAs targeting in the proximity of transcription start sites of APOBEC3B (−1,000 to +100 bp, n = 31) were highly enriched with consistent positive $\log_2$ fold change (LFC) (U2OS: 4.95 ± 1.32; DLD1: 4.15 ± 1.46), confirming the sensitivity of our pooled screening strategy to detect sequences essential for APOBEC3B expression (Figures 2D and 2E). In addition, we identified a number of candidate distal non-coding regulatory regions that coincide with the active histone marks H3K4me3 and H3K27ac (Abraham et al., 2017; Fishilevich et al., 2017; Mei et al., 2017; Rokavec et al., 2017; Walz et al., 2014; Zheng et al., 2019). Interestingly, we also observed a subset of distal candidate regions lacking active chromatin modification. Next, to prioritize candidate regulatory regions for further validation, we identified 216 *cis*-targeting gRNAs from U2OS cells and 465 *cis*-targeting gRNAs from DLD1 cells that showed statistically significant positive enrichment by MAGeCK enrichment analysis (STAR Methods) using four independent post-Cas9 biological replicates with a cutoff of false discovery rate (FDR) < 0.2 (Figure 2F; Table S7). Among these enriched gRNAs, we observed 177 gRNAs adjacent to other candidate gRNAs within 500 bp and identified 27 genomic regions (<2 kb) clustered with multiple candidate gRNAs for further investigation.

In parallel, we performed a CRISPR-Cas9 knockout screen with the same *cis*-targeting gRNA library together with 77 gRNAs targeting EGFP serving as positive controls. As expected, positive control gRNAs were highly enriched in the EGFP$^-$ population (Figure S1F). However, we found that the enrichment of EGFP loss of individual *cis*-targeting gRNAs was less consistent across matched replicates compared with the CRISPRi screen. With the same FDR cutoff, we obtained only 16 gRNAs from the U2OS cell line and none from the DLD1 cell line, which might be because of the fact that the range of Cas9-induced indels is more variable than the effective window of dCas9-KRAB (Gilbert et al., 2013, 2014; Radzisheuskaya et al., 2016). Thus, our result indicated that the CRISPRi strategy outperformed CRISPR-Cas9 knockout in replicate consistency for profiling *cis*-regulatory regions. To provide an overview of the gRNA enrichment from the CRISPR knockout screen, we averaged the LFC of each gRNA within a 1-kb bin in the APOBEC3B ± 300-kb locus (Figure S1E). We still observed the strongest gRNA enrichment mapping to the APOBEC3B promoter and gene body, confirming the sensitivity of the pooled knockout screen to detect the core gene expression elements. We also observed the participation of promoters for the neighboring gene APOBEC3C in regulating APOBEC3B expression in both CRISPR knockout and inhibition (Figure S1G), consistent with previous findings from *cis*-regulatory CRISPR screening (Fulco et al., 2016; Rajagopal et al., 2016).

Validation and Characterization of Key APOBEC3B *cis*-Regulatory Elements

To assess *cis*-regulatory function of candidate regions prioritized from our CRISPRi screen, we selected 18 distinct regions from the candidate regions and distal regions with DNase hypersensitivity for further characterization. For each region, we used three experimental methods to assess regulatory function: dual-gRNA CRISPR deletion (Diao et al., 2017; Gasperini et al., 2017; Liu

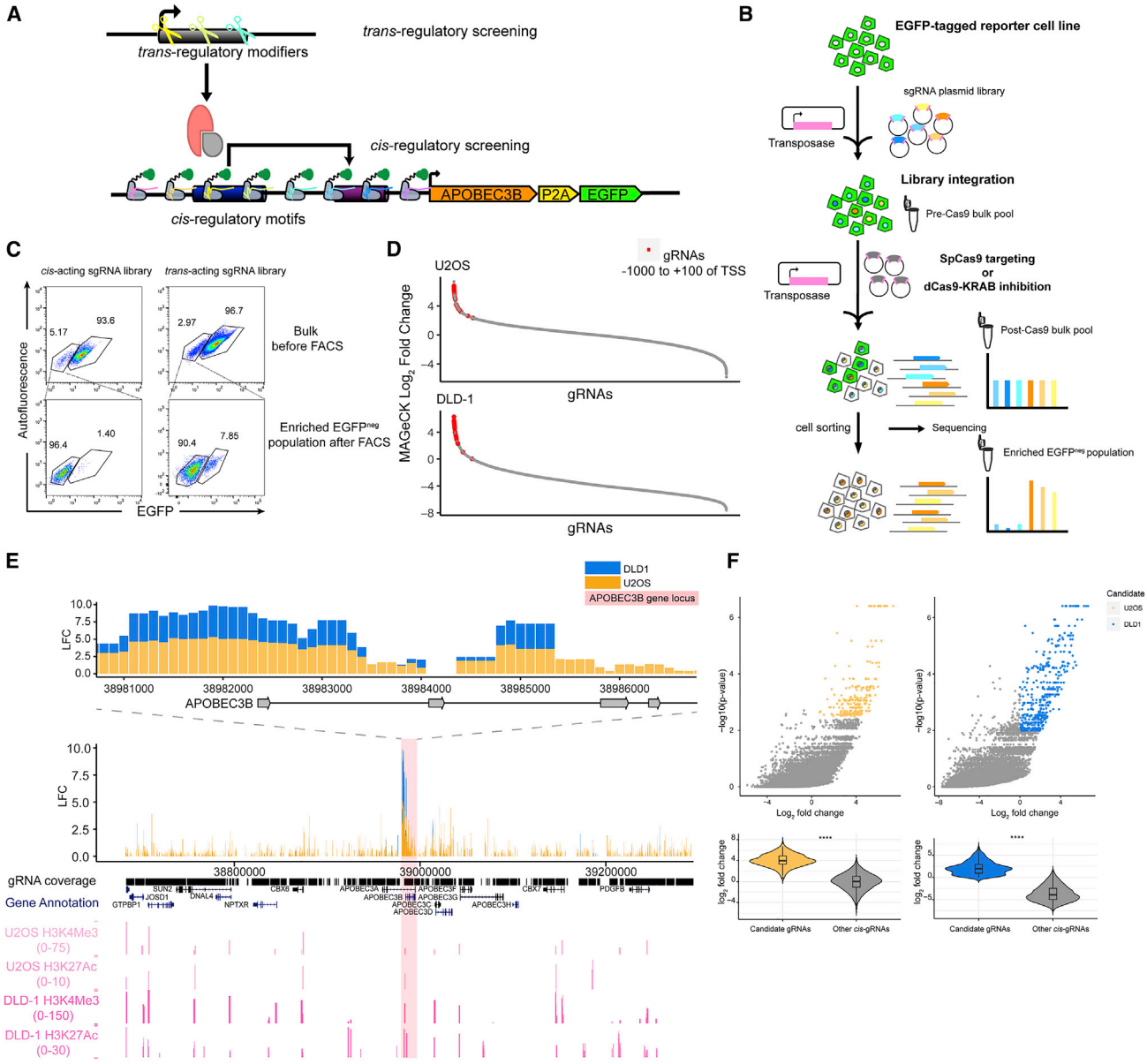


Figure 2. CRISPR-Cas9-Based Screening Identifies In *cis*- and In *trans*-Acting Regulatory Elements of APOBEC3B in Cancer Cells

(A) Schematics of *cis* and *trans* screening design.

(B) Schematic of screening workflow.

(C) Representative FACS plot analyzed from cancer cells undergoing screening before and after enriching the EGFP^{-} cell population.

(D) Dot plot showing the representative distribution of MAGeCK LFC for gRNAs of in *cis* screening library from U2OS and DLD-1 cell lines. Red dots indicate gRNAs targeting APOBEC3B TSS locus (−1,000 to +100 bp).

(E) Stacking plot depicting the LFC of *cis*-targeting gRNAs across the APOBEC3B ± 300 kb locus. A zoomed-in view of the gRNA enrichment covering APOBEC3B gene locus in proximity to promoter is plotted on top; gRNA coverage of in *cis* screening library; gene annotation; normalized signal tracks of H3K4me3; and H3K27Ac ChIP-seq from both U2OS and DLD-1 cell lines.

(F) Volcano plot depicting the statistics cutoff for candidate gRNAs based on MAGeCK pipeline. Violin plot summarizing the MAGeCK LFC of candidate gRNA versus the other *cis*-targeting gRNAs. Statistics analysis is conducted by unpaired t test with Welch's correction, ****$p < 0.0001$.

See also Figure S1 and Table S1.

et al., 2018; Zhu et al., 2016), dual-gRNA CRISPR interference (Fulco et al., 2016; Thakore et al., 2015), and a luciferase reporter assay, all of which are orthogonal to our single-gRNA Cas9-based screen.

We performed gene region knockouts with a pair of gRNAs flanking each of the distal candidate regions (averaged ~2-kb distance) to create site-specific deletion in the U2OS APOBEC3B-EGFP reporter cell line (Table S2). We observed

significant loss of EGFP reporter expression in ~30%–40% of cells with positive control gRNA pairs targeting the APOBEC3B gene promoter or gene body, while non-targeting gRNAs induce EGFP loss in ~1%–3% of cells. We observed loss of EGFP in 4%–18% of cells, with gRNA pairs targeting candidate regulatory regions (Figures 3A and S2A), and 9/18 regions showed statistically significantly increased loss of EGFP compared with cells treated with non-target control gRNA pairs (FDR < 0.05, multi t test). Of note, in addition to the desired deletion, CRISPR-Cas9 may also induce long-range chromosome deletion; thus, additional experiments were performed to further confirm that the validated candidate regions contribute to APOBEC3B gene expression. We measured APOBEC3B mRNA expression in U2OS cells with distal paired gRNA deletions of DisEn21, DisEn23, and DisEn26, which we confirmed to show deletion genotypes (Figure S2B). We sorted the cells into $EGFP^+$ and $EGFP^-$ bins. Noting that only one APOBEC3B allele has GFP knockin, the $EGFP^+$ bin may contain both untargeted and single-allele targeted cells (Figure 3A, DisEn21 [5.6% $EGFP^-$], EisEn23 [4.9% $EGFP^-$], DisEn26 [10.0% $EGFP^-$]). Expression of APOBEC3B is partially impaired in EGFP-positive cells and significantly ablated in the $EGFP^-$ population with the deleted candidate regions (Figure 3B), while the neighboring genes JOSD1 and SUN2 showed no significant changes upon distal targeting.

We additionally assessed the regulatory input of the candidate enhancer regions by using dCas9-KRAB in U2OS APOBEC3B-EGFP cells. We selected 10 candidate regions with chromatin accessibility in U2OS cells for CRISPR interference, using the gene promoter region as positive control and two non-targeting gRNAs as negative controls (Figure 3C). dCas9-KRAB targeting of the APOBEC3B promoter region induced significant, robust loss of EGFP (~21.5% $EGFP^-$, ~4.8% EGFPmed) (Figures S2C and S2D), while negative control targeting caused little EGFP loss (~0.8% $EGFP^-$, ~2.1% EGFPmed). Among all the tested regions, we identified a strong enhancer localized in the second intron of APOBEC3B, the targeting of which produced significant, robust partial reduction of reporter intensity (~12.7% $EGFP^-$, ~15.8% EGFPmed), and targeting of distal enhancer candidates caused modest EGFP loss (~2% $EGFP^-$, ~3% EGFPmed, 6/10 statistically significantly increased over control gRNAs). Comparing the results from dual-gRNA deletion assay and dCas9-KRAB inhibition, we observed that the gene promoter and intronic enhancer contributed most strongly to APOBEC3B expression. Distal enhancer candidates yielded less robust effects on gene regulation in both assays, but there was still concordance in which distal regions had comparatively stronger effects.

We then asked whether these same 10 candidate regulatory regions contain enhancer activity. We cloned 400- to 800-bp candidate regions into a luciferase reporter and measured their luminescence intensity, using a APOBEC3B-adjacent DNase I-sensitive region without statistically significant gRNA enrichment as a negative control (Figure 3D). Compared with the negative control region, which showed no enhancer activity, 8/10 of the tested candidate regions showed significant, robust enhancer activity in both cell lines. We found that the intronic enhancer region exerted ~60- to 100-fold upregulation of luciferase reporter activity. We identified multiple distal regions ~200 kb upstream or downstream of APOBEC3B, such as DisEn4, DisEn5, DisEn21, and DisEn26, which also exhibited strong enhancer activity. One region, DisEn23, significantly repressed luciferase expression, although this region contained gRNA enrichment, and paired gRNA targeting also induced modest EGFP loss in the APOBEC3B-GFP reporter cell line (Figures 3A and 3B). This indicates that not only enhancers but also other types of *cis*-regulatory modules might be involved in APOBEC3B regulation. Taken together, these validations experiments converge on the conclusion that APOBEC3B expression in U2OS and DLD-1 cells is primarily controlled by the promoter and a strong intronic enhancer, while a cohort of at least four distal enhancers provides modest but significant inputs to expression in all assays.

In *trans* TF-wide Screening Identifies Multiple Pro-inflammatory Pathways that Regulate APOBEC3B Expression

We next sought to ask which in *trans*-acting TFs are involved in regulating APOBEC3B gene expression. We selected a collection of 7,210 gRNAs from Brunello CRISPRko library targeting 1,803 annotated human TFs (Lambert et al., 2018). Meanwhile, we recruited 77 gRNAs targeting EGFP as positive control and 100 non-targeting gRNAs as negative control (Figure 1A; STAR Methods). With the same screening strategy as was employed in the *cis*-targeting screen, we assessed loss of APOBEC3B-GFP expression upon CRISPR-Cas9-mediated TF knockout in U2OS and DLD-1 cell lines (Figure 2B). Similar to the *cis*-regulatory screen, we obtained a high coverage rate in the bulk population and obtained high concordance between the biological replicates and cell lines (Figures S3A and S3C). MAGeCK LFC was calculated per gRNA and also per gene for each biological replicate through a standard pipeline (Figure S3B; Table S8; STAR Methods). As expected, gRNAs targeting EGFP were highlighted with top enrichment; meanwhile, non-targeting control gRNAs showed no evidence of enrichment (Figures 4A and S3D). We further prioritized 5%–8% of the TFs through a MAGeCK analysis pipeline based on enrichment LFC and individual gRNA consistency across biological replicates (84 TFs from U2OS and 156 TFs from DLD-1) (Figures 4A and 4B; STAR Methods). A total of 13 TFs were found to be enriched in both cell lines (*NFKB1*, *NFKB2*, *SNAI3*, *CREB3L2*, *IRF3*, *RFX2*, *ST18*, *ZNF365*, *DVL1*, *SHPRH*, *MSX2*, *EZH1*, and *NR1H2*).

To validate the importance of the candidate *trans*-regulatory TFs on APOBEC3B expression, we first performed CRISPR-Cas9 targeting of two of the strongest hits, NFKB1 and NFKB2, using two gRNAs per gene in APOBEC3B-expressing U2OS cells (Figure 5A). Compared with non-targeting controls, we observed a distinct $EGFP^-$ population upon Cas9-mediated targeting in NFKB1 and NFKB2 (Figure 5A); however, there was only a 1.5%–3% increase in the frequency of APOBEC3B-GFP^- cells. We hypothesized that the surprisingly weak effect of these knockouts on APOBEC3B expression, in light of the highly significant identification of NFKB1 and NFKB2 in the in *trans* CRISPR-Cas9 screen (Figure S3E), may be because of functional redundancy between NFKB1 and NFKB2. Thus, to overcome possible genetic redundancy, we treated U2OS cells with a IκB kinase β (IKKβ)-selective inhibitor ML120B, which is known to block NF-κB upstream activation (Nagashima et al., 2006; Wen et al.,

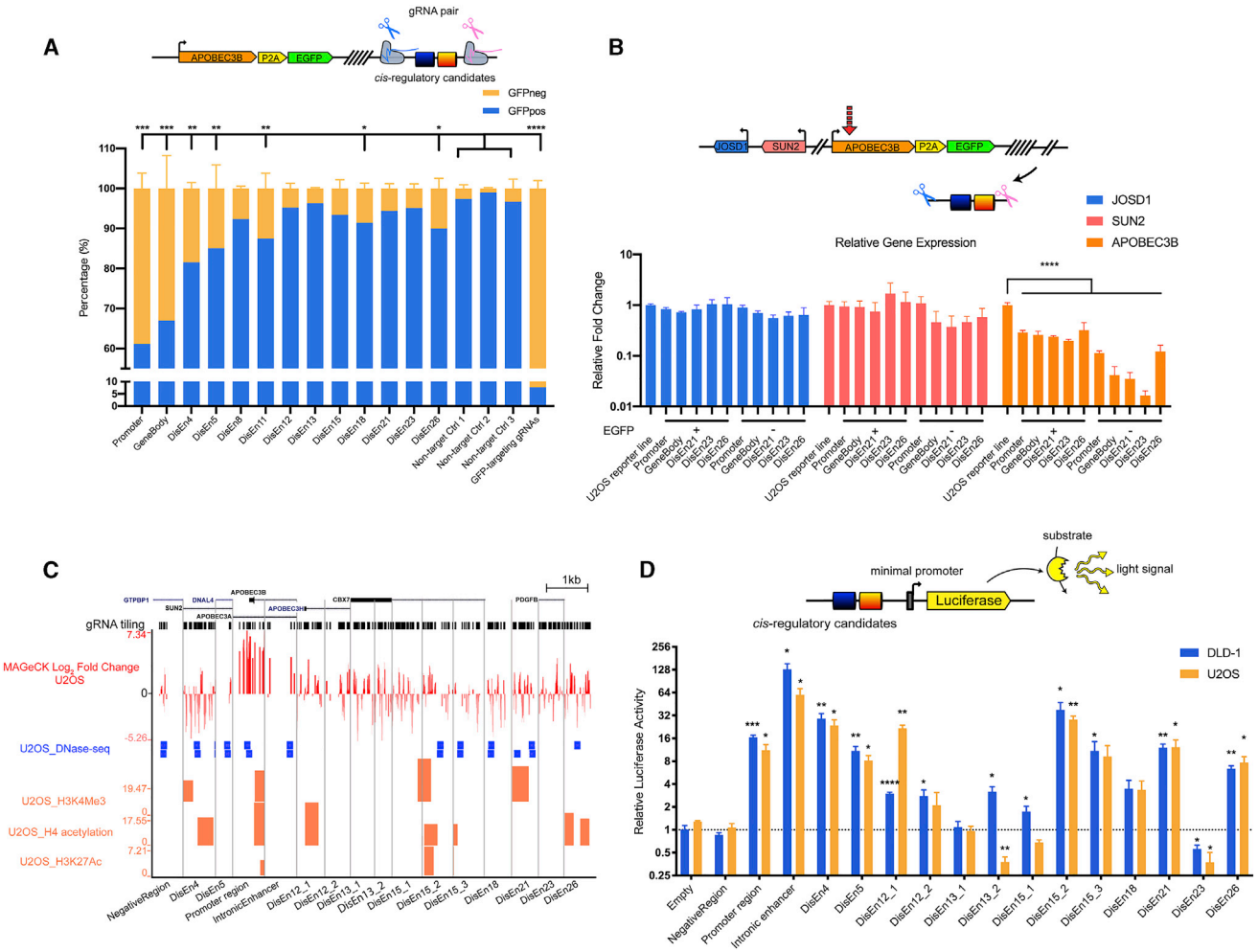


Figure 3. Individual Validation of Selective In *cis* Screening Candidate Regions

(A) FACS quantification of EGFP expression in APOBEC3B reporter cells upon paired gRNA targeting of candidate regions.

(B) Relative gene expression of JOSD1, SUN2, and APOBEC3B in EGFP-positive and -negative populations upon targeting with paired gRNAs for candidate regions.

(C) A multi-region view of the candidate regions showing the following in track order from top to bottom: gene annotation, gRNA library coverage, tracks showing the rolling averages of MAGeCK LFC enrichment from five gRNAs in proximity in U2OS cell line, DNase-seq peak generated from U2OS cell line, and normalized tracks of active histone mark ChIP-seq generated from the U2OS cell line.

(D) Relative luciferase activity for candidate enhancer regions over a negative control region in both U2OS (yellow) and DLD-1 (blue) cell lines.

In all bar plots, data are shown as mean ± SEM. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. (A) Statistics analysis is conducted by one-way ANOVA; $EGFP^-$ population is compared with pool of non-targeting controls, with $Q = 5\%$, $n = 3$ for each paired gRNA targeting. (B) Statistics analysis is conducted by one-way ANOVA, with $Q = 5\%$, $n = 4$. (D) Luciferase readouts from empty vector and each construct are compared with negative region construct within each cell line using one-way ANOVA, with $Q = 5\%$, $n = 4$. See also Figure S2 and Table S1. The complete chromosome coordinates of the regions highlighted in this figure can be found in Table S3.

2006). We observed that, compared with DMSO-treated cells, ML120B-treated cells had ~32.5% as much APOBEC3B expression ($p < 0.05$), whereas expression of neighboring genes APOBEC3C and JOSD1 was unaffected (Figure 5B), ruling out the possibility that the APOBEC3B inhibition was due to a global transcriptional inhibition by drug treatment. Next, we treated HCT116 cells, a low APOBEC3B-expressing cell line (Figures S1A and S1B), with agents that activate NF-κB pathways, including the protein kinase C (PKC) activator phorbol 12-myristate 13-acetate (PMA), pro-inflammatory cytokine tumor necrosis factor α (TNF-α), and chemotherapeutic drug gemcitabine (Figure 5C) (Arora et al., 2013; Hernández-Vargas et al., 2007; Holden et al., 2008; Nagashima et al., 2006; Schütze et al., 1995). We found that NF-κB activation through each of these agents elicited a significant increase in APOBEC3B gene expression with less robust activation of APOBEC3C and no effect on JOSD1. Analysis of published data corroborates this link (Janus et al., 2018; Zhao et al., 2018), showing that U2OS cells treated with TNF-α or ionizing radiation upregulate APOBEC3B, and such upregulation can be suppressed by siRNA inhibition of

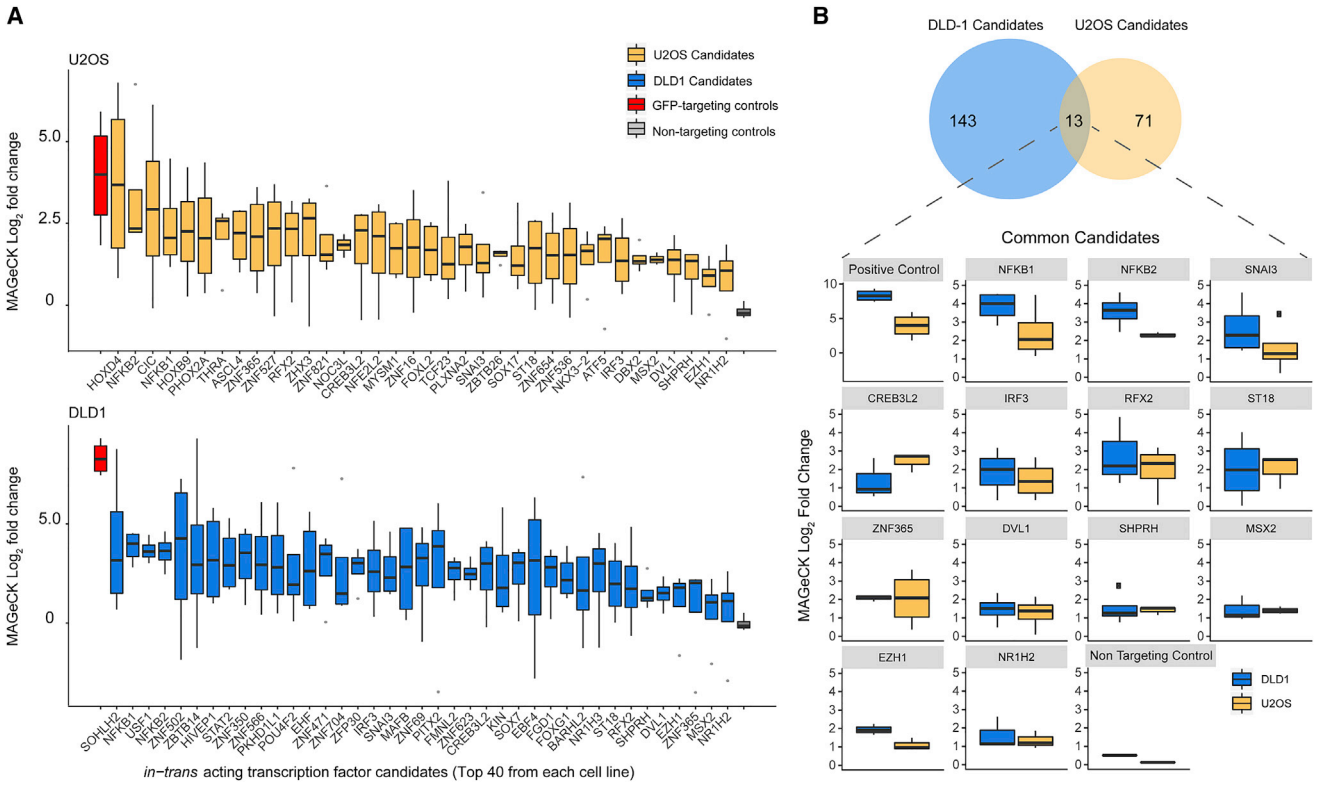


Figure 4. TF-Wide CRISPR Screen Identifies In *trans*-Acting Regulators of APOBEC3B Expression

(A) Boxplot showing the top enriched TF candidates of in *trans* screening library from U2OS (yellow) and DLD-1 (blue) cell lines based on averaged MAGeCK LFC. Positive controls of gRNA targeting EGFP are shown in red, while non-targeting gRNA controls are shown in gray.

(B) Venn diagram on top showing the TFs identified by fulfilling enrichment criteria in U2OS (yellow) and DLD-1 (blue) cell lines. Boxplots below showing the common TF hits and their respective MAGeCK LFC from different experiments (n = 4).

See also Figures S3 and S4.

RELA (Figure S3F). Overall, these data validate that NF-κB signaling is necessary and sufficient for inducing APOBEC3B expression and activation in cancer cells, and that there is redundancy among NF-κB TFs.

We also sought to validate the contribution of other TFs identified in the in *trans* screen. Noting the increased robustness of small-molecule treatment over TF knockout in blocking certain signaling pathways driven by redundant TFs, we treated U2OS cells with five small-molecule inhibitors that target signaling pathways in which at least one TF was a significant in *trans* hit in both cell lines. Treatment with GSK2033 (nuclear receptor antagonist blocking NR1H2 and NR1H3), MRT67307 (IRF signaling inhibitor blocking IRF3 and STAT2), ML385 (Nrf2 inhibitor blocking NFE2L2 and NFE2L3), U0126 (MEK inhibitor blocking ETV7 and EHF), and 666-15 (CREB inhibitor blocking CREB3L2) each induced mild (47%–79% of DMSO control at the most effective dose) dose-dependent reduction of APOBEC3B expression (Figure S4A). All of these pathways have been implicated in the inflammatory response, bolstering the idea that APOBEC3B is regulated by pro-inflammatory signals. Of note, in a collection of RNA-seq data from 1,036 cancer cell lines (Ji et al., 2019), we observed significantly higher expression of APOBEC3B in cancer cell lines with higher inflammatory levels, defined by expression of common targets of the STAT3/NF-κB/AP-1 regulatory network, as compared with cell lines with low inflammatory levels (Figure S4B).

To assess the significance of the in *trans* screening results on cancer-specific upregulation of APOBEC3B, we calculated the correlation between expression of APOBEC3B and every human TF in 1,156 cell lines profiled in the Cancer Cell Line Encyclopedia (CCLE) RNA-seq database (Barretina et al., 2012) (Figure S4C). We find that TFs of the NF-κB and AP-1 families are the most strongly correlated with APOBEC3B across cancer cell lines. Comparing the results of our in *trans* screen with the CCLE expression correlation coefficients, we find that in both cell lines, NFKB1 and NFKB2 stand out as the most critical APOBEC3B regulators (Figures 5E and 5F), consistent with previous work (Leonard et al., 2015; Maruyama et al., 2016; Tilborghs et al., 2017) and expression profiling of cancer biopsies from The Cancer Genome Atlas (TCGA) collection (Weinstein et al., 2013) (Figure S4D). In addition to NF-κB, all five of the pathways represented by in *trans* hits whose effects on APOBEC3B expression we validated above are represented among the TFs most strongly correlated with APOBEC3B expression in CCLE cell lines. Often, the same TFs whose knockout impairs APOBEC3B expression are strongly correlated with APOBEC3B expression (such as NFKB2, NFE2L3, and ETV7), but sometimes

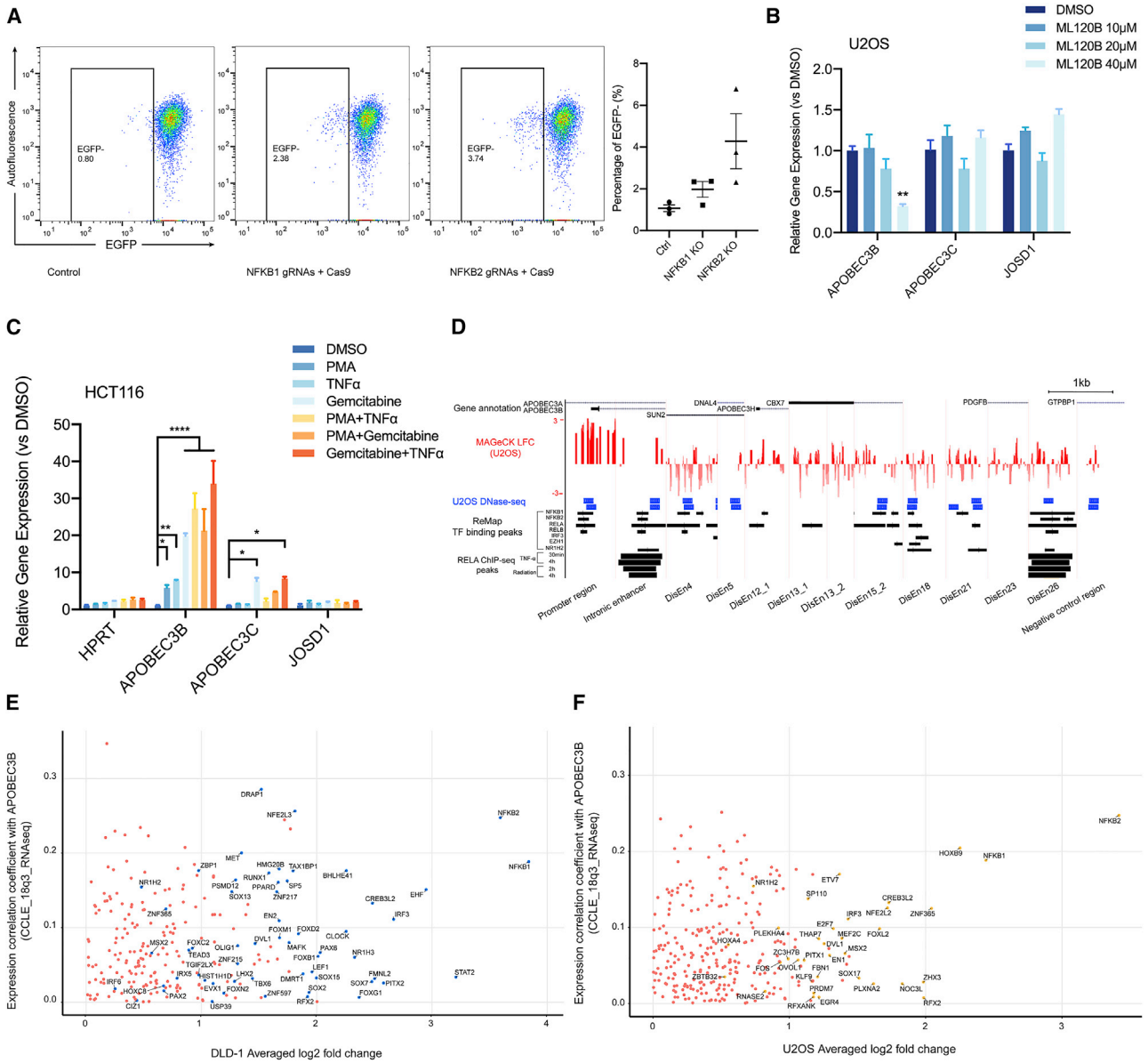


Figure 5. Validation of the Potent Role of NF-κB in Activating APOBE3B Gene Expression

(A) Representative FACS plots of APOBEC3B reporter intensity in U2OS cells with Cas9-mediated non-targeting controls, NFKB1-targeting gRNAs, or NFKB2-targeting gRNAs. A pool of two gRNAs was used for each targeting. Right panel is a bar plot of quantification for each targeting experiment.

(B) qRT-PCR quantifies the APOBEC3B mRNA expression levels in the U2OS cell line in response to different doses of IKKβ-selective inhibitor ML-120B compared with DMSO control.

(C) qRT-PCR quantifies the relative APOBEC3B mRNA expression levels in the HCT116 cell line in response to different upstream activators of NF-κB signaling compared with DMSO control: PMA (PKC activator, 200 ng/mL), TNF-α (pro-inflammatory cytokine, 200 ng/mL), gemcitabine (DNA synthesis inhibitor, 2 μM), and combinatorial treatments.

(D) A multi-region view of the validated enhancer regions with the candidate TF binding peaks. Plots showing the following in track order from top to bottom: gene annotation; MAGeCK LFC enrichment of *cis*-targeting RNAs in proximity from U2OS cells; DNase-seq peak generated from U2OS cell line; ChIP-seq peak datasets of candidate TFs from ReMap database, including NFKB1, NFKB2, RELA, RELB, IRF3, EZH1, and NR1H2; RELA ChIP-seq peaks from U2OS cells with TNF-α or radiation-mediated NF-κB activation (Janus et al., 2018).

(E and F) Dot plot showing the comparison between TF-APOBEC3B expression correlation coefficients and their respective MAGeCK enrichment LFC in U2OS and DLD-1 cell lines. Dots highlighted indicate the candidate TFs derived from each cell line: U2OS (yellow) and DLD-1 (blue).

In all bar plots, data are shown as mean ± SEM. *p < 0.05, **p < 0.01, ****p < 0.0001. (B) Statistics analysis is conducted by one-way ANOVA, with $Q = 5\%$, n = 3. (C) Statistics analysis is conducted between each gene profiling of individual drug treatment versus DMSO control by two-way ANOVA, using the two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli, with $Q = 5\%$, n = 3.

See also Figures S3 and S4.

other related TFs from that pathway are more strongly correlated (such as ETV4 and CREB3) (Table S8). Thus, in *trans* screening effectively identifies TFs and signaling pathways required for robust APOBEC3B expression.

The in *trans* screening platform does not address whether regulation occurs directly at the APOBEC3B locus or indirectly; thus, we examined published NF-κB TF ChIP-seq at the *cis*-regulatory loci we identified through in *cis* CRISPR-Cas9 screening profiles (Chèneby et al., 2018; Ibarra et al., 2016; Janus et al., 2018; Zhao et al., 2018) (Figures 5D and S4E). Consistent with ReMap atlas and previous ChIP-seq profiling upon NF-κB activation, we observed TF binding peaks (NFKB1, NFKB2, RELA, and RELB) at most of the validated *cis*-regulatory regions, including the APOBEC3B promoter upstream, intronic enhancer, DisEn4, DisEn12, DisEn15_2, DisEn18, DisEn21, and DisEn26, but not at DisEn5 or in the negative control regions. Overall, these data demonstrate direct involvement of the NF-κB-mediated inflammatory cascade in regulating APOBEC3B gene activation through regions identified in our in *cis* screening.

Tiling STARR-Seq Identifies AP-1 and NF-κB as the Dominant Regulators of Key Intronic and Distal APOBEC3B Enhancers

To more systematically connect APOBEC3B in *trans* regulators with their *cis*-regulatory target sites, we devised a tiling STARR-seq strategy to fine-map the active TF binding sites within key *cis*-regulatory regions identified in our CRISPR-Cas9 screen (Arnold et al., 2013) (Figure 6A). In total, from 91 prioritized regions, as well as two negative control regions, we designed a collection of 150-bp oligos containing wild-type sequences tiling each candidate region with 30-bp gaps between adjacent tiles (Figure S5A). We further created five mutant versions of each oligo by individually deleting adjacent 30-bp segments. Collectively, the tiling STARR-seq library contained 2,000 oligos with wild-type sequences and 8,094 oligos with sequences containing internal deletions, each represented by unique 15-bp barcodes (Table S6). We employed an optimized STARR-seq strategy that limits cellular transfection-induced artifacts (STAR Methods) (Muerdter et al., 2018).

We performed STARR-seq in U2OS and DLD-1 cells, which express APOBEC3B strongly, and weakly expressing HCT116 cells, each in two biological replicates. The vast majority (∼91%) of oligos were represented by RNA transcripts with five or more different barcodes, and we observed an average of ∼80 barcodes per oligo in DLD-1 cells and ∼40 barcodes per oligo in U2OS and HCT116 cells (Figure S6A). The STARR-seq activity profiles were highly similar between replicates within cell lines (Pearson's r: ∼0.85–0.95) and between cell lines (Pearson's r: ∼0.45–0.67), indicating reproducible patterns of enhancer activity (Figures S5B and S6B; Table S9). We observed no significant peak enrichment from either negative control region in any cell line (Figure 6B, shaded in gray). In the APOBEC3B gene locus, we observed the strongest enrichment of oligos tiling the intronic enhancer regions in both U2OS and DLD-1 cell lines, but not in HCT116 cells (Figures 6B and S6C), in accord with their respective APOBEC3B expression levels. Oligos tiling distal enhancer candidates also elicited different levels of enrichment, with particularly strong enrichment in DisEn11 and DisEn26, whose knockout was among the most deleterious to APOBEC3B expression within distal enhancers (Figures 3A and S6C).

To pinpoint putative TF motifs driving expression from the top-enriched STARR-seq regions, we developed a computational analysis pipeline to identify enriched wild-type oligos in which the deletion of one or two contiguous 30-bp blocks reproducibly disrupted activity (STAR Methods). The analysis pipeline incorporates a sequence scanning setup for known motifs from JASPAR database. Using a database of 572 well-characterized motif position weight matrices, we scan for motif sequences occurring more frequently than expected by chance in a simulated random sequence matching the observed nucleotide distribution of the APOBEC3B region. Strikingly, the 30-bp regions whose deletion most strongly disrupts oligo activity are highly enriched in strong predicted NF-κB and AP-1 TF binding motifs (Figure 6C). ChIP-seq datasets at these sites from ReMap Atlas confirmed the frequent co-occurrence of NF-κB and AP-1 binding at these regulatory regions (Figures 6D and S6D) (Chèneby et al., 2018). For instance, in both intronic enhancer and DisEn26 regions, NF-κB and AP-1 binding motifs were identified in close proximity (Figure 6E). The AP-1 TF family is known to cooperate with NF-κB in triggering pro-inflammatory responses through MAPK cascades (Stein et al., 1993; Zhu et al., 2001), and we found that the expression levels of several AP-1 TFs were strongly correlated with APOBEC3B expression in cancer cells, including FOSL1, which is the single most correlated TF with APOBEC3B expression (Figure S4C). We note that, although the in *trans* screening did identify FOS as a weak candidate regulator in U2OS cells (Figure 5F), the large number of partially redundant AP-1 family members may have hampered our ability to identify the role of AP-1 through gene knockout. Thus, the co-occurrence of AP-1 binding sites alongside NF-κB sites at required APOBEC3B regulatory regions suggests cooperative interactions between these two pathways in regulating APOBEC3B expression.

In addition, among other enriched motifs, we identified a nuclear hormone receptor (NHR) response element at DisEn11 whose deletion ablated its enhancer activity (Figures 6B and 6C). Although we cannot directly link binding of NHR TFs to this site, our *in trans* screen identified two NHRs, NR1H2 and NR1H3, whose mutation and small-molecule inhibition impair APOBEC3B activation, providing a possible binding factor at this site.

AP-1 and NF-κB TFs Must Both Bind to Activate APOBEC3B Expression at the Intronic and Distal Enhancers

To explore the possible cooperation of NF-κB and AP-1 in APOBEC3B expression, we treated U2OS cells with the NF-κB inhibitor ML120B, the AP-1 inhibitor SR11302, or the PI3K inhibitor Ly294002, which is known to block both NF-κB and AP-1 activity in cancer cells through inhibiting the common upstream PI3K/Akt signaling (Kikuchi et al., 2008; Ma et al., 2013; Reddy et al., 1997, 2000). Treatments of APOBEC3B-EGFP U2OS cells with Ly294002, ML120B, and SR11302 all induced significantly decreased reporter intensity by FACS (Figure 7A). We observed a more substantial reduction in double treatment of ML120B and SR11302 and also the Ly294002-treated cells than inhibition of either pathway alone, which we validated at the transcript level,

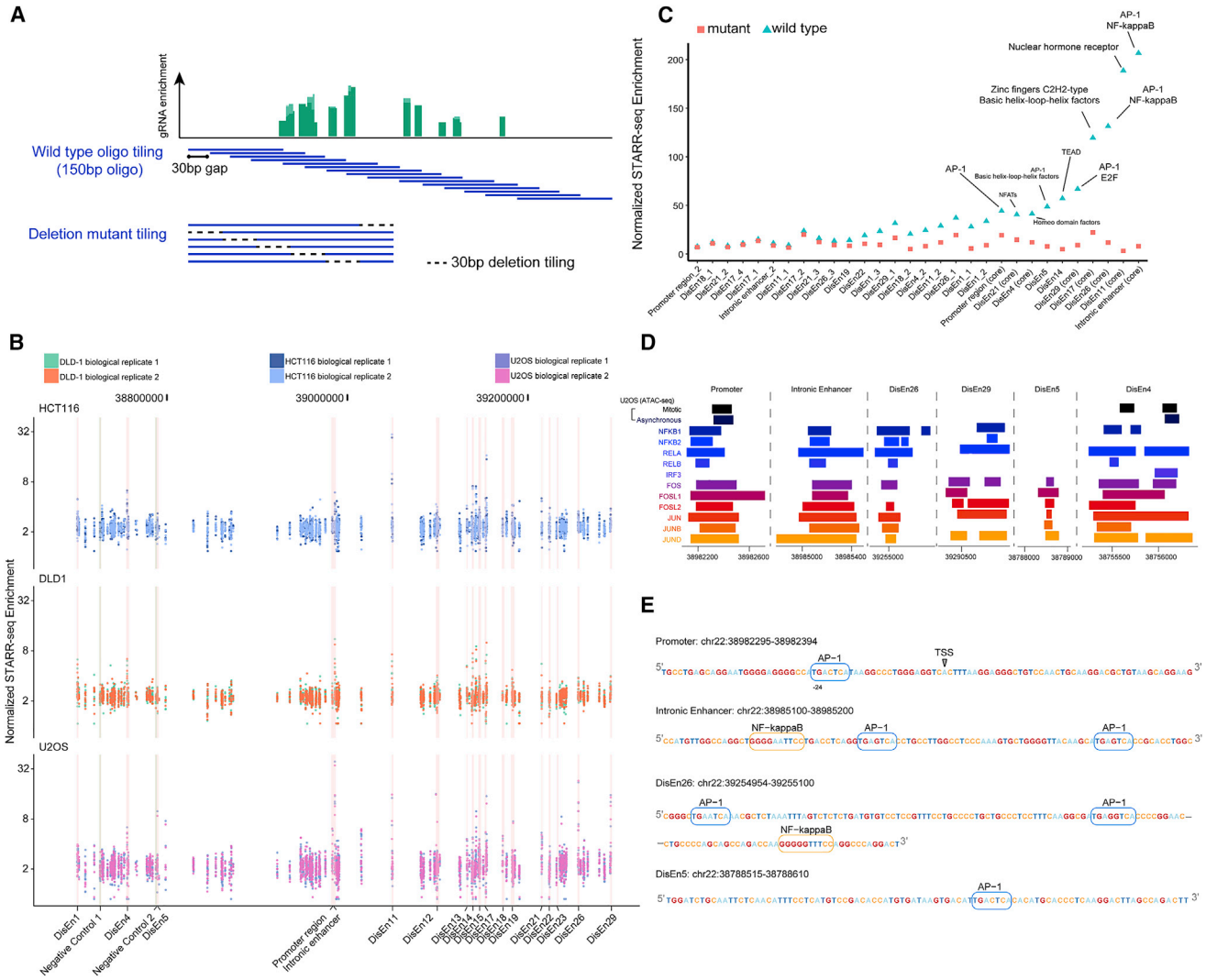


Figure 6. Tiling STARR-Seq Assay Identifies Both Intronic and Distal Enhancers with AP-1 and NF-κB Binding Motifs

(A) Schematic of tiling STARR-seq assay design.

(B) Dot plots showing the normalized enrichment of wild-type STARR-seq oligos from each biological replicate. Negative control regions are shaded in gray; candidate regions previously highlighted from in *cis* screening are shaded in red.

(C) Dot plots showing the normalized enrichment of the key 30-bp regions from STARR-seq. Annotation of the top-scored putative JASPAR motifs were included for regions previously validated with CRISPR screening and knockout validation.

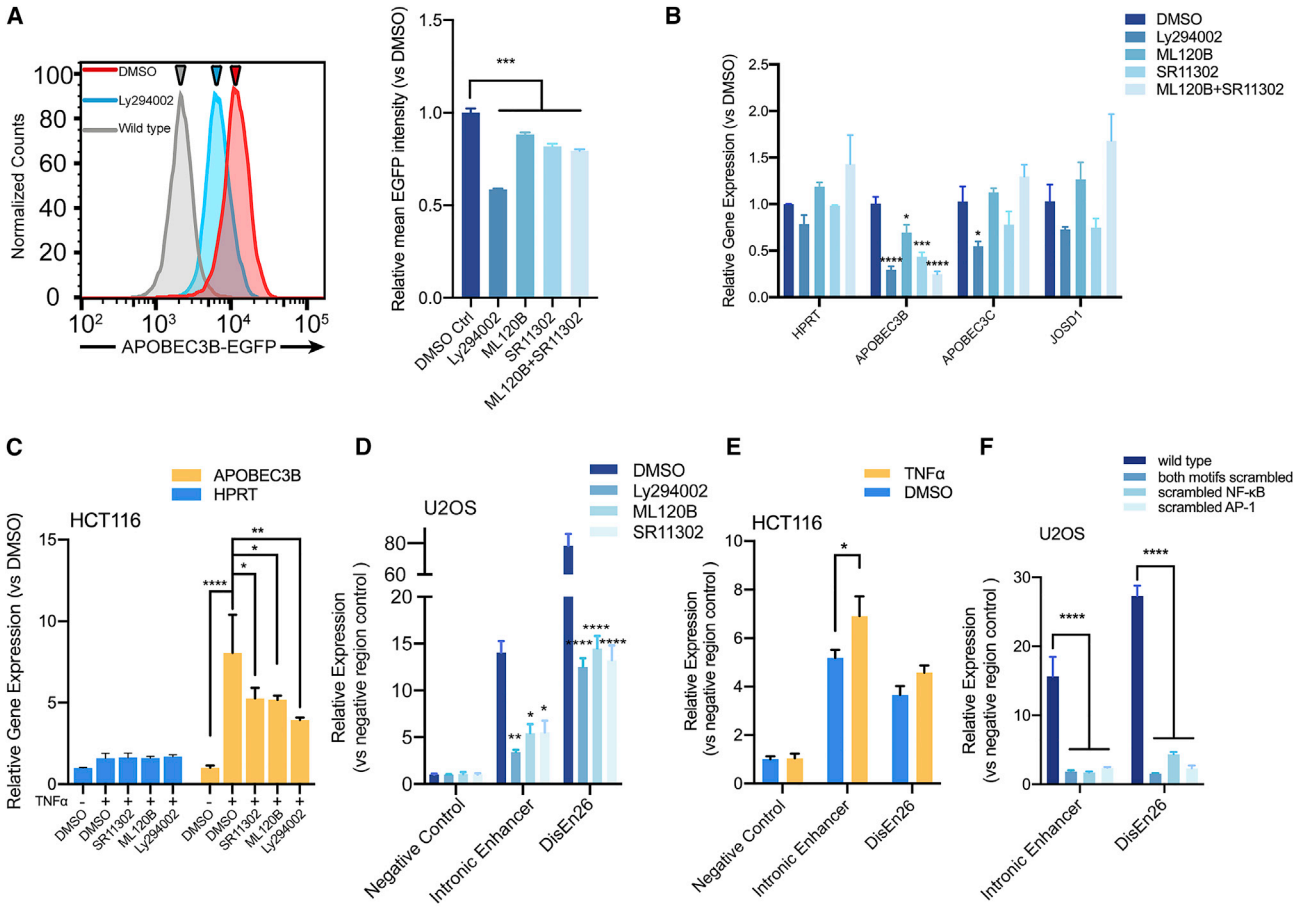
(D) TF occupancy in APOBEC3B promoter region and intronic enhancer and distal enhancer regions. ATAC-seq generated from U2OS cell line from previous literature was plotted to visualize open chromatin (Oomen et al., 2019). ChIP-seq peaks from the ReMap database were plotted within the STARR-seq window with the key 30-bp regions, including TFs from the NF-κB gene family (NFKB1, NFKB2, RELA, RELB) and IRF3 and TFs from the AP-1 gene family (FOS, FOSL1, FOSL2, JUN, JUNB, JUND).

(E) Diagram showing the sequences of NF-κB and AP-1 putative motifs in APOBEC3B promoter, intronic enhancer, and one example of the distal enhancer regions: DisEn26.

See also Figures S5 and S6. The complete chromosome coordinates of the regions highlighted in this figure can be found in Table S3.

indicating a combinatorial effect between NF-κB and AP-1 (Figure 7B). The neighboring expressed gene JOSD1 was not significantly affected by the drug treatments. APOBEC3C transcription was attenuated by Ly294002 and SR11302, but not by ML120B, which is likely due to the fact that the APOBEC3C promoter upstream contains an AP-1 binding site (data not shown).

We also asked whether any other signaling pathways we found to contribute to APOBEC3B expression were independent of AP-1 or PI3K blockage (Figure S7A). In combinatorial treatments with the AP-1 inhibitor, IRF signaling inhibition (MRT67307) more robustly inhibits APOBEC3B expression. A mild combinatorial effect of other inhibitors with AP-1 inhibition was also observed. However, none of these compounds reduced APOBEC3B expression further in combination with PI3K inhibition, suggesting that inhibition of the PI3K pathway predominantly blocks the integrated signaling inputs from all



(legend on next page)

pathways we identified to contribute to APOBEC3B expression.

We then asked whether activation of NF-κB and AP-1 signaling is sufficient to induce APOBEC3B expression in HCT116 cells. We treated HCT116 cells with TNF-α, followed by co-treatment with different inhibitors. We found that both NF-κB and AP-1 inhibitors led to significant reduction in APOBEC3B gene activation, and the PI3K inhibitor Ly294002 caused slightly stronger inhibition (Figure 7C). Similarly, PMA-induced APOBEC3B upregulation in HCT116 cells was inhibited in response to PI3K inhibitor (Figure S7B), but APOBEC3B upregulation by chemotherapeutic agents hydroxyurea and gemcitabine was unaffected by PI3K inhibition. These results suggest that a more complex signaling crosstalk is involved in cytotoxicity-induced APOBEC3B upregulation, consistent with previous findings (Kanu et al., 2016).

To address whether NF-κB and AP-1 act directly on the *cis*-regulatory regions identified using tiling STARR-seq, we first profiled the enhancer activity of *cis*-regulatory regions in response to NF-κB activation in HCT116 cells by performing tiled STARR-seq in the presence of the NF-κB-activating drug gemcitabine. We observed modest intronic and distal enhancer activity (DisEn14 and 12) in untreated HCT-116 cells that was significantly increased upon gemcitabine treatment (Figure S7D, top 10 regions are shown). Among these regions, the intronic enhancer showed the strongest induction upon gemcitabine treatment at three overlapping wild-type STARR-seq oligos, and deletion of a 30-bp region containing both NF-κB and AP1 binding sites eliminated gemcitabine activation from these oligos (Figure S7E). This region drives strong activity in U2OS and DLD1 cell lines as well (Figure 6C). Overall, this experiment links gemcitabine treatment with NF-κB and AP-1 *cis*-regulatory modules.

We next sought to dissect NF-κB and AP-1 pathways and their respective binding motifs in regulating APOBEC3B activation. We found that, in U2OS cells, enhancer activity of NF-κB and AP-1 motif-containing regions within the intronic enhancer and DisEn26 was significantly suppressed by blocking either NF-κB or AP-1 pathways, and Ly294002 again induced a slightly stronger reduction in the activity of both enhancer regions (Figure 7D). These same two regions displayed activated expression in HCT116 cells treated with TNF-α, PMA, hydroxyurea, or gemcitabine, with especially robust activation at the intronic enhancer region (Figures 7E and S7C). To determine whether the NF-κB or AP-1 motifs are indeed responsible for activating gene expression in these *cis*-regulatory regions, we created three mutant versions each of the intronic enhancer and DisEn26 STARR-seq oligos with either or both of the motifs scrambled. Interestingly, the destruction of either motif alone nearly completely extinguished expression from these oligos (Figure 7F), indicating a conditional logic in which NF-κB and AP-1 must both bind to APOBEC3B enhancers to induce expression. This conditional logic was also identified from STARR-seq oligos in which one or the other motif was deleted (Figure S7F). Previous studies have shown that FOS/JUN can physically interact with NF-κB/p65, promoting DNA binding and transactivation (Stein et al., 1993; Thomas et al., 1997). Our data provide a functional demonstration of this cooperative relationship, showing that NF-κB and AP-1 must both bind to adjacent motifs in several enhancer regions to induce APOBEC3B gene expression.

DISCUSSION

In this study, we employ HIGAN, a suite of high-throughput pooled screens (consisting of tiling *cis*-regulatory screening, *trans*-regulatory TF screening, and mutant tiled STARR-seq) that connect in *trans* TF regulators with their in *cis* binding sites to provide in-depth insight into the regulation of the cytidine deaminase APOBEC3B, whose upregulation contributes to cancer progression. Altogether, through connections forged by HIGAN, we show that inflammatory cytokines (canonical NF-κB), PKC activation (alternative NF-κB), and intrinsic genotoxicity (atypical NF-κB) converge through cooperative interactions between NF-κB and AP-1 TFs acting at several *cis*-regulatory regions to induce aberrant activation of the oncogene APOBEC3B (Figure 7G).

Figure 7. AP-1 and NF-κB Orchestrate APOBEC3B Activation Directly through Its Intronic and Distal Regulatory Elements

(A) Representative FACS plots of APOBEC3B reporter intensity in U2OS cells upon drug treatment (left). Bar plot showing the quantification of the relative mean EGFP intensity of the reporter cells upon drug treatment (right): Ly294002 (PI3K inhibitor, 40 μM), SR11302 (AP-1 inhibitor, 40 μM), and ML120B (IKKβ-selective inhibitor, 40 μM) for 48 h.

(B) qRT-PCR quantification of APOBEC3B inhibition upon treatments of NF-κB and AP-1 inhibitors in the U2OS cell line for 24 h.

(C) qRT-PCR quantification depicting the ablation of TNF-α-induced APOBEC3B activation by NF-κB and AP-1 inhibitors in the HCT116 cell line: TNF-α (200 ng/mL), SR11302 (2 μM), ML120B (40 μM), and Ly294002 (20 μM).

(D) qRT-PCR quantification of the relative enrichment of STARR-seq constructs containing negative control region, intronic enhancer, and DisEn26 in response to NF-κB and AP-1 inhibitors in the U2OS cell line: Ly294002 (20 μM), ML120B (40 μM), and SR11302 (40 μM).

(E) qRT-PCR analysis of the relative enrichment of STARR-seq constructs containing negative control region, intronic enhancer, and DisEn26 in response to TNF-α (200 ng/mL) in the HCT116 cell line.

(F) qRT-PCR quantification of the relative enrichment of STARR-seq constructs containing wild-type intronic enhancer and DisEn26 sequences or sequences with scrambled TF binding sites in the U2OS cell line.

(G) A schematic representation of the upstream signaling pathways triggering APOBEC3B expression activation through the binding of NF-κB and AP-1 to its *cis*-regulatory motifs. Pro-inflammatory cytokines, such as TNF-α, and PKC activator PMA signal via canonical NF-κB, PI3K, and MAPK pathways, leading to recruitment of NF-κB and AP-1 TF at APOBEC3B intronic and distal enhancers. Chemotherapeutic agents gemcitabine and hydroxyurea trigger genotoxicity, inducing atypical activation of NF-κB through ATM-NEMO interaction, in turn promoting APOBEC3B activation.

In all bar plots, data are shown as mean ± SEM. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. (A and B) Statistics analysis is conducted by one-way ANOVA, with $Q = 5\%$, $n = 3$. (C–F) Statistics analysis is conducted by two-way ANOVA, using multiple comparisons with two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli, with $Q = 5\%$, $n = 3$ (C, E, and F); $n = 4$ (D). Akt, Protein Kinase B; IKKα, IκB kinase α; IκB, NF-κB inhibitor; IKKβ, IκB kinase β; NEMO, IκB kinase γ; PI3K, phosphatidylinositol 3-kinase. See also Figures S6 and S7.

By pairing a broad CRISPR-Cas9-mediated survey of required regulatory regions with fine-mapping and functional interrogation of regions down to 30-bp resolution using STARR-seq, we were able to identify and confirm functional regulatory motifs in high resolution. Neither assay alone could have yielded such insights, because it is difficult to fine-map and assign regulatory function with CRISPR-Cas9 screening while the lack of native genomic context of plasmid-based STARR-seq assays prevents attribution of *in situ* function to identified regions. By further assessing the requirement of every human TF for APOBEC3B expression and measuring the expression correlation between these TFs and APOBEC3B across thousands of cancer cell lines, we were able to connect TF networks with their native *cis*-regulatory binding sites, clarifying the direct regulatory interactions governing APOBEC3B expression. Putting all of the three functional genomics assays together as an integrated research pipeline, we are able to trace coherent threads in the regulatory logic of APOBEC3B, for example, tracing pro-inflammatory extracellular cues such as TNF-α through to TFs such as NFKB1 and NFKB2 that act cooperatively with AP-1 TFs at two key *cis*-regulatory regions that are required for APOBEC3B expression in two cancer cell lines.

Performing such multi-modal genomics compensates for weaknesses in individual approaches. For example, delivering dCas9-KRAB to intronic regions could impede the transcriptional process, complicating discovery of intronic enhancer elements. STARR-seq thus provides an orthogonal approach to verify the enhancer activity of intronic regions, in this case providing strong support for our identification that APOBEC3B expression is driven by a strong intronic enhancer. In contrast, the in *trans* screening largely misses the importance of AP-1 in APOBEC3B regulation (FOS knockout weakly downregulates APOBEC3B in U2OS cells), presumably because of redundancies in the FOS and JUN families. Because AP-1 binding sites are highly enriched in essential 30-bp STARR-seq tiles and the expression of FOSL1 is the single most correlated TF with APOBEC3B expression, however, we were able to recognize and validate the key role of AP-1 in APOBEC3B expression using a pan-AP-1 small-molecule inhibitor. The STARR-seq assay employed small-molecule inhibitors to block potential false-positive signals triggered by type I IFN induction during cellular transfection, preventing us from assessing the role of IFN-related TFs in this assay. Nevertheless, the in *trans* screen allowed us to identify IRF3 and other IFN-related TFs (IRF5, IRF6, and STAT2 in DLD-1 cells) as APOBEC3B regulators, which we validated with a selective pan-IRF inhibitor MRT67307.

Our work also demonstrates the challenge in building accurate, complete gene-regulatory networks. We are able to confirm statistically significant roles for nine *cis*-regulatory regions using various validation methods, and using STARR-seq, we confirm the activity of >10 separable 30-bp regions within these regions. We implicate dozens of TFs in regulating APOBEC3B (many of which are involved in only one of the two cell lines we test), and by manipulating signaling pathways that use these TFs, we show that at least six signaling pathways play a role in promoting APOBEC3B expression. Moreover, within regulatory regions, there are complex relationships between TFs at individual binding sites, because we show that the APOBEC3B intronic enhancer and a distal enhancer (DisEn26) require temporal synchronization of NF-κB and AP-1 inputs to function. Thus, although we can provide distilled explanations such as that a number of inflammatory signaling pathways contribute in stronger (NF-κB/AP-1) or weaker (IRF3, Nrf2, CREB) ways to APOBEC3B expression, the reality is decidedly more nuanced. APOBEC3B does not possess an abnormally large number of adjacent epigenetically marked *cis*-regulatory regions, so it is likely that most genes have evolved similarly multifaceted regulatory architectures.

In spite of this complexity, with HIGAN, we demonstrate that the most salient APOBEC3B regulatory mechanism in U2OS and DLD-1 cells is a requirement for the simultaneous activity of NF-κB and AP-1 at two of the strongest enhancers, neither of which had been recognized previously as key APOBEC3B regulatory regions (Chou et al., 2017; Mori et al., 2015). A few previous studies have also provided evidence of the interdependency of NF-κB and AP-1 in regulating specific cytokine expression (Thomas et al., 1997; Yasumoto et al., 1992). A recent study showed that the majority (89%) of genomic NF-κB binding sites overlap with AP-1 during breast cancer cell transformation, but there are a large number of AP-1 binding sites without nearby NF-κB binding (Ji et al., 2019). This result is in line with evidence that AP-1 TFs act as pioneer factors, opening chromatin to enable the binding of other TFs at otherwise inaccessible adjacent binding sites (Biddie et al., 2011; Vierbuchen et al., 2017). Thus, we posit that AP-1 binding may act to recruit NF-κB through opening chromatin at NF-κB motifs, enabling the strong transcriptional enhancing activity of NF-κB to activate an inflammation-related gene expression program. This requirement for joint activity of NF-κB and AP-1 at dominant APOBEC3B enhancers may serve to prevent activation of APOBEC3B under conditions of transient pro-inflammatory signaling, reserving the activation of this potent yet dangerous anti-viral protein to cases of severe or prolonged inflammation in which both pathways are activated. The identification of roles for other inflammatory signaling pathways in regulating APOBEC3B, although with minor impacts, provides further explanation for the differential APOBEC3B levels in different tumors, suggesting that robust APOBEC3B upregulation occurs when there is a confluence of inflammatory triggers. For instance, in cancer cell lines with high NFKB2 expression but low APOBEC3B expression, we indeed observe significantly lower expression of a number of pro-inflammatory TFs that we have implicated in APOBEC3B activation, including RELA, FOSL1, IRF3, NFE2L3, CREB3L2, and NR1H2 (Figure S7G).

There are compelling studies to motivate the therapeutic upregulation (Driscoll et al., 2020; Wang et al., 2018) or repression (Law et al., 2016; Sieuwerts et al., 2017) of APOBEC3B as part of cancer therapy. Our study provides valuable information to guide such efforts. For instance, we show that the US Food and Drug Administration (FDA)-approved chemotherapy agents gemcitabine and hydroxyurea dramatically increase APOBEC3B expression, presumably leading to genomic hypermutation and neo-antigen creation, and thus suggesting that these agents might be particularly suited to sensitizing tumor cells for follow-up checkpoint immunotherapy. Meanwhile, the first recently FDA-approved PI3K inhibitor Piqray (alpelisib) has shown

significant improvement of outcomes in a clinical trial for hormone receptor (HR)-positive breast cancer therapy (André et al., 2019). We show that PI3K inhibition potently inhibits APOBEC3B expression, and thus we predict that Piqray treatment should dampen the metastatic potential of these cancers. In sum, the increased molecular understanding of APOBEC3B gene regulation provided by this study has the potential to influence the rationale for certain cancer treatment regimens, and the model for APOBEC3B gene regulation revealed by our HIGAN pipeline provides a paradigm for understanding gene-regulatory mechanisms in an unbiased manner.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
- METHOD DETAILS
  - gRNA library design
  - gRNA library preparation and transfection
  - Genomic DNA extraction
  - Sequencing library preparation
  - Flow cytometry cell sorting and analysis
  - Luciferase assay
  - Tiling STARR-seq plasmid library preparation
  - Deep sequencing to determine STARR-seq barcode-oligo dictionary
  - STARR-seq library transfection and transcript sequencing library preparation
  - Reverse transcription quantitative PCR (RT-qPCR)
  - Drug treatment for cancer cells
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Demultiplexing and read preprocessing
  - gRNA enrichment analysis (MAGeCK)
  - STARR-seq transcript enrichment analysis
  - STARR-seq tiling motif scanning
  - CCLE and TCGA datasets
  - Statistical analyses

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.108426.

ACKNOWLEDGMENTS

The authors thank Olga Krabbe and the Hubrecht Institute flow cytometry core facility for technical assistance. The authors thank Utrecht Sequencing Facility for providing sequencing service and data. Utrecht Sequencing Facility is subsidized by the University Medical Center Utrecht, Hubrecht Institute, and Utrecht University. M.W.S. is funded by a National Science Foundation (NSF) Graduate Research Fellowship. The authors acknowledge funding from National Institutes of Health (NIH) grants 1RO1HG008363 (to D.K.G.), 1R01HG008754 (to D.K.G.), 1R01NS109217 (to D.K.G.), 1K01DK101684 (to R.I.S.), and 1R21OD025309 (to R.I.S.); the Human Frontier Science Program; Brigham Research Institute; Harvard Stem Cell Institute; American Cancer Society (to R.I.S.); and Netherlands Organisation for Scientific Research (NWO) (to R.I.S. and L.L.).

AUTHOR CONTRIBUTIONS

L.L. and R.I.S. conceived the study, designed experiments, analyzed data, and wrote the manuscript. L.L., B.H., and M.W.S. designed and performed bioinformatics analysis. D.K. performed experiments. N.G. and D.K.G contributed to discussion and editing of the manuscript. All authors provided feedback on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 10, 2020
Revised: October 12, 2020
Accepted: November 2, 2020
Published: November 24, 2020

REFERENCES

Abraham, B.J., Hnisz, D., Weintraub, A.S., Kwiatkowski, N., Li, C.H., Li, Z., Weichert-Leahey, N., Rahman, S., Liu, Y., Etchin, J., et al. (2017). Small genomic insertions form enhancers that misregulate oncogenes. Nat. Commun. *8*, 14385.

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell *167*, 1867–1882.e21.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

André, F., Ciruelos, E., Rubovszky, G., Campone, M., Loibl, S., Rugo, H.S., Iwata, H., Conte, P., Mayer, I.A., Kaufman, B., et al.; SOLAR-1 Study Group (2019). Alpelisib for *PIK3CA*-Mutated, Hormone Receptor-Positive Advanced Breast Cancer. N. Engl. J. Med. *380*, 1929–1940.

Arbab, M., Srinivasan, S., Hashimoto, T., Geijsen, N., and Sherwood, R.I. (2015). Cloning-free CRISPR. Stem Cell Reports *5*, 908–917.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, L.M., Rath, M., and Stark, A. (2013). Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. Science *339*, 1074–1077.

Arora, S., Bhardwaj, A., Singh, S., Srivastava, S.K., McClellan, S., Nirodi, C.S., Piazza, G.A., Grizzle, W.E., Owen, L.B., and Singh, A.P. (2013). An undesired effect of chemotherapy: gemcitabine promotes pancreatic cancer cell invasiveness through reactive oxygen species-dependent, nuclear factor κB- and hypoxia-inducible factor 1α-mediated up-regulation of CXCR4. J. Biol. Chem. *288*, 21197–21207.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

Biddie, S.C., John, S., Sabo, P.J., Thurman, R.E., Johnson, T.A., Schiltz, R.L., Miranda, T.B., Sung, M.-H., Trump, S., Lightman, S.L., et al. (2011). Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. Mol. Cell *43*, 145–155.

Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., et al. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature *527*, 192–197.

Chèneby, J., Gheorghe, M., Artufel, M., Mathelier, A., and Ballester, B. (2018). ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. Nucleic Acids Res. *46* (*D1*), D267–D275.

Chou, W.-C., Chen, W.-T., Hsiung, C.-N., Hu, L.-Y., Yu, J.-C., Hsu, H.-M., and Shen, C.-Y. (2017). B-Myb Induces APOBEC3B Expression Leading to Somatic Mutation in Multiple Cancers. Sci. Rep. *7*, 44089.

De Bruin, E.C., McGranahan, N., Mitter, R., Salm, M., Wedge, D.C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A.J., et al. (2014). Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science *346*, 251–256.

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K.C., Huang, H., Liu, T., Marina, R.J., et al. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. Nat. Methods *14*, 629–635.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853–1866.e17.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

Driscoll, C.B., Schuelke, M.R., Kottke, T., Thompson, J.M., Wongthida, P., Tonne, J.M., Huff, A.L., Miller, A., Shim, K.G., Molan, A., et al. (2020). APOBEC3B-mediated corruption of the tumor cell immunopeptidome induces heteroclitic neoepitopes for cancer immunotherapy. Nat. Commun. *11*, 790.

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) *2017*, bax028.

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer–promoter connections with CRISPR interference. Science *354*, 769–773.

Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. Nat. Genet. *51*, 1664–1669.

Gasperini, M., Findlay, G.M., McKenna, A., Milbank, J.H., Lee, C., Zhang, M.D., Cusanovich, D.A., and Shendure, J. (2017). CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. Am. J. Hum. Genet. *101*, 192–205.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Hernández-Vargas, H., Rodríguez-Pinilla, S.M., Julián-Tendero, M., Sánchez-Rovira, P., Cuevas, C., Antón, A., Ríos, M.J., Palacios, J., and Moreno-Bueno, G. (2007). Gene expression profiling of breast cancer cells in response to gemcitabine: NF-kappaB pathway activation as a potential mechanism of resistance. Breast Cancer Res. Treat. *102*, 157–172.

Holden, N.S., Squires, P.E., Kaur, M., Bland, R., Jones, C.E., and Newton, R. (2008). Phorbol ester-stimulated NF-kappaB-dependent transcription: roles for isoforms of novel protein kinase C. Cell. Signal. *20*, 1338–1348.

Ibarra, A., Benner, C., Tyagi, S., Cool, J., and Hetzer, M.W. (2016). Nucleoporin-mediated regulation of cell identity genes. Genes Dev. *30*, 2253–2258.

Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., Salame, T.M., Tanay, A., van Oudenaarden, A., and Amit, I. (2016). Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell *167*, 1883–1896.e15.

Janus, P., Szołtysek, K., Zając, G., Stokowy, T., Walaszczyk, A., Widłak, W., Wojtaś, B., Gielniewski, B., Iwanaszko, M., Braun, R., et al. (2018). Pro-inflammatory cytokine and high doses of ionizing radiation have similar effects on the expression of NF-kappaB-dependent genes. Cell. Signal. *46*, 23–31.

Ji, Z., He, L., Regev, A., and Struhl, K. (2019). Inflammatory regulatory network mediated by the joint action of NF-kB, STAT3, and AP-1 factors is involved in many human cancers. Proc. Natl. Acad. Sci. USA *116*, 9453–9462.

Kanu, N., Cerone, M.A., Goh, G., Zalmas, L.-P., Bartkova, J., Dietzen, M., McGranahan, N., Rogers, R., Law, E.K., Gromova, I., et al. (2016). DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. Genome Biol. *17*, 185.

Kikuchi, J., Kinoshita, I., Shimizu, Y., Oizumi, S., Nishimura, M., Birrer, M.J., and Dosaka-Akita, H. (2008). Simultaneous blockade of AP-1 and phosphatidylinositol 3-kinase pathway in non-small cell lung cancer cells. Br. J. Cancer *99*, 2013–2019.

Klann, T.S., Black, J.B., Chellappan, M., Safi, A., Song, L., Hilton, I.B., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2017). CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. Nat. Biotechnol. *35*, 561–568.

Korkmaz, G., Lopes, R., Ugalde, A.P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R., and Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat. Biotechnol. *34*, 192–198.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665.

Law, E.K., Sieuwerts, A.M., LaPara, K., Leonard, B., Starrett, G.J., Molan, A.M., Temiz, N.A., Vogel, R.I., Meijer-van Gelder, M.E., Sweep, F.C.G.J., et al. (2016). The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. Sci. Adv. *2*, e1601737.

Leonard, B., Hart, S.N., Burns, M.B., Carpenter, M.A., Temiz, N.A., Rathore, A., Vogel, R.I., Nikas, J.B., Law, E.K., Brown, W.L., et al. (2013). APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. Cancer Res. *73*, 7222–7231.

Leonard, B., McCann, J.L., Starrett, G.J., Kosyakovsky, L., Luengas, E.M., Molan, A.M., Burns, M.B., McDougle, R.M., Parker, P.J., Brown, W.L., and Harris, R.S. (2015). The PKC/NF-κB signaling pathway induces APOBEC3B expression in multiple human cancers. Cancer Res. *75*, 4538–4547.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. *15*, 554.

Liu, Y., Cao, Z., Wang, Y., Guo, Y., Xu, P., Yuan, P., Liu, Z., He, Y., and Wei, W. (2018). Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. Nat. Biotechnol. *36*, 1203–1210.

Ma, C., Zuo, W., Wang, X., Wei, L., Guo, Q., and Song, X. (2013). Lapatinib inhibits the activation of NF-κB through reducing phosphorylation of IκB-α in breast cancer cells. Oncol. Rep. *29*, 812–818.

Maruyama, W., Shirakawa, K., Matsui, H., Matsumoto, T., Yamazaki, H., Sarca, A.D., Kazuma, Y., Kobayashi, M., Shindo, K., and Takaori-Kondo, A. (2016). Classical NF-κB pathway is responsible for APOBEC3B expression in cancer cells. Biochem. Biophys. Res. Commun. *478*, 1466–1471.

Mei, S., Qin, Q., Wu, Q., Sun, H., Zheng, R., Zang, C., Zhu, M., Wu, J., Shi, X., Taing, L., et al. (2017). Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. Nucleic Acids Res. *45* (*D1*), D658–D662.

Mori, S., Takeuchi, T., Ishii, Y., and Kukimoto, I. (2015). Identification of APOBEC3B promoter elements responsible for activation by human papillomavirus type 16 E6. Biochem. Biophys. Res. Commun. *460*, 555–560.

Muerdter, F., Boryń, Ł.M., Woodfin, A.R., Neumayr, C., Rath, M., Zabidi, M.A., Pagani, M., Haberle, V., Kazmar, T., Catarino, R.R., et al. (2018). Resolving

systematic errors in widely used enhancer activity assays in human cells. Nat. Methods *15*, 141–149.

Nagashima, K., Sasseville, V.G., Wen, D., Bielecki, A., Yang, H., Simpson, C., Grant, E., Hepperle, M., Harriman, G., Jaffee, B., et al. (2006). Rapid TNFR1-dependent lymphocyte depletion in vivo with a selective chemical inhibitor of IKKbeta. Blood *107*, 4266–4273.

Oomen, M.E., Hansen, A.S., Liu, Y., Darzacq, X., and Dekker, J. (2019). CTCF sites display cell cycle-dependent dynamics in factor binding and nucleosome positioning. Genome Res. *29*, 236–249.

Radzisheuskaya, A., Shlyueva, D., Müller, I., and Helin, K. (2016). Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. Nucleic Acids Res. *44*, e141.

Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M.D., Banerjee, B., Syed, T., Emons, B.J.M., Gifford, D.K., and Sherwood, R.I. (2016). High-throughput mapping of regulatory DNA. Nat. Biotechnol. *34*, 167–174.

Reddy, S.A.G., Huang, J.H., and Liao, W.S.-L. (1997). Phosphatidylinositol 3-kinase in interleukin 1 signaling. Physical interaction with the interleukin 1 receptor and requirement in NFkappaB and AP-1 activation. J. Biol. Chem. *272*, 29167–29173.

Reddy, S.A.G., Huang, J.H., and Liao, W.S.-L. (2000). Phosphatidylinositol 3-kinase as a mediator of TNF-induced NF-κ B activation. J. Immunol. *164*, 1355–1363.

Rokavec, M., Horst, D., and Hermeking, H. (2017). Cellular model of colon cancer progression reveals signatures of mRNAs, miRNA, lncRNAs, and epigenetic modifications associated with metastasis. Cancer Res. *77*, 1854–1867.

Sanjana, N.E., Wright, J., Zheng, K., Shalem, O., Fontanillas, P., Joung, J., Cheng, C., Regev, A., and Zhang, F. (2016). High-resolution interrogation of functional elements in the noncoding genome. Science *353*, 1545–1549.

Schütze, S., Wiegmann, K., Machleidt, T., and Krönke, M. (1995). TNF-induced activation of NF-κ B. Immunobiology *193*, 193–203.

Shen, M.W., Arbab, M., Hsu, J.Y., Worstell, D., Culbertson, S.J., Krabbe, O., Cassa, C.A., Liu, D.R., Gifford, D.K., and Sherwood, R.I. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. Nature *563*, 646–651.

Sher, F., Hossain, M., Seruggia, D., Schoonenberg, V.A.C., Yao, Q., Cifani, P., Dassama, L.M.K., Cole, M.A., Ren, C., Vinjamur, D.S., et al. (2019). Rational targeting of a NuRD subcomplex guided by comprehensive in situ mutagenesis. Nat. Genet. *51*, 1149–1159.

Sieuwerts, A.M., Schrijver, W.A.M.E., Dalm, S.U., de Weerd, V., Moelans, C.B., Ter Hoeve, N., van Diest, P.J., Martens, J.W.M., and van Deurzen, C.H.M. (2017). Progressive APOBEC3B mRNA expression in distant breast cancer metastases. PLoS ONE *12*, e0171343.

Stavrou, S., and Ross, S.R. (2015). APOBEC3 Proteins in Viral Immunity. J. Immunol. *195*, 4565–4570.

Stein, B., Baldwin, A.S., Jr., Ballard, D.W., Greene, W.C., Angel, P., and Herrlich, P. (1993). Cross-coupling of the NF-kappa B p65 and Fos/Jun transcription factors produces potentiated biological function. EMBO J. *12*, 3879–3891.

Supek, F., and Lehner, B. (2017). Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. Cell *170*, 534–547.e23.

Taylor, B.J., Nik-Zainal, S., Wu, Y.L., Stebbings, L.A., Raine, K., Campbell, P.J., Rada, C., Stratton, M.R., and Neuberger, M.S. (2013). DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. eLife *2*, e00534.

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods *12*, 1143–1149.

Thomas, R.S., Tymms, M.J., McKinlay, L.H., Shannon, M.F., Seth, A., and Kola, I. (1997). ETS1, NFkappaB and AP1 synergistically transactivate the human GM-CSF promoter. Oncogene *14*, 2845–2855.

Tilborghs, S., Corthouts, J., Verhoeven, Y., Arias, D., Rolfo, C., Trinh, X.B., and van Dam, P.A. (2017). The role of Nuclear Factor-kappa B signaling in human cervical cancer. Crit. Rev. Oncol. Hematol. *120*, 141–150.

Vierbuchen, T., Ling, E., Cowley, C.J., Couch, C.H., Wang, X., Harmin, D.A., Roberts, C.W.M., and Greenberg, M.E. (2017). AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. Mol. Cell *68*, 1067–1082.e12.

Walz, S., Lorenzin, F., Morton, J., Wiese, K.E., von Eyss, B., Herold, S., Rycak, L., Dumay-Odelot, H., Karim, S., Bartkuhn, M., et al. (2014). Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. Nature *511*, 483–487.

Wang, S., Jia, M., He, Z., and Liu, X.-S. (2018). APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer. Oncogene *37*, 3924–3936.

Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M.; Cancer Genome Atlas Research Network (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. *45*, 1113–1120.

Wen, D., Nong, Y., Morgan, J.G., Gangurde, P., Bielecki, A., Dasilva, J., Keaveney, M., Cheng, H., Fraser, C., Schopf, L., et al. (2006). A selective small molecule IkappaB Kinase β inhibitor blocks nuclear factor kappaB-mediated inflammatory responses in human fibroblast-like synoviocytes, chondrocytes, and mast cells. J. Pharmacol. Exp. Ther. *317*, 989–1001.

Yasumoto, K., Okamoto, S., Mukaida, N., Murakami, S., Mai, M., and Matsushima, K. (1992). Tumor necrosis factor alpha and interferon gamma synergistically induce interleukin 8 production in a human gastric cancer cell line through acting concurrently on AP-1 and NF-kB-like binding sites of the interleukin 8 gene. J. Biol. Chem. *267*, 22506–22511.

Zhao, M., Joy, J., Zhou, W., De, S., Wood, W.H., 3rd, Becker, K.G., Ji, H., and Sen, R. (2018). Transcriptional outcomes and kinetic patterning of gene expression in response to NF-κB activation. PLoS Biol. *16*, e2006347.

Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C.H., Brown, M., Zhang, X., Meyer, C.A., and Liu, X.S. (2019). Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res. *47* (*D1*), D729–D735.

Zhu, C., Gagnidze, K., Gemberling, J.H.M., and Plevy, S.E. (2001). Characterization of an activation protein-1-binding site in the murine interleukin-12 p40 promoter. Demonstration of novel functional elements by a reductionist approach. J. Biol. Chem. *276*, 18519–18528.

Zhu, S., Li, W., Liu, J., Chen, C.-H., Liao, Q., Xu, P., Xu, H., Xiao, T., Cao, Z., Peng, J., et al. (2016). Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. Nat. Biotechnol. *34*, 1279–1286.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial Strains		
One Shot TOP10 Chemically Competent *E. coli*	ThermoFisher	Cat#C404003
NEB® 5-alpha Electrocompetent *E. coli*	NEB	Cat#C2989K
Chemicals		
DPBS, no calcium, no magnesium	GIBCO	Cat#11875-093
RPMI 1640 Medium (ATCC Modification), HEPES, Liquid	GIBCO	Cat#A1049101
McCoy's 5A (Modified) Medium	GIBCO	Cat#26600023
Fetal Bovine Serum	Sigma-Aldrich	Cat#F7524
Opti-MEM I Reduced Serum Medium	GIBCO	Cat#11058021
Penicillin-Streptomycin (10,000 U/mL)	GIBCO	Cat#15140122
Lipofectamine® 3000	ThermoFisher	Cat#15212475
Hygromycin B (50mg/ml)	ThermoFisher	Cat#10687010
Blasticidin S HCl (10mg/mL)	ThermoFisher	Cat#A1113903
Trypsin-EDTA (0.25%), phenol red	ThermoFisher	Cat#25200072
Recombinant RNasin® RNase Inhibitor, 10,000 U	Promega	Cat#N2515
RQ1 RNase-Free DNase, 1,000 U	Promega	Cat#M6101
SuperScript III Reverse Transcriptase 10,000 U	ThermoFisher	Cat#18080085
ProtoScript® First Strand cDNA Synthesis Kit	NEB	Cat#E6300L
AgeI-HF	NEB	Cat#R3552L
SalI-HF	NEB	Cat#R3138L
BbsI-HF	NEB	Cat#R3539S
Gibson Assembly Master Mix	NEB	Cat#E2611L
MinElute PCR Purification Kit (50)	QIAGEN	Cat#28004
DNeasy Blood & Tissue Kit	QIAGEN	Cat#69506
Dynabeads Oligo(dT)25	ThermoFisher	Cat#61005
TRIzol Reagent	ThermoFisher	Cat#15596026
RNeasy Mini Kit (250)	QIAGEN	Cat#74106
QIAquick Gel Extraction Kit	QIAGEN	Cat#28706
NEBNext® High-Fidelity 2X PCR Master Mix	NEB	Cat#M0541L
NEBNext® Ultra II Q5® Master Mix	NEB	Cat#M0544L
Agilent High Sensitivity DNA Kit	Agilent	Cat#5067-4626
Qubit dsDNA HS Assay Kit	ThermoFisher	Cat# Q32851
GeneRuler 1 kb Plus DNA Ladder	ThermoFisher	Cat#10101240
Qubit dsDNA HS Assay Kit	ThermoFisher	Cat#Q32854
SYBR Safe DNA Gel Stain	ThermoFisher	Cat#S33102
BX-795 hydrochloride	Sigma-Aldrich	Cat#SML0694
Imidazolo-oxindole PKR inhibitor C16	Sigma-Aldrich	Cat#I9785
Recombinant human TNFα	InvivoGen	Cat#rcyc-htnfa
ML120B	Sigma-Aldrich	Cat#SML1174
SR11302	Cayman Chemical	Cat#16338
LY294002	Cayman Chemical	Cat#70920
Gemcitabine hydrochloride	Sigma-Aldrich	Cat#G6423
Hydroxyurea	Sigma-Aldrich	Cat#H8627

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CREB Inhibitor, 666-15	Sigma-Aldrich	Cat# 5383410001
GSK2033	Sigma-Aldrich	Cat# SML1617-5MG
ML-385	Cayman Chemical	Cat# 21114-5
U0126	Sigma-Aldrich	Cat# 662005-5MG
MRT67307	Sigma-Aldrich	Cat# 5063060001
Y-27632	Sigma-Aldrich	Cat#688000
NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1)	NEB	Cat#E7335L
DyNAmo Flash SYBR Green qPCR Kit	ThermoFisher	Cat#F-415L
Dual-Luciferase® Reporter Assay System	Promega	Cat#E1910
SYBR® Green I staining reagent DNA free	PanReac AppliChem	Cat#A8511.50625
iQ SYBR® Green Supermix	BioRad	Cat#1708887
Cell Line NucleofectorTM Kit V	Lonza	Cat#VCA-1003
Nuclease-Free Water (not DEPC-Treated)	ThermoFisher	Cat#AM9937
GlycoBlue Coprecipitant (15 mg/mL)	ThermoFisher	Cat#AM9515
PureLink® HiPure Plasmid Filter Maxiprep Kit	ThermoFisher	Cat#K210016
Deposited Data		
RNA-seq and ChIP-seq in U2OS cell line in response to TNFα and ionizing radiation	https://www.ncbi.nlm.nih.gov/gds	GSE110387
RNA-seq and ChIP-seq in lymphoma B cell line in reponse to TNFα	https://www.ncbi.nlm.nih.gov/gds	GSE117259
ChIP-seq peaks from ReMap 2018	http://remap.univ-amu.fr/download_page#remap2018tab	NA
DNase-seq peak generated from U2OS cell line	https://www.ncbi.nlm.nih.gov/gds	GSE87831
ATAC-seq generated from U2OS cell line	https://www.ncbi.nlm.nih.gov/gds	GSE121840
Code for processing *in-cis* tiling library and MaGeCK analysis	https://github.com/maxwshen/mera_tiling	NA
Recombinant DNA		
APOBEC3B *cis*-regulatory gRNA library	GenScript	Custom Design Table S5
Human transcription factor-wide gRNA library	GenScript	Custom Design Table S6
APOBEC3B tiling STARR-seq oligo library	Twist Bioscience	Custom Design Table S7
sgBbsI (p2Tol-U6-2xBbsI-sgRNA-HygR)	Addgene	Cat#71485
spCas9-BlastR (pCBhCas9-BlastR)	Addgene	Cat#71489
hSTARR-seq_ORI	Addgene	Cat#99296
pGL4.23[luc2/minP]	Promega	Cat#E841A
Software and Algorithms		
FlowJo software 10.4	FlowJo	https://www.flowjo.com/
Prism 8.2.0	GraphPad	https://www.graphpad.com/scientific-software/prism/
String 11.0	Reference	https://string-db.org/
MAGeCK	Reference	https://sourceforge.net/p/mageck/wiki/Home/
Rstudio	Rstudio	https://rstudio.com/
Experimental Models: Cell Lines		
U2OS	ATCC	Catalog Number: HTB-96
DLD-1	ATCC	Catalog Number: CCL-221
HCT116	ATCC	Catalog Number: CCL-247
Oligonucleotides		
qPCR primers	This paper	Table S1 Primer collection
APOBEC3B *in-cis* gRNA library	This paper	Table S4 APOBEC3B *in-cis* gRNA library

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TF *in-trans* gRNA library	This paper	Table S5 TF *in-trans* gRNA library
Tiling STARR-seq oligo design	This paper	Table S6 Tiling STARR-seq oligo design

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Richard Sherwood (rsherwood@rics.bwh.harvard.edu)

Materials availability

This study did not generate new unique materials.

Data and code availability

Raw data including CRISPRi fastq sequencing datasets, original gel pictures and bar plots presented in this study are available at Mendeley dataset deposit with https://doi.org/10.17632/b4wczz76gf.1; Code for processing *in-cis* tiling library and MAGeCK is available at: https://github.com/maxwshen/mera_tiling. gRNA oligo design, raw read counts for *in-cis* screen and processed MAGeCK analysis is available in Tables S4 and S7. gRNA oligo design, raw read counts for *in-trans* screen and processed MAGeCK analysis is available in Tables S5 and S8; Tiling STARR-seq library design and normalized expression is available in Tables S6 and S9.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

Human cell lines U2OS (ATCC® HTB-96), DLD-1 (ATCC® CCL-221) and HCT116 (ATCC® CCL-247) were obtained from ATCC and cultured according to protocols recommended by ATCC. U2OS and HCT116 cells were grown at 37°C in GIBCO McCoy's 5A (Modified) Medium (GIBCO 26600023) supplemented with 10% fetal bovine serum. DLD-1 cells were grown at 37°C in GIBCO RPMI 1640 Medium (ATCC Modification) (GIBCO A1049101) supplemented with 10% fetal bovine serum. Cells were regularly tested for mycoplasma. APOBEC3B C-terminal P2A-GFP fusion knockin lines were constructed using a previously published protocol with cell line Nucleofector Kit (Lonza) (Arbab et al., 2015).

METHOD DETAILS

gRNA library design

For *cis*- screening library, gRNAs were designed to tile −300 kb to +300 kb proximal region around the APOBEC3B promoter at 22q13.1 chromosomal region (Table S5). gRNAs were filtered if they contained quadruple homopolymers, resided in genomic repeat regions, or if its SpCas9 cutsite was 3 or fewer nucleotides to another SpCas9 gRNA's cutsite, or if its Azimuth score was below 0.35. If the number of gRNAs remaining was greater than 12,270, we removed gRNAs in ascending order of their distance to another gRNA in the library. Following these steps, gRNAs were designed to match the following properties:

Guide RNAs should have 19-20 bp of homology to the genome immediately preceding the NGG "PAM" sequence:

a. If the genome sequence is GNNNNNNNNNNNNNNNNNNN NGG (GN_{19}NGG), the guide RNA sequence should be GNNNNNNNNNNNNNNNNNNN (GN_{19})
b. If a is not satisfied but GNNNNNNNNNNNNNNNNNNN NGG (GN_{18}NGG) is satisfied, the guide RNA sequence should be GNNNNNNNNNNNNNNNNNN (GN_{18})
c. If a and b are not satisfied, the guide RNA sequence should be GNNNNNNNNNNNNNNNNNNN (GN_{20}) where the genomic sequence is NNNNNNNNNNNNNNNNNNNN NGG (N_{20}NGG) – it does not matter if the first G is in the genome.

For *trans*- screening library, 7210 gRNAs targeting 1803 human annotated transcription factors were collected from human CRISPR Brunello genome-wide knockout pooled libraries (Doench et al., 2016), and then a list of 100 non-target control gRNAs and 77 gRNAs targeting EGFP was also included in the *trans*- screening library (Table S6).

We purchased the gRNA oligos from CustomArray, Inc (GenScript, WA) with the following format:

Oligo Structure: ATATATCTTGTGGAAAGGACGAAACACC[GN_{18-20}]GTTTAAGAGCTATGCTGGAAACAGCATAGC

gRNA library preparation and transfection

Both *in-cis* and *in-trans* screening oligo pools were first PCR with the following primers to extend the homolog arm for cloning:

010415_gRNA60bpFw: TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG 010415_sgRNA_60bp_rv: GTTGATAACGGACTAGCCTTATTTAAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC

All primer sequences are listed in Table S1.
PCR conditions:

NEBNext® High-Fidelity 2X PCR Master Mix: 25uL
Pooled oligo library template: final concentration of 0.04ng/uL
−010415_gRNA_60bp_Fw (10uM): 2.5 uL
−010415_sgRNA_60bp_rv (10uM): 2.5 uL
DMSO: 1.5uL
RNase-free water: 17.5 uL

PCR was run with annealing temperature at 68 degrees for 20 cycles, with extension time for 1min. PCR product was then run on a 2% agarose gel and purified for the specific 137bp band, using the QIAquick gel extraction kit followed by PCR purification kit (-QIAGEN). Backbone plasmid p2TU6sg2XBbsIFE was digested with BbsI-HF (New England Bioscience) and gel purified. Gibson assembly was set up according to the reaction ratios below:

Purified linear vector: 330ng (0.083pmols for 6kb)
Purified PCR inserts:50ng (0.55pmols for 140bp, 6.6x more than vector (5-10x recommend))
2x Gibson assembly Master Mix: 10uL
H2O: fill up to 20uL

Reaction mixture was then incubated at 50 degrees for 1h, and then 4 degrees for Gibson assembly reaction. Assembled plasmids were purified by isopropanol precipitation with GlycoBlue Coprecipitant (Thermo Fisher) and resuspended in Nuclease-free water and transformed into NEB10beta (New England Biolabs) electrocompetent cells according to the manufacturer's directions. Following recovery, a small dilution series were plated to assess transformation efficiency and the remainder was grown in liquid culture in LB medium overnight at 37°C, followed by maxiprep with PureLink® HiPure Plasmid Filter Maxiprep Kit (Thermo Fisher). For stable Tol2 transposon plasmid integration, previous protocol was used with Lipofectamine 3000 (Thermo Fisher) (Shen et al., 2018). Selection with the respective agents at an empirically defined concentration (Blasticidin, Hygromycin) (Thermo Fisher) was performed starting 48 hours after transection and continuing for 1∼2 weeks.

Genomic DNA extraction

For genomic DNA isolation, cell cultures were first trypsinized and collected. Based on the estimated cell number, cells were then split into multiple reactions for further lysis and purification to keep the starting material below the maximum range of purification column, which was subjected to DNeasy Blood & Tissue Kit (QIAGEN) following the manufacturer's protocol. The concentration was then measured using a Nanodrop (Thermo Fisher).

Sequencing library preparation

To amplify the gRNA fragments integrated in the genomic DNA for Illumina high-throughput sequencing, we performed two-step PCR amplification using primers flanking the gRNA protospacers. For each sequencing library, 16ug of genomic DNA was used. PCR1 was performed using the primers with stagger barcodes to maintain sequence diversity across the flow-cell. PCR2 was performed to add full-length Illumina sequencing adapters using custom primers with index for multiplexing. All PCRs were performed using NEBNext polymerase (New England Bioscience). To avoid overamplification, between PCR1 and PCR2, we performed qPCR to gauge how much product we have and calculate the number of cycles of PCR2 to be performed. The pooled samples were sequenced using NextSeq (Illumina) at the Utrecht Sequencing Facility.

All primer sequences are listed in Table S3.

Flow cytometry cell sorting and analysis

FACS of screening was performed on BD FACSJazz cell sorter and BD Influx cell sorter (BD Bioscience). For cells that need to be collected back for culture, cells were sorted in complete cell culture medium with Pen-Strep (GIBCO) and ROCK inhibitor (Y-27632, final concentration of 10uM). For cells subjected to analysis, we performed FACS with analysis buffer that is comprised of 5mM EDTA, 25mM HEPES and 1% BSA in DPBS. FACS analysis was performed with CytoFLEX benchtop flow cytometer (Beckman), and analyzed with FlowJo software.

Luciferase assay

Selected regions were PCR amplified from genomic DNA of U2OS cell line, cloned into pGL4.23[luc2/minP] and verified by sequencing (Table S1 for genomic coordinates and primers). Individual constructs were tested by co-transfecting cells in a 96-well format at ~80% confluency with the respective firefly construct (50ng) and a Renilla control plasmid (10ng) using lipofectamine 3000 (Thermo Fisher). Using the Promega Dual Luciferase Assay kit (Promega), we measured luciferase activity at a microplate luminometer (Centro LB 960, Berthold) and normalized firefly luciferase to Renilla luciferase activity.

Tiling STARR-seq plasmid library preparation

Oligo pool for STARR-seq was purchased from Twist Bioscience (CA) (Table S7). Oligo pool was designed in 150bp format with 25bp homolog arm at both 5′ and 3′ end to facilitate cloning.

Oligo format:

TAGATTGATCTAGAGCATGCACCGG- [150bp oligo] -TGGAGTTCAGACGTGTGCTCTTCCG

STARR-seq oligo pools were amplified by two-step PCR with the following primers to extend the homolog arm for cloning:
PCR1 primers:
20180720_GA_STARR60bp _fw:

ACTCTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATTGATCTAGAGCATGCACCGG

20180720_GA_UMI_Read2_rv:

GTATCTTATCATGTCTGCTCGAAGCGGCCGGCCGAATTCGTCGA-[15nt UMI]-AGATCGGAAGAGCACACGTCTGAACTCCA

PCR2 primers:

20180720_GA_STARR_left_fw: ACTCTTCACTGGAGTTGTCCCAATT
20180720_GA_STARR_rightAmp_rv: GTCCAAACTCATCAATGTATCTTATCATGTCTGCTCGAAGC

The amplified product was then run on a 2% agarose gel and purified as previously described. STARR-seq backbone plasmid: hSTARR-seq_ORI vector (Addgene: #99296) was digested with AgeI-HF and SalI-HF (New England Bioscience) and purified for Gibson assembly reaction.

Gibson assembly was set up according to the reaction ratios below:

Purified linear vector: 130.1ng (0.083pmols for 2537bp)
Purified PCR inserts:106.7ng (0.55pmols for 314bp)
2x Gibson assembly Master Mix: 10uL
H2O: fill up to 20uL

Reaction mix was further purified, electroporated, and maxiprep as previously described.

Deep sequencing to determine STARR-seq barcode-oligo dictionary

We used two-step PCR amplification to prepare sequencing library from plasmid libraries to create barcode-oligo dictionary. PCR1 was performed with the pool of the following forward and reverse primer sets with stagger random nucleotides to maintain sequence diversity across the flow-cell.

Forward primers:

20190114_Nextera_halfread1_staggered_fw1: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG NNN TGATCTAGAGCATGCACCGG
20190114_Nextera_halfread1_staggered_fw2: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG NNNN TGATCTAGAGCATGCACCGG
20190114_Nextera_halfread1_staggered_fw3: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG NNNNN TGATCTAGAGCATGCACCGG
20190114_Nextera_halfread1_staggered_fw4: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG NNNNNN TGATCTAGAGCATGCACCGG

Reverse primers:

20190114_half_Nextera_rv1: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNCATGTCTGCTCGAAGCGGCC
20190114_half_Nextera_rv2: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNCATGTCTGCTCGAAGCGGCC
20190114_half_Nextera_rv3: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNCATGTCTGCTCGAAGCGGCC
20190114_half_Nextera_rv4: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNCATGTCTGCTCGAAGCGGCC

PCR2 was performed with the following primers to extend for full-length adapters.
20190114_P5_Nextera_fw:

AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTC

20180720_P7_Nextera_rv:

CAAGCAGAAGACGGCATACGAGATGTCTCGTGGGCTCGG

Sequencing was performed with NextSeq 2X150 MidOutput (Illumina) for 243bp read-length of Read1, and 73bp read-length of Read2. A custom Read2 sequencing primer was used: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG.

STARR-seq library transfection and transcript sequencing library preparation

For each biological replicate, two 15-cm dishes of cell culture with ~80% confluency was transfected with 40ug STARR-seq plasmid library per plate through Lipofectamine 3000 (Thermo Fisher). After 6hrs incubation with transfection mixture, culture medium was changed back to normal culture medium without antibiotics for 12hrs, meanwhile C16 and BX-795 inhibitors were added to the cells (final concentration 1 μM/inhibitor). Next, total RNAs were harvested through TRIzol-chloroform extraction (Thermo Fisher). 75ug of total RNA was used as starting material to enrich mRNA with Dynabeads® Oligo $(dT)_{25}$ (Thermo Fisher) according to the manufacturer's directions. The poly$(A)^+$ RNA was then treated with RQ1 DNase (Promega), followed by purification with RNeasy MinElute kit (QIAGEN). Reverse transcription was performed using the purified mRNA with SuperScript III (Thermo Fisher), and the following RT primer was used:

CCCTACACGACGCTCTTCCGATCT- [10 random nucleotide UMI]- CTCATCAATGTATCTTATCATGTCTG.

Next, the first-strand cDNA reaction mixture was treated with RNaseA and purified with MinElute PCR purification kit (QIAGEN). We performed two-step PCR amplification to construct transcript sequencing library. PCR1 was performed with the following junction primers to capture cDNA generated from STARR-seq transcripts:

20180720_JuncPCR_fw: TGAGGCACTGGGCAGG*T*G*T*C*C (*phosphorothioate bond modified)
20180720_JuncPCR_rv: ACACTCTTTCCCTACACGACGCTCTTCC

PCR2 was performed to extend the full-length Illumina sequencing adapters using the NEBNext Index Primer Sets 1 (New England Bioscience). Pooled library from different conditions was sequenced with NextSeq 1X75bp high output (Illumina).

All primer sequences are listed in Table S1.

Reverse transcription quantitative PCR (RT-qPCR)

Cell culture that subjected to different treatment were washed with PBS, followed by TRIzol incubation for 10min at room temperature. Cell lysis were then collected for purification or storage at – 80 degrees. We performed phenol-chloroform phase separation on cell lysis and collected the RNA-containing aqueous for ethanol precipitation. The precipitated mixture was then subjected to purification with RNeasy Mini Kit (QIAGEN), following the manufacturer's protocol. RNA was resuspended and quantified with Nanodrop (Thermo Fisher), and subjected to DNase treatment with RQ1 RNase-Free DNase (Promega), following the manufacturer's protocol. Two methods were used to generate cDNA from DNase-treated total RNA. Method 1: for samples to profile gene expression, we used 1~2 ug of DNase-treated RNA for each reverse transcription reaction with random hexamer from ProtoScript® First Strand cDNA Synthesis Kit (NEB); Method 2: for samples to quantify the STARR-seq construct transcriptional levels, we used up to 4ug of DNase-treated RNA for each reverse transcription reaction with Oligo $(dT)_{20}$ primer via SuperScript III Reverse Transcriptase (Thermo Fisher). Afterward, cDNA was subjected to qPCR using DyNAmo Flash SYBR Green qPCR mastermix (Thermo Fisher) or iQ SYBR® Green supermix (BioRad) with CFX Connect Real-Time PCR machine (BioRad). For gene expression profiling, qPCR was performed with gene-specific qPCR primers. For STARR-seq construct, we used STARR-seq junction forward primer and regions-specific reverse primer. Ct readouts of each gene were first normalized with housekeeping gene GAPDH (ΔCt), and the relative expression of individual genes versus the expression levels in control conditions was then calculated with $2^{-\Delta\Delta Ct}$ method. The relative expression levels of STARR-seq constructs were also were calculated by comparing the normalized expression of profiling regions with negative control region using $2^{-\Delta\Delta Ct}$ method.

All qPCR primers were listed in Table S1.

Drug treatment for cancer cells

Stock solutions for small molecules were prepared according to manufacturer's recommendation. For single drug treatment, treatments were performed for 48 h. For dose-dependent treatment, combinatory treatment or rescue experiment, treatments were performed for 24 h. For STARR-seq transfection in combinatory with drug treatment, incubation was performed in line with STARR-seq protocol for 12hrs.

QUANTIFICATION AND STATISTICAL ANALYSIS

Demultiplexing and read preprocessing

For CRISPR screen:

Reads were demultiplexed requiring a perfect match to a designed 8-nt index and up to 1 mismatch to a designed 3-nt to 9-nt barcode residing at the beginning of the read. gRNAs were counted by requiring a perfect match to a designed 19-nt, 20-nt, or 21-nt gRNA in the library. Conditions corresponding to technical replicates were combined by summing their gRNA counts.

For constructing STARR-seq oligo-barcode dictionary:

Pooled reads were demultiplexed and quality filtered. Reads with mean Illumina scores Q < 30 were discarded. Paired-end reads were then used for construction of an oligo-barcode dictionary identifying uniqe plasmid DNA molecules. In the aggregate pooled data, 5041642 barcode-oligo pairs were detected in paired-end reads of plasmid DNA. Each oligo-barcode pair allowed us to associate a known oligo sequence from STARR-seq plasmid library with a randomized barcode, facilitating the later matching with transcript unique molecular identifiers (UMIs) to quantify the expression of each plasmid molecule.

gRNA enrichment analysis (MAGeCK)

MaGeCK was run with default parameters for both *in-cis* and *in-trans* screens. For *cis*-regulatory screen, we collected an average of ~1.6 million NGS reads per library, recovering 80%–100% of gRNAs with over 10 reads in all unsorted replicates. For *trans*-regulatory screen, we collected an average of ~7.7 million NGS reads per library, which managed to maintain a 100% coverage of gRNA diversity in all pre-Cas9 population replicates with a threshold of over 10 reads per gRNA. For the bulk population after Cas9 targeting, we obtained 99–100% coverage for U2OS cell line and 83–96% coverage for DLD-1 cell line. To identify *in-trans* acting candidate APOBEC3B regulators, we performed the following steps: 1) We filtered the genes with at least two gRNAs that showed true abundance in at least two biological replicates for each cell line; 2) Filtered gene lists were then ranked by their MAGeCK LFC relative to their abundance in reporter-negative populations with a threshold >0.85 (Log2) in at least three biological replicates for both cell lines.

STARR-seq transcript enrichment analysis

Unique molecules were identified by the incorporation of 10bp UMIs during reverse transcription of mRNA transcripts from STARR-seq plasmid treated cells. Redundant sequencing reads having identical plasmid barcodes and RNA molecule UMIs were discarded. In total, we identified ~2.1 million unique transcripts from U2OS biological replicates, ~3.5 million unique transcripts from DLD1 biological replicates and ~1.8 million unique transcripts from HCT-116 replicates. Transcript UMI counts were averaged over all barcodes matched with each oligo to compute a mean expression rate of each oligo in each experiment, representing the normalized enrichment of transcripts from each STARR-seq oligo.

STARR-seq tiling motif scanning

Using a database of 572 well-characterized motif position weight matrices, we scanned for putative transcription factor binding sites by scanning for motif sequences using a position weight matrix-based scoring method at all positions of each oligo from STARR-seq library with a window size of 60bp. A motif significance score was computed using a modeled, random distribution of simulated DNA sequences matching the observed nucleotide distribution of the APOBEC3B region. Motif hits were identified permissively, identifying motifs with with motif scores passing either a modeled false discovery threshold of 5%, or identified as 2 standard deviation (STD) significant.

CCLE and TCGA datasets

For gene expression correlation calculation, we used Cancer Cell Line Encyclopedia (CCLE) (DepMap 2018) (18Q3 release). Six cohorts of TCGA Pan-Cancer Atlas (PanCanAtlas) RNA-seq datasets (BRCA_HiSeqV2_PANCAN, BLCA_HiSeqV2_PANCAN, HNSC_HiSeqV2_PANCAN, LUAD_HiSeqV2_PANCAN, LUSC_HiSeqV2_PANCAN, LIHC_HiSeqV2_PANCAN) (Weinstein et al., 2013) were also recruited in this study. Patient IDs with the corresponding tumor and normal tissues can be found in Mendeley dataset deposit with https://doi.org/10.17632/b4wczz76gf.1.

Statistical analyses

Data are presented as means with standard error of the mean (SEM) to indicate the variation within each experiment unless otherwise noted. Statistics analysis was performed in Prism and R. Two-tailed t test was used for the comparison between two different conditions. For experiments with more than two conditions, ANOVA test was used to calculate significance. Annotation for p values in figure legends regardless of statistical test type are: *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.