

Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9

Gozde Korkmaz^{1,4}, Rui Lopes^{1,4}, Alejandro P Ugalde¹, Ekaterina Nevedomskaya², Ruiqi Han¹, Ksenia Myacheva¹, Wilbert Zwart², Ran Elkon¹ & Reuven Agami^{1,3}

Systematic identification of noncoding regulatory elements has, to date, mainly relied on large-scale reporter assays that do not reproduce endogenous conditions. We present two distinct CRISPR-Cas9 genetic screens to identify and characterize functional enhancers in their native context. Our strategy is to target Cas9 to transcription factor binding sites in enhancer regions. We identified several functional enhancer elements and characterized the role of two of them in mediating p53 (*TP53*) and ER α (*ESR1*) gene regulation. Moreover, we show that a genomic CRISPR-Cas9 tiling screen can precisely map functional domains within enhancer elements. Our approach expands the utility of CRISPR-Cas9 to elucidate the functions of the noncoding genome

Enhancers are genomic elements that regulate transcription of distantly located genes through chromatin looping. They function as binding platforms for transcription factors and are characterized by specific chromatin modifications¹. Recent studies have shown that genetic alterations can affect enhancer activity and contribute to tumorigenesis^{2,3}. Moreover, transcription factors and other enhancer-associated proteins are frequently mutated in human tumors, and targeting these proteins with small-molecule inhibitors holds much therapeutic potential^{1,4}. It is estimated that the human genome contains >500,000 putative enhancers, a staggering number that poses a major challenge for the identification of functional regulatory elements. Current methods to systematically identify enhancers are based on massively parallel reporter sequencing. However, the intrinsically artificial nature of these methods is likely to have some effect on their ability to delineate and assess the activity of endogenous enhancers. The recent development of CRISPR-Cas9 (clustered, regularly interspaced, short palindromic repeats (CRISPR) and the CRISPR-associated protein 9 (Cas9)) technology has opened unprecedented opportunities for genome-wide targeted editing in human cells⁵. Previous functional genetic screens using CRISPR-Cas9 have mainly been restricted to protein-coding genes^{6,7}. Here, we apply this technology to identify endogenous enhancer elements and to characterize the domains that are essential for their activity.

As a proof-of-principle demonstration, we focused on two transcription factors, p53 and ER α , which play key roles in cancer initiation and progression. p53 is known as the 'guardian of the genome', and is mutated in more than 50% of all human tumors⁸. Upon oncogene activation, one major function of p53 is to activate an irreversible cell-cycle arrest program named oncogene-induced senescence (OIS). OIS is a powerful tumor-suppressive mechanism; somatic mutations in p53, or in other components of its pathway, can overcome OIS and lead to tumorigenesis⁹. ER α is an estrogen-activated transcription factor that has a mitogenic role in breast cancer cells. The standard of care for ER α -positive breast tumors is treatment with selective ER α modulators and aromatase inhibitors. However, many tumors relapse after treatment and most of them still express ER α (also known as *ESR1*)¹⁰. Recently, both p53 and ER α have been shown to directly bind genomic regions that are characterized largely by features of distal-enhancer regions^{11,12}. This evidence suggests that the identification of p53- and ER α -bound enhancers and their target genes could be instrumental for diagnostics and therapeutics of cancer.

Initially, we set out to establish a genetic screen for p53-bound enhancers that are required for OIS (**Fig. 1a**). To build a CRISPR-Cas9 single guide RNA (sgRNA) library, we followed a strategy that enabled us to target $\approx 90\%$ of p53-bound enhancers (**Fig. 1b** and **Supplementary Table 1**). We cloned the sgRNAs into lentiCRISPRv2 vector¹³ using a pooled strategy and generated a lentiviral library (CRISPR-p53-enhLib). We performed our screen in human BJ cells containing tamoxifen-inducible HRAS^{G12V} (BJ-RAS^{G12V}), which are a well-characterized cell model of OIS^{14,15}. Accordingly, we transduced cells with three independent lentiviral pools of CRISPR-p53-enhLib, as well as with a nontargeting sgRNA pool (negative control) (**Fig. 1c**). After 4 weeks of culturing, we harvested the cells and performed next-generation sequencing to identify the sgRNAs present in the populations (**Fig. 1c** and **Supplementary Fig. 1a**).

Our screen detected eight substantially enriched sgRNAs ($q < 0.1$) in the RAS^{G12V}-induced cell populations (**Fig. 1d** and **Supplementary Table 2**). Notably, two independent sgRNAs targeted a putative enhancer located ~ 10 kb upstream of *CDKN1A* (formerly known as *p21*), which is a key effector of p53-dependent OIS (**Fig. 1e**; p53^{enh3507}). Another top-scoring sgRNA mapped to a known p53-responsive

¹Division of Biological Stress Response, the Netherlands Cancer Institute, Amsterdam, the Netherlands. ²Division of Molecular Pathology, the Netherlands Cancer Institute, Amsterdam, the Netherlands. ³Erasmus MC, Rotterdam University, Rotterdam, the Netherlands. ⁴These authors contributed equally to this work. Correspondence should be addressed to R.E. (r.elkon@nki.nl) and R.A. (r.agami@nki.nl).

Received 11 August 2015; accepted 8 December 2015; published online 11 January 2016; doi:10.1038/nbt.3450

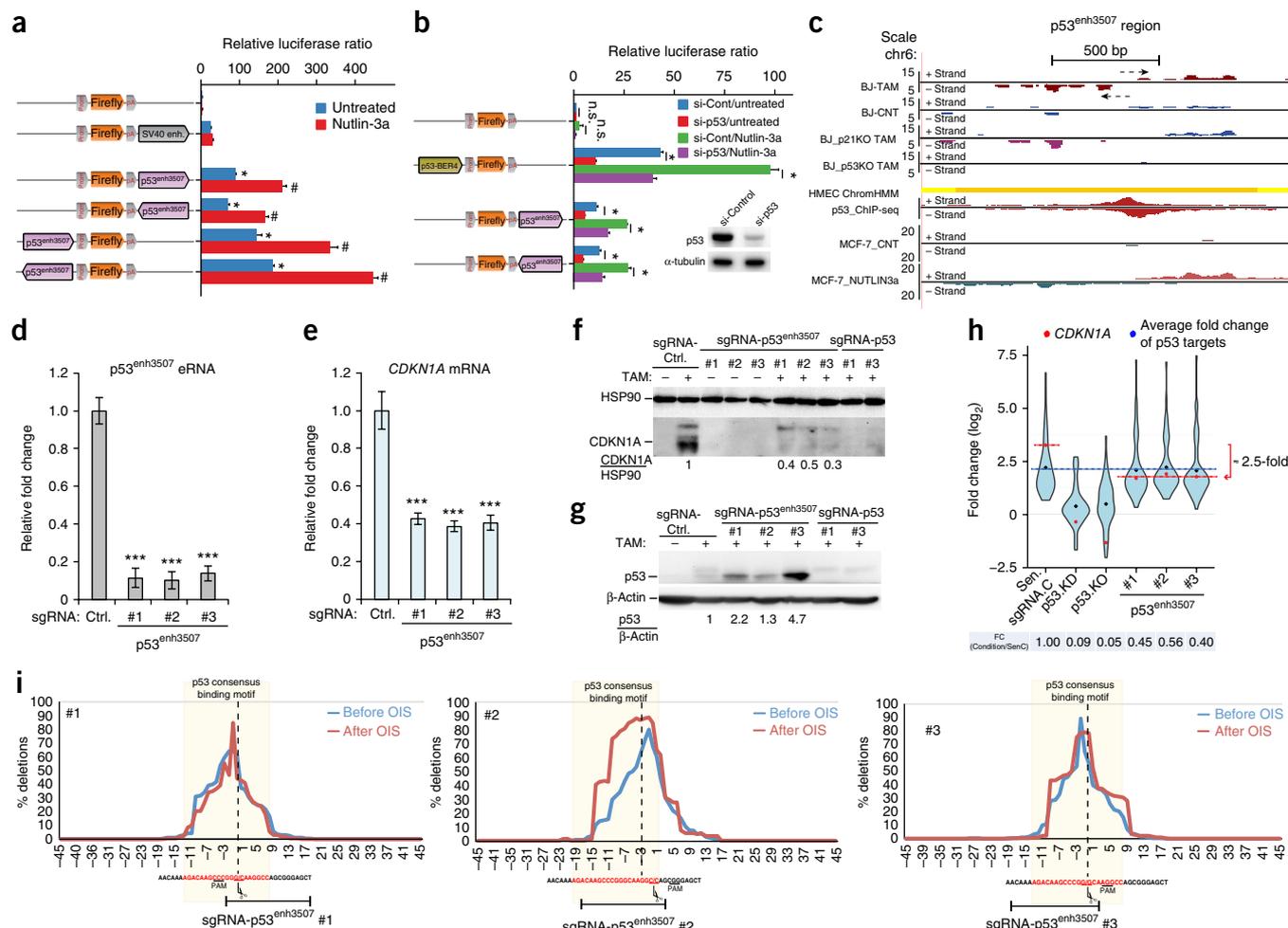
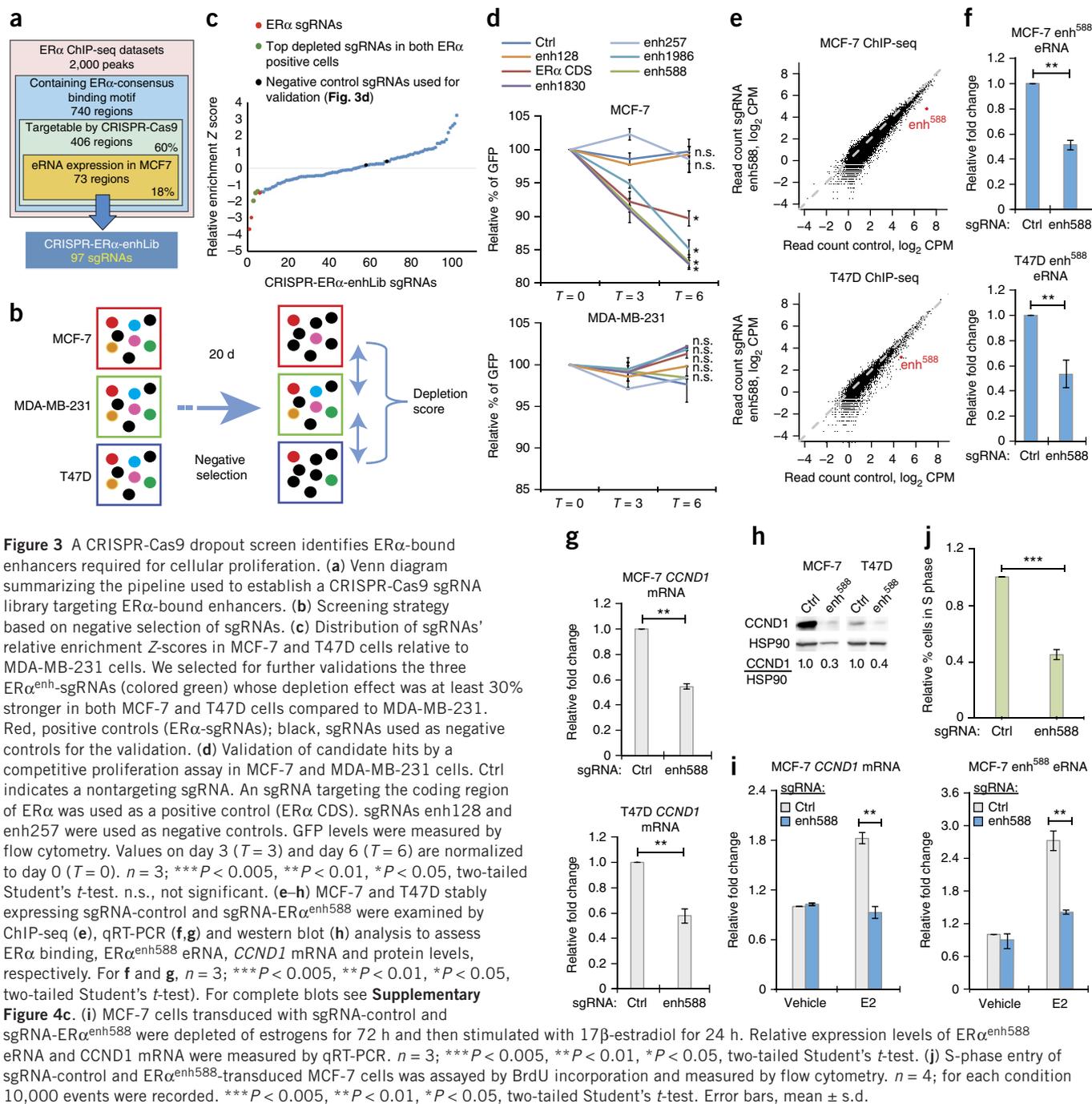


Figure 2 $p53^{enh3507}$ is a p53-dependent enhancer region that regulates *CDKN1A* expression. (a) MCF-7 cells were transfected with the indicated reporter vectors, treated with Nutlin-3a 5–10 h later, and harvested 25–30 h after treatment. The relative luciferase activities (firefly/*Renilla*) were normalized to the control (Ctrl.) reaction. *P*-values for luciferase assay were calculated by two-tailed Student's *t*-test. **P* < 0.005, relative to empty vector; #*P* < 0.005, relative to untreated matching sample. (b) The same assay as in a, only that cells were co-transfected with control, or p53-targeting short interfering RNA (si-Cont.; si-p53). A reporter vector containing the enhancer region p53-BER4 was used as a positive control for p53-dependency. The efficiency of p53 knockdown was determined by immunoblot analysis. *P*-values for luciferase assay were calculated by two-tailed Student's *t*-test. **P* < 0.01, relative to empty vector. (c) GRO-seq analysis detected strong induction of eRNA expression at $p53^{enh3507}$ upon RAS^{G12V} induction in BJ cells. This induction was completely abolished by p53-KO but not affected by CDKN1A-KO. The knockouts of p53 and CDKN1A were verified by western blot analysis (Fig. 2f and Supplementary Fig. 3f). Activation by Nutlin-3a treatment in MCF-7 cells resulted in a strong induction of eRNA in this region. (d,e) qRT-qPCR measurements of either eRNAs transcribed from the $p53^{enh3507}$ region (d) or mRNAs of *CDKN1A* (e). *n* = 3; ****P* < 0.005, **P* < 0.05, two-tailed Student's *t*-test. (f,g) Immunoblot analysis for CDKN1A (f) and p53 (g) proteins in BJ-RAS^{G12V} cells transfected as indicated, and treated with tamoxifen (TAM) for 12–15 d. HSP90 and β -Actin protein levels are shown as a loading control. (h) Using RNA-seq, we identified a set of 54 known direct target genes of p54 that were induced by at least twofold upon HRAS^{G12V} activation in BJ cells (Sen.C), and examined the effect of various vector transductions on the induction of this set of genes: Sen.sgRNAs.c, control sgRNA; p53.KD, knockdown of p53 using siRNA; p53.KO, knockdown of p53 using CRISPR-Cas9 which targets p53; #1, #2 and #3: three independent sgRNA vectors targeting the p53 binding site within $p53^{enh3507}$. Violin plots show the distribution of fold-induction of the set of 54 direct targets of p53 in each condition. Blue diamond indicates average induction of p53 targets; red dot indicates the level of *CDKN1A* mRNA induction. (i) We deep-sequenced a genomic region of 100 nt centered at the p53-consensus binding site of $p53^{enh3507}$ from control and HRAS^{G12V}-induced BJ cells transfected with the indicated sgRNAs. We calculated the prevalence of deletions that occurred at each position within this interval (relative to the total number of reads that contained any deletion). The p53 consensus motif and the location of the sgRNA-mediated CRISPR-Cas9 endonuclease cut are indicated. Error bars, mean \pm s.d.

siRNA-mediated knockdown of p53 (Fig. 2b). We obtained similar results for the $p53^{enh3508}$ enhancer element (Supplementary Fig. 2b,c). Taken together, our results demonstrate that $p53^{enh3507}$ controls OIS, and its transcription-enhancing activity is p53-dependent.

Active enhancer regions produce enhancer-associated RNAs (eRNAs), whose expression levels correlate with enhancer activity¹⁷. Indeed, we performed global run-on sequencing (GRO-seq) in BJ-RAS^{G12V} cells and detected strong induction of eRNA expression

from the $p53^{enh3507}$ region upon activation of OIS (Fig. 2c). CRISPR-Cas9-mediated knockdown of p53, but not of *CDKN1A*, completely abolished eRNA expression from this region, indicating p53-dependent regulation of $p53^{enh3507}$ (Fig. 2c). GRO-seq data of MCF-7 cells treated with Nutlin-3a suggest that this region is an active p53-responsive enhancer in different types of human cells¹⁸ (Fig. 2c). Next, we measured eRNA expression from the $p53^{enh3507}$ region and found that all three sgRNA- $p53^{enh3507}$ caused about a tenfold reduction in eRNA



expression (Fig. 2d), and a corresponding ~2.5-fold reduction in *CDKN1A* mRNA (Fig. 2e) and protein (Fig. 2f) levels. Importantly, the reduction of p53^{enh3507} eRNA and *CDKN1A* mRNA expression occurred in conditions of elevated p53 protein levels and continuous HRAS^{G12V} expression (Fig. 2g and Supplementary Fig. 3a). Finally, we performed RNA-seq in BJ-RAS^{G12V} cells and confirmed that the effect of sgRNA-p53^{enh3507} is specific to *CDKN1A* (Fig. 2h; red dot). In contrast, downregulation of p53 significantly ($P = 6.1 \times 10^{-13}$ for p53KD and $P = 8.3 \times 10^{-8}$ for p53KO; Wilcoxon's test) reduced the expression of the majority of its target genes (Fig. 2h; blue dot). These results indicate that sgRNA-p53^{enh3507} disrupts the enhancing activity of this region and decreases the activation of *CDKN1A* upon OIS.

Cas9-nuclease activity generates DNA double-strand breaks that result in deletions and insertions in the vicinity of the sgRNA recognition site¹⁹. Therefore, we examined the spectrum of deletions caused by sgRNA-p53^{enh3507} in control and induced BJ-RAS^{G12V} cells. We found that sgRNA-p53^{enh3507} caused deletions ranging from 1–15 nucleotides (Fig. 2i). The mutations were restricted to the p53 binding site following OIS induction, indicating that small deletions in key noncoding regulatory sequences are sufficient to cause a phenotypic change *in vivo* (Fig. 2i). We also tested the effect of CRISPR-Cas9-induced mutations on the activity of p53^{enh3507} region in a reporter vector. In all cases, p53-dependent enhancer activity was abolished by the mutations (Supplementary Fig. 3c), indicating that an intact p53 binding site is indispensable for the p53^{enh3507} enhancing function.

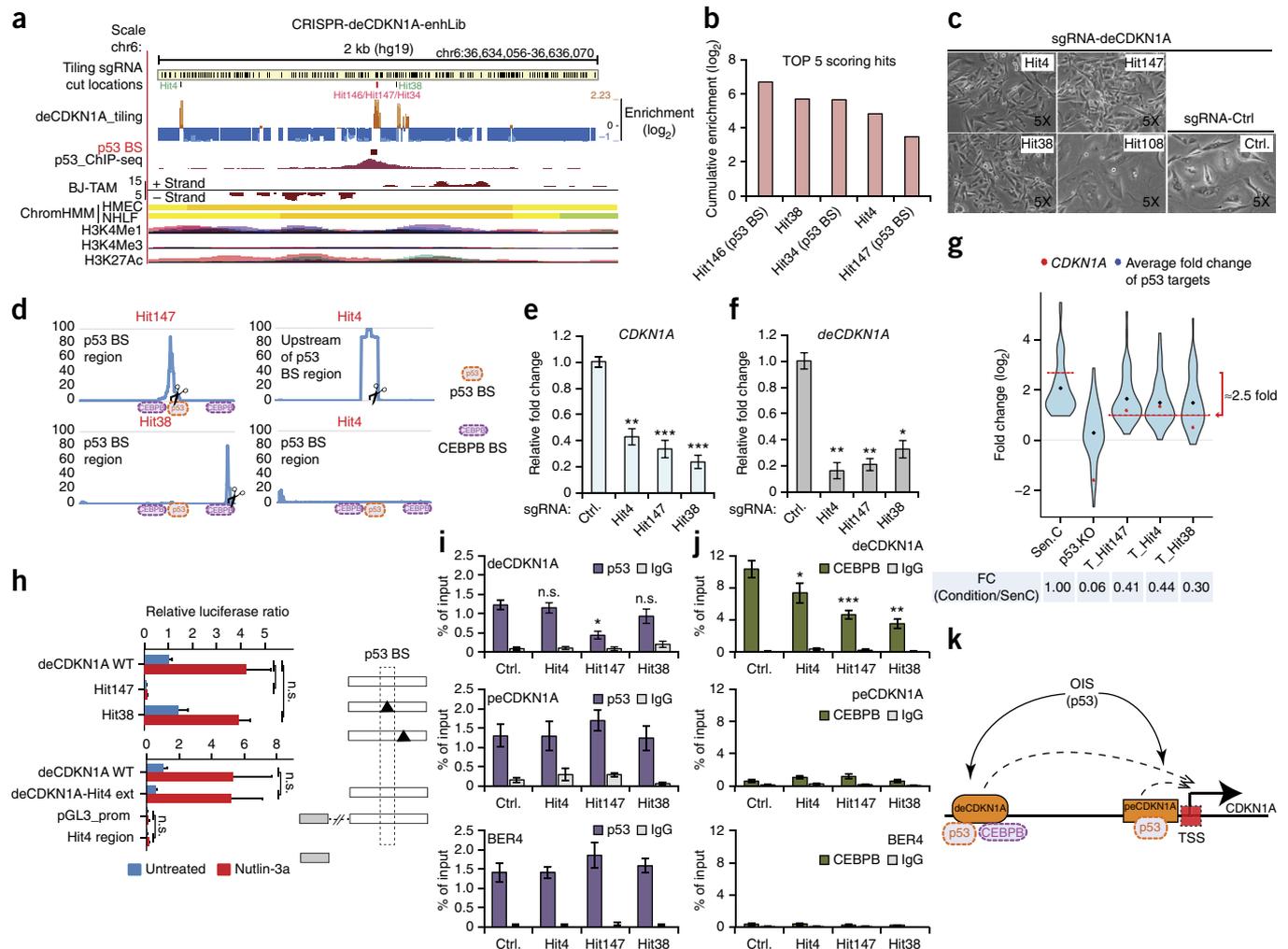


Figure 4 CRISPR-Cas9 tiling screen uncovers novel elements required for enhancer function. **(a)** Top panel, a snapshot of the genomic region surrounding *deCDKN1A*. Each black line indicates one sgRNA (total 197 sgRNAs covering ~2 kb). At the middle, the result of the CRISPR-*deCDKN1A*-enhLib screen is shown. Enrichment score was calculated for each sgRNA based on the ratio between its (normalized) prevalence in the HRAS^{G12V}-induced and control BJ cells. Blue, OIS-depleted sgRNAs; orange, OIS-enriched ones; green, sgRNAs targeting novel candidate regulatory elements; red, positive control sgRNAs that target the p53 binding site (BS). UCSC tracks of p53-ChIP and chromatin annotation are as in **Figure 1e**. HMEC, human mammary epithelial cell; NHLF, human lung fibroblast; ChromHMM, chromatin state segmentation by HMM; H3K4Me1, histone H3 lysine 4 monomethylation; H3K4Me3, histone H3 lysine 4 trimethylation; H3K27Ac, histone H3 lysine 27 acetylation. **(b)** Top five hits from the tiling screen and their cumulative enrichment from three independent screens. **(c)** Light microscopy images of cells transduced with the indicated sgRNAs. **(d)** Sequencing spectrums of cleavage sites of top tiling screen hits. The spectrum of Hit38 was very narrow and resembled the one induced by Hit147 that targets the p53 BS. As the cleavage induced by Hit4 is located outside this genomic interval, there was no marked peak in this region. **(e,f)** qRT-PCR of *CDKN1A* mRNA **(e)** and *deCDKN1A*-eRNA **(f)** of BJ-RAS^{G12V} cells transduced with the indicated sgRNAs, and treated as in **Figure 2d**. $n = 3$; $***P < 0.005$, $**P < 0.01$, $*P < 0.05$, two-tailed Student's *t*-test. **(g)** RNA-seq result depicted as violin plots to show the distribution of p53 targets genes in each condition. Blue diamond indicates average induction of p53 targets; red dot indicates the level of *CDKN1A* mRNA induction. **(h)** Luciferase assay of MCF-7 cells transfected with the indicated reporter vectors and induced with Nutlin-3a. The relative luciferase activities (firefly/*Renilla*) were normalized to the control reaction. All *P*-values for luciferase assay were calculated by Student's *t*-test. **(i,j)** Bar graphs of p53 **(i)** and CEBPB **(j)** binding at *deCDKN1A*, *peCDKN1A* and p53-BER4 regions by ChIP-qPCR. BJ-RAS^{G12V} cells were infected with either sgRNA targeting Ctrl or each newly identified sgRNAs and treated with tamoxifen (TAM) for 12–15 d. $n = 3$; n.s., not significant, $***P < 0.005$, $**P < 0.01$, $*P < 0.05$, two-tailed Student's *t*-test. **(k)** A schematic model showing that p53 binding at both distal and proximal (*deCDKN1A* and *peCDKN1A*, respectively) regions is required for sustained high level of *CDKN1A*, and for OIS. Error bars, mean \pm s.d.

Interestingly, concurrent disruption of p53^{enh3507} and p53^{enh3508} resulted in further reduction of *CDKN1A* expression, suggesting an independent regulation by both enhancer elements (**Supplementary Fig. 3b**; left panel). In comparison, downregulation of p53 resulted in even lower levels of *CDKN1A*, possibly due to indirect effects on *CDKN1A* expression¹⁸. However, BrdU and β -gal assays showed no additional effect by the combined sgRNAs compared with the singles, indicating that

inactivation of one enhancer is sufficient to complete a phenotypic alteration (**Supplementary Fig. 3d,e**). Collectively, these results demonstrate that p53^{enh3507} is an endogenous enhancer of *CDKN1A* and disruption of this region causes bypass of senescence in p53-WT cells. Moreover, cooperative action of p53^{enh3507} (distal enhancer of *CDKN1A*—*deCDKN1A*) and p53^{enh3508} (proximal enhancer of *CDKN1A*—*peCDKN1A*) is required to activate *CDKN1A* expression and initiate OIS.

To demonstrate the generalizability of our screening approach, we designed a dropout screen to identify novel ER α -bound enhancers. First, we selected two breast cancer cell lines (MCF-7 and T47D) that require ER α for cell proliferation, and one (MDA-MB-231) that lacks ER α expression. We generated a sgRNA library to target 73 ER α binding sites according to the strategy shown in **Figure 3a** (CRISPR-ER α -enhLib; **Supplementary Table 3**). For this library, we used eRNA expression measured by GRO-seq in MCF-7 cells as a criteria for active enhancers¹². As positive controls, we included in the library three sgRNAs targeting the coding region of ER α . Accordingly, we transduced the three different cell lines with the CRISPR-ER α -enhLib and allowed the cells to proliferate for 20 d (**Fig. 3b**). After identifying the sgRNAs present in the cell populations, we ranked the hits by negative effect on cell proliferation and selected sgRNAs that were strongly depleted (at least 30%) in both MCF-7 and T47D compared to MDA-MB-231 cells. Using these criteria, we identified two positive controls that target the ER α and three candidate sgRNAs (ER α ^{enh588}, ER α ^{enh1830} and ER α ^{enh1986}) that affect the proliferation of MCF-7 and T47D cells (**Fig. 3c** and **Supplementary Table 4**). We validated the candidates with a competitive proliferation assay in MCF-7 and MDA-MB-231 cells (**Fig. 3d**). As expected, sgRNA-ER α reduced the proliferation of MCF-7 cells but had no effect on MDA-MB-231 cells. Remarkably, all three candidates identified in the screen also significantly decreased the proliferation of only MCF-7 cells. In contrast, two sgRNAs (ER α ^{enh128} and ER α ^{enh257}) that were not depleted in the screen, and a nontargeting sgRNA control, had no significant effect on cell proliferation. These results indicate that our CRISPR-Cas9 dropout screening approach is specific, and led to the identification of three novel enhancers that are required for breast cancer cell proliferation.

One of the validated candidates, ER α ^{enh588}, has not been endogenously characterized to date. A ChIA-PET (chromatin-interaction analysis by paired-end tag) study in MCF-7 cells has previously identified ER α ^{enh588} as a hotspot for ER α binding²⁰. Analyses of this data set showed that ER α ^{enh588} interacts with the promoter region of Cyclin D1 (*CCND1*) (**Supplementary Fig. 4a**), suggesting that ER α ^{enh588} is a putative regulator of *CCND1* expression by ER α . *CCND1* oncogene plays a central role in cell-cycle progression and is overexpressed in more than 50% of breast tumors²¹. We started by cloning ER α ^{enh588} WT region in a reporter vector and verified that it has strong ER α -dependent transcription-enhancing activity, since mutations in the ER α binding site completely abolish the enhancer activity and response to 17 β -estradiol (**Supplementary Fig. 4b**). Next, we examined endogenous ER α binding at ER α ^{enh588} region by ChIP-seq and observed a substantial decrease in both MCF-7 and T47D cells expressing sgRNA-ER α ^{enh588} (**Fig. 3e**). Because *CCND1* is a putative target of ER α ^{enh588}, we verified the endogenous expression of ER α ^{enh588} eRNAs and *CCND1* mRNA and protein in MCF-7 and T47D cells transduced with sgRNA-ER α ^{enh588}. Reassuringly, we found that the expression of eRNA, mRNA and protein is significantly decreased ($P < 0.01$; about twofold) in both cell lines (**Fig. 3f–h**). These results indicate that ER α ^{enh588} is a bona fide enhancer element that regulates the expression of *CCND1*. Next, we assessed the dependency of ER α ^{enh588} and *CCND1* endogenous expression on estrogen signaling by qPCR (**Fig. 3i**). As expected, we confirmed that both ER α ^{enh588} eRNA and *CCND1* mRNA are upregulated in MCF-7 cells upon treatment with 17 β -estradiol. However, we found that sgRNA^{enh588} severely compromises the induction of eRNA expression and completely abolishes *CCND1* mRNA activation in MCF-7 cells. These results suggest that the activation of *CCND1* expression by estrogen in breast cancer cells requires a fully active ER α ^{enh588} enhancer element. Finally, as *CCND1* is a crucial component of the G1-S phase transition, we

examined the phenotypic outcome of disrupting ER α ^{enh588} activity. **Figure 3j** shows that MCF-7 cells transduced with sgRNA^{enh588} display a ~2.5-fold reduction in S-phase entry, compared to control-transduced cells, due to decreased *CCND1* expression.

The genetic code that enables enhancer activity is poorly understood. We reasoned that each enhancer is likely to contain multiple regulatory elements, and adopted CRISPR-Cas9 technology to pinpoint critical domains of enhancers. For that purpose, we performed a pooled high-throughput genetic tiling screen to identify additional elements, apart from the p53 binding site, that are required for the p53^{enh3507} region to regulate *CDKN1A* expression and function in OIS.

We targeted a genomic region of ~2 kb centered at the p53 binding site of *deCDKN1A* (**Fig. 4a**). We identified protospacer-adjacent motifs (PAMs) within this region and designed a library of 197 sgRNAs (CRISPR-deCDKN1A-Lib) that direct CRISPR-mediated cleavage every 10 bp on average (**Supplementary Table 5**). We performed an OIS screen with the CRISPR-deCDKN1A-Lib following the same procedure described in **Figure 1c**, which identified five enriched sgRNAs (**Fig. 4b** and **Supplementary Table 6**). Three of the most-enriched hits targeted the p53 binding site (Hit146, Hit147 and Hit34), and thus served as positive controls and as indicators of the robustness of the assay. The other two hits, Hit4 and Hit38, targeted novel candidate regulatory domains located 0.9 kb upstream and 0.1 kb downstream of the p53 binding site, respectively (**Fig. 4a**, middle panel). For validation, we used individual sgRNAs and confirmed that Hit4 and Hit38, similarly to the positive control Hit147, caused bypass of OIS (**Fig. 4c**). We analyzed the spectrum of mutations generated by these two sgRNAs and verified that it was very similar to the one caused by sgRNAs that directly targeted the p53 binding site (**Figs. 2i** and **4d**). Importantly, sgRNA-Hit4 and sgRNA-Hit38 caused genomic alterations that did not overlap with the p53-binding site within this enhancer region, indicating that they disrupt other enhancer domains required for OIS (**Fig. 4d**).

To characterize the function of these two regulatory domains, we first examined the expression of *CDKN1A* mRNA and *deCDKN1A* eRNA in BJ-RAS^{G12V} cells transduced with sgRNA-Hit4 and sgRNA-Hit38. We observed a significant ($P < 0.005$, and < 0.01 ; and $P < 0.01$ and < 0.05 , respectively) reduction in both mRNA (**Fig. 4e**) and eRNA (**Fig. 4f**) expression levels, which was similar in magnitude to that of the positive control sgRNA-Hit147. Gene expression profiling by RNA-seq also supported a specific effect of these two sgRNAs on *CDKN1A* expression (**Fig. 4g**). However, unlike mutations in the p53 binding site, deletion of the Hit4 or Hit38 region did not affect enhancer function in a reporter assay (**Fig. 4h**), suggesting that these domains only have a functional role in their endogenous context. Next, we analyzed endogenous p53 binding to the *deCDKN1A* region using ChIP-qPCR. As expected, Hit147 significantly ($P < 0.05$) reduced p53 binding to this region, whereas both Hit4 and Hit38 had no significant effect (**Fig. 4i**). For this experiment, we used *peCDKN1A* and p53BER4 regions as negative controls (**Fig. 4i**). We used PROMO²² to identify transcription factors that potentially bind to these regulatory domains and identified a perfect matching consensus binding site for CEBPB at the cleavage site of Hit38 (**Supplementary Fig. 5**). In contrast, no putative transcription factor binding site was predicted for the Hit4 region. Accordingly, we analyzed endogenous CEBPB binding to *deCDKN1A* by ChIP-qPCR and observed a marked reduction in BJ-RAS^{G12V} cells transduced with sgRNA-Hit38 compared to control cells (**Fig. 4j**). These results suggest that the DNA element targeted by sgRNA-Hit38 contributes to CEBPB recruitment to *deCDKN1A* and to its function in OIS (**Fig. 4k**). In addition, Hit147 also decreased CEBPB binding, possibly due to an additional CEBPB

binding site located adjacent to the p53 binding site. Also, Hit4 had a weak but significant ($P < 0.005$, < 0.01 and < 0.05) effect on CEBPB binding at the *deCDKN1A* region (Fig. 4j). Finally, we tested whether Hit4 and Hit38 had a cooperative effect on *CDKN1A* expression but, unlike *deCDKN1A* and *peCDKN1A*, this proved not to be the case (Supplementary Fig. 3b, right panel). Altogether, our results indicate that a CRISPR-Cas9 tiling strategy can precisely pinpoint regulatory domains within enhancer regions.

In summary, we present CRISPR-Cas9-based screens to identify and characterize functional enhancers in human cells. In total, we identified six enhancer elements that potentially control cell proliferation, and characterized two of them in detail—one regulates *CDKN1A* activation during OIS and the other mediates *CCND1* expression in response to ER α signaling (Figs. 2 and 3). We observed different rates of validation of candidates between the two genetic screens presented here, with the ER α -bound enhancer showing higher specificity. We speculate that this difference might be related to the intrinsic nature of the two screens (enrichment vs. dropout) or to the selection procedure of candidates (selecting ER α -bound enhancers based on eRNA expression). Of note, none of the control candidates from either screen showed any phenotypic activity. This evidence suggests that our screening approach has comparable specificity and sensitivity to genetic screens of protein-coding genes performed to present²³.

Recently, a dCas9-LSD1 fusion has been proposed to annotate native enhancers²⁴, yet the sensitivity and specificity of this tool has never been tested in large-scale genetic screens. Our method expands the utility of the CRISPR-Cas9 tool beyond the coding genome and can be applied to systematically identify functional enhancers bound by different sequence-specific transcription factors. At the present date, all CRISPR-Cas9-based systems require a PAM motif to direct DNA cleavage, and therefore cannot guarantee full coverage of the entire human genome. In our approach, about 90% and 60% of the candidate p53- and ER α -bound enhancers were targeted, respectively, but this rate might be different for other transcription factors. However, the use of CRISPR-Cas9 nucleases with altered or reduced PAM specificities²⁵ can increase the coverage of our approach to target enhancers in large-scale genetic screens. The selection of candidate enhancer regions for this study was based on ENCODE, ChIP-seq and GRO-seq data sets, but in principle our approach does not require prior data on transcription factor binding site. As an alternative, active enhancer regions can be detected by transcriptomic profiling of eRNA expression and targeted in an unbiased fashion using a CRISPR-Cas9 tiling approach such as we present here and as others have recently demonstrated for the enhancer of *BCL11A*²⁶.

Our study shows that CRISPR-Cas9 technology is a robust tool to identify and characterize functional enhancers in an unbiased fashion. We envision that our approach will be widely used to unravel the function of the noncoding portion of the human genome under both normal and pathological conditions.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. GEO: [GSE75627](#) (RNA-seq); [GSE75779](#) (ChIP-seq).

Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).

ACKNOWLEDGMENTS

We would like to acknowledge K. Rooijers for help in bioinformatic analyses, insightful ideas and critical discussions. We thank W. de Laat, P. Krijger, B. Evers, M. Amendola and R. Maia for reagents and technical assistance. We also thank all members of the Agami group for helpful discussions. And we thank F. Zhang (MIT) for the gift of plasmid vector lentiCRISPRv2 and D. Root (MIT) for pLX304-GFP. R.L. is a fellow of the Fundação para a Ciência e Tecnologia de Portugal (SFRH/BD/74476/2010; POPH/FSE). This work was supported by funds of The Human Frontier Science Program (LT000640/2013) to A.P.U., and enhReg ERC-AdV and NWO (NGI 93512001/2012) to R.A.

AUTHOR CONTRIBUTIONS

G.K., R.L. and R.A. conceived the project. G.K. and R.L. performed most of the experiments. A.P.U. did all luciferase assays. W.Z. and E.N. carried out and analyzed the ChIP-Seq data, respectively. R.H. helped during the library preparation. K.M. did cloning and validation of candidates from the screens. R.E. conducted all the bioinformatics analyses. G.K., R.L. and R.A. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049–1059 (2015).
- Mansour, M.R. *et al.* Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014).
- Vahedi, G. *et al.* Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* **520**, 558–562 (2015).
- Shi, J. & Vakoc, C.R. The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Mol. Cell* **54**, 728–736 (2014).
- Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–232 (2013).
- Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
- Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
- Muller, P.A. & Vousden, K.H. p53 mutations in cancer. *Nat. Cell Biol.* **15**, 2–8 (2013).
- Schmitt, C.A. *et al.* A senescence program controlled by p53 and p16INK4a contributes to the outcome of cancer therapy. *Cell* **109**, 335–346 (2002).
- Beelen, K., Zwart, W. & Linn, S.C. Can predictive biomarkers in breast cancer guide adjuvant endocrine therapy? *Nat. Rev. Clin. Oncol.* **9**, 529–541 (2012).
- Melo, C.A. *et al.* eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol. Cell* **49**, 524–535 (2013).
- Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).
- Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
- Drost, J. *et al.* BRD7 is a candidate tumour suppressor gene required for p53 function. *Nat. Cell Biol.* **12**, 380–389 (2010).
- Voorhoeve, P.M. *et al.* A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* **124**, 1169–1181 (2006).
- el-Deiry, W.S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).
- Kim, T.K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
- Léveillé, N. *et al.* Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA. *Nat. Commun.* **6**, 6520 (2015).
- Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
- Fullwood, M.J. *et al.* An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
- Arnold, A. & Papanikolaou, A. Cyclin D1 in breast cancer pathogenesis. *J. Clin. Oncol.* **23**, 4215–4224 (2005).
- Messeguer, X. *et al.* PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **18**, 333–334 (2002).
- Sigoillot, F.D. & King, R.W. Vigilance and validation: Keys to success in RNAi screening. *ACS Chem. Biol.* **6**, 47–60 (2011).
- Kearns, N.A. *et al.* Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods* **12**, 401–403 (2015).
- Kleinstiver, B.P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
- Canver, M.C. *et al.* BCL11A enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis. *Nature* **527**, 192–197 (2015).

ONLINE METHODS

Cell lines and chemical reagents. BJ-RAS^{G12V}, HEK293-T, MCF-7, T47D and MDA-MB-231 cells were cultured in DMEM medium (Gibco), supplemented with 10% FCS (Hyclone), and 1% penicillin/streptomycin (Gibco). For the estrogen-depletion experiment, MCF-7 cells were cultured in DMEM phenol red-free medium (Gibco) supplemented with charcoal stripped serum (Gibco). 17 β -estradiol was obtained from Sigma. All cell lines were obtained from the American Type Culture Collection, and they have been tested for mycoplasma contamination.

CRISPR-enhancer library design. For the p53-bound enhancers screen, we first took the union of the results of five publicly available p53 ChIP-seq analyses to create a combined set of 4,237 genomic sites that were bound by p53 in at least one cell line (MCF-7, CAL51 or IMR90) and in response to at least one stress (Nutlin-3a, 5-FU, RITA or ionizing radiation)^{27,28}. We then scanned these sites for occurrences of the p53-binding motif using the *p53scan* tool²⁹ and found that 2,626 sites contained strong matches. To increase the chance that the candidate sites were functional ones, we intersected them with genomic locations of predicted enhancers in six different cell lines. These predictions, which are based on various histone marks, were downloaded from the UCSC Genome Browser (Broad ChromHMM track). 764 sites with a strong match for the p53-binding motif overlapped a predicted enhancer in at least one cell line. Last, we identified the sites that could be targeted by a CRISPR-Cas9 sgRNA that cleaves that DNA within the p53 binding motif, taking into account that Cas9 endonuclease cuts the DNA 3 nt upstream of the PAM (NGG). (In cases that *p53scan* tool predicted an occurrence of the p53 motif that contained a spacer, we required that the cleavage would be out of the spacer). Overall, we designed 1,116 sgRNA vectors that target the p53-binding motif within 685 different genomic binding sites. For the ER α -bound enhancers, we took the top 2,000 ChIP-seq binding sites¹² and identified the ER α consensus motif (up to one mismatch) in 740 of them. 406 of these sites could be targeted by the CRISPR-Cas9 system. As a further step to narrow down the candidate list and focus the screen on active enhancers, we intersected these regions with eRNA expression measured by the same study using GRO-seq¹². Overall, 73 enhancers met these three criteria: (i) ER α binding detected by ChIP-seq; (ii) ER α motif that could be targeted by CRISPR-Cas9; and (iii) bidirectional eRNA expression. These 73 enhancers were targeted by 97 sgRNAs that comprise our CRISPR-ER α -enhLib. A list containing custom sgRNAs designed for this study can be found in **Supplementary Table 7**.

Pooled library cloning. We used standard de-salted DNA oligonucleotides, synthesized and purchased from IDT (Integrated DNA Technologies), to construct sgRNA libraries for p53-bound enhancers (1,116 sgRNAs), ER α -bound enhancers (97 sgRNAs) and deCDKN1A (197 sgRNAs). Complementary single-stranded oligos were phosphorylated and annealed by combining 100 μ M oligos, 1 \times T4 PNK Buffer, 1 mM ATP, 5 U T4 PNK and incubating the reaction at 37 °C/30 min, 95 °C/5 min followed by a ramp down to 25 °C at 5 °C/min. Annealed oligos were pooled into three independent replicates (pool #1, #2, #3), diluted at 1:1,000 in sterile water, and ligated to plasmid vector lentiCRISPRv2 (gift from Feng Zhang (Addgene plasmid #52961)) using the following parameters: 50 ng BsmBI (Fermentas) digested plasmid, 1 μ l diluted oligo duplex, 1 \times Ligation Buffer (Roche), 5 U T4 DNA Ligase (Roche) incubated at RT/30 min. We did five independent ligation reactions per pool, and used them to transform highly competent *Escherichia coli* cells (EletroSHOX - Bioline, BIO-85038) according to the manufacturer's protocol. In order to assess the complexity of our libraries, we plated 1 μ l of cell transformation mixture on Luria-Bertani agar plates containing ampicillin, incubated them overnight at 37 °C, and counted individual bacterial colonies after 16 h. At this point, we estimated that each individual sgRNA is covered >100 \times , ensuring that our libraries have high complexity and are suitable for pooled screening. Transformation mixtures were combined, grew in liquid LB until OD₆₀₀ = 0.8 was reached, and plasmid DNA was harvested using Genopure Plasmid Maxi kit (Roche).

Lentivirus production, purification and transduction. To produce lentivirus, 4 \times 10⁶ HEK293T cells per pool were seeded in ten 100-mm dishes 1 d before transfection. For each dish, we diluted 15 μ g of CRISPR-enhancer plasmid library, 3.5 μ g of pVSV-G, 5 μ g of pMDL RRE and 2.5 μ g of pRSV-REV in 450 μ l

of 0.1 \times TE/H₂O, added 50 μ l of CaCl₂ and incubated 5 min at RT. Plasmid DNA was precipitated by adding 500 μ l 2 \times HBS to the solution while vortexing at full speed. The precipitate was added immediately to the plate and the cells were incubated for 14 h at 37 °C, after which the medium was refreshed. Lentivirus-containing supernatants were collected 60 h post-transfection, filtered through a 0.45 μ m membrane (Milipore Steriflip HV/PVDF) and stored at -80 °C. All cell types and lentivirus batches tested were titrated in order to achieve a multiplicity of infection of 0.4–0.5. Cell lines were infected with lentivirus supernatants supplemented with 8 μ g/ml polybrene (Sigma). At 24 h post-infection, medium was replaced and cells were selected with 2 μ g/ml puromycin (Gibco). Antibiotic selection was stopped as soon as no surviving cells remained in the no-transduction control plate.

CRISPR-Cas9 screen for OIS in BJ-RAS^{G12V} cells. In both OIS screens, we infected ~3,500 BJ-RAS^{G12V} cells per vector, to ensure that every sgRNA was present in the cell population at the start of the experiment. Cells infected with CRISPR-enhancer or nontargeting CRISPR library pools were allowed to proliferate for 48 h after antibiotic selection to clear potentially toxic sgRNAs from the population. At this time point, half of the cells infected with CRISPR-enhancer library pools were harvested ($T = 0$), and the remaining cells were placed in culture and treated with 100 nM tamoxifen (4-hydroxytamoxifen, Sigma) to induce HRAS^{G12V} expression. Cells infected with CRISPR-enhancer and nontargeting CRISPR pools were allowed to proliferate, while we monitored them for senescence or continuous proliferation. After 4 weeks of treatment, we harvested cells infected with CRISPR-enhancer pools ($T = 4$ weeks). Cell pellets harvested at $T = 0$ and $T = 4$ weeks were stored at -80 °C, and processed later on for further analysis. The validation of individual hits identified in both screens was done following the same procedures described above. Enrichment scores were calculated by comparing the normalized frequency of each sgRNA vector present in the cell populations at $T = 0$ with $T = 4$ weeks.

CRISPR-Cas9 dropout screen in breast cancer cells. MCF-7, T47D and MDA-MB-231 cells were infected with two independent pools of CRISPR-ER α -enhLib. We infected ~3,500 cells per vector, to ensure that every sgRNA was present in the cell population at the start of the experiment. Following antibiotic selection, cells were allowed to proliferate for 48 h to clear potentially toxic sgRNAs from the population. At this time point, we harvested half of the cells infected with CRISPR-ER α -enhLib pools ($T = 0$). The remaining cells were placed in culture, allowed to proliferate for 20 d, and then harvested ($T = 20$). Cell pellets were stored at -80 °C, and processed later on for further analysis. Enrichment (depletion) scores were calculated for each sgRNA vector, in each cell line (MCF-7, T47D and MDA-MB-231) by comparing its normalized frequency at $T = 20$ and $T = 0$ pools. Then the differences between these enrichment scores (in log₂) were calculated for MCF-7 and T47D compared to MDA-MB-231 cells (which serve as controls in the screen as they are not dependent on ER α). These “Delta enrichment scores” calculated for MCF-7 and T47D were averaged and standardized (Z-scores). For validation, we selected sgRNA vectors whose repressive effect on proliferation was at least 30% stronger in both MCF-7 and T47D cells compared to MDA-MB-231 cells.

Genomic DNA sequencing to identify sgRNAs. Frozen cell pellets were thawed and genomic DNA (gDNA) was isolated with DNeasy Blood and Tissue kit (Qiagen). Identification of sgRNAs was done by PCR in two steps. For the first PCR, the amount of input gDNA was calculated to achieve >200 \times coverage over the CRISPR-enhancer libraries (assuming that 10⁶ cells contain 6.6 μ g gDNA), which resulted in 2 μ g for CRISPR-p53-enhLib, 200 ng for CRISPR-ER α -enhLib and 300 ng for CRISPR-deCDKN1A-Lib. For each sample, we performed two separate reactions (max. 1 μ g gDNA per reaction) using Phusion DNA polymerase (Thermo Scientific) and combined the resulting amplicons. In the first PCR, we used the following primer sequences to amplify lentiCRISPR-enhancer sgRNAs:

```
PCR1_F1
ACACTCTTCCCTACACGACGCTCTCCGATCTXXXXXXXXGGCTTTA
TATATCTTGTGAAAGGACG (XXXXXX represents a 6-bp barcode)
PCR1_R1
GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTACTGACGGGC
ACCGGAGCCAATTCC
```

A second PCR was performed to attach Illumina adaptors and index samples. The second PCR was done in 50- μ l reaction volume, including 5 μ l of the product from the first PCR, and using the following primers:

PCR2_P5

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC
GCTCTTCCGATCT

PCR2_P7

CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAG
ACGTGTGCTCTTCCGATCT (XXXXXX represents a 6-bp index).

Amplification was carried out with 18 cycles for both first and second PCR. After the second PCR, resulting amplicons were purified using Agencourt AMPure XP beads (Beckman Coulter), quantified in a Bioanalyzer 2100 (Agilent), mixed and sequenced in a HiSeq 2500 (Illumina).

Identification of sgRNAs enriched in BJ-RAS^{G12V} cells. Based on deep sequencing of sgRNAs in BJ control and HRAS^{G12V}-induced populations (done in independent triplicates for each), we estimated the enrichment of each sgRNA vector in HRAS^{G12V} relative to control cells. This was done, by counting the number of reads corresponding to each sgRNA in each population, normalizing these counts to 1 M and taking the ratio of the normalized counts between HRAS^{G12V} and control cells. We averaged these enrichment scores (in log₂ scale) for each sgRNA over the triplicates, and, last, calculate Z-scores (Fig. 1c and Supplementary Table 1). (To avoid inflation of ratios calculated for sgRNAs with low read counts, counts below 20 were set to 20.)

Genomic DNA sequencing to identify CRISPR-induced mutations. Cell pellets were collected and gDNA was isolated with DNeasy Blood and Tissue kit (Qiagen). Amplification of target regions for sequencing was done by PCR in two steps. For each sample, we used 500 ng of gDNA as input for the first PCR (done in duplicate). Resulting amplicons were combined and we used 5 μ l as input for the second PCR. Amplification was carried out with 18 cycles for both first and second PCR. After the second PCR, amplicons were purified using Agencourt AMPure XP beads (Beckman Coulter), quantified in a Bioanalyzer 2100 (Agilent), mixed and sequenced in a HiSeq 2500 (Illumina).

Competitive proliferation assay. MCF-7 and MDA-MB-231 cells were infected with indicated sgRNAs to validate the results of the CRISPR-ER α -enhLib screen. Separately, we generated polyclonal MCF-7 and MDA-MB-231 cells stably expressing GFP using pLX304-GFP³⁰ (gift from David Root; Addgene plasmid # 25890). GFP expressing cells were mixed in a 1:3 ratio with cells containing individual sgRNAs. The percentage of GFP-expressing cells was assessed by flow cytometry at the beginning of the experiment ($T = 0$) and every 72 h onwards ($T = 3$ d and $T = 6$ d). For every condition, 10,000 events were recorded, and the data were analyzed using FlowJo software.

Western blot analysis. Whole-cell lysates were prepared as previously described³¹. Membranes were immunoblotted with the following antibodies: TP53 (DO-1, Santa Cruz; 1:1,000), CDKN1A (Sc-397, Santa Cruz; 1:1,000), HRAS (C-20, Santa Cruz; 1:1,000), Cyclin D1 (M-20, Santa Cruz; 1:1,000), HSP90 (H-114, Santa Cruz; 1:10,000), beta-Actin (C4, Santa Cruz, 1:10,000). Protein bands were visualized using corresponding secondary antibodies (Dako) and ECL reagent (GE Healthcare).

Senescence-associated β -gal assay. BJ-RAS^{G12V} cells were transduced with lentiCRISPRv2 constructs, selected with puromycin, plated in triplicate and treated for 15 d with 100 nM 4-OHT to induce HRAS^{G12V} expression. β -galactosidase activity was determined with Senescence β -galactosidase staining kit (Cell Signaling), and at least 100 cells were analyzed for each condition.

BrdU proliferation assays. Cells were pulsed for 3 h with 30 μ M bromodeoxyuridine (BrdU, Sigma), fixed with ethanol (70% solution), permeabilized, treated with NaOH to denature DNA, incubated with anti-BrdU (GE Healthcare), washed in blocking buffer (PBS, Tween 0.05%, 2% BSA), and finally incubated with anti-rabbit AF488 secondary antibody (Dako). BrdU incorporation was measured either by immunofluorescence (at least 200 cells were scored for each condition) or by flow cytometry (10,000

events were recorded for each sample). Flow cytometry data were analyzed using FlowJo software.

Luciferase reporter assays. Sense and antisense region of *deCDKN1A* and *peCDKN1A* were PCR amplified from gDNA of BJ-RAS^{G12V} cells whereas ER α ^{enh588} was amplified from MCF-7 cells. All regions were cloned into pGL3-promoter vector. Constructs were transfected into MCF-7 cells and treated with 8 μ M Nutlin-3a (Cayman Chemical), 10⁻⁸ M 17 β -estradiol (Sigma) or vehicle for 30 h. Reporter activity was measured 36 h after transfection using Dual-Luciferase system (Promega) according to the manufacturer's instructions.

RNA isolation, reverse-transcription and quantitative real-time PCR (qPCR). Total RNA was extracted using TRIreagent (Bioline) and following the manufacturer's protocol. cDNA was produced with SuperScript III (Invitrogen) using 5 μ g of total RNA per reaction. qPCR reaction was performed with SYBR green I Master mix in a LightCycler 480 (Roche). TATA-binding protein (TBP) was used as an internal control. Primers used in qPCR are listed in Supplementary Table 7.

GRO-seq. GRO-seq was performed as described before with minor modifications. Briefly, 5 \times 10⁶ nuclei were isolated and incubated 5 min at 30 °C with equal volume of reaction buffer (10 mM Tris-Cl pH 8.0, 5 mM MgCl₂, 1 mM DTT, 300 mM KCl, 20 units of SUPERase In, 1% sarkosyl, 500 μ M ATP, GTP and Br-UTP, 0.2 μ M CTP+32P CTP) for the nuclear run-on. The reaction was stopped and total RNA was extracted with Trizol LS (Invitrogen) according to the manufacturer's instructions. RNA was fragmented using fragmentation reagents (Ambion) and the reaction was purified through p-30 RNase-free spin column (BioRad). BrU-labeled RNA was immunoprecipitated with anti-BrdU agarose beads (Santa Cruz), washed one time in binding buffer, one time in low salt buffer (0.2 \times SSPE, 1 mM EDTA, 0.05% Tween-20), one time high-salt buffer (0.25 \times SSPE, 1 mM EDTA, 0.05% Tween-20, 137.5 mM NaCl) and two times in TET buffer (TE with 0.05% Tween-20). RNA was eluted with elution buffer (20 mM DTT, 300 mM NaCl, 5 mM Tris-Cl pH 7.5, 1 mM EDTA and 0.1% SDS) and isolated with Trizol LS. After the binding step, BrU-labeled RNA was treated with tobacco acid pyrophosphatase (TAP, Epicenter) to remove 5'-methyl guanosine cap, followed by T4 polynucleotide kinase (PNK; NEB) to remove 3'-phosphate group. BrU-containing RNA was treated with T4 PNK again at high pH in the presence of ATP to add 5'-phosphate group. The reaction was stopped and RNA was extracted with Trizol LS. Sequencing libraries were prepared using TruSeq Small RNA kit (Illumina) following manufacturer's instructions. Briefly, end-repaired RNA was ligated to RNA 3' and 5' adapters, followed by RT-PCR amplification. cDNA was purified using Agencourt AMPure XP (Beckman Coulter) and amplified by PCR for 12 cycles. Finally, amplicons were cleaned and size-selected using Agencourt AMPure XP (Beckman Coulter), quantified in a Bioanalyzer 2100 (Agilent), and sequenced in a HiSeq 2500 (Illumina). Sequenced reads were aligned to the human genome (hg19) using bowtie2 (ref. 32).

RNA-seq. RNA-seq samples were processed with TruSeq RNA library prep kit v2 (Illumina) and sequenced in a HiSeq 2500 (Illumina). Sequenced reads were aligned to the human genome (hg19) using TopHat2 (ref. 33) and gene expression counts were calculated using HTSeq³⁴ based on Ensembl's human gene annotations (v69)³⁵. Expression levels were normalized using quantile normalization.

ChIP. BJ-RAS^{G12V} (5 \times 10⁶) cells were fixed with 1% formaldehyde at RT/8 min and quenched with 125 mM glycine for 5 min on ice. The cells were centrifuged at 470g/10 min and resuspended in 300 ml of cold lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA and 1% SDS) supplemented with protease inhibitor cocktail (Roche). The suspension was sonicated for 20 min (30 s on/off at maximum power) and diluted with 800 ml of dilution buffer (10 mM Tris-HCl, pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 1% Triton X-100). The lysate was centrifuged at 14,000 r.p.m./10 min and the soluble fraction was transferred to a new tube. For each reaction, 100 ml of chromatin preparation was diluted in 300 ml of dilution buffer and incubated overnight with indicated antibody amount at 4 °C on a rotator. To each ChIP reaction, 30 ml of protein A/G beads, previously blocked (PBS/BSA (0.1%) for 1 h),

was added and incubated 3 h at 4 °C. The immuno-precipitated chromatin was washed 2 × 5 min with dilution buffer and 1 × 5 min with TE (50 mM Tris-HCl pH 8.0 and 10 mM EDTA) and eluted overnight in 300 ml elution buffer (20 mM Tris-HCl pH 7.5, 5 mM EDTA, 50 mM NaCl and 1% SDS) at 65 °C in an orbital shaker. Eluted samples were purified using QIAquick PCR purification kit (Qiagen) and analyzed by real-time qPCR. The following antibodies and amounts were used in this experiment: 3 µg p53 (DO-1, Santa Cruz), 3 µg C/EBP beta (C-19, Santa Cruz). Primers used in ChIP-qPCR are listed in **Supplementary Table 7**.

ChIP-sequencing data analysis. Sequencing reads were aligned to the human genome (hg19) using bwa v 0.7.5 with default parameters. Number of aligned reads per sample can be found in **Supplementary Table 5**. Peaks in control MCF-7 and T47D cell lines were called with MACS³⁶ (default parameters) and DFilter³⁷ (parameters: -bs = 50 -ks = 30 -refine -nonzero) algorithms. In-house MCF-7 mixed input was used for peak calling of MCF-7 cell line; T47D input from a previous study³⁸ was used for calling T47D data. Intersect of the two peak calling algorithms was used for further analysis. 27020 and 6702 ER α peaks were detected in MCF-7 and T47D cell lines, respectively.

27. Botcheva, K., McCorkle, S.R., McCombie, W.R., Dunn, J.J. & Anderson, C.W. Distinct p53 genomic binding patterns in normal and cancer-derived human cells. *Cell Cycle* **10**, 4237–4249 (2011).
28. Rashi-Elkeles, S. *et al.* Parallel profiling of the transcriptome, cistrome, and epigenome in the cellular response to ionizing radiation. *Sci. Signal.* **7**, rs3 (2014).
29. Smeenk, L. *et al.* Characterization of genome-wide p53-binding sites upon stress response. *Nucleic Acids Res.* **36**, 3639–3654 (2008).
30. Yang, X. *et al.* A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods* **8**, 659–661 (2011).
31. Agami, R. & Bernards, R. Distinct initiation and maintenance mechanisms cooperate to induce G1 cell cycle arrest in response to DNA damage. *Cell* **102**, 55–66 (2000).
32. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
33. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
34. Anders, S., Pyl, P.T. & Huber, W. HTseq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
35. Cunningham, F. *et al.* Ensembl 2015. *Nucleic Acids Res.* **43**, D662–D669 (2015).
36. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
37. Kumar, V. *et al.* Uniform, optimal signal processing of mapped deep-sequencing data. *Nat. Biotechnol.* **31**, 615–622 (2013).
38. Ross-Innes, C.S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393. (2012).