

Comprehensive dissection of *cis*-regulatory elements in a 2.8 Mb topologically associated domain in six human cancers

Received: 14 November 2023

Accepted: 22 January 2025

Published online: 13 February 2025

 Check for updates

Christina M. Caragine^{1,2,3,4,5}, Victoria T. Le^{1,2,3,4,5}, Meer Mustafa^{1,2,3,4}, Bianca Jay Diaz^{1,2,3,4}, John A. Morris^{1,2,3,4}, Simon Müller^{1,2,3,4}, Alejandro Mendez-Mancilla^{1,2,3,4}, Evan Geller^{1,2,3,4}, Noa Liscovitch-Brauer^{1,2,3,4} & Neville E. Sanjana^{1,2,3,4} ✉

Cis-regulatory elements (CREs), such as enhancers and promoters, are fundamental regulators of gene expression and, across different cell types, the *MYC* locus utilizes a diverse regulatory architecture driven by multiple CREs. To better understand differences in CRE function, we perform pooled CRISPR inhibition (CRISPRi) screens to comprehensively probe the 2.8 Mb topologically-associated domain containing *MYC* in 6 human cancer cell lines with nucleotide resolution. We map 32 CREs where inhibition leads to changes in cell growth, including 8 that overlap previously identified enhancers. Targeting specific CREs decreases *MYC* expression by as much as 60%, and cell growth by as much as 50%. Using 3-D enhancer contact mapping, we find that these CREs almost always contact *MYC* but less than 10% of total *MYC* contacts impact growth when silenced, highlighting the utility of our approach to identify phenotypically-relevant CREs. We also detect an enrichment of lineage-specific transcription factors (TFs) at *MYC* CREs and, for some of these TFs, find a strong, tumor-specific correlation between TF and *MYC* expression not found in normal tissue. Taken together, these CREs represent systematically identified, functional regulatory regions and demonstrate how the same region of the human genome can give rise to complex, tissue-specific gene regulation.

Nearly 99% of the human genome does not code for proteins^{1–3}. Relative to protein-coding genes, less is known about noncoding regions and, in particular, the modulation of gene expression by noncoding *cis*-regulatory elements (CREs). Historically, it has been a challenge to characterize and identify CREs for a variety of reasons^{4–7}, including the large numbers of CREs distributed throughout the genome, the cell type specificity of CRE activity and their modest impact on gene expression. Large-scale profiling efforts from large consortiums such as the Encyclopedia of DNA Elements (ENCODE) and

The Cancer Genome Atlas (TCGA) have identified thousands of candidate CREs based on biochemical signatures such as histone modifications (e.g. H3K27ac) and chromatin accessibility (e.g. DNase hypersensitivity or ATAC-seq)^{8–11}. However, recent studies have found that, although these biochemical signatures often correlate with gene-regulatory function^{12–15}, they do not always predict functional impact^{16,17}. In addition, some promoters and enhancers have been shown to act redundantly and in combination^{18–22}. A further complication is that some methods to study functional noncoding elements

¹New York Genome Center, New York, NY, USA. ²Department of Biology, New York University, New York, NY, USA. ³Department of Neuroscience and Physiology, New York University School of Medicine, New York, NY, USA. ⁴Perlmutter Cancer Center, New York University School of Medicine, New York, NY, USA. ⁵These authors contributed equally: Christina M. Caragine, Victoria T. Le. ✉ e-mail: neville@sanjanalab.org

do not consider CREs in their native contexts (e.g. massively parallel reporter assays [MPRAs]^{23,24}).

Recent advances in massively-parallel genome perturbations *in situ* have improved our ability to find which CREs modulate gene expression and play a role in disease^{25,26}. Forward genetic pooled screens using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) nucleases in particular have been deployed by many groups to identify CREs in their native context^{27–32}. The improved detection of CREs has even led to the first FDA-approved gene editing therapy for sickle-cell anemia via disruption of a CRE³³. Most studies of CREs using CRISPR, however, do so in a single cell type, and it is unclear whether these elements are cell type specific or shared across multiple cells or tissues. The ENCODE Consortium used primary mouse and human cells and tissues to suggest that most CREs are tissue- or cell-specific: The approximately 700,000 distal enhancers in the human genome identified using a combination of biochemical signatures are found, on average, in only a single tissue or cell type⁸. Some CREs are identified in multiple cell/tissue types, suggesting the presence of core essential CREs that might be used across tissues. This is analogous to core essential genes identified using genome-wide CRISPR screens that are essential in multiple cell lines derived from distinct human tissues^{34,35}.

Despite this highly suggestive evidence that enhancers are mostly cell-type specific, there have been few efforts to explore the *functional* specificity of CREs – either with reporter assays or gene-editing in the native, chromatinized context. Here, we seek to explicitly examine the functional CRE landscape across cells derived from several distinct human cancers with high-resolution. We first catalog biochemical hallmarks of CREs near human oncogenes and use them to rank each oncogene locus by its regulatory diversity in different cancers. At the locus surrounding the oncogene *MYC* which has a diverse *cis*-regulatory landscape, we perform a series of saturation tiling CRISPR screens and identify 32 CREs in 6 cell lines from diverse human tissues. The majority are tissue-specific but several that are shared across different types of tissues. Although CREs at the *MYC* locus have been studied previously in human and mouse models^{36–44}, prior studies have focused typically on one or a few CREs in a single tissue/cell, making comparisons across tissues/cells challenging. Also, many of these studies rely on large deletions that make it challenging to delineate precise CRE boundaries. By deploying a uniform set of ~112,000 perturbations across the 6 cell lines, we identify previously unknown CREs that regulate *MYC* and define their boundaries at nucleotide resolution.

We validate a subset of these CREs in depth, measuring the impact of CRE perturbations on *MYC* and a non-coding RNA in the locus. We harness RNA-targeting Cas13 perturbations^{45,46} to show that the non-coding transcript *CCATI* itself is required for *MYC* regulation. Loss of *CCATI* transcript reduces also reduces looping of the *CCATI* locus with the *MYC* promoter, demonstrating that a noncoding RNA can itself modulate the regulatory impact of a noncoding DNA CRE. Furthermore, restoration of *CCATI* expression *in trans* can partially rescue the growth reduction of *CCATI* repression at the native CRE. We find that identified CREs harbor motifs for lineage-specific transcription factors and are in direct physical contact with the *MYC* promoter via dynamic, cell-type-specific 3D genome loops. Taken together, this comprehensive dissection of functional *MYC* CREs demonstrates the utility of perturbing noncoding CREs across different cell/tissue types to elucidate the complex, context-sensitive regulatory architecture of an important human oncogene.

Results

The *MYC* locus has high regulatory diversity across tissues

We first sought to characterize the diversity of the *cis*-regulatory landscape surrounding key oncogenes across cancer cell types. To do this, we quantified specific biochemical hallmarks: Enhancer-associated histone modifications H3K4me1 and H3K27ac and

chromatin accessibility (DNase I hypersensitivity, DHS) from the Encyclopedia of DNA Elements (ENCODE) flanking each oncogene in the Catalog Of Somatic Mutations In Cancer (COSMIC) database⁴⁷. (Fig. 1a–d). We analyzed data from human tumor-derived cell lines (Supplementary Data 1) representative of 10 different cancer types: melanoma (A375), lung adenocarcinoma (A549), liver carcinoma (HepG2), colorectal adenocarcinoma (HT29), leukemia (K562), lymphoma (Karpas-422), breast adenocarcinoma (MDA-MB-231), pancreatic carcinoma (Panc1), prostate adenocarcinoma (PC-3), and neuroblastoma (SK-N-SH). For 386 genes classified by COSMIC as an oncogene, we examined CREs within a ± 1 Mb window proximal to the gene body (Fig. 1c), as prior studies have indicated that most CREs are within 500 kb to 1 Mb of their target genes^{48,49}.

To assess regulatory diversity, we binned the data for each biochemical hallmark (H3K4me1, H3K27ac, DHS) at several different bin sizes and then, for each bin size, we computed the Euclidean distances between all pairs of cell lines. Greater distances between cell line pairs imply greater regulatory diversity. When examining each biochemical hallmark, we found that *MYC* consistently scored among the 5 top oncogenes (Supplementary Fig. 1a–d). Since the order of different genes tended to vary based on the CRE-associated biochemical hallmark and binning, we sought to identify a robust way to integrate the regulatory diversity scoring across these different analyses. When aggregating rankings via robust rank aggregation (RRA), we found that *MYC* was the oncogene with the greatest regulatory diversity (Fig. 1d). *MYC* is a transcription factor that controls gene expression^{50,51} and cellular proliferation^{52–54} and has been shown to be involved in myriad cancers^{55–62}. Usually tightly controlled at the transcriptional and protein level, *MYC* is estimated to be aberrantly expressed in 50–70% of human cancers^{63,64}.

This indicates that across many different cell types, *MYC* has a much more variable regulatory landscape compared to other oncogenes. For example, the MAPK serine/threonine-protein kinase *ARAF* has very low regulatory diversity across different cancers (Fig. 1e). Other top-ranked genes with high regulatory diversity were *AKT2*, *NFAT2*, and *BCL6* (Supplementary Data 2). Given the high regulatory diversity of *MYC* with many biochemical hallmarks of cell-type-specific enhancers, we decided to systematically interrogate CREs across cell types at the *MYC* locus.

High-throughput CRISPRi screens identify cell-specific CREs

The *MYC* oncogene resides within a 2.8 Mb topologically-associated domain (TAD) that contains only a few other genes. Although the *MYC* TAD is known as a gene desert, it does, however, have an abundance of non-coding RNAs^{65–67}. In order to profile the regulatory landscape around *MYC* across cells derived from different tissues, we first designed a CRISPR inhibition (CRISPRi) tiling library across the 2.8 Mb TAD (Fig. 2a) to identify functional regions associated with cancer-driven cell growth. Although previous studies have used both CRISPR nuclease and CRISPRi to map functional elements in the noncoding genome, we opted to use CRISPRi for these screens because many cancer cell lines tend to have amplifications of oncogenes like *MYC*⁶³. If a CRISPR target site has many copies in the genome, dropout may be due to excessive DNA damage and apoptosis instead of a bona fide regulatory interaction with the target gene^{68,69}. To design the library, we first considered all possible guide RNAs (gRNAs) with an NGG protospacer adjacent motif (PAM) within the TAD (298,839 gRNAs) and then selected a subset of these with a high on-target activity score and no homopolymer repeats of length 4 or more (Supplementary Fig. 2a, b). In total, the *MYC* TAD-tiling CRISPRi library consisted of 112,162 gRNAs with a median spacing of 9 nt; we sought comprehensive coverage of the *MYC* TAD without specific targeting of particular enhancer-associated epigenetic signatures (Supplementary Fig. 2c). We also included in the CRISPRi library design gRNAs targeting 490 essential genes with 4 gRNAs each and gRNAs targeting 613 non-essential genes with 4 gRNAs each (Supplementary Fig. 2d). As a

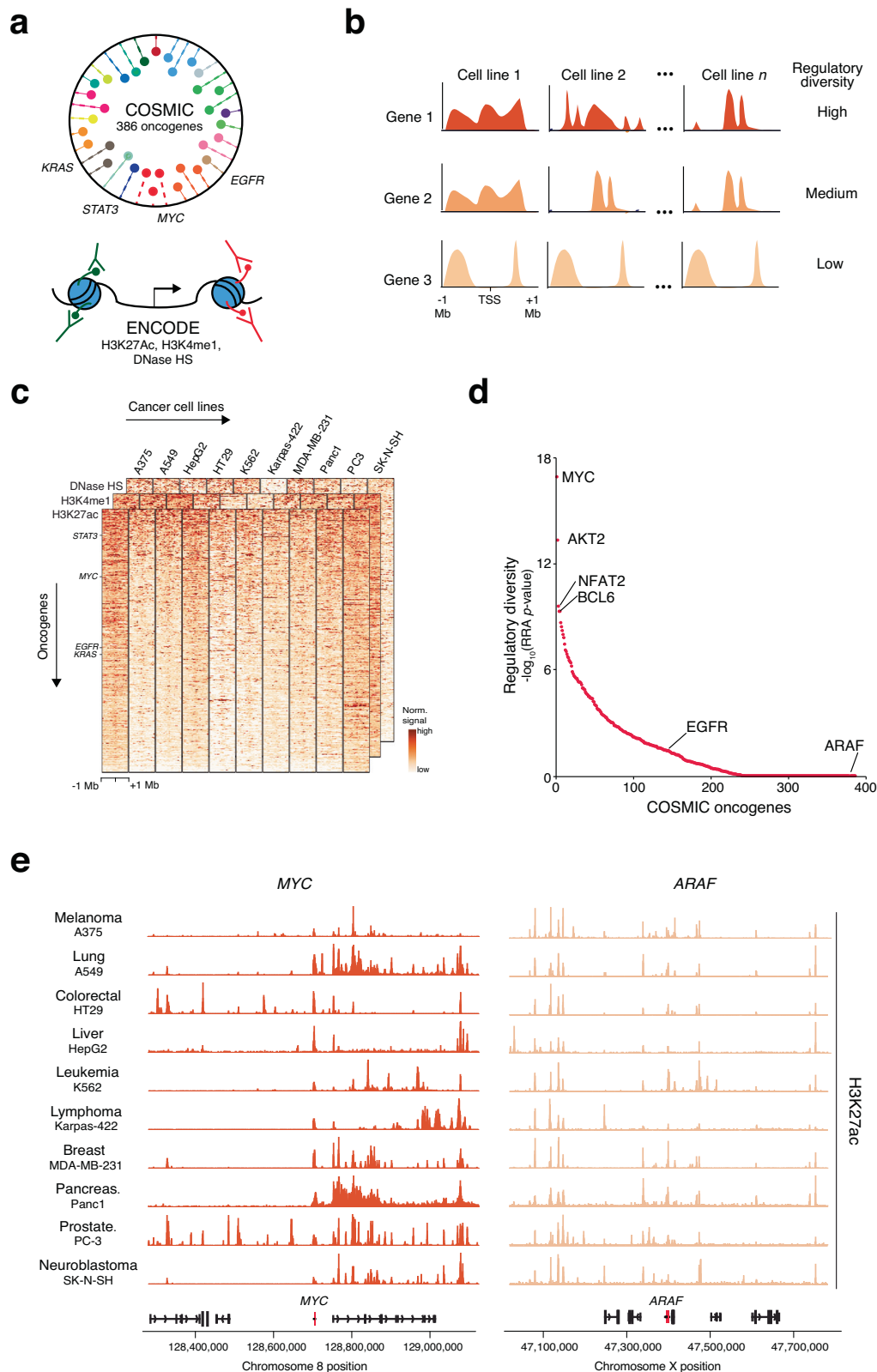


Fig. 1 | Diverse regulatory landscapes of different oncogenes. a Data from the Encyclopedia of DNA Elements (ENCODE) and the Catalog Of Somatic Mutations In Cancer (COSMIC) for quantifying diversity of regulatory landscapes. **b** Schematic of regulatory diverse over different cell lines for each biochemical signatures of enhancer activity. ENCODE data for H3K27ac, H3K4me1 and DNase I hypersensitivity (regulatory features of enhancer activity) were compared among 386 oncogenes (from COSMIC) across different cell types to assess regulatory diversity. We compute an oncogene regulatory diversity (ORD) score for each oncogene; a high score

indicates greater diversity across cell types. **c** Snapshot of ENCODE data for all 386 oncogenes across 10 cell lines and 3 biochemical features, visualizing a 2 Mb window around each oncogene. **d** Ranking of COSMIC oncogenes by ORD across biochemical signatures. The oncogene *MYC* has the greatest regulatory diversity across cell lines. *P* values are calculated by the robust rank aggregation algorithm, see “Methods”. **e** H3K27ac near a high diversity oncogene (*MYC*) and a low diversity oncogene (*ARAF*) across 10 cell lines. Source data are provided as a Source Data file.

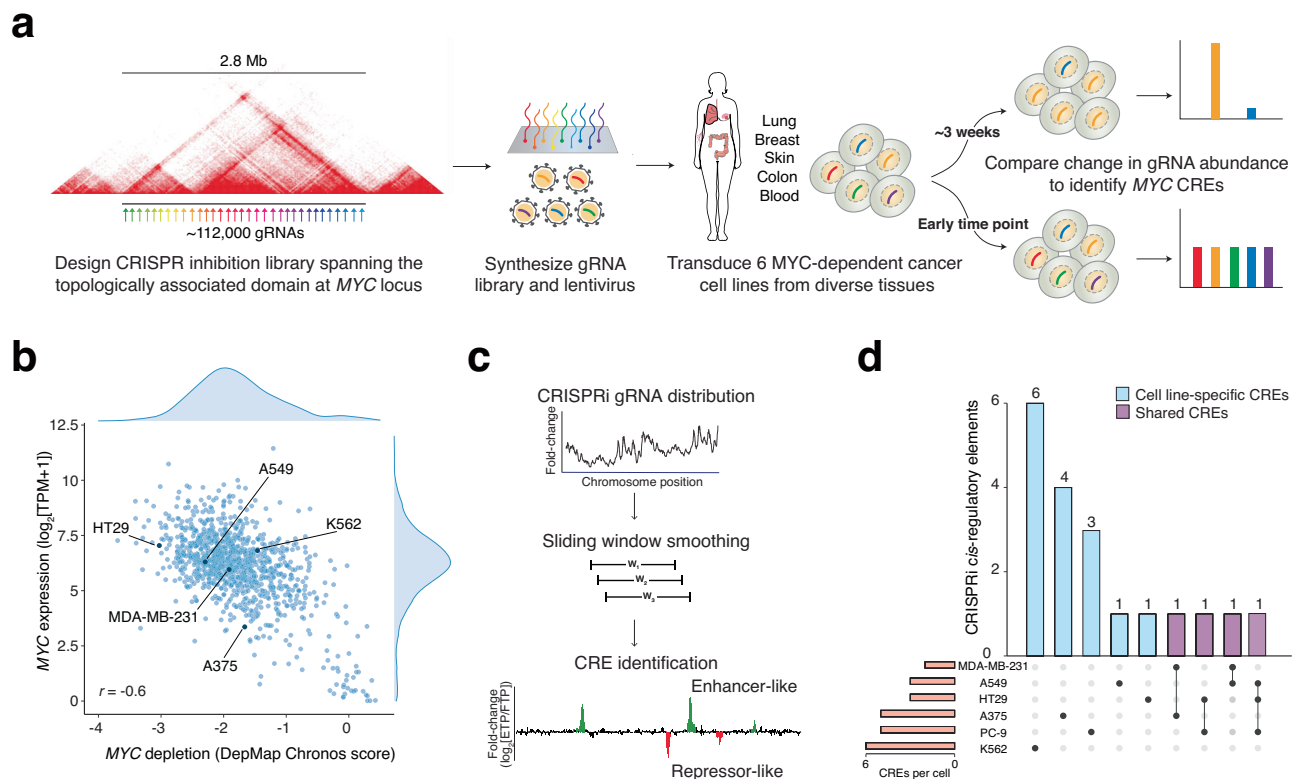


Fig. 2 | Interrogation of shared and unique enhancers across the *MYC* locus via massively-parallel genetic perturbations in six cell lines. **a** CRISPR inhibition (CRISPRi) screen tiling across a 2.8 Mb topologically-associated domain containing the oncogene *MYC* using ~112,000 gRNAs. **b** *MYC* expression and CRISPR screen

depletion (Chronos) scores from Broad DepMap ($n = 1021$ cell lines). **c** Sliding window analysis for identification of *cis*-regulatory elements (CREs). **d** Upset plot of shared and cell-specific enhancer-like CREs across six pooled CRISPRi screens. Source data are provided as a Source Data file.

further negative control, we also included 1000 non-targeting gRNAs from the GeCKOv2 library⁷⁰.

To identify appropriate cell lines for the tiling screen, we chose 6 cell lines (MDA-MB-231, A375, HT29, K562, PC9, A549) across different cancer types (one each of breast, skin, colon and blood and two lung cancer cell lines) that were previously reported to be dependent on *MYC* for their growth^{44,57,71–74} with similar *MYC* copy number (Supplementary Fig. 2e). In prior genome-wide CRISPR-Cas9 screens from the Broad DepMap, *MYC* was found to be highly essential in all 6 cell lines (Chronos score < -1) (Fig. 2b). We confirmed in each cell line that *MYC* transcript knockdown reduced cell growth (46%–96% reduction in growth) (Supplementary Fig. 2f). In total, we screened more than 700,000 genetic perturbations across 6 human cell lines.

We transduced these 6 cell lines with the CRISPRi tiling library at a low multiplicity of infection (MOI 0.2–0.5) to ensure that most cells would receive only one gRNA. After puromycin selection was complete, the cells were cultured for ~4 weeks to ensure sufficient time for *MYC*-driven depletion (or proliferation) of functional *MYC* CREs. We isolated genomic DNA at an early time point and at the end of the 4-week period and, via sequencing, quantified gRNA representation at the start and end of each of the 6 screens. As expected, we saw clear dropout of essential gene-targeting gRNAs in each of the 6 screens (Supplementary Fig. 3a). Overall, the *MYC* locus tiling gRNAs and the non-targeting negative controls had a similar distribution with far less depletion than essential gene-targeting gRNAs.

After calculating the fold change in abundance for each gRNA, we used a sliding window approach to identify significant CREs in each cell line (Fig. 2c, d). We performed these analyses on a per cell line basis, given that sensitivity to *MYC* depletion and growth differed across the cell lines (Supplementary Fig. 2f). To identify an optimal window size, we first examined the depletion of gRNAs at different distances from the

TSS of ~500 known essential genes taken from an overlap of previously identified gene lists^{35,75}. We found that the optimal CRISPRi effect radius (~90% of max) occurred for gRNAs within 180 bp of the TSS (Supplementary Fig. 3b). Given that the median spacing of gRNAs in the library was 9 bp, we used a sliding window of 20 neighboring (sequential) gRNAs for CRE identification (180 bp/9 bp). Next, we performed an unpaired, two-sided *t* test comparing the scores of gRNA sliding windows (of 20 sequential gRNAs each) with those of a set of scrambled sequence (non-targeting negative control) gRNAs included in the pooled screens⁷⁶ (Benjamini-Hochberg FDR < 0.05 across all screens).

With this procedure, we identified unique peaks in each cell line that showed significant depletion in the *MYC* TAD, which we hypothesized were putative enhancer-like CREs (Fig. 3a). Depending on the cell line, we found that 0.2–1% of windows were significantly depleted (Supplementary Fig. 3c). In total, we identified 32 CREs across the 6 cell lines (Fig. 3b). As expected, given the regulatory diversity of the *MYC* locus (Fig. 1d), CREs were mostly unique between cell lines (Fig. 2d, Supplementary Fig. 4a). The CREs were variable in size consisting of between 160 and 1133 gRNAs (Supplementary Fig. 4b, c), and spanning ~400 bp to 12.8 kb in length (Supplementary Fig. 4d). Only 5 CREs are found in more than one cell line and only 2 CREs are found in more than two cell lines—in agreement with the high level of diversity in the regulatory landscape of the *MYC* TAD across cell lines (Fig. 1d).

Of the 32 CREs that we identified, eight of them overlap a previously identified *MYC* enhancer, such as *MYC*-LASE or *BENC* or the super enhancer region around *CCAT1*^{77–79} (Supplementary Fig. 5a–d). Most of the established *MYC* enhancers were identified using large deletions in human or mouse models⁴⁴, and for one cell line (K562) using a similar tiling CRISPR screen²⁸. The other 24 CREs have not previously been described or have not been described in the same tissue/cancer. This highlights the tremendous potential for discoveries

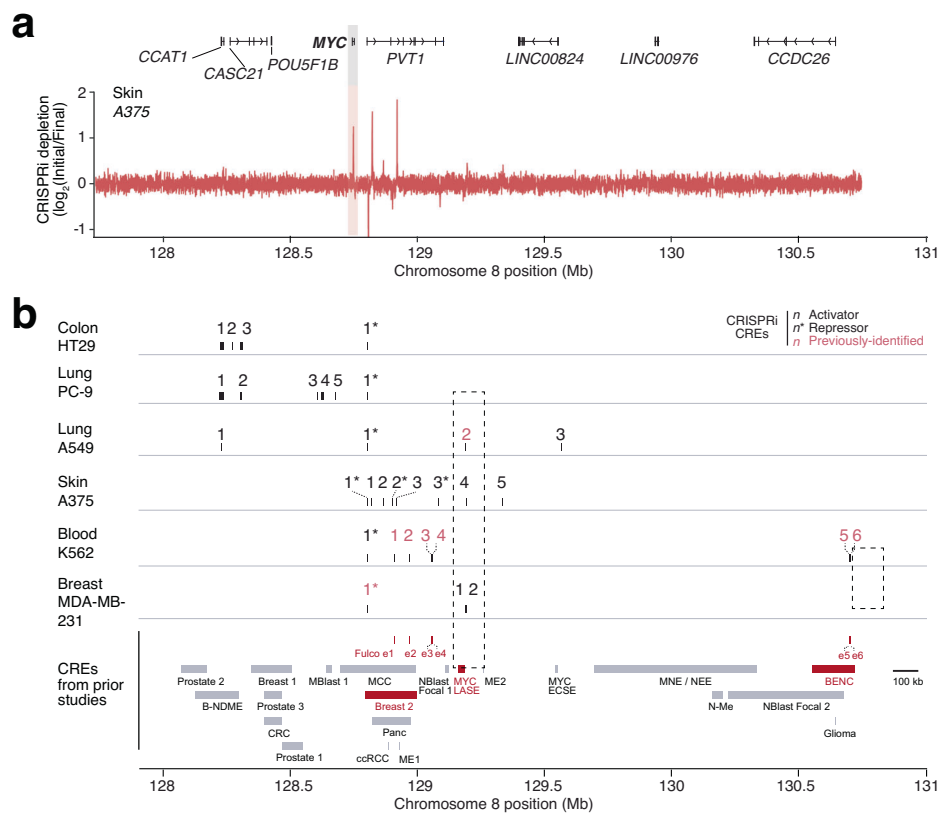


Fig. 3 | Cis-regulatory elements (CREs) in six different cancer cell lines using tiling CRISPRi screens. a CRISPRi log₂ fold-change in A375 melanoma cells for -112,000 gRNAs tiling the *MYC* TAD. Positive values indicate regions/perturbations that are highly depleted. The shaded area indicates the *MYC* promoter. **b** Detected enhancers (corresponding to depleted gRNAs over the -4 week cell culture period) and repressors (enriched gRNAs) in each CRISPRi screen for each cell line.

Repressor CREs are denoted by an asterisk after the CRE identifier. CREs from prior studies are those referenced in a review of *MYC* TAD CREs by Lancho et al.⁴⁴, and from Fulco et al.²⁸. For the CRISPRi screen CREs, those in red text match a previously *MYC* enhancer found in the same tissue/cancer type. Source data are provided as a Source Data file.

using uniform tiling of perturbations in different cell lines from different tissues—even for a relatively well-studied genetic locus like *MYC*.

We aligned the CREs from each cell line to previously identified enhancers⁴⁴; many of these enhancers were found via large genomic deletions or chromatin immunoprecipitation of enhancer-associated hallmarks (Fig. 3b, Supplementary Data 3). For these enhancers, it is not known whether all bases in these regions are required for CRE activity. In most cases, we find the CREs from the CRISPRi screens are smaller and provide better localization of key functional elements. For example, within *MYC*-LASE, we identify two separate CREs, one in a lung cancer and another in breast cancer. In MDA-MB-231 breast cancer cells, we find that both enhancers are present, whereas in other cells we find either one enhancer or the other is present but not both (e.g. in A549 and A375 cells). This highlights the ability of the CRISPR tiling approach to further elucidate functional subregions within even known enhancers, dissecting differential usage of distinct CREs contained within a larger enhancer cluster like *MYC*-LASE.

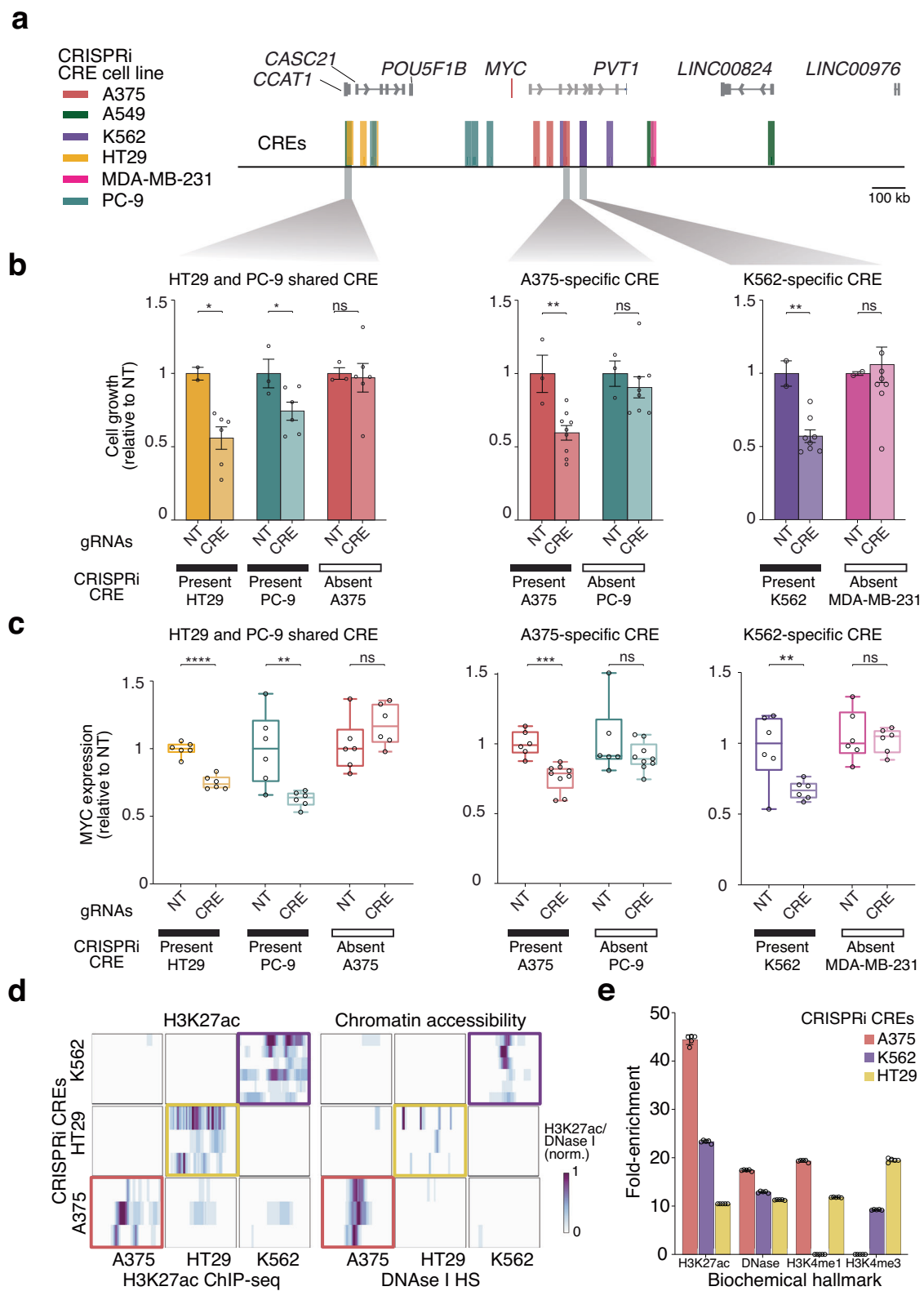
For detected CREs in K562 (leukemia), we found good correlation with a previous CRISPRi screen from Fulco et al.²⁸. (Supplementary Fig. 6a, b). We identified 6 prior enhancers with similar relative growth differences and an additional repressor-like region at the promoter of the noncoding RNA (ncRNA) *PVT1* (Supplementary Fig. 6c). This difference is possibly due to the Fulco et al. screen targeting specific regions within the *MYC* TAD instead of tiling the entire TAD.

CREs in the *MYC* locus regulate cell growth and *MYC* expression

We next sought to validate and measure changes in cell proliferation upon perturbation of identified CREs. We selected the top 3 CREs

across all screens (highest CRISPRi depletion scores) and targeted them in all cell lines where they were present and, as a negative control, a cell line where the CRE did not have a significant depletion (Fig. 4a). We targeted the center of each peak by selecting 2–3 distinct gRNAs located within 90 bp of the depletion maxima of each peak. After selection and expansion for -2 weeks, we measured cell growth via ATP quantification using a highly-sensitive luminescence assay. Compared to non-targeting (negative control) gRNAs, we found that targeting each CRE in the cell line(s) where identified in the primary screen resulted in a marked decrease in cell growth ($p < 0.05$, unpaired t test) (Fig. 4b). However, targeting the same region in cells where the TAD tiling screen did not identify any significant CREs resulted in no significant change in growth. For example, targeting a K562-specific (leukemia) CRE in MDA-MB-231 (breast) cells does not result in any difference in growth compared to non-targeting gRNAs. Also, for the shared CRE found in both HT29 (colon) and PC9 (lung) cells, we see significant decreases (relative to non-targeting) in cell growth in both of those cell lines but not in A375 (melanoma) cells, where this region was not significantly depleted in the *MYC* TAD tiling screen.

Given that CRE inhibition is expected to have more modest effects than complete gene loss, we measured *MYC* expression in cells transduced with the same viral constructs as in the cell growth assay. We found significant decreases in *MYC* expression upon CRE-targeting (compared to non-targeting gRNAs) only in those cell lines where the CRE was present in the TAD tiling screen (Fig. 4c). We find strong concordance between decreased *MYC* expression and decreased growth, suggesting that CREs identified in the TAD tiling screen work through modulation of *MYC* expression.



In cell lines with individually validated CREs that modulate growth and decrease *MYC* expression (K562, A375 and HT29/PC9), we next sought to understand the role of common biochemical hallmarks of enhancer activity: in particular, histone modifications and chromatin accessibility. Using windows surrounding each CRISPRi CRE in these cell lines, we examined the same genomic regions in other cell lines that did not possess those CREs (Fig. 4d). We found that these

biochemical hallmarks were present more often in those cells with a CRISPRi CRE at the same genomic region. To quantitatively test which biochemical hallmarks of enhancer activity predict functional CREs from the CRISPRi screens and to what degree, we used the Genomic Association Tester (GAT)⁸⁰, which assesses overlap via sampling of genomic intervals followed by empirical *p*-value determination. Through this, we found that these CREs significantly overlap with

Fig. 4 | CREs at the *MYC* locus tend to be cell-type specific and harbor H3K27ac and open chromatin biochemical signatures. **a** Locations of indicated CREs within the *MYC* TAD. **b** Cell proliferation (CellTiter Glo) for CRISPRi perturbations of CREs identified in the TAD-wide screen ($n = 3$ biological replicates of 2–3 gRNAs per CRE or non-targeting control). Error bars represent mean \pm s.e.m. The p values are determined by unpaired, one-sided t tests. * $p < 0.05$, ** $p < 0.01$. Peak p values: HT29/PC9 (HT29: 0.01, PC9: 0.028, A375: 0.84), A375 (A375: 0.004, PC9: 0.49), K562 (K562: 0.001, MDA-MB-231: 0.41). **c** *MYC* expression after CRISPRi perturbations of CREs identified in the TAD-wide screen ($n = 3$ biological replicates of 2–3 gRNAs per CRE or non-targeting control). Boxplot whiskers extend from minimum to maximum, the box extends from 25th to 75th percentile and the mid-line

represents the median. The p values are determined by unpaired, one-sided t tests. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Peak p values: HT29/PC9 (HT29: 0.000004, PC9: 0.004, A375: 0.08), A375 (A375: 0.002, PC9: 0.49), K562 (K562: 0.009, MDA-MB-231: 0.25). **d** Heatmap of H3K27ac and chromatin accessibility from specific cell lines (x -axis) for CREs found in the indicated cell lines (y -axis). Regions range from 115 to 215 kb. H3K27ac and chromatin accessibility biochemical signatures are rarely found in cell lines where the CRE is not functional. **e** Overlap of CREs with enhancer and promoter-associated biochemical signatures ($n = 5$ association tests with 2000 samples per test). Error bars represent mean \pm SD of association tests. Source data are provided as a Source Data file.

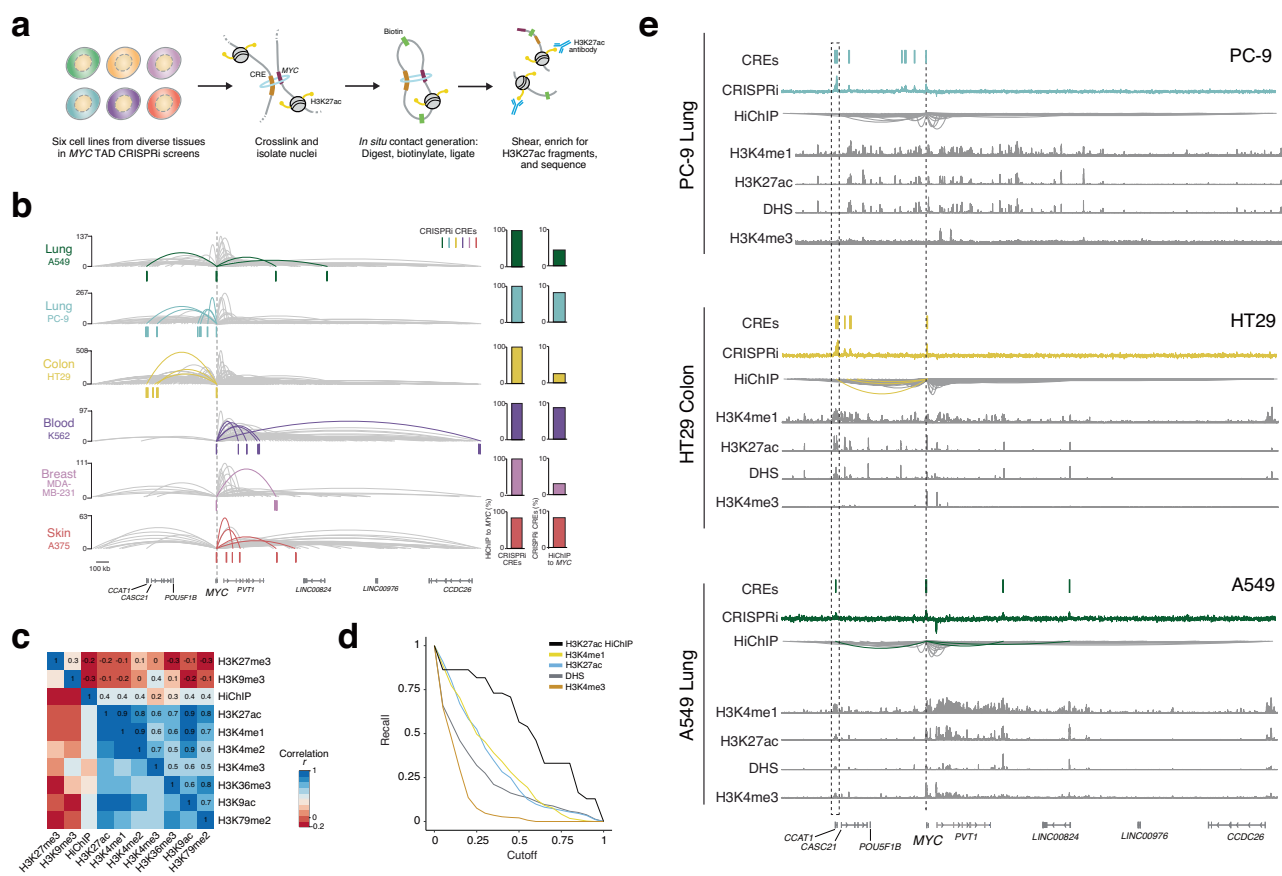


Fig. 5 | Cell type specific CREs have significant physical (HiChIP) interaction with the *MYC* promoter. **a** Schematic of H3K27ac HiChIP workflow. **b** (left) H3K27ac HiChIP contacts at 20 kb resolution between *MYC* and other elements in the TAD. Highlighted in color are the significant CREs identified in each CRISPRi screen and their HiChIP contacts with *MYC*. (right) The percent of CRISPRi-identified CREs that contact *MYC* (first column) and the percent of total contacts with *MYC* that include CRISPRi CREs (second column). **c** Pearson correlation of

H3K27ac HiChIP from A549 cells with histone modifications. **d** Recall analysis of gRNA sequences within CREs by the indicated biochemical signature. Plot represents an average across all cell lines. **e** Overlap of H3K27ac HiChIP and CRISPRi perturbation screens for A549 (lung), PC-9 (lung) and HT-29 (colon). These three cell lines share a CRE (dotted box) at the promoter of *CCAT1*. The single dotted line indicates *MYC*. Source data are provided as a Source Data file.

H3K27ac, H3K4me1 and chromatin accessibility (Fig. 4e, Supplementary Fig. 7a). In general, we find that, despite H3K4me1 being more widespread overall in the *MYC* TAD, there is greater and/or more consistent overlap with H3K27ac and DHS (Supplementary Fig. 7b, c). For example, for CREs located within the introns of the noncoding RNA *PVT1*, we find strong overlap between H3K27ac, chromatin accessibility and CRISPRi CREs (Supplementary Fig. 7d).

CREs show significant interaction with the *MYC* promoter and are transcribed

Transcriptional regulation of target genes by CREs is thought to be mediated by physical contact or looping between the CRE and the proximal promoter region. To determine whether the CREs identified in

the CRISPRi screens were in contact with the *MYC* promoter, we performed H3K27ac HiChIP⁸¹ on the six screened cell lines (Fig. 5a, b). Since H3K27ac was present at all CRISPRi CREs, we focused on 3D contacts with H3K27ac-decorated enhancer regions (Supplementary Fig. 7c). We identified millions of unique enhancer contacts in each cell line and found a high correlation between biological replicates of the same cell line but lower correlations between different cell lines (Supplementary Fig. 8a–c). As expected, the two lung-derived cell lines (PC9 and A549) had a much higher correlation ($r = 0.75$) compared to all other pairs of cell lines ($0.51 \leq r \leq 0.66$). We also found that the HiChIP had a positive correlation for known enhancer marks like H3K27ac and H3K4me1 and a negative correlation with repressive marks like H3K27me3 and H3K9me3 (Fig. 5c, Supplementary Fig. 8d, e).

Using the HiChIP contact mapping, we examined whether the CRISPRi CREs contacted the *MYC* promoter. We found that all CRISPRi CREs, except for one, were in contact with the *MYC* promoter (Fig. 5b). Surprisingly, we found that the opposite was not true: That is, of all contacts with the *MYC* promoter, less than 10% of them were identified as CREs in the CRISPRi screens. Notably, CRISPRi CREs have a ~2-fold greater contact strength/frequency than other HiChIP contacts with the *MYC* promoter (Supplementary Fig. 8f, g). We find that the CREs identified by the screens are more likely to contact the *MYC* promoter than any other enhancer associated biochemical features (Supplementary Fig. 8h). In line with these observations, we found that HiChIP –which combines H3K27ac enhancer activity with physical contact– better predicts functional CREs than histone marks or chromatin accessibility alone (Fig. 5d, Supplementary Fig. 9). This agrees with prior work proposing that enhancer activity and 3D contact (“activity-by-contact”) are required for functional CREs⁸².

Over the past two decades, several studies have established pervasive transcription of enhancer RNAs (eRNAs) from active promoters, enhancers and *cis*-regulatory elements^{83–85}. We sought to understand whether CRISPRi CREs were also transcribed. Using an eRNA dataset that captures the 5' ends of elongating transcripts without the need for pausing RNA polymerase or labeled nucleotides (NET-CAGE)⁸⁶, we observe that most, though not all, CREs generate eRNAs and that they tend to be strand-specific with one strand typically expressed much more highly than the other (Supplementary Fig. 10a, b), in line with prior observations of eRNAs^{87,88}. Ranking eRNAs across the *MYC* TAD by expression level, we find that CRE-containing regions tend to rank highly, although lower than promoters of genes like *MYC*, *PVT1* and *CCDC26* (Supplementary Fig. 10c, d). However, there is not a linear relationship between CRE functional activity and transcription of corresponding eRNAs: For example, in K562, there are several CREs identified within the *PVT1* gene body but the strongest CRISPRi CRE (CRE 2) drives significantly less eRNA transcription than another CRE (CRE 3) with about half the functional activity in the screen (Supplementary Fig. 10a, b). This analysis reveals that a subset of CREs detected in the CRISPRi screens have strong eRNA transcription; however, eRNA transcription is not directly linked to functional impact on cell growth.

The noncoding RNA *CCATI* is required for the activity of its CRE

Across all screened cell lines, we find that many CREs identified in the CRISPRi screens overlap promoters of ncRNAs (Supplementary Fig. 11a). One key mechanistic question is whether other CREs in this region regulate *MYC* via their DNA elements (enhancer activity), or through their RNA transcripts, or a combination of mechanisms. For example, the ncRNA *CCATI* overlaps significant CREs in HT29, PC9 and A549 pooled screens and is most highly expressed in those cell lines (Supplementary Fig. 11b). It is located within a previously identified super enhancer and associated with positive regulation of *MYC*⁸⁹. There is biochemical evidence that suggests both DNA and RNA regulatory activity: In HT29, which has the highest *CCATI* expression of the screened cell lines, we found a significant physical interaction between the *CCATI* promoter CRE and *MYC* itself (H3K27ac HiChIP), which was also observed in a previous study using promoter-capture Hi-C⁹⁰ (Supplementary Fig. 11c).

Recently, Xue and colleagues developed RNA in situ conformation sequencing (RIC-seq) to identify interactions between RNA transcripts, including noncoding RNAs⁹¹. Analysis of RIC-seq data from HeLa cells (which also express *CCATI*) shows strong RNA interactions between the *CCATI* transcript and the *MYC* transcript, suggesting that the RNA itself might be important for enhancer activity (Supplementary Fig. 11c). To dissect the role of the noncoding RNA transcript from the activity of the DNA elements of the CRE, we used three distinct approaches: targeting the CRE using DNA-targeting (Cas9) and RNA-targeting (Cas13) CRISPRs, performing physical genome conformation

mapping (H3K27ac HiChIP) after *CCATI* RNA knockdown and examining changes in CRE looping, and measuring whether the impact of *CCATI* silencing could be rescued by overexpression of *CCATI* in trans.

Cas13 is a recently characterized RNA-targeting CRISPR nuclease that cleaves single-stranded RNA transcripts in a strand-specific manner and, as recently shown, it is effective in knockdown of both coding and noncoding transcripts^{46,92}. We designed optimized Cas13 guide RNAs that target the *CCATI* transcript and compared the reduction in *CCATI* expression to CRISPRi at the *CCATI* CRE^{45,93} (Supplementary Fig. 12a). We found that both CRISPRi and Cas13 reduced *CCATI* transcript levels ~60–70% (Supplementary Fig. 12b) and that *CCATI* knockdown with Cas13 decreased cell growth (Supplementary Fig. 12c) similarly to CRISPRi at the *CCATI* CRE (Fig. 4b).

To investigate the functional role of *CCATI* RNA and possible changes in 3D genome conformation, we targeted the *CCATI* transcript with Cas13 and performed H3K27ac HiChIP. We found that the *CCATI* RNA impacts DNA looping between the *CCATI* and *MYC* promoters (Supplementary Fig. 12d). The DNA looping between *CCATI* and *MYC* is reduced by 60% upon *CCATI* RNA knockdown (Supplementary Fig. 12e). In fact, across the entire genome, the *CCATI*-*MYC* HiChIP loop has the largest decrease in strength and the HiChIP loops with the next largest losses of interaction after *CCATI* RNA knockdown (*FAM91A1-ANXA13* and *TRAPPC9-CHRAC1*) are both on chromosome 8 – the same chromosome as the *MYC* TAD (Supplementary Fig. 12f, g). Further, *ANXA13*, *TRAPPC9*, and *CHRAC1* are associated with tumorigenesis, metastasis, and/or poor outcomes in colon cancer^{94–98}, suggesting that the *CCATI* lncRNA might promote changes in genome structure for other nearby colon cancer-associated loci.

In light of these results that *CCATI* RNA plays a role at the DNA level (i.e. looping), we wondered whether an exogenous *CCATI* RNA could rescue the decreased growth phenotype observed after endogenous *CCATI* silencing. To do this, we first generated an expression vector expressing a *CCATI* lncRNA under a strong constitutive promoter. After *CCATI* silencing via CRISPRi, HT29 cell growth was decreased to 65% compared to the population that was not transduced (Supplementary Fig. 12h). We found the cells with *CCATI* silenced and expressing a (non-targetable) *CCATI* transgene were rescued but not completely (84% of control cell growth). This suggests that upon *CCATI* knockdown, *CCATI* RNA delivered in trans can restore at least partial function.

A recent study found that many transcription factors (TFs) also harbor RNA-binding domains and this suggests a possible mechanism by which ncRNAs regulate target genes such as *MYC*⁹⁹. These results suggest that the *CCATI* RNA plays an essential role in driving the activity of the CRE. More broadly, the overlap between many CRISPRi-identified CREs at the *MYC* locus and ncRNA promoters suggests that this might be a more general principle of gene regulation.

CREs across cell lines harbor binding sites for unique TFs

The CRISPRi screens yielded 32 distinct CREs across the 6 cell lines that potentially act to transcriptionally regulate *MYC* through interactions with transcription factors (TFs). Multiple experimental^{100–102} and computational¹⁰³ studies have attempted to characterize master TFs that drive growth and constitute a cell's identity. We hypothesized that the unique functional CREs in cells derived from different tissue types might be bound by characteristic TFs of those tissues or cancers (Fig. 6a). To test this, we performed an association test between the 32 CREs and each of the 746 human TF motifs in the JASPAR database. Briefly, we compared the overlap of these motifs to a null distribution generated using scrambled motifs (Supplementary Data 4). After correction for multiple comparisons, we found that ~70% of enriched TF motifs were unique to each cell line and not shared (Fig. 6b, c, Supplementary Fig. 13a). This result agrees with our prior finding that most CREs are cell-specific (Fig. 2d). The enrichment of tissue-specific TFs in CRISPRi CREs was also apparent when examining the top gene

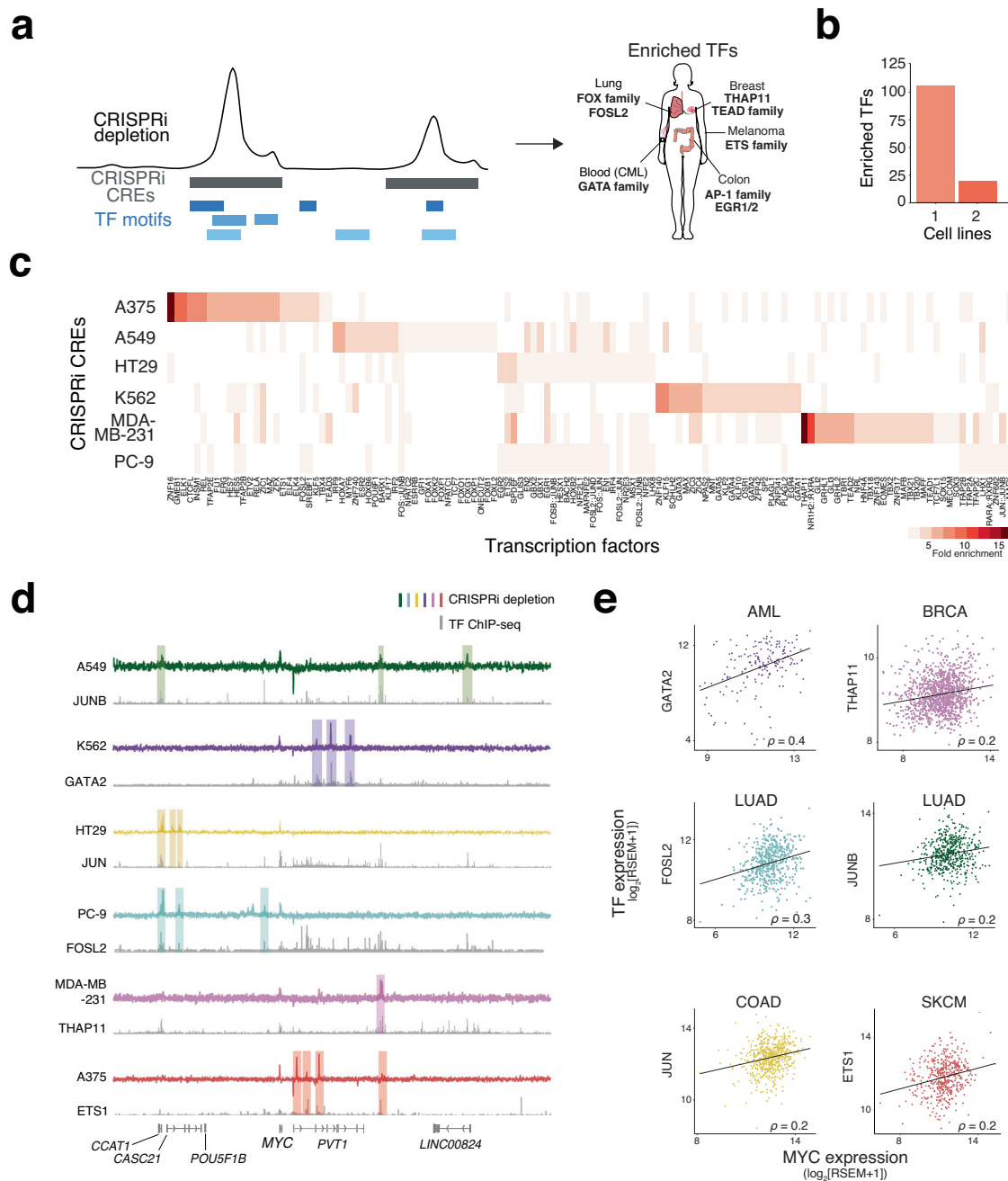


Fig. 6 | Predicted transcription factors at CREs in the *MYC* locus are tissue specific. **a** Identification of transcription factor binding at CREs and summary of key transcription factors found. **b** Number of transcription factors significantly enriched in the CREs from 1 or more cell lines. **c** Fold enrichment for top 25 transcription factors in each cell line. **d** Transcription factor ChIP-seq for selected transcription factors from panel c. JUNB, JUN, and FOSL2 ChIP-seq are from A549, GATA2 ChIP-seq is from K562, THAP11 ChIP-seq is from HepG2 and ETS1 ChIP-seq is

from SCC25. **e** Correlation in primary tumors from The Cancer Genome Atlas (TCGA) between *MYC* expression and top-ranked transcription factors from panel (c). TCGA tumor types are matched to the same cell/tissue type as CRISPRi cell lines in which the transcription factor is enriched ($n = 2855$ TCGA primary tumors; LUAD: $n = 512$, COAD: $n = 594$, SKCM: $n = 472$, AML: $n = 175$, BRCA: $n = 1102$ tumors). Source data are provided as a Source Data file.

ontology (GO) terms associated with these TFs (Supplementary Fig. 13b), such as morphogenesis of lung (in A549) and of mammary gland (in MDA-MB-231). For a subset of these enriched TFs where ChIP-seq data was available, we found excellent correspondence between motif-based analysis and experimentally measured TF binding (Fig. 6d, Supplementary Data 4).

Given the enrichment for specific TFs in CREs, we reasoned that if these TFs regulate *MYC*, then we would also predict that tumors with greater expression of these TFs should also have higher levels of *MYC*. To test this, we quantified expression of TFs in primary tumors of the

same tissue of origin as the cell line containing CREs where the TF motif was enriched, along with *MYC* ($n = 2948$ blood, breast, colon, lung and skin primary tumors from The Cancer Genome Atlas [TCGA]). We identified several CRE-enriched TFs with a positive correlation between TF expression and *MYC* expression ($0.2 < \rho < 0.4$) (Fig. 6e, Supplementary Fig. 14a–f, Supplementary Data 4, 5).

This suggests that these TFs may act through the binding sites identified via ChIP or motif matching in these CREs to regulate *MYC*, although it is possible that these TFs drive *MYC* expression in both cancerous and normal tissue. To identify those TFs that regulate *MYC*

specifically in malignancy, we assembled ~2000 normal (non-cancerous) primary tissue transcriptomes from the Genotype-Tissue Expression (GTEx) project and TCGA and analyzed correlations between TF and *MYC* expression. We found 97 TF-tissue pairs where the TF correlation with *MYC* was higher in tumor samples than in normal tissue samples ($n=1932$ blood, breast, colon, lung and skin tissues) (Supplementary Fig. 15a, Supplementary Data 4) and also identified 37 TF-tissue pairs where the TF correlation with *MYC* was fivefold higher in tumor samples than in normal samples (Supplementary Fig. 15b). These TFs represent distinct transcription factor families identified in each cell line, and potentially can be viable targets for mitigation of cancer growth, given the difficulty of developing drugs that target *MYC* directly⁶⁴.

Discussion

Recent advances in genome editing technology, namely CRISPR systems, have enabled researchers to manipulate genomic elements in order to understand which ones impact key phenotypes and how precisely they work. To date, however, most studies of CREs have been limited to a single cell type, and it is unclear whether these elements are unique to the cell type studied or are shared among many distinct cell types. To answer this question, we have used pooled CRISPR inhibition (CRISPRi) screens to identify and characterize the CREs of the oncogene *MYC* across six cancer cell lines in a systematic, high-resolution manner. These CREs modulate cell growth and do so via changes in *MYC*. This is likely via enhancer activity: Nearly all are in physical contact with the *MYC* promoter, despite being located up to 1.8 Mb away, and contain binding sites for characteristic, lineage-associated transcription factors for each respective tissue/cancer.

Over the past decade, many studies have identified enhancers that regulate *MYC*, such as the E1-E5 cluster, N-ME and BENC in leukemias^{36,37,42}, LASE and ECSE in epithelial cancers³⁹, and others^{38,40,43}. Several of them have relied on indirect evidence of enhancer activity^{42,43}, such as ChIP-seq of H3K27ac, MED1 or BRD4, and those that have used perturbations to show a causal relationship with gene expression have done it with only one or a handful of perturbations^{36–41}. For these studies, it can be challenging to compare between those that use different kinds of perturbations (e.g. mutagenesis, large deletions, and tumor somatic mutations). A hallmark of our study is the uniform set of ~112,000 perturbations across 6 cell lines.

We also use exclusively human models: The majority of prior studies that use genetic perturbations do so in mouse^{36–38,40}, which has poor noncoding sequence conservation with human (<50%)¹⁰⁴. These mouse models typically have low resolution, since they employ large deletions induced via *Cre* recombination. Orthologous noncoding regions of the genome do not always have conserved function when tested with perturbation assays: For example, we previously showed that CRISPR perturbations of the mouse ortholog of the well-studied DHS + 58 enhancer of *BCL11A*—a therapeutic target for hemoglobin disorders—do not modulate fetal hemoglobin in the same manner as the human enhancer²⁹. Thus, for the study of human cancer and gene regulation, the use of human models is preferable where possible.

Our study uses a single method (tiling CRISPR silencing) to investigate CREs systematically and thoroughly in the *MYC* TAD across many different cell lines, yielding a more comprehensive enhancer discovery at the *MYC* locus. This unified approach can help develop unique insights: For example, we discover that a previously known colon cancer enhancer is not only present in the colon cancer cell line, but also in a lung cancer cell line. The scale of these noncoding screens at the *MYC* locus is vast: We perturb every base pair in the *MYC* TAD across 6 human cell lines to systematically discover known and new enhancers.

We find that CRISPRi CREs are strongly associated with certain biochemical features, such as enhancer-associated histone marks and chromatin accessibility. However, neither these marks nor 3D proximity with a gene target (as measured by H3K27ac HiChIP), are sufficient to conclude that a CRE is functional. In a subset of CRISPRi-identified CREs,

we performed an in-depth validation showing that they alter cell growth and *MYC* expression. In cell lines without these CRISPRi CREs, genetic perturbations have no effect on either cell growth or gene expression. One important limitation of this work is that each CRE was examined individually and, thus, it is difficult to disentangle whether these CREs may act redundantly or synergistically. Future work with simultaneous perturbation of multiple CREs or a combination of activation and inhibition at different CREs could help address this question.

Similarly, we cannot be certain whether *MYC* is the only target of CREs identified in the functional screens. A limitation of the CRISPRi screens is that it can only indicate that a region has an effect on cell growth when perturbed, and it is not possible to see which genes are specifically affected. An important future direction would be to directly measure gene expression when perturbing each region (e.g. ECCITE-seq, Perturb-seq, STING-seq) in order to better understand which genes are being targeted. Although genome-scale Perturb-seq data exists for protein-coding genes¹⁰⁵, a similar atlas for CREs would be an enabling resource for enhancer genomics.

The *MYC* TAD is a unique region: It contains several noncoding RNAs but only two protein coding genes (*MYC* and *POU5F1B*). The majority of CREs identified in the CRISPRi screens overlap promoters of ncRNAs, which pose a unique challenge in understanding whether DNA-based CREs or ncRNA transcripts are key drivers of *MYC*. One ncRNA promoter that we found overlapping CRISPRi CREs in three cell lines is *CCATI*, a highly expressed lncRNA located within a super enhancer and associated with positive regulation of *MYC*⁸⁹. Using a combination of DNA- and RNA-targeting CRISPRs, native genome structural mapping, and rescue of ncRNA expression in trans, we found that the *CCATI* locus regulates *MYC* not only via the enhancer's DNA elements but also via the *CCATI* transcript itself. It remains an open question whether this ncRNA activity is specific to *CCATI* or a more general principle of ncRNAs at CREs. Future noncoding RNA-directed perturbations—using CRISPR systems like Cas13 or RNA interference methods^{45,46,106,107}—will be needed to pursue these mechanisms underlying gene regulation.

The dysregulation of *MYC* is a hallmark of many different cancer types^{55,108} and *MYC* has been considered “undruggable” due to lack of targetable active binding sites for ligands⁶⁴. Targeting individual CREs, as we have shown, has the potential to attenuate gene expression by as much as 50%, which might be sufficient for selective eradication of cancer cells, and, in fact, preferable, considering that *MYC* is necessary for normal cell growth as well¹⁰⁹. In the future, single and combinatorial CRE targeting might yield effective strategies to selectively target this important oncogene in a cell-type specific manner.

Methods

Oncogene regulatory diversity analysis

Histone ChIP-seq and DNase-seq were downloaded as BAM files from the Encyclopedia of DNA Elements (ENCODE⁸) (see Supplementary Data 1). For each cancer cell line, we downloaded H3K27ac, DNase-seq and H3K4me1 datasets. We generated Bigwig coverage files for each dataset using deeptools 3.4.2¹¹⁰, with normalization for sequencing depth using the RPGC method. We normalized resulting bigwigs by the maximum read value and fed into genomation¹¹¹ for scoring. The oncogenes profiled were selected from the Catalog Of Somatic Mutations In Cancer (COSMIC) database⁴⁷ (accessed Nov. 7, 2017) as those having the labels of “oncogene” or “oncogene, fusion”. In total, this yielded 386 oncogenes. We defined windows of 2 Mb surrounding each oncogene's transcription start site (TSS), with 1 Mb on each side. This range was chosen because prior studies have indicated that most enhancers are located within 0.5–1 Mb from target gene TSSs^{48,49}.

Using the windows and bigwig files, we ran genomation's ScoreMatrixList to obtain the feature signal (a vector of values r for each 100 kb bin over the 2 Mb region surrounding the oncogene's TSS). To understand the diversity of the DNA landscape surrounding

oncogenes, we computed the summed, pairwise Euclidean distances between each oncogene's feature signal \mathbf{r} across all cell lines: $ORD = \sqrt{\sum_{i,j}^{cell\ types} (r_i - r_j)^2}$, where i and j are unique pairwise cancer cell line comparisons. For example, using H3K27ac ChIP-seq data for the 2 Mb surrounding *MYC* in HCT116, A375 and PC3 cell lines, the sum of all the distances between each of these vectors in a pairwise manner results in the ORD score for *MYC* H3K27ac. We call this the oncogene regulatory diversity (ORD) score. This is done for each of the 386 oncogenes for 3 enhancer biochemical signatures (H3K27ac, DHS and H3K4me1), across 12 cancer cell types (Supplementary Fig. 1).

For each feature, we rank the oncogenes based on ORD score, with the highest score resulting in the top rank. The overall rank for all oncogenes was determined by aggregating the ranks from each feature using Robust Rank Aggregation¹¹². Windows containing regions that overlap with other oncogene TSSs were kept, due to previous findings that promoters can also act as enhancers for other genes (e.g. shadow enhancers)^{28,113,114}. Ranks for all bin sizes can be found in Supplementary Data 2.

Cell culture conditions

We acquired cell lines from ATCC (except HEK293FT, which were from ThermoFisher) and maintained with the media listed (Supplementary Data 6), along with puromycin concentrations for selection. All cell lines were cultured at 37 °C, 5% CO₂, and ambient oxygen levels.

MYC knockdown using Cas13

Monoclonal doxycycline-inducible *RfxCas13d* MDA-MB-231, K562, A549, and HT29 cells were generated by transducing cells with a lentivirus produced using pLentiRNACRISPR_007⁴⁵ (Addgene 138149) at a low multiplicity of infection (MOI < 0.1) and selected with 5 µg/ml of blasticidin S (A.G. Scientific B-1247). Single-cell colonies were isolated by low-density plating and then expression of HA-tagged Cas13 was confirmed by immunoblot using an anti-HA peptide antibody (Cell Signaling Technology 2367S). *RfxCas13d*-NLS A375 monoclonal cells were obtained from Wessels and Méndez-Mancilla et al.⁴⁵.

We cloned a nontargeting guide RNA and a guide RNA targeting the *MYC* transcript (Supplementary Data 7) into an *RfxCas13d* guide-only vector that also expresses GFP and puromycin resistance, pLentiRNAguide_004¹¹⁵ (Addgene 223175). To produce lentivirus, we seeded 900,000 HEK293FT cells per well (6-well dish) and transfected them with 5.5 µl polyethylenimine linear MW 25000 (Polysciences), 1 µg gRNA transfer plasmid (guide RNA in pLentiRNAguide_004), 0.8 µg psPAX2 (Addgene 12260) and 0.5 µg pMD2.G (Addgene 12259). Three days post-transfection, the viral supernatant was collected, filtered through a 0.45-µm filter, and monoclonal Cas13 cell lines were transduced at a low MOI (~0.5). After 24 h, we replated cells in 24-well plates with media containing 1 µg/mL doxycycline to induce *RfxCas13d* expression (Supplementary Data 6). Plates were imaged using an Incucyte SX3 at 10× magnification and acquiring 9 images per well every 4 h for 72 h.

Cis-regulatory element library design and cloning

We selected all Cas9-targetable sites contain the protospacer-adjacent motif (PAM) NGG on both strands within the topologically associated domain containing *MYC* (genome: hg19, coordinates: chr8:127,734,994-130,746,056). We eliminated gRNAs with the Doench on-target score < 50 (as previously described¹¹⁶), as well as those with homopolymer repeats (>4), yielding 112,162 gRNAs a median of 9 bp apart (Supplementary Fig. 3).

As a positive control, we included ~2000 essential gene targeting gRNAs (targeting 490 essential genes) from the TKOv3 genome-wide library⁷⁵. Essential genes were identified as genes that overlapped between two studies of core essential genes across multiple cell types^{35,75}. As negative controls, we included approximately 2000 nontargeting

gRNAs and 2000 nonessential regulatory element targeting gRNAs²⁸. The gRNAs were synthesized as a pool (Twist Biosciences). We amplified and cloned the Twist pool as previously described¹¹⁷ using Gibson assembly into an all-in-one KRAB-dCas9 vector with the F + E modified sgRNA scaffold (pCC_09, Addgene 139094)¹¹⁸ and concentrated the ligation products using isopropanol precipitation. We electroporated the concentrated ligation product with Endura cells (LGCBiosearch 60242-2) in multiple transformations and then pooled together and plated on square 245 × 245 mm BioAssay plates (Corning 431111) with LB / agar 100 µg/ml Carbenicillin. Using dilution plating, we measured a representation of 500 colonies per gRNA. The bacteria were harvested after 14 h of incubation at 37 °C, scraped with 5-10 ml of LB media and processed with a PlasmidPlus maxi prep kit (Qiagen 12965). We used 1 maxi prep column per 0.25 g of bacteria.

We then sequence verified the library. To prepare the library for sequencing, we performed a 2-step PCR reaction (nested PCR) with primers listed in Supplementary Data 7, using 10 ng of the library with NEBNext High Fidelity PCR Master Mix (NEB M0541L). The PCR reaction first amplified the region on the plasmid containing the guide, and second, added on Illumina adaptors and unique barcodes. The protocol was as follows: 98 °C for 30 s, then 10x: [98 °C for 10 s, 66 °C for 30 s, 72 °C for 15 s], 72 °C for 2 min, incubate at 4 °C. Full library sequences and details can be found in Supplementary Data 8.

Pooled lentiviral production and screening

HEK293FT cells were plated in Dulbecco's Modified Eagle Medium (DMEM) + 10% FBS (D10) in a T225 flask so that the following day, cells are 90% confluent. Half of the media was removed from the flask, and cells in each flask were transfected with 25 µg of transfer vector (e.g. CRE-targeting library), 13.6 µg pMD2.G (envelope plasmid), and 20 µg psPAX2 (packaging plasmid) using 2.5 mL Opti-MEM and 136 µl polyethylenimine linear 25 K (Polysciences 23966). The day after, fresh D10 media is added to full volume. Then, 48 h later, we collected the viral supernatant and put it immediately on ice. We concentrated the supernatant by centrifugation at 100,000 × g (Thermo Sorvall LYNX) for 2 h at 4 °C. The resulting pellet was resuspended in cold DMEM and stored at -80 °C until use.

We determined the appropriate titer of virus before each experimental transduction. We transduced 3 M cells with a standard spinfection protocol with different dilutions of virus in a 12-well plate as well as a no virus control well. After adding virus, we spun the cells at 2000 rpm for 2 h at 37 °C (Beckman Coulter Allegra X-14R) and incubated overnight. The next day, we plated half of the cells in each well into two new wells of a 6-well plate. In one set of wells, we added the appropriate puromycin concentration for the cell line (Supplementary Data 6). After all the cells in the no virus well had died, cells in the corresponding wells (with puromycin) were counted to determine the viral volume that results in 20%–40% cell survival, corresponding to a multiplicity of infection (MOI) of 0.2–0.5.

We cultured each cell line in the appropriate media (Supplementary Data 6) and transduced ~2 × 10⁸ of them with the CRISPR lentiviral library via spinfection with the viral volume determined from the previous spinfection. As before, after adding virus, we spun cells at 2000 rpm for 2 h at 37 °C and incubated them overnight. After 24 h, we passaged cells in 5-layer flasks (NEST Scientific 731002) and selected with the appropriate puromycin concentration (Supplementary Data 6) for 4 weeks (28 days) and split at ~80% confluence. Also, at the first passage post-selection, we spun down cells and froze for genomic DNA extraction. The overall representation was ~500 cells per construct with ~83% of surviving cells receiving a single gRNA construct (single infection percentage calculated via Poisson estimation¹¹⁹).

CRISPRi pooled screens library preparation and data analysis

We isolated genomic DNA from cells using a modified salting-out precipitation¹¹⁷. The gRNA readout was performed using two rounds of

PCR. For PCR1, we used 10 μg of gDNA in each 100 μl reaction. We pooled the PCR1 products and used the mixture for a second PCR reaction. This second PCR adds on Illumina sequencing adaptors and barcodes. We ran 1 PCR2 reaction for every 10,000 gRNA in the library for a total of 12 PCR2 reactions per biological conditions. We performed PCR1 reactions using TaqB polymerase (Enzymatics P7250L) and PCR2 reactions with Q5 (NEB M0491). We pooled and purified PCR2 reactions with a QIAquick PCR purification kit. We ran the resulting product through a 2% E-gel, from which we extracted a band of 250–270 bp and purified it with a QiaQuick Gel Extraction kit (Qiagen 28704). We quantified the concentration of the gel-extracted PCR products using Qubit dsDNA HS Assay Kit (Thermo Fisher Q32851), then diluted and sequenced it on an Illumina MiSeq using a v3 kit (Illumina MS-102-3001). This was followed by sequencing on an Illumina NextSeq using a NextSeq 500 High Output v2 kit (150 cycles, 20024907).

We demultiplexed the samples using bcl2fastq (Illumina), trimmed off adapters and aligned to hg19 with bowtie2 with options $-v 1 -m 1$. We normalized the resulting reads by the sample read depth (each read divided by the total number of reads). Then, for each screen, we calculated a depletion CRISPRi score for each gRNA: \log_2 fold change between the early and final timepoints. We then smoothed the signal of each screen by taking the median value of a sliding window (20 gRNAs). To identify significant windows, an unpaired, two-sided T test was performed by comparing the scores of gRNA sliding windows with those of the nontargeting negative control gRNAs. Significant gRNAs are those that score below an adjusted FDR threshold of 0.05 (Benjamini-Hochberg) across all screens (Supplementary Fig. 5). All fold change values for all screens can be found in Supplementary Data 9. We assessed the empirical false discovery rate (FDR) of screen windows by comparing T statistics generated from the screen vs. nontargeting guide RNAs to those from negative control regions vs. nontargeting guide RNAs. These negative control regions span three negative control regions that are located far from known essential genes²⁸. This established a threshold based on an FDR of 0.05. We identified significant CRISPRi screen peaks (CREs) by merging significant gRNAs that were within 500 nt of each other. A valid peak, therefore, was defined as one with gRNAs no more than 500 nt apart from each other, and at least 10 gRNAs within it.

Individual CRE peak targeting, proliferation and digital and quantitative PCR

We identified the top 3 peaks across all screens that showed the highest CRISPRi depletion scores. From these peaks (located in the cell lines A375, HT29/PC9, and K562), we chose 3 guide RNAs located at the maxima of each peak. For HT29/PC9 and K562 peaks, we cloned guide sequences into pCC₀₉ and lentivirally transduced the constructs into cell cultures in 2 infection replicates. After 24 h, we replaced the media with puromycin media (Supplementary Data 6). We titered viral volumes such that 20–40% of cells survived after puromycin selection. After selection and expansion for ~2 weeks in puromycin, we plated cells for DNA/RNA extraction for digital PCR (dPCR) and Cell Titer Glo. For the A375 peak, we cloned guide sequences into pCC₀₉ and lentivirally transduced the constructs into cell cultures in 3 infection replicates. After 24 h, we replaced the media with puromycin media (Supplementary Data 6). We titered viral volumes such that 20–40% of cells survived after puromycin selection. After selection, cells were cultured for 3 days and collected for qPCR. After selection and expansion for ~2 weeks in puromycin, we plated cells for Cell Titer Glo.

For qPCR and dPCR, we extracted RNA with Trizol (Thermo Fisher 15596026), then column purified it with a Direct-zol RNA Microprep kit (Zymo R2050). We quantified RNA concentration by spectrophotometry (NanoDrop). To measure gene expression, we performed dPCR (Qiagen QIAcuity 4) with Cys/lowa Black RQ target gene probes (IDT), FAM/ZEN/lowa Black FQ for the B2M normalizer (IDT), and Luna Universal One-Step RT qPCR Master Mix kit (NEB E3005L) and Tween-

20 (Sigma-Aldrich P1379). All dPCR primer and probe sequences are in Supplementary Data 7. For each sample, we first normalized the target gene expression by B2M expression per sample. Then, to compare expression across samples, we normalized each sample by the average value of the non-targeting (negative control) gRNAs. For qPCR, we used Taqman Universal PCR Master Mix (Thermo 4304437), HPRT1 endogenous control (VIC/MGB probe) (Thermo 4326321E) and for gene probes we used Taqman Gene Expression assay (Thermo 4453320 and 4331182).

For Cell Titer Glo, we plated ~2000 cells in 100 μl of media in, flat, black-walled 96-well plates (Corning 3904). We waited for the fastest growing cells to reach ~80% confluency, then lysed all cells using a 1:4 dilution of Cell Titer Glo and PBS. We placed the plates on an orbital shaker for 2 min, and, after 5 more minutes, we quantified the luminescence signal with a Tecan Infinite 200 Pro Plate Reader.

Association of CRISPRi screen with enhancer-associated features

We downloaded all tracks from sources indicated in Supplementary Data 1. We aligned the tracks to hg19 with bowtie2 2.2.8²⁰. The resulting bam files were sorted and removed of duplicates. We then generated bigwig coverage files using deeptools 3.4.2, with normalization for sequencing depth using the RPGC method. We then used them as input for genomation for scoring and heatmap generation. Heatmap windows were defined as significant enhancer peaks found across each cell line. These peaks were the same as those defined in *CRISPRi pooled screens library preparation and data analysis*.

We used genomic association tool (GAT) 1.3.4 to associate significant gRNAs in each cell line (as defined in *CRISPRi pooled screens library preparation and data analysis*) with called peaks for corresponding histone mark and accessibility tracks. We downloaded significant peaks for histone and accessibility tracks from sources indicated (Supplementary Data 1). The workspace for GAT was defined as the entirety of the MYC TAD / CRISPRi screen region. GAT was run 5x with 2000 iterations per run, for a total of 10,000 iterations per association test.

H3K27ac HiChIP sample preparation and analysis

For H3K27ac HiChIP libraries³¹, we grew NIH3T3 (mouse) cells and human cell lines A375, A549, HT29, K562, MDA-MB-231 and PC-9 in the appropriate media (Supplementary Data 6). We fixed cells in 1% formaldehyde for 10 min and quenched to a final concentration of 125 nM glycine. We mixed 2 million fixed mouse cells with 10 million of each fixed human cancer cell line. The cells were lysed in 0.5% SDS, quenched with 10% Triton X-100, and digested with *Mbol* (NEB R0147M). The DNA overhangs were blunted, biotinylated (ThermoFisher 19524016), and ligated. Nuclei were spun down, resuspended in nuclear lysis buffer and sonicated using a Covaris LE220 with the following conditions: Fill level 10, PIP 450, Duty factor 30, CPB 200. We incubated the sheared DNA with Dynabeads Protein A (ThermoFisher 10001D) for 2 h at 4 °C. We then placed the tubes on a magnet and the supernatant was kept. We performed immunoprecipitation with a cross-species reactive H3K27ac antibody (Active Motif 39133). The samples were incubated with the antibody overnight at 4 °C. We then washed, eluted and treated the samples and treated with Proteinase K. We purified the samples using Zymo DNA Clean & Concentrator. Biotin capture was performed with Dynabeads M-280 Streptavidin (ThermoFisher 11205D), followed by library preparation. We purified the amplified libraries with Illumina Sample Purification Beads. We sequenced the libraries using paired-end reads with either Illumina NovaSeq 6000 S2 Reagent Kit v1.5 (200 cycles, 20028315) or NextSeq 500 High Output v2 kit (150 cycles, 20024907) to generate 100–200 million read pairs per sample.

We mapped HiChIP paired end reads to hg19 using HiC-Pro¹²¹. Default settings were using to remove duplicate reads, identify valid

interactions, and generate contact maps. We identified statistically significant contacts using FitHiChIP¹²² at a 1% FDR. We used H3K27ac ChIP-seq data as a reference set of peaks in the FitHiChIP pipeline.

CCAT1 DNA (CRISPRi) and RNA (Cas13) perturbations

For RNA targeting, a doxycycline-inducible *RfxCas13d* HT29 cell line was generated as described in *MYC knockdown using Cas13*. We then designed a series of optimized Cas13 gRNAs that target the second exon of the CCAT1 transcript using cas13design (<https://cas13design.nygenome.org/>). Guide sequences were cloned into pLentiRNA-Guide_001 (Addgene 138150)⁷⁰. All constructs were confirmed by Sanger sequencing.

We lentivirally transduced guide constructs into the TetO-*RfxCas13d*-NLS HT29 cells with 3 biological infection replicates. After 24 h, we replaced the media with puromycin media (Supplementary Data 6). We titered viral volumes such that 20–40% of cells survived after puromycin selection. Then, after completed puromycin selection (72 h), we added 1 µg/ml doxycycline (Sigma D9891) to induce *RfxCas13d* expression. After 24 h (with doxycycline), we extracted DNA/RNA qPCR.

For DNA targeting, we chose 3 guide RNA sequences located at the maxima of HT29-1 (Supplementary Data 3) identified in the CRISPRi screen, which is located in the promoter of *CCAT1*. These gRNAs are the same ones used in the individual guide CRE targeting experiments (Supplementary Data 7). We cloned guide sequences into pCC_09 and lentivirally transduced the constructs (in parallel with the above Cas13 experiment) into HT29 cells in 3 biological infection replicates. After 24 h, we replaced the media with puromycin media (Supplementary Data 6). We titered viral volumes such that 20–40% of cells survived after puromycin selection. After selection and expansion for 3 days in puromycin, we plated cells for DNA/RNA extraction for qPCR.

H3K27ac HiChIP after CCAT1 knockdown

TetO-*RfxCas13d*-NLS HT29 cells were transduced with either a nontargeting gRNA or one of two *CCAT1* targeting gRNAs (Supplementary Data 7) cloned into pLentiRNA-Guide_001 (Addgene 138150). The cells were selected in 1.5 µg/ml puromycin for 2 weeks to ensure complete selection. Then, we added 1 µg/ml doxycycline to induce *RfxCas13d* expression and performed H3K27ac HiChIP 72 h after induction as described in *H3K27ac HiChIP sample preparation and analysis*. We mapped HiChIP paired end reads to hg19 using HiC-Pro¹²¹. Default settings were using to remove duplicate reads, identify valid interactions, and generate contact maps.

We then randomly downsampled valid interactions for the nontargeting gRNA and Cas13 *CCAT1* knockdown gRNA 2 to match the number of valid interactions from the Cas13 *CCAT1* knockdown gRNA 1 (-16 million valid interactions). We used FitHiChIP¹²² to determine loop significance. We merged loops from *CCAT1* knockdown guide RNA 1 and guide RNA 2 (Supplementary Data 7) and took the average of the contact count and significance to create a combined *CCAT1* knockdown profile. We then identified statistically significant loops in each condition (nontargeting or *CCAT1* knockdown) using the same threshold as in the prior HiChIP (1% FDR), keeping all loops that were significant in either condition. Loop strengths were normalized so that nontargeting and *CCAT1* knockdown conditions had the same mean across all loops. Bootstrap error in contact count was calculated by re-sampling 90% of loops (without replacement) and computing the standard deviation from 1000 bootstrap resamples. Significance was calculated from the z-score of the distribution of loop strength differences (*CCAT1* knockdown–nontargeting control) assuming a normal distribution. For the volcano plot visualization, the log-fold change was computed comparing the *CCAT1* knockdown loop strength to the nontargeting control loop strength with a pseudocount added to both quantities before the fold-change calculation. We chose a pseudocount equivalent to the minimum number of contacts (8).

CCAT1 rescue after CRISPRi knockdown

For CCAT1 rescue experiments, HT29 cells were transduced with KRAB-dCas9 vector (Addgene 170067) and selected with 5 µg/ml of blasticidin for 5 days; we termed these cells HT29-CRISPRi. To make a constitutively expressed CCAT1 rescue vector that co-expresses mKate2, we first amplified CCAT1 exon 1 and exon 2 from HT29 genomic DNA (see Supplementary Data 7 for PCR primers). Using Gibson cloning, we inserted the CCAT1 exons into a backbone containing a CMV promoter and bGH poly(A) signal (pCMV-PEmax, Addgene 174820) that was digested with *NotI* and *PmeI* restriction enzymes (Thermo). Next, we used this intermediate vector as a template for PCR cloning into a lentiviral vector: The CMV-CCAT1 cassette was amplified (see Supplementary Data 7 for PCR primers) and inserted via Gibson cloning in reverse orientation into pLentiEGFPdestabilized⁴⁵ (Addgene 138152) digested with *EcoRI* and *KpnI* restriction enzymes (Thermo). We digested this vector with *Clal* and *BamHI* restriction enzymes and cloned mKate2 into the vector using Gibson assembly. The mKate2¹²³ cassette was synthesized as a human codon-optimized gBlock (IDT). We termed this CCAT1 rescue plasmid pLentiCCAT1-Red (Addgene 226521).

To produce lentivirus, we seeded 9 million HEK293FT cells into a T75 flask and transfected them with 45.6 µl polyethylenimine linear MW 25000 (Polysciences), 8.3 µg pLentiCCAT1-Red, 6.64 µg psPAX2 and 4.56 µg pMD2.G. Three days post-transfection, the viral supernatant was collected, filtered through a 0.45-µm filter. We then transduced HT29-KRAB-dCas9 cells with pLentiCCAT1-Red (MOI < 0.1). After 5 days, we sorted transduced cells (mKate2-positive cells) using a Sony SH800 cell sorter. We termed these cells HT29-CRISPRi-CCAT1rescue.

To generate CRISPRi guide plasmids with mStayGold (a green fluorescent protein), we first cloned the U6-sgRNA(F + E) cassette from pCC_01 (Addgene 139086) into pLentiRNA-Guide_003⁹³ (Addgene 192505) using *PacI* and *NheI* restriction sites and termed the plasmid pLentiSpCas9guide (GFP-P2A-Puro). We then synthesized a human codon-optimized mStayGold¹²⁴ as a gBlock (IDT) and cloned it into pLentiSpCas9guide using *AgeI* and *BamHI* restriction sites and T4 ligase (New England Biolabs); we termed this plasmid lentiGuideFE-mSG-Puro (Addgene, 226522). We then cloned guide RNAs (CRISPRi nontargeting guide and Peak validation sgRNA CRISPRi HT29/PC9 CRE #1, see Supplementary Data 7) into this plasmid using *BsmBI* restriction sites. Lentivirus was produced as in *MYC knockdown using Cas13*.

HT29-CRISPRi and HT29-CRISPRi-CCAT1rescue were each transduced with the guide RNA constructs in lentiGuideFE-mSG-Puro (MOI - 0.5). After 2 days, all cells were plated onto a 96-well plate at 4000 cells per well. Plates were imaged using an Incucyte SX3 at 10× magnification with 4 images acquired per well every 4 h. For each condition, the relative growth rate was calculated as the mStayGold (green) area divided by the total cell confluence area and then normalized to the cells transduced with nontargeting gRNAs.

Enhancer RNA analyses

We downloaded NET-CAGE data from Hirabayashi, et al.⁸⁶. We compared NET-CAGE signal within the identified CREs by summing the total NET-CAGE reads. We ranked 20 kb (HiChIP) bins and DHS peaks within the screen region by the sum of NET-CAGE reads.

Transcription factor motif prediction

We downloaded predicted transcription factor motif binding sites from JASPAR (Supplementary Data 1). After choosing sites with a score (weight) greater than 400, we separated each transcription factor's binding sites into a separate bed file. Then, we used genomic association tester (GAT) 1.3.4 to associate significant gRNAs in each cell line (as defined in *CRISPRi pooled screens library preparation and data analysis*) with the transcription factor motifs in each transcription factor's file. We defined the workspace for GAT as the entirety of the MYC TAD / screen area. GAT was run 5x with 2000 iterations per run,

for a total of 10,000 iterations per association test, for a total of 10,000 iterations per association test. GO term enrichment was performed with Metascape v3.5 against a background of all *Homo sapiens* transcription factors in JASPAR.

TF ChIP-seq and TCGA analysis

We downloaded TF ChIP-seq data from ENCODE (Supplementary Data 1) as processed bigWigs. We accessed TCGA pan-cancer RNA-seq data via the cBioPortal dashboard. We generated MYC-TF correlation plots in RStudio, which also gave us correlation values and p values. TFs chosen are among the top 50 TFs ranked by fold enrichment from GAT (See *Transcription factor motif prediction*) and represent TF families highlighted by GAT in each cell line.

Statistical analysis

Data between two groups were compared using a two-tailed unpaired Student's *t* test. Unless otherwise indicated, a *P*-value less than or equal to 0.05 was considered statistically significant for all analyses, and not corrected for multiple comparisons. In cases where multiple comparison corrections were necessary, we adjusted the *P* value using the Benjamini-Hochberg method. All group results are represented as mean \pm s.e.m, if not stated otherwise. Statistical analyses were performed in Prism (GraphPad) and RStudio (RStudio PBC).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

H3K27ac HiChIP data generated in this study is available have been deposited in NCBI's Gene Expression Omnibus repository under GEO Series accession number [GSE278245](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE278245). Source data are provided with this paper.

References

- Ohno, S. An argument for the genetic simplicity of man and other mammals. *J. Hum. Evol.* **1**, 651–662 (1972).
- Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).
- Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- Rands, C. M., Meader, S., Ponting, C. P. & Lunter, G. 8.2% of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genet.* **10**, e1004525 (2014).
- Graur, D. An upper limit on the functional fraction of the human genome. *Genome Biol. Evol.* **9**, 1880–1885 (2017).
- Portales-Casamar, E. et al. PAZAR: a framework for collection and dissemination of cis-regulatory sequence annotation. *Genome Biol.* **8**, R207 (2007).
- Abascal, F. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
- Gorkin, D. U. et al. An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583**, 744–751 (2020).
- Chang, K. et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
- Gopi, L. K. & Kidder, B. L. Integrative pan cancer analysis reveals epigenomic variation in cancer type and cell specific chromatin domains. *Nat. Commun.* **12**, 1419 (2021).
- Wang, Z. et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).
- Heintzman, N. D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
- Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
- Yan, J. et al. Systematic analysis of binding of transcription factors to noncoding variants. *Nature* **591**, 147–151 (2021).
- Zhang, T., Zhang, Z., Dong, Q., Xiong, J. & Zhu, B. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biol.* **21**, 45 (2020).
- Rickels, R. et al. Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat. Genet.* **49**, 1647–1653 (2017).
- Sur, I. & Taipale, J. The role of enhancers in cancer. *Nat. Rev. Cancer* **16**, 483 (2016).
- Kvon, E. Z., Waymack, R., Gad, M. & Wunderlich, Z. Enhancer redundancy in development and disease. *Nat. Rev. Genet.* **22**, 324–336 (2021).
- Cannavò, E. et al. Shadow enhancers are pervasive features of developmental regulatory networks. *Curr. Biol.* **26**, 38–51 (2016).
- Kim, D. S. et al. The dynamic, combinatorial cis-regulatory lexicon of epidermal differentiation. *Nat. Genet.* **53**, 1564–1576 (2021).
- Bergman, D. T. et al. Compatibility rules of human enhancer and promoter sequences. *Nature* **607**, 176–184 (2022).
- Patwardhan, R. P. et al. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012).
- Melnikov, A. et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).
- 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).
- Sanjana, N. E. et al. High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545–1549 (2016).
- Fulco, C. P. et al. Systematic mapping of functional enhancer–promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
- Canver, M. C. et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192–197 (2015).
- Diao, Y. et al. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res.* **26**, 397–405 (2016).
- Rajagopal, N. et al. High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* **34**, 167–174 (2016).
- Chen, P. B. et al. Systematic discovery and functional dissection of enhancers needed for cancer cell fitness and proliferation. *Cell Reports* **41**, 111630 (2022).
- Frangoul, H. et al. CRISPR-Cas9 gene editing for sickle cell disease and β -Thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2020).
- Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
- Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
- Bahr, C. et al. A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* **553**, 515–520 (2018).
- Herranz, D. et al. A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.* **20**, 1130–1137 (2014).
- Dave, K. et al. Mice deficient of Myc super-enhancer region reveal differential control mechanism between normal and pathological growth. *eLife* **6**, e23382 (2017).

39. Zhang, X. et al. Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.* **48**, 176–182 (2016).
40. Uslu, V. V. et al. Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat. Genet.* **46**, 753–758 (2014).
41. Sur, I. K. et al. Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338**, 1360–1363 (2012).
42. Shi, J. et al. Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes Dev.* **27**, 2648–2662 (2013).
43. Lin, C. Y. et al. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* **530**, 57–62 (2016).
44. Lancho, O. & Herranz, D. The MYC enhancer-ome: Long-range transcriptional regulation of MYC in cancer. *Trends Cancer* **4**, 810–822 (2018).
45. Wessels, H.-H. et al. Massively parallel Cas13 screens reveal principles for guide RNA design. *Nat. Biotechnol.* **38**, 722–727 (2020).
46. Guo, X. et al. Transcriptome-wide Cas13 guide RNA design for model organisms and viral RNA pathogens. *Cell Genomics* **1**, 100001 (2021).
47. Tate, J. G. et al. COSMIC: The catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
48. Consortium, Gt. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
49. Lettice, L. A. et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
50. Eilers, M. & Eisenman, R. N. Myc's broad reach. *Genes Dev.* **22**, 2755–2766 (2008).
51. Perna, D. et al. Genome-wide mapping of Myc binding and gene regulation in serum-stimulated fibroblasts. *Oncogene* **31**, 1695–1709 (2012).
52. Laurenti, E., Wilson, A. & Trumpp, A. Myc's other life: Stem cells and beyond. *Curr. Opin. Cell Biol.* **21**, 844–854 (2009).
53. Zanet, J. et al. Endogenous Myc controls mammalian epidermal cell size, hyperproliferation, endoreplication and stem cell amplification. *J. Cell Sci.* **118**, 1693–1704 (2005).
54. Wang, R. et al. The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity* **35**, 871–882 (2011).
55. Gabay, M., Li, Y. & Felsher, D. W. MYC activation is a hallmark of cancer initiation and maintenance. *Cold Spring Harb. Perspect. Med.* **4**, a014241 (2014).
56. Shou, Y. et al. Diverse karyotypic abnormalities of the c-myc locus associated with c-myc dysregulation and tumor progression in multiple myeloma. *Proc. Natl Acad. Sci. USA* **97**, 228–233 (2000).
57. He, T.-C. et al. Identification of c-MYC as a target of the APC pathway. *Science* **281**, 1509–1512 (1998).
58. Palomero, T. et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc. Natl Acad. Sci. USA* **103**, 18261–18266 (2006).
59. Sharma, V. M. et al. Notch1 contributes to mouse T-cell leukemia by directly inducing the expression of c-myc. *Mol. Cell. Biol.* **26**, 8022–8031 (2006).
60. Weng, A. P. et al. c-Myc is an important direct target of Notch1 in T-cell acute lymphoblastic leukemia/lymphoma. *Genes Dev.* **20**, 2096–2109 (2006).
61. Beroukhi, R. et al. The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
62. Vita, M. & Henriksson, M. The Myc oncoprotein as a therapeutic target for human cancer. *Semin. Cancer Biol.* **16**, 318–330 (2006).
63. Schaub, F. X. et al. Pan-cancer alterations of the MYC oncogene and its proximal network across the cancer genome atlas. *Cell Syst.* **6**, 282–300 (2018).
64. Lombart, V. & Mansour, M. R. Therapeutic targeting of 'undrugable' MYC. *EBioMedicine* **75**, 103756 (2022).
65. Arman, K. & Möröy, T. Crosstalk between MYC and lncRNAs in hematological malignancies. *Front. Oncol.* **10**, 2135 (2020).
66. Iaccarino, I. lncRNAs and MYC: An intricate relationship. *Int. J. Mol. Sci.* **18**, 1497 (2017).
67. Tseng, Y.-Y. & Bagchi, A. The PVT1-MYC duet in cancer. *Mol. Cell. Oncol.* **2**, e974467 (2015).
68. Aguirre, A. J. et al. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* **6**, 914–929 (2016).
69. Munoz, D. M. et al. CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov.* **6**, 900–913 (2016).
70. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
71. Arango, D. et al. c-Myc overexpression sensitises colon cancer cells to camptothecin-induced apoptosis. *Br. J. Cancer* **89**, 1757–1765 (2003).
72. Zhu, L. et al. Targeting c-Myc to overcome acquired resistance of EGFR mutant NSCLC cells to the third-generation EGFR tyrosine kinase inhibitor, osimertinib. *Cancer Res.* **81**, 4822–4834 (2021).
73. Zhang, Y. et al. BRD4 modulates vulnerability of triple-negative breast cancer to targeting of integrin-dependent signaling pathways. *Cell. Oncol.* **43**, 1049–1066 (2020).
74. Delgado, M. D. & León, J. Myc roles in hematopoiesis and leukemia. *Genes Cancer* **1**, 605–616 (2010).
75. Hart, T. et al. Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3: Genes, Genomes, Genetics* **7**, 2719–2727 (2017).
76. Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
77. Yeager, M. et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39**, 645–649 (2007).
78. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
79. Ryan, R. J. et al. AB cell regulome links notch to downstream oncogenic pathways in small B cell lymphomas. *Cell Rep.* **21**, 784–797 (2017).
80. Heger, A., Webber, C., Goodson, M., Ponting, C. P. & Lunter, G. GAT: A simulation framework for testing the association of genomic intervals. *Bioinformatics* **29**, 2046–2048 (2013).
81. Gryder, B. E., Khan, J. & Stanton, B. Z. Measurement of differential chromatin interactions with absolute quantification of architecture (AQuA-HiChIP). *Nat. Protoc.* **15**, 1209–1236 (2020).
82. Fulco, C. P. et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
83. Szutorisz, H., Dillon, N. & Tora, L. The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem. Sci.* **30**, 593–599 (2005).
84. Kim, T.-K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
85. De Santa, F. et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
86. Hirabayashi, S. et al. NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements. *Nat. Genet.* **51**, 1369–1379 (2019).

87. Cardiello, J. F., Sanchez, G. J., Allen, M. A. & Dowell, R. D. Lessons from eRNAs: understanding transcriptional regulation through the lens of nascent RNAs. *Transcription* **11**, 3–18 (2020).
88. Azofeifa, J. G., Allen, M. A., Hendrix, J. R., Rubin, J. D. & Dowell, R. D. Enhancer RNA profiling predicts transcription factor activity. *Genome Res.* **28**, 334–344 (2018).
89. Xiang, J.-F. et al. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* **24**, 513–531 (2014).
90. Orlando, G. et al. Promoter capture Hi-C-based identification of recurrent noncoding mutations in colorectal cancer. *Nat. Genet.* **50**, 1375–1380 (2018).
91. Cai, Z. et al. RIC-seq for global in situ profiling of RNA–RNA spatial interactions. *Nature* **582**, 432–437 (2020).
92. Konermann, S. et al. Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell* **173**, 665–676 (2018).
93. Wessels, H.-H. et al. Prediction of on-target and off-target activity of CRISPR–Cas13d guide RNAs using deep learning. *Nat. Biotechnol.* **1–10** <https://doi.org/10.1038/s41587-023-01830-8> (2023).
94. Jiang, G. et al. Annexin A13 promotes tumor cell invasion in vitro and is associated with metastasis in human colorectal cancer. *Oncotarget* **8**, 21663 (2017).
95. Zhang, Y. et al. Elevated NIBP/TRAPPC9 mediates tumorigenesis of cancer cells through NF κ B signaling. *Oncotarget* **6**, 6160 (2015).
96. Qin, M. et al. NIK-and IKK β -binding protein promotes colon cancer metastasis by activating the classical NF- κ B pathway and MMPs. *Tumor Biol.* **37**, 5979–5990 (2016).
97. Li, Y. et al. The emerging role of ISWI chromatin remodeling complexes in cancer. *J. Exp. Clin. Cancer Res.* **40**, 1–27 (2021).
98. Wang, M. et al. CHRAC1 promotes human lung cancer growth through regulating YAP transcriptional activity. *Carcinogenesis* **43**, 264–276 (2022).
99. Oksuz, O. et al. Transcription factors interact with RNA to regulate genes. *Mol. Cell* **83**, 2449–2463 (2023).
100. Mansour, M. R. et al. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014).
101. Durbin, A. D. et al. Selective gene dependencies in MYCN-amplified neuroblastoma include the core transcriptional regulatory circuitry. *Nat. Genet.* **50**, 1240–1246 (2018).
102. Sanda, T. et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* **22**, 209–221 (2012).
103. Reddy, J. et al. Predicting master transcription factors from pan-cancer expression data. *Sci. Adv.* **7**, eabf6123 (2022).
104. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
105. Replogle, J. M. et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575.e28 (2022).
106. Elbashir, S. M. et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
107. Özcan, A. et al. Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. *Nature* **597**, 720–725 (2021).
108. Meyer, N. & Penn, L. Z. Reflecting on 25 years with MYC. *Nat. Rev. Cancer* **8**, 976–990 (2008).
109. Florea, V. et al. c-Myc is essential to prevent endothelial pro-inflammatory senescent phenotype. *PLoS One* **8**, e73146 (2013).
110. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
111. Akalin, A., Franke, V., Vlahoviček, K., Mason, C. E. & Schübeler, D. genomation: a toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics* **31**, 1127–1129 (2015).
112. Kolde, R., Laur, S., Adler, P. & Vilo, J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28**, 573–580 (2012).
113. Cho, S. W. et al. Promoter of lncRNA gene PVT1 is a tumor-suppressor DNA boundary element. *Cell* **173**, 1398–1412.e22 (2018).
114. Dao, L. T. M. & Spicuglia, S. Transcriptional regulation by promoters with enhancer function. *Transcription* **9**, 307–314 (2018).
115. Hart, S. et al. Precise RNA targeting with CRISPR–Cas13d. *Nat. Biotechnol.* **1–11** <https://doi.org/10.1038/s41587-025-02558-3> (2025).
116. Doench, J. G. et al. Rational design of highly active sgRNAs for CRISPR–Cas9 – mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
117. Canver, M. C. et al. Integrated design, execution, and analysis of arrayed and pooled CRISPR genome-editing experiments. *Nat. Protoc.* **13**, 946–986 (2018).
118. Legut, M. et al. High-throughput screens of PAM-Flexible Cas9 variants for gene knockout and transcriptional modulation. *Cell Rep.* **30**, 2859–2868 (2020).
119. Joung, J. et al. Genome-scale CRISPR–Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
120. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
121. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
122. Bhattacharyya, S., Chandra, V., Vijayanand, P. & Ay, F. Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat. Commun.* **10**, 4221 (2019).
123. Shcherbo, D. et al. Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J.* **418**, 567–574 (2009).
124. Ando, R. et al. StayGold variants for molecular fusion and membrane-targeting applications. *Nat. Methods* **21**, 648–656 (2024).

Acknowledgements

We thank the entire Sanjana laboratory for support and advice. We also thank E. Khurana and S.-s. C. Huang for advice and feedback on enhancer analyses and the New York University (NYU) Biology Genomics Core for sequencing resources. N.E.S. is supported by the NIH/National Human Genome Research Institute (DP2HG010099, R01HG012790), the NIH/National Cancer Institute (NCI) (R01CA279135 and R01CA218668), the NIH/National Institute of Allergy and Infectious Diseases (NIAID) (R01AI176601), the NIH/National Heart Lung and Blood Institute (R01HL168247), the National Science Foundation (EAGER 2409037), the Melanoma Research Alliance and the Sidney Kimmel Foundation, the Simons Foundation for Autism Research, the MacMillan Center for the Study of the Noncoding Cancer Genome, New York University, and New York Genome Center funds.

Author contributions

C.M.C., V.T.L., M.M., B.J.D., and N.E.S. designed the study. V.T.L., M.M., and B.J.D. performed the CRISPRi screens. C.M.C. and V.T.L. performed subsequent experiments and analysis. C.M.C. performed and analyzed HiChIP experiments. E.G. assisted with qPCR experiments. N.L.B. assisted with transcription factor analysis. J.A.M. performed GTEx and TCGA analyses. A.M.M., S.M. and C.M.C. produced monoclonal cell lines. C.M.C., V.T.L., and N.E.S. wrote the manuscript. N.E.S. supervised the study.

Competing interests

N.E.S. is an adviser to Qiagen and a co-founder and adviser of TruEdit Bio and OverT Bio. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-56568-5>.

Correspondence and requests for materials should be addressed to Neville E. Sanjana.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025