3D genomic analysis reveals novel enhancer-hijacking caused by complex structural alterations that drive oncogene overexpression

Katelyn L. Mortenson[1], Courtney Dawes[2], Emily R. Wilson[1], Nathan E. Patchen[2], Hailey E. Johnson[1,3], Jason Gertz[1], Swneke D. Bailey[4,5], Yang Liu[2], Katherine E. Varley[1,*], Xiaoyang Zhang[1,*]

1. Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah, USA
2. Department of Biochemistry, University of Utah, Salt Lake City, Utah, USA
3. Department of Cell Biology and Physiology, Brigham Young University, Provo, Utah, USA
4. Cancer Research Program, Research Institute of the McGill University Health Centre, Montreal, QC, Canada.
5. Departments of Surgery and Human Genetics, McGill University, Montreal, QC, Canada.

*. Corresponding authors:
Katherine E. Varley (Katherine.Varley@hci.utah.edu) and Xiaoyang Zhang (Xiaoyang.Zhang@hci.utah.edu).

ABSTRACT

Cancer genomes are composed of many complex structural alterations on chromosomes and extrachromosomal DNA (ecDNA), making it difficult to identify non-coding enhancer regions that are hijacked to activate oncogene expression. Here, we describe a 3D genomics-based analysis called HAPI (Highly Active Promoter Interactions) to characterize enhancer hijacking. HAPI analysis of HiChIP data from 34 cancer cell lines identified enhancer hijacking events that activate both known and potentially novel oncogenes such as *MYC, CCND1*, *ETV1*, *CRKL*, and *ID4*. Furthermore, we found enhancer hijacking among multiple oncogenes from different chromosomes, often including *MYC*, on the same complex amplicons such as ecDNA. We characterized a *MYC-ERBB2* chimeric ecDNA, in which *ERBB2* heavily hijacks *MYC*'s enhancers. Notably, CRISPRi of the *MYC* promoter led to increased interaction of *ERBB2* with *MYC* enhancers and elevated *ERBB2* expression. Our HAPI analysis tool provides a robust strategy to detect enhancer hijacking and reveals novel insights into oncogene activation.

INTRODUCTION

Somatic structural variants, including deletions, duplications, inversions, and translocations, dramatically alter genomic structures in cancer cells. These events can be readily detected by DNA imaging, whole-genome/long-read sequencing, or Hi-C-based 3D genomic assays[1-5]. Understanding the functional consequences of structural variants is key to uncovering the mechanisms of tumorigenesis and developing therapeutic strategies. For instance, structural variants can create a fusion gene with oncogenic properties such as *BCR-ABL1* and *EML4-ALK*[6,7]. However, most of the breakpoints of cancer-associated structural variants fall in noncoding regions[8]. These variants can place enhancer elements adjacent to oncogenes such as *MYC, MYCN*, *MYB*, *CCND1, EVI1,* and *GFI1,* and activate their expression, events known as enhancer hijacking[9-17].

Several analysis approaches have been developed to identify enhancer hijacking. PANGEA and CESAM identify candidate genes based on the correlation of their expression with adjacent structural alterations across tumor samples, which require large cohorts of data and do not utilize chromatin interaction evidence of enhancer hijacking[18,19]. Neoloop reconstructs Hi-C maps in structurally altered cancer genomes to identify aberrant chromatin loops and their associated genes[20], however, it does not prioritize the gene targets and the underlying functional enhancers. As structural variants often affect many enhancer-promoter interactions, most of which are likely passenger events[2,3], it is essential to prioritize the functional events. Furthermore, recent work has revealed that complex structural variants such as chromothripsis lead to the formation of circular extrachromosomal DNA (ecDNA)[21]. These events often cause extensive alterations of enhancer-promoter interactions, posing another challenge to characterize the underlying mechanisms of oncogene regulation[22-24].

Here, we propose a 3D genomics-based strategy to identify and characterize enhancer hijacking events based on two assumptions: 1) oncogenes subject to enhancer hijacking should be highly regulated by enhancers; 2) the hijacked enhancers should contribute a substantial proportion of an oncogene's enhancer usage. To test this, we utilized H3K27ac HiChIP data to map enhancer-promoter interactions and measured their interaction intensities. We applied our strategy to HiChIP data from 34 cancer and two normal cell lines. This approach identified known and novel enhancer hijacking events, prioritized their target genes, and defined the underlying enhancers. We also revealed novel enhancer hijacking events associated with ecDNAs or complex chromosomal amplicons. In these events, multiple oncogenes from different chromosomal regions are translocated near each other, amplified together, and hijack each other's enhancers to boost their overexpression.

RESULTS

HAPI analysis identifies genes that are highly interactive with enhancers.

We first developed an analysis method, Highly Active Promoter Interactions (HAPI), that utilizes H3K27ac HiChIP data to quantify all enhancer interactions of each gene's promoter in the genome, including intra- and inter-chromosomal interactions. This method ranks genes by 1) their enhancer contact value defined as the number of interacting enhancers and 2) their interaction intensity as defined by the number of HiChIP paired-end tags (PETs). We define highly interactive (HAPI) genes as those that exceed the inflection points of both the ranked enhancer contact value and ranked interaction intensity in a manner similar to previously published methods that define super-enhancers[25] (Figure 1A), as exemplified in the prostate cancer cell line LNCaP (Figure 1B and S1A-B).

We applied HAPI analysis to H3K27ac HiChIP data that we and others generated in 34 cancer cell lines representing multiple cancer types (Supplementary Table 1), which identified on average 538 HAPI genes in each cell line (Supplementary Data 1). In LNCaP cells, HAPI genes are more likely to interact with super-enhancers compared to non-HAPI enhancer-connected genes as expected[26,27] (Figure S1C), but there are additional highly interactive HAPI genes that do not utilize super-enhancers (Figure S1C). Conversely, super-enhancers are not necessarily looped to HAPI genes in LNCaP cells, as most of the super enhancers interact with both HAPI and non-HAPI genes, or only non-HAPI genes (Figure S1D). Overexpressed oncogenes such as *MYC*, *FOXA1*, *TBX3*, and *PCAT1* were identified as HAPI genes in LNCaP cells. Notably, HAPI genes are not necessarily genes with the highest expression, as highly expressed housekeeping genes *ACTB*, *HPRT1*, and *GAPDH* fell below the HAPI

cutoff (Figure 1B and S1E). Furthermore, when we investigated whether copy number amplification had an impact HAPI gene metrics, we found that only a small proportion of the amplified genes are identified as HAPI genes in LNCaP cells (Figure S1F). We did not normalize copy number variants in this analysis, as we and others have shown that focal copy number amplifications contribute to enhancer-promoter interactions for oncogenes – a signal that we want to preserve in our analysis[28-33].

We performed unsupervised clustering of the cell lines based on a combination of the enhancer contact and interaction intensity values for each identified HAPI gene (Figure 1C, see Methods for more details). The results showed that cell lines from the same lineages tend to cluster together, highlighting the cancer type-specificity of HAPI genes. We also found significant enrichments of previously annotated oncogenes[34-36], and genes involved in cancer-related pathways in our identified HAPI gene lists (Figure 1D and Supplementary Table 2). Overall, HAPI analysis identified known cancer-related genes that are highly interactive with enhancers in a cancer-type specific manner.

HAPI analysis identifies known and novel enhancer-hijacking events.

HAPI analysis not only defines genes that are highly interactive with enhancers but also determines the location of their enhancers. This enables us to determine whether the enhancers are in the same genomic region as the HAPI gene or are hijacked from other genomic regions likely through structural alterations (Figure 2A). These include both 'trans' hijacking events that relocate enhancers from other chromosomes to a HAPI gene and 'cis' hijacking events that relocate distant enhancers (>2Mb, the upper limit of typical TAD domains[37]) on the same chromosome.

We analyzed H3K27ac HiChIP data of the 34 cancer cell lines to determine the origin of enhancers for HAPI genes. For HAPI genes with over 25% of enhancer activity coming from either trans- or abnormal cis- interactions were treated as candidate genes driven by enhancer hijacking. In total, we identified enhancer hijacking genes in 29 of the 34 cell lines (Figure 2B). In this study, we focus on trans-enhancer hijacking events because they are more likely caused by structural alterations than cis-enhancer hijacking events that may be caused by other types alterations such as epigenetic loss of TAD boundaries[38,39]. We found trans-enhancer hijacking events in 16 of the cancer cell lines, but in neither of the two normal cell lines we examined (Figure 2C). The calculated enhancer origins, both trans and cis, for all HAPI genes are listed in Supplementary Data 1.

In addition to enhancer origins, we also examined copy numbers of the HAPI genes using published SNP-array data[40]. We first focused on HAPI genes that exhibit neutral copies or modest amplifications ($\log_2(CN)<2$) (Figure 2D-E). This analysis identified previously reported trans-enhancer hijacking events activating oncogenes *CCND1* in the B-cell lymphoma cell line REC1[10,12] and *ETV1* in the prostate cancer cell lines LNCaP and MDAPCA2B[20,41,42]. We also found novel trans-enhancer hijacking events activating known oncogenes including *MYC* in the small cell lung cancer cell line NCIH146, *NOTCH1* in the T-cell lymphoblastic leukemia cell line CUTLL1, and *CCND1* in the breast cancer cell line ZR751. Furthermore, HAPI analysis identified potentially novel oncogenes that hijack enhancers in trans, such as *CRKL* in the colorectal cancer cell line HT29 and *FOXJ2* in the small cell lung cancer cell line NCIH2171 (Figure 2D and S2A). Finally, we found *cis*-enhancer hijacking events activating potentially novel oncogenes such as *ID4* in the prostate cancer cell line VCaP and *TXLNA* in the small cell lung cancer cell line CORL88 (Figure 2E), which are likely caused by large deletions based on the known patterns of structural variants in the HiChIP signal (Figure S2B-D). Overall, cell lines with these enhancer-hijacking events are associated with higher expression of the involved genes as compared to other cell lines of the same lineage (Figure 2D-E and S2A).

Functional validation of identified enhancer hijacking events.

We then chose two representative enhancer hijacking genes identified from HAPI analysis for functional validation. We identified *ETV1* (chromosome 7) as a trans-enhancer hijacking gene in both LNCaP and MDAPCA2B cells, with most of its trans enhancer activity coming from chromosome 14 (Figure 3A). This is caused by a cryptic insertion in LNCaP and a balanced translocation in MDAPCA2B[43] that connect the two chromosomal regions. A translocation linking these two regions is also identified in a primary prostate cancer

patient tumor sample from Pan-Cancer Analysis of Whole Genomes (PCAWG) (Figure S3A). Previous studies have identified various potential enhancers within regions of this translocation, but a consensus regarding which one regulates *ETV1* has not yet been reached[20,42]. Based on H3K27ac HiChIP and ChIP-seq signal, we observed four enhancers e1-e4, upstream of *FOXA1* on chromosome 14, hijacked to the *ETV1* promoter in LNCaP and MDAPCA2B cells (Figure 3A). CRISPRi assays in LNCaP showed that repression of e3 and e4 caused the most significant decreases in expression of *ETV1* (Figure 3B). H3K27ac ChIP-seq performed after CRISPRi confirmed the on-target effects (Figure S3B). As a control, repression of the four enhancers did not affect *ETV1* expression in 22RV1 cells that lack the translocation (Figure S3C). Interestingly, these enhancers have little or modest regulatory effects on the nearby gene *FOXA1* or *TTC6* on their original chromosome in LNCaP cells (Figure 3B and S3D), indicating that these two genes may be dependent on additional enhancers.

ETV1 was reported as an androgen-responsive gene in LNCaP cells promoting cell invasion[44]. We observed binding of androgen receptor (AR) to both e3 and e4 but not the promoter of *ETV1* (Figure 3A and S3E). Additionally, we observed a 40% upregulation of *ETV1* expression in response to Dihydrotestosterone (DHT) treatment in LNCaP cells. In contrast, *ETV1* expression is minimally affected by DHT in the untranslocated 22RV1 cells, where AR binds the e3 and e4 enhancers in their original locus (Figure S3F). These data suggest that the androgen-induced *ETV1* upregulation is caused by these two hijacked AR-bound enhancers.

The second representative enhancer hijacking gene we selected from HAPI analysis for functional validation was *CCND1*. *CCND1* (chromosome 11) is known to hijack enhancers from chromosome 14 in the B-cell lymphoma cell line REC1[10,12]. We found that the same oncogene hijacks different enhancers from chromosome 8 in the breast cancer cell line ZR751 (Figure 3C). WGS analysis revealed a translocation that places these enhancers upstream of *CCND1* in ZR751 (Figure 3C). In primary breast cancer patient tumor samples from PCAWG, although we did not observe the same translocation event, we found that the *CCND1* locus is highly subject to translocations with diverse regions of chromosome 8 (Figure S3G). We applied CRISPRi to repress five of the strongest enhancers (e1-e5 in Figure 3C) individually or simultaneously in ZR751. While targeting individual enhancers yields little (e1 and e2) or modest effect (e3, e4, and e5) on *CCND1* expression, targeting the combination of all enhancers with CRISPRi caused a 45% reduction in *CCND1* expression (Figure 3D). Repressing most of these enhancers caused significant decreases in expression of their endogenous target gene *DUSP4* on chromosome 8 (Figure 3D). These results together demonstrate that these enhancers activate *CCND1,* in a combinatorial manner, due to enhancer hijacking while also acting in their canonical role of activating *DUSP4*. *CCND1* is overexpressed in 50% of breast cancers, which has been partially attributed to gene amplification[45]. It is possible that enhancer hijacking may serve as an additional mechanism for *CCND1* overexpression in breast cancer.

The hijacked enhancers on chromosome 8 in ZR751 exhibit H3K27ac enrichment in the breast cancer cell line T47D without the translocation (Figure S4A), but little enrichment in the B-cell lymphoma line REC1 in which *CCND1* hijacks enhancers from chromosome 14 (Figure 3C). The reciprocal pattern was observed for the enhancers hijacked in REC1 (Figure S4A). These results indicate that the enhancers involved in hijacking are already active in cancer cells without the translocation and are cancer-type specific. To support this finding, we utilized TCGA ATAC-seq data to estimate potential regulatory activity of translocated enhancers we identified in several breast, prostate, and colon cancer cell lines. Overall, these enhancers have increased ATAC-seq signal in tumors of their respective cancer types compared to others (Figure S4B). These data together suggest that cancer-type specific enhancers may be subject to translocations to activate oncogene expression.

HAPI analysis identifies enhancer hijacking associated with known ecDNAs.

Some of the identified enhancer hijacking genes exhibit high copies (log2(CN)>2). Recent studies have revealed that oncogenes amplified in the form of ecDNAs tend to exhibit higher copies compared to their chromosomal amplifications[46]. As illustrated in Figure 4A, the high copy number of ecDNAs, along with increased chromatin accessibility, results in the overexpression of oncogenes such as *EGFR* in glioblastoma[46]. In addition, enhancer hijacking, as a result of complex rearrangements on ecDNAs, also contributes to oncogene expression as exemplified by *MYCN* in neuroblastoma[15]. Among our analyzed cell lines, the colorectal cancer cell line

COLO320DM and the gastric adenocarcinoma cell line SNU16 are known to contain ecDNAs amplifying the *MYC* oncogene[47]. HAPI analysis identified enhancer hijacking events in both the cell lines, which involve multiple oncogenes on the same ecDNA.

COLO320DM contains an ecDNA harboring *MYC,* a *PVT1-MYC* fusion, and several other genes including *CDX2* and *PDX1*[47]. As a HAPI gene, *MYC* has most of its enhancer activity contributed in *cis*. Interestingly, *CDX2,* a colorectal cancer lineage-survival oncogene[48,49] from chromosome 13, is identified as a trans-enhancer hijacking HAPI gene (Figure 4B), with over 95% of enhancer activity in trans. Indeed, most of the enhancer activity for *CDX2* comes from the *MYC* locus on the ecDNA (Figures 4C-E). These results suggest two separate mechanisms through which the ecDNA causes overexpression of oncogenes in COLO320DM cells – while *MYC* uses its co-amplified endogenous enhancers, *CDX2* hijacks *MYC* enhancers on the same ecDNA to enhance its expression, representing an "opportunistic" enhancer hijacking event (Figure 4D).

SNU16 contains multiple ecDNA isoforms harboring the *MYC* and *FGFR2* oncogenes separately or together[47,50]. The *MYC* ecDNA harbors three other genes from chromosome 11, including *APIP*, *PDHX*, and *CD44,* which are involved in the development of gastric adenocarcinoma[51,52]. HAPI analysis identified these three genes as enhancer hijacking genes (Figure 4F), which have 34%-42% of enhancer activity contributed from enhancers near *MYC* on chromosome 8 that are now on the same ecDNA (Figures 4G-I). Vice versa, *MYC* also has most of the enhancer activity contributed in trans from other chromosomal regions on the same ecDNA (Figure 4G-I). These reciprocal enhancer-oncogene interactions represent a "mutualistic" enhancer hijacking event on ecDNAs (Figure 4H). Interestingly, although most of the *FGFR2* ecDNA does not include regions from other chromosomes based on previous reports[47], it has around half of enhancer activity contributed in trans (Figure 4F-G). This is likely caused by the reported intermolecular interactions between *MYC* and *FGFR2* ecDNAs as well as the presence of chimeric *MYC-FGFR2* ecDNA in a small proportion of the ecDNAs in this cell line[47].

Enhancer hijacking among multiple oncogenes is prevalent on ecDNAs and complex chromosomal amplicons.

We reasoned that enhancer hijacking, together with high copy numbers of the involved genes, may indicate a unique signature for chimeric ecDNAs or complex chromosomal amplicons that contain multiple DNA segments in our HAPI analysis results. Indeed, we found three additional cell lines NCIH2170, NCIH446, and MCF7 that have clusters of highly amplified HAPI genes hijacking enhancers from different chromosomes (Figure 5A-C). We analyzed their WGS data using AmpliconArchitect, an algorithm to reconstruct amplicons[53], which showed extensive rearrangements among the amplified regions harboring the enhancer-hijacking genes (Figure S5A-C).

In the lung squamous cell carcinoma cell line NCIH2170, we identified a cluster of trans enhancer hijacking genes from chromosomes 8 and 17 (Figure 5A). AmpliconArchitect analysis of WGS data predicted an ecDNA containing most of these HAPI genes and their hijacked enhancers (Figure S6A-B). In particular, we found that *ERBB2* (chromosome 17) and its surrounding oncogenes *MIEN1* and *IKZF3* strongly interact with *MYC*'s enhancers (Figure 5D-E). *ERBB2* was known to be an amplified and overexpressed gene in this cell line[54,55]. Our analysis showed that, in addition to the copy number amplification, hijacking *MYC* enhancers on the same ecDNA may also contribute to *ERBB2* and its surrounding oncogenes' overexpression.

Additionally, in the small-cell lung cancer cell line NCIH446, HAPI analysis identified several genes including known oncogenes *MYC* and *NFIB* with trans enhancer hijacking from regions of chromosomes 8 and 9 (Figure 5F-G). WGS analysis showed translocations among these gene loci (Figure S6C). A recent study reported that the *MYC* locus in NCIH446 resides on the chromosome based on DNA Fluorescent In Situ Hybridization (FISH) results[56], which together with our results suggest that complex chromosomal amplicons are also capable of causing multiple oncogenes to hijack each other's enhancers.

Lastly, we identified 36 HAPI genes with trans enhancer hijacking in the ERα-positive breast cancer cell line MCF7, most of which are clustered within genomic regions of chromosomes 17 and 20 (Figure S7A). These include known oncogenes *TBX2* and *BRIP1* from chromosome 17, and *ZNF217* from chromosome 20.

Accordingly, ampliconArchitect analysis predicted an ecDNA harboring most of these HAPI genes and their hijacked enhancers (Figures S7B-C). These oncogenes are highly interactive with each other's enhancers, indicating extensive enhancer hijacking spanning the ecDNA segments (Figures S7D).

In the aforementioned five cell lines that harbor ecDNAs or chromosomal complex amplicons, we observed both "opportunistic" (COLO320DM and NCIH2170) and "mutualistic" (MCF7, SNU16 and NCIH446) enhancer hijacking phenomena among multiple oncogenes on the same amplicon, which are associated with overexpression of the associated oncogenes (Figure S8A). Using publicly available AmpliconArchitect results from WGS data of primary patient tumors included in the PCAWG project[57], we identified 1299 amplicons containing previously annotated oncogenes. We found that most of these amplicons, either in the extrachromosomal or chromosomal forms, harbor additional DNA segments originating from other chromosomal regions. In 37.6% of these complex amplicons, we found multiple oncogenes from separate DNA segments that are now on the same amplicon (Figure 5H). One amplicon is predicted to be a *MYC-ERBB2* chimeric ecDNA, similar to what we report in NCIH2170 cells (Figure S8B). These analyses suggest that many of these amplicons could enable extensive enhancer hijacking among their harbored oncogenes in patient tumors.

Functional exploration of chimeric ecDNA-associated enhancer hijacking.

To validate the presence of chimeric ecDNAs that contain multiple oncogenes from different chromosomes, we performed DNA FISH in NCIH2170 cells using two probes targeting the *MYC* and *ERBB2* regions separately. DNA FISH on metaphase spreads showed the presence of ecDNAs containing both *MYC* and *ERBB2* across examined cells (Figure 6A). However, some of the ecDNAs contain only *MYC* or *ERBB2*, suggesting the intracellular heterogeneity of ecDNA compositions similar to what was reported previously in SNU16[47]. DNA FISH in interphase nuclei showed similar overlap between *MYC* and *ERBB2* signal across the examined cells (Figure S9A). We also observed the presence of ecDNA aggregates in some of the interphase nuclei, indicative of ecDNA hubs as previously reported in other cancer cell lines[47] (Figure S9B). These results demonstrate the presence of highly abundant *MYC-ERBB2* chimeric ecDNA in NCIH2170 cells.

The presence of chimeric *MYC* ecDNAs may provide novel therapeutic opportunities. For instance, although MYC as a transcription factor is difficult to target by small molecules, ERBB2 is considered highly targetable. Indeed, NCIH2170 exhibits the strongest sensitivity to HER2 inhibitors lapatinib and neratinib among all the lung squamous cancer cell lines tested in the Cancer Therapeutics Response Portal (CTRP) CTD[2] project (Figure 6B).

We then sought to assess the regulatory function of enhancers on the *MYC-ERBB2* ecDNA. Due to the large number of potential *MYC* enhancers that are shared by these genes (Figure 5D), individually repressing these enhancers may not reflect their combinatorial activity. As an alternative strategy, we used CRISPRi to repress the *MYC* promoter and released all *MYC* enhancers from *MYC* to assess their effect on expression of *ERBB2* and its neighboring genes that also use these enhancers (illustrated in Figure 6C). HiChIP assays showed that CRISPRi resulted in an overall decrease in interactions between the *MYC* promoter and its enhancers, as expected (Figure 6D-F). In contrast, *ERBB2* gained interactions with these enhancers after CRISPRi of the *MYC* promoter (Figure 6E-F). RNA-seq showed that CRISPRi represses *MYC* expression and causes a significant increase in expression of *ERBB2* (2.7 folds) and its neighboring genes *MIEN1*, *PGAP3*, and *GRB7* (Figure 6G). For the remaining affected genes, we found enrichments of MYC up- and down-regulated gene sets, suggesting that they are secondary effects of *MYC* repression (Figure S9C). All these results demonstrate that, on the chimeric *MYC-ERBB2* ecDNA in NCIH2170 cells, while *MYC* retains its interaction with its endogenous enhancers, *ERBB2* and its neighboring genes opportunistically hijack *MYC* enhancers on the same ecDNA for their transcriptional activation.

DISCUSSION

Enhancer hijacking caused by structural alterations is an oncogenic event in cancer; however, it remains difficult to detect due to the complexity of cancer genomes. Here, we presented a two-step strategy, namely HAPI, to robustly detect enhancer hijacking events based on: 1) oncogenes subject to enhancer hijacking should be highly

regulated by enhancers, 2) the hijacked enhancers should contribute to a substantial proportion of an oncogene's enhancer activity. This assumption is supported by recent work using CRISPR to generate *de novo* enhancer translocations to the *MYC* locus, which showed that the combination of enhancer activity and contact from the translocated regions determines *MYC* activation[58]. Although our analysis was based on H3K27ac HiChIP data, we expect it to be compatible with other types of high-resolution enhancer-promoter contact mapping assays such as PLAC-seq, Micro-C, and ChIA-PET[59-61].

We first found hundreds of HAPI genes in each cell line whose promoters are highly interactive with enhancers. Previous work focused on neural cells found that genes that specify lineage development exhibit this pattern[62]. Our analysis in cancer cell lines found that oncogenes are also enriched in this category. We then prioritized HAPI genes that have a substantial amount of enhancer activity originating from other chromosomes or abnormally distant intra-chromosomal regions. This led to the identification of known and potentially novel enhancer-hijacking events regardless of the underlying structural variants such as balanced translocation, cryptic insertion, large deletion, as well as ecDNA formation. Although several of the identified oncogenes such as *MYC*, *ETV1*, and *CCND1* have been previously linked to enhancer hijacking, we identified their novel hijacked enhancers that are cancer type specific. The analysis also identified potentially novel enhancer-hijacking oncogenes such as *CRKL* and *ID4*. We have made the HAPI analysis tool publicly available (URL: https://codeocean.com/capsule/2804329) and anticipate its broader application in cancer driver discovery and mechanism investigation.

In addition to individual enhancer-hijacking events caused by simple DNA rearrangements, we identified nested enhancer-hijacking on complex amplicons, such as ecDNAs, that involve multiple oncogenes originating from different chromosomal regions. Our analysis of previously characterized as well as our newly identified ecDNAs shows that ecDNAs may contain multiple oncogenes from different chromosomes, which we refer to as chimeric ecDNAs. These structures were predicted by WGS analysis and validated by either optical imaging from previous work[47,63] or dual-color DNA FISH in this study. We revealed two types of enhancer hijacking events associated with chimeric ecDNAs, as exemplified by *MYC* ecDNAs in this study. For the *MYC* ecDNAs found in COLO320DM and NCIH2170 cells, while *MYC* mainly uses its own enhancers, the additional oncogenes such as *CDX2* and *ERBB2* that are translocated to the *MYC* ecDNAs heavily use *MYC* enhancers for their activation, representing an "opportunistic" enhancer hijacking event. For the *MYC* ecDNA found in SNU16 cells, *MYC* and the co-amplified *CD44* gene use each other's enhancers in a more balanced manner, representing a potentially "mutualistic" enhancer hijacking event. These results provide important insights about another layer of complexity for oncogene amplification and overexpression.

Chimeric ecDNAs may be prevalent in cancer. A recent work reported that 33.1% of ecDNAs found in Barret's oesophagus and esophageal adenocarcinomas contain multiple oncogenes on the same molecule based on WGS analysis[64]. In addition to ecDNAs, chromosomal translocations often precede amplifications, also causing co-amplification two oncogenes from distinct chromosomal regions[65]. Characterization of these chromosomal or extrachromosomal complex amplicons may reveal novel therapeutic strategies. As exemplified in this study, although *MYC* is a notoriously difficult drug target, *ERBB2* that is co-amplified with MYC is highly targetable.

Our observation of potentially "mutualistic" or "opportunistic" enhancer hijacking events on different chimeric ecDNAs has important implications for the future of CRIPSR-based therapeutics being developed to silence oncogenes. Targeting the shared enhancers on chimeric ecDNAs that contribute to overexpression of multiple oncogenes is an appealing therapeutic strategy to repress multiple oncogenes simultaneously. However, we also demonstrate that repressing the *MYC* promoter on a chimeric ecDNA released the shared enhancers from *MYC* and increased *ERBB2's* opportunistic enhancer hijacking and expression. This could be a strategy to further sensitize the tumor to HER2 inhibitors, or it could be disastrous in increasing growth factor driven proliferation. These examples illustrate the importance of identifying and characterizing these complex regulatory events involving multiple oncogenes in order to develop effective CRISPR-based therapeutic strategies.

One limitation of our work is that we primarily used cell lines as our experimental models because they are relatively homogenous and easy to engineer. We used WGS data from primary human tumors in PCAWG to support our conclusions about the presence of complex rearrangements involving multiple oncogenes but the functional validation of enhancer hijacking in primary human tumors requires further development of 3D genomics assays and analysis methods that account for their intrinsic heterogeneity. In addition, we focused on trans-enhancer hijacking events that are caused by structural alterations. Many cis-enhancer hijacking events may be caused by other types of genomic alterations such as point mutations and small indels as well as epigenetic alterations such as DNA methylation that are capable of disrupting TAD domains, which fall outside the scope of this manuscript. Future efforts integrating additional model systems and genomic/epigenomic analysis may address these limitations.

In summary, applying a two-step HAPI analysis, we identified and characterized enhancer-hijacking events that are caused by chromosomal and extrachromosomal structural alterations in cancer. Our findings reveal enhancer-hijacking events underlying the activation of multiple oncogenes that reside together on complex amplicons such as ecDNAs.

METHODS

Cell lines

Prostate cancer cell lines LNCaP, MDAPCA2B, 22Rv1, and VCaP, ERα-positive breast cancer cell lines ZR751 and MCF7, lymphoma cell lines REC1 and DOHH2, lung squamous cell line NCIH2170, and small cell lung cancer cell lines NCIH446 and NCIH1876 were obtained from the Cancer Cell Line Encyclopedia (CCLE)[66,67]. The lymphoma cell line CUTLL1 was obtained from Sigma. VCaP, MCF7, and MDAPCA2B cells were grown in DMEM media supplemented with 10% FBS. NCIH1876 cells were grown in DMEM/F12 media supplemented with 4 μg/ml Hydrocortisone, 5 ng/ml murine EGF, 1× Insulin-Transferrin-Selenium, 10 nM of beta-estradiol, and 10% FBS. The rest of the cell lines were grown in RPMI media supplemented with 10% FBS. The identities of the cell lines were verified by either SNP fingerprinting as previously described in the CCLE project[66,67] or short tandem repeat analysis through IDEXX. All the cell lines were tested negative for mycoplasma.

HiChIP and loop calling

HiChIP assays were performed for the following cell lines: CUTLL1, REC1, VCaP, MDAPCA2B, 22Rv1, ZR751, MCF7, DOHH2, NCIH446, NCIH1876, and NCIH2170 as described in the published protocol[68] with minor modifications described previously[69]. Cross-linked chromatin was digested with the MboI restriction enzyme, filled with dNTPs including biotin labeled dATP, ligated with T4 DNA ligase, and sonicated to obtain chromatin fragments averaging ~1 kb in length using Qsonica (Q800). To enrich for chromatin interactions at active regulatory elements, antibodies of H3K27ac (Abcam, ab4729, rabbit polyclonal, 7.5 μg/HiChIP) were used for DNA fragment capture. Magnetic streptavidin beads were applied to pull down fully ligated DNA fragments and HiChIP libraries were prepared using Illumina Tagment DNA Enzyme and Buffer Kit. Sequencing was done with Illumina NextSeq or NovaSeq. HiChIP data from the other cell lines included in the study was obtained from the NCBI Gene Expression Omnibus (GEO) (Supplementary Table 1).

The paired-end HiChIP sequencing reads were aligned to human genome hg19 with the HiC-Pro pipeline[70]. Hichipper was then used to call chromatin loops where DNA anchors were assigned, and the number of PETs connecting the anchors[71]. Loop anchors were either called by hichipper based on enrichment of HiChIP sequencing reads or derived from broadpeaks called by MACS2 using publicly available H3K27ac ChIP-seq data in the corresponding cell lines, as indicated in Supplementary Table 1. If multiple biological replicates of HiChIP data are available, the reads of replicates were merged prior to analysis. Chromatin loops were removed if their PETs <3 or anchors overlap with ENCODE blacklist regions[72]. HiChIP data were visualized using the R package gTrack[3] (URL: https://github.com/mskilab/gTrack) or RCircos[73]. The gene positions in gTrack plots are based on the GENCODE database.

HAPI Analysis

We grouped the loops, called by hichipper, into 'normal,' 'abnormal cis', and 'trans' groups. Intra-chromosomal chromatin interactions spanning more than 5 kb and less than 2 mb were considered normal loops. Intra-chromosomal interactions, whose anchors are separated by at least 2 mb were considered abnormal long-range cis chromatin loops. Inter-chromosomal interactions were treated as trans chromatin loops. The H3K27ac HiChIP anchors were then intersected with +/-2.5 kb of transcription start sites (NCBI RefSeq gene list downloaded from UCSC browser in 2019) using bedtools[74] to identify promoters and enhancers. We consider a HiChIP loop that has a promoter in only one of its anchors as an enhancer-promoter loop. As promoters may act as enhancers or repressors for another gene[75-77], we excluded promoter-promoter interactions for our analysis.

To identify highly interactive genes, all types of enhancer-promoter interactions (normal, abnormal cis, and trans) described above were combined. Then for every gene, two scores were calculated to account for the number and strength of interactions between a promoter and its enhancers. The first score, defined as an enhancer contact value, is the number of enhancer anchors that are looped to the promoter of a gene. The second score, defined as interaction intensity value, is the summation of the paired-end tags (PETs) that support the interactions of the promoter and its enhancer anchors. In the case of genes that have multiple promoters, the promoter with the most PETs stemming from it is used as the primary anchor for enhancer-promoter interactions. Genes are then ranked

based on these scores individually and are presented in hockey stick plots. The inflection points of these plots, defined as the point where the line with a slope of ((max value-min value)/number of values) is tangent to the curve, are the cutoff values to define genes with a significant enhancer contacts values or interaction intensity. Genes with both scores above their respective cutoffs are defined as HAPI genes.

To determine what types of enhancers are interacting with HAPI genes, the trans and long-range cis interaction data sets with enhancer contact values and interaction intensity calculated for each gene as described above are used. Then a percent enhancer contribution is calculated for each HAPI gene using the following formula for both trans and long-range cis interaction data:

$$\text{Enhancer Contribution} = \frac{\log 10(\text{cis OR trans PETs}) * \text{cis OR trans anchors}}{(\log 10(\text{local PETs}) * \text{local anchors}) + (\log 10(\text{cis PETs}) * \text{cis anchors}) + (\log 10(\text{trans PETs}) * \text{trans anchors})}$$

Using this formula, genes with a higher trans enhancer contribution have a higher number of interacting enhancers and stronger interaction intensity coming from a different chromosome. Similarly, genes with a higher abnormal enhancer contribution have a higher number of enhancers and stronger interaction intensity coming from regions greater than 2 mb away on the same chromosome. We focus on the genes that have more than 25% of enhancer activity contributed in trans or cis.

To calculate the contributions of enhancers hijacked from the *MYC* locus to a target gene (Figures 4D, 4H, 5E, and 5G), we used the same formula but classified the enhancers into "own enhancers" (cis-interacting enhancers within 2 mb of the target gene's promoter), "hijacked MYC enhancers" (trans-interacting enhancers hijacked from regions within 2 mb of the MYC promoter), and "other enhancers".

Copy number analysis

SNP-array-based copy number segment data of cancer cell lines were downloaded from the Broad Institute CCLE Portal (URL: https://data.broadinstitute.org/ccle_legacy_data/dna_copy_number/CCLE_copynumber_2013-12-03.seg.txt; DOI: https://doi.org/10.1038/ng.3984; downloaded in November, 2020). For the CUTLL1 cell line that is not included in the CCLE copy number dataset, we used Neoloop[20] to infer the copy number profile using HiChIP reads.

Clustering analysis

An interaction score was calculated for all HAPI genes in each cell line by multiplying the $\log_{10}$ of the interaction intensity (PETs) by the enhancer contact value (interaction score = $\log_{10}$(interaction intensity) * (enhancer contact value)). The clustering analysis was based on Spearman pairwise sample correlation. Heatmap was generated using the heatmap.2 function in the gplots R package.

Pathway analysis and oncogene enrichment analysis

For HAPI genes that are shared by at least half of the cell lines in each cancer type, we performed pathway analysis and oncogene enrichment analysis. For stomach adenocarcinoma, since only one cell line (SNU16) was included in our analysis, all its HAPI genes are used to represent this cancer type. GSEA analysis (MSigDB)[78,79] was performed to identify the top KEGG legacy pathways[80,81], Reactome pathways[82], and Gene Ontology molecular functions[83,84] enriched in the shared HAPI genes. As the GSEA 'compute overlaps' tool requires no more than 500 genes as the input, for prostate cancer cell lines that have 548 shared HAPI genes identified, we ranked the genes based on their interaction scores (averaged across the corresponding cell lines) and used the top 500 genes for the analysis. The shared HAPI genes and other genes in each cancer type were also compared to a list of known oncogenes merged from three previously published oncogene data sets[34-36]. Fisher exact tests were used to calculate oncogene enrichment in the HAPI genes.

RT-qPCR

The Zymo Quick-RNA miniprep kit paired with on-column DNase I treatment was used to extract total RNA. To make cDNA for RT-qPCR, 1µg of RNA was used with the NEB LunaScript SuperMix kit. Real-time PCR was

454 completed with NEB Luna Universal qPCR Master Mix on a Bio-Rad CFX96 qPCR instrument with technical
455 replicates. The qPCR signal was first normalized to the internal reference gene *CTCF* and then the non-targeting
456 sgRNA (sgNT1). RT-qPCR primers are listed in Supplementary Table 3.
457
458 **RNA-seq**
459 CCLE RNA-seq data were downloaded from the DepMap portal (URL: https://depmap.org/portal/; downloaded
460 in May, 2022)[66]. Additional RNA-seq data for the CUTLL1 cell line was downloaded from GSE61999 and
461 aligned to hg19 using Bowtie2[85] and gene expression was calculated using RSEM[86]. The log2(TPM+1) values of
462 CUTLL1 were merged with the CCLE dataset for comparison.
463
464 For determining androgen dependence of *ETV1* expression in LNCaP and 22RV1, RNA-seq data for both cell
465 lines grown in androgen-depleted media treated with DHT or vehicle were downloaded from GSE92574. Counts
466 per million reads were used to determine the $\log_2$ fold change in expression between control and treated cells. We
467 generated RNA-seq data for the CRISPRi experiment in NCIH2170 cells and performed edgeR analysis[87] to
468 identify differentially expressed genes (FDR<0.05; Fold change>1.5).
469
470 **ChIP-seq**
471 ChIP-seq was performed in LNCaP-dCas9-KRAB-MeCP2 cells treated with enhancer-targeting or non-targeting
472 sgRNAs, as previously described[69]. Five million cells were crosslinked with 1% formaldehyde (diluted in 1×
473 PBS) and lysed with Lysis Buffer I (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP40) and then Lysis Buffer II (1×
474 PBS, 1% NP40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate) supplemented with protease inhibitors.
475 Chromatin extract was sonicated with Bioruptor (20 min; pulse: 30 sec on/30 sec off; high amplitude) and
476 immunoprecipitated with H3K27ac antibody (Abcam, ab4729, rabbit polyclonal, 4 µg/ChIP) premixed with
477 Dynabeads A and G. ChIP-seq libraries were prepared using NEBNext DNA Ultra II library prep kit and
478 sequenced by Illumina MiSeq.
479
480 Remaining ChIP-seq data used in the study were obtained from previously published datasets (Supplementary
481 Table 1). ChIP-seq reads and genomic input reads, if available, were aligned to the hg19 using Bowtie2[85].
482 Samtools[88] was used to sort and index aligned reads followed by MACS2[89] to call significant peaks (q value <
483 0.05) and calculate the ChIP-seq signal. Super-enhancers were called using H3K27ac ChIP-seq signal using the
484 Homer pipeline[90]. We focused on super-enhancers that contain at least one enhancer anchor of a HiChIP loop
485 and linked them to gene promoters utilizing the pairToBed function of bedtools[74].
486
487 **CRISPR-mediated enhancer repression**
488 To repress enhancers via CRISPR, we used a lentiviral dCas9-KRAB-MeCP2 vector that was previously
489 generated[28]. Cells were infected with the vector and selected with blasticidin (10 µg/ml) for >5 days to stably
490 express dCas9-KRAB-MeCP2. sgRNAs were designed to target the peaks of DNase I hypersensitive regions and
491 the trough of H3K27ac ChIP-seq profiles for either enhancer or promoter regions of interest. The promoter- or
492 enhancer-targeting sgRNAs and non-targeting sgRNAs with no genome recognition sites were cloned into
493 LentiGuide-Puro (Addgene: 52963). The cells stably expressing dCas9-KRAB-MeCP2 were infected with these
494 vectors and then selected with puromycin (2 µg/ml) for at least 3 days before extracting RNA. All sgRNA
495 sequences used are listed in Supplementary Table 3.
496
497 **WGS data processing and structural variant calling**
498 WGS data for LNCaP, MCF7, ZR751, NCIH446, NCIH2170, COLO320DM, and SNU16 were downloaded from
499 published studies (Supplementary Table 1). Raw FASTQ files were aligned to hg19 using BWA-MEM[91] and the
500 aligned reads were sorted and indexed with Samtools[88]. Structural variants and their breakpoints were called using
501 GRIDSS[92,93]. The output VCF files were converted to bedpe files and filtered by QUAL ≥ 1000, AS>0, and
502 RAS>0 for translocation visualization using the Bioconductor package StructuralVariantAnnotation[94].
503
504 **PCAWG structural variants**

PCAWG consensus call sets of structural variants in tumor samples were downloaded from the ICGC Data Portal (URL: https://dcc.icgc.org/releases/PCAWG/consensus_sv; release date: Nov 25, 2019) and presented using RCircos[73].

For PCAWG complex amplicon analysis, we used the publicly available AmpliconArchitect results released from the AmpliconRepository (URL: https://ampliconrepository.org; downloaded in October 2023). We used the "aggregated results" file in the "PCAWG – full" dataset released on August 29, 2023, which listed the DNA segments included in each amplicon. To identify DNA segments harboring oncogenes, we used a list of known oncogenes merged from three previously published oncogene data sets[34-36].

ecDNA calling

To call ecDNAs in CCLE cell lines we employed the AmpliconSuite pipeline (URL: https://github.com/jluebeck/AmpliconSuite-pipeline). Briefly, WGS data were aligned to hg19 using BWA[91] and amplified regions were called with cnvkit[95]. Seed regions for AmpliconArchitect[53] were identified using the default settings of the prepareAA script. AmpliconArchitect[53] was used to identify ecDNAs. AmpliconClassifier[64] was used to determine if the amplicons represent ecDNAs and CycleViz (URL: https://github.com/jluebeck/CycleViz) was used to visualize the reconstructed ecDNA.

TCGA ATAC analysis

We downloaded the TCGA ATAC-seq data[96] from the UCSC Xena Browser (URL: https://xenabrowser.net/datapages/?dataset=TCGA_ATAC_peak_Log2Counts_dedup_sample&host=https%3A%2F%2Fatacseq.xenahubs.net&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443; version: 2018-06-21), which contains log2-transformed normalized ATAC-seq insertion counts for each tested sample across the identified pan-cancer ATAC peaks. To compare chromatin openness of the hijacked enhancers across different cancer types, we first took enhancers that are hijacked by HAPI genes in each selected cell line, and then summed the log2 normalized ATAC-seq counts from ATAC peaks overlapping with the hijacked enhancers to determine its overall chromatin openness for each sample included in the tested cancer types.

Dual-color DNA FISH

Metaphase arrest of NCIH2170 cells was performed by incubation with KaryoMAX (Gibco) at 0.1 µg/mL overnight. Single cell suspension of the metaphase arrested cells was resuspended in 1xPBS followed by incubation with 75 mM KCL for 15-30 min. The cells were fixed with 3:1 methanol to glacial acetic acid overnight and then washed three times with the fixative before being dropped on a coverslip. The coverslip was dried for 5 min at 60°C before being immersed in 2xSSC in preparation for DNA FISH.

Coverslips with fixed cells in metaphase were washed twice for 5 min with 2xSSC. The coverslips were then washed with 50% formamide/1×SSC for 4 min and dehydrated through a series of ethanol incubations (70%, 85%, and 100%) for 2 min each. Next, the coverslips were placed on slides containing a 1:1 ratio of ERBB2 to MYC FISH probes diluted in the manufacturer's hybridization buffer (Empire Genomics). The coverslips were sealed with rubber cement and denatured in a heated incubator at 82°C for 1 min, followed by incubation in a humidified chamber at 37°C overnight. The coverslip was washed once with 0.4×SSC/0.3% Igepal at 73°C and twice with 2×SSC/0.1% Igepal each for 2min. A final two washes with 2xSSC were performed before mounting the coverslip using Prolong Antifade Mountants with DAPI (Thermo Fisher).

For interphase DNA FISH, NCIH2170 cells were seeded on collagen-coated coverslips and fixed with 4% PFA in 1×PBS for 10 min, followed by quenching for 10 min with 0.1 M Glycine in 1×PBS. The cell membrane was then permeabilized for 10 min with 0.5% Triron-X 100. An additional three 5 min rinses with 1xPBS proceeded RNase A treatment at 0.1 ug/mL for 1 hour at 37°C, followed by three more rinses with 1×PBS before a 10 min incubation with 0.7% Triton X-100 in 0.1 M HCl on ice. After three rinses with PBS, the coverslips were incubated for 30 min in 2x SSC with 50% formamide. During the incubation, the hybridization buffer containing 2×SSC, 1 µl each of MYC and ERBB2 FISH probes (Empire Genomics), 50% (v/v) formamide, 2% dextran sulfate, and 0.2 µl 10 mg/ml salmon sperm DNA was prepared and 10 µl was placed on the slide. The coverslips

were then placed on the hybridization buffer on the slide and sealed with rubber cement. Denaturation was performed in a heated incubator at 78°C for 4 min followed by incubation overnight at 37°C in a humidified chamber. The coverslips were then washed three times with 50% formamide in 2×SSC for 5 min each at 37°C. An additional three washes with 2×SSC at 37°C were performed before mounting with Prolong Antifade Mountants with DAPI (Thermo Fisher).

All DNA FISH samples were imaged using a Nikon Eclipse Ti-2 microscope coupled with a CSU-W1 spinning disk confocal scanner unit (Yokogawa). Specifically, the fluorescent cell samples were excited using three laser lines: 405 nm, 488 nm, and 561 nm. The fluorescence signals were captured and detected using a Plan Apo 60x objective (Nikon, NA 1.40) and a high-speed Kinetix sCMOS camera (Photometrics), respectively. This microscope utilizes the Nikon Perfect Focus System, an interferometry-based focus lock that allows the capture of multipoint images without loss of focus. All imaging data were examined and quantified using Fiji and Cell Profiler.

Virtual 4C Analysis

For HiChIP data from NCIH2170 cells with and without CRISPRi-mediated *MYC* promoter repression, we presented and compared their virtual 4C signal anchored at *MYC* or *ERBB2* promoters. Briefly, we first identified all valid read pairs (HiC-Pro output) that have one end mapped to a 2kb window centered at *MYC* or *ERBB2* TSS. We then used the other end of the read pairs to generate a bedgraph file at a 10bp resolution. The bedgraph signal was then scaled based on the total number of PETs identified in each condition. As most of enhancers used by *MYC* and *ERBB2* come from the chromosome 8 locus surrounding *MYC*, we presented Virtual 4C signal at five regions from the *MYC* locus with a range of enhancer activity as well as the averaged Virtual 4C signal for the top 50 HiChIP signal-derived peaks at the *MYC* locus that are ranked based on their interactions with the *MYC* promoter.

DATA AVAILABILITY

The publicly available H3K27ac HiChIP data used in this study were obtained from the NCBI Gene Expression Omnibus (GEO) at GSE227242 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE227242], GSE97585 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE97585], GSE157381 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE157381], GSE147854 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE147854], GSE166232 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE166232], GSE151002 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE151002], GSE159985 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE159985], and GSE188401 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE188401] (specific cell lines are described in Supplementary Table 1). The publicly availably ChIP-seq data used in this study were downloaded from GEO at GSE73994 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE73994], GSE63109 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE63109], GSE69558 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE69558], GSE105760 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE105760], GSE105627 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE105627], GSE85158 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE85158], GSE57436 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE57436], GSE86743 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE86743], GSE73319 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE73319], GSE115123 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE115123], and GSE159972 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE159972] for H3K27ac, and GSE174109 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE174109] and GSE94013 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE94013] for AR ChIP-seq (specific cell lines are described in Supplementary Table 1). TCGA copy number segment data for all applicable cell lines were downloaded from

the Broad Institute Portal (URL: https://data.broadinstitute.org/ccle_legacy_data/dna_copy_number/CCLE_copynumber_2013-12-03.seg.txt). CCLE RNA-seq data were downloaded from the DepMap portal (URL: https://depmap.org/portal/). RNA-seq data for CUTLL1 was downloaded from GSE61999 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE61999]. Publicly available WGS data for cell lines were downloaded from NCBI SRA: PRJNA361316 [https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA361316], PRJNA523380 [https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA523380], and PRJNA506071 [https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA506071] (specific cell lines are described in Supplementary Table 1). PCAWG consensus call sets of structural variants in tumor samples were downloaded from the ICGC Data Portal (URL: https://dcc.icgc.org/releases/PCAWG/consensus_sv). Complex amplicon data from PCAWG were downloaded from the AmpliconRepository (https://ampliconrepository.org). Publicly available TCGA ATAC-seq data were downloaded from the UCSC Xena Browser (URL: https://xenabrowser.net/datapages/?dataset=TCGA_ATAC_peak_Log2Counts_dedup_sample&host=https%3A%2F%2Fatacseq.xenahubs.net&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443). The ChIP-seq, HiChIP, and RNA-seq data generated in this study have been deposited to GEO under the series GSE228247 [https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE228247]. The remaining data are available within the Article, Supplementary Information, Supplementary Data, or Source Data file. Source data are provided with this paper.

CODE AVAILABILITY

HAPI analysis is available for public use with test examples as a module in Code Ocean at https://codeocean.com/capsule/2804329

632 **REFERENCES**

633 1 Zhang, X. & Meyerson, M. Illuminating the noncoding genome in cancer. *Nat Cancer* **1**, 864-872 (2020).
634 https://doi.org:10.1038/s43018-020-00114-3

635 2 Li, Y. *et al.* Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112-121
636 (2020). https://doi.org:10.1038/s41586-019-1913-9

637 3 Hadi, K. *et al.* Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer
638 Genome Graphs. *Cell* **183**, 197-210 e132 (2020). https://doi.org:10.1016/j.cell.2020.08.006

639 4 Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. *Cell* **153**, 17-37 (2013).
640 https://doi.org:10.1016/j.cell.2013.03.002

641 5 Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719-724 (2009).
642 https://doi.org:10.1038/nature07943

643 6 Soda, M. *et al.* Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer.
644 *Nature* **448**, 561-566 (2007). https://doi.org:10.1038/nature05945

645 7 Bartram, C. R. *et al.* Translocation of c-ab1 oncogene correlates with the presence of a Philadelphia
646 chromosome in chronic myelocytic leukaemia. *Nature* **306**, 277-280 (1983).
647 https://doi.org:10.1038/306277a0

648 8 Rheinbay, E. *et al.* Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **578**,
649 102-111 (2020). https://doi.org:10.1038/s41586-020-1965-x

650 9 Groschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2
651 deregulation in leukemia. *Cell* **157**, 369-381 (2014). https://doi.org:10.1016/j.cell.2014.02.019

652 10 Ryan, R. J. *et al.* Detection of Enhancer-Associated Rearrangements Reveals Mechanisms of Oncogene
653 Dysregulation in B-cell Lymphoma. *Cancer Discov* **5**, 1058-1071 (2015). https://doi.org:10.1158/2159-
654 8290.CD-15-0370

655 11 Northcott, P. A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature*
656 **511**, 428-434 (2014). https://doi.org:10.1038/nature13379

657 12 Tsujimoto, Y. *et al.* Clustering of breakpoints on chromosome 11 in human B-cell neoplasms with the
658 t(11;14) chromosome translocation. *Nature* **315**, 340-343 (1985). https://doi.org:10.1038/315340a0

659 13 Battey, J. *et al.* The human c-myc oncogene: structural consequences of translocation into the IgH locus
660 in Burkitt lymphoma. *Cell* **34**, 779-787 (1983). https://doi.org:10.1016/0092-8674(83)90534-2

661 14 Dubois, F. P. B. *et al.* Structural variants shape driver combinations and outcomes in pediatric high-grade
662 glioma. *Nat Cancer* **3**, 994-1011 (2022). https://doi.org:10.1038/s43018-022-00403-z

663 15 Helmsauer, K. *et al.* Enhancer hijacking determines extrachromosomal circular MYCN amplicon
664 architecture in neuroblastoma. *Nat Commun* **11**, 5823 (2020). https://doi.org:10.1038/s41467-020-19452-
665 y

666 16 Bandopadhayay, P. *et al.* MYB-QKI rearrangements in angiocentric glioma drive tumorigenicity through
667 a tripartite mechanism. *Nat Genet* **48**, 273-282 (2016). https://doi.org:10.1038/ng.3500

668 17 Drier, Y. *et al.* An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma.
669 *Nat Genet* **48**, 265-272 (2016). https://doi.org:10.1038/ng.3502

670 18 Weischenfeldt, J. *et al.* Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2
671 in enhancer hijacking. *Nat Genet* **49**, 65-74 (2017). https://doi.org:10.1038/ng.3722

672 19 He, B. *et al.* Diverse noncoding mutations contribute to deregulation of cis-regulatory landscape in
673 pediatric cancers. *Sci Adv* **6**, eaba3064 (2020). https://doi.org:10.1126/sciadv.aba3064

674 20 Wang, X. *et al.* Genome-wide detection of enhancer-hijacking events from chromatin interaction data in
675 rearranged genomes. *Nat Methods* **18**, 661-668 (2021). https://doi.org:10.1038/s41592-021-01164-w

676 21 Shoshani, O. *et al.* Chromothripsis drives the evolution of gene amplification in cancer. *Nature* **591**, 137-
677 141 (2021). https://doi.org:10.1038/s41586-020-03064-z

678 22 Hung, K. L., Mischel, P. S. & Chang, H. Y. Gene regulation on extrachromosomal DNA. *Nat Struct Mol*
679 *Biol* **29**, 736-744 (2022). https://doi.org:10.1038/s41594-022-00806-7

680 23 Wu, S., Bafna, V., Chang, H. Y. & Mischel, P. S. Extrachromosomal DNA: An Emerging Hallmark in
681 Human Cancer. *Annu Rev Pathol* **17**, 367-386 (2022). https://doi.org:10.1146/annurev-pathmechdis-
682 051821-114223

683 24 Yi, E., Chamorro Gonzalez, R., Henssen, A. G. & Verhaak, R. G. W. Extrachromosomal DNA
684 amplifications in cancer. *Nat Rev Genet* **23**, 760-771 (2022). https://doi.org:10.1038/s41576-022-00521-
685 5

686 25 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity
687 genes. *Cell* **153**, 307-319 (2013). https://doi.org:10.1016/j.cell.2013.03.035

688 26 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947 (2013).
689 https://doi.org:10.1016/j.cell.2013.09.053

690 27 Loven, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-
691 334 (2013). https://doi.org:10.1016/j.cell.2013.03.036

692 28 Liu, Y. *et al.* A predominant enhancer co-amplified with the SOX2 oncogene is necessary and sufficient
693 for its expression in squamous cancer. *Nat Commun* **12**, 7139 (2021). https://doi.org:10.1038/s41467-021-
694 27055-4

695 29 Zhang, X. *et al.* Identification of focally amplified lineage-specific super-enhancers in human epithelial
696 cancers. *Nat Genet* **48**, 176-182 (2016). https://doi.org:10.1038/ng.3470

697 30 Zhang, X. *et al.* Somatic Superenhancer Duplications and Hotspot Mutations Lead to Oncogenic
698 Activation of the KLF5 Transcription Factor. *Cancer Discov* **8**, 108-125 (2018).
699 https://doi.org:10.1158/2159-8290.CD-17-0532

700 31 Takeda, D. Y. *et al.* A Somatically Acquired Enhancer of the Androgen Receptor Is a Noncoding Driver
701 in Advanced Prostate Cancer. *Cell* **174**, 422-432 e413 (2018). https://doi.org:10.1016/j.cell.2018.05.037

702 32 Viswanathan, S. R. *et al.* Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by
703 Linked-Read Genome Sequencing. *Cell* **174**, 433-447 e419 (2018).
704 https://doi.org:10.1016/j.cell.2018.05.036

705 33 Morton, A. R. *et al.* Functional Enhancers Shape Extrachromosomal Oncogene Amplifications. *Cell* **179**,
706 1330-1341 e1313 (2019). https://doi.org:10.1016/j.cell.2019.10.039

707 34 Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* **2017**, 1-16
708 (2017). https://doi.org:10.1200/PO.17.00011

709 35 Liu, Y., Sun, J. & Zhao, M. ONGene: A literature-based database for human oncogenes. *J Genet Genomics*
710 **44**, 119-121 (2017). https://doi.org:10.1016/j.jgg.2016.12.004

711 36 Sondka, Z. *et al.* The COSMIC Cancer Gene Census: describing genetic dysfunction across all human
712 cancers. *Nat Rev Cancer* **18**, 696-705 (2018). https://doi.org:10.1038/s41568-018-0060-1

713 37 Dekker, J. & Heard, E. Structural and functional diversity of Topologically Associating Domains. *FEBS*
714 *Lett* **589**, 2877-2884 (2015). https://doi.org:10.1016/j.febslet.2015.08.044

715 38 Flavahan, W. A. *et al.* Altered chromosomal topology drives oncogenic programs in SDH-deficient
716 GISTs. *Nature* **575**, 229-233 (2019). https://doi.org:10.1038/s41586-019-1668-3

717 39 Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**,
718 110-114 (2016). https://doi.org:10.1038/nature16490

719 40 Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR–
720 Cas9 essentiality screens in cancer cells. *Nature Genetics* **49**, 1779-1784 (2017).
721 https://doi.org:10.1038/ng.3984

722 41 Dixon, J. R. *et al.* Integrative detection and analysis of structural variation in cancer genomes. *Nat Genet*
723 **50**, 1388-1398 (2018). https://doi.org:10.1038/s41588-018-0195-8

724 42 Parolia, A. *et al.* Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer.
725 *Nature* **571**, 413-418 (2019). https://doi.org:10.1038/s41586-019-1347-4

726 43 Tomlins, S. A. *et al.* Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions
727 in prostate cancer. *Nature* **448**, 595-599 (2007). https://doi.org:10.1038/nature06024

728 44 Cai, C. *et al.* ETV1 is a novel androgen receptor-regulated gene that mediates prostate cancer cell invasion.
729 *Mol Endocrinol* **21**, 1835-1846 (2007). https://doi.org:10.1210/me.2006-0480

730 45 Ortiz, A. B. *et al.* Prognostic significance of cyclin D1 protein expression and gene amplification in
731 invasive breast carcinoma. *PLoS One* **12**, e0188068 (2017). https://doi.org:10.1371/journal.pone.0188068

732 46 Wu, S. *et al.* Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**,
733 699-703 (2019). https://doi.org:10.1038/s41586-019-1763-5

734 47 Hung, K. L. *et al.* ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature* **600**, 731-
735 736 (2021). https://doi.org:10.1038/s41586-021-04116-8

736 48 Salari, K. *et al.* CDX2 is an amplified lineage-survival oncogene in colorectal cancer. *Proc Natl Acad Sci*
737 *U S A* **109**, E3196-3205 (2012). https://doi.org:10.1073/pnas.1206004109

738 49 Werling, R. W., Yaziji, H., Bacchi, C. E. & Gown, A. M. CDX2, a highly sensitive and specific marker
739 of adenocarcinomas of intestinal origin: an immunohistochemical survey of 476 primary and metastatic
740 carcinomas. *Am J Surg Pathol* **27**, 303-310 (2003). https://doi.org:10.1097/00000478-200303000-00003

741 50 Hung, K. L. *et al.* Targeted profiling of human extrachromosomal DNA by CRISPR-CATCH. *Nat Genet*
742 **54**, 1746-1754 (2022). https://doi.org:10.1038/s41588-022-01190-0

743 51 Hong, S. H. *et al.* APIP, an ERBB3-binding partner, stimulates erbB2-3 heterodimer formation to promote
744 tumorigenesis. *Oncotarget* **7**, 21601-21617 (2016). https://doi.org:10.18632/oncotarget.7802

745 52 Hou, W., Kong, L., Hou, Z. & Ji, H. CD44 is a prognostic biomarker and correlated with immune
746 infiltrates in gastric cancer. *BMC Med Genomics* **15**, 225 (2022). https://doi.org:10.1186/s12920-022-
747 01383-w

748 53 Deshpande, V. *et al.* Exploring the landscape of focal amplifications in cancer using AmpliconArchitect.
749 *Nat Commun* **10**, 392 (2019). https://doi.org:10.1038/s41467-018-08200-y

750 54 Estrada-Bernal, A. *et al.* Tarloxotinib Is a Hypoxia-Activated Pan-HER Kinase Inhibitor Active Against
751 a Broad Range of HER-Family Oncogenes. *Clin Cancer Res* **27**, 1463-1475 (2021).
752 https://doi.org:10.1158/1078-0432.CCR-20-3555

753 55 Ise, N., Omi, K., Nambara, D., Higashiyama, S. & Goishi, K. Overexpressed HER2 in NSCLC is a
754 Possible Therapeutic Target of EGFR Inhibitors. *Anticancer Res* **31**, 4155-4161 (2011).

755 56 Pongor, L. S. *et al.* Extrachromosomal DNA Amplification Contributes to Small Cell Lung Cancer
756 Heterogeneity and Is Associated with Worse Outcomes. *Cancer Discov* **13**, 928-949 (2023).
757 https://doi.org:10.1158/2159-8290.CD-22-0796

758 57 Aaltonen, L. A. *et al.* Pan-cancer analysis of whole genomes. *Nature* **578**, 82-93 (2020).
759 https://doi.org:10.1038/s41586-020-1969-6

760 58 Xu, Z. *et al.* Structural variants drive context-dependent oncogene activation in cancer. *Nature* **612**, 564-
761 572 (2022). https://doi.org:10.1038/s41586-022-05504-4

762 59 Hsieh, T. H. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**,
763 108-119 (2015). https://doi.org:10.1016/j.cell.2015.05.048

764 60 Fang, R. *et al.* Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq.
765 *Cell Res* **26**, 1345-1348 (2016). https://doi.org:10.1038/cr.2016.137

766 61 Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-
767 64 (2009). https://doi.org:10.1038/nature08497

768 62 Song, M. *et al.* Cell-type-specific 3D epigenomes in the developing human cortex. *Nature* **587**, 644-649
769 (2020). https://doi.org:10.1038/s41586-020-2825-4

770 63 Li, Y., Zhou, S., Schwartz, D. C. & Ma, J. Allele-Specific Quantification of Structural Variations in
771 Cancer Genomes. *Cell Syst* **3**, 21-34 (2016). https://doi.org:10.1016/j.cels.2016.05.007

772 64 Luebeck, J. *et al.* Extrachromosomal DNA in the cancerous transformation of Barrett's esophagus. *Nature*
773 **616**, 798–805 (2023). https://doi.org/10.1038/s41586-023-05937-5

774 65 Lee, J. J. *et al.* ERα-associated translocations underlie oncogene amplifications in breast cancer. *Nature*
775 **618**, 1024-1032 (2023). https://doi.org:10.1038/s41586-023-06057-w

776 66 Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**,
777 503-508 (2019). https://doi.org:10.1038/s41586-019-1186-3

67 Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607 (2012). https://doi.org:10.1038/nature11003

68 Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **13**, 919-922 (2016). https://doi.org:10.1038/nmeth.3999

69 Liu, Y. *et al.* Chromatin Looping Shapes KLF5-Dependent Transcriptional Programs in Human Epithelial Cancers. *Cancer Res* **80**, 5464-5477 (2020). https://doi.org:10.1158/0008-5472.CAN-20-1287

70 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259 (2015). https://doi.org:10.1186/s13059-015-0831-x

71 Lareau, C. A. & Aryee, M. J. hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat Methods* **15**, 155-156 (2018). https://doi.org:10.1038/nmeth.4583

72 Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354 (2019). https://doi.org:10.1038/s41598-019-45839-z

73 Zhang, H., Meltzer, P. & Davis, S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics* **14**, 244 (2013). https://doi.org:10.1186/1471-2105-14-244

74 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010). https://doi.org:10.1093/bioinformatics/btq033

75 Oudelaar, A. M. *et al.* A revised model for promoter competition based on multi-way chromatin interactions at the alpha-globin locus. *Nat Commun* **10**, 5412 (2019). https://doi.org:10.1038/s41467-019-13404-x

76 Cho, S. W. *et al.* Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* **173**, 1398-1412 e1322 (2018). https://doi.org:10.1016/j.cell.2018.03.068

77 Wijgerde, M., Grosveld, F. & Fraser, P. Transcription complex stability and chromatin dynamics in vivo. *Nature* **377**, 209-213 (1995). https://doi.org:10.1038/377209a0

78 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011). https://doi.org:10.1093/bioinformatics/btr260

79 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005). https://doi.org:10.1073/pnas.0506580102

80 Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* **51**, D587-D592 (2023). https://doi.org:10.1093/nar/gkac963

81 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000). https://doi.org:10.1093/nar/28.1.27

82 Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498-d503 (2020). https://doi.org:10.1093/nar/gkz1031

83 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000). https://doi.org:10.1038/75556

815 84 Gene Ontology, C. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325-
816 D334 (2021). https://doi.org:10.1093/nar/gkaa1113

817 85 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359
818 (2012). https://doi.org:10.1038/nmeth.1923

819 86 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a
820 reference genome. *BMC Bioinformatics* **12**, 323 (2011). https://doi.org:10.1186/1471-2105-12-323

821 87 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential
822 expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2009).
823 https://doi.org:10.1093/bioinformatics/btp616

824 88 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
825 https://doi.org:10.1093/bioinformatics/btp352

826 89 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
827 https://doi.org:10.1186/gb-2008-9-9-r137

828 90 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory
829 elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).
830 https://doi.org:10.1016/j.molcel.2010.05.004

831 91 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform.
832 *Bioinformatics* **26**, 589-595 (2010). https://doi.org:10.1093/bioinformatics/btp698

833 92 Cameron, D. L. *et al.* GRIDSS2: comprehensive characterisation of somatic structural variation using
834 single breakend variants and structural variant phasing. *Genome Biol* **22**, 202 (2021).
835 https://doi.org:10.1186/s13059-021-02423-x

836 93 Cameron, D. L. *et al.* GRIDSS: sensitive and specific genomic rearrangement detection using positional
837 de Bruijn graph assembly. *Genome Res* **27**, 2050-2060 (2017). https://doi.org:10.1101/gr.222109.117

838 94 StructuralVariantAnnotation: Variant annotations for structural variants (2019).

839 95 Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection
840 and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* **12**, e1004873 (2016).
841 https://doi.org:10.1371/journal.pcbi.1004873

842 96 Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**,
843 eaav1898 (2018). https://doi.org:10.1126/science.aav1898
844

845
846

ACKNOWLEDGEMENTS

We thank Matthew Meyerson, John Pulice, and Dean Tantin for insightful discussions.

The support and resources from the Center for High Performance Computing and the HSC Imaging Core at the University of Utah are gratefully acknowledged.

Funding for this work was provided by R00CA215244 and R37CA263505 from the National Cancer Institute (XZ), R35GM150941 from the National Institute of General Medical Sciences (YL), 132596-RSG-18-197-01-DMC from the American Cancer Society (KEV), Huntsman Cancer Institute's National Cancer Institute Cancer Center Support Grant (P30CA042014) and the Huntsman Cancer Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

KLM, ERW, HEJ, SDB, and XZ performed the computational analysis for the manuscript, and KLM, CD, NEP, and XZ performed the wet-laboratory experiments. KEV and XZ designed the study. JG, SDB, YL, KEV, and XZ supervised the study. KLM, KEV, and XZ wrote the manuscript with input from other co-authors.

COMPETING INTERESTS

The authors declare no competing interest regarding this work.

867 **FIGURE LEGENDS**

868 **Figure 1: HAPI analysis identifies cancer-related genes that are highly interactive with enhancers.**
869 **A.** Schematic models illustrating the identification of HAPI genes based on enhancer interactions.
870 **B.** Plotted are the number of enhancers (x-axis) and interaction intensity (y-axis) for each gene in LNCaP
871 cells. HAPI cutoff is defined by the inflection point of each scoring metric.
872 **C.** Spearman correlation clustering of cancer cell lines based on normalized interaction scores of their
873 associated HAPI genes. BRCA: breast cancer; COAD: colorectal adenocarcinoma; LUSC: lung squamous
874 cell carcinoma; PRAD: prostate adenocarcinoma; SCLC: small cell lung cancer; STAD: stomach
875 adenocarcinoma.
876 **D.** The percentage of HAPI genes (shared by at least half of the cell lines in each cancer type) that are
877 previously annotated as oncogenes. The numbers outside and within each bar indicate the total number of
878 shared HAPI (red) or other (grey) genes and the ones annotated as oncogenes, respectively. P-values are
879 derived from two-sided Fisher's exact tests.
880

881 **Figure 2: HAPI analysis identifies known and novel enhancer-hijacking events.**
882 **A.** Schematic models illustrating the identification of enhancer hijacking HAPI genes based on the enhancer
883 origin.
884 **B.** The number of cancer cell lines that are identified to contain trans- and cis-enhancer hijacking events in
885 our cohort.
886 **C.** The number of identified trans-enhancer hijacking genes found in each cancer cell line in our cohort. Two
887 immortalized epithelial cell lines are used as negative controls.
888 **D.** Plotted are the calculated trans-enhancer contribution (x-axis) and the copy number estimated by SNP-
889 array data (y-axis) for each HAPI gene in selected cell lines. Besides are the expression levels for the
890 highlighted gene in all CCLE cell lines (violin plot), the trans-enhancer hijacking cell line (red), and other
891 cell lines of the same lineage (blue).
892 **E.** Same as D but with regards to two cis-enhancer hijacking genes in VCaP and CORL88 cells.
893

894 **Figure 3: Functional validation of identified trans-enhancer hijacking events.**
895 **A.** HiChIP and ChIP-seq signal at chromosomal regions containing *ETV1* and the hijacked enhancers in
896 MDAPCA2B and LNCaP cells. The H3K27ac ChIP-seq signal in MDAPCA2B cells is derived from
897 HiChIP reads. In LNCaP cells, WGS-identified breakpoints support translocations underlying the
898 enhancer-hijacking events. The DNA position of the left breakpoint on chromosome 7 in LNCaP cells is
899 noted.
900 **B.** RT-qPCR measuring expression changes of *ETV1* and *FOXA1* after CRISPRi of each enhancer e1-e4 or
901 the promoter of *ETV1* in LNCaP cells. Expression levels were normalized to cells treated with a non-
902 targeting sgRNA (NT1). N = 3 biological replicates. Data are presented as mean values +/- SEM. P-values
903 were derived from two-sided t-tests. Source data are provided as a Source Data file.
904 **C.** HiChIP and ChIP-seq signal at chromosomal regions containing *CCND1* and the hijacked enhancers in
905 REC1 and ZR751 cells. In ZR751 cells, WGS-identified breakpoints support translocations underlying
906 the enhancer-hijacking events.
907 **D.** RT-qPCR measuring expression changes of *CCND1* and *DUSP4* after CRISPRi of each enhancer e1-e5,
908 the promoter of *CCND1*, or a combination of the five enhancers in ZR751 cells. Expression levels were
909 normalized to cells treated with a non-targeting sgRNA (NT1). N = 3 biological replicates. Data are
910 presented as mean values +/- SEM. P-values were derived from two-sided t-tests. Source data are provided
911 as a Source Data file.
912

913 **Figure 4: HAPI analysis identifies enhancer hijacking on known ecDNAs.**
914 **A.** Schematic models illustrating the identification of enhancer hijacking events on ecDNAs that involve
915 multiple oncogenes from different chromosomal regions.
916 **B.** Plotted are the calculated trans-enhancer contribution (x-axis) and the copy number estimated by SNP-
917 array data (y-axis) for each HAPI gene in COLO320DM cells.

C. HiChIP, ChIP-seq, and SNP-array-based copy number signal at the DNA segments of the ecDNA found in COLO320DM cells. The HiChIP data and the ecDNA segments in COLO320DM were published by Hung et al[47].

D. Enhancer distribution of *MYC* and the additional oncogenes harbored in the ecDNA found in COLO320DM cells. Enhancers within 2mb of distance to the promoter of *MYC* or any other gene in the endogenous chromosomal locus will be considered as the gene's 'own' enhancers, otherwise will be considered as 'other' enhancers. Source data are provided as a Source Data file.

E. Circos plot: layer 1=chromosome, 2=CNV, 3=enhancers (grey) and super-enhancers (black), 4=HAPI TSS, 5=ecDNA segments, center=E-P loops (black) and WGS-predicted translocations (red) in COLO320DM cells.

F-I. Same as B-E but with regards to the ecDNAs in SNU16 cells. The HiChIP data and the ecDNA segments in SNU16 were published by Hung et al[47]. Note that the *APIP* and *PDHX* genes share the same promoter anchor for the HiChIP analysis. Source data for the panel H are provided as a Source Data file.

Figure 5: Enhancer hijacking is prevalent on complex amplicons.

A-C. Plotted are the calculated trans-enhancer contribution (x-axis) and the copy number estimated by SNP-array data (y-axis) for each HAPI gene in NCIH2170, NCIH446, and MCF7 cells.

D. HiChIP, H3K27ac ChIP-seq (derived from HiChIP reads), and copy number profiles at DNA segments of the *MYC* ecDNA in NCIH2170 cells.

E. Enhancer distribution of *MYC* and the additional oncogenes harbored in the ecDNA found in NCIH2170 cells. Source data are provided as a Source Data file.

F. HiChIP, H3K27ac ChIP-seq, and copy number profiles at DNA segments of the *MYC* amplicon in NCIH446 cells.

G. Enhancer distribution of *MYC* and the additional oncogenes harbored in the amplicon found in NCIH446 cells. Source data are provided as a Source Data file.

H. Summary of AmpliconArchitect data from PCAWG tumor samples: most of oncogene amplicons contain additional trans or cis DNA segments. 37.6% of these complex amplicons contain multiple oncogenes originating from different DNA segments.

Figure 6: Functional exploration of the *MYC-ERBB2* chimeric ecDNA.

A. Representative images of metaphase chromosome spreads showing highly amplified copies of *MYC* (red) and *ERBB2* (green) colocalizing outside of chromosomal DNA (blue, DAPI staining) in NCIH2170 cells. In total, 20 metaphase spreads of high quality were collected from 7 separate images, all indicating the presence of *MYC-ERBB2* chimeric ecDNAs in high abundance. Scale bar, 5 µm. Four enlarged regions of interest (ROI) were highlighted. Scale bar, 2 µm.

B. AUC values representing the sensitivity of lung squamous carcinoma cell lines to Neratinib and Lapatinib. Source data are provided as a Source Data file.

C. Schematic illustration of the experimental design of the *MYC*-promoter CRISPRi assay.

D. Virtual 4C signal (derived from HiChIP data) showing interactions with the *MYC* promoter at the chromosome 8 locus (now on ecDNA) in NCIH2170 cells (sgNT2, replicate 1).

E. Virtual 4C signal between the promoter of *MYC* (upper) or *ERBB2* (lower) and five highlighted enhancer regions representing different levels of interaction intensity in NCIH2170 cells with and without *MYC* promoter repression (sg*MYC*-prom vs. sgNT2). SEM: standard error of the mean. N=2 biological replicates.

F. The averaged virtual 4C signal between the top 50 *MYC*-interacting regions from the *MYC* locus (ranked by their interactions with the *MYC* promoter) and the promoter of *MYC* (left) or *ERBB2* (right).

G. A volcano plot presenting RNA-seq results from NCIH2170 cells with and without *MYC* promoter repression. Two separate sgRNAs were used for each condition (sgNT1/sgNT2, sgMYC-prom#1/ sgMYC-prom#2). Highlighted are the genes on the *MYC-ERBB2* ecDNA that are significantly differentially expressed.

Figure 1

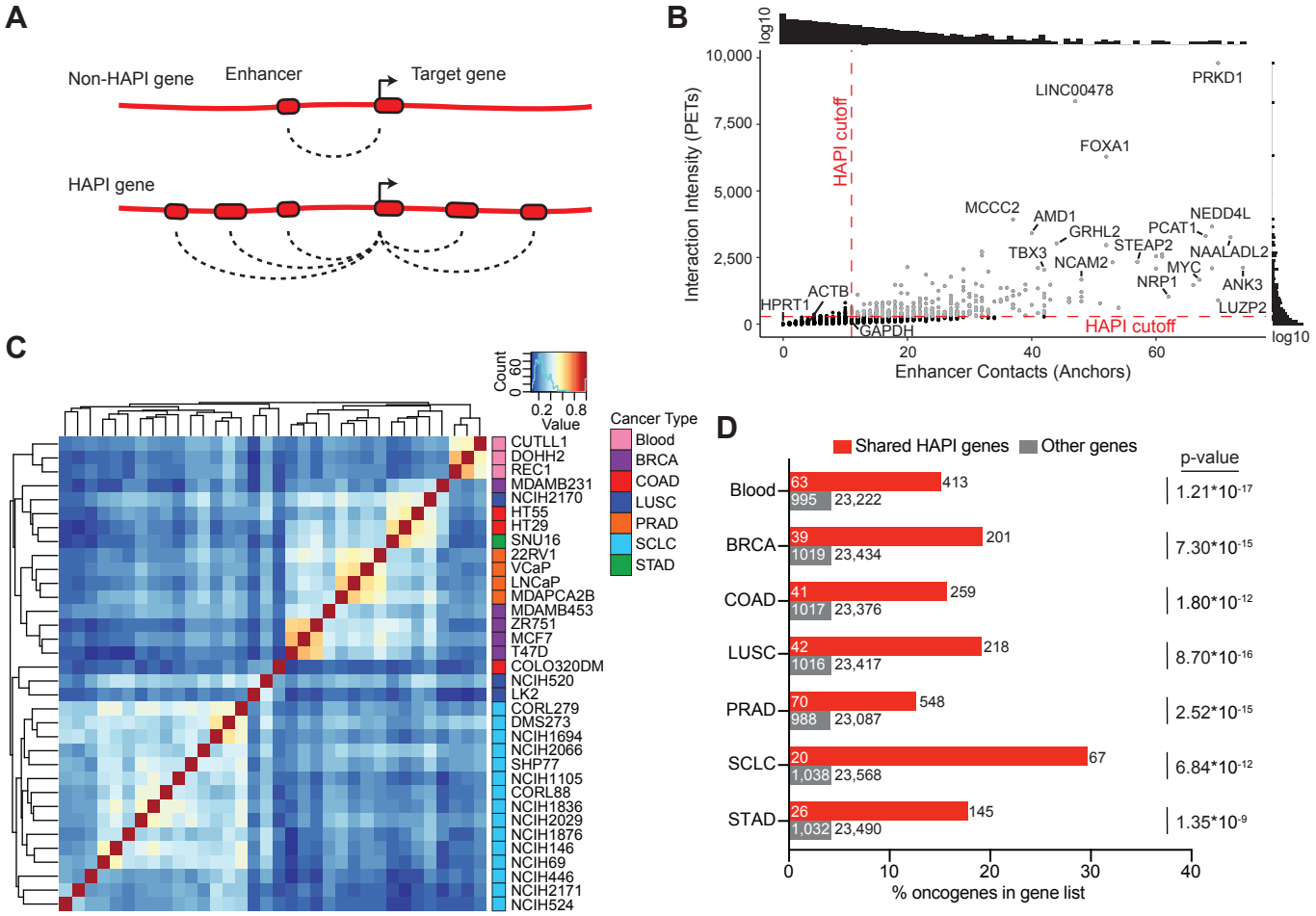

A
Non-HAPI gene
Enhancer
Target gene
HAPI gene
B
log10
Interaction Intensity (PETs)
10,000
7,500
5,000
2,500
0
HAPI cutoff
PRKD1
LINC00478
FOXA1
MCCC2
AMD1
GRHL2
PCAT1
NEDD4L
STEAP2
TBX3
NCAM2
NAALADL2
MYC
NRP1
ANK3
ACTB
HPRT1
GAPDH
LUZP2
HAPI cutoff
0
20
40
60
log10
Enhancer Contacts (Anchors)
C
Count
60
0
0.2
0.8
Value
Cancer Type
Blood
BRCA
COAD
LUSC
PRAD
SCLC
STAD
CUTLL1
DOHH2
REC1
MDAMB231
NCIH2170
HT55
HT29
SNU16
22RV1
VCaP
LNCaP
MDAPCA2B
MDAMB453
ZR751
MCF7
T47D
COLO320DM
NCIH520
LK2
CORL279
DMS273
NCIH1694
NCIH2066
SHP77
NCIH1105
CORL88
NCIH1836
NCIH2029
NCIH1876
NCIH146
NCIH69
NCIH446
NCIH2171
NCIH524
D
Shared HAPI genes
Other genes
p-value
Blood
63
995
413
23,222
1.21*10-17
BRCA
39
1019
201
23,434
7.30*10-15
COAD
41
1017
259
23,376
1.80*10-12
LUSC
42
1016
218
23,417
8.70*10-16
PRAD
70
988
548
23,087
2.52*10-15
SCLC
20
1,038
67
23,568
6.84*10-12
STAD
26
1,032
145
23,490
1.35*10-9
0
10
20
30
40
% oncogenes in gene list

Figure 2

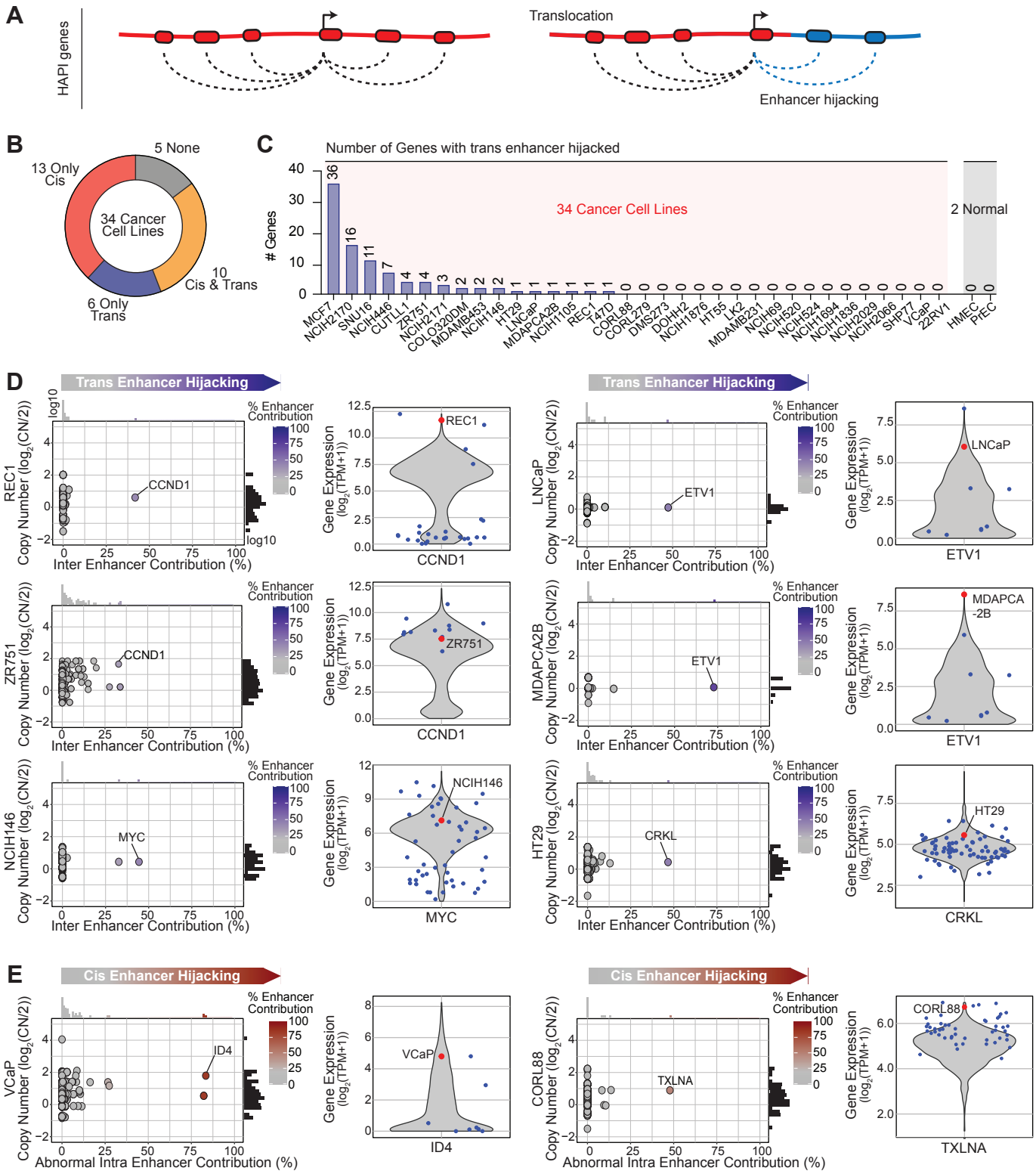

A
HAPI genes
Translocation
Enhancer hijacking
B
5 None
13 Only Cis
34 Cancer Cell Lines
10 Cis & Trans
6 Only Trans
C
Number of Genes with trans enhancer hijacked
Genes
34 Cancer Cell Lines
2 Normal
MCF7
NCIH2170
SNU16
NCIH446
CUTLL1
ZR751
NCIH2171
COLO320DM
MDAMB453
NCIH146
HT29
LNCaP
MDAPCA2B
NCIH1105
REC1
T47D
CORL88
CORL279
DMS273
DOHH2
NCIH1876
HT55
LK2
MDAMB231
NCIH69
NCIH520
NCIH524
NCIH1694
NCIH1836
NCIH2029
NCIH2066
SHP77
VCaP
22RV1
HMEC
PrEC
D
Trans Enhancer Hijacking
Trans Enhancer Hijacking
% Enhancer Contribution
Inter Enhancer Contribution (%)
Copy Number (log2(CN/2))
Gene Expression (log2(TPM+1))
REC1
CCND1
LNCaP
ETV1
ZR751
MDAPCA -2B
NCIH146
MYC
HT29
CRKL
E
Cis Enhancer Hijacking
Cis Enhancer Hijacking
Abnormal Intra Enhancer Contribution (%)
VCaP
ID4
CORL88
TXLNA

Figure 3

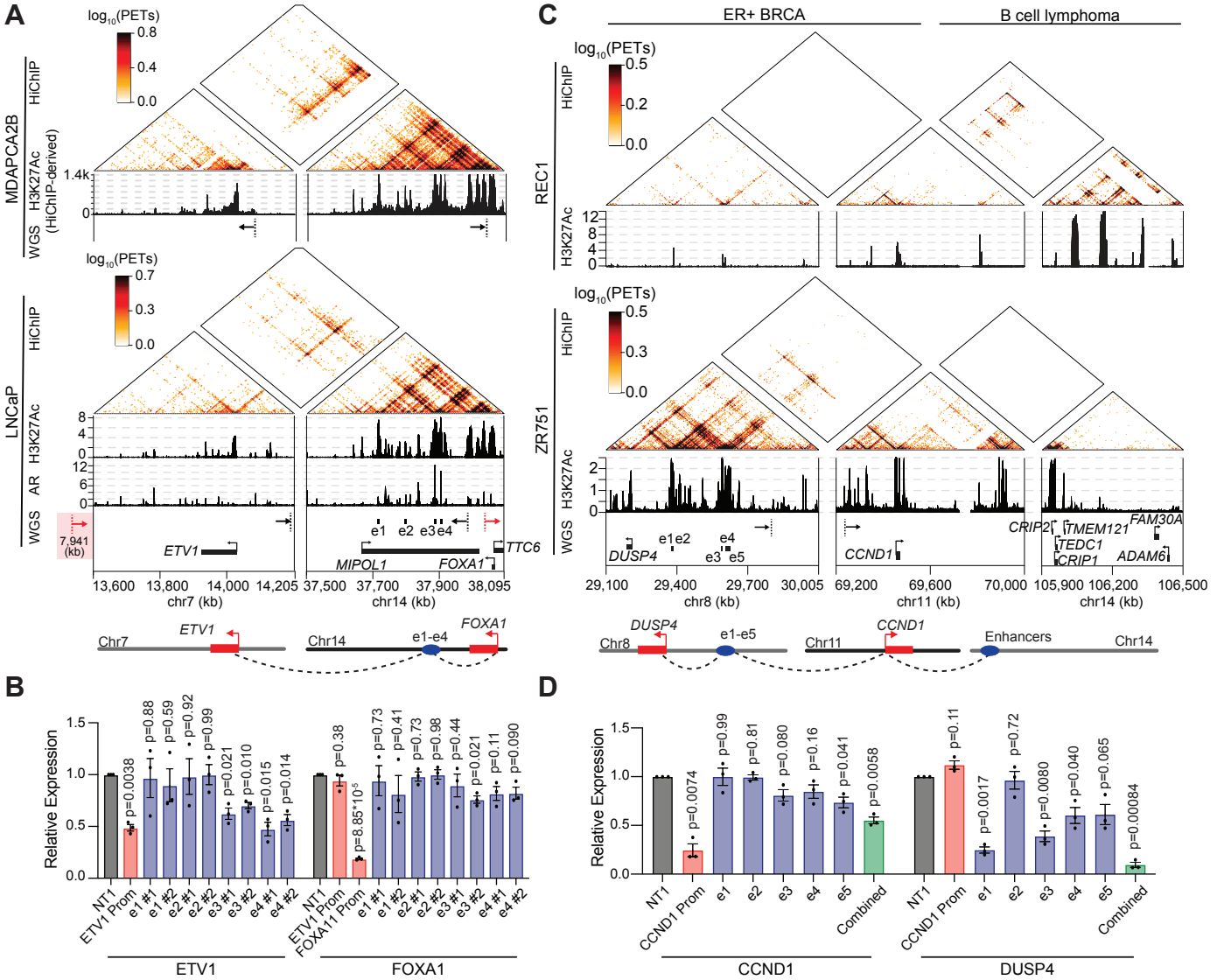

A
MDAPCA2B
HiChIP
H3K27Ac (HiChIP-derived)
WGS
log10(PETs)
LNCaP
AR
ETV1
MIPOL1
FOXA1
TTC6
e1 e2 e3 e4
7,941 (kb)
chr7 (kb)
chr14 (kb)
Chr7
Chr14
e1-e4
B
Relative Expression
NT1
ETV1 Prom
FOXA11 Prom
ETV1
FOXA1
C
ER+ BRCA
B cell lymphoma
REC1
ZR751
DUSP4
CCND1
CRIP2
TMEM121
FAM30A
TEDC1
CRIP1
ADAM6
e1e2
e3 e5
chr8 (kb)
chr11 (kb)
Chr8
Chr11
e1-e5
Enhancers
D
CCND1 Prom
Combined

Figure 4

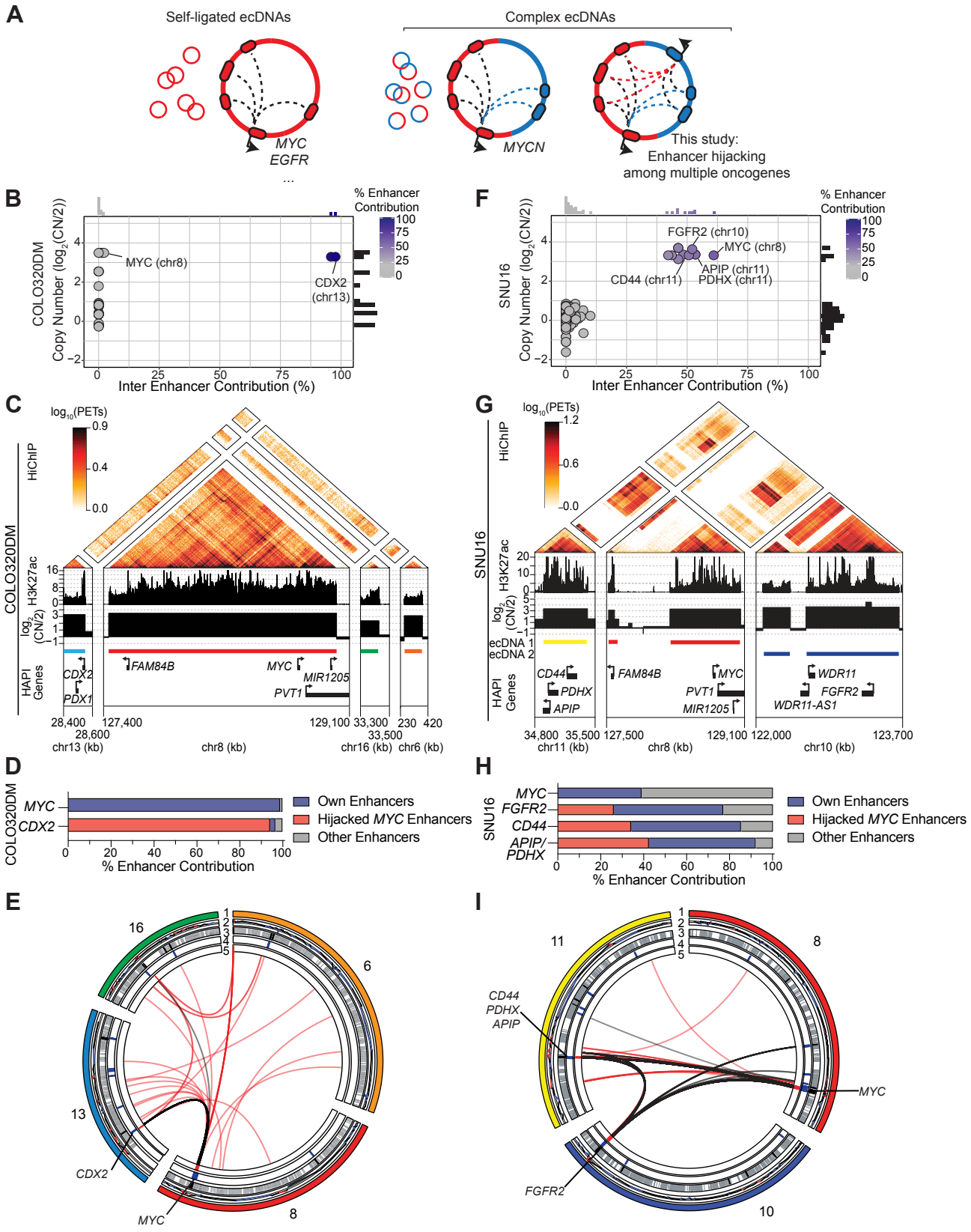

A
Self-ligated ecDNAs
Complex ecDNAs
MYC
EGFR
...
MYCN
This study:
Enhancer hijacking
among multiple oncogenes
B
COLO320DM
Copy Number (log2(CN/2))
MYC (chr8)
CDX2
(chr13)
% Enhancer
Contribution
Inter Enhancer Contribution (%)
F
SNU16
FGFR2 (chr10)
MYC (chr8)
APIP (chr11)
PDHX (chr11)
CD44 (chr11)
C
log10(PETs)
HiChIP
H3K27ac
log2
(CN/2)
HAPI
Genes
CDX2
PDX1
FAM84B
MYC
MIR1205
PVT1
chr13 (kb)
chr8 (kb)
chr16 (kb)
chr6 (kb)
G
ecDNA 1
ecDNA 2
CD44
PDHX
APIP
WDR11
FGFR2
WDR11-AS1
chr11 (kb)
chr10 (kb)
D
Own Enhancers
Hijacked MYC Enhancers
Other Enhancers
% Enhancer Contribution
H
E
I

Figure 5

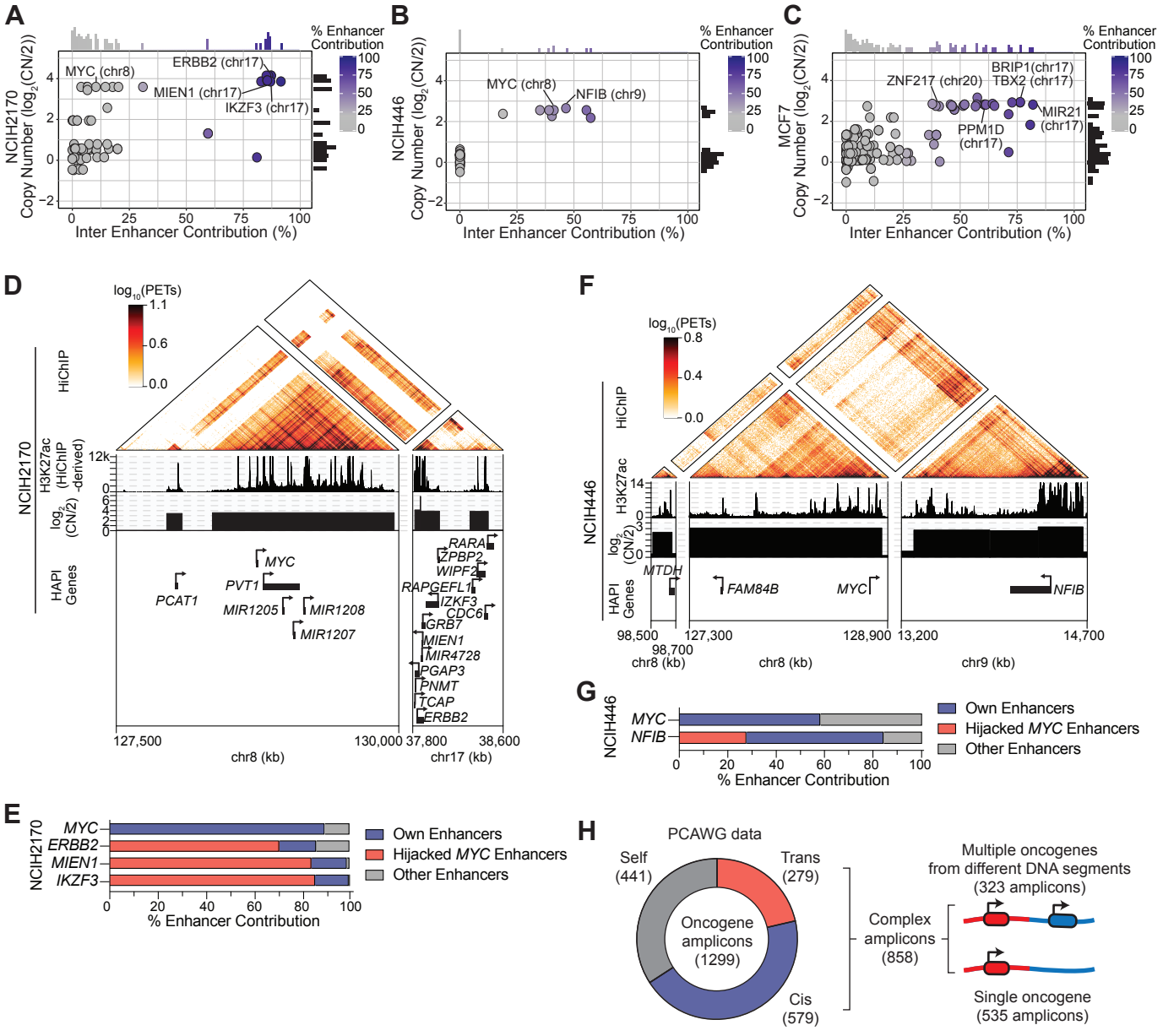
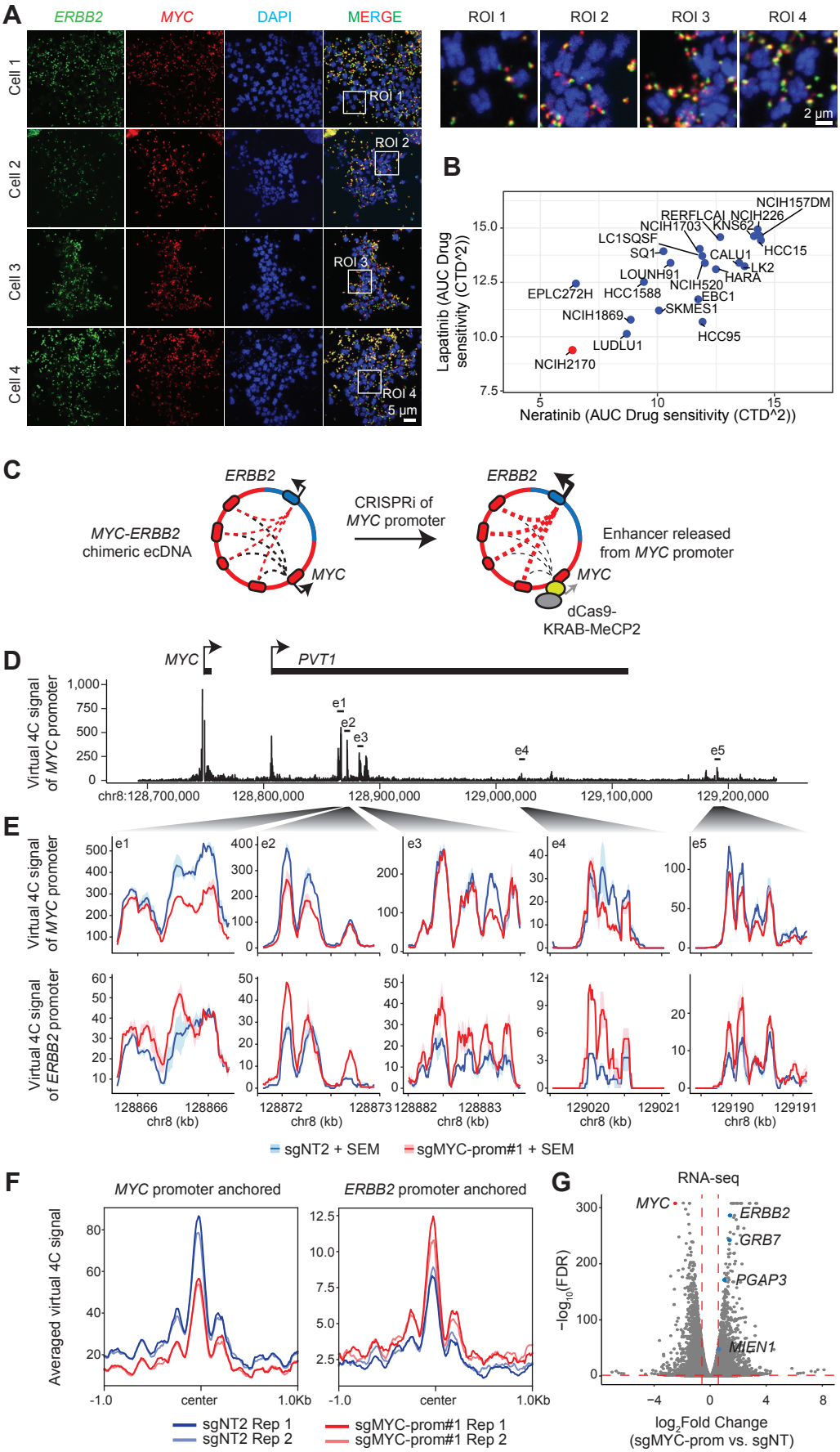

A
NCIH2170
Copy Number (log2(CN/2))
Inter Enhancer Contribution (%)
% Enhancer Contribution
MYC (chr8)
ERBB2 (chr17)
MIEN1 (chr17)
IKZF3 (chr17)
B
NCIH446
MYC (chr8)
NFIB (chr9)
C
MCF7
ZNF217 (chr20)
BRIP1(chr17)
TBX2 (chr17)
MIR21 (chr17)
PPM1D (chr17)
D
log10(PETs)
HiChIP
H3K27ac (HiChIP -derived)
log2 (CN/2)
HAPI Genes
MYC
PVT1
PCAT1
MIR1205
MIR1208
MIR1207
RARA
ZPBP2
WIPF2
RAPGEFL1
IZKF3
CDC6
GRB7
MIEN1
MIR4728
PGAP3
PNMT
TCAP
ERBB2
127,500
130,000
chr8 (kb)
37,800
38,600
chr17 (kb)
F
NCIH446
MTDH
FAM84B
MYC
NFIB
98,500
98,700
127,300
128,900
13,200
14,700
chr8 (kb)
chr9 (kb)
G
MYC
NFIB
Own Enhancers
Hijacked MYC Enhancers
Other Enhancers
E
MYC
ERBB2
MIEN1
IKZF3
H
PCAWG data
Self (441)
Trans (279)
Cis (579)
Oncogene amplicons (1299)
Complex amplicons (858)
Multiple oncogenes from different DNA segments (323 amplicons)
Single oncogene (535 amplicons)

Figure 6


A
ERBB2
MYC
DAPI
MERGE
ROI 1
ROI 2
ROI 3
ROI 4
Cell 1
Cell 2
Cell 3
Cell 4
5 µm
2 µm
B
Lapatinib (AUC Drug sensitivity (CTD^2))
Neratinib (AUC Drug sensitivity (CTD^2))
NCIH157DM
RERFLCAI
NCIH226
NCIH1703
KNS62
LC1SQSF
SQ1
CALU1
HCC15
LK2
LOUNH91
HARA
NCIH520
EPLC272H
HCC1588
EBC1
SKMES1
NCIH1869
HCC95
LUDLU1
NCIH2170
C
MYC-ERBB2 chimeric ecDNA
ERBB2
MYC
CRISPRi of MYC promoter
Enhancer released from MYC promoter
dCas9-KRAB-MeCP2
D
MYC
PVT1
Virtual 4C signal of MYC promoter
e1
e2
e3
e4
e5
chr8:128,700,000
128,800,000
128,900,000
129,000,000
129,100,000
129,200,000
E
Virtual 4C signal of ERBB2 promoter
chr8 (kb)
sgNT2 + SEM
sgMYC-prom#1 + SEM
F
MYC promoter anchored
ERBB2 promoter anchored
Averaged virtual 4C signal
-1.0
center
1.0Kb
sgNT2 Rep 1
sgNT2 Rep 2
sgMYC-prom#1 Rep 1
sgMYC-prom#1 Rep 2
G
RNA-seq
MYC
ERBB2
GRB7
PGAP3
MIEN1
−log10(FDR)
log2Fold Change (sgMYC-prom vs. sgNT)