Article

ecDNA hubs drive cooperative intermolecular oncogene expression

https://doi.org/10.1038/s41586-021-04116-8

Received: 16 November 2020

Accepted: 8 October 2021

Published online: 24 November 2021

Check for updates

King L. Hung[1,22], Kathryn E. Yost[1,22], Liangqi Xie[2,3,4,22], Quanming Shi[1], Konstantin Helmsauer[5], Jens Luebeck[6,7], Robert Schöpflin[8,9,10], Joshua T. Lange[11,12,13], Rocío Chamorro González[5], Natasha E. Weiser[1,13], Celine Chen[5], Maria E. Valieva[8,9], Ivy Tsz-Lo Wong[12,13], Sihan Wu[14], Siavash R. Dehkordi[7], Connor V. Duffy[1], Katerina Kraft[1], Jun Tang[12,13], Julia A. Belk[13,15], John C. Rose[1], M. Ryan Corces[1], Jeffrey M. Granja[1], Rui Li[1], Utkrisht Rajkumar[7], Jordan Friedlein[16], Anindya Bagchi[16], Ansuman T. Satpathy[13], Robert Tjian[3,4], Stefan Mundlos[8,9,17], Vineet Bafna[7], Anton G. Henssen[5,18,19,20], Paul S. Mischel[12,13], Zhe Liu[2] & Howard Y. Chang[1,21]✉

Extrachromosomal DNA (ecDNA) is prevalent in human cancers and mediates high expression of oncogenes through gene amplification and altered gene regulation[1]. Gene induction typically involves *cis*-regulatory elements that contact and activate genes on the same chromosome[2,3]. Here we show that ecDNA hubs—clusters of around 10–100 ecDNAs within the nucleus—enable intermolecular enhancer–gene interactions to promote oncogene overexpression. ecDNAs that encode multiple distinct oncogenes form hubs in diverse cancer cell types and primary tumours. Each ecDNA is more likely to transcribe the oncogene when spatially clustered with additional ecDNAs. ecDNA hubs are tethered by the bromodomain and extraterminal domain (BET) protein BRD4 in a *MYC*-amplified colorectal cancer cell line. The BET inhibitor JQ1 disperses ecDNA hubs and preferentially inhibits ecDNA-derived-oncogene transcription. The BRD4-bound *PVT1* promoter is ectopically fused to *MYC* and duplicated in ecDNA, receiving promiscuous enhancer input to drive potent expression of *MYC*. Furthermore, the *PVT1* promoter on an exogenous episome suffices to mediate gene activation in *trans* by ecDNA hubs in a JQ1-sensitive manner. Systematic silencing of ecDNA enhancers by CRISPR interference reveals intermolecular enhancer–gene activation among multiple oncogene loci that are amplified on distinct ecDNAs. Thus, protein-tethered ecDNA hubs enable intermolecular transcriptional regulation and may serve as units of oncogene function and cooperative evolution and as potential targets for cancer therapy.

Circular ecDNA that encodes oncogenes is a prevalent feature of cancer genomes and a potent driver of cancer progression[4–8]. ecDNAs (including double minutes) are covalently closed, double-stranded and range from around 100 kilobases to several megabases in size[1,9–12]. Lacking centromeres, ecDNAs are randomly segregated into daughter cells during cell division, which enables the rapid accumulation and selection of ecDNA variants that confer a fitness advantage[5,13–15]. ecDNAs can reintegrate into chromosomes[16–20] and may therefore also act as precursors to some chromosomal amplifications. ecDNAs possess highly accessible chromatin[1,21] and co-amplify enhancer elements[22,23], suggesting that oncogene amplicons may be shaped by regulatory dependencies to amplify transcription. ecDNAs cluster with one another during cell division or after DNA damage[24–26]; but the biological consequences of ecDNA clustering are poorly understood.

[1]Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA. [2]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. [3]Department of Molecular and Cell Biology, Li Ka Shing Center for Biomedical and Health Sciences, CIRM Center of Excellence, University of California, Berkeley, Berkeley, CA, USA. [4]Howard Hughes Medical Institute, Berkeley, CA, USA. [5]Department of Pediatric Oncology and Hematology, Charité—Universitätsmedizin Berlin, Berlin, Germany. [6]Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA, USA. [7]Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA, USA. [8]Development and Disease Research Group, Max Planck Institute for Molecular Genetics, Berlin, Germany. [9]Institute for Medical and Human Genetics, Charité—Universitätsmedizin Berlin, Berlin, Germany. [10]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany. [11]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA. [12]ChEM-H, Stanford University, Stanford, CA, USA. [13]Department of Pathology, Stanford University, Stanford, CA, USA. [14]Children's Medical Center Research Institute, University of Texas Southwestern Medical Center, Dallas, TX, USA. [15]Department of Computer Science, Stanford University, Stanford, CA, USA. [16]Tumor Initiation and Maintenance Program, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA. [17]Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité—Universitätsmedizin Berlin, Berlin, Germany. [18]Experimental and Clinical Research Center (ECRC), Max Delbrück Center for Molecular Medicine and Charité—Universitätsmedizin Berlin, Berlin, Germany. [19]German Cancer Consortium (DKTK), partner site Berlin, and German Cancer Research Center DKFZ, Heidelberg, Germany. [20]Berlin Institute of Health, Berlin, Germany. [21]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. [22]These authors contributed equally: King L. Hung, Kathryn E. Yost, Liangqi Xie. ✉e-mail: howchang@stanford.edu

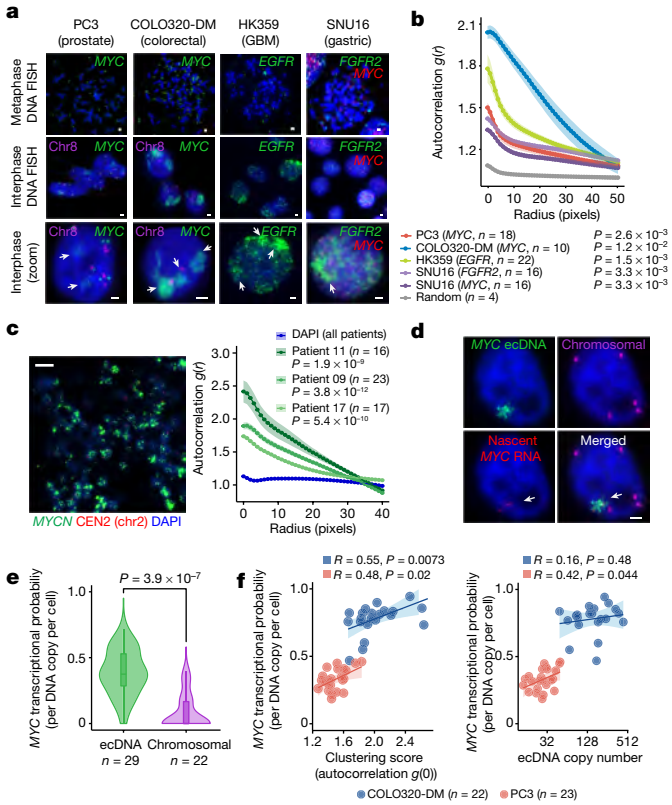


Fig. 1 | ecDNA imaging correlates ecDNA clustering with transcriptional bursting. a, Representative FISH images of interphase ecDNA clustering. A chromosomal control was included for PC3 and COLO320-DM. GBM, glioblastoma. Scale bars, 2 µm. **b**, Interphase ecDNA clustering by autocorrelation $g(r)$ (Methods). Data are mean ± s.e.m. *P* values determined by two-sided Wilcoxon test at $r = 0$ as compared to random distribution. **c**, Left, representative FISH image showing ecDNA clustering in a primary neuroblastoma tumour from patient 09 (*MYCN* ecDNA and chromosomal control). CEN2, chr2 chromosome enumeration probe. Scale bar, 10 µm. Right, ecDNA clustering in three primary tumours using autocorrelation. Data are mean ± s.e.m. *P* values determined by two-sided Wilcoxon test at $r = 0$ as compared to DAPI. **d**, Representative image from combined DNA FISH for ecDNA, chromosomal control and nascent RNA FISH in PC3 cells. Scale bar, 2 µm. **e**, *MYC* transcription probability measured by joint DNA and RNA FISH (RNA normalized to DNA copy number; box centre line, median; box limits, upper and lower quartiles; box whiskers, 1.5× interquartile range). *P* values determined by two-sided Wilcoxon test. **f**, Correlation between *MYC* transcription probability and ecDNA copy number or clustering (joint DNA and RNA FISH; clustering scores are autocorrelation at $r = 0$; Pearson's *R*, two-sided test).

ecDNA hubs amplify oncogene expression

We visualized ecDNA localization in interphase nuclei by DNA fluorescence in situ hybridization (FISH)[27] using probes that target ecDNA-amplified oncogenes in multiple cell lines, including PC3 (*MYC*-amplified), COLO320-DM (*MYC*-amplified), HK359 (*EGFR*-amplified) and SNU16 (*MYC*- and *FGFR2*-amplified)[1] (Fig. 1a, Extended Data Fig. 1a). DNA FISH on metaphase spreads revealed tens to hundreds of individual ecDNAs per cell located outside chromosomes (Fig. 1a, Methods). In a subset of cell lines, we used two-colour DNA FISH to interrogate a non-ecDNA neighbouring control locus (Extended Data Fig. 1a); chromosomal oncogene copies appear as paired dots whereas ecDNAs have a single colour, as expected (Fig. 1a, Extended Data Fig. 1b). In all of the ecDNA-positive cancer cells that we assessed, the ecDNA FISH signal was locally concentrated in interphase nuclei despite arising from tens to hundreds of individual ecDNA molecules, suggesting that ecDNAs strongly cluster with one another—a feature we term ecDNA hubs (Fig. 1a). ecDNA hubs occupied a much larger space than chromosomal signals and are larger than diffraction-limited spots (around 0.3 µm), suggesting that they consist of many clustered ecDNA molecules. Quantification using an autocorrelation function $g(r)$ (Methods) showed a significant increase in clustering over short distances (0–40 pixels, 0–1.95 µm; Fig. 1b, Extended Data Fig. 1c) compared to random distribution. In three primary neuroblastoma tumours with *MYCN* amplifications, we also observed ecDNA hubs in the vast majority of cancer cells[28] (Fig. 1c, Extended Data Fig. 1d, e). These results suggest that ecDNA clustering occurs across various cancer types with different oncogene amplifications and in primary tumours.

Next, we visualized actively transcribing *MYC* alleles by joint DNA and nascent RNA FISH in PC3 and COLO320-DM cells (Fig. 1d, Extended Data Fig. 1a, f–h) and computed the probability of *MYC* transcription from each ecDNA molecule (Methods). The majority of nascent *MYC* mRNA transcripts came from ecDNA hubs rather than from the chromosomal locus, even after accounting for copy number (Fig. 1d, e). ecDNA clustering was significantly correlated with increased *MYC* transcription, and ecDNA clustering was a better predictor of *MYC* transcription probability as compared to copy number (Fig. 1f). Furthermore, ecDNAs in hubs are more transcriptionally active compared to singleton ecDNAs (Extended Data Fig. 1i). Thus, each ecDNA molecule is more likely to transcribe the oncogene when more ecDNAs are present in hubs.

BRD4 links ecDNA hubs and transcription

MYC is flanked by super-enhancers marked by histone H3 acetylation at lysine 27 (H3K27ac) and BET proteins such as BRD4[29,30]. *MYC* transcription is highly sensitive to BET protein displacement by the inhibitor JQ1[31,32]. To examine *MYC* ecDNAs in live cells, we inserted a Tet-operator (TetO) array into *MYC* ecDNAs in COLO320-DM cells and labelled ecDNAs with TetR-eGFP or TetR-eGFP(A206K) to minimize GFP dimerization (Extended Data Fig. 2a–d, Methods). Live-cell imaging revealed multiple dynamic nuclear foci corresponding to clustered ecDNAs (Extended Data Fig. 2e–i, Supplementary Video 1). Epitope tagging of endogenous BRD4 revealed that BRD4 is highly enriched in TetO-labelled ecDNA hubs (Fig. 2a, Extended Data Fig. 2j–l). Assay of transposase-accessible chromatin by sequencing (ATAC–seq) and chromatin immunoprecipitation and sequencing (ChIP–seq) of H3K27ac and BRD4 showed that H3K27ac peaks, which mark active ecDNA enhancers, are indeed also occupied by BRD4 (Fig. 2b, Extended Data Fig. 3a–c).

To determine the role of BET proteins in ecDNA-derived transcription, we focused on the isogenic colorectal cancer cell lines COLO320-DM (*MYC* ecDNA) and COLO320-HSR (chromosomal *MYC* amplicon or homogeneously staining region; HSR)[18], which were derived from the same patient tumour (Extended Data Fig. 3a). Treatment with 500 nM JQ1 dispersed ecDNA hubs in COLO320-DM cells after 6 h, splitting large ecDNA hubs into multiple small ecDNA signals including singleton ecDNAs and abolishing the most-clustered ecDNA hubs (autocorrelation $g(r) \geq 2$) (Fig. 2c, d, Extended Data Fig. 3d–f). Treatment with JQ1 did not alter the spatial distribution of covalently linked *MYC* copies in COLO320-HSR cells as expected (Fig. 2c, d). ecDNA dispersal by JQ1 appears to be highly specific; inhibition of transcription by either the RNA polymerase II inhibitor α-amanitin or 1,6-hexanediol[33] did not affect ecDNA hubs (Extended Data Fig. 3g–j).

JQ1 potently inhibited ecDNA-derived oncogene transcription. Treatment with JQ1 reduced the *MYC* transcription probability per ecDNA copy by fourfold, as shown by joint nascent RNA and DNA FISH (Fig. 2e, Extended Data Fig. 3g). Because BET proteins are also involved in *MYC* transcription from chromosomal DNA, we compared the effect of JQ1 on COLO320-DM versus COLO320-HSR cells. BRD4 ChIP–seq showed that treatment with JQ1 equivalently dislodged BRD4 genome-wide in these isogenic cells (Extended Data Fig. 3k).

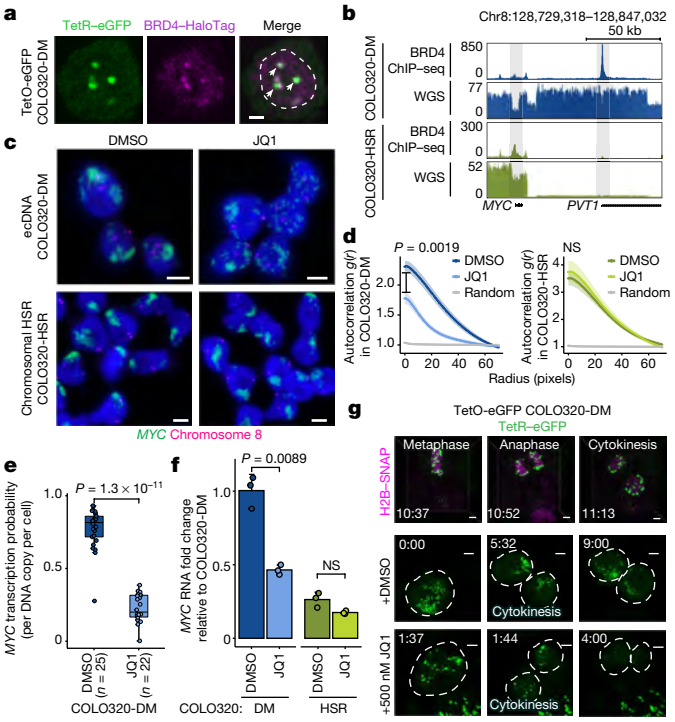


Fig. 2 | BET proteins mediate ecDNA hub formation and transcription. **a**, Representative live-cell images of ecDNA and BRD4–HaloTag signals in TetO–eGFP COLO320-DM cells (independently repeated twice; dashed line indicates nuclear boundary). Scale bar, 2 µm. **b**, BRD4 ChIP–seq and WGS at the *MYC* locus in COLO320-DM and COLO320-HSR cells. **c**, Representative DNA FISH images for cells treated with dimethyl sulfoxide (DMSO) or 500 nM JQ1 for 6 h. Scale bars, 5 µm. **d**, Clustering measured by autocorrelation $g(r)$ for ecDNAs in COLO320-DM cells and HSRs in COLO320-HSR cells treated with DMSO or 500 nM JQ1 for 6 h. Data are mean ± s.e.m. *P* values determined by two-sided Wilcoxon test at $r = 0$. NS, not significant. COLO320-DM: $n = 18$ (DMSO, JQ1) and $n = 10$ (random); COLO320-HSR: $n = 10$ (all groups). **e**, *MYC* transcription probability in COLO320-DM cells treated with DMSO or 500 nM JQ1 for 6 h (joint DNA and RNA FISH; RNA normalized to ecDNA copy number; box plot parameters as in Fig. 1). *P* values determined by two-sided Wilcoxon test. **f**, *MYC* RNA measured by reverse transcription–quantitative PCR (RT–qPCR) for COLO320-DM and COLO320-HSR cells treated with either DMSO or 500 nM JQ1 for 6 h. Data are mean ± s.d. between three biological replicates. *P* values determined by two-sided Student's *t*-test. **g**, Representative live-cell images of TetR-eGFP-labelled ecDNAs in TetO–eGFP COLO320-DM cells treated with DMSO or 500 nM JQ1 at the indicated time points through cell division (independently repeated twice for each condition). H2B–SNAP (top) labels histone H2B in mitotic chromosomes. ecDNAs appear to be tethered to chromosomes. Scale bars, 3 µm (top and bottom rows); 5 µm (middle row).

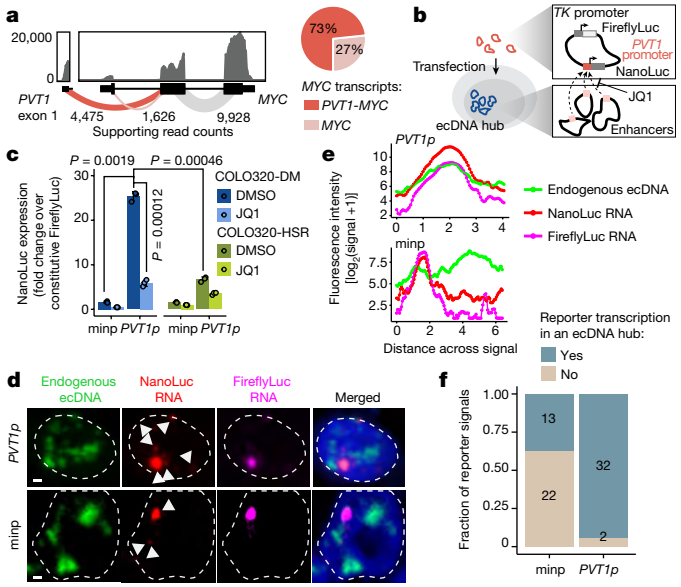


Fig. 3 | Intermolecular activation of an episomal luciferase reporter in ecDNA hubs. **a**, Left, bulk RNA-seq from COLO320-DM cells with exon–exon junction spanning read counts shown. Right, relative abundance of full-length *MYC* and fusion *PVT1-MYC* transcripts using read count supporting either junction. **b**, *PVT1*-promoter-driven luciferase reporter system. **c**, Luciferase reporter activity driven by either a minimal promoter (minp) or the *PVT1* promoter (*PVT1p*) with DMSO or JQ1 treatment (500 nM, 6 h). Data are mean ± s.d. between 3 biological replicates. *P* values determined by two-sided Student's *t*-test (Bonferroni adjusted). **d**, Representative images of *PVT1p* or minp reporter transcriptional activity and endogenous ecDNA hubs in COLO320-DM cells visualized by DNA and RNA FISH (independently repeated 3 times). White arrowheads highlight NanoLuc RNA foci. Scale bars, 1 µm. **e**, Fluorescence intensities on a line drawn across the centre of the largest NanoLuc RNA signal in images in **d**. **f**, Number of nuclear NanoLuc signals that colocalize with ecDNA hubs.

Nonetheless, treatment with 500 nM JQ1 preferentially lowered the level of *MYC* mRNA in COLO320-DM cells, a dose that had no significant effect on the level of *MYC* mRNA in COLO320-HSR cells (Fig. 2f). JQ1 dose titration showed that there was a modest preferential killing of COLO320-DM cells over COLO320-HSR cells (Extended Data Fig. 3l–n). A survey of six additional compounds that target transcription or histone modifications found that only BET inhibitors selectively inhibited *MYC* expression in ecDNA$^+$ cells, and that MS645—a bivalent BET bromodomain inhibitor[34]—reduced ecDNA transcription and clustering similarly to JQ1 (Extended Data Fig. 3o–q). Live-cell imaging with TetO-GFP COLO320-DM cells showed that ecDNA hubs resolve into smaller particles during mitosis (Fig. 2g, Supplementary Videos 1, 2). After partitioning, ecDNAs re-form large hubs; notably, ecDNA hub assembly after mitosis is blocked by JQ1 (Fig. 2g, Supplementary Video 3). Together, these results suggest a unique dependence on the bromodomain–H3K27ac interaction of BET proteins for ecDNA hub formation, maintenance and oncogene transcription in COLO320-DM cells.

PVT1-MYC hijacks ecDNA enhancer input

To link ecDNA structure to the regulation of *MYC* transcription, we reconstructed the COLO320-DM ecDNA using five orthogonal approaches and report what is—to our knowledge—the largest ecDNA structure that has so far been assembled. We identified complex structural rearrangements using (1) whole-genome sequencing (WGS)[35]; (2) nanopore-based single-molecule sequencing; and (3) large DNA contig assembly by optical mapping[36] (Extended Data Fig. 4a–d). In addition, we performed targeted ecDNA digestion using CRISPR–Cas9 followed by pulsed-field gel electrophoresis (PFGE) and deep sequencing of megabase-sized DNA fragments to obtain sequence multiplicity information that was highly concordant with optical mapping ecDNA contigs (Extended Data Fig. 4e, f). Using these first four methods, we reconstructed a 4.328-Mb ecDNA that contains multiple copies of a *PVT1-MYC* fusion[37,38], a canonical *MYC* sequence, and sequences from multiple chromosomal origins (chromosomes 6, 8, 13 and 16) (Extended Data Fig. 4e). Finally, we used DNA FISH to confirm the colocalization of *PLUT*, *PCAT1* and *MYC* genes on ecDNAs as predicted by the reconstruction (Extended Data Fig. 4g).

The *PVT1-MYC* fusion makes up more than 70% of *MYC* transcripts in COLO320-DM cells and consists of the promoter and exon 1 of the long non-coding RNA gene *PVT1* fused to exons 2 and 3 of *MYC* (which encode a functional MYC protein isoform[39]), replacing the promoter and exon 1 of *MYC* (Fig. 3a). Consistently, total *MYC* RNA transcripts were reduced

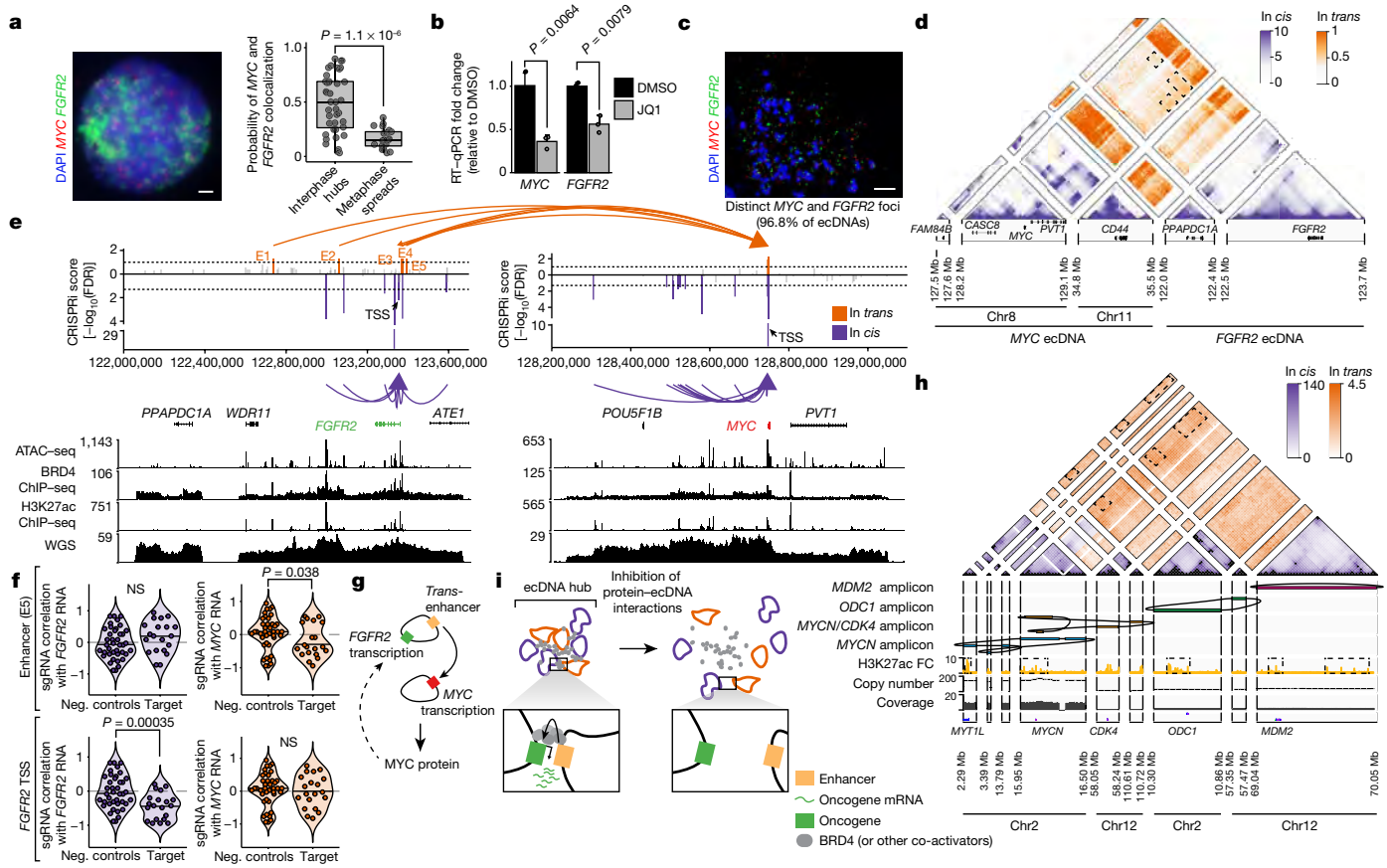


Fig. 4 | ecDNA hubs mediate intermolecular enhancer–gene interactions. **a**, Left, representative DNA FISH image showing clustering of *MYC* and *FGFR2* ecDNAs in interphase SNU16 cells. Scale bar, 2 µm. Right, *MYC* and *FGFR2* colocalization in SNU16 cells (box plot parameters as in Fig. 1). *P* value determined by two-sided Wilcoxon test. **b**, Oncogene RNA measured by RT–qPCR in SNU16 cells treated with DMSO or 500 nM JQ1 for 6 h. Data are mean ± s.d. between 3 biological replicates. *P* values determined by two-sided Student's *t*-test. **c**, Representative metaphase FISH image in SNU16-dCas9-KRAB cells. Quantification summarizes 29 cells from one experiment. Scale bar, 10 µm. **d**, H3K27ac HiChIP contact matrix (10-kb resolution; KR-normalized read counts) in SNU16-dCas9-KRAB cells showing *cis* (purple) and *trans* (orange) interactions. **e**, Top, significance of enhancer CRISPRi effects on oncogene repression (Benjamini–Hochberg-adjusted; $n = 40$ negative control sgRNAs, $n = 20$ target sgRNAs; Methods, Extended Data Fig. 8). Dashed lines mark a false discovery rate (FDR) < 0.05 for *cis* interactions and FDR < 0.1 for *trans* interactions; significant enhancers are coloured and connected to target genes by loops (E1, FDR = 0.048; E2, FDR = 0.052; E3, FDR = 0.048; E4, FDR = 0.052; E5, FDR = 0.052). All datasets contain two independent experiments except the in-*trans* dataset for the *MYC*-targeting sgRNA pool, which contains one independent experiment. Bottom, ATAC–seq, BRD4 ChIP–seq, H3K27ac ChIP–seq and WGS tracks. **f**, Correlations between individual sgRNAs and oncogene expression (Methods). *P* values determined by lower-tailed *t*-test as compared to negative controls. Each dot represents an independent sgRNA ($n = 40$ negative control sgRNAs, $n = 20$ target sgRNAs). TSS, transcription start site. **g**, Cross-regulation between *MYC* and *FGFR2* elements in ecDNA hubs. **h**, Top to bottom: Hi-C contact map (KR-normalized read counts in 25-kb bins) showing *cis* and *trans* contacts, reconstructed amplicons, H3K27ac ChIP–seq (mean fold change (FC) over input), copy number and WGS in TR14 cells. **i**, ecDNA hub model for intermolecular cooperation.

by CRISPR interference (CRISPRi) of the *PVT1* promoter (Extended Data Fig. 4h). Multiple *PVT1-MYC* fusion copies share a common breakpoint, indicative of a common origin (Extended Data Fig. 4i). We observed strong BRD4 binding at the *PVT1* promoter in COLO320-DM cells, but not in COLO320-HSR cells (Fig. 2b). As the *PVT1* promoter can be activated by MYC[40], we hypothesize that *PVT1-MYC* fusion enables positive feedback of *MYC* expression and circumvents competition between the *PVT1* and *MYC* promoters, which is normally observed on the unrearranged chromosome[41]. Notably, *PVT1* rearrangement and gene fusion are observed in several human cancers and drive gene overexpression[42].

We next identified ecDNA regulatory elements that are associated with high oncogene expression. Paired single-cell ATAC–seq and RNA sequencing (RNA-seq) from 72,049 COLO320-DM and COLO320-HSR cells identified 47 ecDNA regulatory elements associated with high *MYC* expression after correcting for *MYC* copy number (Extended Data Fig. 5, Methods). Enhancer connectome analysis using H3K27ac HiChIP, a protein-directed 3D genome conformation assay[43], revealed that multiple enhancers make significant contact with the *PVT1* or *PVT1-MYC* promoter (Extended Data Figs. 6a, b, 5f, g). Whereas the canonical *MYC* promoter participates in several focal enhancer contacts, the HiChIP signal at the *PVT1* promoter is increased across the entire amplified region (Extended Data Fig. 6a). CRISPRi targeting of six enhancers individually with high BRD4 occupancy on ecDNA did not significantly reduce bulk *MYC* mRNA levels (Extended Data Fig. 4i), probably owing to combinatorial and compensatory enhancer–gene interactions. These results indicate that the *PVT1* promoter, now driving *MYC* oncogene expression on ecDNA, receives broad and combinatorial enhancer input within ecDNA hubs.

Gene activation in *trans* in ecDNA hubs

We next interrogated whether ecDNA molecules cooperate in spatial proximity to achieve gene transcription. We constructed a plasmid containing the 2-kb *PVT1* promoter driving NanoLuc luciferase (PVT1p-nLuc) and with a constitutive thymidine kinase promoter (*TKp*) driving Firefly luciferase as an internal control (Fig. 3b). In COLO320-DM

cells, *PVT1p* was highly active (around 25-fold) compared to *TKp* or a minimal promoter (minp-nLuc; Fig. 3c). Of note, *PVT1p* conferred significantly greater (around fourfold) induction in ecDNA$^+$ COLO320-DM cells than in isogenic ecDNA$^-$ COLO320-HSR cells (Fig. 3c), whereas minimal promoter and *MYC* promoter activity was comparable between the isogenic cell lines (Extended Data Fig. 6c). Treatment with JQ1 at a low dose that disperses ecDNA hubs strongly reduced *PVT1p*-mediated transcription in COLO320-DM cells (around fivefold repression) but had a more modest effect in COLO320-HSR cells (around twofold) (Fig. 3c). Joint DNA FISH and nascent RNA FISH showed that *PVT1p* conferred increased NanoLuc transcription when colocalized with ecDNA hubs compared to the minimal promoter (Fig. 3d–f, Extended Data Fig. 6d). Addition of a *cis*-enhancer to the plasmid increased both *PVT1p*- or *MYCp*-driven NanoLuc activity and *TKp*-driven Firefly luciferase activity (Extended Data Fig. 6e, f). Finally, *MYCp* or incorporation of a *cis*-enhancer to the plasmid reduced the difference in reporter sensitivity to JQ1 in COLO320-DM versus COLO320-HSR cells (Extended Data Fig. 6g). Together, these experiments suggest intermolecular enhancer–promoter activation in ecDNA hubs and identify *PVT1p* as a DNA element that is capable of activation in ecDNA hubs in *trans*.

Intermolecular regulation among ecDNAs

We next investigated whether intermolecular enhancer–gene interactions can be precisely mapped and perturbed. We focused on a human gastric cancer cell line, SNU16, which contains two distinct ecDNA types: a *MYC* amplicon derived from chromosomes 8 and 11 and an *FGFR2* amplicon derived from chromosome 10. These ecDNAs intermingle in hubs, as demonstrated by two-colour interphase FISH (Figs. 1a, b, 4a). Treatment with JQ1 reduced the ecDNA-derived transcription of both *MYC* and *FGFR2* (Fig. 4b). We generated a subclone, SNU16-dCas9-KRAB, with stable expression of dCas9-KRAB and reduced ecDNA structural heterogeneity as confirmed by metaphase FISH (96.8% distinct *MYC* and *FGFR2* ecDNAs), WGS and H3K27ac HiChIP analyses (Fig. 4c, Extended Data Fig. 7a–c). H3K27ac HiChIP showed intermolecular contacts between *FGFR2* and *MYC* ecDNAs with a lower contact frequency relative to *cis* interactions but enriched for focal interactions (Fig. 4d). CRISPRi targeting of candidate regulatory elements (20 guides per element; 2,747 guides in total[44]; Extended Data Fig. 8a–c, Methods) identified functional elements linked to the expression of *MYC* or *FGFR2* both in *cis* (oncogene located on the same ecDNA) and in *trans* (oncogene located on a distinct ecDNA) (Methods, Fig. 4e, f, Extended Data Fig. 8d). As a positive control, CRISPRi of the *MYC* and *FGFR2* promoters strongly reduced corresponding gene expression. CRISPRi of the *FGFR2* promoter had no effect on *MYC* expression, indicating that downregulation of FGFR2 protein does not affect *MYC* expression (Fig. 4e, f). Notably, we identified five enhancers on the *FGFR2* ecDNA that activate *MYC* in *trans*, but no *MYC* ecDNA enhancers that activate *FGFR2* (Fig. 4e, f, Extended Data Fig. 8e). Perturbations of in-*trans* interactions resulted in similar significance levels to perturbations of several in-*cis* interactions on the *MYC* ecDNA (Fig. 4e). We validated that *FGFR2 trans*-enhancers are not covalently linked to the *MYC* gene on 98–100% of ecDNA molecules by dual-colour metaphase DNA FISH and in vitro CRISPR–Cas9 digestion (Extended Data Fig. 9). CRISPRi of the *MYC* promoter reduced the expression of both *MYC* and *FGFR2*, suggesting that the MYC protein may act as a transcriptional activator of *FGFR2*[45] (Fig. 4e, g, Extended Data Fig. 8f). These data suggest that *FGFR2* and *MYC* ecDNAs have been co-selected so that enhancers on both amplicons cooperatively activate *MYC* expression. The MYC protein then, in turn, activates *FGFR2* expression (Fig. 4g). There is little overlap between *cis*- and *trans*-regulatory elements, supporting our conclusion that intermolecular enhancer elements directly modify gene expression in *trans* rather than through downstream effects.

Finally, to assess intermolecular ecDNA interactions in an independent cancer type, we used nanopore sequencing and WGS to identify four distinct oncogene amplicons in TR14, a neuroblastoma cell line, which also contains ecDNA hubs (Extended Data Fig. 10a, b). Hi-C analysis revealed *trans* interactions, such as those between the *MYCN* and *ODC1* amplicons, which are not brought together by structural variants (Fig. 4h, Extended Data Fig. 10c–e). *Trans* Hi-C contacts are enriched at sites marked by H3K27ac, which may represent regulatory elements that enable intermolecular cooperation (Fig. 4h, Extended Data Fig. 10f–h). Together, these results suggest that intermolecular enhancer-gene activation in ecDNA hubs occurs for diverse oncogene loci and multiple cancer types.

Discussion

Local congregation of ecDNA in ecDNA hubs promotes novel intermolecular enhancer–gene interactions and oncogene overexpression (Fig. 4i). Unlike chromosomal transcription hubs, which favour local *cis*-regulatory elements and span 100–300 nm[46], ecDNA hubs can span more than 1,000 nm and involve *trans* regulatory elements located on distinct ecDNA molecules. This discovery has profound implications with regard to how ecDNAs undergo selection and how the rewiring of oncogene regulation on ecDNA contributes to transcription. First, *trans*-activation between ecDNAs suggests that oncogene–enhancer co-selection may occur both on individual ecDNAs and on the repertoire of ecDNAs in a cell. Thus, individual ecDNA molecules may not be required to contain all necessary regulatory elements as a diverse repertoire of regulatory elements is accessible in a hub[47]. This type of evolutionary dynamics has been documented in viruses, in which the cooperation of a mixture of specialized variants outperforms a pure wild-type population[48,49]. Moreover, mutations on individual molecules may be better tolerated, which may increase ecDNA sequence diversity. Finally, ecDNA hubs promote variable enhancer usage as cluster ecDNA molecules can 'sample' various enhancers through new enhancer–promoter interactions, including ectopic enhancer–promoter interactions between ecDNAs that arise from distinct chromosomes, as in SNU16 cells.

The recognition that ecDNA hubs promote oncogene transcription may provide therapeutic opportunities. Whereas chromosomal DNA amplicons such as HSRs are covalently linked, ecDNA hubs are held together by proteins. In COLO320-DM cells, we show that inhibition of BET proteins by JQ1 disaggregates ecDNA hubs and reduces ecDNA-derived *MYC* expression. Although *MYC* and *MYCN* are regulated by BET proteins[31,50], other ecDNA oncogene amplifications may exploit their endogenous enhancer mechanisms in ecDNA hubs and may rely on other gene-specific protein factors. Future studies may identify proteins that mediate ecDNA transcriptional activity in various cancer types and will be highly informative for potential therapeutic efforts.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-04116-8.

1. Wu, S. et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699–703 (2019).
2. Gorkin, D. U., Leung, D. & Ren, B. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* **14**, 762–775 (2014).
3. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019).
4. Bailey, C., Shoura, M. J., Mischel, P. S. & Swanton, C. Extrachromosomal DNA—relieving heredity constraints, accelerating tumour evolution. *Ann. Oncol.* **31**, 884–893 (2020).
5. Kim, H. et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 (2020).
6. Turner, K. M. et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).

7. Verhaak, R. G. W., Bafna, V. & Mischel, P. S. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat. Rev. Cancer* **19**, 283–288 (2019).
8. Cox, D., Yuncken, C. & Spriggs, A. I. Minute chromatin bodies in malignant tumours of childhood. *Lancet* **286**, 55–58 (1965).
9. van der Bliek, A. M., Lincke, C. R. & Borst, P. Circular DNA of 3T6R50 double minute chromosomes. *Nucleic Acids Res.* **16**, 4841–4851 (1988).
10. Hamkalo, B. A., Farnham, P. J., Johnston, R. & Schimke, R. T. Ultrastructural features of minute chromosomes in a methotrexate-resistant mouse 3T3 cell line. *Proc. Natl Acad. Sci.* **82**, 1126–1130 (1985).
11. Maurer, B. J., Lai, E., Hamkalo, B. A., Hood, L. & Attardi, G. Novel submicroscopic extrachromosomal elements containing amplified genes in human cells. *Nature* **327**, 434–437 (1987).
12. VanDevanter, D. R., Piaskowski, V. D., Casper, J. T., Douglass, E. C. & Von Hoff, D. D. Ability of circular extrachromosomal DNA molecules to carry amplified MYCN protooncogenes in human neuroblastomas in vivo. *J Natl Cancer Inst.* **82**, 1815–1821 (1990).
13. Nathanson, D. A. et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science* **343**, 72–76 (2014).
14. Ståhl, F., Wettergren, Y. & Levan, G. Amplicon structure in multidrug-resistant murine cells: a nonrearranged region of genomic DNA corresponding to large circular DNA. *Mol. Cell. Biol.* **12**, 1179–1187 (1992).
15. Vicario, R. et al. Patterns of HER2 gene amplification and response to anti-HER2 therapies. *PLoS ONE* **10**, e0129876 (2015).
16. Carroll, S. M. et al. Double minute chromosomes can be produced from precursors derived from a chromosomal deletion. *Mol. Cell. Biol.* **8**, 1525–1533 (1988).
17. Kitajima, K., Haque, M., Nakamura, H., Hirano, T. & Utiyama, H. Loss of irreversibility of granulocytic differentiation induced by dimethyl sulfoxide in HL-60 sublines with a homogeneously staining region. *Biochem. Biophys. Res. Commun.* **288**, 1182–1187 (2001).
18. Quinn, L. A., Moore, G. E., Morgan, R. T. & Woods, L. K. Cell lines from human colon carcinoma with unusual cell products, double minutes, and homogeneously staining regions. *Cancer Res.* **39**, 4914–4924 (1979).
19. Storlazzi, C. T. et al. Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure. *Genome Res.* **20**, 1198–1206 (2010).
20. Wahl, G. M. The importance of circular DNA in mammalian gene amplification. *Cancer Res.* **49**, 1333–1340 (1989).
21. Kumar, P. et al. ATAC–seq identifies thousands of extrachromosomal circular DNA in cancer and cell lines. *Sci. Adv.* **6**, eaba2489 (2020).
22. Morton, A. R. et al. Functional enhancers shape extrachromosomal oncogene amplifications. *Cell* **179**, 1330–1341 (2019).
23. Helmsauer, K. et al. Enhancer hijacking determines extrachromosomal circular *MYCN* amplicon architecture in neuroblastoma. *Nat. Commun.* **11**, 5823 (2020).
24. Itoh, N. & Shimizu, N. DNA replication-dependent intranuclear relocation of double minute chromatin. *J. Cell Sci.* **111**, 3275–3285 (1998).
25. Kanda, T., Sullivan, K. F. & Wahl, G. M. Histone–GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr. Biol.* **8**, 377–385 (1998).
26. Oobatake, Y. & Shimizu, N. Double-strand breakage in the extrachromosomal double minutes triggers their aggregation in the nucleus, micronucleation, and morphological transformation. *Genes Chromosomes Cancer* **59**, 133–143 (2020).
27. Beliveau, B. J. et al. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl Acad. Sci. USA* **109**, 21301–21306 (2012).
28. Koche, R. P. et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat. Genetics* **52**, 29–34 (2019).
29. Parker, S. C. J. et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl Acad. Sci. USA* **110**, 17921–17926 (2013).
30. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
31. Lovén, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
32. Filippakopoulos, P. et al. Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
33. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
34. Ren, C. et al. Spatially constrained tandem bromodomain inhibition bolsters sustained repression of BRD4 transcriptional activity for TNBC cell growth. *Proc. Natl Acad. Sci. USA* **115**, 7949–7954 (2018).
35. Deshpande, V. et al. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat. Commun.* **10**, 392 (2019).
36. Luebeck, J. et al. AmpliconReconstructor integrates NGS and optical mapping to resolve the complex structures of focal amplifications. *Nat. Commun.* **11**, 4374 (2020).
37. Schwab, M., Klempnauer, K. H., Alitalo, K., Varmus, H. & Bishop, M. Rearrangement at the 5′ end of amplified c-*myc* in human COLO 320 cells is associated with abnormal transcription. *Mol. Cell. Biol.* **6**, 2752–2755 (1986).
38. L'Abbate, A. et al. Genomic organization and evolution of double minutes/homogeneously staining regions with *MYC* amplification in human cancer. *Nucleic Acids Res.* **42**, 9131–9145 (2014).
39. Hann, S. R., King, M. W., Bentley, D. L., Anderson, C. W. & Eisenman, R. N. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* **52**, 185–195 (1988).
40. Carramusa, L. et al. The PVT-1 oncogene is a Myc protein target that is overexpressed in transformed cells. *J. Cell. Physiol.* **213**, 511–518 (2007).
41. Cho, S. W. et al. Promoter of lncRNA gene *PVT1* is a tumor-suppressor DNA boundary element. *Cell* **173**, 1398–1412 (2018).
42. Tolomeo, D., Agostini, A., Visci, G., Traversa, D. & Storlazzi, C. T. *PVT1*: a long non-coding RNA recurrently involved in neoplasia-associated fusion transcripts. *Gene* **779**, 145497 (2021).
43. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919 (2016).
44. Fulco, C. P. et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
45. Park, J. et al. A reciprocal regulatory circuit between CD44 and FGFR2 via c-myc controls gastric cancer cell growth. *Oncotarget* **7**, 28670–28683 (2016).
46. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).
47. Zhu, Y. et al. Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. *Cancer Cell* **39**, 694–707 (2021).
48. Xue, K. S., Hooper, K. A., Ollodart, A. R., Dingens, A. S. & Bloom, J. D. Cooperation between distinct viral variants promotes growth of H3N2 influenza in cell culture. *Elife* **5**, e13974 (2016).
49. Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344–348 (2006).
50. Henssen, A. et al. Targeting MYCN-driven transcription by BET-bromodomain inhibition. *Clin. Cancer Res.* **22**, 2470–2481 (2016).

Methods

Cell culture

The TR14 neuroblastoma cell line was a gift from J. J. Molenaar. Cell line identity for the master stock was verified by STR genotyping (IDEXX BioResearch). All remaining cell lines used were obtained from ATCC and verified by FISH or genomic sequencing. TR14 cells were cultured in RPMI-1640 medium (Thermo Fisher Scientific) with 1% penicillin–streptomycin (pen-strep), and 10% fetal calf serum (FCS). COLO320-DM, COLO320-HSR and HCC1569 cells were maintained in Roswell Park Memorial Institute 1640 (RPMI; Life Technologies, 11875-119) supplemented with 10% fetal bovine serum (FBS; Hyclone, SH30396.03) and 1% pen-strep (Thermo Fisher Scientific, 15140-122). PC3 cells were maintained in Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific, 11995073) supplemented with 10% FBS and 1% pen-strep. HK359 cells were maintained in DMEM/nutrient mixture F-12 (DMEM/F12 1:1; Gibco, 11320-082), B-27 supplement (Gibco, 17504044), 1% pen-strep, GlutaMAX (Gibco, 35050061), human epidermal growth factor (EGF, 20 ng ml^{-1}; Sigma-Aldrich, E9644), human fibroblast growth factor (FGF, 20 ng ml^{-1}; Peprotech) and heparin (5 μg ml^{-1}; Sigma-Aldrich, H3149-500KU). SNU16 cells were maintained in DMEM/F12 supplemented with 10% FBS and 1% pen-strep. All cells were cultured at 37 °C with 5% CO_2. All cell lines tested negative for mycoplasma contamination.

Metaphase chromosome spread

Cells in metaphase were prepared by KaryoMAX (Gibco) treatment at 0.1 μg ml^{-1} for 3 h. Single-cell suspension was then collected and washed by PBS, and treated with 75 mM KCl for 15–30 min. Samples were then fixed by 3:1 methanol:glacial acetic acid, v/v and washed an additional three times with the fixative. Finally, the cell pellet resuspended in the fixative was dropped onto a humidified slide. The distribution of ecDNA counts in metaphase for COLO320-DM, PC3 and HK359 have been described previously[1,6]. We find that the majority of cells examined in metaphase are ecDNA$^+$, with a small proportion of HSR$^+$ cells: COLO320-DM: 80% (80/100 cells) ecDNA$^+$, 14% (14/100 cells) HSR$^+$, 6% (6/100 cells) ecDNA$^+$HSR$^+$; PC3: 80% (43/54 cells) ecDNA$^+$, 11% (6/54 cells) HSR$^+$, 9% (5/54 cells) ecDNA$^+$HSR$^+$; SNU16-dCas9-KRAB: 100% (29/29 cells) ecDNA$^+$.

Metaphase DNA FISH

Slides containing fixed cells in interphase or metaphase were briefly equilibrated by 2× SSC, followed by dehydration in 70%, 85% and 100% ethanol for 2 min each. FISH probes in hybridization buffer (Empire Genomics) were added onto the slide, and the sample was covered by a coverslip, denatured at 75 °C for 1 min on a hotplate and hybridized at 37 °C overnight. The coverslip was then removed, and the sample was washed once by 0.4× SSC with 0.3% IGEPAL, and twice by 2× SSC with 0.1% IGEPAL, for 2 min each. DNA was stained with DAPI and washed with 2× SSC. Finally, the sample was mounted by mounting medium (Molecular Probes) before imaging.

Interphase DNA FISH

The Oligopaint FISH probe libraries were constructed as described previously[51]. Each oligo consists of a 40-nucleotide (nt) homology to the hg19 genome assemble designed from the algorithm developed from the laboratory of T. Wu (https://oligopaints.hms.harvard.edu/). Each library subpool consists of a unique set of primer pairs for orthogonal PCR amplification and a 20-nt T7 promoter sequence for in vitro transcription and a 20-nt region for reverse transcription. Individual Oligopaint probes were generated by PCR amplification, in vitro transcription and reverse transcription, in which ssDNA oligos conjugated with ATTO488 and ATTO647 fluorophores were introduced during the reverse transcription step. The Oligopaint-covered genomic regions (hg19) used in this study are as follows: chr8:116,967,673–118,566,852 (hg19_COLO_nonecDNA_1.5Mbp), chr8:127,435,083–129,017,969 (hg19_COLO_ecDNA_1.5Mbp), chr8:128,729,248–128,831,223 (hg19_PC3_ecDNA1_100kb). A ssDNA oligo pool was ordered and synthesized from Twist Bioscience. Fifteen-millimetre #1.5 round glass coverslips (Electron Microscopy Sciences) were pre-rinsed with anhydrous ethanol for 5 min, air-dried, and coated with L-poly lysine solution (100 μg ml^{-1}) for at least 2 h. Fully dissociated COLO320-DM or PC3 cells were seeded onto the coverslips and recovered for at least 6 h before experiments. Cells were fixed with 4% (v/v) methanol-free paraformaldehyde diluted in 1× PBS at room temperature for 10 min. Then cells were washed twice with 1× PBS and permeabilized in 0.5% Triton-X100 in 1× PBS for 30 min. After washing twice in 1× PBS, cells were treated with 0.1 M HCl for 5 min, followed by three washes with 2× SSC and 30 min incubation in 2× SSC + 0.1% Tween20 (2× SSCT) + 50% (v/v) formamide (EMD Millipore, S4117). For each sample, we prepare 25 μl hybridization mixture containing 2× SSCT+ 50% formamide +10% dextran sulfate (EMD Millipore, S4030) supplemented with 0.5 μl 10 mg ml^{-1} RNaseA (Thermo Fisher Scientific, 12091-021) + 0.5 μl 10 mg ml^{-1} salmon sperm DNA (Thermo Fisher Scientific, 15632011) and 20 pmol probes with distinct fluorophores. The probe mixture was thoroughly mixed by vortexing, and briefly microcentrifuged. The hybridization mix was transferred directly onto the coverslip, which was inverted facing a clean slide. The coverslip was sealed onto the slide by adding a layer of rubber cement around the edges. Each slide was denatured at 78 °C for 4 min followed by transferring to a humidified hybridization chamber and incubated at 42 °C for 16 h in a heated incubator. After hybridization, samples were washed twice for 15 min in pre-warmed 2× SSCT at 60 °C and then were further incubated at 2× SSCT for 10 min at room temperature, at 0.2× SSC for 10 min at room temperature, and at 1× PBS for 2 × 5 min with DNA counterstaining with DAPI. Then coverslips were mounted on slides with Prolong Diamond Antifade Mountant (Thermo Fisher Scientific P36961) for imaging acquisition.

DNA FISH of primary neuroblastoma samples was performed on 4-μm sections of FFPE blocks. Slides were deparaffinized, dehydrated and incubated in pre-treatment solution (Dako) for 10 min at 95–99 °C. Samples were treated with pepsin solution for 2 min at 37 °C. For hybridization, the ZytoLight SPEC MYCN/2q11 Dual Color Probe (ZytoVision) was used. Incubation took place overnight at 37 °C, followed by counterstaining with 4,6-diamidino-2-phenylindole (DAPI).

Nascent RNA FISH

To quantify the *MYC* gene expression on the ecDNAs, we ordered the RNA FISH probes conjugated with a Quasar 570 dye (Biosearch Technologies ISMF-2066-5) targeting the intronic region of the human (hg19) *MYC* gene for detection of nascent RNA transcript. We also ordered the RNA FISH probes conjugated with a Quasar 670 dye targeting the exonic region of the human *MYC* gene (Biosearch Technologies VSMF-2231-5) for detection of both mature and nascent RNA transcripts. For simultaneous detection of both ecDNA and MYC transcription, 125 nM RNA FISH probes was mixed with the DNA FISH probes (100-kb probe instead of the 1.5-Mbp probe) together in the hybridization buffer with RNase inhibitor (Thermo Fisher Scientific, cat# AM2694) and incubated at 37 °C overnight for around 16 h. After hybridization, samples were washed twice for 15 min in pre-warmed 2× SSCT at 37 °C and then were further incubated at 2× SSCT for 10 min at room temperature, at 0.2× SSC for 10 min at room temperature, and at 1× PBS for 2 ×5 min with DNA counterstaining with DAPI. Then coverslips were mounted on slides with Prolong Diamond Antifade Mountant for imaging acquisition.

Microscopy

DNA FISH images were acquired either with conventional fluorescence microscopy or confocal microscopy. Conventional fluorescence microscopy was performed using an Olympus BX43 microscope, and images were acquired with a QiClick cooled camera. Confocal microscopy was performed using a Leica SP8 microscope with lightning deconvolution

(UCSD School of Medicine Microscopy Core) or a ZEISS LSM 880 inverted confocal microscope. *Z*-stacks were acquired over an average depth of approximately 8 μm, with roughly 0.6 μm step size.

DNA and RNA FISH images were acquired on the ZEISS LSM 880 Inverted Confocal microscope attached with an Airyscan 32 GaAsP PMT area detector. Before imaging, the beam position was calibrated centring on the 32-detector array. Images were taken under the Airyscan SR mode with a Plan Apochromat 63×/NA 1.40 oil objective in a lens immersion medium having a refractive index 1.515 at 30 °C. We used 405 nm (excitation wavelength) and 460 nm (emission wavelength) for the DAPI channel, 488 nm (excitation wavelength) and 525 nm (emission wavelength) for the ATTO488 channel, 561 nm (excitation wavelength) and 579 nm (emission wavelength) for the Quasar570 channel and 633 nm (excitation wavelength) and 654 nm (emission wavelength) for the ATTO647 channel. *Z*-stacks were acquired with the optimal *z*-sectioning thickness around 200 nm, followed by post-processing using the provided algorithm from the ZEISS LSM880 platform.

DNA FISH images for primary neuroblastoma samples were collected for 50 non-overlapping tumour cells using a fluorescence microscope (BX63 Automated Fluorescence Microscope, Olympus). Computer-based documentation and image analysis was performed with the SoloWeb imaging system (BioView) *MYCN* amplification (MYCN FISH+) was defined as MYCN/2q11.2 ratio > 4.0, as described in the International Neuroblastoma Risk Group (INRG) report[52]. The tumour samples profiled present with multiple *MYCN* foci visible as in interphase, providing evidence that amplified *MYCN* is extrachromosomal in origin, as is the case for approximately 90% of neuroblastoma cases[28,53–55].

Metaphase DNA FISH image analysis

Colocalization analysis for two-colour metaphase FISH data for *MYC*, *PCAT1* and *PLUT* ecDNAs in COLO320-DM cells described in Extended Data Fig. 4g was performed using Fiji (v.2.1.0/1.53c)[56]. Images were split into the two FISH colours + DAPI channels, and signal threshold set manually to remove background fluorescence. Overlapping FISH signals of the same colour were segmented using watershed segmentation. Colocalization was quantified using the ImageJ-Colocalization Threshold program and individual and colocalized FISH signals were counted using particle analysis.

Colocalization analysis for two-colour metaphase FISH data for *MYC* and *FGFR2* ecDNAs in SNU16 cells described in Fig. 4c, Extended Data Fig. 7a was performed using ecSeg (https://github.com/UCRajkumar/ecSeg, not versioned)[57]. In brief, ecSeg takes as input metaphase FISH images containing DAPI and up to two colours of DNA FISH. ecSeg uses the DAPI signal to classify signals as nuclear (arising from interphase nuclei), chromosomal (arising from metaphase chromosome) or extrachromosomal. It then quantifies DNA FISH signal and colocalization segmented by whether the signal is present on chromosomal or extrachromosomal DNA.

Interphase DNA FISH clustering analysis

To analyse the clustering of ecDNAs, we applied the autocorrelation function as described previously[58] in MATLAB (2019). $g(r)$ estimates the probability of detecting another ecDNA signal at increasing distances from the viewpoint of an index ecDNA signal and is equal to 1 for a uniform, random distribution. Specifically, the pair autocorrelation function $g(\vec{r})$ was calculated by the fast Fourier transform (FFT) method described by the equations below.

$$g(\vec{r}) = \frac{\mathrm{FFT}^{-1}(|\mathrm{FFT}(I)|^2)}{\rho^2 N(\vec{r})}$$

$$N(\vec{r}) = \mathrm{FFT}^{-1}(|\mathrm{FFT}(\mathrm{Mask})|^2)$$

$N(\vec{r})$ is the autocorrelation of a mask matrix that has the value of 1 inside the nucleus used for normalization. The fast Fourier transform and its inverse (FFT and FFT^{-1}) were computed by the fft2() and ifft2() functions in MATLAB, respectively. Autocorrelation functions were calculated first by converting the Cartesian coordinates to polar coordinates by the MATLAB cart2pol() function, binning by radius and by averaging within the assigned bins. For comparing autocorrelation with transcription probability, the value of the autocorrelation function at radius of 0 pixels ($g(0)$) was used to represent the degree of spatial clustering. The $g(0)$ values were also used for calculating statistical significance among groups. For samples from patients with neuroblastoma, we avoided cells that lack ecDNA FISH signal (normal cells in the same tissue section may not have ecDNA amplification) for analysis and used the DAPI channel from the same cells as a control.

Colocalization analysis for SNU16 *MYC* and *FGFR2* ecDNAs in Fig. 4a was performed using confocal images of both metaphase and interphase nuclei from the same slides. Images were split into the two FISH colours, and background fluorescence was removed manually for each channel. Colocalization for each nucleus was quantified using the ImageJ Colocalization Threshold program. Analysis was performed across all *z*-stacks for each nucleus. The Manders coefficient (fraction of *MYC* signal colocalized compared to the total *MYC* signal) was used to quantify colocalization.

ecDNA DNA FISH and nascent RNA FISH image analysis

To characterize the shape and size of ecDNA hubs, we used a synthetic model—Surface Objects from Imaris (v.9.1, Bitplane)—and applied a Gaussian filter (σ = 1 voxel in xy) and background subtraction for optimal segmentation and quantification of ecDNA hubs. ecDNA hubs containing connected voxels were sorted by size and singleton ecDNAs were separated from ecDNA hubs (minimum of two ecDNA molecules).

To measure the number of ecDNA or nascent transcripts, we localized the voxels that correspond to the local maximum of identified DNA or RNA FISH signal using the Imaris spots function module. We validated the accuracy of interphase ecDNA counting by comparing to quantification of ecDNA number by metaphase FISH as well as copy number estimated by WGS (Extended Data Fig. 1f). The copy number distribution from WGS is comparable to that from interphase DNA FISH. Although copy number estimates from WGS and interphase FISH are slightly higher than those quantified by metaphase FISH imaging, this may reflect the fact that individual ecDNAs can contain multiple copies of *MYC*.

WGS

WGS data from COLO320-DM, COLO320-HSR and PC3 cells were generated by a previously published study[1] and raw FASTQ reads obtained from the NCBI Sequence Read Archive (SRA), under BioProject accession number PRJNA506071. Reads were trimmed of adapter content with Trimmomatic[59] (v.0.39), aligned to the hg19 genome using BWA-MEM (0.7.17-r1188), and PCR duplicates removed using Picard's MarkDuplicates. WGS data from SNU16 cells were generated by a previously published study[60] and aligned reads in BAM format from the NCBI SRA, under BioProject accession numbers PRJNA523380. WGS data from HK359 cells were generated by a previously published study[6] and aligned reads in BAM format obtained from the NCBI SRA, under BioProject accession number PRJNA338012. Coverage for WGS was 22× for COLO320-DM, 26× for COLO320-HSR, 1.6× for PC3, 1.2× for HK359 and 7.3× for SNU16.

Generation of the ecDNA-TetO array and BRD4-HaloTag knock-in for live-cell imaging

Single guide RNA (sgRNA) was designed by E-CRISP (http://www.e-crisp.org/E-CRISP/designcrispr.html) targeting about 0.5 kb upstream of the *MYC* TSS or N-terminal *BRD4* gene. The sgRNA sequences are listed in Supplementary Table 2. The sgRNA was cloned into the modified pX330 (Addgene, 42230) construct co-expressing wild type SpCas9 and a PGK-Venus cassette. Around 500-bp homology arms were

PCR-amplified from COLO320-DM cells and cloned into a pUC19 donor vector together with around 96 copies of the TetO array and a blasticidin selection cassette (Addgene, 118713) for the ecDNA-TetO array or with HaloTag (Addgene, 139747) for BRD4. Two micrograms of the donor vector and 1 μg of the sgRNA vector were transfected into COLO320-DM cells by Lipofectamine 3000. For the ecDNA-TetO array, blasticidin (10 μg ml^{-1}) selection was applied after 7 days. For the BRD4-HaloTag knock-in, 100 nM HaloTag ligand JF549 (a gift from the laboratory of L. Lavis) was applied to the cells followed by washing and fluorescence-activated cell sorting (FACS). Individual clones were selected, genotyped by PCR and verified by Sanger sequencing before being tested for imaging. To detect TetO-array-labelled ecDNA molecules, we used the TetR-eGFP construct as described previously[61]. To reduce the dimerization potential associated with wild type eGFP, we generated the A206K point mutation according to a previous report[62]. Tet-eGFP-labelled hubs have a slightly smaller size compared to monomeric TetR-eGFP(A206K)-labelled hubs, potentially owing to eGFP dimerization effects (Extended Data Fig. 2c), but the number of ecDNA hubs per cell is not significantly different with Tet-eGFP versus TetR-eGFP(A206K) (Extended Data Fig. 2d).

Live-cell-imaging microscopy

We transiently expressed TetR-eGFP or TetR-eGFP(A206K)[61,62] and performed imaging experiments two days after transfection. To image BRD4, we stained the cells with 200 nM HaloTag ligand JF646 for 30 min followed by washing 3 times in culture medium, each for 10 min.

To monitor ecDNA dynamics within the nucleus, the COLO320-DM TetO-eGFP cell line was transfected with the PiggyBac vector expressing H2B–SNAPf and the super PiggyBac transposase (2:1 ratio) as described previously[51]. Stable transfectants were selected by 500 μg ml^{-1} G418 and sorted by flow cytometry. Cells were seeded in the 8-well Lab-Tek chambered coverglass for long-term time-lapse imaging throughout the cell cycle. Before imaging, COLO320-DM TetO-eGFP cells were stained with 25 nM SNAP ligand JF669[63] (a gift from the laboratory of L. Lavis) at 37 °C for 30 min followed by 3 washes with regular medium for a total of 30 min. Then cells were transferred to an imaging buffer containing 10% serum in the 1× Opti-Klear live-cell imaging buffer pre-warmed at 37 °C. Cells were imaged at the Zeiss LSM880 microscope pre-stabilized at 37 °C for 2 h. We illuminated the sample with 1% 488-nm laser and 0.75% 633-nm laser with the EC Plan-Neofluar 40×/1.30 Oil lens, beam splitter MBS 488/561/633 and filters BP 495–550 + LP 570. *Z*-stack images were acquired with a 0.3 μm *z*-step size with 3-min intervals between each volumetric imaging for up to 12 h. TetO-labelled ecDNA was similarly analysed as described in the previous DNA and RNA FISH section. For BRD4 and PVT1p-nLuc colocalization analysis, a straight line was drawn across the centre of the objects in a 2D plane and the fluorescent intensity was profiled along the line path.

JQ1 treatment

Cells were then treated for 6 h with 500 nM JQ1 in DMSO unless otherwise indicated (Sigma-Aldrich SML1524) or an equivalent volume of DMSO.

ChIP–seq library preparation

Three to five million cells per replicate were fixed in 1% formaldehyde for 10–15 min at room temperature with rotation and then quenched with 0.125 M glycine for 10 min at room temperature with rotation. For COLO320-DM and COLO320-HSR BRD4 ChIP, five million cells per replicate were fixed for 15 min; for all conditions, three million cells per replicate were fixed for 10 min. Fixed cells were pelleted at 800*g* for 5 min at 4 °C and washed twice with cold PBS before storing at −80 °C. Pellets were thawed and membrane lysis performed in 5 ml LB1 (50 mM HEPES pH 8.0, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1 mM PMSF and Roche protease inhibitors 11836170001) for 10 min at 4 °C with rotation. Nuclei were pelleted at 1,350*g* for 5 min at 4 °C and lysed in 5 ml LB2 (10 mM Tris-Cl pH 8.0, 5 M, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1 mM PMSF, Roche protease inhibitors) for 10 min at room temperature with rotation. Chromatin was pelleted at 1,350*g* for 5 min at 4 °C and resuspended in 1 ml of TE buffer + 0.1% SDS before sonication on a Covaris E220. Samples were clarified by spinning at 16,000*g* for 10 min at 4 °C. Supernatant was transferred to a new tube and diluted with 1 volume of IP dilution buffer (10 mM Tris pH 8.0, 1 mM EDTA, 200 mM NaCl, 1 mM EGTA. 0.2% Na-DOC, 1% Na-laurylsarcosine and 2% Triton X-100). Following the addition of 20 ng spike-in chromatin (Active Motif 61686) and 2 μg spike-in antibody (Active Motif 53083), 50 μl of sheared chromatin was reserved as input and ChIP performed overnight at 4 °C with rotation with 7.5 μg of antibody per immunoprecipitation: H3K27Ac (Abcam ab4729), BRD4 (Bethyl Laboratories A301-985A100).

One hundred microlitres of Protein G Dynabeads per ChIP were washed 3 times in 0.5% BSA in PBS and then bound to antibody-bound chromatin for 4 h at 4 °C with rotation. Antibody-bound chromatin was washed on a magnet 5 times with RIPA wash buffer (50 mM HEPES pH 8.0, 500 mM LiCl, 1 mM EDTA, 1% NP-40 and 0.7% Na-deoxycholate) and once with 1 ml TE buffer (10 mM Tris-Cl pH 8.0 and 1 mM EDTA) with 500 mM NaCl. Washed beads were resuspended in 200 ml ChIP elution buffer (50 mM Tris-Cl pH 8.0, 10 mM EDTA, 1% SDS) and chromatin was eluted following incubation at 65 °C for 15 min. Supernatant and input chromatin were removed to fresh tubes and reverse cross-linked at 65 °C overnight. Samples were diluted with 200 ml TE buffer and treated with 0.2 mg ml^{-1} RNase A (Qiagen 19101) for 2 h at 37 °C, then 0.2 mg ml^{-1} Proteinase K (New England Biolabs P8107S) for 30 min at 55 °C. DNA was purified using the ChIP DNA Clean & Concentrator kit (Zymo Research D5205). ChIP sequencing libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs E7645S) with dual indexing (New England Biolabs E7600S) following the manufacturer's instructions. ChIP–seq libraries were sequenced on an Illumina HiSeq 4000 with paired-end 76 bp read lengths.

ChIP–seq data processing

Paired-end reads were aligned to the hg19 genome using Bowtie2[64] (v.2.3.4.1) with the very-sensitive option after adapter trimming with Trimmomatic[59] (v.0.39). Reads with MAPQ values less than 10 were filtered using SAMtools (v.1.9) and PCR duplicates removed using Picard's MarkDuplicates (v.2.20.3-SNAPSHOT). MACS2[65] (v.2.1.1.20160309) was used for peak-calling with the following parameters: macs2 callpeak -t chip_bed -c input_bed -n output_file -f BED -g hs -q 0.01 --nomodel --shift 0. A reproducible peak set across biological replicates was defined using the IDR framework (v.2.0.4.2). Reproducible peaks from all samples were then merged to create a union peak set. ChIP–seq signal was converted to bigwig format for visualization using deepTools bamCoverage[66] (v.3.3.1) with the following parameters: --bs 5 --smoothLength 105 --normalizeUsing CPM --scaleFactor 10. Enrichment of ChIP signal at peaks was performed using deepTools computeMatrix on ChIP signal in bigwig format containing the ratio of BRD4 ChIP signal over input calculated using deepTools bamCoverage[66] (v.3.3.1) with the following parameters: --operation ratio --bs 5 --smoothLength 105.

RT–qPCR

RNA was extracted using RNeasy Plus mini Kit (Qiagen 74136). Purified RNA was quantified by Nanodrop (Thermo Fisher Scientific). For RT–qPCR, 50 ng of RNA, 1× Brilliant II qRT-PCR mastermix with 1 μl RT/RNase block (Agilent 600825), and 200 nM forward and reverse primer were used. Each Ct value was measured using Lightcycler 480 (Roche) and each mean dCt was averaged from a duplicate RT–qPCR reaction and performed in biological triplicate. Relative *MYC* RNA level (RT–qPCR primers MYC_exon3_fw and MYC_exon3_rv) was calculated by the ddCt method compared to 18S and GAPDH controls (RT–qPCR primers GAPDH_fw, GAPDH_rv, 18S_fw, 18S_rv). *P* values were calculated using a Student's *t*-test by comparing the relative fold change of biological triplicates. Primer sequences are listed in Supplementary Table 1.

Drug treatments

Approximately 0.6×10^6 COLO320-DM or COLO-320-HSR cells were plated in 6-well plates and cultured under standard conditions for 24 h. Cells were then treated for 6 h with one of the following: 500 nM JQ1 (Sigma-Aldrich SML1524), 500 nM MS645 (Sigma Aldrich SML2549), 1 μM THZ-1 (Selleck Chemicals S7549), 20 μM SGC-SCP30 (Selleck Chemicals S7256), 10 μM OICR-9429 (Selleck Chemicals S7833), 50 μM MI-3 (Selleck Chemicals S7619), 2 μM trichostatin A (Selleck Chemicals S1045), or DMSO. Experiments were performed in biological triplicates. RT–qPCR was performed as above in technical triplicates.

Cell viability assay

Cells were plated in 96-well plates at 25,000 cells per well in triplicate and incubated either with JQ1 (Sigma-Aldrich SML1524) at the indicated concentrations or an equivalent volume of DMSO for 48 h. Cell viability was measured using the CellTiterGlo assay kit (Promega G7572) in triplicate with luminescence measured on SpectraMax M5 plate reader with an integration time of 1 s per well. Luminescence was normalized to the DMSO-treated controls and *P* values calculated using a Student's *t*-test comparing biological triplicates.

Cell proliferation assay

Cells were plated in 96-well plates at 10,000 cells per well and incubated either with JQ1 (Sigma-Aldrich SML1524) at the indicated concentrations or an equivalent volume of DMSO. Every 24 h, cells were collected and counted on a Countess 3 Automated Cell Counter (Thermo Fisher Scientific) with Trypan Blue used to assess cell viability. *P* values were calculated using a Student's *t*-test comparing biological triplicates.

COLO320-DM WGS sequencing and data processing

Genomic DNA was sheared on a Covaris S2 (Covaris) and libraries were made using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB). Indexed libraries were pooled, and paired end sequenced (2 × 75 bp) on an Illumina NextSeq 500 sequencer. Read data were processed in BaseSpace (https://basespace.illumina.com). Reads were aligned to the *Homo sapiens* genome (hg19) using BWA aligner ve.0.7.13 (https://github.com/lh3/bwa) with default settings. Coverage for ultra-low WGS for COLO320-DM was 0.3×.

COLO320-DM nanopore sequencing and data processing

Genomic DNA from COLO320-DM cells was extracted using a MagAttract HMW DNA Kit (Qiagen 67563) and prepared for long-read sequencing using a Ligation Sequencing Kit (Oxford Nanopore Technologies SQK-LSK109) according to the manufacturer's instructions. Sequencing was performed on a MinION (Oxford Nanopore Technologies). Coverage for long-read nanopore sequencing for COLO320-DM was 0.5× genome-wide and 50× for the *MYC* amplicon.

Bases were called from FAST5 files using Guppy (Oxford Nanopore Technologies, v.2.3.7). Reads were then aligned using NGMLR[67] (v.0.2.7) with the following parameters: -x ont–no-lowqualitysplit. Structural variants were called using Sniffles[67] (v.1.0.11) using the following parameters: -s 1 --report_BND --report_seq.

COLO320-DM optical mapping data collection and processing

Ultra-high molecular weight (UHMW) DNA was extracted from frozen cells preserved in DMSO following the manufacturer's protocols (Bionano Genomics). Cells were digested with Proteinase K and RNAse A. DNA was precipitated with isopropanol and bound with nanobind magnetic disks. Bound UHMW DNA was resuspended in the elution buffer and quantified with Qubit dsDNA assay kits (Thermo Fisher Scientific).

DNA labelling was performed following the manufacturer's protocols (Bionano Genomics). Standard Direct Labeling Enzyme 1 (DLE-1) reactions were carried out using 750 ng of purified UHMW DNA. The fluorescently labelled DNA molecules were imaged sequentially across nanochannels on a Saphyr instrument. A genome coverage of approximately 400× was achieved.

De novo assemblies of the samples were performed with Bionano's de novo assembly pipeline (Bionano Solve v.3.6) using standard haplotype aware arguments. With the Overlap-Layout-Consensus paradigm, pairwise comparison of DNA molecules having 248× coverage against the reference was used to create a layout overlap graph, which was then used to generate the initial consensus genome maps. By realigning molecules to the genome maps (*P* value cut-off of less than 10^{-12}) and by using only the best matched molecules, a refinement step was done to refine the label positions on the genome maps and to remove chimeric joins. Next, during an extension step, the software aligned molecules to genome maps ($P < 10^{-12}$), and extended the maps based on the molecules aligning past the map ends. Overlapping genome maps were then merged ($P < 10^{-16}$). These extension and merge steps were repeated five times before a final refinement ($P < 10^{-12}$) was applied to 'finish' all genome maps.

In vitro ecDNA digestion and PFGE

Genomic DNA from COLO320-DM cells was embedded in agarose beads as previously described[68]. In brief, molten 1% certified low-melt agarose (Bio-Rad, 1613112) in PBS and mineral oil (Sigma Aldrich, 69794) was equilibrated to 45 °C. Fifty million cells were pelleted, washed twice with cold 1× PBS, resuspended in 2 ml PBS, and briefly heated to 45 °C. Two millilitres of agarose solution was added to cells followed by the addition of 10 ml mineral oil. The mixture was swirled rapidly to create an emulsion, then poured into cold PBS with continuous stirring to solidify agarose beads. The resulting mixture was centrifuged at 500*g* for 10 min; the supernatant was removed and beads were resuspended in 10 ml PBS and centrifuged in a clean conical tube. Supernatant was removed, beads were resuspended in buffer SDE (1% SDS, 25 mM EDTA at pH 8.0) and placed on a shaker for 10 min. Beads were pelleted again, resuspended in buffer ES (1% *N*-laurolsarcosine sodium salt solution, 25 mM EDTA at pH 8.0, 50 μg ml^{-1} proteinase K) and incubated at 50 °C overnight. On the following day, proteinase K was inactivated with 25 mM EDTA with 1 mM PMSF for 1 h at room temperature with shaking. Beads were then treated with RNase A (1 mg ml^{-1}) in 25 mM EDTA for 30 min at 37 °C, and washed with 25 mM EDTA with a 5-min incubation.

To perform in vitro Cas9 digestion, 50–100 μl agarose beads containing DNA were washed three times with 1× NEBuffer 3.1 (New England BioLabs) with 5-min incubations. Next, DNA was digested in a reaction with 30 nM sgRNA (Synthego) and 30 nM spCas9 (New England BioLabs, M0386S) after pre-incubation of the reaction mix at room temperature for 10 min. Cas9 digestion was performed at 37 °C for 4 h, followed by overnight digestion with 3 μl proteinase K (20 mg ml^{-1}) in a 200-μl reaction. Proteinase K was inactivated with 1 mM PMSF for 1 h with shaking. Beads were then washed with 0.5× TAE buffer three times with 10-min incubations. Beads were loaded into a 1% certified low-melt agarose gel (Bio-Rad, 1613112) in 0.5× TAE buffer with ladders (CHEF DNA Size Marker, 0.2–2.2 Mb, *Saccharomyces cerevisiae* ladder: Bio-Rad, 1703605; CHEF DNA Size Marker, 1–3.1 Mb, *Hansenula wingei* ladder: Bio-Rad, 1703667) and PFGE was performed using the CHEF Mapper XA System (Bio-Rad) according to the manufacturer's instructions and using the following settings: 0.5× TAE running buffer, 14 °C, two-state mode, run time duration of 16 h 39 min, initial switch time of 20.16 s, final switch time of 2 min 55.12 s, gradient of 6 V cm^{-1}, included angle of 120°, and linear ramping. Gel was stained with 3× Gelred (Biotium) with 0.1 M NaCl on a rocker for 30 min covered from light and imaged. Bands were then extracted and DNA was purified from agarose blocks using β-agarase I (New England BioLabs, M0392L) following the manufacturer's instructions.

To sequence the resulting DNA, we first transposed it with Tn5 transposase produced as previously described[69], in a 50-μl reaction with TD buffer[70], 50 ng DNA and 1 μl transposase. The reaction was performed at 37 °C for 5 min, and transposed DNA was purified using a MinElute

PCR Purification Kit (Qiagen, 28006). Libraries were generated by five rounds of PCR amplification using NEBNext High-Fidelity 2X PCR Master Mix (NEB, M0541L), purified using a SPRIselect reagent kit (Beckman Coulter, B23317) at 1.2× volumes and sequenced on the Illumina Miseq platform.

COLO320-DM reconstruction strategy

Owing to the large size of the COLO320DM ecDNA (4.3 Mb), we used a scaffolding strategy based on a manual combination of results from multiple data sources. All data that required alignment back to a reference genome used hg19.

The first source of data used was the copy-number aware breakpoint graph detected by AmpliconArchitect (v.1.2)[35] (AA) generated from low-coverage WGS data. The AA graph specified copy numbers of amplicon segments as well as genomic breakpoints between them. AA was run with default settings and seed regions were identified using the PrepareAA pipeline (v.0.931.0, https://github.com/jluebeck/PrepareAA) with CNVKit (v.0.9.6)[71]. The AA graph file was cleaned with the PrepareAA 'graph_cleaner.py' script to remove edges which conform to sequencing artifact profiles—namely, very short everted (inside-out read pair) orientation edges. Such spurious edges appear as numerous short brown 'spikes' in the AA amplicon image. Second, we used optical map (OM) contigs (Bionano Genomics) which we incorporated with the AA breakpoint graph. We used AmpliconReconstructor (v.1.01)[36] (AR) to scaffold together individual breakpoint graph segments against the collection of OM contigs. We ran AR with the–noConnect flag set and otherwise default settings. Third, we used the OM alignment tool FaNDOM (v.0.2)[72] (default settings) to correct and infer additional OM contig reference alignments and junctions missed by AA and AR. OM contigs identified three additional breakpoint edges, which were subsequently added into the AA graph file. Finally, we incorporated fragment size and sequencing data from PFGE experiments, identifying from the separated bands the estimated length and identity of genomic segments between CRISPR cut sites.

We explored the various ways in which the overlapping OM scaffolds could be joined while conforming to the PFGE fragment sizes and identities of the genomic regions suggested from the PFGE data. We selected a candidate structure that was concordant with the PFGE cut data expected fragment sizes, as well as intra-fragment sequence identity and multiplicity of copy count as suggested by AA analysis of the sequenced PFGE bands. The reconstruction used all but five discovered genomic breakpoint edges inside the DM region. The remaining five edges were scaffolded by two different OM contigs and each scaffold individually suggested a separate site of structural heterogeneity within the ecDNA as compared against the reconstruction.

We required that the entirety of the significantly amplified amplicon segments was used in the reconstruction. We estimated that at the baseline, genomic segments appearing once in the reconstruction existed with a copy number between 170 and 190. In the final structure, all amplicon segments with a copy number greater than 40 were used. In addition, when segments were repeated inside the reconstruction, we ensured that the multiplicities of the amplicon segments suggested the reconstruction matched the multiplicities of the amplicon segments as reported by WGS.

For fine-mapping analysis of the *PVT1-MYC* breakpoint, reads that align to both *PVT1* and *MYC* were extracted from WGS short-read sequencing, which identified 10 unique reads supporting the breakpoint. Multiple sequence alignment was performed with ClustalW (v.2.1) for visualization.

RNA-seq library preparation

COLO320-DM cells were transfected with Alt-R S.p. Cas9 Nuclease V3 (IDT, 1081058) complexed with a non-targeting control sgRNA (Synthego) with a Gal4 sequence following Synthego's RNP transfection protocol using the Neon Transfection System (Thermo Fisher Scientific, MPK5000). A total of 500,000 to 1 million cells were collected, and RNA was extracted using the RNeasy Plus Mini Kit (Qiagen, 74136). Genomic DNA was removed from samples using the TURBO DNA-free kit (Thermo Fisher Scientific, AM1907), and RNA-seq libraries were prepared using the TruSeq Stranded mRNA Library Prep (Illumina, 20020595) following the manufacturer's protocol. RNA-seq libraries were sequenced on an Illumina HiSeq 4000 with paired-end 75 bp read lengths.

RNA-seq data processing

Paired-end reads were aligned to the hg19 genome using STAR-Fusion[73] (v.1.6.0) and the genome build GRCh37_gencode_v19_CTAT_lib_Mar272019.plug-n-play. The numbers of reads supporting the *PVT1-MYC* fusion transcript were obtained from the 'star-fusion.fusion_predictions.abridged.tsv' output file and the junction read counts and spanning fragment counts were combined. Reads supporting the canonical *MYC* exon 1–2 junction were obtained using the Gviz (v.1.30.3) package in R (v.3.6.1)[74] in a sashimi plot.

Lentivirus production

Lentiviruses were produced as previously described[41]. In brief, 4 million HEK293Ts per 10 cm plate were plated the evening before transfection. Helper plasmids, pMD2.G and psPAX2, were transfected along with the vector plasmid using Lipofectamine 3000 (Thermo Fisher Scientific, L3000) according to the manufacturer's instructions. Supernatants containing lentivirus were collected 48 h later, filtered with a 0.45-μm filter and concentrated using Lenti-X concentrator (Clontech, 631232) and stored at 80 °C.

Stable CRISPR cell line generation

The pHR-SFFV-dCas9-BFP-KRAB (Addgene, 46911) plasmid was modified to dCas9-BFP-KRAB-2A-Blast as previously described[41]. Lentivirus was produced using the modified vector plasmid. Cells were transduced with lentivirus, incubated for 2 days and selected with 1 μg ml^{-1} blasticidin for 10–14 days, and BFP expression was analysed by flow cytometry. To generate stable, monoclonal dCas9-KRAB cell lines, single BFP-positive cell clones were sorted into 96-well plates and expanded. Vector expression was validated by flow cytometry.

CRISPRi in COLO320-DM cells

sgRNAs that target the *MYC* and *PVT1* promoters were previously published[41]. sgRNAs that target enhancers were designed using the Broad Institute sgRNA designer online tool (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). An additional guanine was appended to each of the protospacers that do not start with a guanine. sgRNAs were cloned into either mU6(modified)-sgRNA-Puromycin-mCherry or mU6(modified)-sgRNA-Puromycin-EGFP previously generated[41] and lentiviruses were produced. To evaluate the effects of CRISPRi on gene expression, cells were transduced with sgRNA lentiviruses, incubated for 2 days and selected with 0.5 μg ml^{-1} puromycin for 4 days, and the expression of BFP, GFP and/or mCherry were assessed by flow cytometry. Cells were collected for RT–qPCR assays as described above. All guide sequences are in Supplementary Table 2.

Single-cell paired RNA and ATAC–seq library preparation

Single-cell paired RNA and ATAC–seq libraries for COLO320-DM and COLO320-HSR were generated on the 10x Chromium Single-Cell Multiome ATAC + Gene Expression platform following the manufacturer's protocol and sequenced on an Illumina NovaSeq 6000.

Single-cell RNA and ATAC–seq data processing and analysis

A custom reference package for hg19 was created using cellranger-arc mkref (10x Genomics, v.1.0.0). The single-cell paired RNA and ATAC–seq reads were aligned to the hg19 reference genome using cellranger-arc count (10x Genomics, v.1.0.0).

Subsequent analyses on RNA were performed using Seurat (v.3.2.3)[75], and those on ATAC–seq were performed using ArchR (v.1.0.1)[76]. Cells with more than 200 unique RNA features, less than 20% mitochondrial RNA reads and fewer than 50,000 total RNA reads were retained for further analyses. Doublets were removed using ArchR.

Raw RNA counts were log-normalized using Seurat's NormalizeData function and scaled using the ScaleData function, and the data were visualized on a uniform manifold approximation and projection (UMAP) using the first 30 principal components. Dimensionality reduction for the ATAC–seq data were performed using iterative latent semantic indexing (LSI) with the addIterativeLSI function in ArchR. To impute accessibility gene scores, we used addImputeWeights to add impute weights and plotEmbedding to visualize scores. To compare the accessibility gene scores for *MYC* with *MYC* RNA expression, getMatrixFromProject was used to extract the gene score matrix and the normalized RNA data were used.

To identify variable ATAC–seq peaks on COLO320-DM and COLO320-HSR amplicons, we first calculated amplicon copy numbers on the basis of background ATAC–seq signals as previously described, using a sliding window of 5 Mb moving in 1-Mb increments across the reference genome[77]. We used the copy number *z*-scores calculated for the chr8:124,000,001–129,000,000 interval for estimating copy numbers of *MYC*-bearing ecDNAs in COLO320-DM and *MYC*-bearing chromosomal HSRs in COLO320-HSR. We then incorporated these estimated copy numbers into the variable peak analysis as follows. COLO320-DM and COLO320-HSR cells were separately assigned into 20 bins on the basis of their RNA expression of *MYC*. Next, pseudo-bulk replicates for ATAC–seq data were created using the addGroupCoverages function grouped by *MYC* RNA quantile bins. ATAC–seq peaks were called using addReproduciblePeakSet for each quantile bin, and peak matrices were added using addPeakMatrix. Differential peak testing was performed between the top and the bottom RNA quantile bins using getMarkerFeatures. An FDR cut-off of 1×10^{-15} was imposed. The mean copy number *z*-score for each quantile bin was then calculated and a copy-number fold change between the top and bottom bin was computed. Finally, we filtered on significantly differential peaks that are located in chr8:127,432,631–129,010,071 and have fold changes above the calculated copy number fold change multiplied by 1.5.

HiChIP library preparation

One to four million cells were fixed in 1% formaldehyde in aliquots of one million cells each for 10 min at room temperature. HiChIP was performed as previously described[43,78] using antibodies against H3K27ac (Abcam ab4729; 2 µg antibody for one million cells, 7.5 µg antibody for four million cells) with the following optimizations[79]: SDS treatment at 62 °C for 5 min; restriction digest with MboI for 15 min; instead of heat inactivation of MboI restriction enzyme, nuclei were washed twice with 1× restriction enzyme buffer; biotin fill-in reaction incubation at 37 °C for 15 min; ligation at room temperature for 2 h. HiChIP libraries were sequenced on an Illumina HiSeq 4000 with paired-end 76 bp read lengths.

HiChIP data processing

HiChIP data were processed as described previously[43]. In brief, paired-end reads were aligned to the hg19 genome using the HiC-Pro pipeline (v.2.11.0)[80]. Default settings were used to remove duplicate reads, assign reads to MboI restriction fragments, filter for valid interactions and generate binned interaction matrices. The Juicer (v.1.5) pipeline's HiCCUPS tool and FitHiChIP (v.8.0) were used to identify loops[81,82]. Filtered read pairs from the HiC-Pro pipeline were converted into .hic format files and input into HiCCUPS using default settings. Dangling end, self-circularized, and re-ligation read pairs were merged with valid read pairs to create a one-dimensional signal bed file. FitHiChIP was used to identify 'peak-to-all' interactions at 10-kb resolution using peaks called from the one-dimensional HiChIP data. A lower distance threshold of 20 kb was used. Bias correction was performed using coverage specific bias. HiChIP contact matrices stored in .hic files were visualized in R (v.4.0.3) using gTrack (v.0.1.0) at 10-kb resolution following Knight-Ruiz normalization. We also compared HiChIP contract matrices following ICE and OneD normalization following copy number correction using the dryhic R package (v.0.0.0.9100)[83]. Virtual 4C plots were generated from dumped matrices generated with Juicer Tools (1.9.9). The Juicer Tools tools dump command was used to extract the chromosome of interest from the .hic file. The interaction profile of a 10-kb bin containing the anchor was then plotted in R (v.4.0.3) after normalization by the total number of valid read pairs and smoothing with the rollmean function from the zoo package (v.1.8-9).

Reporter plasmid construction and transfection

We constructed a plasmid containing the 2-kb *PVT1* promoter (chr8:128,804,981–128,806,980, hg19) or the *MYC* promoter (chr8:128,745,990–128,748,526, hg19) driving NanoLuc luciferase (PVT1p-nLuc) and a constitutive thymidine kinase (TK) promoter driving Firefly luciferase as an internal control (Fig. 3b). In brief, pGL4-tk-luc2 (Promega) was digested with KpnI and PciI. A sequence containing multiple cloning sites (GTACCTGAGCTCGCTAGCCTCGAGAAGATCTGCGTACGGTCGAC), NanoLuc and BGH polyA sequence were inserted in tandem into the vector using Gibson assembly (NEBuilder DNA assembly mix). Next, the PVT1 promoter or the MYC promoter was inserted into the vector using NheI and SalI digestion to generate the final reporter construct. For the negative control, a minimal promoter (TAGAGGGTATATAATGGAAGCTCGACTTCCAGCTT) was used in place of the PVT1 promoter. For constructing plasmids with a *cis*-enhancer, an enhancer (chr8:128,347,148–128,348,310, hg19; positive H3K27ac mark and looping to the *PVT1* promoter in HiChIP, overlapping with *BRD4* ChIP peak and ATAC–seq peak in COLO320-DM) was inserted directly 5′ to the promoter into the region with multiple cloning sites. To assess luciferase reporter expression, COLO320-DM or COLO320-HSR cells were seeded into a 24-well plate with 75,000 cells per well. Reporter plasmids were transfected into cells the next day with Lipofectamine 3000 following the manufacturer's protocol, using 0.25 µg DNA per well. Two days later, cells were treated with either JQ1 (500 nM) or DMSO for 6 h before collection. Luciferase levels were quantified using the Nano-Glo Dual reporter luciferase assay (Promega). The reporter level was calculated as the ratio of NanoLuc reading over Firefly reading using Tecan M1000. Mean and standard errors were calculated based on three biological replicates with three technical replicate each.

To analyse the spatial relationship of NanoLuc activity with ecDNA hubs in situ, we designed and ordered the RNA FISH probe sets for NanoLuc luciferase gene (30 probes mix) and Firefly luciferase gene (47 probes mix) conjugated with the Quasar 570 dye and Quasar 670 dye, respectively (Biosearch Technologies). We transfected 0.5 µg *PVT1* promoter or minimal promoter reporter plasmid into COLO320-DM cells seeded on the 12-mm #1.5 round glass coverslips (Electron Microscopy Sciences). Two days after transfection, DNA and RNA FISH were performed as described in the 'Nascent RNA FISH' section except that a 1.5-Mb probe conjugated with Atto488 was applied together with the NanoLuc Quasar 570 probe and Firefly Quasar 670 probe. We applied the same Gaussian smoothing with Gaussian filter ($\sigma = 1$ voxel in xy) and background subtraction in all images for proper segmentation of the active transcription sites of luciferase genes. The size of the active transcription sites was estimated from the diameter of the sphere with identical volume of the segmented objects and the luciferase transcription activity was quantified from the sum of the fluorescence intensity within the segmented transcription sites. The ecDNA hubs were similarly segmented and the binary overlap between the two surfaces was used to determine the spatial relationship between the luciferase gene transcription sites and ecDNA hubs.

SNU16-dCas9-KRAB WGS and data processing

DNA was extracted from collected cells using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. Libraries

were prepared using a modified Nextera library preparation protocol. Eight nanograms of input DNA was combined with 1× TD buffer[70] and 1 µl transposase[69] (40 nM final) in a reaction volume of 50 µl and incubated at 37 °C for 5 min. Transposed DNA was purified using a MinElute PCR Purification Kit (Qiagen) according to the manufacturer's instructions. Libraries were generated by five rounds of PCR amplification, purified using SPRIselect reagent kit (Beckman Coulter, B23317) at 1.2× volumes and sequenced on an Illumina HiSeq 6000 with paired end 2 ×150 bp reads. Coverage for SNU16-dCas9-KRAB WGS was 12×.

Reads were trimmed of adapter content with Trimmomatic[59] (v.0.39), aligned to the hg19 genome using BWA-MEM (0.7.17-r1188), and PCR duplicates removed using Picard's MarkDuplicates (version 2.20.3-SNAPSHOT). Regions of copy number alteration were identified using ReadDepth (version 0.9.8.5) with parameters recommended by AmpliconArchitect (version 1.0), and amplicon reconstruction performed using the default parameters. Structural variant junctions were extracted from the edges_cnseg.txt output files and used for visualization.

ATAC–seq library preparation and data processing

ATAC–seq library preparation was performed as previously described[70] and sequenced on the NovaSeq 6000 platform (Illumina) with 2 × 75 bp reads. Adapter-trimmed reads were aligned to the hg19 genome using Bowtie2 (2.1.0). Aligned reads were filtered for quality using SAMtools (v.1.9), duplicate fragments were removed using Picard (v.2.21.9-SNAPSHOT) and peaks were called using MACS2 (v.2.1.0.20150731) with a *q*-value cut-off of 0.01 and with a no-shift model. Peaks from replicates were merged, read counts were obtained using bedtools (v.2.17.0) and normalized using DESeq2 (v.1.26.0).

To identify accessible elements in *MYC* and *FGFR2* ecDNAs in SNU16, we filtered on all ATAC–seq peaks within known ecDNA-amplified regions (chr8:128,200,000–129,200,000 for the *MYC* ecDNA, chr10:122,000,000–123,680,000 for the *FGFR2* ecDNA) for which the normalized read counts (using the 'counts' function in DESeq2 with normalized = TRUE) exceeded a manually determined threshold (500 for the *MYC* amplicon, 1,000 for the *FGFR2* amplicon). Peaks that met all criteria for two technical replicates were included as candidate DNA elements in the CRISPRi study.

CRISPRi screen

After generation of monoclonal SNU16-dCas9-KRAB cells, *MYC* and *FGFR2* ecDNAs in single clones were assessed using metaphase FISH. A clone with distinct *MYC* and *FGFR2* amplicons on the vast majority of ecDNAs was selected for CRISPRi experiments.

For the pooled experiments in SNU16-dCas9-KRAB, sgRNAs targeting ATAC–seq peaks were designed using the Broad Institute sgRNA designer online tool. An additional guanine was appended to each of the protospacers. Pooled sgRNA cloning was performed as described previously[84]. In brief, sgRNA sequences were designed with flanking Esp3I digestion sites and two nested PCR handles. Oligos were amplified by PCR and then cloned into the lentiGuidePuro vector modified to express a 2A-GFP fusion in frame with puromycin. The vector was pre-digested and then sgRNA cloning was done by one-step digestion–ligation of the insert. One microlitre of this reaction was transformed via electroporation and purified with maxiprep. sgRNA representation was confirmed by sequencing.

SNU16-dCas9-KRAB cells were transduced with the lentiviral guide pool at an effective multiplicity of infection of 0.2. Cells were incubated for 2 days, selected with puromycin for 4 days, and rested for 3–5 days in culture medium without puromycin. Twenty million cells were fixed and a two-colour RNA flowFISH was performed for *ACTB* and either *MYC* or *FGFR2* using the PrimeFlow RNA Assay Kit (Thermo Fisher Scientific) following the manufacturer's protocol and corresponding probe sets (*MYC*: VA1-6000107-PF; *FGFR2*: VA1-14785-PF; *ACTB*: VA6-10506-PF). *ACTB* labels a housekeeping control gene to control for noise in RNA flowFISH due to variable staining intensity. Cells were sorted by FACS using the gating strategy shown in Extended Data Fig. 8c and as previously described[44]. The oncogene (*MYC* or *FGFR2*) was labelled with Alexa Fluor 647 and *ACTB* was labelled with Alexa Fluor 750. On the basis of the assumption that the expression of the housekeeping gene is not correlated with the oncogene, any correlation in fluorescence intensities between the *ACTB* and the oncogene was attributed to flowFISH staining efficiency and manually regressed using the FACS compensation tool. The degree of compensation was determined so that the top and bottom 25% of cells based on Alexa Fluor 647 signal intensity deviated no more than 15% from the population mean in Alexa Fluor 750 signal intensity. After compensation, we gated on cells with positive *ACTB* labelling and sorted cells into six bins using Alexa Fluor 647 MFI corresponding to the following percentile ranges: 0–10% (bin 1), 10–20% (bin 2), 35–45% (bin 3), 55–65% (bin 4), 80–90% (bin 5), 90–100% (bin 6). FACS data were analysed using FlowJo (10.7.0).

Cells were pelleted at 800*g* for 5 min and resuspended in 100 µl lysis buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS). The lysate was incubated at 65 °C for 10 min for reverse cross-linking and cooled to 37 °C. RNase A (10 mg ml^{-1}) was added at 1:50 by volume and incubated at 37 °C for 30 min. Proteinase K (20 mg ml^{-1}) was added at 1:50 by volume and samples were incubated at 45 °C overnight. Genomic DNA was extracted using a Zymo DNA miniprep kit. Libraries were prepared using three rounds of PCR as previously described[84]. Amplified product sizes were validated on a gel, and the final products were purified using an SPRIselect reagent kit (Beckman Coulter, B23318) at 1.2× sample volumes following the manufacturer's protocol. Libraries were sequenced on an Illumina Miseq with paired-end 75 bp read lengths. Read 1 was used for downstream analysis.

Relative abundances of sgRNAs were measured using MAGeCK (v.0.5.9.4)[85]. sgRNA counts were obtained using the 'mageck count' command. For samples with PCR replicates, if a PCR replicate has fewer than 1,000 total sgRNAs passing filter (raw counts > 20), the replicate was excluded. Next, each sgRNA count was divided by total sgRNA counts for each library and multiplied by one million to give a normalized count (counts per million, CPM). For samples with PCR replicates, the mean CPM was calculated for each sgRNA. sgRNAs that have CPMs lower than 20 in the unsorted cells were classified as dropouts and removed from the analysis. We then calculated the $\log_2$-transformed fold change of each sgRNA in each sorted cell bin over unsorted cells by dividing the respective CPMs followed by log-transformation. sgRNA enrichment was then quantified as previously described[84]. In brief, the $\log_2$-transformed fold change in the high-expression bin was subtracted from that in the low-expression bin ($\log_2$(low/high)) for each sgRNA. The resulting $\log_2$(low/high) values were averaged for each candidate regulatory element and *z*-scores were calculated using the formula $z = (x - m)/\text{s.e.}$, where x is the mean $\log_2$(low/high) of the candidate element, m is the mean log2(low/high) of negative control sgRNAs, and s.e. is the standard error calculated from the standard deviation of negative control sgRNAs divided by the square root of the number of sgRNAs targeting the candidate element in independent biological replicates. *Z*-scores were used to compute upper-tail *P* values using the normal distribution function, which were adjusted with p.adjust in R (v.3.6.1) using the Benjamini–Hochberg procedure to produce FDR values. For assessing sgRNA correlations across all six sorted bins for individual elements, we computed Spearman coefficients for all individual sgRNAs across the six fluorescence bins using $\log_2$-transformed fold changes over unsorted cells. All sgRNA sequences used in the CRISPRi experiments in SNU16-dCas9-KRAB cells are listed in Supplementary Table 3.

TR14 amplicon reconstruction

We obtained WGS data for TR14 cells as follows. DNA was extracted from collected cells (NucleoSpin Tissue kit, Macherey-Nagel GmbH & Co. KG). Libraries were prepared (NEBNext Ultra II FS DNA Library Prep Kit for Illumina, New England BioLabs) and sequenced on the NovaSeq

6000 platform (Illumina) with 2 × 50 bp reads. Adapters were trimmed with BBMap 38.58. Reads were then aligned to hg19 using BWA-MEM 0.7.15[86] with default parameters and duplicate reads were removed (Picard 2.20.4). Coverage was computed in 20-bp bins, normalized as CPM, using deepTools 3.3.0[66]. Copy-number variation was called using QDNAseq 1.22.0[87], binning primary alignments with MAPQ ≥ 20 in 10-kb bins, default filtering and additional filtering of bins with more than 5% Ns in the reference. Bins were corrected for GC content and normalized. Segmentation was performed using the CBS method with no transformation of the normalized counts and parameter $\alpha = 0.05$.

Genomic DNA from TR14 cells was extracted using a MagAttract HMW DNA Kit and fragments larger than 10 kb were selected using the Circulomics SRE kit (Circulomics). Libraries were prepared using a Ligation Sequencing Kit and sequenced on a R9.4.1 MinION flowcell (FLO-MIN106). Reads were aligned to hg19 using NGMLR v.0.2.7. Structural variants were called using Sniffles v.1.0.11 and parameters –min_length 15–genotype–min_support 3–report_seq.

To reconstruct the coarse structure of oncogene amplifications in TR14, we compiled all Sniffles structural variants larger than 10 kb with a minimum read support of 15 into one genome graph using gGnome 0.1[88]. In such a graph, nodes represent genomic segments and edges indicate adjacency in the reference genome or resulting from structural variation. Non-amplified segments (that is, mean Illumina WGS coverage less than 10-fold the median chromosome 2 coverage) were discarded from the graph. Strong clusters in the genome graph were identified, partitioning the graph into groups of segments that could be reached from one another. We identified the clusters containing the four amplified oncogenes (*MYCN*, *CDK4*, *MDM2* and *ODC1*) and manually selected circular paths through each cluster that could account for the main copy number steps around the oncogenes. We used gTrack (https://github.com/mskilab/gTrack) for visualization. Hi-C data were used to validate these reconstructions, confirming that all strong off-diagonal signals indicative of structural rearrangements were captured by the reconstruction. Previous studies suggest that the identified amplicons exist as ecDNA[89,90].

Hi-C

Hi-C libraries were prepared as described previously[23]. Samples were sequenced with Illumina Hi-Seq according to standard protocols in 100-bp paired-end mode at a depth of 433.7 million read pairs. FASTQ files were processed using the Juicer pipeline v.1.19.02, CPU version[91], which was set up with BWA v.0.7.17[86] to map short reads to reference genome hg19, from which haplotype sequences were removed and to which the sequence of Epstein-Barr virus (NC_007605.1) was added. Replicates were processed individually. Mapped and filtered reads were merged afterwards. A threshold of MAPQ ≥ 30 was applied for the generation of Hi-C maps with Juicer tools v.1.7.5[91]. Knight-Ruiz normalization per hg19 chromosome was used for Hi-C maps[82,92]; interaction across different chromosome pairs should therefore only carefully be interpreted.

For TR14, we created a custom genome containing the amplicon reconstructions in addition to standard chromosomes. The sequences of amplicons were composed from hg19 on the basis of the order and orientation of their chromosomal fragments. The original fragment locations on hg19 were masked to allow unambiguous mapping. Note, by this also Hi-C reads from wild-type alleles are mapping to the amplicon sequences leading to a mix of signal, depending on the fraction of amplicons and wild-type allele. After mapping, we kept only amplicons and removed all other chromosomes to create Hi-C maps and apply GW_KR normalization using Juicer Tools v.1.19.02[91].

TR14 interaction analysis

TR14 H3K27ac ChIP–seq raw data were downloaded from the Gene Expression Omnibus (GEO) (GSE90683)[93]. We trimmed adapters with BBMap 38.58 and aligned the reads to hg19 using BWA-MEM 0.7.15[86] with default parameters. Coverage tracks were created by extending reads to 200 bp, filtering using the ENCODE DAC blacklist and normalizing to CPM in 10-bp bins with deepTools 3.3.0[66]. Enhancers were called using LILY (https://github.com/BoevaLab/LILY, not versioned)[93] with default parameters.

The *HPCAL1* enhancer region was defined by two LILY-defined boundary enhancers as chr2:10,424,449–10,533,951. A virtual 4C track was generated by the mean genome-wide interaction profile (KR-normalized Hi-C signal in 5-kb bins) across all overlapping 5-kb bins.

For the aggregate analysis of the effect of H3K27 acetylation on interaction, all 5-kb bin pairs located on different amplicons were analysed for their KR-normalized Hi-C signal depending on the mean H3K27ac fold change over input of each of the two bins. We used a 5-fold change threshold to distinguish low- from high-H3K27ac bins.

Reporting summary

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

Data availability

ChIP–seq, HiChIP, Hi-C, RNA-seq and single-cell multiomics (10x Chromium Single Cell Multiome ATAC + Gene Expression) data generated in this study have been deposited in the GEO and are available under accession number GSE159986. Nanopore sequencing data, WGS data, sgRNA sequencing data and targeted ecDNA sequencing data after CRISPR–Cas9 digestion and PFGE generated in this study have been deposited in the SRA and are available under accession number PRJNA670737. Optical mapping data generated in this study have been deposited in GenBank with BioProject code PRJNA731303. The following publicly available data were also used in this study: TR14 H3K27ac ChIP–seq[93] (GEO: GSE90683); COLO320-DM, COLO320-HSR and PC3 WGS[1] (SRA: PRJNA506071); SNU16 WGS[60] (SRA: PRJNA523380); and HK359 WGS[6] (SRA: PRJNA338012). Microscopy image files are available on figshare at https://doi.org/10.6084/m9.figshare.c.5624713.

Code availability

Custom code used in this study is available at https://github.com/ChangLab/ecDNA-hub-code-2021.

51. Xie, L. et al. 3D ATAC-PALM: super-resolution imaging of the accessible genome. *Nat. Methods* **17**,430–436 (2020).
52. Ambros, P. F. et al. International consensus for neuroblastoma molecular diagnostics: report from the International Neuroblastoma Risk Group (INRG) Biology Committee. *Br. J. Cancer* **100**, 1471–1482 (2009).
53. Balaban-Malenbaum, G. & Gilbert, F. Double minute chromosomes and the homogeneously staining regions in chromosomes of a human neuroblastoma cell line. *Science* **198**, 739–741 (1977).
54. Marrano, P., Irwin, M. S. & Thorner, P. S. Heterogeneity of *MYCN* amplification in neuroblastoma at diagnosis, treatment, relapse, and metastasis. *Genes Chromosomes Cancer* **56**, 28–41 (2017).
55. Villamón, E. et al. Genetic instability and intratumoral heterogeneity in neuroblastoma with *MYCN* amplification plus 11q deletion. *PLoS ONE* **8**, e53740 (2013).
56. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
57. Rajkumar, U. et al. EcSeg: semantic segmentation of metaphase images containing extrachromosomal DNA. *Iscience* **21**, 428–435 (2019).
58. Veatch, S. L. et al. Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting. *PLoS ONE* **7**, e31457 (2012).
59. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114 (2014).
60. Ghandi, M. et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
61. Normanno, D. et al. Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat. Commun.* **6**, 7357 (2015).
62. Mirkin, E. V., Chang, F. S. & Kleckner, N. Protein-mediated chromosome pairing of repetitive arrays. *J. Mol. Biol.* **426**, 550–557 (2014).
63. Grimm, J. B. et al. A general method to optimize and functionalize red-shifted rhodamine dyes. *Nat. Methods* **17**, 815–821 (2020).
64. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
65. Zhang, Y. et al. Model-based analysis of ChIP–seq (MACS). *Genome Biol.* **9**, R137 (2008).

66. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
67. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).
68. Overhauser, J. in *Pulsed-Field Gel Electrophoresis, Methods in Molecular Biology* Vol. 12 (eds. Burmeister, M. & Ulanovsky, L.) 129–134 (Humana Press, 1992).
69. Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
70. Corces, M. R. et al. An improved ATAC–seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
71. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
72. Raeisi Dehkordi, S., Luebeck, J. & Bafna, V. FaNDOM: fast nested distance-based seeding of optical maps. *Patterns* **2**, 100248 (2021).
73. Haas, B. J. et al. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol.* **20**, 213 (2019).
74. Hahne, F. & Ivanek, R. in *Statistical Genomics, Methods in Molecular Biology* Vol. 1418 (eds. Mathé, E. & Davis, S.) 335–351 (Humana Press, 2016).
75. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
76. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
77. Satpathy, A. T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
78. Mumbach, M. R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**, 1602–1612 (2017).
79. Mumbach, M. R. et al. HiChIRP reveals RNA-associated chromosome conformation. *Nat. Methods* **16**, 489–492 (2019).
80. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
81. Bhattacharyya, S., Chandra, V., Vijayanand, P. & Ay, F. Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat. Commun.* **10**, 4221 (2019).
82. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
83. Vidal, E. et al. OneD: increasing reproducibility of Hi-C samples with abnormal karyotypes. *Nucleic Acids Res.* **46**, e49 (2018).
84. Flynn, R. A. et al. Discovery and functional interrogation of SARS-CoV-2 RNA–host protein interactions. *Cell* **184**, 2394–2411 (2021).
85. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
86. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
87. Scheinin, I. et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* **24**, 2022–2032 (2014).
88. Hadi, K. et al. Distinct classes of complex structural variation uncovered across thousands of cancer genome graphs. *Cell* **183**, 197–210 (2020).
89. Blumrich, A. et al. The *FRA2C* common fragile site maps to the borders of *MYCN* amplicons in neuroblastoma and is associated with gross chromosomal rearrangements in different cancers. *Hum. Mol. Genet.* **20**, 1488–1501 (2011).
90. Gogolin, S. et al. CDK4 inhibition restores G_1-S arrest in *MYCN*-amplified neuroblastoma cells in the context of doxorubicin-induced DNA damage. *Cell Cycle* **12**, 1091–1104 (2013).
91. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
92. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047 (2013).
93. Boeva, V. et al. Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* **49**, 1408–1413 (2017).

Acknowledgements We thank members of the Chang, Liu, Mischel, and Bafna laboratories for discussions; R. Zermeno, M. Weglarz and L. Nichols at the Stanford Shared FACS Facility for assistance with cell sorting experiments; X. Ji, D. Wagh and J. Coller at the Stanford Functional Genomics Facility for assistance with high-throughput sequencing; and A. Pang of Bionano Genomics for assistance with optical mapping. H.Y.C. was supported by NIH R35-CA209919 and RM1-HG007735; K.L.H. was supported by a Stanford Graduate Fellowship; and K.E.Y. was supported by the National Science Foundation Graduate Research Fellowship Program (NSF DGE-1656518), a Stanford Graduate Fellowship and a NCI Predoctoral to Postdoctoral Fellow Transition Award (NIH F99CA253729). Cell sorting for this project was done on instruments in the Stanford Shared FACS Facility. Sequencing was performed by the Stanford Functional Genomics Facility (supported by NIH grants S10OD018220 and 1S10OD021763). Microscopy was performed on instruments in the UCSD Microscopy Core (supported by NINDS NS047101). A.G.H. is supported by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) (398299703) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 949172). Z.L. is a Janelia Group Leader, and H.Y.C. and R.T. are Investigators of the Howard Hughes Medical Institute.

Author contributions K.L.H., K.E.Y. and H.Y.C. conceived the project. K.L.H. performed and analysed CRISPRi and in vitro ecDNA digestion and PFGE experiments, and analysed single-cell multiomics, RNA-seq and ATAC–seq experiments. K.E.Y. performed and analysed metaphase DNA FISH imaging, ChIP–seq, HiChIP, WGS, COLO320-DM nanopore sequencing and JQ1 perturbation experiments. L.X. performed and analysed interphase DNA and RNA FISH imaging, TetO-eGFP cell line generation and live-cell imaging, and PVT1p-nLuc imaging experiments. Q.S. performed and analysed all luciferase reporter experiments, with the exception of PVT1p-nLuc RNA and DNA FISH, and helped with CRISPRi experiments. K.H. and R.S. analysed TR14 Hi-C data and amplicon reconstruction. J.L. and S.R.D. analysed COLO320-DM WGS, nanopore sequencing, optical mapping data and amplicon reconstruction. J.T.L., S.W., C.C. and J.T. performed and analysed DNA FISH imaging. R.C.G. generated Hi-C, DNA FISH, WGS and nanopore sequencing data for TR14. N.E.W. performed and analysed small-molecule inhibitor experiments and DNA FISH imaging after MS645 treatment. M.E.V. performed Hi-C experiments and data analysis for TR14. I.T.-L.W. performed metaphase DNA FISH imaging. C.V.D. performed and analysed ChIP–seq experiments. K.K. performed HiChIP experiments. J.A.B. helped with CRISPRi experimental design and cloning of the sgRNA pool. R.L. performed RNA-seq experiments. U.R. analysed metaphase DNA FISH data. J.F. generated COLO320-DM WGS data. M.R.C. and J.M.G. wrote the HiChIP data processing pipeline. M.R.C., J.C.R., A.B., A.T.S., R.T., S.M., V.B., A.G.H., P.S.M., Z.L. and H.Y.C. guided data analysis and provided feedback on experimental design. K.L.H., K.E.Y. and H.Y.C. wrote the manuscript with input from all authors.

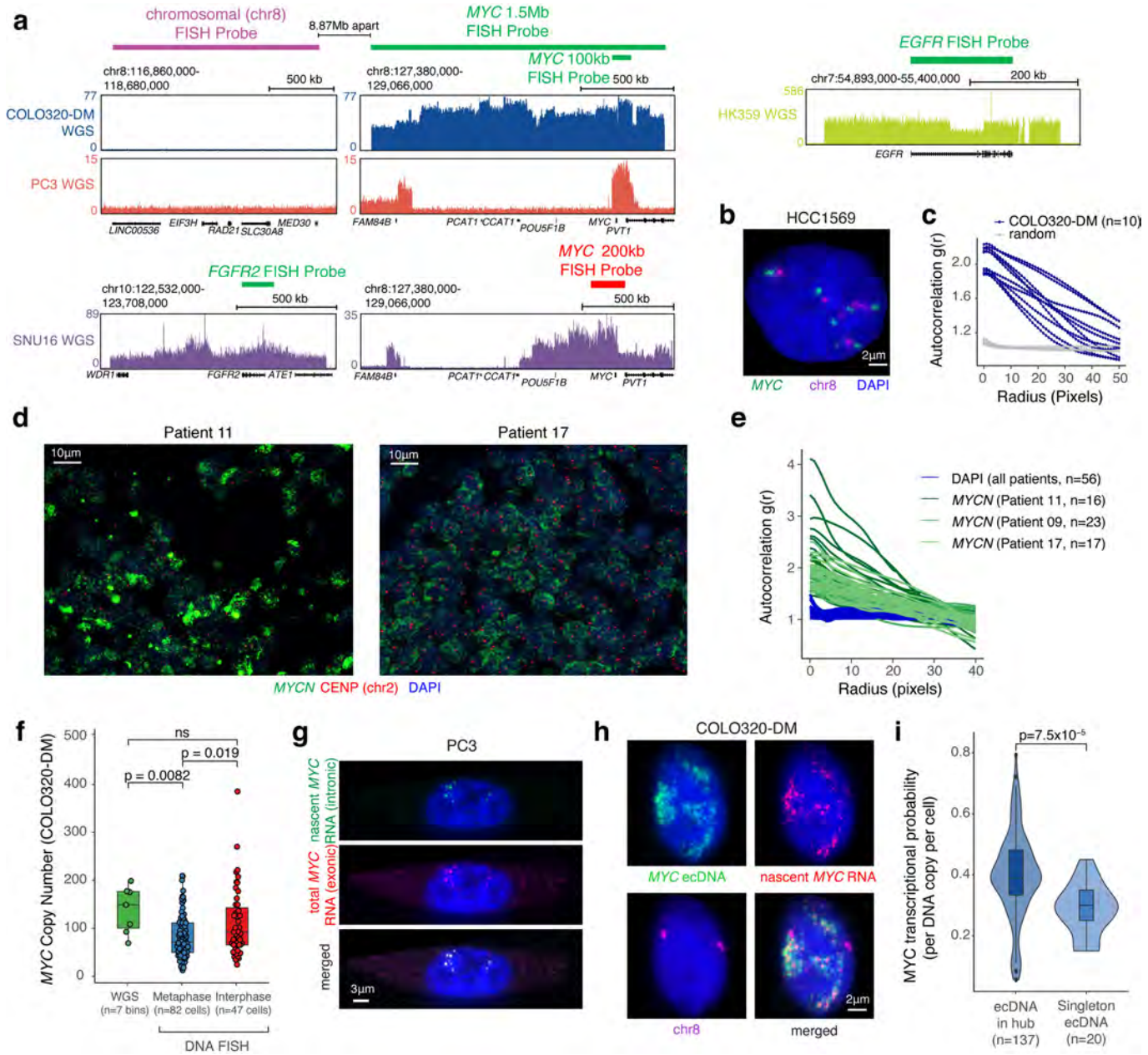
Competing interests H.Y.C. is a co-founder of Accent Therapeutics, Boundless Bio and Cartography Biosciences, and an advisor of 10x Genomics, Arsenal Biosciences and Spring Discovery. P.S.M. is a co-founder of Boundless Bio. He has equity and chairs the scientific advisory board, for which he is compensated. V.B. is a co-founder and advisor of Boundless Bio. A.T.S. is a founder of Immunai and Cartography Biosciences. K.E.Y. is a consultant for Cartography Biosciences.

Additional information

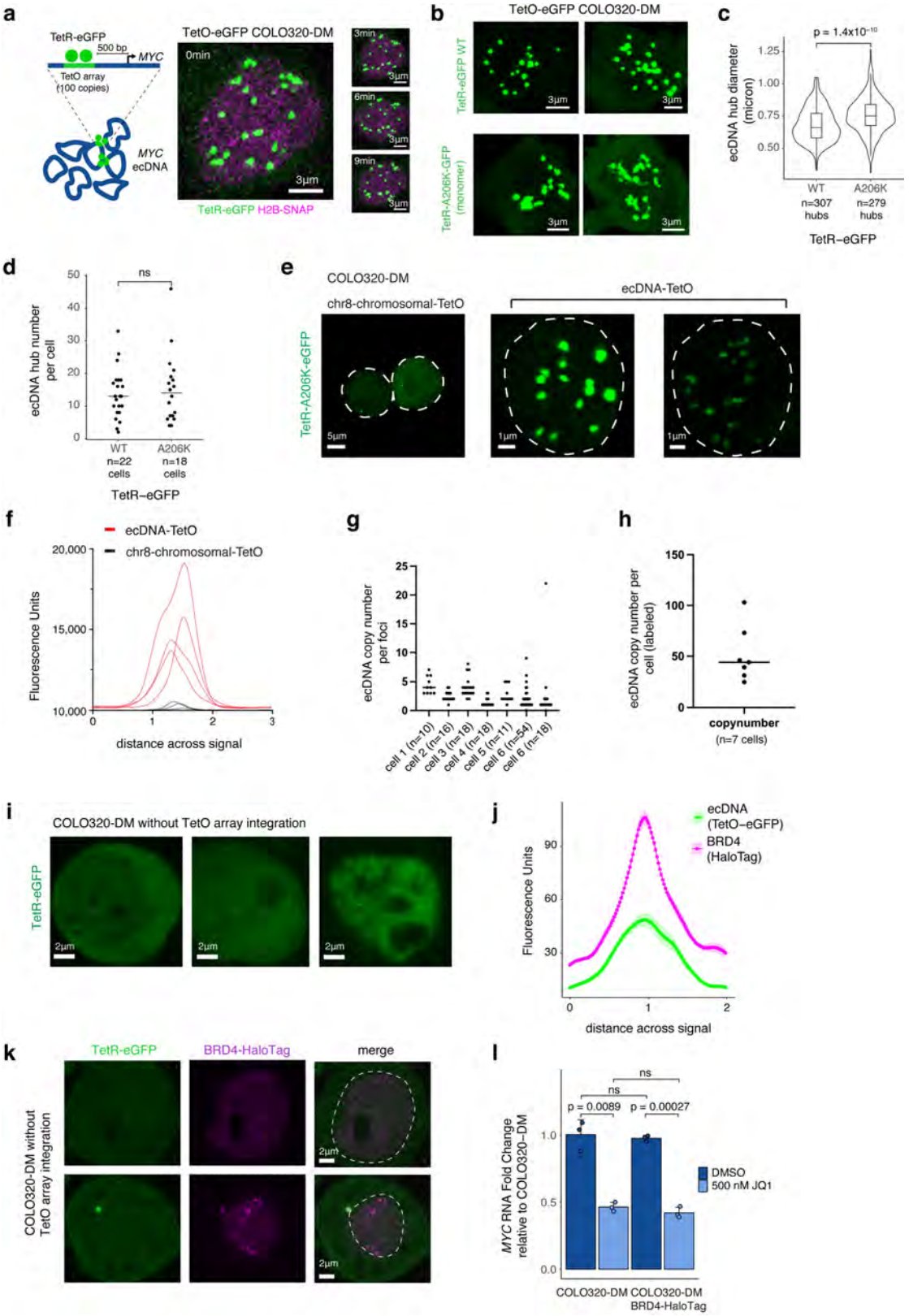
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-04116-8.

Correspondence and requests for materials should be addressed to Howard Y. Chang.

Peer review information *Nature* thanks Charles Lin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

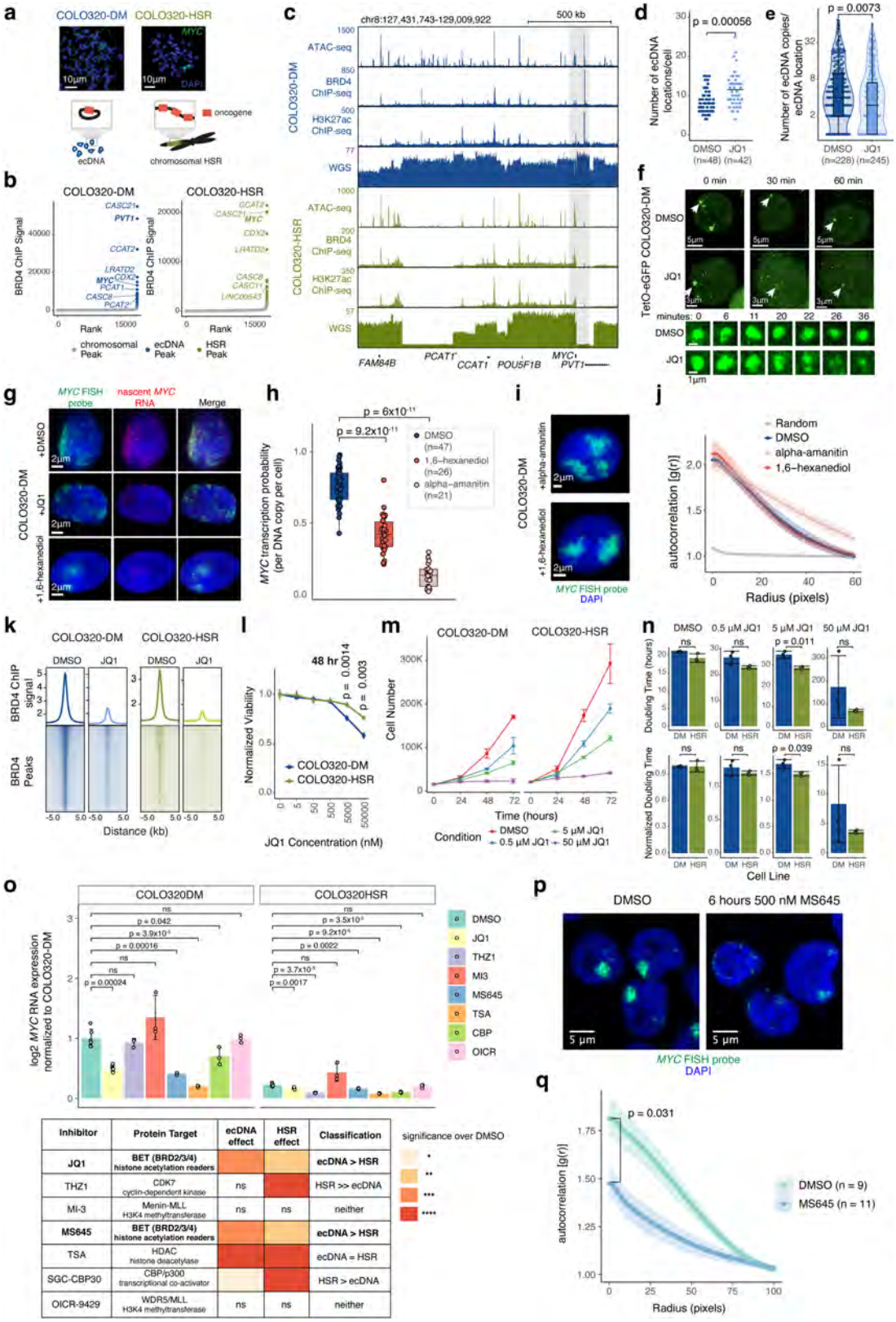


Extended Data Fig. 1 | ecDNA FISH strategies and copy number estimation. **a**, WGS tracks with DNA FISH probe locations. For COLO320-DM and PC3, a 1.5 Mb *MYC* FISH probe (Fig. 1a, b), a 100 kb *MYC* FISH probe (Fig. 1d–f), or a 1.5 Mb chromosome 8 FISH probe was used. Commercial probes were used in SNU16 and HK359 cells. **b**, Representative DNA FISH image using chromosomal and 1.5 Mb *MYC* probes in non-ecDNA amplified HCC1569 showing paired signals as expected from the chromosomal loci. **c**, ecDNA clustering of individual COLO320-DM cells by autocorrelation $g(r)$. **d**, Representative FISH images showing ecDNA clustering in primary neuroblastoma tumours (patients 11 and 17). **e**, ecDNA clustering of individual primary tumour cells from all three patients using autocorrelation $g(r)$. **f**, Comparison of *MYC* copy number in COLO320-DM calculated based on WGS (n=7 genomic bins overlapping with DNA FISH probes), metaphase FISH (n=82 cells) and interphase FISH (n=47 cells). P-values determined by two-sided Wilcoxon test. **g**, Representative images of nascent *MYC* RNA FISH showing overlap of nascent RNA (intronic) and total RNA (exonic) FISH probes in PC3 cells (independently repeated twice). **h**, Representative images from combined DNA FISH for *MYC* ecDNA (100 kb probe) and chromosomal DNA with nascent *MYC* RNA FISH in COLO320-DM cells (independently repeated four times). **i**, *MYC* transcription probability measured by nascent RNA FISH normalized to DNA copy number by FISH comparing singleton ecDNAs to those found in hubs in COLO320-DM (box centre line, median; box limits, upper and lower quartiles; box whiskers, 1.5x interquartile range). To control for noise in transcriptional probability for small numbers of ecDNAs, we randomly re-sampled RNA FISH data grouped by hub size and calculated transcription probability. The violin plot represents transcriptional probability per ecDNA hub based on the hub size matched sampling. P-value determined by two-sided Wilcoxon test.



Extended Data Fig. 2 | See next page for caption.

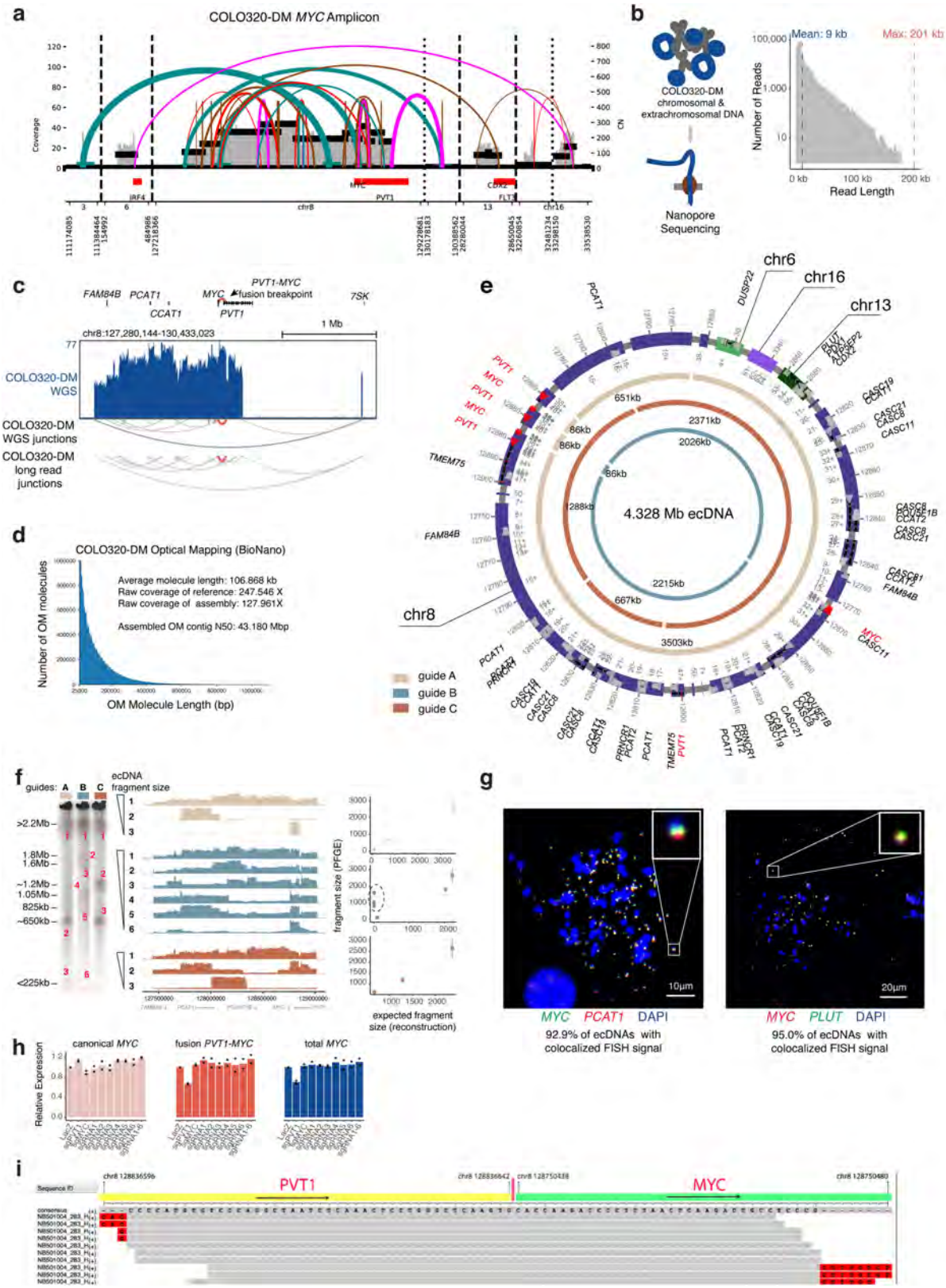
Extended Data Fig. 2 | Generation of TetR-GFP COLO320-DM cells for ecDNA imaging in live cells. a, ecDNA imaging based on TetO array knock-in and labelling with TetR-eGFP (left). Representative images of TetR-eGFP signal in TetO-eGFP COLO320-DM cells at indicated timepoints in a time course (right; independently repeated twice). **b**, GFP signal in ecDNA-TetO COLO320-DM cells. TetR-eGFP and monomeric TetR-eGFP(A206K)-labelled ecDNA hubs appear to be smaller in living cells than in DNA FISH studies of fixed cells, probably because the TetO array is not integrated in all ecDNA molecules and there are potential differences caused by denaturation during DNA FISH and eGFP dimerization. **c**, ecDNA hub diameter in microns (box centre line, median; box limits, upper and lower quartiles; box whiskers, 1.5x interquartile range). Tet-eGFP-labelled hubs are slightly smaller than monomeric TetR-eGFP(A206K)-labelled hubs, potentially due to eGFP dimerization effects (Methods). P-value determined by two-sided Wilcoxon test. **d**, ecDNA hub number per cell. Line represents median. P-value determined by two-sided Wilcoxon test. **e**, TetR-eGFP signal in chr8-chromosomal-TetO (chr8:116,860,000–118,680,000, left) and ecDNA-TetO (TetO-eGFP COLO320-DM, right) COLO320-DM cells. **f**, Fluorescence intensity for chr8-chromosomal-TetO and ecDNA-TetO foci. **g**, **h**, Inferred ecDNA copy number per foci (g; n = number of foci/cell) and per cell (h; n = number of cells) for ecDNA-TetO labelled cells based on summed fluorescence intensity relative to chr8-chromosomal-TetO foci. Line represents median. **i**, Representative images of TetR-GFP signal in parental COLO320-DM without TetO array integration which shows minimal TetR-GFP foci. **j**, Mean fluorescence intensities for ecDNA (TetO-eGFP) and BRD4 (HaloTag) foci across a line drawn across the centre of the largest ecDNA (TetO-eGFP) signal. Data are mean ± SEM for n=5 ecDNA foci. **k**, Representative image of TetR-eGFP signal in COLO320-DM cells without TetO array integration overlaid with BRD4-HaloTag signal. Dashed line indicates nucleus boundary. We noted cytoplasmic TetR-eGFP signal in a subset of COLO320-DM cells without TetO array integration but it did not colocalize with BRD4-HaloTag. **l**, *MYC* RNA measured by RT–qPCR for parental COLO320-DM and BRD4-HaloTag COLO320-DM cells treated with DMSO or 500 nM JQ1 for 6 h which shows similar levels of *MYC* transcription and sensitivity to JQ1 inhibition following epitope tagging of BRD4. Data are mean ± SD between 3 biological replicates. *P* values determined by two-sided Student's *t*-test.



Inhibitor	Protein Target	ecDNA effect	HSR effect	Classification
JQ1	BET (BRD2/3/4) histone acetylation readers			ecDNA > HSR
THZ1	CDK7 cyclin-dependent kinase	ns		HSR >> ecDNA
MI-3	Menin-MLL H3K4 methyltransferase	ns	ns	neither
MS645	BET (BRD2/3/4) histone acetylation readers			ecDNA > HSR
TSA	HDAC histone deacetylase			ecDNA = HSR
SGC-CBP30	CBP/p300 transcriptional co-activator			HSR > ecDNA
OICR-9429	WDR5/MLL H3K4 methyltransferase	ns	ns	neither

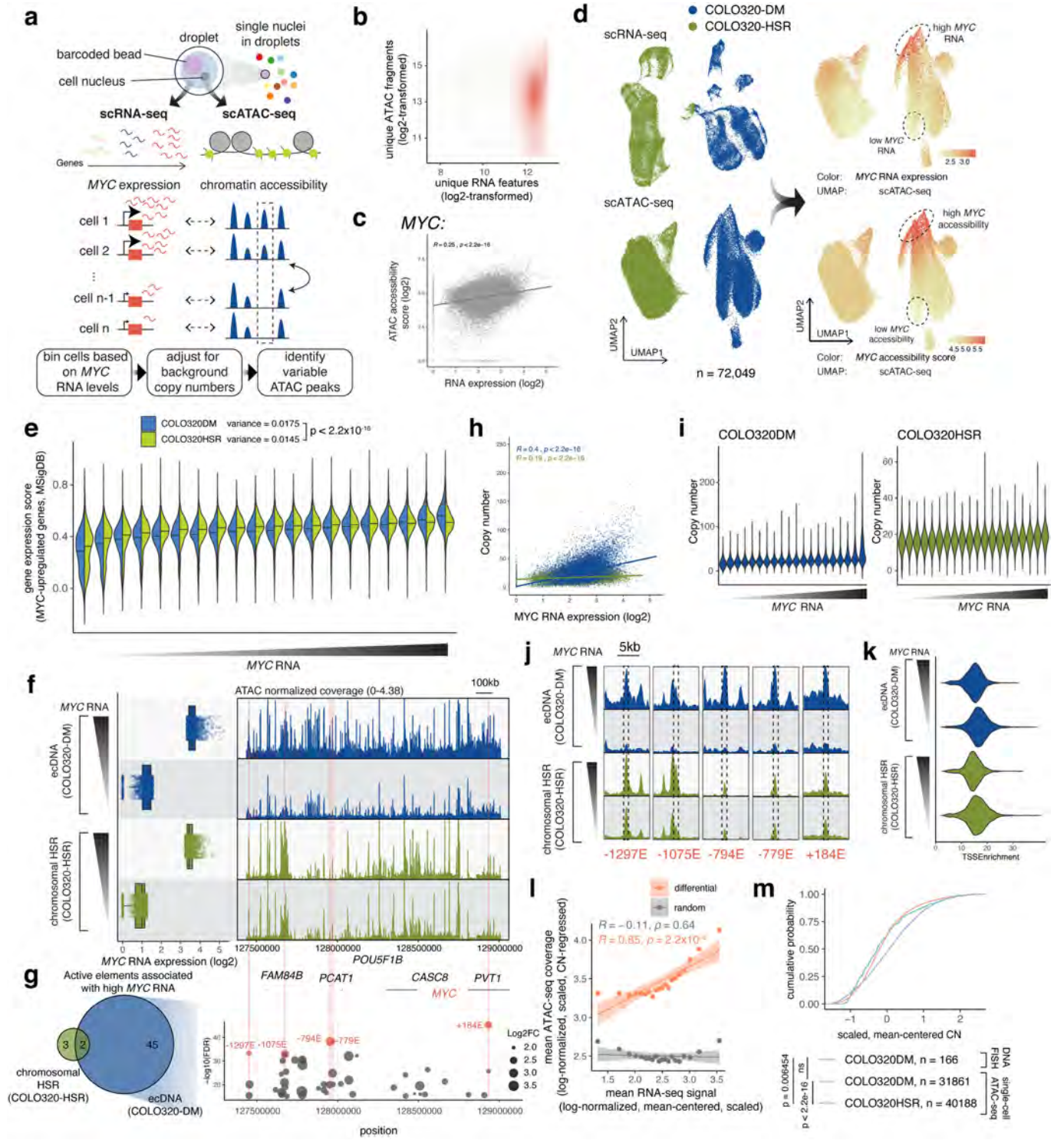
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | BET inhibition leads to ecDNA hub dispersal. **a**, Representative metaphase FISH images and schematic showing ecDNA in COLO320-DM and chromosomal HSRs in COLO320-HSR (independently repeated twice for COLO320-DM and not repeated for COLO320-HSR). **b**, Ranked BRD4 ChIP–seq signal. Peaks in ecDNA or HSR amplifications are highlighted and labelled with nearest gene. **c**, ATAC–seq, BRD4 ChIP–seq, H3K27ac ChIP–seq and WGS at amplified MYC locus. **d**, Number of ecDNA locations (including ecDNA hubs with >1 ecDNA and singleton ecDNAs) from interphase FISH imaging for individual COLO320-DM cells after treatment with DMSO or 500 nM JQ1 for 6 h. N = number of cells quantified per condition. P-value determined by two-sided Wilcoxon test. **e**, ecDNA copies in each ecDNA location from interphase FISH imaging in COLO320-DM after treatment with DMSO or 500 nM JQ1 for 6 h (box centre line, median; box limits, upper and lower quartiles; box whiskers, 1.5x interquartile range). N = number of ecDNA locations quantified per condition. P-value determined by two-sided Wilcoxon test. **f**, Representative live images of TetR-eGFP-labelled ecDNA after treatment with DMSO or 500 nM JQ1 at indicated timepoints in a time course (top; independently repeated twice) and ecDNA hub zoom-ins (bottom). **g**, Representative image from combined DNA/RNA FISH in COLO320-DM cells treated with DMSO, 500 nM JQ1, or 1% 1,6-hexanediol for 6 h. **h**, *MYC* transcription probability measured by dual DNA/RNA FISH after treatment with DMSO, 1% 1,6-hexanediol, or 100 µg/mL alpha-amanitin for 6 h (box centre line, median; box limits, upper and lower quartiles; box whiskers, 1.5x interquartile range; n = number of cells). P-values determined by two-sided Wilcoxon test. **i**, Representative DNA FISH images for MYC ecDNA in interphase COLO320-DM treated with either 1% 1,6-hexanediol or 100 µg/mL alpha-amanitin for 6 h. **j**, ecDNA clustering in interphase cells by autocorrelation $g(r)$ for COLO320-DM treated with DMSO, 1% 1,6-hexanediol, or 100 µg/mL alpha-amanitin for 6 h. Data are mean ± SEM (n = 10 cells quantified per condition). **k**, Averaged BRD4 ChIP–seq signal and heat map over all BRD4 peaks for cells treated with DMSO or 500 nM JQ1 for 6 h. **l**, Cell viability measured by ATP levels (CellTiterGlo) after treatment with different JQ1 concentrations for 48 h normalized to DMSO-treated cells. Data are mean ± SD between 3 biological replicates. *P* values determined by two-sided Student's *t*-test. **m**, Cell proliferation after treatment with different JQ1 concentrations over 72 h. Data are mean ± SD between 3 biological replicates. **n**, Cell doubling times after treatment with different JQ1 concentrations over 72 h in hours (top) or after normalization to DMSO-treated cells (bottom). Data are mean ± SD between 3 biological replicates. *P* values determined by two-sided Student's *t*-test. **o**, *MYC* RNA measured by RT–qPCR after treatment with indicated inhibitors for 6 h (top; each point represents a biological replicate, n=6 for DMSO and JQ1 treatments, n=3 for all other drug treatments). Data are mean ± SD. *P* values determined by two-sided Student's *t*-test. Details of inhibitor panel, protein target, significance of effect on *MYC* transcription, and comparison of effect on ecDNA and HSR transcription (bottom). **p**, **q**, Representative DNA FISH images (**p**) and clustering by autocorrelation $g(r)$ (**q**) for *MYC* ecDNAs in COLO320-DM treated with DMSO or 500 nM MS645 for 6 h. Data are mean ± SEM. P-value determined by two-sided Wilcoxon test at radius = 0.



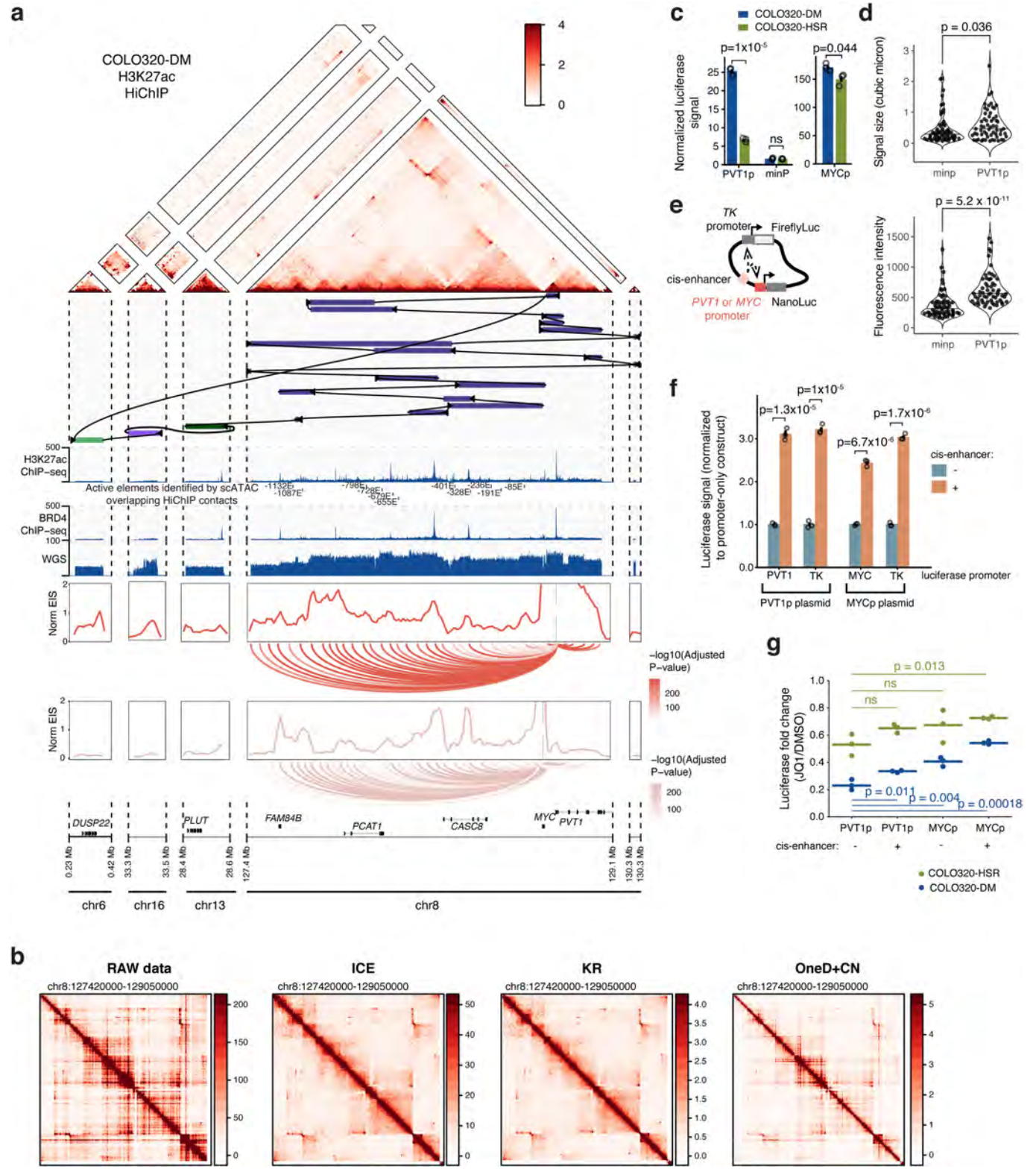
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Reconstruction of COLO320-DM ecDNA amplicon structure. a, Structural variant (SV) view of AmpliconArchitect (AA) reconstruction of the *MYC* amplicon in COLO320-DM cells. **b**, Nanopore sequencing of COLO320-DM cells (left) and distribution of read lengths. **c**, WGS for COLO320-DM with junctions detected by WGS and nanopore sequencing. **d**, Molecule lengths used for optical mapping and statistics. **e**, Reconstructed COLO320-DM ecDNA after integrating WGS, optical mapping, and in-vitro ecDNA digestion. Chromosomes of origin and corresponding coordinates (hg19) are labelled. Three inner circular tracks (light tan, slate and brown in colour; guides A, B and C, respectively) representing expected fragments as a result of Cas9 cleavage using three distinct sgRNAs and their expected sizes. Guide sequences are in Supplementary Table 2 (PFGE_guide_A-C). **f**, In-vitro Cas9 digestion of COLO320-DM ecDNA followed by PFGE (left). Fragment sizes were determined based on *H. wingei* and *S. cerevisiae* ladders. Uncropped gel image is in Supplementary Fig. 1. Middle panel shows short-read sequencing of the *MYC* ecDNA amplicon for all isolated fragments, ordered by fragment size. Right panel shows concordance of expected fragment sizes by optical mapping reconstruction, and observed fragment sizes by in-vitro Cas9 digestion (discordant fragments circled). Each sgRNA digestion was performed in one independent experiment. **g**, Metaphase FISH images showing colocalization of *MYC*, *PCAT1* and *PLUT* as predicted by optical mapping and in-vitro digestion. N = 20 cells and 1,270 ecDNAs quantified for *MYC*/*PCAT1* DNA FISH and n = 15 cells and 678 ecDNAs for *MYC*/*PLUT* DNA FISH from one experiment. **h**, RNA expression measured by RT–qPCR for indicated transcripts in COLO320-DM cells stably expressing dCas9-KRAB and indicated sgRNAs (n=2 biological replicates). Canonical *MYC* was amplified with primers MYC_exon1_fw and MYC_exon2_rv; fusion *PVT1-MYC* was amplified with PVT1_exon1_fw and MYC_exon2_rv; total MYC was amplified with total_MYC_exon2_fw and total_MYC_exon2_rv. All primer sequences are in Supplementary Table 1 and guide sequences are in Supplementary Table 2. **i**, Alignment of junction reads at the *PVT1-MYC* breakpoint.



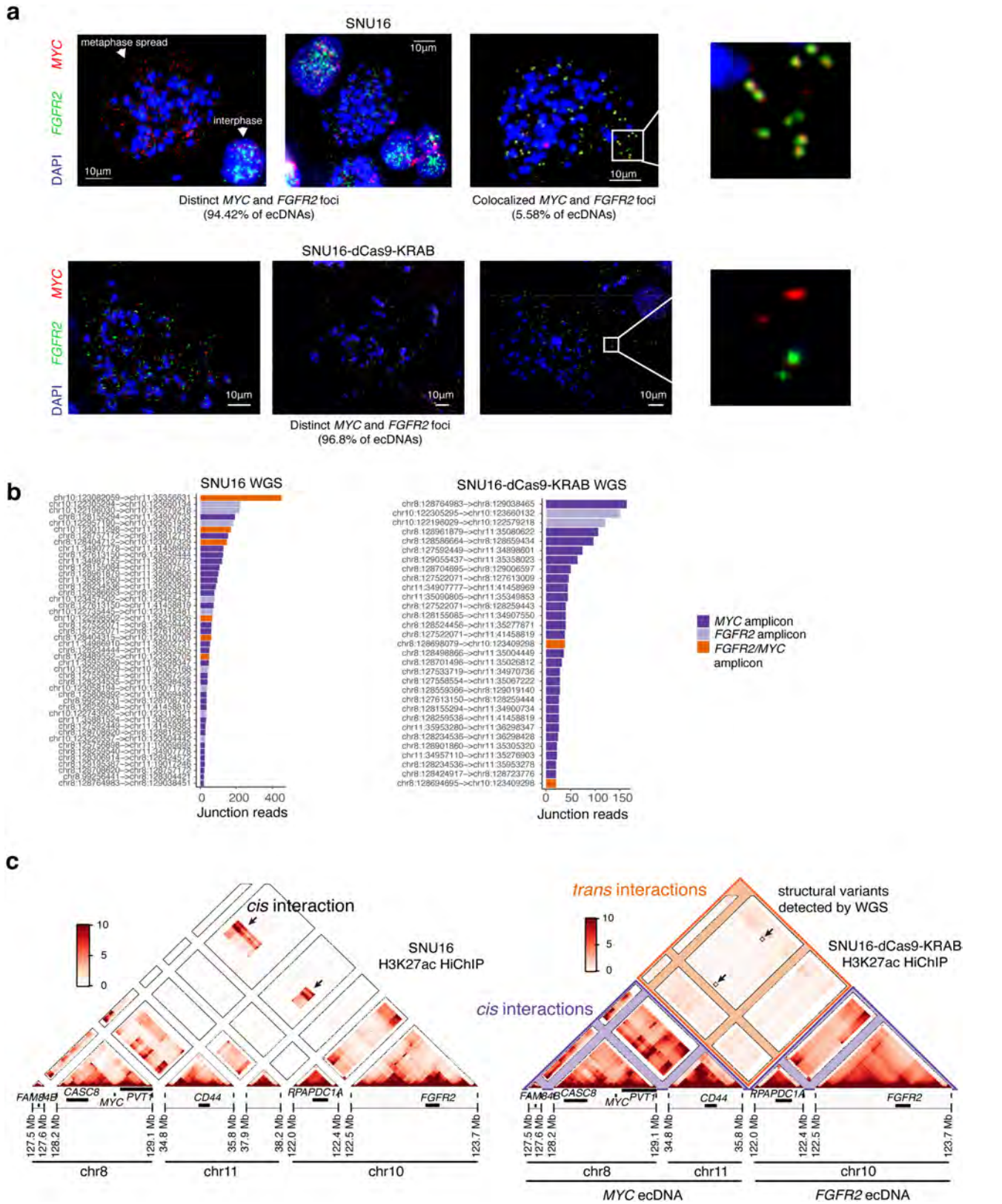
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Single-cell multiomic analysis reveals combinatorial and heterogeneous ecDNA regulatory element activities associated with *MYC* expression. a, Joint single-cell RNA and ATAC–seq for simultaneously assaying gene expression and chromatin accessibility and identifying regulatory elements associated with *MYC* expression. **b**, Unique ATAC–seq fragments and RNA features for cells passing filter (both log2-transformed). **c**, Correlation between *MYC* accessibility score and normalized RNA expression. **d**, UMAP from the RNA or the ATAC–seq data (left). Log-normalized and scaled *MYC* RNA expression (top right) and *MYC* accessibility scores (bottom right) were visualized on the ATAC–seq UMAP, showing cell-level heterogeneity in *MYC* RNA-seq and ATAC-seq signals in ecDNA-containing COLO320-DM. **e**, Gene expression scores (calculated using Seurat in R) of *MYC*-upregulated genes (Gene Set M6506, Molecular Signatures Database; MSigDB) across all *MYC* RNA quantile bins. Horizontal line marks median. Population variances for all individual cells are shown (top). P-value determined by two-sided F-test. **f**, *MYC* expression levels of top and bottom bins (left). Normalized ATAC–seq coverages are shown (right). **g**, Number of variable elements identified on COLO320-DM ecDNAs compared to chromosomal HSRs in COLO320-HSR (left). 45 variable elements were uniquely observed on ecDNA. All variable elements on ecDNA are shown on the right (y-axis shows -log10(FDR) and dot size represents log2 fold change. Five most significantly variable elements are highlighted and named based on relative position in kb to the *MYC* TSS (negative, 5′; positive, 3′). **h**, Correlation between estimated *MYC* copy numbers and normalized log2-transformed *MYC* expression of all individual cells showing a high level of copy number variability associated with increased expression, in particular for COLO320-DM. **i**, Estimated *MYC* amplicon copy number of all cell bins separated by *MYC* RNA expression. **j**, Zoom-ins of the ATAC–seq coverage of each of the five most significantly variable elements identified in **g** (marked by dashed boxes). **k**, Similar distributions of TSS enrichment in the high and low cell bins, indicating differences in accessibility at variable elements are not an artifact of differences in data quality. **l**, Mean copy number regressed, log-normalized, scaled ATAC–seq coverage of the differential peaks against mean *MYC* RNA (log-normalized, mean-centred, scaled) for each cell bin in orange. Same number of random non-differential peaks from the same amplicon interval and shown in grey. Error bands show 95% confidence intervals for the linear models. **m**, Cumulative probability of *MYC* amplicon copy number distributions (mean-centred, scaled) of single-cell ATAC–seq data and DNA FISH data, suggesting that copy number estimates from single cell ATAC-seq data reflect heterogeneity in ecDNA copy number as measured by DNA FISH. P-values determined by Kolmogorov-Smirnov test (1,000 bootstrap simulations).



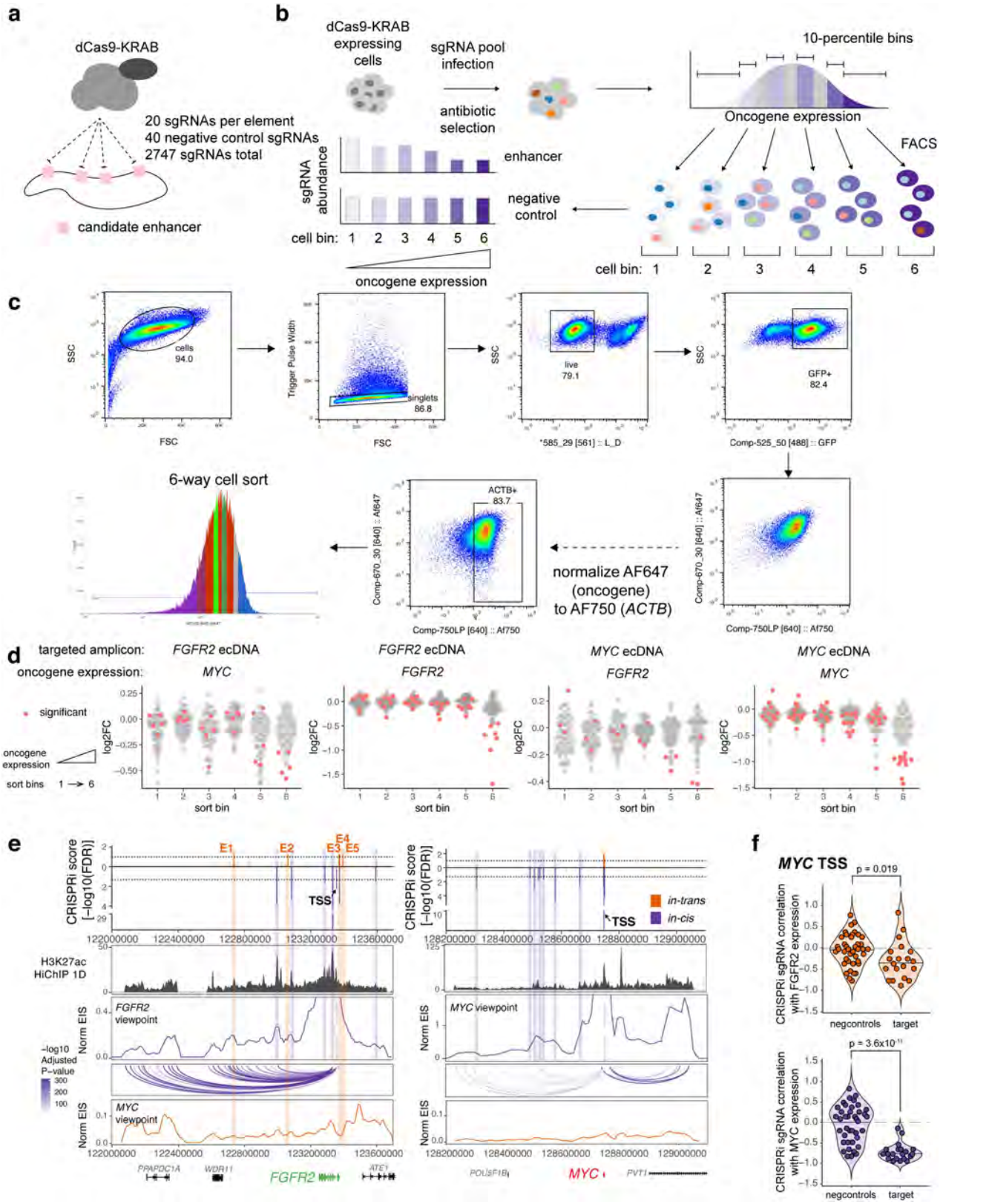
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Endogenous enhancer connectome of COLO320-DM *MYC* ecDNA amplicon and effect of promoter sequence, *cis* enhancers, and BET inhibition on episomal reporter activation. a, Top to bottom: COLO320-DM H3K27ac HiChIP contact map (KR-normalized read counts, 10-kb resolution), reconstructed COLO320-DM amplicon, H3K27ac ChIP–seq signal, BRD4 ChIP–seq signal, WGS coverage, interaction profile of *PVT1* (top, dark pink) and *MYC* (bottom, light pink) promoters at 10-kb resolution with FitHiChIP loops shown below, coloured by adjusted p-value. Active elements identified by scATAC and overlapping H3K27ac HiChIP contacts named by genomic distance to MYC start site: −1132E, −1087E, −679E, −655E, −401E, −328E, −85E. **b**, Comparison of HiChIP matrix normalization methods for COLO320-DM H3K27ac HiChIP at 10-kb resolution. HiChIP signal is robust to different normalization methods. **c**, Quantification of NanoLuc luciferase signal for plasmids with *PVT1p*-, *minp*-, or *MYCp*-driven NanoLuc reporter expression. Luciferase signal was calculated by normalizing NanoLuc readings to Firefly readings. Bar plot shows mean ± SEM. *P* values were calculated using a two-sided Student's *t*-test (n=3 biological replicates). **d**, Violin plots showing mean fluorescence intensities and signal sizes of the NanoLuc reporter RNA in PVT1p-reporter and minp-reporter transfected cells. P-values were calculated using a two-sided Wilcoxon test. **e**, Schematic of *PVT1* promoter-driven luciferase reporter plasmid with a *cis*-enhancer. Details of *cis*-enhancer are in Methods. **f**, Bar plot showing luciferase signal driven by PVT1p, MYCp or the constitutive TKp with or without a cis-enhancer (mean ± SEM). All values are normalized to the corresponding promoter-only construct without a cis-enhancer. *P* values were calculated using a two-sided Student's *t*-test (n=3 biological replicates). **g**, Dot plots showing fold change in luciferase signal (Firefly-normalized NanoLuc signal) in JQ1-treated over DMSO-treated COLO320-DM and COLO320-HSR cells after transfection with the PVT1p or the MYCp plasmid with or without a cis-enhancer. *P* values were calculated using a two-sided Student's *t*-test (n=3 biological replicates).



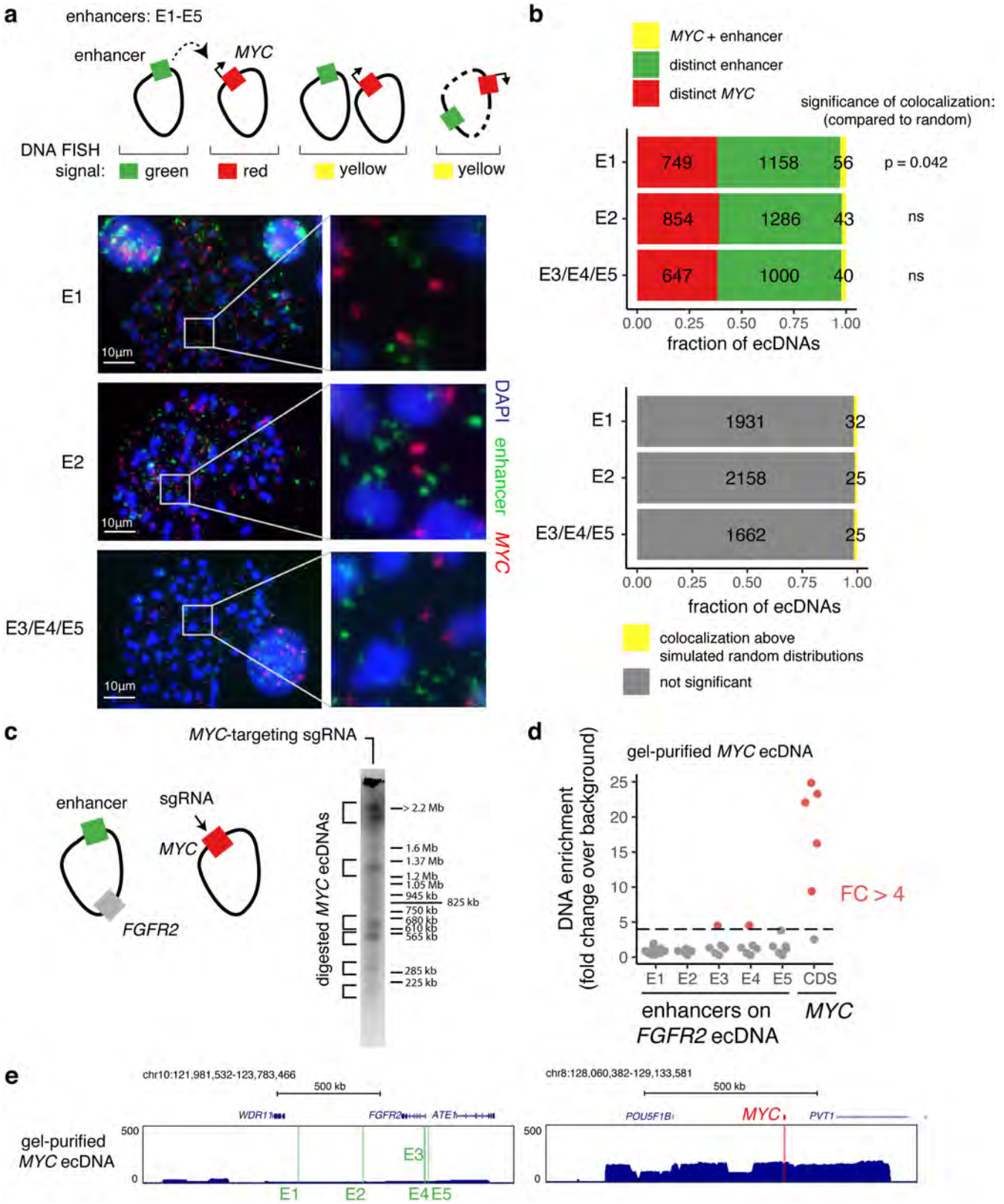
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Generation of monoclonal SNU16-dCas9-KRAB with reduced ecDNA fusions. a, Representative DNA FISH images showing extrachromosomal single-positive *MYC* and *FGFR2* amplifications (top left and top middle) and double-positive *MYC* and *FGFR2* amplifications in metaphase spreads in parental SNU16 cells (top right) with zoom in (top right). N = 42 cells and 8,222 ecDNAs. Representative DNA FISH images showing distinct extrachromosomal *MYC* and *FGFR2* amplifications in metaphase spreads in SNU16-dCas9-KRAB cells (bottom). N = 29 cells and 3,893 ecDNAs. **b**, Ranked plot showing number of junction reads supporting each breakpoint in AmpliconArchitect. Breakpoints are coloured based on whether they span regions from the same amplicon (*MYC*/*FGFR2*) or regions from two distinct amplicons. **c**, HiChIP contact matrices at 10-kb resolution with KR normalization for parental SNU16 cell line (left) and SNU16-dCas9-KRAB cell line (right). Contact matrix for parental cells contains regions of increased *cis*-contact frequency between chr8 and chr10 as indicated, as compared to SNU16-dCas9-KRAB cells with highly reduced contact frequency between chr8 and chr10. Regions of increased focal interaction overlapping low frequency structural rearrangements between chr8 and chr10 described in **b** indicated with boxes.



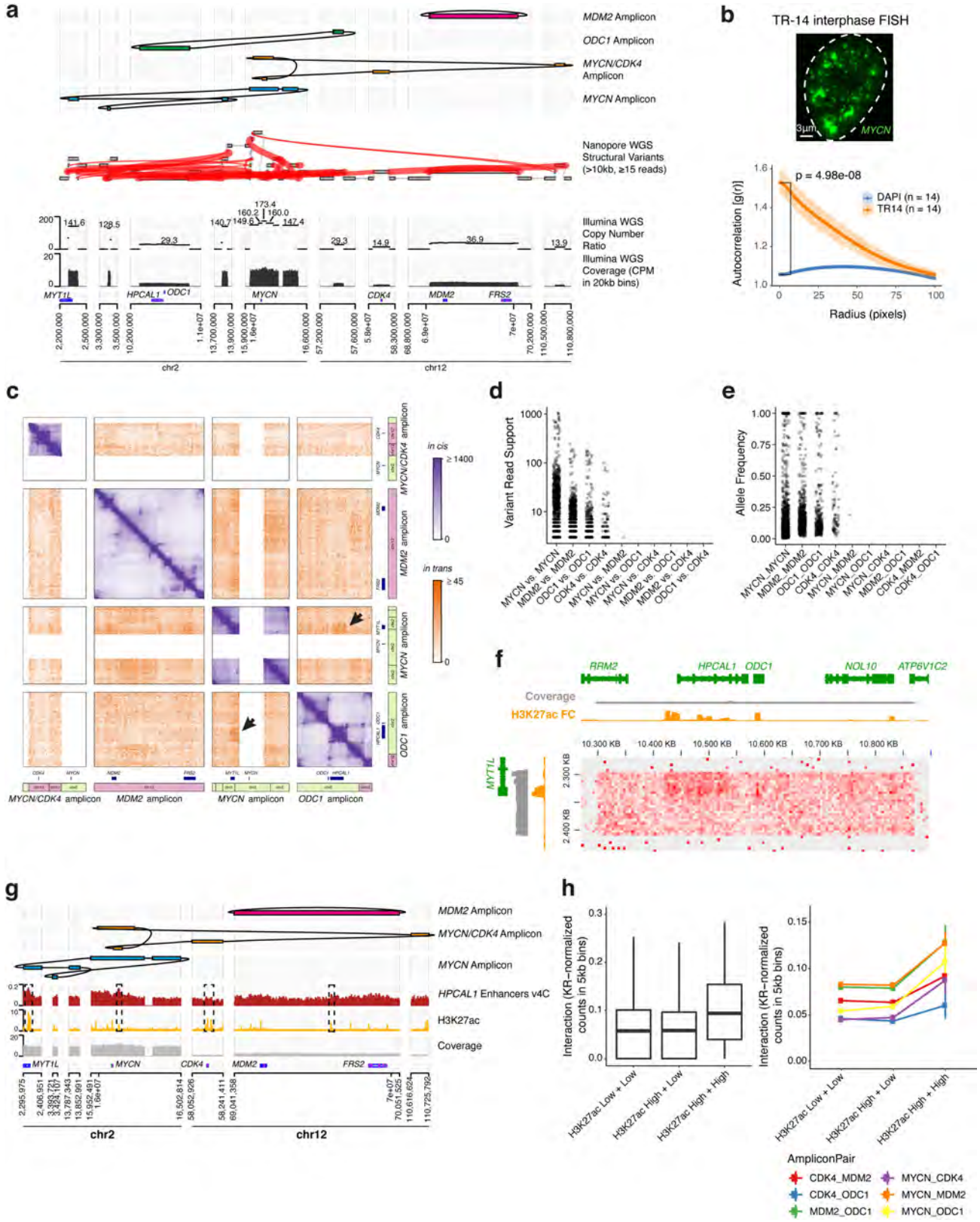
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Perturbations of ecDNA enhancers by CRISPRi reveal functional intermolecular enhancer–gene interactions. a, CRISPRi experiments perturbing candidate enhancers in SNU16-dCas9-KRAB cells. Single-guide RNAs (sgRNAs) were designed to target candidate enhancers on *FGFR2* and *MYC* ecDNAs based on chromatin accessibility. **b**, Experimental workflow for pooled CRISPRi repression of putative enhancers. Stable SNU16-dCas9-KRAB cells were generated from a single cell clone. Cells were transduced with a lentiviral pool of sgRNAs, selected with antibiotics and oncogene RNA was assessed by flowFISH. Cells were sorted into six bins by fluorescence-activated cell sorting (FACS) based on oncogene expression. sgRNAs were quantified for cells in each bin. **c**, FACS gating strategy. **d**, Log2 fold changes of sgRNAs for each candidate enhancer element compared to unsorted cells for CRISPRi libraries targeting either *MYC* or *FGFR2* ecDNAs, followed by cell sorting based on expression levels of *MYC* or *FGFR2*. Each dot represents the mean log2 fold change of 20 sgRNAs targeting a candidate element. Elements negatively correlated with oncogene expression as compared to the negative control sgRNA distributions in the same pools are marked in red. **e**, Bar plot showing significance of CRISPRi repression of candidate enhancer elements as in Fig. 4e (top). Significant in-*trans* and in-*cis* enhancers are coloured as indicated. SNU16-dCas9-KRAB H3K27ac HiChIP 1D signal track and interaction profiles of *FGFR2* and *MYC* promoters at 10-kb resolution with *cis* FitHiChIP loops shown below. Interaction profiles in *cis* shown in purple and in *trans* shown in orange. **f**, Spearman correlations of individual sgRNAs that target MYC TSS across fluorescence bins corresponding to *MYC* and *FGFR2* expression. P values using the lower-tailed t-test comparing target sgRNAs with negative control sgRNAs (negcontrols) are shown. Each dot represents an independent sgRNA.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Intermolecular enhancers and *MYC* are located on distinct molecules for the vast majority of ecDNAs. a, Top: two-colour DNA FISH on metaphase spreads for quantifying the frequency of colocalization of the *MYC* gene and intermolecular enhancers shown in Fig. 4e. Above-random colocalization would indicate fusion events. Bottom: representative DNA FISH images. DNA FISH probes target the following hg19 genomic coordinates: E1, chr10:122,635,712–122,782,544 (RP11-95I16; n = 11 cells); E2, chr10:122,973,293–123,129,601 (RP11-57H2; n = 12 cells); E3/E4/E5, chr10:123,300,005–123,474,433 (RP11-1024G22; n = 10 cells). **b**, Top: numbers of distinct and colocalized FISH signals. To estimate random colocalization, 100 simulated images were generated with matched numbers of signals and mean simulated frequencies were compared with observed colocalization. P values determined by two-sided t-test (Bonferroni-adjusted). Bottom: number of colocalized signals significantly above random chance. Colocalization above simulated random distributions is the sum of colocalized molecules in excess of random means in all FISH images in which total colocalization was above the random mean plus 95% confidence interval (100 simulated images per FISH image). **c**, In vitro Cas9 digestion of *MYC*-containing ecDNA in SNU16-dCas9-KRAB followed by PFGE (one independent experiment). Fragment sizes were determined based on *H. wingei* and *S. cerevisiae* ladders. Uncropped gel image is in Supplementary Fig. 1. *MYC* CDS guide corresponds to guide B in Supplementary Table 2. **d**, Enrichment of enhancer DNA sequences in isolated *MYC* ecDNAs bands from **c** over background (DNA isolated from a separate PFGE lane in the corresponding size range resulting from undigested genomic DNA) based on normalized reads in 5kb windows. Each dot represents DNA from a distinct gel band. Red indicates fold change above 4. **e**, Sequencing track for a gel-purified *MYC* ecDNA showing enrichment of the *MYC* amplicon and depletion of the *FGFR2* amplicon containing enhancers E1-E5.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Reconstruction of four distinct amplicons in TR14 neuroblastoma cell line and intermolecular amplicon interaction patterns associated with H3K27ac marks. a, Top to bottom: long read-based reconstruction of four different amplicons; genome graph with long read-based structural variants of >10kb size and >20 supporting reads indicated by red edges; copy number variation and coverage from short-read whole-genome sequencing, positions of the selected genes. **b**, A representative DNA FISH image of *MYCN* ecDNAs in interphase TR14 cells (top) and ecDNA clustering compared to DAPI control in the same cells assessed by autocorrelation $g(r)$ (bottom). Data are mean ± SEM (n = 14 cells). **c**, Custom Hi-C map of reconstructed TR14 amplicons. The *MYCN*/*CDK4* amplicon and the *MYCN* ecDNA share sequences, which prevented an unambiguous short-read mapping in these regions and appear as white areas. *Trans* interactions appear locally elevated between *MYCN* ecDNA and *ODC1* amplicon (indicated by arrows). *Cis*- and *trans*-contact frequencies are coloured as indicated. **d**, Read support for structural variants identified by long read sequencing linking amplicons. Only one structural variant between distinct amplicons (*MYCN* and *MDM2* amplicons) was identified with 3 supporting reads. **e**, Variant allele frequency for structural variants linking amplicons. **f**, *Trans*-interaction pattern between enhancers on a *MYCN* amplicon fragment (vertical) and an *ODC1* amplicon fragment (horizontal). Short-read WGS coverage (grey), H3K27ac ChIP–seq track showing mean fold change over input in 1kb bins (yellow) and Hi-C contact map showing (KR-normalized counts in 5kb bins). **g**, Top to bottom: three amplicon reconstructions, virtual 4C interaction profile of the enhancer-rich *HPCAL1* locus on the *ODC1* amplicon with loci on other amplicons (red), and H3K27ac ChIP–seq (fold change over input; yellow). **h**, *Trans* interaction between different amplicons (KR-normalized counts in 5kb bins) depending on H3K27ac signal of the interaction loci (left; box centre line, median; box limits, upper and lower quartiles; box whiskers, 1.5x interquartile range). *Trans* interaction (KR-normalized counts in 5kb bins) separated by amplicon pair (right). H3K27ac High vs. Low denotes at least vs. less than 3-fold mean enrichment over input in 5kb bins. N = 114,636 H3K27ac Low + Low pairs, n = 11,990 H3K27ac High + Low pairs, n = 296 H3K27ac High + High pairs.

nature research

Corresponding author(s): Howard Y. Chang

Last updated by author(s): Sep 21, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Whole Genome Sequencing
Whole genome sequencing (WGS) data from COLO320-DM, COLO320-HSR and PC3 cells were generated by a previously published study and raw fastq reads obtained from the NCBI Sequence Read Archive, under BioProject accession PRJNA506071. Reads were trimmed of adapter content with Trimmomatic (version 0.39), aligned to the hg19 genome using bwa mem (0.7.17-r1188), and PCR duplicates removed using Picard's MarkDuplicates. WGS data from SNU16 cells was generated by a previously published study and aligned reads in bam format from the NCBI Sequence Read Archive, under BioProject accession PRJNA523380. WGS data from HK359 cells was generated by a previously published study and aligned reads in bam format obtained from the NCBI Sequence Read Archive, under BioProject accession PRJNA338012.

ChIP-seq Data Processing
Paired-end reads were aligned to the hg19 genome using Bowtie264 (version 2.3.4.1) with the --very-sensitive option following adapter trimming with Trimmomatic59 (version 0.39). Reads with MAPQ values less than 10 were filtered using samtools (version 1.9) and PCR duplicates removed using Picard's MarkDuplicates (version 2.20.3-SNAPSHOT).

COLO320-DM WGS Data Processing
Reads were aligned to Homo sapiens genome (hg19) using BWA aligner version 0.7.13 (https://github.com/lh3/bwa) with default settings.

COLO320-DM Nanopore Data Processing
Bases were called from fast5 files using guppy (Oxford Nanopore Technologies, version 2.3.7). Reads were then aligned using NGMLR (version 0.2.7) with the following parameters: -x ont --no-lowqualitysplit.

COLO320-DM Optical Mapping Data Processing
De novo assemblies of the samples were performed with Bionano's de novo assembly pipeline (Bionano Solve v3.6) using standard haplotype aware arguments. With the Overlap-Layout-Consensus paradigm, pairwise comparison of DNA molecules having 248X coverage against the reference was used to create a layout overlap graph, which was then used to generate the initial consensus genome maps. By realigning

molecules to the genome maps (P value cut off of <10-12) and by using only the best matched molecules, a refinement step was done to refine the label positions on the genome maps and to remove chimeric joins. Next, during an extension step, the software aligned molecules to genome maps (P<10-12), and extended the maps based on the molecules aligning past the map ends. Overlapping genome maps were then merged (P<10-16). These extension and merge steps were repeated five times before a final refinement (P<10-12) was applied to "finish" all genome maps.

RNA-seq Data Processing
Paired-end reads were aligned to the hg19 genome using STAR-Fusion (version 1.6.0) and the genome build GRCh37_gencode_v19_CTAT_lib_Mar272019.plug-n-play. Number of reads supporting the PVT1-MYC fusion transcript were obtained from the "star-fusion.fusion_predictions.abridged.tsv" output file and the junction read counts and spanning fragment counts were combined. Reads supporting the canonical MYC exon 1-2 junction were obtained using the Gviz (version 1.30.3) package in R (version 3.6.1) in a sashimi plot.

Single-cell RNA and ATAC-seq Data Processing
A custom reference package for hg19 was created using cellranger-arc mkref (10x Genomics, version 1.0.0). The single-cell paired RNA and ATAC-seq reads were aligned to the hg19 reference genome using cellranger-arc count (10x Genomics, version 1.0.0).

HiChIP Data Processing
HiChIP data were processed as described previously. Briefly, paired end reads were aligned to the hg19 genome using the HiC-Pro pipeline (version 2.11.0). Default settings were used to remove duplicate reads, assign reads to MboI restriction fragments, filter for valid interactions, and generate binned interaction matrices.

SNU16-dCas9-KRAB WGS Data Processing
Reads were trimmed of adapter content with Trimmomatic (version 0.39), aligned to the hg19 genome using bwa mem (0.7.17-r1188), and PCR duplicates removed using Picard's MarkDuplicates (version 2.20.3-SNAPSHOT).

ATAC-seq Data Processing
Adapter-trimmed reads were aligned to the hg19 genome using Bowtie2 (2.1.0). Aligned reads were filtered for quality using samtools (version 1.9) and duplicate fragments were removed using Picard (version 2.21.9-SNAPSHOT).

TR14 WGS Data Processing
Adapters were trimmed with BBMap 38.58. Reads were then aligned to hg19 using BWA-MEM 0.7.15 with default parameters and duplicate reads were removed (Picard 2.20.4).

TR14 Nanopore Data Processing
Reads were aligned to hg19 using NGMLR v0.2.7.

Hi-C Data Processing
FASTQ files were processed using the Juicer pipeline v1.19.02, CPU version, which was set up with BWA v0.7.17 to map short reads to reference genome hg19, from which haplotype sequences were removed and to which the sequence of Epstein-Barr virus (NC_007605.1) was added. Replicates were processed individually. Mapped and filtered reads were merged afterwards. A threshold of MAPQ≥30 was applied for the generation of Hi-C maps with Juicer tools v1.7.5.

TR14 ChIP-seq Data Analysis
TR14 H3K27ac ChIP-seq raw data were downloaded from Gene Expression Omnibus (GSE90683). We trimmed adapters with BBMap 38.58 and aligned the reads to hg19 using BWA-MEM 0.7.15 with default parameters.

Data analysis

Metaphase DNA FISH Image Analysis
Colocalization analysis for two-color metaphase FISH data for MYC, PCAT1 and PLUT ecDNAs in COLO320-DM described in Extended Data Figure 4g was performed using Fiji (version 2.1.0/1.53c). Images were split into the two FISH colors + DAPI channels, and signal threshold set manually to remove background fluorescence. Overlapping FISH signals were segmented using watershed segmentation. Colocalization was quantified using the ImageJ-Colocalization Threshold program and individual and colocalized FISH signals were counted using particle analysis. Colocalization analysis for two-color metaphase FISH data for MYC and FGFR2 ecDNAs in SNU16 described in Figure 4c and Extended Data Figure 7a was performed using ecSeg (https://github.com/UCRajkumar/ecSeg, not versioned). Briefly, ecSeg takes as input metaphase FISH images containing DAPI and up to two colors of DNA FISH. ecSeg uses the DAPI signal to classify signals as nuclear (arising from interphase nuclei), chromosomal (arising from metaphase chromosome), or extrachromosomal. It then quantifies DNA FISH signal and colocalization segmented by whether the signal is present on chromosomal or extrachromosomal DNA.

Interphase DNA FISH Clustering Analysis
To analyze the clustering of ecDNAs, we applied the autocorrelation function as described previously in Matlab (2019). Colocalization analysis for SNU16 MYC and FGFR2 ecDNAs in Figure 4a was performed using confocal images of both metaphase and interphase nuclei from the same slides. Images were split into the two FISH colors, and background fluorescence was removed manually for each channel. Colocalization for each nucleus was quantified using the ImageJ-Colocalization Threshold program. Analysis was performed across all z-stacks for each nucleus. Manders coefficient (fraction of MYC signal colocalized compared to total MYC signal) was used to quantify colocalization.

ecDNA DNA FISH and nascent RNA FISH Image Analysis
To characterize the ecDNA hub shape and size, we employed the synthetic model—Surfaces object from Imaris (version 9.1, Bitplane) and applied a Gaussian filter (σ = 1 voxel in xy) and background subtraction for optimal segmentation and quantification of ecDNA hubs. ecDNA hubs containing connected voxels were sorted by size and singleton ecDNAs were separated from ecDNA hubs (minimal two ecDNA molecules). To measure the number of ecDNA or nascent transcripts, we localized the voxels corresponding to the local maximum of identified DNA or RNA FISH signal using the Imaris spots function module. We validated the accuracy of interphase ecDNA counting by comparing to quantification of ecDNA number by metaphase FISH as well as copy number estimated by whole genome sequencing (Extended Data Figure 1f). The copy number distribution from whole genome sequencing is comparable to that from interphase DNA FISH. While copy number estimates from WGS and interphase FISH are slightly higher than those quantified by metaphase FISH imaging, this may reflect the fact that individual ecDNAs can contain multiple copies of MYC.

ChIP-seq Data Analysis
MACS2 (version 2.1.1.20160309) was used for peak calling with the following parameters: macs2 callpeak -t chip_bed -c input_bed -n output_file -f BED -g hs -q 0.01 --nomodel --shift 0. A reproducible peak set across biological replicates was defined using the IDR framework (version 2.0.4.2). Reproducible peaks from all samples were then merged to create a union peak set. ChIP-seq signal was converted to bigwig format for visualization using deepTools bamCoverage (version 3.3.1) with the following parameters: --bs 5 --smoothLength 105 --normalizeUsing CPM --scaleFactor 10. Enrichment of ChIP signal at peaks was performed using deepTools computeMatrix on ChIP signal in bigwig format containing the ratio of BRD4 ChIP signal over input calculated using deepTools bamCoverage (version 3.3.1) with the following parameters: --operation ratio --bs 5 --smoothLength 105.

COLO320-DM Nanopore sequencing Data Analysis
Structural variants were called using Sniffles (version 1.0.11) using the following parameters: -s 1 --report_BND --report_seq.

COLO320-DM reconstruction strategy
Due to the large size of the COLO320DM ecDNA (4.3 Mbp), we used a scaffolding strategy based on manual combination of results from multiple data sources. All data which required alignment back to a reference genome used hg19. The first source of data used was the copy-number aware breakpoint graph detected by AmpliconArchitect (version 1.2) (AA) generated from low-coverage WGS data. The AA graph specified copy-numbers of amplicon segments as well as genomic breakpoints between them. AA was run with default settings and seed regions were identified using the PrepareAA pipeline (version 0.931.0, https://github.com/jluebeck/PrepareAA) with CNVKit (version 0.9.6). The AA graph file was cleaned with the PrepareAA “graph_cleaner.py” script to remove edges which conform to sequencing artifact profiles - namely, very short everted (inside-out read pair) orientation edges. Such spurious edges appear as numerous short brown 'spikes' in the AA amplicon image. Second, we utilized optical map (OM) contigs (Bionano Genomics, USA) which we incorporated with the AA breakpoint graph. We used AmpliconReconstructor (version 1.01) (AR) to scaffold together individual breakpoint graph segments against the collection of OM contigs. We ran AR with the --noConnect flag set and otherwise default settings. Third, we utilized the OM alignment tool FaNDOM (version 0.2) (default settings) to correct and infer additional OM contig reference alignments and junctions missed by AA and AR. OM contigs identified three additional breakpoint edges, which were subsequently added into the AA graph file. Lastly, we incorporated fragment size and sequencing data from PFGE experiments, identifying from the separated bands the estimated length and identity of genomic segments between CRISPR cut sites. We explored the various ways the overlapping OM scaffolds could be joined while conforming to the PFGE fragment sizes and identities of the genomic regions suggested from the PFGE data. We selected a candidate structure which was concordant with the PFGE cut data expected fragment sizes, as well as intra-fragment sequence identity and multiplicity of copy count as suggested by AA analysis of the sequenced PFGE bands. The reconstruction used all but five discovered genomic breakpoint edges inside the DM region. The remaining five edges were scaffolded by two different OM contigs and each scaffold individually suggested a separate site of structural heterogeneity within the ecDNA as compared against the reconstruction. We required that the entirety of the significantly amplified amplicon segments was used in the reconstruction. We estimated that at the baseline, genomic segments appearing once in the reconstruction existed with a copy number between 170-190. In the final structure, all amplicon segments with copy number >40 were used. Additionally, when segments were repeated inside the reconstruction, we ensured that the multiplicities of the amplicon segments suggested the reconstruction matched the multiplicities of the amplicon segments as reported by WGS. For fine mapping analysis of the PVT1-MYC breakpoint, reads that align to both PVT1 and MYC were extracted from WGS short read sequencing which identified 10 unique reads support the breakpoint. Multiple sequence alignment was performed with ClustalW (version 2.1) for visualization.

Single-cell RNA and ATAC-seq Data Analysis
Subsequent analyses on RNA were performed using Seurat (version 3.2.3), and those on ATAC-seq were performed using ArchR (version 1.0.1). Cells with more than 200 unique RNA features, less than 20% mitochondrial RNA reads, less than 50,000 total RNA reads were retained for further analyses. Doublets were removed using ArchR. Raw RNA counts were log-normalized using Seurat’s NormalizeData function, scaled using the ScaleData function, and the data were visualized on a UMAP using the first 30 principal components. Dimensionality reduction for the ATAC-seq data were performed using Iterative Latent Semantic Indexing (LSI) with the addIterativeLSI function in ArchR. To impute accessibility gene scores, we used addImputeWeights to add impute weights and plotEmbedding to visualize scores. To compare the accessibility gene scores for MYC with MYC RNA expression, getMatrixFromProject was used to extract the gene score matrix and the normalized RNA data were used. To identify variable ATAC-seq peaks on COLO320-DM and COLO320-HSR amplicons, we first calculated amplicon copy numbers based on background ATAC-seq signals as previously described, using a sliding window of five megabases moving in one-megabase increments across the reference genome. We used the copy number z scores calculated for the chr8:124000001-129000000 interval for estimating copy numbers of MYC-bearing ecDNAs in COLO320-DM and MYC-bearing chromosomal HSRs in COLO320-HSR. We then incorporated these estimated copy numbers into the variable peak analysis as follows. COLO320-DM and COLO320-HSR cells were separately assigned into 20 bins based on their RNA expression of MYC. Next, pseudo-bulk replicates for ATAC-seq data were created using the addGroupCoverages function grouped by MYC RNA quantile bins. ATAC-seq peaks were called using addReproduciblePeakSet for each quantile bin, and peak matrices were added using addPeakMatrix. Differential peak testing was performed between the top and the bottom RNA quantile bins using getMarkerFeatures. A false discovery rate cutoff of 1e-15 was imposed. The mean copy number z score for each quantile bin was then calculated and a copy number fold change between the top and bottom bin was computed. Finally, we filtered on significantly differential peaks that are located in chr8:127432631-129010071 and have fold changes above the calculated copy number fold change multiplied by 1.5.

HiChIP Data Analysis
The Juicer (version 1.5) pipeline's HiCCUPS tool and FitHiChIP (version 8.0) were used to identify loops. Filtered read pairs from the HiC-Pro pipeline were converted into .hic format files and input into HiCCUPS using default settings. Dangling end, self-circularized, and re-ligation read pairs were merged with valid read pairs to create a 1D signal bed file. FitHiChIP was used to identify “peak-to-all” interactions at 10 kb resolution using peaks called from the one-dimensional HiChIP data. A lower distance threshold of 20 kb was used. Bias correction was performed using coverage specific bias. HiChIP contact matrices stored in .hic files were visualized in R (version 4.0.3) using gTrack (version 0.1.0) at 10 kb resolution following Knight-Ruiz normalization. We also compared HiChIP contract matrices following ICE and OneD normalization following copy number correction using the dryhic R package (version 0.0.0.9100). Virtual 4C plots were generated from dumped matrices generated with Juicer Tools (1.9.9). The Juicer Tools tools dump command was used to extract the chromosome of interest from the .hic file. The interaction profile of a 10-kb bin containing the anchor was then plotted in R (version 4.0.3) following normalization by the total number of valid read pairs and smoothing with the rollmean function from the zoo package (version 1.8-9).

SSNU16-dCas9-KRAB WGS Data Analysis
Regions of copy number alteration were identified using ReadDepth (version 0.9.8.5) with parameters recommended by AmpliconArchitect (version 1.0), and amplicon reconstruction performed using the default parameters. Structural variant junctions were extracted from the edges_cnseg.txt output files and used for visualization.

ATAC-seq Data Analysis
Peaks were called using MACS2 (version 2.1.0.20150731) with a q-value cut-off of 0.01 and with a no-shift model. Peaks from replicates were merged, read counts were obtained using bedtools (version 2.17.0) and normalized using DESeq2 (version 1.26.0). To identify accessible elements in MYC and FGFR2 ecDNAs in SNU16, we filtered on all ATAC-seq peaks within known ecDNA-amplified regions (chr8:128200000-129200000 for the MYC ecDNA, chr10:122000000-123680000 for the FGFR2 ecDNA) whose normalized read counts (using the "counts" function in DESeq2 with normalized = TRUE) exceeded a manually determined threshold (500 for the MYC amplicon, 1000 for the FGFR2 amplicon). Peaks that met all criteria for two technical replicates were included as candidate DNA elements in the CRISPR interference study.

CRISPR Interference Screen Data Analysis
Relative abundances of sgRNAs were measured using MAGeCK (version 0.5.9.4). sgRNA counts were obtained using the "mageck count" command. For samples with PCR replicates, if a PCR replicate has fewer than 1000 total sgRNAs passing filter (raw counts > 20), the replicate was excluded. Next, each sgRNA count was divided by total sgRNA counts for each library and multiplied by one million to give a normalized count (count per million, CPM). For samples with PCR replicates, mean CPM was calculated for each sgRNA. sgRNAs that have CPMs lower than 20 in the unsorted cells were classified as dropouts and removed from the analysis. We then calculated the log2 fold change of each sgRNA in each sorted cell bin over unsorted cells by dividing the respective CPMs followed by log-transformation. sgRNA enrichment was then quantified as previously described. Briefly, the log2 fold change in the high expression bin was subtracted from that in the low expression bin [log2(low/high)] for each sgRNA. The resulting log2(low/high) values were averaged for each candidate regulatory element and z scores were calculated using the formula z = (x-m)/S.E., where x is the mean log2(low/high) of the candidate element, m is the mean log2(low/high) of negative control sgRNAs, and S.E. is the standard error calculated from the standard deviation of negative control sgRNAs divided by the square root of the number of sgRNAs targeting the candidate element in independent biological replicates. Z scores were used to compute upper-tail p values using the normal distribution function, which were adjusted with p.adjust in R using the Benjamini-Hochberg Procedure to produce false discovery rate (FDR) values. For assessing sgRNA correlations across all six sorted bins for individual elements, we computed Spearman coefficients for all individual sgRNAs across the six fluorescence bins using log2 fold changes over unsorted cells.

TR14 Amplicon Reconstruction
WGS coverage was computed in 20bp bins, normalized as counts per million, using using deepTools 3.3.0. Copy number variation was called using QDNAseq 1.22.0, binning primary alignments with MAPQ≥20 in 10kb bins, default filtering and additional filtering of bins with more than 5% Ns in the reference. Bins were corrected for GC content and normalized. Segmentation was performed using the CBS method with no transformation of the normalized counts and parameter alpha=0.05. Structural variants were called on Nanopore long read data using Sniffles v1.0.11 and parameters --min_length 15 --genotype --min_support 3 --report_seq. To reconstruct the coarse structure of oncogene amplifications in TR14, we compiled all Sniffles structural variants larger than 10kb with a minimum read support of 15 into one genome graph using gGnome 0.1, nodes representing genomic segments connected by reference or structural variant edges. Non-amplified segments (i.e. mean Illumina WGS coverage less than 10-fold the median chromosome 2 coverage) were discarded from the graph. Strong clusters in the genome graph were identified, partitioning the graph into groups of segments that could be reached from one another. We identified the clusters containing the four amplified oncogenes (MYCN, CDK4, MDM2, ODC1) and manually selected circular paths through each cluster that could account for the main copy number steps around the oncogenes. We used gTrack (https://github.com/mskilab/gTrack) for visualization. Hi-C data were used to validate these reconstructions, confirming that all strong off-diagonal signal indicative of structural rearrangements were captured by the reconstruction. Previously studies suggest that the identified amplicons exist as extrachromosomal DNA.

Hi-C Data Analysis
Knight-Ruiz normalization per hg19 chromosome was used for Hi-C maps, interaction across different chromosome pairs should therefore only carefully be interpreted. For TR14, we created a custom genome containing additionally the amplicon reconstructions. The sequences of amplicons were composed from hg19 based on the order and orientation of their chromosomal fragments. The original fragment locations on hg19 were masked to allow unambiguous mapping. Note, by this also Hi-C reads from wildtype alleles are mapping to the amplicon sequences leading to a mix of signal, depending on the fraction of amplicons and wildtype allele. After mapping, we kept only amplicons and removed all other chromosomes to create Hi-C maps and apply GW_KR normalization using Juicer Tools v1.19.02.

TR14 Interaction analysis
H3K27ac ChIP-seq coverage tracks were created by extending reads to 200bp, filtering using the ENCODE DAC blacklist and normalizing to counts per million in 10bp bins with deepTools 3.3.0. Enhancers were called using LILY (https://github.com/BoevaLab/LILY, not versioned) with default parameters. The HPCAL1 enhancer region was defined by two LILY-defined boundary enhancers as chr2:10424449-10533951. A virtual 4C track was generated by the mean genome-wide interaction profile (KR-normalized Hi-C signal in 5kb bins) across all overlapping 5kb bins. For the aggregate analysis of the effect of H3K27 acetylation on interaction, all 5kb bin pairs located on different amplicons were analyzed for their KR-normalized Hi-C signal depending on the mean H3K27ac fold-change over input of each of the two bins. We used 5-fold change threshold to distinguish low- from high-H3K27ac bins.

Custom code used in this study is available at https://github.com/ChangLab/ecDNA-hub-code-2021.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

ChIP-seq, HiChIP, Hi-C, RNA-seq, and single cell multiome ATAC + gene expression data generated in this study have been deposited in GEO and are available under accession number GSE159986. Nanopore sequencing data, whole genome sequencing data, sgRNA sequencing data, and targeted ecDNA sequencing data following CRISPR-Cas9 digestion and PFGE generated in this study has been deposited in SRA and are available under accession number PRJNA670737. Optical mapping data generated in this study has been deposited in GenBank with Bioproject code PRJNA731303. The following publicly available data was also used in this study: TR14

H3K27ac ChIP-seq (GEO: GSE90683); COLO320-DM, COLO320-HSR and PC3 WGS (SRA: PRJNA506071); SNU16 WGS (SRA: PRJNA523380); HK359 WGS (SRA: PRJNA338012).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. Experiments were performed with two or more replicates to capture variability. Imaging quantifications included 9 or more cells for assessing differences between treatments to capture cell-to-cell variability. For the CRISPRi studies in SNU16-dCas9-KRAB, 20 independent guides for each candidate enhancer were included for the perturbation analysis as determined by previously published studies outlined in Methods. Sample size was determined by previous studies to ensure data reproducibility.
Data exclusions	For the CRISPRi studies in SNU16-dCas9-KRAB, relative abundances of sgRNAs were measured using MAGeCK85. sgRNA counts were obtained using the "mageck count" command. For samples with PCR replicates, if a PCR replicate has fewer than 1000 total sgRNAs passing filter (raw counts > 20), the replicate was excluded. Next, each sgRNA count was divided by total sgRNA counts for each library and multiplied by one million to give a normalized count (count per million, CPM). For samples with PCR replicates, mean CPM was calculated for each sgRNA. sgRNAs that have CPMs lower than 20 in the unsorted cells were classified as dropouts and removed from the analysis.
Replication	Experiments were performed with two or more replicates to capture variability. All replication attempts were successful.
Randomization	All experiments used cultured cell lines. As the techniques used did not involve live organisms, randomization was not relevant to this study.
Blinding	All data were collected using instruments without bias. Because these data were generated using objective quantifications, researchers assessing results were not blinded for the experimental design. Blinding is not relevant to this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	HiChIP: H3K27ac (Abcam ab4729), 2 µg antibody per HiChIP (1 million cells), 7.5 µg antibody per HiChIP (4 million cells) ChIP-seq: spike-in antibody (Active Motif 53083), 2 µg antibody per ChIP H3K27Ac (Abcam ab4729), 7.5 µg of antibody per ChIP BRD4 (Bethyl Laboratories A301-985A100), 7.5 µg of antibody per ChIP
Validation	All antibodies were validated by the manufacturers (validation described below). spike-in antibody (Active Motif 53083): The Spike-in antibody recognizes a Drosophila-specific histone variant, H2Av. Because of the specificity of the Spike-in Antibody for the Spike-in Chromatin modification, there is no cross-reactivity with mammalian samples leading to reduced background signal. The Spike-in Antibody shows minimal cross reactivity with mammalian samples. When the Spike-in Antibody was tested in ChIP-Seq with human chromatin, there is little to no signal detected. This demonstrates the specificity of the spike-in normalization strategy.

H3K27Ac (Abcam ab4729): Chromatin was prepared from HeLa (Human epithelial cell line from cervix adenocarcinoma) cells according to the Abcam X-ChIP protocol. Cells were fixed with formaldehyde for 10 minutes. The ChIP was performed with 25 µg of chromatin, 2 µg of ab4729 (blue), and 20 µl of Protein A/G sepharose beads. No antibody was added to the beads control (yellow). The immunoprecipitated DNA was quantified by real time PCR (Taqman approach). Primers and probes are located in the first kb of the transcribed region.

BRD4 (Bethyl Laboratories A301-985A100): Detection of human BRD4 by western blot of immunoprecipitates.Samples: Whole cell lysate (1.0 mg per IP reaction; 20% of IP loaded) from HeLa cells prepared using NETN lysis buffer. Antibodies: Affinity purified rabbit anti-BRD4 antibody A301-985A100 (lot A301-985A100-7) used for IP at 3 µg per reaction. BRD4 was also immunoprecipitated by rabbit anti-BRD4 antibody BL5482 and rabbit anti-BRD4 recombinant monoclonal antibody [BL-149-2H5] (A700-004). For blotting immunoprecipitated BRD4, A700-004 was used at 1:1000. Detection: Chemiluminescence with an exposure time time of 1 seconds.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	COLO320-DM, COLO320-HSR, HCC1569, SNU16, HK359 and PC3 cells were purchased from ATCC. The TR14 neuroblastoma cell line was a gift from J. J. Molenaar (Princess Máxima Center for Pediatric Oncology, Utrecht, Netherlands).
Authentication	Cell lines obtained from ATCC were not authenticated. TR-14 cell line identity for the master stock was verified by STR genotyping (IDEXX BioResearch, Westbrook, ME).
Mycoplasma contamination	Cells were tested negative for mycoplasma.
Commonly misidentified lines (See ICLAC register)	None of the cell lines used are registered by ICLAC as commonly misidentified.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Data generated in this study have been deposited in GEO and are available under accession number GSE159986: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159986
Files in database submission	Raw fastq files and bigwig signal tracks
Genome browser session (e.g. UCSC)	No longer applicable

Methodology

Replicates	Two biological replicates
Sequencing depth	All ChIP-seq libraries were sequenced with paired end 75 bp reads. COLO320DM_DMSO_H3K27ac_rep1: 35028073 total read pairs, 57976491 uniquely mapped reads COLO320DM_DMSO_H3K27ac_rep2: 36231441 total read pairs, 59443223 uniquely mapped reads COLO320DM_DMSO_Brd4_rep1: 37942699 total read pairs, 56562296 uniquely mapped reads COLO320DM_DMSO_Brd4_rep2: 40976059 total read pairs, 58580199 uniquely mapped reads COLO320DM_JQ1_Brd4_rep1: 35791866 total read pairs, 51099627 uniquely mapped reads COLO320DM_JQ1_Brd4_rep2: 36509133 total read pairs, 51776818 uniquely mapped reads COLO320DM_DMSO_input: 37562377 total read pairs, 59101674 uniquely mapped reads COLO320DM_JQ1_input: 48141792 total read pairs, 75922546 uniquely mapped reads COLO320HSR_DMSO_H3K27ac_rep1: 30147193 total read pairs, 52767607 uniquely mapped reads COLO320HSR_DMSO_H3K27ac_rep2: 29068492 total read pairs, 50143394 uniquely mapped reads COLO320HSR_DMSO_Brd4_rep1: 39478660 total read pairs, 51644801 uniquely mapped reads COLO320HSR_DMSO_Brd4_rep2: 36985925 total read pairs, 49640593 uniquely mapped reads COLO320HSR_JQ1_Brd4_rep1: 44614061 total read pairs, 55518894 uniquely mapped reads COLO320HSR_JQ1_Brd4_rep2: 41801618 total read pairs, 53234955 uniquely mapped reads COLO320HSR_DMSO_input: 40993674 total read pairs, 65005687 uniquely mapped reads COLO320HSR_JQ1_input: 42605487 total read pairs, 67286304 uniquely mapped reads SNU16_H3K27ac_rep1: 27016766 total read pairs, 47315915 uniquely mapped reads SNU16_H3K27ac_rep2: 25527024 total read pairs, 44235670 uniquely mapped reads SNU16_Brd4_rep1: 26859602 total read pairs, 32000619 uniquely mapped reads SNU16_Brd4_rep2: 26495646 total read pairs, 31723816 uniquely mapped reads SNU16_input: 26889359 total read pairs, 45300478 uniquely mapped reads
Antibodies	ChIP-seq:

Antibodies	spike-in antibody (Active Motif 53083), 2 µg antibody per ChIP H3K27Ac (Abcam ab4729), 7.5 µg of antibody per ChIP BRD4 (Bethyl Laboratories A301-985A100), 7.5 µg of antibody per ChIP
Peak calling parameters	MACS2 76 (version 2.1.1.20160309) was used for peak calling with the following parameters: macs2 callpeak -t chip_bed -c input_bed -n output_file -f BED -g hs -q 0.01 --nomodel --shift 0. A reproducible peak set across biological replicates was defined using the IDR framework (version 2.0.4.2). Reproducible peaks from all samples were then merged to create a union peak set. ChIP-seq signal was converted to bigwig format for visualization using deepTools bamCoverage 77 (version 3.3.1) with the following parameters: --bs 5 --smoothLength 105 --normalizeUsing CPM --scaleFactor 10. Enrichment of ChIP signal at peaks was performed using deepTools computeMatrix.
Data quality	We used IDR to identify reproducible peaks between biological replicates, which identified the following number of peaks with an IDR <0.05: COLO320DM_DMSO_H3K27ac: 16077 unique, reproducible peaks COLO320DM_DMSO_Brd4: 18805 unique, reproducible peaks COLO320DM_JQ1_Brd4: 4252 unique, reproducible peaks COLO320HSR_DMSO_H3K27ac: 31115 unique, reproducible peaks COLO320HSR_DMSO_Brd4: 2305 unique, reproducible peaks COLO320HSR_JQ1_Brd4: 313 unique, reproducible peaks SNU16_H3K27ac: 38962 unique, reproducible peaks SNU16_Brd4: 4070 unique, reproducible peaks
Software	Paired-end reads were aligned to the hg19 genome using Bowtie2 74 (version 2.3.4.1) with the --very-sensitive option following adapter trimming with Trimmomatic 75 (version 0.39). Reads with MAPQ values less than 10 were filtered using samtools and PCR duplicates removed using Picard's MarkDuplicates. MACS2 76 (version 2.1.1.20160309) was used for peak calling with the following parameters: macs2 callpeak -t chip_bed -c input_bed -n output_file -f BED -g hs -q 0.01 --nomodel --shift 0. A reproducible peak set across biological replicates was defined using the IDR framework (version 2.0.4.2). Reproducible peaks from all samples were then merged to create a union peak set. ChIP-seq signal was converted to bigwig format for visualization using deepTools bamCoverage 77 (version 3.3.1) with the following parameters: --bs 5 --smoothLength 105 --normalizeUsing CPM --scaleFactor 10. Enrichment of ChIP signal at peaks was performed using deepTools computeMatrix.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	RNA FISH flow was performed for MYC and FGFR2 using the PrimeFlow™ RNA Assay Kit (Thermo Fisher) following the manufacturer's protocol.
Instrument	Influx
Software	Flow cytometry data were analyzed using FlowJo (10.7.0).
Cell population abundance	Live singlets which expressed GFP (encoded by sgRNA lentiviral construct) were selected for the 6-way cell sort based on Alexa Fluor 647 (MYC or FGFR2) level. Each bin was set at 10% of the filtered population.
Gating strategy	Cells were gated in forward/side scatter, singlets were gated using tirgger pulse width and forward scatter, and live cells were selected using negativity in the PE channel for the live/dead stain. GFP+ cells were selected for expression of sgRNA construct, MYC or FGFR2 RNA populations were gated based on MFI in the alexa fluor 647 channel. The oncogene (MYC/FGFR2) was labeled with Alexa Fluor 647 and ACTB was labeled with Alexa Fluor 750. Based on the assumption that the expression of the housekeeping gene is not correlated with the oncogene, any correlation in fluorescence intensities between the ACTB and the oncogene was attributed to flowFISH staining efficiency and manually regressed using the FACS compensation tool. The degree of compensation was determined so that the top and bottom 25% of cells based on Alexa Fluor 647 signal intensity deviated no more than 15% from the population mean in Alexa Fluor 750 signal intensity. After compensation, we gated on cells with positive ACTB labeling and sorted cells into six bins using Alexa Fluor 647 MFI corresponding to the following percentile ranges: 0-10% (bin 1), 10-20% (bin 2), 35-45% (bin 3), 55-65% (bin 4), 80-90% (bin 5), 90-100% (bin 6).

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.