Aus der Klinik für Pädiatrie mit Schwerpunkt Onkologie und Hämatologie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

MYCN amplification in neuroblastoma: sequence, chromatin
landscape, and genome remodeling

MYCN Amplifikation im Neuroblastom: Sequenz,
Chromatinlandschaft und genomische Remodellierung

zur Erlangung des akademischen Grades
Medical Doctor - Doctor of Philosophy (MD/PhD)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Konstantin Helmsauer

Datum der Promotion: 30.11.2023

Table of Contents

Figures

Abbreviations

4C-seq	Chromosome conformation capture circular sequencing
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
ChIP-seq	Chromatin immunoprecipitation sequencing
CRC	Core regulatory circuit
dmin	Double minute chromosome
ecDNA	Extrachromosomal DNA
FISH	Fluorescence in situ hybridization
H3K27ac	Histone H3 lysine residue 27 acetylation
H3K4me1	Histone H3 lysine residue 4 monomethylation
HSR	Homogeneously staining region
MRD	Minimal residual disease
TAD	Topologically associated domain

Abstract

MYCN amplification drives half of all high-risk neuroblastomas and confers dismal prognosis. It occurs when *MYCN* and other chromosomal fragments are joined together to form an amplicon sequence that can be found tens to hundreds of times per cell. In most cases, these copies come as extrachromosomal DNA (ecDNA) in addition to chromosomes. The sequence and chromatin landscape of the *MYCN* amplicon has not been systematically mapped, gene regulation in the context of amplification has not been explored, and the consequences of extrachromosomal amplification are incompletely understood. Here, we analyze Illumina and Nanopore whole-genome sequencing, RNA-seq, ATAC-seq, ChIP-seq, and Hi-C data of primary neuroblastomas and neuroblastoma cell lines. We map *MYCN*-driving enhancers in neuroblastoma and identify co-amplification of local enhancers as the main principle governing the non-coding contents of *MYCN* amplification. We find that loss of local enhancers can be compensated for by the incorporation of distal enhancers and the formation of new chromatin domains through structural variation on the amplicon. We show that *MYCN* amplification in neuroblastoma is associated with clusters of interchromosomal rearrangements at the *MYCN* locus indicative of complex amplicon structure or re-integration of ecDNA into chromosomes. This can affect gene expression and is prognostically relevant. Finally, we describe heterogeneous ecDNA populations and show that ecDNA can engage in *in trans* interactions forming transcriptionally active hubs of individual amplicon copies. Our results demonstrate how non-coding elements shape amplification structure and function. Together, they suggest a model of amplification as building blocks for genome remodeling and nuclear reorganization.

Zusammenfassung

MYCN Amplifikation charakterisiert etwa jedes zweite Hochrisiko-Neuroblastom und geht mit besonders aggressivem klinischem Verhalten einher. In solchen Tumoren sind *MYCN* und andere chromosomale Fragmente zu einer Amplikonsequenz zusammengefügt, die dann meist viele dutzendmal kopiert pro Zelle vorliegt. Typischerweise liegen die Amplikonkopien als ringförmige extrachromosomale DNA (ecDNA) zusätzlich zu den Chromosomen vor. Wie Sequenz und Epigenetik des *MYCN*-Amplikons zur Genregulation im Kontext von Amplifikation beitragen, wurde bisher noch nicht systematisch untersucht. Auch die Auswirkungen extrachromosomaler Amplifikation auf das chromosomale Genom und dessen Organisation im Zellkern sind nur unvollständig bekannt. In der vorliegenden Arbeit analysieren wir Illumina und Nanopore Whole Genome Sequencing, RNA-seq, ATAC-seq, ChIP-seq und Hi-C-Daten von primären Neuroblastomen und Zelllinien. Wir lokalisieren *MYCN*-regulierende Enhancer und identifizieren die Ko-Amplifikation von lokalen Enhancern als Prinzip, das die Sequenz von *MYCN* Amplifikation entscheidend mitbestimmt. Wir stellen fest, dass der Verlust lokaler Enhancer durch den Einbau distaler Enhancer und die Bildung neuer Chromatindomänen durch strukturelle Veränderungen auf dem Amplikon kompensiert werden kann. Außerdem zeigen wir, dass *MYCN*-Amplifikation mit Clustern von interchromosomalen Bruchpunkten am *MYCN*-Lokus einhergeht, die auf komplexe ecDNA-Strukturen oder die Re-Integration von ecDNA in Chromosomen hinweisen. Dies kann die Genexpression in der Bruchpunktumgebung beeinflussen und wird als klinisch prognostischer Faktor identifiziert. Zuletzt beschreiben wir heterogene ecDNA-Populationen, die Organisation von ecDNA in transkriptionell aktiven ‚ecDNA hubs' im Zellkern und zeigen, dass ecDNA *in trans* interagieren kann. Unsere Ergebnisse zeigen, wie nicht-kodierende DNA-Elemente die Struktur und Funktion von Genamplifikationen mitbestimmen. Davon ausgehend schlagen wir ein Modell von Genamplifikationen als Bausteine zur Umordnung des Genoms und der Zellkernorganisation vor.

1   Introduction

Neuroblastoma is an embryonal tumor arising from the developing sympathetic nervous system. It occurs primarily in infants and toddlers with a median age at diagnosis of 1.5 years (1). Its clinical course is highly variable: Low-risk tumors require minimal treatment. Some may even regress spontaneously. High-risk neuroblastomas, in contrast, have a dismal prognosis despite aggressive multimodal therapy and account for around 10% of childhood cancer deaths (2). To this day, high-risk neuroblastoma remains one of the major challenges in pediatric oncology.

44% of high-risk neuroblastomas are characterized by *MYCN* amplification, where cells harbor tens to hundreds of copies of the *MYCN* gene, a developmental transcription factor and member of the *MYC* family of proto-oncogenes (3-6). With a 5-year-survival rate of 55% (6), this is a clinically relevant marker of poor prognosis. *MYCN* amplification is central to establishing and maintaining neuroblastoma cell identity: *MYCN* overexpression initiates neuroblastoma tumorigenesis in mice (7, 8). Once *MYCN*-amplified, neuroblastoma cells depend on sustained *MYCN* expression and differentiate or become apoptotic upon *MYCN* loss (9, 10). As a transcription factor, MYCN forms heterodimers with MAX and preferentially binds at E-box motifs (CACGTG) throughout the genome, driving several aspects of malignancy at once, e.g. increased proliferation and cell cycle progression, replicative immortality, invasion, metastasis, and immune escape (as reviewed in (11) and (12)). At oncogenic expression levels, it also serves as a global transcriptional amplifier by invading pre-established developmental enhancers (13). In neuroblastoma, MYCN has been proposed as a member of the noradrenergic core regulatory circuit (CRC), a positive feedback loop of transcription factors including MYCN, PHOX2B, HAND2, GATA3, ISL1, TBX2, and ASCL1 (14-17). These factors drive each other's expression and together establish noradrenergic neuroblastoma cell identity, which has originally been defined by its specific enhancer landscape (14, 15). Beyond neuroblastoma, *MYCN* amplification is among the most common copy number aberrations in pediatric cancer in general (18). It is particularly prevalent in Wilms' tumor, medulloblastoma, glioblastoma, and rhabdomyosarcoma and is often associated with particularly aggressive clinical behavior (18-22).

Focal amplifications are a distinctive feature of cancer genomes and affect 52% of cases in a recent pan-cancer study (23). The amplicon, i.e. the amplified DNA sequence, usually comprises a proto-oncogene and up to a few megabases of surrounding sequence (24). Structurally, there are two

canonical forms of amplification: As extrachromosomal DNA (ecDNA; also termed double minute chromosomes, dmin), copies of the amplicon sequence are found on chromatin circles in addition to the chromosomal genome. Alternatively, amplification can occur intrachromosomally as homogeneously staining regions (HSRs), repetitive arrays of individual copies integrated into a chromosome. The origins of ecDNA and HSRs are incompletely understood. In cell lines, ecDNA formation has been observed when chromothripsis shatters a chromosome and some of the resulting fragments fuse in a circular fashion (25, 26). This can explain why focal amplifications are often structurally complex, containing distant parts of the genome, and are often associated with chromothripsis in primary tumors (27). In primary neuroblastomas, *MYCN* amplification usually occurs through ecDNA (28) but most *MYCN*-amplified neuroblastomas show no signs of chromothripsis at all (29), which suggests that other mechanisms likely contribute to ecDNA generation in neuroblastoma. HSRs, in turn, can form through breakage-fusion bridge cycles where telomere crisis leads to repeated sister chromatid fusion followed by chromosome breakage (26). Alternatively, HSRs can arise when ecDNA re-integrates into chromosomes (26, 30-32). Vice versa, HSRs can also give rise to new ecDNA (26), and both HSRs and ecDNA can coexist in individual cells (31, 33), raising the question how the form of amplification impacts tumor biology.

Amplification as ecDNA offers unique advantages during tumor evolution. ecDNA gets replicated during the cell cycle but does not harbor a centromere. During mitosis, it is therefore randomly distributed among daughter cells which may inherit vastly different numbers of ecDNA copies. This way, extrachromosomal amplification drives copy number heterogeneity in the cell population and enables fast copy number changes during tumor evolution (33, 34). Indeed, extrachromosomal amplification is selected for in experimental setups requiring particularly high or variable gene expression (26, 31). HSRs are commonly seen as a more stable form of amplification. Although intrachromosomal copy number can also evolve over time, HSRs seem to be the preferred form of amplification in stable environments (31). Beyond copy number plasticity, both ecDNA and HSRs are prone to further mutations and rearrangements (23, 26). Emerging amplification, changes to amplicon sequence, or amplicon copy number have repeatedly been linked to the acquisition of drug resistance (31, 34-36). This illustrates that amplified loci likely are among the most plastic parts of a cancer genome and play a key role during tumor evolution.

Much of the genetics and epigenetics of oncogene amplification, however, is yet to be understood: the functional relevance of amplicon sequence, gene regulation in the context of

extrachromosomal and intrachromosomal amplification, and the effects of amplification on genome architecture and nuclear organization. Here, we present three publications that address parts of these questions. In Helmsauer et al. (37), we map the content and structure of the *MYCN* amplicon and the enhancer landscape around *MYCN* in neuroblastoma. We characterize how the epigenetics of *MYCN* regulation shape the structure of amplification. This reveals that specific enhancers are systematically included on the amplicon, how enhancer hijacking contributes to the structure of amplification, and how new chromatin domains can emerge on extrachromosomal and intrachromosomal amplicons. In Koche et al. (38), we explore whole-genome sequencing and RNA-seq data of primary neuroblastomas. We identify clusters of rearrangements connecting ecDNA with distant genomic loci, representing complex ecDNA structures or re-integration into chromosomal loci. Such rearrangements can disrupt tumor suppressor genes or increase oncogene expression nearby. Clinically, we find that this pattern of structural variation defines a novel subgroup of ultra-high risk neuroblastomas. This leads us to propose ecDNA as a driver of genome remodeling. Finally, in Hung et al. (39), we show that ecDNA spatially clusters together in the nucleus. This allows for *in trans* interaction between individual ecDNA elements suggesting a novel mechanism for oncogene activation in cancer.

2 Materials & Methods

The following section summarizes the Materials & Methods of the presented publications (37-39) relevant to the scope of this thesis. We refer to the respective publications for a detailed account of all Materials & Methods.

2.1 Helmsauer et al. Enhancer hijacking determines extrachromosomal circular *MYCN* amplicon architecture in neuroblastoma

2.1.1 *Cell lines*

We obtained neuroblastoma cell lines CHP-212, IMR-5/75, NGP, and Kelly as a gift from Carol J. Thiele, from the German Collection of Microorganisms and Cell Cultures or the American Type Culture Collection. Cell line identification was performed using short tandem repeat genotyping (Genetica DNA Laboratories, Burlington, NC, USA, and IDEXX BioResearch, Westbrook, ME, USA), and *Mycoplasma sp.* contamination was ruled out by the Lonza MycoAlert assay (Lonza Group Ltd., Basel, Switzerland). We cultured cells in Roswell Park Memorial Institute 1640 medium (RPMI-1640; Thermo Fisher Scientific Inc., Waltham, MA, USA) with 1% Penicillin/Streptomycin (Thermo Fisher Scientific Inc., Waltham, MA, USA) and 10% Fetal Calf Serum (Thermo Fisher Scientific Inc., Waltham, MA, USA).

2.1.2 *Fluorescence in situ hybridization*

Metaphase-arrested cells were prepared using Colcemid (20 µl/2 ml; Roche AG, Basel, Switzerland). Cells were suspended in 5ml 0.4% potassium chloride for 10min, fixed in 1ml potassium chloride and 1ml methanol/acetic acid 3:1, and washed three times (2ml, 5ml, and 5ml methanol/acetic acid followed by 200g/10min centrifugation each). Metaphase spreads were prepared by dropping the cell suspension on slides. Slides were washed in PBS buffer, pepsin-digested (0.04% pepsin in 0.001M hydrochloric acid, 10min), washed again in 0.5x saline sodium citrate wash buffer, dehydrated in ascending ethanol solutions (70%, 80%, 100%; 3min each), and dried. 10µl Vysis LSI N-MYC probe (Abbott Laboratories, Abbott Park, IL, USA) was added. Samples were denatured at 75°C for 10min and hybridized at 37°C overnight. After washing with 0.4x saline sodium citrate buffer / 0.3% IGEPAL CA-630 (Sigma-Aldrich, Inc., St. Louis, MO, USA) for 3min at 60 °C and 2x saline sodium citrate buffer / 0.1% IGEPAL CA-630 for 3min at

room temperature, 5µl 4',6-Diamidino-2-phenylindol (DAPI; Vector Laboratories, Inc., Newark, CA, USA) were added.

2.1.3   RNA-seq data analysis

We downloaded RNA-seq data for neuroblastoma cell lines from Gene Expression Omnibus (GSE90683) (14). We used FASTQC 0.11.8 for quality control. We trimmed adapters with BBMap 38.58, aligned reads to GRCh37 using STAR 2.7.1 (40) with default parameters, used featureCounts from Subread 1.6.4 (41) to obtain read counts for genes as defined in Ensembl release 75, and accounted for library size and composition using sizeFactor-normalization (DESeq2 1.22.2) (42).

2.1.4   ChIP-seq data analysis

We refer to the enclosed publication for details on chromatin immunoprecipitation sequencing (ChIP-seq) library preparation and sequencing (37). Additional ChIP-seq data were obtained from Gene Expression Omnibus (GSE90683, GSE24447, and GSE28874) (14, 43). Data were quality controlled with FASTQC 0.11.8, adapters were trimmed with BBMap 38.58, and reads were aligned to hg19 with BWA-MEM 0.7.15 (default parameters) (44). Duplicate reads were removed using Picard 2.20.4. We considered RSC and NSC metrics using Phantompeakqualtools 1.2.1 (45) as metrics for ChIP-seq data quality. We used deepTools 3.3.0 (46) to extend reads to 200bp for single-end libraries and to fragment size for paired-end libraries, to discard reads mapping to ENCODE DAC blacklisted regions (47), and obtain read counts per million reads in all 10bp bins. ChIP-seq peaks were identified with MACS2 2.1.2 using default parameters (48). Super enhancers were defined by LILY (https://github.com/BoevaLab/LILY) (14) based on histone H3 lysine residue 27 acetylation (H3K27ac) ChIP-seq signal using default parameters.

2.1.5   ATAC-seq data analysis

We refer to the enclosed publication for details on library preparation and sequencing for assay for transposase-accessible chromatin using sequencing (ATAC-seq) experiments (37). We downloaded additional ATAC-seq data from Gene Expression Omnibus (GSE80154) (13). Read trimming, mapping, and duplicate removal was performed as described above for ChIP-seq data. We extended reads to fragment size and discarded reads that mapped to ENCODE DAC blacklisted regions (47). ATAC-seq signal was then quantified as read counts per million reads in

10bp bins using deepTools 3.3.0 (46). Peaks were identified with MACS2 2.1.2 using default parameters (48).

2.1.6 4C-seq and Hi-C data analysis

We refer to the enclosed publication for details on Chromosome conformation capture circular sequencing (4C-seq) and Hi-C library preparation and sequencing (37). 4C-seq reads were processed and mapped to GRCh37 as described before (49). Coverage was normalized as read counts per million mapped reads in a sliding window of 10 fragments. For Hi-C data analysis, we used Juicer 1.5.6 (50) with BWA 0.7.17 (44) to map and process Hi-C data. We used an adapted version of hg19 as the reference genome, with haplotype sequences discarded and the Epstein-Barr virus genome (NC_007605.1) added. Reads were mapped and filtered per replicate and merged afterward. We chose MAPQ≥30 to discard multi-mapping reads. Hi-C maps were normalized using the Knight-Ruiz method. A virtual 4C-seq track for the *MYCN* promoter was created by calculating the mean Knight-Ruiz-normalized Hi-C signal across three 5kb bins around the *MYCN* promoter (chr2:16,075,000-16,090,000).

2.1.7 Illumina whole-genome sequencing

DNA was obtained from harvested cells using the NucleoSpin Tissue kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). We then used the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England BioLabs Inc., Ipswich, MA, USA) to prepare the sample for sequencing on a NovaSeq S1 flow cell (2x150bp paired-end mode; Illumina Inc., San Diego, CA, USA). We performed adapter trimming, alignment, and duplicate removal as described above for ChIP-seq experiments. We called copy number variants with Control-FREEC 11.4 (51) using its default settings and identified structural variants using SvABA 1.1.1 (52) in germline mode as no normal control was available for cell lines. Variants mapping to blacklisted regions provided on https://data.broadinstitute.org/snowman/svaba_exclusions.bed were discarded. We used the unfiltered call set for further analysis.

2.1.8 Nanopore whole-genome sequencing

High molecular weight DNA was obtained from harvested cells using the MagAttract HMW DNA Kit (Qiagen N.V., Venlo, Netherlands). Additionally, we enriched the sample for fragments >10kb with the Circulomics SRE kit (Circulomics Inc., Baltimore, MD, USA). We used the Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies Ltd., Oxford, UK) to create

sequencing libraries which were then sequenced on the R9.4.1 MinION platform (FLO-MIN106, Oxford Nanopore Technologies Ltd., Oxford, UK). For the NGP cell line, we used the NucleoSpin Tissue kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) for DNA extraction and the ONT Rapid Kit for library preparation (SQK-RBK004, Oxford Nanopore Technologies Ltd., Oxford, UK). The quality of raw data was assessed using NanoPlot 1.0.0 (53). We created FASTQ files from signal-level data using Guppy 2.3.7 (Oxford Nanopore Technologies Ltd., Oxford, UK) with default settings. Methylated CpG sites were identified using Megalodon 0.1.0 (Oxford Nanopore Technologies Ltd., Oxford, UK). We used Flye 2.4.2 (54) for *de novo* assembly of amplicons (metagenomics mode, genome size 1Gb) and mapped contigs to hg19 using minimap2 2.16 (parameter settings: -ax asm5) (55).

2.1.9 Enhancer identification

Based on the assessment of the overall distribution of *MYCN* expression across non-*MYCN*-amplified neuroblastoma cell lines, we chose a sizeFactor-normalized expression of 100 to distinguish *MYCN*-expressing from non-*MYCN*-expressing cell lines. We then inspected a ±500kb window around *MYCN* and identified all regions with an H3K27ac peak in more than half of *MYCN*-expressing-non-amplified cell lines. Regions less than 2kb apart were joined. We computed the maximum difference in H3K27ac signal fold change between *MYCN*-expressing-non-amplified, and non-expressing cell lines for each region to estimate its potential contribution to *MYCN* expression. Based on this computation, we chose the five regions with the highest maximum difference fold change as candidate *MYCN*-driving enhancers. We used the JASPAR2018 (56) and JASPAR2020 (57) databases and TFBSTools 1.20.0 (matchPWM function with min.score='85%') (58) to identify potential transcription factor binding sites in candidate enhancers. To identify CRC-driven super enhancers, we merged super enhancer calls across *MYCN*-expressing-non-amplified cell lines and filtered for those that overlapped with a GATA3, HAND2, or PHOX2B ChIP-seq peak in the neuroblastoma cell line CLB-GA.

2.1.10 Copy number data analysis

We obtained a compilation of array-based copy number datasets for high-risk neuroblastomas from https://github.com/padpuydt/copynumber_HR_NB/ (59) and considered all samples annotated as *MYCN*-amplified that showed signs of *MYCN* amplification in the copy number track. To create co-amplification profiles, we counted the number of samples with amplifications in 10kb bins across the genome. Co-amplification was statistically assessed by deriving empirical one-

sided P-values from 10,000 synthetic datasets under a null hypothesis of random positioning of the *MYCN* amplicon. To this end, we randomly shifted each patient's copy number profile on chromosome 2 such that *MYCN* was still fully amplified. This allowed us to randomize positioning and retain relevant copy number statistics including patient-specific fragment counts, sizes, and local correlations. P-values were Benjamini-Hochberg-corrected for multiple comparisons.

2.1.11 Amplicon reconstruction

Starting with short read-based structural rearrangements from SvABA, we discarded small rearrangements (≤1kb) and rearrangements in ENCODE blacklisted regions (47). To discard rearrangements that were only present on a subset of amplicon copies, we filtered for breakpoints with more than 50 variant-supporting reads. We represented these data as a genome graph using gGnome 0.1 (60) with genomic segments as nodes that were connected either by a rearrangement (alternate edge) or by adjacency in the reference genome (reference edge). To focus on amplifications, only nodes with a mean coverage of 10 times the median coverage of chromosome 2 or higher were considered. To make sure that highly amplified breakpoints must be taken into account during amplicon reconstruction, we removed reference edges if the alternate edge was among the highest allele-depth quartile of edges. After pruning the graph like this, we searched for circular walks starting at *MYCN* and chose the walk with the most nodes without traversing any node twice. We used gTrack (https://github.com/mskilab/gTrack) to graphically display genome graphs.

2.1.12 Data and code availability

Sequencing data were submitted to the Sequencing Read Archive (Accession PRJNA622577). Additional data were downloaded from Gene Expression Omnibus (Accessions GSE90683, GSE80152, GSE24447, and GSE28874) and from https://github.com/padpuydt/ copynumber_HR_NB/. Processed ChIP-seq and ATAC-seq have been made available through CyVerse (https://data.cyverse.org/dav-anon/iplant/home/konstantin/helmsaueretal/). We provide a custom UCSC genome browser track hub (https://de.cyverse.org/dl/d/27AA17DA-F24C-4BF4-904C-62B539A47DCC/hub.txt). Code is available at https://github.com/henssenlab/ MYCNAmplicon.

2.2 Koche et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma

2.2.1 Patient samples

We obtained primary tumor and blood samples of patients diagnosed with neuroblastoma between 1991 and 2016 who participated in the clinical trials NB97, NB2004, and NB2004-HR of the German Society of Pediatric Oncology and Hematology. Data collection and use were approved by the ethics committees of Charité – Universitätsmedizin Berlin and of the medical faculty of Universität zu Köln. Informed consent was obtained from all patients or their legal guardians. *MYCN* status (copy number, dmin vs. HSR) was obtained from FISH staining during routine pathological workup. Only material with at least 60% pathologist-determined tumor cell content was used for sample preparation.

2.2.2 RNA interference experiments

The neuroblastoma cell line Kelly was obtained and cultured as described above. pLKO.1 vectors for shRNA-targeting of *DCLK1* (TRCN0000002145, TRCN0000002146) and *GFP* RNA were obtained from the RNAi Consortium (Broad Institute, Cambridge, MA, USA) and transduction was performed as described in Henssen et al. (61). Transduced cells were counted using a TC20 Automated Cell Counter (Bio-Rad Laboratories, Inc., Hercules, CA, USA) and seeded in 24-well plates (5000 cells per well). After 7 days, wells were imaged and surface coverage was assessed using the ImageJ plugin ColonyArea (62). Experiments were performed in biological triplicates, we tested for differences between experimental conditions using Student's t-test.

2.2.3 PCR and Sanger sequencing

For PCR experiments, we used 50–100 ng of genomic DNA, 0.4U Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Inc., Waltham, MA, USA), 0.5 μM forward and reverse primers (see attached publication for primer sequences), 200 μM deoxyribonucleotide triphosphates (Bio-Budget Technologies GmbH, Krefeld, Germany) and 4 μl 5x Phusion Green buffer (Thermo Fisher Scientific, Inc., Waltham, MA, USA). PCR products were run on 1% agarose gel and purified using the PureLink PCR Purification Kit (Thermo Fisher Scientific, Inc., Waltham, MA, USA). Capillary Sanger sequencing was performed externally (Eurofins Scientific SE, Luxemburg, Luxemburg).

2.2.4   Whole-genome sequencing and structural variant detection

We refer to the enclosed publication for a detailed description of whole-genome sequencing data generation and processing (38). In brief, we acquired tumor-normal whole-genome sequencing data for 37 patients using the HiSeq X Ten system (2x150bp paired-end mode; Illumina Inc., San Diego, CA, USA). In addition, paired tumor-normal whole-genome sequencing data for 56 patients were downloaded from the European Genome-phenome Archive (Accession EGAS00001001308) (63). Reads were aligned to hg19 using BWA-MEM 0.7.15 and BWA-ALN 0.5.9 (44). Duplicate reads were marked using Biobambam 2.0.79 (64). Copy number variation was called using Control-FREEC 10.6 (51) and ASCAT 4.0.1 (65). Amplification was defined as a total copy number of 9 or higher. Structural rearrangements were called using Novobreak 1.1.3 (66), SvABA 1.1.1 (52), Delly2 0.7.7 (67), BRASS 6.0.5 (https://github.com/cancerit/BRASS) and SMuFin 0.9.4 (68). Variant calls were filtered (BRASS BAS score 99 or higher; at least 6 supporting reads with MAPQ score higher than 60 at each breakpoint). Structural variant calls from at least four different callers were obtained. Variants were merged if breakpoints were located within 500bp of each other. Six samples were discarded due to their high number of rearrangements. One region (chr2:33,000,000-34,000,000) was excluded due to its high number of false-positive breakpoints upon visual inspection.

Structural variant data from a whole-genome sequencing dataset of 546 pediatric cancer samples (18) were downloaded from https://hgserver1.amc.nl/cgi-bin/r2/main.cgi?&dscope=DKFZ_PED &option=about_dscope.

2.2.5   RNA-seq

RNA was extracted from 37 primary neuroblastoma samples (matching the whole-genome sequenced samples) using the TRIzol reagent (Thermo Fisher Scientific Inc., Waltham, MA, USA). We assessed RNA purity using a NanoDrop 2000 spectrometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and RNA integrity using the 2100 Bioanalyzer system (Agilent Technologies Inc., Santa Clara, CA, USA) or the 4200 TapeStation instrument (Agilent Technologies Inc., Santa Clara, CA, USA). An RNA integrity number of 8 or above was required for further processing. Depletion of ribosomal RNA was performed as described previously (69). Libraries were prepared using the TruSeq Stranded mRNA kit (Illumina Inc., San Diego, CA, USA) and sequenced on the HiSeq 4000 platform (2x150bp paired-end mode; Illumina Inc., San Diego, CA, USA). In addition, RNA-seq data for 54 primary neuroblastomas were downloaded from the European Genome-phenome Archive (Accession EGAS00001001308) (63), matching almost all samples in the downloaded whole-genome sequencing data. We used STAR 2.7.0 (40)

for alignment using transcript annotations from Gencode 27, obtained read counts using featureCounts 1.6.1 (41), and normalized read counts to transcripts per million.

2.2.6   Circle-seq

We refer to the enclosed publication (38) and a protocol deposited online (70) for details on Circle-seq library preparation and sequencing. In brief, high-molecular weight DNA was extracted using the MagAttract HMW DNA Kit (Qiagen N.V., Venlo, Netherlands). To enrich for extrachromosomal circular DNA, the sample was treated with an exonuclease (Plasmid-Safe ATP-Dependent DNAse; Epicentre, Madison, WI, USA). The remaining DNA was then amplified by rolling-circle amplification (REPLI-g Mini Kit7; Qiagen N.V., Venlo, Netherlands). Libraries were prepared using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England BioLabs, Inc., Ipswich, MA, USA) and sequenced on the MiSeq platform (2x150bp paired-end mode; Illumina, Inc., San Diego, CA, USA), HiSeq 4000 platform (2x125bp paired-end mode; Illumina, Inc., San Diego, CA, USA) or NextSeq 500 platform (2x150 bp paired-end mode; Illumina, Inc., San Diego, CA, USA). Reads were aligned to hg19 (BWA 0.7.15) (44). Duplicate reads were removed (Picard 2.16.0). High-coverage regions were called using findPeaks from Homer 4.11 (71). For the edges of each of these regions, the number of outward-facing split reads or read pairs was counted to determine the number of circle-supporting reads. Based on a background distribution of outward-facing split reads and read pairs in whole-genome sequencing data, only those high-coverage regions with a significantly enriched number of outward-facing split reads or read pairs (empirical P=0.01) at their edges were considered as putative circle regions for further analysis. The same circle-calling algorithm was run on whole-genome sequencing data to screen for circular DNA in whole-genome sequencing libraries. EccDNA and EcDNA was defined by non-overlap and overlap with amplification respectively.

2.2.7   Clusters of interchromosomal rearrangements

Clusters of interchromosomal rearrangements were defined as all genomic regions harboring at least three interchromosomal rearrangement breakpoints in a 4 Mb sliding window. If two chromosomes were connected by more than four rearrangements, these rearrangements were discarded during cluster detection. Rearrangements between two 10Mb regions were merged for cluster detection. We discarded chromosomes with more than 25 inter-chromosomal rearrangements. We evaluated this method using 500 synthetic datasets with randomized

breakpoints across the mappable genome. For a threshold of three rearrangements within 4Mb, we estimated a false discovery rate of 0.13.

We tested for an association of clusters of interchromosomal rearrangements (based on the union set of rearrangement calls by all structural variant callers) and circularized regions. The relative overlap of all clusters with circularized regions was computed for each sample. To test this overlap for statistical significance, we created 2000 synthetic datasets with random cluster positioning in the genome (excluding poorly or non-assembled regions and the ENCODE DAC blacklist (47)) and computed their relative overlap with circular DNA in matching patients. Comparing the real overlap to the overlap distribution for randomly positioned clusters, we derived one-sided empirical P-values. As this was performed for different circle classes (amplified ecDNA vs. non-amplified eccDNA) and circle calling methods (Circle-seq, whole-genome sequencing) separately, we accounted for multiple comparisons by Benjamini-Hochberg correction.

We then investigated whether breakpoint partners of breakpoints in clusters of rearrangements were enriched in the vicinity of specific types of cancer-related genes as defined by the COSMIC data base (all COSMIC genes, COSMIC oncogenes, and COSMIC tumor suppressor genes (72)). We calculated the median distance to the closest member of the respective gene set. We tested against the null hypothesis of independent positioning of breakpoints and genes and derived an empirical P-value from the distribution of median distances in 500 synthetic datasets with breakpoint positions randomly distributed across the non-blacklisted genome. We used Benjamini-Hochberg correction to account for individually testing the null hypothesis for the three types of genes.

Gene expression was quantified relative to other samples in the cohort using the modified z-score (the distance to the median normalized by the median absolute deviation from the median) of expression in transcripts per million. To screen for breakpoint-associated differential expression, we focused on protein-coding genes within 2Mb of each breakpoint.

We compared the survival of different subgroups using the log-rank test. To define *MYCN*-associated clusters, we considered all clusters within ±1Mb around *MYCN*.

2.2.8 Data and Code Availability

Sequencing data are available at the European Genome-phenome Archive (Accession numbers EGAS00001001308 and EGAS00001004022). We provide visualized somatic structural variation data at https://kons.shinyapps.io/trees/. Code and further data are available at https://github.com/henssenlab/TreeShapedRearrangements.

2.3 Hung et al. ecDNA hubs drive cooperative intermolecular oncogene expression

2.3.1 TR-14 cell line

The neuroblastoma cell line TR-14 was a gift from J. J. Molenaar (Princess Máxima Center for Pediatric Oncology, Utrecht, Netherlands). Cells were maintained in Roswell Park Memorial Institute 1640 medium (RPMI-1640; Thermo Fisher Scientific Inc., Waltham, MA, USA) with 1% Penicillin/Streptomycin (Thermo Fisher Scientific Inc., Waltham, MA, USA) and 10% Fetal Calf Serum (Thermo Fisher Scientific Inc., Waltham, MA, USA). We used short tandem repeat genotyping (IDEXX BioResearch, Westbrook, ME, USA) to confirm cell line identity. Contamination with *Mycoplasma sp.* was ruled out using the Lonza MycoAlert system (Lonza Group Ltd., Basel, Switzerland).

2.3.2 Illumina whole-genome sequencing

Illumina whole-genome sequencing libraries were prepared from DNA extracted with the NucleoSpin Tissue kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England BioLabs, Inc., Ipswich, MA, USA). Libraries were sequenced on the NovaSeq 6000 platform (2x150bp reads; Illumina, Inc., San Diego, CA, USA). We trimmed adapters (BBMap 38.58), aligned reads to hg19 (BWA-MEM 0.7.15 with default parameters) (44) and removed duplicate reads (Picard 2.20.4). We computed coverage as read counts per million reads in 20bp bins (deepTools 3.3.0) (46). Copy number variation profiles were created with QDNAseq 1.22.0 (73) by binning primary alignments of MAPQ≥20 in 10kb bins, default filtering, additional filtering of bins with more than 5% undefined bases, GC-correction, and normalization. Then, copy number segments were called with the circular binary segmentation method implemented in QDNAseq (no transformation of the binned signal, parameter alpha=0.05).

2.3.3 Nanopore whole-genome sequencing

For Nanopore whole-genome sequencing, high-molecular DNA was extracted using the MagAttract HMW DNA Kit (Qiagen N.V., Venlo, Netherlands) and enriched for molecules >10kb (Circulomics SRE kit, Circulomics Inc., Baltimore, MD, USA). A sequencing library was created with the Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies Ltd., Oxford, UK) and then sequenced on an R9.4.1 MinION flowcell (FLO-MIN106; Oxford Nanopore Technologies Ltd., Oxford, UK). We used Guppy 2.3.7 (Oxford Nanopore Technologies Ltd.,

Oxford, UK) with default parameters for basecalling. We mapped reads to hg19 using NGMLR 0.2.7 (74) and called structural variants using Sniffles 1.0.11 (74) with parameters --min_length 15 --genotype --min_support 3 --report_seq.

2.3.4 Reconstruction of amplicon repertoire

To reconstruct the structure of all amplicons, we filtered Nanopore structural variants for size (>10kb) and variant read support (≥15). From this set, we created a genome graph using gGnome 0.1 (60). This graph treated genomic regions as nodes and structural variants as edges. We filtered non-amplified regions (mean short-read whole-genome sequencing coverage less than ten times the median chromosome 2 coverage) and partitioned the graph into clusters of regions that were mutually connected by a path in the graph. We identified which clusters contained the amplified oncogenes *MYCN*, *CDK4*, *MDM2*, and *ODC1*. Then, we selected circular paths traversing the regions in those clusters that could account for the observed copy number profile. gTrack (no version number available; https://github.com/mskilab/gTrack) was used to display genome graphs.

2.3.5 Hi-C and ChIP-seq data analysis

Hi-C libraries were prepared as described in Helmsauer et al. (37) and sequenced on the Illumina Hi-Seq platform (2x100bp paired-end mode; Illumina, Inc., San Diego, CA, USA) at a depth of 433.7 million read pairs. We used Juicer 1.19.02 (50) with BWA 0.7.17 (44), mapping short reads to hg19 (with haplotype sequences removed and the Epstein-Barr virus genome (NC_007605.1) added). Replicates were independently mapped, filtered by MAPQ≥30, and then merged. Per-chromosome Knight-Ruiz normalization was performed.

We obtained TR-14 H3K27ac ChIP-seq data from Gene Expression Omnibus (GSE90683) (14). We trimmed adapters using BBMap 38.58. Reads were mapped to hg19 with BWA-MEM 0.7.15 (44) using default parameters. We used deepTools 3.3.0 (46) to compute ChIP-seq coverage by extending reads to 200bp, removing reads mapping to ENCODE DAC blacklisted regions (47), and normalizing read counts to counts per million in 10bp bins. We used LILY (14) with default parameters to call enhancers and super-enhancers. We chose LILY-defined enhancers as boundaries to define the extent of the *HPCAL1* gene- und enhancer-rich region (chr2:10424449-10533951).

To investigate the overall effect of the H3K27ac signal on *in trans* interaction strength, all 5kb bin pairs on different amplicons were inspected for their Knight-Ruiz-normalized Hi-C signal. We used a threshold of mean ChIP-seq fold change over input larger than 3 to define low-H3K27ac

and high-H3K27ac bins. We used the Welch two-sample t-test to compare interaction of bin pairs of high-H3K27ac+high-H3K27ac signal against all other pairs of bins.

3 Results

In the following, key results of the publications presented in this thesis (37-39) are summarized, focusing on the structure, sequence, and epigenetics of *MYCN* amplification in neuroblastoma.

3.1 Helmsauer et al. Enhancer hijacking determines extrachromosomal circular *MYCN* amplicon architecture in neuroblastoma

In Helmsauer et al. (37), we characterized the chromatin landscape around *MYCN* in neuroblastoma. We first catalogued *MYCN*-driving enhancers in non-*MYCN*-amplified neuroblastoma, comparing H3K27ac ChIP-seq data for five *MYCN*-expressing and seven non-*MYCN*-expressing neuroblastoma cell lines (Fig. 1a). Computing the difference in mean H3K27ac fold change between the two groups, we identified five intergenic regions, e1-e5, whose H3K27ac status correlated with *MYCN* expression (Fig. 1b). Consistent with their role as potential enhancers, four of these regions were bound by the noradrenergic CRC transcription factors PHOX2B, GATA3, and HAND2 in ChIP-seq data for the neuroblastoma cell line CLB-GA (Fig. 1b). To investigate the role of these enhancers in the context of *MYCN* amplification, we mapped the extent of the
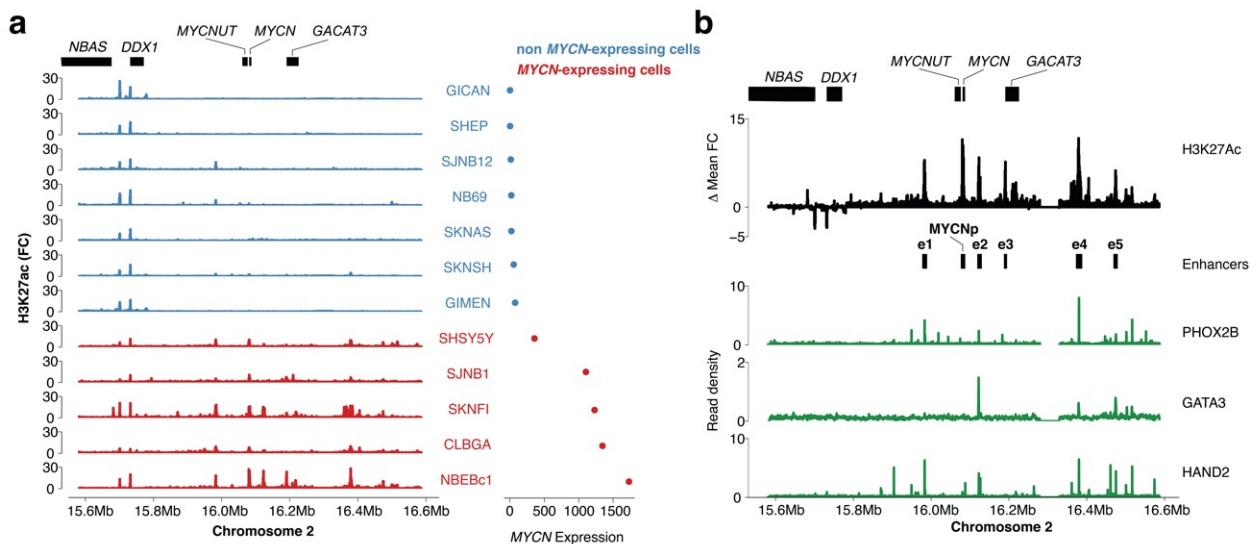
Figure 1 Mapping *MYCN*-driving enhancers in neuroblastoma a H3K27ac ChIP-seq signal (FC, fold change over input) and RNA-seq based *MYCN* expression (size-factor normalized read counts) for neuroblastoma cell lines without *MYCN* amplification. **b** Top: Difference in mean H3K27ac ChIP-seq fold change over input between *MYCN*-expressing and non-*MYCN*-expressing cell lines (Δ Mean FC). The *MYCN* promotor (MYCNp) and five expression-associated enhancers (e1-e5) are marked. Bottom: CRC transcription factor ChIP-seq (PHOX2B, GATA3, and HAND2) in the neuroblastoma cell line CLB-GA. Adapted from Helmsauer et al. (37).
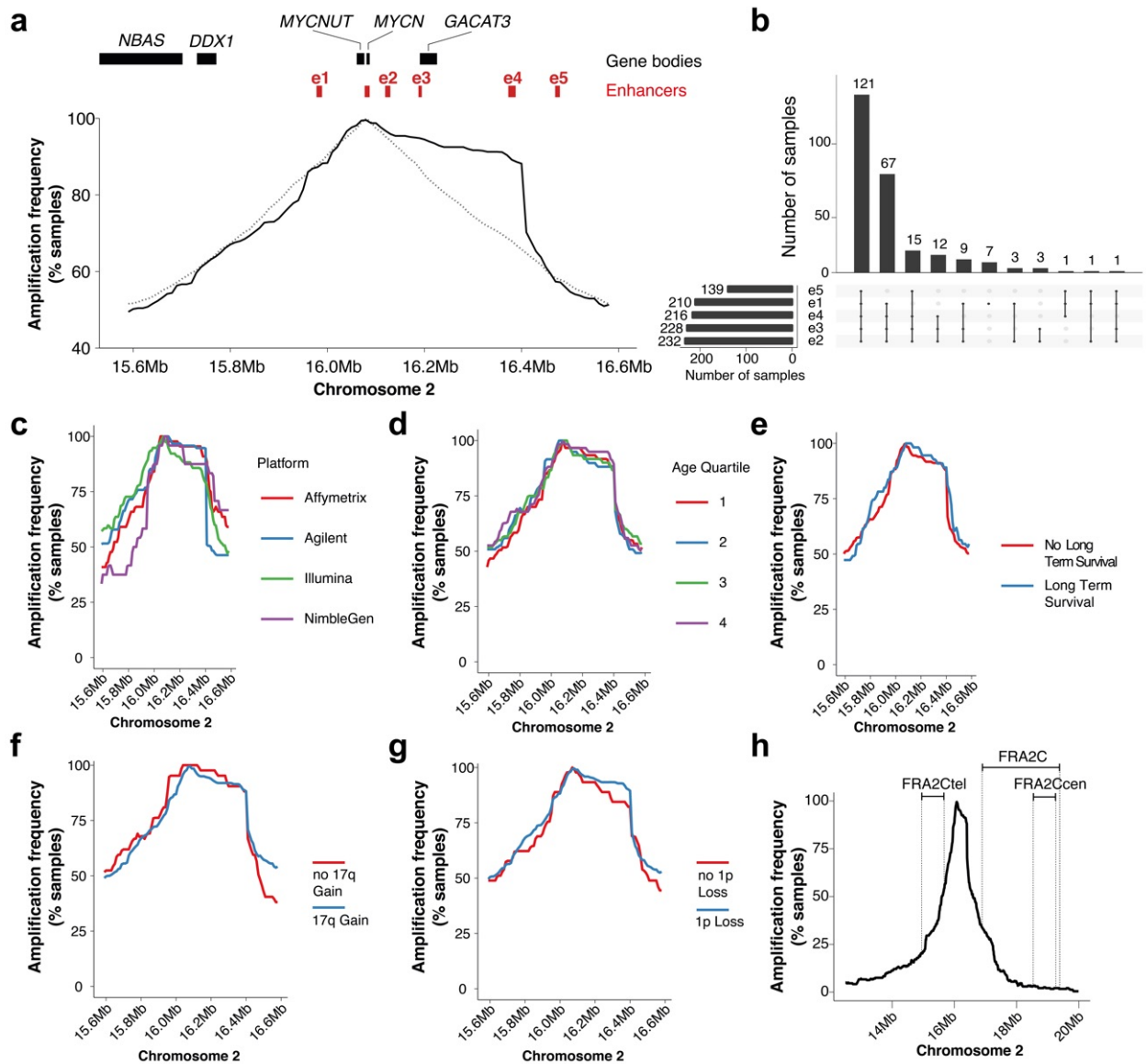
Figure 2 Systematic inclusion of the e4 enhancer on the *MYCN* amplicon a Co-amplification profile around *MYCN* for 240 *MYCN*-amplified neuroblastomas (solid line) and expected profile for random *MYCN* amplicon boundaries (dashed line). **b** Upset plot summarizing co-amplification patterns of local enhancers. **c-g** Co-amplification profile by experimental platform, age, long-term survival, 17q gain, and 1p loss. **h** Co-amplification profile and chromosomal fragile sites proximal to the *MYCN* locus. Adapted from Helmsauer et al. (37).

24

MYCN amplicon in a copy-number dataset of 240 *MYCN*-amplified neuroblastomas. This revealed a highly asymmetric co-amplification pattern where a 290kb region downstream of *MYCN* was co-amplified with *MYCN* in 90% of cases (Fig. 2a). This pattern was robust across different experimental platforms (Affymetrix SNP array, Agilent aCGH platform, Illumina SNP array, NimbleGen aCGH platform), clinical variables (age, long-term survival), and genetic subtypes (17q gain, 1p loss; Fig. 2c-g). The amplification boundaries could not be accounted for by genomic fragile sites (Fig. 2h). However, taking the local enhancer landscape into consideration showed that the e4 enhancer was co-amplified with *MYCN* in 90% of cases (Fig. 2a,b), accounting for the asymmetric co-amplification pattern. Randomization analysis confirmed that this was 1.3-fold more than expected by a random distribution of amplicon boundaries around *MYCN*, establishing statistically significant co-amplification (empirical P=0.0003). This suggests a selective pressure to include certain enhancers, particularly e4, on the *MYCN* amplicon.

We then examined the epigenetic characteristics of the e4 locus when it is co-amplified with *MYCN* using the neuroblastoma cell lines Kelly and NGP. ATAC-seq and ChIP-seq experiments showed that e4 was characterized by open chromatin and enhancer histone marks H3K27ac (histone H3 lysine residue 27 acetylation) and H3K4me1 (histone H3 lysine residue 4 monomethylation) in this context (Fig. 3). Additionally, the locus spatially interacted with the *MYCN* gene as measured by *MYCN* promotor-anchored 4C-seq (Fig. 3), substantiating that e4 retains the characteristics of an active enhancer in the context of amplification.

With the vast majority but not all *MYCN* amplicons containing e4, we asked how the selective pressure to co-amplify this enhancer could be evaded in some cases. Analyzing copy number data showed that samples without e4 co-amplification were almost twice as likely to harbor additional non-contiguous amplified segments than samples with e4 co-amplification (66.7% vs. 35.7%, Fisher's exact test P=0.003; Fig. 4), indicating more complex amplicon structures when the amplicon lacked the local enhancer. This led us to hypothesize that e4 loss can only result in a functional amplicon if other genomic regions are co-amplified and contribute distal enhancers to the amplicon.

We, therefore, inspected the amplicon structure in two *MYCN*-amplified neuroblastoma cell lines that did not co-amplify the e4 enhancer. One of them, CHP-212, contained *MYCN*-ecDNA as
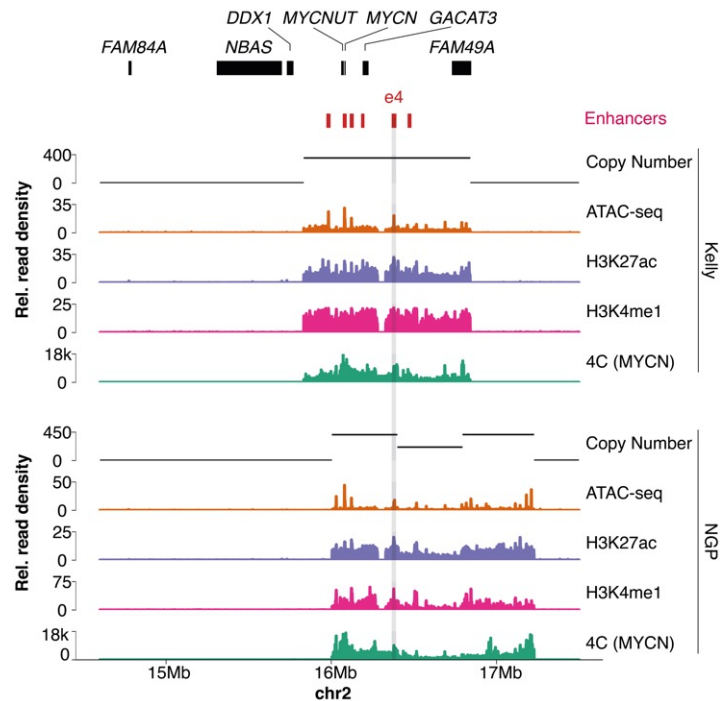
Figure 3 e4 is an active enhancer on the *MYCN* amplicon Copy number variation, ATAC-seq, H3K27ac and H3K4me1 ChIP-seq, and *MYCN* promoter-anchored 4C-seq signal for the neuroblastoma cell lines Kelly and NGP around the *MYCN* locus. Adapted from Helmsauer et al. (37).
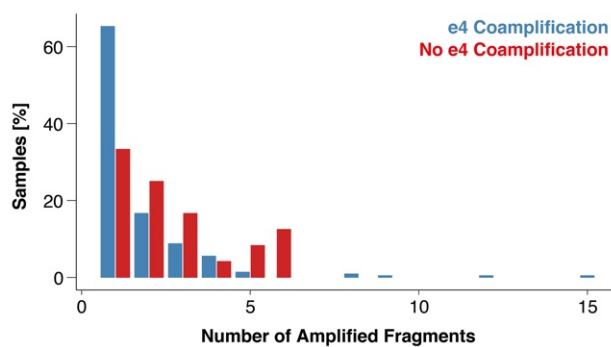

Figure 4 e4 loss is associated with more complex amplification Number of non-contiguously amplified genomic fragments depending on e4 coamplification status in a copy number dataset of 240 *MYCN*-amplfied neuroblastomas. Adapted from Helmsauer et al. (37).
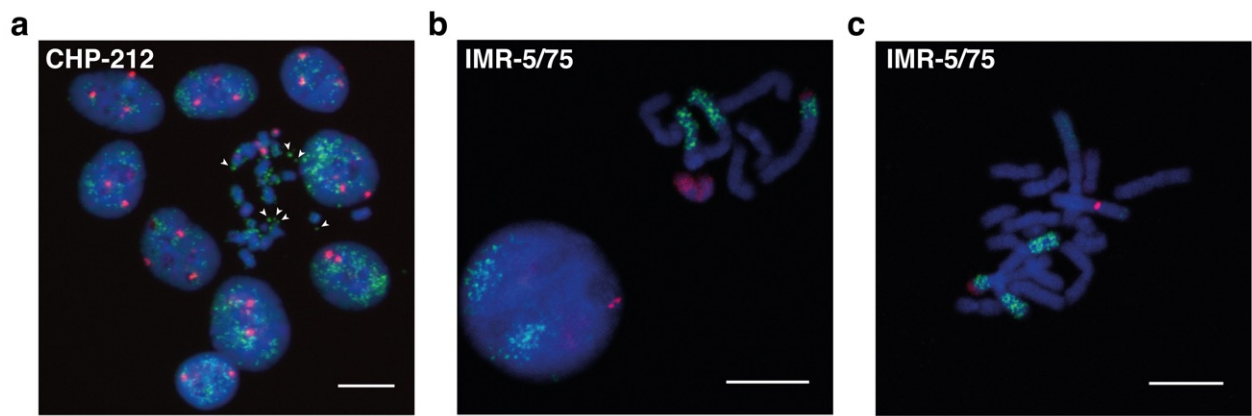
Figure 5 *MYCN* ecDNA and HSRs in CHP-212 and IMR-5/75 FISH for CHP-212 (**a**) and IMR-5/75 (**b-c**) metaphase spreads. *MYCN* probe in green, chromosome 2 centromere probe in red. Scale bars indicate 10μm. Adapted from Helmsauer et al. (37).

indicated by FISH (Fig. 5a). The other one, IMR-5/75, contained *MYCN*-HSRs (Fig. 5b,c). We reconstructed the structures of both *MYCN* amplicons using structural variant and copy number analysis based on short-read whole-genome sequencing data (Fig. 6a, e). We also resolved both loci using *de novo* assembly of long-read Nanopore sequencing reads, validating our reconstructions (Fig. 6d, h). In IMR-5/75, we identified one complex amplicon structure common to all HSRs. It consisted of five distal fragments fused to the *MYCN* locus (Fig. 6a,b). This included a segment of the *ALK* gene containing a super-enhancer in *MYCN*-expressing cell lines which, like e4, was bound by CRC transcription factors in CLB-GA (Fig. 7a). In CHP-212, we also found that distal CRC-driven super-enhancers were juxtaposed with *MYCN* by structural variation on the amplicon (Fig. 6e,f). ATAC-seq and ChIP-seq showed that the super enhancers were highly accessible and marked by H3K27ac on both amplicons (Fig. 7a,b). Taken together, this suggests that enhancer hijacking can compensate for the loss of key local enhancers and underscores the relevance of CRC-driven enhancers in the formation of a functional *MYCN* amplicon.

To examine the three-dimensional conformation of the CHP-212 and IMR-5/75 amplicons, we performed Hi-C and mapped Hi-C reads to the reconstructed amplicons (Fig. 6c, g). This validated the proposed amplicon structures showing no strong off-diagonal signal indicative of major structural rearrangements missing from the reconstructions. Corner signals supported circularity in CHP-212 and tandem duplication-like structure in IMR-5/75. Hi-C also revealed that topologically associated domains (TADs) formed both on ecDNA in CHP-212 and the HSRs in IMR-5/75. In particular, structural rearrangements led to the formation of new chromatin domains (neo-TADs) that brought the hijacked enhancers in close spatial proximity with the *MYCN*
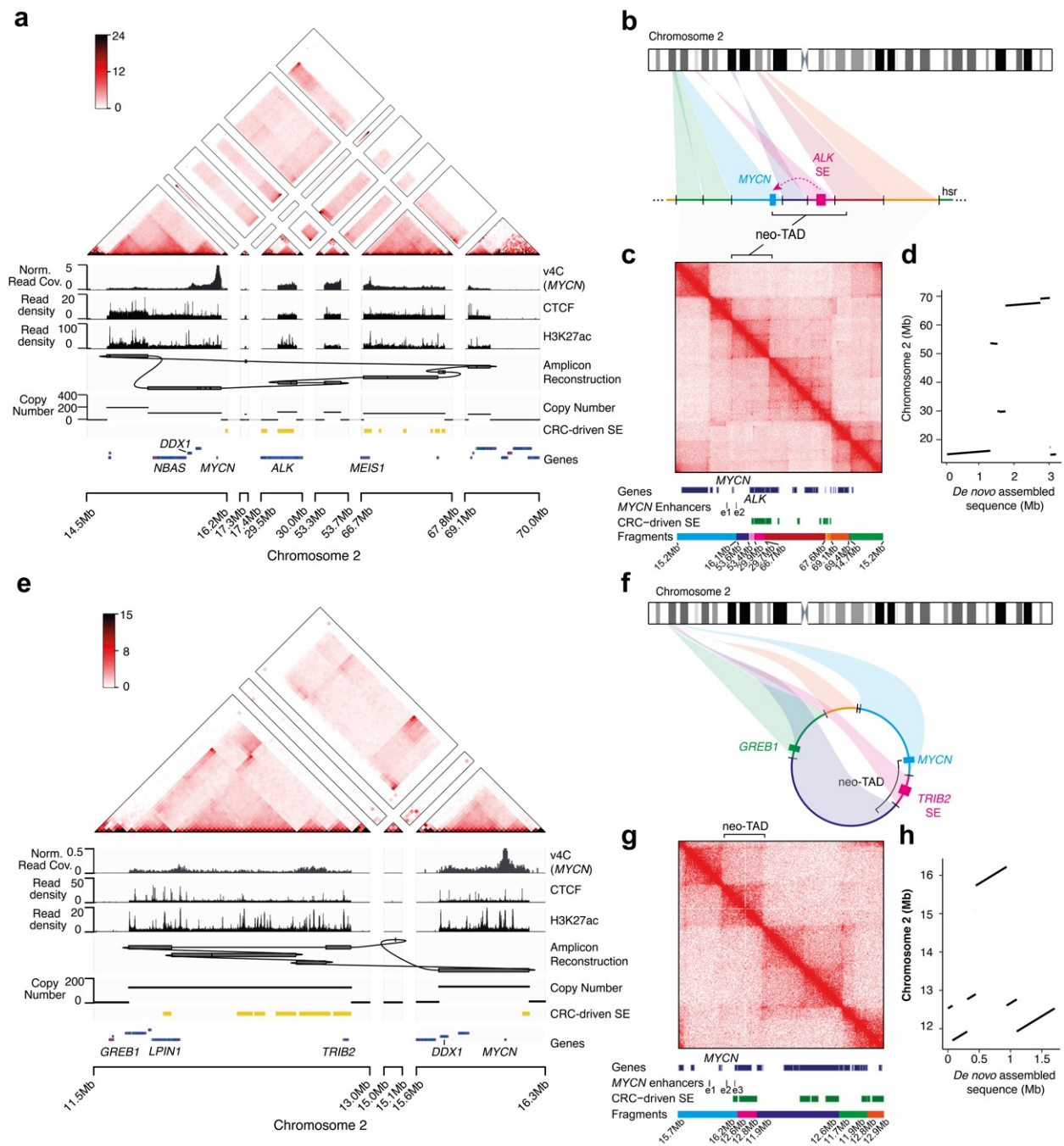
Figure 6 Reconstruction, epigenetic landscape and three-dimensional conformation of structurally complex amplicons a,e Reconstruction of the *MYCN* amplicon for IMR-5/75 (**a**) and CHP-212 (**e**) with Hi-C signal, virtual 4C-seq signal anchored at the *MYCN* locus (v4C), CTCF and H3K27ac ChIP-seq signal, copy number profile, and CRC-driven superenhancers (SE) for the genomic fragments forming the amplicon. **b,f** Schematics illustrating amplicon architectures in IMR-5/75 (**b**) and CHP-212 (**f**). **c,g** Mapping of Hi-C reads to the reconstructed IMR-5/75 (**c**) and CHP-212 (**g**) amplicons. **d,h** Mapping of de novo assembled IMR-5/75 (**d**) and CHP-212 (**h**) amplicons to the reference genome. Adapted from Helmsauer et al. (37).
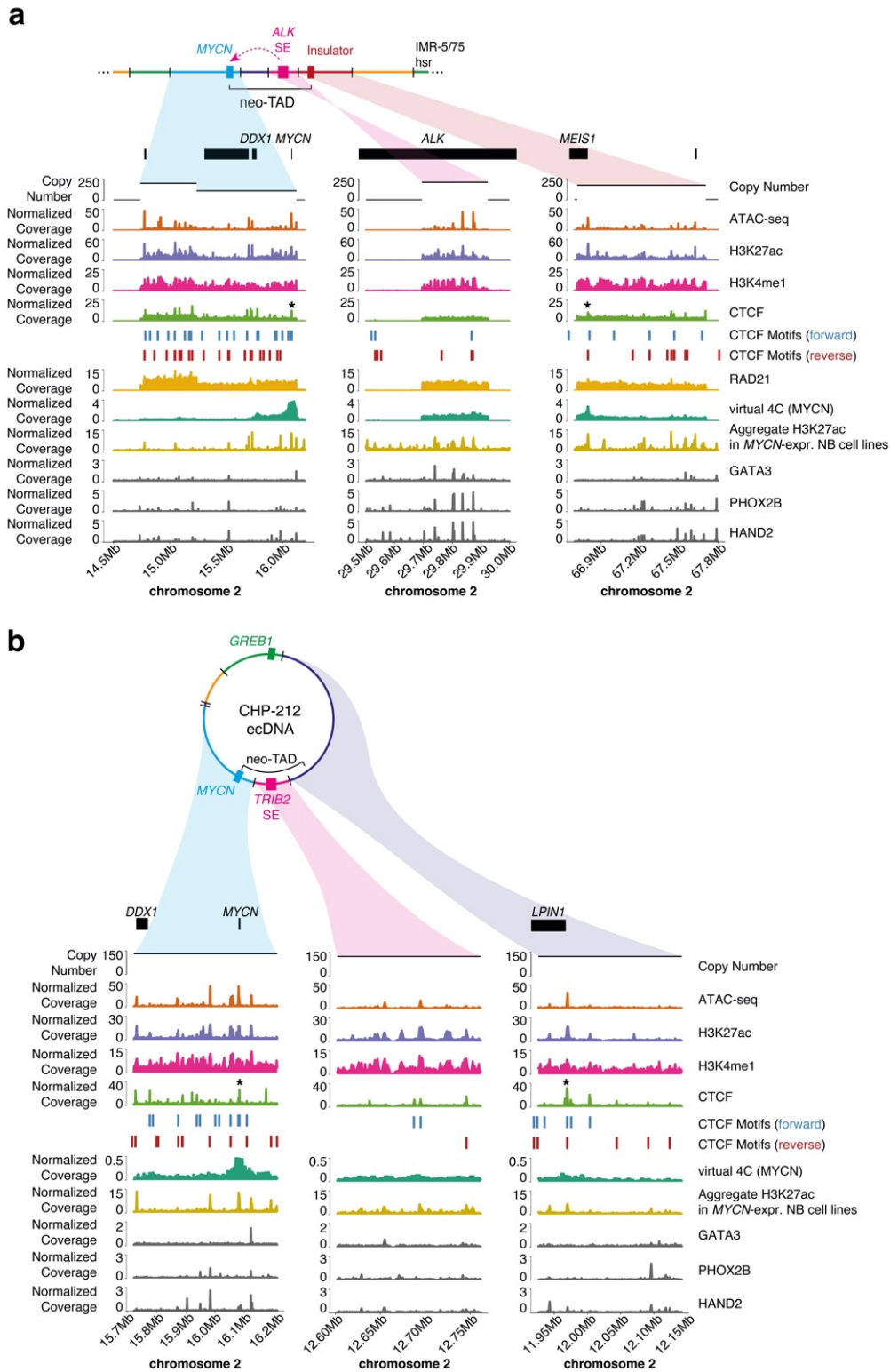
Figure 7 New gene-regulatory neighborhoods on the *MYCN* amplicon Fragments forming neo-TADs around *MYCN* in IMR-5/75 (**a**) and CHP-212 (**b**) are depicted. Top to bottom: copy number variation, ATAC-seq signal, H3K27ac, and H3K4me1, and CTCF ChIP-seq signal, CTCF binding motifs, Rad21 ChIP-seq signal, *MYCN*-anchored virtual 4C-seq signal (i.e. mean Knight-Ruiz normalized read counts anchored at chr2:16,075,000-16,085,000), aggregate H2K27ac ChIP-seq signal in *MYCN*-expressing non-amplified neuroblastoma cell lines, and ChIP-seq signal for three CRC transcription factors (GATA3, PHOX2B, and HAND2) in CLB-GA. Neo-TAD insulators are marked by an asterisk. Adapted from Helmsauer et al. (37).

29
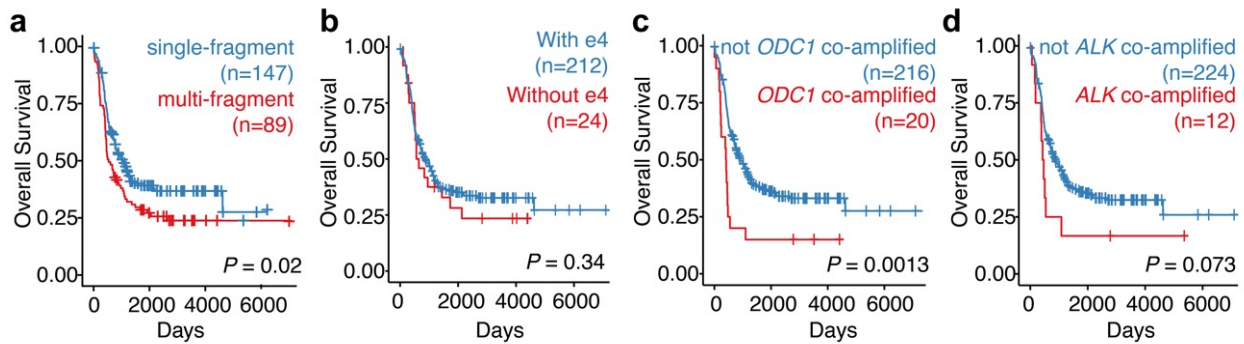
Figure 8 Prognostic relevance of amplicon architecture in *MYCN*-amplfied neuroblastoma Kaplan Meier plots for overall survival depending on **a** single- vs. multi-fragment amplification, **b** e4 enhancer co-amplification, **c** *ODC1* co-amplification, and **d** *ALK* co-amplification. P-values are based on the Log-Rank Test. Adapted from Helmsauer et al. (37).

promoter. The boundaries of these chromatin domains were marked by CTCF ChIP-seq peaks (Fig. 7a, b). Notably, these insulator elements originated from distal parts of the genome, different from the *MYCN* locus and the super enhancer loci. This illustrated that not only enhancers but also other gene regulatory elements were hijacked to form new oncogene-driving chromatin domains. More generally, both ecDNA and HSRs adhered to basic rules of chromatin topology likely contributing to *MYCN* expression in the context of amplification.

Finally, we explored the clinical relevance of the contents of *MYCN* amplification in neuroblastoma. In line with previous analyses (75), we found that complex amplification, i.e. more than one contiguous amplified fragment in the reference genome, was associated with worse prognosis within *MYCN*-amplified neuroblastoma (Hazard Ratio 1.5, 95%-confidence interval [1.1-2], Log-Rank Test *P*=0.02; Fig. 8a). We did not find evidence that the presence or absence of e4 on the amplicon affected survival (Hazard Ratio 1.3, 95%-confidence interval [0.78-2.1], Log-Rank Test *P*=0.34; Fig. 8d). However, some of the prognostic relevance of complex amplification likely stems from the co-amplification of specific oncogenes that define prognostic subgroups. *ODC1* co-amplification, for instance, occurred in 9% of *MYCN*-amplified neuroblastomas and was significantly associated with worse prognosis within this group (Hazard Ratio 2.3, 95%-confidence interval [1.4-3.7], Log-Rank Test *P*=0.0013; Fig. 8b). A similar, albeit not statistically significant, trend was observed for *ALK* co-amplification (5% of cases; Hazard Ratio 1.8, 95%-confidence interval [0.94-3.4], Log-Rank Test *P*=0.073; Fig. 8c). Taken together, amplicon gene content correlated with clinical prognosis. Whether local or distal enhancers drove the amplified *MYCN* gene, however, was not clinically relevant.

3.2 Koche et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma

In Koche et al. (38), we inspected whole-genome sequencing data for a cohort of 91 primary neuroblastomas to characterize genomic changes related to ecDNA. Here, structural variant analysis revealed a novel pattern of structural rearrangements defined by clusters of interchromosomal rearrangements. Each of these clusters connected one genomic region with loci on several other chromosomes (Fig. 9a). We identified this pattern in 28.5% (26/91) of neuroblastomas and also found clusters of interchromosomal rearrangements in 8.9% (49/546) of cases in an independently processed pediatric pan-cancer dataset (18), showing that the pattern can be found across pediatric cancers and is not specific to our method of structural variant analysis.
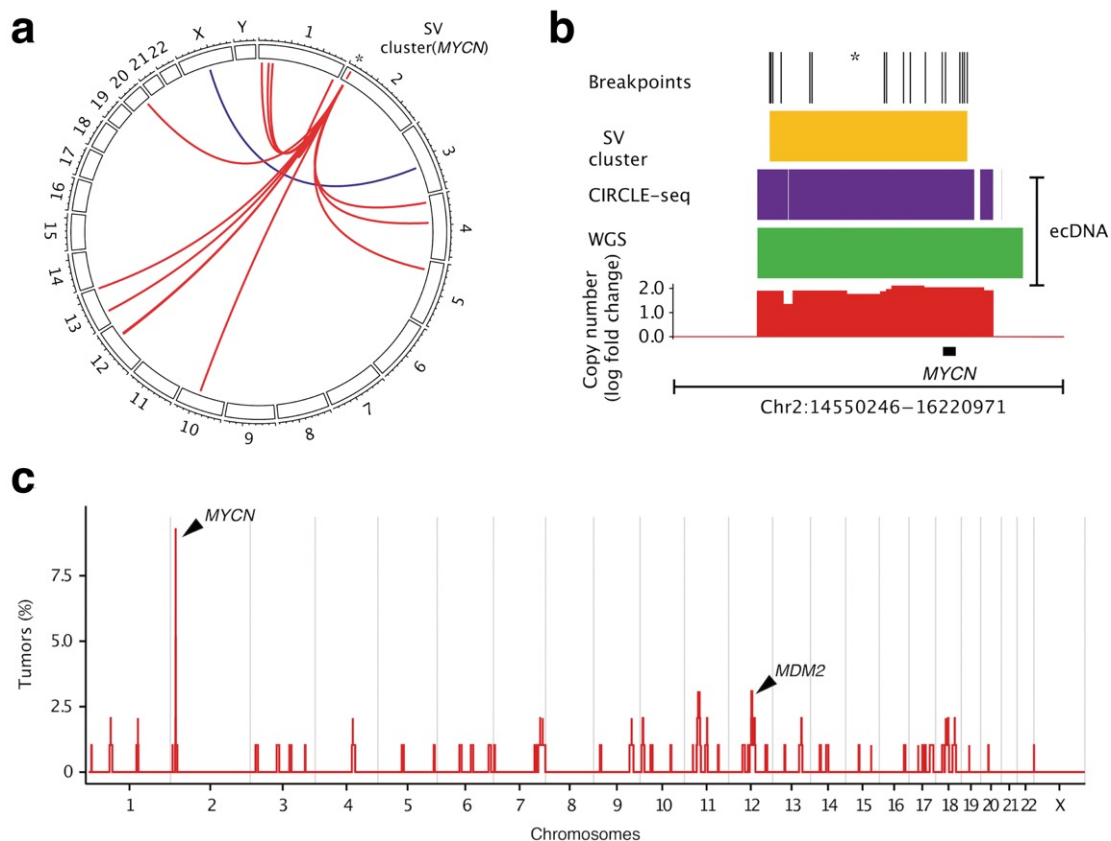
Figure 9 Clusters of interchromosomal rearrangements a Exemplary Circos plot depicting interchromosomal structural rearrangements in one primary neuroblastoma (CB2013). A cluster of interchromosomal rearrangements at the *MYCN* locus with its associated rearrangements is marked in red. **b** Breakpoint distribution, cluster region, Circle-seq ecDNA calls, whole-genome sequencing ecDNA calls and copy number profile for the cluster of interchromosomal rearrangements in CB2013. **c** Genome-wide recurrence of clusters of interchromosomal rearrangements in 91 primary neuroblastomas. Adapted from Koche et al. (38).

The clusters recurrently originated in the vicinity of amplified oncogenes like *MYCN* or *MDM2* (Fig. 9b,c), raising the question whether they were associated with ecDNA. We therefore integrated clusters of interchromosomal rearrangements with Circle-seq data, a method to enrich for extrachromosomal circular DNA using exonuclease treatment and rolling circle amplification before sequencing (Fig. 9b). 5 out of 6 clusters in patients, for which matching Circle-seq data was available, overlapped with Circle-seq-defined ecDNA. The proportional overlap of clusters and Circle-seq defined ecDNA was significantly higher than expected by chance (empirical P=9.995e-4). HSRs are rare in primary neuroblastoma, with FISH indicating that 89.5% (17/19) of *MYCN* amplification occurred as ecDNA in our cohort. Therefore, we also called potential ecDNA from whole-genome sequencing data for the whole cohort. Again, we found significant overlap between ecDNA and clusters of interchromosomal rearrangements (empirical P=9.995e-4). With ecDNA as the origin of many clusters of interchromosomal rearrangements, some of the rearrangements might represent ecDNA containing fragments from different parts of the genome. Other rearrangements could arise from the integration of ecDNA into chromosomal loci, raising the question of the functional consequences of cluster-associated rearrangements.

We therefore investigated the partner sites of clustered rearrangements, i.e. the breakpoint partners distal to the cluster region. Across the whole cohort, we found that partner sites were significantly enriched in the vicinity of cancer-relevant genes (empirical P=0.033) and tumor suppressor genes (empirical P=0.033) as defined by the COSMIC gene census (72). In one tumor, for instance, we identified a potential integration of *MYCN*-containing ecDNA into the chromosomal *DCLK1* locus. We validated this integration using PCR, Sanger sequencing, and visual inspection of breakpoint-spanning reads (Fig. 10a). Matching RNA-seq data showed that the integration went along with the disruption of *DCLK1* expression in this tumor (Fig. 10b). With *DCLK1*-knockdown increasing clonogenicity of the Kelly neuroblastoma cell line in at least one RNA interference experiment (Fig. 10c), ecDNA integration could have been a cancer-driving event in this tumor. In addition to gene disruption, we also found upregulation of gene expression in the vicinity of rearrangement target sites, in some cases affecting oncogenes like *TERT* (Fig. 10d), which could be due to co-amplification of genes on complex ecDNA or ecDNA-derived enhancers driving chromosomal genes in the vicinity of an integration site.
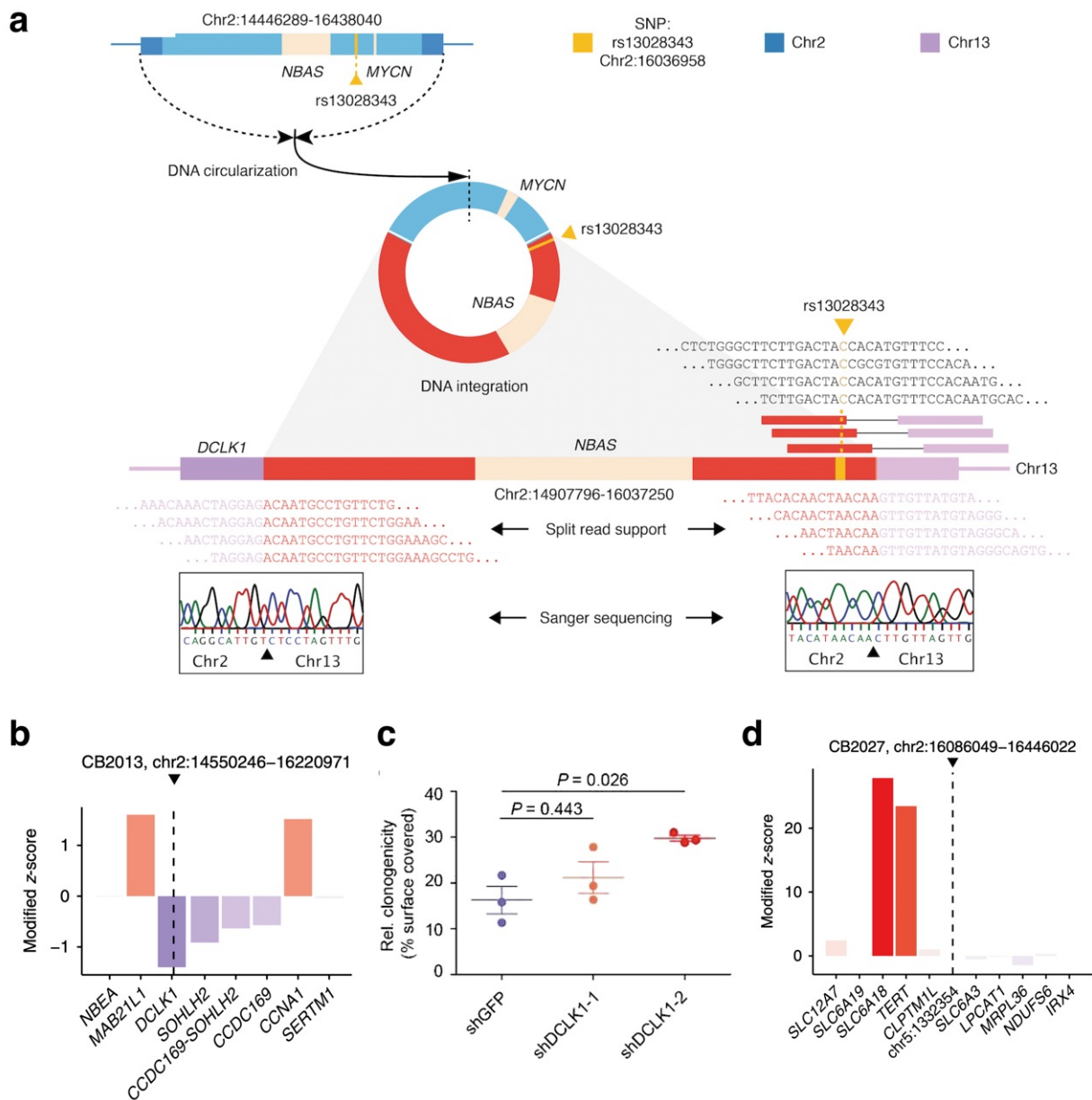
Figure 10 Clusters of interchromosomal rearrangements are associated with altered gene expression
a A case of ecDNA re-integration in one primary neuroblastoma (CB2013). Parts of a *MYCN* ecDNA (red) have been integrated into the *DCLK1* locus on chromosome 13 (purple). Split-read support of the re-integration and its validation through Sanger sequencing are depicted below. **b** Expression of *DCLK1* and other genes around the re-integration site (dashed line) for CB2013 relative to the other neuroblastomas in the cohort. **c** Clonogenicity, quantified as percent surface coverage of the cell culture plate, after experimental knockdown of *DCLK1* using RNA interference in the Kelly cell line. P-values based on Student's t-test. **d** Gene expression in the vicinity of another partner site of a clustered rearrangement distal to the cluster region in another primary neuroblastoma. The expression of nearby genes, including *TERT*, is quantified relative to other neuroblastomas in the cohort. Adapted from Koche et al. (38).

33
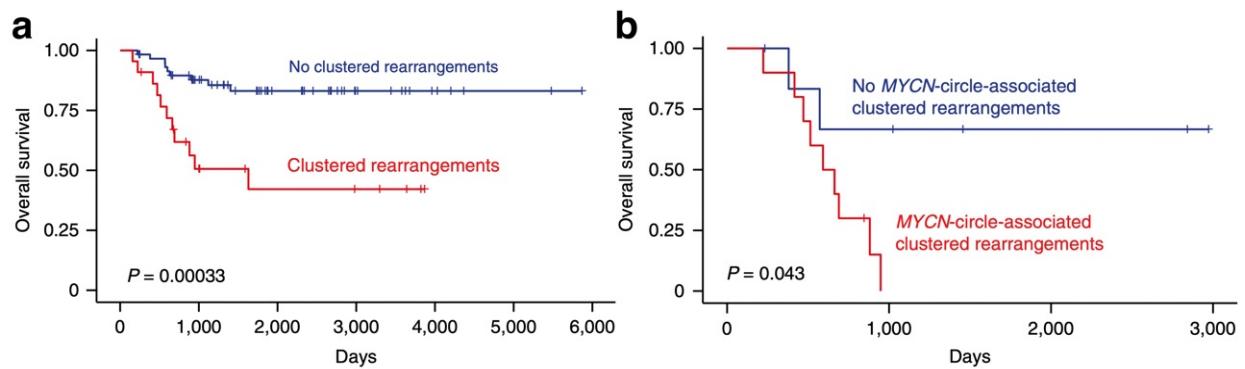
Figure 11 Clusters of interchromosomal rearrangements are associated with poor prognosis Kaplan-Meier plots for overall survival depending on **a** the presence of clusters of interchromosomal rearrangements across the entire cohort and **b** the presence of clusters of interchromosomal rearrangements at the *MYCN* locus for *MYCN*-amplified neuroblastomas. P-values are based on the log-rank test. Adapted from Koche et al. (38).

Tumors whose ecDNA is involved in genome remodeling through clusters of interchromosomal rearrangements, might possess an additional mechanism of genome remodeling and might therefore be able to adapt more readily to changing environments, resulting in worse clinical outcomes. Indeed, across our entire cohort, clusters of interchromosomal rearrangements were associated with clearly worse clinical prognosis (Fig. 11a). Crucially, this prognostic relevance also holds within *MYCN*-amplified neuroblastoma (Fig. 11b) providing one possible explanation for clinical heterogeneity within this subgroup which previously could not be genetically accounted for.

3.3 Hung et al. ecDNA hubs drive cooperative intermolecular oncogene expression

How ecDNA organizes in the nucleus to enable strong and stable expression of oncogenes has been hardly understood. In Hung et al. (39), FISH targeting ecDNA in cell lines and primary tumors revealed that ecDNA is not uniformly distributed in the nucleus but clusters together in ecDNA hubs. Hub formation shown to drive oncogene expression and could be disrupted by treatment with the bromodomain inhibitor JQ1. In SNU16, a gastric cancer cell line that harbors independent *MYC*- and *FGFR2*-ecDNA, interfering with enhancer activity on *FGFR2*-ecDNA using CRISPR interference led to decreased expression of *MYC*. Further analysis showed that this effect was unlikely to be mediated by decreased *FGFR2* expression because experimentally lowering *FGFR2* expression did not lower *MYC* expression. This suggests that enhancers on *FGFR2*-ecDNA potentially activate *MYC* through intermolecular *in trans* contacts within ecDNA hubs. As these experiments were unrelated to the author's contributions, we refer to the enclosed publication (39) for a detailed account of the findings.
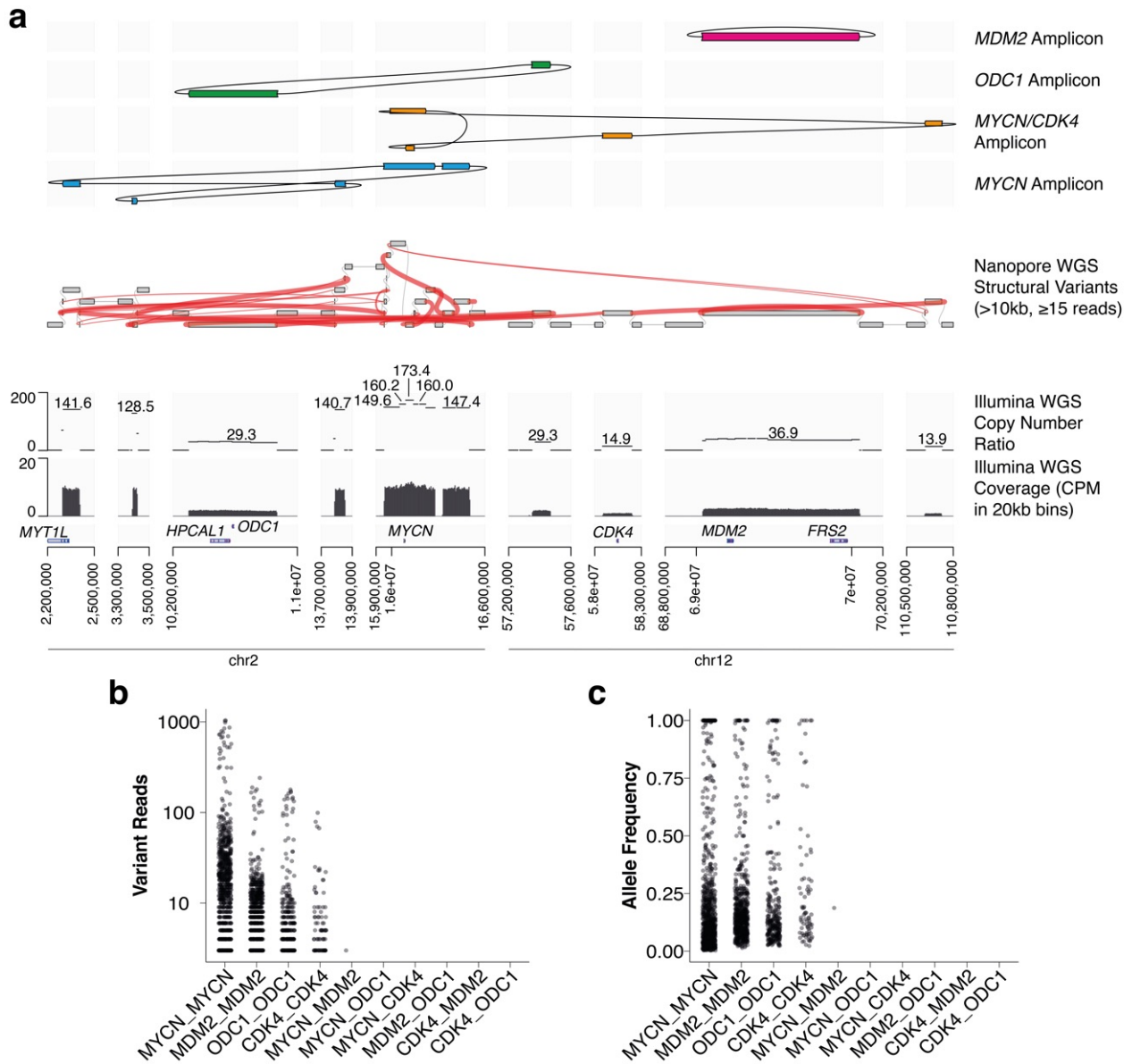
Figure 12 Four independent amplicons in the TR-14 cell line a Amplicon reconstruction, Nanopore whole genome sequencing-based structural variants (size >10kb, >20 supporting reads), short-read copy number profile and short-read coverage, and selected genes for the *MYCN*-amplified neuroblastoma cell line TR-14. **b** Number of variant-supporting reads and **c** estimated allele frequency for inter-amplicon structural variants identified in Nanopore whole-genome sequencing data. Adapted from Hung et al. (39).
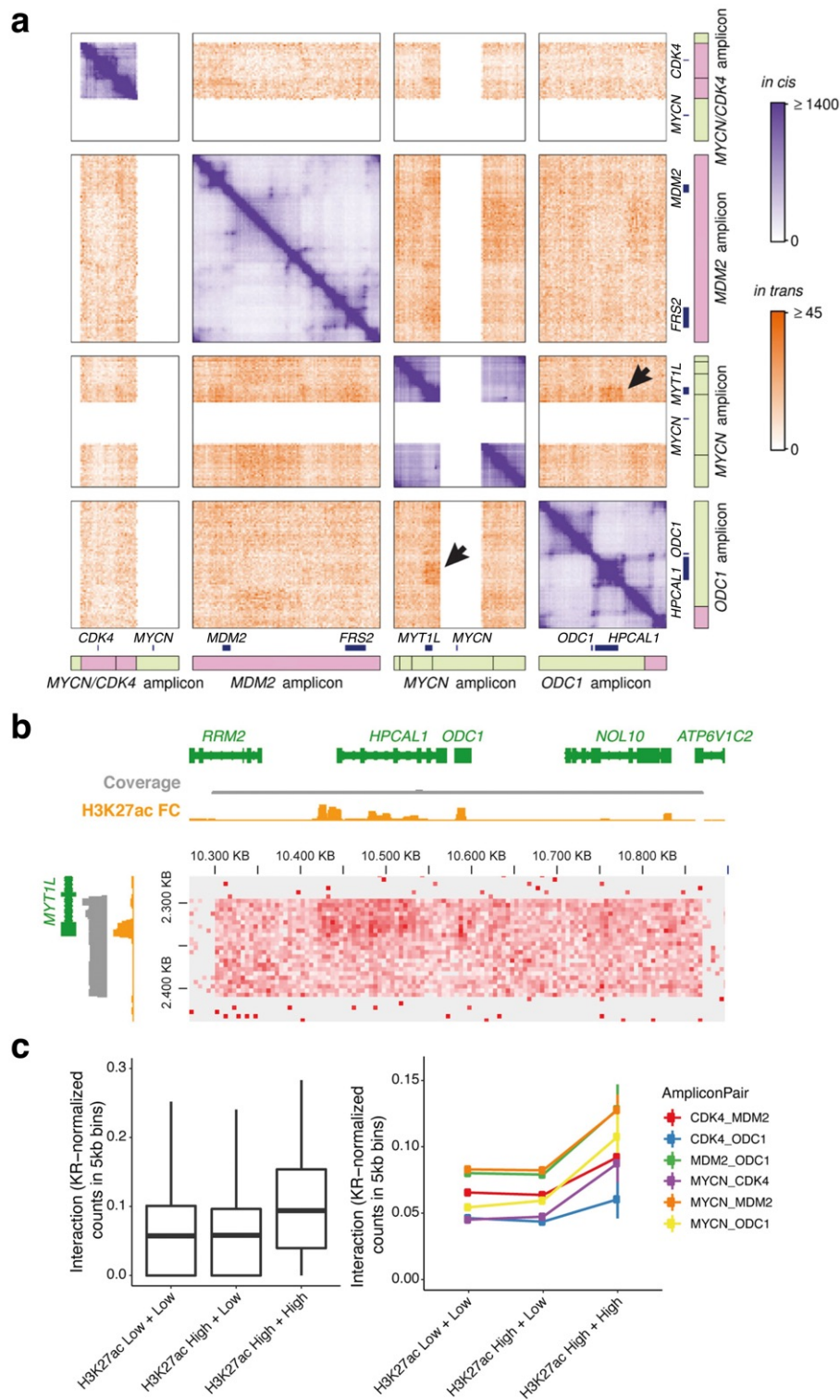
Figure 13 *In trans* inter-amplicon interaction a Mapping of Hi-C reads to the reconstructed amplicons. Arrows mark a potential spatial *in trans* interaction between the *MYCN* and *ODC1* amplicon. **b** Hi-C *in trans* interaction between loci on the *MYCN* amplicon (vertical) and the *ODC1* amplicon (horizontal), see arrow-marked interaction above. Corresponding H3K27ac ChIP-seq tracks (H3K27ac FC, H3K27ac ChIP-seq fold-change over input) are depicted in orange, whole-genome sequencing coverage in gray. **c** Left: Boxplots of Hi-C *in trans* interaction (KR-normalized counts in 5kb bins) depending on the H3K27ac status of both involved loci. H3K27ac 3-fold change over input was chosen to dichotomize bins into H3K27ac High vs. Low. N=114,636 H3K27ac Low+Low pairs, N=11,990 H3K27ac High+Low pairs, N=296 H3K27ac High+High pairs. Right: *In trans* interaction by H3K27ac status and amplicon pair. Means and 95% confidence intervals are depicted. Adapted from Hung et al. (39).

36

Our contribution to Hung et al. (39) consisted of the characterization of such *in trans* interactions in the neuroblastoma cell line TR-14. We acquired short-read and long-read whole-genome sequencing to reconstruct the genome-wide landscape of amplification (Fig. 12a). This way, we identified four independent circular amplicons harboring the neuroblastoma oncogenes *MYCN*, *CDK4*, *MDM2,* and *ODC1* (Fig. 12a). Interestingly, we found two different types of *MYCN* amplicons: Most *MYCN* copies were located on an amplicon joining four fragments from across chromosome 2 ("*MYCN* amplicon"). A minority of *MYCN* amplicons, however, was structurally different: It contained two copies of the *MYCN* locus and two additional fragments of chromosome 12, including the *CDK4* oncogene ("*MYCN/CDK4* amplicon"; Fig. 12a). We then acquired Hi-C data which validated the reconstruction and confirmed that both *MYCN* amplicons are structurally independent (Fig. 13a). This demonstrates that diverse populations of amplicons can exist within the same cell line, including different types of *MYCN* amplicons.

With a population of several distinct amplicons, we asked whether small subpopulations of fused amplicons existed. To this end, we used long read-based structural variant analysis to search for amplicon-connecting rearrangements. Among all pairs of amplicons, we identified only one breakpoint connecting the *MYCN* and the *MDM2* amplicon. This breakpoint was supported by the minimal number of only three variant-supporting reads, resulting in a locally estimated allele frequency of 19% (Fig. 12b,c). We therefore cannot rule out that a small number of *MYCN* and *MDM2* copies might be located on fused amplicons. For all other amplicons, however, we did not find such evidence for subpopulations of directly fused amplicons through amplicon-connecting structural variants.

This allowed us to use TR-14 as a model system to investigate *in trans* interaction between different amplicons. Using Hi-C, we found evidence for *in trans* interactions between loci on different amplicons (Fig. 13a). Integrating H3K27ac ChIP-seq data, we showed that this included an interaction between a strong enhancer on the *MYCN* amplicon and the gene- and enhancer-rich *HPCAL1/ODC1* locus on the *ODC1* amplicon (Fig. 13a,b). Across amplicons, pairs of loci that were both characterized by high H3K27ac signal were more prone to interact *in trans* than other pairs of loci (Welch two sample t-test P=7.031e-15; Fig. 13c). Taken together, this suggests that individual copies of amplified genomic material organize in localized hubs in the nucleus. This creates opportunities for *in trans* interactions between amplified enhancers and genes, potentially contributing to the excessive transcription of oncogenes from ecDNA.

4 Discussion

Here, we first describe enhancer co-amplification as a determinant of the non-coding contents of *MYCN* amplification in neuroblastoma. We find that complex amplification structure is in part due to enhancer hijacking that drives the amplified oncogene and creates new chromatin domains on extrachromosomal and intrachromosomal amplicons. Second, we identify clustered interchromosomal rearrangements around amplified oncogenes and propose that these reflect complex ecDNA structure or ecDNA re-integration into chromosomal loci, contributing to local gene overexpression or disruption. Third, we show that ecDNA organizes as hubs in the nucleus contributing to oncogene overexpression. Specifically, we describe how hub formation allows for *in trans* enhancer contacts between ecDNA elements.

Cancer genomes are characterized by highly rearranged haplotypes formed by multiple structural variants at once. Until recently, however, cancer genomics has largely been limited to inspecting individual chromosomal breakpoints in isolation. This hampered the investigation of the chromatin landscape and three-dimensional conformation of complex genomic loci in cancer. In Helmsauer et al. (37) and Hung et al. (39), we leverage breakpoint read support and surrounding copy number to integrate several breakpoints and resolve the coarse structure of such loci, similar to other recent approaches (24, 60, 76-78). Validation using Hi-C showed that our method yields correct reconstructions. In parts, such robustness is achieved by adjustment of algorithm parameters to the coverage and noise levels of different datasets. Other approaches to amplicon reconstruction, like Episomizer (78), have relied on similar sample-level adjustments. As the main limitation, however, such tuning does not scale well to high sample numbers. AmpliconArchitect (24) or JaBbA (60) address this challenge by statistically inferring haplotypes but can be quickly led astray by false-negative or false-positive structural variant calling. In conjunction with orthogonal validation using optical mapping (79) or Hi-C, however, we think that such approaches will enable large scale characterization of cancer haplotypes in the future. Our work also demonstrates how long-read sequencing and *de novo* assembly can be used to investigate amplicon sequences. This does not only avoid the high false positive and false negative rates of short-read sequencing-based structural variant analysis. Assembly-based approaches can also be successful where short-read sequencing yields ambiguous reconstructions or where amplicons harbor non-reference sequence (e.g. co-amplified viral sequences as reported in (24)). We expect that breakpoint-based locus

reconstruction and long read-based locus assembly will benefit the investigation of complex loci in cancer genomes in the future.

Building on amplicon reconstruction and large-scale copy number data sets, we identify enhancer co-amplification as a major principle governing the extent of *MYCN* amplification. Shortly before the publication of our findings, Morton et al. (80) also found asymmetric co-amplification patterns across different oncogenes and cancer entities, including *EGFR* in glioblastoma and *MYCN* in Wilms' tumor and neuroblastoma. Crucially, Morton et al. (80) provide experimental evidence for the role of amplicon enhancers: Repressing co-amplified *EGFR* enhancers in glioblastoma cell lines by CRISPR interference reduced *EGFR* expression and cell proliferation. Recently, similar experimental evidence has been reported for *SOX2* amplification in squamous cancers by Liu et al. (81). In contrast to Morton et al. (80) and Liu et al. (81), we show that local enhancers are sometimes replaced by distal enhancers on the amplicon. In our model, simple amplification must include local enhancers whereas the loss of local enhancers requires enhancer hijacking to form functioning amplicons. Dubois et al. (82) have just corroborated such a model in pediatric high-grade glioma showing that amplified *MYCN* frequently hijacks a strong *ID2*-associated ectopic enhancer, forming a complex amplicon that includes only small parts of the *MYCN* neighborhood. If the ectopic enhancer is not included, however, amplification includes much larger parts of the local TAD, including several local enhancers. Overall, these data provide strong evidence for enhancer co-amplification as a principle behind the non-coding contents of amplification.

The enhancer content, abundance, and individual redundance of ecDNA made us speculate about ecDNA as building blocks for genome remodeling. It is well known that ecDNA integrates into chromosomes forming HSRs which is commonly regarded as stabilizing copy number across cells in the context of a stable cellular environment (31). However, this cannot explain why amplified oncogenes are sometimes found at multiple loci scattered throughout the genome as multiple HSRs (32, 83) or as a non-canonical form of amplification termed 'jumping translocations' (84). Our findings explain such phenomena in terms of tumor evolution: following ecDNA integration, genes at the integration sites can be disrupted and genes in the neighborhood can be activated by ecDNA-derived gene-regulatory elements. Our work is limited to a single sample per tumor, but we hypothesize that future data (e.g. multi-region sequencing or primary vs. relapse samples) will demonstrate ongoing ecDNA integration during tumor evolution. We expect that further research will also identify the underlying molecular mechanisms of ecDNA integration and its exact genetic and epigenetic consequences.

In our work, we find that certain basic rules of chromatin topology also apply to ecDNA and HSRs: In line with recent work by Wu et al. (85), we confirm that TADs form on ecDNA. In addition, we show that this is also true for HSRs, even though HSRs differ from chromosomal chromatin by losing territoriality and forming abnormally large intranuclear structures (86). Our work in Hung et al. (39), however, uncovers how ecDNA might differ from the chromosomal genome, providing evidence that ecDNA enhancers engage in *in trans* chromatin interactions. Simultaneously with our findings, Yi et al. (87) describe ecDNA hubs, independently validating this concept using CRISPR-based tagging of ecDNA and live-cell imaging. They then use immunostaining to show that ecDNA hubs co-localize with RNA polymerase II and significantly overlap with PML and Cajal bodies in the nucleus, indicative of transcriptional activity. This work is complemented by Zhu et al. (88) who identify enriched *in trans* interaction between enhancers on ecDNA and chromosomal loci based on ChIA-PET and ChIA-Drop. This leads them to propose a model of ecDNA as 'mobile enhancers'. This way, ecDNA could potentially affect chromosomal transcription and drive oncogenic expression programs. Such a model also offers an elegant explanation as to why a recent pan-cancer analysis identified amplicons that included enhancer regions but lacked any protein-coding genes (88, 89). We think that there is emerging evidence for a role of *in trans* interaction and hub formation in the nuclear organization of ecDNA. Understanding the underlying mechanisms might pave the way to ecDNA-directed therapy in the future.

Our results in all three publications are based on epigenomic and genomic data of primary tumors and cell lines. One shared limitation of these findings is their reliance on correlational evidence. This is mainly due to a current lack of experimental models for extra- and intrachromosomal amplification, ecDNA integration and complex circle formation. Whereas single-nucleotide variants, deletions, inversions, or translocations can be engineered experimentally (90), this is far from trivial for amplifications. In specific cell lines, it has been shown that *DHFR* or *BRAF* amplification can be induced through treatment with methotrexate or vemurafenib/selumetinib respectively (26, 31). Although such approaches lead to ecDNA or HSR formation, they remain limited to single genes and do not offer experimental control over the copy number, contents, or structure of amplification. Recently, it has been shown that chromatin circularization can be experimentally induced by simultaneously targeting the boundaries of a chromosomal region using CRISPR/Cas9 (91). So far, however, no successful models of oncogene amplification based on this method have been reported. Once established, such an experimental strategy can not only

benefit the understanding of basic ecDNA biology. It might also allow to engineer more realistic cancer models. Genetically engineered mouse models of oncogene amplification typically rely on an overexpressed transgene that has been integrated into a chromosomal locus. This is in stark contrast to the actual biology of oncogene amplification. Such models do not, for instance, recapitulate gene regulatory factors acting upon the overexpressed gene. Furthermore, they do not adequately account for copy number heterogeneity, chromosomal integration, or ecDNA hub formation. Combining emerging experimental strategies with our findings on the architectural principles behind ecDNA structure might allow to engineer ecDNA in the future. This could be one step towards more realistic preclinical cancer models.

We hope that a better understanding of amplicon structure and function will also translate into diagnostic and therapeutic applications. Recently, it has been reported that complex amplification patterns are associated with particularly aggressive *MYCN*-amplified neuroblastoma (75). Our work extends this finding by demonstrating the prognostic relevance of clusters of interchromosomal rearrangements and of the co-amplification of specific oncogenes. Both findings are based on the explorative analysis of existing data. Validation by independent, larger, and prospective cohorts is needed. However, we hope that such genetic markers of ultra-high-risk disease can one day help guide differentiated treatment of *MYCN*-amplified neuroblastoma.

Amplicon architecture might also play a role in disease monitoring. Minimal residual disease (MRD) detection based on genomic breakpoints is already an invaluable tool guiding clinical decision-making in several hematologic malignancies. Neuroblastoma often relapses in the bone marrow which can be surveilled for MRD. Our work illustrates that *MYCN* amplification is structurally complex, with at least one breakpoint at the ecDNA circle junction but often many more. We demonstrate how reconstructing the amplicon can identify a set of linked genomic breakpoints likely formed during amplicon formation which are therefore shared among most copies. As neuroblastoma typically relies on sustained *MYCN* overexpression (9, 10), the *MYCN* amplicon is unlikely to be lost during tumor evolution although exact data on this is missing. While the amplicon could in principle get rearranged over time and acquire additional breakpoints, it is also unlikely that all breakpoints get lost through local deletions over time. This makes amplicon breakpoints promising targets for MRD detection in neuroblastoma, an approach that our group is currently evaluating clinically (Szymansky et al., in preparation). As amplicon breakpoints are one to two orders of magnitude more abundant than single-copy genomic mutations, they might also be particularly suitable targets for MRD detection from liquid biopsies (blood, cerebrospinal fluid,

…). This could extend MRD surveillance to a broad range of non-hematologic malignancies in the future.

We and others have shown that ecDNA drives aberrant oncogene expression, can rapidly change in copy number, participate in genome remodeling, and transform into intrachromosomal amplification forming HSRs. This confers adaptive advantages during tumor evolution. Whether and how ecDNA also leads to specific therapeutic vulnerabilities remains one of the major questions in the field. Our findings on the sensitivity of ecDNA-containing cells to JQ1 in Hung et al. (39) are a first hint that some vulnerabilities might be related to the nuclear organization of ecDNA. Other vulnerabilities might result from the specific ecDNA content. Genes that are co-amplified with the oncogene itself, e.g., because they are close to one of the co-amplified enhancers, might confer a phenotype that can be therapeutically exploited as recently proposed by our group (92). Other therapeutic opportunities might arise from more general features of ecDNA. For instance, it has been found that ecDNA is prone to be included in micronuclei (26, 93, 94) which are relatively unstable and have been shown to stimulate cytoplasmic DNA sensing, e.g. by the cGAS/STING pathway (95). ecDNA-containing tumors have likely evolved to bypass such systems. It is tempting to speculate that this might render them particularly vulnerable to specific interventions, e. g. oncolytic virus therapy, or to a re-instantiation of cytoplasmic DNA sensing pathways. In parallel to the investigation of mechanistically plausible vulnerabilities, ecDNA- vs. HSR-specific vulnerabilities might also be exposed through genetic and pharmacological screens of known ecDNA- and HSR-containing cell lines. Screens can be performed in established models of intra- and extrachromosomal amplification, but existing pharmacogenomic databases (e.g. (96)) can also be leveraged by using metaphase FISH to annotate ecDNA and HSR content in pharmacologically characterized cell lines. We envision that such complementary approaches will pave the way for ecDNA- and HSR-directed therapeutics in the years to come.

5 References

1.   Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L, Weiss WA. Neuroblastoma. Nat Rev Dis Primers. 2016;2:16078.

2.   Smith MA, Seibel NL, Altekruse SF, Ries LA, Melbert DL, O'Leary M, Smith FO, Reaman GH. Outcomes for children and adolescents with cancer: challenges for the twenty-first century. J Clin Oncol. 2010;28(15):2625-34.

3.   Schwab M, Alitalo K, Klempnauer KH, Varmus HE, Bishop JM, Gilbert F, Brodeur G, Goldstein M, Trent J. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. Nature. 1983;305(5931):245-8.

4.   Kohl NE, Kanda N, Schreck RR, Bruns G, Latt SA, Gilbert F, Alt FW. Transposition and amplification of oncogene-related sequences in human neuroblastomas. Cell. 1983;35(2 Pt 1):359-67.

5.   Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. Science. 1984;224(4653):1121-4.

6.   Campbell K, Gastier-Foster JM, Mann M, Naranjo AH, Van Ryn C, Bagatell R, Matthay KK, London WB, Irwin MS, Shimada H, Granger MM, Hogarty MD, Park JR, DuBois SG. Association of MYCN copy number with clinical features, tumor biology, and outcomes in neuroblastoma: A report from the Children's Oncology Group. Cancer. 2017;123(21):4224-35.

7.   Weiss WA, Aldape K, Mohapatra G, Feuerstein BG, Bishop JM. Targeted expression of MYCN causes neuroblastoma in transgenic mice. EMBO J. 1997;16(11):2985-95.

8.   Althoff K, Beckers A, Bell E, Nortmeyer M, Thor T, Sprussel A, Lindner S, De Preter K, Florin A, Heukamp LC, Klein-Hitpass L, Astrahantseff K, Kumps C, Speleman F, Eggert A, Westermann F, Schramm A, Schulte JH. A Cre-conditional MYCN-driven neuroblastoma mouse model as an improved tool for preclinical studies. Oncogene. 2015;34(26):3357-68.

9.   Kang JH, Rychahou PG, Ishola TA, Qiao J, Evers BM, Chung DH. MYCN silencing induces differentiation and apoptosis in human neuroblastoma cells. Biochem Biophys Res Commun. 2006;351(1):192-7.

10.   Nara K, Kusafuka T, Yoneda A, Oue T, Sangkhathat S, Fukuzawa M. Silencing of MYCN by RNA interference induces growth inhibition, apoptotic activity and cell differentiation in a neuroblastoma cell line with MYCN amplification. Int J Oncol. 2007;30(5):1189-96.

11.   Huang M, Weiss WA. Neuroblastoma and MYCN. Cold Spring Harb Perspect Med. 2013;3(10):a014415.

12.   Rickman DS, Schulte JH, Eilers M. The Expanding World of N-MYC-Driven Tumors. Cancer Discov. 2018;8(2):150-63.

13.   Zeid R, Lawlor MA, Poon E, Reyes JM, Fulciniti M, Lopez MA, Scott TG, Nabet B, Erb MA, Winter GE, Jacobson Z, Polaski DR, Karlin KL, Hirsch RA, Munshi NP, Westbrook TF, Chesler L, Lin CY, Bradner JE. Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma. Nat Genet. 2018;50(4):515-23.

14.   Boeva V, Louis-Brennetot C, Peltier A, Durand S, Pierre-Eugene C, Raynal V, Etchevers HC, Thomas S, Lermine A, Daudigeos-Dubus E, Geoerger B, Orth MF, Grunewald TGP, Diaz E, Ducos B, Surdez D, Carcaboso AM, Medvedeva I, Deller T, Combaret V, Lapouble E, Pierron G, Grossetete-Lalami S, Baulande S, Schleiermacher G, Barillot E, Rohrer H, Delattre O, Janoueix-Lerosey I. Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. Nat Genet. 2017;49(9):1408-13.

15.   van Groningen T, Koster J, Valentijn LJ, Zwijnenburg DA, Akogul N, Hasselt NE, Broekmans M, Haneveld F, Nowakowska NE, Bras J, van Noesel CJM, Jongejan A, van Kampen AH, Koster L, Baas F, van Dijk-Kerkhoven L, Huizer-Smit M, Lecca MC, Chan A, Lakeman A, Molenaar P, Volckmann R, Westerhout EM, Hamdi M, van Sluis PG, Ebus ME, Molenaar JJ, Tytgat GA, Westerman BA, van Nes J, Versteeg R. Neuroblastoma is composed of two super-enhancer-associated differentiation states. Nat Genet. 2017;49(8):1261-6.

16.   Decaesteker B, Denecker G, Van Neste C, Dolman EM, Van Loocke W, Gartlgruber M, Nunes C, De Vloed F, Depuydt P, Verboom K, Rombaut D, Loontiens S, De Wyn J, Kholosy WM, Koopmans B, Essing AHW, Herrmann C, Dreidax D, Durinck K, Deforce D, Van Nieuwerburgh F, Henssen A, Versteeg R, Boeva V, Schleiermacher G, van Nes J, Mestdagh P, Vanhauwaert S, Schulte JH, Westermann F, Molenaar JJ, De Preter K, Speleman F. TBX2 is a neuroblastoma core regulatory circuitry component enhancing MYCN/FOXM1 reactivation of DREAM targets. Nat Commun. 2018;9(1):4866.

17.   Wang L, Tan TK, Durbin AD, Zimmerman MW, Abraham BJ, Tan SH, Ngoc PCT, Weichert-Leahey N, Akahane K, Lawton LN, Rokita JL, Maris JM, Young RA, Look AT, Sanda T. ASCL1 is a MYCN- and LMO1-dependent member of the adrenergic neuroblastoma core regulatory circuitry. Nat Commun. 2019;10(1):5622.

18.   Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, Johann PD, Balasubramanian GP, Segura-Wang M, Brabetz S, Bender S, Hutter B, Sturm D, Pfaff E, Hubschmann D, Zipprich G, Heinold M, Eils J, Lawerenz C, Erkek S, Lambo S, Waszak S, Blattmann C, Borkhardt A, Kuhlen M, Eggert A, Fulda S, Gessler M, Wegert J, Kappler R, Baumhoer D, Burdach S, Kirschner-Schwabe R, Kontny U, Kulozik AE, Lohmann D, Hettmer S,

Eckert C, Bielack S, Nathrath M, Niemeyer C, Richter GH, Schulte J, Siebert R, Westermann F, Molenaar JJ, Vassal G, Witt H, ICGC PedBrain-Seq Project, ICGC MMML-Seq Project, Burkhardt B, Kratz CP, Witt O, van Tilburg CM, Kramm CM, Fleischhack G, Dirksen U, Rutkowski S, Fruhwald M, von Hoff K, Wolf S, Klingebiel T, Koscielniak E, Landgraf P, Koster J, Resnick AC, Zhang J, Liu Y, Zhou X, Waanders AJ, Zwijnenburg DA, Raman P, Brors B, Weber UD, Northcott PA, Pajtler KW, Kool M, Piro RM, Korbel JO, Schlesner M, Eils R, Jones DTW, Lichter P, Chavez L, Zapatka M, Pfister SM. The landscape of genomic alterations across childhood cancers. Nature. 2018;555(7696):321-27.

19.   Pfister S, Remke M, Benner A, Mendrzyk F, Toedt G, Felsberg J, Wittmann A, Devens F, Gerber NU, Joos S, Kulozik A, Reifenberger G, Rutkowski S, Wiestler OD, Radlwimmer B, Scheurlen W, Lichter P, Korshunov A. Outcome prediction in pediatric medulloblastoma based on DNA copy-number aberrations of chromosomes 6q and 17q and the MYC and MYCN loci. J Clin Oncol. 2009;27(10):1627-36.

20.   Williams RD, Al-Saadi R, Chagtai T, Popov S, Messahel B, Sebire N, Gessler M, Wegert J, Graf N, Leuschner I, Hubank M, Jones C, Vujanic G, Pritchard-Jones K, Children's Cancer and Leukaemia Group, SIOP Wilms Tumour Biology Group. Subtype-specific FBXW7 mutation and MYCN copy number gain in Wilms' tumor. Clin Cancer Res. 2010;16(7):2036-45.

21.   Williamson D, Lu YJ, Gordon T, Sciot R, Kelsey A, Fisher C, Poremba C, Anderson J, Pritchard-Jones K, Shipley J. Relationship between MYCN copy number and expression in rhabdomyosarcomas and correlation with adverse prognosis in the alveolar subtype. J Clin Oncol. 2005;23(4):880-8.

22.   Mackay A, Burford A, Carvalho D, Izquierdo E, Fazal-Salom J, Taylor KR, Bjerke L, Clarke M, Vinci M, Nandhabalan M, Temelso S, Popov S, Molinari V, Raman P, Waanders AJ, Han HJ, Gupta S, Marshall L, Zacharoulis S, Vaidya S, Mandeville HC, Bridges LR, Martin AJ, Al-Sarraj S, Chandler C, Ng HK, Li X, Mu K, Trabelsi S, Brahim DH, Kisljakov AN, Konovalov DM, Moore AS, Carcaboso AM, Sunol M, de Torres C, Cruz O, Mora J, Shats LI, Stavale JN, Bidinotto LT, Reis RM, Entz-Werle N, Farrell M, Cryan J, Crimmins D, Caird J, Pears J, Monje M, Debily MA, Castel D, Grill J, Hawkins C, Nikbakht H, Jabado N, Baker SJ, Pfister SM, Jones DTW, Fouladi M, von Bueren AO, Baudis M, Resnick A, Jones C. Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma. Cancer Cell. 2017;32(4):520-37 e5.

23.   Rosswog C, Bartenhagen C, Welte A, Kahlert Y, Hemstedt N, Lorenz W, Cartolano M, Ackermann S, Perner S, Vogel W, Altmuller J, Nurnberg P, Hertwig F, Gohring G, Lilienweiss E,

Stutz AM, Korbel JO, Thomas RK, Peifer M, Fischer M. Chromothripsis followed by circular recombination drives oncogene amplification in human cancer. Nat Genet. 2021;53(12):1673-85.

24.	Deshpande V, Luebeck J, Nguyen ND, Bakhtiari M, Turner KM, Schwab R, Carter H, Mischel PS, Bafna V. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. Nat Commun. 2019;10(1):392.

25.	Ly P, Brunner SF, Shoshani O, Kim DH, Lan W, Pyntikova T, Flanagan AM, Behjati S, Page DC, Campbell PJ, Cleveland DW. Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. Nat Genet. 2019;51(4):705-15.

26.	Shoshani O, Brunner SF, Yaeger R, Ly P, Nechemia-Arbely Y, Kim DH, Fang R, Castillon GA, Yu M, Li JSZ, Sun Y, Ellisman MH, Ren B, Campbell PJ, Cleveland DW. Chromothripsis drives the evolution of gene amplification in cancer. Nature. 2021;591(7848):137-41.

27.	The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature. 2020;578(7793):82-93.

28.	Villamon E, Berbegall AP, Piqueras M, Tadeo I, Castel V, Djos A, Martinsson T, Navarro S, Noguera R. Genetic instability and intratumoral heterogeneity in neuroblastoma with MYCN amplification plus 11q deletion. PLoS One. 2013;8(1):e53740.

29.	Molenaar JJ, Koster J, Zwijnenburg DA, van Sluis P, Valentijn LJ, van der Ploeg I, Hamdi M, van Nes J, Westerman BA, van Arkel J, Ebus ME, Haneveld F, Lakeman A, Schild L, Molenaar P, Stroeken P, van Noesel MM, Ora I, Santo EE, Caron HN, Westerhout EM, Versteeg R. Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature. 2012;483(7391):589-93.

30.	Lin CC, Alitalo K, Schwab M, George D, Varmus HE, Bishop JM. Evolution of karyotypic abnormalities and C-MYC oncogene amplification in human colonic carcinoma cell lines. Chromosoma. 1985;92(1):11-5.

31.	Song K, Minami JK, Huang A, Dehkordi SR, Lomeli SH, Luebeck J, Goodman MH, Moriceau G, Krijgsman O, Dharanipragada P, Ridgley T, Crosson WP, Salazar J, Pazol E, Karin G, Jayaraman R, Balanis NG, Alhani S, Sheu K, Ten Hoeve J, Palermo A, Motika SE, Senaratne TN, Paraiso KH, Hergenrother PJ, Rao PN, Multani AS, Peeper DS, Bafna V, Lo RS, Graeber TG. Plasticity of extrachromosomal and intrachromosomal BRAF amplifications in overcoming targeted therapy dosage challenges. Cancer Discov. 2022;12(4):1046-69.

32.	Storlazzi CT, Lonoce A, Guastadisegni MC, Trombetta D, D'Addabbo P, Daniele G, L'Abbate A, Macchia G, Surace C, Kok K, Ullmann R, Purgato S, Palumbo O, Carella M, Ambros PF, Rocchi M. Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure. Genome Res. 2010;20(9):1198-206.

33.   Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, Li B, Arden K, Ren B, Nathanson DA, Kornblum HI, Taylor MD, Kaushal S, Cavenee WK, Wechsler-Reya R, Furnari FB, Vandenberg SR, Rao PN, Wahl GM, Bafna V, Mischel PS. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. Nature. 2017;543(7643):122-5.

34.   Lange JT, Rose JC, Chen CY, Pichugin Y, Xie L, Tang J, Hung KL, Yost KE, Shi Q, Erb ML, Rajkumar U, Wu S, Taschner-Mandl S, Bernkopf M, Swanton C, Liu Z, Huang W, Chang HY, Bafna V, Henssen AG, Werner B, Mischel PS. The evolutionary dynamics of extrachromosomal DNA in human cancers. Nat Genet. 2022;54(10):1527-33.

35.   Nathanson DA, Gini B, Mottahedeh J, Visnyei K, Koga T, Gomez G, Eskin A, Hwang K, Wang J, Masui K, Paucar A, Yang H, Ohashi M, Zhu S, Wykosky J, Reed R, Nelson SF, Cloughesy TF, James CD, Rao PN, Kornblum HI, Heath JR, Cavenee WK, Furnari FB, Mischel PS. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. Science. 2014;343(6166):72-6.

36.   Kaufman RJ, Brown PC, Schimke RT. Amplified dihydrofolate reductase genes in unstably methotrexate-resistant cells are associated with double minute chromosomes. Proc Natl Acad Sci U S A. 1979;76(11):5669-73.

37.   Helmsauer K, Valieva ME, Ali S, Chamorro González R, Schöpflin R, Roefzaad C, Bei Y, Dorado García H, Rodriguez-Fos E, Puiggros M, Kasack K, Haase K, Keskeny C, Chen CY, Kuschel LP, Euskirchen P, Heinrich V, Robson MI, Rosswog C, Toedling J, Szymansky A, Hertwig F, Fischer M, Torrents D, Eggert A, Schulte JH, Mundlos S, Henssen AG, Koche RP. Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma. Nat Commun. 2020;11:5823.

38.   Koche RP, Rodriguez-Fos E, Helmsauer K, Burkert M, MacArthur IC, Maag J, Chamorro R, Munoz-Perez N, Puiggros M, Dorado García H, Bei Y, Roefzaad C, Bardinet V, Szymansky A, Winkler A, Thole T, Timme N, Kasack K, Fuchs S, Klironomos F, Thiessen N, Blanc E, Schmelz K, Künkele A, Hundsdörfer P, Rosswog C, Theissen J, Beule D, Deubzer H, Sauer S, Toedling J, Fischer M, Hertwig F, Schwarz RF, Eggert A, Torrents D, Schulte JH, Henssen AG. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. Nat Genet. 2020;52(1):29-34.

39.   Hung KL, Yost KE, Xie L, Shi Q, Helmsauer K, Luebeck J, Schöpflin R, Lange JT, Chamorro González R, Weiser NE, Chen C, Valieva ME, Wong IT, Wu S, Dehkordi SR, Duffy CV, Kraft K, Tang J, Belk JA, Rose JC, Corces MR, Granja JM, Li R, Rajkumar U, Friedlein J, Bagchi A, Satpathy AT, Tjian R, Mundlos S, Bafna V, Henssen AG, Mischel PS, Liu Z, Chang

HY. ecDNA hubs drive cooperative intermolecular oncogene expression. Nature. 2021;600:731-6.

40. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

41. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.

42. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

43. Rada-Iglesias A, Bajpai R, Prescott S, Brugmann SA, Swigut T, Wysocka J. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. Cell Stem Cell. 2012;11(5):633-48.

44. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95.

45. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 2012;22(9):1813-31.

46. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160-5.

47. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep. 2019;9(1):9354.

48. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012;7(9):1728-40.

49. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schopflin R, Kraft K, Kempfer R, Jerkovic I, Chan WL, Spielmann M, Timmermann B, Wittler L, Kurth I, Cambiaso P, Zuffardi O, Houge G, Lambie L, Brancati F, Pombo A, Vingron M, Spitz F, Mundlos S. Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature. 2016;538(7624):265-9.

50.   Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 2016;3(1):95-8.

51.   Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Barillot E. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012;28(3):423-5.

52.   Wala JA, Bandopadhayay P, Greenwald NF, O'Rourke R, Sharpe T, Stewart C, Schumacher S, Li Y, Weischenfeldt J, Yao X, Nusbaum C, Campbell P, Getz G, Meyerson M, Zhang CZ, Imielinski M, Beroukhim R. SvABA: genome-wide detection of structural variants and indels by local assembly. Genome Res. 2018;28(4):581-91.

53.   De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34(15):2666-9.

54.   Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540-6.

55.   Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094-100.

56.   Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Cheneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, Mathelier A. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. 2018;46(D1):D260-6.

57.   Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranasic D, Santana-Garcia W, Tan G, Cheneby J, Ballester B, Parcy F, Sandelin A, Lenhard B, Wasserman WW, Mathelier A. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2020;48(D1):D87-92.

58.   Tan G, Lenhard B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. Bioinformatics. 2016;32(10):1555-6.

59.   Depuydt P, Koster J, Boeva V, Hocking TD, Speleman F, Schleiermacher G, De Preter K. Meta-mining of copy number profiles of high-risk neuroblastoma tumors. Sci Data. 2018;5:180240.

60.   Hadi K, Yao X, Behr JM, Deshpande A, Xanthopoulakis C, Tian H, Kudman S, Rosiene J, Darmofal M, DeRose J, Mortensen R, Adney EM, Shaiber A, Gajic Z, Sigouros M, Eng K, Wala JA, Wrzeszczynski KO, Arora K, Shah M, Emde AK, Felice V, Frank MO, Darnell RB, Ghandi

M, Huang F, Dewhurst S, Maciejowski J, de Lange T, Setton J, Riaz N, Reis-Filho JS, Powell S, Knowles DA, Reznik E, Mishra B, Beroukhim R, Zody MC, Robine N, Oman KM, Sanchez CA, Kuhner MK, Smith LP, Galipeau PC, Paulson TG, Reid BJ, Li X, Wilkes D, Sboner A, Mosquera JM, Elemento O, Imielinski M. Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. Cell. 2020;183(1):197-210 e32.

61.   Henssen AG, Koche R, Zhuang J, Jiang E, Reed C, Eisenberg A, Still E, MacArthur IC, Rodriguez-Fos E, Gonzalez S, Puiggros M, Blackford AN, Mason CE, de Stanchina E, Gonen M, Emde AK, Shah M, Arora K, Reeves C, Socci ND, Perlman E, Antonescu CR, Roberts CWM, Steen H, Mullen E, Jackson SP, Torrents D, Weng Z, Armstrong SA, Kentsis A. PGBD5 promotes site-specific oncogenic mutations in human tumors. Nat Genet. 2017;49(7):1005-14.

62.   Guzman C, Bagga M, Kaur A, Westermarck J, Abankwa D. ColonyArea: an ImageJ plugin to automatically quantify colony formation in clonogenic assays. PLoS One. 2014;9(3):e92444.

63.   Peifer M, Hertwig F, Roels F, Dreidax D, Gartlgruber M, Menon R, Kramer A, Roncaioli JL, Sand F, Heuckmann JM, Ikram F, Schmidt R, Ackermann S, Engesser A, Kahlert Y, Vogel W, Altmuller J, Nurnberg P, Thierry-Mieg J, Thierry-Mieg D, Mariappan A, Heynck S, Mariotti E, Henrich KO, Gloeckner C, Bosco G, Leuschner I, Schweiger MR, Savelyeva L, Watkins SC, Shao C, Bell E, Hofer T, Achter V, Lang U, Theissen J, Volland R, Saadati M, Eggert A, de Wilde B, Berthold F, Peng Z, Zhao C, Shi L, Ortmann M, Buttner R, Perner S, Hero B, Schramm A, Schulte JH, Herrmann C, O'Sullivan RJ, Westermann F, Thomas RK, Fischer M. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. Nature. 2015;526(7575):700-4.

64.   Tischler G, Leonard S. biobambam: tools for read pair collation based algorithms on BAM files. Source Code for Biology and Medicine. 2014;9(1):13.

65.   Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B, Perou CM, Borresen-Dale AL, Kristensen VN. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A. 2010;107(39):16910-5.

66.   Chong Z, Ruan J, Gao M, Zhou W, Chen T, Fan X, Ding L, Lee AY, Boutros P, Chen J, Chen K. novoBreak: local assembly for breakpoint detection in cancer genomes. Nat Methods. 2017;14(1):65-7.

67.   Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28(18):i333-9.

68.   Moncunill V, Gonzalez S, Bea S, Andrieux LO, Salaverria I, Royo C, Martinez L, Puiggros M, Segura-Wang M, Stutz AM, Navarro A, Royo R, Gelpi JL, Gut IG, Lopez-Otin C, Orozco M,

Korbel JO, Campo E, Puente XS, Torrents D. Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. Nat Biotechnol. 2014;32(11):1106-12.

69.   Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, Sivachenko A, Thompson DA, Wysoker A, Fennell T, Gnirke A, Pochet N, Regev A, Levin JZ. Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods. 2013;10(7):623-9.

70.   Henssen A, MacArthur IC, Koche R, Dorado García H. Purification and Sequencing of Large Circular DNA from Human Cells. Protocol Exchange. 2019. Available from: https://doi.org/10.1038/protex.2019.006.

71.   Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-89.

72.   Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer. 2018;18(11):696-705.

73.   Scheinin I, Sie D, Bengtsson H, van de Wiel MA, Olshen AB, van Thuijl HF, van Essen HF, Eijk PP, Rustenburg F, Meijer GA, Reijneveld JC, Wesseling P, Pinkel D, Albertson DG, Ylstra B. 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. 2014;24(12):2022-32.

74.   Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods. 2018;15(6):461-8.

75.   Depuydt P, Boeva V, Hocking TD, Cannoodt R, Ambros IM, Ambros PF, Asgharzadeh S, Attiyeh EF, Combaret V, Defferrari R, Fischer M, Hero B, Hogarty MD, Irwin MS, Koster J, Kreissman S, Ladenstein R, Lapouble E, Laureys G, London WB, Mazzocco K, Nakagawara A, Noguera R, Ohira M, Park JR, Potschger U, Theissen J, Tonini GP, Valteau-Couanet D, Varesio L, Versteeg R, Speleman F, Maris JM, Schleiermacher G, De Preter K. Genomic Amplifications and Distal 6q Loss: Novel Markers for Poor Survival in High-risk Neuroblastoma Patients. J Natl Cancer Inst. 2018;110(10):1084-93.

76.   Sanborn JZ, Salama SR, Grifford M, Brennan CW, Mikkelsen T, Jhanwar S, Katzman S, Chin L, Haussler D. Double minute chromosomes in glioblastoma multiforme are revealed by precise reconstruction of oncogenic amplicons. Cancer Res. 2013;73(19):6036-45.

77. L'Abbate A, Tolomeo D, Cifola I, Severgnini M, Turchiano A, Augello B, Squeo G, D'Addabbo P, Traversa D, Daniele G, Lonoce A, Pafundi M, Carella M, Palumbo O, Dolnik A, Muehlematter D, Schoumans J, Van Roy N, De Bellis G, Martinelli G, Merla G, Bullinger L, Haferlach C, Storlazzi CT. MYC-containing amplicons in acute myeloid leukemia: genomic structures, evolution, and transcriptional consequences. Leukemia. 2018;32(10):2152-66.

78. Xu K, Ding L, Chang TC, Shao Y, Chiang J, Mulder H, Wang S, Shaw TI, Wen J, Hover L, McLeod C, Wang YD, Easton J, Rusch M, Dalton J, Downing JR, Ellison DW, Zhang J, Baker SJ, Wu G. Structure and evolution of double minutes in diagnosis and relapse brain tumors. Acta Neuropathol. 2019;137(1):123-37.

79. Luebeck J, Coruh C, Dehkordi SR, Lange JT, Turner KM, Deshpande V, Pai DA, Zhang C, Rajkumar U, Law JA, Mischel PS, Bafna V. AmpliconReconstructor integrates NGS and optical mapping to resolve the complex structures of focal amplifications. Nat Commun. 2020;11(1):4374.

80. Morton AR, Dogan-Artun N, Faber ZJ, MacLeod G, Bartels CF, Piazza MS, Allan KC, Mack SC, Wang X, Gimple RC, Wu Q, Rubin BP, Shetty S, Angers S, Dirks PB, Sallari RC, Lupien M, Rich JN, Scacheri PC. Functional Enhancers Shape Extrachromosomal Oncogene Amplifications. Cell. 2019;179(6):1330-41 e13.

81. Liu Y, Wu Z, Zhou J, Ramadurai DKA, Mortenson KL, Aguilera-Jimenez E, Yan Y, Yang X, Taylor AM, Varley KE, Gertz J, Choi PS, Cherniack AD, Chen X, Bass AJ, Bailey SD, Zhang X. A predominant enhancer co-amplified with the SOX2 oncogene is necessary and sufficient for its expression in squamous cancer. Nat Commun. 2021;12(1):7139.

82. Dubois FPB, Shapira O, Greenwald NF, Zack T, Wala J, Tsai JW, Crane A, Baguette A, Hadjadj D, Harutyunyan AS, Kumar KH, Blattner-Johnson M, Vogelzang J, Sousa C, Kang KS, Sinai C, Wang DK, Khadka P, Lewis K, Nguyen L, Malkin H, Ho P, O'Rourke R, Zhang S, Gold R, Deng D, Serrano J, Snuderl M, Jones C, Wright KD, Chi SN, Grill J, Kleinman CL, Goumnerova LC, Jabado N, Jones DTW, Kieran MW, Ligon KL, Beroukhim R, Bandopadhayay P. Structural variants shape driver combinations and outcomes in pediatric high-grade glioma. Nat Cancer. 2022;3(8):994-1011.

83. Emanuel BS, Balaban G, Boyd JP, Grossman A, Negishi M, Parmiter A, Glick MC. N-myc amplification in multiple homogeneously staining regions in two human neuroblastomas. Proc Natl Acad Sci U S A. 1985;82(11):3736-40.

84. Berger R, Bernard OA. Jumping translocations. Genes Chromosomes Cancer. 2007;46(8):717-23.

85.   Wu S, Turner KM, Nguyen N, Raviram R, Erb M, Santini J, Luebeck J, Rajkumar U, Diao Y, Li B, Zhang W, Jameson N, Corces MR, Granja JM, Chen X, Coruh C, Abnousi A, Houston J, Ye Z, Hu R, Yu M, Kim H, Law JA, Verhaak RGW, Hu M, Furnari FB, Chang HY, Ren B, Bafna V, Mischel PS. Circular ecDNA promotes accessible chromatin and high oncogene expression. Nature. 2019;575(7784):699-703.

86.   Solovei I, Kienle D, Little G, Eils R, Savelyeva L, Schwab M, Jager W, Cremer C, Cremer T. Topology of double minutes (dmins) and homogeneously staining regions (HSRs) in nuclei of human neuroblastoma cell lines. Genes Chromosomes Cancer. 2000;29(4):297-308.

87.   Yi E, Gujar AD, Guthrie M, Kim H, Zhao D, Johnson KC, Amin SB, Costa ML, Yu Q, Das S, Jillette N, Clow PA, Cheng AW, Verhaak RG. Live-cell imaging shows uneven segregation of extrachromosomal DNA elements and transcriptionally active extrachromosomal DNA hubs in cancer. Cancer Discov. 2022;12(2):468-83.

88.   Zhu Y, Gujar AD, Wong CH, Tjong H, Ngan CY, Gong L, Chen YA, Kim H, Liu J, Li M, Mil-Homens A, Maurya R, Kuhlberg C, Sun F, Yi E, deCarvalho AC, Ruan Y, Verhaak RGW, Wei CL. Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. Cancer Cell. 2021;39(5):694-707 e7.

89.   Kim H, Nguyen NP, Turner K, Wu S, Gujar AD, Luebeck J, Liu J, Deshpande V, Rajkumar U, Namburi S, Amin SB, Yi E, Menghi F, Schulte JH, Henssen AG, Chang HY, Beck CR, Mischel PS, Bafna V, Verhaak RGW. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. Nat Genet. 2020;52(9):891-7.

90.   Sanchez-Rivera FJ, Jacks T. Applications of the CRISPR-Cas9 system in cancer biology. Nat Rev Cancer. 2015;15(7):387-95.

91.   Møller HD, Lin L, Xiang X, Petersen TS, Huang J, Yang L, Kjeldsen E, Jensen UB, Zhang X, Liu X, Xu X, Wang J, Yang H, Church GM, Bolund L, Regenberg B, Luo Y. CRISPR-C: circularization of genes and chromosome by CRISPR in human cells. Nucleic Acids Res. 2018;46(22):e131.

92.   Bei Y, Bramé L, Kirchner M, Fritsche-Guenther R, Kunz S, Bhattacharya A, Köppke J, Proba J, Wittstruck N, Sidorova OA, González RC, Garcia HD, Brückner L, Xu R, Giurgiu M, Rodriguez-Fos E, Koche R, Schmitt C, Schulte JH, Eggert A, Haase K, Kirwan J, Hagemann AIH, Mertins P, Dörr JR, Henssen AG. Amplicon structure creates collateral therapeutic vulnerability in cancer. bioRxiv [Preprint]. 2022;2022.09.08.506647v2. Available from: https://doi.org/10.1101/2022.09.08.506647.

93.   Von Hoff DD, McGill JR, Forseth BJ, Davidson KK, Bradley TP, Van Devanter DR, Wahl GM. Elimination of extrachromosomally amplified MYC genes from human tumor cells reduces their tumorigenicity. Proc Natl Acad Sci U S A. 1992;89(17):8165-9.

94.   Ambros IM, Rumpler S, Luegmayr A, Hattinger CM, Strehl S, Kovar H, Gadner H, Ambros PF. Neuroblastoma cells can actively eliminate supernumerary MYCN gene copies by micronucleus formation--sign of tumour cell revertance? Eur J Cancer. 1997;33(12):2043-9.

95.   Mackenzie KJ, Carroll P, Martin CA, Murina O, Fluteau A, Simpson DJ, Olova N, Sutcliffe H, Rainger JK, Leitch A, Osborn RT, Wheeler AP, Nowotny M, Gilbert N, Chandra T, Reijns MAM, Jackson AP. cGAS surveillance of micronuclei links genome instability to innate immunity. Nature. 2017;548(7668):461-5.

96.   Smirnov P, Kofia V, Maru A, Freeman M, Ho C, El-Hachem N, Adam GA, Ba-Alawi W, Safikhani Z, Haibe-Kains B. PharmacoDB: an integrative database for mining in vitro anticancer drug screening studies. Nucleic Acids Res. 2018;46(D1):D994-1002.

6 Affidavit & Contributions

Eidesstattliche Versicherung

„Ich, Konstantin Helmsauer, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „*MYCN* amplification in neuroblastoma: sequence, chromatin landscape, and genome remodeling" / „*MYCN* Amplifikation im Neuroblastom: Sequenz, Chromatinlandschaft und genomische Remodellierung" selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe. Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet. Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht. Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.og) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte. Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe. Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst."

Berlin, 05.10.2022   _____

Anteilserklärung an den erfolgten Publikationen

Konstantin Helmsauer hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: **Helmsauer K***, Valieva ME*, Ali S*, Chamorro Gonzalez R, Schopflin R, Roefzaad C, Bei Y, Dorado Garcia H, Rodriguez-Fos E, Puiggros M, Kasack K, Haase K, Keskeny C, Chen CY, Kuschel LP, Euskirchen P, Heinrich V, Robson MI, Rosswog C, Toedling J, Szymansky A, Hertwig F, Fischer M, Torrents D, Eggert A, Schulte JH, Mundlos S, Henssen AG, Koche RP. Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma, *Nature Communications*, 2020. *= Ko-Erstautorschaft

Beitrag im Einzelnen:

Konstantin Helmsauer war hauptverantwortlich für die Publikation Helmsauer et al. (2020) unter der Betreuung von Anton Henssen, Richard Koche und Stefan Mundlos.

Er führte dabei die Rohdatenverarbeitung (Datenverwaltung, Qualitätskontrolle, Basecalling, Methylierungscalling, Mapping, Peak Calling, Copy Number Analysis, Structural Variant Analysis, De Novo Assembly) von Illumina und Nanopore Whole-Genome Sequencing, RNA-seq, ChIP-seq, ATAC-seq und Array-CGH Daten durch. Die Verarbeitung und Analyse von Hi-C Daten führte Konstantin Helmsauer gemeinsam mit Robert Schöpflin durch. Die Verarbeitung und Analyse von 4C Daten erfolgte gemeinsam mit Verena Heinrich. Die Verarbeitung und Analyse von Nanopore-Methylierungsdaten erfolgte gemeinsam mit Richard Koche.

Konstantin Helmsauer analysierte und interpretierte die prozessierten Hochdurchsatzdaten sowie einen öffentlich zugänglichen Copy Number Datensatz. Insbesondere analysierte er Ko-Amplifikationsprofile und entwickelte einen Ansatz zur statistischen Analyse von Ko-Amplifikationen. Er entwickelte den verwendeten Algorithmus zur Rekonstruktion von Amplifikations-Strukturen auf Basis von Whole-Genome Sequencing Daten. Er analysierte klinische Daten.

Konstantin Helmsauer erstellte Fig. 1-3, 4a-b, 4d, 4e-f, 4h, 6 und 7 sowie Supplementary Fig. 1-7, 8a und 9. Konstantin Helmsauer war maßgeblich an der Konzeption der Studie unter der Supervision von Richard Koche, Anton Henssen und Stefan Mundlos beteiligt. Er schrieb den ersten Manuskriptentwurf. Er koordinierte die Fertigstellung des Manuskripts unter Berücksichtigung der Beiträge der Ko-Autor*innen. Er war hauptverantwortlich für die Überarbeitung des Manuskripts und Beantwortung von Gutachter*innenfragen im Rahmen des Peer Review.

Publikation 2: Koche RP*, Rodriguez-Fos E*, **Helmsauer K***, Burkert M, MacArthur IC, Maag J, Chamorro R, Munoz-Perez N, Puiggros M, Dorado Garcia H, Bei Y, Roefzaad C, Bardinet V, Szymansky A, Winkler A, Thole T, Timme N, Kasack K, Fuchs S, Klironomos F, Thiessen N, Blanc E, Schmelz K, Kunkele A, Hundsdorfer P, Rosswog C, Theissen J, Beule D, Deubzer H, Sauer S, Toedling J, Fischer M, Hertwig F, Schwarz RF, Eggert A, Torrents D, Schulte JH, Henssen AG, Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma, *Nature Genetics*, 2020. *=Ko-Erstautorschaft

Beitrag im Einzelnen:

Konstantin Helmsauer analysierte Whole-Genome Sequencing, Circle-seq, RNA-seq und klinische Daten. Insbesondere entwickelte er den Ansatz zur automatisierten Bestimmung von Bruchpunkt-Clustern und analysierte deren genomweite Verteilung. Er entwickelte einen statistischen Analyseansatz zum Test auf Assoziation mit ecDNA und eccDNA. Er analysierte die Verteilung von Cluster-assoziierten Bruchpunkten, entwickelte einen statistischen Analyseansatz zum Test auf Assoziation mit krebsrelevanten Genen und analysierte RNA-seq Daten in Hinblick auf transkriptionelle Konsequenzen von Cluster-assoziierten Bruchpunkten. Konstantin Helmsauer analysierte klinische Daten zur prognostischen Bedeutung von Bruchpunkt-Clustern. Konstantin Helmsauer erstellte Fig. 3a-c, 4a-e und Supplementary Fig. 4j-l, 7, 9 und 12. Er trug zur Konzeption der Studie bei. Er hatte wesentlichen Anteil an der Erstellung des Manuskripts (Erstellung von Methodenabschnitten, Überarbeitung des ersten Manuskriptentwurfes) sowie an der Beantwortung von Gutachter*innenfragen im Peer Review.

Publikation 3: Hung KL, Yost KE, Xie L, Shi Q, **Helmsauer K**, Luebeck J, Schöpflin R, Lange JT, Chamorro González R, Weiser NE, Chen C, Valieva ME, Wong IT, Wu S, Dehkordi SR, Duffy CV, Kraft K, Tang J, Belk JA, Rose JC, Corces MR, Granja JM, Li R, Rajkumar U, Friedlein J, Bagchi A, Satpathy AT, Tjian R, Mundlos S, Bafna V, Henssen AG, Mischel PS, Liu Z, Chang HY. ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature*, 2021.

Beitrag im Einzelnen:

Konstantin Helmsauer führte die Rohdatenverarbeitung (Datenverwaltung, Qualitätskontrolle, Basecalling, Mapping, Peak Calling, Copy Number Analysis, Structural Variant Analysis) von Illumina und Nanopore Whole-Genome Sequencing und ChIP-seq Daten für die Neuroblastom-Zelllinie TR-14 durch. Die Verarbeitung und Analyse von Hi-C Daten führte Konstantin Helmsauer gemeinsam mit Robert Schöpflin durch. Konstantin Helmsauer entwickelte den Ansatz

zur Rekonstruktion heterogener Amplikonpopulationen in der Zelllinie TR-14. Er analysierte den Zusammenhang von lokalen Chromatineigenschaften und deren Einfluss auf potentielle intermolekulare Kontakte auf Basis von Hi-C Daten für die Zelllinie TR-14.

Konstantin Helmsauer erstellte die Abbildungen Fig. 4h, Extended Data Figure 10a und Extended Data Figure 10c-h. Er erstellte Teile des Methodenabschnitts des Manuskripts und trug zur Revision von Manuskriptentwürfen und der Beantwortung von Gutachter*innenfragen im Peer Review bei.

_____

Unterschrift, Datum und Stempel des/der erstbetreuenden Hochschullehrers/in

_____

Unterschrift des Doktoranden/der Doktorandin

7 Publications

7.1 Helmsauer et al. Enhancer hijacking determines extrachromosomal circular *MYCN* amplicon architecture in neuroblastoma

The following section contains a reproduction of the article Helmsauer K, Valieva ME, Ali S, et al. Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma. Nat Commun. 2020;11:5823. https://doi.org/10.1038/s41467-020-19452-y. Reproduced with permission from Springer Nature.

Journal Data Filtered By: **Selected JCR Year: 2018** Selected Editions: SCIE,SSCI
Selected Categories: **"MULTIDISCIPLINARY SCIENCES"** Selected Category
Scheme: WoS
Gesamtanzahl: 69 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NATURE	745,692	43.070	1.285010
2	SCIENCE	680,994	41.037	1.070190
3	National Science Review	1,842	13.222	0.006500
4	Science Advances	21,901	12.804	0.110010
5	Nature Communications	243,793	11.878	1.103290
6	Nature Human Behaviour	1,230	10.575	0.006550
7	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	661,118	9.580	1.022190
8	Science Bulletin	3,569	6.277	0.009840
9	Scientific Data	3,240	5.929	0.015610
10	Frontiers in Bioengineering and Biotechnology	1,994	5.122	0.006540
11	Journal of Advanced Research	2,691	5.045	0.004780
12	Research Synthesis Methods	1,932	5.043	0.005420
13	GigaScience	2,674	4.688	0.012510
14	Annals of the New York Academy of Sciences	46,385	4.295	0.025840
15	Scientific Reports	302,086	4.011	1.061540
16	Journal of the Royal Society Interface	12,933	3.224	0.029190
17	NPJ Microgravity	203	3.111	0.000670
18	PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A-MATHEMATICAL PHYSICAL AND ENGINEERING SCIENCES	19,227	3.093	0.028200

1

Selected JCR Year: 2018; Selected Categories: "MULTIDISCIPLINARY SCIENCES"

Check for updates

Enhancer hijacking determines extrachromosomal circular *MYCN* amplicon architecture in neuroblastoma

Konstantin Helmsauer [1,14], Maria E. Valieva [2,3,14], Salaheddine Ali [2,3,4,14], Rocío Chamorro González [1], Robert Schöpflin[2,3], Claudia Röefzaad[5], Yi Bei [1], Heathcliff Dorado Garcia [1], Elias Rodriguez-Fos [6], Montserrat Puiggròs [6], Katharina Kasack[7], Kerstin Haase [1], Csilla Keskeny[1,7], Celine Y. Chen [1], Luis P. Kuschel [8], Philipp Euskirchen [7,8,9], Verena Heinrich[10], Michael I. Robson [2], Carolina Rosswog[11], Joern Toedling [1], Annabell Szymansky [1], Falk Hertwig[1], Matthias Fischer [11], David Torrents[6,12], Angelika Eggert[1,7,9], Johannes H. Schulte [1,7,9], Stefan Mundlos[2,3,4,15], Anton G. Henssen [1,5,7,9,15 ✉] & Richard P. Koche [9,13,15 ✉]

MYCN amplification drives one in six cases of neuroblastoma. The supernumerary gene copies are commonly found on highly rearranged, extrachromosomal circular DNA (ecDNA). The exact amplicon structure has not been described thus far and the functional relevance of its rearrangements is unknown. Here, we analyze the *MYCN* amplicon structure using short-read and Nanopore sequencing and its chromatin landscape using ChIP-seq, ATAC-seq and Hi-C. This reveals two distinct classes of amplicons which explain the regulatory require-ments for *MYCN* overexpression. The first class always co-amplifies a proximal enhancer driven by the noradrenergic core regulatory circuit (CRC). The second class of *MYCN* amplicons is characterized by high structural complexity, lacks key local enhancers, and instead contains distal chromosomal fragments harboring CRC-driven enhancers. Thus, ectopic enhancer hijacking can compensate for the loss of local gene regulatory elements and explains a large component of the structural diversity observed in *MYCN* amplification.

[1] Department of Pediatric Oncology/Hematology, Charité—Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. [2] RG Development & Disease, Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195 Berlin, Germany. [3] Institute for Medical and Human Genetics, Charité—Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. [4] Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité—Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. [5] Experimental and Clinical Research Center (ECRC), Max Delbrück Center for Molecular Medicine and Charité—Universitätsmedizin Berlin, Lindenberger Weg 80, 13125 Berlin, Germany. [6] Barcelona Supercomputing Center, Joint BSC-CRG-IRB Research Program in Computational Biology, Jordi Girona 29, 08034 Barcelona, Spain. [7] German Cancer Consortium (DKTK), partner site Berlin, and German Cancer Research Center DKFZ, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. [8] Department of Neurology with Experimental Neurology, Charité—Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. [9] Berlin Institute of Health, Anna-Louisa-Karsch-Str. 2, 10178 Berlin, Germany. [10] Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195 Berlin, Germany. [11] Department of Experimental Pediatric Oncology, University Children's Hospital of Cologne and Center for Molecular Medicine Cologne (CMMC), University of Cologne, Kerpener Straße 62, 50937 Köln, Germany. [12] Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain. [13] Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, 430 East 67th Street, New York, NY 10065, USA. [14] These authors contributed equally: Konstantin Helmsauer, Maria Valieva, Salaheddine Ali. [15] These authors jointly supervised this work: Stefan Mundlos, Anton G. Henssen, Richard P. Koche. ✉email: henssenlab@gmail.com; kocher@mskcc.org

Oncogene amplification is a hallmark of cancer genomes. It leads to excessive proto-oncogene overexpression and is a key driver of oncogenesis. The supernumerary gene copies come in two forms: (i) self-repeating arrays on a chromosome (homogeneously staining regions, HSR) and (ii) many individual circular DNA molecules (extrachromosomal DNA, ecDNA, alias double minute chromosomes, dmin)[1]. EcDNA can arise during genome reshuffling events like chromothripsis and are subsequently amplified[2,3]. This partially explains why ecDNA can consist of several coding and non-coding distal parts of one or more chromosomes[4]. Over time, amplified DNA acquires additional internal rearrangements as well as coding mutations, which can confer adaptive advantages such as resistance to targeted therapy[5–7]. EcDNA reintegration into chromosomes can lead to intrachromosomal amplification as HSRs[8,9] and act as a general driver of genome remodeling[10]. Our knowledge of the functional relevance of non-coding regions co-amplified on ecDNA, however, is currently limited.

MYCN amplification is a prototypical example of a cancer-driving amplification. The developmental transcription factor was identified as the most commonly amplified gene in a recent pediatric pan-cancer study[11]. Its most prominent role is in neuroblastoma, a pediatric malignancy of the sympathetic nervous system. *MYCN* amplification characterizes one in six cases and confers dismal prognosis[12]. In contrast to long-term survival of more than 80% for non-amplified cases, 5-year overall survival is as low as 32% for *MYCN*-amplified neuroblastoma[12]. In these cases, *MYCN* amplification is likely an early driver of neuroblastoma formation. Indeed, *MYCN* overexpression is sufficient to induce neuroblastic tumor formation in mice[13,14]. Despite its central role in neuroblastoma biology, the epigenetic regulation of *MYCN* is incompletely understood.

Recently, studies have identified a core regulatory circuit (CRC) including half a dozen transcription factors that drive a subset of neuroblastomas with noradrenergic cell identity, including most *MYCN*-amplified cases[15–18]. The epigenetic landscape around *MYCN* is less well characterized. In part, this is due to the structural complexity of *MYCN* amplicons and difficulties in interpreting epigenomic data in the presence of copy-number variation. Recent evidence has emerged suggesting that local enhancers may be required for proto-oncogene expression on amplicons[19]. Structural rearrangements can also juxtapose ectopic enhancers to proto-oncogenes and thereby drive aberrant expression, a phenomenon known as enhancer hijacking in several pediatric tumors[20–24]. Here, we seek out to identify key regulatory elements near *MYCN* in neuroblastoma by integrating short- and long-read genomic and epigenomic data from neuroblastoma cell lines and primary tumors. We investigate the activity of regulatory elements in the context of *MYCN* amplification and characterize the relationship between amplicon structure and epigenetic regulation. This reveals the retention of local CRC-driven enhancers on the *MYCN* amplicon in the majority of cases. When such local elements are not co-amplified, however, amplicons are structurally complex and distal elements are combined to form novel gene-regulatory neighborhoods.

Results

Defining the local enhancer landscape of *MYCN*. Acetylation at the 27th lysine residue of the histone H3 protein (H3K27ac) characterizes active chromatin at promoters and enhancers[25]. In order to identify candidate active regulatory elements near *MYCN*, we examined public H3K27ac chromatin immunoprecipitation and sequencing (ChIP-seq) and RNA sequencing (RNA-seq) data from 25 neuroblastoma cell lines[15]. ChIP-seq data for amplified genomic regions are characterized by a very low signal-to-noise ratio, which has complicated their interpretation in the past[16]. We therefore focused our analysis on 12 cell lines lacking *MYCN* amplifications but expressing *MYCN* at different levels, allowing for the identification of *MYCN*-driving enhancers in neuroblastoma. Comparison of composite H3K27ac signals of *MYCN*-expressing vs. non-expressing cell lines identified at least five putative enhancer elements (e1–e5) that were exclusively present in the vicinity of *MYCN* in cells expressing *MYCN*, thus likely contributing to *MYCN* regulation (Fig. 1 and Supplementary Fig. 1a). Consistent with differential RNA expression, a strong differential H3K27ac peak was identified spanning the *MYCN* promoter and gene body (MYCNp; Fig. 1). The identified
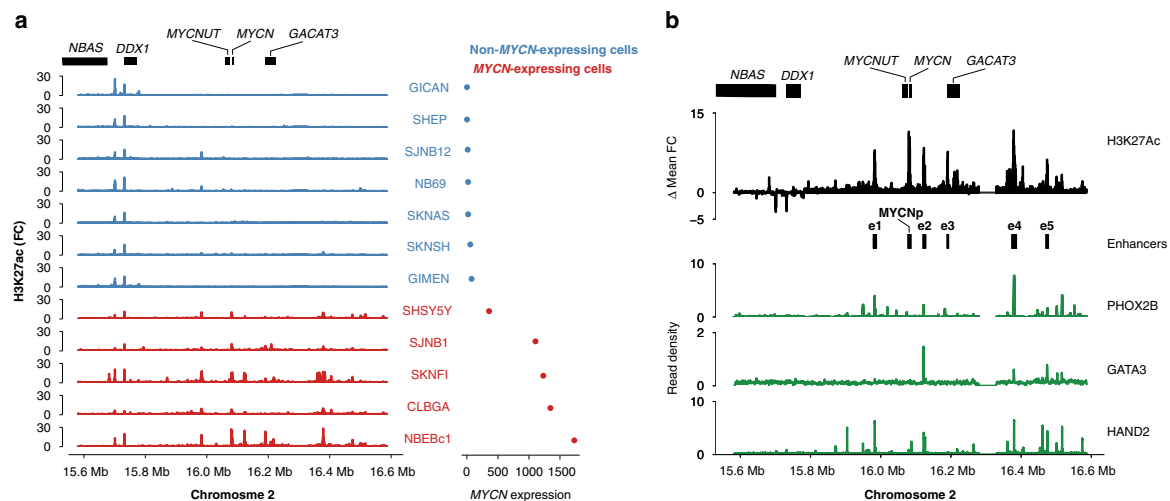

Fig. 1 Five enhancers are specifically found in *MYCN*-expressing neuroblastoma cells. a H3K27ac ChIP-seq fold change over input (left) and size-factor normalized *MYCN* expression as determined from RNA-seq for 12 non-MYCN-amplified neuroblastoma cell lines. **b** Aggregated H3K27ac signal of *MYCN*-expressing compared to non-expressing cells (top, black; MYCNp, *MYCN* promoter; e1–e5, *MYCN*-specific enhancers). PHOX2B, GATA3, and HAND2 core regulatory circuit transcription factor ChIP-seq in an *MYCN*-expressing neuroblastoma cell line (green, CLB-GA). Source data are provided as a Source Data file.

61

enhancers were not active in developmental precursor cells such as embryonic stem cells, neuroectodermal cells, neural crest cells, or fetal adrenal cells (Supplementary Fig. 1b), suggesting these enhancers were specific for later stages of sympathetic nervous system development or neuroblastoma. Transcription factor ChIP-seq in *MYCN*-expressing cells confirmed that four of the enhancers (e1, e2, e4, and e5) were bound by each of three noradrenergic neuroblastoma core regulatory circuit transcription factors (PHOX2B, HAND2, GATA3; Fig. 1b). All but enhancer e3 harbored binding motifs for the remaining members of the CRC (ISL1, TBX2, ASCL1; Supplementary Fig. 1c) for which ChIP-seq data were unavailable. Additionally, all enhancers contained binding motifs for TEAD4, a transcription factor implicated in a positive feedback loop with MYCN in *MYCN*-amplified neuroblastoma[26]. Two of the enhancers (e1 and e2) also harbored canonical E-boxes, suggesting binding of MYCN at its own enhancers (Supplementary Fig. 1c). Taken together, a common set of CRC-driven enhancers is found uniquely in *MYCN*-expressing neuroblastoma cells, indicating that *MYCN* expression is regulated by the CRC.

Enhancer selection explains MYCN amplicon boundaries. *MYCN* is expressed at the highest levels in neuroblastomas harboring *MYCN* amplifications, with a strong effect of genomic copy number on expression levels (Supplementary Fig. 1d, e). It is unclear, however, to what extent enhancers are required for sustained *MYCN* expression on *MYCN*-containing amplicons. To address this, we mapped amplified genomic regions in a meta-dataset of copy-number variation in 240 *MYCN*-amplified neuroblastomas[27]. This revealed an asymmetric pattern of *MYCN* amplification (Fig. 2a and Supplementary Fig. 2). Intriguingly, a 290 kb region downstream of *MYCN* was co-amplified in more than 90% of neuroblastomas, suggesting that *MYCN* amplicon boundaries were not randomly distributed, which is in line with

recent reports using a smaller tumor cohort[19]. Notably, the consensus amplicon boundaries did not overlap with common fragile sites (Supplementary Fig. 2g), challenging a previous association found in 24 neuroblastoma cell lines and tumors[28]. Regions of increased chromosomal instability alone are therefore unlikely to explain amplicon boundaries. Strikingly, several *MYCN*-specific enhancers were found to be commonly co-amplified (Fig. 2b). The distal *MYCN*-specific CRC-driven enhancer, e4, was part of the consensus amplicon region in 90% of cases. Randomizing amplicon boundaries around *MYCN* showed that e4 co-amplification was significantly enriched on *MYCN* amplicons (empirical $P = 0.0003$). Co-amplification frequency quickly dropped downstream of e4, suggesting that *MYCN*-specific, CRC-driven enhancers are a determinant of *MYCN* amplicon structure and may be required for *MYCN* expression, even in the context of high-level amplification.

Considering that *MYCN* is amplified in many pediatric cancer entities that differ in chromatin landscape, we hypothesized that *MYCN* amplicon structure should also differ between cancer entities. To test this, we inspected the amplicon architecture in a cohort of sonic hedgehog-driven medulloblastomas (SHH-MB) and Group 4 medulloblastomas (GROUP4-MB)[29], which often harbor *MYCN* amplifications and are commonly thought to originate from different precursor cell types[30]. In line with our model of tissue-specific enhancer co-amplification, *MYCN* amplicon structure differed between medulloblastomas and neuroblastomas (Supplementary Fig. 3a). *MYCN* amplicon distributions also differed between SHH-MB and GROUP4-MB (Supplementary Fig. 3b). A SHH-MB-specific super-enhancer (SE) > 350 kb downstream of *MYCN* was co-amplified in 8/9 cases, indicating selection. GROUP4-MB lack *MYCN*-driving SEs and are characterized by several enhancers close to *MYCN*. At least one of these local enhancers was co-amplified in 11/12 cases. Thus, tissue-specific enhancers are a determinant of *MYCN*
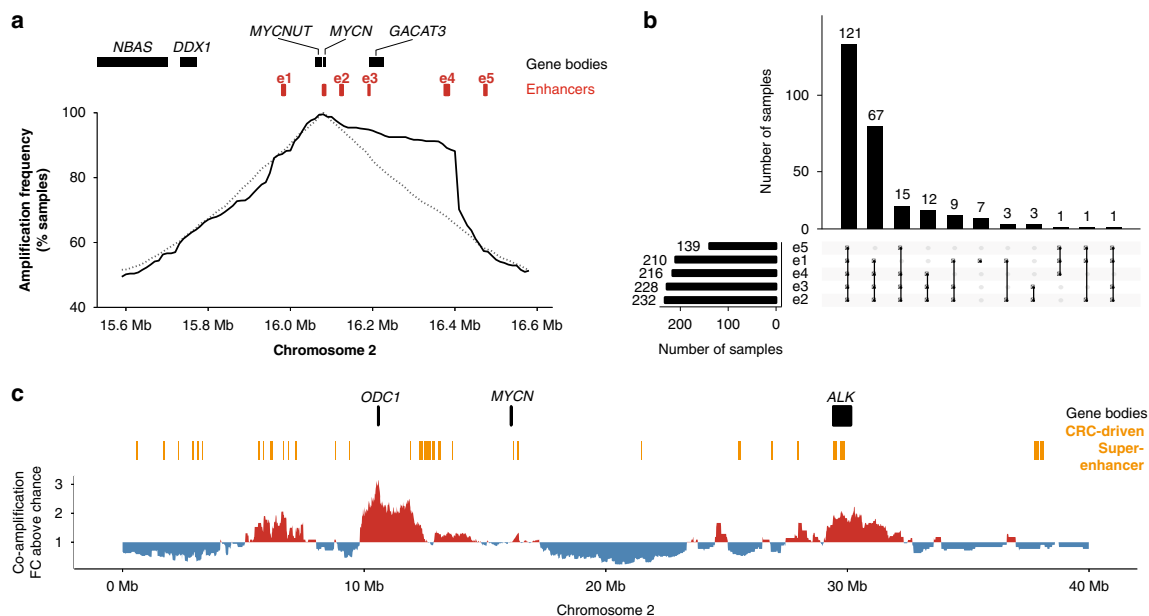
Fig. 2 Local enhancer e4 is significantly co-amplified with *MYCN*. a Co-amplification frequency of the immediate *MYCN* neighborhood measured using copy-number profiles from 240 *MYCN*-amplified neuroblastomas (solid line) compared to the expected co-amplification frequencies for randomized *MYCN*-containing amplicons (dashed line). **b** Upset plot showing the co-amplification patterns of all five *MYCN*-specific local enhancers identified in neuroblastoma. **c** Enrichment for co-amplification with *MYCN* of genomic regions on 2p (red, co-amplification more frequent than expected by chance; blue, co-amplification less frequent than expected by chance). Source data are provided as a Source Data file.

amplicon structure and may be required for *MYCN* expression in various tumor entities.

Distal super-enhancer co-amplification with *MYCN*. We and others have previously described chimeric *MYCN* amplicons[10] containing distal chromosomal fragments. We therefore systematically inspected *MYCN*-distal regions on chromosome 2 for signs of co-amplification. Distinct regions were statistically enriched for co-amplification with *MYCN* (Fig. 2c). In line with previous reports[31], significant co-amplification of 19 protein-coding genes, including known neuroblastoma drivers such as *ODC1*, *GREB1,* and *ALK* occurred in *MYCN*-amplified neuroblastoma. Notably, co-amplification of distal CRC-driven SEs occurred in 23.3% of samples. Seven specific CRC-driven SEs were significantly co-amplified more often than expected by chance. Most of these SEs were found in gene-rich regions, making it difficult to discern whether genes or regulatory elements were driving co-amplification. One significantly co-amplified CRC-driven SE, however, was found in a gene-poor region in 2p25.2, where most co-amplified segments did not overlap protein-coding genes (Fig. 2c). This led us to ask whether hijacking of such distal regulatory elements could explain co-amplification with *MYCN*.

Enhancers remain functional on *MYCN* amplicons. Based on our amplicon boundary analysis, two classes of *MYCN* amplicons could be distinguished in neuroblastoma: (i) amplicons containing local *MYCN*-specific enhancers, including e4 (here referred to as class I amplicons; Fig. 3a) and (ii) amplicons lacking local *MYCN*-specific enhancers, and at least lacking e4 (referred to as class II amplicons; Fig. 3b). To determine whether co-amplified enhancers were active, we acquired genomic (long- and short-read whole-genome sequencing) and epigenomic (Assay for Transposase-Accessible Chromatin using sequencing, ATAC-seq, and mono-methylation at the fourth lysine residue of the histone H3, H3K4me1, and H3K27ac ChIP-seq) data for two neuroblastoma cell lines with class I amplicons (Kelly and NGP) and two neuroblastoma cell lines with class II amplicons (IMR-5/75 and CHP-212). Notably, H3K27ac signal-to-noise ratio was lower on *MYCN* amplicons than in non-amplified regions. While the fraction of reads in peaks was similar across amplicons and randomly drawn regions, we observed more peaks on the amplicon than for non-amplified regions (Supplementary Fig. 4). These peaks were characterized by a lower relative signal compared to the amplicon background signal, indicating a larger variety of active regulatory regions on different *MYCN* amplicons. Using nanopore long-read-based de novo assembly, we reconstructed the *MYCN* neighborhood, confirming that *MYCN* and e4 were not only co-amplified in class I amplicons, but also lacked large rearrangements, which could preclude enhancer–promoter interaction (Supplementary Figs. 5 and 6). Enhancer e4 was characterized by increased chromatin accessibility and active enhancer histone marks as determined by ATAC-seq, H3K4me1, and H3K27ac ChIP-seq (Fig. 3c). Importantly, 4C chromatin conformation capture analysis showed that e4 spatially interacted with the *MYCN* promoter on the amplicon (Fig. 3c). Thus, e4 presents as a functional enhancer and appears to contribute to *MYCN* expression, even in the context of class I *MYCN* amplification.

Enhancer hijacking compensates for local enhancer loss. In contrast to class I amplicons, class II amplicons lacked key local enhancers and nevertheless expressed relatively high levels of *MYCN* per gene copy, raising the possibility of alternative routes of *MYCN* regulation (Supplementary Fig. 7). The lack of a strong

local regulatory element on class II amplicons and our observation of frequent co-amplification of distal SE (Fig. 2c) led us to hypothesize that ectopic enhancers might be recruited to enable *MYCN* expression in class II amplicons. In agreement with our hypothesis, primary neuroblastomas with class II amplicons were more likely to harbor complex amplifications containing more than one amplified fragment in the genome (66.7% vs. 35.7%, Fisher's exact test $P = 0.003$; Fig. 3e). In this largely array-based dataset, we cannot exclude fragments that are not structurally fused to the *MYCN* locus. However, it is unlikely that highly amplified loci have very similar copy number if they are not part of a common amplicon. We therefore filtered for fragments with highly similar copy number as *MYCN* (log ratio difference ≤0.1) and again found increased amplicon complexity for class II (class II 36.0% vs. class I 11.6%, Fisher's exact test $P = 0.003$). All but one class II amplicon co-amplified at least one CRC-driven enhancer element distal of *MYCN*. Some of these enhancers were recurrently found on class II amplicons, including an enhancer 1.2 Mb downstream of *MYCN* that was co-amplified in 20.8% (5/24) of *MYCN*-amplified neuroblastomas, 2.1-fold higher than expected for randomized amplicons that include *MYCN* but not e4 (Fig. 3f). Thus, class II *MYCN* amplicons are characterized by high structural complexity, allowing for the replacement of local enhancers through hijacking of distal CRC-driven enhancers.

To determine the structure and epigenetic regulation of class II amplicons in detail, we inspected long-read-based de novo assemblies and short-read-based reconstructions of IMR-5/75 and CHP-212 *MYCN* amplicons. High-throughput chromosome conformation capture (Hi-C) was performed and validated the reconstructions, recapitulating the order and orientation of the joined fragments. IMR-5/75 was characterized by a linear HSR class II *MYCN* amplicon, not including e3–e5 (Fig. 3b). Inspection of the IMR-5/75 *MYCN* amplicon structure revealed that the amplicon consisted of six distant genomic regions, which were joined together to form a large and complex chimeric amplicon (Fig. 4a–d). One of the fragments was likely included as a tandem duplication on the amplicon (Supplementary Fig. 8a). In line with enhancer hijacking, a segment of *ALK* containing a large SE, marked by H3K27ac and chromatin accessibility as measured using ATAC-seq, was juxtaposed with *MYCN* on the chimeric amplicon. Similar to e4, this enhancer was bound by adrenergic CRC factors in non-amplified cells (Supplementary Fig. 9a). In CHP-212, *MYCN* is amplified on ecDNA, as confirmed by fluorescence in situ hybridization (Supplementary Fig. 10). Both de novo assembly and short-read-based reconstruction of the amplicon confirmed the circular *MYCN* amplicon structure independently (Fig. 4f–h). Similar to IMR-5/75, distal fragments containing CRC-driven SEs were joined to the *MYCN* neighborhood (Fig. 4e, f and Supplementary Fig. 9b).

Neo-topologically associated domains (TADs) form on chimeric *MYCN* amplicons. To analyze the three-dimensional conformation of circular and linear amplicons we mapped Hi-C reads to the reconstructed amplicon (Fig. 4c, g). Notably, high-frequency interactions in the corners of the maps opposite to the main diagonal confirmed the circularity of CHP-212 amplicon and tandem duplication-type amplification in IMR-5/75. On a more local level, Hi-C can be used to characterize TADs, i.e. regions of increased spatial interaction which contribute to gene control and arise through chromatin loops anchored at CTCF-marked insulator elements[32]. In IMR-5/75 and CHP-212, we observed insulated TADs as in the rest of the genome, suggesting that general rules of chromatin topology are retained on ecDNA and HSRs. Due to the rearrangements in CHP-212, the *MYCN* gene became part of a new chromatin domain (neo-TAD) where
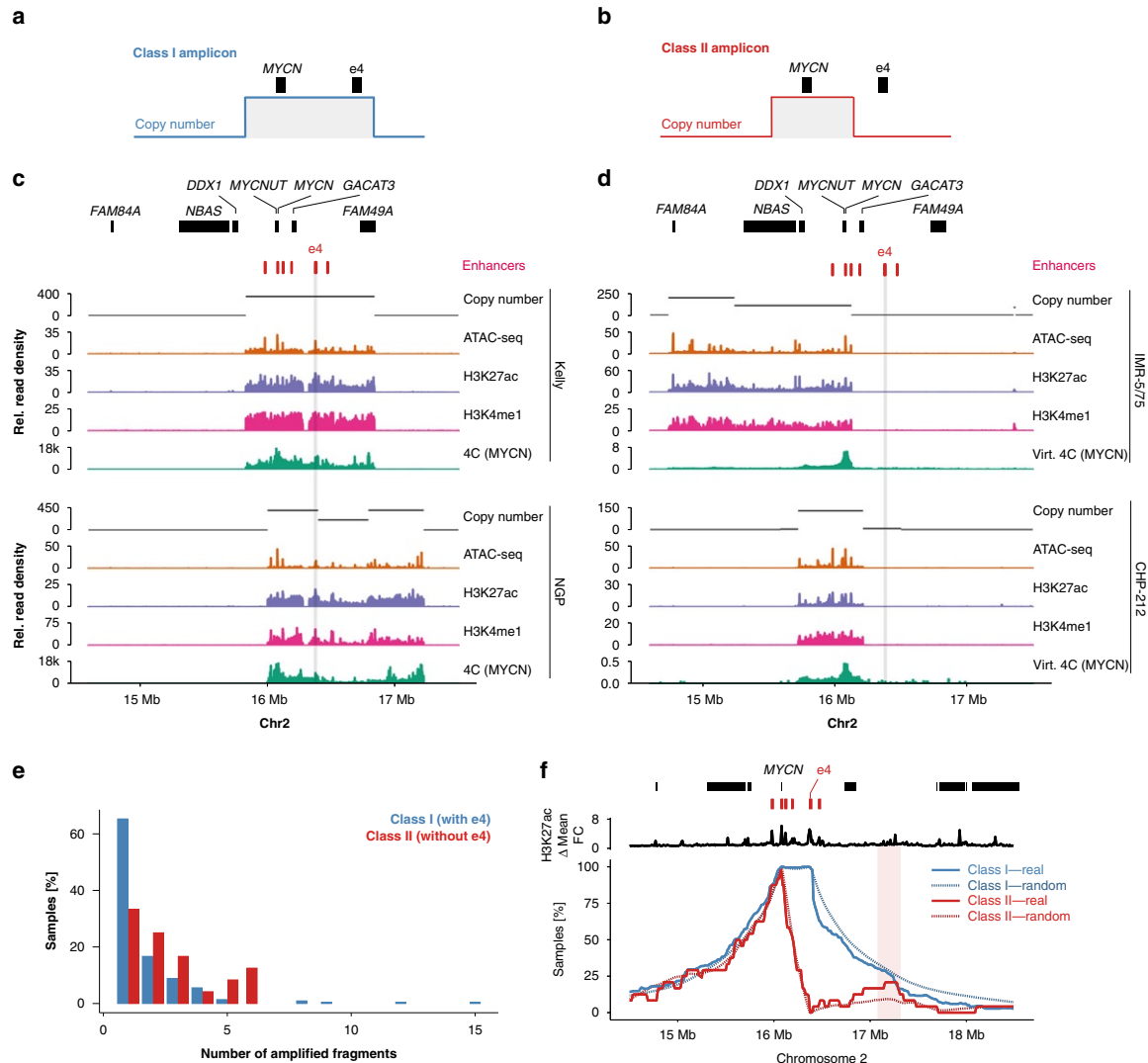
Fig. 3 Two classes of MYCN amplicons can be identified in neuroblastoma. Schematic representation of class I (**a**) and class II (**b**) *MYCN* amplicons. **c** Copy-number profile (black), ATAC-seq (orange), H3K27ac ChIP-seq (purple), H3K4me1 ChIP-seq (pink), and 4C (*MYCN* promoter as the viewpoint; green) for two neuroblastoma cell lines with class I amplicons, co-amplifying the e4. **d** Copy-number profile (black), ATAC-seq (orange), H3K27ac ChIP-seq (purple), H3K4me1 ChIP-seq (pink), and virtual 4C (*MYCN* locus as the viewpoint; green) for two neuroblastoma cell lines class II amplicons, not co-amplifying e4. **e** Number of non-contiguous amplified fragments in class I samples ($N = 216$) vs. class II samples ($N = 24$). **f** Amplicon boundary frequency relative to gene and enhancer positions in class I vs. class II amplicons compared to random amplicon boundary frequencies. Source data are provided as a Source Data file.

genes, enhancers, and insulators from distal parts of the genome form a new spatially interacting neighborhood. *MYCN* itself was located at the intersection of two smaller sub-TADs. The first sub-TAD originated from the wild-type genome as an intact unit. The second sub-TAD resulted from the fusion of the *MYCN* locus with another region from a distal part of chromosome 2 (chr2:12.6–12.8 Mb) containing CRC-driven SEs (Fig. 4g and Supplementary Fig. 9b). The fused segments were part of one TAD and not separated by a boundary, which enables the interaction of *MYCN* with the ectopic SEs. A similar situation was observed for the linear amplicon in IMR-5/75, where frequent contacts between *MYCN* and SEs from the genomic regions juxtaposed to *MYCN*, containing intronic parts of *ALK*, were

detected using Hi-C (Fig. 4c and Supplementary Figs. 8b and 9a). Notably, hijacked SEs covered 46% and 44% of the neo-TAD for IMR-5/75 and CHP-212, respectively. In both cell lines, additional fragments of chromosome 2 were fused to the SE-containing region. These contained neo-TAD boundaries as determined by Hi-C (Fig. 4d, g). All neo-TAD boundaries were marked by CTCF ChIP-seq peaks, with canonical forward–reverse motif orientations in IMR-5/75 (Supplementary Fig. 9a). In CHP-212, no unambiguous CTCF motif orientations at the downstream neo-TAD border were identified (Supplementary Fig. 9b). In both cases, however, the new insulators originated from genomic locations other than the *MYCN* fragment and the SE-containing fragments. In addition to the
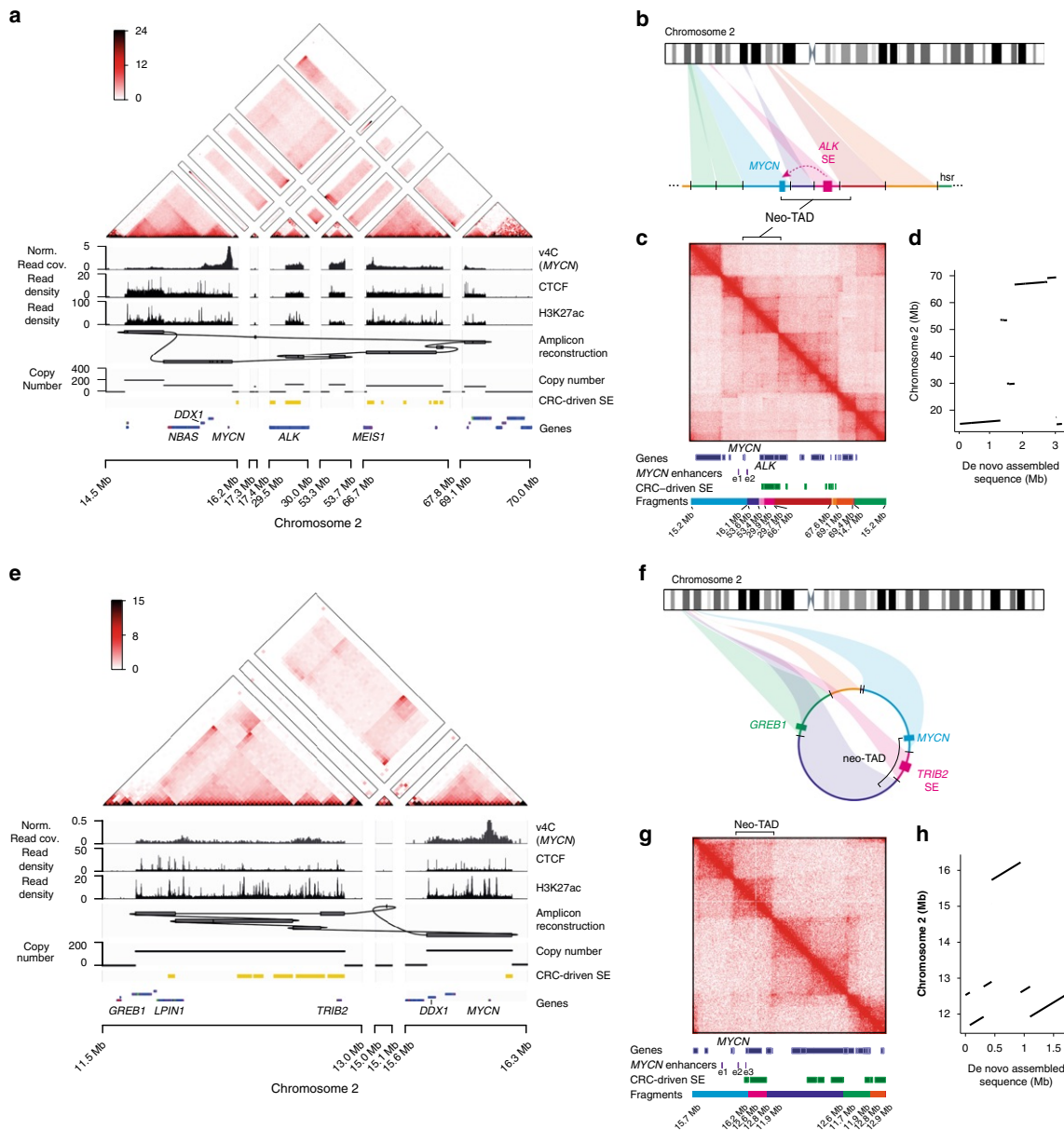
Fig. 4 Reconstruction and epigenetic markup of class II *MYCN* amplicons. a, e Short-read-based reconstruction and epigenomic characterization of the *MYCN* amplicon in IMR-5/75 (**a**) and CHP-212 (**e**) cells. Top to bottom: Hi-C map (color indicating Knight–Ruiz normalized read counts in 25 kb bins), virtual 4C (*MYCN* viewpoint, v4C), CTCF ChIP-seq, H3K27Ac ChIP-seq, Amplicon reconstruction, copy-number profile, super-enhancer locations (yellow), gene positions (blue). **b, f** Schematic representation of the class II amplicon described in **a, e**, showing ectopic enhancers and insulator reshuffling leading to locally disrupted regulatory neighborhoods on the HSR in IMR-5/75 (**b**) and on ecDNA in CHP-212 (**f**). **c, g** Alignment of Hi-C reads to the reconstructed *MYCN* amplicon in IMR-5/75 (**c**) and CHP-212 (**g**) and positions of genes, local *MYCN* enhancers and CRC-driven super-enhancers on the amplicon. **d, h** Mapping of the long-read sequencing-based de novo assembly of the *MYCN* amplicon in IMR-5/75 (**d**) and CHP-212 (**h**) on chromosome 2. Source data are provided as a Source Data file.

observed TAD structures, weaker off-diagonal interactions were visible, suggesting a heterogeneous group of structurally different variants of the original amplicon. Nevertheless, the TAD structure, boundaries, and loops were clearly visible on the reconstructed Hi-C map (Fig. 4c). Thus, hijacking of ectopic enhancers and insulators can compensate for the loss of endogenous regulatory elements on intra- and extrachromosomal circular class II *MYCN* amplicons via the formation of neo-TADs, which may explain the higher structural complexity of *MYCN* amplicons lacking endogenous enhancers.

65
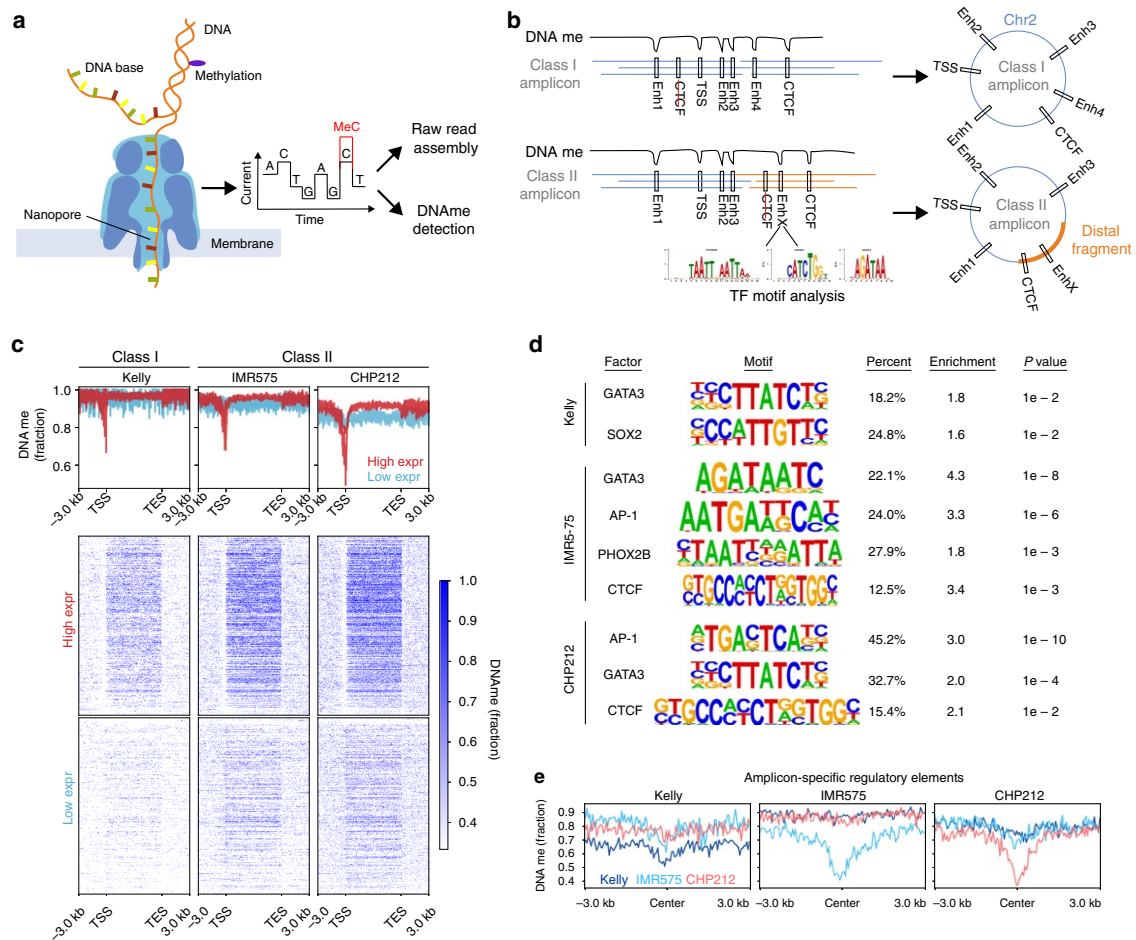
Fig. 5 Nanopore sequencing characterizes DNA methylation on *MYCN* amplicons. a Schematic of experimental approach. **b** Schematic representation of how nanopore sequencing facilitates de novo amplicon assembly and can be used to simultaneously to detect regulatory elements through DNA methylation analysis. **c** Composite DNA methylation signal detected using nanopore sequencing over genes expressed at high (HighExpr) vs. low levels (LowExpr). **d** Motif analysis based on accessibility in regulatory elements co-amplified on *MYCN* amplicons (unadjusted *P* values from one-sided binomial test against nucleotide composition-matched background sequences). **e** Amplicon-specific methylation pattern detected in three neuroblastoma cell lines (Kelly, IMR-5/75, CHP-212) using nanopore sequencing-based DNA methylation analysis.

Nanopore sequencing characterizes amplicon methylation. In addition to allowing the alignment-free de novo assembly of the *MYCN* amplicon in several samples (Fig. 4b–d, f–h and Supplementary Figs. 5 and 6), nanopore sequencing also allows for the direct measurement of DNA methylation without the need for bisulfite conversion (Fig. 5a)[33]. While DNA methylation at regulatory elements is often associated with repression, a trough in DNA methylation may indicate a transcription factor-binding event, a poised or active gene-regulatory element, or a CTCF-occupied insulator element (Fig. 5b). In theory, nanopore sequencing and assembly might allow for the simultaneous inference of both structure and regulatory landscape (Fig. 5b). Prior to evaluating the *MYCN* amplicons, the DNA methylation landscape of highly expressed and inactive genes demonstrated the expected distribution of decreased methylation at active promoters and increased methylation within active gene bodies (Fig. 5c). In order to assess the DNA methylation status of putative regulatory elements near *MYCN*, we first used the amplicon-enriched ATAC-seq peaks to classify relevant motif

signatures (Fig. 5d). While *MYCN* was surrounded by the expected CRC-driven regulatory elements at the overlapping core enhancers as well as some CTCF sites, both their number and location varied, indicating sample-specific sites of regulation. Indeed, DNA methylation decreased in accordance with sites specific to a given sample (Fig. 5e), opening up the possibility of using these data to infer regulatory elements in patient samples, when no orthogonal epigenomic data are available.

Class II amplicons clinically phenocopy class I amplicons. *MYCN*-amplified neuroblastoma is characterized by significant clinical heterogeneity, which cannot entirely be explained by genetic differences. Whether the structure of the *MYCN* amplicon itself could account for some of this variation is currently unknown. In line with previous reports[31], higher counts of amplified fragments were associated with a more malignant clinical phenotype (Fig. 6a). Co-amplification of *ODC1*, a gene located 5.5 Mb upstream of *MYCN* and co-amplified in 9% (21/
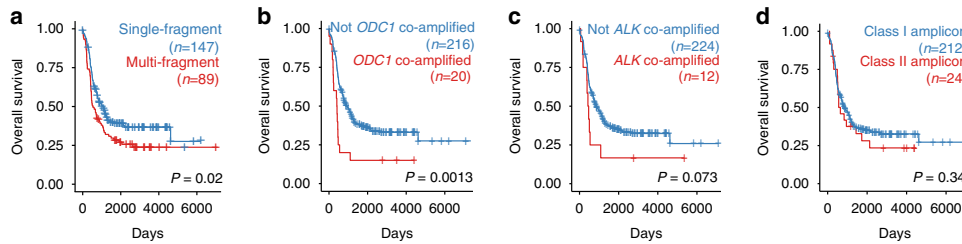
Fig. 6 Class II amplicons clinically phenocopy class I amplicons. Kaplan Meier survival analysis of patients with *MYCN*-amplified neuroblastoma, comparing single-fragment vs. multi-fragment amplification (**a**), co-amplification of *ODC1* vs. no co-amplification (**b**), co-amplification of *ALK* vs. no co-amplification (**c**), and class I amplicons vs. class II amplicons (**d**; $N = 236$ *MYCN*-amplified neuroblastomas; *P* value based on two-sided log-rank test). Source data are provided as a Source Data file.
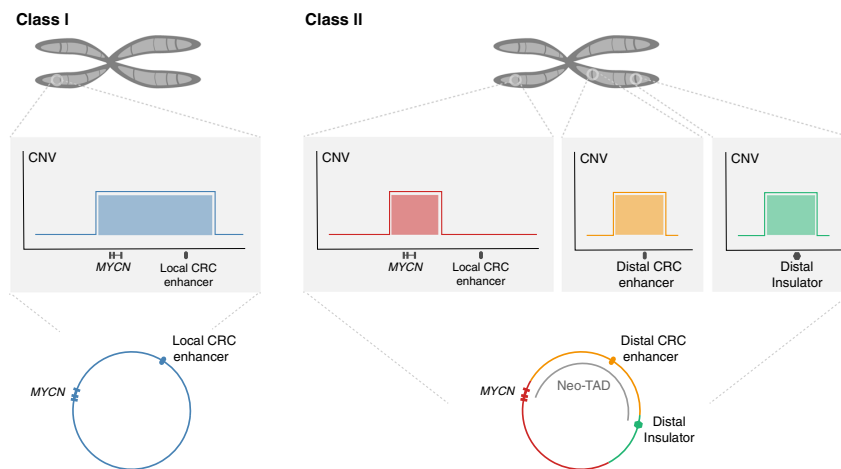
Fig. 7 Enhancer co-amplification determines *MYCN* amplicon patterns. In most cases, *MYCN* and its local gene-regulatory neighborhood including a CRC-driven super-enhancer is amplified (Class I). If the local neighborhood is not co-amplified, amplicons are more complex and recruit distal gene-regulatory elements (Class II).

240) of *MYCN*-amplified neuroblastomas (Fig. 2c), defined an ultra-high-risk genetic subgroup of *MYCN*-amplified neuroblastoma (hazard ratio (HR) 2.3 (1.4–3.7), log-rank test $P = 0.001$; Fig. 6b). Similarly, *ALK* co-amplification, present in 5% (12/240) of *MYCN*-amplified tumors, was also associated with adverse clinical outcome (HR 1.8 (0.94–3.4), log-rank test $P = 0.073$; Fig. 6c). In contrast, differences in the *MYCN* amplicon enhancer structure, i.e. class I vs. class II amplification, did not confer prognostic differences (HR 1.3 (0.78–2.1), log-rank test $P = 0.34$; Fig. 6d). We therefore conclude that chimeric co-amplification of proto-oncogenes partly explains the malignant phenotype of neuroblastomas with complex *MYCN* amplicons, whereas enhancer hijacking in class II amplicons does not change clinical behavior, fully phenocopying class I *MYCN* amplicons.

Discussion

Here, we show that neuroblastoma-specific CRC-driven enhancers contribute to *MYCN* amplicon structure in neuroblastoma and retain the classic features of active enhancers after genomic amplification. While most *MYCN* amplicons contain local enhancers, ectopic enhancers are regularly incorporated into chimeric amplicons lacking local enhancers, leading to enhancer hijacking (Fig. 7).

A large subset of neuroblastomas was recently found to be driven by a small set of transcription factors that form a self-sustaining CRC, defined by their high expression and presence of super-enhancers[15–18]. The extent to which *MYCN* itself is directly regulated by CRC factors was previously unclear, complicated by the challenge of interpreting epigenomic data on amplicons[16]. Our results provide empiric evidence that *MYCN* is driven by CRC factors, even in the context of *MYCN* amplification. This could mechanistically explain the previous observation that genetic depletion of CRC factors represses *MYCN* expression even in *MYCN*-amplified cells[16]. The finding that ectopic enhancers driven by the CRC are juxtaposed to *MYCN* on amplicons that lack local enhancers further strengthens the relevance of the CRC in *MYCN* regulation.

In line with our observation of local enhancer co-amplification, Morton et al.[19] recently described that local enhancers are significantly co-amplified with other proto-oncogenes in other cancer entities. They showed that experimentally interfering with local *EGFR* enhancers in *EGFR*-amplified glioblastoma impaired oncogene expression and cell viability in *EGFR*-amplified as well as non-amplified cases. Consistent with our findings, the authors identified a region overlapping e4 that was significantly co-amplified in *MYCN*-amplified neuroblastomas, corresponding to class I amplicons observed in our cohort. In contrast to Morton et al.[19], who suggest that the inclusion of local enhancers is necessary for proto-oncogene expression on amplicons, we show that exceptions to this rule occur in a significant subset of *MYCN*-amplified neuroblastomas. In such cases, amplicons characterized by highly

complex chimeric structure enable the reshuffling of ectopic enhancers and insulators to form neo-TADs that can compensate for disrupted local neighborhoods through enhancer hijacking.

More generally, we show that TADs also form on ecDNA, in parallel with recent findings by Wu et al.[34]. We extend this observation to HSRs, which form extremely expanded stretches of chromatin in interphase nuclei and lose chromosomal territoriality[35]. Gene activation by enhancer adoption requires the fusion of distant DNA fragments and the formation of new chromatin domains, called neo-TADs[36]. In some cases, this fusion requires a convergent directionality of CTCF sites in order to form a new boundary and drive aberrant gene expression[37]. This has been explained by a model of blocked loop extrusion at forward–reverse oriented CTCF sites[32]. We found convergent CTCF for the neo-TAD in IMR-5/75 but not necessarily for the one in CHP-212. However, non-convergent CTCF sites have been consistently reported before and characterize at least one in ten CTCF-mediated chromatin loops in the wild-type genome[38,39]. Although the exact underpinnings are not yet clear, CTCF convergence is likely not required in some genomic contexts, which could be the case in CHP-212 and other ecDNA amplicons.

Reconstruction of amplicons has previously relied on combining structural breakpoint coordinates to infer the underlying structure. This regularly resulted in ambiguous amplicon reconstructions, which had to be addressed by secondary data such as chromium linked reads or optical mapping[4,6,34]. We demonstrate the feasibility of long-read de novo assembly for the reconstruction of amplified genomic neighborhoods. De novo assembly was able to reconstruct entire ecDNA molecules and confirm the tandem duplicating nature of HSRs. Integrating de novo assembly with methylation data from nanopore sequencing reads will likely benefit further studies of other proto-oncogene-containing amplicons by enabling the characterization of the interplay between structure and regulation in highly rearranged cancer genomes.

Functional studies have shown that both *ODC1* and *ALK* are highly relevant in neuroblastoma[40,41]. Co-amplification with *MYCN* has been reported before[31], but to our knowledge the clinical relevance of co-amplification had not been determined so far. Similar to our previous observations of *PTP4A2* co-amplification on chimeric ecDNA[10], we demonstrate here that proto-oncogenes reside side-by-side on the same ecDNAs, sometimes even sharing the same regulatory neighborhood. It is tempting to speculate that this structural coupling of genes could confer MYCN-independent but *MYCN*-amplicon-specific, collateral therapeutic vulnerabilities in *MYCN*-amplified tumors.

We conclude that the structure of genomic amplifications can be explained by a selective pressure to amplify oncogenes together with suitable non-coding regulatory elements. CRC-driven enhancers are required for successful *MYCN* amplification and remain functional throughout this process. Even though the majority of amplicons contain endogenous enhancers, these can be functionally replaced by ectopic CRC-driven enhancers that are juxtaposed to the oncogene through complex chimeric amplicon formation. We envision that our findings also extend to oncogene amplifications in other cancers and will help identify functionally relevant loci among the diverse array of complex aberrations that drive cancer.

Methods

Cell lines. Neuroblastoma cell lines were a gift from F. Speleman (Cancer Research Institute Ghent, Ghent, Belgium; NGP), F. Westermann (German Cancer Research Center, Heidelberg, Germany; IMR-5/75), obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ GmbH, Braunschweig, Germany; Kelly), or obtained from the American Type Culture Collection (ATCC, Manassas, VA; CHP-212). Cell line identity was verified by STR genotyping (Genetica DNA Laboratories, Burlington, NC and IDEXX BioResearch, Westbrook, ME) and absence of Mycoplasma sp. contamination was determined with a Lonza

MycoAlert system (Lonza Group Ltd, Basel, CH). All cell lines were cultured in RPMI-1640 medium (Thermo Fisher Scientific, Inc., Waltham, MA) with 1% Penicillin/Streptomycin and 10% FCS.

RNA-seq. Public RNA-seq data were downloaded from Gene Expression Omnibus (GSE90683)[15]. FASTQ files were quality controlled (FASTQC 0.11.8) and adapters were trimmed (BBMap 38.58). We mapped reads to GRCh37 (STAR 2.7.1 (ref. [42]) with default parameters), counted them per gene (Ensembl release 75, feature-Counts from Subread package 1.6.4 (ref. [43])), and normalized for library size and composition (sizeFactors from DESeq2 1.22.2 (ref. [44])).

ChIP-seq. For the cell lines CHP-212, NGP, and Kelly, $5–10 \times 10^6$ cells were digested with Trypsin–EDTA 0.05% (Gibco) for 10 min at 37 °C. The cells were mixed with 10% FCS–PBS, and a single-cell suspension was obtained using a 40-µm cell strainer. After centrifugation, cells were resuspended in 10% FCS–PBS again and fixed in 1% paraformaldehyde (PFA) for 10 min at room temperature. The reaction was quenched with 2.5 M glycine (Merck) on ice and centrifuged at 400g for 8 min. We resuspended cell pellets in lysis buffer (50 mM Tris, pH 7.5; 150 mM NaCl; 5 mM EDTA; 0.5% NP-40; 1.15% Triton X-100; protease inhibitors (Roche), 5 mM Na-butarate), and nuclei were pelleted again by centrifugation at 750g for 5 min. For sonication, nuclei were resuspended in sonication buffer (10 mM Tris–HCl, pH 8.0; 100 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.1% Na-deoxycholate; 0.5% *N*-lauroylsarcosine; protease inhibitors (Roche complete), 5 mM Na-butarate). Chromatin was sheared using a Diagenode Bioruptor (35–40 cycles with a 30 s on/off pulse and HI power mode) until reaching a fragment size of 200–500 base pairs (bp). Lysates were clarified from sonicated nuclei, and protein–DNA complexes were immunoprecipitated overnight at 4 °C with the respective antibody. A total of 10–15 µg chromatin was used for each replicate of histone ChIP and 20–25 µg of transcription factor ChIP. For the immunoprecipitation[45] in 1200 µl precipitation buffer (10 mM Tris–HCl, pH 8.0; 100 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.1% Na-deoxycholate; 0.5% *N*-lauroylsarcosine; protease inhibitors (Roche complete), 5 mM Na-butarate, 1% Triton X-100), Anti-H3K27ac (Diagenode c15410174; lot A7071-001P; dilution 1:500), anti-H3K4m1 (Abcam; ab8895; lot GR141677-1; dilution 1:1200), anti-RAD21 (Abcam; ab992; lot GR221348-8; dilution 1:150) and anti-CTCF (Active Motif; 613111; lot 34614003; dilution 1:150) antibodies were used. Sequencing libraries were prepared using standard Nextera adapters (Illumina) according to the supplier's recommendations. Twenty-five million reads per sample were sequenced on a HiSeq 4000 sequencer (Illumina) in 75 bp single read mode.

Additional public ChIP-seq FASTQ files were downloaded from Gene Expression Omnibus (GSE18927, GSE90683, GSE24447, and GSE28874)[15,46] and from ArrayExpress (E-MTAB-6570)[17]. FASTQ files were quality controlled (FASTQC 0.11.8) and adapters were trimmed (BBMap 38.58). Reads were then aligned to hg19 (BWA-MEM 0.7.15 (ref. [47]) with default parameters) and duplicate reads removed (Picard 2.20.4). We generated BigWig tracks by extending reads to 200 bp for single-end libraries and extending to fragment size for paired-end libraries, filtering by ENCODE DAC blacklist and normalizing to counts per million in 10 bp bins (deepTools 3.3.0 (ref. [48])). Peaks were called using MACS2 (2.1.2)[49] with default parameters. Super-enhancers were called for H3K27ac data using LILY[15] (https://github.com/BoevaLab/LILY) with default parameters. ChIP-seq data were quality controlled using RSC and NSC (Phantompeakqualtools 1.2.1). CTCF motifs within CTCF ChIP-seq peaks were identified using JASPAR2018 (ref. [50]) and the TFBSTools (1.20.0)[51] function matchPWM with min.score = "75%". Copy-number ratio was estimated by binning ChIP-seq input reads (primary alignments of mapping quality 20 or higher) in 1 kb bins, correcting for GC content, normalization, and segmentation using QDNAseq (1.22.0)[52].

ATAC-seq. ATAC-seq samples were processed as reported in Buenrostro et al.[53] with some adaptations: After a treatment of $5–10 \times 10^6$ cells with Trypsin–EDTA 0.05% (Gibco) for 10 min at 37 °C, a 40-µm cell strainer was used to obtain a single-cell suspension. 5×10^5 cells were washed with cold 1× PBS and lysed with freshly prepared lysis buffer (10 mM Tris–Cl pH 7.4, 10 mM NaCl, 3 mM MgCl$_2$, 0.1% (v/v) Igepal CA-630) by pipetting six times up and down and a subsequent incubation on ice for 1 min. After a centrifugation at 500g for 5 min at 4 °C the pellet was resuspended with gentle mixing in transposition reaction mix (25 µl 2× TD, 2.5 µl TDE1 and 22.5 µl H$_2$O, Illumina). Immediately after the transposition reaction, the DNA was purified using a MinElute PCR Purification Kit (Qiagen). The transposed DNA was amplified using a Nextera PCR Kit (Illumina) according to the supplier's recommendation. The maximum number of cycles was determined with qPCR to reduce PCR bias. For sequencing, libraries were generated using Illumina/Nextera adapters and size selected (100–1000 bp) with AMPure Beads (Beckman Coulter). Approximately 100 million 75 bp paired-end reads were acquired per sample on the HiSeq 4000 system (Illumina). Additional public ATAC-seq FASTQ files were downloaded from Gene Expression Omnibus (GSE80154)[54]. Adapter trimming, alignment, and duplicate removal as for ChIP-seq. We generated BigWig tracks by extending paired-end reads to fragment size, filtering by the ENCODE DAC blacklist and normalizing to counts per million in 10 bp bins (deepTools 3.3.0 (ref. [48])). Peaks were called using MACS2 (2.1.2)[49] with default parameters.

Hi-C. 3C libraries for Hi-C and 4C were prepared from confluent neuroblastoma cells according to the cell culture section above. Hi-C experiments were performed as duplicates. $5-10 \times 10^6$ cells were washed twice with PBS and digested with Trypsin–EDTA 0.05% (Gibco) for 10 min at 37 °C. A 40-µm cell strainer was used to obtain single cells. The cell suspension was pelleted at 300g for 5 min and resuspended with cold 10% FCS. Subsequently, the cells were fixed by adding an equal volume of 4% formaldehyde (Sigma-Aldrich). The suspension was mixed for 10 min while shaking at room temperature in 50 ml tubes. Exactly after 10 min the fixation was quenched with 500 µl 1.425 M glycine (Merck) on ice. The suspension was pelleted at 400g for 8 min and resuspended in cold lysis buffer (50 mM Tris, pH 7.5; 150 mM NaCl; 5 mM EDTA; 0.5% NP-40; 1.15% Triton X-100; protease inhibitors (Roche)). After a washing step with cold 1× PBS and centrifugation at 750g for 5 min, the pellet was washed with 1× DpnII buffer (NEB) and resuspended in 50 µl 0.5% SDS and incubated for 10 min at 62 °C. After that 145 µl water and 25 µl 10% Triton (Sigma) was added to quench the SDS followed by a incubation at 37 °C for 30 min. For the restriction enzyme digestion, 25 µl DpnII buffer and 100 U DpnII was added. The digestion reaction was incubated for 2 h at 37 °C, after 1 h another 10 U were added and then heat inactivated at 65 °C for 20 min.

The digested sticky ends were filled up with 10 mM dNTPs (without dATP) and 0.4 mM biotin-14-dATP (Life Technologies) and 40 U DNA Pol I, Large Klenow (NEB) at 37 °C for 90 min. Biotinylated blunt ends were then ligated using a ligation reaction (663 µl water, 120 µl 10× NEB T4 DNA ligase buffer (NEB), 100 µl 10% Triton X-100 (Sigma), 12 µl 10 mg/ml BSA, and 2400 U of T4 DNA ligase (NEB)) overnight at 16 °C with slow rotation.

For the 3C library preparation, DNA was sheared using a Covaris sonicator (duty cycle: 10%; intensity: 5; cycles per burst: 200; time: six cycles of 60 s each; set mode: frequency sweeping; temperature: 4–7 °C). After sonication, religated DNA was pulled down using 150 µl of 10 mg/ml Dynabeads Streptavidin T1 beads (Thermo Fisher) according to the supplier's recommendation. Sheared and pulled down DNA was treated using a 100 µl end-repair reaction (25 mM dNTPs, 50 U NEB PNK T4 Enzyme, 12 U NEB T4 DNA polymerase, 5 U NEB DNA pol I, Large (Klenow) Fragment, 10× NEB T4 DNA ligase buffer with 10 mM ATP) and incubated for 30 min at 37 °C.

Universal sequencing adaptor were added using the NEBnext Ultra DNA Library Kit (NEB) according to the supplier's recommendation. The PCR cycle number was adjusted to 4–12 based on the initial DNA concentration. The final libraries were purified using AMPure Beads (Beckman Coulter) and samples were sequenced with Ilumina Hi-Seq technology according to the standard protocols and 75 bp (shallow CHP-212 Hi-C, deep IMR-5/75) and 150 bp (shallow IMR-5/75 Hi-C) paired-end mode. Around 100 million reads were generated per IMR-5/75 replicate (deep IMR-5/75 Hi-C) and around 5–25 million reads per replicate were generated for shallow CHP-212 and shallow IMR-5/75 Hi-C.

FASTQ files were processed using the Juicer pipeline v1.5.6, CPU version[55], which was set up with BWA v0.7.17 (ref. [47]) to map short reads to reference genome hg19, from which haplotype sequences were removed and to which the sequence of Epstein–Stein–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 (ref. [55]). Knight–Ruiz normalization was used for Hi-C maps[38,56]. In cases with copy-number variation within the amplicon, we visually compared unnormalized, Knight–Ruiz-normalized and local iterative correction-normalized[57] maps to confirm the robustness of our conclusions across different normalization approaches (Supplementary Fig. 8). Virtual 4C signal for the *MYCN* locus was generated by the mean Knight–Ruiz-normalized Hi-C signal across three 5 kb bins (chr2: 16,075,000–16,090,000).

4C-seq. For 4C-seq libraries, a starting material of 5×10^6–1×10^7 cells were used. The fixation and lysis were performed as described in the "Hi-C" section. After the first digestion with DpnII (NEB), sticky ends were religated in a 50 ml falcon tube (700 µl 10 ligation buffer (Fermentas), 7 ml H_2O, 50 U T4 DNA ligase (Thermo); overnight at 16 °C) and DNA de-cross linked and cleaned as described in the "HiC" section. Subsequently, a second digestion (150 µl sample, 50 µl 10× Csp6I buffer (Thermo), 60 U Csp6I (Thermo) 295 µl H_2O; overnight at 37 °C) and another re-ligation was performed. For the *MYCN* promoter viewpoint, DNA was purified using a PCR clean up Kit (Qiagen) and 1.6 µg DNA was amplified by PCR (Primer 1 5′-GCAGAATCGCCTCCG-3′, Primer 2 5′-CCTGGCTCTGCTTCCTAG-3′). For the library reaction, primers were modified with TruSeq adapters (Illumina): Adapter1 5′-CTACACGACGCTCTTCCGATCT-3′ and Adapter2 5′-CAGAC GTGTGCTCTTCCGATCT-3′. The input of a single 4C PCR reaction was between 50 and 200 ng depending on the complexity. The reaction was performed in a 50 µl volume using the Expand Long Template System (Roche) and 29 reaction cycles. After the PCR all reactions were combined and the DNA purified with a PCR clean up Kit (Qiagen). All samples were sequenced with the HiSeq 4000 (Illumina) technology according to the standard protocols and with around 20 million single-end reads per sample.

Reads were pre-processed, filtered for artefacts, and mapped to the reference genome GRCh37 using BWA-MEM as described earlier[36]. After removing the viewpoint fragment as well as 1.5 kb up- and downstream of the viewpoint the raw read counts were normalized per million mapped reads (RPM) and a window of 10 fragments was chosen to smooth the profile.

Whole-genome sequencing. Cells were harvested and DNA was extracted using the NucleoSpin Tissue kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Libraries for whole-genome sequencing were prepared with the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England BioLabs, Inc., Ipswich, MA). Libraries were sequenced on a MGISEQ-2000 (NGP; MGI Tech Co. Ltd, Shenzhen, China), HiSeq X (IMR-5/75, Kelly; Illumina, Inc., San Diego, CA), and NovaSeq 6000 (CHP-212; Illumina, Inc., San Diego, CA) with 2×150 bp paired-end reads. Quality control, adapter trimming, alignment, duplicate removal as for ChIP-seq data. Copy-number variation was called (Control-FREEC[58] 11.4 with default parameters). Structural variants were called using SvABA[59] (1.1.1) in germline mode and discarding regions in a blacklist provided by SvABA (https://data. broadinstitute.org/snowman/svaba_exclusions.bed).

Nanopore sequencing. Cells were harvested and high molecular weight DNA was extracted using the MagAttract HMW DNA Kit (Qiagen N.V., Venlo, Netherlands). Size selection was performed to remove fragments <10 kilobases (kb) using the Circulomics SRE kit (Circulomics Inc., Baltimore, MD). DNA content was measured with a Qubit 3.0 Fluorometer (Thermo Fisher) and sample quality control was performed using a 4200 TapeStation System (Agilent Technologies, Inc., Santa Clara, CA). Libraries were prepared using the Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies Ltd, Oxford, UK) and sequenced on a R9.4.1 MinION flowcell (FLO-MIN106, Oxford Nanopore Technologies Ltd, Oxford, UK). Quality control was performed using NanoPlot 1.0.0 (ref. [60]). For the NGP cell line, DNA was extracted with the NucleoSpin Tissue kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) and libraries were prepared using the ONT Rapid Kit (SQK-RBK004, Oxford Nanopore Technologies Ltd, Oxford, UK). Guppy 2.3.7 (Oxford Nanopore Technologies Ltd, Oxford, UK) was used for basecalling with default parameters. For de novo assembly, Flye 2.4.2 (ref. [61]) was run in metagenomics assembly mode on the unfiltered FASTQ files with an estimated genome size of 1 Gb. Contigs were mapped back to hg19 using minimap2 2.16 (ref. [62]) with parameter -ax asm5. Assembly results were visualized with Bandage 0.8.1 (ref. [63]) and Ribbon 1.0 (ref. [64]). CpG methylation was called from the unfiltered raw FAST5 files using Megalodon 0.1.0 (Oxford Nanopore Technologies Ltd, Oxford, UK). Motif signatures were derived with Homer (4.9.1)[65] using the binomial test against nucleotide composition-matched background sequences. CpG methylation composite profiles were created by averaging signal in 50 bp bins using computeMatrix in deepTools (3.3.0)[48].

Fluorescence in situ hybridization. Cells were grown to 200,000 per well in six-well plates and metaphase-arrested using Colcemid (20 µl/2 ml; Roche #10295892001) for 30 min–3 h, trypsinized, centrifuged (200g/10 min), washed, and pelleted. Five milliliters of 0.4% KCl (4 °C; Roth #6781.1) was added to the pellet and incubated for 10 min. One milliliter KCl and 1 ml MeOH/acetic acid 3:1 (Roth #4627.2, #KK62.1) was added drop-wise. In all, 2/5/5 ml of MeOH/acetic acid was added in between centrifugation steps (200g/10 min), respectively. Suspension was dropped on a slide from a height of 40 cm. Slides were washed with PBS (Gibco, #70011036) and digested for 10 min in 0.04% pepsin solution in 0.001 N HCl. Slides were washed in 0.5× SSC, dehydrated with 70%/80%/100% EtOH (3 min each), and air-dried. Ten microliters of the probe (Vysis LSI N-MYC; #07J72-001; Lot #472123; Abbott Laboratories, Abbott Park, IL) were added and coverslips fixed on the slide. Slides were incubated at 75 °C for 10 min and at 37 °C overnight. The coverslip was removed and the slide was washed in 0.4× SSC/0.3% IGEPAL (CA-630, #18896; Sigma-Aldrich Inc.) for 3 min at 60 °C and 2× SSC/0.1% IGEPAL for 3 min at RT. Five microliters DAPI (Vectashield, #H-1200, Vector) was added. A coverslip was added and fixed with nail polish.

Enhancer calling. *MYCN*-expressing cell lines were defined as cell lines with size-Factor normalized expression of 100 or above based. We identified enhancer candidate regions in a ±500 kb window around *MYCN*. We focused on regions with a H3K27ac peak in the majority of *MYCN*-expressing, non-*MYCN*-amplified cell lines, i.e. three or more. If the gap between two such regions was less than 2 kb, they were joined. These regions were then ranked by the maximum difference in H3K27ac signal fold change between non-amplified, *MYCN*-expressing, and non-expressing cell lines. We chose the five highest-ranking regions as candidate regulatory elements. Enhancer regions were screened for transcription factor-binding sequences from the JASPAR2018 (ref. [50]) and JASPAR2020 (ref. [66]) database using the TFBSTools 1.20.0 (ref. [51]) function matchPWM with min.score = "85%". CRC-driven super-enhancers were defined as all regions with a LILY-defined super-enhancer in *MYCN*-expressing, non-*MYCN*-amplified cell lines that overlapped with a GATA3, HAND2, or PHOX2B peak in CLB-GA.

Analysis of neuroblastoma copy-number data. Public data were downloaded from https://github.com/padpuydt/copynumber_HR_NB/ (ref. [27]). Samples that were described as *MYCN*-amplified in the metadata but did not show *MYCN* amplification in the copy-number profile were excluded. In order to generate an aggregate copy-number profile, the genome was binned in 10 kb bins and number of samples with overlapping amplifications was counted per bin. Randomized copy-number profiles were generated by randomly sampling one of the original copy-number profiles on chromosome 2 and randomly shifting it such that *MYCN* is still fully included within

an amplified segment. For class I-specific shuffling, e4 had to be included as well; for class II-specific shuffling, e4 was never included on the randomly shifted amplicon. Empirical *P* values for significant co-amplification were derived by creating 10,000 randomized datasets with each amplicon randomly shifted and comparing the observed co-amplification frequency to the distribution of co-amplification frequencies in the randomized data. Empirical *P* values were always one-sided and adjusted for multiple comparisons using the Benjamini–Hochberg procedure.

Analysis of medulloblastoma copy number and ChIP-seq data. Medulloblastoma Affymetrix SNP6 data (10 cell lines, 1087 patient samples) were downloaded from Gene Expression Omnibus (GSE37385)[29] and processed using rawcopy 1.1 (ref. [67]) with default parameters. Segments with a log2 ratio ≥1.8 were classified as amplifications. The genome was binned in 10 kb bins and the number of samples with overlapping amplifications was counted per bin to generate composite copy-number plots.

Medulloblastoma H3K27ac ChIP-seq BigWig files and super-enhancer regions were downloaded from https://pecan.stjude.cloud/dataset/northcott (ref. [30]). The medulloblastoma subgroup-wise average H3K27ac signal was computed in 1 kb bins.

Amplicon reconstruction. All unfiltered SvABA structural variant calls were filtered to exclude regions from the ENCODE blacklist[68] and small rearrangements of 1 kb or less. As we were only aiming at the rearrangements common to all amplicons, we only considered breakpoints with more than 50 variant-support reads ("allele depth"). gGnome[69] was used to represent these data as a genome graph with nodes being breakpoint-free genomic intervals and edges being rearrangements ("alternate edge") or connections in the reference genomes ("reference edge"). We considered only nodes with high copy number, i.e. with a mean whole-genome sequencing coverage of at least 10-fold the median coverage of chromosome 2. Then, reference edges were removed if its corresponding alternate edge was among the 25% highest allele-depth edges. The resulting graph was then searched for the circular, *MYCN*-containing walk that included the highest number of nodes without using any node twice. We used gTrack (https://github.com/mskilab/gTrack) for visualization. For custom Hi-C maps of reconstructed amplicon sequences of CHP-212 and IMR-5-75, respectively, the corresponding regions from chromosome 2 were copied, ordered, oriented, and compiled according to the results from the amplicon reconstruction and added to the reference genome. Additionally, these copied regions were masked with "N" at the original locations on chromosome 2 to allow a proper mapping of reads to the amplicon sequence. The contribution of Hi-C di-tags from these regions on chromosome 2 to the amplicon Hi-C map is expected be minor, because the copy number of amplicons is much higher than the number of wild-type alleles.

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

Data availability

Sequencing data generated for this study are available at the Sequence Read Archive under accession PRJNA622577. Copy-number data for high-risk neuroblastoma were downloaded from https://github.com/padpuydt/copynumber_HR_NB/ (ref. [27]). Public data supporting the findings of this manuscript were downloaded from the Gene Expression Omnibus under accessions GSE90683, GSE80152, GSE24447, GSE37385, GSE18927, and GSE28874 and from ArrayExpress under accession E-MTAB-6570. Medulloblastoma ChIP-seq data were downloaded from https://pecan.stjude.cloud/dataset/northcott. BigWig und narrowPeak files can be downloaded from https://data.cyverse.org/dav-anon/iplant/home/konstantin/helmsaueretal/. An accompanying UCSC genome browser track hub is provided for ChIP-seq and ATAC-seq data visualization (https://de.cyverse.org/d/d/27AA17DA-F24C-4BF4-904C-62B539A47DCC/hub.txt). All other data are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

Code availability

Code is available at https://github.com/hennssenlab/MYCNAmplicon.

References

1. Turner, K. M. et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).
2. Zhang, C. Z. et al. Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
3. Ly, P. et al. Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. *Nat. Genet.* **51**, 705–715 (2019).
4. Deshpande, V. et al. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat. Commun.* **10**, 392 (2019).
5. Nathanson, D. A. et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science* **343**, 72–76 (2014).
6. Xu, K. et al. Structure and evolution of double minutes in diagnosis and relapse brain tumors. *Acta Neuropathol.* **137**, 123–137 (2019).
7. deCarvalho, A. C. et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat. Genet.* **50**, 708–717 (2018).
8. 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).
9. Wahl, G. M. The importance of circular DNA in mammalian gene amplification. *Cancer Res.* **49**, 1333–1340 (1989).
10. Koche, R. P. et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat. Genet.* **52**, 29–34 (2020).
11. Gröbner, S. N. et al. The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).
12. Cohn, S. L. et al. The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report. *J. Clin. Oncol.* **27**, 289–297 (2009).
13. Weiss, W. A., Aldape, K., Mohapatra, G., Feuerstein, B. G. & Bishop, J. M. Targeted expression of MYCN causes neuroblastoma in transgenic mice. *EMBO J.* **16**, 2985–2995 (1997).
14. Althoff, K. et al. A Cre-conditional MYCN-driven neuroblastoma mouse model as an improved tool for preclinical studies. *Oncogene* **34**, 3357–3368 (2015).
15. Boeva, V. et al. Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* **49**, 1408–1413 (2017).
16. Durbin, A. D. et al. Selective gene dependencies in MYCN-amplified neuroblastoma include the core transcriptional regulatory circuitry. *Nat. Genet.* **50**, 1240–1246 (2018).
17. Decaesteker, B. et al. TBX2 is a neuroblastoma core regulatory circuitry component enhancing MYCN/FOXM1 reactivation of DREAM targets. *Nat. Commun.* **9**, 4866 (2018).
18. Wang, L. et al. ASCL1 is a MYCN- and LMO1-dependent member of the adrenergic neuroblastoma core regulatory circuitry. *Nat. Commun.* **10**, 5622 (2019).
19. Morton, A. R. et al. Functional enhancers shape extrachromosomal oncogene amplifications. *Cell* **179**, 1330–1341 (2019).
20. Northcott, P. A. et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).
21. Peifer, M. et al. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526**, 700–704 (2015).
22. Valentijn, L. J. et al. TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. *Nat. Genet.* **47**, 1411–1414 (2015).
23. Weischenfeldt, J. et al. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat. Genet.* **49**, 65–74 (2017).
24. Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
25. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
26. Rajbhandari, P. et al. Cross-cohort analysis identifies a TEAD4-MYCN positive feedback loop as the core regulatory element of high-risk neuroblastoma. *Cancer Discov.* **8**, 582–599 (2018).
27. Depuydt, P. et al. Meta-mining of copy number profiles of high-risk neuroblastoma tumors. *Sci. Data* **5**, 180240, https://doi.org/10.1038/sdata.2018.240 (2018).
28. 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).
29. Northcott, P. A. et al. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488**, 49–56 (2012).
30. Lin, C. Y. et al. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* **530**, 57–62 (2016).
31. Depuydt, P. et al. Genomic amplifications and distal 6q loss: novel markers for poor survival in high-risk neuroblastoma patients. *J. Natl Cancer Inst.* **110**, 1084–1093 (2018).
32. Robson, M. I., Ringel, A. R. & Mundlos, S. Regulatory landscaping: how enhancer-promoter communication is sculpted in 3D. *Mol. Cell* **74**, 1110–1122 (2019).
33. Simpson, J. T. et al. Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14**, 407–410 (2017).
34. Wu, S. et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699–703 (2019).
35. Solovei, I. et al. Topology of double minutes (dmins) and homogeneously staining regions (HSRs) in nuclei of human neuroblastoma cell lines. *Genes Chromosomes Cancer* **29**, 297–308 (2000).
36. Franke, M. et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).

37. Despang, A. et al. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* **51**, 1263–1271 (2019).

38. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

39. Guo, Y. et al. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).

40. Hogarty, M. D. et al. ODC1 is a critical determinant of MYCN oncogenesis and a therapeutic target in neuroblastoma. *Cancer Res.* **68**, 9735–9745 (2008).

41. Gamble, L. D. et al. Inhibition of polyamine synthesis and uptake reduces tumor progression and prolongs survival in mouse models of neuroblastoma. *Sci. Transl. Med.* **11**, eaau1099 (2019).

42. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

43. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

44. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

45. Lee, T. I., Johnstone, S. E. & Young, R. A. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protoc.* **1**, 729–748 (2006).

46. Rada-Iglesias, A. et al. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11**, 633–648 (2012).

47. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

48. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

49. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).

50. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).

51. Tan, G. & Lenhard, B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32**, 1555–1556 (2016).

52. 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).

53. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).

54. Zeid, R. et al. Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma. *Nat. Genet.* **50**, 515–523 (2018).

55. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).

56. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA J. Numer Anal.* **33**, 1928–1047 (2013).

57. Servant, N., Varoquaux, N., Heard, E., Barillot, E. & Vert, J. P. Effective normalization for copy number variation in Hi-C data. *BMC Bioinformatics* **19**, 313 (2018).

58. Boeva, V. et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).

59. Wala, J. A. et al. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res.* **28**, 581–591 (2018).

60. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).

61. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).

62. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

63. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350–3352 (2015).

64. Nattestad, M., Aboukhalil, R., Chin, C. S. & Schatz, M. C. Ribbon: intuitive visualization for complex genomic variation. *Bioinformatics* btaa680, https://doi.org/10.1093/bioinformatics/btaa680 (2020).

65. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

66. Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).

67. Mayrhofer, M., Viklund, B. & Isaksson, A. Rawcopy: Improved copy number analysis with Affymetrix arrays. *Sci. Rep.* **6**, 36158 (2016).

68. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).

69. Hadi, K. et al. Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. *Cell* **183**, 197–210 (2020).

Author contributions

All authors contributed to the study design and collection and interpretation of the data. M.V. and S.A. acquired ChIP-seq, ATAC-seq, 4C, and Hi-C data. R.C.G., L.P.K., and P.E. acquired nanopore sequencing data. K.K. acquired Illumina whole-genome sequencing data. C. Röefzaad and C. Rosswog performed FISH experiments. R.S., V.H., and K. Helmsauer analyzed 4C and Hi-C data. K. Helmsauer analyzed ChIP-seq, ATAC-seq, and RNA-seq data. E.R-F., M.P.M., J.T., and K. Helmsauer analyzed Illumina whole-genome sequencing data. K. Helmsauer and R.P.K. analyzed nanopore sequencing data. M.F., F.H., A.S., and J.H.S. collected and prepared patient samples. M.V., S.A., R.C.G., Y.B., H.D.G., C.K., C.Y.C., and K. Haase performed experiments and analyzed data. K. Haase, M.F., F.H., A.S., M.R., D.T., A.E., and J.H.S. contributed to study design. K. Helmsauer, M.V., S.A., S.M., A.G.H., and R.P.K. led the study design, performed data analysis and wrote the manuscript, to which all authors contributed.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-19452-y.

Correspondence and requests for materials should be addressed to A.G.H. or R.P.K.

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

Reprints and permission information is available at http://www.nature.com/reprints

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

Supplementary Information for Helmsauer et al. *"Enhancer hijacking determines extrachromosomal MYCN amplicon architecture in neuroblastoma"*
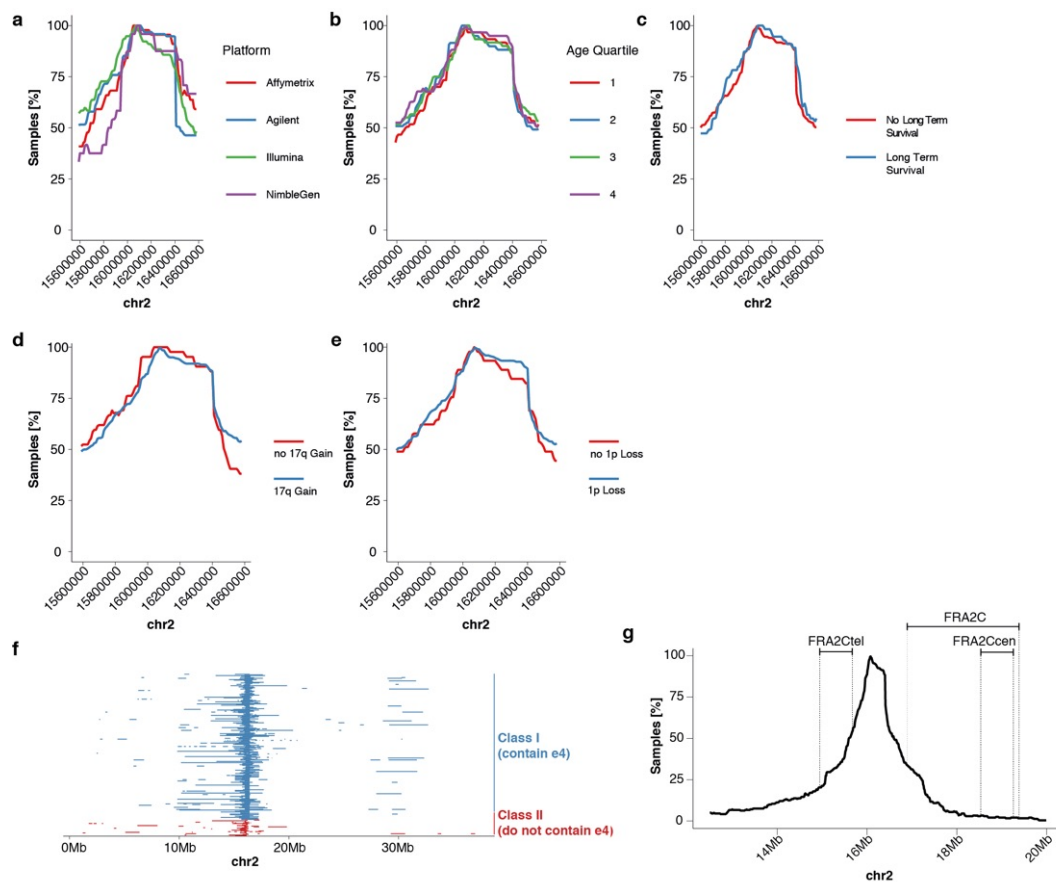
Supplementary Figures

- Supplementary Fig. 1. Enhancer profiling of the *MYCN* locus
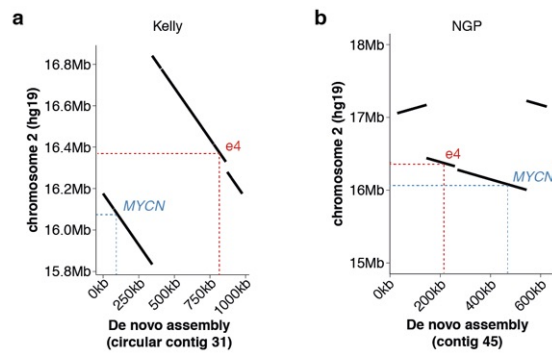- Supplementary Fig. 2. Amplification patterns across clinical and experimental covariates
- Supplementary Fig. 3. Super enhancer co-amplification confers different *MYCN* amplicon structures of medulloblastoma subgroups
- Supplementary Fig. 4. H3K27 acetylation on the *MYCN* amplicon
- Supplementary Fig. 5. Long-read sequencing enables de novo assembly of *MYCN* neighborhoods
- Supplementary Fig. 6. De novo assembly confirms co-amplification of *MYCN* and e4 in Kelly and NGP
- Supplementary Fig. 7. *MYCN* expression on two amplicon classes
- Supplementary Fig. 8. A tandem duplication on the *MYCN* amplicon in IMR-5/75
- Supplementary Fig. 9. Enhancer hijacking and neo-TAD formation on the *MYCN* amplicon
- Supplementary Fig. 10. *MYCN* fluorescence in situ hybridization in neuroblastoma cell lines
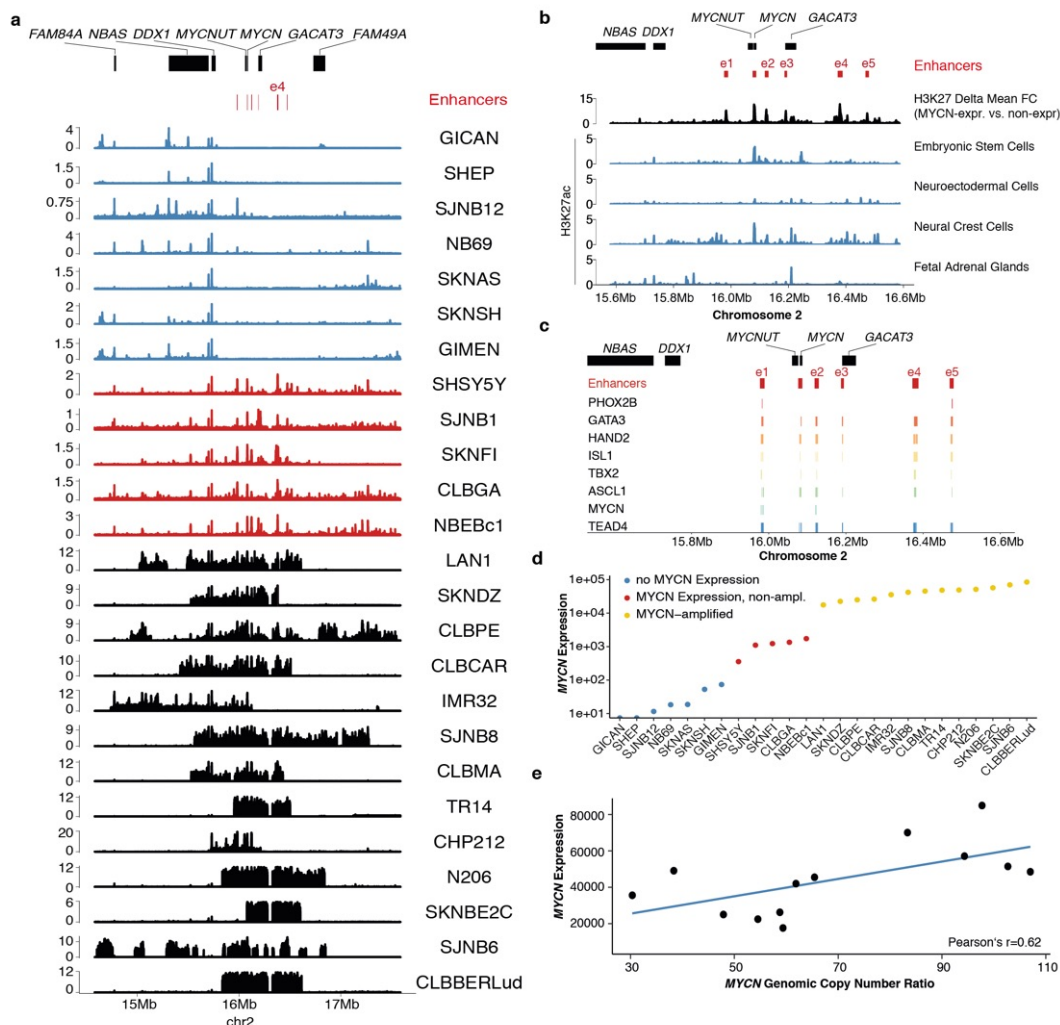
Supplementary Fig. 1. Enhancer profiling of the *MYCN* locus. a H3K27ac ChIP-seq signal (counts per million in 10bp bins, smoothed in 1kb bins) for seven non-*MYCN*-expressing neuroblastoma cell lines (blue), five *MYCN*-expressing non-*MYCN*-amplified neuroblastoma cell lines (red) and 13 MYCN-amplified neuroblastoma cell lines (black). **b** Differential composite H3K27ac signal for *MYCN*-non-expressing vs. *MYCN*-expressing non-*MYCN*-amplified cells (difference in the group-wise mean fold change H3K27ac vs. input; black) and H3K27ac ChIP-seq (counts per million in 10bp bins, smoothed in 1kb bins) signal for *in vitro* differentiated developmental cell types (embryonic stem cells, neuroectodermal cells, neural crest cells; blue) and a fetal adrenal cell sample (blue). **c** Core regulatory circuit factor (PHOX2B, GATA3, HAND2, ISL1, TBX2, ASCL1), MYCN and TEAD4 binding motif positions in the *MYCN*-driving enhancers e1-e5. Only motif hits within the enhancer regions are depicted. **d** *MYCN* expression for *N*=25 neuroblastoma cell lines as determined by RNA-seq (one sequencing experiment per cell line; size factor-normalized read counts) classified into no *MYCN* expression (size factor normalized expression lower than 100), *MYCN*-expressing non-*MYCN*-amplified cells (size factor normalized expression 100 or more) and *MYCN*-amplified cell lines. **e** *MYCN* expression by *MYCN* genomic copy number ratio determined from ChIP-seq input data for *N*=13 *MYCN*-amplified neuroblastoma cell lines (one sequencing experiment per cell line). Source data are provided as a Source Data file.

73

Supplementary Fig. 2. Amplification patterns across clinical and experimental covariates. **a-e** Percent co-amplification in 10kb bins for *MYCN*-amplified neuroblastoma (n=240) split by the experimental method to measure genomic copy-number (**a**, Affymetrix SNP array, Agilent aCGH platform, Illumina SNP array, NimbleGen aCGH platform), the age quartile of patients (**b**, 1=lowest quartile, 4=highest quartile), long-term survival (**c,** defined as survival beyond five years post diagnosis) and the genetic factors 17q gain (**d**) and 1p loss (**e**). **f** Amplified regions on chromosome 2 (0Mb-40Mb) for primary neuroblastoma (n=240), colored by amplicon class (class I amplicons including e4 vs. class II amplicons not including e4) **g** Percent co-amplification in 10kb bins around *MYCN* for *MYCN*-amplified neuroblastoma (n=240) and position of common fragile sites on chromosome 2 between 12Mb and 20Mb. Source data are provided as a Source Data file.
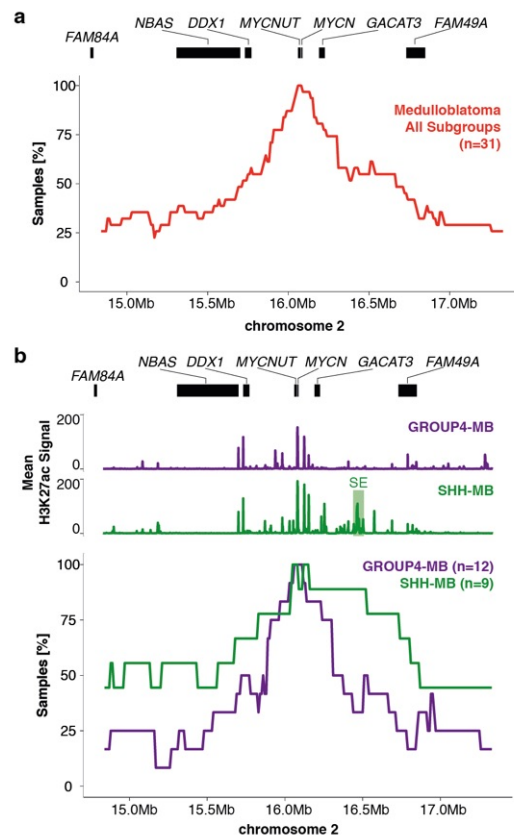
Supplementary Fig. 3. Super enhancer co-amplification confers different *MYCN* amplicon structures of medulloblastoma subgroups. a Percent co-amplification in 10kb bins for *MYCN*-amplified medulloblastoma across all medulloblastoma subgroups (n=31). **b** Mean H3K27ac ChIP-seq signal for GROUP4-MB (n=11) and SHH-MB (n=5) and percent co-amplification in 10kb bins for *MYCN*-amplified medulloblastoma for GROUP4-MB (n=12) and SHH-MB (n=9). The SHH-MB-specific super enhancer is marked by SE. Only medulloblastoma subgroups with more than five *MYCN* amplicons were considered for aggregate copy number profiles. Source data are provided as a Source Data file.
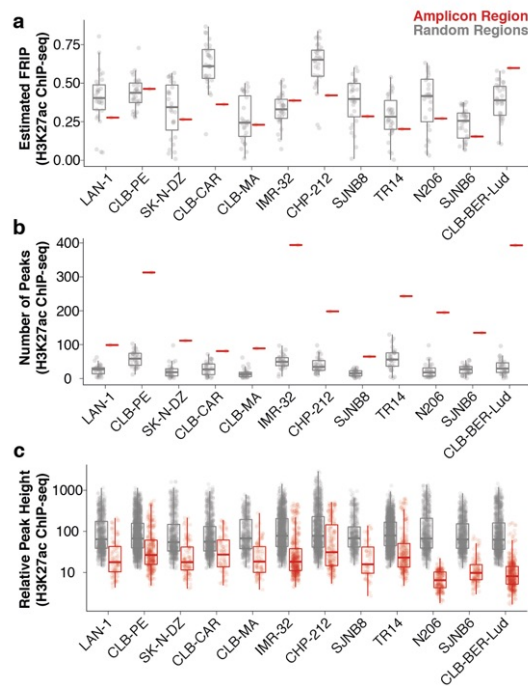
Supplementary Fig. 4. H3K27 acetylation on the *MYCN* amplicon. a Estimated fraction of H3K27ac reads in peaks for amplified regions vs. randomly drawn genomic regions of matching size (n=30) in 12 *MYCN*-amplified neuroblastoma cell lines. **b** Number of peaks for amplified regions (red) vs. randomly drawn genomic regions of matching size (grey, n=30) in 12 *MYCN*-amplified neuroblastoma cell lines. **c** Relative H3K27ac peak heights (compared to amplicon background) for amplified regions (red) vs. randomly drawn genomic regions of matching size (grey, n=30). In all boxplots, boxes depict the median, the upper quartile boundary and the lower quartile boundary. Whiskers extends to the largest data point within 1.5-fold of the inter-quartile range. Source data are provided as a Source Data file.
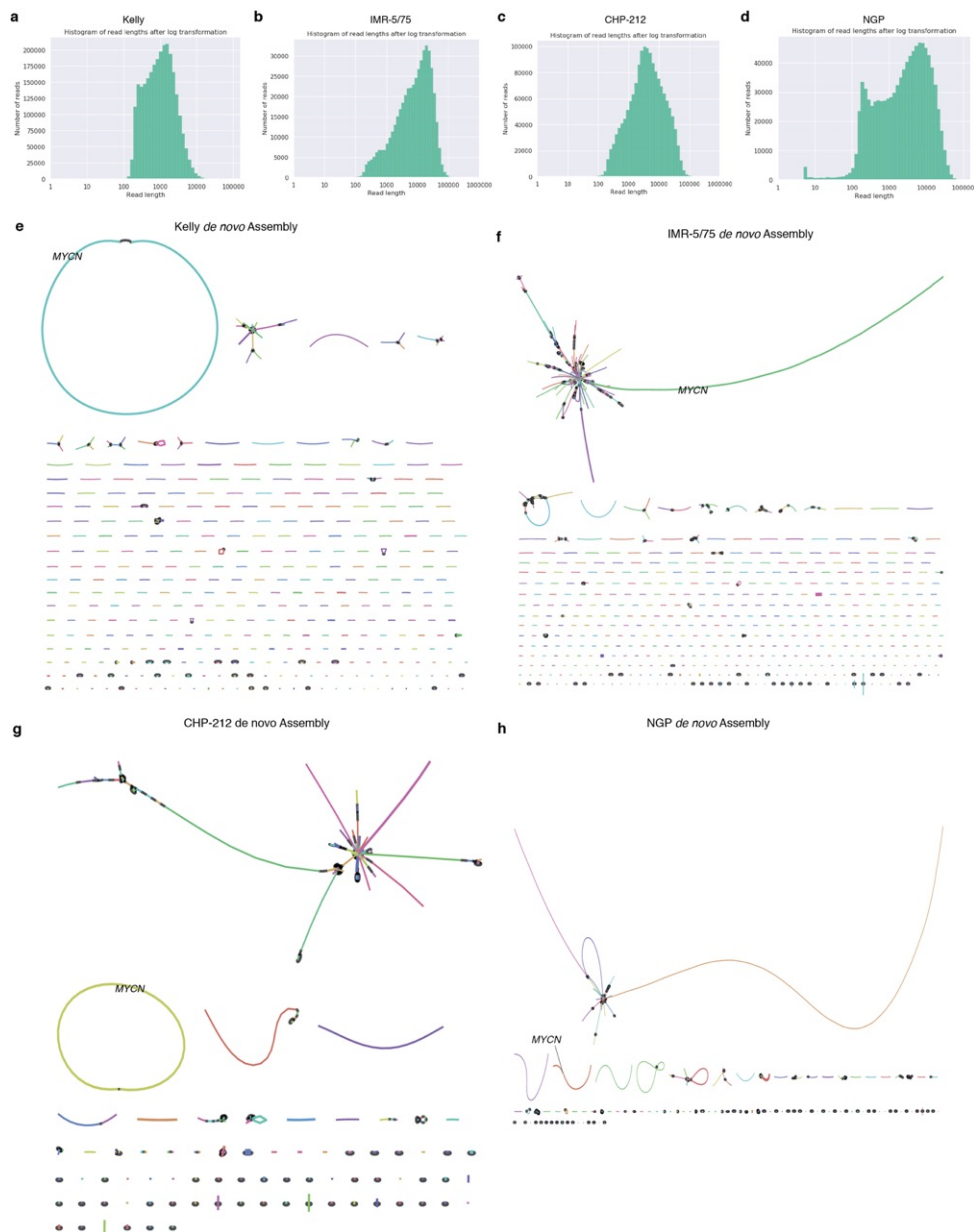
Supplementary Fig. 5. Long-read sequencing enables de novo assembly of *MYCN* neighborhoods. a-d Nanopore read length distribution (log-transformed) for the neuroblastoma cell lines Kelly (**a**, n=2,654,406), IMR-5/75 (**b**, n=474,980), CHP-212 (**c**, n=1,554,048) and NGP (**d**, n=952,031). **e** Nanopore long read-based de novo assembly of Kelly cells yields 477 contigs and an overall assembly N50 of 24,929 bp. BLAST analysis locates *MYCN* on a circular 975,932 bp contig. **f** Nanopore long read-based de novo assembly of IMR-5/75 cells yields 6,265 contigs and an overall assembly N50 of 91,273 bp. BLAST analysis locates *MYCN* on a linear 3,201,197 bp contig. **g** Nanopore long read-based de novo assembly of CHP-212 cells yields 21,264 contigs and an overall assembly N50 of 113,845 bp. BLAST analysis locates *MYCN* on a circular 1,705,218 bp contig. **h** Nanopore long read-based de novo assembly of NGP cells yields 6,550 contigs and an overall assembly N50 of 60,981 bp. BLAST analysis locates *MYCN* on a linear 623,907 bp contig.

Supplementary Fig. 6. De novo assembly confirms co-amplification of *MYCN* and e4 in Kelly and NGP. Mapping of the de novo assembled *MYCN*-containing contig to hg19 in Kelly (**a**) and NGP (**b**) cells. Positions of *MYCN* and e4 are marked on the contig and in the reference genome. Note that the Kelly contig is circular such that the shortest distance from e4 to *MYCN* spans the contig circle junction.
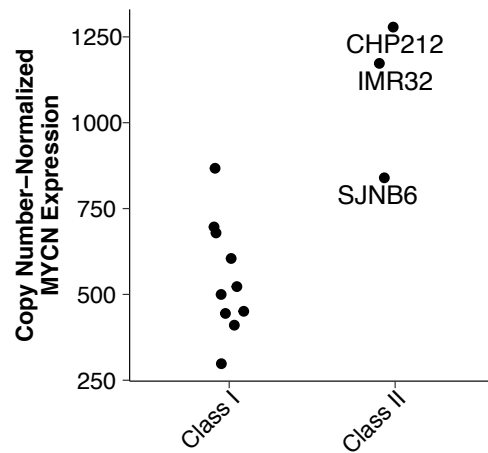
Supplementary Fig. 7. *MYCN* expression on two amplicon classes. *MYCN* copy number was estimated from ChIP-seq input signal for *N*=13 *MYCN*-amplified neuroblastoma cell lines, three of which were classified as class II (CHP-212, IMR-32, SJNB6). Size factor-normalized RNA-seq read counts from one sequencing experiment per cell line were divided by copy number ratios to estimate *MYCN* expression per genomic copy. This revealed relatively high expression per copy for class II amplicons although low sample size precludes further interpretation. Source Data are provided as a Source Data file.
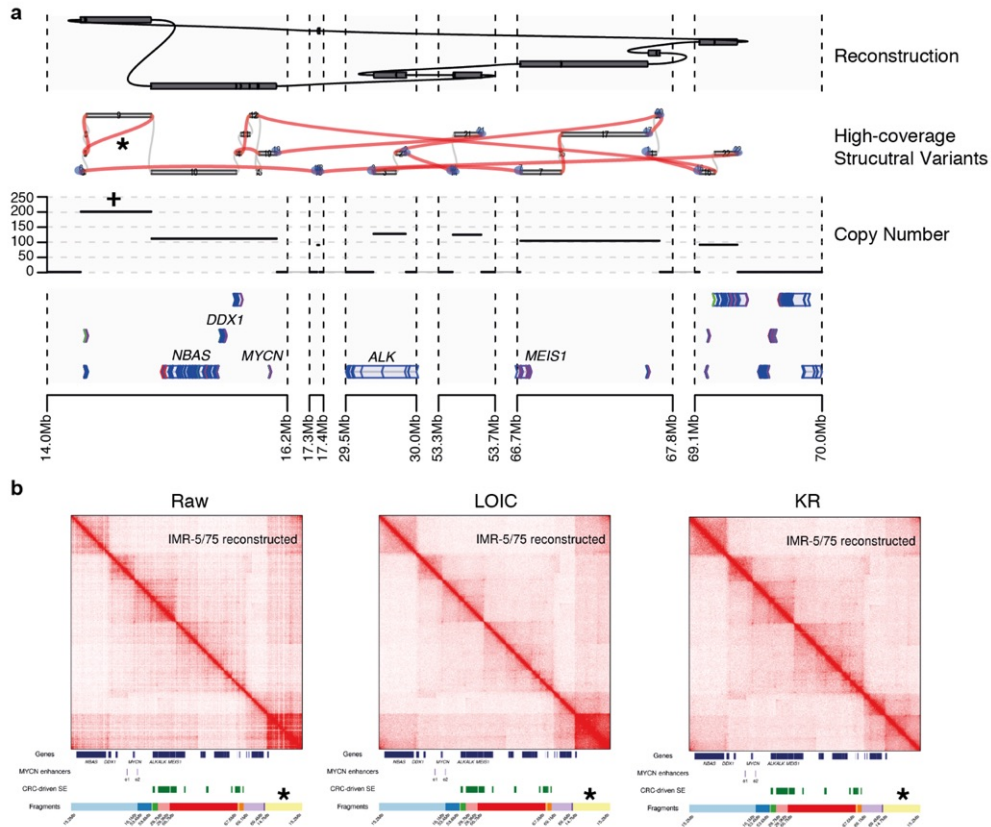
Supplementary Fig. 8. A tandem duplication on the *MYCN* amplicon in IMR-5/75. a Concordant copy-number (marked by a plus sign) and structural variation (marked by an asterisk) data indicates the presence of a tandem duplication on the *MYCN* amplicon in IMR-5/75. **b** Comparison of unnormalized, local iterative correction-normalized (LOIC; taking copy-number variation within the amplicon explicitly into account), as well as Knight-Ruiz-normalized (KR) Hi-C maps for the reconstructed amplicon shows that the finding of a neo-TAD is robust across different normalization approaches. The tandem duplication is marked by an asterisk.
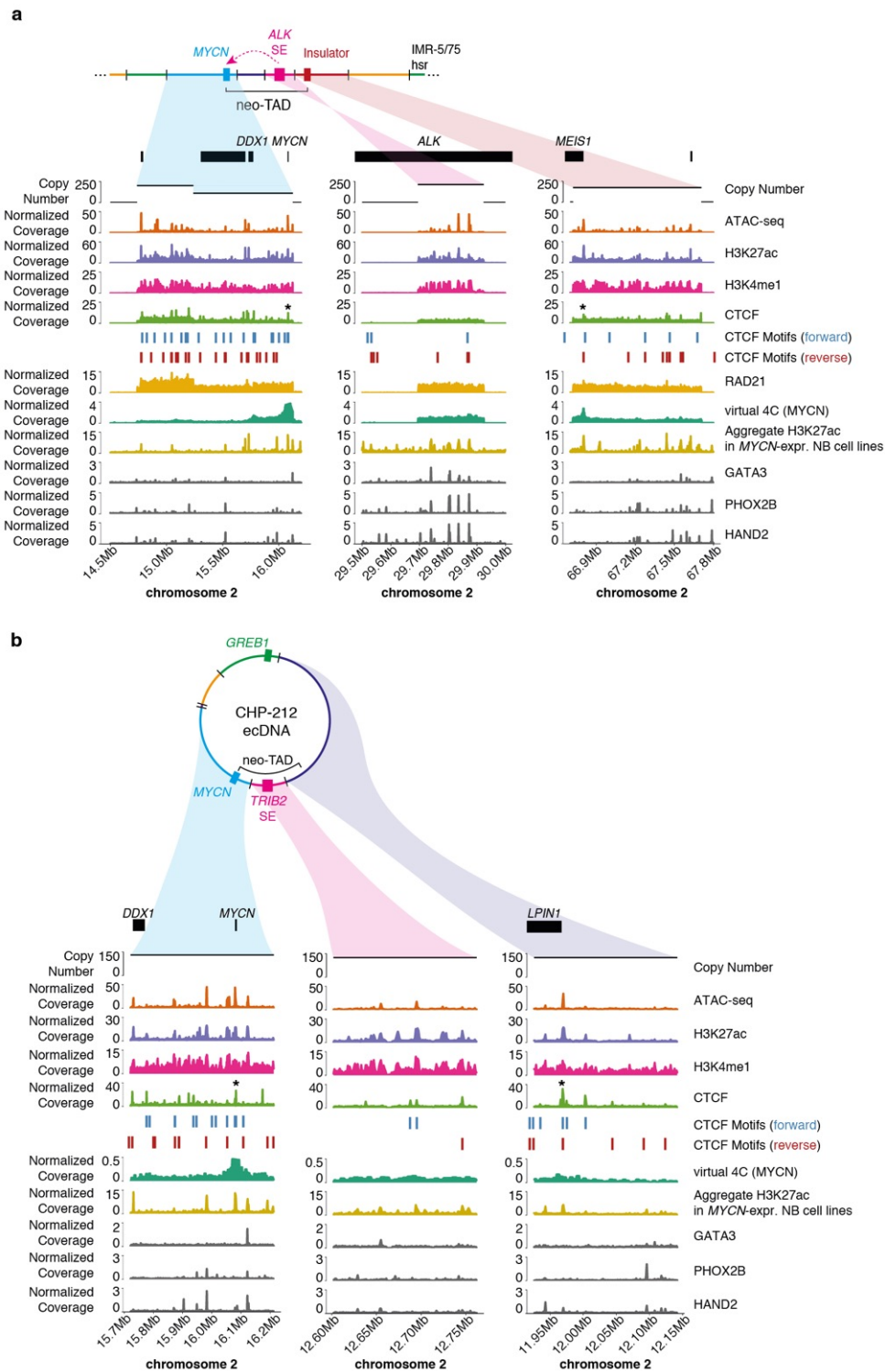
Supplementary Fig. 9. Enhancer hijacking and neo-TAD formation on the *MYCN* amplicon. Regions that contribute to neo-TAD formation are depicted for IMR-5/75 (**a**) and CHP-212 (**b**). Top to bottom: Gene bodies, Copy Number (black), ATAC-seq (orange), H3K27ac ChIP-seq (purple), H3K4me1 ChIP-seq (pink), CTCF ChIP-seq (light green), CTCF-

peak overlapping CTCF forward (blue) vs. reverse (red) binding motifs, RAD21 ChIP-seq (yellow; only available for IMR-5/75), virtual 4C with *MYCN* as the viewpoint (mean Knight-Ruiz normalized interaction frequency of three 5kb bins (chr2:16,075,000-16,085,000) around *MYCN*; dark green), the aggregate H3K27ac signal over 7 *MYCN*-expressing non-*MYCN*-amplified neuroblastoma cell lines (mean fold change over input; light brown), GATA3 ChIP-seq, PHOX2B ChIP-seq and HAND2 ChIP-seq (all CRC transcription factors in the neuroblastoma cell lines CLB-GA; grey). ChIP-seq and ATAC-seq is depicted as counts per million in 10bp bins, smoothed in 1kb bins. Neo-TAD boundaries are marked by an asterisk. Source data are provided as a Source Data file.
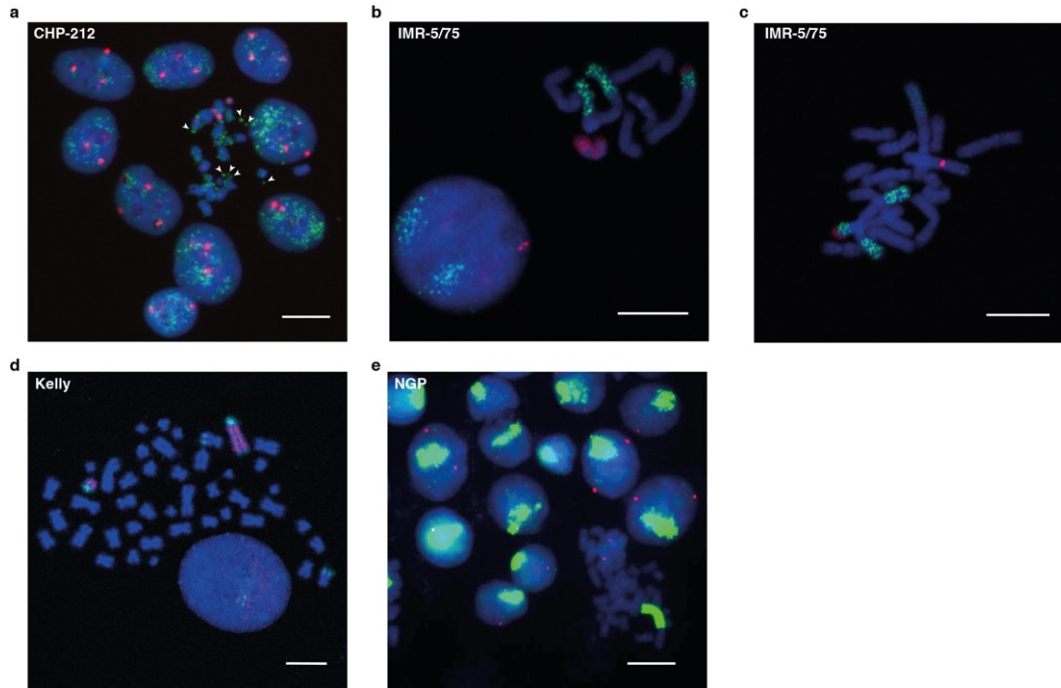
Supplementary Fig. 10. *MYCN* **fluorescence in situ hybridization in neuroblastoma cell lines. a** Fluorescence in situ hybridization (FISH) of CHP-212 metaphase spreads with a *MYCN* probe (green) and a probe for the chromosome 2 centromere (red). Arrowheads point to *MYCN* ecDNA. **b, c** FISH of metaphase spreads in IMR-5/75 with a *MYCN* probe (green), a chromosome 2 centromere (red) and a chromosome 12 paint (red) **d** FISH of metaphase spreads in Kelly cells with a *MYCN* probe (red) and chromosome 17 paint (green). **e** FISH of NGP metaphase spreads with a *MYCN* probe (green) and a probe for the chromosome 2 centromere (red). Scale bars are 10µm. FISH experiments were performed once per cell line.

7.2 Koche et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma

The following section contains a reproduction of the article Koche RP, Rodriguez-Fos E, Helmsauer K, et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. Nat Genet. 2020;52(1):29-34. https://doi.org/10.1038/s41588-019-0547-z. Reproduced with permission from Springer Nature.

Journal Data Filtered By: **Selected JCR Year: 2018** Selected Editions: SCIE,SSCI
Selected Categories: **"GENETICS and HEREDITY"** Selected Category
Scheme: WoS
Gesamtanzahl: 173 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NATURE REVIEWS GENETICS	36,697	43.704	0.079260
2	NATURE GENETICS	93,920	25.455	0.221380
3	TRENDS IN ECOLOGY & EVOLUTION	36,697	15.236	0.035920
4	MOLECULAR BIOLOGY AND EVOLUTION	46,915	14.797	0.092200
5	GENOME BIOLOGY	38,920	14.028	0.133310
6	Genome Medicine	5,264	10.886	0.023220
7	TRENDS IN GENETICS	11,959	10.627	0.021110
8	GENOME RESEARCH	39,240	9.944	0.079580
9	AMERICAN JOURNAL OF HUMAN GENETICS	36,007	9.924	0.063650
10	Annual Review of Genetics	7,478	9.184	0.011380
11	GENES & DEVELOPMENT	54,563	8.990	0.072340
12	GENETICS IN MEDICINE	10,855	8.683	0.035800
13	MOLECULAR THERAPY	16,991	8.402	0.030050
14	Annual Review of Genomics and Human Genetics	2,537	7.914	0.006020
15	ONCOGENE	63,249	6.634	0.074600
16	GENOMICS PROTEOMICS & BIOINFORMATICS	1,612	6.597	0.004790
17	MUTATION RESEARCH-REVIEWS IN MUTATION RESEARCH	3,566	6.081	0.004090
18	JOURNAL OF MEDICAL GENETICS	12,417	5.899	0.016850
19	Molecular Autism	2,107	5.712	0.008000

Selected JCR Year: 2018; Selected Categories: "GENETICS and HEREDITY"

There are amendments to this paper

Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma

Richard P. Koche [1,12]*, Elias Rodriguez-Fos [2,12], Konstantin Helmsauer [3,12], Martin Burkert [4,5,6], Ian C. MacArthur[3], Jesper Maag[1], Rocio Chamorro[3], Natalia Munoz-Perez[3], Montserrat Puiggròs[2], Heathcliff Dorado Garcia [3], Yi Bei [3], Claudia Röefzaad[3], Victor Bardinet[3], Annabell Szymansky[3], Annika Winkler[3], Theresa Thole[3], Natalie Timme[3], Katharina Kasack[7], Steffen Fuchs [3,5,7], Filippos Klironomos [3], Nina Thiessen[5], Eric Blanc[5], Karin Schmelz[3], Annette Künkele[3,5,7], Patrick Hundsdörfer[3,5,7], Carolina Rosswog[7], Jessica Theissen[7], Dieter Beule [5], Hedwig Deubzer[3,7,8], Sascha Sauer[5], Joern Toedling[3], Matthias Fischer [9,10], Falk Hertwig[3,7], Roland F. Schwarz [6,7], Angelika Eggert[3,5,7], David Torrents[2,11,12], Johannes H. Schulte[3,5,7,12] and Anton G. Henssen [3,5,7,8,12]*

Extrachromosomal circularization of DNA is an important genomic feature in cancer. However, the structure, composition and genome-wide frequency of extrachromosomal circular DNA have not yet been profiled extensively. Here, we combine genomic and transcriptomic approaches to describe the landscape of extrachromosomal circular DNA in neuroblastoma, a tumor arising in childhood from primitive cells of the sympathetic nervous system. Our analysis identifies and characterizes a wide catalog of somatically acquired and undescribed extrachromosomal circular DNAs. Moreover, we find that extrachromosomal circular DNAs are an unanticipated major source of somatic rearrangements, contributing to oncogenic remodeling through chimeric circularization and reintegration of circular DNA into the linear genome. Cancer-causing lesions can emerge out of circle-derived rearrangements and are associated with adverse clinical outcome. It is highly probable that circle-derived rearrangements represent an ongoing mutagenic process. Thus, extrachromosomal circular DNAs represent a multihit mutagenic process, with important functional and clinical implications for the origins of genomic remodeling in cancer.

Recent studies have shown that circular DNA is more prevalent in human tissues than previously anticipated[1–5]. Based on size and copy number, at least three classes of circular DNA exist in human cells: (1) small extrachromosomal circular DNA (including microDNA; referred to as eccDNA throughout the text)[3,6]; (2) large, copy number–amplified extrachromosomal circular DNA (ecDNA)[1], and (3) ring and/or neochromosomes[7,8]. ecDNA can lead to oncogene amplification and is a powerful driver of intratumoral heterogeneity[1,9–12]. Whether ecDNA has other cancer-causing functions is unknown, and the impact circularization has on genome remodeling is unclear.

Neuroblastoma is one of the first tumor entities where extrachromosomal oncogene circularization in the form of *MYCN* proto-oncogene double-minute chromosomes was detected[10,13]. Since the first descriptions in 1965 (refs. [14,15]), the extent of DNA circularization has not been accurately quantified in neuroblastoma. We hypothesized that DNA circularization could represent a genome-wide, driving mutagenic process in neuroblastoma with functional consequences beyond oncogene amplification. We set out to systematically describe the spectrum and impact of circular DNA in neuroblastoma by using different genomic and transcriptomic approaches (Supplementary Fig. 1).

Since DNA circularity can be computationally inferred from whole-genome sequencing (WGS) data[3,16,17], we applied an algorithm using paired-end read orientation to detect circularity to WGS from 93 neuroblastomas paired with normal blood specimens (Fig. 1a,b). This approach detected a large tumor-specific circular DNA catalog, including *MYCN* double-minute chromosomes, mitochondrial DNA and many previously undescribed ecDNAs and eccDNAs (Fig. 1c,d and Supplementary Fig. 2a,b). This suggests a greater prevalence and complexity of circular DNA in neuroblastoma than previously anticipated.

To achieve complementary and more sensitive detection and characterization of circular DNA in neuroblastoma, we adapted and modified the Circle sequencing (Circle-seq) method (Supplementary Figs. 1 and 2c,d)[6]. We achieved specific DNA circle enrichment through >10[10]-fold depletion of linear genomic DNA (gDNA; Fig. 1c and Supplementary Figs. 2c and 3a–c). Applying Circle-seq to endonuclease-treated gDNA significantly reduced read mapping to circularized genomic regions by 474-fold ($P = 7.566 \times 10^{-11}$, Welch's *t*-test; Fig. 1c and Supplementary Fig. 3d,e), confirming specific enrichment of circular DNA.

Fig. 1 | A genome-wide map of circular DNA in neuroblastoma. a, Schematic representation of sequencing reads as predicted for circular genomic regions. Background indicated noncircular genome. **b**, Schematic representation of sequencing read positions on circular DNA. **c**, Genome tracks comparing sequencing read densities on an ecDNA as detected via WGS (only circle-specific head-to-tail reads are depicted), Circle-seq followed by Illumina paired-end sequencing (ILM) and SMRT-seq in neuroblastoma cells. DNA digestion with an exonuclease and/or endonuclease is indicated (+/−). The dashed blue line indicates the predicted circle junction. Interruption of read density profile is due to lack of read alignment (y axis: 0–30 reads). **d**, Chromosome ideogram with genome-wide somatic circular DNA density as inferred from WGS (red) compared to Circle-seq (blue). M, circular mitochondrial DNA. **e**, Number of ecDNAs and eccDNAs per neuroblastoma ($n = 21$ tumors, $n = 96,436$ eccDNAs, $n = 14$ ecDNAs). **f**, Size distribution of ecDNAs and eccDNAs identified using Circle-seq in neuroblastomas ($n = 21$ tumors, $n = 96,436$ eccDNAs, $n = 14$ ecDNAs). **g**, Alternative B-allele frequencies (BAF) in the sequencing reads from Circle-seq ($n = 21$ tumors) and WGS ($n = 93$ tumors). **h**, Density of circular DNA detected using Circle-seq over genic compared to gene-surrounding regions in *MYCN*-amplified and nonamplified neuroblastomas ($n = 7$ *MYCN*-amplified tumors, $n = 14$ nonamplified tumors). The lines represent the mean signal and the shaded area represents the s.e.m. TES, transcription end site; TSS, transcription start site. **i**, Fraction of genomic regions affected by eccDNA compared to ecDNA ($n = 21$ tumors).

Sequence composition was analyzed and genomic origin inferred combining massive parallel paired-end sequencing with long-read Nanopore and single-molecule real-time sequencing (SMRT-seq). Circular head-to-tail junctions predicted computationally were confirmed by PCR and Sanger sequencing (Supplementary Fig. 3a–c). De novo sequence assembly of long reads spanning the entirety of circles allowed further physical confirmation of their circular structure in 65% of cases (Supplementary Fig. 4a–c). Circle-seq confirmed 100% of ecDNAs and 30% of eccDNAs predicted from WGS and identified on average 0.82 ecDNAs and 5,673 eccDNAs per neuroblastoma (Fig. 1c–e and Supplementary Fig. 4d–f). Although ecDNA was accurately predicted from WGS with high sensitivity (100%), our results highlight the advantages of using additional and more sensitive approaches, such as Circle-seq, to obtain a comprehensive characterization of circular DNAs in tumors.

The structure of circularized genomic loci in neuroblastoma varied considerably, with mean sizes of 680,200 base pairs (bp; ecDNA) and 2,403 bp (eccDNA) in tumors, reproducing the oscillating length distribution observed in lymphoma cancer cell lines[3] (Fig. 1f

and Supplementary Fig. 4g–j). In agreement with cytogenetic reports[18], no ring chromosomes were detected in neuroblastoma. Notably, both ecDNAs and eccDNAs were of monoallelic origin, as determined by haplotype phasing (Fig. 1g). Inspection of circle junction sequences (ecDNA and eccDNA) indicated the probable mechanism(s) of generation, since 2.8% contained nontemplate insertions indicative of nonhomologous end joining repair or replication-associated mechanisms (Supplementary Fig. 4k). In line with reports in human lymphoma cell lines[19], 6.3% of circle junctions contained sequence microhomologies (minimally 5 bp), suggesting the involvement of microhomology-mediated DNA repair (Supplementary Fig. 4l). Notably, eccDNA and ecDNA were significantly enriched in genic regions, particularly in *MYCN*-amplified neuroblastomas (Fig. 1h and Supplementary Fig. 5a–c). Whereas ecDNAs regularly contained entire genes (62.5%), eccDNAs mostly included fractions of genes (Fig. 1i). Our genome-wide map of circular DNA in neuroblastoma shows that DNA circularization is not restricted to proto-oncogenes but also affects various coding and noncoding regions with yet unknown functional consequences.
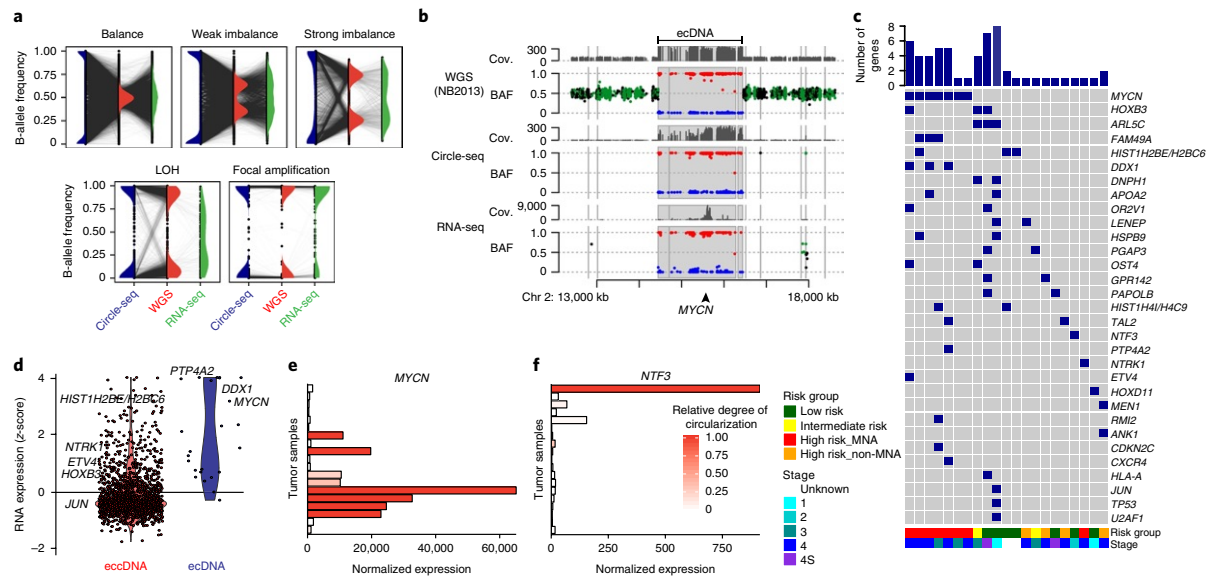
Fig. 2 | Monoallelic large circular DNAs are an origin of oncogene amplification and overexpression in neuroblastoma. a, B-allele frequency of all circular DNAs involving genes (both ecDNA and eccDNA) detected using Circle-seq (blue) compared to the corresponding genomic loci in WGS (red) and mRNA expressed from genes affected by DNA circularization measured using RNA-seq (green; the gray lines indicate the corresponding measurements from Circle-seq, WGS and RNA-seq, $n = 18$ tumors). **b**, Genome track with phased reads from WGS of NB2013, Circle-seq and RNA-seq at the region of extrachromosomal circularization on chromosome 2 affecting *MYCN*. The blue and red colored dots represent reads from different haplotypes. Cov., read coverage. **c**, Genes (rows) affected by circularization in neuroblastoma samples (columns) as detected using Circle-seq ($n = 21$ tumors). **d**, Relative mRNA expression (*z*-scores) of genes affected by DNA circularization in the form of eccDNA ($n = 1,696$) compared to ecDNA ($n = 24$) as measured using total RNA-seq ($n = 21$ tumors). **e,f**, Normalized gene expression (mRNA) for *MYCN* proto-oncogene, basic helix-loop-helix transcription factor (**e**) and *NTF3* (**f**) in neuroblastomas (MNA, *MYCN*-amplified neuroblastoma). The degree of gene circularization is indicated in red (see color scale).

Extrachromosomal circularization and amplification are associated with increased oncogene expression. It is unclear whether circularization itself or subsequent circle copy number amplification drives overexpression. The majority of genomic amplifications (85.7%) identified using WGS coincided with ecDNAs, as confirmed by Circle-seq, suggesting that ecDNAs contribute to genomic amplifications. Moreover, haplotype phasing showed that ecDNAs were exclusively derived from the amplified allele, confirming extrachromosomal circularization as a potential driver of high-level focal genomic amplifications (Fig. 2a,b). Notably, circle length was significantly associated with a higher copy number of circularized regions (Supplementary Fig. 5d; $P < 1 \times 10^{-4}$), implicating circle length as a determining factor for subsequent amplification/propagation of circular DNA (Supplementary Fig. 5d–f). In agreement with its prominent role in neuroblastoma genesis, *MYCN* was the most recurrently extrachromosomally amplified and overexpressed gene in our cohort (Fig. 2b–e and Supplementary Fig. 5a–c). Other cancer-related genes listed in the Catalogue Of Somatic Mutations In Cancer (COSMIC) database[20] were also circularized in tumors and neuroblastoma cell lines, including the *JUN* and *MDM2* proto-oncogenes and *SOX11* and *TAL2* transcription factor genes (Fig. 2c and Supplementary Fig. 5a–c). However, the genomic copy number of oncogenes contained in the majority of eccDNAs was not altered (Supplementary Fig. 5g,h), suggesting that extrachromosomal circularization may be required but insufficient for oncogene amplification.

To determine the consequences of DNA circularization on gene expression, we performed RNA sequencing (RNA-seq) on our neuroblastoma cohort. Whereas differences in gene expression were not observed for most genes affected by circularization in the form of small eccDNA (Fig. 2d and Supplementary Fig. 5i–j), massive

increases in expression occurred for a small subset of genes entirely incorporated on circularized DNA and amplified as ecDNA (Fig. 2d–f). For example, *NTF3*, a gene encoding a neurotrophic factor with known importance in neuroblastoma[21], was strongly expressed from amplified ecDNA (Fig. 2f). Allele-specific messenger RNA expression (allele-specific expression (ASE)) analysis confirmed that increased gene expression originated from the circular allele (Fig. 2a,b). In contrast, ASE from copy number–neutral extrachromosomal circles did not differ from noncircular counterparts (Supplementary Fig. 5g,i,j; binomial test for equal probability, $P = 0.24$), suggesting that DNA circularization was insufficient to induce high-level gene expression. Thus, even though DNA circularization is a major route to gene amplification, it appears insufficient alone (without combined amplification) to increase gene expression. Given this observation, we hypothesized that circular DNA may have additional, cancer-relevant functions.

The genome-wide frequency and functional impact of circle-derived structural rearrangements, such as chimeric circle formation (circular DNA including parts from different chromosomes)[17,22], and circular DNA reintegration[23], in neuroblastomas are currently unknown. We hypothesized that beyond their ability to drive gene amplification, circular DNAs may serve as substrates for oncogenic genome remodeling. We sought evidence of genomic rearrangements at circularization loci (ecDNA and eccDNA) in WGS data (Supplementary Fig. 1). Strikingly, most intrachromosomal and interchromosomal rearrangements detected in neuroblastoma genomes coincided with regions of extrachromosomal circularization, supporting the idea of circle-mediated genome remodeling (Supplementary Fig. 6a,b). Visual inspection of Circos plots from each tumor showed that interchromosomal rearrangements at circularization loci often formed a tree-shaped pattern, defined as
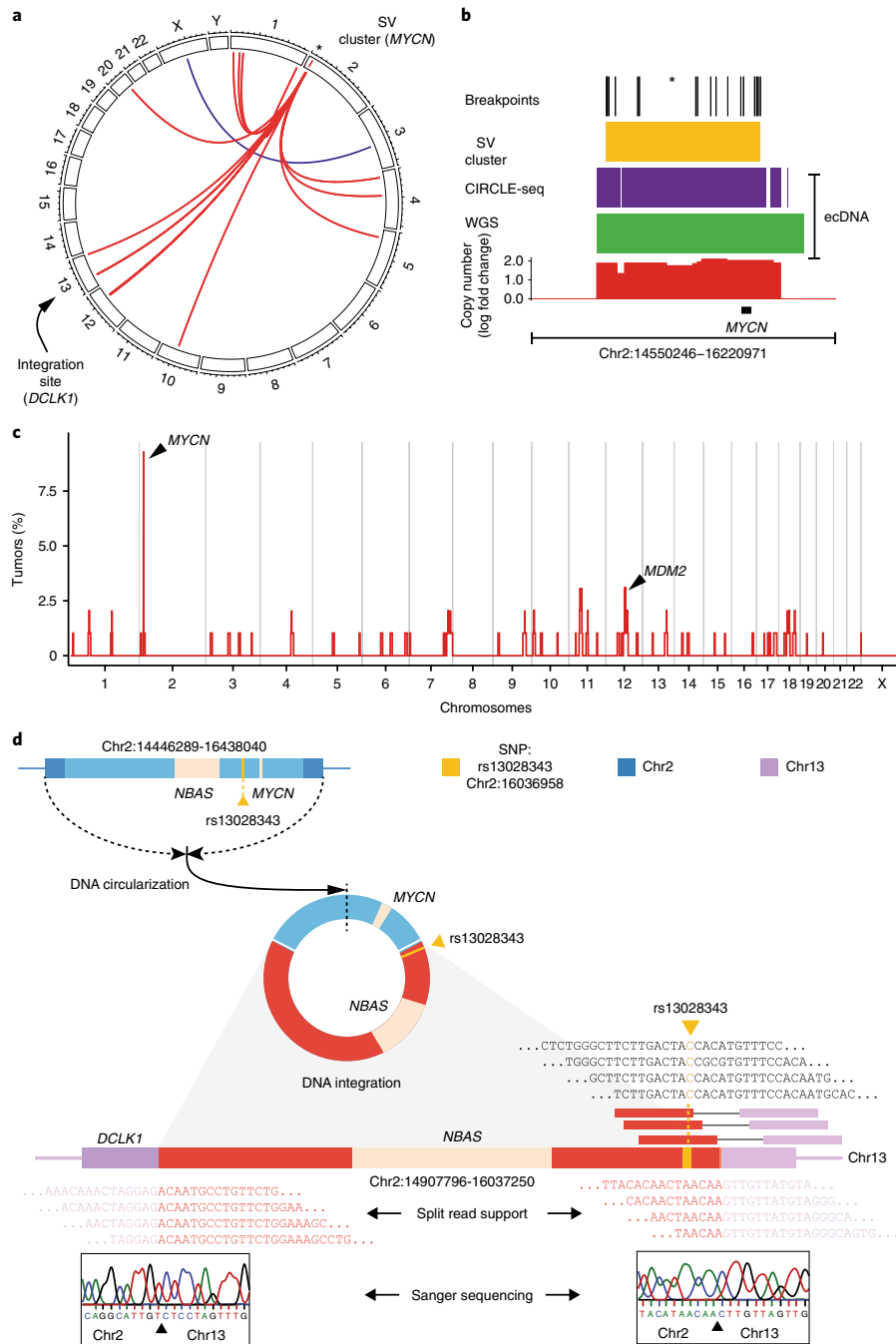
Fig. 3 | The majority of structural rearrangements involve sites of DNA circularization and form clustered rearrangement patterns in neuroblastoma. a, Circos plot of interchromosomal rearrangements identified using five variant detection algorithms in one neuroblastoma genome (CB2013), shown exemplarily. The tree-shaped clustered rearrangement pattern (red), originating at a region of *MYCN* circularization (asterisk) is highlighted (SV, structural variants). **b**, Detailed view of genomic breakpoint localizations (black) at the base of the tree-shaped rearrangement cluster (SV cluster) for the neuroblastoma shown in **a** define a region of clustered breakpoints (yellow) and overlaps with the region of DNA circularization, as detected using Circle-seq (pink) and WGS (green). The copy number changes are highlighted in red. **c**, Genome-wide frequency of tree-shaped clusters of rearrangements in 93 primary neuroblastoma samples. The pattern is recurrently identified on chromosome 2 (at the *MYCN* locus), and chromosomes 11 and 12 (at the *MDM2* locus). **d**, Schematic representation of circle integration in an example of neuroblastoma (CB2013). The genomic region, including *MYCN* (blue), is circularized; parts of the extrachromosomal circle are integrated (red) into chromosome 13 (pink) leading to a disruption of *DCLK1*. The sequencing reads supporting a circle-specific SNP as well as split reads supporting circle integration are shown below. Sanger sequencing of integration breakpoints is shown in the boxes.
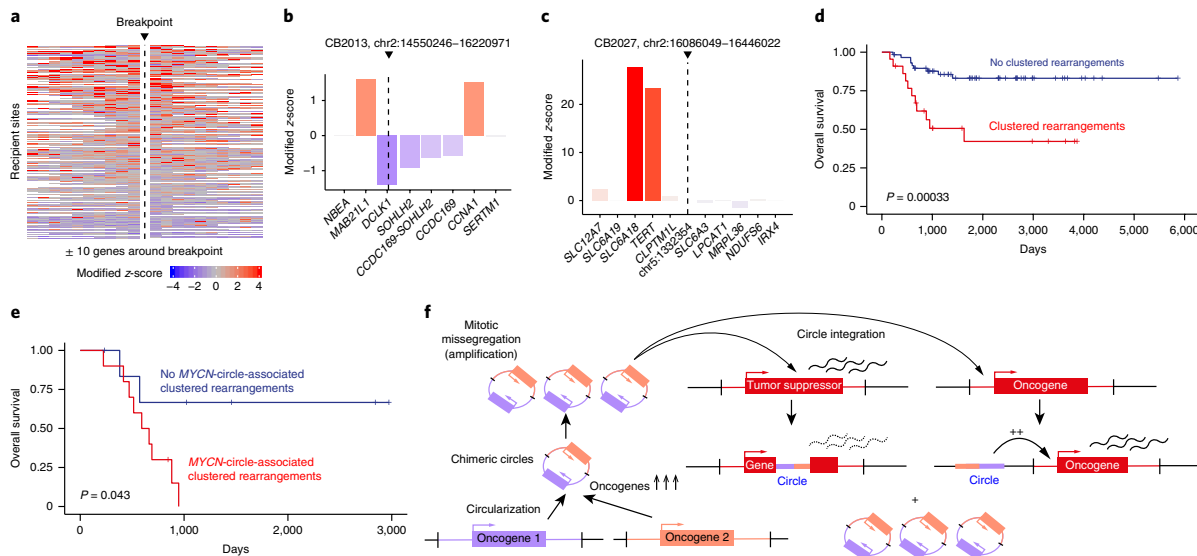
Fig. 4 | Rearrangement of circular DNAs drives transcriptional deregulation and dismal prognosis in neuroblastoma. a, Heatmap showing differential expression of up to ten genes located both upstream and downstream or a maximal distance of 2 Mb from each circle-derived rearrangement breakpoint (*n* = 259 breakpoints, *n* = 24 tumors). **b,c**, The modified *z*-scores for the expression of the cancer-relevant genes *DCLK1* (**b**) and *TERT* (**c**) affected by circle-derived rearrangements are shown for two representative genomic loci (in two neuroblastomas). **d**, Kaplan–Meier analysis comparing the neuroblastoma patient survival of patients with neuroblastomas affected by circle-derived clustered rearrangements (*n* = 22 patients) to patients with tumors lacking such rearrangements (*n* = 59 patients, *P* = 0.00033, two-sided log-rank test). **e**, Kaplan–Meier analysis comparing neuroblastoma patient survival with *MYCN*-amplified tumors affected by *MYCN*-circle-derived clustered rearrangements (*n* = 10) to patients with tumors lacking such rearrangements (*n* = 7, *P* = 0.043, two-sided log-rank test). **f**, Schematic diagram of the proposed mechanism of circle-mediated genome remodeling.

clusters of at least three interchromosomal rearrangements with the same origin and branches reaching other distant genomic regions (Fig. 3a,b and Supplementary Fig. 7a–l). Tree-shaped rearrangement cluster origins significantly overlapped with ecDNAs, with hot spots on chromosomes 2 (including *MYCN*) and 12 (Fig. 3c and Supplementary Fig. 7i). Only 10.5% of *MYCN*-amplified neuroblastomas displayed homogenously staining regions (Supplementary Table 1), consistent with their rarity in neuroblastomas[14,24,25]. Thus, the majority of *MYCN*-derived tree-shaped rearrangements did not represent homogenously staining regions. Tree-shaped rearrangement patterns indicative of circle-derived rearrangements were detected in 9% of pediatric tumors in the analysis of an independent dataset of structural rearrangements in 546 pediatric cancer genomes[26], confirming that this pattern is neither entity-specific nor dependent on variant detection methods (Supplementary Fig. 7j). Our data reveal an unanticipated association between circular DNA and somatic genomic rearrangements in neuroblastoma.

We reasoned that circle-derived tree-shaped rearrangements could either represent chromosomal circle integrations or the formation of chimeric circles, incorporating different chromosomal parts. To test this, we inspected the rearrangement recipient sites for signs of extrachromosomal circularization and integration and performed de novo assembly of circular DNAs (ecDNA and eccDNA). Extrachromosomal circular DNAs (identified using Circle-seq) appeared in 5.5% of rearrangement recipient sites (tree branch intercepts), indicating chimeric circle formation (Supplementary Fig. 6). This was confirmed by long-read Nanopore sequencing and assembly-based circle reconstruction, determining chimeric structures in 2.1% of eccDNAs and 84% of ecDNAs with on average 2.2 and 4.8 chimeric segments, respectively. Chromosomal circle integration was defined as interchromosomal rearrangements connecting extrachromosomal circles with intrachromosomal sites (that is, not detected by Circle-seq). The majority of rearrangement recipient sites (83.3%) were classified as circle

integrations (Fig. 3d and Supplementary Fig. 6), which were validated by visual inspection of split reads, allele-specific PCR and Sanger sequencing (Fig. 3d and Supplementary Fig. 8). Phased heterozygous SNPs near integration breakpoints further confirmed extrachromosomal DNA circles as the origin of the integrations (Fig. 3d). Thus, circle-derived, tree-shaped rearrangement clusters represent (1) formation of chimeric circles and (2) chromosomal circle integrations.

To test the functional impact of circle-derived, tree-shaped rearrangements in neuroblastoma, we inspected the rearrangement recipient sites for the presence of cancer-relevant genes and changes in gene expression (Fig. 4a). Circle integration sites and sites included in chimeric circles were significantly enriched for cancer-relevant genes (*P* = 0.033) and particularly for tumor suppressor genes (*P* = 0.033), whose expression varied from tumors where the same gene was not involved in circle-derived rearrangements (Fig. 4b,c and Supplementary Fig. 9). For example, integration of an extrachromosomal circle fragment into the *DCLK1* gene (shown in Fig. 3d) led to loss of heterozygosity (LOH) and was associated with significant repression of *DCLK1* expression (Fig. 4b). In agreement with a tumor suppressor function in neuroblastoma, low *DCLK1* expression was associated with adverse patient prognosis and short hairpin RNA-mediated DCLK1 knockdown significantly increased clonogenicity in neuroblastoma cell lines (Supplementary Fig. 10a–i). Notably, circle integration also occurred proximal to the *TERT* gene and was associated with enhanced *TERT* expression (Fig. 4c). It is tempting to speculate that enhancer hijacking[27] or disruption of other *cis*-regulatory elements could explain such expression changes. Chimeric circle formation, on the other hand, often resulted in simultaneous amplification of multiple proto-oncogenes and aberrant circle-specific fusion transcript expression in a subset of cases (Supplementary Fig. 11). Thus, circle-derived rearrangements can contribute to aberrant expression of cellular tumor suppressors and proto-oncogenes.

Seemingly genetically identical *MYCN*-amplified neuroblastomas can produce strong clinical heterogeneity, representing a conundrum in the field. We hypothesized that circle-derived oncogenic lesions could functionally cooperate with extrachromosomal circular *MYCN* amplification, explaining some of the clinical heterogeneity observed. Indeed, the presence of circle-derived rearrangements was associated with adverse patient outcome (Fig. 4d). In line with our hypothesis, patients with *MYCN*-amplified neuroblastomas and circle-derived rearrangement clusters involving *MYCN* had significantly worse overall survival compared to patients with *MYCN*-amplified tumors lacking such rearrangements (Fig. 4e). Contrastingly, the number of rearrangements in *MYCN*-amplified tumors did not correlate with survival (Supplementary Fig. 12a–c). This implicates circle-derived rearrangements as clinically relevant genomic alterations in neuroblastoma.

Our work provides a comprehensive map of extrachromosomal DNA circularization in neuroblastoma, revealing this mutagenic process to be more frequent than previously anticipated. We demonstrate that the majority of genomic rearrangements in neuroblastoma involve circular DNA, challenging our current understanding about cancer genome remodeling. Such rearrangements have previously gone largely undetected or underestimated in WGS analyses because integrative, sequencing-based methods identifying circular DNA in tumor samples were lacking. In contrast to previous cytogenetic reports describing homogenously staining region-based circle integration and chimeric circle formation as a means of stable gene amplification, we conclude that extrachromosomally circularized DNA can actively contribute to genome remodeling with important functional and clinical consequences (Fig. 4f). It is tempting to speculate that factors exist, such as recently described oncogenic transposases[28–30], that could induce a mutator phenotype in the presence of circular DNA, driving circle-mediated genome remodeling. We envision that our findings extend to other cancers and that further detailed analyses of circle-derived rearrangements will shed new insights into our understanding of cancer genome remodeling.

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/s41588-019-0547-z.

References

1. Turner, K. M. et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).
2. Møller, H. D. et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat. Commun.* **9**, 1069 (2018).
3. Shibata, Y. et al. Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* **336**, 82–86 (2012).
4. Pennisi, E. Circular DNA throws biologists for a loop. *Science* **356**, 996 (2017).
5. Verhaak, R. G. W., Bafna, V. & Mischel, P. S. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution.*Nat. Rev. Cancer* **19**, 283–288 (2019).
6. Møller, H. D., Parsons, L., Jørgensen, T. S., Botstein, D. & Regenberg, B. Extrachromosomal circular DNA is common in yeast. *Proc. Natl Acad. Sci. USA* **112**, E3114–E3122 (2015).
7. Tjio, J. H. & Levan, A. The chromosome number of man. *Hereditas* **42**, 1–6 (1956).
8. Garsed, D. W. et al. The architecture and evolution of cancer neochromosomes. *Cancer Cell* **26**, 653–667 (2014).
9. Rausch, T. et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with *TP53* mutations. *Cell* **148**, 59–71 (2012).
10. Kohl, N. E. et al. Transposition and amplification of oncogene-related sequences in human neuroblastomas. *Cell* **35**, 359–367 (1983).
11. deCarvalho, A. C. et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat. Genet.* **50**, 708–717 (2018).
12. Nikolaev, S. et al. Extrachromosomal driver mutations in glioblastoma and low-grade glioma. *Nat. Commun.* **5**, 5690 (2014).
13. Schwab, M. et al. Amplified DNA with limited homology to *myc* cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* **305**, 245–248 (1983).
14. 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).
15. Cox, D., Yuncken, C. & Spriggs, A. I. Minute chromatin bodies in malignant tumours of childhood. *Lancet* **1**, 55–58 (1965).
16. Sanborn, J. Z. et al. Double minute chromosomes in glioblastoma multiforme are revealed by precise reconstruction of oncogenic amplicons. *Cancer Res.* **73**, 6036–6045 (2013).
17. Deshpande, V. et al. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat. Commun.* **10**, 392 (2019).
18. Avet-Loiseau, H. et al. Morphologic and molecular cytogenetics in neuroblastoma. *Cancer* **75**, 1694–1699 (1995).
19. Dillon, L. W. et al. Production of extrachromosomal microDNAs is linked to mismatch repair pathways and transcriptional activity. *Cell Rep.* **11**, 1749–1759 (2015).
20. Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
21. Bouzas-Rodriguez, J. et al. Neurotrophin-3 production promotes human neuroblastoma cell survival by inhibiting TrkC-induced apoptosis. *J. Clin. Invest.* **120**, 850–858 (2010).
22. Xu, K. Structure and evolution of double minutes in diagnosis and relapse brain tumors. *Acta Neuropathol.* **137**, 123–137 (2019).
23. 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).
24. Villamón, E. et al. Genetic instability and intratumoral heterogeneity in neuroblastoma with *MYCN* amplification plus 11q deletion. *PLoS ONE* **8**, e53740 (2013).
25. 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).
26. Gröbner, S. N. et al. The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).
27. Northcott, P. A. et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).
28. Henssen, A. G. et al. PGBD5 promotes site-specific oncogenic mutations in human tumors. *Nat. Genet.* **49**, 1005–1014 (2017).
29. Henssen, A. G. et al. Genomic DNA transposition induced by human PGBD5. *eLife* **4**, e10565 (2015).
30. Henssen, A. G. et al. Forward genetic screen of human transposase genomic rearrangements. *BMC Genomics* **17**, 548 (2016).

Methods

Reagents. The synthetic oligonucleotides listed in Supplementary Table 2 were obtained from Eurofins Genomics and were salt-free purified. pLKO.1 shRNA vectors targeting *DCLK1* (TRCN0000002145, TRCN0000002146) and control short hairpin green fluorescent protein were obtained from the RNAi Consortium (Broad Institute).

Cell culture. Human tumor cell lines were obtained from the DSMZ-German Collection of Microorganisms and Cell Cultures (Leibniz Institute), from ATCC or were a gift from C. J. Thiele. The identity of all cell lines was verified by short tandem repeat STR genotyping (Genetica DNA Laboratories and/or IDEXX BioResearch). Absence of *Mycoplasma* contamination was determined with a MycoAlert system (Lonza). Neuroblastoma cell lines were cultured in RPMI 1640 medium (Thermo Fisher Scientific) supplemented with penicillin, streptomycin and 10% FCS. To assess the number of viable cells, cells were trypsinized, resuspended in medium and sedimented at 500*g* for 5 min. Cells were then resuspended in medium, mixed in a 1:1 ratio with 0.02% Trypan Blue Solution (Thermo Fisher Scientific) and counted with a TC20 Automated Cell Counter (Bio-Rad Laboratories). Lentiviral production and transduction were performed as described previously[28]. Clonogenicity was assessed as described previously[28]. Kelly and IMR-5 cells were plated in 24-well microplates at a concentration of 5,000 cells per well and incubated for 7 d. Clonogenicity was quantified using methods described previously[31].

Protein blotting. Protein blotting was performed as described previously[28] using antibodies directed against mouse anti-β-actin (clone 8H10D10; Cell Signaling Technology), mouse anti-α-tubulin (clone DM1A; Cell Signaling Technology) and rabbit anti-DCLK1/DCAMKL1 (clone D2U3L; Cell Signaling Technology).

PCR and Sanger sequencing. PCR reactions were performed on 50–100 ng of gDNA using 0.4U Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific), 0.5 μM forward and reverse primers (Supplementary Table 2), 200 μM deoxyribonucleotide triphosphates (Bio-Budget Technologies) and 4 μl 5× Phusion Green buffer (Thermo Fisher Scientific). PCR products were resolved on 1% agarose gels. PCR amplicons were purified using the PureLink PCR Purification Kit (Thermo Fisher Scientific). Sanger sequencing was carried out by capillary sequencing using standard procedures (Eurofins Genomics).

Quantitative PCR (qPCR). qPCR was performed using 50 ng or 1.5 μl of template DNA and 0.5 μM primers with SYBR Green PCR Master Mix (Thermo Fisher Scientific) in FrameStar 96-well PCR plates (4titude). Reactions were run and monitored on a StepOnePlus Real-Time PCR System (Thermo Fisher Scientific) and Ct values were calculated with the StepOne Plus software v.2.3 (Thermo Fisher Scientific).

Circular DNA isolation, purification and sequencing. Circular DNA isolation and purification was performed on the samples described in Supplementary Tables 3 and 4 similarly to previous reports of Circle-seq[6]. A detailed step-by-step protocol for circular DNA isolation has been deposited on the *Nature* Protocol Exchange server[32]. DNA content was measured with a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific) and a Qubit 3.0 Fluorometer (Thermo Fisher Scientific). Amplified circular DNA was sheared to an average fragment size of 150–200 bp using an S220 focused ultrasonicator (Covaris). Libraries for next-generation sequencing were prepared using the NEBNext Ultra DNA Library Kit for Illumina according to the manufacturer's protocol (New England Biolabs). Libraries were sequenced on MiSeq instruments with 2×150 bp paired-end reads, HiSeq 4000 instruments with 2×125 bp paired-end reads or NextSeq 500 instruments with 2×150 bp paired-end reads (all Illumina). SMRT-seq was performed on a PacBio RS II instrument according to the manufacturer's protocol (Pacific Biosciences). Nanopore sequencing was performed on a MinION instrument according to the manufacturer's protocol (Oxford Nanopore).

Circle-seq analysis. Reads were 3′ trimmed for both quality and adapter sequences, with reads removed if the length was less than 20 nucleotides. Burrows–Wheeler Aligner MEM v.0.7.15 with default parameters was used to align the reads to human reference assembly hg19; PCR and optical duplicates were removed with Picard v.2.16.0. The aligned BAM files were then analyzed in two ways. First, all read pairs and split reads containing any outward-facing read orientation, indicating potential circles, were placed in a new BAM file. Second, genomic segments enriched for signal over background were detected in the 'all reads' BAM file using variable-width windows from Homer v.4.11 findPeaks (http://homer.ucsd.edu/), and the edges of these enriched regions were intersected with the 'circle only' BAM file to quantify the number of circle-supporting reads. To determine the thresholds for significance of real circles versus background noise, matched WGS data were used to determine the background distribution of circle-oriented reads in non-circle-enriched regions that were matched for length and nucleotide composition. An empirical *P* value of 0.01 was used to filter putative circles and regions passing this filter were then used for downstream analysis.

Circle analysis in WGS data. Alignments to hg19 were created as outlined earlier, with read trimming, Burrows–Wheeler Aligner MEM and duplicate removal. Discovery of putative tumor-specific circular DNA relied on the filtering of false positives from genomic sequence as well as circles from normal tissue. This was classified with the following approach: (1) alignments with an outward-facing read orientation served as markers of putative circle boundaries projected onto the linear genome; (2) all such regions were merged if their edges occurred within 500 bp on both ends; (3) regions not meeting the empirically defined background threshold were filtered out (*P* < 0.01; see Circle-seq analysis); (4) lastly, these putative circles were classified as tumor-specific once filtered against circles discovered in the matched normal genome (using steps 1–3). To allow for the detection of copy number–neutral DNA circles, copy number information was not used for this analysis. We confirmed that tandem duplications identified using variant calling algorithms did not identify the same number of circular DNA from the WGS data (Supplementary Fig. 4).

De novo assembly of extrachromosomal circular DNA. De novo assembly of long-read data (SMRT and Nanopore) was accomplished using two approaches. First, for long-read data alone, the Flye v.2.5 assembler (http://github.com/fenderglass/Flye) was used in '-meta' mode with circle junctions evaluated after polishing. Second, for hybrid assemblies using both long and short read data, Unicycler v.0.4.7 (http://github.com/rrwick/Unicycler) was used with racon v.1.3.3 and SPAdes v.3.13.0 and polished with Pilon v.1.23. In all cases, circle assembly was inspected visually using Bandage v.0.8.1 (http://rrwick.github.io/Bandage/). Genic overlap with de novo assemblies was evaluated in two ways. First, by building a BLAST database of all assembled contigs and scoring matches to human genes with at least 70% of gene length covered. Second, each contig, independent of genic overlap, was mapped to hg19 using minimap2 v.2.17 (http://github.com/lh3/minimap2).

SMRT-seq analysis. Reads from the SMRT-seq data were aligned to hg19 using the Burrows–Wheeler Aligner MEM with the, pacbio flag (-k17 -W40 -r10 -A1 -B1 -O1 -E1 -L0). Since these data are single-ended, outward-facing read pairs cannot be used; thus, classification of circle junctions depended on split reads. Segments of the genome enriched for circular DNA were discovered by scanning 10-kilobase (kb) windows and calculating the false discovery rate (FDR)-adjusted *P* value from the Poisson distribution of the randomized reads.

Circle classification. Genome-wide distribution was calculated by dividing each chromosome into 1-megabase (Mb) bins and overlapping with quality-filtered circles. The number of circle reads overlapping each bin was divided by the total number of circle reads per patient, calculated separately for Circle-seq and WGS data. Genic circles were classified with bedtools v.2.25.0 intersect (http://bedtools.readthedocs.io/) against all protein-coding genes, with gene bodies covered at least 20% being used for downstream analysis. Recurrence across samples was calculated from a high-confidence set of genic circles created from genes with at least four circle-supporting reads covering at least 80% of the shortest transcript. Patients with matched Circle-seq, WGS and RNA-seq (*n* = 16) were used to investigate the relationship between circles, amplification and expression with a focus on circles with genic overlap. Correlation plots were computed per patient based on circle coverage, RNA expression and copy number variation fold change. Concordance between gene expression and circles was discovered by converting normalized read counts to *z*-scores and correlating with circle coverage across patient samples. For further methods, see the Supplementary Note.

Circle chimerism. Circle chimerism was evaluated using split reads from Nanopore sequencing (*n* = 21) that either bridged another chromosome or linked to a region separated by at least 4 Mb on the same chromosome. A minimum of 5 reads at a mapping quality (MAPQ) > 30 were required for a region to be considered chimeric; all such regions within the circle length ±500 bp were merged using pgltools v.1.2.0 (http://github.com/billgreenwald/pgltools). The resulting chimeric circles were further used as a secondary metric to evaluate the FDR of clustered tree-shaped rearrangement contacts in the WGS data.

Structural variant detection. Copy number variation was detected using Control-FREEC[33] v.10.6 with contamination adjustment based on a contamination of 0.4 (that is, samples are 60% tumor), a minimalSubclonePresence of 0.244 and with ASCAT v.4.0.1 using default parameters[33,34]. Regions in the genome with a total copy number ≥9 were considered amplified regions following COSMIC copy number variant definition[20]. Amplifications were intersected with regions of circularization using the bedtools v.2.25.0; circular DNAs identified over these amplified regions were classified as ecDNAs. All remaining circular DNAs were classified as eccDNAs. Structural variation was done on matched tumor/normal genomes using novoBreak v.1.1.3 (ref. [35]), SvABA v.1.1.1 (ref. [36]), Delly2 v.0.7.7 (ref. [37]), BRASS v.6.0.5 (https://github.com/cancerit/BRASS) and SMUFIN v.0.9.4 (ref. [38]) using default parameters. From 97 initial neuroblastoma genomes, 4 of them (NBL47, NBL53, NBL54 and NBL61) were excluded from the analysis due to their abnormal high number of breakpoints and amplified regions. The 93 genomes left were analyzed with at least 4 variant callers each. Focusing on interchromosomal

rearrangements, merging and filtering of the results from different variant calling algorithms was performed. Filtering for all variants was performed with a Brass Assembly Score (BAS) ≥99 and at least 6 variant-supporting reads with an MAPQ > 60. All rearrangements that did not have a minimum of 6 aligned supporting reads with an MAPQ > 60 at each breakpoint were discarded. For the merging of interchromosomal rearrangements, all results from different variant callers were joined after filtering. Variants with breakpoints within a window of 500 bp where collapsed. Only intrachromosomal variants supported by at least two different callers were included. Two additional samples (NBL49 and NBL50), which had exceptionally high numbers of rearrangements (z-score > 2) were discarded. A 1-Mb genomic region was blacklisted due to its high number of recurrent, visually confirmed false positive breakpoints (z-score > 2 within the 10 highest-ranking bins). Structural variant calls from an independent cohort of WGS data of 546 pediatric cancer genomes was obtained from the DKFZ Pediatric Pan Cancer dataset (https://hgserver1.amc.nl/cgi-bin/r2/main.cgi?&dscope=DKFZ_PED&option=about_dscope). For further methods, see the Supplementary Note.

Regions of clustered rearrangements. A region of clustered, tree-shaped rearrangement pattern was defined as having three or more interchromosomal rearrangements within a 4-Mb sliding window. The outermost breakpoints defined the boundaries of a cluster region. When five or more interchromosomal rearrangements connected the same two chromosomes, these were flagged and not considered for cluster detection. When 2 or more interchromosomal rearrangements connected 2 regions <10 Mb in size, only one rearrangement was counted for cluster detection. All chromosomes with >25 interchromosomal rearrangements were not considered. All structural variants detected in our dataset, as well as regions of clustered rearrangements detected using the methods described, can be visually inspected in an openly accessible website[39]. To estimate the FDRs, we randomly redistributed breakpoints of each sample across the mappable genome before counting the number of rearrangements within 4-Mb sliding windows. Five hundred such randomized datasets were created. The FDR was estimated as the mean fraction of rearrangement cluster-positive samples in this randomized dataset. For the chosen threshold of 3 or more rearrangements, the estimated FDR was 0.13. The analysis of circle integration was carried out by detecting the rearrangements connecting a circularized region with a candidate insertion site. Integration sites were defined by two main characteristics: both recipient breakpoints being located on the same chromosome and at a distance between breakpoints smaller than the circularized region inserted. Visual inspection of BAM files was performed for each candidate integration site. For further methods, see the Supplementary Note.

Circle length analysis. To identify the length preferences for circles depending on the copy number state of the underlying genomic segment, we derived a zero-sum score, following common enrichment test strategies such as gene set enrichment analysis[40,41]. For a given copy number category (balanced, weak imbalance, strong imbalance, LOH and focal amplification), each circle was assigned a score of $1/k$ if the circle belonged to the category and $-1/(n-k)$ otherwise, where k is the total number of circles in that category and n is the total number of circles. Circles were ranked by length and cumulative scores along the list were calculated. The absolute maximum cumulative score was tested against 10,000 random permutations of the ranked list to determine the approximate enrichment P values. For further methods, see the Supplementary Note.

Circle breakpoint analysis. Base-pair accurate circle junctions were reassembled using SvABA v.1.1.1 with default parameters and only read pairs and split reads containing any outward-facing read orientation as input. Each precise head-to-tail rearrangement call was considered a circle junction. Homology and insertion sequences were taken from the SvABA output directly.

To screen for motifs enriched at circle junction breakpoints, hg19 reference sequences for 41-bp windows around each circle junction breakpoint were obtained. MEME v.5.0.2 (parameters -objfun de -revcomp -nmotifs 5) was used to assess these sequences for motif enrichment with respect to a set of 1 million length-matched sequences randomly sampled from hg19 (excluding poorly or nonassembled regions and the ENCODE DAC blacklist). We compared reference sequence-derived microhomology lengths for actual breakpoints versus a random permutation of breakpoint partners using a two-sided t-test. For further methods, see the Supplementary Note.

Structural variant breakpoint analysis. Base-pair accurate structural rearrangement calls from the merged structural variant set were considered for detailed breakpoint analysis. The hg19 reference sequence was obtained for a 61-bp window around each breakpoint. MEME v.5.0.2 (parameters -objfun de -revcomp -nmotifs 10) was used to identify motifs that were enriched with regard to a set of 1 million length-matched sequences randomly sampled from hg19 (excluding poorly or nonassembled regions and the ENCODE DAC blacklist). Differential enrichment was equally assessed to compare subsets of rearrangements (clustered rearrangements versus nonclustered rearrangements, circle–circle versus other, circle–genome versus other, genome–genome versus other). Only SvABA rearrangement could be readily analyzed for homology and inserted sequences at breakpoints. We compared reference sequence-derived microhomology lengths for

actual breakpoints versus a random permutation of breakpoint partners using a two-sided t-test. For further methods, see the Supplementary Note.

Statistical analysis. The enrichment of rearrangements in circularization loci was done using a two-sample test for equality of proportions with continuity correction. The enrichment of interchromosomal rearrangement breakpoint clusters within circularized regions was assessed using the union of interchromosomal rearrangements detected by all variant callers and at regions of circularization determined using Circle-seq and WGS separately. The relative overlap of each region of clustered breakpoints with circularized regions in the respective sample was computed. The distribution of overlap was then compared to the distribution expected by chance. For each region of clustered rearrangements, 2,000 random intervals of matching length were randomly positioned over a masked genome that excluded poorly or nonassembled regions and the ENCODE DAC blacklist. The relative overlap of each random interval with circular DNA in the matching patient was then assessed. A hypothesis test was derived from considering the mean relative overlap for the set of observed cluster regions with regard to the distribution of the mean relative overlap for the 2,000 synthetic sets of cluster regions. The one-sided empirical P value was calculated and Benjamini–Hochberg-corrected for multiple comparisons (circle classes and circle calling methods). We investigated the distance of distal breakpoints of tree-shaped clustered rearrangements. We tested whether these breakpoints were closer to certain classes of genes than expected by chance. We looked at three gene classes: all COSMIC v.87 genes versus only COSMIC v.87 oncogenes versus only COSMIC v.87 tumor suppressor genes. For each breakpoint, we calculated the distance to the closest gene of the particular gene class and calculated the class-wise median of distances. Each median was assigned a one-tailed P value based on the distribution of medians in 500 synthetic datasets with breakpoint positions randomly drawn from the nonblacklisted genome. P values were corrected for multiple testing using the Benjamini–Hochberg method. To assess gene expression changes around rearrangement breakpoints, expression of protein-coding genes within 2 Mb of each breakpoint were analyzed. The differential RNA expression of genes in each sample compared to the rest of the cohort was quantified and the modified z-score of their transcripts per million was calculated. Two-sided log-rank tests were used for survival analysis across subgroups. To assess the effect of rearrangement clusters at the *MYCN* amplicon locus, *MYCN*-associated clusters were defined as all clusters that overlapped the ±1-Mb window around the *MYCN*. All violin plots depict the smoothed distribution using a Gaussian kernel with bandwidth selected according to Silverman's rule. The box plots depict the first and third quartiles, segmented by the median; the whiskers depict the points within the 1.5× interquartile range beyond the box edges. All cell culture experiments were conducted at least three independent times, unless otherwise stated. For further details, see the Nature Research Reporting Summary. For further methods, see the Supplementary Note.

Patient samples and clinical data access. This study comprised the analyses of tumor and blood samples of patients diagnosed with neuroblastoma between 1991 and 2016. Patients were registered and treated according to the trial protocols of the German Society of Pediatric Oncology and Hematology (GPOH). This study was conducted in accordance with the World Medical Association Declaration of Helsinki (2013) and good clinical practice; informed consent was obtained from all patients or their guardians. The collection and use of patient specimens was approved by the institutional review boards of Charité-Universitätsmedizin Berlin and the Medical Faculty, University of Cologne. Specimens and clinical data were archived and made available by Charité-Universitätsmedizin Berlin or the National Neuroblastoma Biobank and Neuroblastoma Trial Registry (University Children's Hospital Cologne) of the GPOH. The *MYCN* gene copy number was determined as a routine diagnostic method using FISH. DNA and total RNA were isolated from tumor samples with at least 60% tumor cell content as evaluated by a pathologist. For further methods, see the Supplementary Note.

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

Data availability

The WGS and RNA-seq data that support the findings of this study have been deposited with the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/) under accession nos. EGAS00001001308 and EGAS00001004022. The Circle-seq data that support the findings of this study are available from the corresponding author upon request. Source data for Fig. 1 are available online.

Code availability

The scripts used to analyze the sequencing data have been uploaded to www.github.com/henssenlab. Data on tree-shaped rearrangements can be accessed and visualized online (https://kons.shinyapps.io/trees/).

References

31. Guzmán, C., Bagga, M., Kaur, A., Westermarck, J. & Abankwa, D. ColonyArea: an ImageJ plugin to automatically quantify colony formation in clonogenic assays. *PLoS ONE* **9**, e92444 (2014).

32. Henssen, A. G., MacArthur, I., Koche, R. & Dorado-García, H. Purification and sequencing of large circular DNA from human cells. *Protoc Exch* (2019); https://doi.org/10.1038/protex.2019.006

33. Boeva, V. et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).

34. Van Loo, P. et al. Allele-specific copy number analysis of tumors. *Proc. Natl Acad. Sci. USA* **107**, 16910–16915 (2010).

35. Chong, Z. et al. novoBreak: local assembly for breakpoint detection in cancer genomes. *Nat. Methods* **14**, 65–67 (2017).

36. Wala, J. A. et al. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res.* **28**, 581–591 (2018).

37. Rausch, T. et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).

38. Moncunill, V. et al. Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. *Nat. Biotechnol.* **32**, 1106–1112 (2014).

39. Helmsauer, K. *Tree-shaped Rearrangement Patterns in Pediatric Cancer Genomes* (2019); https://kons.shinyapps.io/trees

40. Mootha, V. K. et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).

41. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

Author contributions

R.P.K., E.R.F., K.H., J.M., F.H., I.C.M., R.C., A.W., M.B., M.P., C. Röefzaad, T.T., H.D., R.F.S., C. Rosswog, J. Theissen, V.B., N.M.P., H.D.G., Y.B., A.S., N. Timme, K.K., S.F., N. Thiessen, E.B., K.S, A.K., P.H., J. Toedling, M.F., D.B., S.S., A.E., D.T., J.H.S. and A.G.H. contributed to the study design and the collection and interpretation of the data. R.P.K. performed the analysis of Circle-seq and WGS. E.R.F. performed the data analysis of the WGS data. I.C.M., R.C., N.M.P. and H.D.G. performed the circular DNA extraction and PCR-based validation. V.B., C. Röefzaad, A.W., P.H., K.S., M.F., A.S. and F.H. collected and prepared the patient samples. C. Rosswog. and J. Theissen performed and analyzed the FISH. M.B. and R.F.S. performed the haplotype phasing analyses. M.P. and J. Toedling. analyzed the tumor genome sequencing data. M.B. performed the allele-specific analysis of Circle-seq, WGS and RNA-seq. S.F., F.K., R.P.K, K.H. and J.M. performed the RNA-seq data analysis. P.H., H.D.G., N.M.P., A.S., D.B. and K.S. performed the experiments and analyzed the data. R.P.K., A.K., A.E. and J.H.S. contributed to the study design. R.P.K. and A.G.H. led the study design, performed the data analysis and wrote the manuscript. All authors contributed to manuscript drafting.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0547-z.

Correspondence and requests for materials should be addressed to R.P.K. or A.G.H.

Reprints and permissions information is available at www.nature.com/reprints.

Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma

Richard P. Koche [1,12]*, Elias Rodriguez-Fos [2,12], Konstantin Helmsauer [3,12], Martin Burkert [4,5,6], Ian C. MacArthur[3], Jesper Maag[1], Rocio Chamorro[3], Natalia Munoz-Perez[3], Montserrat Puiggròs[2], Heathcliff Dorado Garcia [3], Yi Bei [3], Claudia Röefzaad[3], Victor Bardinet[3], Annabell Szymansky[3], Annika Winkler[3], Theresa Thole[3], Natalie Timme[3], Katharina Kasack[7], Steffen Fuchs [3,5,7], Filippos Klironomos [3], Nina Thiessen[5], Eric Blanc[5], Karin Schmelz[3], Annette Künkele[3,5,7], Patrick Hundsdörfer[3,5,7], Carolina Rosswog[7], Jessica Theissen[7], Dieter Beule [5], Hedwig Deubzer[3,7,8], Sascha Sauer[5], Joern Toedling[3], Matthias Fischer [9,10], Falk Hertwig[3,7], Roland F. Schwarz [6,7], Angelika Eggert[3,5,7], David Torrents[2,11,12], Johannes H. Schulte[3,5,7,12] and Anton G. Henssen [3,5,7,8,12]*

[1]Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Barcelona Supercomputing Center, Joint Barcelona Supercomputing Center-Centre for Genomic Regulation-Institute for Research in Biomedicine Research Program in Computational Biology, Barcelona, Spain. [3]Department of Pediatric Oncology/Hematology, Charité-Universitätsmedizin Berlin, Berlin, Germany. [4]Department of Biology, Humboldt University, Berlin, Germany. [5]Berlin Institute of Health, Berlin, Germany. [6]Max Delbrück Center for Molecular Medicine, Berlin, Germany. [7]German Cancer Consortium (DKTK), partner site Berlin, and German Cancer Research Center (DKFZ), Heidelberg, Germany. [8]Experimental and Clinical Research Center, Max Delbrück Center for Molecular Medicine and Charité-Universitätsmedizin Berlin, Berlin, Germany. [9]Department of Experimental Pediatric Oncology, University Children's Hospital of Cologne, Cologne, Germany. [10]Center for Molecular Medicine Cologne, Medical Faculty, University of Cologne, Cologne, Germany. [11]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain. [12]These authors contributed equally: Richard P. Koche, Elias Rodriguez-Fos, Konstantin Helmsauer, David Torrents, Johannes H. Schulte, Anton G. Henssen. *e-mail: kocher@mskcc.org; henssenlab@gmail.com

Supplementary information for Koche et al. *"Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma"*
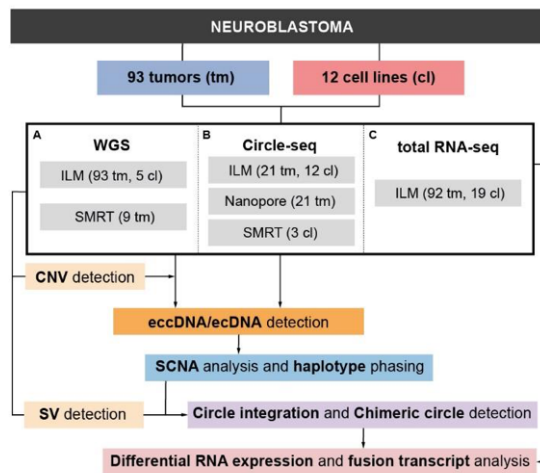
Content:

Supplementary Figures
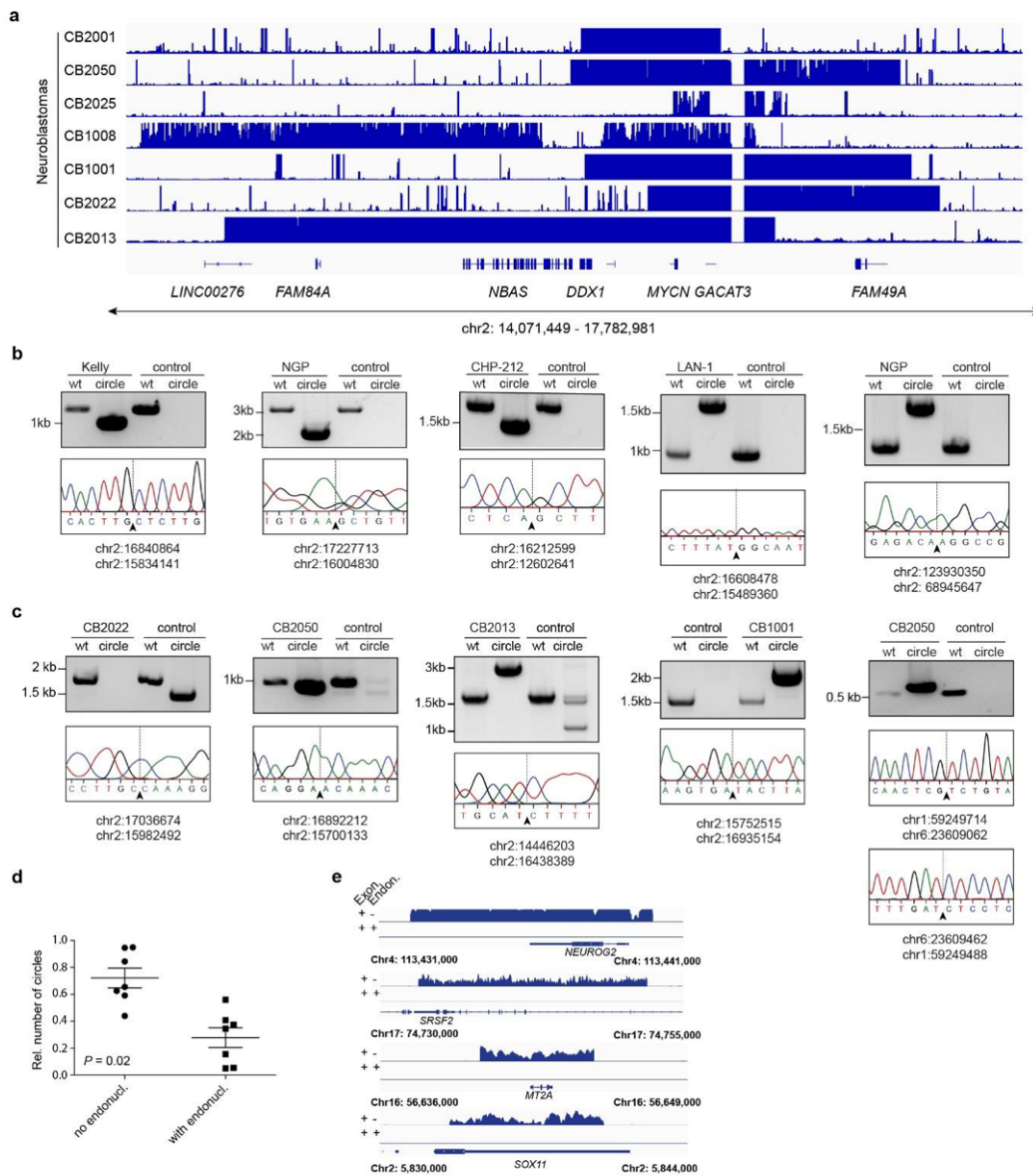
Supplementary Tables

Supplementary Note

1

Supplementary Fig. 1. Flow chart of the data analysis strategy. (WGS: Whole genome sequencing; ILM: Illumina sequencing; SMRT: Single molecule real-time sequencing; SCNA: Somatic copy number alterations; SV: Structural variant; RNA-seq: RNA sequencing).

2
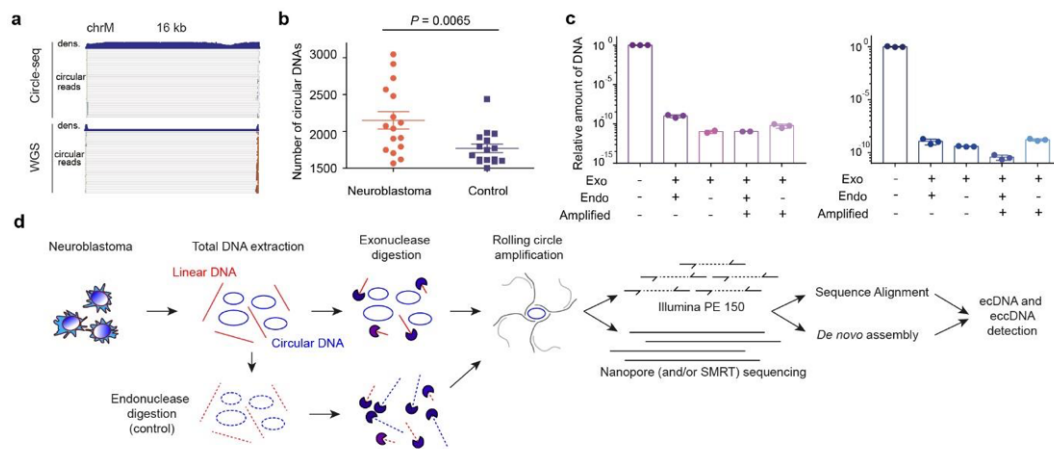
Supplementary Fig. 2. Circle-seq enables efficient enrichment of extrachromosomal circular DNA. a, Circle read density from Circle-seq (top) and whole genome sequencing (bottom) from mitochondrial DNA (ChrM) in one exemplary patient. **b,** Number of circular DNAs detected in 16 neuroblastoma tumors-normal pairs (center line indicates mean, error bars indicate standard deviation from the mean, two sided t-test, *P*=0.0065). **c,** Relative amount of linear genomic DNA (β-globin) before and after exonuclease +/- endonuclease treatment as measured using quantitative PCR (qPCR) in two independent cell lines (right and left; center indicates mean, error bars represent standard deviation of at least two independent qPCR measurements). **d,** Detailed schematic of the Circle-seq method (A step by step protocol is available online)[1].

3

Supplementary Fig. 3. Circle-seq reliably detects extrachromosomal circular DNA. a, Extrachromosomal circular DNA read density at chromosome 2 near *MYCN* in primary neuroblastomas as detected using Circle-seq. Extrachromosomal circle junction and wild-type allele-specific PCR in neuroblastoma cell lines **(b)** and matched tumor and normal primary patient specimens (**c**) repeated independently at least three times. **d,** Relative number of extrachromosomal circular DNAs detected using Circle-seq before and after treatment with a rare-cutting endonuclease in 7 independent neuroblastoma cell lines (Error bars represent standard deviation, mean is indicated by horizontal line, unpaired two-sided t-test, *P*=0.02). **e,** Exemplary genome tracks at sites of extrachromosomal circularization detected via Circle-seq in each tumor before and after treatment with a rare-cutting endonuclease (number of reads in spanning all circular DNAs was reduced 474 fold, $P = 7.566 \times 10^{-11}$, Welch two sample two sided t-test).
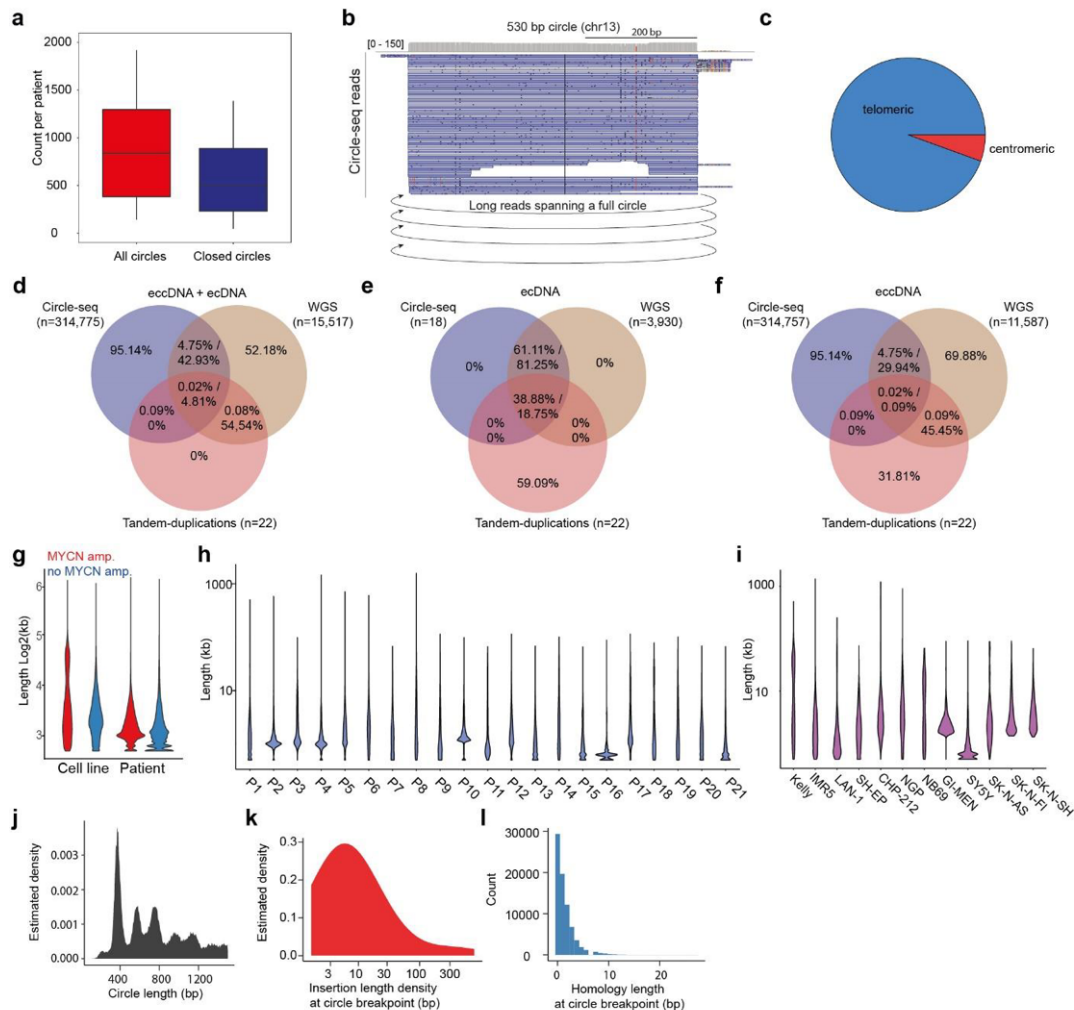
Supplementary Fig. 4. Combining whole-genome sequencing with Circle-seq enables the characterization of extrachromosomal circular DNAs. a, Circular DNAs detected by Nanopore long-read sequencing (*N*=19,947) and with fully closed circles by junction-spanning reads (*N*=12,985; 65%) in 21 patient samples. **b**, Exemplary genome track at site of extrachromosomal circularization with long reads spanning an entire circular DNA multiple times, physically confirming its circular structure. **c**, Fraction of unmappable Nanopore long reads, which are mapped to telomeric and centromeric sequences after *de novo* assembly. Intersection between all circular DNAs **(d)**, ecDNAs **(e)** and eccDNAs **(f)** detected using Circle-seq compared to circular DNAs inferred from whole-genome sequencing and regions recognized as tandem-duplications by 5 variant callers in 16 neuroblastoma patients. Size distribution of extrachromosomal circular DNAs identified using Circle-seq in 21 primary neuroblastomas **(g+h)** and 12 neuroblastoma cell lines **(g+i)**. **j**, Size distribution of small extrachromosomal circles less than 1,500bp in length (*N*=59,560) with basepair-accurate breakpoint reconstruction using Circle-seq in 17 primary neuroblastomas. Density estimate using a Gaussian kernel with standard deviation set to 3. **k**, Length distribution of sequence insertions (*N*=2,145) at basepair-accurate reconstructions of circle junctions in 17 primary neuroblastomas. Density estimate using a Gaussian kernel with standard deviation set to 1. **l,** Length distribution of homologous sequences (microhomologies) at base-pair accurate reconstructions of circle junctions in 17 primary neuroblastomas (*N*=76,220).

5
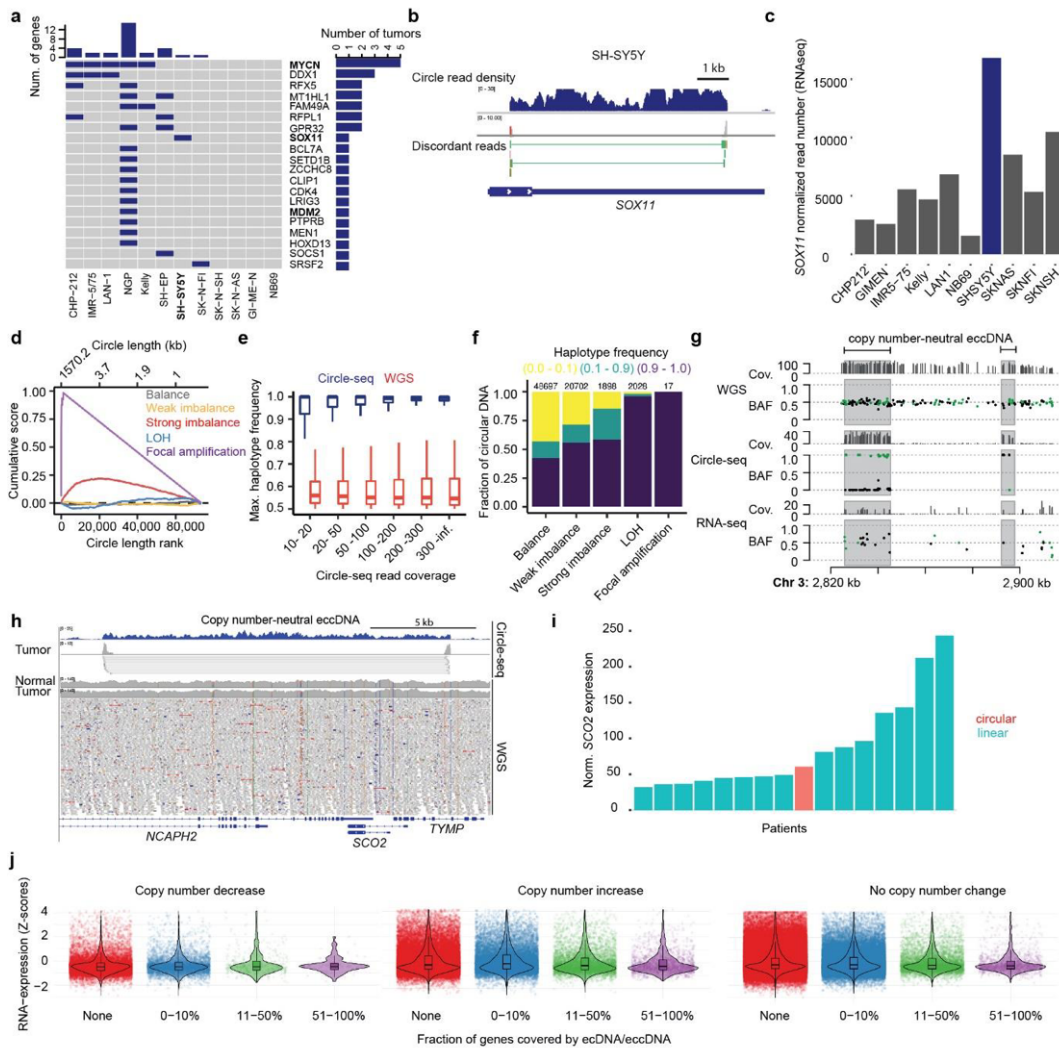
Supplementary Fig. 5. Copy number-neutral extrachromosomal circular DNAs are not associated with changes in gene expression. a, Cancer-relevant genes (rows) circularized in neuroblastoma cell lines (columns) as detected using Circle-seq (*N*=12 cell lines). **b,** Circle read density and genome track at *SOX11* gene. **c,** mRNA expression of *SOX11* in 12 neuroblastoma cell lines. **d**, Cumulative enrichment score of genomic copy number states within a ranked list of extrachromosomal circle sizes (*N*=73,342). Large extrachromosomal circular DNAs are significantly associated with focal copy number amplifications (pink, P_{emp}=0, empirical nominal one-sided *P*-value of absolute maximum cumulative scores from 10,000 random permutations of copy number scores. None of the random absolute maximum cumulative scores was greater than the observed score of 0.98; LOH: Loss of heterozygosity). **e**, Maximum haplotype frequency in Circle-seq compared to WGS at different Circle-seq coverages of 73,342 circles across samples (Box indicates first, second and third quartile. Whiskers extend to lowest and highest value at max. 1.5 × interquartile distance from first and third quartile). **f**, Fraction of extrachromosomal circular DNA with distinct haplotype preferences depending on their copy number status. **g**, Genome track at the site of a copy number neutral extrachromosomal circular DNA showing the coverage of haplotype specific, phased, reads from Circle-seq, whole genome sequencing (WGS) and RNA sequencing. **h**, Genome track at site of a copy number neutral extrachromosomal circular DNA affecting *SCO2*. **i**, Normalized

6

SCO2 RNA expression in a subset of patient tumors. **j,** Degree of gene circularization and gene copy number differences (increase, *N*=26,374; decrease, *N*=5,656; no change, *N*=58,862) compared to RNA expression differences (z-scores).
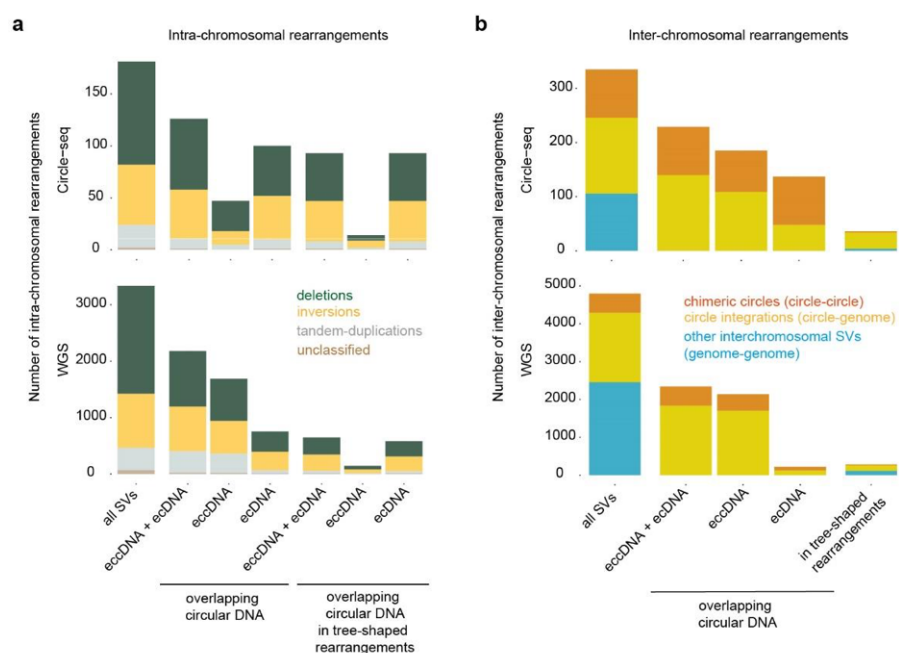

Supplementary Fig. 6. The majority of somatic structural rearrangements in neuroblastoma involve extrachromosomal circular DNA. Intra- (**a**) and inter-chromosomal (**b**) somatic rearrangements and their association with extrachromosomal circular DNAs (ecDNAs compared to eccDNAs) as detected by Circle-seq (*N*=16 tumors) and WGS (*N*=93 tumors) define two classes of circle-associated rearrangements, chimeric circles (connecting extrachromosomal circular DNA to extrachromosomal circular DNA, circle-circle) and circle integrations (connecting extrachromosomal circular DNA to chromosomal linear DNA, circle-genome).
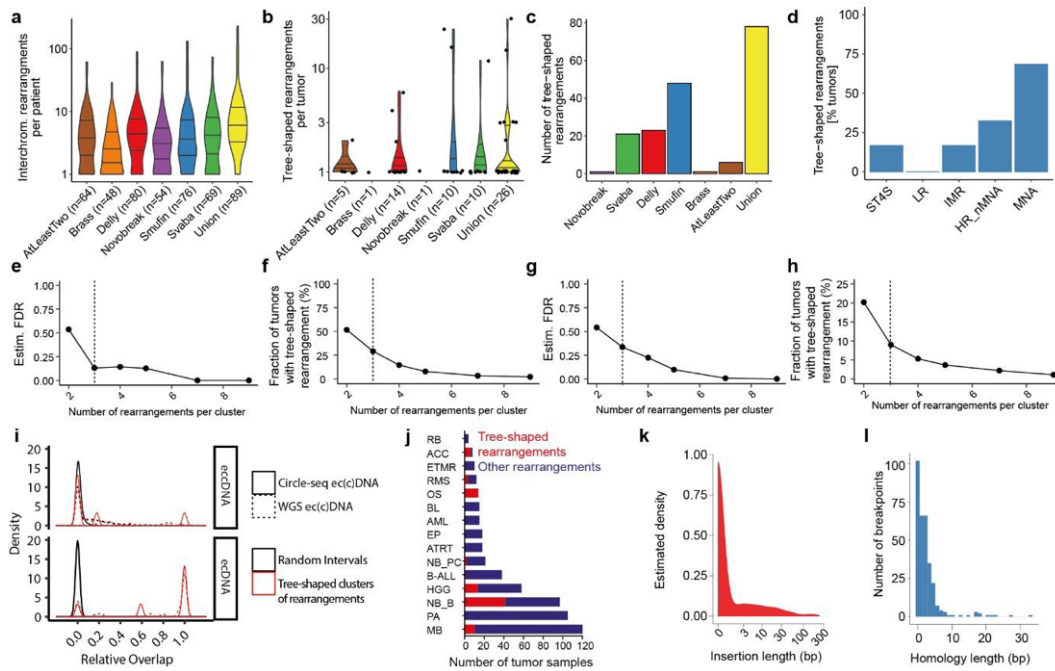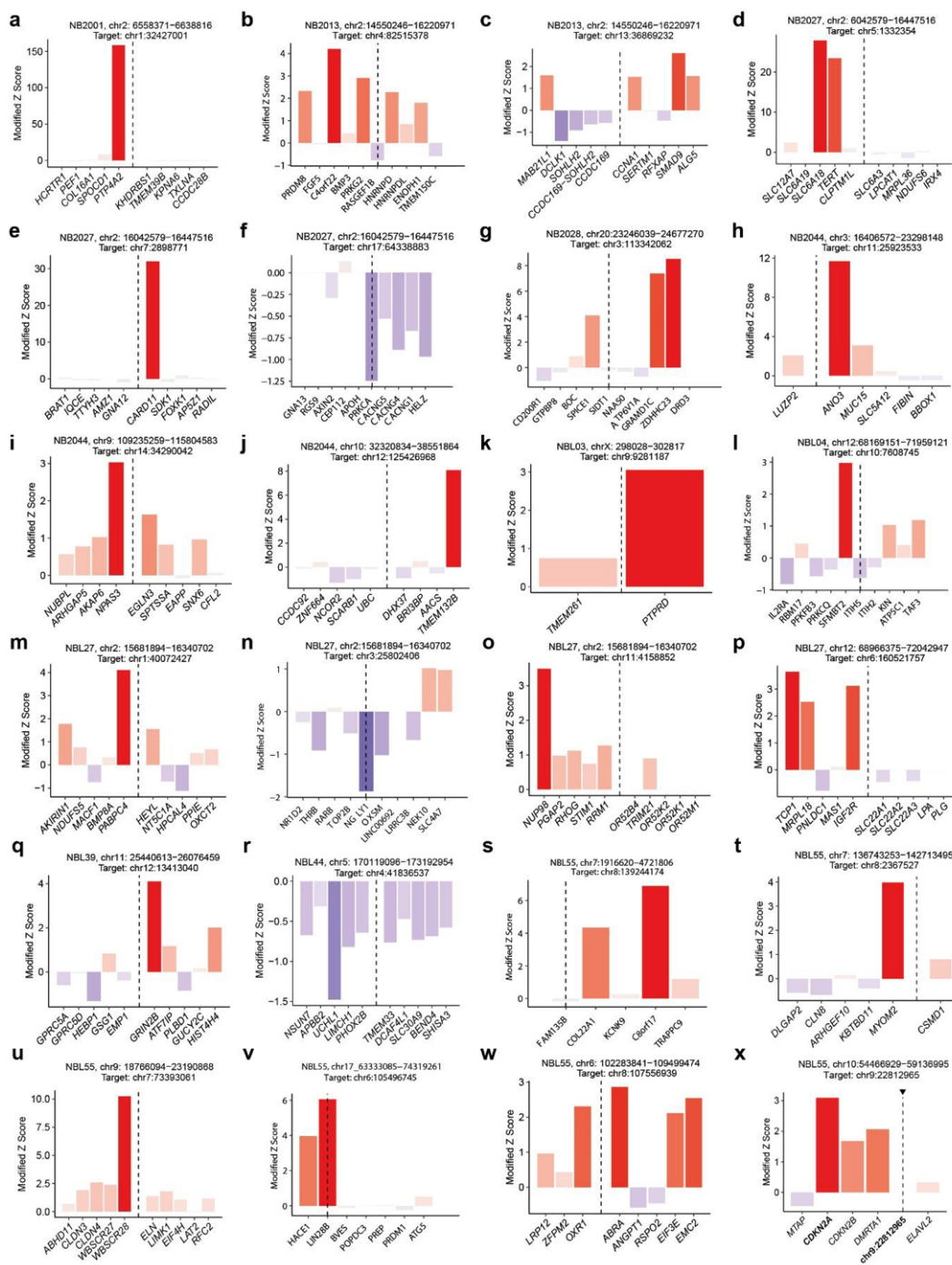
Supplementary Fig. 7. Clustered tree-shaped circle-associated rearrangements can be detected in neuroblastoma genomes. a, Number of inter-chromosomal rearrangements per tumor detected by each structural variant caller, at least two variant callers (AtLeastTwo) compared to the union of rearrangements detected by all variant calling algorithms (Bars indicate maximum, minimum and quartile boundaries). **b,** Number of tree-shaped rearrangement patterns per tumor detected in 91 neuroblastoma genomes by five different variant callers (Bars indicate maximum, minimum and quartile boundaries). **c,** Total number of tree-shaped rearrangements detected in 91 neuroblastoma genomes by five different variant callers. **d,** Fraction of neuroblastomas with at least one tree-shaped rearrangement pattern in different neuroblastoma subgroups (ST4S = stadium 4S, LR = low risk, IMR = intermediate risk, HR_nMNA = high-risk non-*MYCN*-amplified, MNA = *MYCN*-amplified). **e**, Estimated false discovery rate (FDR) for the detection of tree-shaped rearrangement patterns based on the number of rearrangements set as a threshold to define such a pattern in the neuroblastoma cohort (*N*=91). **f**, Fraction of neuroblastoma genomes with tree-shaped rearrangement patterns depending on the number of rearrangements set as a threshold to detect such a pattern in the neuroblastoma cohort (*N*=91). **g**, Estimated FDR for the detection of tree-shaped rearrangement patterns in a publicly available dataset from 546 pediatric cancers comprising 15 types[2]. **h**, Fraction of pediatric cancer genomes with tree-shaped rearrangement patterns depending on the number of rearrangements set as a threshold to detect such a pattern in 546 pediatric cancers. **i,** Relative overlap of circular DNAs (ecDNA and eccDNA) with tree-shaped clustered rearrangements compared to overlap of randomized regions as measured using Circle-seq compared to WGS. All overlaps except eccDNA x Circle-seq are significantly above chance (empirical p-values based on 2000 randomized datasets, one-tailed test, Benjamini-Hochberg-corrected; *N*=6 cluster regions for Circle-seq data, *N*=78 cluster regions for WGS data; *P*=1.0 for Circle-seq eccDNA, *P*=9.995e-4 for Circle-seq ecDNA, *P*=0.0227 for WGS eccDNA and *P*=9.995e-4 for WGS ecDNA). **j,** Frequency of tree-shaped rearrangements indicative of circle-associated rearrangements in our cohort of 91 neuroblastomas (NB_B) and a publicly available dataset from 546 pediatric cancers comprising 15 types[2] (RB = retinoblastoma, ACC = adrenocortical carcinoma, ETMR = embryonal tumor with multilayered rosettes, RMS = rhabdomyosarcoma, OS = osteosarcoma, BL = Burkitt lymphoma, AML = acute myeloid

8

leukemia, EP = ependymoma, ATRT = atypical teratoid rhabdoid tumor, NB_PC = neuroblastoma Heidelberg cohort, B-ALL = B-cell acute lymphoblastic leukemia, HGG = high-grade glioma, PA = pilocytic astrocytoma, MB = medulloblastoma). **k**, Length distribution of sequence insertions at accurately reconstructable breakpoints (N=320) of rearrangement breakpoints. 14.5% of breakpoints showed small insertions of at least 5bp. **l,** Length distribution of homologous sequences at accurately reconstructable rearrangement breakpoints (N=320). Microhomologies of at least 5 bp are found at 10.0% of rearrangement breakpoint junctions. Mean homology length was found to be 5-times longer for real SV than for a set of randomly permuted breakpoint pairs (both groups N=320; Group means 2.47bp vs. 0.49bp; two-sided unequal variances t-test, t=-8.64, df=342.12, P=2.2e-16).

9
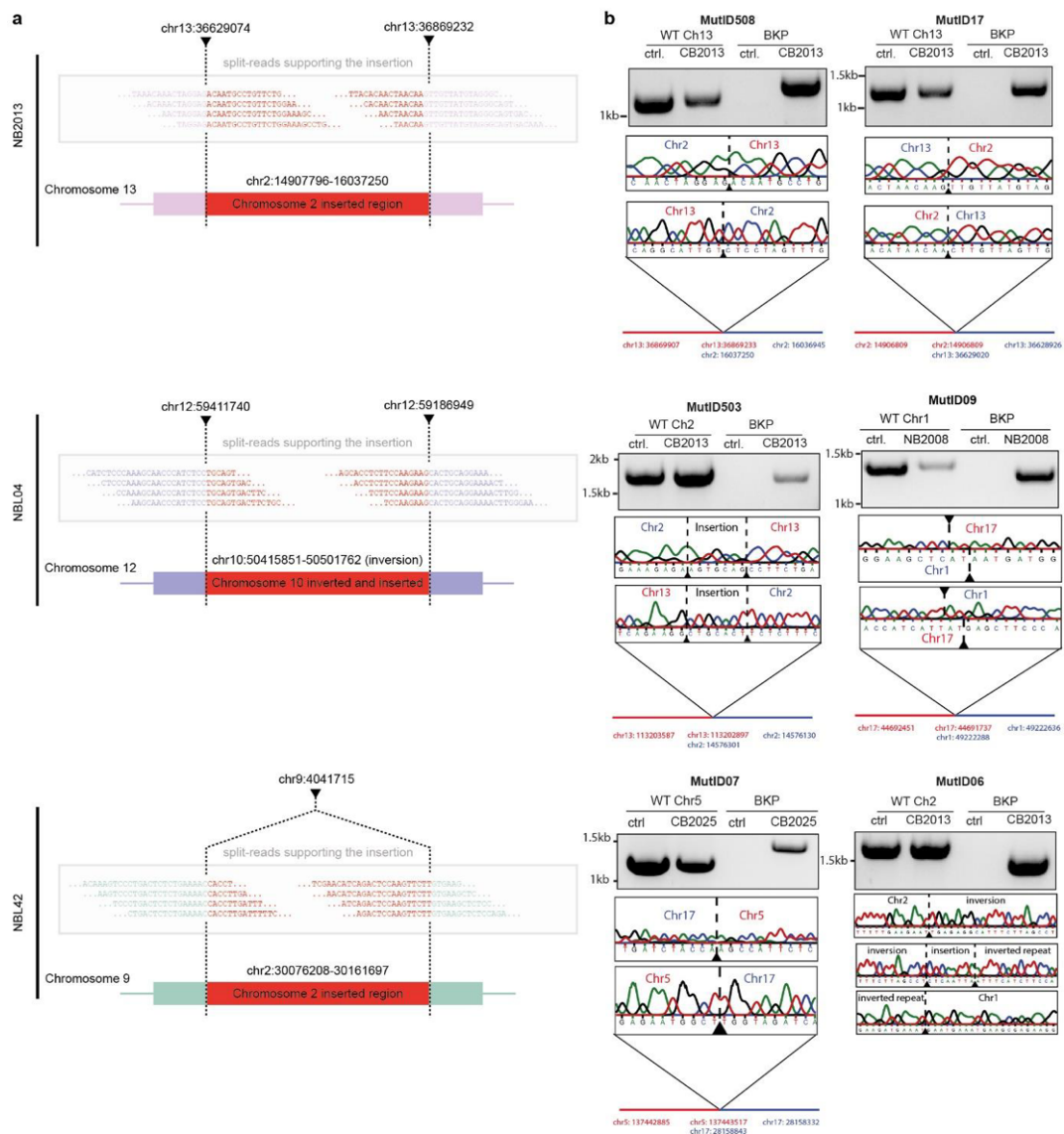
Supplementary Fig. 8. Validation of circle-associated rearrangements by allele-specific PCR and Sanger sequencing. a, Split read support for variant breakpoints of exemplary circle-associated rearrangements. **b,** Validation of circle integration as assessed using variant and wild-type allele-specific PCR in matched tumor and normal primary patient specimens (repeated at least three independent times).

Supplementary Fig. 9. Cancer-driving lesions can emerge out of circle-associated tree-shaped rearrangement clusters. Modified z-scores for the mRNA expression of a subset of breakpoint-neighboring genes at different genomic loci in different neuroblastomas affected by circle-associated tree-shaped rearrangements. The genomic interval indicates the rearrangement cluster. Targets indicate the breakpoint partner connected to the rearrangement cluster **(a-z)**.
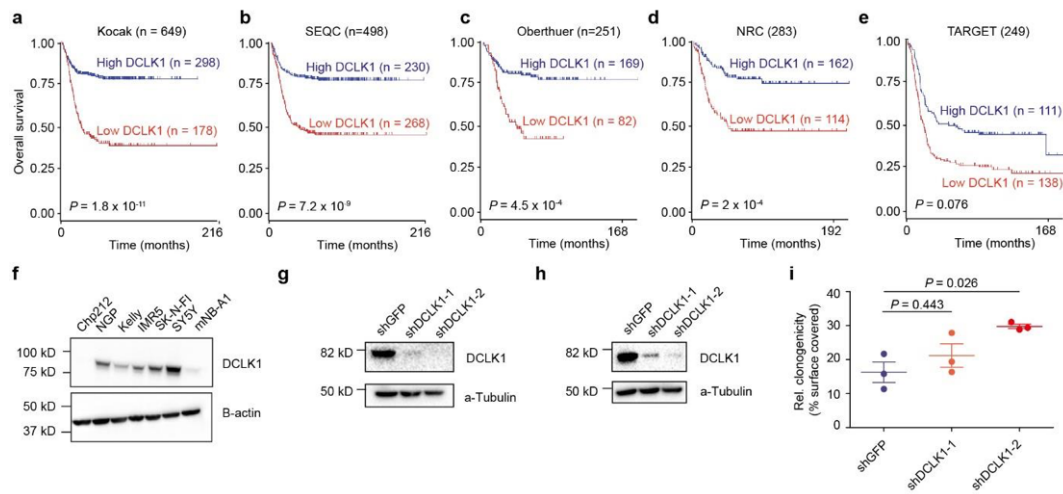
Supplementary Fig. 10. Cancer-driving lesions with clinical relevance can emerge out of circle integration. Low *DCLK1* mRNA expression in neuroblastomas correlates with adverse clinical outcome in five independently published gene expression datasets (statistical difference was calculated using a two-sided log rank test corrected after Bonferroni) **(a-e)**. **f,** DCLK1 protein expression in a panel of 7 neuroblastoma cell lines as measured using western immunoblotting (measured at least three times). DCLK1 protein expression as measured using western immunoblotting after shRNA-mediated knock-down of *DCLK1* in IMR5 (**g**) and Kelly cells (**h**) (repeated independently three times). **i**, Surface area of tissue culture plate covered by cell colonies after shRNA-mediated *DCLK1* knock-down compared to shGFP expressing cells (Center line indicates mean, error bars represent standard deviation of three independent cell culture plates, ANOVA test between groups followed by post-hoc pairwise comparisons using Tukey's Honestly Significant Difference test, *P*=0.443 for shGFP vs. shDCLK1-1, *P*=0.026 for shGFP vs. shDCLK1-2 and *P*=0.136 for shDCLK1-1 vs. shDCLK1-2).
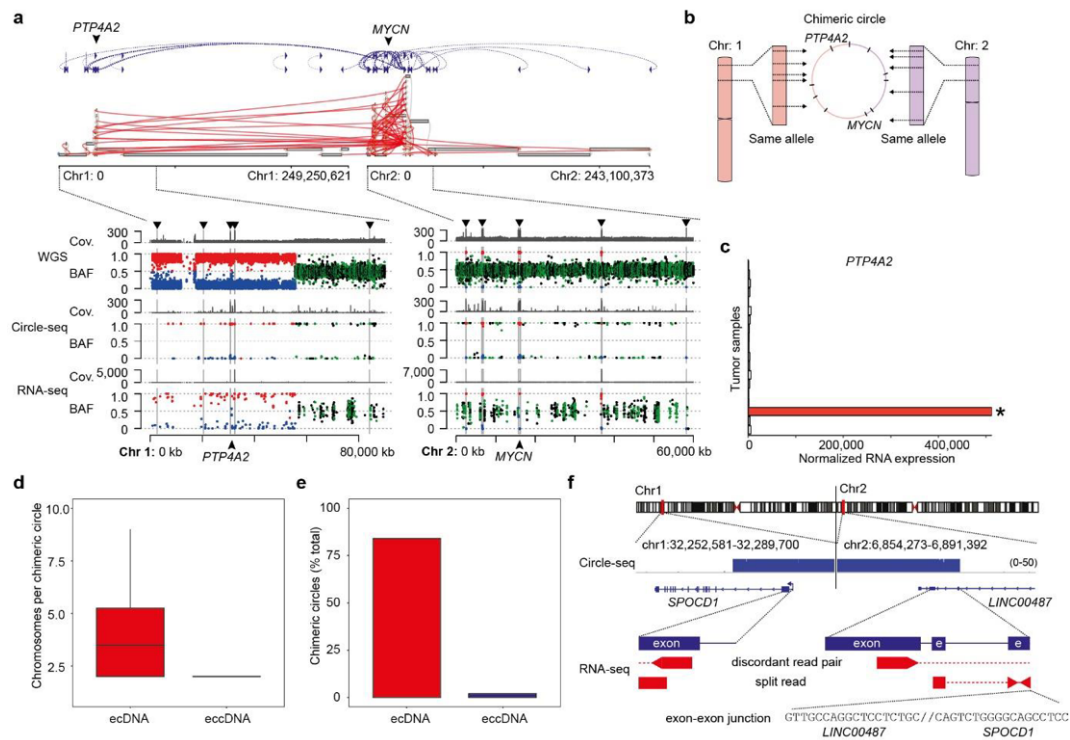
Supplementary Fig. 11. Extrachromosomal chimeric circles can lead to co-amplification and overexpression of oncogenes and expression of aberrant circle-specific fusion transcripts. a, Genome track with genomic copy number alterations at chromosome 1 (top left) and chromosome 2 (top right) connected through formation of a chimeric extrachromosomal circle (top, blue lines). Coverage and B-allelic frequency of reads from regions connected between chromosome 1 and 2 as detected by whole genome sequencing, Circle-seq and RNA sequencing (bottom). **b**, Schematic of chimeric circle formation depicted in **a**. **c**, Normalized gene expression (mRNA) for protein tyrosine phosphatase type IVA, member 2 (*PTP4A2*) in 21 neuroblastomas (tumor affected by chimeric circle shown in (a) is marked by asterisk). **d,** Number of chromosomes included per chimeric extrachromosomal circular DNA (*N*=19 ecDNA; *N*=3,514, eccDNA). **e,** Fraction of extrachromosomal circular DNA that are of chimeric structure (*N*=16/19, ecDNA; *N*=3,514/167,793, eccDNA). **f**, RNA sequencing split reads indicating the expression of an aberrant fusion transcript on the chimeric extrachromosomal circular DNA.

13

107
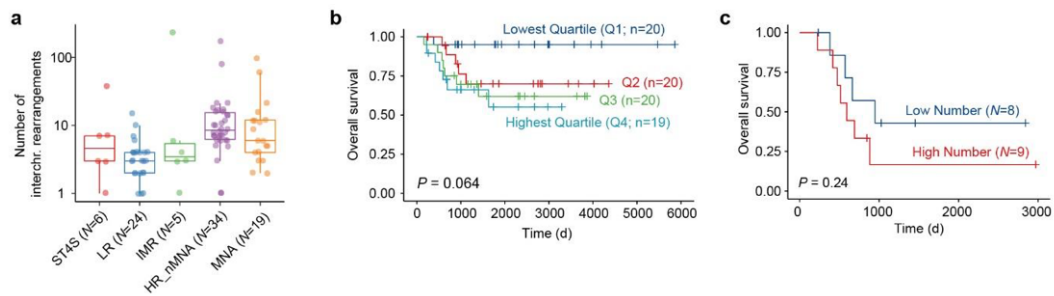
Supplementary Fig. 12. Higher number of inter-chromosomal rearrangements does not distinguish clinically distinct subgroups of *MYCN*-amplified neuroblastoma. a, Number of inter-chromosomal rearrangements in neuroblastoma from different clinical risk groups (Tukey-style boxplots with box encompassing the second and third quartile, whiskers include data points within 1.75 times the interquartile range.). **b**, Kaplan Meier analysis of patient survival comparing patients with neuroblastomas affected by different numbers of somatic inter-chromosomal rearrangements (Q1-4: quartile of numbers of inter-chromosomal rearrangements; two-sided log rank test, P=0.064). **c**, Kaplan Meier analysis comparing patient survival with *MYCN*-amplified neuroblastomas (N=17) and high numbers of inter-chromosomal rearrangements to neuroblastomas with low numbers of inter-chromosomal rearrangements (Low number of rearrangements are defined as numbers below the median; two-sided log-rank test, P=0.24).

Supplementary Table 1. Analysis of the *MYCN* amplicon state (homogenously stating region, HSR, extrachromosomal circular double minute chromosome, Dmin) as measured using FISH and interpreted by the German clinical diagnostic reference center and tumor bank for neuroblastoma.

Sample ID	MYCN Amplification[4]	Amplicon state	MYCN-associated clustered rearrangements
CB2013	Yes	HSR/Dmin	Yes
CB2022	Yes	Dmin	Yes
CB2025	Yes	HSR/Dmin	Yes
CB2027	Yes	Dmin	Yes
CB2034	Yes	Dmin	Yes
CB2030	Yes	Dmin	Yes
CB2050	Yes	Dmin	Yes
NBL26	Yes	Dmin	Yes
NBL27	Yes	Dmin	Yes
NBL29	Yes	Dmin	Yes
NBL30	Yes	Dmin	Yes
NBL33	Yes	Dmin	Yes
NB2001	Yes	Dmin	No
NB2024	Yes	Dmin	No
NBL28	Yes	Dmin	No
NBL31	Yes	Dmin	No
NBL32	Yes	Dmin	No
NBL34	Yes	Dmin	No
NBL42	Yes	Dmin	No

Supplementary Table 2. Sequences of oligonucleotide primers used for PCR and Sanger DNA sequencing

Name	Description	Sequence (5'-3')
MYCN_1_F1	PCR of circle junction in KELLY	TGGCTAGCAGAAACTCACCAG
MYCN_1_R3	PCR of circle junction in KELLY	CAGGGGGCATTTACTTTCAA
MYCN_1_WT_R1	PCR of wild type locus in KELLY	GCACAGAGGGACCAGCTTAC
MYCN_5_F1	PCR of circle junction in NGP	CTGTCCTGCCTCTTTTCTGG
MYCN_5_R1	PCR of wild-type locus and circle junction in NGP	GCGTCAGTTAGGGCATTTTC
MYCN_5_WT_F2	PCR of wild type locus in NGP	GCCAGTGCTGAAAAAGGAAG
MYCN_2_F2	PCR of circle junction in CB2050	TCACTCCAGTCCCAGGTCTG
MYCN_2_R3	PCR of circle junction in CB2050	CCCGGGTCTGTGATGAAGAG
MYCN_2_WT_R1	PCR of wild type locus in CB2050	GTCCTCCTCTTACTACCTTGGC
JUN_2_F2	PCR of circle junction in CB2052	CGCTAGCTCTGGGCAGTTAG
JUN_2_R1	PCR of circle junction in CB2052	TTTCGGGAGTGTCCAGAGAG
JUN_2_WT_R1	PCR of wild type locus in CB2052	CAGCCACCGTCACTAGACAG
MYCN_8_WT_F1	PCR of wild type locus in CB2013	TCCTCACAACCCAGTACTCA
MYCN_8_ F2	PCR of circle junction in CB2013	ACATCAAACTCAGGGCCAAC
MYCN_8_ R1	PCR of wild type locus and circle junction in CB2013	TACAAAAGGACGGACACGGAG
Beta_Globin_F	qPCR of human *HBB* on gDNA	TATTGGTCTCCTTAAACCTGTCTTG
Beta_Globin_R	qPCR of human *HBB* on gDNA	CTGACACAACTGTGTTCACTAGC
MutID_503_Ch2_BKP1_F2	PCR of wild type locus and translocation in CB2013	CCTCTCTCCAGTCAAGCCAC
MutID_503_Ch2_BKP1_R2	PCR of wild type locus only in CB2013	CTCTTGGGTGTGTAGGGGTATC
MutID_503_Ch13_BKP2_R2	PCR of translocation only in CB2013	CGGTATCCAAGTTCCCCACC
MutID_06_Ch2_BKP2_F2	PCR of wild type locus only in CB2013	TCAGGCTTTGGACAGGAGTG
MutID_06_Ch2_BKP2_R2	PCR of wild type locus and translocation in CB2013	ACCTGTACCTGAACACAGGC
MutID_06_Ch1_BKP1_R2	PCR of translocation only in CB2013	TGGCTGGTACCGGTTGTTC

MutID_17_Ch13_BKP1_F1	PCR of wild type locus and translocation in CB2013	CCTGACACAAAGGGAAACAGC
MutID_17_Ch13_BKP1_R1	PCR of wild type locus only in CB2013	GTGTTCTCCAGGAACCCGTC
MutID_17_Ch2_BKP2_F1	PCR of translocation only in CB2013	CTAGGGCGTCCAGTATCTTGC
MutID_508_Ch13_BKP2_F1	PCR of wild type locus only in CB2013	GGCCTTACTTTTGGCCCACTA
MutID_508_Ch13_BKP2_R1	PCR of wild type locus and translocation in CB2013	TACGGAACTGGAAGCCAAAC
MutID_508_Ch2_BKP1_F1	PCR of translocation only in CB2013	GGTGGCCAACGATTACCAAG
MYCN_9_F1	PCR of wild type locus in VM1 patient sample	CATGGGTCACACTCTCAACC
MYCN_9_WT_R1	PCR of wild type locus in VM1 patient sample	TCTGCCTCCATCTGCGTATG
MYCN_9_F2	PCR of circle junction in VM1 patient sample	TCACTTCTAGGAGCAGCACC
MYCN_9_R1	PCR of circle junction in VM1 patient sample	CCACGGTCTCCCTCTGATAC
MYCN_11_WT_F2	PCR of wild type locus in NB2022	CCCGTCTTCCCCTCTAGAAC
MYCN_11_R1	PCR of wild type locus and circle junction in NB2022	TTTTGCTAGGGGACGGTAGG
MYCN_11_F1	PCR of circle junction in NB2022	TCATACCAGGGCCCACTTAAG
18S_rRNA_F	Control PCR for ribodepletion	GTAACCCGTTGAACCCCATT
18S_rRNA_R	Control PCR for ribodepletion	CCATCCAATCGGTAGTAGCG
28S_rRNA_F	Control PCR for ribodepletion	GTGTTAGGACCCGAAAGATGGT
28S_rRNA_R	Control PCR for ribodepletion	TAGTCTTTCGCCCCTATACCCA
ActinB_F	Control PCR for ribodepletion	CCAACCGCGAGAAGATGA
ActinB_R	Control PCR for ribodepletion	CCAGAGGCGTACAGGGATAG

Supplementary Table 3. Annotation of cell lines analyzed with Circle-seq

Cell line	Species	Cell Type	MYCN Amplification[1]	Number of circular DNAs
LAN-1	Human	Neuroblastoma	Yes	864
KELLY	Human	Neuroblastoma	Yes	3302
SK-N-SH	Human	Neuroblastoma	No	1617
IMR-5	Human	Neuroblastoma	Yes	2629
SK-N-FI	Human	Neuroblastoma	No	1471
SH-SY5Y	Human	Neuroblastoma	No	14848
SK-N-AS	Human	Neuroblastoma	No	209
GI-ME-N	Human	Neuroblastoma	No	2124
NB69	Human	Neuroblastoma	No	260
CHP-212	Human	Neuroblastoma	Yes	2701
SH-EP	Human	Neuroblastoma	No	22804
NGP	Human	Neuroblastoma	Yes	9019

[1]MYCN status assessed by FISH

18

Supplementary Table 4. Clinical annotation of patient samples analyzed with Circle-seq

Sample ID	INSS[1]	Age[2]	INRG[3]	MYCN Amplification[4]	Number of circular DNAs
CB1008	IV	2	High	Yes	23753
CB1010	IV	6	High	No	41371
CB1001	III	4	High	Yes	19574
CB1003	IV	3	High	Yes	16865
CB2051	*na*	*na*	Low	No	4923
CB2052	*na*	*na*	Low	No	42749
CB2001	IV	2	High	Yes	33718
CB2050	IV	2	High	Yes	16719
CB2025	III	2	High	Yes	5348
CB2026	IV	5	High	No	7135
CB2011	IVS	<1	Low	No	11223
CB2013	IV	1	High	Yes	7221
CB2014	I	<1	Low	No	10671
CB2022	IV	2	High	Yes	36636
CB2036	IV	2	High	No	12785
CB2006	IVS	<1	Low	No	20932
CB2049	IV		High	No	40286
CB2004	I	<1	Low	No	21278
CB2007	III	2	Intermediate	No	34446
CB2008	III	2	Intermediate	No	15324
CB2056	II		Low	No	36374

[1]INSS: International Neuroblastoma Staging System

[2]Age in years

[4]MYCN status assessed by FISH

[3]International Neuroblastoma Risk Group

na: not available

Supplementary Note

Analysis of published gene expression arrays.

The R2 visualization and analysis platform (http://r2.amc.nl/) was used to reanalyze published gene expression and clinical data from neuroblastoma patients[3-6].

Whole genome sequencing

Whole-genome sequencing of 37 matched neuroblastoma tumor-normal pairs was performed on the Illumina HiSeq X-Ten platform with a paired-end read length of 2 × 150bp at the sequencing facility of the German Cancer Research Center (DKFZ, Heidelberg, Germany). Whole genome sequencing of 56 matched tumor-normal pairs was downloaded (https://www.ebi.ac.uk/ega/) under accession number EGAS00001001308[7]. The quality of the raw data was assured using *FastQC*[8]. The reads were aligned to the human genome reference, assembly hg19, using BWA-MEM[9]. Raw sequencing reads were aligned to the human genome (NCBI build 37/hg19) by using BWA-MEM mem version 0.7.15 and BWA-ALN version 0.5.9, and BAM files were generated using SAMtools version 1.9[9]. Optical and PCR duplicates were marked using bammarkduplicates2 from Biobambam version 2.0.79.

RNA sequencing

RNA of 37 primary neuroblastomas was isolated with Trizol according to the manufacturer's protocol (Thermo Fisher). RNA purity was analyzed on a Nanodrop 2000 spectrometer and RNA integrity assessed on a Bioanalyzer 2100 or TapeStation4200 as per manufacture's instructions. Only samples with an RNA integrity number of 8 or above were included. Depletion of ribosomal RNA (rRNA) was performed by enzymatic digestion as previously described[10]. Ribo-depletion of 95-99% was confirmed using RT-qPCR. The ribo-depleted RNA was used for generation of RNA sequencing libraries using the TrueSeq Stranded mRNA kit according to the manufacturer's protocol (Illumina). Total RNA of 37 primary tumor samples was sequenced on a HiSeq4000 (Illumina, San Diego, USA) with a paired-end read length of 2 × 150bp at the sequencing core facility of the Max Delbrueck Center for Molecular Medicine (MDC, Berlin, Germany). On average 100 million reads were generated per sample. The raw RNA sequencing data of 54 previously published neuroblastomas were downloaded from the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/) under accession number EGAS00001001308[11]. The quality of the raw data was assured using *FastQC*[8]. The reads were aligned to the human genome reference, with the annotation of known transcripts taken from

Gencode (v27)[11], using the *STAR* aligner[12]. Read counts per gene were determined using the *featureCounts* software[13]. Raw read counts were normalized and further analyzed using the *DESeq2* R/Bioconductor package[14].

Single nucleotide polymorphism (SNP) calling and phasing

Single nucleotide polymorphism (SNP) detection and phasing was performed similarly as previously described[15]. For haplotype-specific analyses, we considered 84,801,880 germline variants as reported by 1000 Genomes[16] phase 3. Only bi-allelic SNPs on chromosomes 1-22 with minor allele frequency (MAF) $\geqq$ 1% were retained for downstream analyses. SNPs were additionally filtered for mappability by excluding positions with score less than one in the UCSC HG19 CRG Align 50 score[17], yielding a total of 9,866,569 variant sites after filtering. Allele-specific read counts and B-allele frequency (BAF) ratios were determined for each variant in each sample using bcftools 1.8 (http://www.htslib.org/doc/bcftools.html). SNPs were genotyped on WGS samples from normal tissue using the multiallelic-caller from bcftools 1.8. Genotypes were then merged and phased by Eagle2 (version 2.4)[18] using phased 1000 Genomes Project genotypes as reference.

Allele-specific expression analysis

Allele-specific RNA read counts were determined by GATK[19] (version 3.5.0) ASEReadCounter from RNA-seq alignments at heterozygous SNPs following established protocols[20]. Sites (i.e. SNP in a sample) with less than 8 total or less than 2 allelic reads were removed. Additionally, only sites that qualified as bi-allelic according to a statistical test were retained: A binomial test on the minimum allele count = min(alt, ref) with probability sum(non_ref_alt)/sum(raw_depth) was applied, where ref and alt are the reference and alternative allele counts, and non_ref_alt and raw_depth the non-reference/non-alternative allele count and raw read depth per site respectively. Sites for which the null hypothesis was rejected (FDR 0.05, Benjamini-Hochberg) were classified as bi-allelic. The reference allele bias was estimated by averaging over the reference allele fraction ref / (ref + alt) of all ASE sites from balanced copy-number regions per sample. Allelic-expression imbalance (AEI) was assessed through a binomial test against the calculated reference bias at 5% FDR after Benjamini-Hochberg correction for multiple testing. The ASE ratio was calculated as max(ref, alt) / (ref + alt). Allelic expression preference of mono-allelic circles in copy-number balanced regions was determined by a statistical test on the circles' ASE haplotype state. Circles in copy-number imbalanced regions (see Methods, Allele-specific-copy-number analysis) were

21

removed. From the remaining circles only mono-allelic circles were retained, which were defined as circles with Circle-seq maximum haplotype frequency > 0.9 (see Methods, Assignment of CN states to circles). Circles with identical RNA counts for both haplotypes were removed and remaining circles were annotated with two different states dependent on whether the majority of ASE counts came from the circularized haplotype or not. Finally a binomial test on the circles was conducted, parameterized for equal probability to draw a circle from one of the two states under the null hypothesis. Read ratios were computed for each heterozygous SNP and corrected for reference bias. A coverage log ratio per SNP was calculated between tumor and normal WGS and normalized by total SNP coverage per assay.

Allele-specific copy-number analysis

Allele-specific read counts and BAF ratios were determined for variants in each sample using bcftools 1.8 (http://www.htslib.org/doc/bcftools.html) mpileup. Allele-specific copy-number (CN) profiles and estimates for purity and ploidy of tumors were obtained from BAFs and log ratios of coverage between tumor and normal samples (logR) using a modified version of ASCAT 2.5[21]. We only considered SNPs with a minimum total coverage of 10 reads from tumor and normal WGS samples. All CN segments were inspected manually for quality. For samples with estimated tumor purity less than 60% (CB2006, CB2007 and CB2014) CN calling was rerun with manually selected purity and ploidy values based on inspection of the goodness-of-fit plots and in agreement with pathology. Copy-number segments were assigned to categories as follows: Balanced: total copy-number $\geqq 0$ and majorCN = minorCN, where majorCN and minorCN are the copy-numbers of major- and minor allele respectively; Weak imbalance, total copy-number $\geqq 0$ and copy-number ratio = majorCN / (majorCN + minorCN) $\leqq \frac{2}{3}$; Strong imbalance: as weak imbalance, but copy-number ratio > $\frac{2}{3}$; LOH: minorCN = 0 and majorCN > 0; Focal amplification: copy-number segment smaller than 3 Mb , logr_seg > 0.8 and median(logr_seg) - median(logr_chr) > 0.8, where logr_seg and logr_chr are coverage log ratios of SNPs on the segment and its chromosome of origin respectively.

Assignment of CN states to circles

The copy-number state of a circle was defined as the state of the copy-number segment with the largest overlap. The phase of SNPs overlapping imbalanced copy-number segments (majorCN > minorCN) was defined such that haplotype 1 was assigned to the minor allele and haplotype 2 to the major allele according to the tumor WGS BAF at that SNP. Only heterozygous SNPs with a Circle-seq coverage of more than 10 reads and circles with at least

one SNP fulfilling this requirement were included in the allele-specific analysis. Circle-seq haplotype counts were defined as the sum over allelic-depths of the same haplotype for SNPs overlapping the circle. The haplotype frequency was calculated as hc2 / (hc1 + hc2), where hc1 and hc2 are haplotype counts of haplotype 1 and 2 respectively. The maximum haplotype frequency per circle was calculated as max(hc1, hc2) / (hc1 + hc2).

Supplement-only references

1   MacArthur, I., Koche, R., Dorado-García, H. & Henssen, A. G. Purification and Sequencing of Large Circular DNA from Human Cells. *Nature Protocol Exchange*, doi:10.1038/protex.2019.006 (2019).
2   Grobner, S. N. *et al*. The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321-327, doi:10.1038/nature25480 (2018).
3   Kocak, H. *et al*. Hox-C9 activates the intrinsic pathway of apoptosis and is associated with spontaneous regression in neuroblastoma. *Cell Death Dis* **4**, e586, doi:10.1038/cddis.2013.84 (2013).
4   Consortium, S. M.-I. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol* **32**, 903-914, doi:10.1038/nbt.2957 (2014).
5   Oberthuer, A. *et al*. Customized oligonucleotide microarray gene expression-based classification of neuroblastoma patients outperforms current clinical risk stratification. *J Clin Oncol* **24**, 5070-5078, doi:10.1200/JCO.2006.06.1879 (2006).
6   Pugh, T. J. *et al*. The genetic landscape of high-risk neuroblastoma. *Nat Genet* **45**, 279-284, doi:10.1038/ng.2529 (2013).
7   Peifer, M. *et al*. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526**, 700-704, doi:10.1038/nature14980 (2015).
8   Andrews, S. FastQC: a quality control tool for high throughput sequence data. *Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc* (2010).
9   Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997* (2013).
10   Adiconis, X. *et al*. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods* **10**, 623-629, doi:10.1038/nmeth.2483 (2013).
11   Harrow, J. *et al*. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-1774, doi:10.1101/gr.135350.111 (2012).
12   Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
13   Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
14   Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
15   Jamal-Hanjani, M. *et al*. Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med* **376**, 2109-2121, doi:10.1056/NEJMoa1616288 (2017).
16   Genomes Project, C. *et al*. A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

23

17   Haeussler, M. *et al.* The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res* **47**, D853-D858, doi:10.1093/nar/gky1095 (2019).

18   Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448, doi:10.1038/ng.3679 (2016).

19   McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

20   Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol* **16**, 195, doi:10.1186/s13059-015-0762-6 (2015).

21   Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A* **107**, 16910-16915, doi:10.1073/pnas.1009843107 (2010).

24

7.3 Hung et al. ecDNA hubs drive cooperative intermolecular oncogene expression

The following section contains a reproduction of the article Hung KL, Yost KE, Xie L, et al. ecDNA hubs drive cooperative intermolecular oncogene expression. Nature. 2021;600:731-6. https://doi.org/10.1038/s41586-021-04116-8. Reproduced with permission from Springer Nature.

Journal Data Filtered By: **Selected JCR Year: 2019** Selected Editions: SCIE,SSCI
Selected Categories: **"MULTIDISCIPLINARY SCIENCES"** Selected Category
Scheme: WoS
Gesamtanzahl: 71 Journale

| Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NATURE	767,209	42.778	1.216730
2	SCIENCE	699,842	41.845	1.022660
3	National Science Review	2,775	16.693	0.009760
4	Science Advances	36,380	13.116	0.172060
5	Nature Human Behaviour	2,457	12.282	0.014190
6	Nature Communications	312,599	12.121	1.259510
7	Science Bulletin	5,172	9.511	0.014150
8	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	676,425	9.412	0.931890
9	Journal of Advanced Research	3,564	6.992	0.005470
10	GigaScience	4,068	5.993	0.016410
11	Scientific Data	5,761	5.541	0.028720
12	Research Synthesis Methods	2,572	5.299	0.006440
13	ANNALS OF THE NEW YORK ACADEMY OF SCIENCES	45,596	4.728	0.026370
14	FRACTALS-COMPLEX GEOMETRY PATTERNS AND SCALING IN NATURE AND SOCIETY	2,156	4.536	0.002210
15	iScience	1,410	4.447	0.004140
16	GLOBAL CHALLENGES	481	4.306	0.001440
17	Scientific Reports	386,848	3.998	1.231180
18	JOURNAL OF KING SAUD UNIVERSITY SCIENCE	1,640	3.819	0.002020
19	Journal of the Royal Society Interface	13,762	3.748	0.027670

Selected JCR Year: 2019; Selected Categories: "MULTIDISCIPLINARY SCIENCES"

Article

ecDNA hubs drive cooperative intermolecular oncogene expression
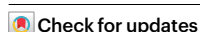
King L. Hung[1,22], Kathryn E. Yost[1,22], Liangqi Xie[2,3,4,22], Quanming Shi[1], Konstantin Helmsauer[5], Jens Luebeck[6,7], Robert Schöpflin[8,9,10], Joshua T. Lange[11,12,13], Rocío Chamorro González[5], Natasha E. Weiser[1,13], Celine Chen[5], Maria E. Valieva[8,9], Ivy Tsz-Lo Wong[12,13], Sihan 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
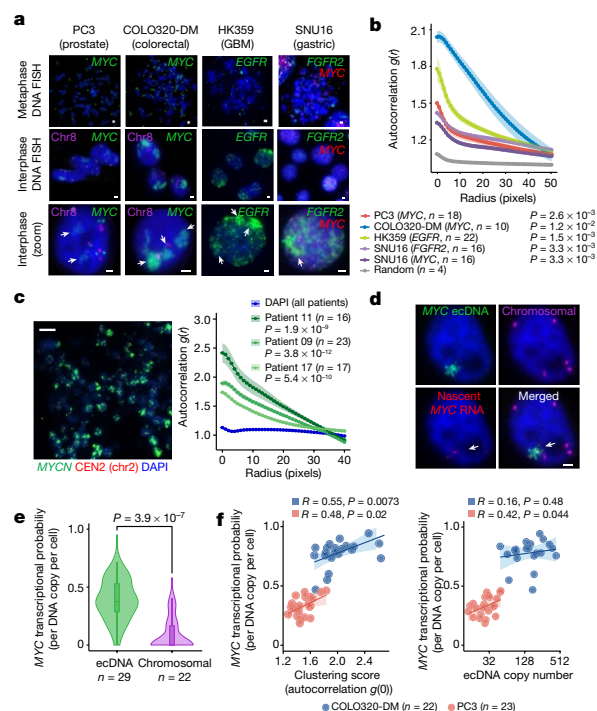
Fig. 1 | ecDNA imaging correlates ecDNA clustering with transcriptional bursting. a, Representative FISH images of interphase ecDNA clustering. A chromosomal control was included for PC3 and COLO320-DM. GBM, glioblastoma. Scale bars, 2 μm. **b**, Interphase ecDNA clustering by autocorrelation $g(r)$ (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).

121
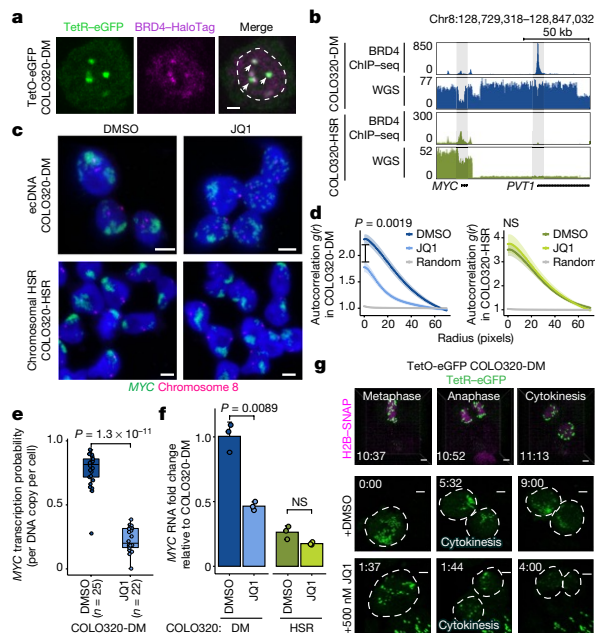
Fig. 2 | BET proteins mediate ecDNA hub formation and transcription.
a, Representative live-cell images of ecDNA and BRD4–HaloTag signals in TetO–eGFP COLO320-DM cells (independently repeated twice; dashed line indicates nuclear boundary). Scale bar, 2 μm. **b**, BRD4 ChIP–seq and WGS at the *MYC* locus 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).

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
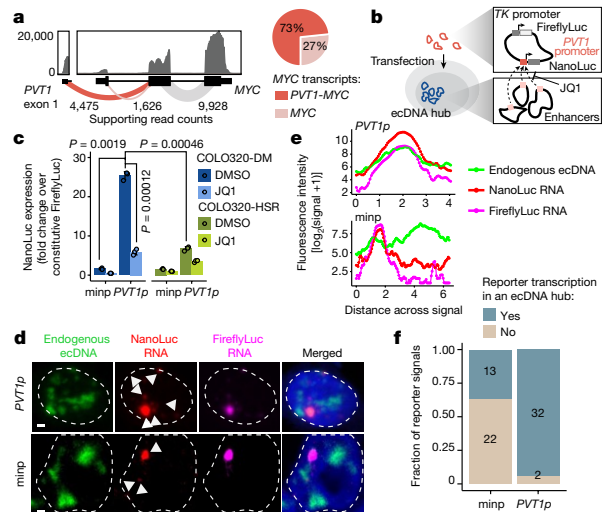

Fig. 3 | Intermolecular activation of an episomal luciferase reporter in ecDNA hubs. a, Left, bulk RNA-seq from COLO320-DM cells with exon–exon junction spanning read counts shown. Right, relative abundance of full-length *MYC* and fusion *PVT1-MYC* transcripts using read count supporting either junction. **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.

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

122
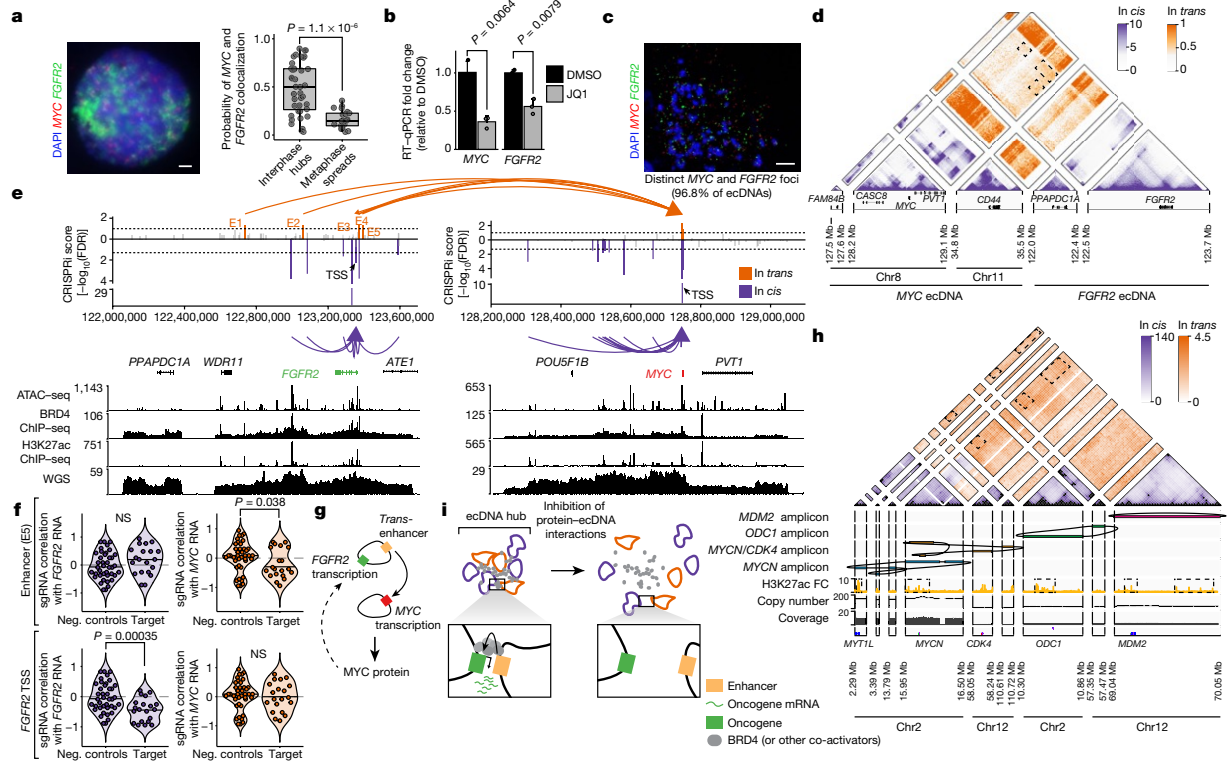
Fig. 4 | ecDNA hubs mediate intermolecular enhancer–gene interactions.
a, Left, representative DNA FISH image showing clustering of *MYC* and *FGFR2* ecDNAs in interphase SNU16 cells. Scale bar, 2 μm. Right, *MYC* and *FGFR2* colocalization in SNU16 cells (box plot parameters as in Fig. 1). *P* value 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).

124

Article

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).

125

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

Article

(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 ($\sigma = 1$ 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 800g 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 1,350g 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 1,350g 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 16,000g 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.

Article

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 $500g$ 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).

Article

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 Normalize-Data 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, get-MatrixFromProject 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 (GTACCTGAGCTCGCTAGCCTCGAGA-AGATCTGCGTACGGTCGAC), 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 (Bio-search 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 800g 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 log₂-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 log₂-transformed 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 (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 log₂-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

Article

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 flow-cell (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.

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).

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.
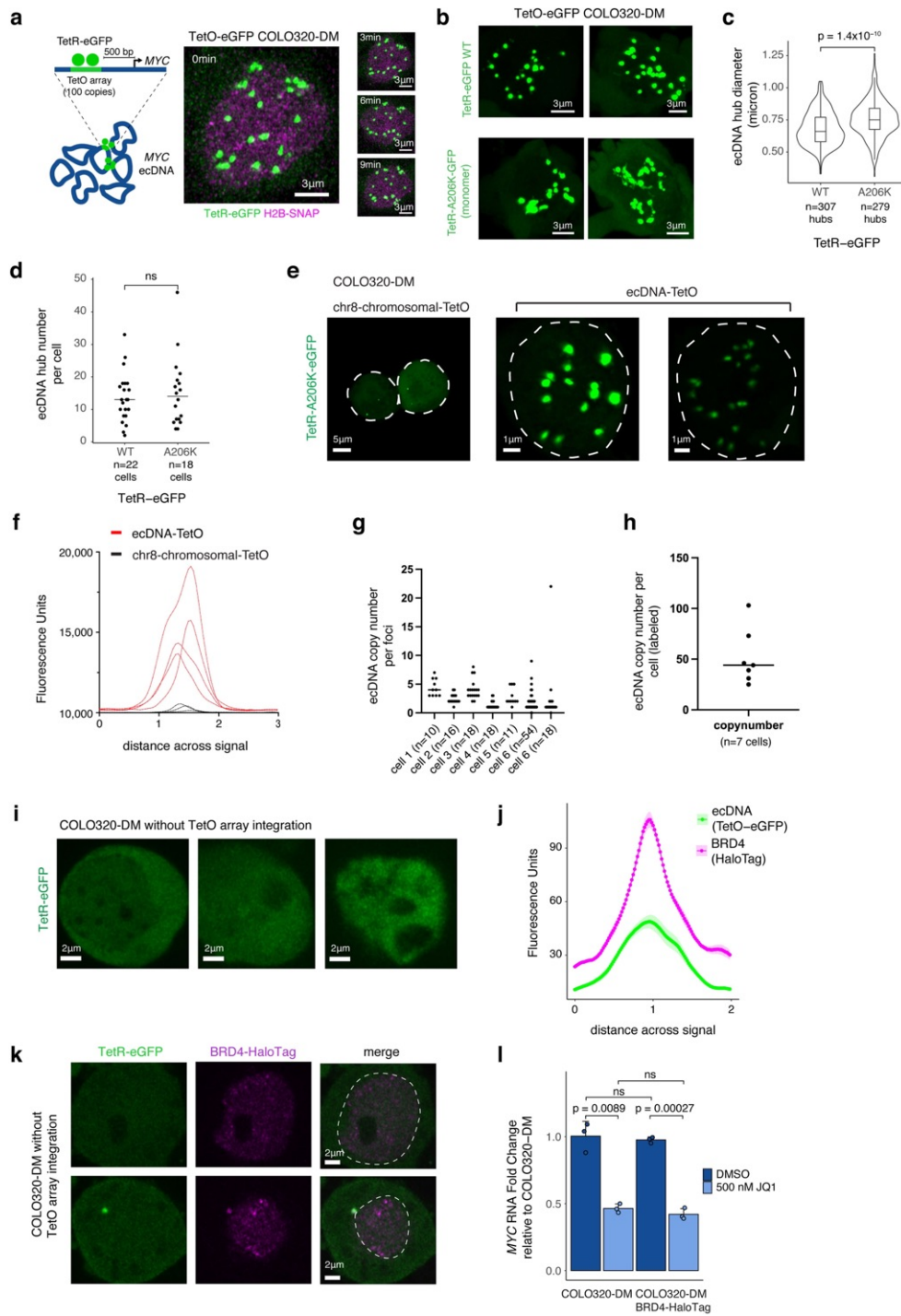**Reprints and permissions information** is available at http://www.nature.com/reprints.
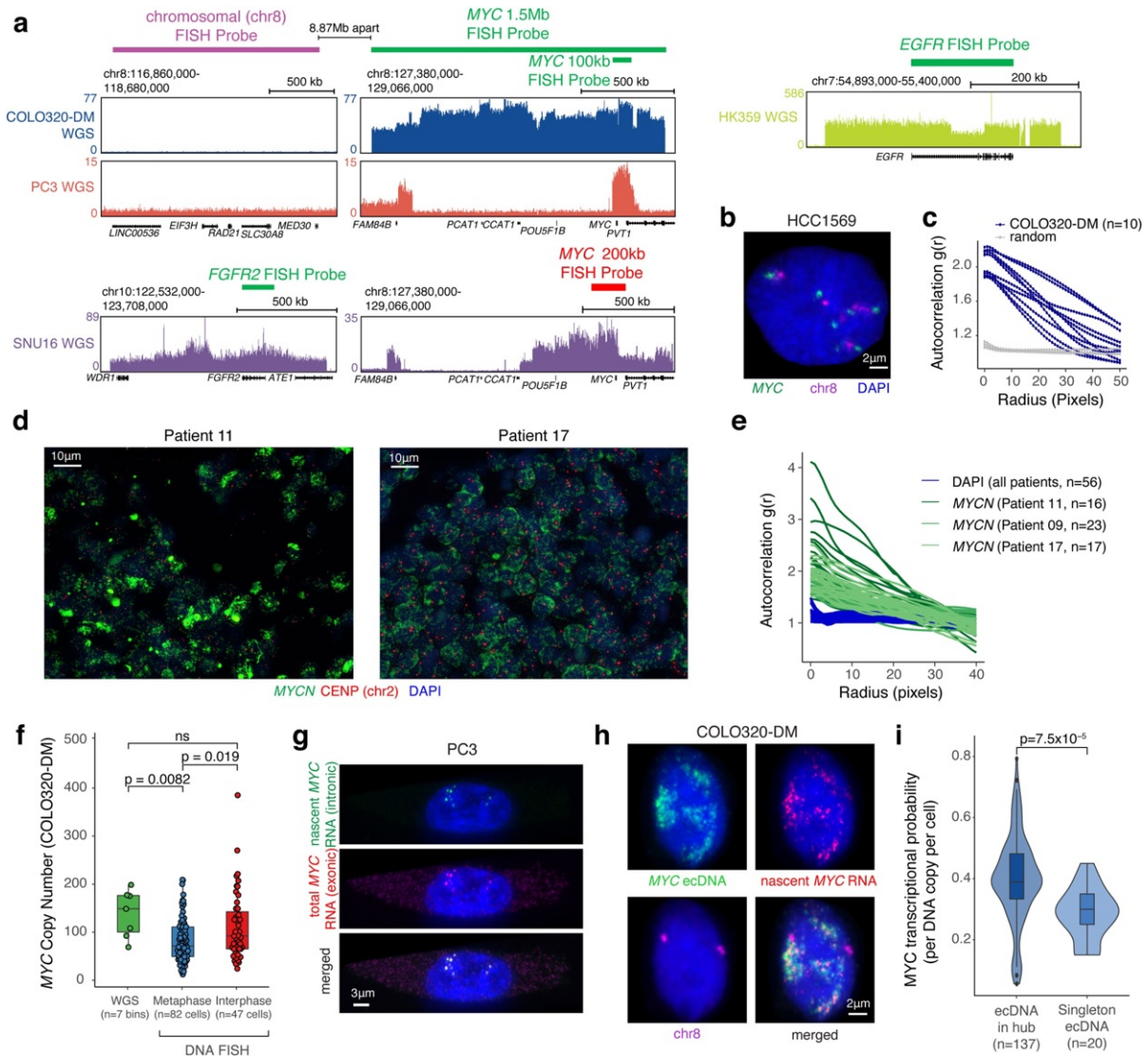
Extended Data Fig. 1 | ecDNA FISH strategies and copy number estimation. a, WGS tracks with DNA FISH probe locations. For COLO320-DM and PC3, a 1.5 Mb *MYC* FISH probe (Fig. 1a, b), a 100 kb *MYC* FISH probe (Fig. 1d–f), or a 1.5 Mb chromosome 8 FISH probe was used. Commercial probes were used in SNU16 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.

a TetR-eGFP, 500 bp, MYC, TetO array (~100 copies), MYC ecDNA

TetO-eGFP COLO320-DM

0min, 3min, 6min, 9min

TetR-eGFP H2B-SNAP

b TetO-eGFP COLO320-DM

TetR-eGFP WT

TetR-A206K-GFP (monomer)

c p = 1.4×10^-10

ecDNA hub diameter (micron)

WT n=307 hubs, A206K n=279 hubs

TetR−eGFP

d ns

ecDNA hub number per cell

WT n=22 cells, A206K n=18 cells

TetR−eGFP

e COLO320-DM

chr8-chromosomal-TetO, ecDNA-TetO

TetR-A206K-eGFP

f ecDNA-TetO, chr8-chromosomal-TetO

Fluorescence Units

distance across signal

g ecDNA copy number per foci

cell 1 (n=10), cell 2 (n=16), cell 3 (n=18), cell 4 (n=18), cell 5 (n=11), cell 6 (n=54), cell 6 (n=18)

h ecDNA copy number per cell (labeled)

copynumber (n=7 cells)

i COLO320-DM without TetO array integration

TetR-eGFP

j ecDNA (TetO−eGFP), BRD4 (HaloTag)

Fluorescence Units

distance across signal

k TetR-eGFP, BRD4-HaloTag, merge

COLO320-DM without TetO array integration

l ns, ns

MYC RNA Fold Change relative to COLO320-DM

p = 0.0089, p = 0.00027

DMSO, 500 nM JQ1

COLO320-DM, COLO320-DM BRD4-HaloTag

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

136

Article

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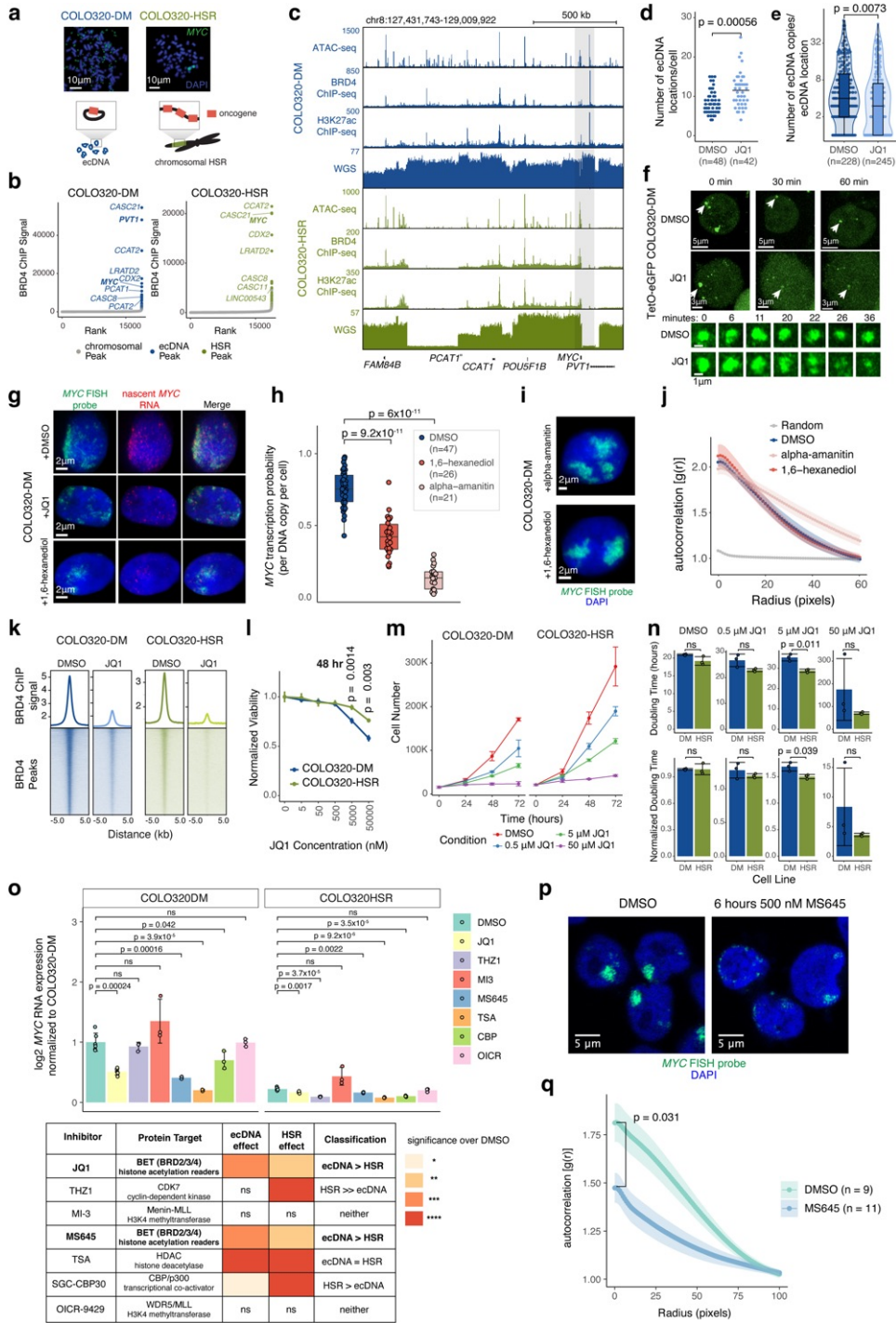
Extended Data Fig. 3 | See next page for caption.

Article

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.

139
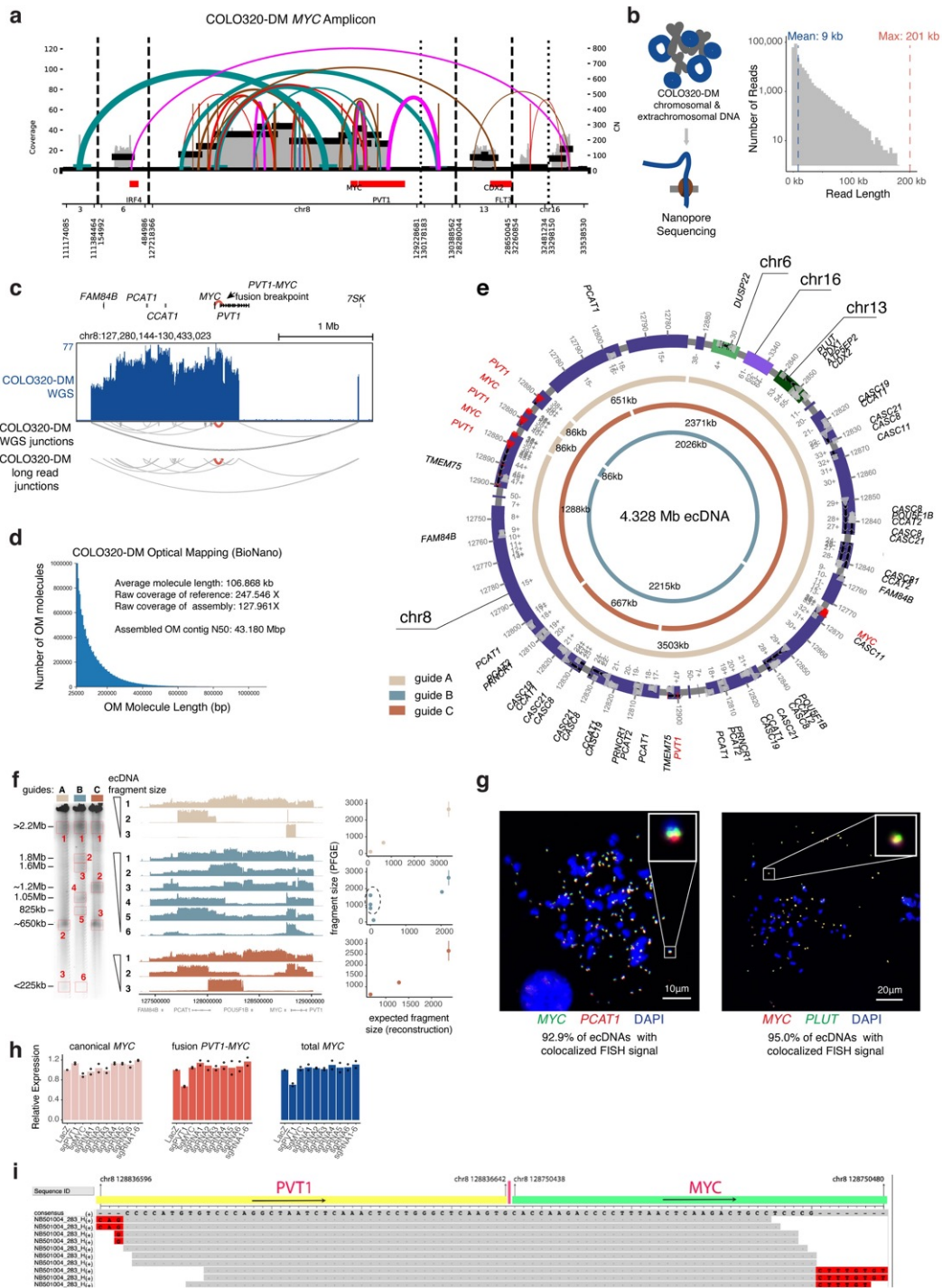
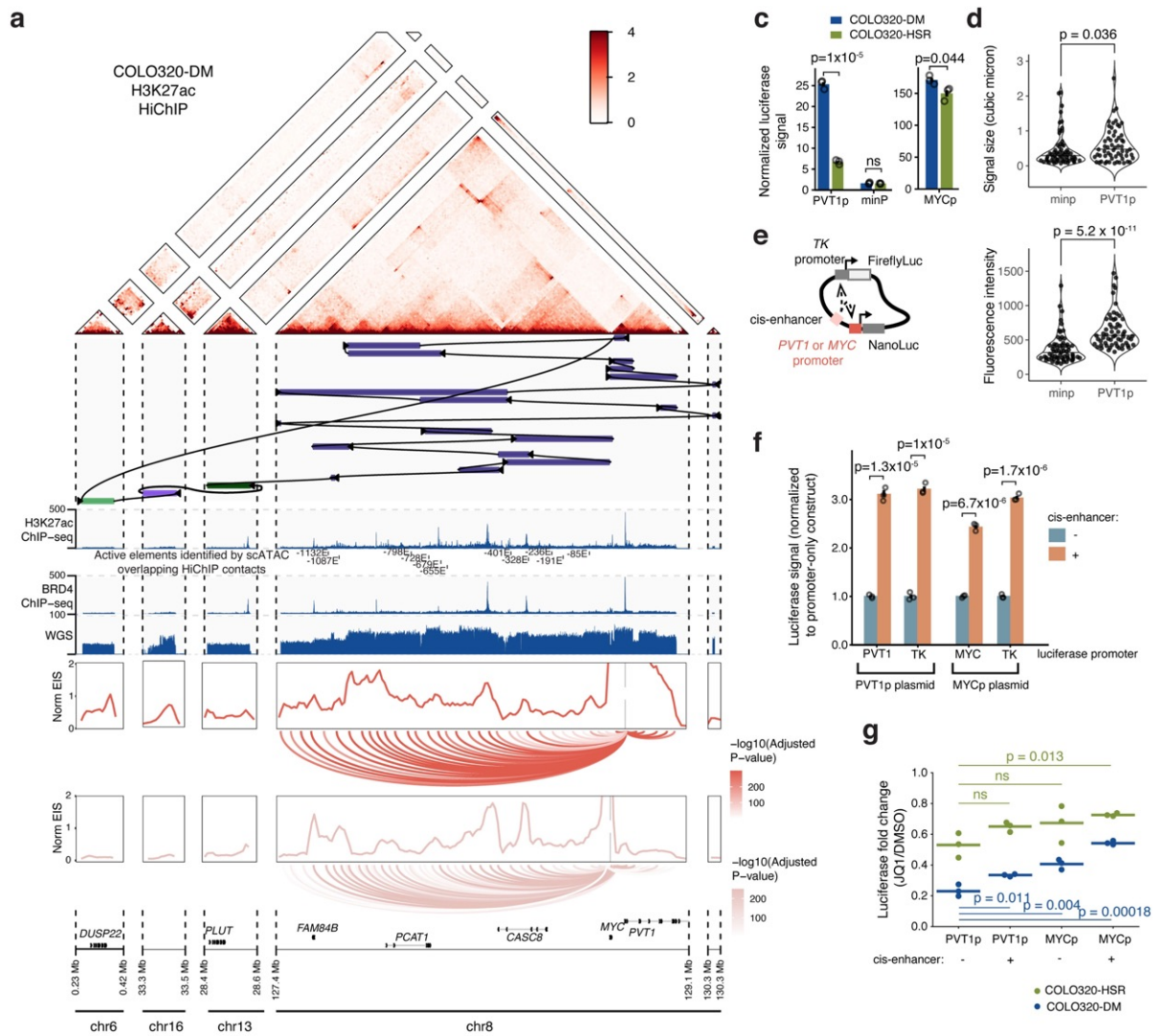**Extended Data Fig. 4** | See next page for caption.

Article

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

Article

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).
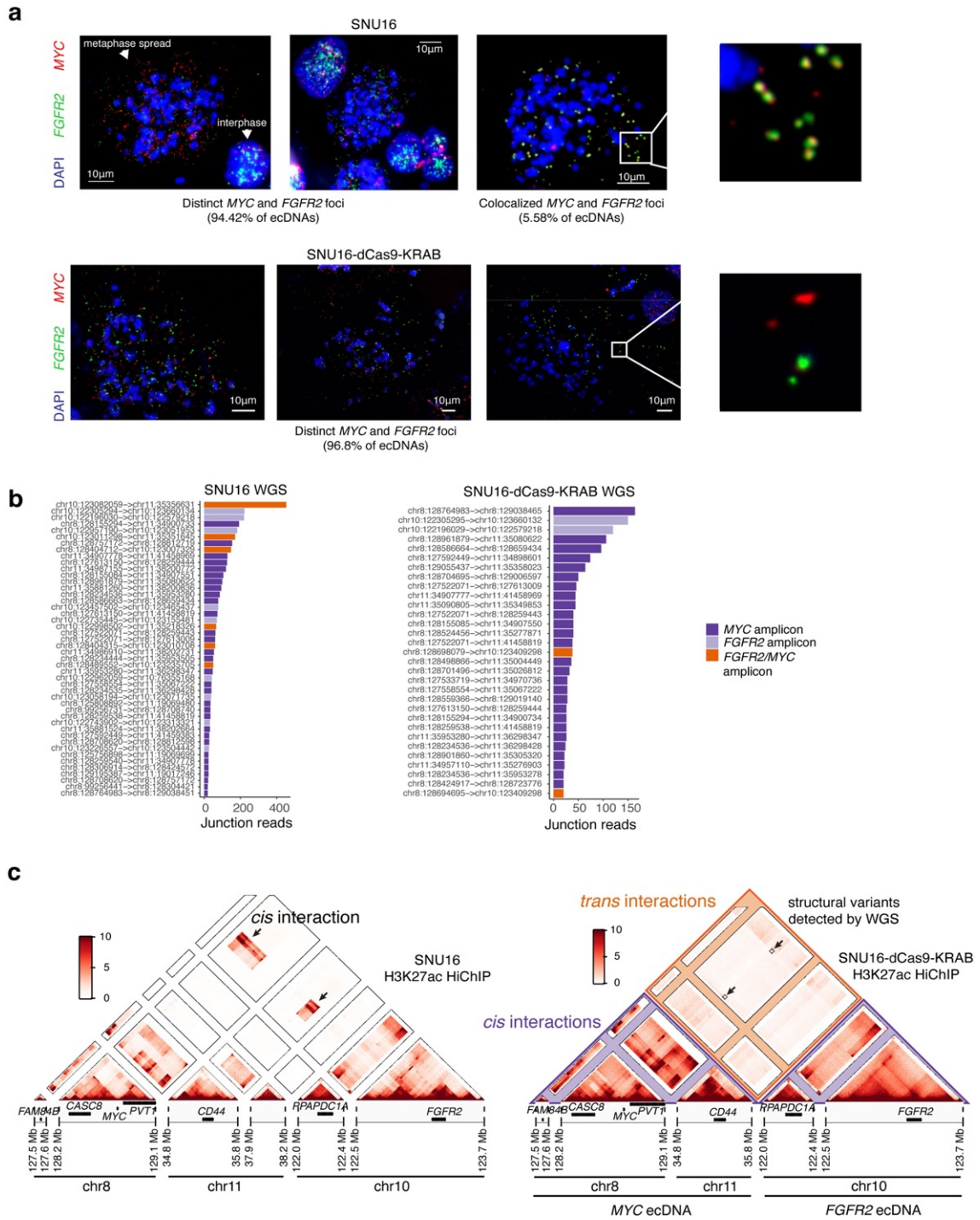
a

COLO320-DM
H3K27ac
HiChIP

H3K27ac ChIP-seq
Active elements identified by scATAC overlapping HiChIP contacts

BRD4 ChIP-seq
WGS

Norm EIS

−log10(Adjusted P−value)

Norm EIS

−log10(Adjusted P−value)

DUSP22   *PLUT*   *FAM84B*   *PCAT1*   *CASC8*   *MYC* *PVT1*

0.23 Mb 0.42 Mb   33.3 Mb 33.5 Mb   28.4 Mb 28.6 Mb   127.4 Mb   129.1 Mb 130.3 Mb 130.3 Mb

chr6   chr16   chr13   chr8

c
COLO320-DM
COLO320-HSR

Normalized luciferase signal

p=1×10⁻⁵   ns   p=0.044

PVT1p   minP   MYCp

d
p = 0.036
Signal size (cubic micron)

minp   PVT1p

e
TK promoter → FireflyLuc
cis-enhancer
PVT1 or *MYC* promoter → NanoLuc

p = 5.2 × 10⁻¹¹
Fluorescence intensity

minp   PVT1p

f
Luciferase signal (normalized to promoter-only construct)

p=1.3x10⁻⁵   p=1x10⁻⁵   p=6.7x10⁻⁶   p=1.7x10⁻⁶

cis-enhancer:
−
+

PVT1   TK   MYC   TK   luciferase promoter
PVT1p plasmid   MYCp plasmid

g
Luciferase fold change (JQ1/DMSO)

p = 0.013
ns
ns

p = 0.011   p = 0.004   p = 0.00018

PVT1p   PVT1p   MYCp   MYCp
cis-enhancer: −   +   −   +

COLO320-HSR
COLO320-DM

b

RAW data	ICE	KR	OneD+CN
chr8:127420000-129050000	chr8:127420000-129050000	chr8:127420000-129050000	chr8:127420000-129050000

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

Article

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).

a

SNU16

Distinct *MYC* and *FGFR2* foci
(94.42% of ecDNAs)

Colocalized *MYC* and *FGFR2* foci
(5.58% of ecDNAs)

SNU16-dCas9-KRAB

Distinct *MYC* and *FGFR2* foci
(96.8% of ecDNAs)

b

SNU16 WGS

SNU16-dCas9-KRAB WGS

Junction reads

- *MYC* amplicon
- *FGFR2* amplicon
- *FGFR2/MYC* amplicon

c

cis interaction

SNU16
H3K27ac HiChIP

trans interactions — structural variants detected by WGS

cis interactions

SNU16-dCas9-KRAB
H3K27ac HiChIP

chr8   chr11   chr10

chr8   chr11   chr10

MYC ecDNA   *FGFR2* ecDNA

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

146

Article

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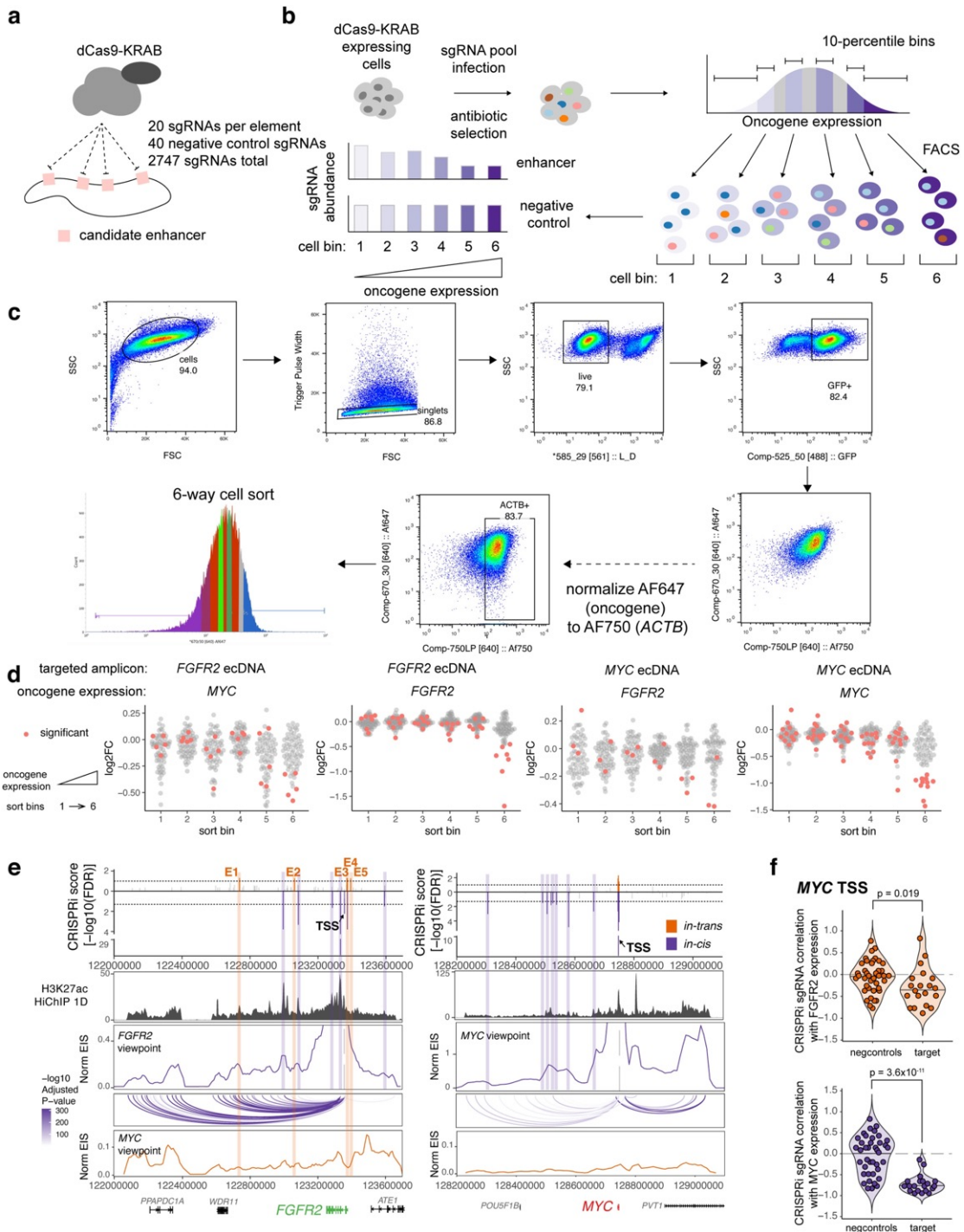
Extended Data Fig. 8 | See next page for caption.

Article

a enhancers: E1-E5

b

c

d

e

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

Article

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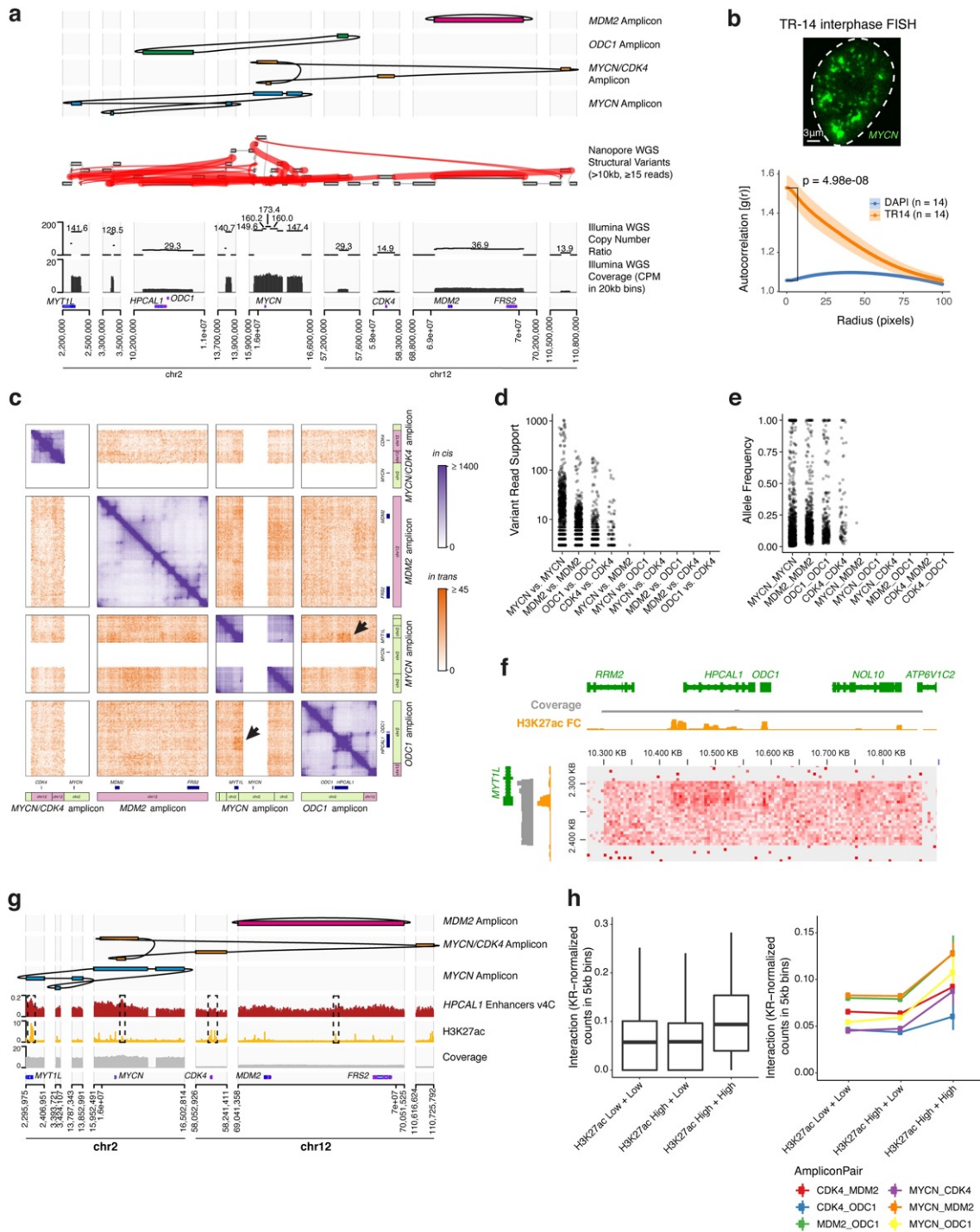
Extended Data Fig. 10 | See next page for caption.

Article

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.

Supplementary information

ecDNA hubs drive cooperative intermolecular oncogene expression

In the format provided by the
authors and unedited

Supplementary Tables

Supplementary Table 1.

primer	sequence	assay
MYC_exon1_fw	AGGCTCTCCTTGCAGCTGCTTA	RT_qPCR
MYC_exon2_rv	GCTAACGTTGAGGGGCATCGTC	RT_qPCR
PVT1_exon1_fw	GAAAGGATGTTGGCGGTCCCTG	RT_qPCR
total_MYC_exon2_fw	TCCACCTCCAGCTTGTACCT	RT_qPCR
total_MYC_exon2_rv	CGTCGAGGAGAGCAGAGAAT	RT_qPCR
GAPDH_fw	GGAGCGAGATCCCTCCAAAAT	RT_qPCR
GAPDH_rv	GGCTGTTGTCATACTTCTCATGG	RT_qPCR
18S_fw	TAAAGGAATTGACGGAAGGGCA	RT_qPCR
18S_rv	ATCTGTCAATCCTGTCCGTGTC	RT_qPCR
MYC_exon3_fw	GGCTCCTGGCAAAAGGTCA	RT_qPCR
MYC_exon3_rv	CTGCGTAGTTGTGCTGATGT	RT_qPCR
FGFR2_fw	GGTCGTTTCATCTGCCTGGT	RT_qPCR
FGFR2_rv	CCTTCCCGTTTTTCAGCCAC	RT_qPCR

Supplementary Table 2.

ID	gRNA_sequence	gRNA_info
1	GCCTCCGGGCAGAGCGCGTG	sgPVT1, targets PVT1 TSS in **Extended Data Figure 4i**
2	GGAATAGGGGGCTTCGCCTC	sgMYC, targets MYC TSS in **Extended Data Figure 4i**
3	GATGGATGGACCACAACAGGG	sgRNA1, targets enhancer in **Extended Data Figure 4i**
4	GGTTTCCTTATCTATGAACCG	sgRNA2, targets enhancer in **Extended Data Figure 4i**
5	GAGTGCATTAGAGGTACACAG	sgRNA3, targets enhancer in **Extended Data Figure 4i**
6	GATAAAAAGTGCCCGACAATG	sgRNA4, targets enhancer in **Extended Data Figure 4i**
7	GGGTTGGACTGATGACCTCAG	sgRNA5, targets enhancer in **Extended Data Figure 4i**
8	GAGGTTTGGAGACATACACAT	sgRNA6, targets enhancer in **Extended Data Figure 4i**
9	GAACGACTAGTTAGGCGTGTA	sgGal4, non-targeting guide in RNA-seq experiment
10	GGAGAGCTTGTGGACCGAGC	MYC ecDNA TetO insertion
11	GGGATCACTAGCATGTCTG	BRD4 HaloTag knockin
12	CCAGCAATCGTTAACCACTG	PFGE_guide_A, used in in vitro Cas9 ecDNA digestion
13	CTTCGGGGAGACAACGACGG	PFGE_guide_B, used in in vitro Cas9 ecDNA digestion
14	GTCTTGCTAGACTATACCAG	PFGE_guide_C, used in in vitro Cas9 ecDNA digestion

Supplementary Table 1: Primer sequences. All primers used for RT-qPCR are listed.

Supplementary Table 2: sgRNA sequences. sgRNAs used in the COLO320-DM CRISPR interference experiments and ecDNA editing for TetO insertion are listed.

Supplementary Table 3: FlowFISH CRISPRi sgRNA sequences. All sgRNA sequences used in the CRISPRi experiments in SNU16-dCas9-KRAB are listed.

8 Curriculum vitae

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

9 Publication List

PEER-REVIEWED ARTICLES
* = co-first authorship

Koche RP*, Rodriguez-Fos E*, Helmsauer K*, Burkert M, MacArthur IC, Maag J, Chamorro R, Munoz-Perez N, Puiggros M, Dorado García H, Bei Y, Roefzaad C, Bardinet V, Szymansky A, Winkler A, Thole T, Timme N, Kasack K, Fuchs S, Klironomos F, Thiessen N, Blanc E, Schmelz K, Künkele A, Hundsdörfer P, Rosswog C, Theissen J, Beule D, Deubzer H, Sauer S, Toedling J, Fischer M, Hertwig F, Schwarz RF, Eggert A, Torrents D, Schulte JH, Henssen AG. **Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma.** Nat Genet. 2020;52(1):29-34.

Helmsauer K*, Valieva ME*, Ali S*, Chamorro González R, Schöpflin R, Roefzaad C, Bei Y, Dorado García H, Rodriguez-Fos E, Puiggros M, Kasack K, Haase K, Keskeny C, Chen CY, Kuschel LP, Euskirchen P, Heinrich V, Robson MI, Rosswog C, Toedling J, Szymansky A, Hertwig F, Fischer M, Torrents D, Eggert A, Schulte JH, Mundlos S, Henssen AG, Koche RP. **Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma.** Nat Commun. 2020;11:5823.

Hung KL, Yost KE, Xie L, Shi Q, Helmsauer K, Luebeck J, Schöpflin R, Lange JT, Chamorro González R, Weiser NE, Chen C, Valieva ME, Wong IT, Wu S, Dehkordi SR, Duffy CV, Kraft K, Tang J, Belk JA, Rose JC, Corces MR, Granja JM, Li R, Rajkumar U, Friedlein J, Bagchi A, Satpathy AT, Tjian R, Mundlos S, Bafna V, Henssen AG, Mischel PS, Liu Z, Chang HY. **ecDNA hubs drive cooperative intermolecular oncogene expression.** Nature. 2021;600:731-6.

Dorado Garcia H, Pusch F, Bei Y, von Stebut J, Ibanez G, Guillan K, Imami K, Gurgen D, Rolff J, Helmsauer K, Meyer-Liesener S, Timme N, Bardinet V, Chamorro Gonzalez R, MacArthur IC, Chen CY, Schulz J, Wengner AM, Furth C, Lala B, Eggert A, Seifert G, Hundsoerfer P, Kirchner M, Mertins P, Selbach M, Lissat A, Dubois F, Horst D, Schulte JH, Spuler S, You D, Dela Cruz F, Kung AL, Haase K, DiVirgilio M, Scheer M, Ortiz MV, Henssen AG. **Therapeutic targeting of ATR in alveolar rhabdomyosarcoma.** Nat Commun. 2022;13(1):4297.

CONFERENCE ABSTRACTS

Helmsauer K, Valieva M, Ali S, Chamorro Gonzalez R, Schöpflin R, Schulte JH, Mundlos S, Koche RP, Henssen AG. **Enhancer hijacking on the MYCN amplicon in neuroblastoma.** 33. Jahrestagung der Kind-Philipp-Stiftung für pädiatr. onkolog. Forschung. Klin Padiatr. 2020;232(3):e4. (Conference cancelled due to COVID-19)

Henssen AG, Helmsauer K, Valieva M, Ali S, Eggert A, Schulte JH, Mundlos S, Koche RP. **Enhancer hijacking determines intra- and extrachromosomal circular MYCN amplicon architecture in neuroblastoma.** Advances in Neuroblastoma Research, 2021.

Henssen AG, Fos E, Helmsauer K, Eggert A, Torrents D, Schulte JH, Koche RP. **Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma.** Advances in Neuroblastoma Research, 2021.

10 Acknowledgements