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 Röfzaad 5, Yi Bei 1, Heathcliff Dorado Garcia 1, Elias Rodriguez-Fos 6, Montserrat Puiggrós 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 1,5,7,9,15,23 & 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 requirements 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.
ncogene 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). EcDNA can arise during genome reshuffling events like chromothripsis and are subsequently amplified. This partially explains why ecDNA can consist of several coding and non-coding distal parts of one or more chromosomes. Over time, amplified DNA acquires additional internal rearrangements as well as coding mutations, which can confer adaptive advantages such as resistance to targeted therapy. EcDNA reintegration into chromosomes can lead to intrachromosomal amplification as HSRs and act as a general driver of genome remodeling. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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-amplifying vs. non-amplifying 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.
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 neuroblastoma26. Two of the enhancers (e1 and e2) also harbored canonical E-boxes, suggesting binding of MYCN at its own enhancer (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 neuroblastomas27. 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 cohort19. 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 tumors28. 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 types30. 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.
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 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, 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. 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. 5a). 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 neuroblastoma (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 threedimensional 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. In IMR-5/75 and CHP-212, we observed insulator 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
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
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.

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). e, g Alignment of Hi-C reads to the reconstructed MYCN amplicon in IMR-5/75 (e) 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.
**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 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 reports31, 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/
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 1 amplicons observed in our cohort. In contrast to Morton et al., 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 territory35. Gene activation by enhancer adoption requires the fusion of distant DNA fragments and the formation of new chromatin domains, called neo-TADs36. In some cases, this fusion requires a convergent directionality of CTCF sites in order to form a new boundary and drive aberrant gene expression37. This has been explained by a model of blocked loop extrusion at forward–reverse oriented CTCF sites38. 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 CTFC-mediated chromatin loops in the wild-type genome8,39. 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 chromosomal linked reads or optical mapping46,34. 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 neuroblastoma40,41. Co-amplification with MYCN has been reported before31, but to our knowledge the clinical relevance of co-amplification had not been determined so far. Similar to our previous observations of PTPA2 co-amplification on chimeric ecDNA9, 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 fusions, 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. Spelman (Cancer Research Institute Ghent, Ghent; Belgium; NCP), F. Wessernagl (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 (Genetics 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)43. FASTQ files were quality controlled (FASTQC 0.11.8) and adapters were trimmed (BbMap 38.58). We mapped reads to GRCh37 (STAR 2.7.1f (ref. 42) with default parameters), counted them per gene (Ensembl release 75, FeatureCounts from Subread package 1.64 (ref. 43)), and normalized for library size and composition (sizeFactors from DESeq2 1.2.2 (ref. 44)).

**ChIP-seq.** For the cell lines CHP-212, NGP, and Kelly, 5–10 × 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 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-buturate), 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-buturate, 1% Triton X-100), Anti-H3K27ac (Dianova, 1:500; diluted 1:500; dilution 1:500), anti-H3K4me1 (Upstate, 1:2,000; diluted 1:200; dilution 1:200). Sequencing libraries were prepared using standard Nextera adaptors (Illumina) according to the supplier’s recommendations. Twenty-five million reads per sample were sequenced on a HiSeq 4000 on a single read mode.

Additional public ChIP-seq FASTQ files were downloaded from Gene Expression Omnibus (GSE18927, GSE90683, GSE24447, and GSE28874)45,46 and from ArrayExpress (E-MTAB-6570)47. 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. 48)) with default parameters and duplicate reads removed (Picard 2.2.0). 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. 49)). Peaks were called using MACS2 (2.1.2)49 with default parameters. Super-enhancers were called for H3K27ac data using LILY15 (https://github.com/BoevaLab/LILY) with default parameters. ChIP-seq data were quality controlled using RSC and NSC (Phantompeakcloudtools 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 0.75, 1 mM EDTA; 0.5 mM EGTA; 0.1% Na-deoxycholate; 0.5% N-lauroylsarcosine; protease inhibitors (Roche complete), 5 mM Na-buturate, 1% Triton X-100, Anti-H3K27ac (Dianova, 1:500; diluted 1:500; dilution 1:500), anti-H3K4me1 (Upstate, 1:2,000; diluted 1:200; dilution 1:200). Sequencing libraries were prepared using standard Nextera adaptors (Illumina) according to the supplier’s recommendations. Twenty-five million reads per sample were sequenced on a HiSeq 4000 on a single read mode.

**ATAC-seq.** ATAC-seq samples were processed as reported in Buenrostro et al.53 with some adaptations: After a treatment of 5–10 × 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 × 10^6 cells were washed with cold 1× PBS and lysed with fragmentation lysis buffer (10 mM Tri-Gly pH 7.5; 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 25.4 µl H2O, 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 adaptors 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)44. Adapter trimming, alignment, and duplicate removal as for ChIP-seq. We generated BigWig tracks by extending paired-end reads to fragment size, filtering by ENCODE DAC blacklist and normalizing to counts per million in 10 bp bins (deepTools 3.3.0 (ref. 49)). 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 described5. DNA was sheared to 500 bp in 1 ml sample with the M220 Ultrasonicator (Covaris) and digested with Tsp50I–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. After the washing step with cold 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 at 2 h at 37 °C, after which 10 U were added and then kept 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 ligation using a ligation reaction (663 nucleotides, 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 °C–7 °C). After sonication, the samples were purified using Ampure Beads (Beckman Coulter) and samples were sequenced with Illumina Hi-Seq technology according to the standard protocols and 75 bp (simple CHIMP-212 Hi-C, deep IMR-5 or 75/5 bp (simple IMR-5/75 Hi-C)) pair-end mode. Around 100 million reads were generated per IMR-5/75 replicate (deep IMR-5/75 Hi-C) and around 25–35 million reads per replicate were generated for shallow CHIMP-212 and shallow IMR-5/75 Hi-C.

Fluorescent in situ hybridization. Cells were grown to 200,000 per well in six-well plates and metaphase-arrested using Colcemid (20 µ/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 #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 PBST (0.4× SSC/0.3% IGEPAL for 3 min at RT. Five microliters DAPI (Vectashield, #H-1200, Vector) was added.

Enhancer calling. MYCN-expressing cell lines were defined as cell lines with size-factor normalized expression of 100 or above. 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 regions. Enhancer regions were screened to be overlapping with enhancer-promoter sequences from the JASPAR2018 (ref. 50) and JASPAR2020 (ref. 66) database using the TFBSTools 1.20.0 (ref. 31) function matchPWM with min.score = 85%. CRC-driven super-enhancers were defined as all regions with a LIFY-defined super-enhancer in MYCN-expressing, non-MYCN-amplified cell lines that overlapped with a GATA3, HAND2, or PHOX2B peak in CLIA-BG.

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 over 50% 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.

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, Ipswich, MA). Libraries were sequenced on a MGISEQ-2000 (NGP; MGI Tech Co Ltd, Shenzhen, China), HiSeq X (Illumina, Inc., San Diego, CA), and NovaSeq 6000 (CHP-212; Illumina, Inc., San Diego, CA) with 2 × 150 bp paired-end reads. Quality control was performed using the demultiplexing and alignment software FastQer (broadinstitute.org/snowman/svaba_exclusions.bed). Copy-number variation was called by GATK. 11.4 with default parameters. Structural variants were called using SplitSeg (11.1.1) in germline mode and discarding regions in a blacklist provided by SvABA (https://data.broadinstitute.org/snowman/svaba_exclusions.bed).
an amplified segment. For class II-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 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) and processed using rawcopy 1.1 (ref. 9) 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/](https://pecan.stjude.cloud/). An accompanying UCSC Track is available at [https://cyverse.org/dav-anon/iplant/home/konstantin/helmsaueretal/](https://cyverse.org/dav-anon/iplant/home/konstantin/helmsaueretal/). A fragment of the UCSC tracks for some 2. Then, reference edges were removed if its corresponding alternate edge was filtered as amplification.

**Amplicon reconstruction**. All unfiltered SvABA structural variant calls were filtered to exclude regions from the ENCODE blacklist and small rearrangements of 1 kbp 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 (“alleles depth”). Gnome was used to represent these data as a genome graph with nodes being breakpoint-free genomic intervals and edges being rearrangements (alternates) or connections derived by dividing the reference genome (“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 amplon sequences of CHIP-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 amplion sequence. The contribution of Hi-C di-tags from these regions on chromosome 2 to the amplion Hi-C map is expected to 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/](https://github.com/padpuydt/copynumber_HR_NB/) (ref. 27). Public under accession PRJNA622577. Copy-number data for high-risk neuroblastoma were downloaded from [https://github.com/henssenlab/MYCNAmplicon](https://github.com/henssenlab/MYCNAmplicon). All unfiltered SvABA structural variant calls were filtered as amplification.

**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. Decaesteker, B. 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*, 1333–1430 (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*, 2985–2997 (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 circuits. *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 adrenomedullary neuroblastoma core regulatory circuitry. *Nat. Commun. 10*, 5622 (2019).

19. Morton, A. R. et al. Functional enhancers shape extrachromosomal oncogene amplification. *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. Valentin, L. J. et al. TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. *Nat. Commun. 8*, 789 (2017).

23. Wenschenfeldt, 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. Hinzu, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science 351*, 1454–1458 (2016).

25. Croyongth, 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](https://doi.org/10.1038/sdata.2018.240) (2018).

28. Blümrich, A. et al. The FRG2C common fragile site maps to the borders of MYCN amplics 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 rearrangements 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., Ring, 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 (dmns) and homogeneously staining regions (HSRs) in nuclei of human neuroblastoma cell lines. *Genes Chromosomes Cancer 29*, 126–138 (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 microarray-based analysis of protein location. *Nat. Protoc.* **7**, 729–748 (2006).

40. Hogarty, M. D. et al. ODC1 is a critical determinant of MYCN oncogenesis and a therapeutic target in neuroblastoma. *Cell Stem Cell* **6**, 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**, eaau1999 (2019).

42. 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).

43. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DSeq2. *Genome Biol.* **15**, S50 (2014).

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

45. Despang, A. et al. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Cell Stem Cell* **11**, 633–648 (2012).

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

47. Ramirez, F. et al. deepTools2: a next generation web server for deep-seq sequencing data analysis. *Bioinformatics* **32**, 1355–1356 (2016).

48. Liang, J. M. et al. 3D Hi-C analysis of genome architecture. *Nat. Genet.* **47**, 215–223 (2015).

49. Liang, J. M. et al. 3D Hi-C analysis of genome architecture. *Nat. Genet.* **47**, 215–223 (2015).

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 RBioConductor package for transcription factor binding site analysis. *Bioinformatics* **32**, 1355–1356 (2016).

52. Scheinert, L. 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 inversion shapes MYCN-dependent transcriptional enhancer/promoter function. *Cell Syst.* **26**, 1740 (2016).

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. *Highly selective and instructive roles of TAD architecture.* *Nat. Genet.* **50**, 515–523 (2018).

57. Schep, N. et al. DNA methylation and chromatin interactions in human cell lines. *Nat. Genet.* **46**, 1266–1269 (2018).

58. Kloth, M., Wu, A., Wain, J., Lin, Y. & Cattanach, B. NCI-60: A 3D model of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

59. Guo, Y. et al. CRISPR inversion of CTCF sites alters genome topology and microarray-based analysis of protein location. *Nat. Protoc.* **7**, 729–748 (2006).

60. Mayrhofer, M., Viklund, B. & Isaksson, A. Rawcopy: Improved copy number normalization for copy number variation in Hi-C data. *Genome Res.* **29**, 29–39 (2019).

61. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-tolerant 13,504–13,504 (2018).

62. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *Highly selective and instructive roles of TAD architecture.* *Nat. Genet.* **50**, 515–523 (2018).

63. 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).

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

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

66. Wright, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *Highly selective and instructive roles of TAD architecture.* *Nat. Genet.* **50**, 515–523 (2018).

67. 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).

68. 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).

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

70. 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).

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

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

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

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

75. Heine, 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).

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

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