DISSERTATION

Dissecting extrachromosomal circular DNA heterogeneity in single cells with scEC&T-seq

Untersuchung der Heterogenität extrachromosomaler zirkulärer DNA in Einzelzellen mit scEC&T-seq

zur Erlangung des akademischen Grades
Doctor of Philosophy (PhD)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

María del Rocío Chamorro González

Erstbetreuer: Prof. Dr. Anton G. Henssen

Datum der Promotion: 30.06.2024

Table of contents

List of figures

List of abbreviations

DM	Double minute
ecDNA	Extrachromosomal DNA
eccDNA	Extrachromosomal circular DNA
FACS	Fluorescence-activated cell sorting
FISH	Fluorescence in-situ hybridization
mtDNA	Mitochondrial DNA
RCA	Rolling circle amplification
scEC&T-seq	Single-cell extrachromosomal circular DNA and transcriptome
spcDNA	Small polydispersed circular DNA
SV	Structural variant
WGS	Whole genome sequencing

Abstract

Extrachromosomal DNA circularization is a common event in cancer that can occur in various forms. Many small circular DNAs, generated through DNA fragmentation processes such as DNA damage and apoptosis, exist in both normal and cancerous tissues. Other large, cancer-specific circular DNAs, commonly referred to as ecDNA, serve as potent vehicles for oncogene amplification in aggressive tumors. Extrachromosomal elements undergo random segregation during mitosis, facilitating rapid intercellular heterogeneity and enabling tumors to evolve and evade therapy at an accelerated pace. A better understanding of circular DNA dynamics and their impact on intercellular heterogeneity in cancer is crucial. Therefore, there is a need for novel single-cell methodologies that enable the profiling of all types of circular DNAs in single cancer cells. In this doctoral thesis, we introduce scEC&T-seq (single-cell extrachromosomal circular DNA and transcriptome sequencing), a new single-cell method that enables the simultaneous detection of circular DNA and full-length mRNA from the same single cell. Applying scEC&T-seq to neuroblastoma cell lines and primary tumors, we characterized hundreds of circular DNA elements per single cell. While large, ecDNAs were recurrently identified in most cancer cells, small circular DNAs were mostly private and not shared between individual cells, indicating differences in their selection and propagation. scEC&T-seq successfully recapitulated the complexity of ecDNA structures in single cells and characterized intercellular differences in ecDNA structure. This enabled the inference of ecDNA structural dynamics over time in primary neuroblastoma tumors. Additionally, the parallel interrogation of ecDNA and gene expression data revealed that intercellular differences in ecDNA content drive oncogene expression differences in single cells. We anticipate that applying scEC&T-seq to samples from different biological contexts will unravel new insights into the role of both large and small extrachromosomal circular DNAs in cancer and beyond.

Zusammenfassung

Extrachromosomale DNA-Zirkularisierung ist ein häufiges Ereignis bei Krebs, das in verschiedenen Formen auftreten kann. Viele kleine zirkuläre DNAs, die durch DNA-Fragmentierungsprozesse entstehen, existieren sowohl in gesundem Gewebe als auch in bösartigem Gewebe. Andere große, tumorspezifische zirkuläre DNAs, allgemein als ecDNA bezeichnet, dienen als effektives Mittel zur Verstärkung von Onkogenen in aggressiven Tumoren. Aufgrund ihres Fehlens von Zentromeren segregieren sich zirkuläre DNAs zufällig auf Tochterzellen während der Mitose. Dies kann eine schnelle interzelluläre Heterogenität fördern, wodurch Tumoren die Fähigkeit erhalten, sich schnell zu entwickeln und der Therapie zu entkommen. Das Verständnis der Dynamik zirkulärer DNA und ihres Beitrags zur interzellulären Heterogenität bei Krebs ist jedoch nach wie vor von großer Bedeutung, da Methoden für eine unvoreingenommene Charakterisierung sowohl großer als auch kleiner zirkulärer DNAs in einzelnen Zellen fehlen. In diesem Projekt haben wir uns zum Ziel gesetzt, scEC&T-seq (single-cell extrachromosomal circular DNA and transcriptomic sequencing) zu etablieren, eine Methode für die parallele Detektion zirkulärer DNAs und vollständiger mRNA in einzelnen Zellen. Durch Anwendung von scEC&T-seq auf Neuroblastom-Zelllinien und primäre Tumore haben wir hunderte zirkuläre DNA-Elemente pro Einzelzelle charakterisiert. Während große, Onkogene enthaltende ecDNAs wiederholt in den meisten Krebszellen identifiziert wurden, waren kleine zirkuläre DNAs hauptsächlich auf einzelne Zellen beschränkt, was auf Unterschiede in ihrer Auswahl und Vermehrung hinweist. scEC&T-seq konnte erfolgreich die komplexe Struktur von ecDNAs in einzelnen Zellen wiedergeben, und lieferte ausreichende Details, um Unterschiede in der ecDNA-Struktur zwischen den Zellen zu charakterisieren. Dies ermöglichte die Ableitung von ecDNA-Strukturdynamiken im Laufe der Zeit in primären Tumoren. Zusätzlich offenbarte die gleichzeitige Untersuchung von ecDNA- und Genexpressionsdaten, dass interzelluläre Unterschiede im ecDNA-Gehalt Unterschiede in der Onkogenexpression in einzelnen Zellen beeinflussen. Wir gehen davon aus, dass die Anwendung von scEC&T-seq auf Proben aus verschiedenen biologischen Kontexten neue Erkenntnisse über die Rolle großer und kleiner extrachromosomaler zirkulärer DNAs bei Krebs und darüber hinaus ermöglichen wird.

1. Introduction

While the majority of the human genome consists of large linear strands of DNA that are packaged in chromosomes, extrachromosomal circular DNA elements have long been identified in the nucleus of human cells(1-6). These closed DNA circles are derived from the linear chromosomal genome becoming independent structures in the cell's nucleus. Recent reports have shown the unexpectedly high prevalence of circular DNA in human tissues(2, 4, 5). Since their discovery more than five decades ago(7, 8), many forms of circular DNA have been described in human cells. Based on differences in size, content, and copy number, they can be classified into two main categories: i. small (<100kb), circular DNA, commonly referred to as eccDNA, this category includes: microDNAs(5), apoptosis-derived circular DNAs(9), small polydispersed circular DNAs(10) (spcDNA), and telomeric circles(11); ii. large (>100kb), cancer-specific, copy number-amplified extrachromosomal DNA, known as ecDNA(3). Although in the past decade, important progress has been made in characterizing both small and large circular DNAs, many questions about their function and relevance in normal and cancer development remain unanswered.

In the cancer field, cancer-specific ecDNAs are of particular interest due to their described role as powerful drivers of oncogene amplification(1). These mega-base-sized circular DNAs were initially discovered as double minute chromosomes (DM) in the 1960s through microscopic examination of stained metaphases of neuroblastoma cell lines(7). They were soon found to carry copies of proto-oncogenes and other functional regulatory elements and, by accumulating in the cell, they led to abnormally high oncogene expression levels(1, 3, 12-14). Recently, a re-evaluation of ecDNA in large-scale DNA sequencing datasets revealed that extrachromosomal amplification is a more frequent event in cancer than previously anticipated(15). Most importantly, this study revealed that patients with ecDNA-driven tumors present a decreased survival compared to those carrying other types of focal amplifications(15, 16). The aggressive behavior of ecDNA-containing tumors may be attributed to their capacity for rapid adaptation to selective pressures(12, 17, 18). This rapid adaptation has been linked to ecDNA's ability to be replicated and, since it lacks centromeres, randomly segregated to daughter cells in mitosis(12, 19-21). This missegregation is expected to yield high ecDNA copy number heterogeneity in the cancer population, which is suggested to enable more rapid changes in DNA content,

such as changes in oncogene copy number, and support rapid adaptation to new selective pressures(12). In addition to diversity in copy number, ecDNAs can exhibit diverse structural compositions(22-26). EcDNAs can present complex structures resulting from the rearrangement of multiple genomic regions from the same or different chromosomes(16, 23, 27). Recent studies have also shown that in some tumor populations, extrachromosomal elements can be highly heterogeneous. In certain cases, multiple ecDNA structures may be present, sometimes carrying different oncogenes, and they can become more complex over time and in response to therapy(23, 28, 29). Even in some cases, ecDNAs carrying no oncogene but only functional regulatory elements, such as enhancers, have been identified to co-exist with other oncogene-containing ecDNAs(20). The molecular mechanisms generating this diversity in ecDNA content, its extent, and its functional impact on tumor evolution are, however, not yet completely understood.

Besides ecDNA, other smaller types of extrachromosomal circular DNAs have been described in cancer cells(30). Their discovery was parallel to that of ecDNAs(8). Unlike ecDNAs, which are specific to cancer cells, small circular DNAs have been identified in all tested eukaryotic cells, both malignant and non-malignant(6). Compared to large ecDNAs, small circular DNA elements are not present in high copy numbers (non-amplified) (2, 5) and their mechanism of inheritance is largely unknown. Moreover, their sequence is often too short to contain full genes or promoter regions(2, 5). Some reports indicate that small circular DNA formation is linked to DNA damage and aberrant repair, as observed by the elevated levels of small circular DNAs in genetically unstable cells or after DNA damage induction(10, 31). A recent study has demonstrated that apoptosis can significantly increase the formation of small circular DNAs through a process that depends on DNase γ endonuclease activity and Ligase 3-mediated DNA ligation(9). These increased numbers of small circular DNAs can also stimulate innate immunity(9). Although cancers can harbor thousands of small circular DNAs of varying sizes and genetic content(6), their role in cancer biology has been understudied and it is currently poorly understood, with most of the research focusing on oncogene-containing ecDNAs.

Our research group recently described the landscape of DNA circularization in neuroblastoma(4), a tumor entity in which extrachromosomal amplification of the *MYCN* oncogene, which was first described in the 60s(7), is frequently identified and is associated with unfavorable outcomes and risk of relapse(32-34). Our findings revealed the presence of

ecDNA in approximately 30% of neuroblastoma tumors, with *MYCN* being the most commonly amplified gene in this context(4). In addition, we identified thousands of other small, non-amplified circular DNAs per tumor of yet unknown functional relevance(4). The prevalence and diversity of both large ecDNAs and small circular DNAs observed in neuroblastoma tumors raised several important questions, including: (1) How does the overall content of circular DNA vary in cell populations? For instance, how are (i) small, non-amplified circular DNAs (ii) or ecDNA copy numbers distributed in a cell population? (2) How are multiple ecDNA structures distributed among cancer cells? Do these structures co-exist within cells, or are they only present in a subset of cells but not others? (3) What impact does this heterogeneity have on cell phenotype?

To answer these relevant questions, gaining a better understanding of circular DNA's contribution to intercellular heterogeneity is crucial. While recent advances have been made, current methodologies have limitations that hinder a comprehensive understanding of the contribution of both large and small circular DNAs to intercellular tumor heterogeneity. Most studies have relied on either cytogenetic identification of ecDNAs using fluorescence in situ hybridization (FISH), or computational inference of ecDNA sequence from bulk whole-genome sequence (WGS) data(1). These methods are, however, limited in their ability to deconvolute ecDNA diversity and can't detect non-amplified, small circular DNAs. Other targeted exonuclease-based circular DNA enrichment protocols, such as Circle-seq(4, 35), have allowed the detection of both small and large circular DNAs in tumors(4). However, this data only provides an average of circular DNA content across a population of cells and does not reflect intercellular heterogeneity. A promising alternative to bulk sequencing methods is single-cell sequencing approaches. Current single-cell genomic approaches, however, present similar limitations only relying on high copy numbers for circular DNA detection(36). To overcome these limitations, in this PhD, we set up to establish scEC&T-seq (single-cell extrachromosomal circular DNA and transcriptome sequencing), a new method for sequencing and characterization of all extrachromosomal circular DNA types, independent of their size, content, and copy number, combined with parallel detection of full-length mRNA transcripts in single cells. We demonstrate its utility for profiling intercellular heterogeneity in single cancer cells containing both structurally complex multi-fragmented ecDNAs and small circular DNAs, while interrogating the effects of this heterogeneity on gene expression.

2. Methods

This methods section has been adapted from Chamorro González *et al*.(37)

Cell culture

The two human cancer cell lines used in this study were obtained from the American Type Culture Collection (CHP-212; ATCC; Manassas, VA, USA) or provided by J. J. Molenaar (TR14; Princess Máxima Center for Pediatric Oncology, Utrecht, Netherlands). STR genotyping (Genetica DNA Laboratories and IDEXX BioResearch) was used to confirm the identity of both cell lines, and absence of *Mycoplasma sp.* contamination was verified with a Lonza MycoAlert system (Lonza). Cells were grown in RPMI-1640 medium (Thermo Fisher Scientific) supplemented with 1% penicillin/streptomycin (PS; Thermo Fisher Scientific), and 10% fetal calf serum (FCS; Thermo Fisher Scientific).

Patient samples and clinical data access

Tumor tissue and blood samples of patients diagnosed with neuroblastoma between 1991 and 2022 were used in this study. Informed consent from patients or their legal guardians was obtained in all cases. The institutional review boards of Charité-Universitätsmedizin Berlin and the Medical Faculty, University of Cologne approved the collection and use of patient specimens for this study. Access to specimens and clinical data was granted by Charité-Universitätsmedizin Berlin or the National Neuroblastoma Biobank and Neuroblastoma Trial Registry (University Children's Hospital Cologne) of the German Society of Pediatric Oncology and Hematology (GPOH).

Single-cell or single-nuclei suspension preparation

Nuclei isolation from tissue samples was performed by tissue homogenization using a pre-cooled glass dounce homogenizer (Wheaton) in 1 mL of ice-cold EZ PREP buffer (Sigma). The tissue was homogenized using ten strokes with the loose pestle, followed by five additional strokes with the tight pestle, then filtered through a BD Falcon tube with a 35µm cell strainer cap (Becton Dickinson). Intact nuclei were stained and counted with 0.02% trypan Blue (Thermo Fisher Scientific). To isolate peripherial blood mononuclear cells (PBMCs) from patient blood samples, density gradient centrifugation with Ficoll-Plaque^{TM} Plus (Cytiva) was used. Blood samples were mixed 1:1 with free-calcium PBS

and slowly added to 12mL of Ficoll (Cytiva). After centrifugation for 30 min at 200g without breaking, an upper layer containing PBMCs was isolated and washed in 40 mL of PBS. After washing, PBMCs were collected by centrifugation at 500g for 5 minutes and resuspended in 10% DMSO in FCS.

Fluorescence-activated cell sorting (FACS)

For single-cell sorting, one to ten million neuroblastoma cells or PBMCs were stained with Propidium Iodide (PI, Thermo Fisher Scientific) in 1× PBS, and viable cells were selected based on the forward and side scattering properties as well as PI staining. PBMCs suspensions were additionally stained with a 1:400 dilution of anti-human CD3 (Ax700, Biolegend). Nuclei suspensions were stained with DAPI (Thermo Fisher Scientific, final concentration 2 µM). Viable cells, CD3-positive PBMCS or DAPI-positive nuclei were sorted using a FACSAria Fusion flow cytometer (Biosciences) into 2.5 µl of RLT Plus buffer (Qiagen) in low binding 96-well plates (4titude) sealed with foil (4titude) and stored at −80 °C until processing.

scEC&T sequencing

A comprehensive protocol for scEC&T-seq can be accessed on “Nature protocol exchange”(37). To separate DNA and mRNA, the procedure described by Macaulay *et al*. was followed(38). Briefly, oligo-dT primers (5′-biotin-triethyleneglycol-AAGCAGTGGTATCAACGCAGAGTACT30VN-3′) conjugated to streptavidin-coupled magnetic beads (Dynabeads® MyOne Streptavidin C1, Invitrogen) were used to capture polyadenylated mRNA. 10 µl of the conjugated beads were added to the cell lysate and mixed at 800 rpm for 20 min (MixMate, Eppendorf). The captured mRNA was separated from the genomic DNA (gDNA), present in the supernatant, using a magnet (Alpaqua). The mRNA-bound beads were washed three times at RT in 200 µl of a mixture of 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 0.05% Tween-20 and 0.2× RNAse inhibitor (SUPERasin, Life Technologies). Each washing step involved mixing the beads for 5 min at 2,000 rpm using a MixTape (Eppendorf). The supernatant from each wash was pooled with the original supernatant using the same pipette tips to minimize DNA loss. The mRNA captured on the beads was then eluted into 10 µl of a reverse-transcription master mix containing 10 U/µl SuperScript II reverse transcriptase (Life Technologies), 1× Su-

perscript II First-Strand Buffer (Life Technologies), 1 U/µl RNAse inhibitor (Life Technologies), 2.5 mM DTT (Life Technologies), 6 mM $MgCl_2$ (Life Technologies), 1 M betaine (Sigma),), dNTP mix 1mM each (dATP, dCTP, dGTP and dTTP) (Thermo Scientific) 1 µM Template-Switching Oligo (TSO, 5′-AAGCAGTGGTATCAACGCAGAGTACrGrG+G-3′, where "r" indicates a ribonucleic acid base and "+" indicates a locked nucleic acid base) and nuclease-free water (Life Technologies) up to the final volume (10 µl). The reverse transcription reaction was carried out on a thermocycler at 42 °C for 60 min followed by 10 cycles of 2 min at 50 °C and 2 min at 42 °C and a final 10-min incubation at 60 °C. Immediately after reverse transcription, the cDNA was PCR-amplified by adding 12 µl of PCR master mix including 1× KAPA HiFi HotStart ReadyMix (Kapa) and 0.1 µM ISPCR primer (5′-AAGCAGTGGTATCAACGCAGAGT-3, 10 mM) to the 10 µl of the reverse-transcription reaction mixture. The PCR reaction was performed on a thermocycler and consisted of 7 cycles of 3 min at 98 °C; followed by 18 cycles of 15 s at 98 °C, 20 s at 67 °C, and 6 min at 72 °C; and a final 5-min incubation at 72 °C. The amplified cDNA was purified using AMPure beads (Beckman Coulter) at a volumetric ratio of 1:0.9 and eluted into 20 µl of elution buffer (Buffer EB, Qiagen). For circular DNA isolation, a previously described method used in bulk populations was employed(35, 39). The isolated DNA was incubated with a 1:0.8 volumetric ratio of Ampure Beads (Beckman Coulter) for 20 min at RT with mixing at 800 rpm (MixMate, Eppendorf). The DNA was directly eluted from the beads into 20 µl of an exonuclease digestion master mix consisting of 20 units of Plasmid-Safe ATP dependent DNase (Epicentre), 1mM ATP (Epicentre), 1× Plasmid-Safe buffer (Epicentre). In a subset of cases for the endonuclease controls, 1 µl of the endonuclease MssI/PmeI (New England Biolabs, 20 U/µl) was added. Linear DNA digestion was performed for either 1 or 5 days at 37 ºC. For the 5-day digestion regimen, 10 units of Plasmid-Safe DNase and 4 µl of ATP (25mM) were added every 24 hours to sustain enzymatic digestion. After digestion, the exonuclease was inactivated by heating at 70 ºC for 30 min. To purify the exonuclease-resistant DNA, 32 µl of PEG buffer (18% (w/v) PEG (Sigma), 25 M NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.05% Tween -20) were added and incubated for 20 min at RT with mixing at 800 rpm (MixMate, Eppendorf). Rolling circle amplification of exonuclease-digested DNA was performed by Multiple Displacement Amplification (MDA) with the REPLIg single-cell kit (Qiagen). Following incubation with the AMPure beads, the beads were washed with 80% ethanol twice and eluted directly into 7 µl of denaturation buffer (D2) from the REPLIg single-cell kit (Qiagen). The reaction was incubated at 65 ºC for 10 min. After incubation, the denaturation

was stopped by adding 3 µl of STOP solution. The MDA reaction was then initiated by adding 40 µl of reaction master mix and was incubated at 30 ºC for 8 hours. The DNA polymerase was inactivated by heating at 65 ºC for 3 min. The amplified circular DNA was then purified using Ampure beads (Beckman Coulter) at a volumetric ratio of 1:0.8 and eluted in 100 µl of elution buffer (Buffer EB, Qiagen). For library preparation, 20 ng of amplified cDNA or circular DNA was used with the NEBNext Ultra II FS (New England Biolabs), following the manufacturer's instructions. Sample barcoding was performed using unique dual index primer pairs (New England Biolabs), and libraries were pooled. Sequencing was performed on a HiSeq 4000 instrument (Illumina) or a NovaSeq 6000 instrument using 2× 150bp paired-end reads for Circle-seq libraries and 2× 75 bp paired-end reads for RNA-seq libraries.

Circle-seq and RNA-seq read alignments

Circle-seq sequencing reads were trimmed using Trim Galore (version 0.6.4)(40) and subsequently aligned to the human genome build 19 (hg19) using bwa mem (version 0.7.17)(41). Using hisat2 (version 2.2.1)(42), RNA-seq data were aligned against a transcriptome reference created from human genome build 19 (hg19) and ENCODE annotation v19(43). Genes and isoforms were then quantified using rsem (version 1.3.1)(44) incorporating a single cell prior.

Single-cell Circle-seq analysis

Circular DNA analysis was performed as previously described in Koche *et al*.(39). In short, sequencing reads were 3′ trimmed for quality, and adapter sequences with reads shorter than 20 nucleotides were removed. The alignment of reads to the human reference assembly hg19 was performed using Bwa mem (version 0.7.15) with default parameters. PCR and optical duplicates were removed using Picard (version 2.16.0). The classification of putative circles was accomplished in two steps. First, a new bam file was generated including all split reads and read pairs containing an outward-facing read orientation. Second, in the bam file including all reads, regions with enriched signal over background with a false rate discovery (FDR) < 0.001 were detected using variable-width windows from Homer v.4.11 findPeaks (http://homer.ucsd.edu/). The edges of these enriched regions were then intersected with the circle-supporting reads from the previously created bam file including split reads and outward-facing reads. The threshold for circle

detection was determined empirically based on a positive control set of circular DNAs obtained from bulk sequencing data. Regions enriched and intersected by at least two circle-supporting reads were classified as circular regions.

QC filtering of single-cell Circle-seq data

Mitochondrial DNA was used as an internal control to evaluate circular DNA enrichment. Cells that exhibited less than 10 reads per base pair sequence-read depth over mitochondrial DNA or less than 85% genomic bases captured in mitochondrial DNA were excluded from further analyses. The chosen cut-off values were determined based on the maximal read depth values detected in endonuclease controls. Since mitochondrial DNA is not present in nuclei, single-nuclei Circle-seq data was filtered only based on RNA QC, meaning that those cells that did not pass RNA QC, were also excluded from further Circle-seq analyses.

Circular DNA recurrence analysis

Using bedtools multicov (https://bedtools.readthedocs.io), read counts from scCircle-seq bam files were quantified in 100 kb bins across all chromosomes from the genome assembly hg19. To account for differences in sequencing depth among cells, counts were normalized to library size. Only 100-kb bins that contained circle read enrichment with $p < 0.05$ compared to the background read distribution were considered. Based on their genomic coordinates, bins were then classified into three groups: i) ecDNA if the region overlapped the reconstructed ecDNA assembled from bulk WGS sequencing data, ii) chrM, mitochondrial DNA, iii) all other sites of the genome. Recurrence was then assessed by plotting the fraction of cells containing a detected circular region in each of the three abovementioned categories.

Identification of Structural Variants (SV)

Structural variant calling from single-cell Circle-seq data was performed using lumpy-sv[55] (version 0.2.14) and SvABA[56] (version 1.1.0). For SV detection in pseudobulk and WGS data, lumpy-sv(45) (version 0.3.1) and SvABA(46) (version 1.1.0) were used, both with standard parameters. To merge all alignment files from the same cell line into one pseudobulk alignment, Samtools(47) (version 1.11) was used. The pseudobulk merged bam

file was then downsampled to ten percent its original size to achieve coverage close to bulk sequencing.

QC filtering and clustering of scRNA-seq data

Seurat(48) (version 4.10) was used for single-cell RNA-seq analyses. Cells from cell lines with less than 5,000 features and T-cells and nuclei with less than 2,000 features were considered low quality and filtered out for further analyses. Cells or nuclei with high expression of mitochondrial genes, exceeding 15% in cells and 2.5% in nuclei, were also excluded. Normalization was performed using transcripts per million (TPM), then a pseudo count of one was added and natural-log transformation was applied. For cell clustering, the first 5 PCs were used for FindNeighbors and RunUMAP as the first 4 PC were found to be significant. The resolution for FindClusters was set to 0.5.

Differential expression and pathway enrichment analyses

First, CHP-212 cells were classified based on their very small circular DNA (<3kb) content. In each cell, the number of circular DNAs shorter than 3kb was counted and divided by the total number of circular DNAs. CHP-212 cells were then ranked and grouped by taking the top and bottom 40% of the ranked list, defined as “high” and “low” groups. The logarithmic fold change of gene expression between the two groups was determined using the FindMarkers function from the Seurat R package(48) (version 4.10). No logarithmic fold change threshold was applied and a minimum detection rate per gene of 0.05 was used. For unsupervised gene set enrichment analyses (GSEA), the R package clusterprofiler*(49)* (version 4.0.5) was used. Gene sets from gseGO with a minimum of 3 genes and a maximum of 800 genes were included in the analysis.

Correlation of scCircle-seq and scRNA-seq coverage

Using Counts Per Million (CPM) normalization with bamCoverage(50), we determined the read coverage across ecDNA regions in single-cell Circle-Seq and RNA-seq bam files. Coverage correlation was analyzed by fitting a linear model.

Fusion gene identification

Single-cell RNA-seq fastq files were merged and aligned with STAR(51) (version 2.7.9a) to the reference decoy GRCh37/hs37d5, with the gene annotation GENCODE19, enabling chimeric alignment (--chimOutType WithinBAM SoftClip). Arriba(52) (version2.1.0) with custom parameters (-F 150 -U 700) was used to call and visualize fusion genes. Only fusions with equal or more than 50× total coverage across the breakpoint, and at least 30% of the mapped reads classified as split or discordant reads, were considered.

Reconstruction of ecDNA

ecDNA reconstructions for CHP-212 and TR14 had been previously published in Helmsauer *et al*.(53) and Hung *et al*.(23), respectively. For neuroblastoma tumors, WGS nanopore data was generated as described in Helmsauer *et al*.(53). Basecalling and demultiplexing was performed with Guppy (version 5.0.14). Quality filtering of reads was done using NanoFilt(54) (version 2.8.0), and alignment against GRCh37/hg19 reference genome was performed using ngmlr(55) (version 0.2.7). Sniffles(55) (version 1.0.12) was used for SV calling and binned coverage was obtained using deepTools(50) (version 3.5.1) bamCoverage. For ecDNA reconstruction, a set of confident SV calls was compiled (VAF > 0.2 and supporting reads >= 50X). A genome graph was built using gGnome[62] (version 0.1) and manually curated. Reconstructions were visualized using gTrack (version 0.1.0; https://github.com/mskilab/gTrack), including the GRCh37/hg19 reference genome and GENCODE 19 track.

ecDNA co-ocurrence analysis in TR14 cells

Circle-enriched regions in single cells were defined as previously described. In each single cell, we search for overlaps between circle-enriched regions and ecDNA regions (*MYNC, CDK4, MDM2*) as defined by ecDNA reconstructions in bulk. To find overlaps, we used the function findOverlaps from the R package Genomic Ranges(56) (version 1.44.0). “Presence” or “absence” of each ecDNA was defined based on whether an overlap with circle-enriched regions was found for each of the three *MYNC, CDK4, MDM2* ecDNAs independently, excluding the common regions present in both *MYCN* and *CDK4* ecDNAs.

3. Results

Circular DNA enrichment and sequencing approaches rely on the digestion of linear DNA by exonucleases, followed by enrichment of undigested circular DNA through rolling circle amplification (RCA) using the phi29 enzyme(4, 9, 35). We reasoned that circular DNA isolation methods could be scaled down to single cells and combined with Smart-seq2(38, 57) for parallel sequencing of circular DNA and mRNA in single cells. (Fig 1a). In short, in scEC&T-seq, individual cells are isolated using fluorescence-activated cell sorting (FACS), and the mRNA of each cell is physically separated from the DNA by poly(A)-capture with oligo-dT primers coupled to magnetic beads. The captured mRNA is then reverse transcribed into cDNA and amplified following the Smart-seq2 protocol(38). The separated DNA from the same cells is subjected to exonuclease digestion and the remaining DNA after digestion is amplified by RCA. Both the cDNA and exonuclease-digested DNA are sequenced using paired-end short-read Illumina and, in some cases, also long-read nanopore sequencing (Fig. 1a).

To benchmark the protocol, we applied scEC&T-seq to two well-characterized, ecDNA-containing neuroblastoma cell lines: CHP-212 and TR14. We tested both short (1-day) and long (5-days) exonuclease digestion regimens. As negative controls, some cell's DNA was left undigested. Additionally, the DNA of a subset of cells was subjected to PmeI endonuclease digestion prior to exonuclease treatment, rendering most circular DNAs susceptible to exonuclease digestion (Fig. 1b). Circular DNA regions were computationally inferred using split, circle-junction spanning reads detection, as previously described in Koche *et al.*, 2020(4). Mitochondrial DNA (mtDNA) enrichment was first evaluated to assess the performance of the method, as mtDNA is circular, extrachromosomal, present in all cells, and susceptible to PmeI endonuclease digestion. We observed a significantly higher fraction of reads mapping to mtDNA after a prolonged 5-day exposure to exonuclease digestion compared to non-digested controls (p-value < 2.22e-16, two-sided Welch's t-test, Fig 1c). In addition to mtDNA enrichment, we detected a significantly higher fraction of reads mapping to all other computationally inferred circular DNA regions after short, 1-day exonuclease digestion and even higher after long, 5-days digestion (p-value < 2.22e-16, two-sided Welch's t-test, Fig. 1d). Endonuclease digestion with PmeI prior to exonuclease digestion almost completely depleted reads mapping to mtDNA and all other circular DNA regions containing PmeI cutting sites (p-value < 2.22e-16, two-

sided Welch's t-test, Fig 1c,e). These results demonstrate significant and specific enrichment of circular DNA by scEC&T-seq.

In parallel, the amplified cDNA from each single cell was subjected to paired-end Illumina sequencing (Fig 1a). In summary, we identified an average of 9,058 +- 1,163 (mean +- sd) transcripts per cell. The majority of cells (90,3 %) passed quality control standards, defined by low mitochondrial gene expression levels (<15%) and detection of more than 5,000 features. Differential transcription analyses allowed the separation of single cells into clusters defined by their cell line of origin (Fig. 1f). Altogether, these results demonstrate the effectiveness of scEC&T-seq in achieving circular DNA enrichment and parallel mRNA profiling in single cells.

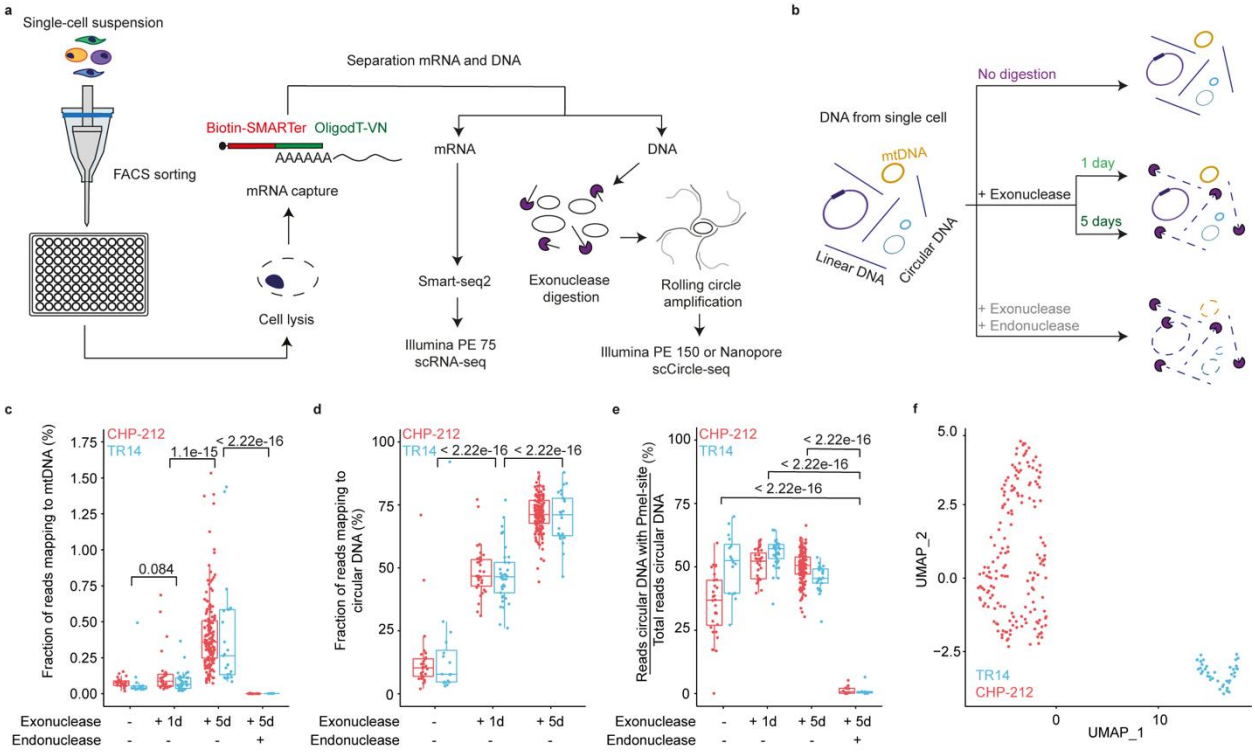


Figure 1: scEC&T-seq enables parallel detection of extrachromosomal circular DNAs and transcriptomes in single cancer cells. a, Schematic representation of the scEC&T-seq method. **b,** Scheme outlining all experimental conditions tested and their expected outcomes. **c-e,** Percentage of reads mapping to mtDNA (c), all circular DNA regions (d), and circular DNA regions containing PmeI targetting sites in each experimental condition described in (b) in CHP-212 (red) and TR14 (blue) cells. The sample size is the same across conditions: no digestion (*n* = 16 TR14 cells, *n* = 28 CHP-212 cells), 1-day exonuclease digestion (*n* = 37 TR14 cells, *n* = 31 CHP-212 cells), 5-days exonuclease digestion (*n* = 25 TR14 cells, *n* = 150 CHP-212 cells) and PmeI endonuclease digestion prior to 5-days exonuclease digestion (*n* = 6 TR14 cells, *n* = 12 CHP-212 cells). **f,** UMAP visualization displaying clusters of transcriptionally similar cells colored by cell line identity (*n* = 171 CHP-212 cells in red, *n* = 42 TR14 cells in blue). All statistical analyses correspond to two-sided Welch's t-test and p-values are shown. Figure modified from Chamorro González *et al*.(37).

We next focused on the regions of the genome that were identified as circular DNAs. Consistently with bulk population studies, we observed that DNA circularization occurs genome-wide in single cancer cells(4) (Fig. 2a). The number of individual circular DNAs identified per single neuroblastoma cell was variable and ranged between 97 and 1,939 (median = 702). The circular DNA size distribution was similar across single cancer cells and to that previously observed in bulk cancer populations(4). As expected, mega-base-sized circular DNA regions containing hallmark neuroblastoma genes were recurrently identified in most TR14 and CHP-212 cells (Fig. 2a). These circular regions included the *MYCN* oncogene, detected in both cell lines, as well as the *CDK4* and *MDM2* oncogenes, which were found exclusively in TR14 cells. These large, oncogene-containing circular DNAs were highly amplified as evidenced by increased coverage in WGS data, and were, therefore, classified as ecDNA (Fig 2a). Besides ecDNAs, hundreds of other circular DNA regions were identified per cell. Most of these circular DNAs were small (<100kb), non-amplified, and did not harbor oncogenes, therefore, they were classified as small circular DNAs (Fig 2b). We next investigated the recurrence of each circular DNA type and observed that while large, oncogene-containing ecDNAs were recurrently identified in most cells, all other small circular DNAs were mostly private to each cell and rarely shared among single cells (Fig 2c,d).

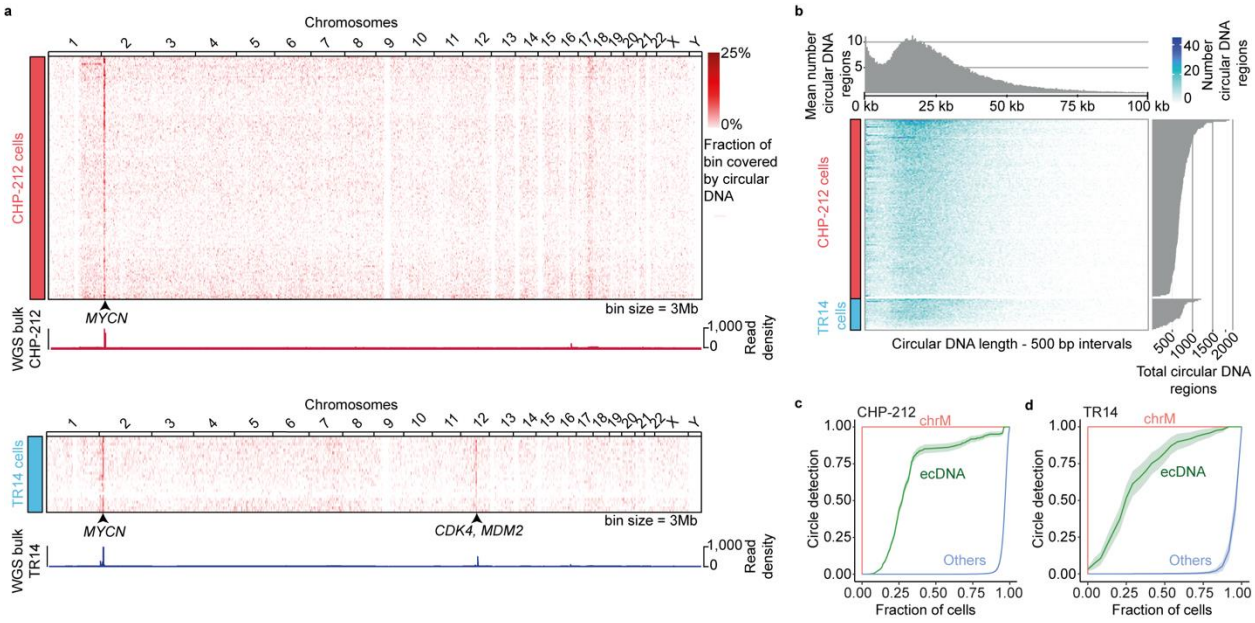


Figure 2: ecDNAs harboring oncogenes are recurrently detected in most neuroblastoma cells. a, Genome-wide circular DNA density in CHP-212 and TR14 cells (Top heat map: *n* = 150 CHP-212 cells; bottom heat map: *n* = 25 TR14 cells, bin size = 3Mb), and WGS tracks displaying read density in bulk populations. **b,** Number and size of individual circular DNA regions (<100kb) identified in CHP-212 and TR14 neuroblastoma single cells (*n* = 150 CHP-212 cells, *n* = 25 TR14 cells; bin size = 500 bp) with density distribution for circular DNA length (top) and total individual circular DNA counts (right). **c, d,** Fraction of cells containing a detected circular DNA from each circular DNA type in

CHP-212 (c, *n* = 150 CHP-212 cells) and TR14 cells (d, *n* = 25 TR14 cells). ecDNA was defined as large, copy-number amplified circular DNAs (green); mitochondrial DNA or chrM (red); “Others” are defined as all other circular DNAs (blue). Figure modified from Chamorro González *et al*.(37).

We then focused on circular DNA regions classified as ecDNA. EcDNAs often exhibit complex structures, with multiple regions from the same or different chromosomes rearranged to form interconnected circular structures(15). Therefore, we next evaluated whether scEC&T-seq can accurately capture the structural complexity of ecDNA in single cells. To do this, we compared scEC&T-seq results with previously described ecDNA reconstructions from WGS data of CHP-212 and TR14 bulk cell populations(16) (Fig. 3a,b). Indeed, scEC&T-seq successfully captured multi-fragment ecDNAs in nearly all single cells, reproducing the known element structures identified in bulk populations (Fig. 3a,b). In CHP-212 single cells, the *MYCN*-containing ecDNA was composed of six distinct chromosomal fragments, which also contained other genes such as *LPIN1*, *DDX1* and *TRIB2* (Fig. 3a). In TR14, three independent ecDNAs were detected. Two of these ecDNAs were multi-fragmented, and some overlapped in their location (Fig. 3b). One ecDNA that harbored *MDM2*, on the other hand, was very simple in structure and only contained one fragment (Fig. 3b). As observed in CHP-212 cells, genes other than *CDK4*, *MYCN* and *MDM2* were co-amplified alongside the oncogenes on these ecDNAs (Fig. 3b). Next, we focus on breakpoint sites interconnecting ecDNA segments and search for reads supporting each of these structural variants (SV) in single cells. At least one variant-supporting read per ecDNA breakpoint was detected in about 30% of single cells. Furthermore, using local sequence assembly-based methods, we computationally detected all ecDNA breakpoints as structural variants in merged scEC&T-seq data. We then evaluated whether inter-connected ecDNA segments could result in aberrant fusion gene expression. Indeed, novel fusion transcripts from merged scEC&T-seq data could be identified (Fig. 3c). Thus, scEC&T-seq enables the detection of complex multi-fragment ecDNA, and the ability to measure full transcripts from the same cells allows for the analysis of resulting fusion gene expression.

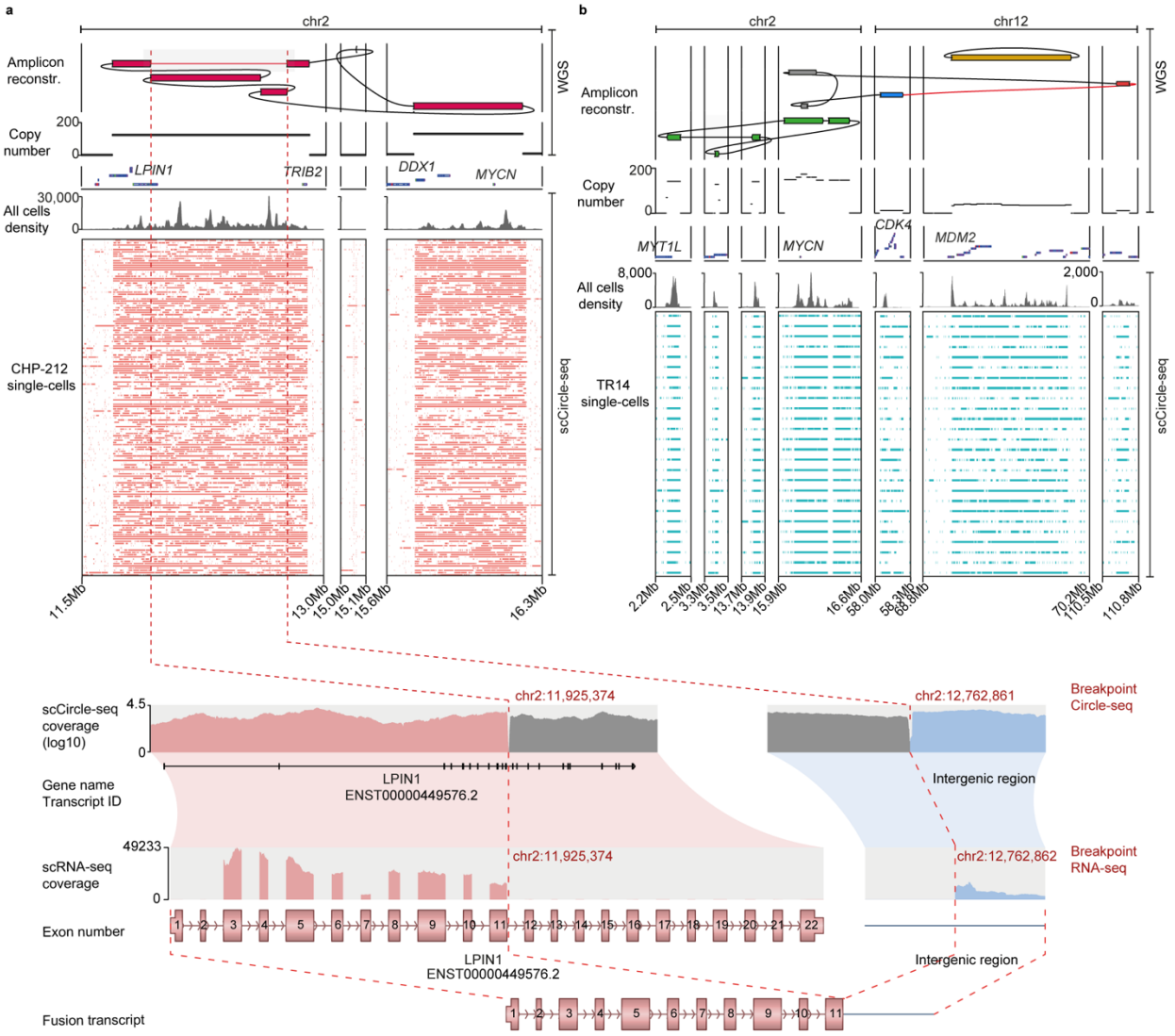


Figure 3: scEC&T-seq captures the complex structure of multi-fragmented ecDNAs in neuroblastoma cells. a, b, ecDNA reconstructions from WGS data in bulk populations and scEC&T-seq read coverage over ecDNA segments in CHP-212 (a; *n* = 150) and TR14 cells (b; *n* = 25). Top to bottom: ecDNA reconstruction, copy number profile, gene annotations, merged scEC&T-seq read density over ecDNA segments, coverage over ecDNA segments in single cells (rows). **c,** Fusion transcript detected by scEC&T-seq resulting from a breakpoint of the newly interconnected *MYCN* ecDNA structure in CHP-212 cells. Top to bottom: single-cell Circle-seq (scCircle-seq) read coverage over the breakpoint region in merged CHP-212 single cells (log-scaled), transcript annotations, single-cell RNA-seq (scRNA-seq) read coverage over the fused transcripts in merged CHP-212 single cells, native transcripts representations, fusion transcript representation. The interconnected genomic segments in *MYCN* ecDNA that give rise to the fusion gene are indicated by a red dotted line and the SV involved is highlighted in red. Figure modified from Chamorro González *et al*.(37).

We next focused on describing intercellular differences in ecDNA content in single cells. As three independent ecDNAs were identified in TR14 (Fig 3b and 4a), we investigated their co-occurrence in individual cells. We observed that most cells carried all three independent ecDNAs, with only a subset of cells carrying either only the *MDM2*-containing ecDNA or both the *MDM2* and *MYCN*-containing ecDNAs (Fig. 4b,c).

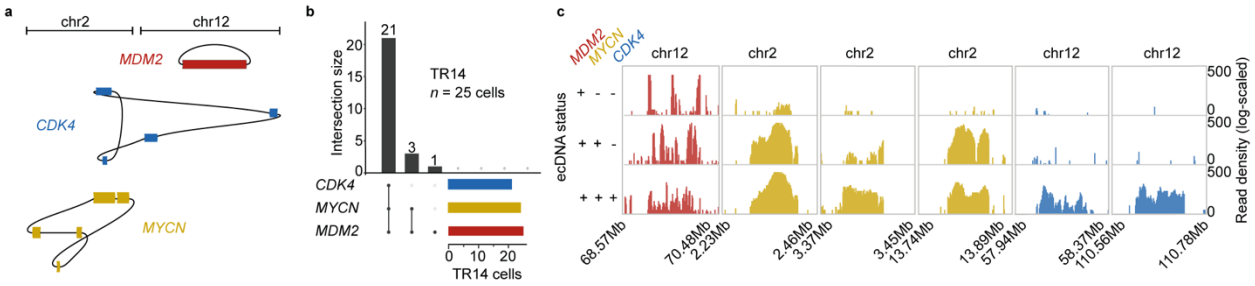


Figure 4: scEC&T-seq reveals co-ocurrence of independent ecDNAs in most TR14 cells. a, Schematic representation of the three ecDNAs identified in TR14 cells: *MYCN* (yellow), *CDK4* (blue) and *MDM2* (red) ecDNAs. **b,** Upset plot describing the co-occurrence of ecDNAs in TR14 cells (*n* = 25 TR14 cells). **c,** Log-scaled read densities over all ecDNA segments in three exemplary TR14 cells. Figure modified from Chamorro González *et al*.(37).

Next, as scEC&T-seq allows for measuring both mRNA and ecDNA from the same single cells, we investigated whether ecDNA copy number heterogeneity can influence the expression of genes present on ecDNA. As linear DNA is not captured by our method, absolute copy number and cell ploidy estimations are not feasible. Instead, we compared relative differences in ecDNA content among single cells as measured by normalized number of reads mapping to ecDNA-specific regions. Indeed, ecDNA-specific scCircle-seq read counts positively correlated with scRNA-seq read counts of genes on ecDNA suggesting that ecDNA copy number is a main driver of oncogene expression differences at the single-cell level (Fig 5a-d). These findings indicate that scEC&T-seq can effectively describe intercellular ecDNA content heterogeneity in single cells and evaluate its impact on gene expression.

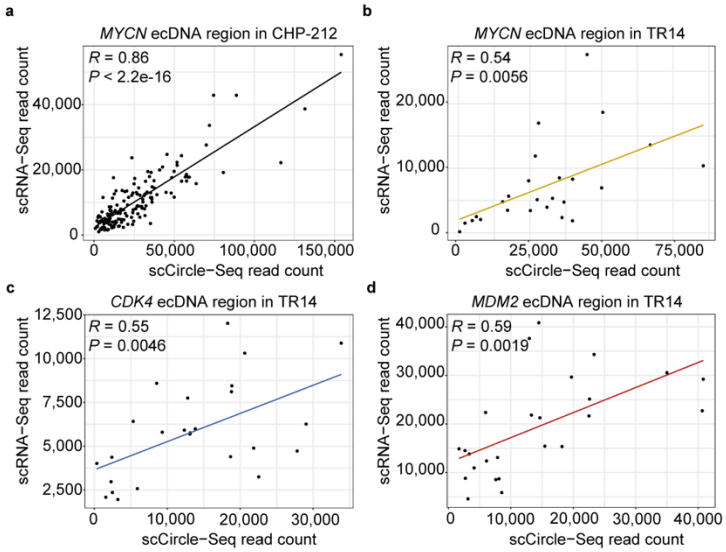


Figure 5: ecDNA relative copy number is positively correlated with transcript levels of genes encoded on ecDNA. a, b, Pairwise comparison between ecDNA-specific and mRNA read counts over the *MYCN* ecDNA region in CHP-212 single cells (e) and in TR14 single cells (f). **c, d,** Pairwise comparison between ecDNA-specific and mRNA read counts over the *CDK4* (g) and *MDM2* (h) ecDNA regions in TR14 single cells. All statistical analyses correspond

to two-sided Pearson correlation and p-values and regression coefficients are shown. Figure modified from Chamorro González *et al*.(37).

As previously shown in figure 2, in addition to identifying oncogene-containing ecDNAs, we also discovered hundreds of other small circular DNAs per single cancer cell (Fig. 2). We reasoned that parallel detection of circular DNA and mRNA in single cells may enable us to study pathways associated with their presence, which may either point to mechanisms of circular DNA generation or to the biological impact these circular DNAs have on cells. The size distribution of circular DNA in single cancer cells indicated the existence of at least two subtypes of small circular DNAs based on size (Fig. 2b). For this analysis, we focused primarily on one subtype of very small circular DNAs (less than 3kb) that had been previously described and referred to in the literature as microDNAs or apoptosis-derived eccDNAs, among other names(9, 58). To assess the relative abundance of these very small circular DNAs, we calculated their fraction within each CHP-212 cell. Based on this value, we categorized cells as either "high" or "low" in terms of the relative content of very small circular DNAs. Differential expression and pathway enrichment analyses were then performed to compare the "high" and "low" cell categories. We found that multiple pathways involving DNA fragmentation were associated with higher content of this circular DNA subtype (Fig. 6a-c). This is consistent with previous studies that have linked apoptosis, DNA repair, and DNA damage response to an increased abundance of these small circular DNAs(9, 31, 59-61).

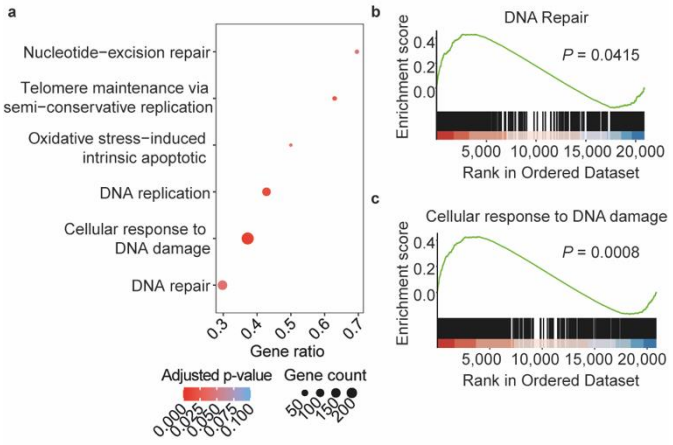


Figure 6: scEC&T-seq identifies pathways associated with high content of very small circular DNAs. a, Significantly enriched pathways associated with high relative content of very small circular DNAs in CHP-212 cells. Gene counts and adjusted p-values are shown. **b,c,** Gene Set Enrichment Analysis (GSEA) plots of genes involved in DNA repair (b; Adjusted p-value = 0.0415) and in cellular response to DNA damage (c; Adjusted p-value = 0.0008). The Bejamini-Hochberg method was used for p-value adjustment. Figure modified from Chamorro González *et al*.(37).

After validating scEC&T-seq in cell line cells, we tested its ability to profile intercellular circular DNA heterogeneity in primary patient samples. We successfully performed scEC&T-seq in single nuclei derived from two neuroblastoma primary tumors, as well as in single T-cells isolated from the blood of two different neuroblastoma patients. Genome-wide DNA circularization was detected in both single T-cells and tumor nuclei, consistent with our observations in cell lines (Fig. 7a). Previous reports have suggested that genetically unstable cells exhibit higher numbers of circular DNAs(6). Consistent with these findings, we observed a significantly larger number of individual circular DNA regions in cancer cells from both tumors and cell lines compared to non-malignant T-cells. (Fig. 7b). In line with their known oncogenic role, *MYCN* ecDNAs were exclusively identified in tumor nuclei and were absent in normal T-cells (Fig. 7a). These results demonstrate that scEC&T-seq is suitable for reproducibly profiling circular DNA content, regardless of the input sample.

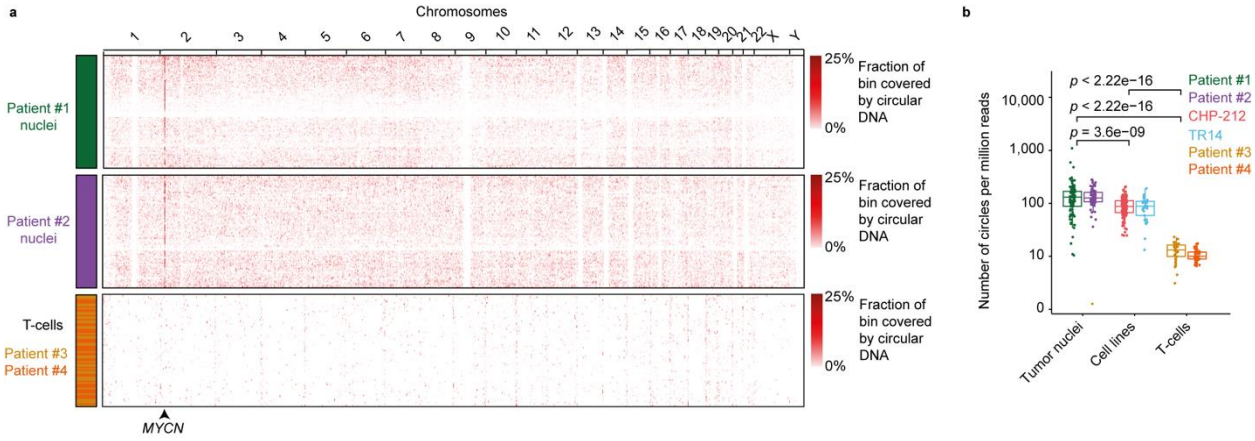


Figure 7: scEC&T-seq successfully profiles circular DNA at the single-cell level in primary patient samples. a, Genome-wide circular DNA density in neuroblastoma primary tumor nuclei and normal T-cells (*n* = 93 patient #1, green; *n* = 86 patient #2, purple; *n* = 38 patient #3, yellow; *n* = 41 patient #4, orange; bin sizes = 3 Mb). **b**, Normalized number of circular DNAs detected in primary tumor nuclei (*n* = 93 nuclei patient #1, *n* = 86 nuclei patient #2), neuroblastoma cell line cells (*n* = 25 TR14 cells, *n* = 150 CHP-212 cells), and normal T-cells (*n* = 38 patient #3, *n* = 41 patient #4). *P*-values were calculated using two-sided Welch's *t*-test and are shown. Boxplot's boxes represent 25th and 75th percentile with center bar as median value and whiskers represent furthest outlier ≤1.5× the interquartile range from the box. Figure modified from Chamorro González *et al*.(37).

Finally, we investigated whether scEC&T-seq could provide insights into structural ecDNA heterogeneity in tumors. We first reconstructed the consensus ecDNA structure in primary tumors using bulk WGS data and compare them to our scEC&T-seq data (Fig. 8a). In one patient (#2), the reconstructed *MYCN* ecDNA structure was complex, composed of five fragments derived from chromosome 2 and interconnected by four structural variants (SV1-SV4; Fig. 8a). While most cells from patient #2 displayed the consensus

ecDNA structure reconstructed in bulk (94.2%), a small subset of cells showed signs of structural heterogeneity across the 6kb deletion flanked by structural variants SV2 and SV3 (Fig. 8a,b). This subset of cells showed sequencing read coverage across the 6kb region, while all others were consistent with bulk predictions and presented a deletion with no reads mapping to it (Fig. 8a,b). We then search for split-read support for the structural variants associated with the 6kb deletion (SV2 and SV3) in this subset of cells (Fig. 8c). Based on the presence or absence of split-read support for SV2 and SV3 variants, we differentiated two subpopulations in this subset of cells: subclone #1, in which no split reads supporting SV2 or SV3 were identified; and subclone #2, in which reads supporting SV2 and/or SV3 were identified (Fig. 8c). The remaining cells, defined by the presence of the 6kb deletion, were classified as subclone #3 (Fig. 8c). To resolve the structure of the ecDNA variants present in each identified subclone, we analysed split-read support for all structural variants predicted in the consensus ecDNA structure from bulk, including SV1 and SV4 (Fig. 8d). Subclone #1, showed split-read support exclusively for structural variant SV1. This suggests a simple ecDNA structure made up of one unique chromosome 2 fragment connected head-to-tail by SV1, which explains the absence of support for the 6kb deletion (Fig. 8d). We named this simple ecDNA variant ecDNA v1 (Fig. 8e). Subclone #2, showed split-read support for SV1-3, but not for SV4 (Fig. 8d). This indicates that another ecDNA variant, lacking a large 180kb deletion explained by SV4 in the consensus ecDNA structure, exists in subclone #2 cells, which we named ecDNA v2 (Fig. 8e). Finally, subclone #3, showed support for all structural variants SV1-4, indicating the presence of the consensus ecDNA variant, which we termed ecDNA v3 (Fig. 8d,e). While subclones #1 and #3 are pure populations presenting ecDNA variants v1 or v3, respectively, subclone #2 could also potentially be the result of a mixed population of cells containing ecDNA variants v2 and, additionally, v1.

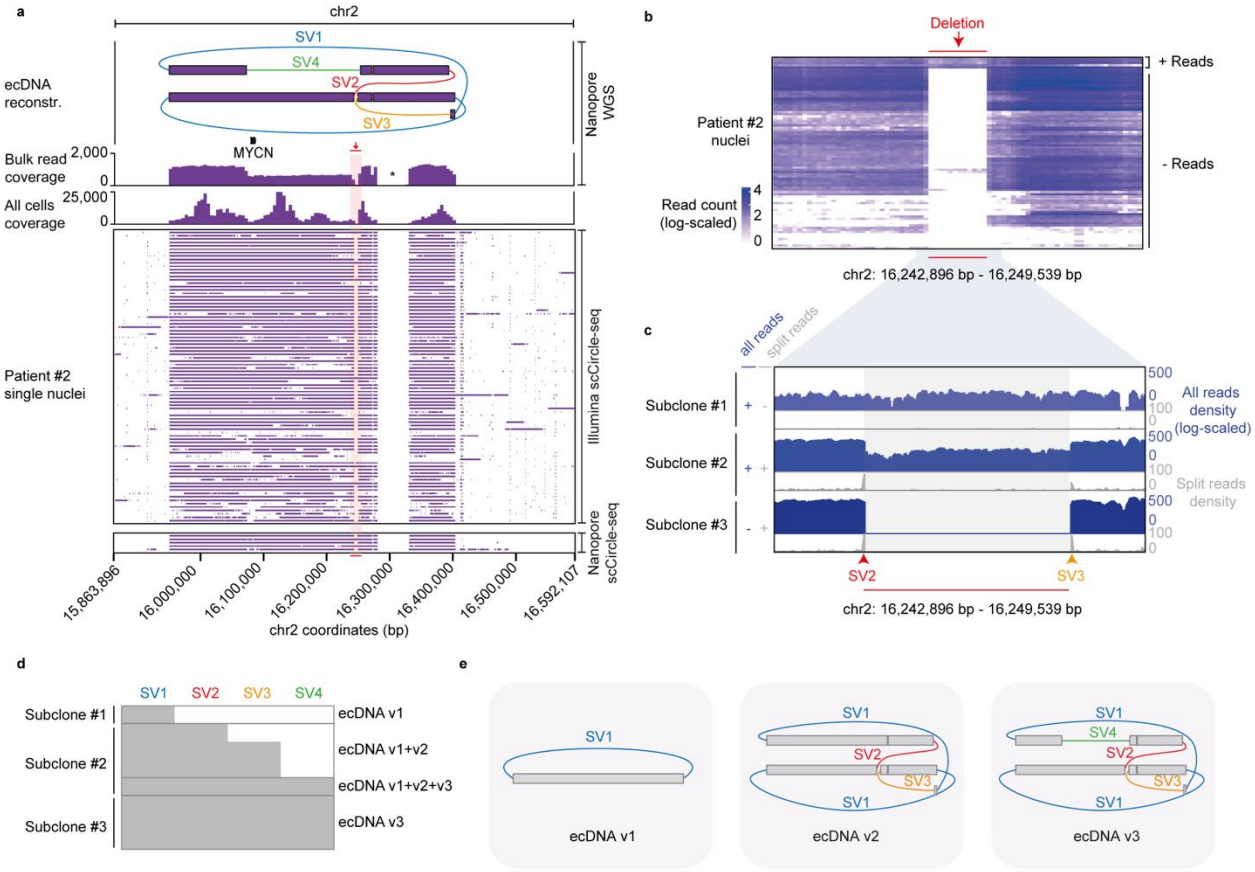


Figure 8: scEC&T-seq depicts structural ecDNA heterogeneity at the single-cell level in neuroblastoma tumors. **a,** ecDNA reconstruction from bulk WGS data and nanopore long-read or illumina short-read scEC&T-seq coverage over ecDNA segments in patient #2 (*n* = 86 patient #2 nuclei). Top to bottom: graph representing the reconstructed ecDNA (structural variants are colored, SV#1-4), gene annotation, read density over ecDNA segments in bulk WGS data, read density over ecDNA segments in merged scEC&T-seq data, long-read or short-read scEC&T-seq coverage over ecDNA segments in single nuclei (rows). The 6 kb deletion is marked in red. Unmappable region of the reference genome (hg19) is indicated by an asterisk. **b,** Heatmap of the log-scaled number of reads in 500 bp bins over the 6 kb deletion on the reconstructed ecDNA across single nuclei in patient #2 (*n* = 86 patient #2 nuclei). **c,** Exemplary genome tracks of clone variants in patient #2. Log-scaled total read density (blue) and circle edge-supporting read density (grey) are shown. **d,** Detection of reads supporting the SV#1-4 present in the consensus ecDNA reconstructed in bulk in 8 exemplary single cells representing each clone variant (1-3) (>= 1 read supporting the SV, grey; 0 reads supporting the SV, white). **e,** Schematic representation of the three ecDNA variants v1-3 identified by scEC&T-seq. Figure modified from Chamorro González *et al*.(37).

The simplest sequence of events that could explain the presence of all different ecDNA variants identified in patient #2's cells should start with the excision and circularization of the chromosome 2 region containing *MYCN* and its four local enhancers (e1-e4), generating ecDNA v1 (Fig. 8e and Fig. 9). Next, the fusion of two ecDNA v1 molecules would generate the more complex ecDNA variant v2, giving rise to structural variants SV2 and SV3 and the 6kb deletion (Fig. 8e and Fig. 9). This new variant, ecDNA v2, would present two copies of the *MYCN* oncogene and all four enhancers. Finally, an additional large deletion on ecDNA v2 would create SV4 leading to the last and most prevalent ecDNA variant, v3, resulting in the loss of one copy of *MYCN* and enhancers e2 and e3 (Fig. 8e

and Fig. 9). These findings demonstrate that using scEC&T-seq, not only we can describe intercellular structural ecDNA heterogeneity in primary neuroblastomas, but also infer the evolutionary dynamics of the ecDNA structure in the cancer population.

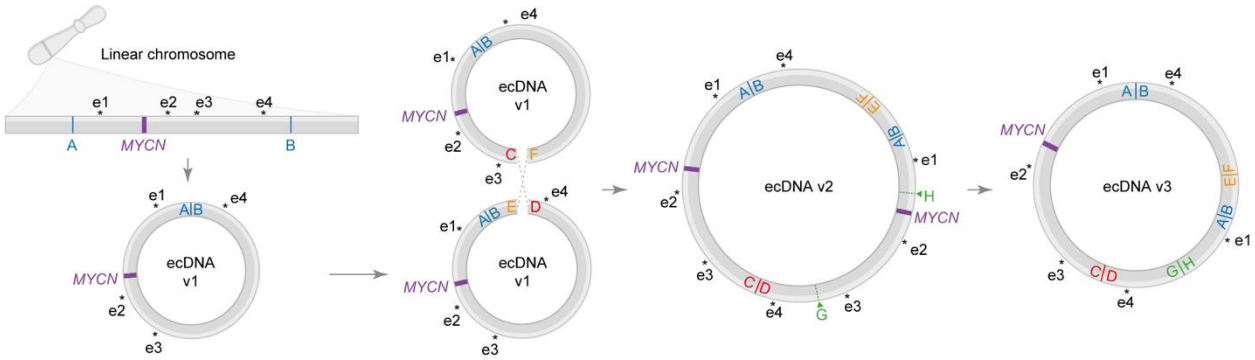


Figure 9: scEC&T-seq enables the inference of ecDNA structural dynamics in primary tumors. Schematic representation of the predicted structural evolution of ecDNA in patient #2. The location of the *MYCN* oncogene and its four local enhancers (e1-e5) are indicated. Figure modified from Chamorro González *et al*.(37).

4. Discussion

The increasing evidence of the impact of extrachromosomal circular DNA on tumor heterogeneity and cancer evolution emphasizes the need for methods to interrogate it at single-cell resolution. In this project, we developed scEC&T-seq, a method for parallel profiling of extrachromosomal circular DNAs and transcriptomes in single cancer cells. We showed its efficacy in characterizing all types of circular DNA elements in single cells, regardless of their size or genomic origin. The integrated analysis of circular DNA and gene expression data allowed the interrogation of the transcriptional consequences of ecDNA copy number heterogeneity. We further demonstrated its potential to unravel principles of ecDNA structural evolution by its ability to describe differences in ecDNA structure at the single-cell level. We anticipate that scEC&T-seq will serve as a useful tool to enhance our understanding of the biology of both cancer-specific ecDNAs and small circular DNAs.

We successfully applied scEC&T-seq to neuroblastoma cell lines and tumors, in which we identified hundreds of circular DNAs per single cell, most of which were small and didn’t carry oncogenes. Our data revealed that while oncogene-containing ecDNAs were recurrently detected in most cells, the majority of small circular DNAs were exclusive to individual cells. A recent study that used mathematical modeling to explore the impact of random segregation of extrachromosomal elements in cancer populations suggested that only positively selected extrachromosomal elements are retained over time, while those under neutral selection are eventually lost(12). The high recurrence of ecDNA, in line with their recognized oncogenic role, supports the notion that ecDNA confers a positive selective advantage to the cancer population. In contrast, although the functional relevance of small circular DNAs in cancer cannot be completely excluded, their low recurrence suggests they may not contribute to cancer cell fitness to the same degree as ecDNAs. This also implies the existence of yet unknown requirements for the selection, propagation, and maintenance of these small circular DNAs. Beyond cancer, the functional relevance of small circular DNAs in normal development and other disease contexts also remains largely unknown. Our study suggests that integrating small circular DNA detection and mRNA sequencing may facilitate the discovery of molecular mechanisms associated with the presence of small circular DNA in various biological contexts.

Multiple ecDNA structures can co-exist within a tumor(20, 23, 25), however, little is known about their distribution in the cancer population. In the neuroblastoma TR14 cell line, for instance, three independent ecDNAs carrying different oncogenes (*MYCN*, *CDK4* or *MDM2)* have been identified(23). By applying scEC&T-seq to TR14 cells, we observed that while a small subset of cells carried only one or two of the ecDNAs, the majority of cells harbored all three different ecDNAs. These findings suggest that maintaining all three independent ecDNAs may confer a strong fitness advantage. Recent investigations have demonstrated that distinct ecDNAs can congregate and interact in nuclear hubs, enabling intermolecular gene activation, where enhancer elements on one ecDNA activate genes on another ecDNA(20, 23). These interactions could potentially explain the advantage of maintaining a mixed population of cooperating ecDNAs within the same cell. However, further investigations are required to determine whether positive selection alone is sufficient to ensure the co-maintenance of randomly segregating ecDNAs. It is tempting to speculate that co-segregation or co-selection mechanisms might exist to sustain their co-inheritance.

The uneven segregation of ecDNA is expected to result in high intercellular variation in copy number. Consistently, our scEC&T-seq data revealed intercellular heterogeneity in ecDNA copy number in neuroblastoma populations. Interestingly, we observed a positive correlation between ecDNA copy number and the expression levels of genes carried on ecDNA. While it is plausible that enhancer-oncogene interactions of ecDNAs within nuclear hubs may also contribute to regulating gene expression levels(23), our results provide evidence that ecDNA copy number heterogeneity is a major driver of oncogene expression differences in single cells. In a separate study, we analyzed additional single-cell datasets from *MYCN*-amplified neuroblastoma tumors and revealed significant heterogeneity in *MYCN* expression at the single-cell level(62). We observed that cells exhibiting high *MYCN* expression levels also showed elevated expression of *MYCN* target genes(62). This indicates that heterogeneity in oncogene expression, which can potentially be driven by *MYCN*-ecDNA copy number, is biologically functional and directly impacts gene expression patterns associated with *MYCN*.

Furthermore, we demonstrated the capability of scEC&T-seq to reconstruct complex ecDNAs structures and profile intercellular differences in ecDNA structure. In a neuroblastoma tumor, for instance, scEC&T-seq identified three different ecDNA structural variants

(v1, v2, v3), enabling the inference of ecDNA structural evolutionary dynamics within the tumor. The mechanisms of ecDNA structural dynamics described in this study are consistent with recent models that propose circular recombination as one of the primary mechanisms driving ecDNA evolution in tumors(63). The reason behind the selective advantage conferred by the more complex ecDNA structure (v3) can only be speculated. One possible explanation is the change in enhancer:oncogene stoichiometry, where the loss of one copy of the *MYCN* oncogene after the large deletion may be compensated by more efficient use of enhancers on ecDNA v3. This means ecDNA v3 has six enhancers to regulate one copy of the gene, compared to ecDNA v2, which has eight enhancers to regulate two copies of the gene. Exploring the structural dynamics of ecDNA using scEC&T-seq holds promise for future investigations aimed at addressing critical questions regarding the origin and evolution of ecDNA.

Current single-cell methods are limited in their ability to characterize extrachromosomal circular DNAs in single cells. While certain single-cell DNA and RNA sequencing methods, such as single-cell ATAC-seq(36) and G&T-seq(38), can detect ecDNAs due to their high level of amplification, they are unable to identify non-amplified, small circular DNAs and only rely on the computational inference of circularity to distinguish ecDNA from linear amplifications. In contrast, scEC&T-seq relies on specific circular DNA enrichment through exonuclease-mediated depletion of linear DNA. Therefore, it represents the first single-cell method capable of characterizing potentially all extrachromosomal circular DNA types in single cells, regardless of their size or amplification status. As a drawback, the elimination of linear DNA prevents the estimation of ecDNA absolute copy number and the tracing of circular DNA re-integration. One additional current limitation of scEC&T-seq is its low throughput. While droplet-based single-cell methods can assay thousands of cells in one run, scEC&T-seq is a plate-based sequencing method, therefore, the number of cells that can be processed in parallel is limited to hundreds. This also translates into more elaborate and expensive laboratory procedures. Future investigations focusing on increasing the method’s throughput will, therefore, be critical. Nonetheless, these limitations come with a unique set of advantages that make scEC&T-seq a valuable tool for investigating intercellular heterogeneity of extrachromosomal circular DNA. For instance, scEC&T-seq generates independent sequencing libraries per cell, enabling re-sequencing for higher sequencing coverage when a detailed analysis of a particular cell is re-

quired. Additionally, the isolated cDNA and DNA from each cell can be subjected to additional testing and sequencing using other technologies, such as nanopore long-read sequencing. The use of longer reads might facilitate the analysis of complex circular DNA structures in single cells. Unlike droplet-based methods, scEC&T-seq allows the detection of full-length transcripts, enabling the identification of fusion transcripts directly derived from ecDNA breakpoints, as demonstrated in this study. Thus, scEC&T-seq provides a much higher level of detail compared to high-throughput methods, enabling in-depth assessment of ecDNA sequence, structure, and transcription at the single-cell level.

Over the past decade, the rediscovery of ecDNA as a major driver of oncogene amplification in aggressive tumors has made it an attractive target for therapeutic intervention(1). However, the high level of heterogeneity and complexity associated with extrachromosomal amplifications pose a significant challenge. Despite significant progress in recent years, our understanding of the intercellular heterogeneity and dynamics of ecDNA during tumor development and adaptation to therapy is still incomplete. In part, many questions remain unresolved due to our limited ability to deconvolute ecDNA diversity from bulk sequencing data or imaging-based approaches. Therefore, single-cell sequencing methods are crucial for improving the characterization of intercellular ecDNA heterogeneity in tumors and gaining insights into the mechanisms underlying ecDNA dynamics and evolution in cancer. We envision that applying scEC&T-seq to longitudinal tumor samples before and after treatment will provide crucial data to further investigate the role of ecDNA-driven intercellular heterogeneity, its dynamics, and its functional consequences in cancer evolution. Integrating this information with data from other single-cell techniques has the potential to inform and shape future therapeutic approaches targeting ecDNA-driven tumors. Besides cancer-specific ecDNAs, scEC&T-seq allows for profiling additional small, non-amplified circular DNA elements in human cells, which was previously unattainable with other single-cell methods. As a result, scEC&T-seq holds the potential to shed light on unresolved questions regarding the maintenance, selection, and function of small circular DNAs in cancer and other biological contexts. Consequently, we anticipate that our method will serve as a valuable resource for future research across diverse fields, extending beyond extrachromosomal oncogene amplification and encompassing many currently unanswered biological questions.

5. Reference List

1. Yi E, Chamorro González R, Henssen AG, Verhaak RGW. Extrachromosomal DNA amplifications in cancer. Nature Reviews Genetics. 2022.
2. Møller HD, Mohiyuddin M, Prada-Luengo I, Sailani MR, Halling JF, Plomgaard P, Maretty L, Hansen AJ, Snyder MP, Pilegaard H, Lam HYK, Regenberg B. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. Nat Commun. 2018;9(1):1069.
3. 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.
4. Koche RP, Rodriguez-Fos E, Helmsauer K, Burkert M, MacArthur IC, Maag J, Chamorro R, Munoz-Perez N, Puiggròs M, Dorado Garcia H, Bei Y, Röefzaad 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.
5. Shibata Y, Kumar P, Layer R, Willcox S, Gagan JR, Griffith JD, Dutta A. Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. Science. 2012;336(6077):82-6.
6. Paulsen T, Kumar P, Koseoglu MM, Dutta A. Discoveries of Extrachromosomal Circles of DNA in Normal and Tumor Cells. Trends Genet. 2018;34(4):270-8.
7. Cox D, Yuncken C, Spriggs AI. MINUTE CHROMATIN BODIES IN MALIGNANT TUMOURS OF CHILDHOOD. Lancet. 1965;1(7402):55-8.
8. Hotta Y, Bassel A. MOLECULAR SIZE AND CIRCULARITY OF DNA IN CELLS OF MAMMALS AND HIGHER PLANTS. Proc Natl Acad Sci U S A. 1965;53(2):356-62.
9. Wang Y, Wang M, Djekidel MN, Chen H, Liu D, Alt FW, Zhang Y. eccDNAs are apoptotic products with high innate immunostimulatory activity. Nature. 2021;599(7884):308-14.
10. Cohen S, Regev A, Lavi S. Small polydispersed circular DNA (spcDNA) in human cells: association with genomic instability. Oncogene. 1997;14(8):977-85.
11. Henson JD, Cao Y, Huschtscha LI, Chang AC, Au AY, Pickett HA, Reddel RR. DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. Nat Biotechnol. 2009;27(12):1181-5.
12. 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.
13. 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.
14. Wahl GM. The importance of circular DNA in mammalian gene amplification. Cancer Res. 1989;49(6):1333-40.

15. Kim H, Nguyen N-P, 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.
16. Helmsauer K, Valieva ME, Ali S, Chamorro González R, Schöpflin R, Röefzaad C, Bei Y, Dorado Garcia H, Rodriguez-Fos E, Puiggròs 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(1):5823.
17. 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.
18. Nikolaev S, Santoni F, Garieri M, Makrythanasis P, Falconnet E, Guipponi M, Vannier A, Radovanovic I, Bena F, Forestier F, Schaller K, Dutoit V, Clement-Schatlo V, Dietrich P-Y, Antonarakis SE. Extrachromosomal driver mutations in glioblastoma and low-grade glioma. Nat Commun. 2014;5(1):5690.
19. Levan A, Levan G. Have double minutes functioning centromeres? Hereditas. 1978;88(1):81-92.
20. 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 RGW. Live-cell imaging shows uneven segregation of extrachromosomal DNA elements and transcriptionally active extrachromosomal DNA hubs in cancer. Cancer Discovery. 2021:candisc.1376.2021.
21. Lundberg G, Rosengren AH, Håkanson U, Stewénius H, Jin Y, Stewénius Y, Påhlman S, Gisselsson D. Binomial Mitotic Segregation of MYCN-Carrying Double Minutes in Neuroblastoma Illustrates the Role of Randomness in Oncogene Amplification. PLOS ONE. 2008;3(8):e3099.
22. Owen SC, Jens L, Sameena W, Ashutosh T, Meghana P, Shanqing W, Jon DL, Joshua TL, Ivy Tsz-Lo W, Siavash RD, Sahaana C, Miriam A, Yingxi L, Edwin J, James TR, Sunita S, Denise MM, Nicole C, Michael L, John RC, Scott LP, Jeremy R, Richard HS, Hannah C, Jesse D, Paul SM, Ernest F, Robert JW-R, Vineet B, Jill PM, Lukas C. The landscape of extrachromosomal circular DNA in medulloblastoma. bioRxiv. 2021:2021.10.18.464907.
23. 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-L, 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.
24. deCarvalho AC, Kim H, Poisson LM, Winn ME, Mueller C, Cherba D, Koeman J, Seth S, Protopopov A, Felicella M, Zheng S, Multani A, Jiang Y, Zhang J, Nam D-H, Petricoin EF, Chin L, Mikkelsen T, Verhaak RGW. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. Nat Genet. 2018;50(5):708-17.
25. Hung KL, Luebeck J, Dehkordi SR, Colón CI, Li R, Wong IT-L, Coruh C, Dharanipragada P, Lomeli SH, Weiser NE, Moriceau G, Zhang X, Bailey C, Houlahan KE, Yang W, González RC, Swanton C, Curtis C, Jamal-Hanjani M, Henssen AG, Law

JA, Greenleaf WJ, Lo RS, Mischel PS, Bafna V, Chang HY. Targeted profiling of human extrachromosomal DNA by CRISPR-CATCH. Nat Genet. 2022;54(11):1746-54.
26. Deshpande V, Luebeck J, Nguyen N-PD, 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.
27. 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.
28. 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.
29. Luebeck J, Ng AWT, Galipeau PC, Li X, Sanchez CA, Katz-Summercorn AC, Kim H, Jammula S, He Y, Lippman SM, Verhaak RGW, Maley CC, Alexandrov LB, Reid BJ, Fitzgerald RC, Paulson TG, Chang HY, Wu S, Bafna V, Mischel PS. Extrachromosomal DNA in the cancerous transformation of Barrett's oesophagus. Nature. 2023;616(7958):798-805.
30. Noer JB, Hørsdal OK, Xiang X, Luo Y, Regenberg B. Extrachromosomal circular DNA in cancer: history, current knowledge, and methods. Trends Genet. 2022;38(7):766-81.
31. Paulsen T, Malapati P, Shibata Y, Wilson B, Eki R, Benamar M, Abbas T, Dutta A. MicroDNA levels are dependent on MMEJ, repressed by c-NHEJ pathway, and stimulated by DNA damage. Nucleic Acids Res. 2021;49(20):11787-99.
32. Schwab M, Alitalo K, Klempnauer K-H, 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.
33. Seeger RC, Brodeur GM, Sather H, Dalton A, Siegel SE, Wong KY, Hammond D. Association of Multiple Copies of the N-myc Oncogene with Rapid Progression of Neuroblastomas. New England Journal of Medicine. 1985;313(18):1111-6.
34. 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.
35. Moller HD, Parsons L, Jorgensen TS, Botstein D, Regenberg B. Extrachromosomal circular DNA is common in yeast. Proc Natl Acad Sci U S A. 2015;112(24):E3114-22.
36. Kashima Y, Sakamoto Y, Kaneko K, Seki M, Suzuki Y, Suzuki A. Single-cell sequencing techniques from individual to multiomics analyses. Experimental & Molecular Medicine. 2020;52(9):1419-27.
37. Chamorro González R, Conrad T, Stöber MC, Xu R, Giurgiu M, Rodriguez-Fos E, Kasack K, Brückner L, van Leen E, Helmsauer K, Dorado Garcia H, Stefanova ME, Hung KL, Bei Y, Schmelz K, Lodrini M, Mundlos S, Chang HY, Deubzer HE, Sauer S, Eggert A, Schulte JH, Schwarz RF, Haase K, Koche RP, Henssen AG. Parallel sequencing of extrachromosomal circular DNAs and transcriptomes in single cancer cells. Nat Genet. 2023;55(5):880-90.
38. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, Goolam M, Saurat N, Coupland P, Shirley LM, Smith M, Van der Aa N, Banerjee R, Ellis PD, Quail MA, Swerdlow HP, Zernicka-Goetz M, Livesey FJ, Ponting CP, Voet T. G&T-seq: parallel

sequencing of single-cell genomes and transcriptomes. Nature Methods. 2015;12(6):519-22.
39. 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. Nat Genet. 2019.
40. Felix Krueger FJ, Phil Ewels, Ebrahim Afyounian, & Benjamin Schuster-Boeckler. FelixKrueger/TrimGalore: v0.6.7 - DOI via Zenodo (0.6.7). Zenodo. 2021.
41. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997. 2013.
42. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology. 2019;37(8):907-15.
43. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan J S, Jolanki O, Tanaka FY, Cherry J M. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 2018;46(D1):D794-D801.
44. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12(1):323.
45. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biology. 2014;15(6):R84.
46. 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.
47. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10(2).
48. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-87.e29.
49. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16(5):284-7.
50. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160-5.
51. 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.
52. Uhrig S, Ellermann J, Walther T, Burkhardt P, Fröhlich M, Hutter B, Toprak UH, Neumann O, Stenzinger A, Scholl C, Fröhling S, Brors B. Accurate and efficient detection of gene fusions from RNA sequencing data. Genome Res. 2021;31(3):448-60.
53. 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. Nat Commun. 2020;11(1):5823.
54. 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.
55. 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.
56. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for Computing and Annotating Genomic Ranges. PLOS Computational Biology. 2013;9(8):e1003118.
57. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nature Protocols. 2014;9(1):171-81.
58. Paulsen T, Shibata Y, Kumar P, Dillon L, Dutta A. Small extrachromosomal circular DNAs, microDNA, produce short regulatory RNAs that suppress gene expression independent of canonical promoters. Nucleic Acids Res. 2019;47(9):4586-96.
59. Dillon LW, Kumar P, Shibata Y, Wang YH, Willcox S, Griffith JD, Pommier Y, Takeda S, Dutta A. Production of Extrachromosomal MicroDNAs Is Linked to Mismatch Repair Pathways and Transcriptional Activity. Cell Rep. 2015;11(11):1749-59.
60. Sunnerhagen P, Sjöberg RM, Karlsson AL, Lundh L, Bjursell G. Molecular cloning and characterization of small polydisperse circular DNA from mouse 3T6 cells. Nucleic Acids Res. 1986;14(20):7823-38.
61. Huang C, Jia P, Chastain M, Shiva O, Chai W. The human CTC1/STN1/TEN1 complex regulates telomere maintenance in ALT cancer cells. Exp Cell Res. 2017;355(2):95-104.
62. Maja CS, Rocío Chamorro G, Lotte B, Thomas C, Nadine W, Annabell S, Angelika E, Johannes HS, Richard PK, Anton GH, Roland FS, Kerstin H. Intercellular extrachromosomal DNA copy number heterogeneity drives cancer cell state diversity. bioRxiv. 2023:2023.01.21.525014.
63. Rosswog C, Bartenhagen C, Welte A, Kahlert Y, Hemstedt N, Lorenz W, Cartolano M, Ackermann S, Perner S, Vogel W, Altmüller J, Nürnberg P, Hertwig F, Göhring G, Lilienweiss E, Stütz 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.

6. Statutory Declaration

“I, Maria del Rocío Chamorro González, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic “Dissecting extrachromosomal circular DNA heterogeneity in single cells with scEC&T-seq / Untersuchung der Heterogenität extrachromosomaler zirkulärer DNA in Einzelzellen mit scEC&T-seq“, independently and without the support of third parties, and that I used no other sources and aids than those stated.

All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons (cf. declaration of contribution). I have correctly marked all texts or parts of texts that were generated in collaboration with other persons.

My contributions to any publications to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; http://www.icmje.org) on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty.

The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me.”

Date Signature

7. Declaration of own contribution

María del Rocío Chamorro González contributed the following to the below listed publications:

Publication 1: Rocío Chamorro González, Thomas Conrad, Maja C. Stöber, Robin Xu, Madalina Giurgiu, Elias Rodriguez-Fos, Katharina Kasack, Lotte Brückner, Eric van Leen, Konstantin Helmsauer, Heathcliff Dorado Garcia, Maria E. Stefanova, King L. Hung, Yi Bei, Karin Schmelz, Marco Lodrini, Stefan Mundlos, Howard Y. Chang, Hedwig E. Deubzer, Sascha Sauer, Angelika Eggert, Johannes H. Schulte, Roland F. Schwarz, Kerstin Haase, Richard P. Koche and Anton G. Henssen. Parallel sequencing of extrachromosomal circular DNAs and transcriptomes in single cancer cells. *Nature Genetics* (2023).

Contribution: Contributed to study design, collection, and interpretation of the data. Performed all the single-cell experiments from sample preparation to library preparation at the MDC Genomics Technology Platform. Performed nanopore sequencing experiments in bulk populations from sample preparation to library sequencing. Performed bulk Circle-seq experiments. All libraries were sequenced at the MDC Genomics Technology Platform, except for nanopore libraries. Performed single-cell and bulk Circle-seq data analysis, including the analyses presented in Figure 1c-f; Figure 2a,b; Figure 3a,b (excluding the ecDNA reconstructions); Figure 4b,c; Figure 5a,b; Figure 6b,c; Figure 7a,b,c,e (excluding the ecDNA reconstructions); and Supplementary Figures 1,4,6,8,14,16(a,f,g),18. Contributed to the study design, led the project, and wrote the manuscript related to this publication together with Prof. med. Dr. Anton G. Henssen and Dr. Richard Koche. Prepared all the figures presented in the manuscript.

_____

Signature, date and stamp of first supervising university professor / lecturer

_____

Signature of doctoral candidate

8. Printing copy(s) of the publication(s)

nature genetics

Technical Report

https://doi.org/10.1038/s41588-023-01386-y

Parallel sequencing of extrachromosomal circular DNAs and transcriptomes in single cancer cells

Received: 20 December 2021

Accepted: 28 March 2023

Published online: 4 May 2023

Check for updates

Rocío Chamorro González[1,2], Thomas Conrad[3], Maja C. Stöber[4,5,6], Robin Xu[1,2], Mădălina Giurgiu[1,2,7], Elias Rodriguez-Fos[1,2], Katharina Kasack[8], Lotte Brückner[2,9], Eric van Leen[1,2], Konstantin Helmsauer[1,2], Heathcliff Dorado Garcia[1,2], Maria E. Stefanova[10,11], King L. Hung[12], Yi Bei[1,2], Karin Schmelz[1], Marco Lodrini[1], Stefan Mundlos[10,11,13], Howard Y. Chang[12,14], Hedwig E. Deubzer[1,2,15,16], Sascha Sauer[4], Angelika Eggert[1,15,16], Johannes H. Schulte[1,15,16], Roland F. Schwarz[4,17,18], Kerstin Haase[1,2,15], Richard P. Koche[19,20] & Anton G. Henssen[1,2,9,15,20] ✉

Extrachromosomal DNAs (ecDNAs) are common in cancer, but many questions about their origin, structural dynamics and impact on intratumor heterogeneity are still unresolved. Here we describe single-cell extrachromosomal circular DNA and transcriptome sequencing (scEC&T-seq), a method for parallel sequencing of circular DNAs and full-length mRNA from single cells. By applying scEC&T-seq to cancer cells, we describe intercellular differences in ecDNA content while investigating their structural heterogeneity and transcriptional impact. Oncogene-containing ecDNAs were clonally present in cancer cells and drove intercellular oncogene expression differences. In contrast, other small circular DNAs were exclusive to individual cells, indicating differences in their selection and propagation. Intercellular differences in ecDNA structure pointed to circular recombination as a mechanism of ecDNA evolution. These results demonstrate scEC&T-seq as an approach to systematically characterize both small and large circular DNA in cancer cells, which will facilitate the analysis of these DNA elements in cancer and beyond.

Measuring multiple parameters in the same cells is key to accurately understand biological systems and their changes during diseases[1]. In the case of circular DNAs, it is critical to integrate DNA sequence information with transcriptional output measurements to assess their functional impact on cells. At least three types of circular DNAs can be distinguished in human cells[2–5]: (1) small circular DNAs (<100 kb)[6], which have been described under different names including eccDNAs[6], microDNAs[4], apoptotic circular DNAs[6], small polydispersed circular DNAs[7] and telomeric circular DNAs or C-circles[8]; (2) T cell receptor excision circles (TRECs)[9]; and (3) large (>100 kb), oncogenic, copy number-amplified circular extrachromosomal DNAs[10,11] (referred to as ecDNA and visible as double minute chromosomes during metaphase[12]). Despite our increasing ability to characterize multiple features in single cells[13], an in-depth characterization of circular DNA content, structure and sequence in single cells remains elusive with current approaches.

In cancer, oncogene amplifications on ecDNA are of particular interest because they potently drive intercellular copy number heterogeneity through their unique ability to be replicated and unequally segregated during mitosis[14–19]. This heterogeneity enables tumors to

A full list of affiliations appears at the end of the paper. ✉e-mail: henssenlab@gmail.com

Technical Report https://doi.org/10.1038/s41588-023-01386-y

adapt and evade therapies[2,20–22]. Indeed, patients with ecDNA-harboring cancers have adverse clinical outcomes[11]. Recent investigations indicate that enhancer-containing ecDNAs interact with each other in nuclear hubs[17,23] and can influence distant chromosomal locations in *trans*[23,24]. This suggests that even ecDNAs not harboring oncogenes may be functional[23,24]. Furthermore, we recently revealed that tumors harbor an unanticipated repertoire of smaller, copy number-neutral circular DNAs of yet unknown functional relevance[3].

In this study, we report single-cell extrachromosomal circular DNA and transcriptome sequencing (scEC&T-seq), a method that enables parallel sequencing of all circular DNA types, independent of their size, content and copy number, and full-length mRNA in single cells. We demonstrate its utility for profiling single cancer cells containing both structurally complex multifragmented ecDNAs and small circular DNAs.

Results

scEC&T-seq detects circular DNA and mRNA in single cells

Current state-of-the-art circular DNA purification approaches involve three sequential steps, that is, isolation of DNA followed by removal of linear DNA through exonuclease digestion and enrichment of circular DNA by rolling circle amplification[3,6,25]. We reasoned that this approach may be scaled down to single cells and when combined with Smart-seq2 (ref. 26) may allow the parallel sequencing of circular DNA and mRNA. To benchmark our method in single cells, we used neuroblastoma cancer cell lines, which we had previously characterized in bulk populations[3]. We used FACS to separate cells into 96-well plates (Fig. 1a, Supplementary Fig. 1a,b and Supplementary Table 1). DNA was separated from polyadenylated RNA, which was captured on magnetic beads coupled to single-stranded sequences of deoxythymidine (Oligo dT) primers, similarly to previous approaches[27]. DNA was subjected to exonuclease digestion, as successfully performed in bulk cell populations in the past, to enrich for circular DNA[3,6,25] (Fig. 1b). DNA subjected to PmeI endonuclease before exonuclease digestion served as a negative control[3]. In a subset of cases, DNA was left undigested as an additional control (Fig. 1b). The DNA remaining after the different digestion regimens was amplified. The amplified DNA was subjected to Illumina paired-end sequencing and in some cases to long-read Nanopore sequencing (Fig. 1a). The sequence composition of circular DNAs was analyzed and genomic origin was inferred in circularized regions using previously established computational algorithms for circular DNA analysis[3].

To evaluate the performance of our scEC&T-seq method, we first assessed mitochondrial DNA (mtDNA) detection and enrichment because mtDNA is present in all cells, is digested by PmeI and, due to its circularity and extrachromosomal nature, serves as a positive control. A significantly higher percentage of reads mapping to mtDNA was detected after longer exposure of the DNA of single cells to exonuclease ($P < 2.2 \times 10^{-16}$, two-sided Welch's *t*-test; Fig. 1c,d and Supplementary Fig. 1c,d). This was also the case for all other circular DNA elements ($P < 2.2 \times 10^{-16}$, two-sided Welch's *t*-test; Fig. 1e), indicating significant enrichment of circular DNA. Significant enrichment of ecDNA regions, that is, large (>100 kb) circular DNAs containing oncogenes, was observed after 1-day exonuclease digestion ($P = 2.10 \times 10^{-5}$, two-sided Welch's *t*-test; Supplementary Fig. 1e). This enrichment was not as pronounced as that of smaller circular DNAs after prolonged 5-day exonuclease digestion, suggesting that ecDNA may be less stable in the presence of exonuclease compared to smaller circular DNAs, or that small circular DNAs are more efficiently amplified by φ29 polymerase (Supplementary Fig. 1e,f). PmeI endonuclease incubation before 5-day exonuclease digestion significantly reduced reads mapping to mtDNA by 404.8 fold ($P < 2.2 \times 10^{-16}$, two-sided Welch's *t*-test; Fig. 1c,d and Supplementary Fig. 1c). Similar depletion was observed for reads mapping to circular DNAs containing PmeI recognition sites, confirming specific enrichment of circular DNA through our scEC&T-seq protocol ($P < 2.2 \times 10^{-16}$, two-sided Welch's *t*-test; Fig. 1f and Supplementary Fig. 1g,h). Significant concordance between Illumina- and Nanopore-based detection of circular DNAs suggested reproducible detection independent of sequencing technology (two-sided Pearson correlation, $R = 0.95$, $P < 2.2 \times 10^{-16}$; Supplementary Fig. 2a–d). Thus, scEC&T-seq enables the isolation and sequencing of circular DNAs from single cells.

The separated mRNA from the same cells was processed using Smart-seq2 (ref. 26,27) (Fig. 1a and Supplementary Note 1). We detected on average 9,058 ± 1,163 (mean ± s.d.) full mRNA transcripts from different genes per cell (Supplementary Fig. 3a–c and Supplementary Table 2). Unsupervised clustering separated both cell line populations (Supplementary Fig. 3d,e). To test whether scEC&T-seq provided high-quality mRNA sequencing data, we assessed cell cycle signature gene expression and classified single cells into three cell cycle phases (G1, S, G2/M; Supplementary Fig 3f). The cell cycle distributions inferred from scEC&T-seq matched those measured using FACS-based cell cycle analysis, confirming its accuracy (Supplementary Fig. 3g). Thus, scEC&T-seq not only enables the enrichment and detection of circular DNAs, but also allows parallel measurement of high-quality, full transcript mRNA in single cancer cells.

scEC&T-seq detects recurrent ecDNAs in single cells

Only circular DNAs conferring a fitness advantage are expected to be clonally present in a cancer cell population[22]. We recently found that tumors on average harbor more than 1,000 individual circular DNAs, most of which are small (<100 kb), lack oncogenes and do not contribute to oncogene amplification[3]. Their intercellular differences, however, remain unexplored and it is still unclear whether small circular DNAs can confer a fitness advantage and are clonally propagated in cancer cells[10]. Consistent with our previous reports in bulk populations[3], the average number of individual circular DNA regions identified using scEC&T-seq varied between 97 and 1,939 (median = 702) per single cell in neuroblastoma cell lines (Fig. 2a). The circular DNA size distribution and genomic origin was similar between single cells and mirrored the distribution observed in bulk sequencing[3] (minimum = 30 bp, maximum = 1.2 Mb, median = 21,483 kb; Fig. 2a and Supplementary Fig. 4a,b). All analyzed cells were alive at the time of sorting (Supplementary Fig. 1a,b) and most (>95%) circular DNAs detected in single cells were larger than apoptotic circular DNAs, suggesting that most circular DNAs were not a result of apoptosis, as suggested by other reports[6] (Fig. 2a and Supplementary Fig. 4a). Thus, each cancer cell contains a wide spectrum of individual circular DNAs from different genomic contexts.

As expected, most small circular DNAs did not harbor oncogenes[10]. The overall proportion of small circular DNAs detected recurrently in cells was low (Fig. 2b–d and Supplementary Fig. 4c). This indicates that only a small subset of small circular DNAs is clonally propagated in cancer cells. In line with their known oncogenic role in cancer and the positive selective advantages, amplified, oncogene-containing ecDNAs were recurrently detected in cells (Fig. 2b–d), which was validated by FISH (Fig. 2b and Supplementary Fig. 5a–c). Even though the functional relevance of small circular DNAs cannot be excluded, the observed high subclonality suggests that they do not contribute to cancer cell fitness to the same extent as clonal oncogene-amplifying ecDNA.

Complex multifragmented ecDNAs are detectable in single cells

We and others recently showed that ecDNAs are complex structures, sometimes containing rearranged fragments from different chromosomes[23,28–30]. Considering that scEC&T-seq was able to recurrently detect megabase-sized ecDNAs harboring the oncogenes *MYCN*, *CDK4* or *MDM2* (Fig. 2b), we asked whether scEC&T-seq could provide insights into ecDNA structures. Indeed, scEC&T-seq captured multifragment ecDNAs in almost all single cells recapitulating the

Technical Report https://doi.org/10.1038/s41588-023-01386-y

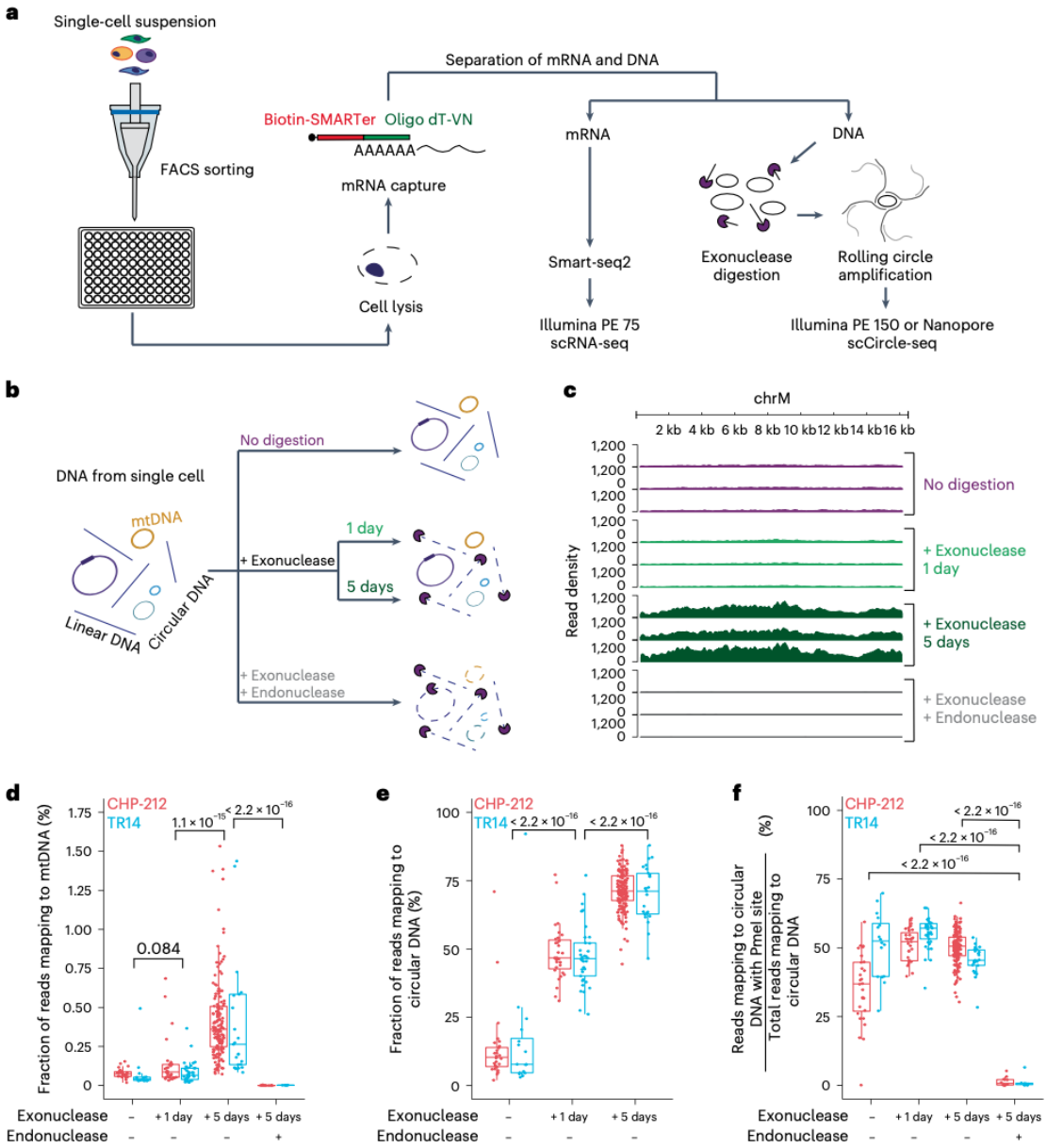


Fig. 1 | scEC&T-seq enables enrichment and detection of circular DNA in single cells. a, Schematic of the scEC&T-seq method. **b**, Schematic representation of the experimental conditions and expected outcomes. **c**, Genome tracks comparing read densities on mtDNA (chrM) in three exemplary CHP-212 cells for each experimental condition tested. Top to bottom, No digestion (purple), 1-day exonuclease digestion (light green), 5-day exonuclease digestion (dark green) and endonuclease digestion with PmeI before 5-day exonuclease digestion (gray). **d**, Fraction of sequencing reads mapping to mtDNA in each experimental condition in CHP-212 (red) and TR14 (blue) cells. **e**, Fraction of sequencing reads mapping to circular DNA regions identified by scEC&T-seq in each experimental condition in CHP-212 and TR14 cells. **f**, Fraction of sequencing reads mapping to circular DNA regions with the endonuclease PmeI targeting the sequence identified by scEC&T-seq in each experimental condition in CHP-212 and TR14 cells. **d**–**f**, Sample size is identical across conditions: no digestion ($n = 16$ TR14 cells, $n = 28$ CHP-212 cells); 1-day exonuclease digestion ($n = 37$ TR14 cells, $n = 31$ CHP-212 cells); 5-day exonuclease digestion ($n = 25$ TR14 cells, $n = 150$ CHP-212 cells); and endonuclease digestion with PmeI before 5-day exonuclease digestion ($n = 6$ TR14 cells, $n = 12$ CHP-212 cells). All statistical analyses correspond to a two-sided Welch's t-test. P values are shown. In all boxplots, the boxes represent the 25th and 75th percentiles with the center bar as the median value and the whiskers representing the furthest outlier ≤1.5× the interquartile range (IQR) from the box.

previously described element structures found in bulk populations[23,28] (Fig. 3a,b). At least one variant-supporting read per ecDNA breakpoint was detectable in approximately 30% of single cells (Supplementary Table 3). Further quantification of ecDNA junction-spanning reads and computational structural variant (SV) detection both from short- and long-read sequencing confirmed the interconnectedness of segments (Supplementary Fig. 6a–p and Supplementary Tables 4 and 5). Such SVs can lead to fusion transcript expression on ecDNA[3]. Indeed, fusion transcripts could be identified in single cells using scEC&T-seq (Fig. 3c and Supplementary Fig. 7). Thus, scEC&T-seq is sufficiently sensitive to detect ecDNA-associated SVs and resulting fusion gene expression in single cells.

https://doi.org/10.1038/s41588-023-01386-y

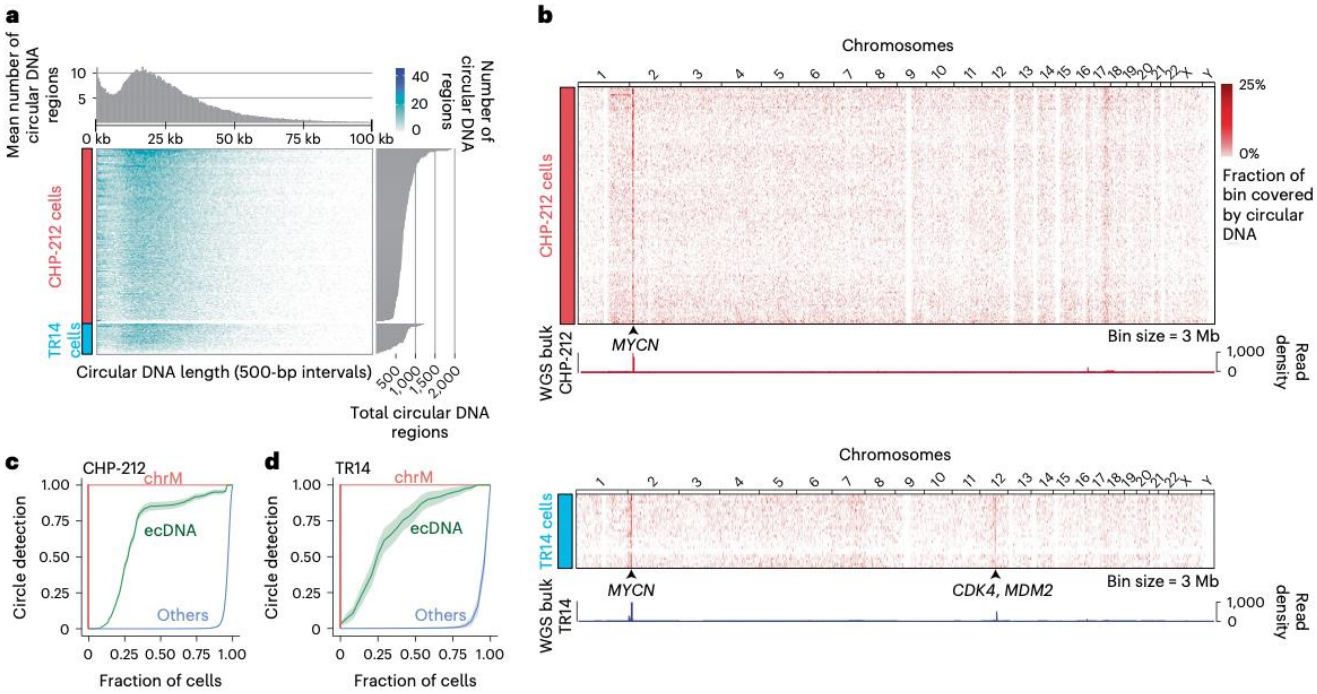


Fig. 2 | Oncogene-containing ecDNAs are recurrently identified in neuroblastoma single cells. a, Heatmap displaying the number and length of individual circular DNA regions (<100 kb) identified by scEC&T-seq in CHP-212 and TR14 neuroblastoma single cells (n = 150 CHP-212 cells, n = 25 TR14 cells; bin size = 500 bp) with density distribution for circular DNA sizes (top) and overall circular DNA counts (right). **b**, Heatmap of genome-wide circular DNA density in CHP-212 and TR14 neuroblastoma single cells (top: n = 150 CHP-212 cells, bin size = 3 Mb; bottom: n = 25 TR14 cells, bin size = 3 Mb), and genome tracks displaying genome-wide read density from WGS in bulk cell populations. The location of the *MYCN* gene in chromosome 2 is shown. **c**, **d**, Recurrence analysis in CHP-212 (n = 150) (**c**) and TR14 (n = 25) (**d**) cells displayed as the fraction of cells containing a detected circular DNA from each circular DNA type. ecDNA was defined as circular DNAs overlapping with copy number-amplified regions identified in bulk sequencing (green) and mtDNA or chrM (red). 'Others' are defined as all other small circular DNAs (blue). Data are presented as the mean ± s.e.m.

Intercellular differences in ecDNA content drive expression differences

The unequal mitotic segregation of ecDNA implies that ecDNA copy number can vary greatly between single cells[17,22]. In most single cells, multifragment ecDNAs did not differ in structure and composition (Fig. 3a,b), suggesting that ecDNA is structurally stable in cultured cell lines. As predicted by their binomial mitotic segregation and the conferred strong fitness advantage[2,17], most single TR14 cells contained all three independent oncogene-harboring ecDNAs also detected in bulk populations (Fig. 3b and Fig. 4a). However, a small number of cells only contained a subset of independent ecDNAs (Fig. 4a–c). This suggests that ecDNA content variation serves as a source of population heterogeneity. Intriguingly, *MDM2*-harboring ecDNAs were detected in all single cells, whereas *CDK4*- and *MYCN*-harboring ecDNAs were absent in some cells (Fig. 4b,c), suggesting that yet undefined biological principles of ecDNA segregation may exist. Next, we asked whether ecDNA copy number heterogeneity influenced the expression of genes encoded on ecDNA. We confirmed that the distribution of relative ecDNA copy number was consistent with copy number distributions measured using FISH (Supplementary Fig. 8a–h). Phasing of SNPs suggested that ecDNAs are of mono-allelic origin in each single cancer cell (Supplementary Fig. 9a,b), confirming previous observation in bulk cell populations[3]. Consistent with copy number-driven differences in gene expression, relative ecDNA copy number was positively correlated with the mRNA read counts of genes contained on ecDNAs in the same single cells (Fig. 4d–h). Even though enhancer interactions in clustered ecDNA may also contribute to intercellular ecDNA expression variability[23], we provide evidence that ecDNA copy number heterogeneity is a major determinant of intercellular differences in oncogene expression.

scEC&T-seq detects single-nucleotide variants on ecDNA and mtDNA

Single-nucleotide variants (SNVs) are important drivers of intercellular heterogeneity and tumor evolution[31]. Furthermore, SNVs can be tracked in cells, allowing their use for lineage tracing applications[32]. To test whether scEC&T-seq could be used to detect SNVs, we applied SNV detection algorithms on merged single-cell scEC&T-seq data and compared the detected SNVs to those identified in the whole-genome sequences of bulk populations. Most SNVs detected using scEC&T were also detected in whole genomes (>69.5%). Because scEC&T-seq also detects mtDNA (Fig. 2c,d), we hypothesized that heteroplasmic mitochondrial mutations may enable lineage tracing, as demonstrated in other single-cell assays in the past[32] (Fig. 1c,d and Supplementary Fig. 1c). Indeed, unsupervised hierarchical clustering by homoplasmic mtDNA variants accurately genotyped cells (Supplementary Fig. 10a). Heteroplasmic SNVs on mtDNA revealed high intercellular heterogeneity, and unsupervised hierarchical clustering on individual single cells grouped them, which indicates subclonality and may allow lineage tracing (Supplementary Fig. 10b and Supplementary Fig. 11a,b). Thus, scEC&T-seq can detect heteroplasmic variants in mtDNA and ecDNA, allowing for a wide range of SNV-based applications and analyses, including lineage inference.

Distinct pathways are active in cells with high small circular DNA content

Whereas the origin and functional consequences of large oncogene-containing ecDNA elements has been studied in some detail in the past[33,34], it is largely unclear how small circular DNAs are formed and how they influence the behavior of cells. Recent work suggests that some small circular elements are formed during apoptosis[6]. Other reports

https://doi.org/10.1038/s41588-023-01386-y

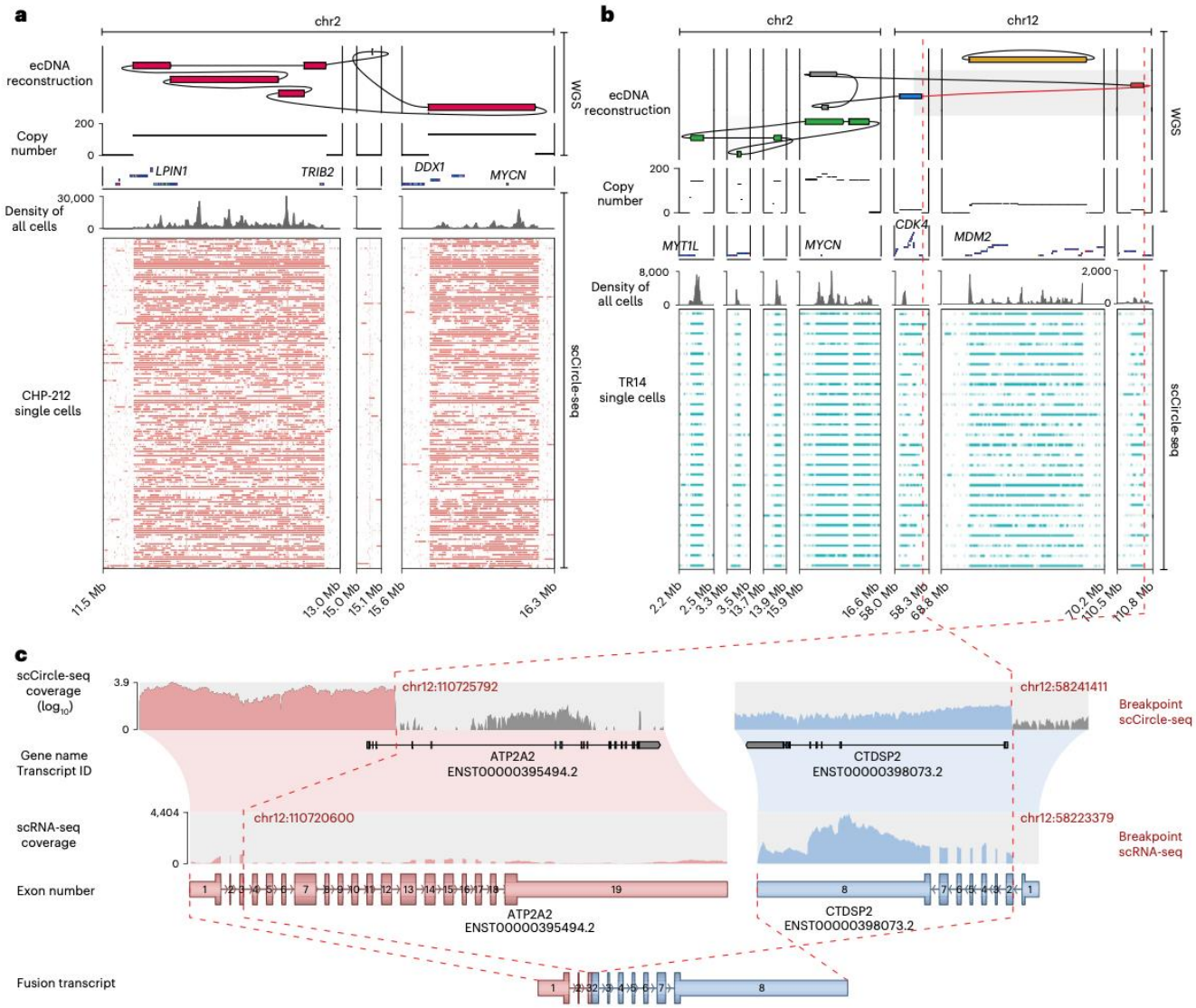


Fig. 3 | scEC&T-seq captures the complex structure of multifragmented ecDNAs in single neuroblastoma cells. a, **b**, Long- and short-read-based ecDNA reconstructions derived from WGS data in bulk cell populations and read coverage over the ecDNA fragments across single cells in CHP-212 ($n = 150$) (**a**) and TR14 ($n = 25$) cells (**b**) as detected by scEC&T-seq. Top to bottom, ecDNA amplicon reconstruction, copy number profile, gene annotations, read density over the ecDNA region in merged single cells and coverage over the ecDNA region in single cells (rows). **c**, Exemplary fusion transcript detected by scEC&T-seq resulting from the rearrangement of chromosomal segments in the *CDK4* ecDNA in TR14. Top to bottom, scCircle-seq read coverage over the breakpoint region in merged TR14 single cells (log-scaled), transcript annotations, scRNA-seq read coverage over the fused transcripts in merged TR14 single cells, native transcript representations and fusion transcript representation. The interconnected genomic segments in *CDK4* ecDNA that give rise to the fusion gene are indicated by a red dashed line.

provide evidence for the involvement of aberrant DNA damage repair in their generation[35]. In line with previous reports[36], we identified the presence of microhomology at circular breakpoints of small circular DNAs, suggesting that microhomology-mediated repair may be involved in their generation (Supplementary Fig. 12). The bimodal size distribution identified in single cells (Fig. 2a) suggested that at least two types of small circular DNAs exist in cells. Very small circular DNAs (<3 kb) were found in all analyzed single cells (Fig. 2a and Fig. 5a). No difference was observed in the fraction of very small circular DNAs between cells at different cell cycle phases (Fig. 5b), raising the question whether such small circular DNAs can be replicated. To identify the pathways associated with the high contents of these very small circular DNAs, we compared RNA expression of cells with a high relative amount of such small circular DNAs to that of cells with low relative content (Fig. 5a). Twenty pathways were significantly positively enriched in cell transcriptomes with high very small circular DNA content (Fig. 5c–e and Supplementary Table 6). In agreement with previous studies, DNA damage and repair pathways[35,37,38], apoptosis[6] and telomere maintenance[39] were significantly enriched in cells with a high relative content of this smaller subtype of circular DNA (Fig. 5c–e). This demonstrates that scEC&T-seq can help address long-standing questions about the origin and functional consequences of small circular DNAs.

Small circular DNA breakpoints frequently overlap with CCCTC-binding factor sites

Chromatin conformation and accessibility can influence DNA damage susceptibility[40]. We hypothesized that small circular DNAs

Technical Report https://doi.org/10.1038/s41588-023-01386-y

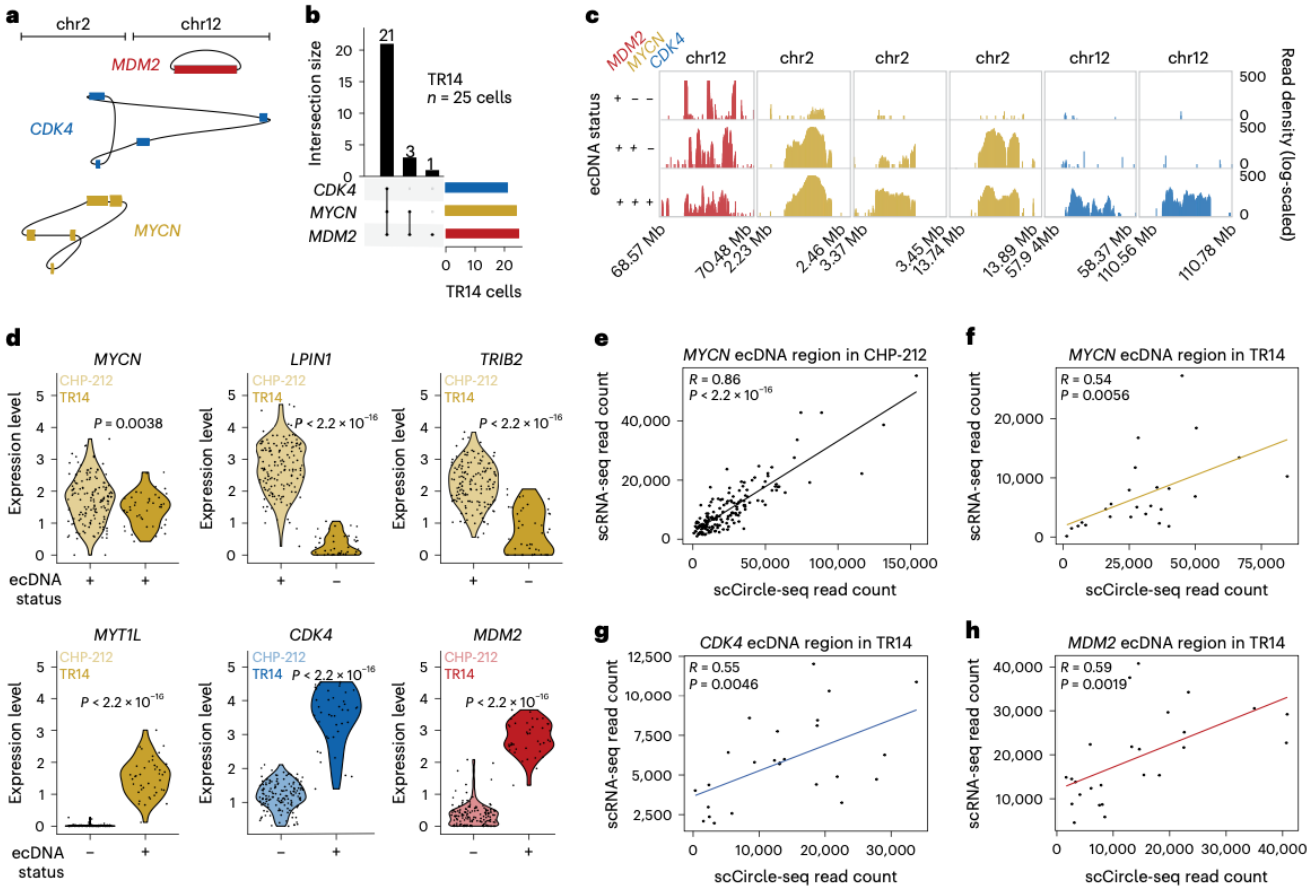


Fig. 4 | Intercellular differences in ecDNA content drive gene expression differences. a, Schematic representation of the three independent ecDNAs identified in TR14: *MYCN* ecDNA (yellow); *CDK4* ecDNA (blue); and *MDM2* ecDNA (red). **b**, UpSet plot displaying the co-occurrence of the three ecDNAs identified in TR14 (*MDM2*, *CDK4*, *MYCN*) in single cells ($n = 25$ TR14 cells). **c**, Genome tracks with read densities (log-scaled) over reconstructed ecDNA regions in three exemplary TR14 cells showing different ecDNAs detected. **d**, Violin plots of mRNA expression levels in TR14 and CHP-212 single cells (two-sided Welch's *t*-test; $P = 0.0038$ (*MYCN*), $P < 2.2 \times 10^{-16}$ (*LPIN1*, *TRIB2*, *CDK4*, *MDM2*, *MYT1L*)); $n = 171$ CHP-212 cells, $n = 42$ TR14 cells. **e**, **f**, Pairwise comparison between ecDNA and mRNA read counts from scEC&T-seq over the reconstructed *MYCN* ecDNA region in CHP-212 single cells (two-sided Pearson correlation, $P < 2.2 \times 10^{-16}$, $R = 0.86$, $n = 150$ cells) (**e**) and in TR14 single cells (two-sided Pearson correlation, $P = 0.0056$, $R = 0.54$, $n = 25$ cells) (**f**). **g**, **h**, Pairwise comparison between ecDNA and mRNA read counts from scEC&T-seq over the reconstructed *CDK4* (**g**) and *MDM2* (**h**) ecDNAs in TR14 single cells (two-sided Pearson correlation, $P = 0.0046$, $R = 0.55$ for *CDK4* and $P = 0.0019$, $R = 0.59$ for *MDM2*, $n = 25$ TR14 cells).

may be a product of DNA damage at sites of differential chromatin accessibility or conformation. To test this hypothesis, we measured the relative enrichment of CCCTC-binding factor (CTCF) chromatin immunoprecipitation followed by sequencing (ChIP–seq) and assay for transposase-accessible chromatin using sequencing (ATAC–seq) peaks in regions of small circular DNAs compared to other sites in the genome, respectively. Small circular DNAs detected using scEC&T-seq in single CHP-212 cells and those detected using Circle-seq in the bulk cell populations were used for this analysis (Supplementary Fig. 13a–d). Intriguingly, circular DNA breakpoints were significantly enriched at CTCF binding sites both in single cells and in bulk cell populations. This enrichment was even more striking considering that regions from which small circular DNAs originated were significantly depleted at sites of high ATAC–seq signals (Supplementary Fig. 13e). This suggests that CTCF binding sites and non-accessible chromatin, which is abundant at CTCF bindings sites[41], may be susceptible to breakage and circular DNA formation. To control for background ChIP–seq signals, we measured the enrichment of H3K4me1, H3K27ac and H3K27me3 ChIP–seq peaks at sites of small circular DNA formation. In all cases, small circular DNAs were found at significantly lower frequency at these sites than expected for randomly distributed regions (Supplementary Fig. 13f–h), confirming the specificity of CTCF enrichment and indicating that sites marked by H3K4me1, H3K27ac and H3K27me3 may be protected from breakage and circularization. Considering the role of CTCF in regulating the three-dimensional structure of chromatin through mediation of chromatin loop formation[41], our data raise the possibility that DNA breaks during CTCF-mediated loop extrusion may represent a mechanism of small circular DNA formation.

scEC&T-seq profiles circular DNA in primary neuroblastomas

We next applied scEC&T-seq to single nuclei from two neuroblastomas and live T cells isolated from the blood samples of two patients (Fig. 6a, Supplementary Figs. 14a,b and 15a–t and Supplementary Note 1). The number of individual circular DNA elements identified in cancer cells was significantly higher compared to that of normal T cells and cell line cells, suggesting that DNA circularization is more frequent in tumors than in untransformed cells or cells in culture (Fig. 6b). Circular DNA size distributions and relative genomic content were comparable to those observed in cell lines, suggesting that scEC&T-seq reproducibly captures circular DNA regardless of the input material (Fig. 6b and Supplementary Figs. 4a and 16a). In agreement with our observations in cell lines, the proportion of recurrently identified small circular DNAs was

 https://doi.org/10.1038/s41588-023-01386-y

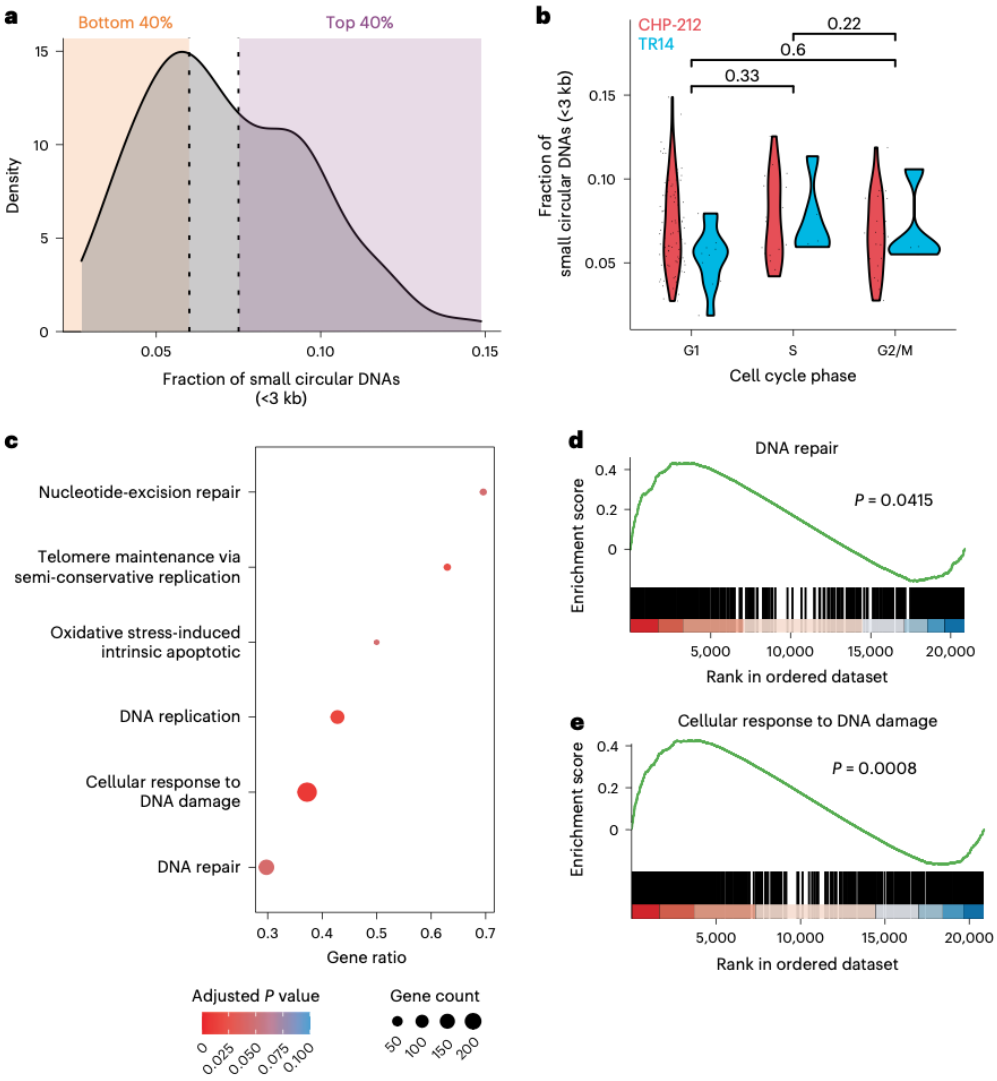


Fig. 5 | High relative content of small circular DNAs is associated with DNA damage response pathway activation. a, Density plot of relative small circular DNA (<3 kb) content in CHP-212 single cells ($n = 129$). For differential expression analyses, cells were divided in two categories: 'low' (orange area, bottom 40%) and 'high' (purple area, top 40%). **b**, Violin plot comparing the relative number of small circular DNAs (<3 kb) at different cell cycle phases in CHP-212 (red, $n = 129$) and TR14 (blue, $n = 20$) single cells. A two-sided Welch's *t*-test was used among the indicated conditions. *P* values are shown. **c**, Cellular processes significantly enriched in CHP-212 cells with high relative very small circular DNA content. Adjusted *P* values and gene counts are shown. **d**, Gene set enrichment analysis (GSEA) plot of genes involved in DNA repair (adjusted $P = 0.0415$). **e**, GSEA plot of genes involved in the cellular response to the DNA damage stimulus (adjusted $P = 0.0008$). *P* values were adjusted using the Bejamini–Hochberg method.

low (Supplementary Fig. 16b–d). Large, oncogene-containing ecDNAs, on the other hand, were recurrently identified in tumor nuclei but not in T cells (Fig. 6c and Supplementary Fig. 16b–d), in agreement with their oncogenic role. *MYCN*-containing ecDNAs were detectable in almost all cancer nuclei from both patients, which was confirmed with FISH (Supplementary Fig. 16e–g). As observed in cell lines, intercellular differences in *MYCN* transcription positively correlated with relative ecDNA content (Supplementary Fig. 16h,i). Thus, scEC&T-seq can be successfully applied to human tumors.

scEC&T-seq enables inference of ecDNA structural dynamics

Recent studies of cancer genomes have described structurally complex ecDNAs[3,11,18,19,28,29,42]; however, due to the analysis of bulk cell populations, they were limited in their ability to infer structural ecDNA heterogeneity. Both analyzed neuroblastomas contained large and structurally complex *MYCN*-containing ecDNAs, as confirmed using long-read Nanopore sequencing of the same single nuclei and by whole-genome sequencing (WGS) of bulk cell populations (Fig. 7a and Supplementary Fig. 17a). Whereas the ecDNA structure in patient no. 1 was so complex that it was not fully computationally reconstructed (Supplementary Fig. 17b), the *MYCN*-containing ecDNA in the other patient (patient no. 2) was structurally composed of five individual genomic fragments, all derived from chromosome 2, which were connected by four SVs (nos. 1–4) in a manner that was simple enough to be reliably reconstructed in single cells (Fig. 7a). We hypothesized that the assessment of intercellular ecDNA structural heterogeneity in this patient could facilitate the inference of ecDNA structural dynamics. Indeed, ecDNA considerably structurally differed between a subset of single cells (Fig. 7a,b). SV no. 1 was present in all single cells, suggesting it occurred before the other SVs and may represent the initial variant

Technical Report https://doi.org/10.1038/s41588-023-01386-y

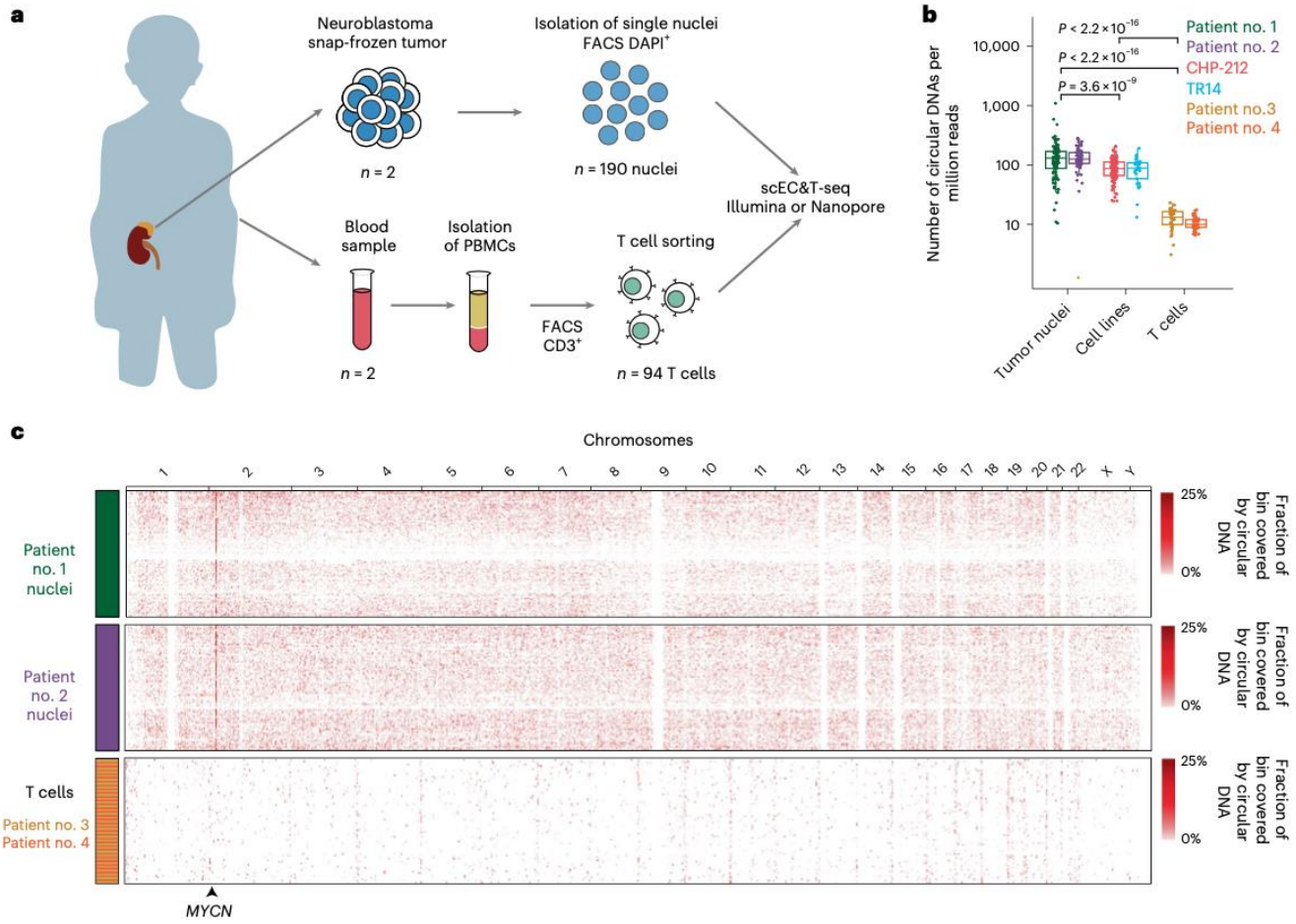


Fig. 6 | scEC&T-seq detects circular DNAs in primary neuroblastomas at the single-cell level. a, Schematic diagram describing tumor and blood sample processing. **b**, Number of individual circular DNA regions normalized by library size detected in primary tumor nuclei ($n = 93$ nuclei patient no. 1, $n = 86$ nuclei patient no. 2), neuroblastoma cell line single cells ($n = 25$ TR14 cells, $n = 150$ CHP-212 cells) and nonmalignant single T cells ($n = 38$ patient no. 3, $n = 41$ patient no. 4). *P* values were calculated using a two-sided Welch's *t*-test and are shown. The boxes in the boxplots represent the 25th and 75th percentiles with the center bar as the median value and the whiskers representing the furthest outlier ≤1.5× the IQR from the box. **c**, Heatmap of the genome-wide circular DNA density in neuroblastoma primary tumors and normal T cells ($n = 93$ patient no. 1, green; $n = 86$ patient no. 2, purple; $n = 38$ patient no. 3, yellow; $n = 41$ patient no. 4, orange; bin sizes = 3 Mb). The location of the *MYCN* gene in chr2 is shown.

leading to circularization (Fig. 7b–d). SVs nos. 2–4, on the other hand, were not detected in a subset of cells. Moreover, SV no. 2 and SV no. 3 indicated the presence of a 6-kb deletion and SV no. 4 supported the presence of a larger deletion (approximately 180 kb) on the ecDNA, both of which were present in most but not all single cells (94.2%; Fig. 7c,d). Analysis of split reads at the breakpoints of SV nos. 2 and 3, that is, the edges of the 6-kb deletion, and coverage across this deletion in single cells, suggested the presence of three different subclonal cell populations we termed subclone nos. 1–3. Clone no. 1 contained an intact ecDNA lacking deletions. Clone no. 2 harbored a mixed population of ecDNAs with and without deletions (Fig. 7b–e). In clone no. 3, the detected SVs and sequencing coverage indicated the presence of a pure population of ecDNAs harboring both deletions and all SVs (Fig. 7c–e). The simplest sequence of mutational events that would result in the observed intercellular structural ecDNA heterogeneity starts with a simple excision of an ecDNA containing *MYCN* and neighboring chromosomal regions, that is, SV no.1 generating ecDNA variant no. 1 found in clone no. 1 (Fig. 7e,f). This is followed by the fusion of two simple ecDNA no. 1 variants generating a more complex rearranged ecDNA variant no. 2 that includes the small 6-kb deletion and SV nos. 2 and 3 in addition to SV no. 1 (Fig. 7e,f). Such circular recombination is in agreement with recent models based on WGS[43]. An additional large deletion on this ecDNA would create ecDNA variant 3 with all SV nos. 1–4 and both deletions (Fig. 7e,f). The predominance of ecDNA variant 3 in these neuroblastoma cells suggests that it may confer a positive selective advantage. Our proof-of-principle demonstration that scEC&T-seq can help infer ecDNA structural dynamics illustrates that scEC&T-seq may facilitate future studies addressing important open questions about the origin and evolution of ecDNA.

Enhancers are coamplified with oncogenes on ecDNA in single cells

Regulatory elements are commonly amplified on ecDNA, have an essential role in the transcriptional regulation of oncogenes on ecDNA and are assumed to be under strong positive selection[28,29]. Indeed, at least one of the recently described *MYCN*-specific enhancer elements[28,29] was recurrently detected on ecDNAs harboring *MYCN* in over 82.7% of neuroblastoma single cells (Fig. 7f and Supplementary Fig. 18a). Interestingly, the deletion detected in patient no. 2, that is, ecDNA variant 3, is predicted to result in the loss of one of two *MYCN*

Technical Report https://doi.org/10.1038/s41588-023-01386-y

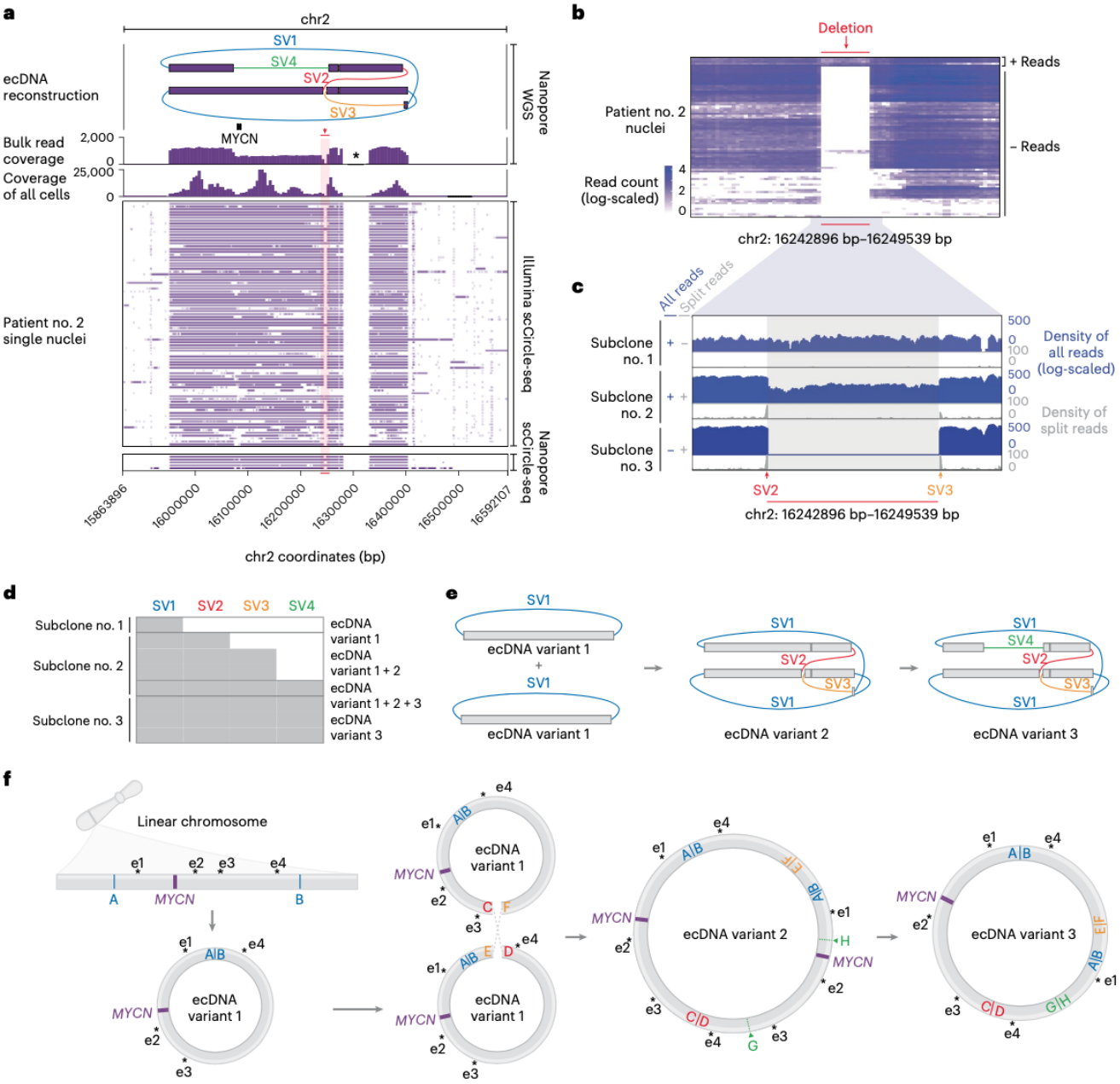


Fig. 7 | scEC&T-seq profiles intercellular structural ecDNA heterogeneity in neuroblastomas. a, Long read-based ecDNA reconstructions derived from WGS data in bulk populations and read coverage over the ecDNA fragments across single nuclei in patient no. 2 ($n = 86$ nuclei) as detected by long-read or short-read scEC&T-seq. Top to bottom, ecDNA amplicon reconstruction (the SVs on ecDNAs are colored; SV nos. 1–4), gene annotation, read density over the ecDNA region in bulk long-read Nanopore WGS data, read density over the ecDNA region in merged single nuclei and coverage over the ecDNA region in single nuclei (rows) as detected by long-read or short-read scEC&T-seq. The 6-kb deletion is highlighted in red. The single asterisk indicates the unmappable region of the reference genome (hg19). **b**, Heatmap of the total number of reads (log-scaled) in a 500-bp window over the identified 6-kb deletion on ecDNA across single nuclei in patient no. 2 ($n = 86$ nuclei). **c**, Exemplary genome tracks of the three identified clone variants in patient no. 2 based on the absence or presence of the 6-kb deletion on the ecDNA element. The log-scaled total read density is shown in blue and the circle edge-supporting read density is shown in gray. **d**, Detection of SV nos. 1–4 supporting the multifragmented ecDNA element in eight exemplary single cells representing the three identified clone variant groups (≥1 read supporting the SV, gray; 0 reads supporting the SV, white). **e**, Schematic representation of ecDNA variants 1–3 detected in **d**. **f**, Schematic interpretation of the evolution of the ecDNA structure in patient no. 2 based on the identified ecDNA variants in the scEC&T-seq data. The position of the *MYCN* oncogene and its local enhancer elements (e1–e5), indicated by the single asterisks, in each ecDNA variant is shown.

gene copies, including regulatory elements e2 and e3 present on ecDNA variant 2 (Fig. 7f). This raises the possibility that the change in enhancer:oncogene stoichiometry (6:1 in variant 3 versus 8:2 in variant 2), that is, the presence of one instead of two oncogene copies on an ecDNA, may be beneficial for oncogene expression because it may allow a more efficient use of enhancers on the ecDNA. Such mechanisms may explain the observed predominance of ecDNA variant no. 3 in the tumor cell population.

Recent reports suggest that ecDNAs not harboring oncogenes but containing enhancer elements exist and can enhance transcriptional output on linear chromosomes or on other ecDNAs in *trans* as part of ecDNA hubs[17,23]. To identify such ecDNA elements, we analyzed H3K4me1, H3K27ac, H3K27me3 ChIP–seq and ATAC–seq data from neuroblastoma cells and searched for ecDNAs including these regions but not harboring oncogenes. No ecDNA only harboring enhancer elements was recurrently identified in single neuroblastoma cells. All recurrently detected ecDNAs contained at least one oncogene. However, a large set of nonrecurrent small circular DNAs were identified that only contained genomic regions with regulatory elements (Supplementary Fig. 18b). The lack of recurrence of these circular DNA elements, however, suggests that they are not maintained in these cancer cells or do not confer positive selective advantages. Thus, scEC&T-seq allows the detection of noncoding circular DNAs and enables future investigations of their role in transcriptional regulation in cancer.

Discussion

We have shown that by parallel sequencing of circular DNA and mRNA from single cancer cells, scEC&T-seq not only readily distinguishes the transcriptional consequences of ecDNA-driven intercellular oncogene copy number heterogeneity, but also has the potential to uncover principles of ecDNA structural evolution. We believe that the integrated analysis of a cell's circular DNA content and transcriptome through scEC&T-seq will enable a more complete understanding of the extent, function, heterogeneity and evolution of circular DNAs in cancer and beyond.

scEC&T-seq complements recently published methods for single-cell DNA and single-cell RNA sequencing (scRNA-seq)[23,27], which cannot readily distinguish linear intra- from extrachromosomal circular amplicons. Even though scEC&T-seq is compatible with automation, the elaborate circular DNA enrichment procedures only allow low throughput, which drives costs per cell and currently represents a limitation of this method. However, compared to droplet-based microfluidic single-cell technologies, plate-based scEC&T-seq generates a uniform number of reads per cell and produces independent sequencing libraries available for selection and resequencing, which is advantageous when high sequencing coverage is needed. Indeed, we showed that scEC&T can be combined with different sequencing technologies. The level of detail provided by scEC&T-seq far exceeds that of high-throughput methods. Pairing our method with other single-cell technologies, for example, Strand-seq[44], and processing approaches, for example, single-cell tri-channel processing[45], may increase the spectrum of somatic variation detected by scEC&T-seq.

Performing scEC&T-seq in single cancer cells allowed us to profile their circular DNA content independently of copy number and circular DNA size. Small circular DNAs were identified in live single cells, suggesting that apoptosis is not the only mechanism of their generation. Whereas oncogene-containing ecDNAs were clonally present in single cells, small circular DNAs were exclusive to single cells. This not only indicates that small circular DNAs probably do not confer a selective advantage to cancer cells, but also suggests the existence of yet unknown prerequisites for selection, propagation and maintenance of these circular DNAs.

The robust demonstration of integrating circular DNA and mRNA sequencing in single cancer cells indicates that the same approach can be applied to a diverse range of biological systems to further explore the diversity and invariance of circular DNA in single cells. Thus, we anticipate that our method will be a resource for future research in many fields beyond cancer biology and suggest that it has the potential to address many currently unresolved biological questions regarding circular DNA.

Online content

Any methods, additional references, Nature Portfolio 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/s41588-023-01386-y.

References

1. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257–272 (2019).
2. Turner, K. M. et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).
3. Koche, R. P. et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat. Genet.* **52**, 29–34 (2020).
4. Shibata, Y. et al. Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* **336**, 82–86 (2012).
5. Møller, H. D. et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat. Commun.* **9**, 1069 (2018).
6. Wang, Y. et al. eccDNAs are apoptotic products with high innate immunostimulatory activity. *Nature* **599**, 308–314 (2021).
7. Cohen, S., Regev, A. & Lavi, S. Small polydispersed circular DNA (spcDNA) in human cells: association with genomic instability. *Oncogene* **14**, 977–985 (1997).
8. Henson, J. D. et al. DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. *Nat. Biotechnol.* **27**, 1181–1185 (2009).
9. Okazaki, K., Davis, D. D. & Sakano, H. T cell receptor β gene sequences in the circular DNA of thymocyte nuclei: direct evidence for intramolecular DNA deletion in V-D-J joining. *Cell* **49**, 477–485 (1987).
10. Verhaak, R. G. W., Bafna, V. & Mischel, P. S. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat. Rev. Cancer* **19**, 283–288 (2019).
11. Kim, H. et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 (2020).
12. Cox, D., Yuncken, C. & Spriggs, A. I. Minute chromatin bodies in malignant tumours of childhood. *Lancet* **1**, 55–58 (1965).
13. Lee, J., Hyeon, D. Y. & Hwang, D. Single-cell multiomics: technologies and data analysis methods. *Exp. Mol. Med.* **52**, 1428–1442 (2020).
14. Levan, A. & Levan, G. Have double minutes functioning centromeres? *Hereditas* **88**, 81–92 (1978).
15. Barker, P. E., Drwinga, H. L., Hittelman, W. N. & Maddox, A. M. Double minutes replicate once during S phase of the cell cycle. *Exp. Cell Res.* **130**, 353–360 (1980).
16. Mark, J. Double-minutes—a chromosomal aberration in Rous sarcomas in mice. *Hereditas* **57**, 1–22 (1967).
17. Yi, E. et al. Live-cell imaging shows uneven segregation of extrachromosomal DNA elements and transcriptionally active extrachromosomal DNA hubs in cancer. *Cancer Discov.* **12**, 468–483 (2022).
18. Yi, E., Chamorro González, R., Henssen, A. G. & Verhaak, R. G. W. Extrachromosomal DNA amplifications in cancer. *Nat. Rev. Genet.* **23**, 760–771 (2022).
19. van Leen, E., Brückner, L. & Henssen, A. G. The genomic and spatial mobility of extrachromosomal DNA and its implications for cancer therapy. *Nat. Genet.* **54**, 107–114 (2022).
20. 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).
21. Nathanson, D. A. et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science* **343**, 72–76 (2014).
22. Lange, J. T. et al. The evolutionary dynamics of extrachromosomal DNA in human cancers. *Nat. Genet.* **54**, 1527–1533 (2022).

23. Hung, K. L. et al. ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature* **600**, 731–736 (2021).
24. Zhu, Y. et al. Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. *Cancer Cell* **39**, 694–707 (2021).
25. 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).
26. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
27. Macaulay, I. C. et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12**, 519–522 (2015).
28. Helmsauer, K. et al. Enhancer hijacking determines extrachromosomal circular *MYCN* amplicon architecture in neuroblastoma. *Nat. Commun.* **11**, 5823 (2020).
29. Morton, A. R. et al. Functional enhancers shape extrachromosomal oncogene amplifications. *Cell* **179**, 1330–1341 (2019).
30. Deshpande, V. et al. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat. Commun.* **10**, 392 (2019).
31. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
32. Ludwig, L. S. et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339 (2019).
33. Wahl, G. M. The importance of circular DNA in mammalian gene amplification. *Cancer Res.* **49**, 1333–1340 (1989).
34. Shoshani, O. et al. Chromothripsis drives the evolution of gene amplification in cancer. *Nature* **591**, 137–141 (2021).
35. Dillon, L. W. et al. Production of extrachromosomal microDNAs is linked to mismatch repair pathways and transcriptional activity. *Cell Rep.* **11**, 1749–1759 (2015).
36. Tatman, P. D. & Black, J. C. Extrachromosomal circular DNA from TCGA tumors is generated from common genomic loci, is characterized by self-homology and DNA motifs near circle breakpoints. *Cancers* **14**, 2310 (2022).
37. Paulsen, T. et al. MicroDNA levels are dependent on MMEJ, repressed by c-NHEJ pathway, and stimulated by DNA damage. *Nucleic Acids Res.* **49**, 11787–11799 (2021).
38. Sunnerhagen, P., Sjöberg, R. M., Karlsson, A. L., Lundh, L. & Bjursell, G. Molecular cloning and characterization of small polydisperse circular DNA from mouse 3T6 cells. *Nucleic Acids Res.* **14**, 7823–7838 (1986).
39. Huang, C., Jia, P., Chastain, M., Shiva, O. & Chai, W. The human CTC1/STN1/TEN1 complex regulates telomere maintenance in ALT cancer cells. *Exp. Cell Res.* **355**, 95–104 (2017).
40. Downey, M. & Durocher, D. Chromatin and DNA repair: the benefits of relaxation. *Nat. Cell Biol.* **8**, 9–10 (2006).
41. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
42. Wu, S. et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699–703 (2019).
43. Rosswog, C. et al. Chromothripsis followed by circular recombination drives oncogene amplification in human cancer. *Nat. Genet.* **53**, 1673–1685 (2021).
44. Sanders, A. D., Falconer, E., Hills, M., Spierings, D. C. J. & Lansdorp, P. M. Single-cell template strand sequencing by Strand-seq enables the characterization of individual homologs. *Nat. Protoc.* **12**, 1151–1176 (2017).
45. Sanders, A. D. et al. Single-cell analysis of structural variations and complex rearrangements with tri-channel processing. *Nat. Biotechnol.* **38**, 343–354 (2020).

[1]Department of Pediatric Oncology and Hematology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany. [2]Experimental and Clinical Research Center of the MDC and Charité Berlin, Berlin, Germany. [3]Genomics Technology Platform, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. [4]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. [5]Charité–Universitätsmedizin Berlin, Berlin, Germany. [6]Faculty of Life Science, Humboldt-Universität zu Berlin, Berlin, Germany. [7]Freie Universität Berlin, Berlin, Germany. [8]Fraunhofer Institute for Cell Therapy and Immunology, Branch Bioanalytics and Bioprocesses IZI-BB, Potsdam, Germany. [9]Max-Delbrück-Centrum für Molekulare Medizin, Berlin, Germany. [10]RG Development and Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany. [11]Institute for Medical Genetics, Charité–Universitätsmedizin Berlin, Berlin, Germany. [12]Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA. [13]Berlin-Brandenburg Center for Regenerative Therapies, Charité–Universitätsmedizin Berlin, Berlin, Germany. [14]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. [15]German Cancer Consortium, partner site Berlin, and German Cancer Research Center, Heidelberg, Germany. [16]Berlin Institute of Health, Berlin, Germany. [17]Institute for Computational Cancer Biology, Center for Integrated Oncology, Cancer Research Center Cologne Essen Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [18]Berlin Institute for the Foundations of Learning and Data, Berlin, Germany. [19]Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [20]These authors jointly supervised this work: Richard P. Koche and Anton G. Henssen. ✉e-mail: henssenlab@gmail.com

Technical Report https://doi.org/10.1038/s41588-023-01386-y

Methods

scEC&T sequencing

A detailed, step-by-step protocol of scEC&T-seq is available on the Nature Protocol Exchange[46] and is described below. The duration of the protocol is approximately 8 days per 96-well plate.

Cell culture

Human tumor cell lines were obtained from ATCC (CHP-212) or were provided by J. J. Molenaar (TR14; Princess Máxima Center for Pediatric Oncology). The identity of all cell lines was verified by short tandem repeat genotyping (Genetica DNA Laboratories and IDEXX BioResearch); absence of *Mycoplasma* spp. contamination was determined with a Lonza MycoAlert Detection System. Cell lines were cultured in Roswell Park Memorial Institute 1640 medium (Thermo Fisher Scientific) supplemented with 1% penicillin, streptomycin and 10% FCS. To assess the number of viable cells, cells were trypsinized (Gibco), 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 (Thermo Fisher Scientific) and counted with a TC20 cell counter (Bio-Rad Laboratories).

Preparation of metaphase spreads

Cells were grown to 80% confluency in a 15-cm dish and metaphase-arrested by adding KaryoMAX Colcemid (10 µl ml^{-1}, Gibco) for 1–2 h. Cells were washed with PBS, trypsinized (Gibco) and centrifuged at 200*g* for 10 min. We added 10 ml of 0.075 M KCl preheated at 37 °C, 1 ml at a time, vortexing at maximum speed in between. Afterwards, cells were incubated for 20 min at 37 °C. Then, 5 ml of ice-cold 3:1 MeOH:acetic acid (kept at −20 °C) were added, 1 ml at a time followed by resuspension of the cells by flicking the tube. The sample was centrifuged at 200*g* for 5 min. Addition of the fixative followed by centrifugation was repeated four times. Two drops of cells within 200 µl of MeOH:acetic acid were dropped onto prewarmed slides from a height of 15 cm. Slides were incubated overnight.

FISH

Slides were fixed in MeOH:acetic acid for 10 min at −20 °C followed by a wash of the slide in PBS for 5 min at room temperature. Slides were incubated in pepsin solution (0.001 N HCl) with the addition of 10 µl pepsin (1 g 50 ml^{-1}) at 37 °C for 10 min. Slides were washed in 0.5× saline-sodium citrate (SSC) buffer for 5 min and dehydrated by washing in 70%, 90% and 100% cold ethanol (stored at −20 °C) for 3 min. Dried slides were stained with 10 µl Vysis LSI N-MYC SpectrumGreen/CEP 2 SpectrumOrange Probes (Abbott), ZytoLight SPEC CDK4/CEN 12 Dual Color Probe (ZytoVision) or ZytoLight SPEC MDM2/CEN 12 Dual Color Probe (ZytoVision), covered with a coverslip and sealed with rubber cement. Denaturing occurred in a ThermoBrite system (Abbott) for 5 min at 72 °C followed by 37 °C overnight incubation. The slides were washed for 5 min at room temperature in 2× SSC/0.1% IGEPAL, followed by 3 min at 60 °C in 0.4× SSC/0.3% IGEPAL (Sigma-Aldrich) and an additional wash in 2× SSC/0.1% IGEPAL for 3 min at room temperature. Dried slides were stained with 12 µl Hoechst 33342 (10 µM, Thermo Fisher Scientific) for 10 min and washed with PBS for 5 min. After drying, a coverslip was mounted on the slide and sealed with nail polish. Images were taken using a Leica SP5 Confocal microscope (Leica Microsystems).

Interphase FISH

CHP-212 and TR14 cells for the interphase FISH were grown in 8-chamber slides (Nunc Lab-Tek, Thermo Scientific Scientific) to 80% confluence. Wells were fixed in MeOH:acetic acid for 20 min at −20 °C followed by a PBS wash for 5 min at room temperature. The wells were removed and the slides were digested in pepsin solution (0.001 N HCl) with the addition of 10 µl pepsin (1 g 50 ml^{-1}) at 37 °C for 10 min. After a wash in 0.5× SSC for 5 min, slides were dehydrated by washing in 70%, 90% and 100% cold ethanol stored at −20 °C (3 min in each solution). Dried slides were stained with either 5 µl of Vysis LSI N-MYC SpectrumGreen/CEP 2 SpectrumOrange Probes, ZytoLight SPEC CDK4/CEN 12 Dual Color Probe or ZytoLight SPEC MDM2/CEN 12 Dual Color Probe, covered with a coverslip and sealed with rubber cement. Denaturing occurred in a ThermoBrite system for 5 min at 72 °C followed by 37 °C overnight. Slides were washed for 5 min at room temperature within 2× SSC/0.1% IGEPAL, followed by 3 min at 60 °C in 0.4× SSC/0.3% IGEPAL and a further 3 min in 2× SSC/0.1% IGEPAL at room temperature. Dried slides were stained with 12 µl Hoechst 33342 (10 µM) for 10 min and washed with PBS for 5 min. After drying, a coverslip was mounted on the slide and sealed with nail polish. Images were taken with a Leica SP5 Confocal microscope. For ecDNA copy number estimation, we counted foci using FIJI v.2.1.0 with the function find maxima. Nuclear boundaries were marked as regions of interest. The threshold for signal detection within the regions of interest was determined manually and used for all images analyzed within one group.

Patient samples and clinical data access

This study includes tumor and blood samples of patients diagnosed with neuroblastoma between 1991 and 2022. 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 version) 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 at the 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* copy number was determined using FISH. Tumor samples presented at least 60% tumor cell content as evaluated by a pathologist.

Isolation of nuclei

Tissue samples were homogenized using a precooled glass dounce tissue homogenizer (catalog no. 357538, Wheaton) in 1 ml of ice-cold EZ PREP buffer (Sigma-Aldrich). Ten strokes with a loose pestle followed by five additional strokes with a tight pestle were used for tissue homogenization. To reduce the heat caused by friction, the douncer was always kept on ice during homogenization. The homogenate was filtered through a Falcon tube (Becton Dickinson) with a 35-µm cell strainer cap. The number of intact nuclei was estimated by staining and counting with 0.02% trypan blue (Thermo Fisher Scientific) mixed in a 1:1 ratio.

Isolation of peripheral blood mononuclear cells from blood samples

Peripheral blood mononuclear cells (PBMCs) were isolated using density gradient centrifugation with Ficoll-Plaque PLUS (Cytiva). Whole-blood samples were resuspended 1:1 in calcium-free PBS and slowly added to 12 ml of Ficoll-Plaque PLUS. The sample was centrifuged at 200*g* for 30 min without breaking. The upper layer of PBMCs was isolated and washed into 40 ml of PBS. PBMCs were collected by centrifugation at 500*g* for 5 min and resuspended in 10% dimethylsulfoxide in FCS. The PBMC suspensions were stored at −80 °C until use.

FACS

For single-cell sorting, 1–10 million neuroblastoma cells or PBMCs were stained with propidium iodide (PI) (Thermo Fisher Scientific) in 1× PBS; viable cells were selected based on forward and side scattering properties and PI staining. PBMC suspensions were additionally stained with a 1:400 dilution of anti-human CD3 (Ax700, BioLegend). Nuclei suspensions were stained with DAPI (final concentration 2 µM, Thermo

 https://doi.org/10.1038/s41588-023-01386-y

Fisher Scientific). Viable cells, $CD3^+$ PBMCS or $DAPI^+$ nuclei were sorted using a FACSAria Fusion Flow Cytometer (BD Biosciences) into 2.5 µl of RLT Plus buffer (QIAGEN) in low-binding 96-well plates (4titude) sealed with foil (4titude) and stored at −80 °C until processing.

Genomic DNA and mRNA separation from single cells

Physical separation of genomic DNA (gDNA) and mRNA was performed as described previously in the G&T-seq protocol by Macaulay et al.[27]. All samples were processed using a Biomek FXP Laboratory Automation Workstation (Beckman Coulter). Briefly, polyadenylated mRNA was captured using a modified Oligo dT primer (Supplementary Table 7) conjugated to streptavidin-coupled magnetic beads (Dynabeads MyOne Streptavidin C1, catalog no. 65001, Invitrogen). The conjugated beads were directly added (10 µl) to the cell lysate and incubated for 20 min at room temperature with mixing at 800 rpm (MixMate, Eppendorf). Using a magnet (Alpaqua), the captured mRNA was separated from the supernatant containing the gDNA. The supernatant containing gDNA was transferred to a new 96-well plate (4titude); the mRNA-captured beads were washed three times at room temperature in 200 µl of 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 0.05% Tween 20 and 0.2× RNase inhibitor (SUPERase•In, Thermo Fisher Scientific). For each washing step, the beads were mixed for 5 min at 2,000 rpm in a MixTape (Eppendorf). The supernatant was collected after each wash and pooled with the original supernatant using the same tips to minimize DNA loss.

Complementary DNA generation

The mRNA captured on the beads was eluted into 10 µl of a reverse-transcription master mix including 10 U $µl^{-1}$ SuperScript II Reverse Transcriptase (Thermo Fisher Scientific), 1 U $µl^{-1}$ RNase inhibitor, 1× Superscript II First-Strand Buffer (Thermo Fisher Scientific), 2.5 mM DTT (Thermo Fisher Scientific), 1 M betaine (Sigma-Aldrich), 6 mM $MgCl_2$ (Thermo Fisher Scientific), 1 µM template-switching oligo (Supplementary Table 7), deoxynucleoside triphosphate mix (1 mM each of dATP, dCTP, dGTP and dTTP) (Thermo Fisher Scientific) and nuclease-free water (Thermo Fisher Scientific) up to the final volume (10 µl). Reverse transcription was performed on a thermocycler for 60 min at 42 °C followed by 10 cycles of 2 min at 50 °C and 2 min at 42 °C and ending with one 10-min incubation at 60 °C. Amplification of complementary DNA (cDNA) by PCR was immediately performed after reverse transcription by adding 12 µl of PCR master mix including 1× KAPA HiFi HotStart ReadyMix with 0.1 µM ISPCR primer (10 mM; Supplementary Table 7) directly to the 10 µl of the reverse transcription reaction mixture. The reaction was performed on a thermocycler for seven cycles as follows: 98 °C for 3 min, then 18 cycles of 98 °C for 15 s, 67 °C for 20 s, 72 °C for 6 min and finally 72 °C for 5 min. The amplified cDNA was purified using a 1:0.9 volumetric ratio of Ampure Beads (Beckman Coulter) and eluted into 20 µl of elution buffer (Buffer EB, QIAGEN).

Circular DNA isolation, amplification and purification

The isolated DNA was purified using a 1:0.8 volumetric ratio of Ampure Beads. The sample was incubated with the beads for 20 min at room temperature with mixing at 800 rpm (MixMate). Circular DNA isolation was performed as described previously in bulk populations[3,25]. Briefly, the DNA was eluted from the beads directly into an exonuclease digestion master mix (20 units of Plasmid-Safe ATP-dependent DNase (Epicentre), 1 mM ATP (Epicentre), 1× Plasmid-Safe Reaction Buffer (Epicentre)) in a 96-well plate. In a subset of samples, 1 µl of the endonuclease MssI/PmeI (20 U µl, New England Biolabs) was added. The digestion of linear DNA was performed for 1 or 5 days at 37 °C with 10 U of Plasmid-Safe DNase and 4 µl of ATP (25 mM), which was added again every 24 h to continue the enzymatic digestion. After 1 or 5 days of enzymatic digestion, the exonuclease was heat-inactivated by incubating at 70 °C for 30 min. The exonuclease-resistant DNA was purified and amplified using the REPLIg Single-Cell Kit (QIAGEN) according to the manufacturer's instructions. For this purification step, 32 µl of polyethylene glycol buffer (18% (w/v) (Sigma-Aldrich), 25 M NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.05% Tween 20) were added, mixed and incubated for 20 min at room temperature. After incubation, the beads were washed twice with 80% ethanol and the exonuclease-resistant DNA was eluted directly into the reaction mixture multiple displacement amplification with a REPLIg Single-Cell Kit (QIAGEN). Amplified circular DNA was purified using a 1:0.8 volumetric ratio of Ampure Beads and eluted in 100 µl of elution buffer (Buffer EB, QIAGEN).

Library preparation and sequencing

A total of 20 ng amplified cDNA or circular DNA was used for library preparation using the NEBNext Ultra II FS (New England Biolabs) according to the manufacturer's protocol. Samples were barcoded using unique dual-index primer pairs (New England Biolabs) and libraries were pooled and sequenced on a HiSeq 4000 instrument (Illumina) or a NovaSeq 6000 instrument with 2× 150-bp paired-end reads for circular DNA libraries and 2× 75-bp paired-end reads for cDNA libraries.

Genomic and transcriptomic read alignments

Sequenced reads from the gDNA libraries were trimmed using TrimGalore (v.0.6.4)[47] and mapped to the human genome build 19 (GRCh37/hg19). Alignment was performed with the Burrows–Wheeler Aligner (BWA)-MEM (v.0.7.17)[48]. Following the recommendation of the Human Cell Atlas project[49] (v.2.2.1)[50] was used to align the RNA-seq data obtained from Smart-seq2 (ref. 26) against a transcriptome reference created from the hg19 and ENCODE annotation v.19 (ref. 51). Afterwards, genes and isoforms were quantified using rsem (v.1.3.1)[52] with a single cell prior.

Nanopore scCircle-seq

Before Nanopore sequencing, the amplified circular DNA from single cells was subjected to T7 endonuclease digestion to reduce DNA branching. Then, 1.5 µg of amplified circular DNA were incubated at 37 °C for 30 min with 1.5 µl T7 endonuclease I (10 U $µl^{-1}$, New England Biolabs) in 3 µl of NEBuffer 2 and nuclease-free water up to a final volume of 30 µl. The endonuclease-digested DNA was purified using a 1:0.7 volumetric ratio of Ampure Beads and eluted in 25 µl of nuclease-free water. Libraries were prepared using the ONT Rapid Barcoding Kit (catalog no. SQK-RBK004, Oxford Nanopore Technologies) according to the manufacturer's instructions, and sequenced on an R9.4.1 MinION flowcell (FLO-MIN106, Oxford Nanopore Technologies). A maximum of four samples were multiplexed per run.

Nanopore scCircle-seq data processing

The scCircle-seq Nanopore data were base-called and demultiplexed using Guppy (v.5.0.14; running guppy_basecaller with dna_r9.4.1_450bps_hac model and guppy_barcoder with FLO-MIN106 and default parameters). The obtained reads were quality-filtered using NanoFilt[53] (v.2.8.0) (-l100--headcrop 50--tailcrop 50) and aligned using ngmlr[54] (v.0.2.7) against the GRCh37/hg19 reference genome. To call SVs, we applied Sniffles[54] (v.1.0.12) (--min_homo_af 0.7--min_het_af 0.1--min_length 50--min_support 4); to obtain the binned coverage, we used deepTools[55] (v.3.5.1) bamCoverage. All these steps are available as a Snakemake pipeline (https://github.com/henssen-lab/nano-wgs).

Circle-seq in bulk populations

Circle-seq in bulk populations was performed as described previously[3]. A detailed step-by-step protocol can be found on the Nature Protocol Exchange server.

ChIP–seq

We generated H3K27me3 ChIP–seq data for CHP-212 according to a previously described protocol[28]. Briefly, 5–10 million CHP-212 cells

 https://doi.org/10.1038/s41588-023-01386-y

were fixed in 10% FCS-PBS with 1% paraformaldehyde for 10 min at room temperature. Chromatin was prepared as described previously[28] and sheared until a fragment size of 200–500 bp. H3K27me3–DNA complexes were immunoprecipitated for 15 h at 4 °C with an anti-H3K27me3 polyclonal antibody (catalog no. 07-449, Sigma-Aldrich). In total 10–15 µg of chromatin and 2.5 µg of antibody were used for immunoprecipitation. Libraries for sequencing were prepared using Illumina Nextera adapters according to the recommendations provided. Libraries were sequenced in 50-bp single-read mode in an Illumina HiSeq 4000 sequencer. FASTQ files were quality-controlled with FASTQC (v.0.11.8) and adapters were trimmed using BBMap (v.38.58). Reads were aligned to the hg19 using the BWA-MEM[48] (v.0.7.15) with default parameters. Duplicate reads were removed using Picard (v.2.20.4).

Chromatin marks enrichment analyses

We obtained public CHP-212 copy number variation, ChIP–seq (H3K27ac, H3K4me1, CTCF) and ATAC–seq data[28,56]. For further analysis, we used the processed bigwig tracks, filtered to exclude ENCODE Data Analysis Center (DAC) blacklisted regions and normalized to read counts per million (CPM) in 10-bp bins, and peak calls provided by Helmsauer et al.[28]. To assess the correlation of epigenetic marks with circle regions, we only considered circle regions that did not overlap with copy number variation in CHP-212 or ENCODE DAC blacklisted regions. For H3K27ac, H3K4me1 and H3K27me3 ChIP–seq and ATAC–seq data, we computed the mean CPM signal across all circle regions, weighted by the respective circle sizes. To test for statistical association, we created 1,000 datasets with randomized circle positions within a genome masked for copy number variation in CHP-212 and ENCODE DAC blacklisted regions using regioneR[57] (v.1.24.0). We derived an empirical *P* value from the distribution of mean CPM signal across the randomized circle regions. For CTCF ChIP–seq data, we calculated the percentage of circle edges overlapping with a CTCF peak and assessed statistical significance using the same randomization strategy as described above.

Circle-seq analysis

Extrachromosomal circular DNA analysis was performed as described previously[3]. Reads were 3′-trimmed for both quality and adapter sequences, with reads removed if the length was less than 20 nucleotides. BWA-MEM (v.0.7.15) with default parameters was used to align the reads to the human reference assembly GRCh37/hg19; PCR and optical duplicates were removed with Picard (v.2.16.0). Putative circles were classified with a two-step procedure. First, all split reads and read pairs containing an outward-facing read orientation were placed in a new BAM file. Second, regions enriched for signal over background with a false discovery rate < 0.001 were detected in the 'all reads' BAM file using variable-width windows from Homer v.4.11 findPeaks (http://homer.ucsd.edu/); the edges of these enriched regions were intersected with the circle-supporting reads. The threshold for circle detection was then determined empirically based on a positive control set of circular DNAs from bulk sequencing data. Only enriched regions intersected by at least two circle-supporting reads were classified as circular regions.

Quality-controlled filtering of scCircle-seq data

To evaluate adequate enrichment of circular DNA, we used coverage over mtDNA as the internal control. Cells with fewer than ten reads per base pair sequence-read depth over mtDNA or fewer than 85% genomic bases captured in mtDNA were omitted from further analyses. Cutoff values were chosen based on maximal read depth values detected in endonuclease controls (with PmeI; Supplementary Fig. 1c). For all downstream analyses, we only considered sequencing data from cells digested with exonuclease for 5 days. Because mtDNA is not present in nuclei, we filtered single-nucleus Circle-seq data only based on RNA quality control.

Recurrence analysis from scCircle-seq data

Read counts from putative circles were quantified using bedtools multicov (https://bedtools.readthedocs.io) from single-cell BAM files in 100-kb bins across all canonical chromosomes from genome assembly GRCh37/hg19. Counts were normalized to sequencing depth in each cell and each bin was marked positive if it contained circle read enrichment with $P < 0.05$ compared with the background read distribution. Bins were then classified into three groups based on genomic coordinates: (1) ecDNA if the region overlapped the amplicon assembled from the bulk sequencing data; (2) chrM; and (3) all other sites. Recurrence was then analyzed by plotting the fraction of cells containing a detected circle in each of the three categories.

Phasing of SNPs in scCircle-seq data

Reference phasing was used to assign each SNP to one of the two alleles based on bulk WGS data. Then, single cells were genotyped to compare if the same allele was gained in all of them. For this analysis, we used the known SNPs identified by the 1000 Genomes Project[58] and extracted coverage and nucleotide counts for each annotated position. In regions with allelic imbalance, like the high copy number gains at ecDNA loci, the B-allele frequency of a heterozygous SNP is significantly different from 0.5. Hence, we could assign each SNP in these regions to either the gained or non-gained allele. We then also genotyped all single cells at each known SNP location and visualized the resulting B-allele frequency values while keeping the allele assignment from the bulk WGS data.

Relative copy number estimation ($\log_2$ coverage)

The average coverage over all annotated genes was calculated and genes were split into amplicon and non-amplicon genes based on whether their genomic location overlapped with the identified ecDNA regions per cell. The coverage of all amplicon genes was normalized by the background coverage, that is, the winsorized mean coverage of all non-amplicon genes. A winsorized mean was chosen to account for the fact that the identification of ecDNA regions might have been incomplete; thus, the top and bottom 5% of values were removed from the background coverage. The resulting values were $\log_2$-transformed and used as a proxy for ecDNA copy number.

Identification of SVs in scCircle-seq data

The SV calling for scCircle-seq was done using lumpy-sv[55] (v.0.2.14) and SvABA(v.1.1.0). To our knowledge, no dedicated SV caller for single-cell DNA data is available. However, because of high copy numbers of ecDNA, bulk methods work.

Identification of SVs in WGS bulk data and merged scCircle-seq data

SAMtools[59] (v.1.11) was used to merge all alignment files of the same cell line into one pseudobulk alignment. To achieve a coverage closer to standard bulk sequencing, the resulting BAM file was subsequently downsampled to 10% of its original size using SAMtools. The identification of SVs in WGS and merged scCircle-seq data for the TR14 and CHP-212 cell lines was accomplished using lumpy-sv[60] (v.0.3.1) and SvABA[61] (v.1.1.0), both with standard parameters. The preprocessing of the BAM files, which included lower size (<20 bp) and lower quality reads (MAPQ < 5) filtering, as well as supporting read counts and VAF calculations, was performed using SAMtools[59] (v.1.10). All the analysis steps were completed using the GRCh37/hg19 reference genome. The identification and counts of reads supporting the SV breakpoints were performed considering split and abnormally mapped reads and filtering out duplicated reads and secondary alignments.

Identification of SNVs in bulk WGS data and merged scCircle-seq data

To ensure compatibility with standard mitochondrial variation reporting[62], each single-cell sequencing sample was realigned to GRCh37/

 https://doi.org/10.1038/s41588-023-01386-y

hg19 with a substituted revised Cambridge Reference Sequence mitochondrial reference (GenBank no. NC_012920) using BWA-MEM[63] (v.0.7.17). Duplicate reads were removed using Picard (v.2.23.8). GATK4/Mutect2[64] (v.4.1.9.0) with default parameters was used to call variants in whole-genome bulk and merged scCircle-seq sequencing data (pseudobulk). Only variants on canonical chromosomes (including chrM) and passing GATK4/FilterMutectCalls were retained and subsequently filtered for the regions previously reconstructed for the respective cell lines (Fig. 3a) using bcftools filter with flag-r.

Identification of SNVs in mtDNA

For mitochondrial SNV identification in single cells, we applied a custom pipeline consisting of GATK4/Mutect2 (ref. 64) (v.4.1.9.0) in mitochondria mode and Mutserve[65] (v.2.0.0-rc12), a variant caller optimized to detect heteroplasmic sites in mitochondrial sequencing data, with default parameters. First, variants were called by both callers for each single cell separately. Variants were then filtered in a two-step process: (1) variants were only retained if they have been called in at least two samples by the same caller; and (2) remaining variants were only kept if they were called by both callers. Variants labeled 'blacklist' by Mutserve were removed. To infer the allele frequency for each variant in the final set, each single cell was then subjected to genotyping using alleleCount (v.4.0.2) (https://github.com/cancerit/alleleCount). Only reads uniquely mapping to the mitochondrial reference and with a mapping quality ≥ 30 were kept. For each called alternate allele b at position x, the allele frequency (AF) was calculated as:

$$\mathrm{AF}_{x,b} = \frac{(\text{read count})_{x,b}}{\text{read depth}_x}$$

The resulting single-cell x variant AF matrix was further filtered manually and separately for each cell line. Single cells with fewer than three variants and variants with a maximum column allele frequency < 5%, mean AF (MAF) > 30% and MAF < 0.1% for CHP-212 as well as MAF > 30% and MAF < 0.1% for TR14 were considered uninformative for clustering and removed based on spot checking.

Heatmap visualization of the filtered single-cell x variant AF matrix was generated using the R package ComplexHeatmap[66] (v.2.6.2). Hierarchical clustering was then applied to the single cells using the R package hclust with the agglomeration method parameter 'complete'. Phylogenetic trees were rendered using the R package dendextend (v.1.15.2).

Microhomology detection

Microhomology analysis was performed using NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with the following parameters: blastn -task megablast -word_size = 4 -evalue = 1 -outfmt '6 qseqid length evalue' -subject_besthit -reward = 1 -penalty = -2. These parameters look for a minimum microhomology length of 4 bp, and the standard reward and penalty values for nucleotide match and mismatch. In addition, we only considered significant results with an Expect value < 1. To evaluate the presence of microhomology around the circular DNA junctions, we generated files that include 100 bp around the start and end of the circle (50 bp inside the circular DNA and 50 bp of linear DNA). To be able to perform this analysis, we filtered out all the circles with a length <100 bp. Then, we compared the sequences for each start and end pair (one circle junction), evaluating and retrieving microhomologous sequences around the circular junction. This analysis was repeated for each individual circle in the CHP-212 and TR14 cell lines.

Quality control filtering and clustering of scRNA-seq data

Cells and nuclei were loaded into Seurat[67] (v.4.10); features that were detected in at least three cells were included. Subsequently, cells with 5,000 or more features in cell lines and 2,000 features in T cells and nuclei were selected for further analysis. Cells or nuclei with high expression of mitochondrial genes (>15% in single cells and >2.5% in nuclei) were also excluded. Data were normalized with a scale factor of 10.000 and scaled using default ScaleData settings. To account for gene length and total read count in each cell, the Smart-seq2 data were normalized using transcripts per million; then, a pseudocount of one was added and natural-log transformation was applied. The first four principal components were significant; therefore, the first five principal components were used for FindNeighbors and RunUMAP to capture as much variation as possible as recommended by the Seurat authors. The resolution for FindClusters was set to 0.5.

Cell cycle analyses in scRNA-seq data

Cell cycle phase was assigned to single cells based on the expression of G2/M and S phase markers using the Seurat CellCycleScoring function.

Single-cell differential expression analysis

Very small circular DNAs were defined as circles shorter than 3 kb. To calculate the relative number of this subtype of small circular DNAs per cell, the number of <3 kb circular DNAs was divided by the total number of circles in a cell. The cells were ranked by their relative number and grouped by taking the top and bottom 40% of the ranked list, defined as 'high' and 'low', respectively. Logarithmic fold change of gene expression between the two groups was calculated using the FindMarkers function in the Seurat R package[67] (v.4.10) without logarithmic fold change threshold and a minimum detection rate per gene of 0.05. The R package clusterProfiler[68] (v.4.0.5) was used to perform unsupervised GSEA of gene ontology terms using gseGO and including gene sets with at least three genes and a maximum of 800 genes.

Correlation of scCircle-seq and scRNA-seq coverage

Coverage of ecDNA amplicon regions in the scCircle-seq and scRNA-seq BAM files was calculated with bamCoverage[55] using CPM normalization. Correlation between Circle-seq and RNA-seq coverage was analyzed by fitting a linear model.

Identification of fusion genes

The single-cell, paired-end, RNA-seq FASTQ files were merged (96 cells for TR14 and 192 cells for CHP-212). The obtained merged data were aligned with STAR[69] (v.2.7.9a) to the reference decoy GRCh37/hs37d5, using the GENCODE 19 gene annotation, allowing for chimeric alignment (--chimOutType WithinBAM SoftClip). To call and visualize fusion genes, Arriba[70] (v.2.1.0) was applied, with the custom parameters -F 150 -U 700. The final confident call set included only fusions with (1) total coverage across the breakpoint ≥ 50× and (2) ≥30% of the mapped reads being split or discordant reads. Only fusion genes in the proximity (±10 Mb) of the amplicon boundaries were considered for the downstream analysis.

ecDNA amplicon reconstruction

We used the amplicon reconstructions provided by Helmsauer et al.[28] for CHP-212 and Hung et al.[23] for TR14. Briefly, these reconstructions were obtained by organizing a filtered set of Illumina WGS (CHP-212) and Nanopore WGS (TR14) SV calls as genome graphs using gGnome[71] (v.0.1) (genomic intervals as nodes and reference or SVs as edges). Then, circular paths through these graphs were identified that included the amplified oncogenes and could account for the major copy number steps observed in the respective cell line. For the two patients added to the study, patient no. 1 and patient no. 2, shallow whole-genome Nanopore data were generated as described by Helmsauer et al.[28]. Basecalling, read filtering (NanoFilt −l 300), mapping and SV calling were performed as described previously in the Methods ('Nanopore scCircle-seq data processing'). For ecDNA reconstruction, a set of confident SV calls was compiled (variant AF > 0.2 and supporting reads ≥ 50×). As for CHP-212 and TR14, a genome graph was built using gGnome[61] (v.0.1) and manually curated. To check amplicon structure correctness for the patient samples, in silico-simulated Nanopore reads were sampled from the

reconstructed amplicon using an adapted version of PBSIM2 (ref. 72) (https://github.com/madagiurgiu25/pbsim2) and preprocessed as the original patient samples. Lastly, the SV profiles between original samples and in silico simulation were compared. All reconstructed amplicons were visualized using gTrack (v.0.1.0; https://github.com/mskilab/gTrack), including the GRCh37/hg19 reference genome and GENCODE 19 track.

ecDNAs co-occurrence analysis in TR14 single cells

We used the circle classification algorithm described previously to define circular DNA-enriched regions in single cells. For each single cell, we defined whether the circular DNA-enriched regions overlapped the ecDNA amplicon (*MYNC*, *CDK4*, *MDM2*) assembled from TR14 bulk sequencing data using the function findOverlaps from the R package GenomicRanges[73] (v.1.44.0). Presence or absence of overlap was defined for each of the three *MYNC*, *CDK4*, *MDM2* ecDNAs independently, excluding the amplicon regions shared by *MYCN* and *CDK4* ecDNAs.

Statistics and reproducibility

No statistical method was used to predetermine sample size. No data were excluded from the analyses. Experiments were not randomized and the investigators were not blinded to allocation during the experiments and outcome assessment. The FISH experiments were performed once per cell line and primary tumor.

Reporting summary

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

Data availability

The sequencing data generated in this study are available at the European Genome-phenome Archive under accession no. EGAS00001007026. The ChIP–seq narrowPeak and bigwig files were downloaded from https://data.cyverse.org/dav-anon/iplant/home/konstantin/helmsaueretal/. All other data are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

The data analysis code associated with this publication can be found at https://github.com/henssen-lab/scEC-T-seq.

References

46. González, R. C., Conrad, T., Kasack, K., & Henssen, A. G. scEC &T-seq: a method for parallel sequencing of extrachromosomal circular DNAs and transcriptomes in single human cells. https://doi.org/10.21203/rs.3.pex-2180/v1
47. Krueger, F., James, F., Ewels, P., Afyounian, E. & Schuster-Boeckler, B. TrimGalore v.0.6.7. Zenodo https://zenodo.org/record/5127899#.ZDUyQuzMIqs (2021).
48. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1303.3997 (2013).
49. Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041 (2017).
50. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
51. Davis, C. A. et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
52. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
53. 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).
54. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).
55. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
56. Boeva, V. et al. Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* **49**, 1408–1413 (2017).
57. Gel, B. et al. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289–291 (2016).
58. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
59. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
60. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
61. Wala, J. A. et al. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res.* **28**, 581–591 (2018).
62. Bandelt, H. J., Kloss-Brandstatter, A., Richards, M. B., Yao, Y. G. & Logan, I. The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies. *J. Hum. Genet* **59**, 66–77 (2014).
63. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
64. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
65. Weissensteiner, H. et al. mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Res.* **44**, W64–W69 (2016).
66. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
67. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
68. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
69. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
70. Uhrig, S. et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res.* **31**, 448–460 (2021).
71. Hadi, K. et al. Distinct classes of complex structural variation uncovered across thousands of cancer genome graphs. *Cell* **183**, 197–210 (2020).
72. Ono, Y., Asai, K. & Hamada, M. PBSIM2: a simulator for long-read sequencers with a novel generative model of quality scores. *Bioinformatics* **37**, 589–595 (2021).
73. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).

Acknowledgements

A.G.H. is supported by the Deutsche Forschungsgemeinschaft (DFG) (grant no. 398299703). This project has received funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme (grant no. 949172). A.G.H. is supported by the Deutsche Krebshilfe Mildred Scheel Professorship program no. 70114107. This project was supported by the Berlin Institute of Health (BIH). Computation was performed on the HPC for Research cluster of the BIH. The project that gave rise to these results

received the support of a fellowship from 'la Caixa' Foundation (no. 100010434). The fellowship code is LCF/BQ/EU20/11810051. E.R.-F. is supported by the Alexander von Humboldt Foundation. R.X. is supported by Deutsche Krebshilfe. R.P.K. is a BIH Visiting Professor funded by the Stiftung Charité. M.C.S. is funded by the DFG-funded Research Training Group 2424/CompCancer. R.F.S. is a Professor at the Cancer Research Center Cologne Essen funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia. This work was partially funded by the German Ministry for Education and Research as BIFOLD (Berlin Institute for the Foundations of Learning and Data) (refs. 01IS18025A and 01IS18037A). This work was delivered as part of the eDyNAmiC team supported by the Cancer Grand Challenges partnership funded by Cancer Research UK (H.Y.C. no. CGCATF-2021/100012, A.G.H. no. CGCATF-2021/100017) and the National Cancer Institute (H.Y.C. no. OT2CA278688, A.G.H. no. OT2CA278644).

Author contributions

R.C.G., T.C., R.P.K. and A.G.H. contributed to study design, and data collection and interpretation. R.C.G., T.C. and K.K. performed the single-cell experiments. R.P.K., R.C.G. and K.Haase performed the analysis of the scCircle-seq data and WGS. E.R.-F. and M.G. performed the SV analysis of the single-cell and WGS data. E.V. and M.C.S. performed the scRNA-seq data analysis. M.G. performed the fusion gene detection analyses. R.X. performed the SNV analyses in the single-cell and WGS data. L.B. performed and analyzed the FISH. K.Helmsauer and M.G. performed the amplicon reconstruction analyses. M.E.S. performed the ChIP–seq. K.Helmsauer performed the ChIP–seq analyses. H.D.G., K.S., Y.B., M.L. and K.L.H. performed the experiments and contributed to the data analysis. S.M., H.Y.C., H.E.D., S.S., A.E., J.H.S. and R.F.S. contributed to study design. R.C.G., R.P.K. and A.G.H. led study design, performed the data analysis and wrote the manuscript, to which all authors contributed.

Competing interests

R.P.K. and A.G.H. are founders of Econic Biosciences Ltd.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-023-01386-y.

Correspondence and requests for materials should be addressed to Anton G. Henssen.

Peer review information *Nature Genetics* thanks Andrea Ventura, Jan Korbel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Supplementary information for Chamorro *et al*. "Parallel sequencing of extrachromosomal circular DNAs and transcriptomes in single cancer cells"

Rocío Chamorro González[1,2], Thomas Conrad[18], Maja C. Stöber[3,4,17], Robin Xu[1,2], Madalina Giurgiu[1,2,5], Elias Rodriguez-Fos[1,2], Katharina Kasack[6], Lotte Brückner[2,19], Eric van Leen[1,2], Konstantin Helmsauer[1,2], Heathcliff Dorado Garcia[1,2], Maria E. Stefanova[7,8], King L. Hung[9], Yi Bei[1,2], Karin Schmelz[1], Marco Lodrini[1], Stefan Mundlos[7,8,10], Howard Y. Chang[9,11], Hedwig E. Deubzer[1,2,12,13], Sascha Sauer[3], Angelika Eggert[1,12,13], Johannes H. Schulte[1,12,13], Roland F. Schwarz[14,15,3], Kerstin Haase[1,2,12], Richard P. Koche[16*] and Anton G. Henssen[1,2,12,19*]

[1]Department of Pediatric Oncology and Hematology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany.

[2]Experimental and Clinical Research Center (ECRC) of the MDC and Charité Berlin, Berlin, Germany.

[3]Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany.

[4]Charité—Universitätsmedizin Berlin, Berlin, Germany.

[5]Freie Universität Berlin, Berlin, Germany.

[6]Fraunhofer Institute for Cell Therapy and Immunology, Branch Bioanalytics and Bioprocesses IZI-BB, Potsdam, Germany.

[7]RG Development and Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany

[8]Institute for Medical Genetics, Charité – Universitätsmedizin Berlin, Germany

[9]Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA

[10]Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité – Universitätsmedizin Berlin, Germany

[11]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA

[12]German Cancer Consortium (DKTK), partner site Berlin, and German Cancer Research Center (DKFZ), Heidelberg, Germany.

[13]Berlin Institute of Health, 10178 Berlin, Germany.

[14]Institute for Computational Cancer Biology (ICCB), Center for Integrated Oncology (CIO), Cancer Research Center Cologne Essen (CCCE), Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany.

[15]Berlin Institute for the Foundations of Learning and Data (BIFOLD), Berlin, Germany.

[16]Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, New York, USA.

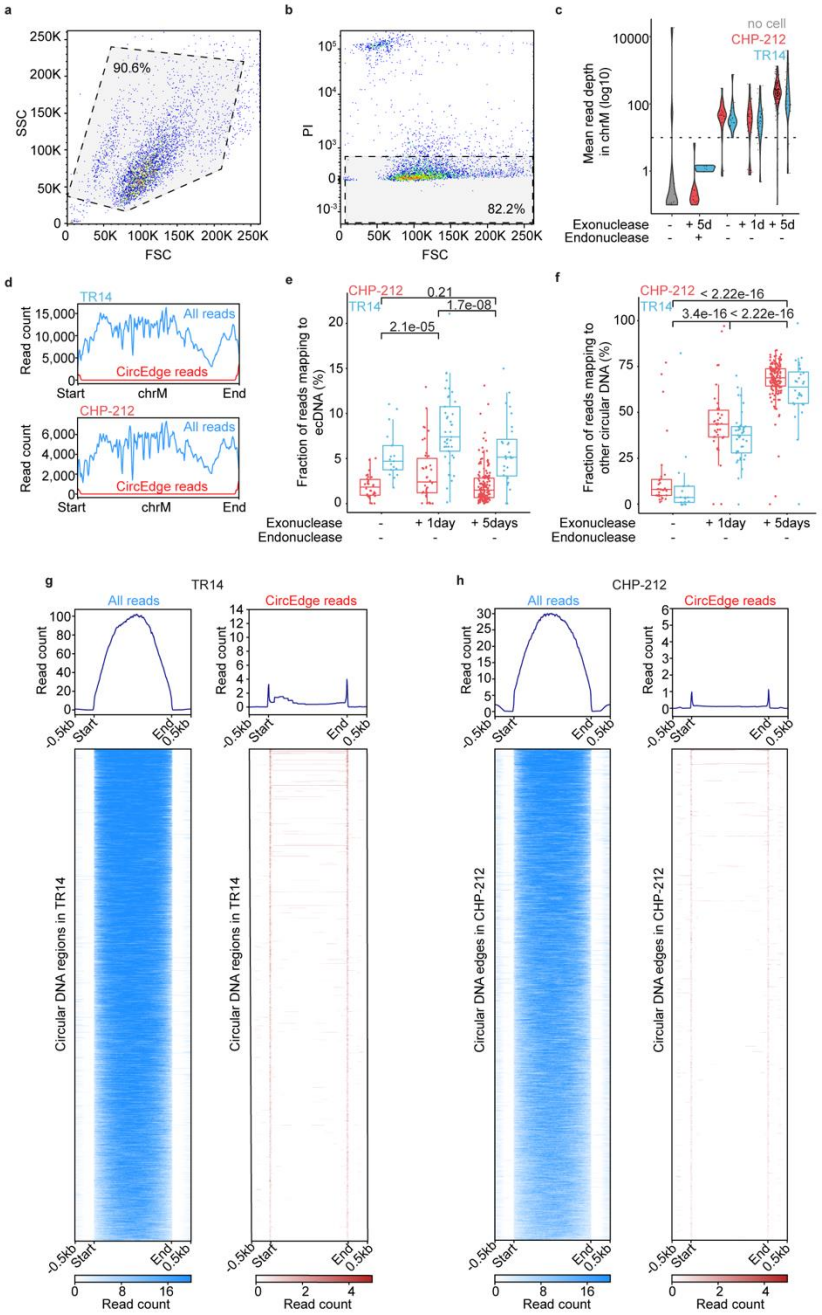
[17]Humboldt-Universität zu Berlin, Faculty of Life Science, 10099 Berlin, Germany.

[18]Genomics Technology Platform, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

[19]Max-Delbrück-Centrum für Molekulare Medizin, Berlin, Germany.

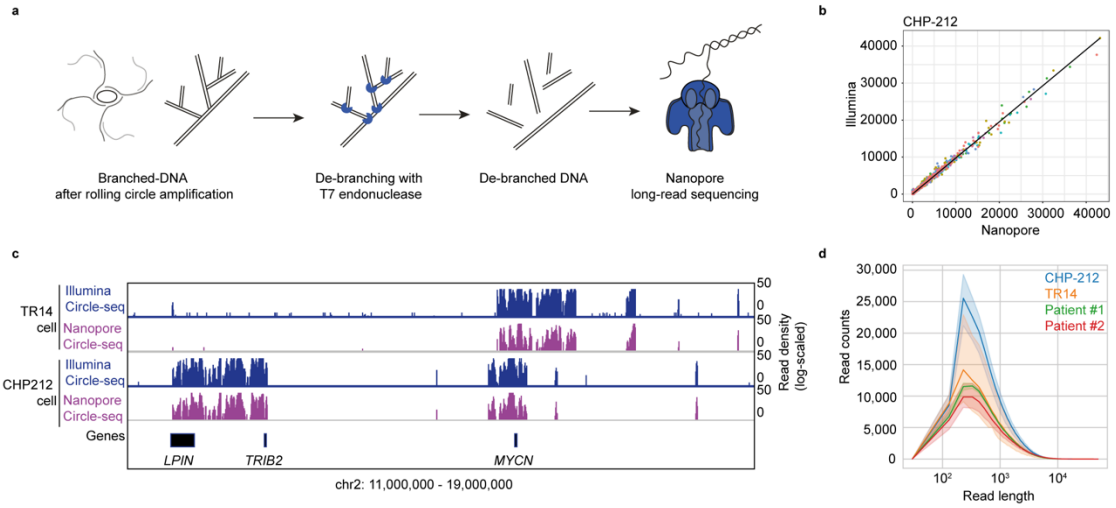
*These authors jointly supervised this work.

Correspondence should be addressed to A.G.H. (henssenlab@gmail.com).

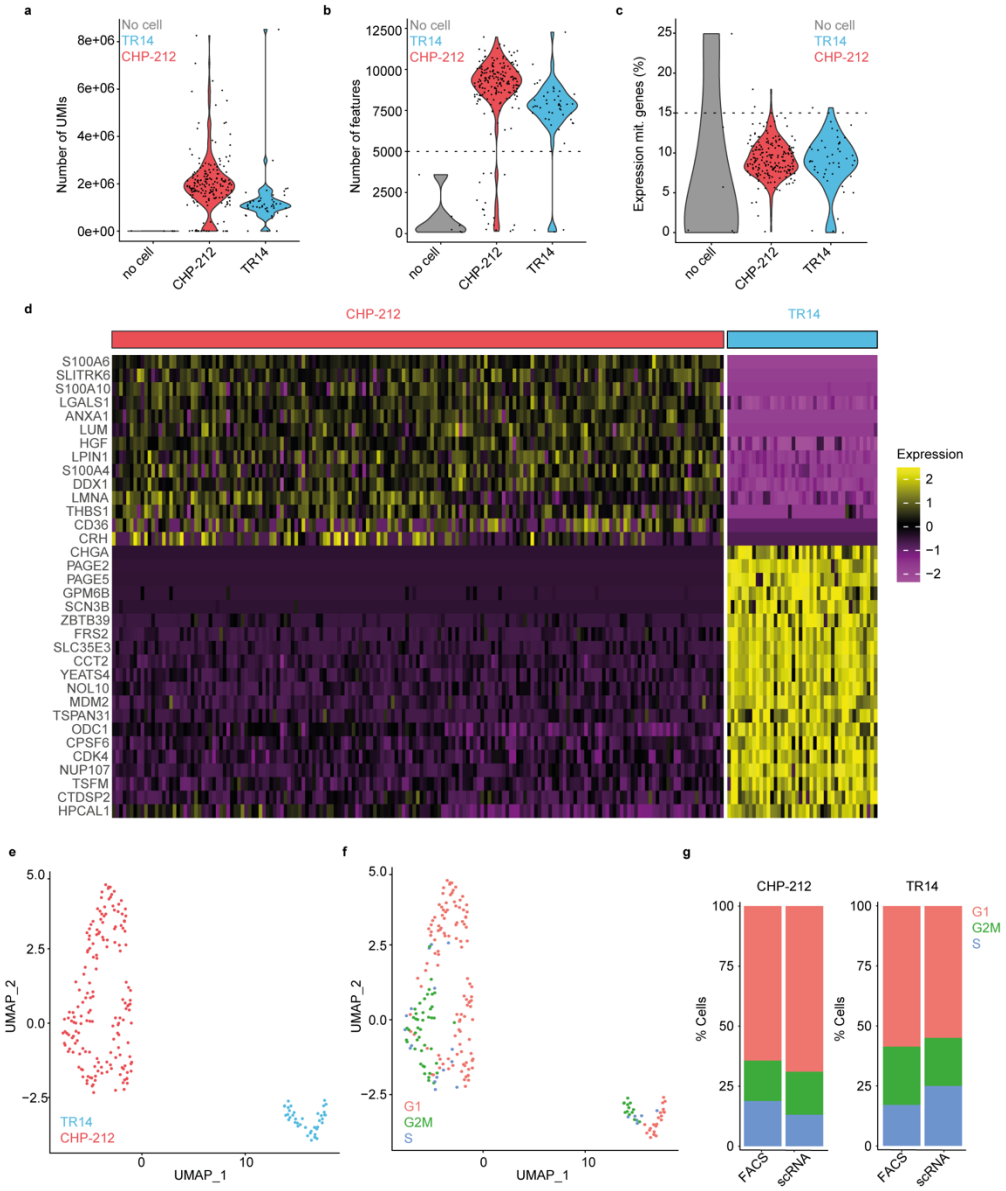


Supplementary Figure 1. Quality control of single cell Circle-seq data (scCircle-seq). a-b, Results from Fluorescence-activated cell sorting (FACS) of CHP-212 cells. Forward scatter (FSC) vs. side scatter (SSC) plot (a) with gate (dotted line) to separate events from debris. Forward scatter (FSC) vs. propidium iodide (PI) plot (b) with gate (dotted line) to separate live cells from dead cells. Cell percentages are shown. **c,** Violin plot showing per base-pair mean read depth in mitochondrial DNA (chrM) in TR14 cells (blue; *n* = 17 non digested cells, *n* = 41 one-day exonuclease digested cells, *n* = 28 five-days exonuclease digested cells, *n* = 6 endonuclease and exonuclease digested cells) and CHP-212 cells (red; *n* = 30 non digested cells, *n* = 38 one-day exonuclease digested cells, *n* = 154 five-days exonuclease digested cells, *n* = 12 endonuclease and exonuclease digested cells) and empty wells (no cell, grey; *n* = 7). For QC filtering of scCircle-seq data, we used a threshold of minimum 10 per base read depth (dotted line). **d**, Total read (blue line) and circle-edge split read count density (red line) over mtDNA (chrM) in merged TR14 cells (top) and merged CHP-212 cells (bottom). **e**, Fraction of sequencing reads mapping to ecDNA regions in each experimental condition in CHP-212 and TR14 cells. **f,** Fraction of sequencing reads mapping

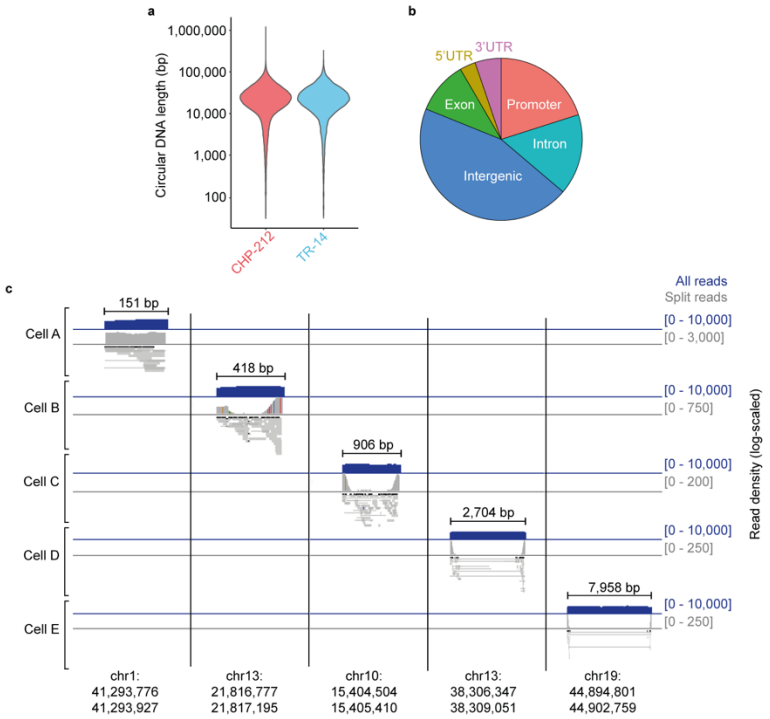
to other circular DNA regions (excluding ecDNA regions and mtDNA) in each experimental condition in CHP-212 and TR14 cells. In figs. e and f, sample size is identical across conditions: no digestion (n = 16 TR14 cells, n = 28 CHP-212 cells), one-day exonuclease digestion (n = 37 TR14 cells, n = 31 CHP-212 cells), five-days exonuclease digestion (n = 25 TR14 cells, n = 150 CHP-212 cells) and endonuclease digestion with PmeI prior to five-days exonuclease digestion (n = 6 TR14 cells, n = 12 CHP-212 cells). **g,h,** Total read and circle-edge split read count density over individual circular DNA regions identified by scEC&T-seq in merged TR14 cells (e) and merged CHP-212 cells (f). Top left and right: cumulative plots of total read count (left) and circle-edge split read count (right) density over all circular DNA regions; bottom left and right: heat map of total read count (left, blue) and circle-edge split read count (right, red) in all identified circular DNA regions. All statistical analyses correspond to two-sided Welch's *t*-test. In all boxplots, boxes represent 25th and 75th percentile with center bar as median value and whiskers represent furthest outlier ≤1.5× the interquartile range from the box.



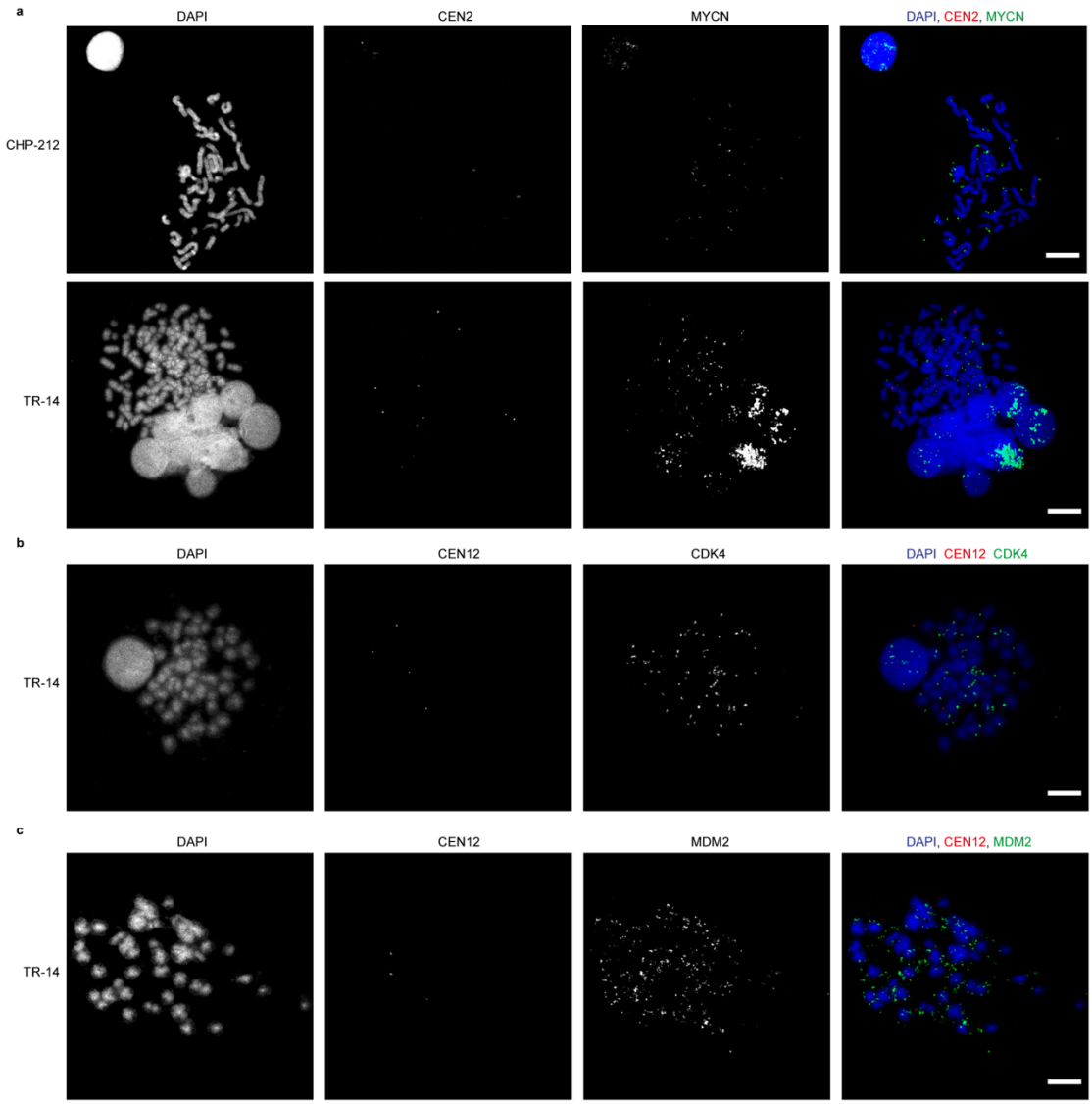
Supplementary Figure 2. Nanopore-based detection of extrachromosomal circular DNAs in single cells. a, Schematic of T7 endonuclease de-branching of rolling-circle amplified DNA prior to nanopore Circle-seq. **b**, Correlation of normalized read counts from Illumina and Nanopore scCircle-seq data from a subset of CHP-212 cells (log-scaled, two-sided Pearson correlation: R = 0.95, P < 2.2e-16). Each color represents a different cell, and each point is a putative circle. **c**, Genome tracks comparing log-scaled read coverage across the *MYCN* ecDNA amplicon regions in Illumina (blue) vs Nanopore (pink) Circle-seq data in two exemplary cells (CHP-212 and TR14). **d,** Read length distribution of Nanopore sCircle-seq data. Individual lines represent the average across single cells grouped by sample (n = 6 CHP-212 cells (blue), n = 3 TR-14 cells (orange), n = 4 patient #1 nuclei (green), n = 5 patient #2 nuclei (red)), whereas the shade stands for 95% confidence interval.



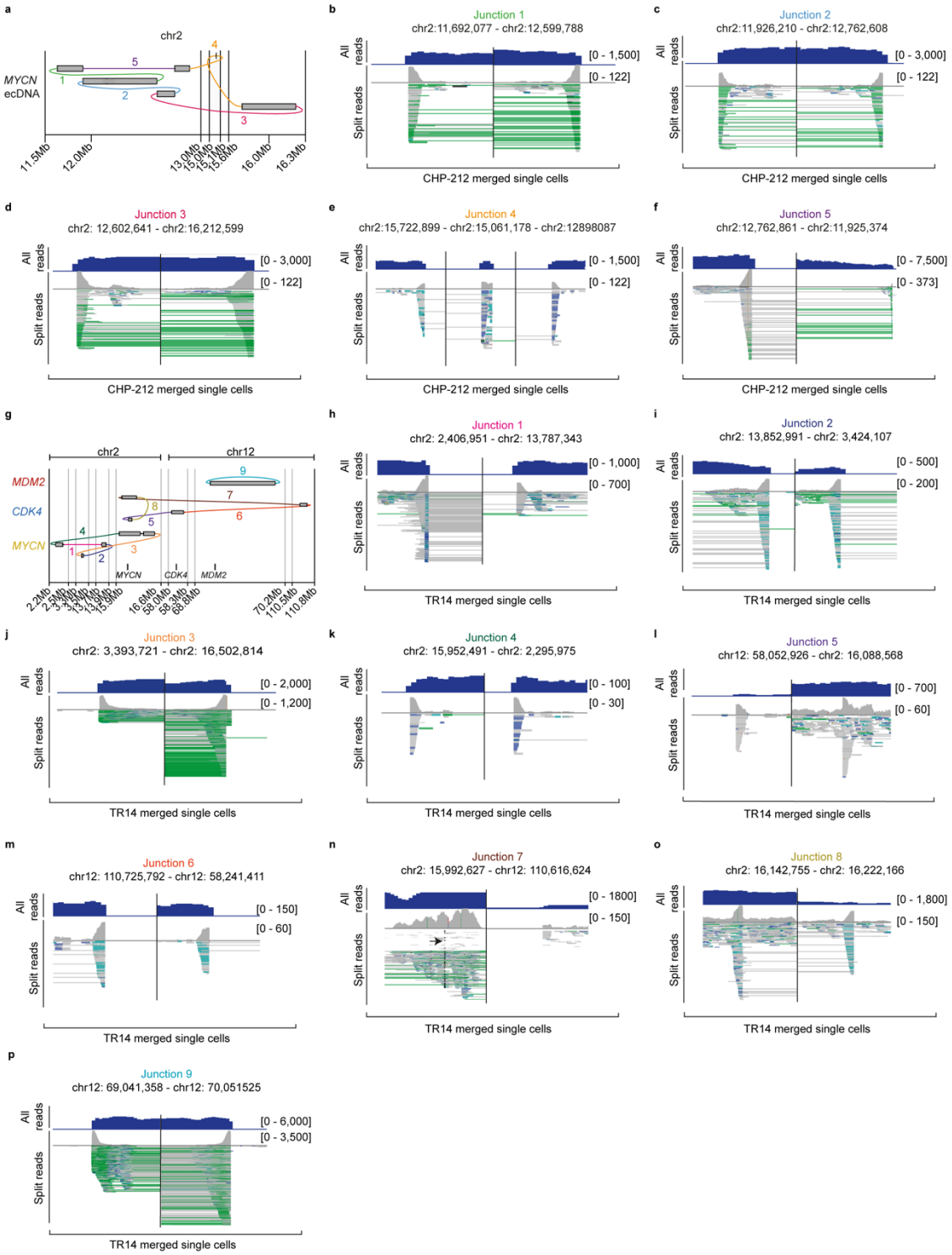
Supplementary Figure 3. **Quality control of scRNA-seq data. a,** Violin plot of number of unique molecular identifiers (UMIs) in CHP-212 (red; *n* = 171) and TR14 (blue; *n* = 42) cells and in negative controls (grey; *n* = 5). **b,** Violin plot of number of features (genes) identified in CHP-212 (red; *n* = 171; mean +/- sd = 9,328 +- 1,006) and TR14 (blue; *n* = 42; mean +/- sd = 7,961 +- 1,124) single cells and in negative controls (grey; *n* = 5). **c,** Violin plot of fraction of expression (%) of mitochondrial genes in CHP-212 (red; *n* = 171) and TR14 (blue; *n* = 42) single-cells and in negative controls (grey; *n* = 5). **d,** Heatmap showing the top 20 differentially expressed genes between CHP-212 and TR14 cells (*n* = 171 CHP-212 cells in red, *n* = 42 TR14 cells in blue). **e,** UMAP visualization showing clusters of transcriptionally similar cells colored by cell line identity (*n* = 171 CHP-212 cells in red, *n* = 42 TR14 cells in blue). **f,** UMAP visualization illustrating cells from **e** colored by predicted cell cycle phase (G1 in red, G2M in green, S in blue). **g,** Bar plots comparing the relative distribution of CHP-212 and TR14 cells (%) across cell cycle phases measured by FACS-based cell cycle analysis (PI) or inferred from scRNA-seq data (G1 in red, G2M in green, S in blue).



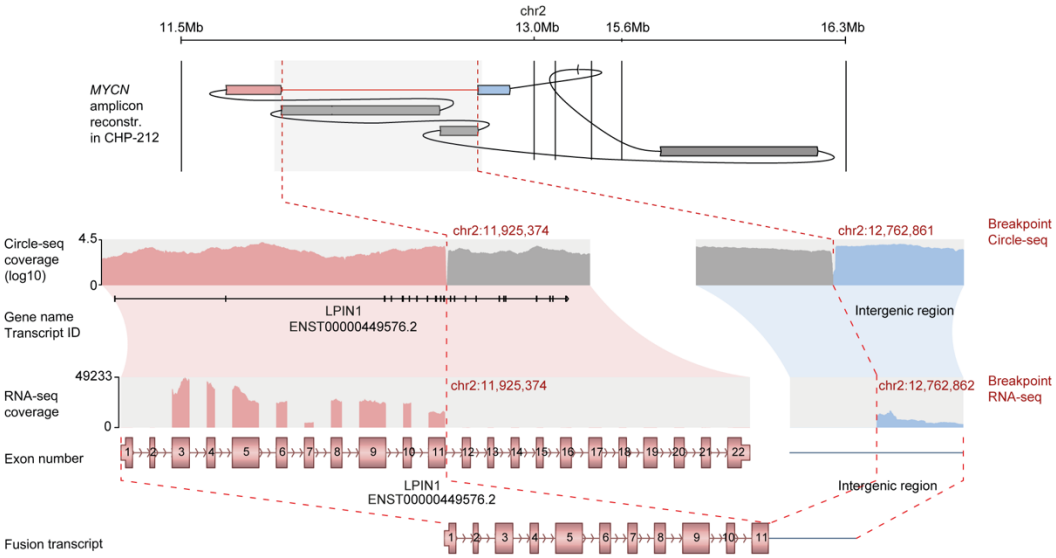
Supplementary Figure 4. The majority of small extrachromosomal circular DNAs are not recurrently identified in single cells a, Size distribution of extrachromosomal circular DNAs identified using scEC&T-seq in CHP-212 and TR14 single cells (log-scaled, *n* = 150 CHP-212 cells, *n* = 25 TR14 cells). **b,** Fraction of genomic regions affected by extrachromosomal DNA circularization in TR14 (*n* = 25 cells) and CHP-212 (*n* = 150 cells) (promoter = 20.11% (red), intron = 16.10% (light blue), intergenic = 44.88% (dark blue), exon = 10.51% (green), 5'UTR = 3.16% (brown), 3'UTR = 5.24% (pink)). **c,** Exemplary genome tracks of non-recurrent extrachromosomal DNA circularization in 5 different CHP-212 single cells. Log-scaled total read density is shown in blue and log-scaled circle edge read density is shown in grey.



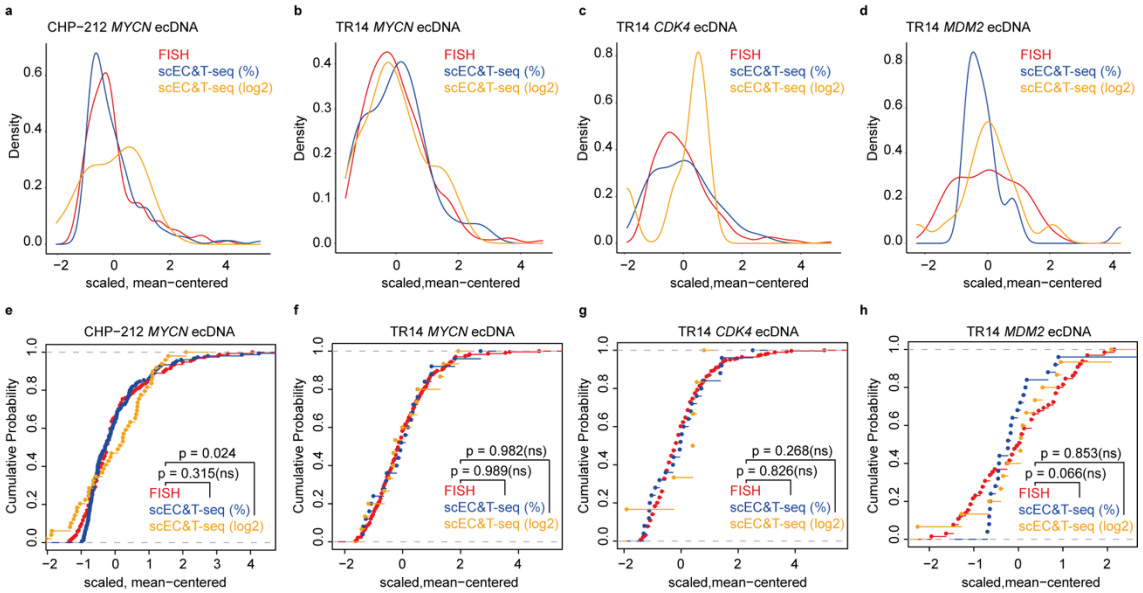
Supplementary Figure 5. Fluorescence in situ hybridization in neuroblastoma cell lines. a, DNA - fluorescence in situ hybridization (FISH) of CHP-212 (top row) and TR14 (bottom row) metaphase spreads with *MYCN* probe (green) and control chromosomal probe for chromosome 2 (CEN2; red). Nuclei and chromosomes are stained with DAPI (blue). Channels left to right: DAPI, CEN2, *MYCN* and merged. **b,** FISH of TR14 metaphase spreads with *CDK4* probe (green) and control chromosomal probe for chromosome 12 (CEN12; red). Nuclei and chromosomes are stained with DAPI (blue). Channels left to right: DAPI, CEN12, *CDK4* and merged. **c,** FISH of TR14 metaphase spreads with *MDM2* probe (green) and control chromosomal probe for chromosome 12 (CEN12; red). Nuclei and chromosomes are stained with DAPI (blue). Channels left to right: DAPI, CEN12, *MDM2* and merged. In all cases, the scale bar is 10 μm. FISH experiments were done once per cell line.



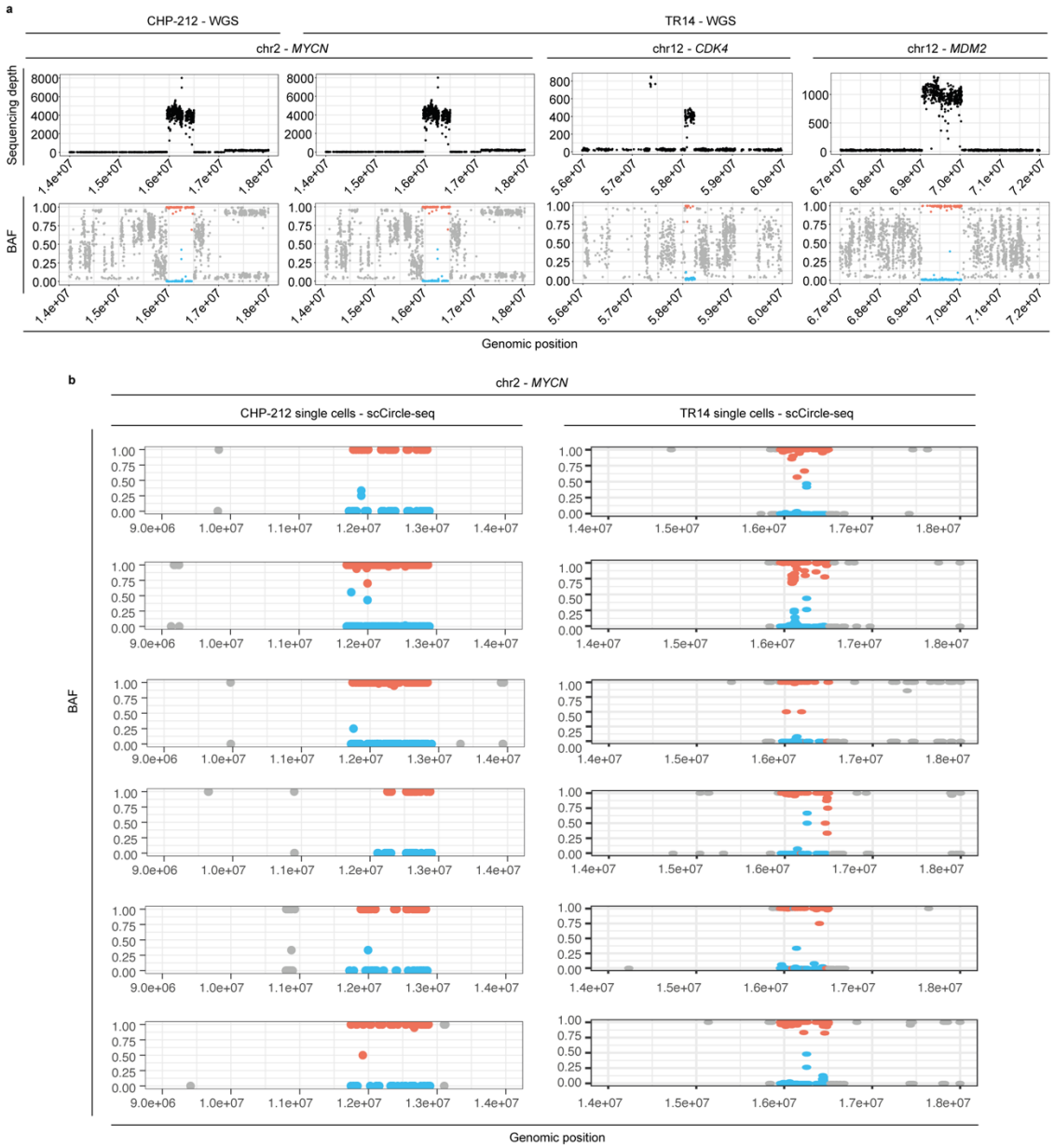
Supplementary Figure 6. Detection of ecDNA junction-supporting split reads in scEC&T-seq data. a, Long and short-read based ecDNA amplicon reconstruction from whole-genome bulk sequencing data in CHP-212. (**b-f),** Total and split read density at predicted junctions in merged CHP-212 scEC&T-seq data. **g,** Long and short-read based ecDNA amplicon reconstructions from whole-genome bulk sequencing data in TR14. (**h-p),** Total and split read density at predicted junctions in merged CHP-212 scEC&T-seq data.



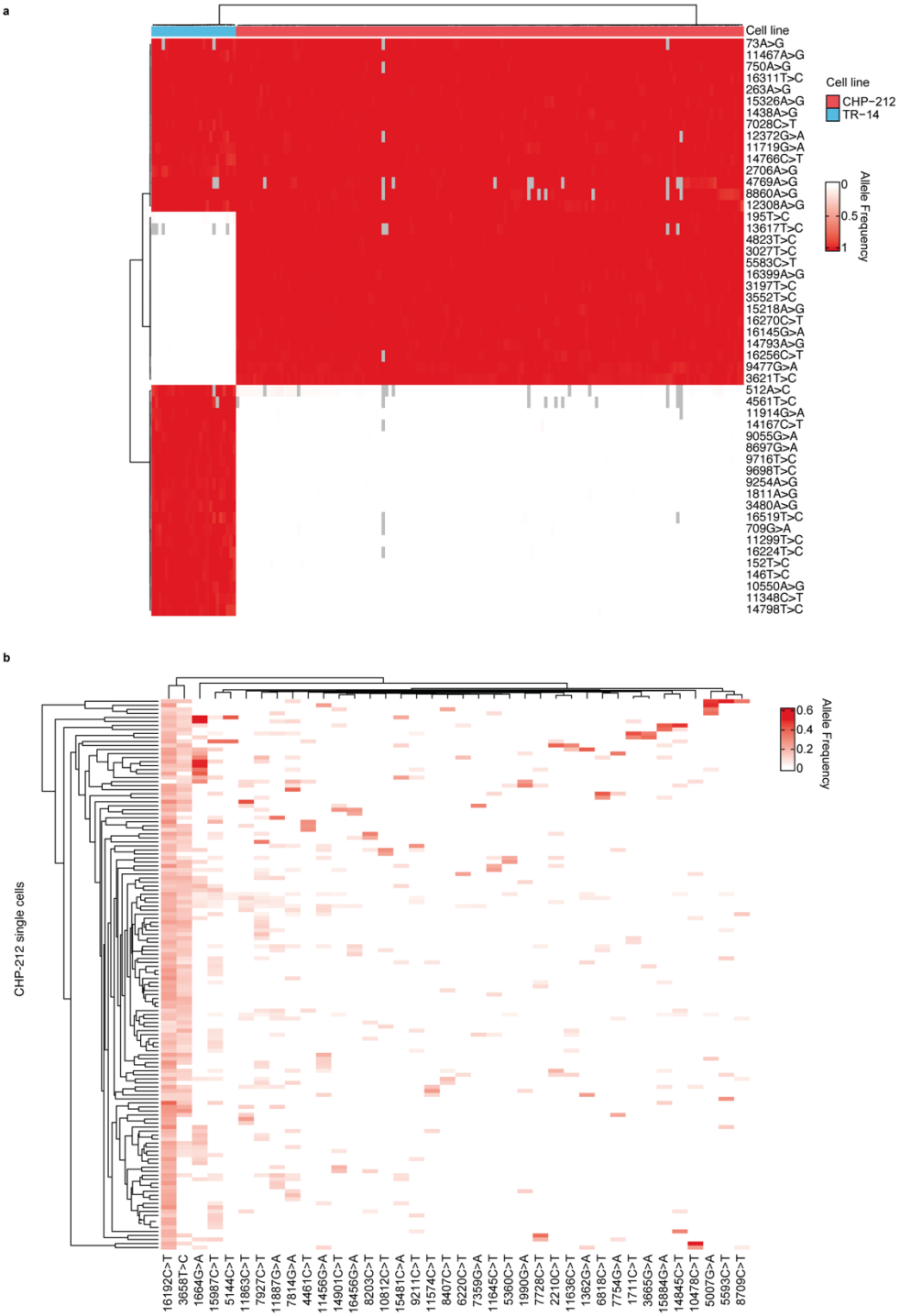
Supplementary Figure 7. Detection of fusion transcript from scRNA-seq data. Example of a fusion transcript resulting from the rearrangement of chromosomal segments in the *MYCN* ecDNA in CHP-212 cells. Top to bottom: amplicon reconstruction from WGS data, Circle-seq read coverage over the breakpoint region in merged CHP-212 single cells (log-scaled), transcript annotations, merged scRNA-seq read coverage over the fused transcripts, native transcripts representations, fusion transcript representation.



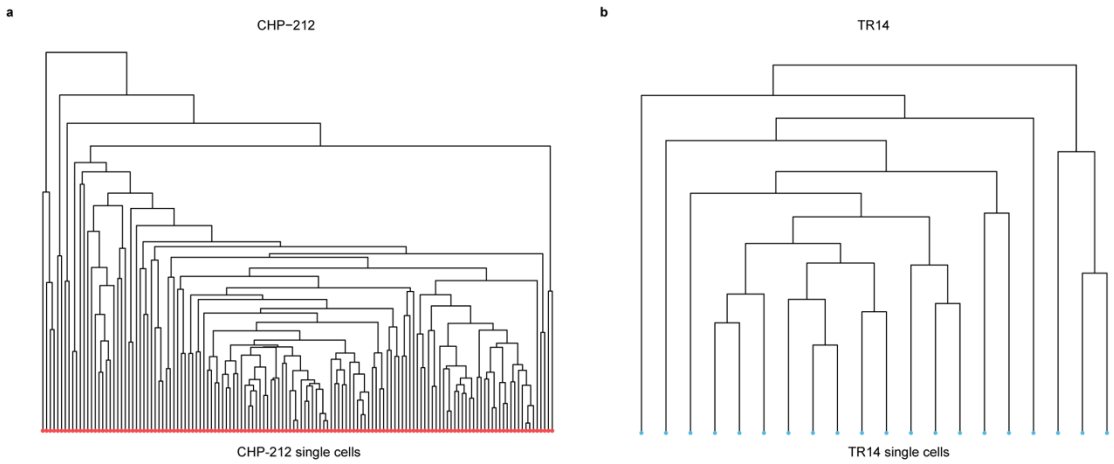
Supplementary Figure 8: Relative ecDNA copy number measured using scEC&T-seq resembles FISH-based copy number estimates. a-d, Density plots displaying the scaled, mean-centered ecDNA relative copy number distributions in CHP-212 (a, *MYCN* ecDNA) and TR14 (b, *MYCN* ecDNA; c, *CDK4* ecDNA; d, *MDM2* ecDNA), as measured by *MYCN* DNA interphase FISH (red, *n* =154 (a), *n* = 232 (b), *n* = 284 (c), *n* = 65(d)), log2 *MYCN* coverage in scEC&T-seq (yellow, *n* = 49 (a), *n* = 15 (b), *n* = 6 (c), *n* = 15(d)) and fraction of ecDNA-specific reads in scEC&T-seq (blue, *n* = 150 (a), *n* = 25 (b-d)). **e-h,** Cumulative probability of scaled, mean-centered ecDNA relative copy number in CHP-212 (a, *MYCN* ecDNA) and TR14 (b, *MYCN* ecDNA; c, *CDK4* ecDNA; d, *MDM2* ecDNA), as measured by *MYCN* DNA interphase FISH (red, *n* = 154 (a), *n* = 232 (b), *n* = 284 (c), *n* = 65(d)), log2 *MYCN* coverage in scEC&T-seq (yellow, *n* = 49 (a), *n* = 15 (b), *n* = 6 (c), *n* = 15(d)) and fraction of ecDNA-specific reads in scEC&T-seq (blue, *n* = 150 (a), *n* = 25 (b-d)). P-values were calculated by Kolmogorov-Smirnov test and are shown. FISH experiments were done once per cell line.



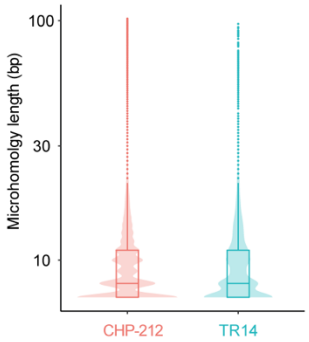
Supplementary Figure 9. Phasing of SNPs in ecDNA loci in scEC&T-seq data indicates ecDNAs are of mono-allelic origin. a, Reference phasing of *MYCN*, *CDK4* and *MDM2* ecDNA loci in bulk WGS data. Shown is the raw sequencing coverage (top) and the B-allele frequency (BAF; bottom) of known SNPs based on the 1000 genomes annotation. SNPs in regions of high-level amplifications can be very clearly assigned to the gained or non-gained allele based on BAF. **b,** Genotyped *MYCN* ecDNA locus in scCircle-seq CHP-212 and TR14 sequencing data (6 exemplary cells in each case are shown). Shown is the B-allele frequency (BAF) of known SNPs based on the 1000 genomes annotation. SNPs that have been reference phased based on bulk sequencing data are colored the same as in (a).



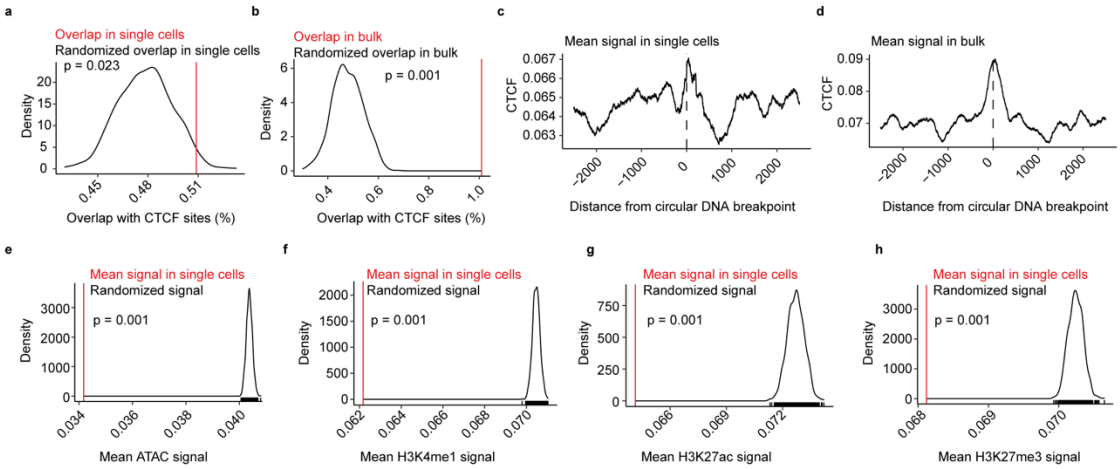
Supplementary Figure 10. scEC&T-seq enables identification of homoplasmic and heteroplasmic variants (SNVs) in mitochrondrial DNA. a, Homoplasmic mitochondrial single nucleotide variants (SNV) detected in CHP-212 and TR14 single cells (*n* = 150 CHP-212 cells, *n* = 25 TR14 cells). Unsupervised hierarchical clustering allows for clear separation of both cell lines based on their haplogroup variants, suggesting usage for population scale phylogeny studies. Sites with read depth ≤ 10 are shown in grey. **b,** Heteroplasmic variants detected in CHP-212 single cells (*n* = 150 CHP-212 cells). Unsupervised hierarchical clustering (y-axis) suggests usage for lineage tracing exploration and applications. Sites with read depth ≤ 10 are shown in grey.



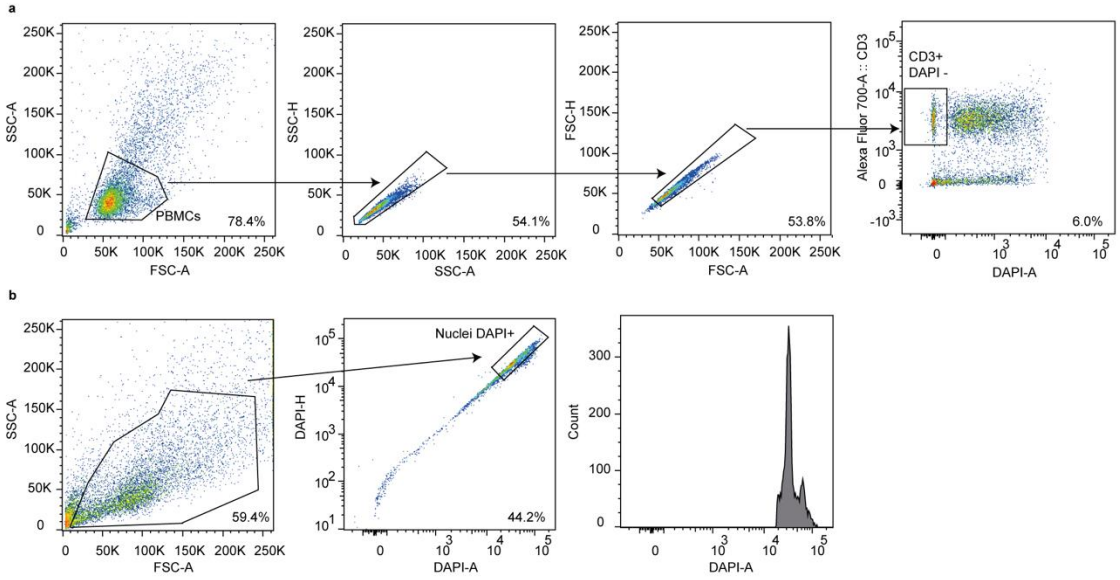
Supplementary Figure 11. **Mitochondrial heteroplasmic SNVs can be used to infer phylogeny. a,b,** Phylogenetic trees inferred from heteroplasmic single-nucleotide variants identified in mitochondrial DNA in CHP-212 (a; *n* = 148) and TR14 (b; *n* = 20) single cells.



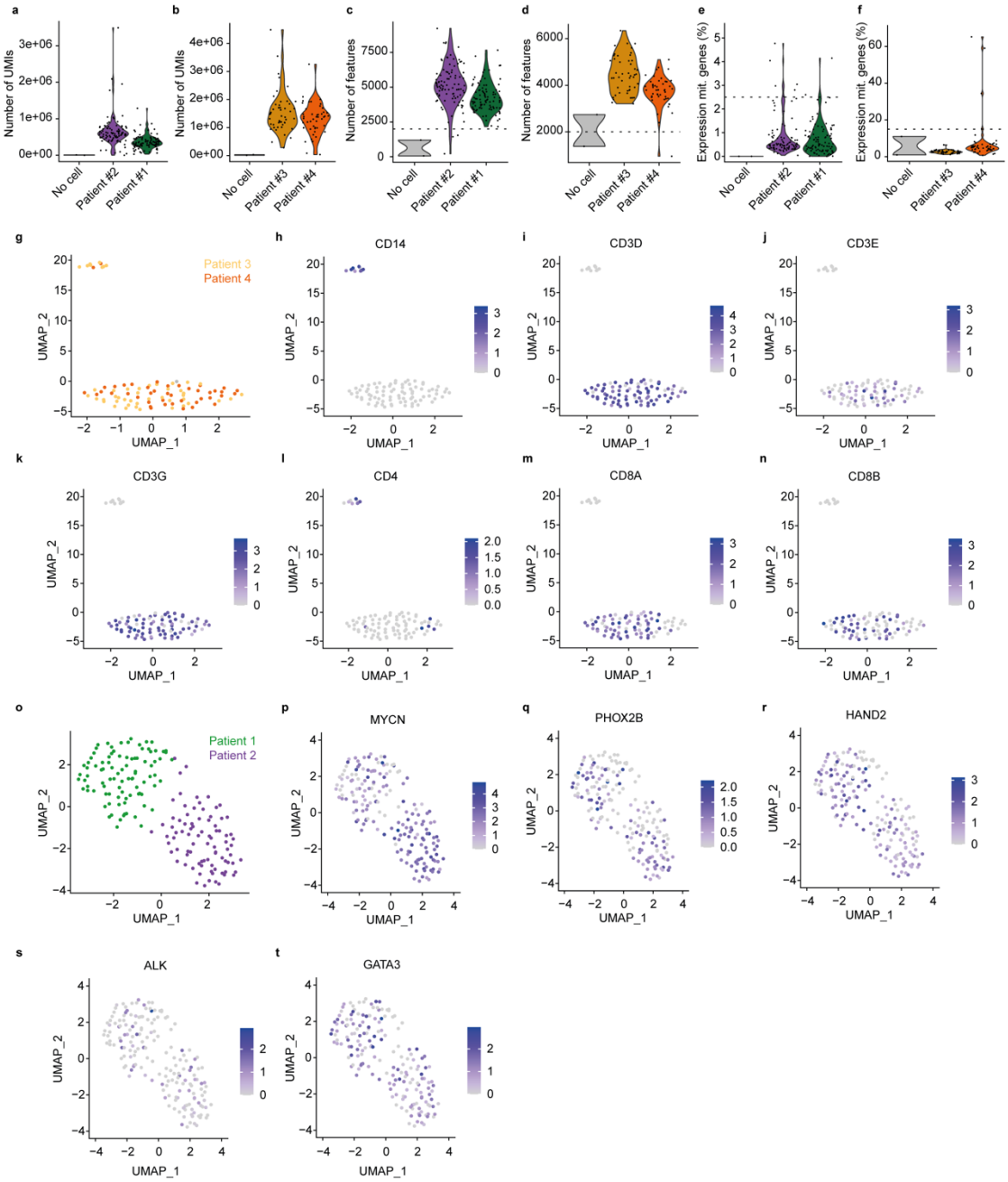
Supplementary Figure 12. **Microhomology detection at circular DNA breakpoints. a,** Length distribution of microhomologies in CHP-212 and TR14 single cells (*n* = 150 CHP-212 cells and *n* = 25 TR14 cells). All boxplot's box represents 25th and 75th percentile with center bar as median value and whiskers represent furthest outlier ≤1.5× the interquartile range from the box.



Supplementary Fig. 13. **Chromatin marks and chromatin accessibility in extrachromosomal circular DNAs in single cells.** **a, b,** Fraction of circular DNA edge regions overlapping with CTCF ChIP-seq peaks in CHP-212 single cells (a) and in bulk CHP-212 Circle-seq (b). Overlap shown in red and randomized overlap in dark**. c, d,** Mean CTCF signal around the edges (dashed line) of circular DNA regions detected in all CHP-212 single cells (c) and in bulk CHP-212 Circle-seq (d). **e-h,** Mean ChIP-seq or ATAC-seq signal across all detected circular DNA regions in all CHP-212 single cells (red) and randomized signal (black): ATAC-seq (e, *P* = 1e-06); H3K4me1 ChIP-seq (f, *P* = 0.001); H3K27ac ChIP-seq (g, *P* = 1e-06); H3K27me3 ChIP-seq (h, *P* = 0.494). Empirical one-sided p-values were used from randomization analyses.

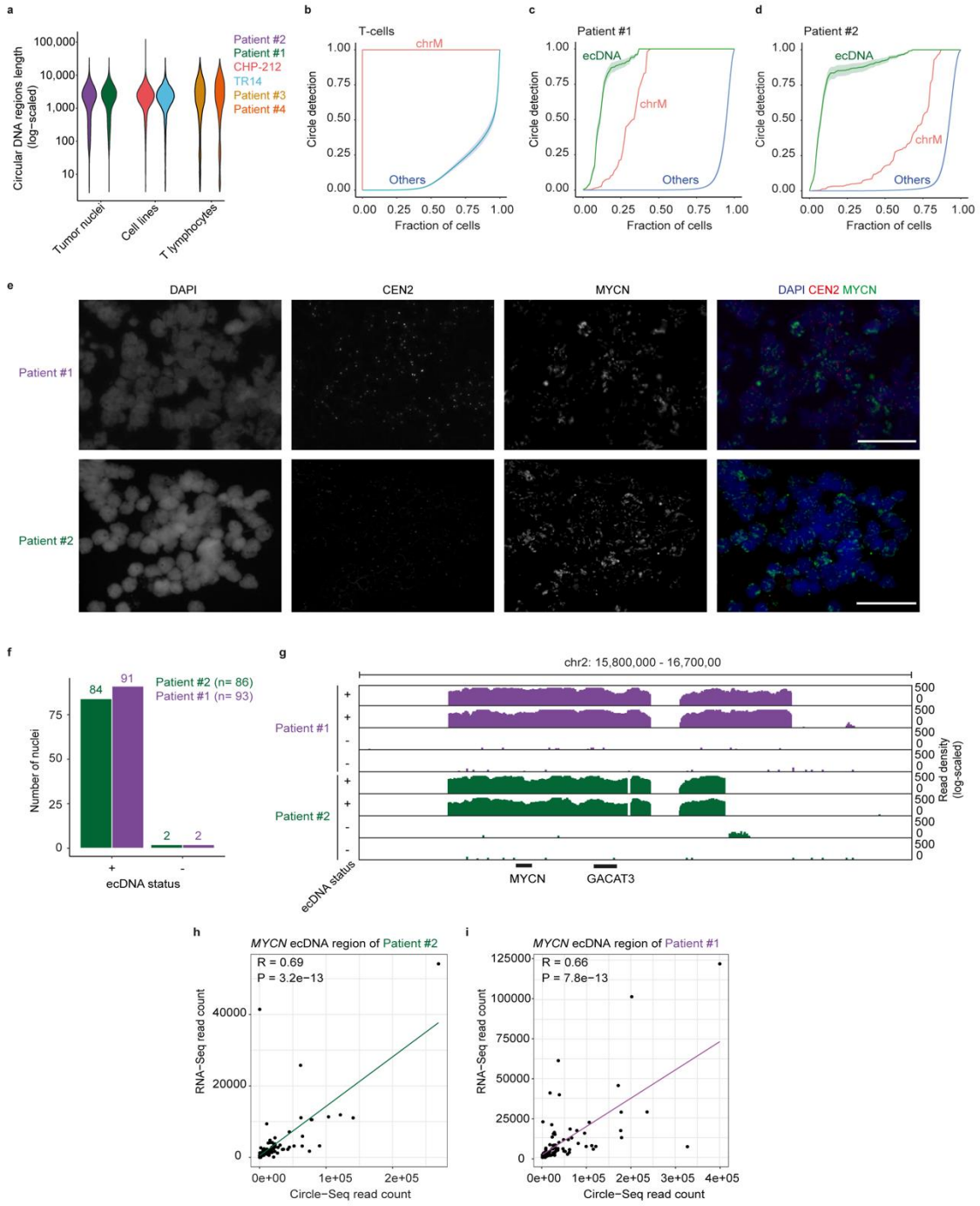


Supplementary Fig. 14: Representative FACS gating strategy for PBMCs and nuclei. a, Representative gating strategy of CD3+ DAPI- live T-cells from PBMCs population derived from patient's blood in patient #3. **b,** Representative gating strategy for nuclei isolated from primary tumor samples. In both cases, Forward scatter (FSC) and side scatter (SSC) were used to separate events from debris. DAPI, in PBMCs, was used to stain and filter out dead cells. DAPI+ nuclei were sorted. Gating strategy and cell percentages are shown in each case.



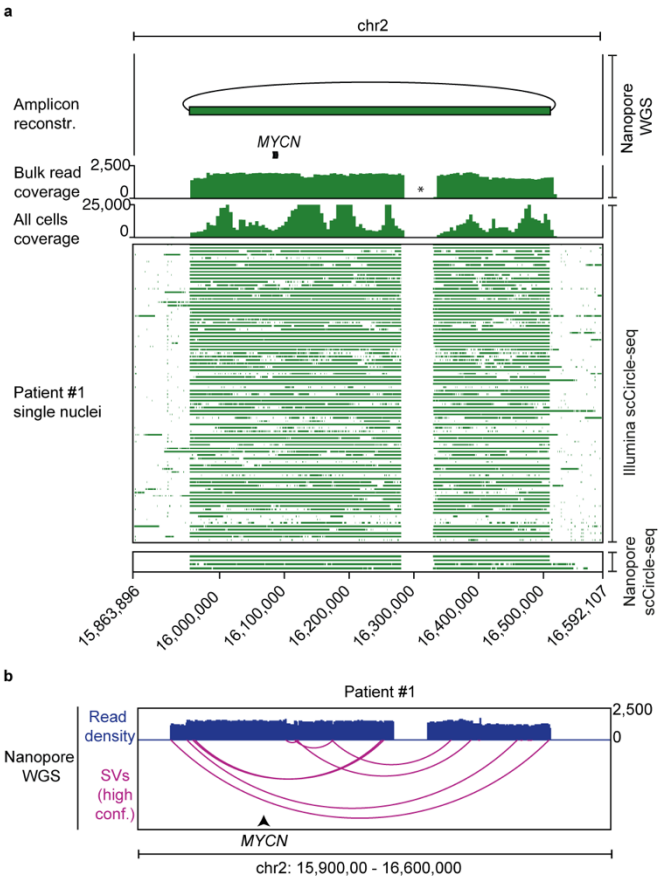
Supplementary Figure 15. Quality control of tumor and T-cell scRNA-seq data. a,b, Violin plot of number of unique molecular identifiers (UMIs) in primary tumor nuclei (a, *n* = 93 nuclei patient #1, green; *n* = 86 nuclei patient #2, purple) and single T-cells (b, *n* = 38 patient #3, yellow; *n* = 41 patient #4, orange). **c,d,** Violin plot of number of features (genes) identified in primary tumor nuclei (c, *n* = 93 nuclei patient #1, green; *n* = 86 nuclei patient #2, purple) and single T-cells (b, *n* = 38 patient #3, yellow; *n* = 41 patient #4, orange). **e,f,** Violin plot of fraction of expression (%) of mitochondrial genes in primary tumor nuclei (e, *n* = 93 nuclei patient #1, green; *n* = 86 nuclei patient #2, purple) and single T-cells (b, *n* = 38 patient #3, yellow; *n* = 41 patient #4, orange). **g,** UMAP visualization showing clusters of transcriptionally similar cells colored by patient identity (*n* = 38 patient #3 shown in yellow, *n* = 41 patient #4 shown in orange). **h-n,** UMAP visualization showing relative expression of marker genes: *CD14* (h), *CD3D* (i), *CD3E* (j), *CD3G* (k), *CD4* (l), *CD8A* (m), *CD8B* (n). **o,** UMAP visualization showing clusters of transcriptionally similar nuclei colored by patient identity (*n* = 93 nuclei patient #1 shown in green, *n* = 86 nuclei patient

#2 shown in purple). **p-t,** UMAP visualization showing relative expression of neuroblastoma marker genes: *MYCN* (p), *PHOX2B* (q), *HAND2* (r), *CD3G* (k), *ALK* (s), *GATA3* (t).

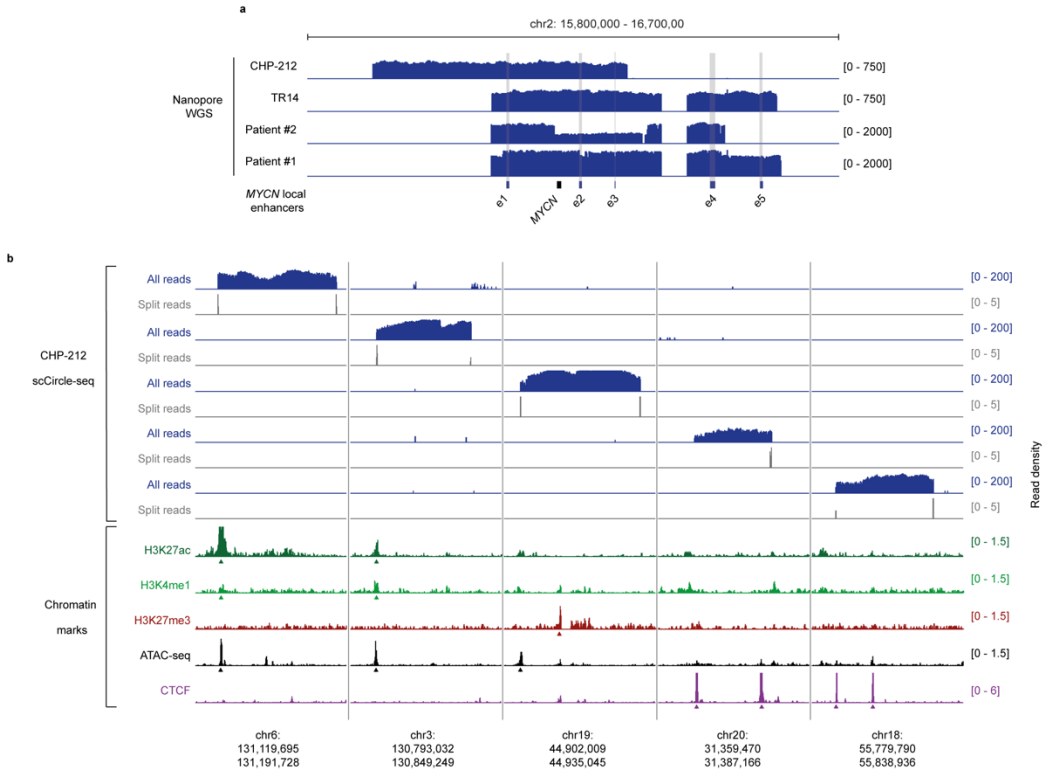


Supplementary Figure 16. Intercellular differences in ecDNA content in primary neuroblastoma tumors drive gene expression differences. a, Log-scaled size distribution of extrachromosomal circular DNAs identified using scEC&T-seq in primary tumor nuclei (*n* = 93 nuclei patient #1 shown in green; *n* = 86 nuclei patient #2 shown in purple), neuroblastoma cell lines (*n* = 150 CHP-212 in red, *n* = 25 TR14 in blue) and single T-cells (*n* = 38 patient #3 shown in yellow; *n* = 41 patient #4 shown in orange).**b-d,** Recurrence analysis in single T-cells (b, *n* = 79 T-cells), primary tumor nuclei from patient #1(c, *n* = 93 nuclei) and pantient #2(d, *n* = 86 nuclei) displayed as fraction of cells containing a detected circular DNA from each circular DNA type. ecDNA was defined as extrachromosomal circular DNAs overlapping with copy number amplified regions identified in bulk sequencing (green); chrM (red); "Others" are defined as all other extrachromosomal circular DNAs (blue). Data are presented as mean values +/- SEM. **e,** Interphase

FISH of patient #1 (top row) and patient #2 (bottom row) with *MYCN* probe (green) and control chromosomal probe for chromosome 2 (CEN2; red). Nuclei are stained with DAPI (blue). Channels left to right: DAPI, CEN2, *MYCN* and merged. Scale bar indicates 50 μm. **f,** Bar plot of number of nuclei based on ecDNA status (n = 93 nuclei patient #1 shown in green; n = 86 nuclei patient #2 shown in purple). **g,** Genome tracks with read densities (log-scaled) over reconstructed *MYCN* ecDNA region in 8 exemplary patient #1(green) and #2(purple) nuclei showing + and - ecDNAs status. **h,i,** Pairwise comparison between ecDNA and mRNA read counts from scEC&T-seq over the reconstructed MYCN ecDNA region in patient #2 single nuclei (h; two-sided Pearson correlation, P = 3.2e-13, R = 0.69, n = 86 patient #2 nuclei) and in patient #1 single nuclei (i; two-sided Pearson correlation, P = 7.6e-13, R = 0.66, n = 93 patient #1 nuclei).



Supplementary Figure 17. Patient #1 presents a complex ecDNA structure with multiple internal rearrangements. a, Simplified long-read based amplicon reconstruction derived from WGS sequencing data in bulk cell populations and read coverage over the ecDNA region across single cells in patient #1 (*n* = 93 nuclei) as detected by scEC&T-seq. Top to bottom: simplified ecDNA amplicon reconstruction, copy number profile, gene annotations, read density over the ecDNA region in merged single cells, coverage over the ecDNA region in single cells (rows) as detected by short or long-read scEC&T-seq. **b,** Genome track of long-read nanopore WGS data displaying read density across the *MYCN* ecDNA region in patient #1, and showing the identified high-confidence SVs within the ecDNA amplicon region.



Supplementary Figure 18. *MYCN* local enhancers are recurrently included in ecDNA in neuroblastoma. a, Genome tracks of long-read nanopore WGS data showing co-amplification in ecDNA of *MYCN* local enhancers (e1-e5) with the *MYCN* proto-oncogene in neuroblastoma cell lines (CHP-212 and TR14) and primary tumors (patient #1 and #2). **b,** Overlap of non-recurrent extrachromosomal circular DNAs with chromatin marks. Top to bottom: 5 exemplary genome tracks of 5 different CHP-212 single cells (Log-scaled total read density in blue and circle edge read density in grey); H3K27ac ChIP-seq (dark green); H3K4me1 ChIP-seq (light green); H3K27me3 ChIP-seq (red); ATAC-seq (black); CTCF ChIP-seq (purple).

Supplementary Table 7: Oligos used in scEC&T-seq protocol

Primer Name	**Primer Sequence**	**Provider**
Oligo-dT	5′-biotin-triethyleneglycol-AAGCAGTGGTATCAAC-GCAGAGTACT30VN-3′,	IDT
Template Switch Oligo (TSO)	5′-AAGCAGTGGTATCAACGCAGAGTACrGrG+G-3′, where “r” indicates a ribonucleic acid base and “+” indicates a locked nucleic acid base	IDT
ISPCR	5′-AAGCAGTGGTATCAACGCAGAGT-3	IDT

Supplementary Note 1: QC results from scEC&T-seq data in cell lines, tumor nuclei and T-cells

Based on the observation that longer exonuclease exposure improved small circular DNA enrichment and recovered large ecDNA, we used sequencing data from cells digested with exonuclease for 5 days for downstream analyses in all cases. In cell lines, a total of 182 cells (154 CHP-212 cells and 28 TR14 cells) were exposed to exonuclease digestion for 5 days. Of those, 175 cells (96.15%; 150 CHP-212 and 25 TR14 cells, Supplementary Table 1) passed scCircle-seq quality control. Of those cells, 149 cells additionally passed scRNA-seq quality control (81.87%; 129 CHP-212 cells and 20 TR14, Supplementary Table 1 and 2). We detected on average 9,058 +- 1,163 full mRNA transcripts from different genes per cell (9,328 +- 1,006 in CHP-212 and 7,961 +- 1,124 in TR14; Supplementary Fig. 3b and Supplementary Table 2). For a total of 94 single T-cells (47 per patient) and 190 primary tumor nuclei (95 per primary tumor sample), the DNA was exposed to exonuclease digestion for 5 days in all cases. Of those, 79 T-cells (84%; 38 patient #3 and 41 patient #4) and 179 nuclei (93 patient #1 and 86 patient #2) passed quality control criteria (Supplementary Table 1 and 2). We identified 3,793 +- 1,055 mRNA transcripts per cancer cell nucleus (Supplementary Fig. 15c; Supplementary Table 2) and 3,177 +- 541 transcripts per T-cell (Supplementary Fig. 15d; Supplementary Table 2). Both CD4+ and CD8+ T-cells were identified based on their expression profiles (Supplementary Fig. 15l-n). Unsupervised clustering of tumor nuclei separated two population based on patient origin (Supplementary Fig. 15o). In both cancer nuclei populations, we observed expression of common neuroblastoma cell markers, including *MYCN, PHOX2B, HAND2, ALK* and *GATA3* (Supplementary Fig. 15p-t).

9. Curriculum Vitae

My curriculum vitae does not appear in the electronic version of my paper for reasons of data protection.

10. Publication list

Rocío Chamorro González, Thomas Conrad, Maja C. Stöber, Robin Xu, Mădălina Giurgiu, Elias Rodriguez-Fos, Katharina Kasack, Lotte Brückner, Eric van Leen, Konstantin Helmsauer, Heathcliff Dorado Garcia, Maria E. Stefanova, King L. Hung, Yi Bei, Karin Schmelz, Marco Lodrini, Stefan Mundlos, Howard Y. Chang, Hedwig E. Deubzer, Sascha Sauer, Angelika Eggert, Johannes H. Schulte, Roland F. Schwarz, Kerstin Haase, Richard P. Koche & Anton G. Henssen. Parallel sequencing of extrachromosomal circular DNAs and transcriptomes in single cancer cells. ***Nature Genetics***, (2023). IF: 41.307

Eunhee Yi*, **Rocío Chamorro González***, Anton G. Henssen & Roel G. W. Verhaak. Extrachromosomal DNA amplifications in cancer. ***Nature Reviews Genetics***, (2022) ***Shared first authorship.** IF: 59.581

King L. Hung, Kathryn E. Yost, Liangqi Xie, Quanming Shi, Konstantin Helmsauer, Jens Luebeck, Robert Schöpflin, Joshua T. Lange, **Rocío Chamorro González**, Natasha E. Weiser, Celine Chen, Maria E. Valieva, Ivy Tsz-Lo Wong, Sihan Wu, Siavash R. Dehkordi, Connor V. Duffy, Katerina Kraft, Jun Tang, Julia A. Belk, John C. Rose, M. Ryan Corces, Jeffrey M. Granja, Rui Li, Utkrisht Rajkumar, Jordan Friedlein, Anindya Bagchi, Ansuman T. Satpathy, Robert Tjian, Stefan Mundlos, Vineet Bafna, Anton G. Henssen, Paul S. Mischel, Zhe Liu & Howard Y. Chang. ecDNA hubs drive cooperative intermolecular oncogene expression. ***Nature***, (2021). IF: 69.504

King L. Hung, Jens Luebeck, Siavash R. Dehkordi, Caterina I. Colón, Rui Li, Ivy Tsz-Lo Wong, Ceyda Coruh, Prashanthi Dharanipragada, Shirley H. Lomeli, Natasha E. Weiser, Gatien Moriceau, Xiao Zhang, Chris Bailey, Kathleen E. Houlahan, Wenting Yang, **Rocío Chamorro González**, Charles Swanton, Christina Curtis, Mariam Jamal-Hanjani, Anton G. Henssen, Julie A. Law, William J. Greenleaf, Roger S. Lo, Paul S. Mischel, Vineet Bafna & Howard Y. Chang. Targeted profiling of human extrachromosomal DNA by CRISPR-CATCH. ***Natur Genetics***, (2022). IF: 41.307

Heathcliff Dorado García, Fabian Pusch, Yi Bei, Jennifer von Stebut, Glorymar Ibáñez, Kristina Guillan, Koshi Imami, Dennis Gürgen, Jana Rolff, Konstantin Helmsauer,

Stephanie Meyer-Liesener, Natalie Timme, Victor Bardinet, **Rocío Chamorro González**, Ian C MacArthur, Celine Y Chen, Joachim Schulz, Antje M Wengner, Christian Furth, Birgit Lala, Angelika Eggert, Georg Seifert, Patrick Hundsoerfer, Marieluise Kirchner, Philipp Mertins, Matthias Selbach, Andrej Lissat, Frank Dubois, David Horst, Johannes H Schulte, Simone Spuler, Daoqi You, Filemon Dela Cruz, Andrew L Kung, Kerstin Haase, Michela DiVirgilio, Monika Scheer, Michael V Ortiz, Anton G Henssen. Therapeutic targeting of ATR in alveolar rhabdomyosarcoma. ***Nat Commun***, (2022). IF: 17.69

Alicia Isabell Tüns, Till Hartmann, Simon Magin, **Rocío Chamorro González**, Anton George Henssen, Sven Rahmann, Alexander Schramm and Johannes Köster. Detection and Validation of Circular DNA Fragments Using Nanopore Sequencing. ***Front Genet,*** (2022). IF: 4.772

Konstantin Helmsauer, Maria E. Valieva, Salaheddine Ali, **Rocío Chamorro González**, Robert Schöpflin, Claudia Röefzaad, Yi Bei, Heathcliff Dorado Garcia, Elias RodriguezFos, Montserrat Puiggròs, Katharina Kasack, Kerstin Haase, Csilla Keskeny, Celine Y. Chen, Luis P. Kuschel, Philipp Euskirchen, Verena Heinrich, Michael I. Robson, Carolina Rosswog, Joern Toedling, Annabell Szymansky, Falk Hertwig, Matthias Fischer, David Torrents, Angelika Eggert, Johannes H. Schulte, Stefan Mundlos, Anton G. Henssen & Richard P. Koche. Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma. ***Nat Commun***, (2020). IF: 14.92

Christina Stangl, Sam de Blank, Ivo Renkens, Liset Westera, Tamara Verbeek, Jose Espejo Valle-Inclan, **Rocio Chamorro González**, Anton G. Henssen, Markus J. van Roosmalen, Ronald W. Stam, Emile E. Voest, Wigard P. Kloosterman, Gijs van Haaften & Glen R. Monroe. Partner independent fusion gene detection by multiplexed CRISPR-Cas9 enrichment and long read nanopore sequencing. ***Nat Commun***, (2020). IF: 14.92

Richard P. Koche*, Elias Rodriguez-Fos*, Konstantin Helmsauer*, Martin Burkert, Ian C. MacArthur, Jesper Maag, **Rocio Chamorro**, Natalia Munoz-Perez, Montserrat Puiggròs, Heathcliff Dorado Garcia, Yi Bei, Claudia Röefzaad, Victor Bardinet, Annabell Szymansky, Annika Winkler, Theresa Thole, Natalie Timme, Katharina Kasack, Steffen Fuchs, Filippos Klironomos, Nina Thiessen, Eric Blanc, Karin Schmelz, Annette Künkele, Patrick Hundsdörfer, Carolina Rosswog, Jessica Theissen, Dieter Beule, Hedwig

Deubzer, Sascha Sauer, Joern Toedling, Matthias Fischer, Falk Hertwig, Roland F. Schwarz, Angelika Eggert, David Torrents*, Johannes H. Schulte* & Anton G. Henssen*. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. ***Nature Genetics***, 2019. 52(1): p. 29-34. IF: 27.61

11. Acknowledgements

First and foremost, I would like to thank my parents for their unconditional love and support throughout my entire life and studies. Thank you for believing in me and blindly supporting me in every step of this journey. ¡Gracias! Os quiero mucho.

Special thanks go to my supervisor, Anton, for his continuous support, encouragement, motivation, and accessibility. Thank you for your guidance and for giving me this great opportunity. I am also immensely grateful to all the members of the Henssen lab for their invaluable support and for cultivating such a fantastic and inclusive atmosphere in the lab. Thanks also to Richard Koche for his incredible support and guidance.

I would also like to extend my heartfelt gratitude to all our collaborators in Berlin and abroad, especially those at the Department of Pediatric Oncology in Charité. A special note of thanks goes to everyone at the Genomics Technology Platform in BIMSB for their warm welcome and assistance, particularly to Thomas Conrad.

Last but never least. I would like to thank my friends. To those who have been by my side for many years, and even though they are far away, still feel very close. And especially to those I met in this city, who made Berlin my home and helped me survive its long, cold winters. You lifted me up in the toughest moments and encouraged me during the most difficult times.