UC San Diego
UC San Diego Electronic Theses and Dissertations

Title

Inter- and intratumoral heterogeneity of circular extrachromosomal DNA in medulloblastoma and other pediatric cancers

Permalink

Author

Chapman, Owen

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Inter- and intratumoral heterogeneity of circular extrachromosomal DNA in medulloblastoma and other pediatric cancers

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioinformatics and Systems Biology (with a specialization in Biomedical Informatics)

by

Owen Chapman

Committee in charge:

    Professor Jill Mesirov, Chair
    Professor Bing Ren, Co-Chair
    Professor Vineet Bafna
    Professor Lukas Chavez
    Professor Paul Mischel

2023

The dissertation of Owen Chapman is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To the anonymous pediatric patients whose tumor tissue I have studied for the

past five years. May your generosity one day save someone's life.

EPIGRAPH

I think that if ever a mortal heard the word of God, it would be in a garden at the cool of the day.

—F. Frankfort Moore

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AA AmpliconArchitect. 3, 4, 37

ACC adrenocortical carcinoma. 73

AFT accelerated failure time. 13, 22, 23

CBTN Children's Brain Tumor Network. 7

CMT1A Charcot-Marie-Tooth disease type 1A. 81

CNA copy-number alteration. 3

CPC choroid plexus carcinoma. 72

CPG craniopharyngioma. 73

CPH Cox Proportional Hazards. 22

ecDNA extrachromosomal DNA. xii, xviii, 2–4, 6, 12, 15, 26, 44

ecDNA+ ecDNA-containing. 2, 4, 8, 13, 26, 73

EPN ependymoma. 73, 79

ETMR embryonal tumor with multilayered rosettes. viii, 72, 77

FFPE formalin fixed paraffin embedded. 4, 37

FISH fluorescence *in situ* hybridization. 4, 29, 43

GG ganglioglioma. 73

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Lin, Yingxi; Dutta, Aditi; Lo, Yan Yuen; Juarez, Edwin; Robinson, James T.; Malicki, Denise M.; Coufal, Nicole G.; Levy, Michael; Hobbs, Charlotte; Scheuermann, Richard H.; Crawford, John R.; Pomeroy, Scott L.; Rich, Jeremy; Zhang, Xinlian; Chang, Howard Y.; Dixon, Jesse R.; Bagchi, Anindya; Deshpande, Aniruddha J.; Carter, Hannah; Fraenkel, Ernest; Mischel, Paul S.; Wechsler-Reya, Robert J.; Bafna, Vineet; Mesirov, Jill P.; Chavez, Lukas. The dissertation author was the primary investigator and author of the manuscript.

Chapter 2, in part, is reproduced from Chapman *et al.* above.

Chapter 3, in full, is reproduced from Chapman *et al.* above.

Chapter 4, in full, is adapted from the following manuscript currently being prepared for publication: "Extrachromosomal DNA promotes oncogene amplification across multiple pediatric cancer types." Sridhar, Sunita; Chapman, Owen S; Dutta, Aditi; Wang, Shanqing; Luebeck, Jens; Bafna, Vineet; Mesirov, Jill P; Chavez, Lukas. The dissertation author was the second investigator and author of the manuscript.

2017	B. A. in Computer Science *cum laude*, Pomona College, Claremont
2018-2020	Graduate Teaching Assistant, University of California San Diego
2020-2023	Graduate Research Assistant, University of California San Diego
2023	Ph. D. in Bioinformatics and Systems Biology, University of California San Diego

PUBLICATIONS

S. Banerjee, T. Wei, J. Wang, J. J. Lee, H. L. Gutierrez, O. Chapman, S. E. Wiley, J. E. Mayfield, V. Tandon, E. F. Juarez, L. Chavez, R. Liang, R. L. Sah, C. Costello, J. P. Mesirov, L. de la Vega, K. L. Cooper, J. E. Dixon, J. Xiao, and X. Lei. Inhibition of dual-specificity tyrosine phosphorylation-regulated kinase 2 perturbs 26s proteasome-addicted neoplastic progression. *Proceedings of the National Academy of Sciences*, 116(49):24881–24891, 2019.

K. Okonechnikov, A. Camgöz, O. Chapman, S. Wani, D. E. Park, J.-M. Hübner, A. Chakraborty, M. Pagadala, R. Bump, S. Chandran, K. Kraft, R. Acuna-Hidalgo, D. Reid, K. Sikkink, M. Mauermann, E. F. Juarez, A. Jenseit, J. T. Robinson, K. W. Pajtler, T. Milde, N. Jäger, P. Fiesel, L. Morgan, S. Sridhar, N. G. Coufal, M. Levy, D. Malicki, C. Hobbs, S. Kingsmore, S. Nahas, M. Snuderl, J. Crawford, R. J. Wechsler-Reya, T. B. Davidson, J. Cotter, G. Michaiel, G. Fleischhack, S. Mundlos, A. Schmitt, H. Carter, K. A. Michealraj, S. A. Kumar, M. D. Taylor, J. Rich, F. Buchholz, J. P. Mesirov, S. M. Pfister, F. Ay, J. R. Dixon, M. Kool, and L. Chavez. 3d genome mapping identifies subgroup-specific chromosome conformations and tumor-dependency genes in ependymoma. *Nature Communications*, 14(1):2300, 2023.

A. Tiwari, U. Paithane, K. Tashiro, O. Saulnier, K. Barbosa, A. Vishvanathan, Q. Trinh, T. Nakashima, T. Mohammad, J. T. Lange, S. Wu, M. B. Masihi, A. Soni, K. Vogel, J. Friedlein, B. Hall, A. Shukla, S. Saha, L. Denis, L. D. Hendrikse, B. James, O. Chapman, J. Luebeck, B. Konety, S. M. Dehm, Y. Marahrens, L. D. Stein, L. Chavez, V. Bafna, P. S. Mischel, H. Suzuki, I. Hassan, A. J. Deshpande, R. J. Wechsler-Reya, M. D. Taylor, and A. Bagchi. The lncrna pvt1 encodes peptides with oncogenic and tumor suppressive activity, 2022.

O. S. Chapman, J. Luebeck, S. Wani, A. Tiwari, M. Pagadala, S. Wang, J. D. Larson, J. T. Lange, I. T.-L. Wong, S. R. Dehkordi, S. Chandran, M. Adam, Y. Lin, E. Juarez, J. T. Robinson, S. Sridhar, D. M. Malicki, N. Coufal, M. Levy, J. R. Crawford, S. L. Pomeroy, J. Rich, R. H. Scheuermann, H. Carter, J. Dixon, P. S. Mischel, E. Fraenkel, R. J. Wechsler-Reya, V. Bafna, J. P. Mesirov, and L. Chavez. The landscape of extrachromosomal circular dna in medulloblastoma. *bioRxiv*, 2021.

S. Sridhar, O. Chapman, A. Dutta, S. Wang, J. Luebeck, V. Bafna, and L. Chavez. Extrachromosomal dna promotes oncogene amplification across multiple pediatric cancer types, 2023.

ABSTRACT OF THE DISSERTATION

Inter- and intratumoral heterogeneity of circular extrachromosomal DNA in medulloblastoma and other pediatric cancers

by

Owen Chapman

Doctor of Philosophy in Bioinformatics and Systems Biology (with a specialization in Biomedical Informatics)

University of California San Diego, 2023

Professor Jill Mesirov, Chair
Professor Bing Ren, Co-Chair

ecDNA is an important driver of aggressive adult cancers, but its role in pediatric tumors is not yet well-understood. To survey the genomic landscape of ecDNA amplifications across pediatric cancers, we applied computational methods to detect ecDNA in the genomes of a cohort of 1670 pediatric tumors from 1481 patients. ecDNA was detected in at least 10% of pediatric ETMR, osteosarcoma, rhabdomyosarcoma, high-grade glioma, retinoblastoma, and medulloblastoma (MB) tumors. To assess the clinical importance of ecDNA in MB, the most common malignant pediatric brain tumor, we applied the same methods to a cohort of 468

MB patients. ecDNA was detected in 18% of MB tumors and carried a threefold greater risk of mortality within 5 years. Affected genomic loci harbor up to hundredfold amplification of oncogenes including *MYC*, *MYCN*, *TERT*, and other novel putative oncogenes. To elucidate the functional regulatory landscapes of ecDNAs in MB, we generated transcriptional (RNA-seq), accessible chromatin (ATAC-seq), and chromatin interaction (Hi-C) profiles of 6 MB tumor samples. In each case, we identified regulatory interactions that cross fusion breakpoints on the ecDNA, representing potential "enhancer rewiring" events which may contribute to transcriptional activation of co-amplified oncogenes. To test this hypothesis, we conducted an *in vitro* CRISPRi screen targeting regulatory regions on the ecDNA of a MB cell line and identified distal enhancers which promote proliferation. Using single-cell sequencing, we have also developed methods to explore intratumoral heterogeneity of ecDNA in a p53-mutant SHH MB patient tumor and its corresponding PDX model. In summary, our study analyzes the frequency, diversity, and functional relevance of ecDNA across MB subgroups and provides strong justification for continued mechanistic studies of ecDNA in MB with the potential to uncover new therapeutic approaches.

Introduction

Cancer is a heterogeneous disease of common phenotypes but diverse genetic roots. To become malignant, a tumor must overcome several molecular hurdles known as the "Hallmarks of Cancer", including sustaining proliferative signalling, evading immune destruction, resisting cell death, and evading growth suppressors, among others [39]. The possible genetic mechanisms for achieving these transformations, however, are extremely diverse. Accordingly, the standard of care for cancer treatment varies by tissue, cell of origin, and genetic, epigenetic, and molecular features of the tumor. The practice of personalized oncology, or the tailoring of therapy to the unique combination of molecular features present in a tumor, has become increasingly feasible in recent decades with the advent of genomic data science. Specific molecular targeting of tumor dependency genes has led to improved prognoses of many molecular subtypes of cancer [147]; however, some tumors evade molecular therapy and prognoses for these patients remain poor.

Cancers harboring circular extrachromosomal DNA (ecDNA) comprise one such class of tumors for which current therapies often fail. ecDNA-containing (ecDNA+) tumors are associated with poorer overall survival across many histological tumor types [55], including gliomas [96], neuroblastoma [58], leukemias [131], lung cancer [103], and breast cancer [150]. *In vitro* experiments have shown that ecDNA amplification mediates drug resistance in colon cancer [92] and glioblastoma [95] cell lines. Recent continued work on the same glioblastoma cell line showed that ecDNA copy number is modulated in response to administration or subsequent withdrawal of targeted therapy, and a similar pattern may be observed in sequential biopsies of ecDNA+ treatment-naïve and post-treatment patient tumors treated with targeted therapies [63]. These results constitute strong evidence that ecDNA contributes to the evolution of therapy resistance.

The mechanistic link between ecDNA and poor prognosis is most likely explained by sequence dynamics of ecDNA molecules under selective pressure. ecDNAs are defined as circular, acentric (ie., lacking a centromere) chromatin bodies hundreds of kilobases to tens of megabases in length [148]. Tumors containing ecDNA commonly have been observed to accumulate ten-

or hundredfold copy-number alteration (CNA) per cell of the circularized sequence, driving high oncogenic expression of one or more genes amplified on the ecDNA [141]. Because ecDNAs lack centromeres, they segregate randomly rather than equally during cell division [63]. This non-Mendelian pattern of inheritance allows a tumor to acquire intratumoral copy-number heterogeneity with each cell division, which in turn enables rapid tumor evolution in response to selective pressure [148, 37, 63].

There is accumulating evidence that ecDNA is subject to unique patterns of epigenetic regulation as well. Experiments in glioblastoma and other cell line models have shown that ecDNAs have highly accessible chromatin [160] and bring functional regulatory elements, or enhancers, in contact with co-amplified oncogene promoters [93, 42]. Furthermore, ecDNA-amplified enhancers have been observed to bind in *trans* to promoters on entirely different molecules, namely chromosomal promoters [166] and other distinct lineages of ecDNA [45]. Thus, co-amplified regulatory elements on ecDNA appear to play previously underappreciated oncogenic roles.

New and emerging technologies have powered these recent advances in ecDNA biology. The growth of large cancer genomics cloud data resources such as the Pan Cancer Analysis of Whole Genomes (PCAWG) [13], St Jude Cloud [86], and the Kids First Data Resource Commons (KF-DRC) [41] have made thousands of cancer genomes available for public research, alongside high-quality clinical metadata. Concurrently, the algorithmic softwares AmpliconArchitect (AA) [22] and AmpliconClassifier [55] have enabled robust and reproducible detection of ecDNA from whole genome sequencing. Finally, new technologies facilitate investigation of ecDNA sequence and gene transcription: CRISPR-CATCH (ecDNA length and circularity) [44], ATAC-seq (accessible chromatin) [12], Hi-C (chromatin interactions) [110], single-cell multiome sequencing (single cell resolution gene regulation) [1], and others. These technologies have enabled investigation of the roles of ecDNA in oncogenesis at ever-greater resolution.

While the frequency of ecDNA is known in most adult cancer types and some pediatric

3

tumor types, only isolated case reports describe ecDNA in rare pediatric cancers. Here, The accumulation of whole genome sequencing (WGS) data in cloud genomic data platforms now offers the opportunity to examine ecDNA in pediatric patient populations at unprecedented resolution. In Chapter 1, I apply AA to a large retrospective cohorts of medulloblastomas to identify and characterize ecDNA across pediatric tumor types. I perform survival analysis with respect to ecDNA and other clinical and genomic metadata to establish association between ecDNA and patient outcomes.

In Chapter 2, I present new methods to characterize intratumoral heterogeneity of ecDNA in medulloblastoma (MB) tumors. Using fluorescence *in situ* hybridization (FISH), we label genomic copies of a marker gene on ecDNA in archival formalin fixed paraffin embedded (FFPE) tissue of patient tumors, then apply an automated computational pipeline to estimate ecDNA copy number across thousands of cells. This analysis establishes that the ecDNA has greater average copy number and greater variance in copy number per cell than other focal amplifications, consistent with previous reports of copy number heterogeneity in cell lines[63]. In addition, we perform high-throughput sequencing of single cells to characterize the transcriptomes of cell types present in ecDNA+ tumors. We show that this approach enables detection and characterization of a distinct ecDNA+ cell population within a heterogeneous tumor. We anticipate that further application of these approaches may illuminate the evolutionary dynamics of ecDNA+ tumors under therapeutic pressure.

In Chapter 3, we map the transcriptomes, accessible chromatin, and chromatin interactions of ecDNA sequences in patient and model tumors. In doing so, we reveal ectopic chroamatin interactions linking accessible transcriptional enhancer elements to highly-expressed gene loci, both co-amplified on the same ecDNA sequences. The frequency and diversity of these putative "enhancer rewiring" events suggests that dysregulation of gene transcription likely contributes to oncogenic activation of ecDNA-amplified genes.

Chapter 1

CLINICAL IMPORT OF CIRCULAR ECDNA IN MEDULLOBLASTOMA AND OTHER PEDIATRIC CANCERS

1.1 Abstract

Motivation — Medulloblastoma (MB) is the most common malignant pediatric brain tumor. The current standard of care for MB consists of maximal safe resection followed by chemotherapy and/or radiation, and carries substantial risk of co-morbidity and secondary neoplasms. So far, targeted molecular therapies have failed to reliably achieve sustained remission. Notably, the genomic features of MB subgroups with the worst prognoses, namely *TP53* inactivation and *MYC* family amplification, are also associated with ecDNA in other tumor types. We therefore asked whether extrachromosomal DNA (ecDNA) may be a prognostic biomarker of patient survival in MB, and investigated the genomes of ecDNA sequences found in MB patient tumors to elucidate their possible oncogenic roles.

Results — In MB, the presence of ecDNA in a tumor sample correlates with other known molecular prognostic indicators, namely *TP53* mutation and *MYC* gene family amplification, and is itself strongly associated with poor progression-free survival. We propose a causal model whereby *TP53* inactivation promotes genome instability, which may affect response to therapy by the formation of ecDNA or by the formation of other structural variants.

Availability and implementation — This chapter summarizes work from the following publication: [14] Code is available at https://github.com/auberginekenobi/medullo-ecdna.

1.2 Introduction

Circular extrachromosomal DNA (ecDNA), defined as circular, acentric megabase-scale genomic DNA found exclusively in cancer cells [148, 3], has been associated with high-copy oncogenic amplification [141], intratumoral heterogeneity [63, 161], evolution of targeted therapy resistance [19, 95], and poor patient outcomes across various tumor types [55, 58]. Although the frequency, sequence content, and prognostic value of ecDNA is well-described in various adult solid tumor types [55] and pediatric neuroblastoma [58], the clinical presentation of ecDNA in

MB has not yet been described. Therefore, there is compelling need to survey the frequency, genomic sequence, and prognostic utility of ecDNA across the molecular subgroups of MB, to determine how extrachromosomal amplification may affect the response of a tumor to the current standards of care. Here we computationally screen for ecDNA across a large cohort of whole genome sequencing (WGS) of pediatric solid tumors from the St. Jude Cloud [86], International Cancer Genome Consortium (ICGC), and Children's Brain Tumor Network (CBTN) [71], and identify key clinical and genomic features of ecDNA in MB.

Medulloblastomas were represented among the first patient case reports describing ecDNA [17, 77], and several patient-derived cell line models of MB are known to contain ecDNA [93]. Few effective targeted molecular treatments exist for MB, and the current standard of care carries substantial risk of developmental disorder, neurological damage, and secondary metastasis [121]. There are four major molecular subgroups of MB: WNT, SHH, Group 3 (G3) and Group 4 (G4) [49]. Prognosis is especially poor for a subset of aggressive MYC-activated Group 3 tumors, and for p53-mutant SHH subgroup tumors [98, 108]. Both *MYC* amplification and p53 mutation have been co-reported alongside ecDNA in MB [119, 111], but the strength of any possible association has not yet been quantified in a patient population. Although the genomic landscape of medulloblastoma subgroups has been largely characterized [98], it remains unknown how frequently ecDNA occurs in MB patients, and whether the presence of ecDNA may affect prognosis or correlate with other molecular features.

1.3   Results

1.3.1   ecDNA amplifies known and putative medulloblastoma oncogenes.

To examine the landscape of ecDNA in medulloblastoma, we accessed WGS data available in three cancer cloud genomics platforms: St Jude Cloud [86], the Children's Brain Tumor Network (CBTN) [71], and the International Cancer Genome Consortium (ICGC) [13]. In

addition, we included 43 tumors from a previous proteomic analysis [2] and 8 tumors from MB patients diagnosed at the Rady Children's Hospital, San Diego. In total, we analyzed a retrospective cohort of WGS data of 481 tumor biopsies from 468 different patients. Using DNA fingerprint analysis, we ensured that the combined cohort WGS data contained no duplicate specimens (see section 1.5). Clinical metadata were available for most patients and included age at diagnosis, sex, MB molecular subgroup, and survival (Suppl. Tbl. 1 from [14]). To detect ecDNA, we applied AmpliconArchitect (AA), an algorithm that detects ecDNA using paired-end WGS data [22]. 102 putative ecDNA sequences were detected in tumor samples from 82 of 468 (18%) patients. By molecular subgroup, ecDNA+ patients were distributed as follows: WNT 0/22, SHH 30/112 (27%), Group 3 19/107 (18%), and Group 4 26/181 (14%) (Fig. 1.1). SHH subgroup tumors were significantly more likely to contain ecDNA than tumors from the other MB subgroups ($\chi^2 = 7.66$, $p = 0.006$). Among the ecDNA-amplified genes occurring in two or more samples in this cohort were known or suspected MB oncogenes *MYC*, *MYCN*, *MYCL1*, *TERT*, *GLI2*, *CCND2* (Cyclin D2) [28], *PPM1D* (WIP1) [156], and *ACVR2B* [91]; members of signaling pathways commonly dysregulated in cancer, such as DNA repair (*RAD51AP1* and *RAD21*); and p53 pathway inhibitors (*PPM1D* [76] and *CDK6* [9]) (Fig. 1.2). Of *MYC* gene family amplifications, 19/23 *MYCN*, 11/18 *MYC*, and 3/3 *MYCL1* were on ecDNA, as were all amplifications of *CCND2*, *GLI2* and *TERT*. These results reveal that known MB oncogenes are frequently amplified as ecDNA in MB patients, potentially providing future therapeutic opportunities to target ecDNA-specific properties such as their generation [126], inheritance [63], micronucleation [125], or aggregation [146, 45].

1.3.2   ecDNA predicts poor prognosis in medulloblastoma.

To evaluate ecDNA as a potential prognostic marker in MB, we performed survival analyses across patients for whom clinical metadata were available. Patients with ecDNA-containing (ecDNA+) tumors had significantly worse overall and progression-free 5-year survival
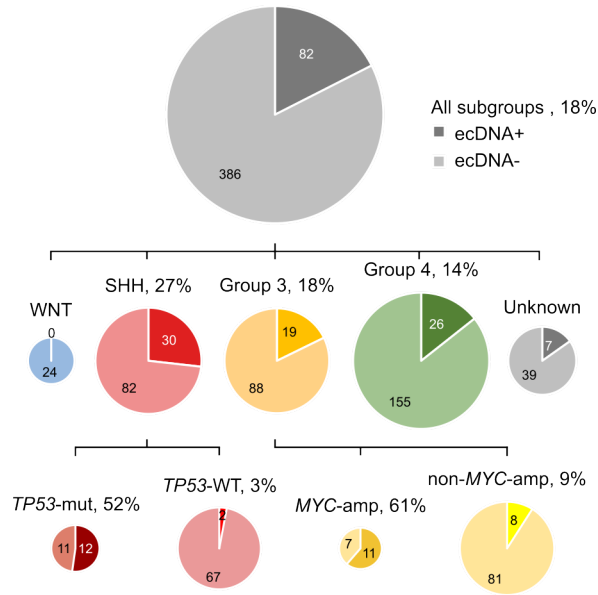
Figure 1.1: **Presence of ecDNA by molecular subgroup across 468 MB patient tumors.** *TP53*-mut/WT: *TP53*-mutant/wild-type. *MYC*-amp: *MYC* copy number ≥ 5 estimated from WGS.

(OS, PFS) compared to patients with ecDNA-negative tumors (ecDNA-) (log-rank test, $p < 0.005$; Fig. 1.3). Stratified by molecular subgroup, ecDNA+ patients had worse overall survival in the SHH, Group 3 and Group 4 MB subgroups ($p < 0.05$ for all subgroups; Fig. 1.4). Survival of WNT subgroup patients was not analyzed because no WNT tumors in our patient cohort were ecDNA+. To determine whether ecDNA+ patients had worse outcomes than patients with tumors harboring other types of focal somatic copy number amplification (fSCNA), we stratified patients by the topology of the amplification(s) present in the tumor genomes, using previously published methods [55]. As expected, ecDNA+ patients had the poorest clinical outcomes, significantly worse compared to patients without fSCNAs or with linear amplifications ($p < 0.005$, Fig. 1.5).

To further estimate the prognostic value of ecDNA, we conducted Cox proportional hazards models controlling for sex, age, and molecular subgroup. Patients with ecDNA+ tumors had greater estimated risk for progression (hazard ratio 2.36, $p < 0.005$) and mortality ($HR = 2.99$, $p < 0.005$) compared to patients with ecDNA- tumors (Fig. 1.6, Supplementary Table 2 of [14]).
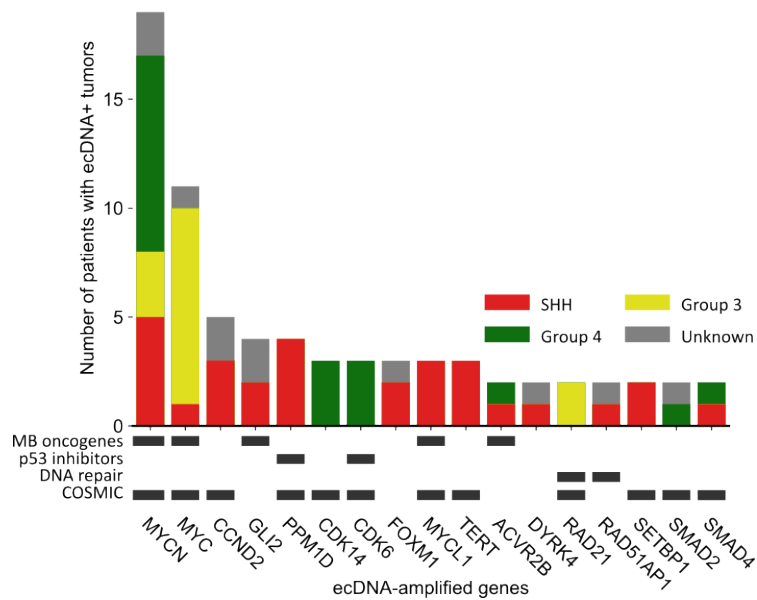
Figure 1.2: **Recurrently amplified genes on ecDNA in MB.** A subset of recurrently ($n \geq 2$) amplified genes on ecDNA in this patient cohort. p53 inhibitors: negative regulators of p53 pathway activity; COSMIC: genes listed as tier 1 or 2 of the COSMIC Cancer Gene Census [127].
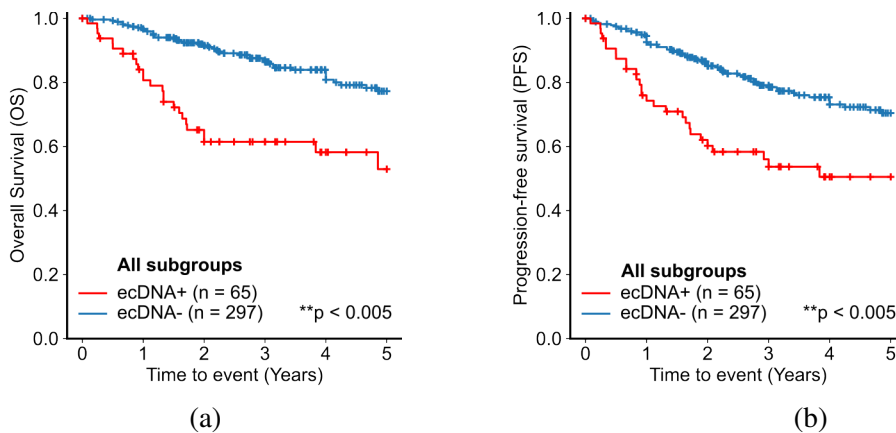
Figure 1.3: **Survival of ecDNA+ and ecDNA- MB patients.** (**a**) 5-year overall survival and (**b**) progression-free survival of ecDNA+ and ecDNA- MB patients.

1.3.3 *TP53* alterations are associated with ecDNA in SHH MB tumors.

The tumor suppressor protein p53 (*TP53*) is involved in DNA damage sensing, cell cycle arrest and apoptosis, and is frequently affected by somatic mutations and pathogenic germline
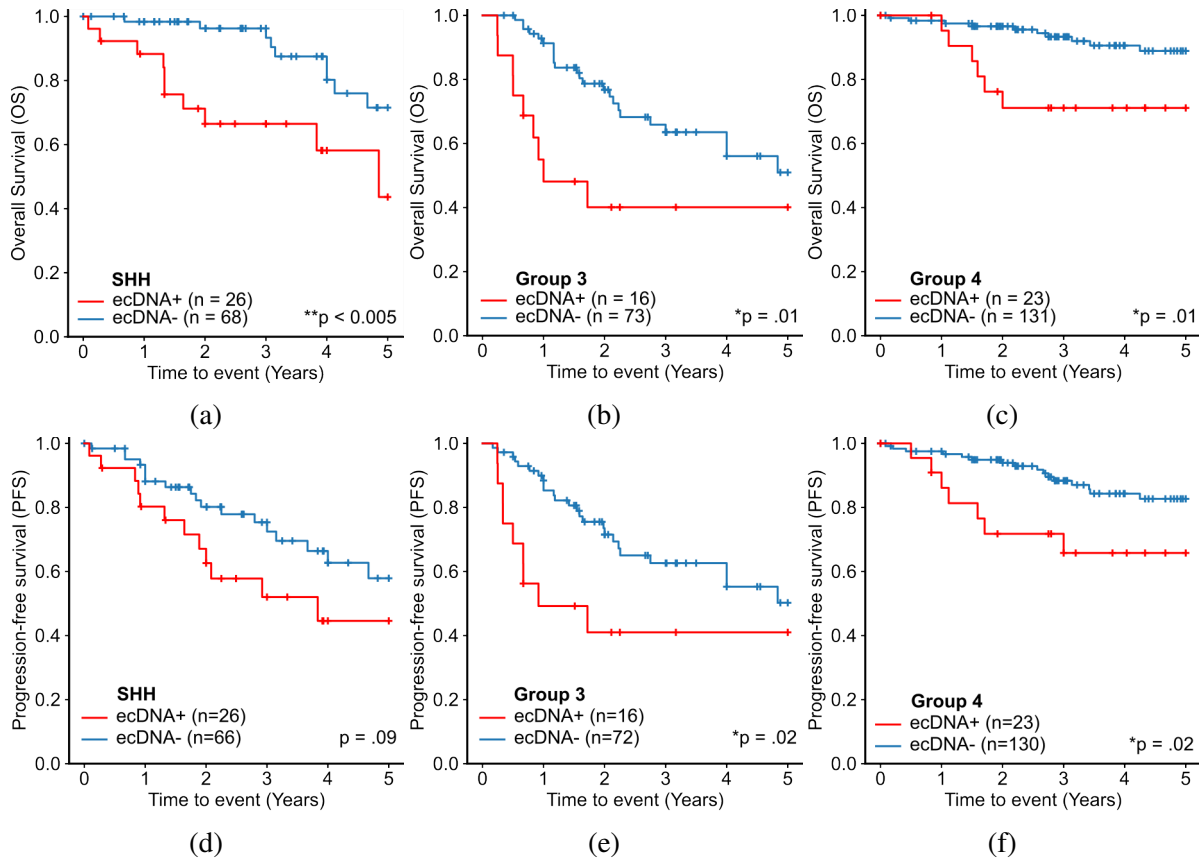
Figure 1.4: **Survival of ecDNA+ and ecDNA- MB patients, stratified by molecular subgroup.** Survival curves indicating (**a-c**) 5-year overall survival and (**d-f**) progression-free survival of ecDNA+ and ecDNA- MB patients. **a-b**: SHH. **c-d**: Group 3. **e-f**: Group 4. WNT subgroup was not analyzed because ecDNA was not detected in WNT MB tumors.

variants in SHH MB [108, 66, 155]. Moreover, SHH-subgroup MBs with inactivating *TP53* mutations are known to be associated with chromothripsis [111], a process of catastrophic shattering of a chromosome that has been shown to precede ecDNA formation in cell line models [126, 142]. To test whether *TP53* mutations were associated with the presence of ecDNA in MB, we accessed the somatic and germline *TP53* mutation status available for 92 SHH MBs. We found significant enrichment of *TP53* alterations in ecDNA+ SHH MB tumors (12 of 23, 52%) compared to ecDNA- SHH MB tumors (2 of 69, 3%; $p = 1.3e - 7$, Fisher exact test; Table 1.1). We did not find a significant association between *TP53* alterations and ecDNA either in the other MB subgroups or across the entire MB cohort, suggesting that in MB, a possible functional
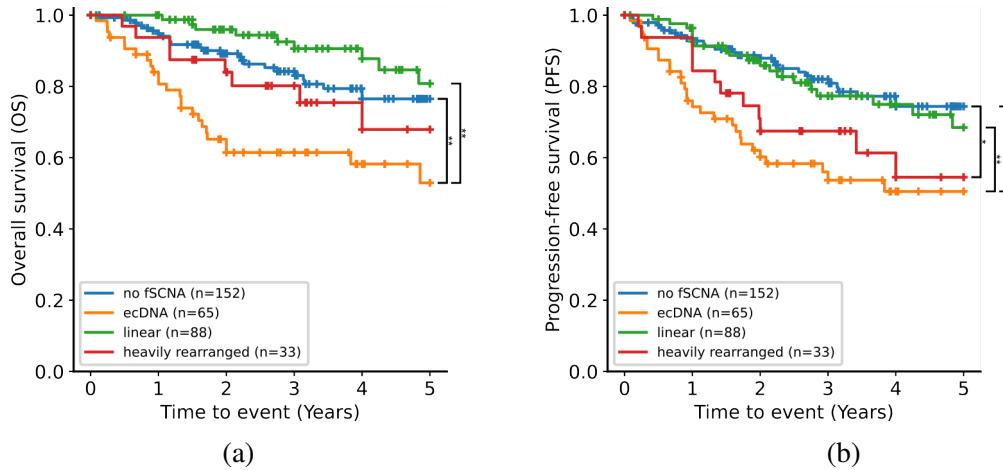
Figure 1.5: **Survival of ecDNA+ and ecDNA- MB patients, stratified by focal somatic amplification(s) present** in the tumor as classified by AmpliconClassifier [55]. ecDNA: tumor harbors a circular extrachromosomal amplification. Heavily rearranged: Tumor harbors an amplification classified as complex non-cyclic, but no cyclic ecDNA or BFB amplifications. Linear: Tumor harbors a linear amplification, but no highly-rearranged, BFB, or ecDNA amplifications. no fSCNA: no focal somatic copy number amplification detected.
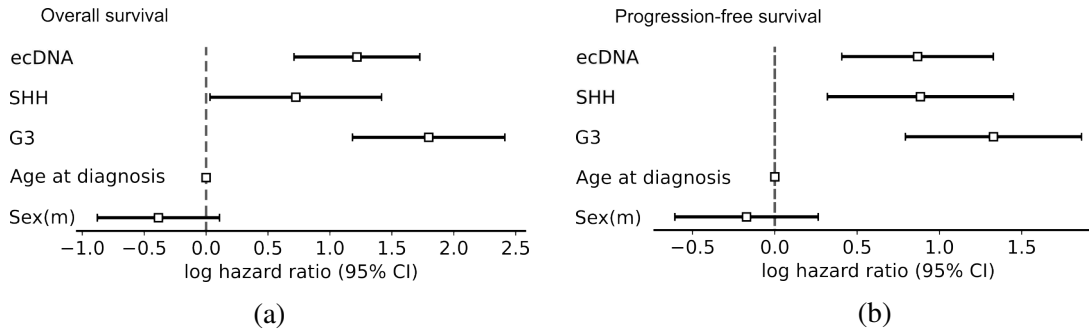
Figure 1.6: **Survival regressions on ecDNA, MB subgroup, age and sex.** Log hazard ratios for ecDNA status, MB subgroup, age and sex estimated by Cox proportional hazards regression on (**a**) overall survival and (**b**) progression-free survival. Model was fitted using maximum likelihood estimation (MLE). Bars indicate 95% confidence intervals. Estimated hazard ratios for ecDNA are $HR_{OS} = 2.99$, $p < 0.005$ and $HR_{PFS} = 2.36$, $p < 0.005$.

relationship between *TP53* alterations and ecDNA is restricted to the SHH subgroup.

We hypothesized that the established effect of p53 mutation on SHH MB patient survival [167] may be mediated, at least partially, by ecDNA (Fig. 1.7). To test this, we conducted mediation analyses using the Baron Kenny approach [8]. Briefly, variable *M* (ie, ecDNA status) is considered a mediator of the effect of independent variable *X* (ie, p53 mutation) on outcome *Y*

12

Table 1.1: *TP53* **mutation is associated with ecDNA in SHH subgroup MBs.** We acquired somatic and germline mutation status of 92 SHH MBs from our patient cohort. Nearly every *TP53*-mutant patient tumor contained ecDNA (12/14). The two without ecDNA showed evidence of breakage-fusion-bridge (BFB) amplification, a hypothesized precursor to ecDNA formation.

	ecDNA+	ecDNA-
$TP53_{mut}$	12	2
$TP53_{WT}$	11	67

(ie, patient survival) if (1) X significantly predicts Y; (2) X significantly predicts M; and (3) in a multiple regression of Y on X and M, M significantly predicts Y and the effect of X on Y is reduced relative to (1) (Fig. 1.7a) [48]. To test (1), we fitted a log-normal accelerated failure time (AFT) regression of progression-free survival (PFS) parameterized on age, sex, molecular subgroup and p53 mutation status, using MLE. Without information on ecDNA, this model estimates that p53 mutation significantly reduces survival time by 77% ($\mu = -1.46$, $p < 0.005$, Wald test; Fig. 1.7b). For (2), we have already shown that p53 mutation strongly predicts the presence of ecDNA given molecular subgroup. To test (3), we refitted our AFT model parameterized on age, sex, molecular subgroup, p53 mutation, and ecDNA status. This model estimates that ecDNA significantly reduces survival time by 56% ($\mu = -0.81$, $p < 0.005$; Fig. 1.7c). The same model estimates an insignificant reduction in survival time for p53 mutation, with a coefficient of lesser magnitude than that of (1) ($\mu = -0.94$, $p = 0.06$). We therefore concluded that the known prognostic effect of p53 mutation on survival may be partially explained by an effect of ecDNA on prognosis and by the frequent co-occurrence of ecDNA in p53-mutant tumors.

To evaluate whether there is a *TP53*-independent effect of ecDNA on survival, we performed Cox regression including *TP53* alteration as a covariate and controlling for collinearity. The effect of ecDNA on survival remains significant but diminished when we include p53 alteration as a covariate in our Cox models ($HR_{PFS} = 1.87, p = 0.01$; $HR_{OS} = 2.32, p < 0.005$; Fig. 1.8), indicating that there is an effect of ecDNA on survival that cannot be explained by p53 mutation alone. Such an effect may be explainable by a p53-independent mechanism of ecDNA
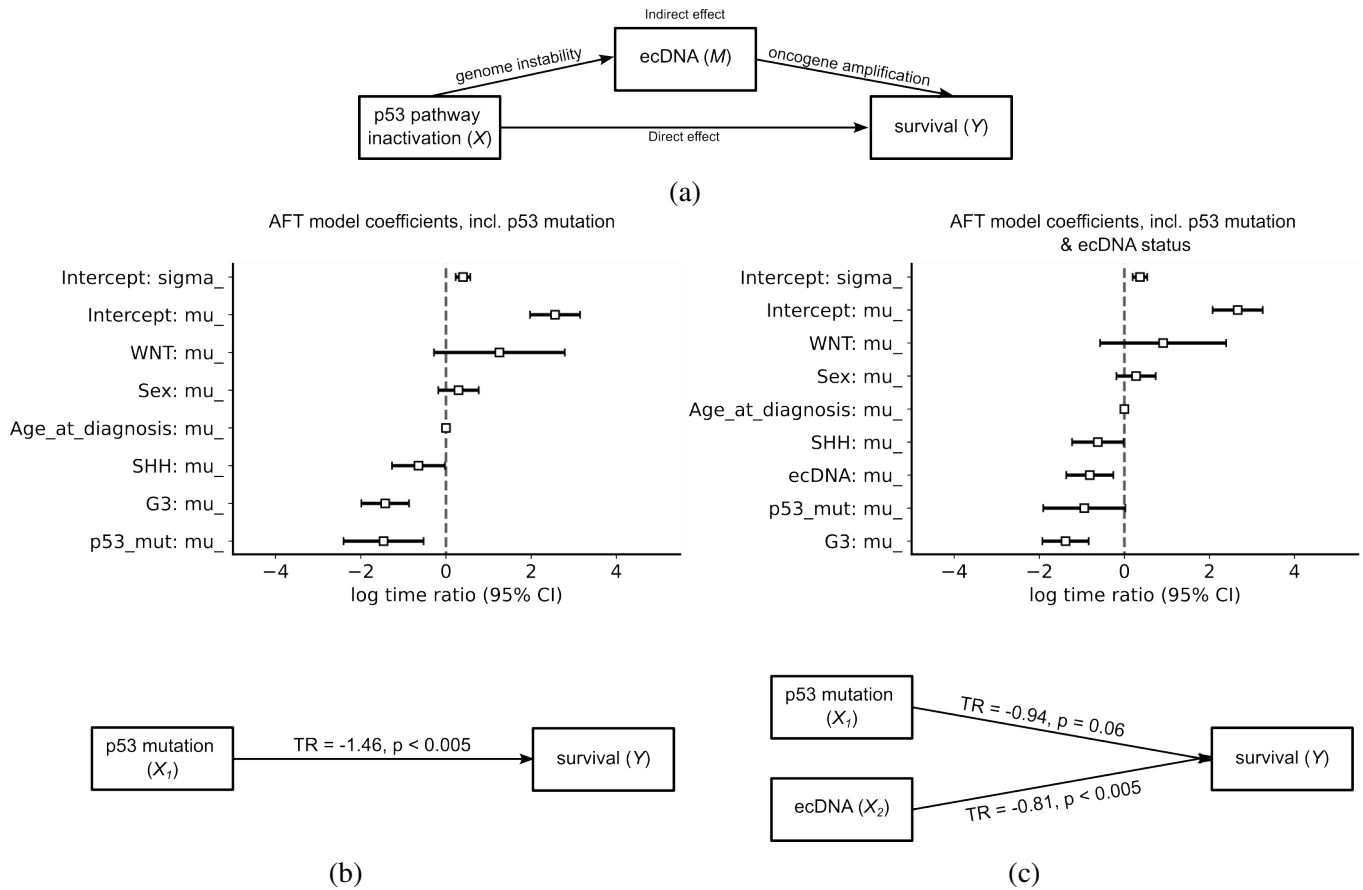
Figure 1.7: **Mediation analysis of p53 mutation, ecDNA, and progression-free survival.**
(**a**) Model diagram of the proposed mediation relationship. p53 inactivation may predispose
a tumor to form ecDNA, which then affects survival; p53 may also have ecDNA-independent
effects on survival, eg. by formation of other structural variants. (**b**) Top: estimated time ratios
for effects of age, sex, molecular subgroup and p53 mutation on PFS. Bottom: effect size and
p-value for p53 mutation are illustrated in the diagram below. (**c**) Top: estimated time ratios
for effects of age, sex, molecular subgroup, p53 mutation, and ecDNA on PFS. Bottom: effect
size and p-values for p53 mutation and ecDNA are illustrated in the diagram below. The effect
size of p53 on survival is decreased when controlling for ecDNA as a covariate, suggesting that
ecDNA partially mediates the effect of p53 on survival.

formation, or by inactivation of the p53 pathway by other means such as *CDKN2A* deletion or

PPM1D, *CDK6*, *MDM4* or *MDM2* amplification [85]. Although causality cannot be inferred

from these data alone, our survival analyses identify *TP53* alteration and ecDNA as clinically

relevant biomarkers for a subset of highly aggressive SHH MB tumors.

14
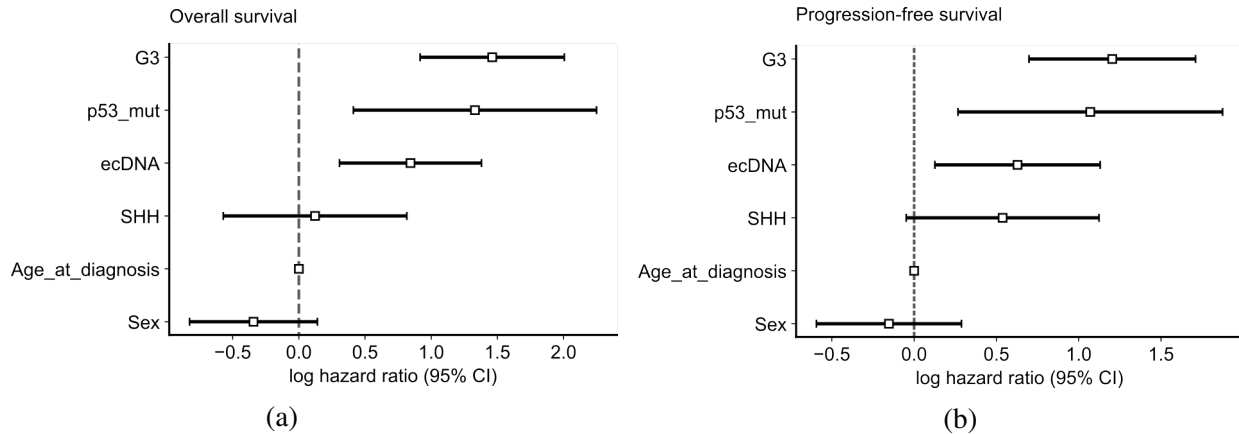
Figure 1.8: **Survival regressions on ecDNA status, MB subgroup, age, sex, and p53 mutation.** Log hazard ratios for ecDNA status, MB subgroup, age, sex, and p53 mutation estimated by Cox proportional hazards regression on (**a**) overall survival and (**b**) progression-free survival. Model was fitted using ridge regression to control for collinearity. Bars indicate 95% confidence intervals. Estimated hazard ratios for ecDNA are $HR_{OS} = 2.32, p < 0.005$; $HR_{PFS} = 1.87, p = 0.01$.

1.4 Discussion

As in other cancers [141, 55, 79], ecDNA frequently amplifies known oncogenic MB driver genes. Our results identify ecDNA as a frequent feature of *MYC*-amplified Group 3 and p53-mutant SHH tumors, which share exceptionally poor prognoses [119, 111] but few other recurrent driver mutations. Recent longitudinal analysis of Barrett's esophagus suggests that *TP53* alteration is an early event in ecDNA-driven malignant transformation [79]. However, the absence of detectable ecDNA in *TP53*-mutant WNT subgroup tumors, and the frequent occurrence of ecDNA in Group 3 tumors with wild type *TP53*, suggest that the mechanisms for the generation and selection of tumor cells with ecDNA are MB subgroup-specific and thus may be modulated by the different cellular contexts of MB progenitor cells.

A long-standing problem in the clinical management of medulloblastoma tumors has been the paucity of effective targeted molecular treatments for MB, especially in relapsed cases. For example, the SMO inhibitor vismodegib, one of few targeted drugs approved for SHH medulloblastoma, is ineffective against *TP53*-mutant, *MYCN*-amplified or *GLI2*-amplified tumors

[115], each a recurrent feature of ecDNA+ MB in our patient cohort. Through a retrospective analysis of one of the largest cohorts of WGS and clinical outcome data from MB patients assembled to date, we demonstrate that the occurrence of ecDNA significantly associates with poor outcome across the entire cohort and within individual MB subgroups. Our survival analysis estimates that relative to ecDNA- patients, ecDNA+ MB patients are more than twice as likely to relapse and three times as likely to die during the follow-up interval. The identification of MB patients with ecDNA is therefore crucial to pave the way for precision medicine approaches targeting ecDNA-specific properties.

1.5   Methods

1.5.1   Statistical methods

Statistical test, test statistic and p-values are indicated where appropriate in the main text. Categorical associations were established using the chi-squared test of independence if $N > 5$ for all categories, and Fisher exact test otherwise. For both tests, the python package scipy.stats v1.5.3 implementation was used [151]. Multiple hypothesis corrections were performed using the Benjamini-Hochberg correction [10] implemented in statsmodels v0.12.0 [123]. All statistical tests described herein were two-sided unless otherwise specified.

1.5.2   Medulloblastoma WGS

Paired-end whole genome sequencing (WGS) data were acquired for each of the sources described below. In total the WGS cohort comprised 468 patients, 4 cell lines and 26 PDX models. Unless otherwise specified, WGS was acquired for 1 tumor biosample per patient.

CBTN – Children's Brain Tumor Network (114 biosamples from 101 patients)

WGS of medulloblastoma tumor biopsies were identified using the Gabriella Miller KidsFirst Data Resource center portal (https://portal.kidsfirstdrc.org/) on 29 May 2020. Patients were originally sequenced as part of the Pediatric Brain Tumor Atlas (PBTA) (https://cbtn.org/pediatric-brain-tumor-atlas). WGS data were preprocessed and aligned using the KidsFirst harmonized WGS pipeline (hg38) and subsequently analyzed using Cavatica (https://cavatica.sbgenomics.com/), on the cloud genomic analysis platform for KidsFirst genomic data. Docker containers containing fingerprint and AmpliconArchitect software were installed on the Cavatica cloud genomics platform (see Methods: "EcDNA detection and classification").

St Jude (79 patients)

WGS of MB biopsies were identified using the St Jude Cloud Data Portal (https://platform.stjude.cloud/data/) on 11 February 2020. WGS was preprocessed and aligned by bwa-mem to hg38 as described in McLeod et al31. Docker containers of fingerprint and AmpliconArchitect software were installed on the DNANexus cloud genomics platform (see Methods: "EcDNA detection and classification").

ICGC - International Cancer Genome Consortium (237 patients)

WGS of MB tumor biopsies was identified using the ICGC Data Portal (https://dcc.icgc.org/) on 12 May 2020 and downloaded from the Collaboratory cloud genomics platform using the ICGC Score download client. These tumor genomes were previously sequenced, aligned and deposited with ICGC in other publications (GRCh37)[98, 13].

Archer *et al.* (43 patients)

All WGS data for these 43 patients were previously published elsewhere[98]. Sequencing data were aligned to human genome reference GRCh37 processed according to the best practice

17

pipelines at the Cancer Genome Analysis group at the Broad Institute[2].

Rady Children's Hospital (8 patients)

Tumor biopsies were collected and consented for research as part of the Rady Molecular Tumor Board (MTB). Paired end reads were acquired from Rady's Children hospital. Sequencing depth for all samples were at least 30x. Raw fastq reads were aligned to UCSC hg38 coordinates using BWA v0.7.17-r1188 [67]. Reads were sorted by samtools v0.1.19 [68], marked for duplicates with Picard Tools v2.12.3, and recalibrated with GATK v3.8-1-0 [83, 21, 145].

Cell lines and PDX models (4 cell lines, 26 PDXs)

Low-coverage WGS for 19 PDX models and 9 corresponding origin human tumors were obtained from a previous publication [117]. An additional 6 PDX biosamples (RCMB25, RCMB32, RCMB56, RCMB57, RCMB58 and RCMB69) were contributed by Wechsler-Reya lab using similar methods to establish PDX lines. Cell lines (D283, D341, D425 and D458) were contributed by Bagchi (SBP), Taylor (U. Toronto) and Wechsler-Reya (SBP) labs. WGS was performed at the UCSD IGM Genomics and Sequencing Core (hg38) for all samples except D283, for which no WGS was obtained. Preprocessing and alignment was performed using the same protocol as the RCH patient biopsies above.

1.5.3   ecDNA detection and classification from bulk WGS

To detect ecDNA, all samples in the WGS cohort were analyzed using AmpliconArchitect [22] v1.2 and AmpliconClassifier [55] v0.4.4. Briefly, the AmpliconArchitect algorithm was performed as follows. Copy number segmentation and estimation were performed using CNVkit v0.9.6 [134]. Segments with copy number ≥ 4 were extracted using AmpliconSuite-pipeline (April 2020 update) as "seed" regions. For each seed, AmpliconArchitect searches the region and nearby loci for discordant read pairs indicative of genomic structural rearrangement. Genomic

segments are defined based on boundaries formed by genomic breakpoint locations (identified by discordant reads) and by modulations in genomic copy number. A breakpoint graph of the amplicon region is constructed using the CN-aware segments and the genomic breakpoints, and cyclic paths are extracted from the graph. Amplicons are classified as ecDNA, breakage-fusion-bridge, complex, linear, or no focal amplification by the heuristic-based companion script, AmpliconClassifier. Biosamples with one or more classifications of "ecDNA" were considered potentially ecDNA+, and all others were considered ecDNA- (Supplementary Table 1c from [14]). We manually reviewed all potential ecDNA+ assembly graphs and reclassified those with inconclusive ecDNA status, which we defined as any of the following:

- Low-copy amplification (< 5) AND no copy number change at discordant read breakpoints

- Cycles consisting of the repetitive region at chr5:820000 (GRCh37).

Code is available at:

- AmpliconSuite-pipeline: https://github.com/jluebeck/AmpliconReconstructor

- AmpliconArchitect: https://github.com/jluebeck/AmpliconArchitect

- AmpliconClassifier: https://github.com/jluebeck/AmpliconClassifier

The ecDNA- status of the D283 cell line was not determined computationally by WGS, but by copy number analysis of DNA methylation, FISH (see Methods: "FISH").

1.5.4 Fingerprinting analysis

To uniquely identify WGS from each patient, we counted reference and alternate allele frequencies at 1000 variable non-pathogenic SNP locations in the human genome according to the 1000 Genomes project [27], and performed pairwise Pearson correlation between all WGS samples. Biospecimens originating from the same patient tumor (eg., primary/relapse

or origin/PDX pairs) were readily distinguishable by high correlation across these sites ($r >$ 0.80). We identified one case in our cohort in which 2 tumor biosamples had highly correlated fingerprints: MDT-AP-1217.bam and ICGC_MB127.bam. We arbitrarily removed ICGC_MB127 from the patient cohort.

1.5.5   Patient metadata, survival, and medulloblastoma subgroup annotation.

Where available, patient samples and models were assigned metadata annotations including age, sex, survival, and MB subgroup based on previously published annotations of the same tumor or model [2, 98, 155, 117, 47, 114, 100]. Sample metadata are also available in some cases from the respective cloud genomics data platform: https://dcc.icgc.org/ (ICGC), https://pedcbioportal.kidsfirstdrc.org/ and https://portal.kidsfirstdrc.org/ (CBTN), and https://pecan.stjude.cloud/ (St Jude). Where primary sources disagreed on a metadata value, that value was reassigned to NA. Patient tumors from the CBTN were assigned molecular subgroups based on consensus of 2 molecular classifiers using RSEM-normalized FPKM data: MM2S [30] and the D3b medulloblastoma classifier at the Children's Hospital of Philadelphia (https://github.com/d3b-center/medullo-classifier-package). To determine molecular subgroup of PDX samples, we generated or obtained from a previous publication [117] DNA methylation profiles (Illumina 450k or EPIC) and classified samples by molecular subgroup according to the DKFZ brain tumor methylation classifier (https://www.molecularneuropathology.org/mnp) [98].

1.5.6   *TP53* mutation annotation

Somatic mutations

Somatic *TP53* mutation information for the ICGC and CBTN cohorts was acquired from a previous publication [155] and from the ICGC and CBTN data portals (see Methods "Patient

metadata, survival, and subgroup annotation". Somatic *TP53* mutation information for the St. Jude cohort was extracted from the standard internal St. Jude variant calling pipeline [86]. Somatic mutations were only considered which were protein-coding and missense, nonsense, insertion or deletion, or affected a splice site junction.

Germline variants

Whole-genome sequencing (WGS) GVCF files were downloaded from the ICGC data portal (https://dcc.icgc.org/), the KidsFirst data portal (https://portal.kidsfirstdrc.org/dashboard) and DNAnexus for St. Jude Pediatric Cancer Genome Project (PCGP). xGVCF files were merged with GLnexus [73] and converted to PLINK format for analysis of ICGC, PCGP and KidsFirst genotypes. PCGP genotypes were converted to hg19 coordinates using liftover. Variants from *TP53* genomic locus (hg19 chr17 7571739-759080) were extracted and annotated with REVEL [46], CADD [113], ClinVar (accessed June 2021) and Variant Effect Predictor (VEP) [84]. REVEL GRCh38 scores were downloaded from https://sites.google.com/site/revelgenomics/ (date accessed: 03/05/2021). CADDv1.6 scores were downloaded from https://cadd.gs.washington.edu/info (date accessed: 03/23/2020). VEP scores were calculated with http://grch37.ensembl.org/index.html release 104 (date accessed: 04/27/21). Clinvar scores were obtained from https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh37/ (Date accessed: 04/27/21). VEP variants considered pathogenic included "frameshift" and "splice" variants. ClinVar annotations considered pathogenic included "frameshift", "stop", "splice", and "deletion", and whose clinical significance were "pathogenic" or "likely pathogenic". CADD "pathogenic" variants had a CADD score of at least 10 (Phred score). REVEL "pathogenic" variants had a REVEL score of at least 0.5. Only variants with minor allele frequency (MAF) less than 5%, according to the gnomAD r2.1.1 database, were analyzed [50].

1.5.7   Survival analyses

Kaplan-Meier (KM), Cox Proportional Hazards (CPH), and accelerated failure time (AFT) analyses were performed with Lifelines v0.26.5 [18]. For all analyses, the sample set was all patients annotated with the included covariates; no imputation was performed.

Kaplan-Meier

Sample sizes were $n = 362$ (65 ecDNA+; 297 ecDNA-). Differential survival was determined by log-rank test. For KM analyses by class of structural variant, samples were classified as previously described [55]. Our sample of patients with breakage-fusion-bridge (BFB) amplification but no ecDNA was too small to test ($n = 2$).

Cox Proportional Hazards on age, sex, molecular subgroup, & ecDNA

Sample was $n = 352$ observations. Model was fitted by maximum likelihood estimation (MLE).

Cox Proportional Hazards on age, sex, molecular subgroup, p53 mutation, & ecDNA

Sample was $n = 322$ observations. Collinearity, *i.e.* strong correlation between predictive variables in a regression model, can result in model instability and unreliable estimation of the collinear coefficients [75]. To address collinearity between ecDNA and p53 status in our model, we performed ridge estimation of model coefficients [149, 162], determining the ridge penalty parameter λ by grid search on 5-fold cross validation of model likelihood on the withheld set.

Accelerated failure time models and mediation analysis

Mediation analysis was performed using the Baron-Kenny framework [8], following recent best practices [64]. Because of the non-collapsibility of hazard ratios, the proportional hazards assumption and CPH model may not be suitable for mediation analysis in which we need

to compare the coefficients with and without the mediator. Thus, we fitted we fitted parametric log-normal AFT regression models as a reasonable alternative to Cox regression. Percent change values reported herein were calculated as

$$\text{PERCENTAGE CHANGE} = 100(e^{\hat{\beta}_k} - 1),$$

where $\hat{\beta}_k$ is the MLE regression coefficient for random variable k [20].

1.6   Acknowledgments

Chapter 2

INTRATUMORAL HETEROGENEITY OF ECDNA AT SINGLE-CELL RESOLUTION

2.1 Abstract

Motivation — Intratumoral copy number heterogeneity of extrachromosomal DNA (ecDNA) in single cells of a tumor is widely believed to contribute to tumor evolution and drug resistance. The capacity to count ecDNA in a large number of single cells would enable the observation of this heterogeneity in sequential tumor biopsies, revealing the population dynamics of ecDNA under therapeutic selection. Numerous algorithms have been developed to estimate genomic copy number from single-cell sequencing data. However, none are designed to quantify ecDNA copy number from single-cell profiles of accessible chromatin (scATAC-seq).

Results — To estimate the copy number of a known ecDNA sequence from scATAC-seq data, we compare sequencing read coverage at the ecDNA locus to that elsewhere in the genome. ecDNA copy number estimated by this approach strongly correlates to transcriptional measures of ecDNA activity. We apply our method to a dataset of multiome single cell ATAC + Gene Expression (scRNA+ATAC-seq) data from the primary human tumor (HT) and patient-derived xenograft (PDX) of a p53-mutant, SHH subgroup medulloblastoma (MB), and show that ecDNA-containing (ecDNA+) cells were strongly selected during PDX model establishment. We orthogonally validate our model by estimating ecDNA copy number in the same tumor using automated image analysis on FISH imaging of interphase nuclei stained for ecDNA.

Availability and implementation — A software implementation of our approach is available at https://github.com/auberginekenobi/ecdna-quant.

2.2 Introduction

Intratumoral copy number heterogeneity of extrachromosomal circular DNA (ecDNA) in single cells of a tumor is widely believed to contribute to tumor evolution, drug resistance, and poor patient outcomes [3]. In the previous chapter, we detected circular ecDNA by assembling cyclic sequences in short-read whole genome sequencing (WGS) of bulk patient tissue, enabling

us to survey a large retrospective cohort of patient data. However, assembly of bulk WGS cannot distinguish variation of ecDNA sequence, copy number, and and gene expression between tumor cells, limiting our capacity to measure intratumoral heterogeneity within a given tumor sample. Therefore, to understand the population dynamics of an ecDNA+ tumor, it is necessary to employ emerging methods to examine ecDNA at single-cell resolution.

To address this challenge, we have applied four orthogonal technologies to the patient-derived xenograft (PDX) MB model tumor RCMB56-pdx. We apply optical genome mapping, Bionano Genomics (OGM) and CRISPR-CATCH [44] to reconstruct high-confidence sequence assemblies of 2 ecDNA lineages originating from distinct portions of the reference human genome, and show that both sequences are conserved in the corresponding patient tumor RCMB56-ht. Next, we analyze FISH microscopy and single-cell sequencing data to estimate ecDNA copy number amplification at single-cell resolution, and show that a clonal expansion of ecDNA+ cells occurred during the engraftment interim between RCMB56-ht and RCMB56-pdx. We anticipate that similar approaches may shed light on clonal dynamics of ecDNA+ cell populations under other selective pressures such as conventional or targeted therapy.

2.3 Results

2.3.1 Linear and circular extrachromosomal amplifications co-exist in the RCMB56 patient tumor and PDX model.

We have previously described a SHH MB primary tumor with heterozygous somatic *TP53* mutation (RCMB56-ht) [117], which we have established as a PDX model (RCMB56-pdx). Analysis of WGS from RCMB56-ht predicted two distinct amplifications, a circular ecDNA of length 3.2Mbp comprising three regions of chr1 (amp1, Fig. 2.1a), and a complex, possibly chromothriptic, 4.5Mbp amplicon comprising 21 segments of chr7 and chr17, with ends

mapping to pericentromeric and peritelomeric regions (amp2, Fig. 2.1b). Analysis of WGS from RCMB56-pdx confirmed that both focal amplifications were unchanged as compared to the original primary human tumor. To assemble high-confidence sequences for the two amplicons, we performed optical genome mapping (OGM, Bionano Genomics) of ultrahigh molecular weight DNA (UHMW DNA) from RCMB56-pdx. Assembly from deep WGS and OGM validated the circular amp1 composed of three DNA segments from chr1 (Fig. 2.1a). This analysis also validated the contiguous chromothriptic amplification comprising 21 segments of chr7 and chr17; however, a circular structure could not be conclusively established from OGM and WGS data (Fig. 2.1b). Copy number of amp1 and amp2 was estimated from WGS data at 20 and 10 in RCMB56-ht, and 30 and 25 in RCMB56-pdx (Suppl. Figs. 5-6 of [14]).
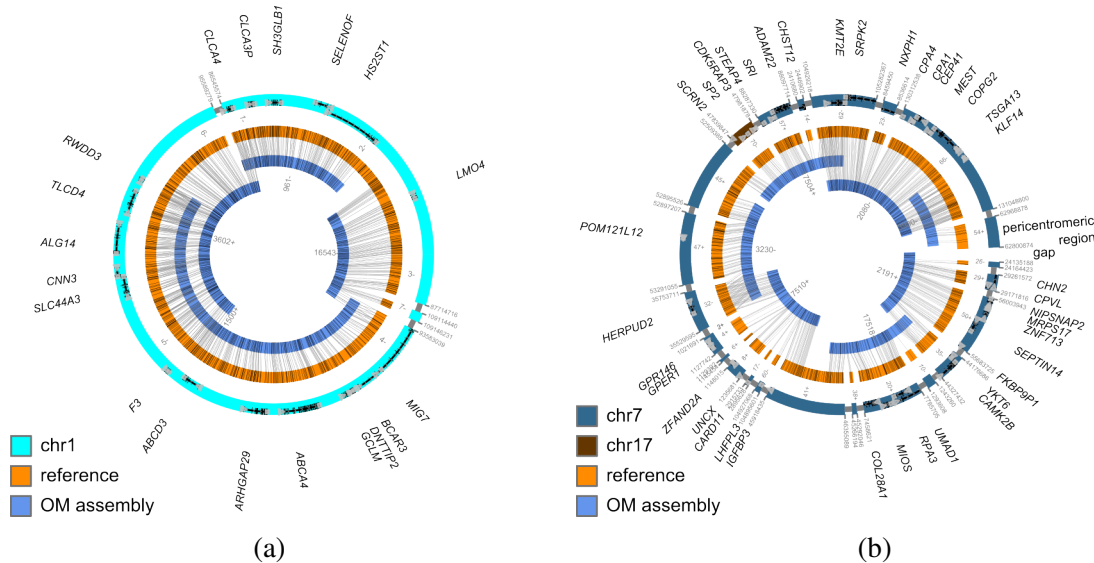
Figure 2.1: **Sequence assemblies of extrachromosomal amplifications in RCMB56.** (**a**) RCMB56 amp1 is a circular amplification composed of 3 segments of chr1p. (**b**) RCMB56 amp2 is a noncircular assembly composed of 21 segments of chr7 and chr17. Ends of the assembly map to pericentromeric (pictured, supported by WGS and OGM data) and peritelomeric (not pictured, supported by WGS data only) regions of chr7. Sequences were assembled from WGS aligned to the hg38 reference genome and scaffolded by OGM data [55].

To further validate the sequence assemblies of amp1 and amp2, we used a recent method for targeted profiling of ecDNA, CRISPR-CATCH [44]. As expected, cutting amp1 in UHMW DNA from RCMB56-pdx produced a single fraction of DNA with the predicted length of ecDNA

amp1. Moreover, short read sequencing maps this DNA to the amp1 sequence identified from bulk sequencing, confirming its circular structure (Fig. 2.2a-b).

Cutting amp2 produced 2 distinct fractions of DNA with lengths substantially shorter than amp2, which mapped to mutually exclusive segments of the amp2 assembly. Not cutting amp2 produced a single fraction with the predicted length of amp2 and mapping to the amp2 sequence (Fig. 2.2c-d). Both observations are consistent with the linear sequence assembly shown in Fig. 2.1a, and inconsistent with a hypothetical sequence linking the two ends of the amp2 assembly.

Visual inspection of the WGS data also revealed a low-copy gain (gain1) of unknown architecture composed of other segments of chr7 (35Mbp) and chr17 (800kbp). DNA fluorescence *in situ* hybridization (FISH) imaging of metaphase cells for marker genes *DNTTIP2* (amp1), *KMT2E* (amp2) and *ETV1* (gain1) indicated that amp1 and amp2 are amplified extrachromosomally (Fig. 2.3). To confirm co-occurrence in the same cells, we performed multi-channel FISH imaging for the same markers in interphase cells. We observed distinct fluorescence spots for each gene, indicating that copies of each amplified gene are located on distinct chromatin bodies (Fig. 2.4).

2.3.2 MB patient tumors with ecDNA exhibit high intratumoral copy number heterogeneity.

Substantial intratumoral copy number heterogeneity is expected in an ecDNA+ tumor due to random segregation of ecDNA during mitosis, driving tumor evolution and treatment resistance [63]. To determine the extent of ecDNA copy number heterogeneity in patient MB tumors, we established an automated image analysis pipeline to estimate the distributions of copy number per cell in interphase FISH microscopy imaging (see Methods 2.5.2) and applied it to four primary MB tumors harboring ecDNA: MB036 (*MYCN*), MB177 (*MYCN*), MB268 (*MDM4*), and RCMB56 (*DNTTIP2*, *KMT2E*, *ETV1*). The estimated copy number per cell of all ecDNA-amplified marker genes had significantly greater mean (Wilcoxon test) and variance
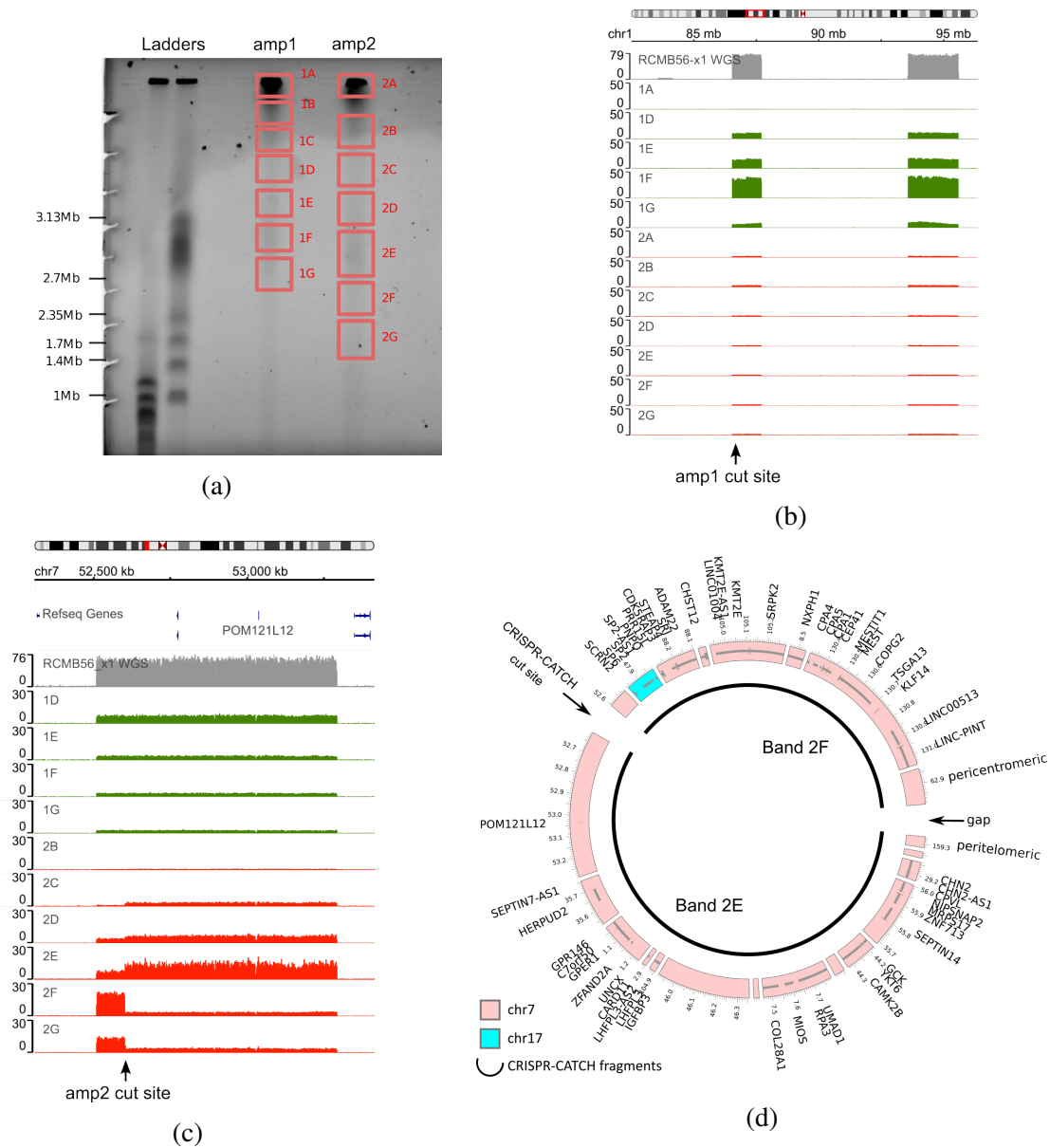
Figure 2.2: **CRISPR-CATCH indicates circular and linear extrachromosomal amplifications in RCMB56 tumors.** (**a**) UHMW DNA from RCMB56-pdx was cut using CRISPR-Cas9 targeting locations on amp1 or amp2, then sorted by size using pulsed-field gel electrophoresis (PFGE). Size-sorted bands were then sequenced and aligned to the human reference genome hg38. Circular and chromosomal DNA do not migrate, and are sampled in bands 1A and 1B. (**b**) WGS depth of sequencing coverage of all bands at the amp1 locus. When amp1 is cut (green tracks), PFGE enriches linear UHMW DNA mapping to amp1 with a unimodal size distribution centered at $s \approx 3$Mbp (band 1F). When amp1 is not cut (red tracks), no enrichment of amp1 sequences is detected. (**c**) WGS sequencing coverage of all bands at the amp2 locus. When amp2 is not cut (green tracks), PFGE enriches linear UHMW DNA mapping to amp2 with size $s \gg 3$Mbp (band 1D). When amp2 is cut (red tracks), PFGE enriches two distinct fractions of UHMW DNA, mapping to mutually exclusive fragments of the amp2 assembly and with lengths $s_1 \approx 2.8$Mbp (band 2E) and $s_2 \approx 2.4$Mbp (band 2F). (**d**) Genomic map of the fragments of amp2 enriched in bands 2E and 2F.

30
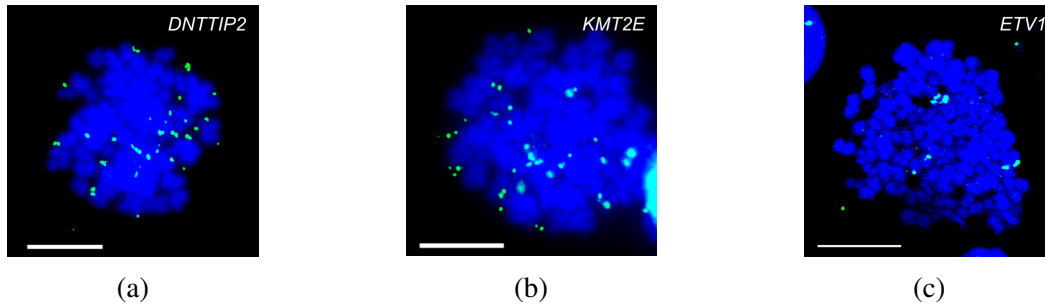
Figure 2.3: **FISH imaging of amplified marker genes in metaphase spreads of RCMB56-pdx nuclei.** (**a**) Marker gene *DNTTIP2* (amp1). (**b**) Marker gene *KMT2E* (amp2). (**c**) Marker gene *ETV1* (gain1). DAPI chromatin stain in blue. Scale bars 10*μm*.
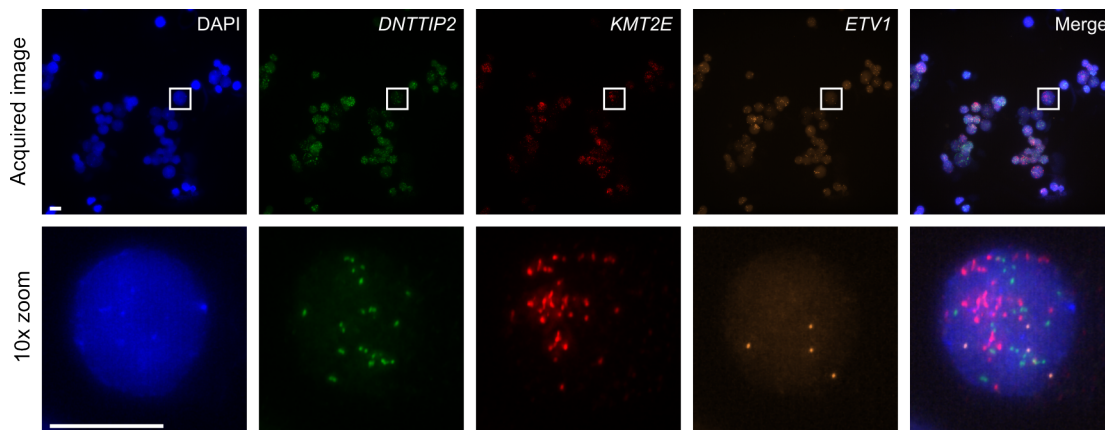
Figure 2.4: **Multi-channel FISH imaging of amplified marker genes in RCMB56-ht interphase cells.** Example image shown on top row, 10x zoom of an example single cell on bottom row. Columns from left to right: DAPI channel (chromatin), *DNTTIP2* (amp1), *KMT2E* (amp2), *ETV1* (gain1), merge of all channels. Marker gene copies are spatially independent, indicating distinct amplifications of all 3 markers in the same cells.

(Levene's test) than the ecDNA- cell line COLO320-HSR (Fig. 2.5) which includes *MYC* on a chromosomal amplification [45]. These results from human MB tumors are consistent with high copy number heterogeneity observed in human cancer cell lines with ecDNA [63]. In each primary tumor analyzed, ecDNA was amplified (copy number greater than 5) in only a subset of cells (22-41%, Supplementary Table 5 of [14]).

To determine whether copy number heterogeneity of an ecDNA+ tumor is accompanied by transcriptional heterogeneity, we analyzed 2,762 single nuclei from frozen tissue of RCMB56-ht using a multiome snRNA+ATAC-seq assay (10X Genomics) to profile transcriptomes and
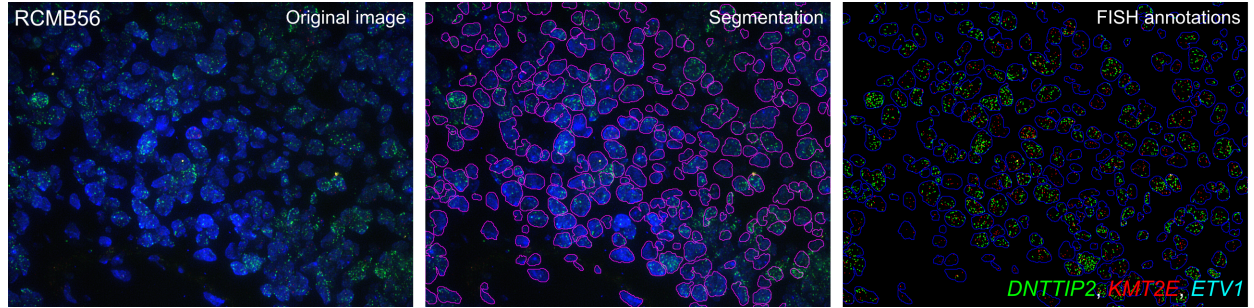
(a)
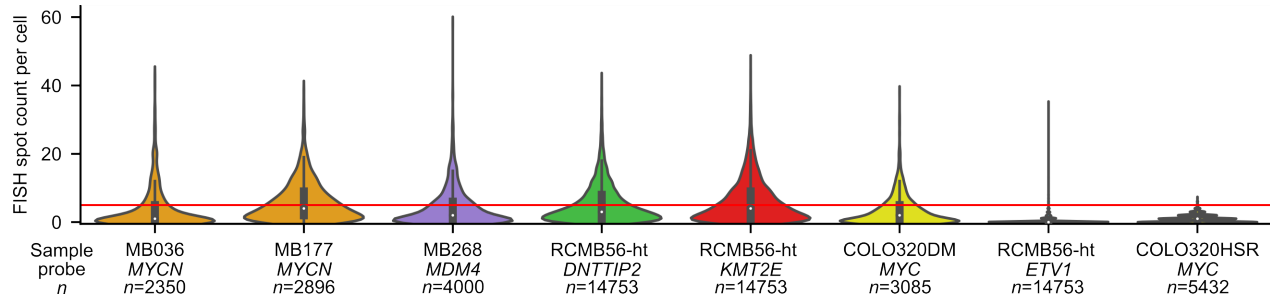
(b)

Figure 2.5: **Copy number estimation of ecDNA in single cells from FISH microscopy.** (**a**) Example output of automated image analysis pipeline. Segmentation: nuclear boundaries were detected from the DAPI channel image using NuSeT [163]. FISH annotations: The number of FISH foci was inferred using ecSeg-i [107]. (**b**) Estimated copy counts for amplified foci on ecDNA and other copy gains. Red line indicates $n \geq 5$, the threshold whereupon a cell was classified as amplified.

accessible chromatin of the same individual cells. Consistent with previous findings in bulk samples [141, 160] RCMB56-ht single nucleus ATAC-seq (snATAC-seq) sequencing coverage was enriched at the amp1 and amp2 loci at the aggregate level, and also in individual cells (Fig. 2.6a). To detect focal amplifications in single nuclei, we performed Monte Carlo permutation tests comparing snATAC-seq read density at the amplicon locus to that at random locations elsewhere in the genome (excluding blacklisted regions, see Methods 2.5.3). Z-score normalized read density at the amp1 and amp2 loci had greater mean and variance than at gain1 (Fig. 2.6b), consistent with the copy number heterogeneity we observed in the interphase FISH data. We conservatively estimate that at least 224 out of 2762 (8%, FDR $q < 0.10$) cells contained amp1 or amp2 (ecDNA+ cells). Of these, both amp1 and amp2 were detected together in only a minority

of cells (72 out of 224, 32%) (Fig. 2.6c). Thus, evidence from quantitative FISH microscopy and multiome single-cell sequencing show that only a fraction of tumor cells in ecDNA+ MB tumors harbor high-copy ecDNA, and that these have highly variable copy numbers of extrachromosomal amplifications.

(a)

(b)

(c)
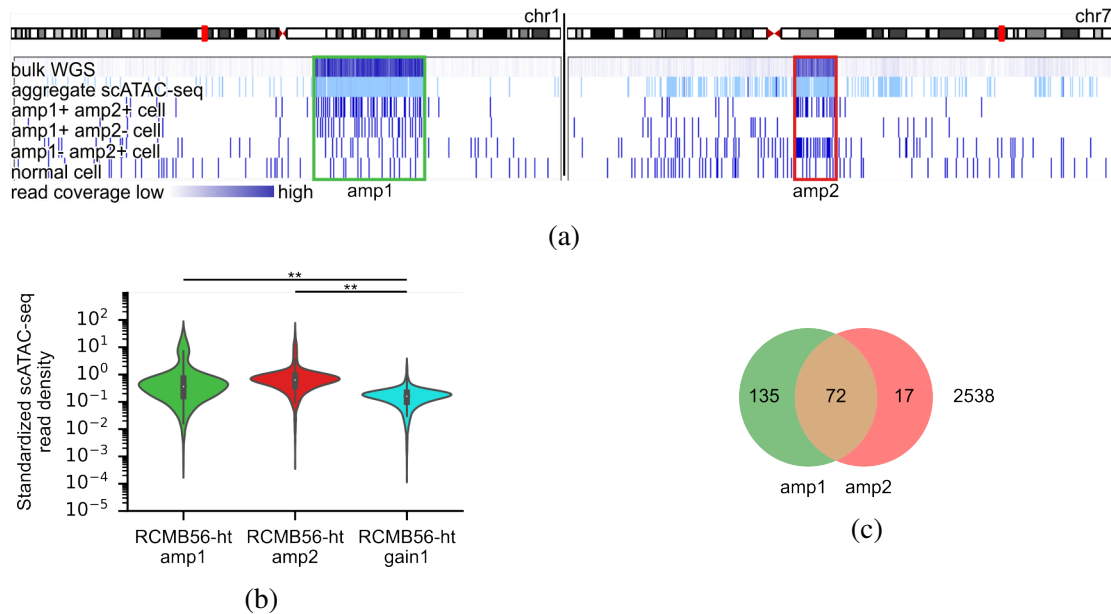
Figure 2.6: **Detection of ecDNA in in single cells from single-cell ATAC-seq of RCMB56-ht.**
(**a**) Mapped read density of sequencing modalities at RCMB56 amp1 and amp2 loci. Tracks from top to bottom: bulk WGS, aggregate scATAC-seq across all cells, and scATAC-seq of individual cells. (**b**) Distribution of Z-score normalized read coverage at the amp1, amp2, and gain1 loci. ** $p < 0.005$, Mann-Whitney test. (**c**) Sets of RCMB56 cells identified as harboring high-copy amplification of amp1 and amp2. Monte Carlo permutation test, $q < 0.05$.

2.3.3   ecDNA+ and ecDNA- tumor cell populations in the same tumor have distinct transcriptional profiles.

Clustering single cells using the Weighted Nearest Neighbors algorithm (WNN, [40]) placed the majority of ecDNA+ cells in a single cluster with distinct transcriptional and epigenetic features (Fig. 2.7). As expected, cells in the ecDNA+ cluster overexpressed *DNTTIP2* ($q < 0.001$, Wilcoxon Rank Sum test) and *KMT2E* ($q < 0.001$), the marker genes amplified on amp1 and

amp2. Compared with other tumor and normal cells, ecDNA+ tumor cells also overexpressed *GLI2* ($q < 0.001$), a known mediator of SHH-mediated transcription and a marker for SHH medulloblastoma (MB), despite *GLI2* not being affected by copy number alteration in this tumor. (Fig. 2.7c, Supplementary Table 6 of [14]). To further investigate the relationship between ecDNA copy number and transcription, we estimated first, ecDNA copy number in single cells (z-scores), and second, the transcriptional activity of genes amplified on ecDNA in each cell (ssGSEA [6] scores, see Methods 2.5.3). As expected, ssGSEA scores were positively correlated with z-scores, indicating increased transcription of ecDNA-amplified genes with increasing ecDNA copy number (Fig. 2.8).

In addition to the ecDNA+ tumor cells, we identified two other clusters of tumor cells not enriched for ecDNA and with low expression of the amp1 and amp2 marker genes, one of which strongly expressed mitochondrial genes ('ecDNA-' and 'ecDNA- MT high'), as well as normal cells such as astrocytes, oligodendrocytes and cells from the hematopoietic lineage (Fig. 2.7c). Normal cell types were annotated by cell cluster-specific expression of known marker genes. Genomic copy number estimation from snRNA-seq confirms that normal cells had stable genomes whereas the tumor cells were characterized by copy-number alterations (Fig. 2.9).

2.3.4   A PDX model resembles ecDNA+ cells of its origin human tumor.

Patient-derived xenografts (PDX) and cell lines have been used as preclinical models for identification and testing of compounds for new targeted therapies for MB [117]. To test whether cells in a PDX faithfully model ecDNA in the corresponding human tumor, we applied multiome scRNA+ATAC-seq to fresh tissue from passage 1 of the RCMB56-pdx model. In contrast to RCMB56-ht, nearly every human cell in the RCMB56-pdx sample contained ecDNA (9608 out of 9,620, 99.96%, $q < 0.10$) (Fig. 4f, Supplementary Fig. 7b). The majority of these contained both amp1 and amp2 (8554 out of 9608, 89%). Batch-correction and dimensionality reduction on the unified set of transcriptomes from both samples, using the Conos package [7], indicates that
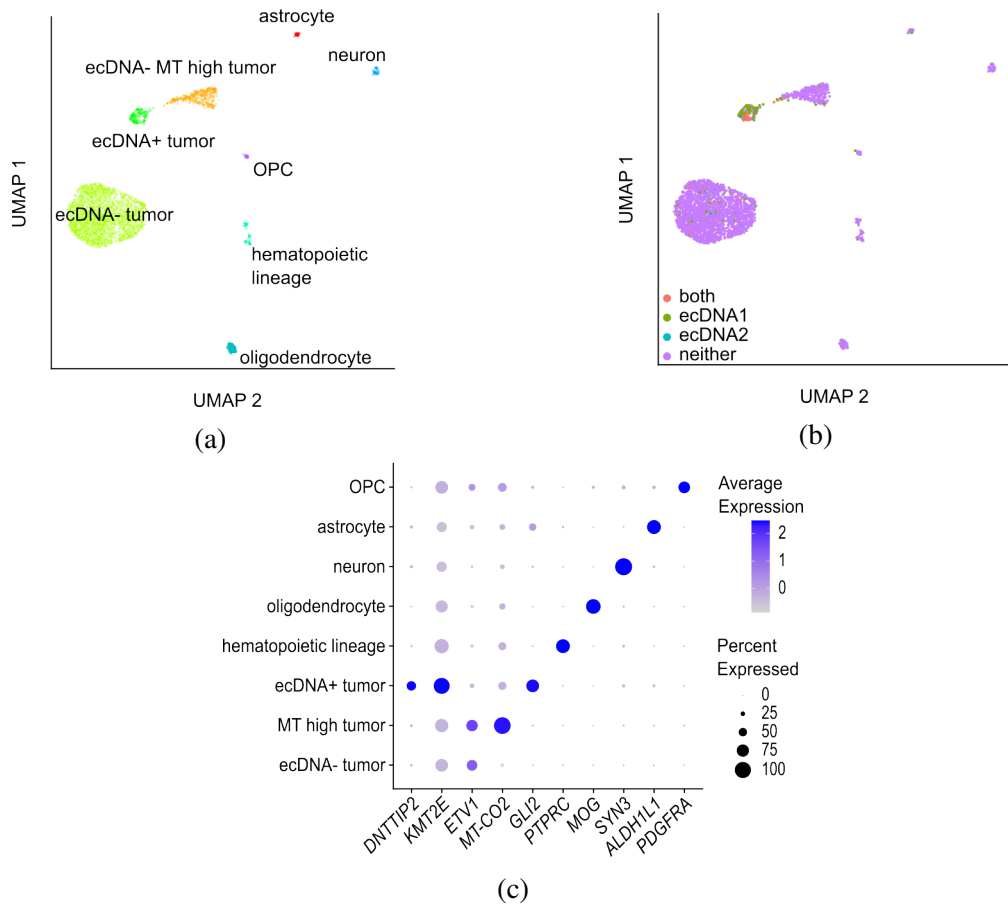
Figure 2.7: **RCMB56 ecDNA+ cells express a distinct transcriptional and epigenetic profile.** (**a**) UMAP projection of 2,762 RCMB56-ht cells were clustered by joint transcriptional and epigenetic profiles using the WNN algorithm [40]. ecDNA- tumor: cells harboring other structural variants but not enriched for ecDNA (see also Fig. 2.9). ecDNA- MT high: ecDNA-tumor cells with high expression of mitochondiral genes. Normal cells were identified by expression of marker genes. (**b**) UMAP projection of RCMB56-ht cells labelled by high-copy focal amplification of ecDNA, identifed by depth of scATAC-seq coverage at ecDNA loci. (**c**) Transcription of marker genes in single cells of RCMB56-ht.

the cells of the PDX are transcriptionally similar to ecDNA+ cells in RCMB56-ht, and distinct from all other RCMB56-ht cells (Fig. 2.10). We then asked whether the human tumor ecDNA+ cell subpopulation harbored additional identifying genomic alterations which could be confirmed in the PDX. Genomic copy number estimation from the scRNA-seq data indicated that tumor cells of the RCMB56-ht ecDNA+ cluster uniquely harbored chr3p deletion and chr3q duplication but not chr5q duplication (Fig. 2.9). Cells from RCMB56-pdx universally harbored the same

35
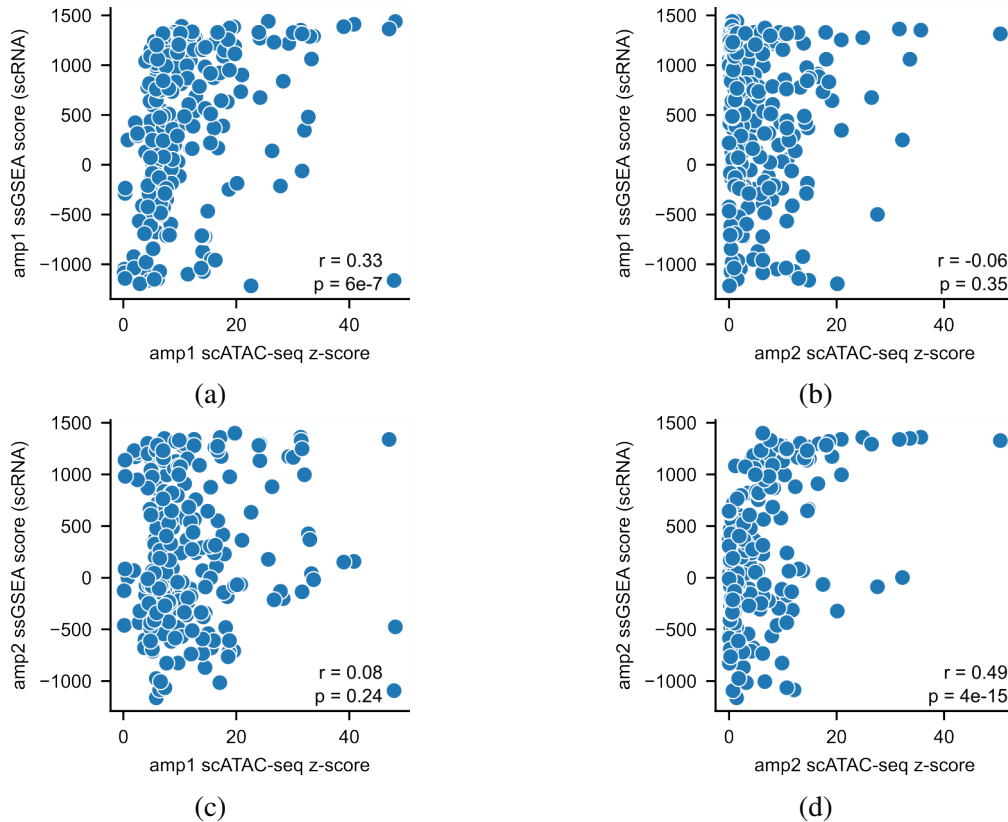
Figure 2.8: **ecDNA copy number is associated with gene expression of ecDNA-amplified genes.** Pearson correlations between copy number at the amplified locus (z-scores) and transcriptional activity of amplified genes (ssGSEA scores). Copy number of amp1 is associated with transcription of amp1-amplified genes, but not with transcription for amp2-amplified genes. Conversely, copy number of amp2 is associated with amp2-amplified, but not amp1-amplified, gene expression.

alterations (Fig. 2.11). These results indicate that the PDX tumor RCMB56-pdx is a clonal expansion of the ecDNA+ cells of the human tumor RCMB56-ht.

2.3.5 The relapsed RCMB56 patient tumor karyotype was present in a subpopulation of ecDNA+ cells in the primary.

Following maximal resection, the RCMB56 primary tumor had been treated according to the standard of care with high-dose craniospinal proton therapy and chemotherapy including cisplatin, cyclophosphamide, and vincristine [117]. Two years after initial diagnosis, however, the
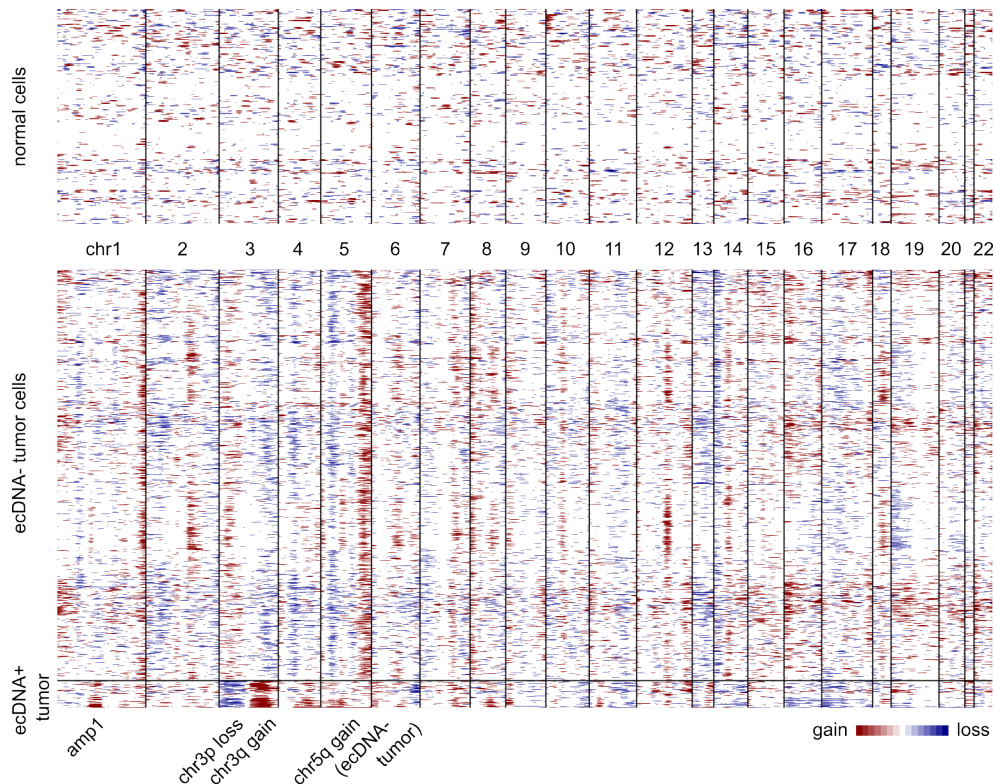
Figure 2.9: **Aneuploidy estimated from single-cell transcriptomes of 2,762 cells from RCMB56-ht** using inferCNV [137]. Diploid cells comprising the astrocyte, neuron, OPC, hematopoietic and oligodendrocyte cell clusters comprised the reference set of normal cells. The cluster enriched for high-copy ecDNA is annotated. This cluster also uniquely harbors chr3p deletion and chr3q gain.

RCMB56 tumor recurred and a second surgical resection was performed. WGS data from formalin fixed paraffin embedded (FFPE) tissue of the relapsed tumor (RCMB56-r1) indicated copy-number amplification at the amp1 locus but not the amp2 locus (data not shown). Computational reconstruction using AmpliconArchitect (AA) confirmed that the amp1 sequence remained largely unchanged. Following WGS, no tissue remained of the relapsed tumor for further analysis. Thus, although we cannot rule out that amp2 was present in an area of the relapsed tumor which was not sampled, we observe amp1 but no evidence of amp2 in the RCMB56 relapsed patient tumor.

(a)

(b)
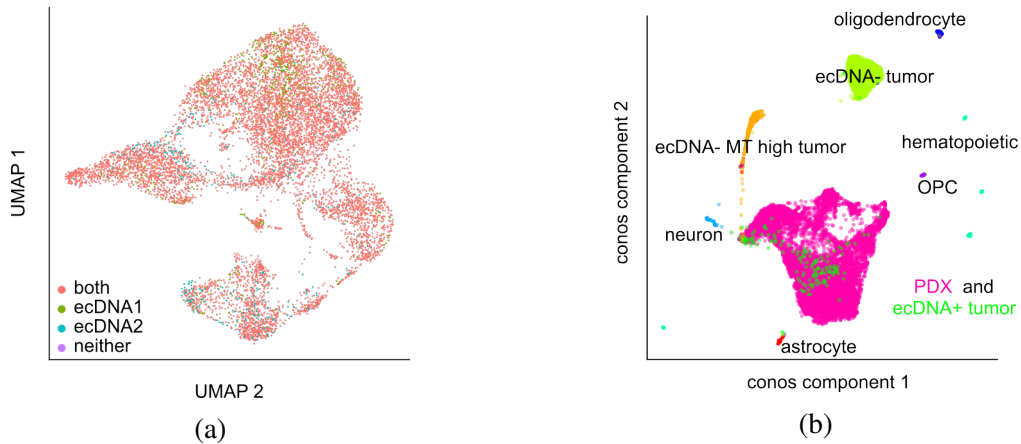
Figure 2.10: **High-copy ecDNA is universally present in 9,620 RCMB56-pdx cells.** (**a**) Low-dimensional UMAP embedding of 9,620 RCMB56-pdx accessible chromatin and transcriptomes, annotated by ecDNA detected by scATAC-seq. (**b**) Low-dimensional embedding of 9,620 RCMB56-pdx and 2,762 RCMB56-ht transcriptomes, integrated and batch-corrected using Conos [7]. PDX cells (pink) overlap ecDNA+ cells from the patient tumor, but are distinct from other RCMB56-ht cell clusters.
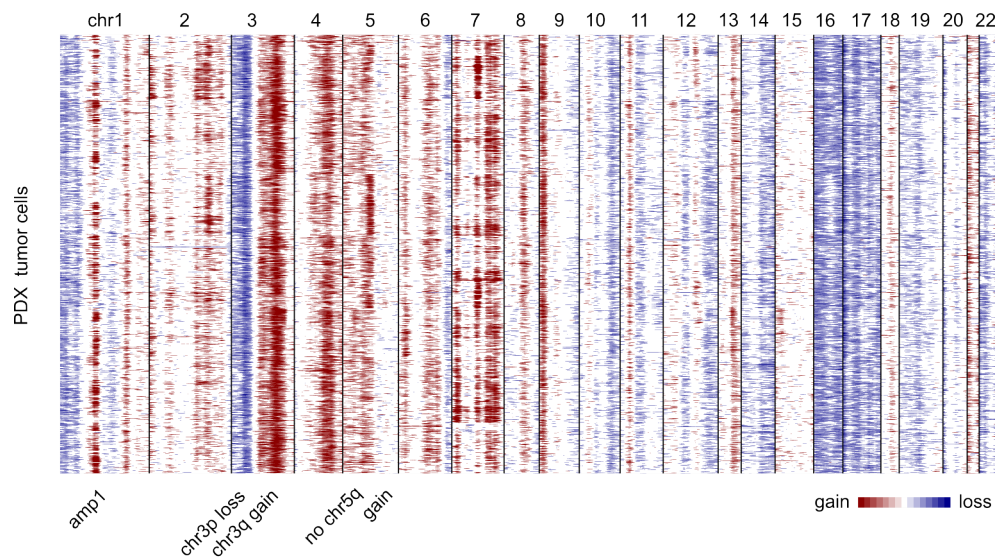
Figure 2.11: **Aneuploidy estimated from single-cell transcriptomes of 9,620 cells from RCMB56-pdx** using inferCNV [137]. PDX cells univerally harbor the chromosomal copy gain (chr3q) and loss (chr3p) which uniquely identify ecDNA+ cells in the RCMB56-ht human tumor (Fig. 2.9).

2.4   Discussion

Analyses of the p53-mutant SHH tumor RCMB56-ht reveal intratumoral heterogeneity not present in subsequent longitudinal biopsies of the same tumor. Sequence assembly, from whole genome short-read sequencing, optical genome mapping on long reads, and CRISPR-CATCH, reconstructs two distinct high-copy extrachromosomal focal amplifications, one from 3 segments of chr1 and the other from 20 segments of chr7 and chr17. Analyses of FISH microscopy and single-cell sequencing of RCBM56-ht tissue concur that both amplicons may be found in the same tumor cells, and that amplicon copy number is highly variable between cells. Further analysis of multi-omic single-cell sequencing data reveals that tumor cells with high-copy ecDNA expressed a distinct transcriptional and epigenetic program, including *GLI2*, a marker of malignant SHH subgroup MB. Thus, the RCMB56 patient tumor comprised a population of cells with heterogeneous ecDNA sequences, ecDNA copy number, and cell states.

Although ecDNA+ and ecDNA- tumor cells coexisted in RCMB56-ht, the presence or absence of ecDNA became fixed in subsequent tumors descended from the primary tumor. In the relapsed tumor RCMB56-r1, resected 2 years after original diagnosis, we find amp1 but no evidence of amp2. In RCMB56-pdx, amp1 and amp2 are near-univerally present, indicating loss of the ecDNA- tumor cell subpopulation during engraftment. Thus, in either RCMB56 tumor, we observe substantial changes of intratumoral heterogeneity of ecDNA following a strong selective pressure. It remains to be seen whether similar changes in ecDNA+ cell populations occur in other patient tumors during tumor progression, or in other PDX models under selective pressure.

Data from multiple modalities confirm the circular extrachromosomal sequence of amp1, but not amp2. Whereas sequence assembly from WGS and OGM data fully reconstruct the circular amp1 sequence, the same data map a linear amp2 sequence assembly, with ends mapping to pericentromeric and peritelomeric regions of chr7. Cutting circular amp1 by CRISPR-CATCH yields linear DNA of the approximate length of the amp1 assembly, but cutting amp2 results in 2

fractions of linear DNA, whose lengths and sequences sum to the length and sequence of the amp2 assembly. In chapter 3, we will observe in Hi-C data that the ends of the amp2 assembly do not co-localize spatially. Nevertheless, amp2 appears a high-copy extrachromosomal amplification in FISH imaging, and the copy number distribution per cell resembles that of circular amp1. The base-resolution sequences of the amp2 assembly ends remain an intriguing unresolved question, and may shed light on whether stable linear extrachromosomal amplification can arise in human tumors.

2.5   Methods

Animals

NOD-SCID IL2Rγ null (NSG) mice used for intracranial human tumor transplantation were purchased from The Jackson Laboratory (#005557). Mice were bred and maintained in the animal facilities at the Sanford Consortium for Regenerative Medicine. All experiments were performed in accordance with national guidelines and regulations, and with the approval of the animal care and use committees at the Sanford Burnham Prebys Medical Discovery Institute and University of California San Diego (San Diego, CA, USA).

Establishment and maintenance of RCMB56-pdx

The RCMB56 patient-derived xenograft model was established by implanting 0.5-1x106 dissociated patient tumor cells directly into the cerebellum of NSG mice. Subsequent tumors were harvested from mice, dissociated and reimplanted into new NSG mice without *in vitro* passaging. *Ex vivo* experiments were performed with PDX RCMB56 cells of *in vivo* passage 1 (x1).

2.5.1 Optical genome mapping data collection and processing

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

DNA labeling was performed following manufacturer's protocols (Bionano Genomics, USA). Standard Direct Labeling Enzyme 1 (DLE-1) reactions were carried out using 750 ng of purified UHMW DNA. The fluorescently labeled DNA molecules were imaged sequentially across nanochannels on a Saphyr instrument. A genome coverage of at least 400x was achieved for all samples.

De novo assemblies of the samples were performed with Bionano's *De Novo* assembly Pipeline (DNP) using standard haplotype aware arguments (Bionano Solve v3.6). With the Overlap-Layout-Consensus paradigm, pairwise comparison of DNA molecules was used to create a layout overlap graph, which was then used to generate the initial consensus genome maps. By realigning molecules to the genome maps ($p < 10^{-12}$) and by using only the best matched molecules, a refinement step was done to refine the label positions on the genome maps and to remove chimeric joins. Next, during an extension step, the software aligned molecules to genome maps ($p < 10^{-12}$), and extended the maps based on the molecules aligning past the map ends. Overlapping genome maps were then merged ($p < 10^{-16}$). These extension and merge steps were repeated five times before a final refinement ($p < 10^{-12}$) was applied to "finish" all genome maps.

2.5.2 ecDNA reconstruction with OGM data

We used an ecDNA reconstruction strategy which incorporated the short-read derived CN-aware breakpoint graph generated by AA[22] with OGM contigs generated by the Bionano *de novo* assembly Pipeline, and in RCMB56 we utilized contigs from both the Bionano DNP as well as the Rare Variant Pipeline (RVP).

We used AmpliconReconstructor (AR) v1.01[78] to scaffold together individual breakpoint graph segments using the collection of OGM contigs. We ran AR with the –noConnect flag set and otherwise default settings. A subset of informative contigs with alignments to multiple graph segments as well as a breakpoint junction were then selected for subsequent scaffolding by AR, using the "–contig_subset" argument of AR's OMPathFinder.py script. For exploration of unaligned regions of OGM contigs used in the reconstructions, we utilized the OGM alignment tool FaNDOM[106] v0.2 (default settings). FaNDOM was used to identify the loose ends of the RCMB56 amp2.

RCMB56 amp1 and D458 were fully reconstructed as described above; however, RCMB56 amp2 required more manual intervention. Due to the fractured nature of the breakpoint graphs in RCMB56 amp2, we searched for CN-aware paths in the AA breakpoint graph (using the plausible_paths.py script from the AmpliconSuite pipeline), then converted these to in silico OGM sequences and aligned paths to OGM contigs directly using AR's SegAligner.

Metaphase spreads

Cell lines were enriched for metaphases by addition of KaryoMAX (Gibco) at 0.1μg/mL for between 2h.-overnight (0.02μg/mL overnight for dissociated PDX cells). Single cell suspensions were then incubated with 75mM KCl for 8-15 minutes at 37°C. Cells were then fixed by carnoy fixative (3:1 methanol:acetic acid) and washed in fixative 3 times. Cells were then dropped onto humidified slides.

FISH

Slides containing fixed cells were briefly equilibrated in 2X SSC, followed by dehydration in 70%, 85%, and 100% EtOH for 2 minutes each. FISH probes (Empire Genomics) diluted in hybridization buffer (EG) were applied to slides and covered with a coverslip. Slides were denatured at 72°C for 1-2 minutes and hybridized overnight at 37°C. The slide was then washed with 0.4x SSC, then 2x SSC-0.1 Tween 20. DAPI was added before washing again and mounting with Prolong Gold.

Microscopy

Conventional fluorescence microscopy was performed using either the Olympus BX43 microscope equipped with a QiClick cooled camera, or the Leica DMi8 widefield fluorescence microscope followed by Thunder deconvolution using a 63X oil objective. Confocal microscopy was performed using a Leica SP8 microscope with lightning deconvolution and white light laser (UCSD School of Medicine Microscopy Core). Excitation wavelengths for multiple color FISH images were set manually based on the optimal wavelength for the individual probes, with care taken to minimize crosstalk between channels. ImageJ was used to uniformly edit and crop images.

2.5.3   Automated FISH analysis

Cell segmentation. We applied NuSeT [163] to perform cell segmentation. Since all interphase tissue images were taken at the same resolution of 0.240 μm, we used the same NuSeT parameters for all images. Parameters were min_score 0.95, nms threshold of 0.01, a nuclei size threshold of 500 and a scale ratio of 0.3. The average number of fish pixels per cell was computed by counting the number of FISH pixels with a pixel intensity above 85 pixels in each cell. *Number of FISH blobs.* We convolved the original image with a normalized gaussian kernel

to determine which pixels have high local intensity. We used a sampled gaussian kernel with a standard deviation of 3 pixels, and a size of 7 by 7 pixels. After convolving, we applied a threshold of 15 / 255 pixel brightness. Then, to filter out low brightness noise, we set a binary threshold that the brightness of these peaks must exceed one standard deviation above the average FISH brightness, and added an additional minimum area requirement. *Amplification mechanism.* We ran ecSeg-i [107] on each segmented cell to determine the amplification mechanism. ecSeg-i produces three probability scores representing the likelihood of the cell having no amplification, ecDNA amplification, or homogeneously staining region (HSR) amplification. We assigned the amplification mechanism with the highest likelihood.

Sample preparation and single-cell multiome sequencing

From the RCMB56 primary human tumor (RCMB56-ht), disassociated cryopreserved cells stored in 10% DMSO/FBS were used. Fresh tumor tissue was acquired from RCMB56 models grown as described above. At least 50mg of tissue (1M cells) was used for both samples. Disassociated cells were prepared for single-nucleus multiome RNA-seq and ATAC-seq (10X Genomics) (snRNA+ATAC) according to the manufacturer's instructions[1]. Sequencing was performed on an Illumina NovaSeq S4 200 to a depth of at least 250M reads for snATAC-seq and 200M reads for snRNA-seq.

Single cell coverage tracks

Sequencing coverage of single cells were visualized in IGV desktop v2.9.297. Bulk WGS coverage (bigwig format) was generated from deduplicated sequencing reads using deeptools v3.5.1[109] bamCoverage at 50bp resolution using default parameters. Single-cell coverage tracks were parsed from CellRanger ARC atac_fragments.txt.gz output format to .bed format using a custom script, then converted to bigwig format using bedtools v2.27.1[105] genomecov and UCSC browser tools[54] bedGraphToBigWig v4. Source code is available at https://github.

com/auberginekenobi/rcmb56-single-cell.

Single cell data processing and clustering

Sequencing data were uniformly processed using CellRanger ARC v2.0.0 (10X) with default parameters, followed by Seurat v4.0.4[40]. Cell barcodes were retained whose sequencing coverage passed the following quality thresholds: ATAC mitochondrial fraction less than 0.1; ATAC read count between 1,000 and 70,000; and RNA read count between 500 (ht) or 1000 (pdx) and 25,000. Doublets were identified and removed using DoubletFinder v2.0[82] using default parameters. Host cells were identified and removed from the PDX sample, using Xenocell v1.0.1[15], as barcodes with less than 90% of uniquely mapping reads mapped to the human genome. Following preprocessing, 2,986 RCMB56-ht and 9,620 RCMB56-pdx cells remained. Single cell transcription data were normalized using regularized negative binomial regression implemented in the sctransform package[38] (SCT) included with Seurat. Clustering was performed independently on each sample using the Weighted Nearest Neighbors algorithm[40] with default parameters. To label cell clusters with cell type identities, differentially expressed genes were found for each cluster using Seurat's FindAllMarkers function with default parameters (Supplementary Table 6 of [14]). Differentially expressed genes were cross-referenced against known cell type marker genes[51]. Only the RCMB56-ht cell clusters were labelled in this way.

Single-cell data integration and batch correction

SCT normalized gene expression data for RCMB56-ht and RCMB56-pdx were integrated and batch-corrected using Conos v1.4.4[7], using default parameters and following the walk-through tutorial available at https://github.com/kharchenkolab/conos/blob/main/doc/walkthrough. md. Low-dimensional embeddings plotted in 2.10b were estimated using the largeVis algorithm implemented in Conos using random seed 47; other random seeds returned similar results. Code is available at https://github.com/auberginekenobi/rcmb56-single-cell.

Identification of ecDNA-containing cells

ecDNA-containing cells were identified by permutation tests comparing snATAC-seq read coverage at the ecDNA regions to read coverage of random regions elsewhere in the genome. Code is available at https://github.com/auberginekenobi/ecdna-quant. Briefly, deduplicated snATAC-seq reads were obtained from the fragments.tsv output of CellRanger ARC and sorted by barcode. For Monte Carlo permutation testing, 1000 random contiguous regions of the genome, excluding centromeres, telomeres, known ecDNA, and low-mappability regions, were generated using bedtools v2.27.1[105]. Read coverage was counted using PyRanges v0.0.112[132] and scaled to region length. For each cell, empirical p-values were estimated as $\hat{p} = (r+1)/(n+1)$, where r is the rank of the test value out of n permutations[97]. Multiple testing correction was performed using the Benjamini-Hochberg correction. Z-scores were calculated using the standard formula, comparing the average read coverage at the ecDNA-amplified region to the mean and variance of the Monte Carlo permutations.

Single cell genomic copy number estimation

Copy number signature was identified in the human tumor and PDX using InferCNV v1.3.3[137]. Code is available at https://github.com/suneetz/githubSingleCell/blob/0243d6a8f8c6a2e96222d659bd5b3071c9ed28a4/infercnvHT. Briefly, counts matrix and annotation files were obtained from CellRanger ARC and Seurat outputs (see Methods 2.5.3). Normal reference cells were defined as ecDNA- cells belonging to cell clusters labelled as normal cell types. Because no normal reference cells were present in RCMB56-pdx, the set of normal cells from RCMB56-ht was spiked into the RCMB56-pdx data prior to running inferCNV. The HMM fitting step was not performed for RCMB56-pdx. All other parameters were default.

Single-sample gene set enrichment analysis (ssGSEA)

ssGSEA is a variation of Gene Set Enrichment Analysis for quantifying aggregate expression of a gene set across the transcriptome of one sample[6]. To quantify transcriptional activity of ecDNA in single cells, we performed ssGSEA of two gene sets comprising every gene amplified on RCMB56 amp1 or amp2, treating each cell as a single sample. The population sample consisted of $n = 247$ ecDNA+ cells from the RCMB56-ht sample. Gene expression values were the SCT-normalized transcription matrix, generated as described above using Seurat v4.0.4. ssGSEA was run using ssGSEA v10.0.11 implemented at https://cloud.genepattern.org[112]. Association with z-score ecNDA copy number estimates was performed using Pearson's R implemented in scipy.stats v1.7.3 and visualized using Seaborn v0.9.0[154] histplot.

2.6 Acknowledgments

the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. DNA methylation array analysis was conducted at the IGM Genomics Center, University of California, San Diego, La Jolla, CA (P30 CA023100). This research was conducted using data made available by The Children's Brain Tumor Network (formerly the Children's Brain Tumor Tissue Consortium). HYC is an Investigator of the Howard Hughes Medical Institute.

Chapter 2, in part, is adapted from the following manuscript currently being prepared for publication: "Circular extrachromosomal DNA promotes inter- and intratumoral heterogeneity in high-risk medulloblastoma." Chapman, Owen; Luebeck, Jens; Sridhar, Sunita; Wong, Ivy T.L.; Dixit, Deobrat; Wang, Shanqing; Prasad, Gino; Rajkumar, Utkrisht; Pagadala, Meghana; Larson, Jon D.; He, Britney J.; Hung, King L.; Lange, Joshua T.; Dehkordi, Siavash R.; Chandran, Sahaana; Adam, Miriam; Morgan, Ling; Wani, Sameena; Tiwari, Ashutosh; Guccione, Caitlin; Lin, Yingxi; Dutta, Aditi; Lo, Yan Yuen; Juarez, Edwin; Robinson, James T.; Malicki, Denise M.; Coufal, Nicole G.; Levy, Michael; Hobbs, Charlotte; Scheuermann, Richard H.; Crawford, John R.; Pomeroy, Scott L.; Rich, Jeremy; Zhang, Xinlian; Chang, Howard Y.; Dixon, Jesse R.; Bagchi, Anindya; Deshpande, Aniruddha J.; Carter, Hannah; Fraenkel, Ernest; Mischel, Paul S.; Wechsler-Reya, Robert J.; Bafna, Vineet; Mesirov, Jill P.; Chavez, Lukas. The dissertation author was the primary investigator and author of the manuscript.

Chapter 3

TRANSCRIPTIONAL REGULATION OF ECDNA-AMPLIFIED GENE EXPRESSION IN MEDULLOBLASTOMA

3.1   Introduction

In the previous chapter, we examined the RCMB56-ht patient tumor and showed that high-copy ecDNA defined a subpopulation of tumor cells expressing a distinct transcriptional program. In this chapter, we examine transcriptional dysregulation which occurs on and around the ecDNA amplified sequences of the RCMB56-pdx and D458 medulloblastoma (MB) models as well as several patient tumors. Gene transcription is frequently regulated by distal regulatory elements (enhancers) located up to 1Mbp away [102]. In some medulloblastomas [99] and other cancers [152], a somatic structural rearrangement (structural variant, SV) may juxtapose an oncogene and an ectopic enhancer to activate oncogenic transcription. Comparing the enhancer landscapes of the Group 3 MB cell lines D458 (*MYC* ecDNA) and D283 (*MYC* HSR), we observe functional differences in the enhancer loci driving oncogene expression, indicating that the enhancer landscape of the structural variant contributes to oncogenic transcriptional dysregulation. These results add to accumulating evidence suggesting co-amplified enhancers as potential therapeutic vulnerabilities in ecDNA-amplified cancers [45].

3.2   Results

3.2.1   ecDNA places oncogenes in new gene regulatory contexts.

It has been shown that some MB tumors are driven by "enhancer hijacking" events, whereby somatic SVs cause an enhancer to be rewired to amplify transcription of *GFI1* family or *PRDM6* genes [100, 98]. Given the extensive genomic rearrangement associated with some MB ecDNA, we investigated whether new DNA interactions between co-amplified non-coding regulatory enhancers and oncogenes emerge on circular ecDNA. To test this hypothesis, we profiled the accessible chromatin of 25 MB tumors (11 ecDNA+, 14 ecDNA-) using ATAC-seq [12], as well as chromatin interactions of 17 MB tumors (8 ecDNA+, 9 ecDNA-) using chromatin

conformation capture (Hi-C) [110]. Consistent with previous reports [160, 60], bulk ATAC-seq read density was markedly enriched across entire ecDNA regions, even for ecDNA with only low-level amplification as estimated by bulk WGS. Hi-C sequencing reads exhibited a similar pattern of enrichment at ecDNA regions (Fig. 3.1).
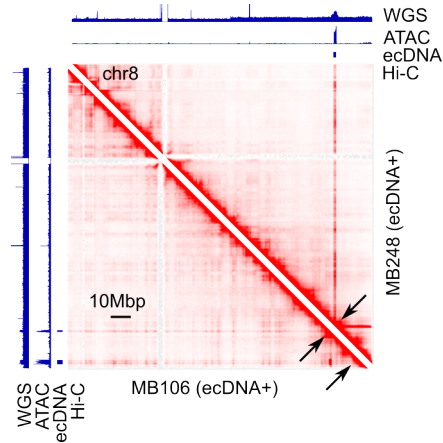
Figure 3.1: **Chromatin interaction maps of chr8 in patient tumors with *MYC* ecDNA.** Chromosome-wide view of intrachromosomal interactions mapping to chr8 of the reference human genome. Whole genome sequencing (WGS), ATAC-seq and Hi-C sequencing depth are shown for two Group 3 MB samples harboring circular amplification of *MYC*. Extrachromosomally amplified loci are indicated by arrows. These tumors are described further in [2].

In half of the analyzed ecDNA+ tumors (D458, MB106, MB268, and RCMB56), we observed clear evidence of aberrant chromatin interactions on ecDNA which spanned DNA breakpoints to link accessible loci and co-amplified genes from distal genomic regions. In the *MYC*-amplified Group 3 MB primary tumor MB106, DNA interactions occurred between the *MYC* locus and two co-amplified accessible regions located 13Mbp away on the reference genome, but 75kbp and 800kbp away on the ecDNA (Fig. 3.2a). Comparing the MB106 Hi-C interactome to that of MB288, a Group 3 MB primary tumor without ecDNA, we found that these chromatin interactions were specific to the MB106 ecDNA (Fig. 3.2b).

In the SHH MB primary tumor MB268, we identified a 10.2Mbp ecDNA amplification which included the p53 regulator *MDM4* [139] and is among the largest ecDNA amplifications identified in a patient tumor to date (Fig. 3.3). *MDM4* is recurrently amplified on glioblastoma

51
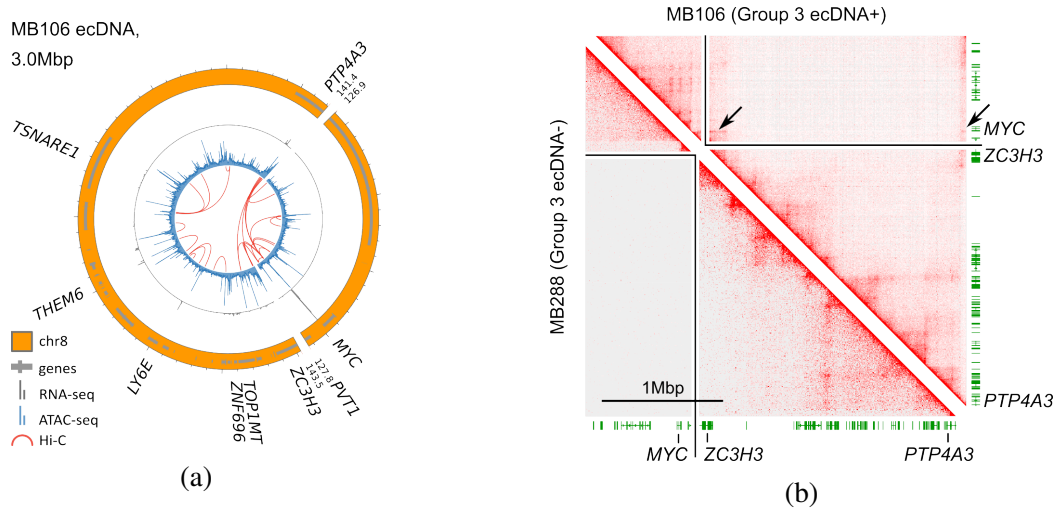
Figure 3.2: **Transcriptional regulatory circuitry of ecDNA in *MYC*-amplified Group 3 patient tumor MB106.** (**a**) Circular multi-omic map of the ecDNA sequence identified in Group 3 MB sample MB106. *MYC* is highly expressed, and interacts with distal accessible loci located elsewhere on the ecDNA sequence. (**b**) Comparison of interaction densities at the MB106 ecDNA locus between MB106 (Group 3 ecDNA+) and MB288 (Group 3 ecDNA-). Ectopic interactions (arrows) between co-amplified segments of chr8 in MB106 are not present in the MB288 tumor with unrearranged *MYC* locus.

ecDNA [22] and is a putative driver event in the MB268 tumor. On the same ecDNA, we also observed aberrant DNA interactions with the immune complement system regulator *CFH* promoter. However, the functional significance of these co-amplified genes and DNA interactions remains unclear.

In two instances, the SHH MB tumor RCMB56-pdx and the Group 3 MB cell line D458, we identified rewired interactions between genomic loci originating from different chromosomes but co-amplified on the same ecDNA. As described above, RCMB56 harbored an ecDNA comprised of segments of chr1, and a complex extrachromosomal amplification comprised of segments of chr7 and chr17. Hi-C data indicated frequent chromatin interaction across breakpoints in each of the two amplicons present in this tumor (Fig. 3.4-3.5). Aberrant chromatin interactions mapping to amp1 targeted accessible regions at the *DNTTIP2*, *SH3GLB1*, and *SELENOF* gene loci (Fig. 3.5a). Aberrant interactions on amp2 included intrachromosomal interactions mapping to *RPA3*, *HERPUD2*, *KLF14* and others; and trans-chromosomal interactions between the *SP2*
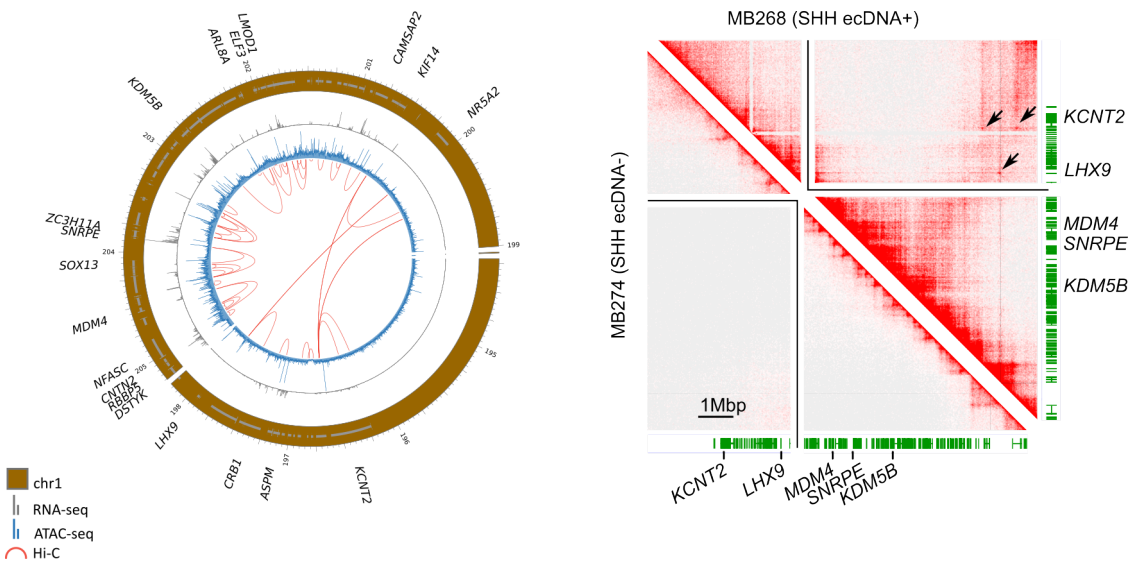
Figure 3.3: **Transcriptional regulatory circuitry of ecDNA in SHH MB tumor MB268.** (**a**) Circular multi-omic map of the ecDNA present in the SHH MB patient tumor MB268. *MDM4* is a negative p53 regulator recurrently amplified in glioblastoma [22], and is a putative driver oncogene on this amplicon. (**b**)Comparison of interaction densities at the MB268 ecDNA locus between MB268 (SHH ecDNA+) and MB274 (SHH ecDNA-). Ectopic interactions (arrows) between co-amplified segments of chr1 in MB268 are not present in unrearranged the MB274 tumor.

locus and the brain-specific long noncoding RNA *LINC0301359*, and from the *PRR15L* promoter to an intragenic region upstream of *SRI* (Fig. 3.5b).

D458 harbors an ecDNA amplification containing MB oncogenes *MYC* and *OTX2*, from chromosomes 8 and 14 respectively. Co-amplification of *MYC* and *OTX2* on the same ecDNA was validated by confocal FISH showing extrachromosomal co-localization of *OTX2* and *MYC* in D458 cells (Fig. 3.6a), and by assembly of the D458 ecDNA from WGS and OGM data (Supplementary Fig. 11 of [14]). *OTX2* is a known regulator of *MYC* transcription and both genes are highly expressed in D458 (Fig. 3.6b). The Hi-C data revealed several interactions of the *MYC* promoter with co-amplified regulatory elements of chr8 and chr14 (Fig. 3.7). In summary, these results show that aberrant enhancer-promoter interactions resulting from structural rearrangements on ecDNA are common in MB tumors.
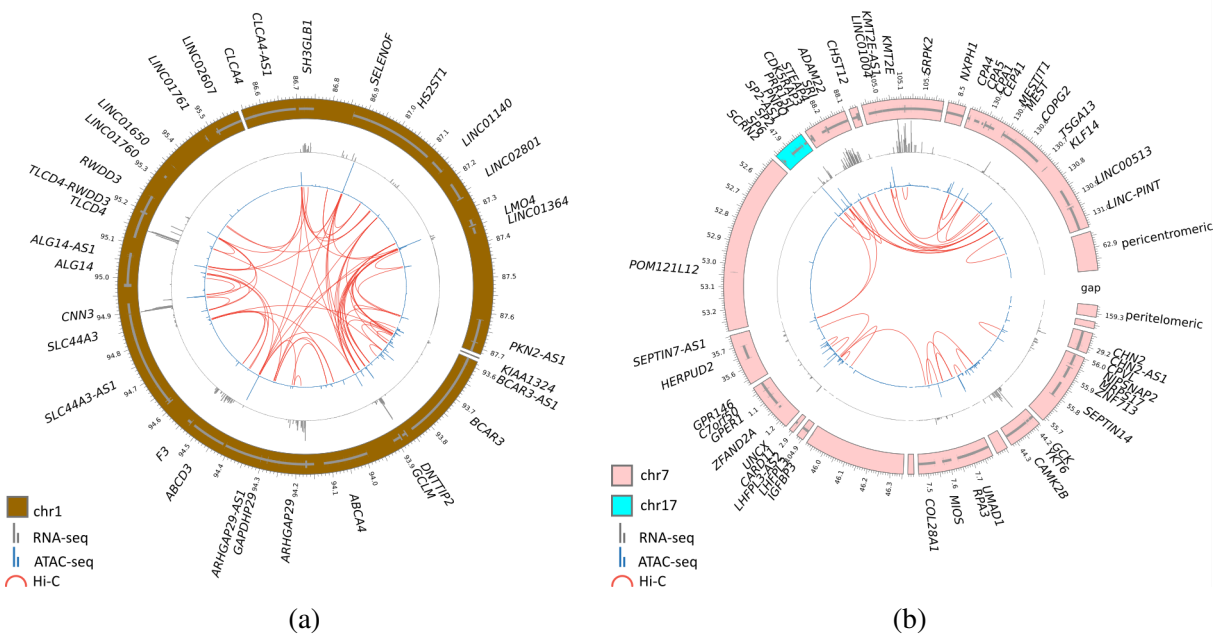
Figure 3.4: **Transcriptional regulatory circuitry of ecDNA sequences in RCMB56 p53-mutant SHH MB tumors.** (**a**) Transcription (RNA-seq, grey), accessible chromatin (blue), and chromatin interactions (Hi-C, red arcs) mapped onto the amp1 assembly (outer track, brown). Chromatin interactions occur across a structural breakpoint between accessible loci near highly-expressed genes such as *DNTTIP2*. (**b**) Transcription, accessible chromatin, and chromatin interactions mapped onto the amp2 assembly. The gap in the assembly is adjacent to pericentromeric and peritelomeric loci. Ectopic interactions putatively target *RPA3*, *HERPUD2*, *KLF14*, *SP2*, and others.

3.2.2 Co-amplified enhancers shape the transcriptional regulatory circuitry of ecDNA.

To test whether the observed enhancer rewiring events in ecDNA have functional roles in tumor cell proliferation, we performed a pooled CRISPRi proliferation screen in the Group 3 MB cell line D458, targeting all 645 accessible loci co-amplified on the ecDNA using 32,530 small guide RNA sequences (sgRNAs). These loci included 10 highly accessible regions from chr14, each overlapping ENCODE candidate cis-regulatory elements (cCREs) [90]. Because enhancer usage is highly conserved in Group 3 MB tumors [72], we performed the same screen in the Group 3 cell line D283, in which *MYC* (but not *OTX2*) is tandem amplified on a 55Mbp HSR of
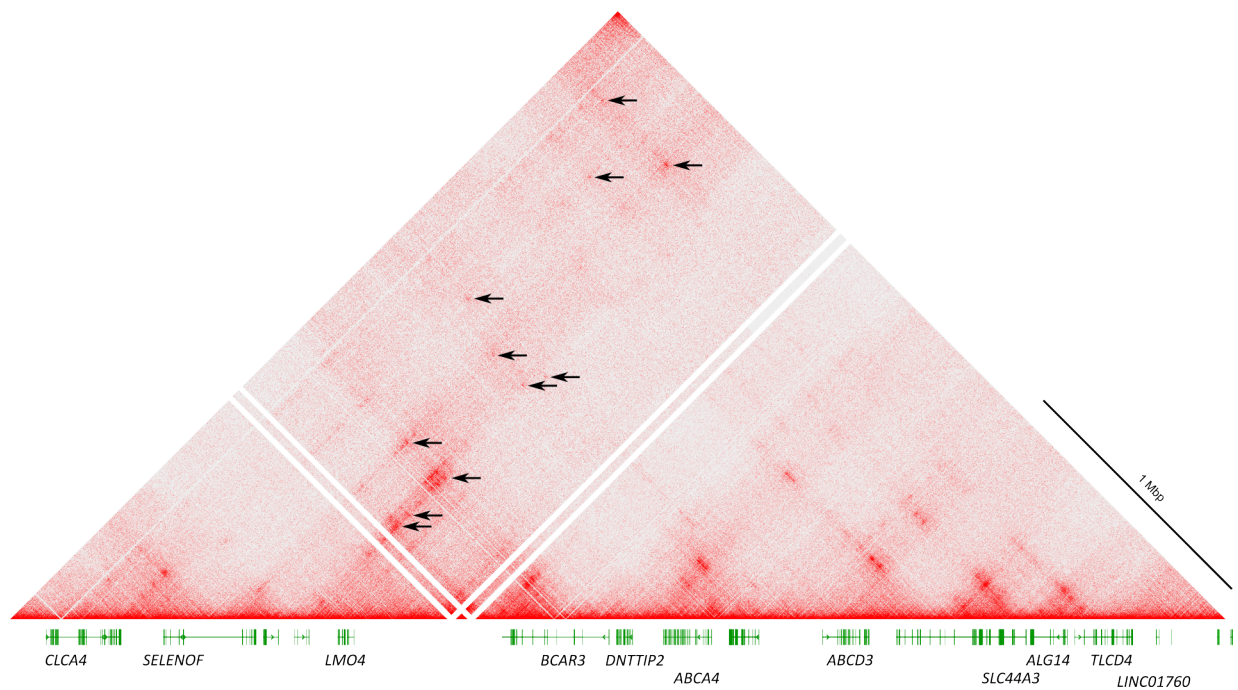
(a)
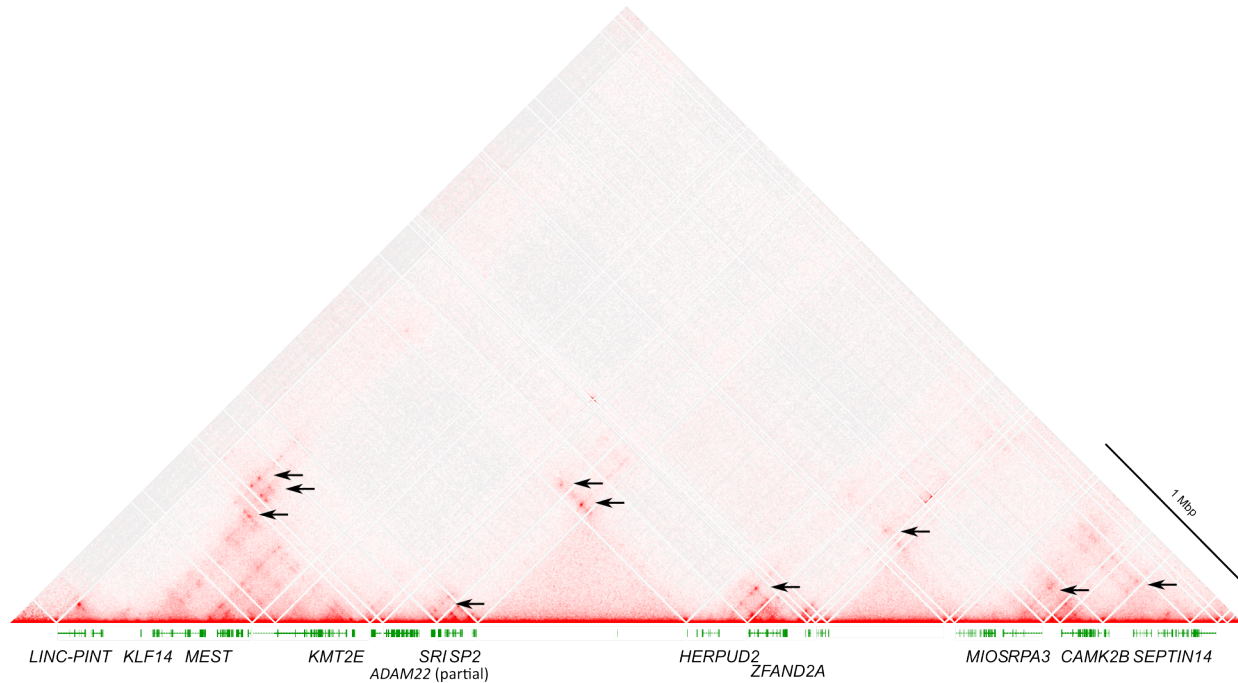
(b)

Figure 3.5: **Enhancer rewiring on ecDNA in RCMB56.** See also Fig. 3.4. (**a**) Chromatin interaction density map of the amp1 assembly. Arrows indicate putative enhancer rewiring events, or chromatin loops which span a breakpoint on the amp1 assembly. (**b**) Chromatin interaction density map of the amp2 assembly. Putative enhancer hijacking events are again indicated by arrows. The two ends of this assembly do not interact (top of the triangle), suggesting that they are spatially distant in the cell.
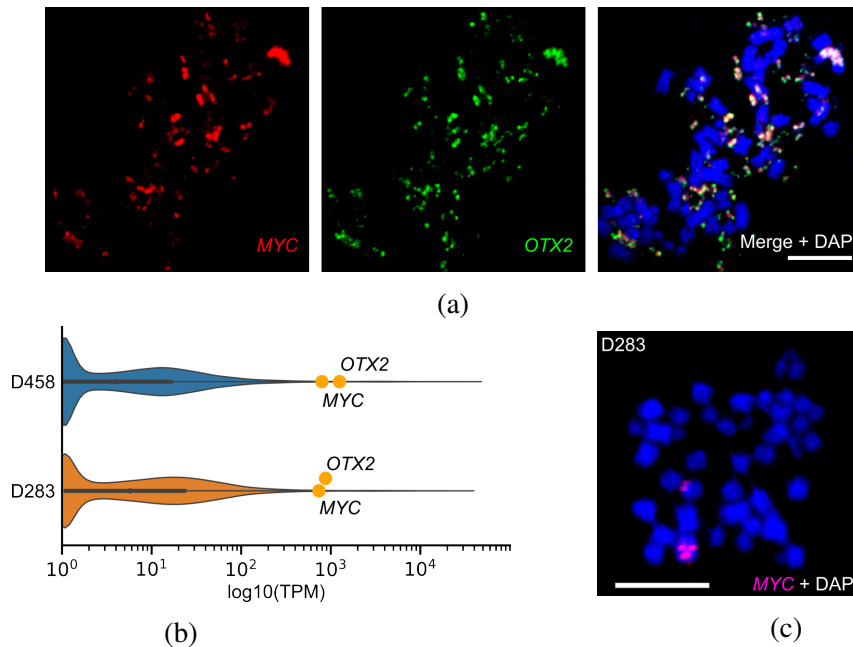
Figure 3.6: **Two Group 3 MB cell lines with intra- and extrachromosomal *MYC* amplifications.** (**a**) Confocal FISH of *MYC* and *OTX2* on a metaphase spread of a D458 cell. (**b**) Gene transcription of all protein-coding genes in D458 and D283 from publicly available data in DepMap [140]. MB Group 3 oncogenes *MYC* and *OTX2* are highlighted. (**c**) FISH in a metaphase spread of a D283 nucleus shows chromosomal homogeneously staining region (HSR) *MYC* amplification. *OTX2* is not amplified in D283.

chr8q (Fig. 3.6c). While the *MYC* promoter was essential in both cell lines, our screen identified 6 functional elements which, upon CRISPRi inhibition, specifically reduced D458 proliferation compared to D283 after 21 days (MAGeCK MLE, $q < 0.05$; Fig. 3.8) [70]. On chr8, these loci included two accessible regions of a known *MYC* superenhancer [72] and the *PVT1* promoter. In D458, much of the superenhancer is duplicated internally on the ecDNA, and *PVT1* is amplified in D458 but not in D283. Conversely, we observed that other accessible regions of the same *MYC* superenhancer were specifically essential for D283 but not for D458, indicating altered regulatory dependencies on the *MYC*-ecDNA compared to the *MYC*-HSR amplification. These results are consistent with a previous comparison of ecDNA+ and ecDNA- glioblastoma cell lines, where the impact of enhancer silencing on proliferation was associated with cell line-specific chromatin accessibility and interactions [93].

(a)

(b)
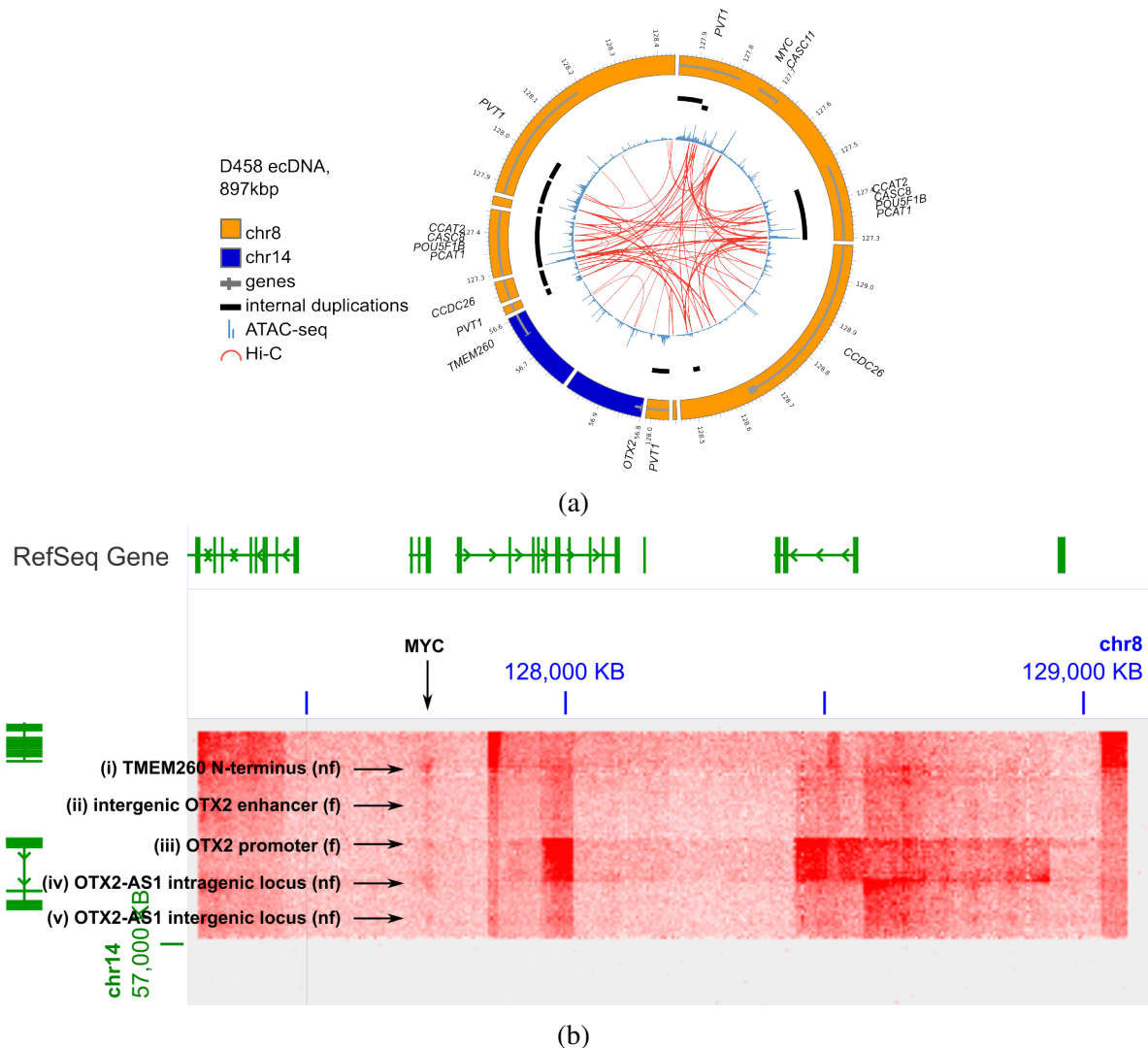
Figure 3.7: **Transcriptional regulatory map of the D458 ecDNA.** (**a**) Circular map of accessible chromatin (blue) and chromatin interactions (red) mapping to the D458 ecDNA sequence assembly. (**b**) Interchromosomal interactions between segments of the D458 ecDNA originating from chr8 and chr14. Arrows highlight putative chromatin loops included in (a). f: functional, as determined by a significant and D458-specific effect on cell proliferation upon CRISPRi inhibition (Fig. 3.8); nf: not identified as functional in the same pooled CRISPRi screen.

The Hi-C data indicated interactions between *MYC* on chr8 and regulatory elements of chr14 co-amplified on the same ecDNA, two of which are essential for D458 proliferation (Fig. 3.7b): a cluster of elements at the *OTX2* locus as well as a distal enhancer [158] located 80kbp downstream of *OTX2* on the reference genome, but inverted on the ecDNA. D283-specific

elements on chr14 included peaks at the N-terminal exon of *OTX2* and another distal enhancer [158] 55kbp from *OTX2* on the reference but also inverted on the ecDNA. To further test the influence on transcription of regulatory regions essential in D458 but not in D283, we performed additional CRISPRi inhibition experiments targeted against the *PVT1* promoter and an accessible region within the internal duplication of the *MYC* superenhancer. Consistent with the result of the CRIPSRi proliferation screen, silencing of the *MYC* superenhancer reduced *MYC* expression for 2 out of 3 tested sgRNAs in D458 but not in D283 (Fig. 3.9). No significant difference was observed in *OTX2* transcription in either cell line. Silencing of the *PVT1* promoter abrogated *PVT1* transcription but not *MYC* or *OTX2*, in D458 but not in D283 (Fig. 3.10). Thus, although proliferation in both Group 3 MB cell lines is driven by *MYC* amplification, the relative importance of co-amplified genes and cis-regulatory elements is specific to the genomic architecture of the amplicon.

3.2.3   ecDNA interacts with chromosomal oncogene loci.

Recent studies have revealed intermolecular enhancer-promoter interactions between ecDNA molecules ('ecDNA hubs') [45] or between the chromosomes and ecDNA ('mobile enhancers') [166]. To test for such intermolecular chromatin interactions in medulloblastoma, we computationally identified interchromosomal loops from Hi-C of the SHH MB tumor RCMB56-pdx, where one loop anchor mapped to the circular ecDNA amp1. This analysis revealed a nexus of interactions mapping from the *ARHGAP29* locus on amp1 to loci elsewhere in the genome with plausible tumorigenic roles, including *MECOM*, *RAD51AP2*, *POU4F1*, and *IGF1R* (Fig. 3.11). However, the functional significance of these intermolecular chromatin interactions in MB remains untested.
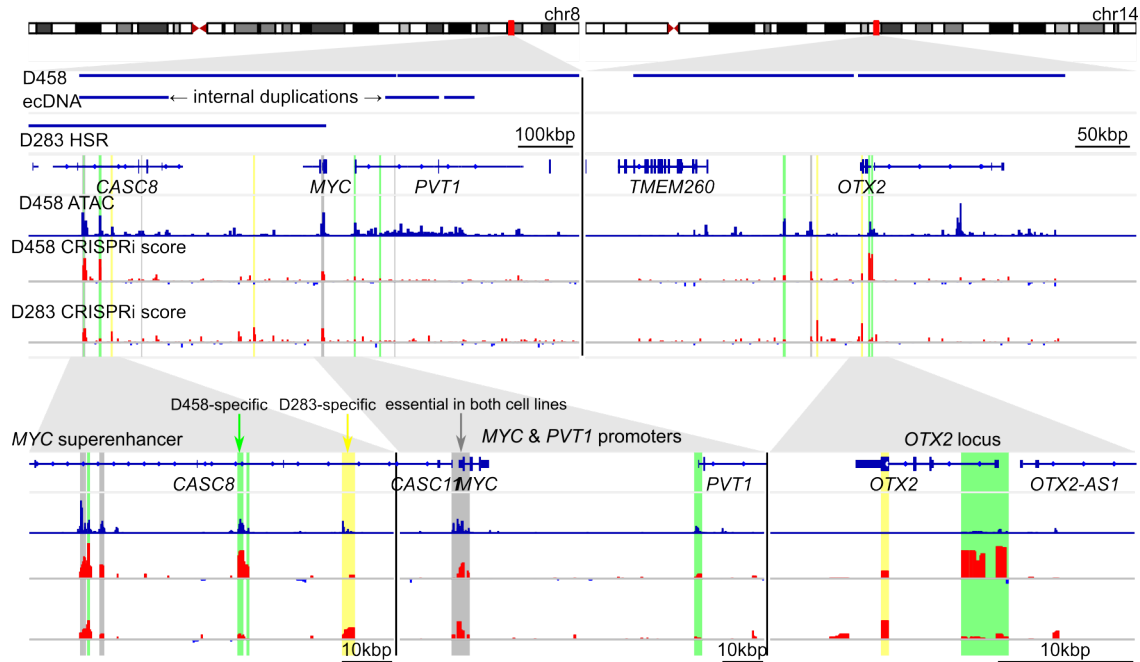
Figure 3.8: **A large CRISPRi screen identifies differentially active enhancers in two *MYC* amplicons.** Pooled CRISPRi screen in the two G3 MB cell lines D458 (*MYC* ecDNA) and D283 (*MYC* HSR) targeted against all accessible loci on the D458 ecDNA as identified by ATAC-seq. Tracks from top to bottom: amplified regions of the D458 ecDNA; amplified regions of the D283 HSR; refSeq genes; D458 bulk ATAC-seq signal; CRISPRi essentiality score tracks for D458 and D283 generated by CRISPR-SURF [43]. Vertical highlighted bars indicate accessible loci which are significantly depleted at T21 relative to T0, and colored by cell line specificity. Grey: essential in D458 and D283 with no significant difference; green: significantly depleted in D458 relative to D283; yellow: significantly depleted in D283 relative to D458. All significance values determined by FDR-corrected MAGeCK-MLE permutation test [70] ($q < 0.05$).

3.3   Discussion

By mapping chromatin interactions occurring on ecDNA in patient and model tumors, we observe ectopic chromatin interactions occurring between genomic loci which are distal on the reference genome but proximal on the ecDNA. In representative tumors of the SHH and Group 3 MB subgroups, these interactions targeted known or putative oncogene loci, and coincided with high gene expression. To determine the extent to which the enhancer landscape of an ecDNA may affect cell proliferation, we performed pooled and targeted CRISPR inhibition experiments in the Group 3 MB cell line D458, which amplifies oncogenes *MYC*, *OTX2*, and *PVT1* on ecDNA.
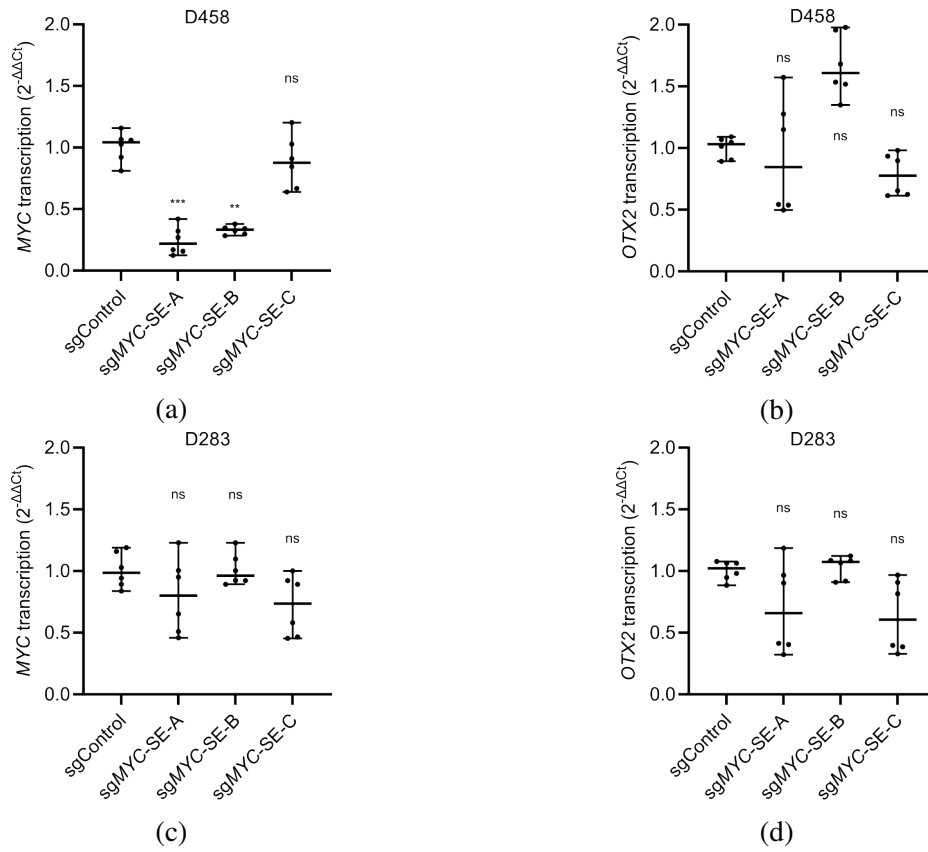
Figure 3.9: **A co-amplified enhancer on the D458 ecDNA promotes *MYC* expression.** Relative expression of *MYC* and *OTX2* in D458 (**a-b**) and D283 (**c-d**), measured by qPCR ($2^{-\Delta\Delta Ct}$), in D458 and D283 cell lines upon CRISPRi targeting of an accessible locus within a known *MYC* superenhancer which promotes D458 proliferation (see also Fig. 3.8). sgNT: nontargeting control; sg*MYC*-SE-A-C: sgRNAs targeting the *MYC* enhancer at D458_peak_30782, positions chr8:127330655, chr8:127330840, and chr8:127330927 respectively. qPCR was performed on all guides in technical triplicate. All co-ordinates refer to the hg38 assembly. ** $p < 0.01$; *** $p < 0.001$; nested ANOVA with Dunnett's correction.

These experiments reveal that a distal *MYC* superenhancer is essential to cell proliferation, and that the relative importance of specific enhancer sub-elements is specific to the sequence of the structural rearrangement. In this illustrative model, the rearranged enhancer landscape of the ecDNA sequence contributes to the cell's proliferative capacity.

Therapeutic targeting of rewired enhancers in patient tumors remains a major unresolved challenge. In our example Group 3 *MYC*-amplified cell lines D283 and D458, the bromodomain inhibitor JQ1 attenuates cell growth by downregulating *MYC* transcription [4]. The mechanism
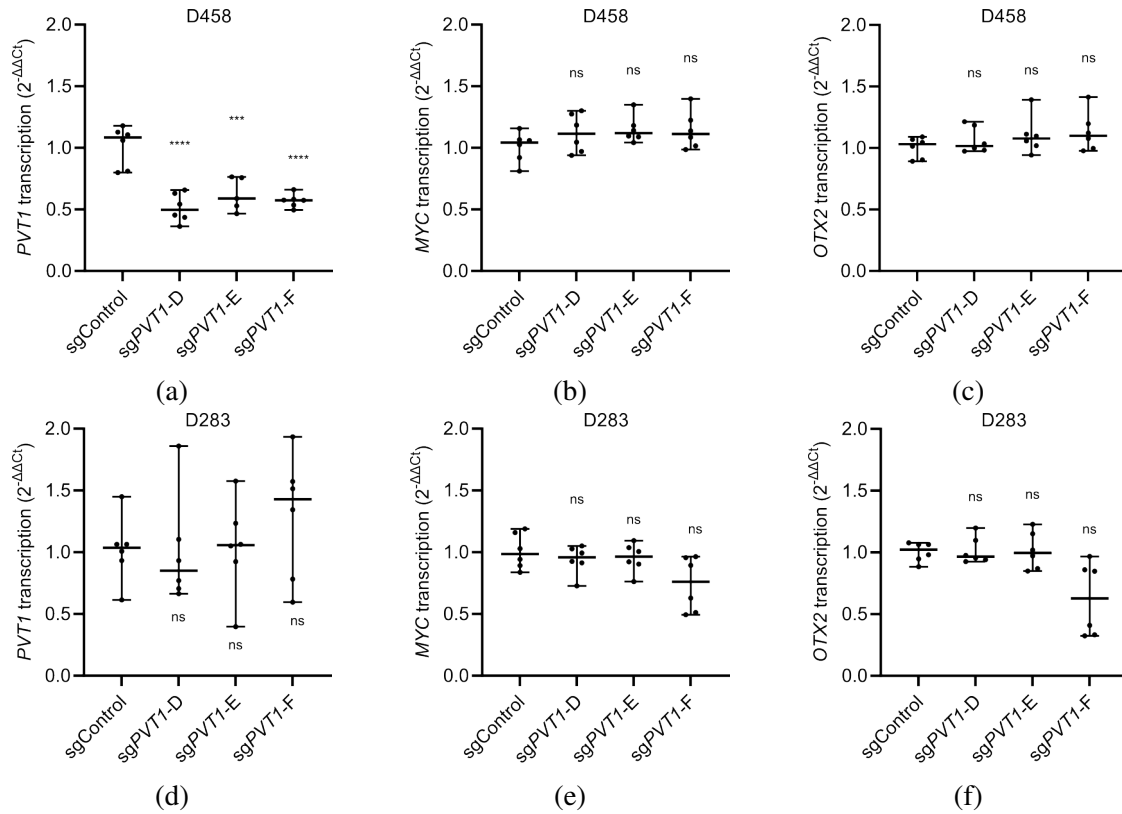
Figure 3.10: *PVT1* **promotes D458 cell proliferation.** Relative expression of *PVT1*, *MYC*, and *OTX2*, measured by qPCR ($2^{-\Delta\Delta Ct}$), in D458 (**a-c**) and D283 (**d-f**) cell lines upon CRISPRi targeting of the *PVT1* promoter which promotes D458 proliferation (see also Fig. 3.8). sgNT: nontargeting control; sg*PVT1*-D-F: sgRNAs targeting the *PVT1* promoter at D458_peak_30920, positions chr8:127794266, chr8:127794773, and chr8:127794945 respectively. qPCR was performed on all guides in triplicate. All co-ordinates refer to the hg38 assembly. *** $p < 0.001$; **** $p < 0.0001$; nested ANOVA with Dunnett's correction.

of action is believed to be disruption of BRD4-mediated binding of transcriptional machinery to *MYC* enhancers, downregulating *MYC* expression. However, clinical trials of BRD inhibitors have so far shown high rates of adverse events and limited efficacy [133]. In addition, further research will be required to resolve and target the exact oncogenic seuqences driving patient tumors such as RCMB56, which does not amplify a known MB oncogene on either ecDNA amplification.

(a)

(b)
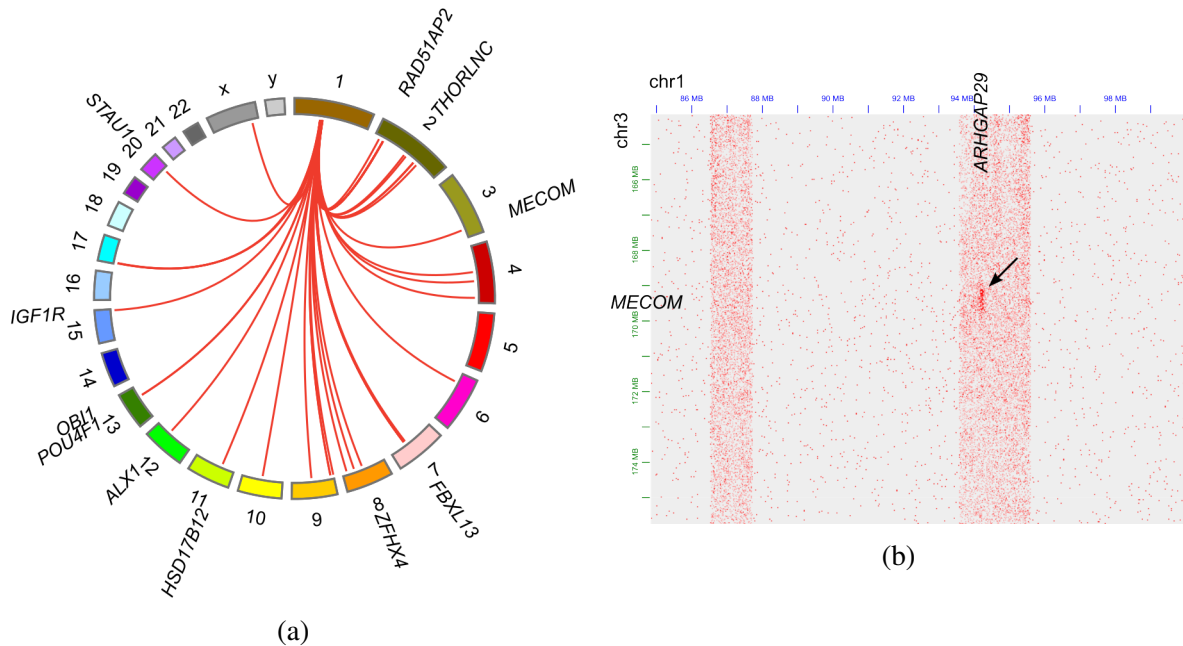
Figure 3.11: **The RCMB56 amp1 ecDNA interacts with chromosomal gene loci.** (**a**) Interchromosomal chromatin interactions detected by FitHiC2 [52] between amp1 and other chromosomes. Interactions mapping to a gene locus are labelled. (**b**) Interaction density map between segments of chr1 and chr3, rendered in Juicebox [24]. Vertical stripes indicate increased contact density between the ecDNA and chr3 over background interactions between chr1 and chr3. Arrow indicates the chromatin loop detected between *ARHGAP29* on amp1 and *MECOM* on chr3.

3.4   Methods

3.4.1   ATAC-seq

Archer *et al.* samples. At least 25mg of pulverized frozen tissue from samples previously included in proteomics analysis5 were sequenced by ATAC-seq according to an established protocol for bulk ATAC-seq of frozen neuronal cells[88]. Reads were aligned, deduplicated and preprocessed according to ENCODE best practices. Samples included in subsequent analyses passed the following quality control thresholds: PCR bottleneck coefficient 1 (PBC1) > 0.7, PCR bottleneck coefficient 2 (PBC2) > 3, non-redundant read fraction (NRF) > 0.8 and TSS enrichment > 2.6. These samples had at least 13 million uniquely mapped single-end reads

(GRCh37). Accessible chromatin regions were identified using MACS2 v2.1.2 [165] using Benjamini-Hochberg corrected p-value threshold $q < 0.05$.

RCMB56 and D458. Frozen dissociated cells were ATAC-sequenced at ActiveMotif, Inc. (San Diego, CA). Briefly, 100K cells were permeabilized and transposase loaded with sequencing adapters was added. Sequencing was performed on Illumina NextSeq 500. Reads were aligned, deduplicated and preprocessed according to ENCODE best practices. Samples had PBC1 > 0.9, PBC2 > 20, NRF > 0.5 and at least 30 million uniquely mapped paired-end reads (hg38). Accessible chromatin regions were identified using MACS2 v2.1.2 [165] using Benjamini-Hochberg corrected p-value threshold $q < 0.05$.

Chromosome conformation capture

Archer et al. samples (except MB248). Hi-C on frozen tumor tissue sample was carried out using protocols previously described for tissue Hi-C experiments88. In brief, frozen tissues are pulverized using a mortar and pestle kept cold on a bed of dry ice into a fine powder. The tissue powder was then transferred to a 15mL conical tube containing 5mLs of DPBS and fixed with 2% formaldehyde for 10 minutes. The fixation was quenched by addition of 0.2M Glycine. The fixed tissue was pelleted by centrifugation, washed 1x with DPBS, and then flash frozen until ready for further processing. For Hi-C experiments, the fixed frozen tissue pellets were first resuspended in 3mLs of lysis buffer (10mM Tris-HCl pH 8.0, 5mM $CaCl_2$, 3mM MgAc, 2mM EDTA, 0.2mM EGTA, 1mM DTT, 0.1mM PMSF, 1X Complete Protease Inhibitors). The sample was transferred to an M-tube and dissociated using a GentleMACS Tissue dissociator (Miltenyi) using the "Protein M-tube" setting. The sample was removed from the M-tube into a 50mL conical. The M-tube was washed with 3mLs of lysis buffer with 0.4% Triton X-100 added, and this wash was combined with the original 3mLs of sample for a total volume of 6mLs with final concentration of 0.2% Triton X-100. The sample was then passed through a 40μM cell strainer. The strainer was washed with an additional 2mLs of lysis buffer with 0.2% Triton X-100.

The sample was then centrifuged and washed with 1mL of lysis buffer with 0.2% Triton X-100. After centrifugation, the sample was resuspended in 0.5% SDS and processed with previously described in situ Hi-C method [110] using the MboI enzyme. Libraries were prepared using the Illumina TruSeq LT sequencing adaptors. Initial QC sequencing was first performed on a MiSeq to assess library quality, and if sufficient, was subject to production scale sequencing on the HiSeq X or NovaSeq platform, respectively.

D458, RCMB56, and MB248. Hi-C experiments were carried out by Arima Genomics, Inc (San Diego, CA) using the Arima-HiC Kit protocol with MboI restriction enzyme. Subsequently, Illumina-compatible sequencing libraries were prepared by shearing the proximally ligated DNA and then size-selecting DNA fragments using SPRI beads. The size-selected fragments containing ligation junctions were enriched using Enrichment Beads (provided in the Arima-HiC Kit), and converted into Illumina-compatible sequencing libraries using the Swift Accel-NGS 2S Plus kit (P/N: 21024) reagents. After adapter ligation, DNA was PCR amplified and purified using SPRI beads. The purified DNA underwent standard QC (qPCR and Bioanalyzer) and sequenced on the NovaSeq following manufacturer's protocols.

3.4.2   Hi-C data processing

Hi-C reads were trimmed using Trimmomatic 0.39 [11] and aligned to the hg38 human genome reference using the HiC-Pro toolkit v2.11.3-beta using default parameters[124]. Visualization and contact normalization was performed with JuiceBox v1.11.08 [24] and the Knight-Ruiz algorithm [56]. Chromatin interactions were called using Juicer Tools GPU HiC-CUPS v1.22.01 [25] using fdr threshold 0.2 and default recommended parameters for Hi-C [110]. From visual inspection, we found that HiCCUPS correctly annotated interactions mapping to ecDNA, except for locus pairs mapping within 50kb of a structural rearrangement. Due to these technical challenges, chromatin interactions described herein were manually curated based on HiCCUPS interaction calls.

3.4.3 Identification of intermolecular chromatin interactions

See also 3.11. To screen for putative intermolecular chromatin interactions originating from possible mobile enhancers on ecDNA [166], we performed loop detection on Hi-C data of RCMB56-pdx using FitHiC v2.0.8 [52] inter-chromosomal mode, at resolution 50kbp and setting no bias upper bound, as recommended by the tool's authors for this task. Interactions with corrected q-values less than 0.05 were selected, then further filtered for loops with one anchor mapping to RCMB56 amp1. To reduce false positive loop calls originating from changes in copy number variation, loops were further filtered to remove those mapping to amp2 or to within 100kbp of a breakpoint on amp1. After filtering, 46 high-confidence loops remained which mapped from amp1 to elsewhere in the reference genome. Genes were associated with a loop if the gene locus overlapped the 50kbp loop anchor. Panel 3.11a was generated using circos v0.69-8 [59].

3.4.4 Pooled CRISPRi proliferation screen

The pooled CRISPRi proliferation screen was designed according to methods described previously [93]. Briefly, sgRNAs targeting the D458 ecDNA regulome were designed using CHOPCHOP [61]. All sgRNAs with no exact homology elsewhere in the genome (ie, mm0 = 0) were included, totaling 32,530 sgRNAs. Targeted regions included 645 accessible regions on the D458 ecDNA, and putative enhancers of *MYC* in medulloblastoma61. As positive controls, sgRNAs targeting the promoter regions of *POLR2A*, *MYC* and *OTX2* were also included. No non-targeting sgRNAs were included; however 4 1kbp loci from chr8 with no evidence of regulatory activity in D458 were also included. Pooled oligonucleotides were generated at Genewiz, Inc. The library was amplified and packaged into lentivirus in 293FT cell line (Invitrogen) by SBP Functional Genomics Core Facility.

D283 Med and D458 Med cells were infected by the lentivirally-packaged sgRNA library

at a MOI of 0.3 at a multiplicity of 1000 cells/sgRNAs. Two independent infections were performed for each cell line as biological replicates (A and B). Forty-eight hours post-infection, infected cells were selected using puromycin (1μg/ml). After 2 days selection, the MOI was measured and 30 million of cells were collected for NGS sequencing (timepoint 0 "T0") as a reference baseline for sgRNA distribution. Another 30 million of cells were re-plated into the culture flasks and maintained for 21 days (timepoint 21 "T21").

To ensure sgRNA representation in each replicate, 30 million cells were maintained throughout the screen during cell passage every 3-4 days. Cells were harvested at T21, and genomic DNA extraction was performed using the NucleoSpin Blood XL kit. The NGS samples were prepared by the SBP Functional Genomics Core Facility using two-step PCR amplification of the integrated sgRNA using the genomic DNA as a template. For the first PCR, "PCR1" sgRNAs were amplified using Herculase II Fusion DNA Polymerases (Agilent 600675) from total genomic DNA isolated from each replicate at T0 and T21. 4μg of genomic DNA was used for each 50ul reaction. PCR1 products were cleaned up using QIAquick PCR Purification Kit (Qiagen). For the second PCR "PCR2", 5 μl of each cleaned PCR1 product was used for each 50μl reaction with primers designed to include Illumina Adaptors and unique barcodes to allow for multiplexed NGS. PCR2 products were bead-purified (AMPure XP, Beckman), and eluted in 30 μl of low-EDTA TE buffer. Final concentrations of the desired products were determined by TapeStation (Agilent) and equimolar amounts from each sample were pooled for Next Generation Sequencing. Following selection, libraries were sequenced on an Illumina NovaSeq 6000 to a minimum depth of 60M reads per sample. Sequencing data QC, alignment and preprocessing was performed using MAGeCK-VISPR v0.4.165 [70] At least 85% of reads per sample were uniquely mappable to a sgRNA sequence, and fewer than 100 sgRNAs (0.3%) were absent from any library. CRISPRi score tracks were generated using CRISPR-SURF v1.06 [43]. To identify D458-specific functional loci, T21 and T0 for both cell lines were compared using the Robust Rank Aggregation (RRA) algorithm of MAGeCK. Essential loci were identified for each

cell line where guides mapping to that locus were depleted at T21 (corrected p-value < 0.05). Independently, D458-specific loci were identified by a linear model parameterized on cell line and timepoint (MAGeCK MLE, corrected p-value < 0.05). The final set of D458-specific functional loci was the intersection of the loci identified by these two analyses. Genomic tracks were plotted in IGV desktop v2.9.2 [116].

3.4.5 Targeted CRISPRi experiments

For CRISPRi experiments, D283 and D458 medulloblastoma cells were lentivirally transduced with dCas9-KRAB-mCherry plasmid [33] (Addgene #60954) to express dCas9 in these cells. Cells stably expressing dCas9 were FACS sorted based on mCherry expression and were transduced with sgRNA vectors. sgRNAs were cloned into the lentiGuide-puro plasmid (Addgene #52963) as described previously [23]. All plasmids were verified by Sanger sequencing. 293T cells (ATCC Cat# CRL-3216) were used to generate lentiviral particles by cotransfecting the packaging vectors psPAX2 and pMD2.G using LipoD293 transfection reagent (SignaGen, SL100668) following the manufacturer's instructions.

3.4.6 Quantitative RT-PCR

Five days after sgRNA transduction, cells from different conditions were collected and Qiagen RNeasy Kit was used to isolate total cellular RNA from cell pellets according to the manufacturer's instructions. iScript cDNA Synthesis Kit (Bio-Rad, catalog 1708890) was used for reverse transcription into cDNA. Quantitative real-time PCR was performed in technical triplicate for 2 bioreplicates of each experimental condition on a Bio-Rad CFX384 Real-Time System using SYBR Green PCR Master Mix (Bio-Rad, catalog 1725270). Gene transcription was estimated from qPCR data using the delta delta Ct method ($Exp = 2^{-\Delta\Delta Ct}$). Testing for change in gene expression was performed using nested ANOVA with Dunnett's multiple comparisons test,

implemented in GraphPad Prism 9.5.2.

3.5   Acknowledgments

Chapter 3, in full, is adapted from the following manuscript currently being prepared for publication: "Circular extrachromosomal DNA promotes inter- and intratumoral heterogeneity in high-risk medulloblastoma." Chapman, Owen; Luebeck, Jens; Sridhar, Sunita; Wong, Ivy

Chapter 4

THE GENOMIC LANDSCAPE OF ECDNA ACROSS PEDIATRIC CANCERS

4.1 Abstract

Circular extrachromosomal DNA (ecDNA) is a driver of oncogene amplification in aggressive cancers. The frequency and diversity of ecDNA has been comprehensively described in many different types of adult cancers and in some selected childhood cancer types, such as neuroblastoma and medulloblastoma. However, the frequency of ecDNA and its association with survival has not been comprehensively analyzed in the spectrum of pediatric cancers. In this study, we used the Amplicon Architect (AA) bioinformatics tool via cloud computing to reconstruct ecDNA sequence structures from Whole Genome Sequencing (WGS) data available in two of the largest publicly accessible pediatric cancer data repositories, the Children's Brain Tumor Network (CBTN) and the St. Jude Cloud cancer genomics platform. In total, we analyzed a retrospective cohort of over 1,600 tumor biopsies from 1,400 different patients, spanning 47 different childhood tumor types. As a result, we identified ecDNA in 15 different pediatric tumor types, with an overall incidence of ecDNA of 10% in the entire patient cohort. We estimate frequencies of

ecDNA in pediatric osteosarcomas (47%), rhabdomyosarcomas (40%), retinoblastomas (19%), ETMRs (all, $n = 4$), high-grade gliomas (20%), and others. Across this cohort, ecDNA was found to be significantly associated with poor survival. Our results show that ecDNA is widely prevalent in pediatric cancers, amplifies known and potentially novel oncogenes, frequently evolves during disease progression, and significantly associates with worse clinical outcomes.

4.2 Introduction

Circular extrachromosomal DNA (ecDNA) has recently emerged as a mechanism for rapid oncogene amplification. Also known as double minutes, ecDNA sequences are typically greater than 1 Mb in size, contain one or more oncogenic sequences, and lack centromeres. ecDNA is subject to non-Mendelian inheritance, leading to unequal segregation of ecDNA during cytokinsesis and consequently intratumoral heterogeneity with every cell division [3]. The mechanisms which promote and maintain ecDNA or which suppress or inhibit ecDNA have yet to be fully elucidated. Several methods have emerged for identification of ecDNA, including microscopic examination with FISH and sequencing-based approaches. Algorithmic assembly from short-read whole genome sequencing has arisen as a cost-effective method to identify ecDNA, and has been applied to estimate the prevalence of ecDNA in adult [55, 103, 79] and some pediatric cancer types [58, 14]. However, we are aware of only isolated meta-analyses [29] and case reports describing ecDNA in other pediatric solid tumors including rhabdomyosarcoma, osteosarcoma, retinoblastoma, and others.

We and others have recently shown that many pediatric cancers harbor unique mutations and driver oncogenes that differ from their adult counterparts [35, 80]. However, ecDNA analysis was not included in these pan-cancer studies, in part because computational methods for reconstructing ecDNA from whole-genome sequencing data were not available at the time. To address this gap, we have now analyzed the frequency of ecDNA in two of the largest publicly available

childhood cancer data repositories. The pediatric pan-cancer study presented here identifies childhood cancer types commonly affected by ecDNA, catalogues ecDNA-amplified genes in primary and recurrent tumors, and examines the association between ecDNA and patient survival. In addition to being an indicator of prognosis, ecDNA provides an opportunity for therapies targeting ecDNA-specific properties, but its real-time clinical applicability has yet to be explored. This study seeks to take the first step in this process by highlighting the spectrum of childhood patients who may benefit from future therapies targeting ecDNA.

4.3 Results

4.3.1 ecDNA occurs in at least 15 distinct pediatric tumor types.

To investigate the prevalence of ecDNA in pediatric tumors, we accessed whole genome sequencing data from two large cloud genomic platforms: St Jude Cloud and Children's Brain Tumor Network (CBTN). In total, we analyzed a retrospective cohort of WGS data of 1670 tumor biopsies from 1481 different patients. Clinical data was available for some patients and included age at diagnosis, sex, survival, tumor tissue site, race/ethnicity, and molecular subgroup for some tumor types. To detect ecDNA, we applied AmpliconArchitect (AA), a tool that detects and reconstructs focal genomic amplifications from whole genome sequencing data [22]. We then applied AmpliconClassifier, which classifies the reconstructed sequence as ecDNA or a separate noncircular amplification. Amplicons were classified as circular, breakage-fusion-bridge, complex noncyclic, or linear (Fig. 4.1a). 174 ecDNA sequences were detected in tumor tissue of 157 out of 1486 patients (10%). This cohort spans 47 different tumor types, 15 of which were found to have ecDNA in at least one sample. The frequencies of ecDNA in these 15 cancer types were as follows: embryonal tumor with multilayered rosettes (ETMR) 4/4 (100%), pineoblastoma (PB) 2/4 (50%), osteosarcoma (OS) 27/57 (47%), rhabdomyosarcoma (RMS) 14/35 (40%), choroid plexus carcinoma (CPC) 1/3 (33%), neuroblastoma (NB) 32/106 (30%),

malignant peripheral nerve sheath tumor (MPNST) 1/4 (25%), high-grade glioma (HGG) 32/157 (20%), retinoblastoma (RB) 6/32 (19%), medulloblastoma (MB) 25/177 (14%), adrenocortical carcinoma (ACC) 2/21 (10%), ependymoma (EPN) 2/73 (3%), craniopharyngioma (CPG) 1/39 (3%), ganglioglioma (GG) 1/44 (2%), and low-grade glioma (LGG) 1/290 (0.3%) (Fig. 4.1b). ecDNA was significantly associated with tumor type ($\chi^2 = 299.96$, p value 2.2×10^{-16}). ecDNA was not identified in the 32 other tumor types in our cohort.

4.3.2   ecDNA status is predictive of 5-year survival across pediatric cancer types.

To evaluate the relationship between ecDNA and survival across pediatric cancer types, we performed Kaplan-Meier regression analyses across 345 patients for whom clinical metadata were available. Patients with ecDNA-containing (ecDNA+) tumors had significantly worse five-year overall and progression free survival compared to patients with ecDNA negative tumors (log-rank-test, $p_{OS} < 0.001$, $p_{PFS} < 0.001$) (Fig. 4.2). Individual survival analyses for each tumor type were not significant, attributable to low statistical power. To further substantiate the association between survival and ecDNA status, we performed a Cox regression analysis with ecDNA as the predictor and age and sex as covariates. ecDNA was significantly associated with poor survival ($HR = 3.17$, $p = 1.24 \times 10^{-9}$). To determine whether ecDNA+ tumors had worse outcomes than those with other forms of focal amplification, we compared survival across different classes of amplification: no amplicon, ecDNA, linear, and highly rearranged amplicons). Patients with cyclic amplification had significantly worse overall survival compared to patients with no amplification and linear amplification (log-rank test, $p_{OS} < 0.001$).

Molecular subgroup annotations were available for 108 for high-grade glioma (HGG) tumors, facilitating comparison of the effect size of ecDNA on survival relative to other molecular determinants of risk. The most current subgrouping of HGG includes classification into histone 3 lysine 27 mutant (H3K27$^{\text{mut}}$), histone 3 glycine 34 mutant (H3G34$^{\text{mut}}$), *IDH* mutant (IDH$^{\text{mut}}$)

(a)

(b)
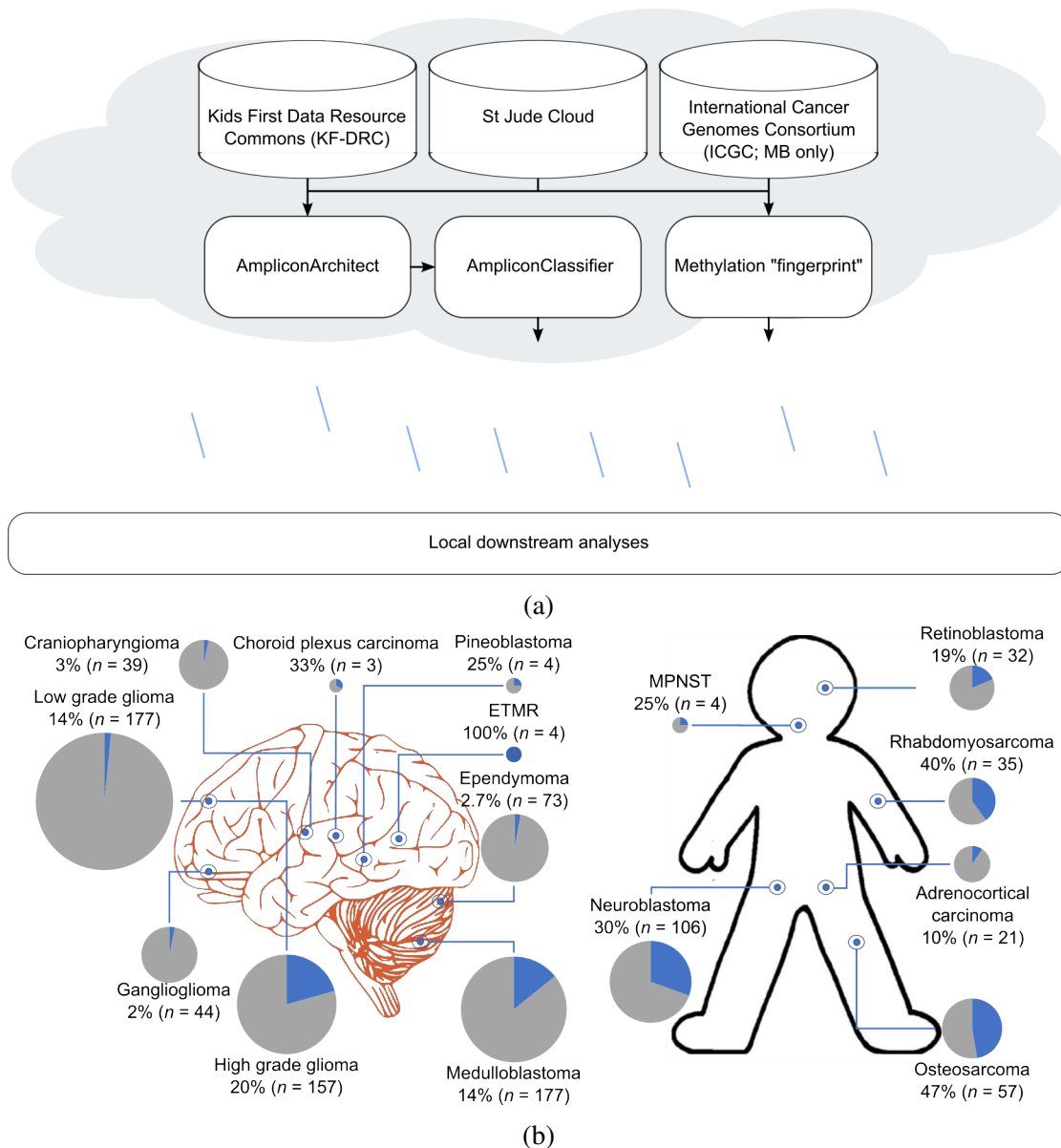
Figure 4.1: **Frequency of ecDNA sequences across a large restrospective cohort of pediatric tumors.** (**a**) Analysis workflow diagram of cloud cancer genomics databases of whole genome sequencing data. (**b**) ecDNA frequencies for select tumor types in this cohort.

and *H3* wild-type *IDH* wild-type (H3/IDH^{mut}) tumors [81]. ecDNA was distributed across these subgroups as follows: H3K27^{mut} (14/56, 25%), H3G34^{mut} (0/4, 0%), IDH^{mut} (2/6, 33%), and H3/IDH^{mut} (11/42, 26%). Cox regression indicates that molecular subgroup, especially H3K27 mutation status, is a major determinant of patient outcome ($HR = 4.39$, $p < 0.001$), consistent

(a)
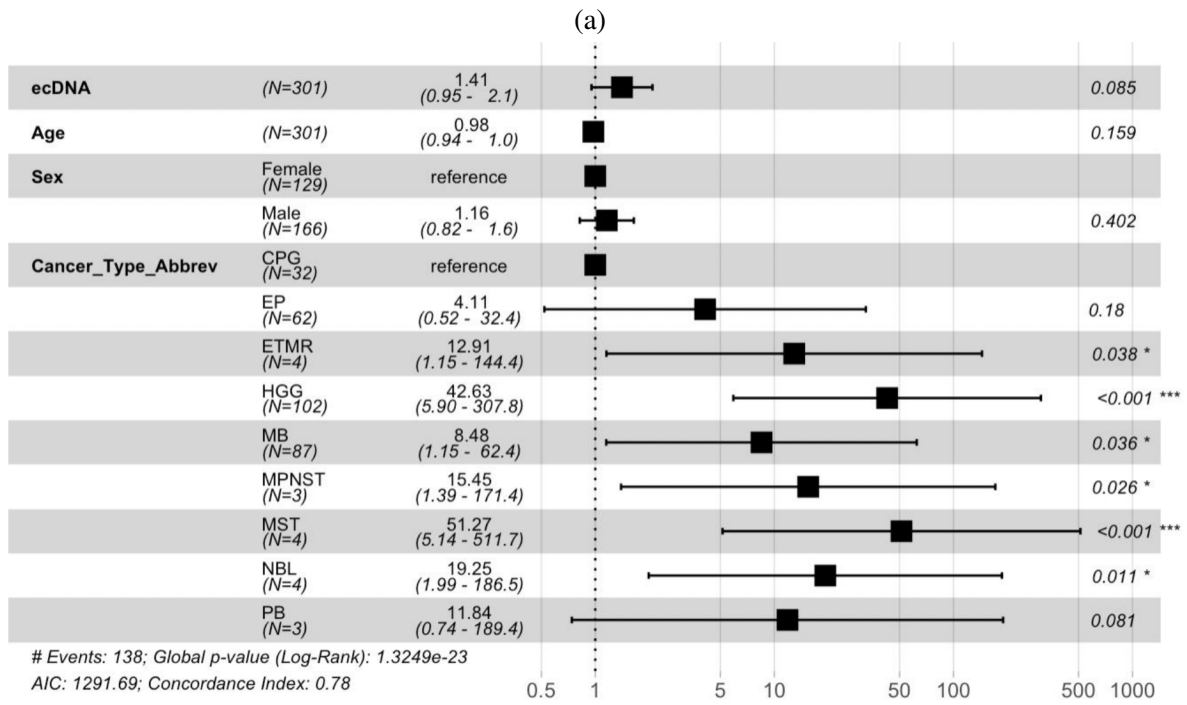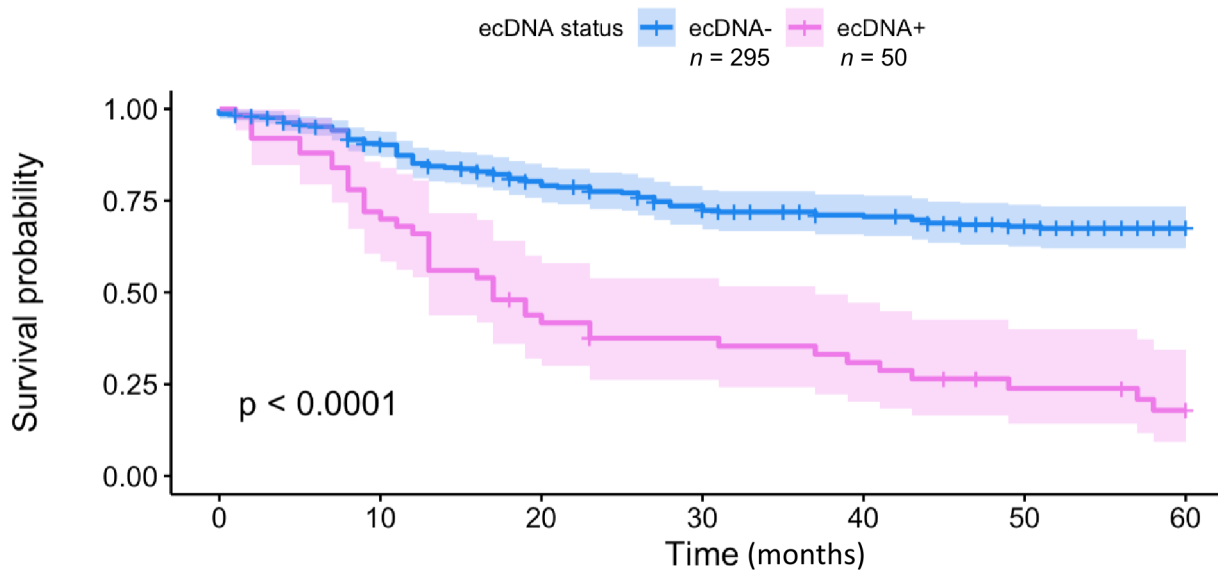
(b)

Figure 4.2: **Survival analysis of a retrospective cohort of 345 patients across _n_ tumor types.**
(**a**) Kaplan-Meier curves indicating 5-year overall survival for ecDNA+ and ecDNA- patients.
(**b**) Cox Proportional Hazards regression coefficients indicating log relative hazards for age, sex,
ecDNA status and tumor type.

with previous results [81]. This model estimates a modest but insignificant effect of ecDNA on patient survival ($HR = 1.13$, $p = 0.6$, Fig. 4.3).
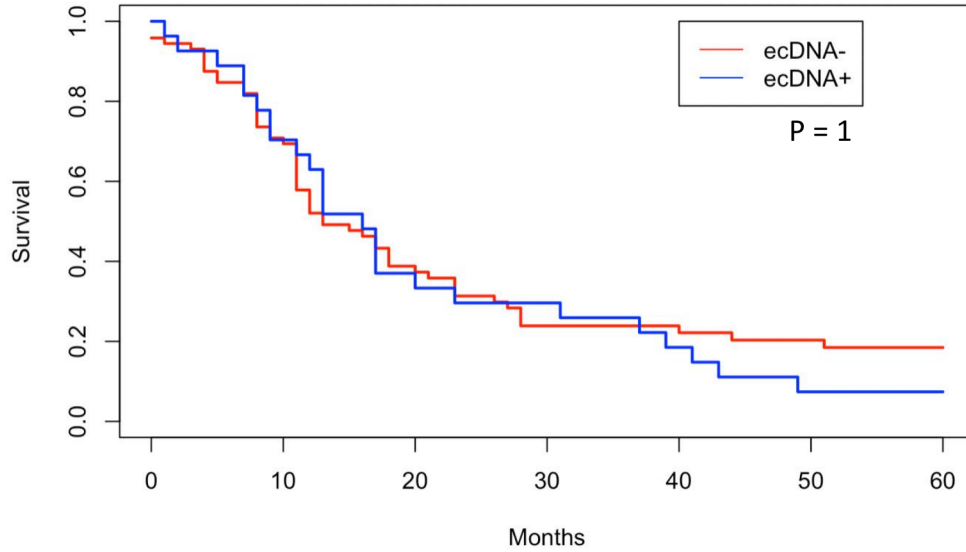
(a)
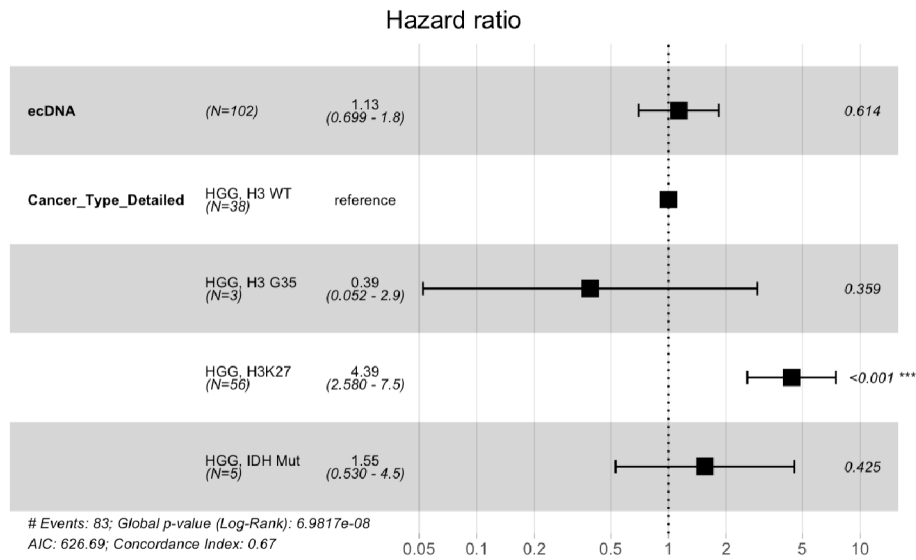
(b)

Figure 4.3: **Survival analysis of a retrospective cohort of 108 pediatric HGG patients.** (a) Kaplan-Meier curves indicating 5-year overall survival for ecDNA+ and ecDNA- patients. (b) Cox Proportional Hazards regression coefficients indicating log relative hazards for ecDNA status and tumor type.

4.3.3 ecDNA most commonly amplifies *MYCN* in neuroblastoma.

The genomic landscape of ecDNA in neuroblastoma has been thoroughly described [58, 164], and the distribution of ecDNA in this neuroblastoma patient cohort is largely consistent with previous observations. We identify circular ecDNA amplification in 32 out of 106 neuroblastomas (30%). ecDNA amplification was almost exclusively detected at the *MYCN* oncogene locus (30 out of 31, 97%).

4.3.4 ecDNA amplifies C19MC-*TTYH1* fusions in embryonal tumor with multilayered rosettes.

Nine out of ten embryonal tumor with multilayered rosettes (ETMR) tumors harbor oncogenic amplification and fusion of *TTYH1* with the C19MC microRNA cluster [69, 62]. Our cohort includes 4 ETMR tumors, all *TTYH1*-C19MC amplified on ecDNA. One of these also had noncircular *MYCN* amplification, suggesting a secondary driver.

4.3.5 *MYC* amplification occurs on ecDNA in pineoblastoma

Pineoblastomas (PBs) comprise five molecular subgroups, including a subgroup harboring recurrent amplification of *MYC* [65]. We identify 2 ecDNA sequences in 4 PB patient tumors, one amplifying *MYC* and the other the C19MC cluster canonically amplified in ETMR. It remains unknown whether the C19MC-amplified PB shares other molecular or clinical features with ETMR tumors.

4.3.6 ecDNA amplifies diverse rhabdomyosarcoma oncogenes including *FOXO1-PAX7* fusions.

ecDNA has been described in rhabdomyosarcoma (RMS) case reports since 1971 [34, 130, 87, 32, 94]. A review of case literature [29] identifies ecDNA in 15% of human rhabdomyosarcomas ($n = 84$). We identify ecDNA in WGS of 14 out of 35 RMS tumors (40%). Amplified oncogenes were *SOX21* (putative, $n = 3$), *MYCN* ($n = 2$), *FGFR1* (2), *MDM2* (2), *NSD3* (2), and *FOXO1-PAX7* fusion (2). One ecDNA sequence did not contain any known oncogenes. Both *MYCN* amplification [157] and *FOXO1* fusion [89] are prognostic indicators of poor RMS patient outcomes; however it remains unknown whether properties of ecDNA contribute to these outcomes.

4.3.7 Half of focal amplifications in osteosarcoma are circular.

Osteosarcomas have among the highest rates of structural variation among pediatric cancer types [159], and over 90% carry inactivating mutations of the p53 pathway [16]. A previous case literature review reported ecDNA in 21% of OS tumors ($n = 146$) [29]. We detect ecDNA in WGS of 27 out of 57 OS patient tumors (47%). The genomic distribution of ecDNA amplifications largely reflected the known regions of recurrent focal amplification in OS, including *MYC* ($n = 4$ patients) *RUNX2* ($n = 3$), and chr17p11.2-p12 ($n = 6$). chr17p11.2-p12 contains unstable low-copy repetitive sequences which are amplified in pediatric osteosarcoma (OS) [143] and MB [14], and adult HGG [144], multiple myeloma (MM) [26], leiomyosarcoma (LMS) and malignant fibrous histiocytoma (MFH) [53]. Genes with putative oncogenic roles within this region include *PMP22*, *COPS3*, *TOP3A*, *MAPK7*, and *MAP2K4* [143].

4.3.8 ecDNA amplifies the known landscape of focal amplifications in pediatric high-grade glioma (HGG)

Among adult patient tumors, ecDNA occurs most frequently in glioblastomas [55]. Reconstruction of circular ecDNA from WGS of a large cohort of adult glioblastomas previously identified recurrent ecDNA amplification of *EGFR*, *MDM2*, *CDK4* and *MYCN* in descending frequency [122]. In a large cohort of pediatric HGG, also included in our dataset, focal amplification was previously reported of oncogenes *PDGFRA*, *MET*, *EGFR*, *MYC*, *MYCN*, Cyclin D family members *CCND1/2/3*, and Cyclin D-dependent kinases *CDK4/6*. The circular structure of one of these amplicons (*EGFR*) was subsequently comprehensively described at diagnosis and relapse [161]. We identify ecDNA in 32 of 157 HGGs (20%). Circular amplicons included all of the oncogenes previously described as amplified in this cohort, establishing circular ecDNA as a frequent form of oncogenic amplification in HGG.

4.3.9 *MYCN* amplification occurs on ecDNA in a high-risk subset of ependymomas.

Amplification of the *MYCN* oncogene defines a subset of ependymoma (EPN) tumors with exceptionally poor prognoses [31]. Given that in neuroblastoma, *MYCN* is almost exclusively amplified on ecDNA and confers poor prognosis, we investigated whether *MYCN* amplification can occur on ecDNA in EPN. We identify 2 ecDNA sequences across 73 EPN tumors. This cohort included a single patient with *MYCN*-amplified spinal EPN, where *MYCN* was amplified on circular ecDNA in both primary and progressive tumors. The other ecDNA comprised amplification of segments of chr13q33, which to our knowledge has not been described in EPN, but is amplified in other cancers [57]. Genes with putative oncogenic roles within this amplicon include *ITGB1L1* [153], *ZIC2* [74] and *ZIC5* [135, 128].

4.3.10   ecDNA amplifies *RB1* fusion genes in RB.

Retinoblastoma (RB) is generally distinguished by early biallelic inactivation of the *RB1* tumor suppressor gene followed by subsequent driver mutations, but 3% of non-hereditary RBs carry wild-type *RB1* and high-copy amplification of *MYCN* [118]. Other high-copy focal amplifications previously reported in RB include *MDM4* and co-amplification of *OTX2* and *AKT1* [120]. Given that *MYCN*-amplified RB bears histological resemblence to neuroblastoma (NB) [118], and that *MYCN* amplifications occur almost exclusively on ecDNA in NB (see 4.3.3), we hypothesized that *MYCN* amplification in RB also occurs on ecDNA. We identify circular ecDNA in 6 of 32 RB patient tumors. Two of these harbored *MYCN* amplification on ecDNA, out of 3 total *MYCN*-amplified RB tumors. Other notable oncogenes amplified on ecDNA included *OTX2* ($n = 1$), *SMARCA5* ($n = 1$), and putative *RB1* fusion genes ($n = 2$): *RB1-THSD1* and *RB1-LHFPL6*. Both amplicons fused the 5' end of *RB1* to the 3' end of the partner gene. No RNA-seq data were available to confirm expression of the putative fusion gene product. However, the recurrent high-copy amplification of truncated *RB1*, and the absence of a high-confidence oncogene on either amplicon, make it tempting to speculate that the *RB1* fusion products provide some other selective advantage beyond *RB1* inactivation.

4.4   Discussion

Our survey of circular ecDNA across a large cohort of many different pediatric tumor types reveals that ecDNA is most frequent in high-risk malignant tumors, and largely absent from low-risk or benign tumors. For example, ecDNA was detected in one in five high-grade gliomas, but only one in 290 low-grade gliomas. The landscape of ecDNA-amplified genes largely reflects previously-known landscapes of gene amplification, indicating that the oncogenic sequence is not only amplified, but circular and acentric as well. Similarly to a recent result in adult cancers, patients with tumors harboring circular acentric amplifications had worse five-year survival than

patients with other forms of amplification.

Recurrently ecDNA-amplified genomic loci in other pediatric tumor types span the spectrum of known oncogenes in pediatric cancer: *MYC* family proto-oncogenes, cyclin-dependent kinases, C19MC, growth factor receptors, and others. We highlight unanswered questions around a few recurrent ecDNA loci as promising areas of future study.

The chromosomal band chr17p11.2-p12 is ecDNA-amplified in 6 OS and 1 MB tumors in our cohort, and its amplification has been described in various adult cancers. Copy number variation in this region has also been shown to cause the childhood developmental disorders Smith-Magenis syndrome (SMS) (deletion) [36] and Charcot-Marie-Tooth disease type 1A (CMT1A) (tandem duplication) [136]. The mechanism underlying both syndromes is believed to be unequal crossing over due to homologous recombination between flanking repeat gene clusters [104]. It remains to be shown whether the ecDNA amplifications at that locus arise by a similar mechanism.

4.5 Methods

4.5.1 Pediatric Tumor WGS

CBTN – Children's Brain Tumor Network (992 biosamples from 831 patients)

WGS of a variety of pediatric tumor biopsies were identified using the Gabriella Miller Kids First Data Resource center portal (https://portal.kidsfirstdrc.org/). Patients were originally sequenced as part of the Pediatric Brain Tumor Atlas (PBTA) (https://cbtn.org/pediatric-brain-tumor-atlas). WGS data were preprocessed and aligned to human reference genome hg38 using the Kids First harmonized WGS pipeline (https://github.com/kids-first/kf-somatic-workflow) and subsequently analyzed using Cavatica (https://cavatica.sbgenomics.com/), on the cloud genomic analysis platform for Kids First genomic data. Docker containers containing fingerprint and

AmpliconArchitect software were installed on the Cavatica cloud genomics platform (see section 4.5.2).

St. Jude – St Jude Cloud Repository (678 biosamples from 655 patients)

WGS of pediatric tumor biopsies were identified using the St Jude Cloud data portal. Patients were sequenced as part of the Pediatric Cancer Genome Project (PCGP) and Real-Time Clinical Genomics (RTCG). AmpliconArchitect software was installed on the SJ Cloud platform (see section 4.5.2).

4.5.2   ecDNA detection and classification from bulk WGS

To detect ecDNA, all samples in the WGS cohort were analyzed using AmpliconArchitect v1.2 [22] and AmpliconClassifier v0.4.4 [55]. Briefly, the AmpliconArchitect algorithm was performed as follows. Copy number segmentation and estimation were performed using CNVkit v0.9.667 [134]. Segments with copy number $CN \geq 4$ were extracted using PrepareAA (April 2020 update) as "seed" regions. For each seed, AmpliconArchitect searches the region and nearby loci for discordant read pairs indicative of genomic structural rearrangement. Genomic segments are defined based on boundaries formed by genomic breakpoint locations (identified by discordant reads) and by modulations in genomic copy number. A breakpoint graph of the amplicon region is constructed using the CN-aware segments and the genomic breakpoints, and cyclic paths are extracted from the graph. Amplicons are classified as cyclic (ecDNA+), breakage-fusion-bridge, complex non-cyclic, linear, or no focal amplification by the heuristic-based companion script, AmpliconClassifier. Biosamples with one or more classifications of "ecDNA"were considered ecDNA+, and all others were considered ecDNA- (Supplementary Table 1). Code is available at:

- PrepareAA: https://github.com/jluebeck/PrepareAA

- AmpliconArchitect: https://github.com/virajbdeshpande/AmpliconArchitect

- AmpliconClassifier: https://github.com/jluebeck/AmpliconClassifier

4.5.3 Patient metadata, survival and subgrouping annotation

Sample metadata was collected where available from the respective cloud genomics data platform: https://pedcbioportal.kidsfirstdrc.org/, https://portal.kidsfirstdrc.org/ (CBTN) and https://pecan.stjude.cloud/ (St Jude).

4.5.4 Survival Analyses

Kaplan-Meier (KM) and Cox Proportional Hazards (CPH) was performed with R with the survfit package.

Kaplan-Meier: Sample size was n = 345 (50 ecDNA+, 295 ecDNA-). Differential survival was determined by log-rank test. Individual analyses by cancer type were too small to test.

Cox Proportional Hazards on age, sex, cancer type and ecDNA: Sample size was n = 301 observations.

Cox Proportional Hazards on age, sex, HGG subgroup and ecDNA: Sample size was n = 102 observations.

4.5.5 Statistical Methods

Statistical test, test statistic and p-values are indicated where appropriate in the main text. Categorical associations were established using the chi-squared test of independence if $N > 5$ for all categories, and the Fisher exact test otherwise. For both tests, the R package stats 3.6.2 was used. Multiple hypothesis corrections were performed using the Bonferroni correction. All statistical tests described herein were two-sided unless otherwise specified.

4.6 Acknowledgments

Chapter 4, in full, is adapted from the following manuscript currently being prepared for publication: "Extrachromosomal DNA promotes oncogene amplification across multiple pediatric cancer types." Sridhar, Sunita; Chapman, Owen S; Dutta, Aditi; Wang, Shanqing; Luebeck, Jens; Bafna, Vineet; Mesirov, Jill P; Chavez, Lukas. The dissertation author was the second investigator and author of the manuscript.

Bibliography

[1] Nuclei isolation from complex tissues for single cell multiome atac + gene expression sequencing. *10X Genomics*, (CG000375 Rev. B), 2021.

[2] T. C. Archer, T. Ehrenberger, F. Mundt, M. P. Gold, K. Krug, C. K. Mah, E. L. Mahoney, C. J. Daniel, A. LeNail, D. Ramamoorthy, P. Mertins, D. R. Mani, H. Zhang, M. A. Gillette, K. Clauser, M. Noble, L. C. Tang, J. Pierre-Francois, J. Silterra, J. Jensen, P. Tamayo, A. Korshunov, S. M. Pfister, M. Kool, P. A. Northcott, R. C. Sears, J. O. Lipton, S. A. Carr, J. P. Mesirov, S. L. Pomeroy, and E. Fraenkel. Proteomics, post-translational modifications, and integrative analyses reveal molecular heterogeneity within medulloblastoma subgroups. *Cancer Cell*, 34(3):396–410.e8, 2018.

[3] V. Bafna and P. S. Mischel. Extrachromosomal dna in cancer. *Annual Review of Genomics and Human Genetics*, 23(1):29–52, 2022.

[4] P. Bandopadhayay, G. Bergthold, B. Nguyen, S. Schubert, S. Gholamin, Y. Tang, S. Bolin, S. E. Schumacher, R. Zeid, S. Masoud, F. Yu, N. Vue, W. J. Gibson, B. R. Paolella, S. S. Mitra, S. H. Cheshier, J. Qi, K. W. Liu, R. Wechsler-Reya, W. A. Weiss, F. J. Swartling, M. W. Kieran, J. E. Bradner, R. Beroukhim, and Y. J. Cho. Bet bromodomain inhibition of myc-amplified medulloblastoma. *Clin Cancer Res*, 20(4):912–25, 2014.

[5] S. Banerjee, T. Wei, J. Wang, J. J. Lee, H. L. Gutierrez, O. Chapman, S. E. Wiley, J. E. Mayfield, V. Tandon, E. F. Juarez, L. Chavez, R. Liang, R. L. Sah, C. Costello, J. P. Mesirov, L. de la Vega, K. L. Cooper, J. E. Dixon, J. Xiao, and X. Lei. Inhibition of dual-specificity tyrosine phosphorylation-regulated kinase 2 perturbs 26s proteasome-addicted neoplastic progression. *Proceedings of the National Academy of Sciences*, 116(49):24881–24891, 2019.

[6] D. A. Barbie, P. Tamayo, J. S. Boehm, S. Y. Kim, S. E. Moody, I. F. Dunn, A. C. Schinzel, P. Sandy, E. Meylan, C. Scholl, S. Fröhling, E. M. Chan, M. L. Sos, K. Michel, C. Mermel, S. J. Silver, B. A. Weir, J. H. Reiling, Q. Sheng, P. B. Gupta, R. C. Wadlow, H. Le, S. Hoersch, B. S. Wittner, S. Ramaswamy, D. M. Livingston, D. M. Sabatini, M. Meyerson, R. K. Thomas, E. S. Lander, J. P. Mesirov, D. E. Root, D. G. Gilliland, T. Jacks, and W. C. Hahn. Systematic rna interference reveals that oncogenic kras-driven cancers require tbk1. *Nature*, 462(7269):108–112, 2009.

[7] N. Barkas, V. Petukhov, D. Nikolaeva, Y. Lozinsky, S. Demharter, K. Khodosevich, and P. V. Kharchenko. Joint analysis of heterogeneous single-cell rna-seq dataset collections. *Nature Methods*, 16(8):695–698, 2019.

[8] R. M. Baron and D. A. Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51:1173–1182, 1986.

[9] F. Bellutti, A.-S. Tigan, S. Nebenfuehr, M. Dolezal, M. Zojer, R. Grausenburger, S. Hartenberger, S. Kollmann, E. Doma, M. Prchal-Murphy, I. Z. Uras, A. Höllein, D. S. Neuberg, B. L. Ebert, A. Ringler, A. C. Mueller, J. I. Loizou, P. W. Hinds, C. Vogl, G. Heller, S. Kubicek, J. Zuber, M. Malumbres, M. Farlik, A. Villunger, K. Kollmann, and V. Sexl. Cdk6 antagonizes p53-induced responses during tumorigenesis. *Cancer Discovery*, 8(7):884, 2018.

[10] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[11] A. M. Bolger, M. Lohse, and B. Usadel. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 2014.

[12] J. D. Buenrostro, B. Wu, U. M. Litzenburger, D. Ruff, M. L. Gonzales, M. P. Snyder, H. Y. Chang, and W. J. Greenleaf. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 523(7561):486–490, 2015.

[13] P. J. Campbell, G. Getz, J. O. Korbel, J. M. Stuart, J. L. Jennings, L. D. Stein, M. D. Perry, H. K. Nahal-Bose, B. F. F. Ouellette, C. H. Li, E. Rheinbay, G. P. Nielsen, D. C. Sgroi, C.-L. Wu, W. C. Faquin, V. Deshpande, P. C. Boutros, A. J. Lazar, K. A. Hoadley, D. N. Louis, L. J. Dursi, C. K. Yung, M. H. Bailey, G. Saksena, K. M. Raine, I. Buchhalter, K. Kleinheinz, M. Schlesner, J. Zhang, W. Wang, D. A. Wheeler, L. Ding, J. T. Simpson, B. D. O'Connor, S. Yakneen, K. Ellrott, N. Miyoshi, A. P. Butler, R. Royo, S. I. Shorser, M. Vazquez, T. Rausch, G. Tiao, S. M. Waszak, B. Rodriguez-Martin, S. Shringarpure, D.-Y. Wu, G. M. Demidov, O. Delaneau, S. Hayashi, S. Imoto, N. Habermann, A. V. Segre, E. Garrison, A. Cafferkey, E. G. Alvarez, J. M. Heredia-Genestar, F. Muyas, O. Drechsel, A. L. Bruzos, J. Temes, J. Zamora, A. Baez-Ortega, H.-L. Kim, R. J. Mashl, K. Ye, A. DiBiase, K.-l. Huang, I. Letunic, M. D. McLellan, S. J. Newhouse, T. Shmaya, S. Kumar, D. C. Wedge, M. H. Wright, V. D. Yellapantula, M. Gerstein, E. Khurana, T. Marques-Bonet, A. Navarro, C. D. Bustamante, R. Siebert, H. Nakagawa, D. F. Easton, S. Ossowski, J. M. C. Tubio, F. M. De La Vega, X. Estivill, D. Yuen, G. L. Mihaiescu, L. Omberg, V. Ferretti, R. Sabarinathan, O. Pich, A. Gonzalez-Perez, A. Taylor-Weiner, M. W. Fittall, J. Demeulemeester, M. Tarabichi, N. D. Roberts, et al. Pan-cancer analysis of whole genomes. *Nature*, 578(7793):82–93, 2020.

[14] O. S. Chapman, J. Luebeck, S. Wani, A. Tiwari, M. Pagadala, S. Wang, J. D. Larson, J. T. Lange, I. T.-L. Wong, S. R. Dehkordi, S. Chandran, M. Adam, Y. Lin, E. Juarez, J. T.

Robinson, S. Sridhar, D. M. Malicki, N. Coufal, M. Levy, J. R. Crawford, S. L. Pomeroy, J. Rich, R. H. Scheuermann, H. Carter, J. Dixon, P. S. Mischel, E. Fraenkel, R. J. Wechsler-Reya, V. Bafna, J. P. Mesirov, and L. Chavez. The landscape of extrachromosomal circular dna in medulloblastoma. *bioRxiv*, 2021.

[15] S. Cheloni, R. Hillje, L. Luzi, P. G. Pelicci, and E. Gatti. Xenocell: classification of cellular barcodes in single cell experiments from xenograft samples. *BMC Medical Genomics*, 14(1):34, 2021.

[16] X. Chen, A. Bahrami, A. Pappo, J. Easton, J. Dalton, E. Hedlund, D. Ellison, S. Shurtleff, G. Wu, L. Wei, M. Parker, M. Rusch, P. Nagahawatte, J. Wu, S. Mao, K. Boggs, H. Mulder, D. Yergeau, C. Lu, L. Ding, M. Edmonson, C. Qu, J. Wang, Y. Li, F. Navid, N. C. Daw, E. Mardis, R. Wilson, J. Downing, J. Zhang, and M. Dyer. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Reports*, 7(1):104–112, 2014.

[17] D. Cox, C. Yuncken, and A. Spriggs. Minute chromatin bodies in malignant tumours of childhood. *The Lancet*, 286(7402):55–58, 1965.

[18] C. Davidson-Pilon. lifelines: survival analysis in python. *Journal of Open Source Software*, 4(40):1317, 2019.

[19] A. C. deCarvalho, H. Kim, L. M. Poisson, M. E. Winn, C. Mueller, D. Cherba, J. Koeman, S. Seth, A. Protopopov, M. Felicella, S. Zheng, A. Multani, Y. Jiang, J. Zhang, D. H. Nam, E. F. Petricoin, L. Chin, T. Mikkelsen, and R. G. W. Verhaak. Discordant inheritance of chromosomal and extrachromosomal dna elements contributes to dynamic disease evolution in glioblastoma. *Nat Genet*, 50(5):708–717, 2018.

[20] J. DeMeritt. Parametric models, 2022.

[21] M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly. A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nat Genet*, 43(5):491–8, 2011.

[22] V. Deshpande, J. Luebeck, N.-P. D. Nguyen, M. Bakhtiari, K. M. Turner, R. Schwab, H. Carter, P. S. Mischel, and V. Bafna. Exploring the landscape of focal amplifications in cancer using ampliconarchitect. *Nature Communications*, 10(1):392, 2019.

[23] D. Dixit, B. C. Prager, R. C. Gimple, H. X. Poh, Y. Wang, Q. Wu, Z. Qiu, R. L. Kidwell, L. J. Y. Kim, Q. Xie, K. Vitting-Seerup, S. Bhargava, Z. Dong, L. Jiang, Z. Zhu, P. Hamerlik, S. R. Jaffrey, J. C. Zhao, X. Wang, and J. N. Rich. The rna m6a reader ythdf2 maintains oncogene expression and is a targetable dependency in glioblastoma stem cells. *Cancer Discov*, 11(2):480–499, 2021.

[24] N. C. Durand, J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, and E. L. Aiden. Juicebox provides a visualization system for hi-c contact maps with unlimited zoom. *Cell systems*, 3(1):99–101, 2016.

[25] N. C. Durand, M. S. Shamim, I. Machol, S. S. Rao, M. H. Huntley, E. S. Lander, and E. L. Aiden. Juicer provides a one-click system for analyzing loop-resolution hi-c experiments. *Cell Syst*, 3(1):95–8, 2016.

[26] S. Fabris, K. Todoerti, L. Mosca, L. Agnelli, D. Ronchetti, D. Intini, M. Lionetti, G. Ciceri, G. L. Deliliers, F. Bertoni, and A. Neri. Molecular and transcriptional characterization of the novel 17p11.2-p12 chromosome amplification in multiple myeloma. *Blood*, 110(11):2486–2486, 2007.

[27] S. Fairley, E. Lowy-Gallego, E. Perry, and P. Flicek. The international genome sample resource (igsr) collection of open human genomic variation resources. *Nucleic Acids Research*, 48(D1):D941–D947, 2020.

[28] A. Garancher, C. Y. Lin, M. Morabito, W. Richer, N. Rocques, M. Larcher, L. Bihannic, K. Smith, C. Miquel, S. Leboucher, N. I. Herath, F. Dupuy, P. Varlet, C. Haberler, C. Walczak, N. El Tayara, A. Volk, S. Puget, F. Doz, O. Delattre, S. Druillennec, O. Ayrault, R. J. Wechsler-Reya, A. Eychène, F. Bourdeaut, P. A. Northcott, and C. Pouponnot. Nrl and crx define photoreceptor identity and reveal subgroup-specific dependencies in medulloblastoma. *Cancer Cell*, 33(3):435–449.e6, 2018.

[29] E. Gebhart. Double minutes, cytogenetic equivalents of gene amplification, in human neoplasia—a review. *Clinical and Translational Oncology*, 7(11):477–485, 2005.

[30] D. M. A. Gendoo and B. Haibe-Kains. Mm2s: personalized diagnosis of medulloblastoma patients and model systems. *Source Code for Biology and Medicine*, 11(1):6, 2016.

[31] D. R. Ghasemi, M. Sill, K. Okonechnikov, A. Korshunov, S. Yip, P. W. Schutz, D. Scheie, A. Kruse, P. N. Harter, M. Kastelan, M. Wagner, C. Hartmann, J. Benzel, K. K. Maass, M. Khasraw, R. Sträter, C. Thomas, W. Paulus, C. P. Kratz, H. Witt, D. Kawauchi, C. Herold-Mende, F. Sahm, S. Brandner, M. Kool, D. T. W. Jones, A. von Deimling, S. M. Pfister, D. E. Reuss, and K. W. Pajtler. Mycn amplification drives an aggressive form of spinal ependymoma. *Acta Neuropathologica*, 138(6):1075–1089, 2019.

[32] R. Gil-Benso, C. López-Ginés, C. Carda, J. López-Guerrero, J. Ferrer, Pellı, x, A. n Pérez, and A. Llombart–Bosch. Cytogenetic and molecular findings related to rhabdomyosarcoma. an analysis of seven cases. *Cancer Genetics and Cytogenetics*, 144(2):125–133, 2003.

[33] L. A. Gilbert, M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C. Guimaraes, B. Panning, H. L. Ploegh, M. C. Bassik, L. S. Qi, M. Kampmann, and J. S. Weissman. Genome-scale crispr-mediated control of gene repression and activation. *Cell*, 159(3):647–61, 2014.

[34] I. Granberg and J. Mark. The chromosomal aberration of double-minutes in a human embryonic rhabdomyosarcoma. *Acta Cytol*, 15(1):42–5, 1971.

[35] S. N. Gröbner, B. C. Worst, J. Weischenfeldt, I. Buchhalter, K. Kleinheinz, V. A. Rudneva, P. D. Johann, G. P. Balasubramanian, M. Segura-Wang, S. Brabetz, S. Bender, B. Hutter, D. Sturm, E. Pfaff, D. Hübschmann, G. Zipprich, M. Heinold, J. Eils, C. Lawerenz, S. Erkek, S. Lambo, S. Waszak, C. Blattmann, A. Borkhardt, M. Kuhlen, A. Eggert, S. Fulda, M. Gessler, J. Wegert, R. Kappler, D. Baumhoer, S. Burdach, R. Kirschner-Schwabe, U. Kontny, A. E. Kulozik, D. Lohmann, S. Hettmer, C. Eckert, S. Bielack, M. Nathrath, C. Niemeyer, G. H. Richter, J. Schulte, R. Siebert, F. Westermann, J. J. Molenaar, G. Vassal, H. Witt, B. Burkhardt, C. P. Kratz, O. Witt, C. M. van Tilburg, C. M. Kramm, G. Fleischhack, U. Dirksen, S. Rutkowski, M. Frühwald, K. von Hoff, S. Wolf, T. Klingebiel, E. Koscielniak, P. Landgraf, J. Koster, A. C. Resnick, J. Zhang, Y. Liu, X. Zhou, A. J. Waanders, D. A. Zwijnenburg, P. Raman, B. Brors, U. D. Weber, P. A. Northcott, K. W. Pajtler, M. Kool, R. M. Piro, J. O. Korbel, M. Schlesner, R. Eils, D. T. W. Jones, P. Lichter, L. Chavez, M. Zapatka, and S. M. Pfister. The landscape of genomic alterations across childhood cancers. *Nature*, 555(7696):321–327, 2018.

[36] A. L. Gropman, W. C. Duncan, and A. C. M. Smith. Neurologic and developmental features of the smith-magenis syndrome (del 17p11.2). *Pediatric Neurology*, 34(5):337–350, 2006.

[37] X. Gu, J. Yu, P. Chai, S. Ge, and X. Fan. Novel insights into extrachromosomal dna: redefining the onco-drivers of tumor progression. *Journal of Experimental & Clinical Cancer Research*, 39(1):215, 2020.

[38] C. Hafemeister and R. Satija. Normalization and variance stabilization of single-cell rna-seq data using regularized negative binomial regression. *Genome Biology*, 20(1):296, 2019.

[39] D. Hanahan and R. Weinberg. Hallmarks of cancer: The next generation. *Cell*, 144(5):646–674, 2011.

[40] Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, and R. Satija. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587.e29, 2021.

[41] A. P. Heath, D. M. Taylor, Y. Zhu, P. Raman, J. Lilly, P. Storm, A. J. Waanders, V. Ferretti, C. Yung, M. Mattioni, B. Davis-Dusenbery, Z. L. Flamig, R. Grossman, S. L. Volchenboum, S. Mueller, J. Nazarian, N. Vasilevsky, M. A. Haendel, and A. Resnick. Abstract 2464: Gabriella miller kids first data resource center: Harmonizing clinical and genomic data to support childhood cancer and structural birth defect research. *Cancer Research*, 79(13 Supplement):2464, 2019.

[42] K. Helmsauer, M. E. Valieva, S. Ali, R. Chamorro González, R. Schöpflin, C. Röefzaad, Y. Bei, H. Dorado Garcia, E. Rodriguez-Fos, M. Puiggròs, K. Kasack, K. Haase, C. Keskeny, C. Y. Chen, L. P. Kuschel, P. Euskirchen, V. Heinrich, M. I. Robson, C. Rosswog, J. Toedling, A. Szymansky, F. Hertwig, M. Fischer, D. Torrents, A. Eggert, J. H. Schulte, S. Mundlos, A. G. Henssen, and R. P. Koche. Enhancer hijacking determines extrachromosomal circular mycn amplicon architecture in neuroblastoma. *Nature Communications*, 11(1):5823, 2020.

[43] J. Y. Hsu, C. P. Fulco, M. A. Cole, M. C. Canver, D. Pellin, F. Sher, R. Farouni, K. Clement, J. A. Guo, L. Biasco, S. H. Orkin, J. M. Engreitz, E. S. Lander, J. K. Joung, D. E. Bauer, and L. Pinello. Crispr-surf: discovering regulatory elements by deconvolution of crispr tiling screen data. *Nature Methods*, 15(12):992–993, 2018.

[44] K. L. Hung, J. Luebeck, S. R. Dehkordi, C. I. Colón, R. Li, I. T.-L. Wong, C. Coruh, P. Dharanipragada, S. H. Lomeli, N. E. Weiser, G. Moriceau, X. Zhang, C. Bailey, K. E. Houlahan, W. Yang, R. C. González, C. Swanton, C. Curtis, M. Jamal-Hanjani, A. G. Henssen, J. A. Law, W. J. Greenleaf, R. S. Lo, P. S. Mischel, V. Bafna, and H. Y. Chang. Targeted profiling of human extrachromosomal dna by crispr-catch. *Nature Genetics*, 2022.

[45] K. L. Hung, K. E. Yost, L. Xie, Q. Shi, K. Helmsauer, J. Luebeck, R. Schöpflin, J. T. Lange, R. Chamorro González, N. E. Weiser, C. Chen, M. E. Valieva, I. T.-L. Wong, S. Wu, S. R. Dehkordi, C. V. Duffy, K. Kraft, J. Tang, J. A. Belk, J. C. Rose, M. R. Corces, J. M. Granja, R. Li, U. Rajkumar, J. Friedlein, A. Bagchi, A. T. Satpathy, R. Tjian, S. Mundlos, V. Bafna, A. G. Henssen, P. S. Mischel, Z. Liu, and H. Y. Chang. ecdna hubs drive cooperative intermolecular oncogene expression. *Nature*, 600(7890):731–736, 2021.

[46] N. M. Ioannidis, J. H. Rothstein, V. Pejaver, S. Middha, S. K. McDonnell, S. Baheti, A. Musolf, Q. Li, E. Holzinger, D. Karyadi, L. A. Cannon-Albright, C. C. Teerlink, J. L. Stanford, W. B. Isaacs, J. Xu, K. A. Cooney, E. M. Lange, J. Schleutker, J. D. Carpten, I. J. Powell, O. Cussenot, G. Cancel-Tassin, G. G. Giles, R. J. MacInnis, C. Maier, C.-L. Hsieh, F. Wiklund, W. J. Catalona, W. D. Foulkes, D. Mandal, R. A. Eeles, Z. Kote-Jarai, C. D. Bustamante, D. J. Schaid, T. Hastie, E. A. Ostrander, J. E. Bailey-Wilson, P. Radivojac, S. N. Thibodeau, A. S. Whittemore, and W. Sieh. Revel: An ensemble method for predicting the pathogenicity of rare missense variants. *The American Journal of Human Genetics*, 99(4):877–885, 2016.

[47] D. P. Ivanov, B. Coyle, D. A. Walker, and A. M. Grabowska. In vitro models of medulloblastoma: Choosing the right tool for the job. *Journal of Biotechnology*, 236:10–25, 2016.

[48] S. Y. Jung, M. Rosenzweig, F. Linkov, A. Brufsky, J. L. Weissfeld, and S. M. Sereika. Comorbidity as a mediator of survival disparity between younger and older women diagnosed with metastatic breast cancer. *Hypertension*, 59(2):205–211, 2012.

[49] K. Juraschka and M. D. Taylor. Medulloblastoma in the age of molecular subgroups: a review. *Journal of Neurosurgery: Pediatrics PED*, 24(4):353–363, 2019.

[50] K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. M. England, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski, C. A. Aguilar Salinas, T. Ahmad, C. M. Albert, D. Ardissino, G. Atzmon, J. Barnard, L. Beaugerie, E. J. Benjamin, M. Boehnke, L. L. Bonnycastle, E. P. Bottinger, D. W. Bowden, M. J. Bown, J. C. Chambers, J. C. Chan, D. Chasman, J. Cho, M. K. Chung, B. Cohen, A. Correa, D. Dabelea, M. J. Daly, D. Darbar, R. Duggirala, J. Dupuis, P. T. Ellinor, R. Elosua, J. Erdmann, T. Esko, M. Färkkilä, J. Florez, A. Franke, G. Getz, B. Glaser, S. J. Glatt, D. Goldstein, C. Gonzalez, L. Groop, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809):434–443, 2020.

[51] M. Karlsson, C. Zhang, L. Méar, W. Zhong, A. Digre, B. Katona, E. Sjöstedt, L. Butler, J. Odeberg, P. Dusart, F. Edfors, P. Oksvold, K. von Feilitzen, M. Zwahlen, M. Arif, O. Altay, X. Li, M. Ozcan, A. Mardinoglu, L. Fagerberg, J. Mulder, Y. Luo, F. Ponten, M. Uhlén, and C. Lindskog. A single–cell type transcriptomics map of human tissues. *Science Advances*, 7(31):eabh2169.

[52] A. Kaul, S. Bhattacharyya, and F. Ay. Identifying statistically significant chromatin contacts from hi-c data with fithic2. *Nature Protocols*, 15(3):991–1012, 2020.

[53] S. Kaur, M. L. Larramendy, H. Vauhkonen, T. Böhling, and S. Knuutila. Loss of tp53 in sarcomas with 17p12 p11 gain. a fine-resolution oligonucleotide array comparative genomic hybridization study. *Cytogenetic and Genome Research*, 116(3):153–157, 2007.

[54] W. J. Kent, A. S. Zweig, G. Barber, A. S. Hinrichs, and D. Karolchik. Bigwig and bigbed: enabling browsing of large distributed datasets. *Bioinformatics (Oxford, England)*, 26(17):2204–2207, 2010.

[55] H. Kim, N.-P. Nguyen, K. Turner, S. Wu, A. D. Gujar, J. Luebeck, J. Liu, V. Deshpande, U. Rajkumar, S. Namburi, S. B. Amin, E. Yi, F. Menghi, J. H. Schulte, A. G. Henssen, H. Y. Chang, C. R. Beck, P. S. Mischel, V. Bafna, and R. G. W. Verhaak. Extrachromosomal dna is associated with oncogene amplification and poor outcome across multiple cancers. *Nature Genetics*, 52(9):891–897, 2020.

[56] P. A. Knight and D. Ruiz. A fast algorithm for matrix balancing. *IMA Journal of Numerical Analysis*, 33(3):1029–1047, 2013.

[57] S. Knuutila, A. M. Björkqvist, K. Autio, M. Tarkkanen, M. Wolf, O. Monni, J. Szymanska, M. L. Larramendy, J. Tapper, H. Pere, W. El-Rifai, S. Hemmer, V. M. Wasenius, V. Vidgren, and Y. Zhu. Dna copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am J Pathol*, 152(5):1107–23, 1998.

[58] R. P. Koche, E. Rodriguez-Fos, K. Helmsauer, M. Burkert, I. C. MacArthur, J. Maag, R. Chamorro, N. Munoz-Perez, M. Puiggròs, H. Dorado Garcia, Y. Bei, C. Röefzaad, V. Bardinet, A. Szymansky, A. Winkler, T. Thole, N. Timme, K. Kasack, S. Fuchs, F. Klironomos, N. Thiessen, E. Blanc, K. Schmelz, A. Künkele, P. Hundsdörfer, C. Rosswog, J. Theissen, D. Beule, H. Deubzer, S. Sauer, J. Toedling, M. Fischer, F. Hertwig, R. F. Schwarz, A. Eggert, D. Torrents, J. H. Schulte, and A. G. Henssen. Extrachromosomal circular dna drives oncogenic genome remodeling in neuroblastoma. *Nat Genet*, 52(1):29–34, 2020.

[59] M. I. Krzywinski, J. E. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. Circos: An information aesthetic for comparative genomics. *Genome Research*, 2009.

[60] P. Kumar, S. Kiran, S. Saha, Z. Su, T. Paulsen, A. Chatrath, Y. Shibata, E. Shibata, and A. Dutta. Atac-seq identifies thousands of extrachromosomal circular dna in cancer and cell lines. *Science Advances*, 6(20):eaba2489, 2020.

[61] K. Labun, T. G. Montague, M. Krause, Y. N. Torres Cleuren, H. Tjeldnes, and E. Valen. Chopchop v3: expanding the crispr web toolbox beyond genome editing. *Nucleic Acids Research*, 47(W1):W171–W174, 2019.

[62] S. Lambo, S. N. Gröbner, T. Rausch, S. M. Waszak, C. Schmidt, A. Gorthi, J. C. Romero, M. Mauermann, S. Brabetz, S. Krausert, I. Buchhalter, J. Koster, D. A. Zwijnenburg, M. Sill, J.-M. Hübner, N. Mack, B. Schwalm, M. Ryzhova, V. Hovestadt, S. Papillon-Cavanagh, J. A. Chan, P. Landgraf, B. Ho, T. Milde, O. Witt, J. Ecker, F. Sahm, D. Sumerauer, D. W. Ellison, B. A. Orr, A. Darabi, C. Haberler, D. Figarella-Branger, P. Wesseling, J. Schittenhelm, M. Remke, M. D. Taylor, M. J. Gil-da Costa, M. Łastowska, W. Grajkowska, M. Hasselblatt, P. Hauser, T. Pietsch, E. Uro-Coste, F. Bourdeaut, J. Masliah-Planchon, V. Rigau, S. Alexandrescu, S. Wolf, X.-N. Li, U. Schüller, M. Snuderl, M. A. Karajannis, F. Giangaspero, N. Jabado, A. von Deimling, D. T. W. Jones, J. O. Korbel, K. von Hoff, P. Lichter, A. Huang, A. J. R. Bishop, S. M. Pfister, A. Korshunov, and M. Kool. The molecular landscape of etmr at diagnosis and relapse. *Nature*, 576(7786):274–280, 2019.

[63] J. T. Lange, C. Y. Chen, Y. Pichugin, L. Xie, J. Tang, K. L. Hung, K. E. Yost, Q. Shi, M. L. Erb, U. Rajkumar, S. Wu, C. Swanton, Z. Liu, W. Huang, H. Y. Chang, V. Bafna, A. G. Henssen, B. Werner, and P. S. Mischel. Principles of ecdna random inheritance drive rapid genome change and therapy resistance in human cancers. *bioRxiv*, 2021.

[64] L. Lapointe-Shaw, Z. Bouck, N. A. Howell, T. Lange, A. Orchanian-Cheff, P. C. Austin, N. M. Ivers, D. A. Redelmeier, and C. M. Bell. Mediation analysis with a time-to-event

outcome: a review of use and reporting in healthcare research. *BMC Medical Research Methodology*, 18(1):118, 2018.

[65] B. K. Li, A. Vasiljevic, C. Dufour, F. Yao, B. L. B. Ho, M. Lu, E. I. Hwang, S. Gururangan, J. R. Hansford, M. Fouladi, S. Nobusawa, A. Laquerriere, M.-B. Delisle, J. Fangusaro, F. Forest, H. Toledano, P. Solano-Paez, S. Leary, D. Birks, L. M. Hoffman, A. Szathmari, C. Faure-Conter, X. Fan, D. Catchpoole, L. Zhou, K. A. P. Schultz, K. Ichimura, G. Gauchotte, N. Jabado, C. Jones, D. Loussouarn, K. Mokhtari, A. Rousseau, D. S. Ziegler, S. Tanaka, S. L. Pomeroy, A. Gajjar, V. Ramaswamy, C. Hawkins, R. G. Grundy, D. A. Hill, E. Bouffet, A. Huang, and A. Jouvet. Pineoblastoma segregates into molecular sub-groups with distinct clinico-pathologic features: a rare brain tumor consortium registry study. *Acta Neuropathologica*, 139(2):223–241, 2020.

[66] F. P. Li, J. Fraumeni, J. F., J. J. Mulvihill, W. A. Blattner, M. G. Dreyfus, M. A. Tucker, and R. W. Miller. A cancer family syndrome in twenty-four kindreds. *Cancer Res*, 48(18):5358–62, 1988.

[67] H. Li and R. Durbin. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, 25(14):1754–1760, 2009.

[68] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–9, 2009.

[69] M. Li, K. F. Lee, Y. Lu, I. Clarke, D. Shih, C. Eberhart, V. P. Collins, T. Van Meter, D. Picard, L. Zhou, P. C. Boutros, P. Modena, M. L. Liang, S. W. Scherer, E. Bouffet, J. T. Rutka, S. L. Pomeroy, C. C. Lau, M. D. Taylor, A. Gajjar, P. B. Dirks, C. E. Hawkins, and A. Huang. Frequent amplification of a chr19q13.41 microrna polycistron in aggressive primitive neuroectodermal brain tumors. *Cancer Cell*, 16(6):533–46, 2009.

[70] W. Li, J. Köster, H. Xu, C.-H. Chen, T. Xiao, J. S. Liu, M. Brown, and X. S. Liu. Quality control, modeling, and visualization of crispr screens with mageck-vispr. *Genome Biology*, 16(1):281, 2015.

[71] J. V. Lilly, J. L. Rokita, J. L. Mason, T. Patton, S. Stefankiewiz, D. Higgins, G. Trooskin, C. A. Larouci, K. Arya, E. Appert, A. P. Heath, Y. Zhu, M. A. Brown, B. Zhang, B. K. Farrow, S. Robins, A. M. Morgan, T. Q. Nguyen, E. Frenkel, K. Lehmann, E. Drake, C. Sullivan, A. Plisiewicz, N. Coleman, L. Patterson, M. Koptyra, Z. Helili, N. Van Kuren, N. Young, M. C. Kim, C. Friedman, A. Lubneuski, C. Blackden, M. Williams, V. Baubet, L. Tauhid, J. Galanaugh, K. Boucher, H. Ijaz, K. A. Cole, N. Choudhari, M. Santi, R. W. Moulder, J. Waller, W. Rife, S. J. Diskin, M. Mateos, D. W. Parsons, I. F. Pollack, S. Goldman, S. Leary, C. Caporalini, A. M. Buccoliero, M. Scagnet, D. Haussler, D. Hanson, R. Firestein, J. Cain, J. J. Phillips, N. Gupta, S. Mueller, G. Grant, M. Monje-Deisseroth, S. Partap, J. P. Greenfield, R. Hashizume, A. Smith, S. Zhu, J. M. Johnston, J. R. Fangusaro, M. Miller, M. D. Wood, S. Gardner, C. L. Carter, L. M. Prolo, J. Pisapia, K. Pehlivan,

A. Franson, T. Niazi, J. Rubin, M. Abdelbaki, D. S. Ziegler, H. B. Lindsay, A. G. Stucklin, N. Gerber, O. M. Vaske, C. Quinsey, B. R. Rood, J. Nazarian, E. Raabe, E. M. Jackson, S. Stapleton, R. M. Lober, D. E. Kram, C. Koschmann, P. B. Storm, R. R. Lulla, M. Prados, A. C. Resnick, and A. J. Waanders. The children's brain tumor network (cbtn) - accelerating research in pediatric central nervous system tumors through collaboration and open science. *Neoplasia*, 35:100846, 2023.

[72] C. Y. Lin, S. Erkek, Y. Tong, L. Yin, A. J. Federation, M. Zapatka, P. Haldipur, D. Kawauchi, T. Risch, H. J. Warnatz, B. C. Worst, B. Ju, B. A. Orr, R. Zeid, D. R. Polaski, M. Segura-Wang, S. M. Waszak, D. T. Jones, M. Kool, V. Hovestadt, I. Buchhalter, L. Sieber, P. Johann, L. Chavez, S. Groschel, M. Ryzhova, A. Korshunov, W. Chen, V. V. Chizhikov, K. J. Millen, V. Amstislavskiy, H. Lehrach, M. L. Yaspo, R. Eils, P. Lichter, J. O. Korbel, S. M. Pfister, J. E. Bradner, and P. A. Northcott. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature*, 530(7588):57–62, 2016.

[73] M. F. Lin, O. Rodeh, J. Penn, X. Bai, J. G. Reid, O. Krasheninina, and W. J. Salerno. Glnexus: joint variant calling for large cohort sequencing. *bioRxiv*, page 343970, 2018.

[74] F. Liu, Z. Shi, W. Bao, J. Zheng, K. Chen, Z. Lin, H.-N. Song, X. Luo, Q. Dong, L. Jiang, Y. Wang, G. Chen, and X. Chen. Zic2 promotes colorectal cancer growth and metastasis through the tgf-β signaling pathway. *Experimental Cell Research*, 415(2):113118, 2022.

[75] G. Liu and S. Piantadosi. Ridge estimation in generalized linear models and proportional hazards regressions. *Communications in Statistics - Theory and Methods*, 46(23):11466–11479, 2017.

[76] X. Lu, T. A. Nguyen, S. H. Moon, Y. Darlington, M. Sommer, and L. A. Donehower. The type 2c phosphatase wip1: an oncogenic regulator of tumor suppressor and dna damage response pathways. *Cancer Metastasis Rev*, 27(2):123–35, 2008.

[77] H. A. Lubs and J. H. Salmon. The chromosomal complement of human solid tumors: Ii. karyotypes of glial tumors. *Journal of Neurosurgery*, 22(2):160 – 168, 1965.

[78] J. Luebeck, C. Coruh, S. R. Dehkordi, J. T. Lange, K. M. Turner, V. Deshpande, D. A. Pai, C. Zhang, U. Rajkumar, J. A. Law, P. S. Mischel, and V. Bafna. Ampliconreconstructor integrates ngs and optical mapping to resolve the complex structures of focal amplifications. *Nature Communications*, 11(1):4374, 2020.

[79] J. Luebeck, A. W. Tian Ng, P. C. Galipeau, X. Li, C. A. Sanchez, A. Katz-Summercorn, H. Kim, S. Jammula, Y. He, S. M. Lippman, R. Verhaak, C. C. Maley, L. B. Alexandrov, B. J. Reid, R. C. Fitzgerald, T. G. Paulson, H. Y. Chang, S. Wu, V. Bafna, and P. S. Mischel. Extrachromosomal dna in the cancerous transformation of barrett's esophagus. *bioRxiv*, page 2022.07.25.501144, 2022.

[80] X. Ma, Y. Liu, Y. Liu, L. B. Alexandrov, M. N. Edmonson, C. Gawad, X. Zhou, Y. Li, M. C. Rusch, J. Easton, R. Huether, V. Gonzalez-Pena, M. R. Wilkinson, L. C. Hermida,

S. Davis, E. Sioson, S. Pounds, X. Cao, R. E. Ries, Z. Wang, X. Chen, L. Dong, S. J. Diskin, M. A. Smith, J. M. Guidry Auvil, P. S. Meltzer, C. C. Lau, E. J. Perlman, J. M. Maris, S. Meshinchi, S. P. Hunger, D. S. Gerhard, and J. Zhang. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*, 555(7696):371–376, 2018.

[81] A. Mackay, A. Burford, D. Carvalho, E. Izquierdo, J. Fazal-Salom, K. R. Taylor, L. Bjerke, M. Clarke, M. Vinci, M. Nandhabalan, S. Temelso, S. Popov, V. Molinari, P. Raman, A. J. Waanders, H. J. Han, S. Gupta, L. Marshall, S. Zacharoulis, S. Vaidya, H. C. Mandeville, L. R. Bridges, A. J. Martin, S. Al-Sarraj, C. Chandler, H. K. Ng, X. Li, K. Mu, S. Trabelsi, D. H. Brahim, A. N. Kisljakov, D. M. Konovalov, A. S. Moore, A. M. Carcaboso, M. Sunol, C. de Torres, O. Cruz, J. Mora, L. I. Shats, J. N. Stavale, L. T. Bidinotto, R. M. Reis, N. Entz-Werle, M. Farrell, J. Cryan, D. Crimmins, J. Caird, J. Pears, M. Monje, M. A. Debily, D. Castel, J. Grill, C. Hawkins, H. Nikbakht, N. Jabado, S. J. Baker, S. M. Pfister, D. T. W. Jones, M. Fouladi, A. O. von Bueren, M. Baudis, A. Resnick, and C. Jones. Integrated molecular meta-analysis of 1,000 pediatric high-grade and diffuse intrinsic pontine glioma. *Cancer Cell*, 32(4):520–537.e5, 2017.

[82] C. S. McGinnis, L. M. Murrow, and Z. J. Gartner. Doubletfinder: Doublet detection in single-cell rna sequencing data using artificial nearest neighbors. *Cell Systems*, 8(4):329–337.e4, 2019.

[83] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome Res*, 20(9):1297–303, 2010.

[84] W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. The ensembl variant effect predictor. *Genome Biology*, 17(1):122, 2016.

[85] R. McLendon, A. Friedman, D. Bigner, E. G. Van Meir, D. J. Brat, G. M. Mastrogianakis, J. J. Olson, T. Mikkelsen, N. Lehman, K. Aldape, W. K. Alfred Yung, O. Bogler, S. VandenBerg, M. Berger, M. Prados, D. Muzny, M. Morgan, S. Scherer, A. Sabo, L. Nazareth, L. Lewis, O. Hall, Y. Zhu, Y. Ren, O. Alvi, J. Yao, A. Hawes, S. Jhangiani, G. Fowler, A. San Lucas, C. Kovar, A. Cree, H. Dinh, J. Santibanez, V. Joshi, M. L. Gonzalez-Garay, C. A. Miller, A. Milosavljevic, L. Donehower, D. A. Wheeler, R. A. Gibbs, K. Cibulskis, C. Sougnez, T. Fennell, S. Mahan, J. Wilkinson, L. Ziaugra, R. Onofrio, T. Bloom, R. Nicol, K. Ardlie, J. Baldwin, S. Gabriel, E. S. Lander, L. Ding, R. S. Fulton, M. D. McLellan, J. Wallis, D. E. Larson, X. Shi, R. Abbott, L. Fulton, K. Chen, D. C. Koboldt, M. C. Wendl, R. Meyer, Y. Tang, L. Lin, J. R. Osborne, B. H. Dunford-Shore, T. L. Miner, K. Delehaunty, C. Markovic, G. Swift, W. Courtney, C. Pohl, S. Abbott, A. Hawkins, S. Leong, C. Haipek, H. Schmidt, M. Wiechert, T. Vickery, S. Scott, D. J. Dooling, A. Chinwalla, G. M. Weinstock, E. R. Mardis, R. K. Wilson, G. Getz, W. Winckler, R. G. W. Verhaak, M. S. Lawrence, M. O'Kelly, J. Robinson, G. Alexe, R. Beroukhim, S. Carter, D. Chiang,

J. Gould, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455(7216):1061–1068, 2008.

[86] C. McLeod, A. M. Gout, X. Zhou, A. Thrasher, D. Rahbarinia, S. W. Brady, M. Macias, K. Birch, D. Finkelstein, J. Sunny, R. Mudunuri, B. A. Orr, M. Treadway, B. Davidson, T. K. Ard, A. Chiao, A. Swistak, S. Wiggins, S. Foy, J. Wang, E. Sioson, S. Wang, J. R. Michael, Y. Liu, X. Ma, A. Patel, M. N. Edmonson, M. R. Wilkinson, A. M. Frantz, T.-C. Chang, L. Tian, S. Lei, S. M. A. Islam, C. Meyer, N. Thangaraj, P. Tater, V. Kandali, S. Ma, T. Nguyen, O. A. Serang, I. McGuire, N. Robison, D. Gentry, X. Tang, L. E. Palmer, G. Wu, E. Suh, L. Tanner, J. McMurry, M. Lear, A. S. Pappo, Z. Wang, C. L. Wilson, Y. Cheng, S. Meshinchi, L. B. Alexandrov, M. J. Weiss, G. T. Armstrong, L. L. Robison, Y. Yasui, K. E. Nichols, D. W. Ellison, C. Bangur, C. G. Mullighan, S. J. Baker, M. A. Dyer, G. Miller, S. Newman, M. Rusch, R. Daly, K. Perry, J. R. Downing, and J. Zhang. St. jude cloud-a pediatric cancer genomic data sharing ecosystem. *Cancer Discovery*, pages CD–20–1230, 2021.

[87] M. Meddeb, A. Valent, G. Danglot, V. C. Nguyen, A. Duverger, F. Fouquet, M. J. Terrier-Lacombe, O. Oberlin, and A. Bernheim. Mdm2 amplification in a primary alveolar rhabdomyosarcoma displaying a t(2;13)(q35;q14). *Cytogenetic and Genome Research*, 73(4):325–330, 1996.

[88] P. Milani, R. Escalante-Chong, B. C. Shelley, N. L. Patel-Murray, X. Xin, M. Adam, B. Mandefro, D. Sareen, C. N. Svendsen, and E. Fraenkel. Cell freezing protocol suitable for atac-seq on motor neurons derived from human induced pluripotent stem cells. *Scientific Reports*, 6(1):25474, 2016.

[89] E. Missiaglia, D. Williamson, J. Chisholm, P. Wirapati, G. Pierron, F. Petel, J.-P. Concordet, K. Thway, O. Oberlin, and K. Pritchard-Jones. Pax3/foxo1 fusion gene status is the key prognostic molecular marker in rhabdomyosarcoma and significantly improves current risk stratification. *J Clin Oncol*, 30(14):1670–1677, 2012.

[90] J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shoresh, J. Adrian, T. Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He, M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X.-O. Zhang, S. I. Elhajjajy, J. Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri, J. Zhang, A. Victorsen, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lécuyer, D. M. Gilbert, J. Dekker, J. Rinn, E. M. Mendenhall, J. R. Ecker, M. Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigó, P. J. Farnham, J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E. Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos, Z. Weng, and E. P. C. The. Expanded encyclopaedias of dna elements in the human and mouse genomes. *Nature*, 583(7818):699–710, 2020.

[91] M. Morabito, M. Larcher, F. M. G. Cavalli, C. Foray, A. Forget, L. Mirabal-Ortega, M. Andrianteranagna, S. Druillennec, A. Garancher, J. Masliah-Planchon, S. Leboucher,

A. Debalkew, A. Raso, O. Delattre, S. Puget, F. Doz, M. D. Taylor, O. Ayrault, F. Bourdeaut, A. Eychène, and C. Pouponnot. An autocrine activinb mechanism drives tgfβ/activin signaling in group 3 medulloblastoma. *EMBO Molecular Medicine*, 11(8):e9830, 2019.

[92] C. Morales, M. J. García, M. Ribas, R. Miró, M. Muñoz, C. Caldas, and M. A. Peinado. Dihydrofolate reductase amplification and sensitization to methotrexate of methotrexate-resistant colon cancer cells. *Mol Cancer Ther*, 8(2):424–32, 2009.

[93] A. R. Morton, N. Dogan-Artun, Z. J. Faber, G. MacLeod, C. F. Bartels, M. S. Piazza, K. C. Allan, S. C. Mack, X. Wang, R. C. Gimple, Q. Wu, B. P. Rubin, S. Shetty, S. Angers, P. B. Dirks, R. C. Sallari, M. Lupien, J. N. Rich, and P. C. Scacheri. Functional enhancers shape extrachromosomal oncogene amplifications. *Cell*, 179(6):1330–1341.e13, 2019.

[94] V. Mysorekar, K. Harish, N. Kilara, M. Subramanian, and A. Giridhar. Embryonal rhabdomyosarcoma of the chest wall: A case report and review of the literature. *Indian Journal of Pathology and Microbiology*, 51(2):274–276, 2008.

[95] D. A. Nathanson, B. Gini, J. Mottahedeh, K. Visnyei, T. Koga, G. Gomez, A. Eskin, K. Hwang, J. Wang, K. Masui, A. Paucar, H. Yang, M. Ohashi, S. Zhu, J. Wykosky, R. Reed, S. F. Nelson, T. F. Cloughesy, C. D. James, P. N. Rao, H. I. Kornblum, J. R. Heath, W. K. Cavenee, F. B. Furnari, and P. S. Mischel. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant egfr dna. *Science*, 343(6166):72, 2014.

[96] S. Nikolaev, F. Santoni, M. Garieri, P. Makrythanasis, E. Falconnet, M. Guipponi, A. Vannier, I. Radovanovic, F. Bena, F. Forestier, K. Schaller, V. Dutoit, V. Clement-Schatlo, P. Y. Dietrich, and S. E. Antonarakis. Extrachromosomal driver mutations in glioblastoma and low-grade glioma. *Nat Commun*, 5:5690, 2014.

[97] B. V. North, D. Curtis, and P. C. Sham. A note on the calculation of empirical p values from monte carlo procedures. *American journal of human genetics*, 71(2):439–441, 2002.

[98] P. A. Northcott, I. Buchhalter, A. S. Morrissy, V. Hovestadt, J. Weischenfeldt, T. Ehrenberger, S. Grobner, M. Segura-Wang, T. Zichner, V. A. Rudneva, H. J. Warnatz, N. Sidiropoulos, A. H. Phillips, S. Schumacher, K. Kleinheinz, S. M. Waszak, S. Erkek, D. T. W. Jones, B. C. Worst, M. Kool, M. Zapatka, N. Jager, L. Chavez, B. Hutter, M. Bieg, N. Paramasivam, M. Heinold, Z. Gu, N. Ishaque, C. Jager-Schmidt, C. D. Imbusch, A. Jugold, D. Hubschmann, T. Risch, V. Amstislavskiy, F. G. R. Gonzalez, U. D. Weber, S. Wolf, G. W. Robinson, X. Zhou, G. Wu, D. Finkelstein, Y. Liu, F. M. G. Cavalli, B. Luu, V. Ramaswamy, X. Wu, J. Koster, M. Ryzhova, Y. J. Cho, S. L. Pomeroy, C. Herold-Mende, M. Schuhmann, M. Ebinger, L. M. Liau, J. Mora, R. E. McLendon, N. Jabado, T. Kumabe, E. Chuah, Y. Ma, R. A. Moore, A. J. Mungall, K. L. Mungall, N. Thiessen, K. Tse, T. Wong, S. J. M. Jones, O. Witt, T. Milde, A. Von Deimling, D. Capper, A. Korshunov, M. L. Yaspo, R. Kriwacki, A. Gajjar, J. Zhang, R. Beroukhim, E. Fraenkel, J. O. Korbel, B. Brors, M. Schlesner, R. Eils, M. A. Marra, S. M. Pfister, M. D. Taylor, and P. Lichter. The whole-genome landscape of medulloblastoma subtypes. *Nature*, 547(7663):311–317, 2017.

[99] P. A. Northcott, C. Lee, T. Zichner, A. M. Stütz, S. Erkek, D. Kawauchi, D. J. H. Shih, V. Hovestadt, M. Zapatka, D. Sturm, D. T. W. Jones, M. Kool, M. Remke, F. M. G. Cavalli, S. Zuyderduyn, G. D. Bader, S. VandenBerg, L. A. Esparza, M. Ryzhova, W. Wang, A. Wittmann, S. Stark, L. Sieber, H. Seker-Cin, L. Linke, F. Kratochwil, N. Jäger, I. Buchhalter, C. D. Imbusch, G. Zipprich, B. Raeder, S. Schmidt, N. Diessl, S. Wolf, S. Wiemann, B. Brors, C. Lawerenz, J. Eils, H.-J. Warnatz, T. Risch, M.-L. Yaspo, U. D. Weber, C. C. Bartholomae, C. von Kalle, E. Turányi, P. Hauser, E. Sanden, A. Darabi, P. Siesjö, J. Sterba, K. Zitterbart, D. Sumerauer, P. van Sluis, R. Versteeg, R. Volckmann, J. Koster, M. U. Schuhmann, M. Ebinger, H. L. Grimes, G. W. Robinson, A. Gajjar, M. Mynarek, K. von Hoff, S. Rutkowski, T. Pietsch, W. Scheurlen, J. Felsberg, G. Reifenberger, A. E. Kulozik, A. von Deimling, O. Witt, R. Eils, R. J. Gilbertson, A. Korshunov, M. D. Taylor, P. Lichter, J. O. Korbel, R. J. Wechsler-Reya, and S. M. Pfister. Enhancer hijacking activates gfi1 family oncogenes in medulloblastoma. *Nature*, 511(7510):428–434, 2014.

[100] P. A. Northcott, D. J. H. Shih, J. Peacock, L. Garzia, A. Sorana Morrissy, T. Zichner, A. M. Stütz, A. Korshunov, J. Reimand, S. E. Schumacher, R. Beroukhim, D. W. Ellison, C. R. Marshall, A. C. Lionel, S. Mack, A. Dubuc, Y. Yao, V. Ramaswamy, B. Luu, A. Rolider, F. M. G. Cavalli, X. Wang, M. Remke, X. Wu, R. Y. B. Chiu, A. Chu, E. Chuah, R. D. Corbett, G. R. Hoad, S. D. Jackman, Y. Li, A. Lo, K. L. Mungall, K. Ming Nip, J. Q. Qian, A. G. J. Raymond, N. Thiessen, R. J. Varhol, I. Birol, R. A. Moore, A. J. Mungall, R. Holt, D. Kawauchi, M. F. Roussel, M. Kool, D. T. W. Jones, H. Witt, A. Fernandez-L, A. M. Kenney, R. J. Wechsler-Reya, P. Dirks, T. Aviv, W. A. Grajkowska, M. Perek-Polnik, C. C. Haberler, O. Delattre, S. S. Reynaud, F. F. Doz, S. S. Pernet-Fattet, B.-K. Cho, S.-K. Kim, K.-C. Wang, W. Scheurlen, C. G. Eberhart, M. Fèvre-Montange, A. Jouvet, I. F. Pollack, X. Fan, K. M. Muraszko, G. Yancey Gillespie, C. Di Rocco, L. Massimi, E. M. C. Michiels, N. K. Kloosterhof, P. J. French, J. M. Kros, J. M. Olson, R. G. Ellenbogen, K. Zitterbart, L. Kren, R. C. Thompson, M. K. Cooper, B. Lach, R. E. McLendon, D. D. Bigner, A. Fontebasso, S. Albrecht, N. Jabado, J. C. Lindsey, S. Bailey, N. Gupta, W. A. Weiss, L. Bognár, A. Klekner, T. E. Van Meter, T. Kumabe, T. Tominaga, S. K. Elbabaa, J. R. Leonard, J. B. Rubin, et al. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature*, 488(7409):49–56, 2012.

[101] K. Okonechnikov, A. Camgöz, O. Chapman, S. Wani, D. E. Park, J.-M. Hübner, A. Chakraborty, M. Pagadala, R. Bump, S. Chandran, K. Kraft, R. Acuna-Hidalgo, D. Reid, K. Sikkink, M. Mauermann, E. F. Juarez, A. Jenseit, J. T. Robinson, K. W. Pajtler, T. Milde, N. Jäger, P. Fiesel, L. Morgan, S. Sridhar, N. G. Coufal, M. Levy, D. Malicki, C. Hobbs, S. Kingsmore, S. Nahas, M. Snuderl, J. Crawford, R. J. Wechsler-Reya, T. B. Davidson, J. Cotter, G. Michaiel, G. Fleischhack, S. Mundlos, A. Schmitt, H. Carter, K. A. Michealraj, S. A. Kumar, M. D. Taylor, J. Rich, F. Buchholz, J. P. Mesirov, S. M. Pfister, F. Ay, J. R. Dixon, M. Kool, and L. Chavez. 3d genome mapping identifies subgroup-specific chromosome conformations and tumor-dependency genes in ependymoma. *Nature Communications*, 14(1):2300, 2023.

[102] L. A. Pennacchio, W. Bickmore, A. Dean, M. A. Nobrega, and G. Bejerano. Enhancers: five essential questions. *Nature Reviews Genetics*, 14(4):288–295, 2013.

[103] L. S. Pongor, C. W. Schultz, L. Rinaldi, D. Wangsa, C. E. Redon, N. Takahashi, G. Fialkoff, P. Desai, Y. Zhang, S. Burkett, N. Hermoni, N. Vilk, J. Gutin, R. Gergely, Y. Zhao, S. Nichols, R. Vilimas, L. Sciuto, C. Graham, J. M. Caravaca, S. Turan, S. Tsai-Wei, V. N. Rajapakse, R. Kumar, D. Upadhyay, S. Kumar, Y. S. Kim, N. Roper, B. Tran, S. M. Hewitt, D. E. Kleiner, M. I. Aladjem, N. Friedman, G. L. Hager, Y. Pommier, T. Ried, and A. Thomas. Extrachromosomal dna amplification contributes to small cell lung cancer heterogeneity and is associated with worse outcomes. *Cancer Discov*, 13(4):928–949, 2023.

[104] L. Potocki, K. S. Chen, S. S. Park, D. E. Osterholm, M. A. Withers, V. Kimonis, A. M. Summers, W. S. Meschino, K. Anyane-Yeboa, C. D. Kashork, L. G. Shaffer, and J. R. Lupski. Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the smith-magenis microdeletion. *Nat Genet*, 24(1):84–7, 2000.

[105] A. R. Quinlan and I. M. Hall. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, 2010.

[106] S. Raeisi Dehkordi, J. Luebeck, and V. Bafna. Fandom: Fast nested distance-based seeding of optical maps. *Patterns*, 2(5):100248, 2021.

[107] U. Rajkumar, K. Turner, J. Luebeck, V. Deshpande, M. Chandraker, P. Mischel, and V. Bafna. Ecseg: Semantic segmentation of metaphase images containing extrachromosomal dna. *iScience*, 21:428–435, 2019.

[108] V. Ramaswamy, M. Remke, E. Bouffet, S. Bailey, S. C. Clifford, F. Doz, M. Kool, C. Dufour, G. Vassal, T. Milde, O. Witt, K. von Hoff, T. Pietsch, P. A. Northcott, A. Gajjar, G. W. Robinson, L. Padovani, N. André, M. Massimino, B. Pizer, R. Packer, S. Rutkowski, S. M. Pfister, M. D. Taylor, and S. L. Pomeroy. Risk stratification of childhood medulloblastoma in the molecular era: the current consensus. *Acta Neuropathologica*, 131(6):821–831, 2016.

[109] F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, and T. Manke. deeptools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*, 44(W1):W160–5, 2016.

[110] S. P. Rao, M. Huntley, N. Durand, E. Stamenova, I. Bochkov, J. Robinson, A. Sanborn, I. Machol, A. Omer, E. Lander, and E. Aiden. A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680, 2014.

[111] T. Rausch, D. W. Jones, M. Zapatka, A. Stütz, T. Zichner, J. Weischenfeldt, N. Jäger, M. Remke, D. Shih, P. Northcott, E. Pfaff, J. Tica, Q. Wang, L. Massimi, H. Witt, S. Bender, S. Pleier, H. Cin, C. Hawkins, C. Beck, A. von Deimling, V. Hans, B. Brors, R. Eils, W. Scheurlen, J. Blake, V. Benes, A. Kulozik, O. Witt, D. Martin, C. Zhang, R. Porat,

D. M. Merino, J. Wasserman, N. Jabado, A. Fontebasso, L. Bullinger, F. G. Rücker, K. Döhner, H. Döhner, J. Koster, J. Molenaar, R. Versteeg, M. Kool, U. Tabori, D. Malkin, A. Korshunov, M. Taylor, P. Lichter, S. Pfister, and J. Korbel. Genome sequencing of pediatric medulloblastoma links catastrophic dna rearrangements with tp53 mutations. *Cell*, 148(1):59–71, 2012.

[112] M. Reich, T. Tabor, T. Liefeld, H. Thorvaldsdóttir, B. Hill, P. Tamayo, and J. P. Mesirov. The genepattern notebook environment. *Cell systems*, 5(2):149–151.e1, 2017.

[113] P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, and M. Kircher. Cadd: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1):D886–D894, 2019.

[114] G. Robinson, M. Parker, T. A. Kranenburg, C. Lu, X. Chen, L. Ding, T. N. Phoenix, E. Hedlund, L. Wei, X. Zhu, N. Chalhoub, S. J. Baker, R. Huether, R. Kriwacki, N. Curley, R. Thiruvenkatam, J. Wang, G. Wu, M. Rusch, X. Hong, J. Becksfort, P. Gupta, J. Ma, J. Easton, B. Vadodaria, A. Onar-Thomas, T. Lin, S. Li, S. Pounds, S. Paugh, D. Zhao, D. Kawauchi, M. F. Roussel, D. Finkelstein, D. W. Ellison, C. C. Lau, E. Bouffet, T. Hassall, S. Gururangan, R. Cohn, R. S. Fulton, L. L. Fulton, D. J. Dooling, K. Ochoa, A. Gajjar, E. R. Mardis, R. K. Wilson, J. R. Downing, J. Zhang, and R. J. Gilbertson. Novel mutations target distinct subgroups of medulloblastoma. *Nature*, 488(7409):43–48, 2012.

[115] G. W. Robinson, B. A. Orr, G. Wu, S. Gururangan, T. Lin, I. Qaddoumi, R. J. Packer, S. Goldman, M. D. Prados, A. Desjardins, M. Chintagumpala, N. Takebe, S. C. Kaste, M. Rusch, S. J. Allen, A. Onar-Thomas, C. F. Stewart, M. Fouladi, J. M. Boyett, R. J. Gilbertson, T. Curran, D. W. Ellison, and A. Gajjar. Vismodegib exerts targeted efficacy against recurrent sonic hedgehog-subgroup medulloblastoma: Results from phase ii pediatric brain tumor consortium studies pbtc-025b and pbtc-032. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 33(24):2646–2654, 2015.

[116] J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. Integrative genomics viewer. *Nature biotechnology*, 29(1):24–26, 2011.

[117] J. M. Rusert, E. F. Juarez, S. Brabetz, J. Jensen, A. Garancher, L. Q. Chau, S. K. Tacheva-Grigorova, S. Wahab, Y. T. Udaka, D. Finlay, H. Seker-Cin, B. Reardon, S. Gröbner, J. Serrano, J. Ecker, L. Qi, M. Kogiso, Y. Du, P. A. Baxter, J. J. Henderson, M. E. Berens, K. Vuori, T. Milde, Y. J. Cho, X. N. Li, J. M. Olson, I. Reyes, M. Snuderl, T. C. Wong, D. P. Dimmock, S. A. Nahas, D. Malicki, J. R. Crawford, M. L. Levy, E. M. Van Allen, S. M. Pfister, P. Tamayo, M. Kool, J. P. Mesirov, and R. J. Wechsler-Reya. Functional precision medicine identifies new therapeutic candidates for medulloblastoma. *Cancer Res*, 80(23):5393–5407, 2020.

[118] D. E. Rushlow, B. M. Mol, J. Y. Kennett, S. Yee, S. Pajovic, B. L. Thériault, N. L. Prigoda-Lee, C. Spencer, H. Dimaras, T. W. Corson, R. Pang, C. Massey, R. Godbout, Z. Jiang, E. Zacksenhaus, K. Paton, A. C. Moll, C. Houdayer, A. Raizis, W. Halliday, W. L.

Lam, P. C. Boutros, D. Lohmann, J. C. Dorsman, and B. L. Gallie. Characterisation of retinoblastomas without rb1 mutations: genomic, gene expression, and clinical studies. *The Lancet Oncology*, 14(4):327–334, 2013.

[119] S. L. Ryan, E. C. Schwalbe, M. Cole, Y. Lu, M. E. Lusher, H. Megahed, K. O'Toole, S. L. Nicholson, L. Bognar, M. Garami, P. Hauser, A. Korshunov, S. M. Pfister, D. Williamson, R. E. Taylor, D. W. Ellison, S. Bailey, and S. C. Clifford. Myc family amplification and clinical risk-factors interact to predict an extremely poor prognosis in childhood medulloblastoma. *Acta Neuropathologica*, 123(4):501–513, 2012.

[120] K. Sakai, N. Kanda, Y. Shiloh, T. Donlon, R. Schreck, J. Shipley, T. Dryja, E. Chaum, R. S. K. Chaganti, and S. Latt. Molecular and cytologic analysis of dna amplification in retinoblastoma. *Cancer Genetics and Cytogenetics*, 17(2):95–112, 1985.

[121] R. Salloum, Y. Chen, Y. Yasui, R. Packer, W. Leisenring, E. Wells, A. King, R. Howell, T. M. Gibson, K. R. Krull, L. L. Robison, K. C. Oeffinger, M. Fouladi, and G. T. Armstrong. Late morbidity and mortality among medulloblastoma survivors diagnosed across three decades: A report from the childhood cancer survivor study. *Journal of Clinical Oncology*, 37(9):731–740, 2019.

[122] J. Z. Sanborn, S. R. Salama, M. Grifford, C. W. Brennan, T. Mikkelsen, S. Jhanwar, S. Katzman, L. Chin, and D. Haussler. Double minute chromosomes in glioblastoma multiforme are revealed by precise reconstruction of oncogenic amplicons. *Cancer Res*, 73(19):6036–45, 2013.

[123] S. Seabold and J. Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[124] N. Servant, N. Varoquaux, B. R. Lajoie, E. Viara, C.-J. Chen, J.-P. Vert, E. Heard, J. Dekker, and E. Barillot. Hic-pro: an optimized and flexible pipeline for hi-c data processing. *Genome Biology*, 16(1):259, 2015.

[125] N. Shimizu, N. Itoh, H. Utiyama, and G. M. Wahl. Selective entrapment of extrachromosomally amplified dna by nuclear budding and micronucleation during s phase. *J Cell Biol*, 140(6):1307–20, 1998.

[126] O. Shoshani, S. F. Brunner, R. Yaeger, P. Ly, Y. Nechemia-Arbely, D. H. Kim, R. Fang, G. A. Castillon, M. Yu, J. S. Z. Li, Y. Sun, M. H. Ellisman, B. Ren, P. J. Campbell, and D. W. Cleveland. Chromothripsis drives the evolution of gene amplification in cancer. *Nature*, 2020.

[127] Z. Sondka, S. Bamford, C. G. Cole, S. A. Ward, I. Dunham, and S. A. Forbes. The cosmic cancer gene census: describing genetic dysfunction across all human cancers. *Nature Reviews Cancer*, 18(11):696–705, 2018.

[128] W. Song, W. Yu, D. Li, C. Cheng, X. Wu, J. Chen, and W. Zhang. Zic5 promotes human hepatocellular carcinoma cell proliferation through upregulating col1a1. *J Gastrointest Oncol*, 13(3):1237–1247, 2022.

[129] S. Sridhar, O. Chapman, A. Dutta, S. Wang, J. Luebeck, V. Bafna, and L. Chavez. Extrachromosomal dna promotes oncogene amplification across multiple pediatric cancer types, 2023.

[130] H. Steilen-Gimbel, K. Remberger, N. Graf, W.-I. Steudel, K. D. Zang, and W. Henn. A novel site of dna amplification on chromosome 1p32-33 in a rhabdomyosarcoma revealed by comparative genomic hybridization. *Human Genetics*, 97(1):87–90, 1996.

[131] C. T. Storlazzi, T. Fioretos, C. Surace, A. Lonoce, A. Mastrorilli, B. Strömbeck, P. D'Addabbo, F. Iacovelli, C. Minervini, A. Aventin, N. Dastugue, C. Fonatsch, A. Hagemeijer, M. Jotterand, D. Mühlematter, M. Lafage-Pochitaloff, F. Nguyen-Khac, C. Schoch, M. L. Slovak, A. Smith, F. Solè, N. Van Roy, B. Johansson, and M. Rocchi. Myc-containing double minutes in hematologic malignancies: evidence in favor of the episome model and exclusion of myc as the target gene. *Hum Mol Genet*, 15(6):933–42, 2006.

[132] E. B. Stovner and P. Sætrom. Pyranges: efficient comparison of genomic intervals in python. *Bioinformatics*, 36(3):918–919, 2020.

[133] Y. Sun, J. Han, Z. Wang, X. Li, Y. Sun, and Z. Hu. Safety and efficacy of bromodomain and extra-terminal inhibitors for the treatment of hematological malignancies and solid tumors: A systematic study of clinical trials. *Front Pharmacol*, 11:621093, 2020.

[134] E. Talevich, A. H. Shain, T. Botton, and B. C. Bastian. Cnvkit: Genome-wide copy number detection and visualization from targeted dna sequencing. *PLoS Computational Biology*, 12(4):e1004873, 2016.

[135] Y.-F. Tan, Y. Zhang, S.-Y. Ge, F. Zhong, C.-Y. Sun, and G.-W. Xia. Ar-regulated zic5 contributes to the aggressiveness of prostate cancer. *Cell Death Discovery*, 8(1):393, 2022.

[136] P. K. Thomas, J. Marques, W., M. B. Davis, M. G. Sweeney, R. H. King, J. L. Bradley, J. R. Muddle, J. Tyson, S. Malcolm, and A. E. Harding. The phenotypic manifestations of chromosome 17p11.2 duplication. *Brain*, 120(3):465–478, 1997.

[137] T. Tickle, I. Tirosh, C. Georgescu, M. Brown, and B. Haas. infercnv of the trinity ctat project. 2019.

[138] A. Tiwari, U. Paithane, K. Tashiro, O. Saulnier, K. Barbosa, A. Vishvanathan, Q. Trinh, T. Nakashima, T. Mohammad, J. T. Lange, S. Wu, M. B. Masihi, A. Soni, K. Vogel, J. Friedlein, B. Hall, A. Shukla, S. Saha, L. Denis, L. D. Hendrikse, B. James, O. Chapman, J. Luebeck, B. Konety, S. M. Dehm, Y. Marahrens, L. D. Stein, L. Chavez, V. Bafna, P. S. Mischel, H. Suzuki, I. Hassan, A. J. Deshpande, R. J. Wechsler-Reya, M. D. Taylor, and A. Bagchi. The lncrna pvt1 encodes peptides with oncogenic and tumor suppressive activity, 2022.

[139] F. Toledo and G. M. Wahl. Mdm2 and mdm4: p53 regulators as targets in anticancer therapy. *The international journal of biochemistry & cell biology*, 39(7-8):1476–1482, 2007.

[140] A. Tsherniak, F. Vazquez, P. G. Montgomery, B. A. Weir, G. Kryukov, G. S. Cowley, S. Gill, W. F. Harrington, S. Pantel, J. M. Krill-Burger, R. M. Meyers, L. Ali, A. Goodale, Y. Lee, G. Jiang, J. Hsiao, W. F. J. Gerath, S. Howell, E. Merkel, M. Ghandi, L. A. Garraway, D. E. Root, T. R. Golub, J. S. Boehm, and W. C. Hahn. Defining a cancer dependency map. *Cell*, 170(3):564–576.e16, 2017.

[141] K. M. Turner, V. Deshpande, D. Beyter, T. Koga, J. Rusert, C. Lee, B. Li, K. Arden, B. Ren, D. A. Nathanson, H. I. Kornblum, M. D. Taylor, S. Kaushal, W. K. Cavenee, R. Wechsler-Reya, F. B. Furnari, S. R. Vandenberg, P. N. Rao, G. M. Wahl, V. Bafna, and P. S. Mischel. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature*, 543(7643):122–125, 2017.

[142] N. T. Umbreit, C.-Z. Zhang, L. D. Lynch, L. J. Blaine, A. M. Cheng, R. Tourdot, L. Sun, H. F. Almubarak, K. Judge, T. J. Mitchell, A. Spektor, and D. Pellman. Mechanisms generating cancer genome complexity from a single cell division error. *Science*, 368(6488):eaba0712, 2020.

[143] M. van Dartel, P. W. A. Cornelissen, S. Redeker, M. Tarkkanen, S. Knuutila, P. C. W. Hogendoorn, A. Westerveld, I. Gomes, J. Bras, and T. J. M. Hulsebos. Amplification of 17p11.2-p12, including pmp22, top3a, and mapk7, in high-grade osteosarcoma. *Cancer Genetics and Cytogenetics*, 139(2):91–96, 2002.

[144] M. van Dartel, S. Leenstra, D. Troost, and T. J. M. Hulsebos. Infrequent but high-level amplification of 17p11.2-p12 in human glioma. *Cancer Genetics and Cytogenetics*, 140(2):162–166, 2003.

[145] G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo. From fastq data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics*, 43(1110):11.10.1–11.10.33, 2013.

[146] E. van Leen, L. Brückner, and A. G. Henssen. The genomic and spatial mobility of extrachromosomal dna and its implications for cancer therapy. *Nature Genetics*, 54(2):107–114, 2022.

[147] N. Vasan, J. Baselga, and D. M. Hyman. A view on drug resistance in cancer. *Nature*, 575(7782):299–309, 2019.

[148] R. G. W. Verhaak, V. Bafna, and P. S. Mischel. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nature reviews. Cancer*, 19(5):283–288, 2019.

[149] P. J. M. Verweij and H. C. Van Houwelingen. Penalized likelihood in cox regression. *Statistics in Medicine*, 13(23-24):2427–2436, 1994.

[150] R. Vicario, V. Peg, B. Morancho, M. Zacarias-Fluck, J. Zhang, A. Martínez-Barriocanal, A. Navarro Jiménez, C. Aura, O. Burgues, A. Lluch, J. Cortés, P. Nuciforo, I. T. Rubio, E. Marangoni, J. Deeds, M. Boehm, R. Schlegel, J. Tabernero, R. Mosher, and J. Arribas. Patterns of her2 gene amplification and response to anti-her2 therapies. *PLoS One*, 10(6):e0129876, 2015.

[151] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3):261–272, 2020.

[152] X. Wang, J. Xu, B. Zhang, Y. Hou, F. Song, H. Lyu, and F. Yue. Genome-wide detection of enhancer-hijacking events from chromatin interaction data in rearranged genomes. *Nat Methods*, 18(6):661–668, 2021.

[153] Z. Wang, L. Fu, J. Zhang, Y. Ge, C. Guo, R. Wang, M. Deng, Q. Wang, and Z. Wang. A comprehensive analysis of potential gastric cancer prognostic biomarker itgbl1 associated with immune infiltration and epithelial-mesenchymal transition. *Biomed Eng Online*, 21(1):30, 2022.

[154] M. L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.

[155] S. M. Waszak, P. A. Northcott, I. Buchhalter, G. W. Robinson, C. Sutter, S. Groebner, K. B. Grund, L. Brugières, D. T. W. Jones, K. W. Pajtler, A. S. Morrissy, M. Kool, D. Sturm, L. Chavez, A. Ernst, S. Brabetz, M. Hain, T. Zichner, M. Segura-Wang, J. Weischenfeldt, T. Rausch, B. R. Mardin, X. Zhou, C. Baciu, C. Lawerenz, J. A. Chan, P. Varlet, L. Guerrini-Rousseau, D. W. Fults, W. Grajkowska, P. Hauser, N. Jabado, Y.-S. Ra, K. Zitterbart, S. S. Shringarpure, F. M. De La Vega, C. D. Bustamante, H.-K. Ng, A. Perry,

T. J. MacDonald, P. Hernáiz Driever, A. E. Bendel, D. C. Bowers, G. McCowage, M. M. Chintagumpala, R. Cohn, T. Hassall, G. Fleischhack, T. Eggen, F. Wesenberg, M. Feychting, B. Lannering, J. Schüz, C. Johansen, T. V. Andersen, M. Röösli, C. E. Kuehni, M. Grotzer, K. Kjaerheim, C. M. Monoranu, T. C. Archer, E. Duke, S. L. Pomeroy, R. Shelagh, S. Frank, D. Sumerauer, W. Scheurlen, M. V. Ryzhova, T. Milde, C. P. Kratz, D. Samuel, J. Zhang, D. A. Solomon, M. Marra, R. Eils, C. R. Bartram, K. von Hoff, S. Rutkowski, V. Ramaswamy, R. J. Gilbertson, A. Korshunov, M. D. Taylor, P. Lichter, D. Malkin, A. Gajjar, J. O. Korbel, and S. M. Pfister. Spectrum and prevalence of genetic predisposition in medulloblastoma: a retrospective genetic study and prospective validation in a clinical trial cohort. *The Lancet. Oncology*, 19(6):785–798, 2018.

[156] J. Wen, J. Lee, A. Malhotra, R. Nahta, A. R. Arnold, M. C. Buss, B. D. Brown, C. Maier, A. M. Kenney, M. Remke, V. Ramaswamy, M. D. Taylor, and R. C. Castellino. Wip1 modulates responsiveness to sonic hedgehog signaling in neuronal precursor cells and medulloblastoma. *Oncogene*, 35(42):5552–5564, 2016.

[157] D. Williamson, Y.-J. Lu, T. Gordon, R. Sciot, A. Kelsey, C. Fisher, C. Poremba, J. Anderson, K. Pritchard-Jones, and J. Shipley. Relationship between mycn copy number and expression in rhabdomyosarcomas and correlation with adverse prognosis in the alveolar subtype. *Journal of Clinical Oncology*, 23(4):880–888, 2005.

[158] M. Wortham, C. Guo, M. Zhang, L. Song, B. K. Lee, V. R. Iyer, T. S. Furey, G. E. Crawford, H. Yan, and Y. He. Chromatin accessibility mapping identifies mediators of basal transcription and retinoid-induced repression of otx2 in medulloblastoma. *PLoS One*, 9(9):e107156, 2014.

[159] G. Wu, A. K. Diaz, B. S. Paugh, S. L. Rankin, B. Ju, Y. Li, X. Zhu, C. Qu, X. Chen, J. Zhang, J. Easton, M. Edmonson, X. Ma, C. Lu, P. Nagahawatte, E. Hedlund, M. Rusch, S. Pounds, T. Lin, A. Onar-Thomas, R. Huether, R. Kriwacki, M. Parker, P. Gupta, J. Becksfort, L. Wei, H. L. Mulder, K. Boggs, B. Vadodaria, D. Yergeau, J. C. Russell, K. Ochoa, R. S. Fulton, L. L. Fulton, C. Jones, F. A. Boop, A. Broniscer, C. Wetmore, A. Gajjar, L. Ding, E. R. Mardis, R. K. Wilson, M. R. Taylor, J. R. Downing, D. W. Ellison, J. Zhang, S. J. Baker, and P. the St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome. The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nature Genetics*, 46(5):444–450, 2014.

[160] S. Wu, K. M. Turner, N. Nguyen, R. Raviram, M. Erb, J. Santini, J. Luebeck, U. Rajkumar, Y. Diao, B. Li, W. Zhang, N. Jameson, M. R. Corces, J. M. Granja, X. Chen, C. Coruh, A. Abnousi, J. Houston, Z. Ye, R. Hu, M. Yu, H. Kim, J. A. Law, R. G. W. Verhaak, M. Hu, F. B. Furnari, H. Y. Chang, B. Ren, V. Bafna, and P. S. Mischel. Circular ecdna promotes accessible chromatin and high oncogene expression. *Nature*, 575(7784):699–703, 2019.

[161] K. Xu, L. Ding, T.-C. Chang, Y. Shao, J. Chiang, H. Mulder, S. Wang, T. I. Shaw, J. Wen, L. Hover, C. McLeod, Y.-D. Wang, J. Easton, M. Rusch, J. Dalton, J. R. Downing, D. W.

Ellison, J. Zhang, S. J. Baker, and G. Wu. Structure and evolution of double minutes in diagnosis and relapse brain tumors. *Acta Neuropathologica*, 137(1):123–137, 2019.

[162] X. Xue, M. Y. Kim, and R. E. Shore. Cox regression analysis in presence of collinearity: an application to assessment of health risks associated with occupational radiation exposure. *Lifetime Data Anal*, 13(3):333–50, 2007.

[163] L. Yang, R. P. Ghosh, J. M. Franklin, S. Chen, C. You, R. R. Narayan, M. L. Melcher, and J. T. Liphardt. Nuset: A deep learning tool for reliably separating and analyzing crowded cells. *PLOS Computational Biology*, 16(9):e1008193, 2020.

[164] E. Yi, R. Chamorro González, A. G. Henssen, and R. G. W. Verhaak. Extrachromosomal dna amplifications in cancer. *Nature Reviews Genetics*, 23(12):760–771, 2022.

[165] Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu. Model-based analysis of chip-seq (macs). *Genome Biol*, 9(9):R137, 2008.

[166] Y. Zhu, A. D. Gujar, C. H. Wong, H. Tjong, C. Y. Ngan, L. Gong, Y. A. Chen, H. Kim, J. Liu, M. Li, A. Mil-Homens, R. Maurya, C. Kuhlberg, F. Sun, E. Yi, A. C. deCarvalho, Y. Ruan, R. G. W. Verhaak, and C. L. Wei. Oncogenic extrachromosomal dna functions as mobile enhancers to globally amplify chromosomal transcription. *Cancer Cell*, 39(5):694–707.e7, 2021.

[167] N. Zhukova, V. Ramaswamy, M. Remke, E. Pfaff, D. J. H. Shih, D. C. Martin, P. Castelo-Branco, B. Baskin, P. N. Ray, E. Bouffet, A. O. von Bueren, D. T. W. Jones, P. A. Northcott, M. Kool, D. Sturm, T. J. Pugh, S. L. Pomeroy, Y.-J. Cho, T. Pietsch, M. Gessi, S. Rutkowski, L. Bognar, A. Klekner, B.-K. Cho, S.-K. Kim, K.-C. Wang, C. G. Eberhart, M. Fevre-Montange, M. Fouladi, P. J. French, M. Kros, W. A. Grajkowska, N. Gupta, W. A. Weiss, P. Hauser, N. Jabado, A. Jouvet, S. Jung, T. Kumabe, B. Lach, J. R. Leonard, J. B. Rubin, L. M. Liau, L. Massimi, I. F. Pollack, Y. Shin Ra, E. G. Van Meir, K. Zitterbart, U. Schüller, R. M. Hill, J. C. Lindsey, E. C. Schwalbe, S. Bailey, D. W. Ellison, C. Hawkins, D. Malkin, S. C. Clifford, A. Korshunov, S. Pfister, M. D. Taylor, and U. Tabori. Subgroup-specific prognostic implications of tp53 mutation in medulloblastoma. *Journal of Clinical Oncology*, 31(23):2927–2935, 2013.