Extrachromosomal DNA associates with poor survival across a broad spectrum of childhood solid tumors

Owen S. Chapman[1,2,3*], Sunita Sridhar[4,5*], Eugene Yui-Ching Chow[2], Rishaan Kenkre[2], Jonathan Kirkland[3], Aditi Dutta[6], Shanqing Wang[6], Wenshu Zhang[3], Miguel Brown[7], Jens Luebeck[6], Yan Yuen Lo[2,8], Elias Rodriguez-Fos[9], Anton G. Henssen[9], Konstantin Okonechnikov[10,11,12], David R. Ghasemi[10,13,14,15], Kristian W. Pajtler[10,11,12,16], Daisuke Kawauchi[1], Vineet Bafna[6,17,18], Megan Paul[4,5], Kevin Yip[2], Jill P. Mesirov[17,19], Lukas Chavez[2,8,19,+]

1. Department of Neuro-oncology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan;
2. Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA;
3. Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA, USA;
4. Rady Children's Hospital San Diego, San Diego, CA, USA;
5. Department of Pediatrics, University of California San Diego, La Jolla, CA, USA;
6. Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA;
7. Children's Hospital of Philadelphia, Philadelphia, PA, USA;
8. Rady Children's Institute for Genomic Medicine, Rady Children's Hospital and Healthcare Center, San Diego, CA, USA;
9. Experimental and Clinical Research Center of the Max Delbrück Center and Charité Berlin, Berlin, Germany;
10. Hopp Children's Cancer Center Heidelberg (KiTZ), Heidelberg, Germany;
11. Division of Pediatric Neurooncology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany;
12. National Center for Tumor Diseases (NCT) Heidelberg, a partnership between DKFZ and Heidelberg University Hospital, Heidelberg, Germany;
13. Department of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany;
14. Research Institute Children's Cancer Center, Hamburg, Germany;
15. Mildred Scheel Cancer Career Center HaTriCS[4], University Medical Center Hamburg-Eppendorf, Hamburg, Germany;
16. Department of Pediatric Oncology, Hematology and Immunology, Heidelberg University Hospital, Germany;
17. Moores Cancer Center, University of California San Diego, La Jolla, CA, USA;
18. Halıcıoğlu Data Science Institute, University of California San Diego, La Jolla, CA, USA;
19. Department of Medicine, University of California San Diego, La Jolla, CA, USA;

[+] Corresponding author

* Co-first authors

ABSTRACT

Circular extrachromosomal DNA (ecDNA) is a common form of oncogene amplification in aggressive cancers. The frequency and diversity of ecDNA has been catalogued in adult and some childhood cancers; however, its role in most pediatric cancers is not well-understood. To address this gap, we accessed large pediatric cancer genomics data repositories and identified ecDNA algorithmically using cloud computing. This retrospective cohort comprises 2,968 patients covering all major childhood solid tumor types. Aggressive tumor types had particularly high incidences of ecDNA. Pediatric patients whose tumors harbored extrachromosomal DNA had poorer five-year overall survival than children whose tumors contain only chromosomal amplifications. We catalogue known and potentially novel oncogenes recurrently amplified on ecDNA and show that ecDNA often evolves during disease progression. These results highlight patient populations that could potentially benefit from future ecDNA-directed therapies. To facilitate discovery, an interactive catalogue of ecDNA in childhood cancer is available at https://ccdi-ecdna.org/.

INTRODUCTION

Cancer can develop through a variety of mechanisms, most commonly from inactivation of tumor suppressor genes or activation of proto-oncogenes. Circular extrachromosomal DNA (ecDNA), also known as double minutes, allows for particularly high amplification of oncogenes. Because they lack centromeres, ecDNA is subject to unequal segregation during cell division leading to increased intra-tumoral heterogeneity[1]. Molecular characterizations have shown that ecDNA in cancer vary greatly in size, most being 0.5 Mb or larger[2] and have greater chromatin accessibility compared to chromosomal DNA, promoting increased transcription[3]. Paired-end whole genome sequencing has emerged as a cost-effective and efficient method to identify oncogenes and other sequences captured on ecDNA[4]. While the analysis of ecDNA was largely overlooked in early large cancer genomics studies[5,6], recent studies have described the prevalence of ecDNA in various adult and some childhood cancer types[7–10]. Recognizing that pediatric cancers have a 14-fold lower somatic mutation frequency than adult cancers as well as unique driver mutations and amplified oncogenes[11,12], we have now comprehensively examined the frequency of ecDNA and its association with survival across a broad spectrum of primary and recurrent pediatric cancers.

RESULTS

Cloud genomic data repositories facilitate analysis of rare pediatric cancers.

To identify ecDNA in pediatric cancers, we analyzed whole genome sequencing (WGS) data available in two large pediatric cancer cloud genomics platforms: the Pediatric Brain Tumor Atlas (PBTA)[13] and St. Jude Cloud (SJC)[14] (**Figure 1**). In total, this retrospective cohort comprised WGS data of 3,631 solid tumor biosamples from 2,968 patients. Clinical metadata, including histological diagnoses and patient survival outcomes, were collated from both platforms. For brain tumors included in the PBTA, histological diagnoses were integrated with

molecular tumor type classifications derived from methylation analyses[15] of the same tumor biopsies. We applied the algorithms AmpliconArchitect (AA)[16] and AmpliconClassifier (AC)[17] to identify ecDNA amplifications from structural and copy number variation in tumor genomes. Briefly, these algorithms perform joint analysis of genomic copy number and structural variation in WGS data to construct a localized genome graph capturing the amplified regions and identify ecDNA as genome cycles satisfying other heuristics including length and copy number. In total, 392 ecDNA sequences were reported in 321 tumor biosamples of 269 patients (9% of patients, **Supplementary Tables 1-3**). No association was observed between presence of ecDNA and age at diagnosis ($R^2 = 0.005$, $p = 0.4$, Student's *t*-test; **Supplementary Figure 1**) or sex ($\chi^2 = 1.3$, $p = 0.3$). As a quality control, we estimated tumor purity and observed that tumor purity is not significantly predictive of ecDNA (likelihood ratio test, p = 0.12, n = 2089; see Methods). Given that previous studies using similar methods did not find ecDNA in adult hematologic malignancies[7,9], we additionally sampled 10% of available pediatric hematologic malignancies (n = 63, **Supplementary Table 4**). No ecDNA was detected in those samples, and we therefore limited the scope of our investigation to solid tumors.

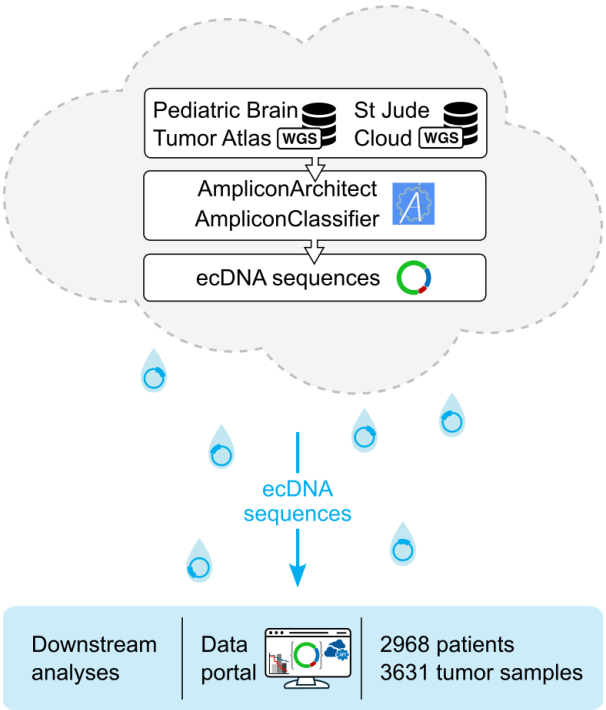


Figure 1: Analysis of ecDNA in pediatric cancers using cloud computing. Whole genome sequencing (WGS) data of solid tumors was analyzed in two cancer genome cloud repositories. ecDNA and other amplifications were identified and annotated with clinical information for downstream analysis.

ecDNA amplifies known and putative oncogenic sequences in pediatric solid tumors.

To identify putative oncogenic drivers amplified on ecDNA in pediatric solid tumors, we first defined recurrently amplified genomic sequences as those amplified on ecDNA in at least three independent cases, representing a 99.9% confidence interval above which the recurrent amplifications are unlikely to have occurred by chance (**Supplementary Note 1**; **Supplementary Figure 2**). 109 distinct genomic sequences were recurrently amplified on ecDNA across all pediatric cancers in this cohort (**Figure 2a**). To identify putative driver genes, we annotated these ecDNA sequences with oncogenes listed in the COSMIC[18] and ONGene[19] databases (**Supplementary Table 5**). The most frequently amplified ecDNA sequences contained oncogenes *MYCN* (66 ecDNA amplifications observed in eight tumor types) and *CDK4* (22 amplifications in six tumor types; **Figure 2b**). Five contiguous recurrently amplified ecDNA sequences contained clusters of three or more oncogenes: chr4q12 (*PDGFRA*, *KIT*, and *KDR*), chr8q24 (*TRIB1*, *MYC*, *PVT1*), chr11q22 (*YAP1*, *BIRC3*, *BIRC2*, *MMP12*), chr12q13-14 (*GLI1*, *DDIT3*, *AGAP2*, *CDK4*), and chr12q15 (*IFNG*, *MDM2, YEATS4*).

82 recurrently amplified ecDNA sequences contained no known oncogene. These ecDNA sequences were relatively short (mean 181kbp, standard deviation 301kbp) and 36 contained uninterrupted gene sequences, nominating 81 protein-coding genes and 235 long noncoding

RNAs (lncRNA) as candidate oncogenic drivers (**Supplementary Table 6**). For example, a peritelomeric recurrently amplified locus at chromosome 1q amplified the full gene sequences of *AKT3* and *ZBTB18* in two high-grade gliomas and a rhabdomyosarcoma (**Supplementary Figure 3**). In mouse models, amplification and overexpression of the paralogous gene *Akt3*

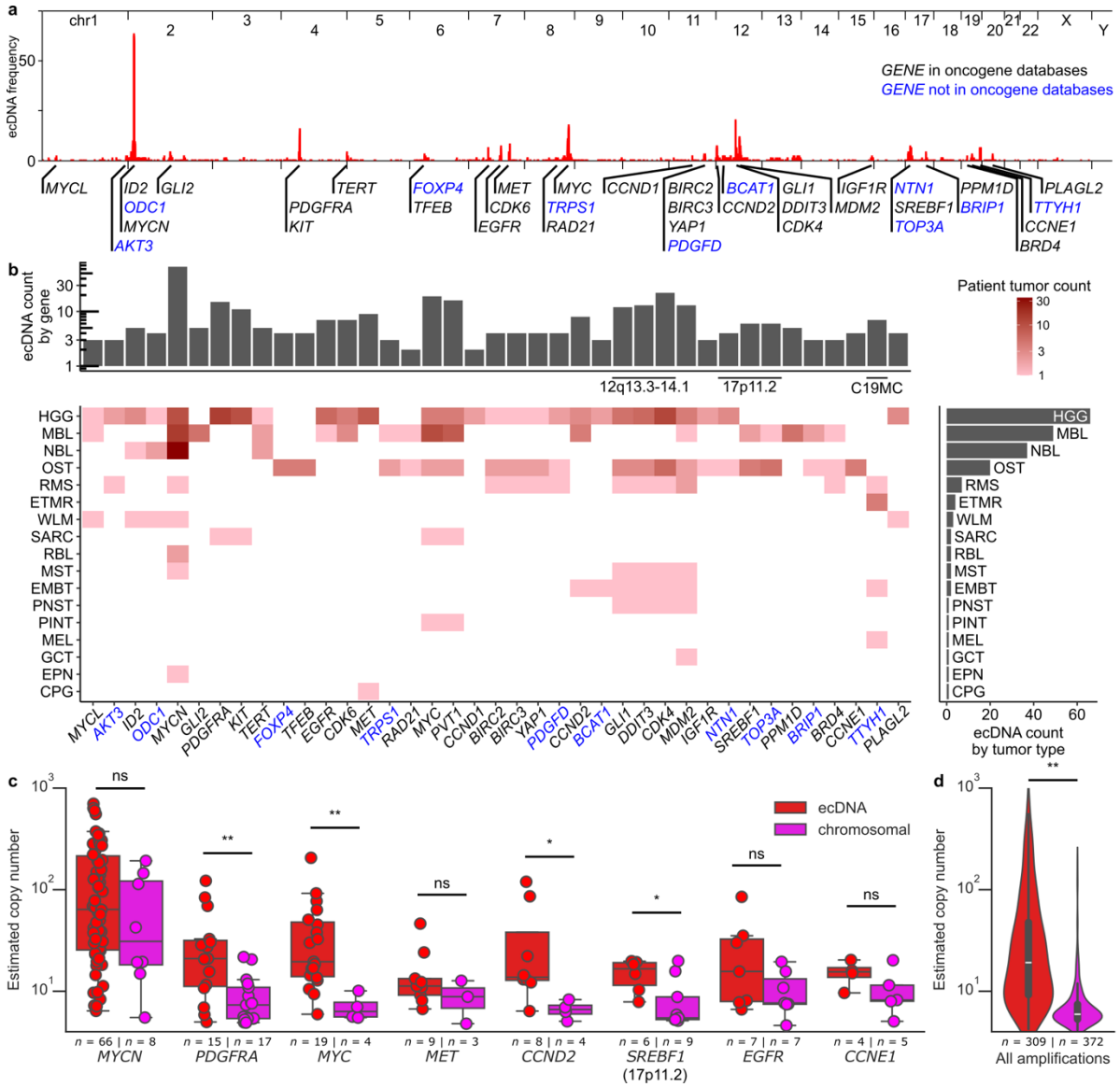


Figure 2: Distribution of recurrent ecDNA amplifications in pediatric tumors. (**a**) Histogram of ecDNA amplifications in pediatric tumors across the reference human genome. 106 contiguous genomic loci were recurrently ($n \geq 3$) amplified containing 49 different oncogenes. Selected genes are shown: black, oncogenes listed in the COSMIC or ONGene databases; blue, recurrently amplified genes with putative oncogenic roles. (**b**) Genes of interest recurrently amplified on ecDNA in pediatric tumors. Heatmap indicates the number of cases with ecDNA amplification of a given gene in that tumor type. Top barplot indicates the number of ecDNA sequences amplifying any of the indicated genes for each cancer type. Right barplot indicates the number of ecDNA sequences amplifying the indicated gene across all tumor types. (**c**) Estimated copy number of extrachromosomal and chromosomal amplifications of selected oncogenes. ** $q < 0.005$; * $q < 0.05$; ns, $q \geq 0.05$; one-sided Mann-Whitney U test with Benjamini-Hochberg FDR correction. (**d**) Estimated copy number of amplicons containing at least 1 gene. Legend and statistical tests are the same as in (c). HGG, high-grade glioma; MBL, medulloblastoma; NBL, neuroblastoma; OST, osteosarcoma; RMS, rhabdomyosarcoma; ETMR, embryonal tumor with multilayered rosettes; WLM, Wilm's tumor; EMBT, miscellaneous embryonal brain tumors; GCT, germ cell tumors; MST, metastatic secondary tumors; RBL, retinoblastoma; CPG, craniopharyngioma; EPN, ependymoma; MEL, melanoma; PINT, pineal tumors; PNST, peripheral nerve sheath tumors; SARC, other miscellaneous sarcomas.

promotes malignant glioma progression[20,21], supporting a functional role for this gene locus in some human pediatric tumors. Other genes repeatedly amplified on ecDNA which may have tumorigenic roles included DNA replication initiation cofactor *TRPS1*[22], amino acid metabolic factor *BCAT1*[23], neurodevelopmental factor *NTN1*[24], and MYC pathway member *ODC1*[25]. Accumulating evidence suggests that ecDNA enables greater copy number amplification than chromosomal amplification[9]. To determine whether amplification by ecDNA is of greater magnitude than other amplifications, we compared estimated copy number for oncogenes amplified in $n \geq 3$ cases on ecDNA and chromosomal amplifications (**Figure 2c**). Median ecDNA copy number was greater for *PDGFRA*, *MYC*, *CCND2*, and chr17p11.2 compared to chromosomal amplification of the same genes in other tumors (one-sided Mann-Whitney *U* with Benjamini-Hochberg FDR correction, $q < 0.05$). Median ecDNA copy number was also greater across all amplifications classified as ecDNA compared to chromosomal (**Figure 2d**), suggesting that oncogenic amplification occurs at greater magnitude in ecDNA than in chromosomal amplifications in pediatric cancers.

ecDNA is associated with poor survival in pediatric cancers.

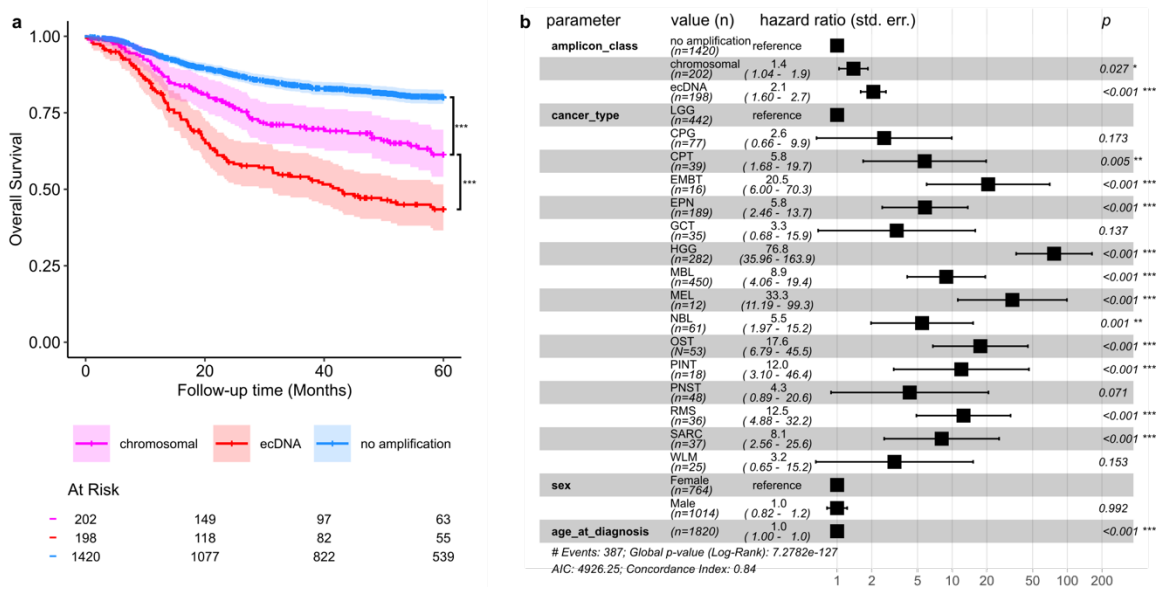


Figure 3: Overall survival of pediatric patients with ecDNA(+) tumors. Sample was limited to tumor types with at least 10 cases, at least 1 ecDNA(+) sample, and at least 1 clinical event. (**a**) Patients with ecDNA(+) tumors have significantly worse outcomes than patients with tumors containing no amplification (adjusted p = 2e-16) or patients with chromosomal amplifications (adjusted p = 2e-4; log-rank test with Benjamini-Hochberg correction). (**b**) Cox regression performed which indicated twofold risk of death for patients with ecDNA compared to no amplification (hazard ratio = 2.1, n = 1820, p < 0.001). ***$q<0.001$; **$q<0.01$; *$q<0.05$; ns, $q\geq0.05$.

To evaluate the relationship between ecDNA and survival, we performed Kaplan-Meier regression on overall survival data available for 1820 patients (**Supplementary Table 1**). To investigate whether ecDNA in pediatric cancers are associated with worse survival than other types of focal amplifications, we stratified ecDNA(-) tumors into those with and without chromosomal amplification. Kaplan-Meier regressions indicated that patients with ecDNA(+) tumors had significantly worse outcomes than patients with tumors containing no amplification (log-rank test, adjusted p = 2e-16) and patients with chromosomal amplifications (adjusted p = 2e-4) (**Figure 3a**). To estimate the magnitude of the association between ecDNA and survival, we performed Cox regression of sex, age at diagnosis, tumor type, and amplification type (ecDNA, chromosomal, or no amplification). Controlling for available variables, ecDNA

157 detected in a patient tumor conferred twofold risk of death within 5 years relative to tumors
158 without amplification (hazard ratio 2.1, $p < 0.001$, log-rank test), and one-and-a-half-fold risk
159 relative to tumors with chromosomal amplification but no ecDNA (HR 1.5, $p = 0.02$, **Figure**
160 **3b**).

161 **ecDNA in aggressive childhood central nervous system tumors**

162 ecDNA has been previously described in some types of pediatric brain tumor types, most
163 extensively in medulloblastoma[8]. We observe ecDNA most frequently in embryonal tumors
164 with multilayered rosettes (4/9, 44%), pediatric high-grade gliomas (78/391, 20%), and
165 medulloblastomas (55/347, 16%); and at low frequencies in pineal tumors (1/28, 4%), choroid
166 plexus tumors (1/56; 2%), craniopharyngiomas (1/101, 1%), ependymomas (3/239, 1%),
167 glioneuronal tumors (1/192, 0.5%), and low-grade gliomas (1/599, 0.2%) (**Figure 4a**).

168 Embryonal tumors with multilayered rosettes (ETMR) have poor outcomes and most are
169 characterized by amplification of a microRNA cluster on chromosome 19 (C19MC) including
170 fusion of the Tweety Family Member 1 (*TTYH1*) gene[26]. Of 9 ETMR cases, 4 had C19MC-
171 amplification, including 3'-truncated *TTYH1*, all of which were extrachromosomal. Although
172 the number of observations is low, these cases suggest that C19MC amplification in ETMR
173 tumors is often extrachromosomal and may contribute to the highly aggressive behavior of
174 this rare tumor type.

175 The clinical outcomes of pediatric high-grade glioma (pHGG) patients differ across
176 molecular subgroups marked by genetic mutations to the histone H3 or isocitrate
177 dehydrogenase (*IDH*) genes[27]. pHGG with H3 mutations are further stratified into those
178 mutated at lysine 27 (H3K27) and at glycine 34 (H3G34). Among 391 pHGG tumors, 78 had
179 ecDNA (20%), including 29 of 103 with wild-type H3 (28%), 29 of 134 H3K27-mutant (22%),
180 2 of 16 H3G34-mutant (13%), and 4 of 10 with *IDH* mutation (40%). Known oncogenes[27–29]
181 recurrently amplified on ecDNA included *PDGFRA* ($n = 14$), *CDK4* (12), *MYCN*, *KIT* (10), *MET*
182 (8), *EGFR* (6), *GLI1* (5), *CDK6, MYC, MDM2* (4), *ID2, CCND2, and PLAGL2* (3). It was
183 previously shown that complex structural variants are associated with poorer overall survival
184 of pHGG patients[30]. However, the relevance of ecDNA versus chromosomal amplification for
185 the survival of pHGG patients in the context of *IDH* or oncohistone mutations remains to be
186 explored in a larger cohort of pHGG tumors.

187 In a previous analysis of 481 medulloblastoma (MBL) tumors, we showed significant
188 associations between ecDNA status and survival across all MBL tumors and in each molecular
189 subgroup[8]. 168 MBL tumors were reanalyzed herein alongside 174 new cases. ecDNA was
190 detected in 16%, modestly lower than that reported previously (18%), which may be attributed
191 to differences in the patient cohorts and analytical methods (**Supplementary Note 2**).

192 The low frequency of ecDNA detected in choroid plexus tumors, pineal tumors,
193 craniopharyngiomas, ependymomas, CNS glioneuronal tumors, and low-grade gliomas
194 (**Supplementary Note 3**) prompts the question of whether ecDNA amplifications are
195 incidental findings or characteristic to well defined rare molecular subtypes. For pineal tumors
196 and ependymomas with available DNA methylation-based classifications, ecDNA
197 amplifications were indeed limited to the recently-described rare molecular subtypes *MYC*-
198 amplified pineoblastoma[31] and *MYCN*-amplified spinal ependymoma[32] respectively. To
199 validate these findings, we obtained new WGS data from 3 *MYCN*-amplified spinal
200 ependymomas. Amplicon analysis identifies ecDNA in each, suggesting that amplification of
201 *MYCN* in this molecular subtype may be commonly driven by ecDNA. Overall, although further
202 studies of larger tumor-specific cohorts are needed, these observations suggest that ecDNA

amplification may be limited to some ultra-rare and highly aggressive subtypes and are otherwise largely absent in these tumor types.

ecDNA in pediatric extracranial solid tumors

ecDNA has been previously described in various extracranial pediatric solid tumors, most extensively in neuroblastoma[10,33], and in case reports and cell lines of retinoblastoma[34,35], rhabdomyosarcoma[28,36–41], and osteosarcoma[42]. We identify patient tumors with ecDNA for each of these extracranial tumor types, as well as in adrenocortical carcinomas, Wilm's tumors, germ cell tumors, malignant peripheral nerve sheath tumors, and various other rare sarcomas (**Figure 4b**).

Consistent with a previous analysis[10], we estimate the prevalence of ecDNA in neuroblastoma at 40/159 (25%). Most patients with ecDNA had *MYCN* amplified on ecDNA (35/40, 88%). Of 5 remaining ecDNA(+) tumors not amplifying *MYCN*, two had ecDNA amplifications of *TERT*, and 3 had low-copy ecDNA containing no known oncogenic sequences.

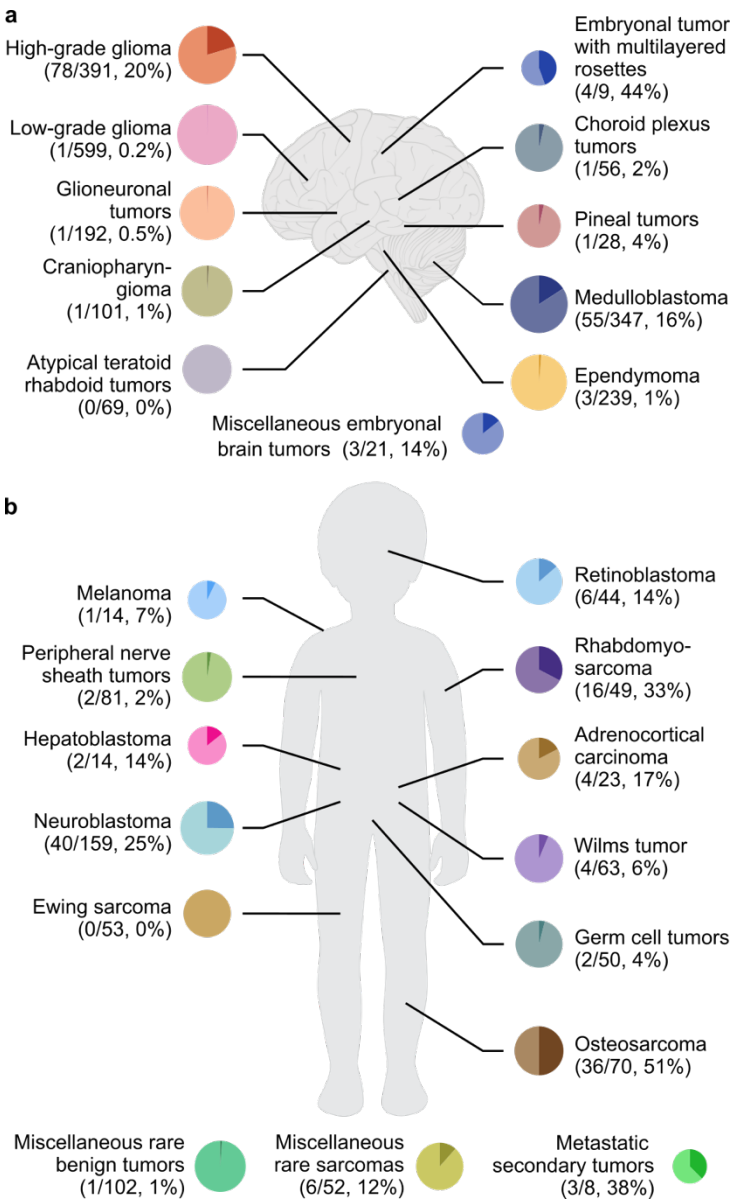


Figure 4: Prevalence of ecDNA across pediatric tumor types. Across 46 pediatric tumor categories, we report ecDNA in 23. Pie chart area is proportional to the log of the sample size for a given tumor.

Although *RB1* gene inactivation is the canonical driver of retinoblastoma, other mechanisms including oncogenic amplification are believed to contribute to tumorigenesis in some cases[43]. ecDNA has previously been described in retinoblastoma cell lines and patient tumors and is particularly associated with *MYCN* amplification[35]. In this cohort, six of 44 (14%) retinoblastoma cases had ecDNA, two of which amplified *MYCN*. A further two ecDNA amplifications disrupted the *RB1* gene to form gene fusions, with lipoma *HMGIC* fusion partner gene family member *LHFPL6* (*RB1-LHFPL6)* or with the *TPTE2P2* pseudogene (*RB1-TPTE2P2)*. In the remaining two tumors, no canonical oncogenes were identified, suggesting that these highly amplified ecDNA sequences may contain previously unknown oncogenic drivers for this tumor type.

Osteosarcomas are frequently characterized by high genomic instability with some genomic regions frequently involved in copy number gain events[44,45]. Prognosis is poor with

few advances in standard treatment for the past 30 years[46]. ecDNA was detected in 36 of 70 cases (51%). We identified recurrent ecDNA amplifications at chromosomes 6p21 (*FOXP4*[47], 4 cases), 12q13 (*GLI1*, *CDK4*, 6 cases), 17p11.2-p12[48,49] (11 cases), and 19q12 (*CCNE1*, 4 cases). 3 tumors had more than one of these recurrently amplified regions on distinct ecDNA sequences.

Double minutes have been described in case reports of rhabdomyosarcoma (RMS) as early as 1971[36,37,39,50,51]. Amplifications of *MYCN*[52] and fusion oncogenes between *FOXO1* and *PAX3* or *PAX7*[53] have been previously identified as prognostic indicators. ecDNA was detected in 16 of 49 RMS cases (33%). Amplified oncogenes included *PAX7-FOXO1* fusion, *MDM2* (*n* = 2), *MYCN*, *FGFR1*, *NCOA1-PAX3* fusion, and *NSD3* (*n* = 1). One tumor harbored ecDNA where none of the amplified genes were annotated in the two queried oncogene databases; however, we suggest the cell cycle regulators *CDK11A* and *CDK11B* as likely oncogenic candidates.

In contrast to a previous adult pan-cancer study[7], we did not observe ecDNA in any of 53 pediatric Ewing sarcomas. On further investigation, we noted that tumors previously annotated as Ewing sarcomas have been subsequently reannotated as osteosarcomas (**Supplementary Table 7**), explaining the discrepancy and suggesting that ecDNA does not play a major role in oncogenesis of pediatric Ewing sarcomas.

We also report ecDNA sequences in 4 adrenocortical carcinomas, 4 Wilms tumors, 2 germ cell tumors, 2 malignant peripheral nerve sheath tumors, 2 hepatoblastomas, one pediatric melanoma, and one hepatic focal nodular hyperplasia. These cases are described in **Supplementary Note 4**.

ecDNA evolves during progression and recurrence of pediatric cancers.

We next sought to understand the differences in prevalence of ecDNA in relapsed compared to primary tumors, and how ecDNA evolves in longitudinal cases of paired primary/relapse samples. For this analysis, we aggregated progressive, recurrence, relapse, and metastasis annotations into one category of secondary samples. Secondary malignancies and autopsy samples were excluded. ecDNA was modestly but nonsignificantly more frequent in secondary (57 of 511, 11%) than primary tumors (209 of 2558, 8%) ($\chi^2 = 3.4; p = 0.06$).

To characterize patterns of ecDNA evolution, we collated 29 longitudinal cases where ecDNA was detected in at least one of sequential tumor samples of the same patient (**Supplementary Table 8**). Here, we included an ecDNA(+) medulloblastoma patient tumor diagnosed at Rady Children's Hospital, San Diego[8] and for which we generated new WGS data from the relapse. In these 29 longitudinal cases, we observed a total of 39 distinct ecDNA sequences. Of these, 13 ecDNA arose *de novo* compared to earlier tumor samples (gain) and 10 ecDNA were not detected in a later tumor sample (loss). Two chromosomal amplifications were identified as ecDNA in a later tumor sample (circularization). We do not observe examples of ecDNA amplification later reclassified as non-ecDNA chromosomal amplification. For extrachromosomal amplifications detected in the same patient at more than one timepoint, we asked whether the ecDNA had undergone structural variation in the interim, using a heuristic similarity metric incorporating Jaccard similarities of the amplified sequences and breakpoints (see Methods). Five such ecDNA amplicons were conserved ($S > 0.7$) whereas the other nine showed structural evolution, including examples where a new oncogene was added (**Figure 5a**) or lost (**Supplementary Figure 4**) in the later tumor. For example, in a pHGG patient (SJHGG019), we observe the oncogenes *EGFR* and *CDK6*, both coded on chr7, co-amplified on the same ecDNA. The ecDNA observed in the relapse tumor of the same

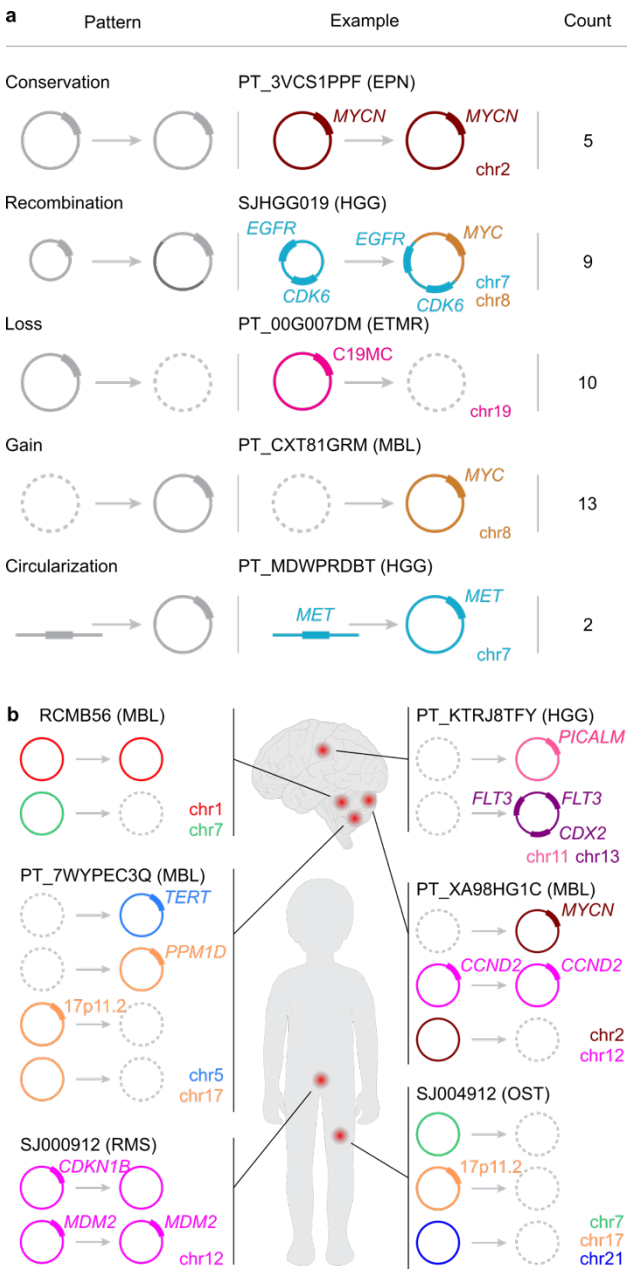


Figure 5: Sequence evolution of ecDNA during disease progression. (**a**) 39 ecDNA sequences were observed longitudinally in 29 tumors. We classified these based on ecDNA sequence over time: conservation (no change), recombination (sequence addition or loss to an extant ecDNA), loss (no amplification observed in the latter sample), gain (no amplification in the earlier sample), or circularization (chromosomal amplification at the same locus in the earlier sample). No examples were observed of linearization (chromosomal amplification at the same locus in the later sample). (**b**) Multiple ecDNA sequences were observed longitudinally in 6 tumors. Amplified oncogenes are annotated. EPN, ependymoma; ETMR, embryonal tumor with multilayered rosettes; HGG, high-grade glioma; MBL, medulloblastoma; RMS, rhabdomyosarcoma; OST, osteosarcoma.

patient contained *MYC*, coded on chr8, as a third oncogene on the same ecDNA together with *EGFR* and *CDK6* (**Figure 5a**). Overall, these results suggest clonal tumor dynamics in which ecDNA may emerge, mutate, remain conserved, or drop out during disease progression.

Patient tumors with more than one ecDNA sequence showed longitudinal variability in the combinations of ecDNA sequences detected. For example, we previously observed two distinct ecDNAs, originating from chromosomes 1 and 7 respectively, present in a primary SHH medulloblastoma tumor (patient ID RCMB56[8]). In WGS data from the recurrent tumor of the same patient, we observe conservation of the ecDNA amplification originating from chromosome 1, whereas the ecDNA amplification originating from chromosome 7 was no longer detected in the relapse tumor. This and other examples (**Figure 5b**) highlight that multiple distinct ecDNA sequences provide tumors with flexibility to amplify different oncogenes and oncogene combinations during disease progression.

DISCUSSION

By leveraging childhood cancer data cloud repositories with whole-genome sequencing data, we estimate frequencies of circular extrachromosomal amplification across a wide spectrum of childhood cancer types. Where grading of the tumor was incorporated into the histological or molecular diagnosis, the higher-grade diagnoses (pediatric high-grade glioma, choroid plexus carcinoma, malignant peripheral nerve sheath tumor, and pineoblastoma) had greater incidence of ecDNA than the lower-grade diagnosis (pediatric low-grade glioma, choroid plexus papilloma, neurofibroma, and low-grade pineal tumors), suggesting that

ecDNA may be more frequent in pediatric tumors with high-grade histology across tissue origins. Notable high-grade tumor types in which we did not observe ecDNA are defined by well described oncogenic mutations which do not require amplification, such as *EWS-FLI* gene fusions in Ewing sarcomas, in which amplification is deleterious[54], or biallelic loss of *SMARCB1/SMARCA4* in atypical teratoid rhabdoid tumors (ATRT)[55]. However, it remains unclear why other oncogenic drivers, such as *ZFTA* fusions in ependymomas, were not observed on ecDNA. Although we did not comprehensively analyze hematological cancers, a pediatric case report[56] and a recent retrospective analysis of adult leukemias[57] recommend this area for further investigation.

Patients with tumors containing ecDNA had significantly poorer survival compared to those with chromosomal amplification and those without amplification, controlling for sex, age and tumor type. This finding goes beyond previous retrospective analyses of pediatric neuroblastomas[10], medulloblastomas[8], and adult cancers[9], as these earlier studies did not establish the relative importance of extrachromosomal versus chromosomal amplification. Our results suggest that both chromosomal and extrachromosomal amplifications have nonredundant prognostic value in pediatric cancers. It remains to be shown whether ecDNA interacts with other known predictors of patient outcomes including extent of resection, which were not universally available for this cohort.

The majority of ecDNA(+) tumors contained at least one well described oncogene, such as *MYC, MYCN*, *CDK4*, and others. In some tumors, multiple oncogenes were amplified on one or multiple ecDNA sequences. Other ecDNA amplifications contained genes not annotated in the COSMIC[18] and ONGene[19] oncogene databases but with plausible oncogenic roles, including *AKT3, NTN1, CDK11A/B* and *SMARCA5* and others. In rhabdomyosarcoma and retinoblastoma, we found recurrent gene fusions on ecDNA. The potential oncogenic roles of still many other ecDNA-amplified sequences without known oncogenes are less clear. For example, relatively little is known of the role of chr17p11.2 amplification, the most frequent ecDNA amplification in osteosarcomas. Intriguingly, this locus is a recombination hotspot in medulloblastoma[58] and in some nonmalignant genetic disorders[59,60]. Given the current poor prognosis of osteosarcomas and the lack of more effective treatment protocols, further analysis of this and other recurrently amplified loci are warranted to identify novel oncogenic drivers and treatment targets.

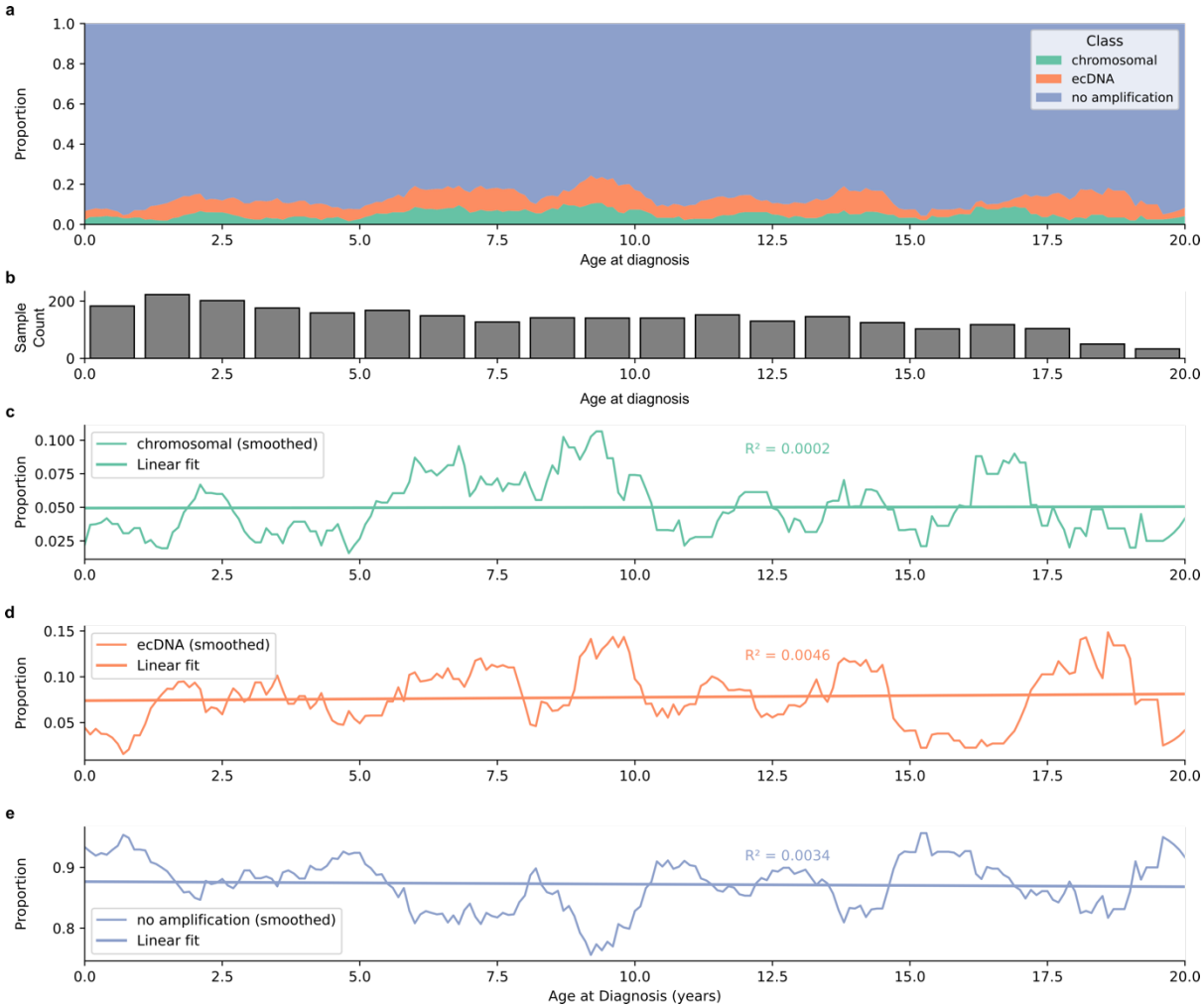
Across a set of paired primary/secondary cases, we observed structural rearrangements of ecDNA sequences. This adds to accumulating evidence[1,8] that ecDNA facilitates tumor evolution in response to therapeutic or other microenvironmental selective pressures. Secondary tumors, including recurrences, progressions and metastases, contained ecDNA more frequently than primary tumors, although the difference was nonsignificant. This trend may stem from the observation that low-grade and benign tumor diagnoses rarely contained ecDNA and rarely progressed. Due to limited treatment information available for this cohort, the role of extrachromosomal DNA evolution during disease progression and the emergence of drug resistance under standard-of-care treatment including chemo- and radiotherapy remains to be further elucidated and functionally tested in pediatric solid tumor models.

Identification of ecDNA in patient tumors holds promise in prognostication and targeted treatment. Although the spectrum of known and putative oncogenic sequences on ecDNA is diverse and sometimes tissue-specific, there is some hope that patients with ecDNA(+) tumors may benefit from future therapies targeting a common vulnerability. For example, a recent study demonstrated that combination treatment of an ecDNA(+) gastric tumor using inhibitors against the ecDNA-amplified *FGFR* oncogene and a CHK1 inhibitor enhances transcription-replication conflict and prevents ecDNA-mediated acquired resistance that otherwise

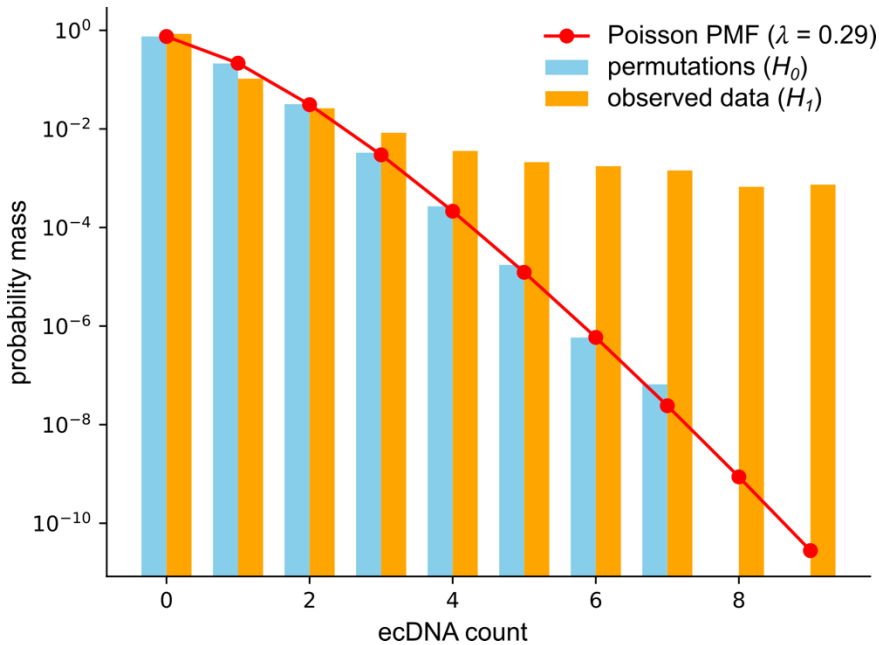
372 develops when using the FGFR inhibitor alone. Thus, transcription-replication conflict
373 emerges as a target for ecDNA-directed therapy, exploiting a synthetic lethality of high ecDNA
374 copy number to treat cancer[61]. Our results show that this approach may also be applicable to
375 some high-risk pediatric tumors. For example, we recently reported preliminary clinical real-
376 world experience using avapritinib, a next-generation tyrosine kinase inhibitor, in pediatric and
377 young adult patients with recurrent/refractory PDGFRA-altered HGG[62]. Avapritinib was well
378 tolerated and led to initial radiographic responses in 3 out of 7 cases. In one patient, we
379 observed *de novo EGFR* amplification in a metastatic lesion that progressed during treatment,
380 indicating *oncogene switching* as a possible resistance mechanism. A promising avenue for
381 future investigation is whether combination therapy with a CHK1 inhibitor may help prevent
382 ecDNA-mediated resistance by exacerbating transcription-replication conflicts, leading to
383 more durable response.

384 Overall, this study highlights the spectrum of pediatric cancer patients with poor prognosis
385 who may benefit from such ecDNA-directed therapies. To facilitate future discovery, the
386 dataset is available through interactive web portals at https://ccdi-ecdna.org/ and
387 https://ampliconrepository.org/project/PedPanCan.

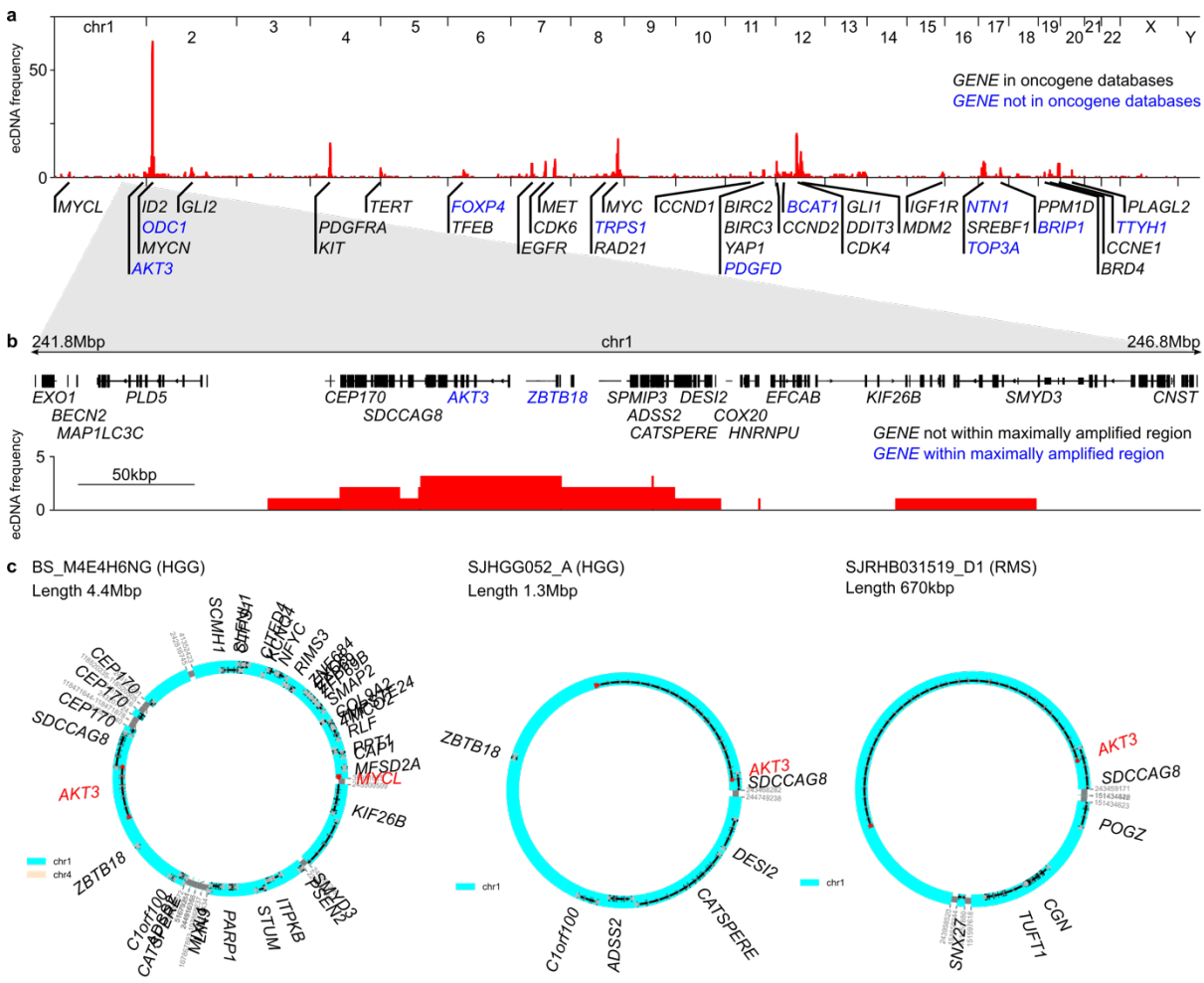
SUPPLEMENTARY FIGURES



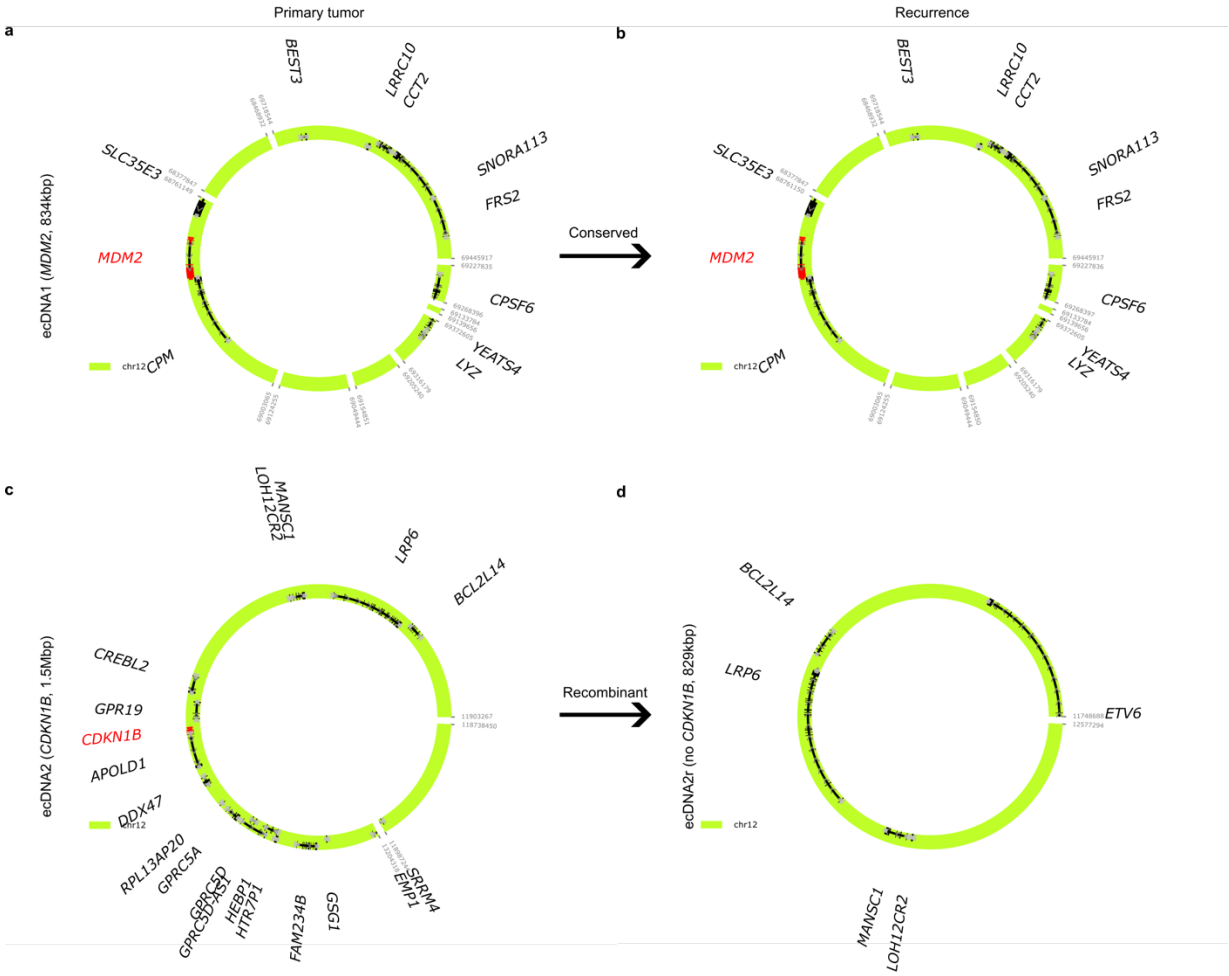
Supplementary Figure 1: Ordinary least squares regression of focal amplifications in patient tumors on age at diagnosis. (**a**) Ages at diagnosis were binned to 0.1-year resolution, and class proportions (ecDNA, chromosomal amplification, or no amplification) were computed for each age group. To reduce sampling noise, a centered 1-year rolling average was applied to the age-wise amplicon class distributions. (**b**) Distribution of sample counts by age. (**c**-**e**) Ordinary least squares linear regressions for proportion of tumors with chromosomal amplification ($R^2 = 2\mathrm{e}-4$, $p = 0.84$, Student's *t*-test), ecDNA ($R^2 = 0.005$, $p = 0.38$), or no amplification ($R^2 = 0.003$, $p = 0.41$).



Supplementary Figure 2: Expected and observed distributions of ecDNA across the human reference genome. A null hypothesis H_0 in which ecDNA is uniformly and randomly distributed across the genome was modelled explicitly as a Poisson distribution with rate parameter λ proportional to the number and average length of ecDNA sequences (red circles) and experimentally by shuffling the locations of ecDNA across the genome (blue bars). 99.9% of the probability mass of H_0 falls within the range $0 \leq n < 3$, such that the chosen threshold $n \geq 3$ corresponds to a p-value threshold of p=0.001. By contrast, the observed distribution of ecDNA (H_1, orange bars) was concentrated at some genomic regions, indicating a nonrandom distribution of ecDNA in pediatric cancer genomes. PMF, probability mass function.



Supplementary Figure 3: A recurrently ecDNA-amplified locus contains putative oncogene *AKT3*. (**a**) Panel reproduced from **Fig. 2a**. Chromosomal co-ordinates of ecDNA amplifications were stacked to generate a histogram of ecDNA amplifications across the genome. Selected recurrently amplified ($n \geq 3$) genes of interest are highlighted. (**b**) Histogram data from (a) zoomed in to show a 5Mbp peritelomeric portion of chr1 containing the *AKT3* locus. The consensus genomic region amplified in all 3 tumors, or maximally amplified region, contains 2 complete protein-coding gene loci: *AKT3* and *ZBTB18*. (**c**) Circular sequence assemblies of the ecDNA sequences containing the full-length *AKT3* locus. Known and putative oncogenes *AKT3* and *MYCL* are highlighted in red.



Supplementary Figure 4: ecDNA amplifications detected in primary and recurrent tumors of rhabdomyosarcoma SJRHB012. (**a**-**b**) A conserved ecDNA sequence containing amplified *MDM2* which does not show evidence of sequence variation between (**a**) diagnosis and (**b**) recurrence biosamples of the rhabdomyosarcoma SJRHB012 (modified Jaccard similarity 0.75). (**c**-**d**) A nonconserved ecDNA sequence in the same tumor. The ecDNA sequence (**c**) in the primary tumor amplified a 1.5Mbp segment of chr12 including oncogene *CDKN1B*. The ecDNA sequence (**d**) in the same tumor at recurrence shared 674kbp of sequence homology with the primary ecDNA including *BCL2L14*, *LRP6*, *MANSC1*, and *LOH12CR2* but not *CDKN1B* (modified Jaccard similarity 0.21). All visualizations were generated using the cyclic sequence which explains the greatest proportion of copy number amplification.

METHODS

Pediatric tumor whole genome sequencing (WGS)

Pediatric tumor WGS was identified from two pediatric cancer genomic data repositories, corresponding to 2398 biosamples from 1874 patients in the Pediatric Brain Tumor Atlas (PBTA) and 1223 biosamples from 1094 patients in the St Jude Cloud (SJC). Inclusion criteria were whole genome sequencing annotated with a tumor diagnosis; nontumor and unannotated samples were excluded. Data preprocessing comprised each institution's standard WGS bioinformatics pipelines.

ecDNA detection and classification from bulk WGS

To detect ecDNA, all samples in the WGS cohort were analyzed using AmpliconArchitect (AA)[1] v1.2 and AmpliconClassifier (AC)[4] v1.1.2. Briefly, the AmpliconArchitect algorithm was performed as follows. Copy number segmentation and estimation were performed using CNVkit v0.9.6[67]. Segments with copy number (CN) ≥ 4.5 were extracted as "seed" regions using AmpliconSuite-pipeline (April 2020 update)[63]. For each seed, discordant read pairs indicative of genomic structural rearrangement were identified within and up to 50kbp distal the query region. Genomic segments were defined by genomic breakpoint locations (identified by discordant reads) and by modulations in genomic copy number. A CN-aware breakpoint graph of the amplicon region was constructed using the genome segments and breakpoints, and cyclic and non-cyclic paths are extracted from the graph. Amplicons were classified as ecDNA(+), breakage-fusion-bridge (BFB), complex non-cyclic, linear, or no focal amplification using the heuristic-based companion script AmpliconClassifier. Biosamples with one or more amplicons classified as ecDNA were labelled ecDNA(+). Biosamples without ecDNA but with one or more amplicons classified as BFB, complex non-cyclic or linear were labelled as chromosomally amplified. All other samples were annotated as having no focal amplifications.

All amplifications classified as ecDNA were manually reviewed for possible false positives, marking for reanalysis amplicons satisfying any of the following criteria:

- amplicon has high variability in depth of coverage of mapped reads, indicating a genomic locus of low mappability;
- cyclic subsequences which were assigned CN < 4.

These samples were reanalyzed using the latest AmpliconSuite versions as of Aug. 2024 (AA v1.4.r2, AC v1.2.0). Samples reclassified in this way were BS_2F8PFTW5, BS_XB34VS6P, BS_G6KYSGQF, BS_ GBSSZBMF, and BS_YMYESCY7.

Patient metadata and tumor type annotations

Clinical metadata including sex, age at diagnosis, histology, and tumor history (primary, secondary, etc.) were available from the respective cloud genomics data platforms https://pedcbioportal.kidsfirstdrc.org/ (PBTA) and https://pecan.stjude.cloud/ (SJC), and from previous peer-reviewed publications[13,30]. Molecular diagnoses were integrated with histological diagnoses for biosamples included in the Pediatric Brain Tumor Atlas (PBTA)[13] in the following priority order: molecular subtype, DKFZ v12 methylation classification (confidence threshold 0.9), DKFZ v11 methylation classification[15] (likewise), harmonized diagnosis, original histological diagnosis. Diagnoses were subsequently collated and harmonized between the data sources to annotate the tumor classifications described herein (**Supplementary Table 8**).

Annotations of tumor purity were available for 2064 biosamples representing 1679 tumors from the PBTA cohort from a previous publication[13] using the THetA2 software[64]. To annotate SJC samples, we applied THetA2 to an additional 1231 samples representing 1100 tumors.

Logistic regressions

To determine whether tumor purity influences ecDNA detection efficiency, we fitted logistic regressions of amplicon class (ecDNA, chromosomal amplification, or no amplification) on sex, age at diagnosis, and cancer type, including and excluding tumor purity (as estimated by ThetA2) as an additional covariate. Logistic regression models were implemented using scikit-learn v1.5.2[65] using the Newton conjugate gradient solver and no penalty term. The models with and without tumor purity as an additional covariate were compared by likelihood ratio test.

Survival analyses

Patient overall survival data were included in PBTA clinical metadata and provided upon request from St. Jude. 237 medulloblastoma tumors from the International Cancer Genome Consortium and analyzed in a previous publication[8] were also included in survival analyses. St. Jude survival data did not include date of most recent follow-up, so survival data were censored at the date of data collection. Tumors with linear, complex noncyclic, or breakage-fusion-bridge amplifications but without ecDNA were classified as chromosomal. Kaplan-Meier and Cox Proportional Hazards regressions were performed in R 4.3.0 with the survival[66] v3.5.7 and survminer v0.4.9 packages.

Kaplan-Meier (KM) regressions. Sample was limited to tumor types with at least 10 cases, at least 1 death and at least 1 ecDNA(+) sample (n = 1820; 202 ecDNA(+), 198 with chromosomal amplification, and 1420 unamplified). Differential survival reported for KM regressions was evaluated by log-rank test.

Cox Proportional Hazards (CPH) regression. Hazard ratios were estimated by CPH regression parameterized on amplification class, tumor type, sex and age at diagnosis, and evaluated by log-rank test. Sample was as in KM above but additionally required annotations for sex and age at diagnosis, resulting in 76 exclusions (n = 1744). Applicability of the proportional hazards assumption was validated by likelihood ratio test ($\alpha > 0.1$) implemented in the `cox.zph` function of the survival package.

Ordinary least-squares linear regression

For regression of amplification status (ecDNA, chromosomal amplification, or no amplification) on age at diagnosis, ages were binned at 0.1-year resolution and a centered 1-year rolling average (smoothening) was applied to reduce sampling noise. For tumors with sampled more than once, the earliest biosample was used for determining amplification status. Ordinary least-squares linear regression models were then fit independently for each class using age as the predictor and smoothened class proportion as the response using scikit-learn[65] v1.5.2. Significance testing was performed using Student's *t*-test implemented in statsmodels[67] v0.14.2.

Other 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, both implemented in scipy.stats[68]

v1.10.2. Multiple hypothesis correction was performed where appropriate using the Benjamini-Hochberg correction implemented in statsmodels[67] v0.14.2 and reported as *q*-values or adjusted *p*-values. All statistical tests described herein were two-sided unless otherwise specified.

Gene fusions

Fusion genes were identified by adjacency on the ecDNA sequence assemblies. For samples from the PBTA subcohort, fusion gene expression was confirmed from RNA-seq data as described in the source publication[13].

Longitudinal subcohort (Supplementary Figure 4; Supplementary Table 8)

Patient tumors were included in the set of longitudinal cases which satisfied the following criteria:

- *St Jude and PNOC cohorts.* For multiple tumors from the same patient, exactly one is annotated as "diagnosis", and the others are labelled progressive, recurrent, or metastasis;
- *PBTA X00 and X01 cohorts.* Multiple tumors from the same patient for which the difference between the dates of diagnosis of the initial and the most recent biosample exceeds 30 days.

Similarity scores reported between amplicons were calculated using AmpliconClassifier feature_similarity.py v1.1.2. The formal definition and motivation of the similarity score is described previously[17].

Circular sequences illustrated in **Supplementary Figures 3 & 4** reflect the highest-copy cyclic sequence detected using AmpliconArchitect containing the oncogene of interest. Figures were generated using CycleViz 0.2.1.

ACKNOWLEDGMENTS

This work is supported by a generous endowment by the Clayes Foundation to the Research Center for Neuro-Oncology and Genomics within the Rady Children's Institute for Genomic Medicine, a Hannah's Heroes St. Baldrick's Scholar Award (L.C.), the Dragon Master Foundation (L.C.), funding from the National Institutes of Health (NIH) National Institute of Neurological Disorders and Stroke Institute R01NS132780 (L.C.), R21NS130137 (L.C.) and R21NS120075 (L.C.), the NIH National Cancer Institute P30CA030199 (L.C. and K.Y.), P30CA030199-42-S1 (L.C. and K.Y.), F31CA271777 (O.S.C.) U01CA253547 (J.P.M.), P30CA051008 (J.P.M.) and U24CA264379 (V.B. and J.P.M.), the NIH National Library of Medicine T15LM011271 (O.S.C.), the Japan Society for Promotion of Science (JSPS) Postdoctoral Research Fellowship F24404 (O.S.C. and D.K.), and a Moores Cancer Center Pilot Grant (L.C., V.B., and J.P.M.). This project is also supported by the UCSD/Rady Children's Hospital pediatric hematology oncology fellowship program (S.S.). D.R.G. would like to gratefully acknowledge support by the Clinician Scientist Program of the Mildred Scheel Cancer Career Center HaTriCS[4] at the University Medical Center Hamburg-Eppendorf and the Else Kröner-Fresenius Stiftung (Grant 2025_EKEA.51). This work used Expanse cluster compute services at the San Diego Supercomputer Center through allocation BIO210026 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296. Support through grant P30CA030199 to the Genomics

core facility at Sanford Burnham Prebys (NCI designated Cancer Center) is gratefully acknowledged. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research was conducted using data made available by The Children's Brain Tumor Network (formerly the Children's Brain Tumor Tissue Consortium) and the St. Jude Cloud. We thank J. H. Zhang, C. McLeod, D. Miller and A. Resnick for facilitating data access, A. Calatroni for his help with statistical analyses, R.J. Nowling for his open-source implementation of the likelihood ratio test, and R. Wachs for figure design assistance.

CONTRIBUTIONS

Conceptualization: S.S., O.S.C., J.P.M., and L.C. Methodology: O.S.C., S.S., J.L., J.P.M., and L.C. Investigation: O.S.C., S.S., E.Y.-C.C., R.K., J.K., A.D., S.W., W.Z., and M.B. Formal analysis: O.S.C., S.S., E. Y.-C.C., and R.K. Data curation: O.S.C., S.S., E. Y.-C.C., J.L., V.B., K.Y., J.P.M., and L.C. Software: O.S.C., S.S., E. Y.-C.C., R.K., J.K., A.D., S.W., W.Z., M.B., J.L., and Y.Y.L. Resources: J.L., K.O., D.N.G., D.K., K.Y., J.P.M., and L.C. Validation: Y.Y.L., E.R.-F., A.G.H., K.O., D.R.G., and K.W.P. Visualization: O.S.C., S.S., and E. Y.-C.C. Writing – original draft: O.S.C. and S.S. Writing – review and editing: O.S.C., S.S., E. Y.-C.C., J.L., D.K., V.B., J.P.M., and L.C. Funding acquisition: O.S.C., D.K., V.B., K.Y., J.P.M., and L.C. Supervision: V.B., M.P., K.Y., J.P.M., and L.C. Project administration: L.C.

DECLARATION OF INTERESTS

The authors declare the following competing interests:

V.B. and J.L. are inventors named in pending patent application no. US 2022/0364182 A1.

V.B. is a co-founder, consultant, SAB member and has equity interest in Boundless Bio, Inc. and Abterra, Inc. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies.

J.L. previously consulted for Boundless Bio, Inc. The terms of this arrangement have been reviewed and approved by the University of California in accordance with its conflict of interest policies.

RESOURCE AVAILABILITY

LEAD CONTACT

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Lukas Chavez (lchavez@sbpdiscovery.org).

MATERIALS AVAILABILITY

This study did not generate new unique reagents.

DATA AND CODE AVAILABILITY

WGS data from the PBTA and St. Jude datasets are under controlled access as implemented by the respective organizations but are available from the following sources upon approval from an institutional data access committee. PBTA patient cohort: Kids First Data Resource Center (https://kidsfirstdrc.org) via the CAVATICA data portal. Inclusion criteria were all patient tumors with WGS in the PBTA-PNOC and PBTA-CBTN datasets. St. Jude patient cohort: St.

Jude Cloud (https://www.stjude.cloud). Inclusion criteria were all patient tumors with WGS from the Pediatric Cancer Genome Project (PCGP, SJC-DS-1001) and Real-Time Clinical Genomics (RTCG, SJC-DS-1007) datasets as of September 2022.

All source metadata are available in the Supplementary Tables and/or from the cited sources. Data and derived visualizations describing all amplifications are accessible from the data portals at https://ccdi-ecdna.org/ and https://ampliconrepository.org/project/PedPanCan.

Source code for the AmpliconArchitect family of software tools is available at:

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

consisting of submodules:

- AmpliconArchitect: https://github.com/AmpliconSuite/AmpliconArchitect
- AmpliconClassifier: https://github.com/AmpliconSuite/AmpliconClassifier
- CycleViz: https://github.com/AmpliconSuite/CycleViz/.

Source code for software applets used to run AmpliconSuite-pipeline on the DNANexus (St. Jude Cloud) and CAVATICA (PBTA) platforms are available at:

- https://github.com/chavez-lab/ampliconsuite-dnanexus-applet
- https://github.com/chavez-lab/ampliconsuite-cwl-workflows

Source code to regenerate analyses and figures is available at https://github.com/auberginekenobi/pedpancan_ecdna.

620 **REFERENCES**

621 1. Lange, J. T. *et al.* The evolutionary dynamics of extrachromosomal DNA in human
622 cancers. *Nat. Genet.* **54**, 1527–1533 (2022).
623 2. Wang, T., Zhang, H., Zhou, Y. & Shi, J. Extrachromosomal circular DNA: a new potential
624 role in cancer progression. *J. Transl. Med.* **19**, 257 (2021).
625 3. Wu, S. *et al.* Circular ecDNA promotes accessible chromatin and high oncogene
626 expression. *Nature* **575**, 699–703 (2019).
627 4. Li, Z., Wang, B., Liang, H. & Han, L. Pioneering insights of extrachromosomal DNA
628 (ecDNA) generation, action and its implications for cancer therapy. *Int. J. Biol. Sci.* **18**,
629 4006–4025 (2022).
630 5. Aaltonen, L. A. *et al.* Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
631 6. Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*
632 **45**, 1113–1120 (2013).
633 7. Kim, H. *et al.* Extrachromosomal DNA is associated with oncogene amplification and
634 poor outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 (2020).
635 8. Chapman, O. S. *et al.* Circular extrachromosomal DNA promotes tumor heterogeneity in
636 high-risk medulloblastoma. *Nat. Genet.* **55**, 2189–2199 (2023).
637 9. Bailey, C. *et al.* Origins and impact of extrachromosomal DNA. *Nature* **635**, 193–200
638 (2024).
639 10. Koche, R. P. *et al.* Extrachromosomal circular DNA drives oncogenic genome
640 remodeling in neuroblastoma. *Nat. Genet.* **52**, 29–34 (2020).
641 11. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers.
642 *Nature* **555**, 321–327 (2018).
643 12. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric
644 leukaemias and solid tumours. *Nature* **555**, 371–376 (2018).

645 13. Shapiro, J. A. *et al.* OpenPBTA: The Open Pediatric Brain Tumor Atlas. *Cell*
646 *Genomics* **3**, 100340 (2023).
647 14. McLeod, C. *et al.* St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing
648 Ecosystem. *Cancer Discov.* **11**, 1082–1099 (2021).
649 15. Capper, D. *et al.* DNA methylation-based classification of central nervous system
650 tumours. *Nature* **555**, 469–474 (2018).
651 16. Deshpande, V. *et al.* Exploring the landscape of focal amplifications in cancer using
652 AmpliconArchitect. *Nat. Commun.* **10**, 392 (2019).
653 17. Luebeck, J. *et al.* Extrachromosomal DNA in the cancerous transformation of
654 Barrett's oesophagus. *Nature* **616**, 798–805 (2023).
655 18. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic*
656 *Acids Res.* **47**, D941–D947 (2019).
657 19. Liu, Y., Sun, J. & Zhao, M. ONGene: A literature-based database for human
658 oncogenes. *J. Genet. Genomics Yi Chuan Xue Bao* **44**, 119–121 (2017).
659 20. Mure, H. *et al.* Akt2 and Akt3 play a pivotal role in malignant gliomas. *Neuro-Oncol.*
660 **12**, 221 (2009).
661 21. Turner, K. M. *et al.* Genomically amplified Akt3 activates DNA repair pathway and
662 promotes glioma progression. *Proc. Natl. Acad. Sci.* **112**, 3421–3426 (2015).
663 22. Yang, J. *et al.* TRPS1 drives heterochromatic origin refiring and cancer genome
664 evolution. *Cell Rep.* **34**, 108814 (2021).
665 23. Peng, H., Wang, Y. & Luo, W. Multifaceted role of branched-chain amino acid
666 metabolism in cancer. *Oncogene* **39**, 6747–6756 (2020).
667 24. Arakawa, H. Netrin-1 and its receptors in tumorigenesis. *Nat. Rev. Cancer* **4**, 978–987
668 (2004).

669 25. Hogarty, M. D. *et al.* ODC1 Is a Critical Determinant of MYCN Oncogenesis and a
670 Therapeutic Target in Neuroblastoma. *Cancer Res.* **68**, 9735–9745 (2008).
671 26. Lambo, S. *et al.* The molecular landscape of ETMR at diagnosis and relapse. *Nature*
672 **576**, 274–280 (2019).
673 27. Mackay, A. *et al.* Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade
674 and Diffuse Intrinsic Pontine Glioma. *Cancer Cell* **32**, 520-537.e5 (2017).
675 28. Diaz, A. K. & Baker, S. J. The Genetic Signatures of Pediatric High-Grade Glioma:
676 No Longer a One-Act Play. *Semin. Radiat. Oncol.* **24**, 240–247 (2014).
677 29. Haase, S. *et al.* Hemispherical Pediatric High-Grade Glioma: Molecular Basis and
678 Therapeutic Opportunities. *Int. J. Mol. Sci.* **21**, 9654 (2020).
679 30. Dubois, F. P. B. *et al.* Structural variants shape driver combinations and outcomes in
680 pediatric high-grade glioma. *Nat. Cancer* **3**, 994–1011 (2022).
681 31. Li, B. K. *et al.* Pineoblastoma segregates into molecular sub-groups with distinct
682 clinico-pathologic features: a Rare Brain Tumor Consortium registry study. *Acta*
683 *Neuropathol. (Berl.)* **139**, 223–241 (2020).
684 32. Raffeld, M. *et al.* High level MYCN amplification and distinct methylation signature
685 define an aggressive subtype of spinal cord ependymoma. *Acta Neuropathol. Commun.* **8**,
686 101 (2020).
687 33. Helmsauer, K. *et al.* Enhancer hijacking determines extrachromosomal circular
688 MYCN amplicon architecture in neuroblastoma. *Nat. Commun.* **11**, 5823 (2020).
689 34. Potluri, V. R., Helson, L., Ellsworth, R. M., Reid, T. & Gilbert, F. Chromosomal
690 abnormalities in human retinoblastoma a review. *Cancer* **58**, 663–671 (1986).
691 35. Inazawa, J. *et al.* Simultaneous existence of double minute chromosomes and a
692 homogeneously staining region in a retinoblastoma cell line (Y79) and amplification of N-
693 myc at HSR. *Cancer Genet. Cytogenet.* **37**, 133–137 (1989).

694 36. Mysorekar, V. V., Harish, K., Kilara, N., Subramanian, M. & Giridhar, A. G.
695 Embryonal rhabdomyosarcoma of the chest wall: a case report and review of the literature.
696 *Indian J. Pathol. Microbiol.* **51**, 274–276 (2008).
697 37. Gil-Benso, R. *et al.* Cytogenetic and molecular findings related to
698 rhabdomyosarcoma. An analysis of seven cases. *Cancer Genet. Cytogenet.* **144**, 125–133
699 (2003).
700 38. Hayashi, Y. *et al.* Characterization of an embryonal rhabdomyosarcoma cell line
701 showing amplification and over-expression of the N-myc oncogene. *Int. J. Cancer* **45**,
702 705–711 (1990).
703 39. Weber-Hall, S. *et al.* Novel formation and amplification of the PAX7-FKHR fusion
704 gene in a case of alveolar rhabdomyosarcoma. *Genes. Chromosomes Cancer* **17**, 7–13
705 (1996).
706 40. Goldstein, M., Meller, I., Issakov, J. & Orr-Urtreger, A. Novel Genes Implicated in
707 Embryonal, Alveolar, and Pleomorphic Rhabdomyosarcoma: A Cytogenetic and
708 Molecular Analysis of Primary Tumors. *Neoplasia* **8**, 332–343 (2006).
709 41. Frascella, E. *et al.* Concomitant Amplif ication and Expression of PAX7-FKHR and
710 MYCN in a Human Rhabdomyosarcoma Cell Line Carrying a Cryptic t(1;13)(p36;q14).
711 *Cancer Genet. Cytogenet.* **121**, 139–145 (2000).
712 42. Fletcher, J. A., Gebhardt, M. C. & Kozakewich, H. P. Cytogenetic aberrations in
713 osteosarcomas: Nonrandom deletions, rings, and double-minute chromosomes. *Cancer*
714 *Genet. Cytogenet.* **77**, 81–88 (1994).
715 43. Gilbert, F., Potluri, V. R., Short, M. P., Kau, C.-L. & Lalatta, F. Retinoblastoma,
716 chromosome abnormalities and oncogene expression. *Ophthalmic Paediatr. Genet.* **8**, 3–
717 10 (1987).

718 44. Atiye, J. *et al.* Gene amplifications in osteosarcoma—CGH microarray analysis.
719 *Genes. Chromosomes Cancer* **42**, 158–163 (2005).
720 45. dos Santos Aguiar, S., de Jesus Girotto Zambaldi, L., dos Santos, A. M., Pinto, W. &
721 Brandalise, S. R. Comparative genomic hybridization analysis of abnormalities in
722 chromosome 21 in childhood osteosarcoma. *Cancer Genet. Cytogenet.* **175**, 35–40 (2007).
723 46. Sayles, L. C. *et al.* Genome-Informed Targeted Therapy for Osteosarcoma. *Cancer*
724 *Discov.* **9**, 46–63 (2019).
725 47. Yang, L. *et al.* FOXP4-AS1 participates in the development and progression of
726 osteosarcoma by downregulating LATS1 via binding to LSD1 and EZH2. *Biochem.*
727 *Biophys. Res. Commun.* **502**, 493–500 (2018).
728 48. Both, J. *et al.* Identification of Novel Candidate Oncogenes in Chromosome Region
729 17p11.2-p12 in Human Osteosarcoma. *PLOS ONE* **7**, e30907 (2012).
730 49. Martin, J. W., Squire, J. A. & Zielenska, M. The Genetics of Osteosarcoma. *Sarcoma*
731 **2012**, e627254 (2012).
732 50. Granberg, I. & Mark, J. The chromosomal aberration of double-minutes in a human
733 embryonic rhabdomyosarcoma. *Acta Cytol.* **15**, 42–45 (1971).
734 51. Steilen-Gimbel, H. *et al.* A novel site of DNA amplification on chromosome 1p32-33
735 in a rhabdomyosarcoma revealed by comparative genomic hybridization. *Hum. Genet.* **97**,
736 87–90 (1996).
737 52. Williamson, D. *et al.* Relationship between MYCN copy number and expression in
738 rhabdomyosarcomas and correlation with adverse prognosis in the alveolar subtype. *J.*
739 *Clin. Oncol.* **23**, 880–888 (2005).
740 53. Sorensen, P. H. B. *et al.* PAX3-FKHR and PAX7-FKHR Gene Fusions Are
741 Prognostic Indicators in Alveolar Rhabdomyosarcoma: A Report From the Children's
742 Oncology Group. *J. Clin. Oncol.* **20**, 2672–2679 (2002).

743 54. Seong, B. K. A. *et al.* TRIM8 modulates the EWS/FLI oncoprotein to promote
744 survival in Ewing sarcoma. *Cancer Cell* **39**, 1262 (2021).
745 55. Holdhof, D. *et al.* Atypical teratoid/rhabdoid tumors (ATRTs) with SMARCA4
746 mutation are molecularly distinct from SMARCB1-deficient cases. *Acta Neuropathol.*
747 *(Berl.)* **141**, 291–301 (2021).
748 56. Murray, J. C. *et al.* Detection of Double Minute Chromosomes in a Child with Acute
749 Lymphoblastic Leukemia. *Cancer Genet. Cytogenet.* **101**, 138–142 (1998).
750 57. Zhang, H. *et al.* Characterization and integrated analysis of extrachromosomal DNA
751 amplification in hematological malignancies. *Neoplasia* **56**, 101025 (2024).
752 58. Mendrzyk, F. *et al.* Isochromosome breakpoints on 17p in medulloblastoma are
753 flanked by different classes of DNA sequence repeats. *Genes. Chromosomes Cancer* **45**,
754 401–410 (2006).
755 59. Potocki, L. *et al.* Molecular mechanism for duplication 17p11.2- the homologous
756 recombination reciprocal of the Smith-Magenis microdeletion. *Nat. Genet.* **24**, 84–87
757 (2000).
758 60. Pentao, L., Wise, C. A., Chinault, A. C., Patel, P. I. & Lupski, J. R. Charcot–Marie–
759 Tooth type 1A duplication appears to arise from recombination at repeat sequences
760 flanking the 1.5 Mb monomer unit. *Nat. Genet.* **2**, 292–300 (1992).
761 61. Tang, J. *et al.* Enhancing transcription–replication conflict targets ecDNA-positive
762 cancers. *Nature* **635**, 210–218 (2024).
763 62. Mayr, L. *et al.* Effective targeting of PDGFRA-altered high-grade glioma with
764 avapritinib. *Cancer Cell* **43**, 740-756.e8 (2025).
765 63. AmpliconSuite: an end-to-end workflow for analyzing focal amplifications in cancer
766 genomes | bioRxiv. https://www.biorxiv.org/content/10.1101/2024.05.06.592768v1.

64. Oesper, L., Satas, G. & Raphael, B. J. Quantifying tumor heterogeneity in whole-genome and whole-exome sequencing data. *Bioinformatics* **30**, 3532–3540 (2014).

65. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

66. Therneau, T. A package for survival analysis in R.

67. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. in *Proceedings of the 9th Python in Science Conference* (eds. Walt, S. van der & Millman, J.) 92–96 (2010). doi:10.25080/Majora-92bf1922-011.

68. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).