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], Riki Goto[1], Wenshu Zhang[3], Miguel Brown[7], Jens Luebeck[6], Hui Hui[3], Jessica Wang[2], Yan Yuen Lo[2,8], Elias Rodriguez-Fos[9], Shanzheng Wang[10,11,12,13], Konstantin Okonechnikov[10,11,12], David R. Ghasemi[10,14,15,16], Kristian W. Pajtler[10,11,12,17], Johannes Gojo[18], Anton G. Henssen[9], Marcel Kool[10], Vineet Bafna[6,19,20], Daisuke Kawauchi[1], Megan Paul[4,5], Kevin Yip[2], Jill P. Mesirov[19,21], Lukas Chavez[2,8,21,+]

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. Faculty of Biosciences, Heidelberg University, Heidelberg, Germany;
14. Department of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany;
15. Research Institute Children's Cancer Center, Hamburg, Germany;
16. Mildred Scheel Cancer Career Center HaTriCS[4], University Medical Center Hamburg-Eppendorf, Hamburg, Germany;
17. Department of Pediatric Oncology, Hematology and Immunology, Heidelberg University Hospital, Germany;
18. Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna, Austria;
19. Moores Cancer Center, University of California San Diego, La Jolla, CA, USA;
20. Halıcıoğlu Data Science Institute, University of California San Diego, La Jolla, CA, USA;
21. Department of Medicine, University of California San Diego, La Jolla, CA, USA;

[+] Corresponding author

* Co-first authors

ABSTRACT

Circular extrachromosomal DNA (ecDNA) is a powerful driver of oncogene amplification and tumor evolution, yet its prevalence, composition, and clinical significance across pediatric cancers remains incompletely understood. Leveraging two major cloud-based genomic data repositories, we analyzed whole genome sequencing data from 3,630 tumor biosamples representing 2,967 children across 39 solid tumor types. ecDNA was identified in 9% of cases and was enriched in high-grade and clinically aggressive malignancies including ETMR, pediatric high-grade glioma, medulloblastoma, neuroblastoma, osteosarcoma and rhabdomyosarcoma. We catalogued 392 ecDNA sequences, revealing recurrent amplification of known oncogenes, diverse gene fusions, and oncogenic loci where recurrent ecDNA involvement underscores their emerging importance in pediatric tumors. Oncogenes amplified on ecDNA reached significantly higher copy number than chromosomal amplifications, and ecDNA was associated with significantly poorer 5-year survival independent of tumor type, age and sex. Longitudinal analyses demonstrated that ecDNA frequently arises, is lost, or undergoes structural remodeling during progression and recurrence, including acquisition or loss of oncogenes on the same circular element. Together, these findings define the landscape, clinical relevance, and evolutionary behavior of ecDNA across childhood cancers, highlight candidate drivers, and identify patient populations that may benefit from emerging ecDNA-targeted therapeutic strategies. An interactive resource is available at https://ccdi-ecdna.org/.

INTRODUCTION

Extrachromosomal DNA (ecDNA, also known as double minutes), are megabase-scale circular focal amplifications of genomic DNA which arise in many human cancers. ecDNA sequences generally contain an oncogene or other functional element conferring a selective fitness advantage, and lack a centromere, leading to unequal segregation during cell division and intra-tumoral heterogeneity. The potent combination of copy number heterogeneity and positive selection drives high-copy oncogenic amplification, rapid tumor evolution, and therapeutic resistance[1–3].

Paired-end whole genome sequencing has emerged as a cost-effective and efficient method to identify oncogenes and other sequences captured on ecDNA. While it was largely overlooked in early pediatric pancancer genomics studies[4,5], recent studies have described the prevalence of ecDNA in various adult and some childhood cancer types[6–10]. Recognizing that pediatric cancers have 14-fold lower somatic mutation frequency than adult cancers as well as unique drivers and amplifications[4,5], 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)[11] and St. Jude Cloud (SJC)[12] (**Figure 1**). In total, this retrospective cohort comprised 3,630 solid tumor biosamples from 2,967 patients with mean WGS coverage of 69x (±15) per sample. Clinical metadata, including histological diagnoses and patient survival

outcomes, were collated from both platforms (**Table 1; Supplementary Table 1**). For brain tumors included in the PBTA, histological diagnoses were integrated with molecular tumor type classifications derived from methylation analyses[13] of the same samples. We applied the AmpliconArchitect (AA)[14] and AmpliconClassifier (AC)[15] algorithms to identify ecDNA amplifications. Briefly, this pipeline performs 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 (**Methods**). 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 measure, we estimated tumor purity and found no significant association with the identification of ecDNA (likelihood ratio test, $p = 0.15$, $n = 2133$). Given that previous studies using similar methods did not find ecDNA in adult hematologic malignancies[6,8], 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.

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

To identify oncogenic drivers amplified on ecDNA in pediatric solid tumors, we defined recurrent genomic intervals as those amplified on ecDNA in at least three independent cases, corresponding to a 99.9% confidence level against chance occurrence (**Supplementary Note 1**; **Supplementary Figure 2**). 109 distinct genomic intervals were recurrently amplified on ecDNA across all pediatric cancers in this cohort (**Figure 2a**). To identify putative driver genes, we annotated these genomic intervals with oncogenes listed in the COSMIC[16] and ONGene[17] databases (**Supplementary Table 5**). Genomic intervals most frequently amplified on ecDNA contained oncogenes *MYCN* (66 ecDNA amplifications observed in eight tumor types) or *CDK4* (22 ecDNA amplifications in six tumor types; **Figure 2b**). Five contiguous intervals recurrently amplified on ecDNA 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*).

Accumulating evidence suggests that ecDNA enables greater oncogene copy number amplification than chromosomal amplification[8]. To determine whether amplification by ecDNA is of greater magnitude than other amplifications, we compared estimated copy number for oncogenes amplified in 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 ($U = 1e5$; $p = 2e\text{-}61$; **Figure 2d**), suggesting that oncogenic amplification occurs at greater magnitude in ecDNA than in chromosomal amplifications in pediatric cancers.

82 genomic intervals recurrently amplified on ecDNA contained no known oncogene. These were relatively short (mean 181kbp, standard deviation 301kbp); however, 38 of these contained uninterrupted gene sequences, nominating 81 protein-coding genes and 235 long noncoding RNAs (lncRNA) as candidates under positive selection (**Supplementary Table 6**). These other recurrently amplified genes exhibited median ecDNA copy numbers comparable

136 to known oncogenes (two-sided Mann-Whitney test; U = 138; p = 0.23; **Supplementary**
137 **Figure 3**). Several recurrently amplified genomic intervals were exclusively co-amplified on
138 ecDNA with known oncogenes on recombinant DNA sequences, suggesting that possible
139 positive selection may be context-dependent[18]. To nominate *bona fide* oncogenic candidates,
140 we therefore further asked which genomic intervals were recurrently amplified on ecDNA yet
141 lacked any co-amplified known oncogenes. As a result, we identified five genomic intervals
142 that fulfilled these criteria, each containing genes with prior evidence supporting tumorigenic
143 roles. For example, a peritelomeric recurrently amplified locus at chromosome 1q amplified
144 the full gene sequences of *AKT3* and *ZBTB18* in two high-grade gliomas and a
145 rhabdomyosarcoma (**Supplementary Figure 4**). In mouse models, amplification and
146 overexpression of the paralogous gene *Akt3* promotes malignant glioma progression[19,20],
147 supporting a functional role for this gene locus in some human pediatric tumors. The other
148 examples contain the DNA replication initiation cofactor *TRPS1*[21], neurodevelopmental factor
149 *NTN1*[22,23], the *C19MC* microRNA cluster frequently amplified in embryonal tumors with
150 multilayered rosettes (ETMR)[24], and the recombination hotspot at chr17p11.2-12 in
151 osteosarcomas[25] and medulloblastomas[26].

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

153 We observed ecDNA most frequently in embryonal tumors with multilayered rosettes (4/9,
154 44%), pediatric high-grade gliomas (78/391, 20%), and medulloblastomas (55/347, 16%); and
155 rarely in pineal tumors (1/28, 4%), choroid plexus tumors (1/56; 2%), craniopharyngiomas
156 (1/101, 1%), ependymomas (3/239, 1%), glioneuronal tumors (1/192, 0.5%), and low-grade
157 gliomas (1/616, 0.2%) (**Figure 3a**).

158 Embryonal tumors with multilayered rosettes (ETMR) have poor outcomes and most are
159 characterized by amplification of a microRNA cluster on chromosome 19 (*C19MC*) including
160 fusion of the *TTYH1* gene[27]. Of 9 ETMR cases, 4 had *TTYH1*::*C19MC* amplification, all
161 extrachromosomal, suggesting that *C19MC* amplification in ETMR tumors may occur
162 predominantly on ecDNA.

163 To evaluate this hypothesis, we examined a validation cohort of 30 additional ETMR tumors
164 (**Supplementary Table 7**). Copy number estimation from bulk WGS indicated *C19MC*
165 amplification (CN > 4.5) in 18 cases and moderate gain (2.5 < CN < 4.5) in an additional 8
166 cases. Using default analysis parameters, 16 of 18 amplifications were classified as ecDNA
167 and 2 as chromosomal amplifications. To improve detection sensitivity, we reanalyzed the 8
168 cases with moderate *C19MC* gain (**Methods**) and identified an additional 5 ecDNA and 1
169 chromosomal amplicon, indicating that ecDNA detection sensitivity may be limited for
170 amplicons with low copy burden.

171 Overall, ecDNA containing *TTYH1*::*C19MC* fusion was detected in 21 of 30 cases in the
172 validation cohort (70%). Chromosomal amplicons containing the fusion were detected in an
173 additional 3 cases (10%). The fraction of *C19MC* amplicons classified as extrachromosomal
174 in the validation set (21/24) did not differ significantly from the discovery set (4/4; p = 1, Fisher
175 exact test), substantiating the high prevalence of extrachromosomal *C19MC* amplification in
176 a larger ETMR cohort. Moreover, ecDNA amplifications of *KANSL1* and *YAP1* were found in
177 non-*C19MC*-amplified tumors in the validation set, suggesting possible alternative driver
178 amplifications of non-*C19MC*-amplified ETMR.

179 The clinical outcomes of pediatric high-grade glioma (pHGG) patients differ across
180 molecular subgroups marked by genetic mutations to the histone H3 or isocitrate
181 dehydrogenase (*IDH*) genes[28]. pHGG with H3 mutations are further stratified into those

182 mutated at lysine 27 (H3K27) and at glycine 34 (H3G34). Among 380 pHGG tumors, 77 had
183 ecDNA (20%), including 29 of 103 with wild-type H3 (28%), 29 of 132 H3K27-mutant (22%),
184 2 of 16 H3G34-mutant (13%), and 4 of 10 with *IDH* mutation (40%). Known oncogenes[28–30]
185 recurrently amplified on ecDNA included *PDGFRA* (*n* = 14), *CDK4* (12), *MYCN*, *KIT* (10), *MET*
186 (8), *EGFR* (6), *GLI1* (5), *CDK6, MYC, MDM2* (4), *ID2, CCND2,* and *PLAGL2* (3).

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[7]. 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[32,33] 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 commonly be driven by ecDNA. Overall, although further
202 studies of larger tumor-specific cohorts are needed, these observations suggest that ecDNA
203 amplification may be limited to some ultra-rare and highly aggressive subtypes and are
204 otherwise largely absent in these tumor types.

205 **ecDNA in pediatric extracranial solid tumors**

206 ecDNA has been previously described in various extracranial pediatric solid tumors, most
207 extensively in neuroblastoma[9,18], and in case reports and cell lines of retinoblastoma[34,35],
208 rhabdomyosarcoma[29,36–41], and osteosarcoma[42]. We identify patient tumors with ecDNA for
209 each of these extracranial tumor types, as well as in adrenocortical carcinomas, Wilms tumors,
210 germ cell tumors, malignant peripheral nerve sheath tumors, and other rare sarcomas (**Figure**
211 **3b**).

212 Consistent with previous analyses[9,10], we estimate the prevalence of ecDNA in
213 neuroblastoma at 40/159 (25%). Most tumors with ecDNA amplified *MYCN* (35/40, 88%). Of
214 5 remaining ecDNA(+) tumors not amplifying *MYCN*, two had ecDNA amplifications of *TERT*,
215 and 3 had low-copy ecDNA containing no known oncogenic sequences.

216 Although *RB1* gene inactivation is the canonical driver of retinoblastoma, other
217 mechanisms including oncogenic amplification are believed to contribute to tumorigenesis in
218 some cases[43]. ecDNA has previously been described in retinoblastoma cell lines and patient
219 tumors and is particularly associated with *MYCN* amplification[35]. In this cohort, six of 44 (14%)
220 retinoblastoma cases had ecDNA, two of which amplified *MYCN*. A further two ecDNA
221 amplifications disrupted the *RB1* gene to form gene fusions, with lipoma *HMGIC* fusion partner
222 gene family member *LHFPL6* (*RB1-LHFPL6)* or with the *TPTE2P2* pseudogene (*RB1-*
223 *TPTE2P2)*. In the remaining two tumors, no canonical oncogenes were identified, suggesting
224 that these highly amplified ecDNA sequences may contain previously unknown oncogenic
225 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 pan-cancer study[6], 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 8**), 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**.

Tumors with ecDNA, chromosomal, or no focal amplification have distinct distributions of somatic small variants.

Somatic variant calling and annotation were performed on paired blood/tumor WGS data from both the SJ and CBTN datasets to assess whether the distribution of somatic mutation (single nucleotide variants (SNVs) and short indels < 50bp) differs in ecDNA-amplified tumors. For this analysis, we defined likely pathogenic (*LP*) somatic variants as somatic mutations of cancer-related genes which altered the protein sequence and were predicted to be functional (**Methods**). Aggregating *LP* mutations at the gene level, we first compared amplified to nonamplified tumors (**Supplementary Figure 5a**). Mutations of *BRAF* and *CTNNB1* were depleted in tumors with amplification, likely reflecting enrichment of *BRAF* and *CTNNB1* mutations in low-grade gliomas, glioneuronal tumors, and *WNT* subgroup medulloblastomas in which focal amplification is rare. Amplified tumors were enriched for mutation of *TP53*, *H3-3A*, *KNL1*, *PDGFRA*, *PABPC1*, *AURKA*, *STAG2*, and *KNSTRN* (n = 7071 *LP* somatic variants across 1798 solid tumors; χ^2 tests with FDR correction, q < 0.10; **Supplementary Table 9**). Notably, *KNL1*, *AURKA, STAG2* and *KNSTRN* are required for proper chromosomal pairing and segregation during mitosis, suggesting a plausible link between defects in this process and the presence of focal amplification in pediatric solid tumors. Due to the large hypothesis space, this analysis was underpowered to examine differences between extrachromosomally and chromosomally amplified tumors, but we note that alterations to *GNAS* were nominally depleted in tumors with ecDNA vs. amplified tumors without ecDNA (**Supplementary Figure 5b**). In an analysis of germline variation using similar methods across a subset of CNS tumors,

271 pathogenic germline variants of *TP53* (Li-Fraumeni Syndrome, LFS) were associated with
272 focal amplification overall but did not show specific enrichment for ecDNA (**Supplementary**
273 **Note 5**).

274 We next asked whether tumor mutation burden (TMB) or likely pathogenic somatic mutation
275 burden (LPTMB) differed across the ecDNA, chromosomally amplified, or nonamplified
276 tumors. To test this, we fitted logistic regressions on tumor type, age at diagnosis, sex and
277 batch (SJ or CBTN). No independent effect of TMB (Wald test; $p = 0.74$; $n = 1749$) or LPTMB
278 ($p = 0.51$) was observed in amplified vs. nonamplified tumors (**Supplementary Figure 5c**);
279 however, within the amplified tumor subset, ecDNA(+) tumors had lower LPTMB than
280 chromosomally amplified ($p = 0.007$; $n = 232$), accounting for independent effects of the above
281 covariates (**Supplementary Figure 5d**). These results indicate that age and tumor type are
282 likely the strongest determinants of overall tumor mutation burden, but that tumors with
283 extrachromosomal amplification have significantly lower LPTMB than tumors with
284 chromosomal amplifications only.

285 **ecDNA is associated with poor survival in pediatric cancers.**

286 To investigate whether ecDNA in pediatric cancers is associated with worse clinical
287 outcome, we performed Kaplan-Meier regression on overall survival data available for 1820
288 patients (**Supplementary Table 1**). Here, we stratified patients into those whose tumors had
289 ecDNA ($n = 216$), chromosomal amplification only ($n = 225$), or no focal amplification ($n =$
290 1452). Kaplan-Meier regression indicated that patients with ecDNA(+) tumors had significantly
291 worse outcomes than patients with tumors containing no amplification (log-rank test, adjusted
292 $p < 2e\text{-}16$) and patients with chromosomal amplifications (adjusted $p = 7e\text{-}6$) (**Figure 4a**). To
293 estimate the magnitude of the association between ecDNA and survival, we performed Cox
294 regression of sex, age at diagnosis, tumor type, and amplification type (ecDNA, chromosomal,
295 or no amplification). Controlling for available covariates, ecDNA detected in a patient tumor
296 conferred twofold risk of death within 5 years relative to tumors without amplification (hazard
297 ratio 2.1, $p = 5e\text{-}9$, Wald test), and one-and-a-half-fold risk relative to tumors with
298 chromosomal amplification but no ecDNA (HR 1.5, $p = 0.02$, **Figure 4b**).

299 Because childhood cancers are rare and highly diverse, survival analyses could not be
300 performed robustly for every tumor type due to limited patient numbers within each category.
301 To address this limitation, we integrated our cohort with previously published whole genome
302 sequencing datasets from studies of medulloblastoma[7], neuroblastoma[10], and pediatric high-
303 grade glioma[54]. This approach enabled us to investigate more thoroughly the association
304 between ecDNA and patient outcomes within each of these three tumor types.

305 Joint modelling of focal amplification, ecDNA, molecular subgroup, and other clinical
306 variables across 445 MBL tumors suggests that ecDNA (HR 2.8, $p = 6e\text{-}4$), rather than focal
307 amplification of any kind (HR 0.9, $p = 0.71$) is a stronger predictor of 5-year survival
308 (**Supplementary Figure 6a-b**). Similarly, survival regressions across 175 NBL tumors
309 estimate poorer 5-year overall survival for ecDNA(+) tumors accounting for chromosomal
310 amplifications (KM: $p = 0.042$, BH-adjusted log-rank test, **Supplementary Figure 6c**; Cox: HR
311 $= 2.8$, $p = 0.048$, **Supplementary Figure 6d**), consistent with a previous report stratifying NBL
312 tumors with focal amplification into high-risk with ecDNA amplification of *MYCN* vs.
313 intermediate-risk with chromosomal amplification elsewhere in the genome[10].

314 Myc family genes *MYC* and *MYCN* are major drivers of aggressive medulloblastoma[55] and
315 neuroblastoma[9,10]. We next investigated whether clinical outcomes may differ by the context
316 of Myc family oncogene amplification, i.e. ecDNA or chromosomal. Across both cancer types,
317 *MYC* and *MYCN* were amplified predominantly on ecDNA (MBL: 34/40, 85%; NBL: 27/31,

87%), leaving few chromosomally amplified cases as internal comparators. As expected, adjusting for Myc family gene amplification reduced the estimated hazard of ecDNA for both medulloblastoma (HR = 1.7, p = 0.14; **Supplementary Figure 7a-b**), and neuroblastoma (HR = 1.2, p = 0.78; **Supplementary Figure 7c-d**). We can therefore confirm that amplification of Myc family genes in these tumors in their predominantly extrachromosomal form carries an exceptionally poor prognosis, consistent with previous reports[10,55]; although the scarcity of chromosomally amplified *MYC*(*N*) tumors limits assessment of whether extrachromosomal Myc family amplification exerts further prognostic influence compared to chromosomal.

Molecular determinants of prognosis in pediatric high-grade gliomas include pathogenic mutations of histone 3 (H3), *IDH*, and *TP53* genes as well as focal amplification of diverse oncogenes. It has been previously shown that complex structural variants are associated with poorer overall survival of pHGG patients, where *TP53* loss and ecDNA were enriched in the complex-SV group[54]. Combining that cohort with pHGG tumors from our pancancer cohort, we confirmed that the strongest predictor of outcome was H3K27M mutation, which distinguishes high-risk diffuse midline gliomas (DMG) with universally poor prognoses (**Supplementary Figure 8a-b**). Further stratification by amplification status showed that the presence of chromosomal or extrachromosomal amplification was nominally but non-significantly associated with poorer survival within 5 years of diagnosis (**Supplementary Figure 8c**). Given the exceptionally poor outcomes associated with H3K27 alteration, we hypothesized that the presence of H3K27M mutation may supersede the prognostic utility of ecDNA in this tumor subtype. In a revised hierarchical model stratified by H3K27 alteration, we estimated that ecDNA is a prognostic indicator specifically in H3K27 wild-type tumors (HR = 1.81; p = 0.02); however, the estimated additional hazard of ecDNA in H3K27-altered tumors was nonsignificant and near-zero (HR = 1.2; p = 0.58; **Supplementary Figure 8d**). These findings indicate a context-dependent role for ecDNA in pHGG prognosis, where it is predictive of worse overall survival except in H3K27-altered tumors.

ecDNA evolves during progression and recurrence of pediatric cancers.

Recognizing that ecDNA may contribute to tumor evolution and treatment resistance, we next compared its prevalence in primary versus relapsed tumors and tracked its dynamics in longitudinal primary-relapse pairs.

For this analysis, we aggregated progression, recurrence, relapse, and metastasis into one category of secondary samples. Secondary malignancies and autopsy samples were excluded, as were tumor types without focal amplification. ecDNA was modestly but significantly more frequent in secondary (57 of 519, 11%) than primary tumors (202 of 2448, 8%), controlling for tumor type (**Figure 5a**; stratified permutation test, p = 0.002). Tumors with chromosomal amplifications alone were not enriched in secondary compared to primary tumors (**Figure 5b**; p = 0.45).

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 10**). Here, we included an ecDNA(+) medulloblastoma patient tumor diagnosed at Rady Children's Hospital, San Diego[7] and for which we generated new WGS data from the relapse. In these 29 longitudinal cases, we observed 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

366 heuristic similarity metric incorporating Jaccard similarities of the amplified sequences and
367 breakpoints (**Methods**). Five such ecDNA amplicons were conserved ($S > 0.7$) whereas the
368 other nine showed structural evolution, including examples where a new oncogene was added
369 (**Figure 5c**) or lost (**Supplementary Figure 9**) in the later tumor. For example, in a pHGG
370 patient (SJHGG019), we observe the oncogenes *EGFR* and *CDK6*, both coded on chr7, co-
371 amplified on the same ecDNA. The ecDNA observed in the relapse tumor of the same patient
372 contained *MYC*, coded on chr8, as a third oncogene on the same ecDNA together with *EGFR*
373 and *CDK6* (**Figure 5c**). Overall, these results suggest clonal tumor dynamics in which ecDNA
374 may emerge, mutate, remain conserved, or be lost during disease progression.

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

384 **DISCUSSION**

385 By leveraging childhood cancer data cloud repositories with whole-genome sequencing
386 data, we estimate frequencies of circular extrachromosomal amplification across a wide
387 spectrum of childhood cancer types. Where grading of the tumor was incorporated into the
388 histological or molecular diagnosis, the higher-grade diagnoses (pediatric high-grade glioma,
389 choroid plexus carcinoma, malignant peripheral nerve sheath tumor, and pineoblastoma) had
390 greater incidence of ecDNA than the lower-grade diagnosis (pediatric low-grade glioma,
391 choroid plexus papilloma, neurofibroma, and low-grade pineal tumors), suggesting that
392 ecDNA may be more frequent in pediatric tumors with high-grade histology across tissue
393 origins. Notable high-grade tumor types in which we did not observe ecDNA are defined by
394 well described oncogenic mutations, such as biallelic loss of *SMARCB1/SMARCA4* in atypical
395 teratoid rhabdoid tumors (ATRT)[56], *EWSR1*::*FLI1* (also commonly *EWS*/*FLI*) in Ewing
396 sarcoma, or *ZFTA*[56] gene fusions in supratentorial ependymoma. Consistent with a
397 "Goldilocks principle", excessive expression of *EWSR1*::*FLI1* is deleterious in Ewing
398 sarcoma[57], rendering high-copy amplification on ecDNA disadvantageous. A similar
399 phenomenon may constrain *ZFTA* fusions in ependymoma, where ecDNA amplification would
400 push fusion oncofactor activity beyond a tolerable range[58]. By contrast, solid tumor types
401 where drivers benefit from gene dosage escalation more frequently contain ecDNA. Although
402 we did not comprehensively analyze hematological cancers, a pediatric case report[59] and a
403 recent retrospective analysis of adult leukemias[60] recommend this area for further
404 investigation.

405 Patients with tumors containing ecDNA had significantly poorer survival compared to those
406 with chromosomal amplification and those without amplification, controlling for sex, age and
407 tumor type. This finding suggests that both focal amplification and extrachromosomal
408 amplification have nonredundant prognostic value in pediatric cancers. The majority of
409 ecDNA(+) tumors contained at least one well described oncogene, such as *MYC, MYCN*,
410 *CDK4*, and others. In some tumors, multiple oncogenes were amplified on one or multiple
411 ecDNA sequences. Other ecDNA amplifications contained genes not annotated in the

412 COSMIC[16] and ONGene[17] oncogene databases but with plausible oncogenic roles, including
413 *AKT3, NTN1,* and others. In rhabdomyosarcoma and retinoblastoma, we found recurrent gene
414 fusions on ecDNA. The potential oncogenic roles of still many other ecDNA-amplified
415 sequences without known oncogenes are less clear. For example, relatively little is known of
416 the role of chr17p11.2 amplification, the most frequent ecDNA amplification in osteosarcomas.
417 Intriguingly, this locus is a recombination hotspot in medulloblastoma[26] and in some
418 nonmalignant genetic disorders[61,62]. Given the current poor prognosis of osteosarcomas and
419 the need for more effective treatment protocols, further analysis of this and other recurrently
420 amplified loci are warranted to identify novel oncogenic drivers and treatment targets.

421 Across a set of paired primary/secondary cases, we observed various structural
422 rearrangements of ecDNA sequences. This adds to accumulating evidence[2,7,63] that ecDNA
423 facilitates tumor evolution in response to therapeutic or other selective pressures. Secondary
424 tumors contained ecDNA more frequently than primary tumors, controlling for tumor type,
425 consistent with a previous analysis of adult tumors[63]. Due to limited treatment information
426 available for this cohort, the role of extrachromosomal DNA evolution in the emergence of
427 drug resistance under standard-of-care treatment including chemo- and radiotherapy remains
428 to be further elucidated and functionally tested in pediatric solid tumor models.

429 Overall tumor mutation burden (TMB) was not associated with amplification or with ecDNA;
430 however, ecDNA-amplified tumors had fewer somatic simple variants with predicted functional
431 consequence targeting cancer genes than chromosomally amplified tumors. Speculatively, a
432 possible explanation for this observation is that ecDNA may be sufficient to drive tumor growth
433 with fewer secondary driver mutations. Focal amplification was specifically associated with
434 somatic alterations to DNA damage response gene *TP53* and to chromosomal pairing and
435 segregation components *KNL1*, *AURKA*, *STAG2*, and *KNSTRN*; however, no specific
436 association was observed for ecDNA vs. chromosomal amplification. Should this observation
437 prove generalizable, a possible explanation is that mechanisms for ecDNA generation
438 associated with *TP53* inactivation such as chromothripsis[64] may be at least as likely to
439 generate chromosomal forms of focal amplification; or that other p53-independent
440 mechanisms of ecDNA generation may yet be found. Future work will be required to
441 substantiate and functionally validate the predictors of ecDNA and their potential underlying
442 mechanisms.

443 Identification of ecDNA in patient tumors holds promise in prognostication and targeted
444 treatment. Although the spectrum of known and putative oncogenic sequences on ecDNA is
445 diverse and tissue-specific, there is some hope that patients with ecDNA(+) tumors may
446 benefit from future therapies targeting a common vulnerability. For example, a recent study
447 demonstrated that combination treatment of an ecDNA(+) gastric tumor using inhibitors
448 against the ecDNA-amplified *FGFR* oncogene and a CHK1 inhibitor enhances transcription-
449 replication conflict and prevents ecDNA-mediated acquired resistance that otherwise
450 develops under FGFR inhibition alone[65]. Our results show that this approach may also be
451 applicable to some high-risk pediatric tumors. For example, we recently reported preliminary
452 clinical real-world experience using avapritinib, a next-generation tyrosine kinase inhibitor, in
453 pediatric and young adult patients with recurrent/refractory PDGFRA-altered HGG[66].
454 Avapritinib was well tolerated and led to initial radiographic responses in 3 out of 7 cases. In
455 one patient, *de novo EGFR* amplification was observed in a metastatic lesion that progressed
456 during treatment, indicating oncogene switching as a possible resistance mechanism. A
457 promising avenue for future investigation is whether combination therapy with an ecDNA-
458 targeting treatment may prevent ecDNA-mediated resistance. ecDNA may also hold promise

459 for detection via liquid biopsy as genes identified on ecDNA including *MYCN*[67], *MYC*[68], and
460 *C19MC*[69] are currently under investigation as liquid biomarkers.

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

465 **TABLES AND FIGURES**

	N = 2,967[1]
Sex	
Female	1,308 (44%)
Male	1,627 (55%)
Not Reported	32 (1.1%)
Age at Diagnosis (years)	
<1	166 (5.8%)
1–4	741 (26%)
5–9	733 (25%)
10–14	693 (24%)
15–18	380 (13%)
19–40	164 (5.7%)
Race	
White	1,904 (64%)
Black Or African American	321 (11%)
Asian	93 (3.1%)
Multiracial	76 (2.6%)
American Indian Or Alaska Native	10 (0.3%)
Native Hawaiian Or Other Pacific Islander	7 (0.2%)
Other	406 (14%)
Not Reported	150 (5.1%)
Ethnicity	
Hispanic Or Latino	310 (10%)
Not Hispanic Or Latino	2,301 (78%)
Not Reported	356 (12%)

466
467

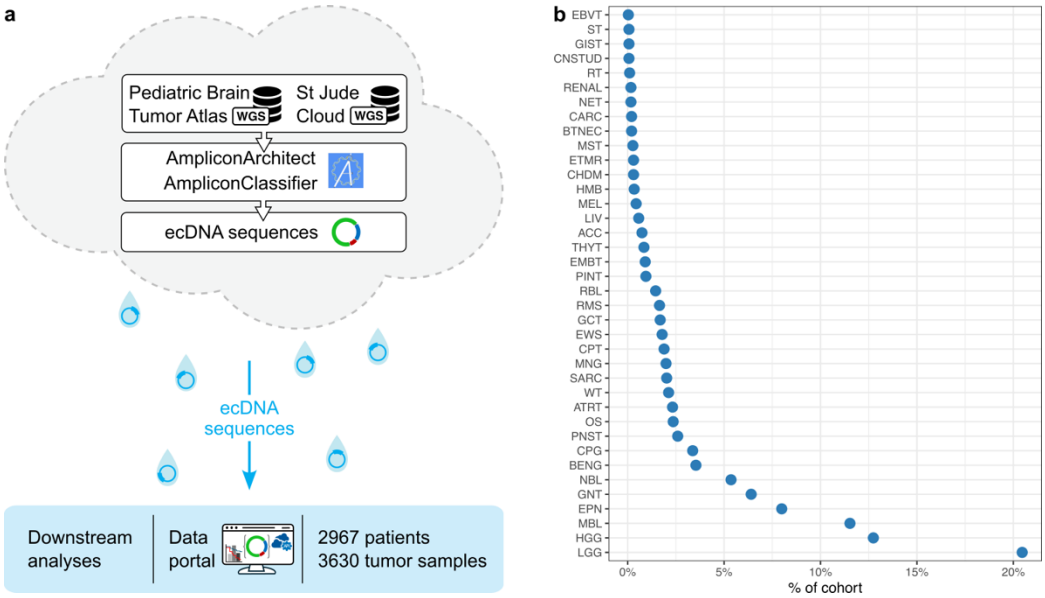


Figure 1: Analysis of ecDNA in pediatric cancers using cloud computing. (**a**) Whole genome sequencing (WGS) data of solid tumors was analyzed in two cancer genome cloud repositories. ecDNA and other amplifications were identified and annotated with clinical information for downstream analysis. (**b**) Tumor biosample distribution by tumor type. Abbreviations are expanded in **Supplementary Table 12**.

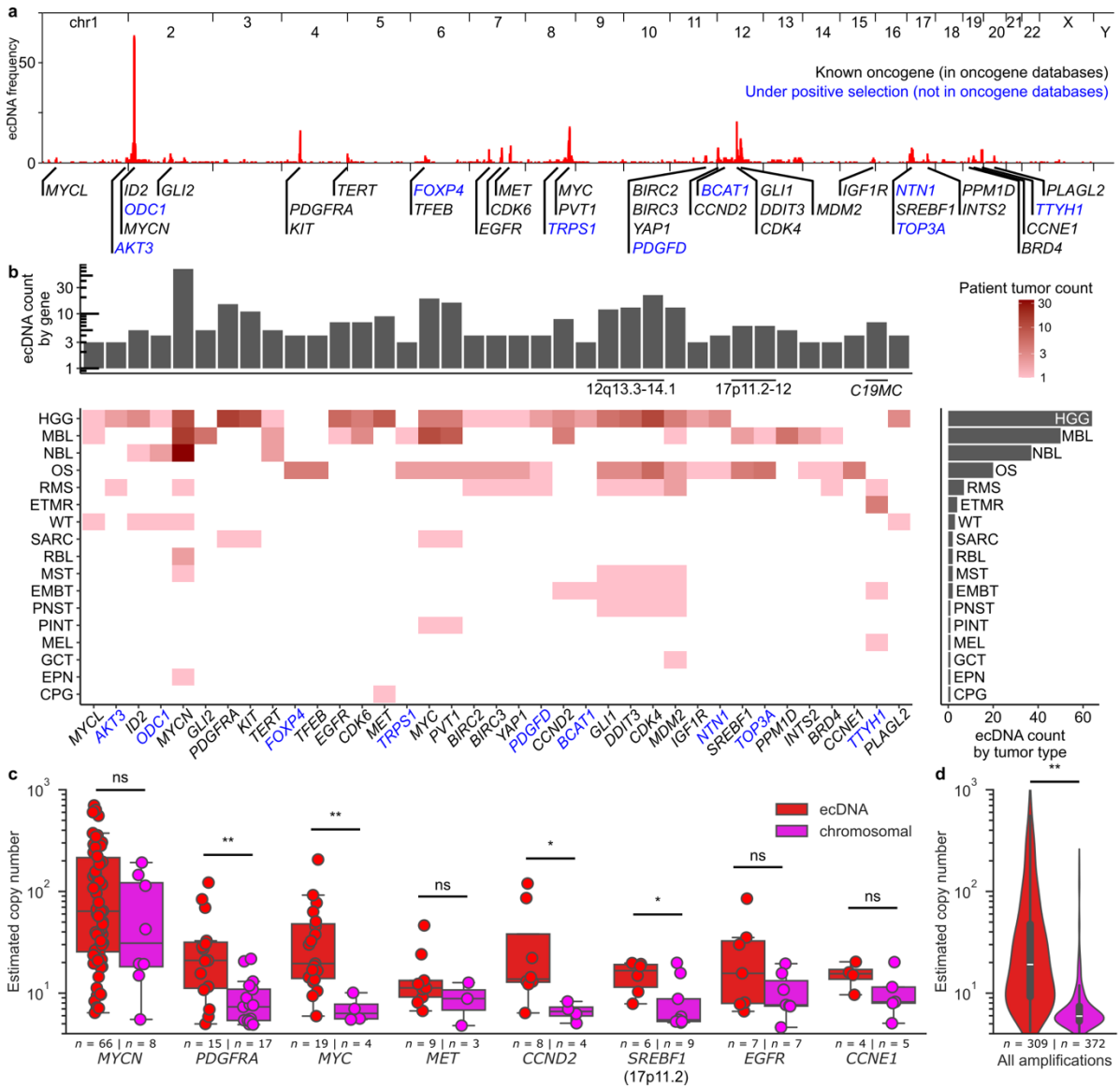
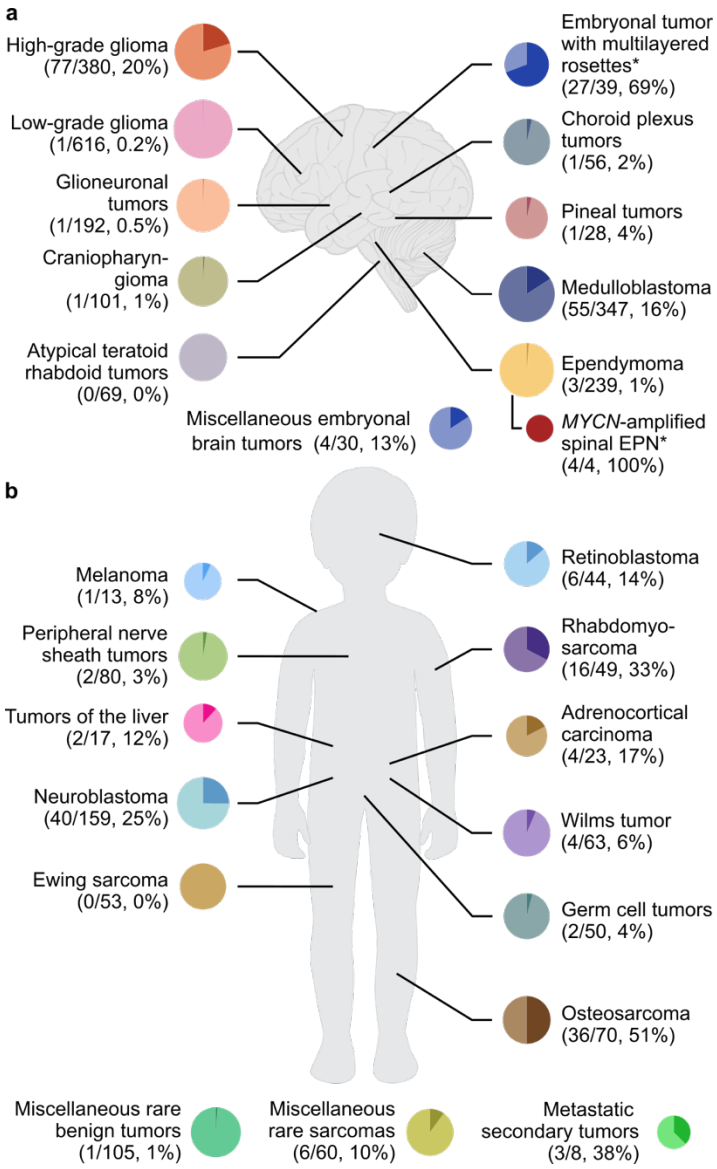


Figure 2: Genomic distribution and copy number of recurrent ecDNA amplifications in pediatric tumors. (**a**) Histogram of ecDNA amplifications in pediatric tumors across the reference human genome. 106 contiguous genomic loci were recurrently ($n \geq 3$) amplified containing 49 different oncogenes. Selected genes are shown: black, oncogenes listed in the COSMIC or ONGene databases; blue, other recurrently amplified genes. (**b**) Amplifications in (**a**) by tumor type. Heatmap indicates the number of cases with ecDNA amplification of a given gene in that tumor type. Right barplot indicates the number of ecDNA sequences amplifying any of the indicated genes for each cancer type. Top 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; OS, osteosarcoma; RMS, rhabdomyosarcoma; ETMR, embryonal tumor with multilayered rosettes; WT, Wilms 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.

495 **Figure 3: Prevalence of ecDNA**
496 **across pediatric tumor types.**
497 Across 39 pediatric tumor categories,
498 we report ecDNA in 23. Pie chart
499 area is proportional to the log of the
500 sample size for a given tumor.
501 *Includes additional samples from an
502 external validation set analyzed with
503 different methods (see **Methods**).
504



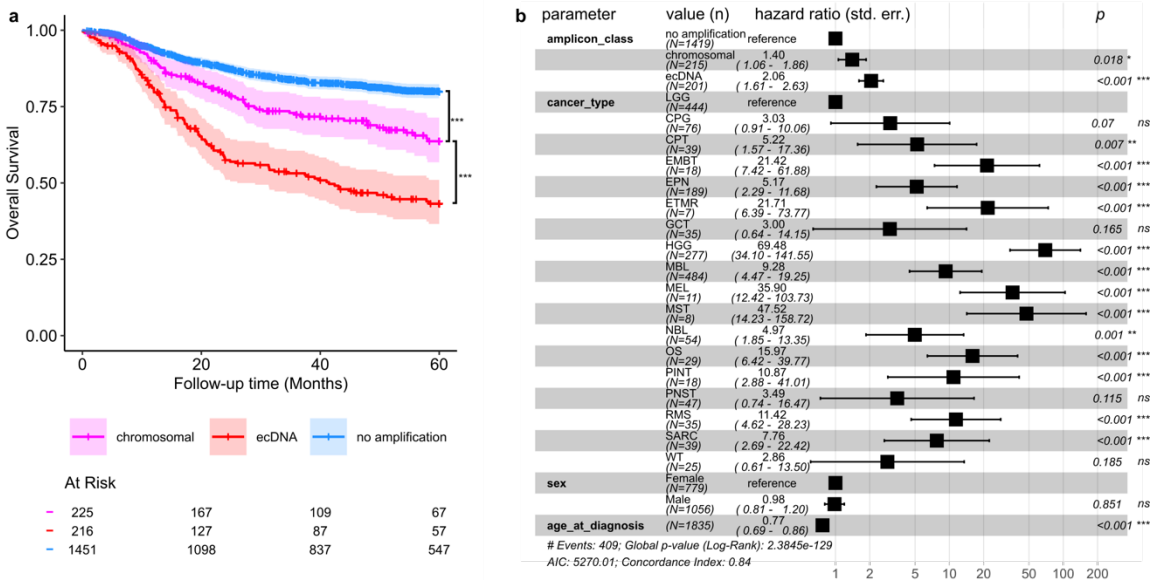
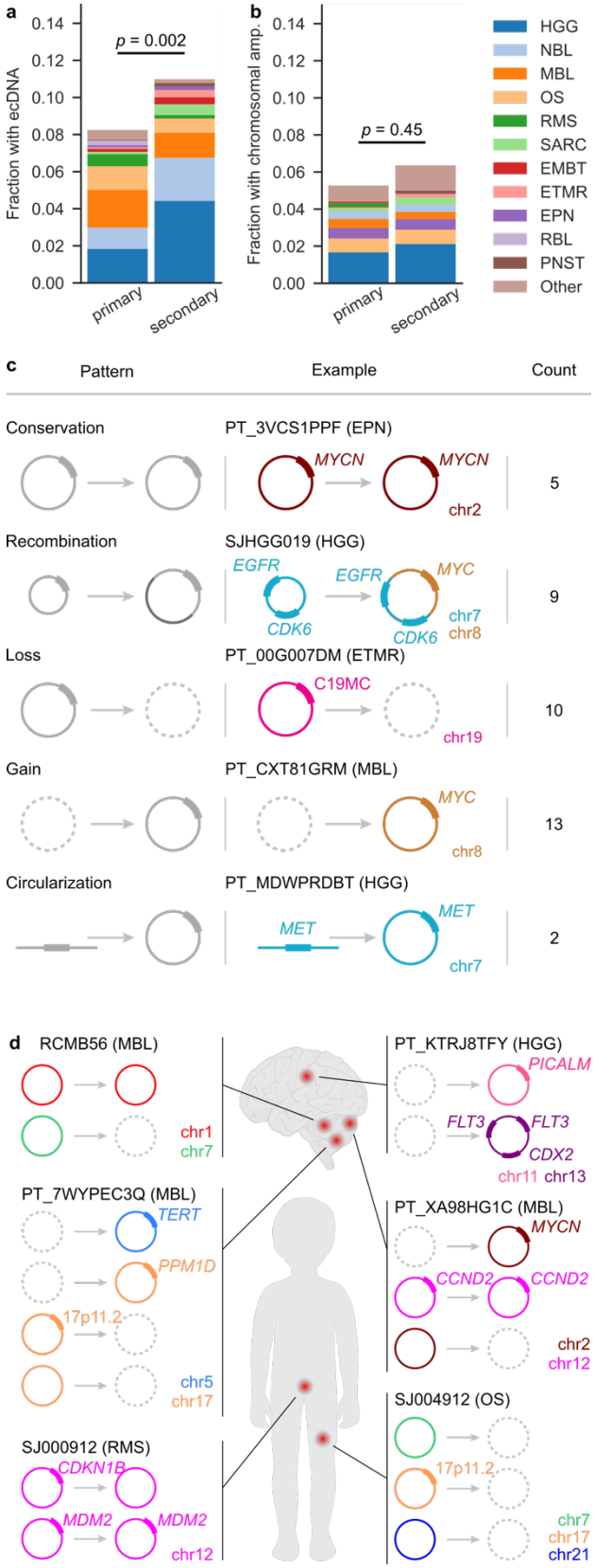
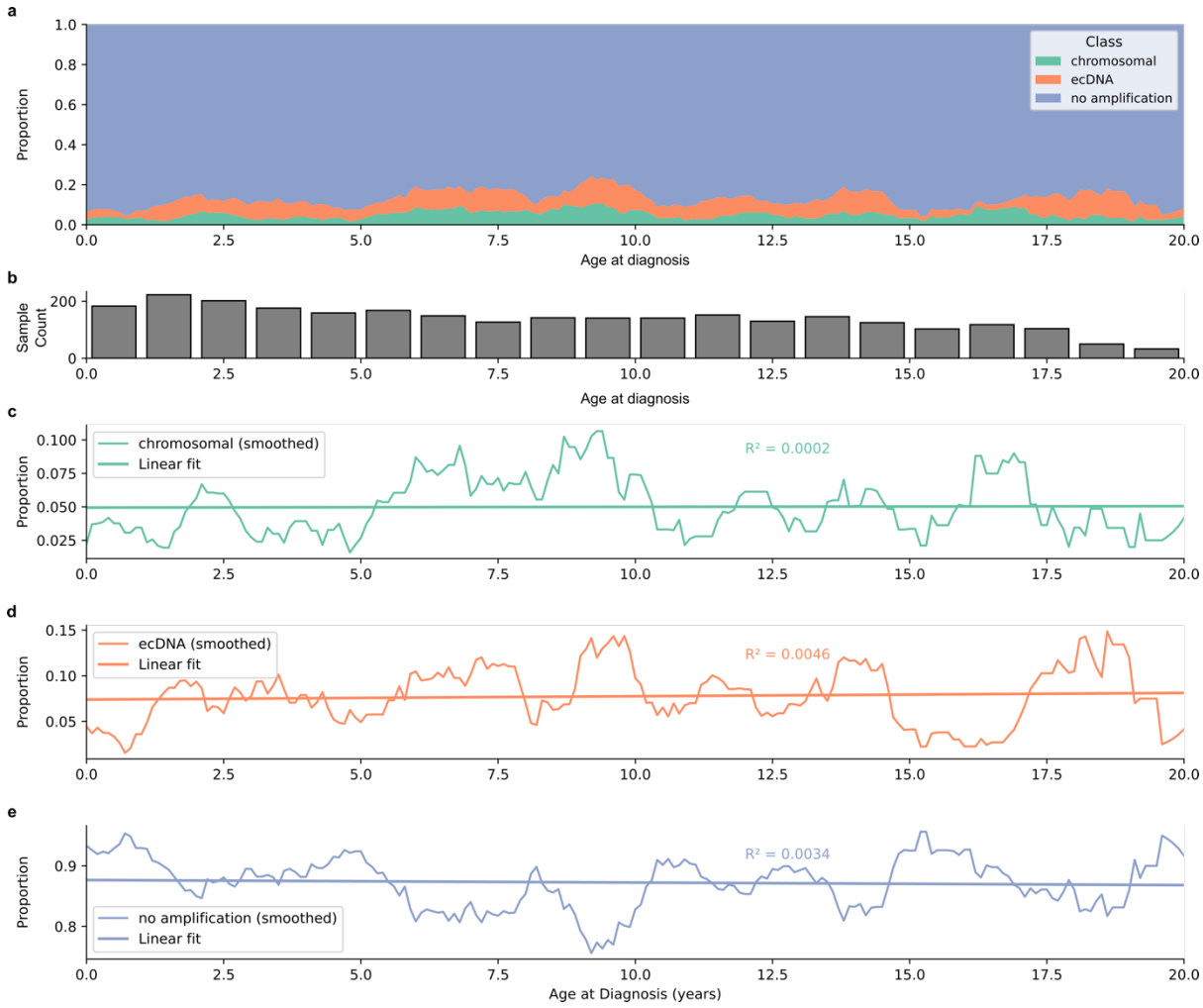


Figure 4: Overall survival of pediatric patients with ecDNA(+) tumors. Sample was limited to tumor types with at least 5 cases, at least 1 ecDNA(+) sample, and at least 1 clinical event. (**a**) Patients with ecDNA(+) tumors have significantly worse outcomes than patients with tumors containing no amplification (adjusted p < 2e-16) or patients with chromosomal amplifications (adjusted p = 7e-6; log-rank test with Benjamini-Hochberg correction). (**b**) Cox regression indicates twofold risk of death for patients with ecDNA compared those without amplifications (hazard ratio 2.1, n = 1836, p = 5e-9) and one-and-a-half-fold relative to patients with chromosomal amplifications only (HR 1.5, p = 0.02). ***$q<0.001$; **$q<0.01$; *$q<0.05$; ns, $q\geq0.05$.

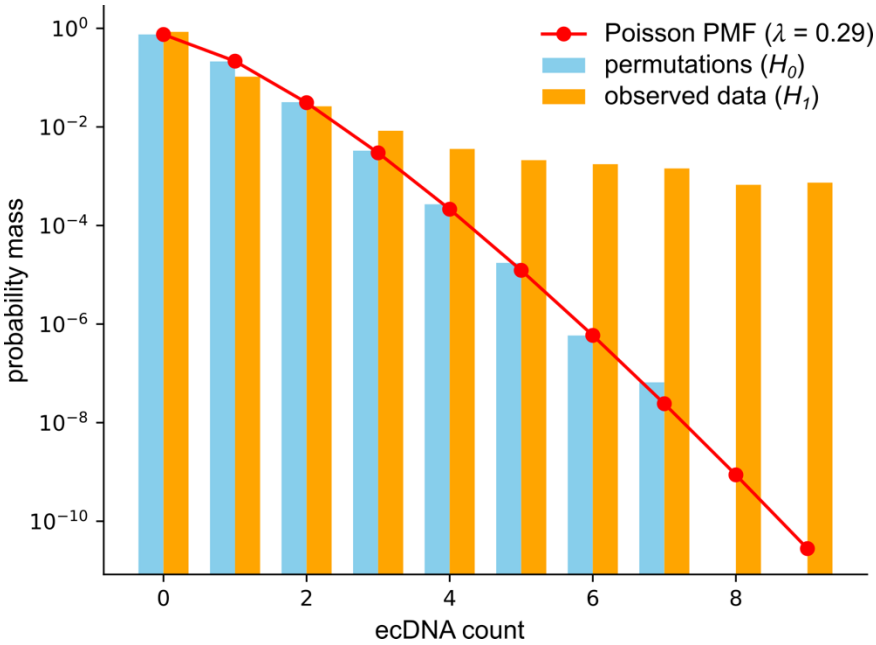
Figure 5: Sequence evolution of ecDNA during disease progression. (**a-b**) Proportion of (**a**) tumors with ecDNA and (**b**) tumors with chromosomal amplification only, by tumor type. *p*-values derived from permutation tests stratified by tumor type. (**c**) 39 ecDNA sequences were observed longitudinally 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). (**d**) 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; OS, osteosarcoma.



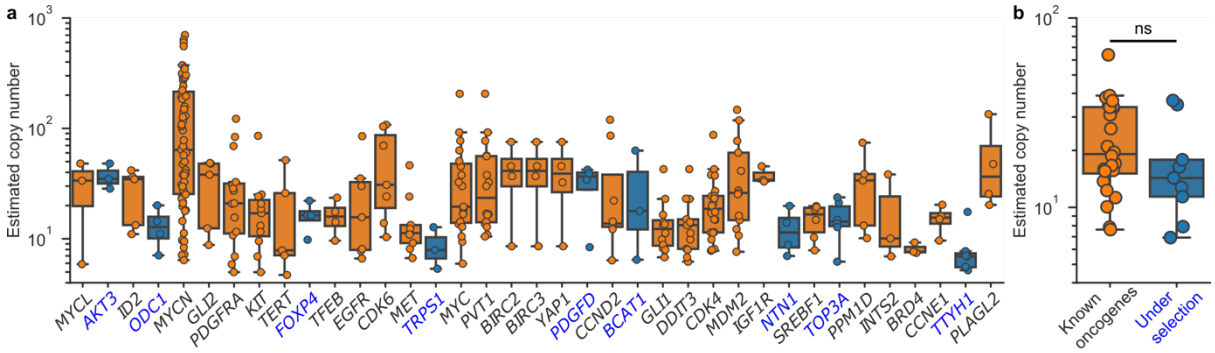
SUPPLEMENTARY FIGURES



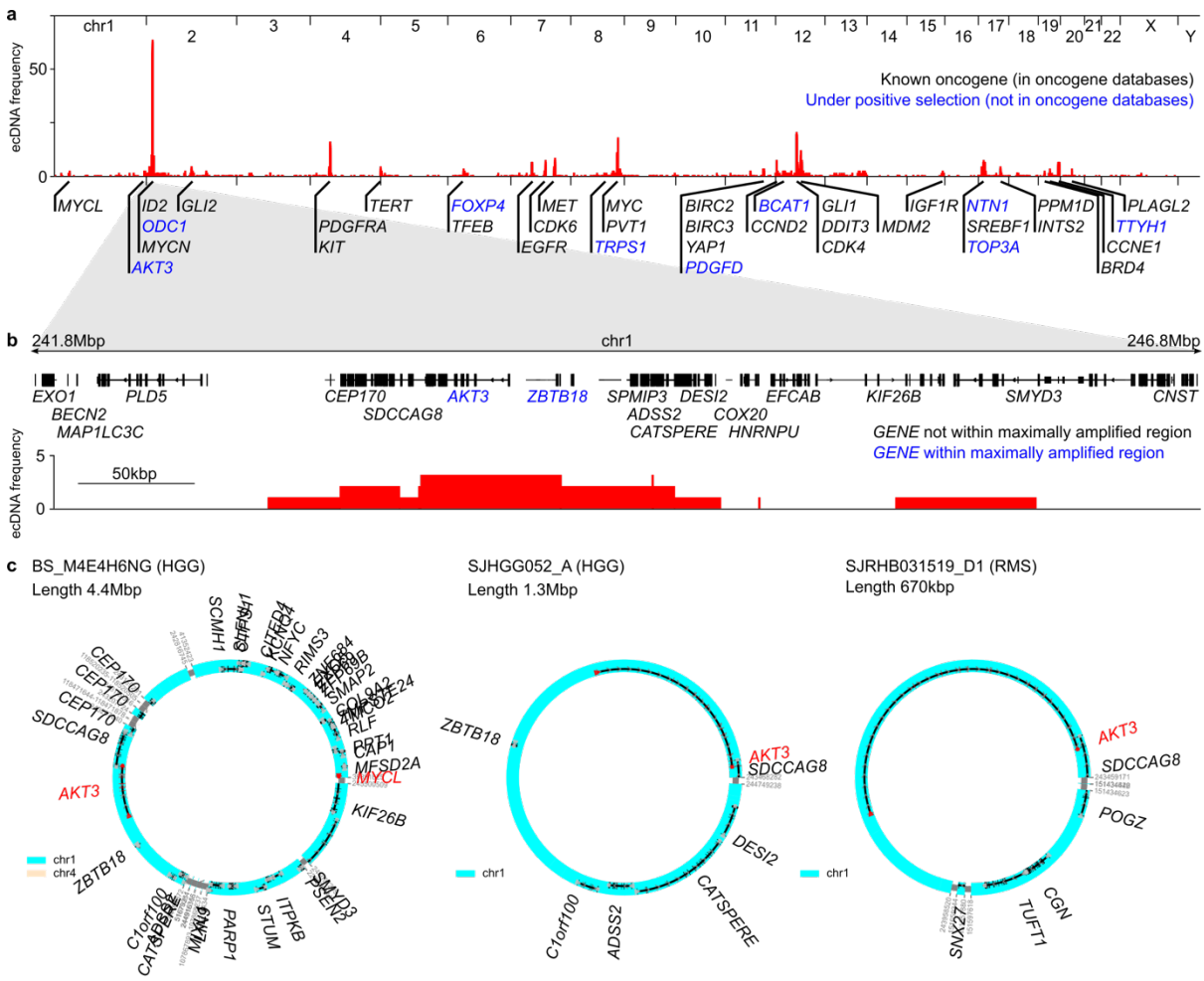
Supplementary Figure 1: Ordinary least squares regression of focal amplifications in patient tumors on age at diagnosis. (**a**) Ages at diagnosis were binned to 0.1-year resolution, and class proportions (ecDNA, chromosomal amplification, or no amplification) were computed for each age group. To reduce 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 ($R^2 = 0.003$, $p = 0.41$).



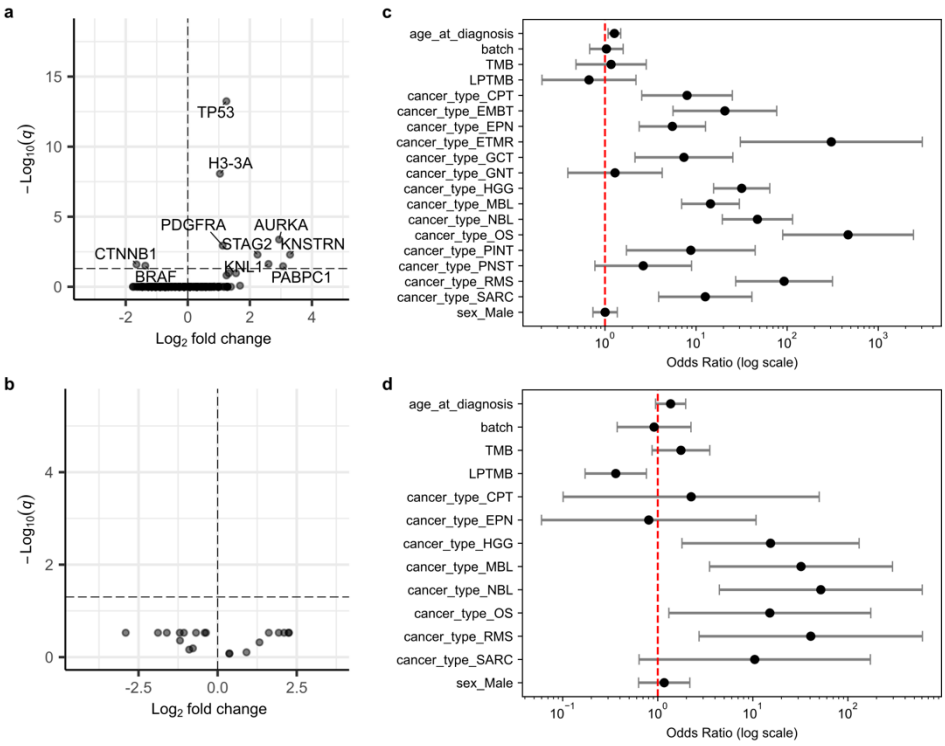
Supplementary Figure 2: Expected and observed distributions of ecDNA across the human reference genome. A null hypothesis H_0 in which ecDNA is uniformly and randomly distributed across the genome was modelled explicitly as a Poisson distribution with rate parameter λ proportional to the number and average 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.



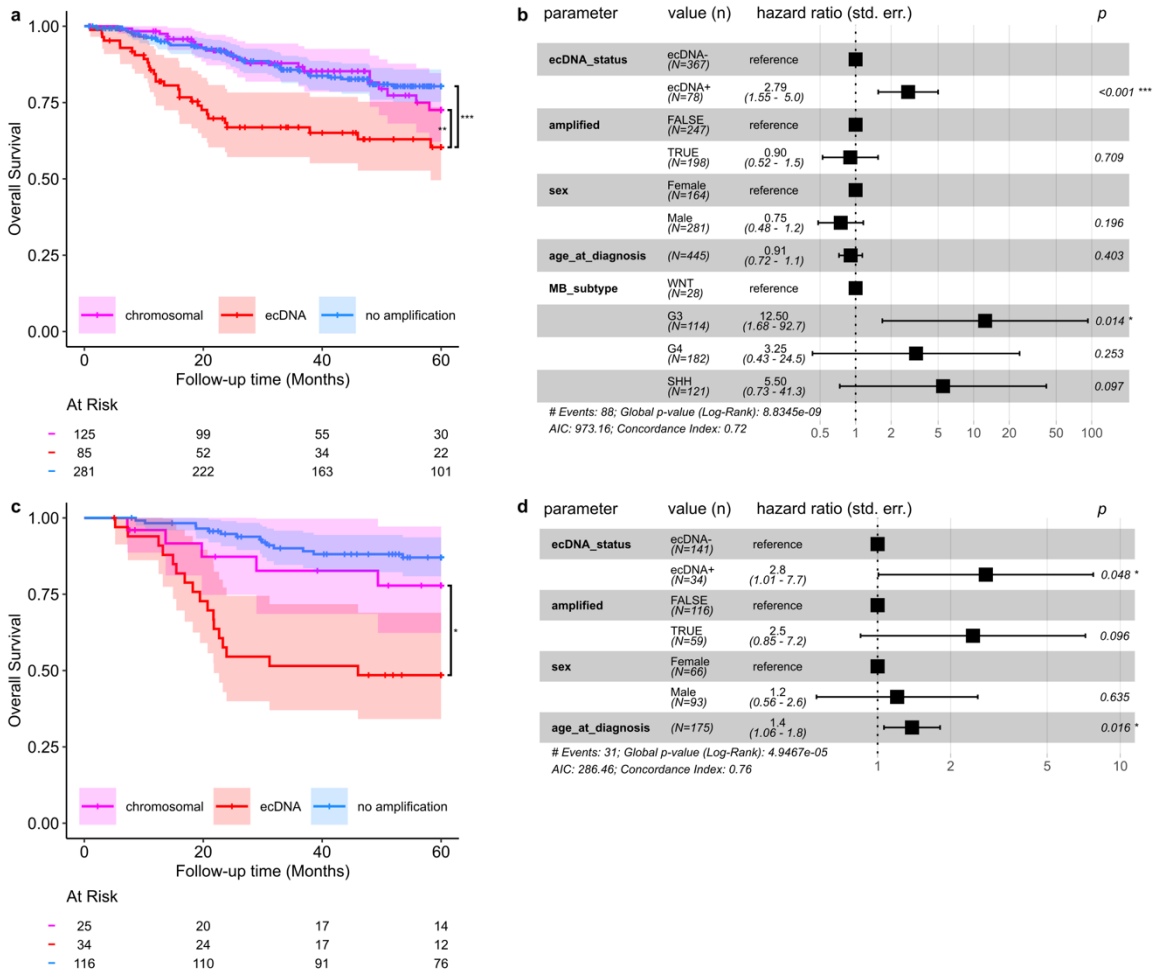
Supplementary Figure 3: Copy number of genes recurrentlu amplified on ecDNA. (**a**) Estimated copy number of each amplification for genes of interest indicated in **Figure 2**. (**b**) Medians of gene copy numbers in (**a**); two-sided Mann-Whitney *U* test; $U = 138$; $p = 0.23$. Medians of *PVT1*, *BIRC2*, *BIRC3* and *DDIT3* were excluded for violations of the indpendence assumption.



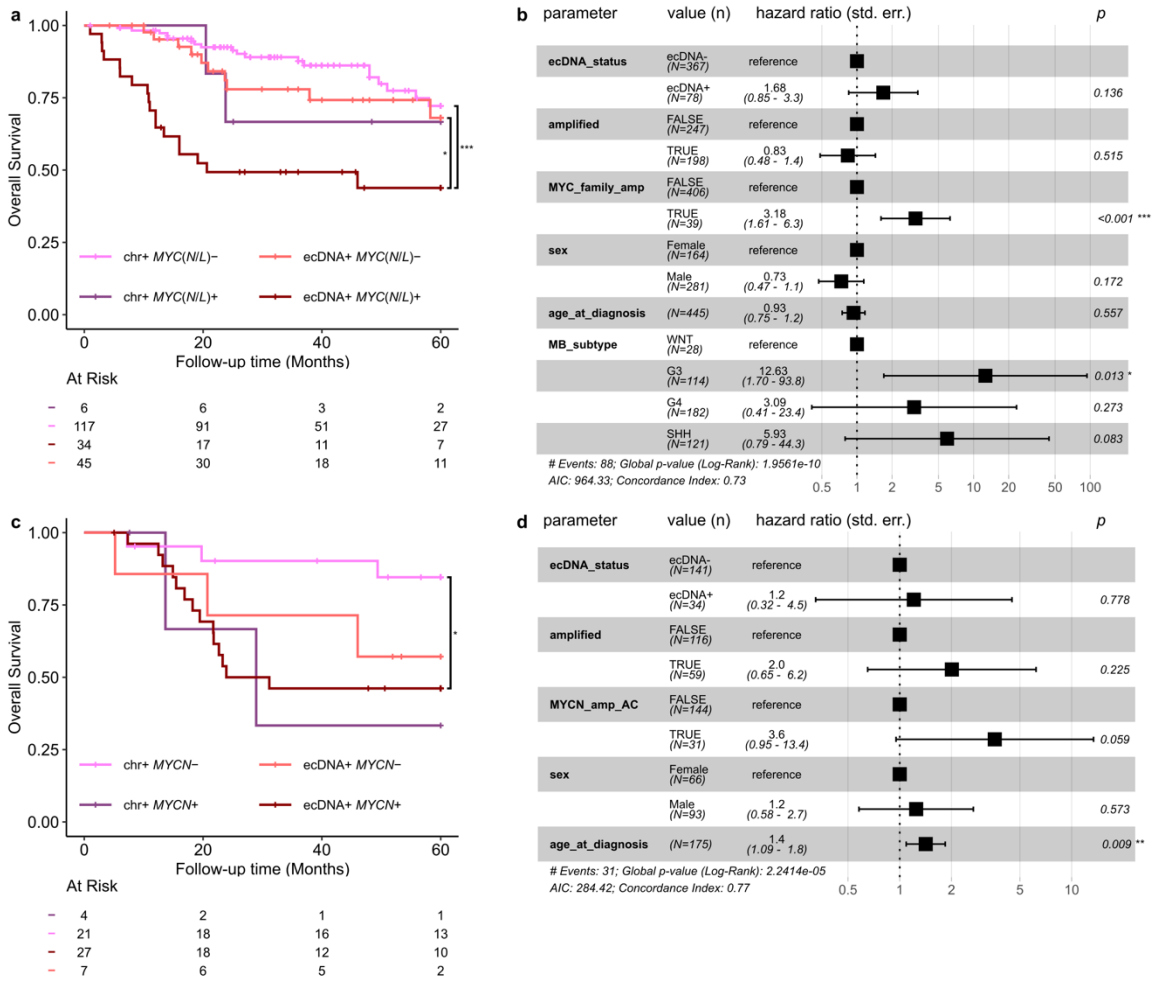
Supplementary Figure 4: A recurrently ecDNA-amplified locus contains putative oncogene *AKT3*. (**a**) Panel reproduced from **Fig. 2a**. Chromosomal co-ordinates of ecDNA amplifications were stacked to generate a histogram of ecDNA amplifications across the genome. Selected recurrently amplified ($n \geq 3$) 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 5: Tumor mutation burden of ecDNA(+) and chromosomally amplified tumors. (**a**-**b**) Associations of "likely pathogenic" somatic simple variants, defined as simple variants with predicted functional consequence targeting cancer-related genes, in (**a**) amplified vs. nonamplified and (**b**) ecDNA(+) vs. chromosomally amplified pediatric tumors (n = 1798). Genes are labelled in which pass the $q < 0.10$ significance threshold (χ^2 test with BH correction) indicated by the horizontal bar. (**c**-**d**) Maximum likelihood logistic regression models of (**c**) amplified vs. nonamplified (n = 1749) and (**d**) ecDNA vs. chromosomal amplification (n = 232) on age at diagnosis, batch, TMB, LPTMB, tumor type and sex. Reference tumor type is low-grade glioma, LGG.

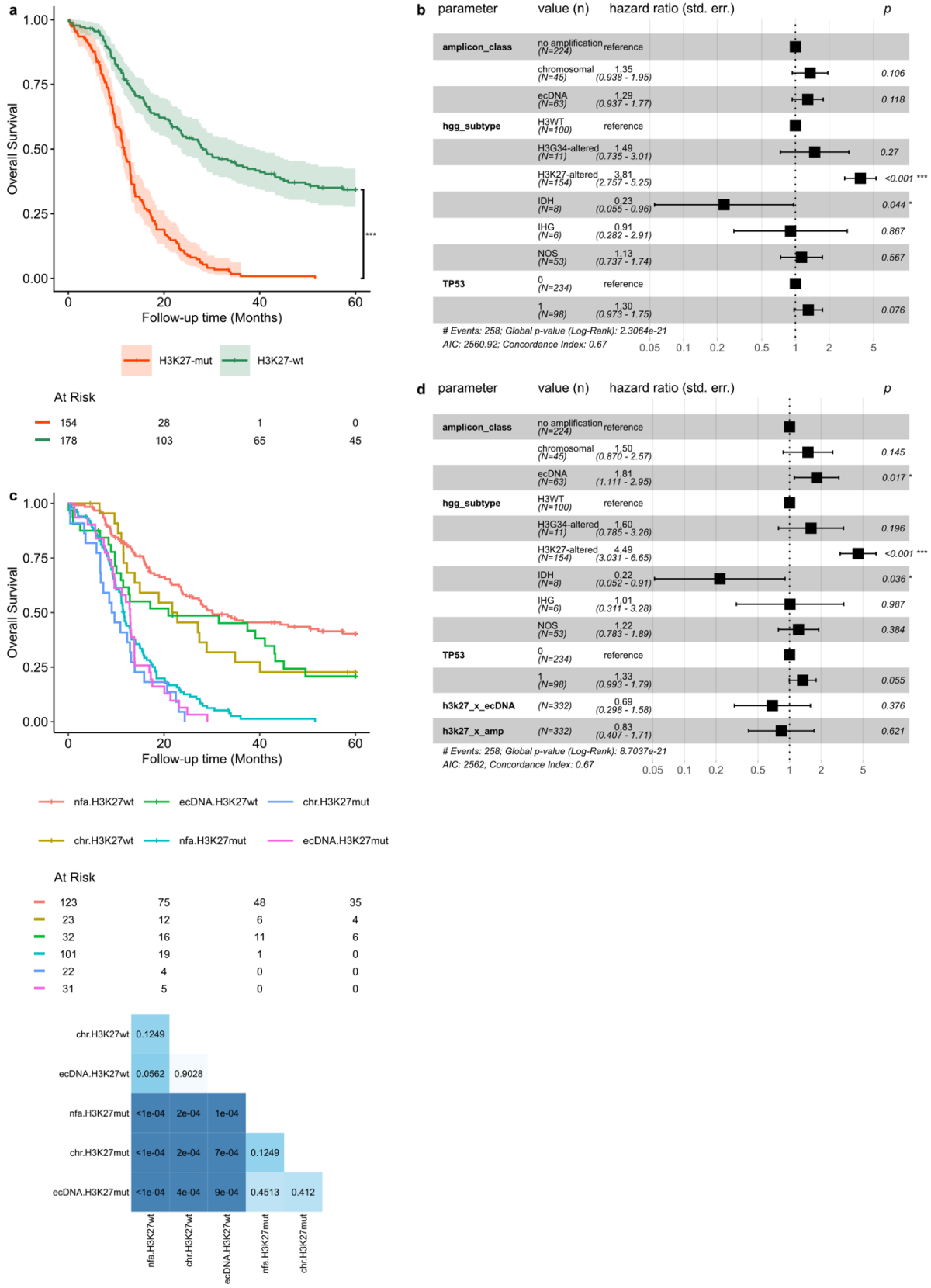


Supplementary Figure 6: Survival of medulloblastoma and neuroblastoma tumors w.r.t. ecDNA. (**a**) Survival of 491 MBL tumors stratified by presence of chromosomal or extrachromosomal amplification. (**b**) Cox proportional hazards regression of 445 MBL tumors from (**a**) with complete metadata for molecular subgroup and clinical parameters. (**c**) Survival of 175 NBL tumors stratified by chromosomal or extrachromosomal amplification. (**d**) Cox proportional hazards regression of NBL tumors from (**c**) on amplification class and clinical parameters. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

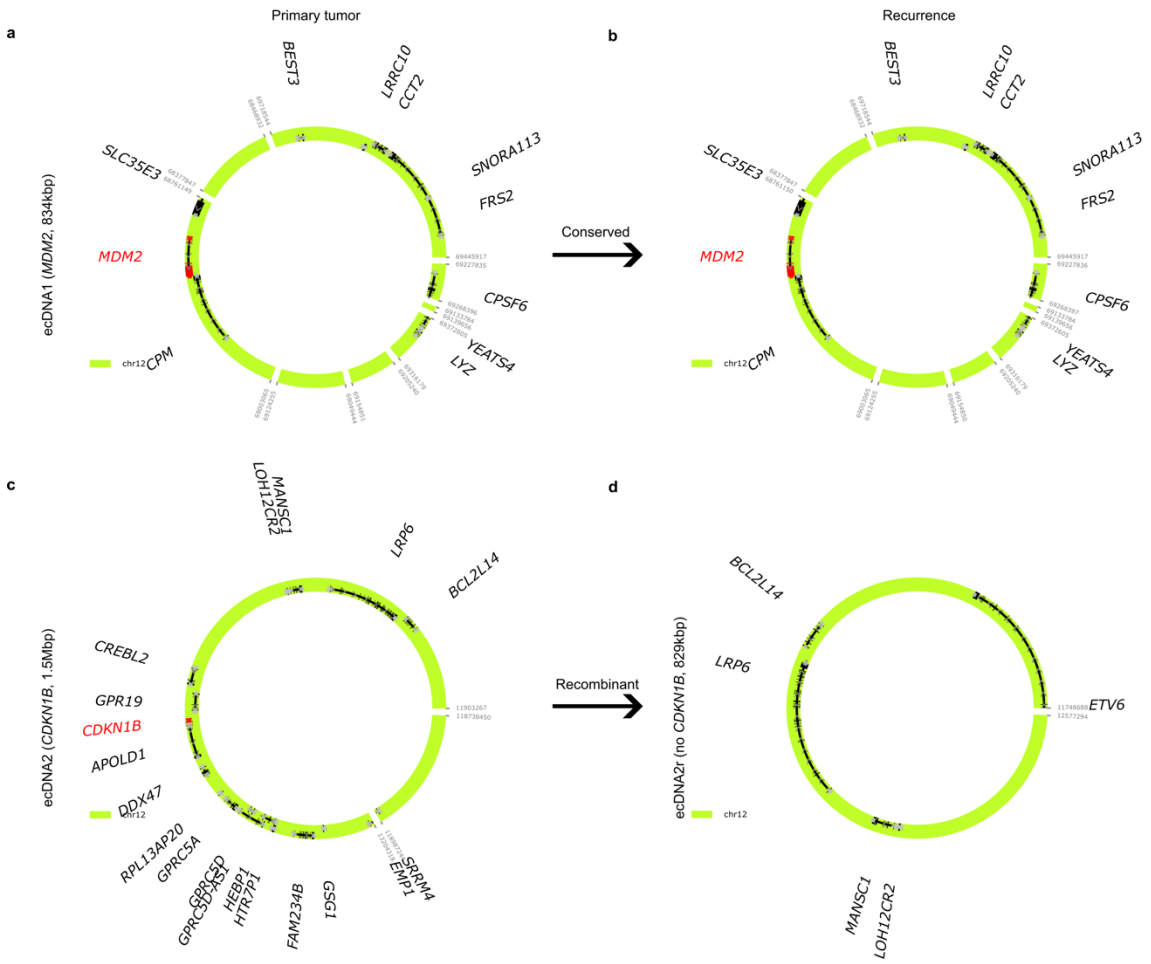


Supplementary Figure 7: Survival of Myc family amplified medulloblastoma and neuroblastoma tumors. (**a**) Survival of 212 MBL tumors with focal amplification stratified by chromosomal/ecDNA and *MYC* gene family amplification. chr+, chromosomal amplification; *MYC*(*N*/*L*)+/−, *MYC* gene family amplified/nonamplified. (**b**) Cox regression as in (**Supplementary Figure 6b**), but including an additional covariate for *MYC* family amplification. (**c**) Survival of 59 NBL tumors with focal amplification, stratified by chromosomal/ecDNA and *MYCN* amplification. (**d**) Cox regression as in (**Supplementary Figure 6d**), but including an additional covariate for *MYCN* amplification. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Supplementary Figure 8: Survival of pediatric high-grade glioma tumors w.r.t. ecDNA. (**a**) Survival of 332 pHGG tumors stratified by chromosomal or extrachromosomal amplification. (**b**) Cox proportional hazards regression of pHGG tumors from (**a**) on amplification class and clinical parameters. (**c**) Survival of pHGG tumors as in (**a**), but stratified additionally on presence/absence of H3K27 alteration. nfa, no focal amplification; chr, chromosomally amplified; H3K27wt, H3K27 wild-type; H3K27mut, H3K27-altered. (**d**) Cox regression as in (**b**), but including additional interaction terms stratifying on amplicon class and H3K27 status. h3k27_x_ecDNA, H3K27 alteration and ecDNA; h3k27_x_amp, H3K27 alteration and focal amplification. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.


a
Overall Survival
1.00
0.75
0.50
0.25
0.00
0
20
40
60
Follow-up time (Months)
***
H3K27-mut
H3K27-wt
At Risk
154 28 1 0
178 103 65 45
b
parameter
value (n)
hazard ratio (std. err.)
p
amplicon_class
no amplification (N=224) reference
chromosomal (N=45) 1.35 (0.938 - 1.95) 0.106
ecDNA (N=63) 1.29 (0.937 - 1.77) 0.118
hgg_subtype
H3WT (N=100) reference
H3G34-altered (N=11) 1.49 (0.735 - 3.01) 0.27
H3K27-altered (N=154) 3.81 (2.757 - 5.25) <0.001 ***
IDH (N=8) 0.23 (0.055 - 0.96) 0.044 *
IHG (N=6) 0.91 (0.282 - 2.91) 0.867
NOS (N=53) 1.13 (0.737 - 1.74) 0.567
TP53
0 (N=234) reference
1 (N=98) 1.30 (0.973 - 1.75) 0.076
Events: 258; Global p-value (Log-Rank): 2.3064e-21
AIC: 2560.92; Concordance Index: 0.67
0.05 0.1 0.2 0.5 1 2 5
c
Overall Survival
1.00
0.75
0.50
0.25
0.00
0
20
40
60
Follow-up time (Months)
nfa.H3K27wt
ecDNA.H3K27wt
chr.H3K27mut
chr.H3K27wt
nfa.H3K27mut
ecDNA.H3K27mut
At Risk
123 75 48 35
23 12 6 4
32 16 11 6
101 19 1 0
22 4 0 0
31 5 0 0
chr.H3K27wt 0.1249
ecDNA.H3K27wt 0.0562 0.9028
nfa.H3K27mut <1e-04 2e-04 1e-04
chr.H3K27mut <1e-04 2e-04 7e-04 0.1249
ecDNA.H3K27mut <1e-04 4e-04 9e-04 0.4513 0.412
nfa.H3K27wt
chr.H3K27wt
ecDNA.H3K27wt
nfa.H3K27mut
chr.H3K27mut
d
parameter
value (n)
hazard ratio (std. err.)
p
amplicon_class
no amplification (N=224) reference
chromosomal (N=45) 1.50 (0.870 - 2.57) 0.145
ecDNA (N=63) 1.81 (1.111 - 2.95) 0.017 *
hgg_subtype
H3WT (N=100) reference
H3G34-altered (N=11) 1.60 (0.785 - 3.26) 0.196
H3K27-altered (N=154) 4.49 (3.031 - 6.65) <0.001 ***
IDH (N=8) 0.22 (0.052 - 0.91) 0.036 *
IHG (N=6) 1.01 (0.311 - 3.28) 0.987
NOS (N=53) 1.22 (0.783 - 1.89) 0.384
TP53
0 (N=234) reference
1 (N=98) 1.33 (0.993 - 1.79) 0.055
h3k27_x_ecDNA (N=332) 0.69 (0.298 - 1.58) 0.376
h3k27_x_amp (N=332) 0.83 (0.407 - 1.71) 0.621
Events: 258; Global p-value (Log-Rank): 8.7037e-21
AIC: 2562; Concordance Index: 0.67
0.05 0.1 0.2 0.5 1 2 5

626



Supplementary Figure 9: 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.

641 **METHODS**

642 **Pediatric tumor whole genome sequencing (WGS)**

643 *Main cohort (PBTA+SJC).* Pediatric tumor WGS was identified from two pediatric cancer
644 genomic data repositories, corresponding to 2397 biosamples from 1873 patients in the
645 Pediatric Brain Tumor Atlas (PBTA) and Pediatric Neuro-Oncology Consortium (PNOC); and
646 1223 biosamples from 1094 patients in the St Jude Cloud (SJC)[12] Pediatric Cancer Genome
647 Project (PCGP; SJC-DS-1001)[70], Clinical Pilot (SJC-DS-1003)[71]; Genomes for Kids (G4K;
648 SJC-DS-1004)[72], Real-Time Clinical Genomics (RTCG; SJC-DS-1007); and Pediatric Brain
649 Tumor Portal (PBTP; SJC-DS-1014)[73]. Inclusion criteria were whole genome sequencing
650 annotated with a solid tumor diagnosis; hematological malignancies, nontumor, and
651 unannotated samples were excluded. Data preprocessing comprised each institution's
652 standard WGS bioinformatics pipelines. Alignments and genomic co-ordinates are w.r.t.
653 human assembly GRCh38 (hg38).

654 **ecDNA detection and classification from bulk WGS**

655 To detect ecDNA, samples in the WGS cohort were analyzed using AmpliconArchitect
656 (AA)[1] v1.2 and AmpliconClassifier (AC)[4] v1.1.2. Briefly, the AmpliconArchitect algorithm was
657 performed as follows. Copy number segmentation and estimation were performed using
658 CNVkit v0.9.6[67]. Segments with copy number (CN) ≥ 4.5 were extracted as "seed" regions
659 using AmpliconSuite-pipeline (April 2020 update)[74]. After confirming that sequencing depth
660 was greater than 10x for all samples (mosdepth v0.3.8[75]), read coverage was downsampled
661 to 10x. Then, for each seed, discordant read pairs indicative of genomic structural
662 rearrangement were identified within and up to 50kbp distal the query region. Genomic
663 segments were defined by genomic breakpoint locations (identified by discordant reads) and
664 by modulations in genomic copy number. A CN-aware breakpoint graph of the amplicon region
665 was constructed using the genome segments and breakpoints, and cyclic and non-cyclic paths
666 are extracted from the graph. Amplicons were classified as ecDNA(+), breakage-fusion-
667 bridge (BFB), complex non-cyclic, linear, or no focal amplification using the heuristic-based
668 companion script AmpliconClassifier. Biosamples with one or more amplicons classified as
669 ecDNA were labelled ecDNA(+). Biosamples without ecDNA but with one or more amplicons
670 classified as BFB, complex non-cyclic or linear were labelled as chromosomally amplified. All
671 other samples were annotated as having no focal amplifications.

672 All amplifications classified as ecDNA were manually reviewed for possible false positives,
673 marking for reanalysis amplicons satisfying any of the following criteria:
674 • amplicon has high variability in depth of coverage of mapped reads, indicating a
675 genomic locus of low mappability;
676 • cyclic subsequences which were assigned CN < 4.
677 These samples were reanalyzed using the latest AmpliconSuite versions as of Aug. 2024 (AA
678 v1.4.r2, AC v1.2.0). Samples reclassified in this way were BS_2F8PFTW5, BS_XB34VS6P,
679 BS_G6KYSGQF, BS_ GBSSZBMF, and BS_YMYESCY7.

680 **Patient metadata and tumor type annotations**

681 Clinical metadata including sex, age at diagnosis, histology, and tumor history (primary,
682 secondary, etc.) were available from the respective cloud genomics data platforms
683 https://pedcbioportal.kidsfirstdrc.org/ (PBTA) and https://pecan.stjude.cloud/ (SJC), and from

684 previous peer-reviewed publications[11,54]. Molecular diagnoses were integrated with
685 histological diagnoses for biosamples included in the Pediatric Brain Tumor Atlas (PBTA)[11] in
686 the following priority order: molecular subtype, DKFZ v12 methylation classification
687 (confidence threshold 0.9), DKFZ v11 methylation classification[13] (likewise), harmonized
688 diagnosis, original histological diagnosis. Diagnoses were subsequently collated and
689 harmonized between the data sources to annotate the tumor types and subtypes described
690 herein (**Supplementary Table 11**). Mutation status of high-grade gliomas for *TP53*, *IDH*, and
691 H3 genes were extracted from diagnosis annotations.

692 Annotations of tumor purity were available for 2064 biosamples representing 1679 tumors
693 from the PBTA cohort from a previous publication[11] using the THetA2 software[76]. To annotate
694 SJC samples, we applied THetA2 to an additional 1231 samples representing 1100 tumors.

695 **Supplementary tumor WGS**

696 *Additional MYCN-amplified spinal ependymomas.* WGS data from 3 pediatric *MYCN*-amplified
697 spinal ependymomas was peprocessed, and aligned according to best practices to human
698 reference GRCh38. Focal amplifications and ecDNA were analyzed as above using
699 AmpliconArchitect v1.5.r2 and AmpliconClassifier v1.3.3. Results were used to validate
700 ecDNA amplification of *MYCN* in spinal ependymoma and were excluded from other analyses.

701 *Additional ETMR tumors.* WGS data of 38 biosamples from 30 ETMR tumors including 29
702 samples from a previous publication[27] were uniformly reprocessed and aligned to human
703 reference GRCh37. Amplicon reconstruction and classification was performed using default
704 paramaterizations for AmpliconArchitect v1.5.r2 and AmpliconClassifier v1.3.7. To improve
705 detection sensitivity of low-copy *C19MC* amplifications in 11 samples, AmpliconArchitect was
706 re-run skipping the default copy number variant detection step and instead specifying the
707 genomic co-ordinates of suspected *C19MC* amplifications according to the following rules: for
708 samples where *C19MC* amplification was detected a different sample from the same tumor,
709 the boundaries of that amplicon were used; and for tumors for which a low-copy gain was
710 identified (copy number estimate from bulk WGS between 2.5 and 4.5), the boundaries of the
711 low-copy gain were used. Results were used to validate ecDNA amplification of *C19MC* in
712 ETMR and excluded from other analyses.

713 *Additional pediatric high-grade gliomas.* Previously published WGS data corresponding to 53
714 pediatric HGG tumors with accompanying clinical metadata including outcomes were
715 accessed from the archival sequencing data repository[54]. Amplicon detection and
716 classification was performed as in the main cohort above. Data were included in pHGG
717 survival analyses and excluded elsewhere.

718 *Additional neuroblastomas.* Previously published amplicon annotatations and associated
719 clinical metadata for 114 neuroblastoma tumors[10] were included in survival analyses of NBL.

720 *Additional medulloblastomas*. Previously published amplicon annotatations and associated
721 clinical metadata for 223 additional medulloblastoma tumors[7] were also included in survival
722 analyses and excluded elsewhere.

723 **Somatic variant data collection and analysis**

724 *Somatic variants.* Consensus simple somatic variant data were retrieved from the respective
725 institutional somatic variant calling and annotation pipelines which passed standard quality
726 control thresholds[11,12]. Annotations were sourced from Variant Effect Predictor (VEP)[77] and

727 Ensembl release v105 (CBTN) or v100 (St Jude). Simple somatic variants were defined as
728 likely pathogenic which satisfied the following criteria:
729 • affecting a known cancer gene, defined as tumor suppressor and oncogenes included
730 in the COSMIC Cancer Gene Census v101;
731 • Ensembl consequence (IMPACT) HIGH or MODERATE;
732 • PolyPhen[78] predicted functionality 'possibly damaging' or 'probably damaging' if
733 available; and
734 • SIFT[79] predicted functionality 'deleterious' or 'deleterious_low_confidence', if
735 available.
736 For association testing, likely pathogenic somatic variants were aggregated at the gene level;
737 ie., biosamples were considered altered at gene *G* which had any likely pathogenic variant
738 mapping to *G*. Association was tested using the χ^2 test for independence with Bonferroni
739 correction where at least 5 samples were altered at a given gene and at least 5 amplified;
740 otherwise the association was considered underpowered and therefore not tested.

741 *Germline variants.* Annotations of pathogenicity for germline variants of the OpenPBTA
742 subcohort were retreived from a previous publication[80]. Association testing w.r.t. focal
743 amplification and ecDNA was performed as described above for somatic variants. Variant
744 enrichment was plotted using EnhancedVolcano v1.24.0.

745 **Logistic regressions**

746 For all logistic regressions, tumor types were included which had at least 5 cases and at least
747 1 ecDNA(+) and 1 chromosomally amplified case. Numerical covariates were z-score
748 normalized according to best practice.

749 *Tumor purity on amplicon class.* To determine whether tumor purity influences ecDNA
750 detection efficiency, we fitted logistic regressions of amplicon class (ecDNA, chromosomal
751 amplification, or no amplification) on sex, age at diagnosis, and cancer type, including and
752 excluding tumor purity (as estimated by ThetA2) as an additional covariate. Logistic regression
753 models were implemented using scikit-learn v1.5.2[81] using the Newton conjugate gradient
754 solver and no penalty term. Models with and without tumor purity as an additional covariate
755 were compared by likelihood ratio test.

756 *Tumor mutation burden on amplicon class.* Regressions were fitted on age at diagnosis, sex,
757 source (CBTN or St Jude), tumor mutation burden (TMB), likely pathogenic mutation burden
758 (LPTMB) and tumor type using statsmodels v0.14.6 and plotted using base matplotlib v3.10.3.

759 **Survival analyses**

760 Patient overall survival data were included in PBTA clinical metadata and provided upon
761 request from St. Jude. 223 medulloblastoma tumors from the International Cancer Genome
762 Consortium and analyzed in a previous publication[7] were also included in survival analyses.
763 St. Jude survival data did not include date of most recent follow-up, so survival data were
764 censored at the date of data collection. Tumors with linear, complex noncyclic, or breakage-
765 fusion-bridge amplifications but without ecDNA were classified as chromosomal. Kaplan-
766 Meier and Cox Proportional Hazards regressions were performed in R 4.5.3 with the survival[82]
767 v3.8.6 and survminer v0.5.2 packages.

768 *Kaplan-Meier (KM) regressions.* Sample was limited to tumor types with at least 5 cases, at
769 least 1 death and at least 1 ecDNA(+) sample (n = 1892; 225 ecDNA(+), 216 with

chromosomal amplification only, and 1451 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 complete annotations for sex and age at diagnosis, resulting in 56 exclusions (n = 1836). 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[81] v1.5.2. Significance testing was performed using Student's *t*-test implemented in statsmodels[83] v0.14.2.

Permutation testing of amplification in primary and secondary tumors

To test whether ecDNA (or chromosomal amplification) is more frequent in secondary than primary tumors, a permutation test was constructed as follows. Let the test statistic $T = f_s - f_p$, where f_s and f_p are the observed frequencies of ecDNA+ tumors among secondary and primary tumors respectively. Then, permutations are constructed as follows. For paired samples, (ie, the same tumor was sampled at diagnosis and again at progression), the primary/secondary labels are swapped with 50% probability. For unpaired samples, for each tumor type, primary/secondary labels are shuffled among samples of that tumor type. The permutation procedure thus samples from a null distribution for which the distribution of ecDNA across tumor types is unchanged, but without any association with primary/secondary status. 1000 permutations were performed and a *p*-value calculated according to standard practice.

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[84] v1.10.2. Multiple hypothesis correction was performed where appropriate using the Benjamini-Hochberg correction implemented in statsmodels[83] v0.14.2 and reported as *q*-values or adjusted *p*-values. All statistical tests described herein were two-sided unless otherwise specified. For all tests performed on patient or biosample sets which require an independence assumption, samples from the same case were deduplicated according to the following priority order:

- any biosample classified ecDNA(+);
- and if any biosample with focal chromosomal amplification;
- the most recent biosample.

812 **Gene fusions**

813 Fusion genes were identified by adjacency on the ecDNA sequence assemblies. For
814 samples from the PBTA subcohort, fusion gene expression was confirmed from RNA-seq data
815 as described in the source publication[11].

816 **Longitudinal subcohort (Supplementary Figure 9; Supplementary Table 10)**

817 Patient tumors were included in the set of longitudinal cases which satisfied the following
818 criteria:
819 - *St Jude and PNOC cohorts.* For multiple tumors from the same patient, exactly one is
820 annotated as "diagnosis", and the others are labelled progressive, recurrent, or
821 metastasis;
822 - *PBTA X00 and X01 cohorts.* Multiple tumors from the same patient for which the
823 difference between the dates of diagnosis of the initial and the most recent biosample
824 exceeds 30 days.

825 Similarity scores reported between amplicons were calculated using AmpliconClassifier
826 feature_similarity.py v1.1.2. The formal definition and motivation of the similarity score is
827 described previously[15].

828 Circular sequences illustrated in **Supplementary Figures 4** & **9** reflect the highest-copy
829 cyclic sequence detected using AmpliconArchitect containing the oncogene of interest.
830 Figures were generated using CycleViz v0.2.1.

831 **ACKNOWLEDGMENTS**

832 This work is supported by a generous endowment by the Clayes Foundation to the Research
833 Center for Neuro-Oncology and Genomics within the Rady Children's Institute for Genomic
834 Medicine, a Hannah's Heroes St. Baldrick's Scholar Award (L.C.), the Dragon Master
835 Foundation (L.C.), the Takeda Science Foundation (O.S.C.), funding from the National
836 Institutes of Health (NIH) National Institute of Neurological Disorders and Stroke Institute
837 R01NS132780 (L.C.), R21NS130137 (L.C.) and R21NS120075 (L.C.), the NIH National
838 Cancer Institute P30CA030199 (L.C. and K.Y.), P30CA030199-42-S1 (L.C. and K.Y.),
839 F31CA271777 (O.S.C.) U01CA253547 (J.P.M.), P30CA051008 (J.P.M.), U24CA220341
840 (J.P.M.), U24CA295532 (J.P.M.), and U24CA264379 (V.B. and J.P.M.), the NIH National
841 Library of Medicine T15LM011271 (O.S.C.), the Japan Society for Promotion of Science
842 (JSPS) 25K23852 (O.S.C.), 24KF0252 (O.S.C.), and Postdoctoral Research Fellowship
843 F24404 (O.S.C. and D.K.), and a Moores Cancer Center Pilot Grant (L.C., V.B., and J.P.M.).
844 This project is also supported by the UCSD/Rady Children's Hospital pediatric hematology
845 oncology fellowship program (S.S.). D.R.G. gratefully acknowledges support by the Clinician
846 Scientist Program of the Mildred Scheel Cancer Career Center HaTriCS[4] at the University
847 Medical Center Hamburg-Eppendorf and the Else Kröner-Fresenius Stiftung (Grant
848 2025_EKEA.51). This work used Expanse cluster compute services at the San Diego
849 Supercomputer Center through allocation BIO210026 from the Advanced Cyberinfrastructure
850 Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by
851 National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and
852 #2138296. Support through grant P30CA030199 to the Genomics core facility at Sanford
853 Burnham Prebys (NCI designated Cancer Center) is gratefully acknowledged. The content is
854 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). This study makes use of data generated by the St. Jude Children's Research Hospital – Washington University Pediatric Cancer Genome Project and Pediatric Brain Tumor Portal; and the St. Jude Children's Research Hospital Genomes for Kids Study, Clinical Pilot and Real-Time Clinical Genomics projects. We thank J.H. Zhang, C. McLeod, D. Miller, Y. Zhu, D. Higgins and A. Resnick for facilitating data access, A. Calatroni for help with statistical analyses, R.J. Nowling for his open-source implementation of the likelihood ratio test, R. Wachs for figure design assistance, K. Ichimura and F.P.B. Dubois for scientific discussion, and anonymous reviewers for thoughtful and critical feedback.

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.Q.W., W.Z., H.H., and J.W. Formal analysis: O.S.C., S.S., E.Y.-C.C., R.K., R.G. Data curation: O.S.C., S.S., E.Y.-C.C., J.L., S.Z.W., E.R.-F., 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., H.H., M.B., J.L., and Y.Y.L. Resources: J.L., K.O., D.R.G., J.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., K.W.P., S.Z.W., J.G., and M.K. 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., D.K., 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 patient solid tumors with WGS in the OpenPBTA, X01, and PNOC datasets. St. Jude patient cohort: St. Jude Cloud (https://www.stjude.cloud). Inclusion criteria were patient solid tumors with WGS from the Pediatric Cancer Genome Project (PCGP; SJC-DS-1001), Clinical Pilot (SJC-DS-1003); Genomes 4 Kids (G4K; SJC-DS-1004), Real-Time Clinical Genomics (RTCG; SJC-DS-1007); and Pediatric Brain Tumor Portal (PBTP; SJC-DS-1014) datasets as of September 2022. WGS data for additional ETMR and spinal ependymomas will be made available upon reasonable request.

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 and other tools 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
- https://github.com/jbk708/theta-applet

Source code for variant calling and annotation pipelines are available at:

- https://github.com/kids-first/kf-somatic-workflow
- https://github.com/kids-first/kf-annotation-tools

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

REFERENCES

1. Turner, K. M. *et al.* Extrachromosomal oncogene amplification drives tumor evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).
2. Lange, J. T. *et al.* The evolutionary dynamics of extrachromosomal DNA in human cancers. *Nat. Genet.* **54**, 1527–1533 (2022).
3. Nathanson, D. A. *et al.* Targeted Therapy Resistance Mediated by Dynamic Regulation of Extrachromosomal Mutant EGFR DNA. *Science* **343**, 72 (2014).
4. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).
5. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371–376 (2018).
6. Kim, H. *et al.* Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 (2020).
7. Chapman, O. S. *et al.* Circular extrachromosomal DNA promotes tumor heterogeneity in high-risk medulloblastoma. *Nat. Genet.* **55**, 2189–2199 (2023).
8. Bailey, C. *et al.* Origins and impact of extrachromosomal DNA. *Nature* **635**, 193–200 (2024).
9. Koche, R. P. *et al.* Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat. Genet.* **52**, 29–34 (2020).
10. Rodriguez-Fos, E. *et al.* Mutational topography reflects clinical neuroblastoma heterogeneity. *Cell Genomics* **3**, (2023).
11. Shapiro, J. A. *et al.* OpenPBTA: The Open Pediatric Brain Tumor Atlas. *Cell Genomics* **3**, 100340 (2023).
12. McLeod, C. *et al.* St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem. *Cancer Discov.* **11**, 1082–1099 (2021).

947 13. Capper, D. *et al.* DNA methylation-based classification of central nervous system
948 tumours. *Nature* **555**, 469–474 (2018).
949 14. Deshpande, V. *et al.* Exploring the landscape of focal amplifications in cancer using
950 AmpliconArchitect. *Nat. Commun.* **10**, 392 (2019).
951 15. Luebeck, J. *et al.* Extrachromosomal DNA in the cancerous transformation of
952 Barrett's oesophagus. *Nature* **616**, 798–805 (2023).
953 16. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic*
954 *Acids Res.* **47**, D941–D947 (2019).
955 17. Liu, Y., Sun, J. & Zhao, M. ONGene: A literature-based database for human
956 oncogenes. *J. Genet. Genomics Yi Chuan Xue Bao* **44**, 119–121 (2017).
957 18. Helmsauer, K. *et al.* Enhancer hijacking determines extrachromosomal circular
958 MYCN amplicon architecture in neuroblastoma. *Nat. Commun.* **11**, 5823 (2020).
959 19. Mure, H. *et al.* Akt2 and Akt3 play a pivotal role in malignant gliomas. *Neuro-Oncol.*
960 **12**, 221 (2009).
961 20. Turner, K. M. *et al.* Genomically amplified Akt3 activates DNA repair pathway and
962 promotes glioma progression. *Proc. Natl. Acad. Sci.* **112**, 3421–3426 (2015).
963 21. Yang, J. *et al.* TRPS1 drives heterochromatic origin refiring and cancer genome
964 evolution. *Cell Rep.* **34**, 108814 (2021).
965 22. Arakawa, H. Netrin-1 and its receptors in tumorigenesis. *Nat. Rev. Cancer* **4**, 978–987
966 (2004).
967 23. Talbot, J. *et al.* Sonic hedgehog medulloblastomas are dependent on Netrin-1 for
968 survival. *Nat. Commun.* **16**, 5137 (2025).
969 24. Korshunov, A. *et al.* Focal genomic amplification at 19q13.42 comprises a powerful
970 diagnostic marker for embryonal tumors with ependymoblastic rosettes. *Acta Neuropathol.*
971 *(Berl.)* **120**, 253–260 (2010).

972 25. Lau, C. C. *et al.* Frequent amplification and rearrangement of chromosomal bands
973 6p12-p21 and 17p11.2 in osteosarcoma. *Genes. Chromosomes Cancer* **39**, 11–21 (2004).
974 26. Mendrzyk, F. *et al.* Isochromosome breakpoints on 17p in medulloblastoma are
975 flanked by different classes of DNA sequence repeats. *Genes. Chromosomes Cancer* **45**,
976 401–410 (2006).
977 27. Lambo, S. *et al.* The molecular landscape of ETMR at diagnosis and relapse. *Nature*
978 **576**, 274–280 (2019).
979 28. Mackay, A. *et al.* Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade
980 and Diffuse Intrinsic Pontine Glioma. *Cancer Cell* **32**, 520-537.e5 (2017).
981 29. Diaz, A. K. & Baker, S. J. The Genetic Signatures of Pediatric High-Grade Glioma:
982 No Longer a One-Act Play. *Semin. Radiat. Oncol.* **24**, 240–247 (2014).
983 30. Haase, S. *et al.* Hemispherical Pediatric High-Grade Glioma: Molecular Basis and
984 Therapeutic Opportunities. *Int. J. Mol. Sci.* **21**, 9654 (2020).
985 31. Li, B. K. *et al.* Pineoblastoma segregates into molecular sub-groups with distinct
986 clinico-pathologic features: a Rare Brain Tumor Consortium registry study. *Acta*
987 *Neuropathol. (Berl.)* **139**, 223–241 (2020).
988 32. Ghasemi, D. R. *et al.* MYCN amplification drives an aggressive form of spinal
989 ependymoma. *Acta Neuropathol. (Berl.)* **138**, 1075–1089 (2019).
990 33. Raffeld, M. *et al.* High level MYCN amplification and distinct methylation signature
991 define an aggressive subtype of spinal cord ependymoma. *Acta Neuropathol. Commun.* **8**,
992 101 (2020).
993 34. Potluri, V. R., Helson, L., Ellsworth, R. M., Reid, T. & Gilbert, F. Chromosomal
994 abnormalities in human retinoblastoma a review. *Cancer* **58**, 663–671 (1986).

995 35. Inazawa, J. *et al.* Simultaneous existence of double minute chromosomes and a
996 homogeneously staining region in a retinoblastoma cell line (Y79) and amplification of N-
997 myc at HSR. *Cancer Genet. Cytogenet.* **37**, 133–137 (1989).
998 36. Mysorekar, V. V., Harish, K., Kilara, N., Subramanian, M. & Giridhar, A. G.
999 Embryonal rhabdomyosarcoma of the chest wall: a case report and review of the literature.
1000 *Indian J. Pathol. Microbiol.* **51**, 274–276 (2008).
1001 37. Gil-Benso, R. *et al.* Cytogenetic and molecular findings related to
1002 rhabdomyosarcoma. An analysis of seven cases. *Cancer Genet. Cytogenet.* **144**, 125–133
1003 (2003).
1004 38. Hayashi, Y. *et al.* Characterization of an embryonal rhabdomyosarcoma cell line
1005 showing amplification and over-expression of the N-myc oncogene. *Int. J. Cancer* **45**,
1006 705–711 (1990).
1007 39. Weber-Hall, S. *et al.* Novel formation and amplification of the PAX7-FKHR fusion
1008 gene in a case of alveolar rhabdomyosarcoma. *Genes. Chromosomes Cancer* **17**, 7–13
1009 (1996).
1010 40. Goldstein, M., Meller, I., Issakov, J. & Orr-Urtreger, A. Novel Genes Implicated in
1011 Embryonal, Alveolar, and Pleomorphic Rhabdomyosarcoma: A Cytogenetic and
1012 Molecular Analysis of Primary Tumors. *Neoplasia* **8**, 332–343 (2006).
1013 41. Frascella, E. *et al.* Concomitant Amplif ication and Expression of PAX7-FKHR and
1014 MYCN in a Human Rhabdomyosarcoma Cell Line Carrying a Cryptic t(1;13)(p36;q14).
1015 *Cancer Genet. Cytogenet.* **121**, 139–145 (2000).
1016 42. Fletcher, J. A., Gebhardt, M. C. & Kozakewich, H. P. Cytogenetic aberrations in
1017 osteosarcomas: Nonrandom deletions, rings, and double-minute chromosomes. *Cancer*
1018 *Genet. Cytogenet.* **77**, 81–88 (1994).

1019 43. Gilbert, F., Potluri, V. R., Short, M. P., Kau, C.-L. & Lalatta, F. Retinoblastoma,
1020 chromosome abnormalities and oncogene expression. *Ophthalmic Paediatr. Genet.* **8**, 3–
1021 10 (1987).
1022 44. Atiye, J. *et al.* Gene amplifications in osteosarcoma—CGH microarray analysis.
1023 *Genes. Chromosomes Cancer* **42**, 158–163 (2005).
1024 45. dos Santos Aguiar, S., de Jesus Girotto Zambaldi, L., dos Santos, A. M., Pinto, W. &
1025 Brandalise, S. R. Comparative genomic hybridization analysis of abnormalities in
1026 chromosome 21 in childhood osteosarcoma. *Cancer Genet. Cytogenet.* **175**, 35–40 (2007).
1027 46. Sayles, L. C. *et al.* Genome-Informed Targeted Therapy for Osteosarcoma. *Cancer*
1028 *Discov.* **9**, 46–63 (2019).
1029 47. Yang, L. *et al.* FOXP4-AS1 participates in the development and progression of
1030 osteosarcoma by downregulating LATS1 via binding to LSD1 and EZH2. *Biochem.*
1031 *Biophys. Res. Commun.* **502**, 493–500 (2018).
1032 48. Both, J. *et al.* Identification of Novel Candidate Oncogenes in Chromosome Region
1033 17p11.2-p12 in Human Osteosarcoma. *PLOS ONE* **7**, e30907 (2012).
1034 49. Martin, J. W., Squire, J. A. & Zielenska, M. The Genetics of Osteosarcoma. *Sarcoma*
1035 **2012**, e627254 (2012).
1036 50. Granberg, I. & Mark, J. The chromosomal aberration of double-minutes in a human
1037 embryonic rhabdomyosarcoma. *Acta Cytol.* **15**, 42–45 (1971).
1038 51. Steilen-Gimbel, H. *et al.* A novel site of DNA amplification on chromosome 1p32-33
1039 in a rhabdomyosarcoma revealed by comparative genomic hybridization. *Hum. Genet.* **97**,
1040 87–90 (1996).
1041 52. Williamson, D. *et al.* Relationship between MYCN copy number and expression in
1042 rhabdomyosarcomas and correlation with adverse prognosis in the alveolar subtype. *J.*
1043 *Clin. Oncol.* **23**, 880–888 (2005).

1044 53. Sorensen, P. H. B. *et al.* PAX3-FKHR and PAX7-FKHR Gene Fusions Are
1045 Prognostic Indicators in Alveolar Rhabdomyosarcoma: A Report From the Children's
1046 Oncology Group. *J. Clin. Oncol.* **20**, 2672–2679 (2002).
1047 54. Dubois, F. P. B. *et al.* Structural variants shape driver combinations and outcomes in
1048 pediatric high-grade glioma. *Nat. Cancer* **3**, 994–1011 (2022).
1049 55. Schwalbe, E. C. *et al.* Molecular and clinical heterogeneity within MYC-family
1050 amplified medulloblastoma is associated with survival outcomes: A multicenter cohort
1051 study. *Neuro-Oncol.* **27**, 222–236 (2025).
1052 56. Holdhof, D. *et al.* Atypical teratoid/rhabdoid tumors (ATRTs) with SMARCA4
1053 mutation are molecularly distinct from SMARCB1-deficient cases. *Acta Neuropathol.*
1054 *(Berl.)* **141**, 291–301 (2021).
1055 57. Seong, B. K. A. *et al.* TRIM8 modulates the EWS/FLI oncoprotein to promote
1056 survival in Ewing sarcoma. *Cancer Cell* **39**, 1262 (2021).
1057 58. Arabzade, A. *et al.* EPEN-04. OVERWHELMING ZFTA-RELA NUCLEAR
1058 LOCALIZATION AS A THERAPEUTIC STRATEGY AGAINST EPENDYMOMA.
1059 *Neuro-Oncol.* **26**, 0 (2024).
1060 59. Murray, J. C. *et al.* Detection of Double Minute Chromosomes in a Child with Acute
1061 Lymphoblastic Leukemia. *Cancer Genet. Cytogenet.* **101**, 138–142 (1998).
1062 60. Zhang, H. *et al.* Characterization and integrated analysis of extrachromosomal DNA
1063 amplification in hematological malignancies. *Neoplasia* **56**, 101025 (2024).
1064 61. Potocki, L. *et al.* Molecular mechanism for duplication 17p11.2- the homologous
1065 recombination reciprocal of the Smith-Magenis microdeletion. *Nat. Genet.* **24**, 84–87
1066 (2000).

1067 62. Pentao, L., Wise, C. A., Chinault, A. C., Patel, P. I. & Lupski, J. R. Charcot–Marie–
1068 Tooth type 1A duplication appears to arise from recombination at repeat sequences
1069 flanking the 1.5 Mb monomer unit. *Nat. Genet.* **2**, 292–300 (1992).
1070 63. Kim, H. *et al.* Mapping extrachromosomal DNA amplifications during cancer
1071 progression. *Nat. Genet.* **56**, 2447–2454 (2024).
1072 64. Shoshani, O. *et al.* Chromothripsis drives the evolution of gene amplification in
1073 cancer. *Nature* **591**, 137–141 (2021).
1074 65. Tang, J. *et al.* Enhancing transcription–replication conflict targets ecDNA-positive
1075 cancers. *Nature* **635**, 210–218 (2024).
1076 66. Mayr, L. *et al.* Effective targeting of PDGFRA-altered high-grade glioma with
1077 avapritinib. *Cancer Cell* **43**, 740-756.e8 (2025).
1078 67. Gotoh, T. *et al.* Prediction of MYCN Amplification in Neuroblastoma Using Serum
1079 DNA and Real-Time Quantitative Polymerase Chain Reaction. *J. Clin. Oncol.* **23**, 5205–
1080 5210 (2005).
1081 68. Stepien, N. *et al.* Proof-of-Concept for Liquid Biopsy Disease Monitoring of MYC-
1082 Amplified Group 3 Medulloblastoma by Droplet Digital PCR. *Cancers* **15**, 2525 (2023).
1083 69. Madlener, S. *et al.* Clinical applicability of miR517a detection in liquid biopsies of
1084 ETMR patients. *Acta Neuropathol. (Berl.)* **145**, 843–846 (2023).
1085 70. Downing, J. R. *et al.* The Pediatric Cancer Genome Project. *Nat. Genet.* **44**, 619–622
1086 (2012).
1087 71. Rusch, M. *et al.* Clinical cancer genomic profiling by three-platform sequencing of
1088 whole genome, whole exome and transcriptome. *Nat. Commun.* **9**, 3962 (2018).
1089 72. Newman, S. *et al.* Genomes for Kids: The Scope of Pathogenic Mutations in Pediatric
1090 Cancer Revealed by Comprehensive DNA and RNA Sequencing. *Cancer Discov.* **11**,
1091 3008–3027 (2021).

1092 73. Wu, G. *et al.* Somatic histone H3 alterations in pediatric diffuse intrinsic pontine
1093 gliomas and non-brainstem glioblastomas. *Nat. Genet.* **44**, 251–253 (2012).
1094 74. AmpliconSuite: an end-to-end workflow for analyzing focal amplifications in cancer
1095 genomes | bioRxiv. https://www.biorxiv.org/content/10.1101/2024.05.06.592768v1.
1096 75. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes
1097 and exomes. *Bioinformatics* **34**, 867–868 (2018).
1098 76. Oesper, L., Satas, G. & Raphael, B. J. Quantifying tumor heterogeneity in whole-
1099 genome and whole-exome sequencing data. *Bioinformatics* **30**, 3532–3540 (2014).
1100 77. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122
1101 (2016).
1102 78. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting Functional Effect of Human
1103 Missense Mutations Using PolyPhen-2. *Curr. Protoc. Hum. Genet. Editor. Board*
1104 *Jonathan Haines Al* **0 7**, Unit7.20 (2013).
1105 79. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense
1106 predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
1107 80. Corbett, R. J. *et al.* Germline pathogenic variation impacts somatic alterations and
1108 patient outcomes in pediatric central nervous system tumors. *Nat. Commun.* **16**, 10282
1109 (2025).
1110 81. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*
1111 **12**, 2825–2830 (2011).
1112 82. Therneau, T. A package for survival analysis in R.
1113 83. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with
1114 Python. in *Proceedings of the 9th Python in Science Conference* (eds Walt, S. van der &
1115 Millman, J.) 92–96 (2010). doi:10.25080/Majora-92bf1922-011.

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