Stratified Medicine Paediatrics: Cell free DNA and serial tumour sequencing identifies subtype specific cancer evolution and epigenetic states

Sally L George[+1,2], Claire Lynn[+1,3], Reda Stankunaite[+1,3,4], Debbie Hughes[4], Carolin M. Sauer[5], Jane Chalker[6], Saira Waqar Ahmed[6], Minou Oostveen[6], Paula Z. Proszek[4], Lina Yuan[4], Ridwan Shaikh[4], Sabri Jamal[4], Ama Brew[4], Jennifer Tall[1], Tony Rogers[1], Steven C. Clifford[8], Josef Vormoor[9], Janet M. Shipley[10], Deborah A. Tweddle[8], Chris Jones[10], Courtney Willis[11], G.A. Amos Burke[12,29], Aditi Vedi[12], Lisa Howell[13], Robert Johnston[14], Helen Rees[15], Madeleine Adams[16], Angela Jesudason[17], Milind Ronghe[18], Martin Elliott[19], Emma Ross[20], Guy Makin[21], Quentin Campbell-Hewson[22], Richard G. Grundy[23], Jennifer Turnbull[23], Shaun Wilson[24], Victoria Lee[25], Juliet C. Gray[26], Sara Stoneham[27], Susanne A. Gatz[28,29], Lynley V. Marshall[1,2], Paola Angelini[2,10], John Anderson[6,7], George D. Cresswell[3,30], Trevor A Graham[3], Bissan Al-Lazikani[31], Isidro Cortés-Ciriano[5], Pamela Kearns[29], J. Ciaran Hutchinson[6], Darren Hargrave[7,32], Thomas S. Jacques[6,7], Michael Hubank[4], Andrea Sottoriva[3,33]✉, Louis Chesler[1,2]✉

[1] Paediatric Oncology Experimental Medicine Centre (POEM), The Institute of Cancer Research, London, UK
[2] Children and Young People's Unit, The Royal Marsden Hospital, London, UK
[3] Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK
[4] Clinical Genomics, Centre for Molecular Pathology, The Royal Marsden Hospital and Institute of Cancer Research, London, UK
[5] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, UK
[6] Department of Histopathology, Great Ormond Street Hospital for Children NHS Foundation Trust, UK
[7] Developmental Biology and Cancer Department, UCL GOS Institute of Child Health, London, UK
[8] Wolfson Childhood Cancer Research Centre, Newcastle University, UK
[9] Princess Máxima Center and University Medical Center, Utrecht, Netherlands
[10] Division of Molecular Pathology, The Institute of Cancer Research, London, UK
[11] Royal Aberdeen Children's Hospital, Aberdeen, UK
[12] Cambridge University Hospital NHS Foundation Trust, Cambridge, UK
[13] Alder-Hey Children's Hospital, Liverpool, UK
[14] Belfast- Royal Belfast Hospital for Sick Children, Belfast, UK
[15] Bristol Royal Hospital for Children, Bristol, UK
[16] Noah's Ark Children's Hospital for Wales, Cardiff, UK
[17] Royal Belfast Hospital for Sick Children, Edinburgh, UK
[18] Royal Hospital for Children, Glasgow, UK
[19] Leeds General Infirmary, Leeds, UK
[20] Leicester Royal Infirmary, Leicester, UK
[21] Royal Manchester Children's Hospital, Manchester, UK
[22] Great North Children's Hospital, Newcastle upon Tyne, UK
[23] Queen's Medical Centre, Nottingham, UK
[24] John Radcliffe Hospital, Oxford, UK
[25] Sheffield Children's Hospital, Sheffield, UK
[26] Southampton General Hospital, Southampton, UK
[27] University College London, London, UK
[28] Birmingham Children's Hospital, Birmingham, UK
[29] Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK
[30] St. Anna Children's Cancer Research Institute, Vienna, Austria
[31] Department of Genomic Medicine and the Therapeutics Discovery Division, MD Anderson Cancer Center, Houston, Texas, USA
[32] Haematology and Oncology Department, Great Ormond Street Hospital, UK
[33] Computational Biology Research Centre, Human Technopole, Milan, Italy

[+]Contributed equally, ✉Corresponding authors: andrea.sottoriva@fht.org, louis.chesler@icr.ac.uk;

Corresponding Authors: Dr. Louis Chesler, Division of Clinical Studies, The Institute of Cancer Research, 15 Cotswold Road, Sutton SM2 5NG, UK, E-mail: louis.chesler@icr.ac.uk, Phone: +442034376122; Dr. Andrea Sottoriva, Computational Biology Research Centre, Human Technopole, Viale Rita Levi-Montalcini, 1, 20157 Milan, Italy. E-mail: andrea.sottoriva@fht.org, Phone: +390230247414.

Running Title
Stratified Medicine Paediatrics

Conflict of interest statement
T. Graham is named as a coinventor on patent applications that describe a method for TCR sequencing (GB2305655.9), and a method to measure evolutionary dynamics in cancers using DNA methylation (GB2317139.0). TG has received an honorarium from Genentech. D. Hargrave has Honoraria/ Advisory/ Consultancy roles for: AstraZeneca/MedImmune, Alexion Pharmaceuticals, Bayer, Biodexa, Roche/Genentech, Novartis. Expert Testimony: AstraZeneca, Novartis. Travel, Accommodations, Expenses; Alexion Pharmaceuticals, Boehringer Ingelheim, Novartis, Roche/Genentech, Novartis. S. Gatz: Advisory role for TESARO/GSK, BAYER, Turning Point Therapeutics, EMD Serono/MERCK KGaA, AMGEN, SANOFI, GILEAD, DAICHI/AstraZeneca, Consultancy agreement with AstraZeneca and Schroedinger

Therapeutics. Research funding from GSK and BAYER (institution) and AstraZeneca (institution and personal). J. Anderson holds stock in Autolus ltd and patents in CAR-T technologies. The other authors have no competing interests to declare.

Abstract

We profiled a large heterogenous cohort of matched diagnostic-relapse tumour tissue and paired plasma-derived cell free DNA (cfDNA) from patients with relapsed and progressive solid tumours of childhood. Tissue and cfDNA sequencing results were concordant, with a wider spectrum of mutant alleles and higher degree of intra-tumour heterogeneity captured by the latter, if sufficient circulating tumour-derived DNA (ctDNA) was present. Serial tumour sequencing identified putative drivers of relapse, with alterations in epigenetic drivers being a common feature. In keeping with epigenetic alterations being a common driver of many childhood cancers, fragmentomics analysis of cfDNA identified tumour-specific epigenetic states and transcription factor binding sites accessible in chromatin. This study leverages a large and well-annotated genomic dataset of aggressive childhood malignancies, identifies genomic and epigenetic drivers of childhood cancer relapse, and highlights the power and practicality of cfDNA analysis to capture both intra-tumoural heterogeneity and the epigenetic state of cancer cells.

Statement of Significance

In tumours of childhood we identify mutations in epigenetic genes as drivers of relapse, with matched cfDNA sequencing showing significant intra-tumour genetic heterogeneity and cell-state specific patterns of chromatin accessibility. This highlights the power of cfDNA analysis to identify both genetic and epigenetic drivers of aggressive disease in paediatric cancers.

Introduction

Cancer remains the leading cause of childhood death due to disease in Europe and North America (1,2). Cancers of childhood are often treated with complex, multi-modal treatment regimens including combinations of chemotherapy, radiotherapy and surgery. Unfortunately, despite these intensive therapies, some paediatric solid tumours are either refractory to treatment or relapse rapidly following an initial response (3,4). In recent years, towards developing more effective therapies, large precision medicine programmes have been developed with the goal of identifying targetable genetic alterations in tumour tissue from patients with either treatment refractory or relapsed cancers of childhood (5-9). However, there is limited understanding of how tumours evolve between diagnosis and relapse, and how representative the tumour biopsy at relapse is of the most aggressive subclones of disease.

Stratified Medicine Paediatrics (SMPaeds) is the UK precision medicine programme offering rapid turnaround clinical-grade sequencing with clinical reporting for patients undergoing biopsy at the time of relapse or treatment-refractory disease. A defined aim of SMPaeds was to identify patients likely to respond to molecularly targeted therapeutic agents, and to discover genomic and epigenetic determinants of cancer relapse. Alongside the clinical programme, an archival tissue sample and a temporally matched research plasma sample was obtained.

Here, we report results of the retrospective sequencing of matched diagnostic-relapse tumour tissues and the largest study of paired tumour cell free DNA (cfDNA) sequencing of relapsed childhood cancer patients to date. Across this heterogeneous cohort of poor-outcome tumours of childhood, we identify both common and cancer specific patterns of tumour evolution and clinically significant rare genetic alterations predictive of very poor outcome disease. In certain cancer subtypes, we highlight the utility of cfDNA to identify clinically actionable mutations and tumour heterogeneity at relapse. Finally, given the limited number of driver alterations identified in relapsed paediatric tumours (10), and the increasing recognition that transcription-factor associated epigenetic dysregulation is a prominent driver of therapy resistance (11,12), we establish the utility and practicality of cfDNA nucleosome positioning analysis to identify cancer-specific DNA fragmentation patterns using low cost, clinical grade, low-coverage whole genome sequencing (lcWGS).

Results

Study Design

Patients with clinically relapsed, treatment-refractory or progressive paediatric-type solid tumours (including lymphomas) who undergo a biopsy as part of standard of care were eligible for enrolment into SMPaeds, which recruited patients from February 2019 to January 2024. The results of the main programme, which returned clinical results on tumour tissue and germline sequencing to practitioners in rapid-turnaround at the time of relapse, will be reported separately. The study mandated collection of an archival tumour sample from the time of diagnosis and a plasma sample in parallel to tissue biopsy at SMPaeds enrolment. Here we present the results of paired NGS panel and lcWGS in diagnosis-relapse tumour pairs and matched tumour-cfDNA pairs from the time of relapse (**Figure 1a**). Archival tissue and cfDNA samples were sequenced retrospectively and results were not reported back to the treating clinicians. Matched relapse-tissue and cfDNA sequencing were performed in 401 patients and matched relapse and archival tissue sequencing was performed in 277 patients (**Figure 1b**). The most frequent diagnoses in the overall study and in the subset analysed were central nervous system (CNS) tumours, sarcomas and neuroblastoma (**Figure 1b, Supplemental Table 1**), reflecting a diagnostic distribution typical of this study population.

The genomic landscape of matched primary and relapse paediatric tumour samples

A comparative overview of all genomic alterations found in matched primary-relapse samples is reported in **Figure 2a**. 41% of sequence alterations were present in both the primary and relapse samples, whereas 37% were only found at relapse (23% were found only in the primary). The proportions of patients who acquired genetic alterations at relapse is summarised in **Supplementary Figure S1a-c**, all sequence variants are presented in **Supplemental Table 2** and fusions in **Supplemental Table 3**. In the whole cohort, and as expected, the most frequently altered gene at the time of relapse was *TP53* (**Figure 2b**), with a high frequency of *TP53* mutations seen in patients with sarcoma (**Figure 2c**). In neuroblastoma, the largest single disease type in the study, the most frequently mutated genes that were enriched at relapse were *ALK* and *NF1,* mirroring findings of previous small, paired sequencing studies (13,14) (**Figure 2d**). In addition to the acquisition of *ALK* mutations at relapse, *ALK* was also the most frequently mutated gene at diagnosis, and ALK mutations were only found in neuroblastoma (**Figure 2d),** highlighting the prominent role of this gene in both neuroblastoma oncogenesis and disease progression. Further subgroup-specific analysis can be found in **Supplementary Figure S2a-d**.

Paediatric cancers at relapse show enrichment of mutations in epigenetic modifiers

At relapse, we identified a small but significant increase in the total number of mutations detected by NGS panel sequencing (Wilcoxon one-sided paired signed rank test, p = 0.000646, n= 276) (**Figure 3a**), and the overall fraction of genome altered (FGA) by lcWGS, in relapse genomes (Wilcoxon one-sided paired signed rank test, p = 0.0015, purity>=10%) (**Figure 3b**). Overall, chromosomal rearrangements were relatively similar, with the exception of a few additional copy number variants (CNVs) acquired at relapse. (**Figure 3c-e**). Disease specific analysis also showed a very limited number of copy number events that were enriched at relapse (**Supplementary Figure S3a-e, Supplementary Figure S4a-e,** and **Supplemental Table 4**). Osteosarcomas had the highest FGA, and lymphoma the lowest, followed by CNS tumours (**Supplementary Figure S5a**). There was no significant difference in FGA between diagnosis and relapse in neuroblastomas (**Supplementary Figure S5b**). Overall, our finding indicate that paediatric cancers show relative genomic stability between diagnosis and relapse.

Using NGS panel sequencing data, the ratio of synonymous to non-synonymous mutations (dN/dS) was calculated to identify non-synonymous and truncating variants that are putative drivers of relapse (**Figure 3f**) (15). *TP53* mutations and truncating *CDKN2A* mutations were positively selected at both diagnosis and relapse (**Figure 3g**). Across the cohort, there was a strong positive selection at relapse for missense and truncating mutations in *CHEK2* (**Figure 3g**). Mutations in genes responsible for epigenetic modification of DNA were also highly enriched at relapse including *ATRX, SETD2, ARID1B and CREBBP* (**Figure 3g**), likely consistent with the role of epigenetic remodelling as a driver of therapy resistance in childhood cancers.

Sequencing of high purity ctDNA at the time of relapse better reflects tumour heterogeneity than tissue sequencing

To systematically evaluate the ability of cfDNA sequencing to detect somatic mutations at relapse in paediatric tumours, panel sequencing was performed on all available time matched tissue and cfDNA samples **(Supplementary Figure S6, Supplemental Table 5)**. We used lcWGS when available to estimate purity of plasma cfDNA – defined as the fraction of circulating tumour derived DNA (ctDNA) in total cfDNA. Out of 397 patients with cfDNA lcWGS and panel sequencing available, 86 had high purity (>10%) ctDNA. Patients with neuroblastoma had higher ctDNA levels, compared to all other cancer types, and patients with CNS tumours had lower ctDNA levels (**Figure 4a**). In the cohort as a whole (n=402 patients), 43% of alterations expected from tissue sequencing were detected in cfDNA, with 89% expected alterations detected in high purity samples and 25% expected alterations detected in low purity samples (**Figure 4b**). Of note, although the overlap between tissue and cfDNA sequencing was lower in low purity samples, cfDNA specific variants could still be identified.

The most common cfDNA unique variants were in *ARID1A, NF1, PPM1D TP53* and *CREBBP* (**Figure 4c**). *TP53* and *PPM1D* mutations are commonly reported in clonal haematopoiesis, therefore the haematopoietic

origin of these variants was excluded by inspecting PanelSeq data from the matching blood cell pellet for each patient. Matched tissue was inspected for all cfDNA-unique variants. Of a total of 98 cfDNA-unique variants, only 6 were detected in the time-matched tissue at extremely low, sub-clonal levels (below the VAF limit for reporting for tissue panel sequencing of 5%). The average VAF of the cfDNA-unique variants was significantly lower than the VAF of variants overlapping between tissue and cfDNA (**Figure 4d**), likely reflecting the sub-clonal nature of cfDNA unique variants. The use of Unique Molecular Identifiers (UMIs) in the sequencing library preparation, stringent variant filtering, and further manual curation to only include cancer related, likely pathogenic variants minimise the risk that these cfDNA unique variants could be technical artefacts.

The highest number of cfDNA-unique variants were detected in neuroblastoma, the tumour type with the highest purity ctDNA (**Figure 4e**)**.** In neuroblastoma, a high frequency of the cfDNA unique variants were in epigenetic genes known to be important in neuroblastoma biology (*ARID1A, ATRX, SMARCA4 (16-18)*) and genes associated with ultra-high-risk disease (*ALK, NF1, PTPN11, FGFR1, TP53, ATM, CREBBP (19)).* cfDNA unique variants in *NF1* were detected in 5/72 patients. Taken together with the enrichment of *NF1* mutations that we also identified in tumour sequencing at relapse (**Figure 2d**), this identifies *NF1* as a major potential target in relapsed and treatment resistant neuroblastoma. In comparison, low ctDNA purity and low concordance between cfDNA and tissue sequencing was seen in patients with brain tumours **(Figure 4f).**

Further cancer type specific comparison of cfDNA versus tissue sequencing were limited by patient numbers. cfDNA specific variants were detected in some sarcoma subgroups and some patients with hepatoblastoma, but not in any patients with Wilms' tumour (**Supplementary Figure S6, Supplementary Figure S7a-d**). Although the numbers of patients with lymphoma were limited, cfDNA sequencing appeared to be of particular utility in these patients. For example, in Hodgkin lymphoma, 4 out of 6 patients had cfDNA unique variants detected (**Figure 4g**). This warrants further investigation given the challenges in producing reliable sequencing data from Hodgkin tumour samples, due to the high levels of stroma and paucity of tumour cells.

Overall, these data show a good correlation between variant detection in high purity ctDNA samples and tissue, and strongly supports the added value of cfDNA profiling in detecting tumour heterogeneity at the time of relapse/tumour progression in certain paediatric cancer subtypes.

ctDNA identifies aggressive sub-clones and can predict relapse in neuroblastoma

Genome wide copy number profiles were analysed for tissue and cfDNA pairs from 350 patients. A good concordance between analytes was observed in high purity pairs (n=80) **(Supplementary Figure S8a-c)**. Although the majority of cfDNA samples were of low purity, 56 tissue samples were also of low purity. In 8 of these patients where no CNVs were detected in the tissue, lcWGS was successful in identifying CNVs in plasma. In one patient with neuroblastoma the relapse biopsy taken from the adrenal tumour did not show clear CCND1 amplification, although did show some CNA's indicative of the presence of tumour tissue. However, the copy number profile of the cfDNA at the time of relapse showed a clear CCND1 amplification more in keeping with the time matched biopsy of a liver metastases (**Figure 5a-b**). Given that only the primary tumour results would have been considered in the clinical arm of this study, this demonstrates the added value of matched cfDNA sequencing at the time of relapse to better reflect genomic changes in other sites of disease. Taken together, our data demonstrates the utility of cfDNA to detect neuroblastoma heterogeneity, including detection of genetic changes in metastatic sites.

Next we asked if cfDNA could better detect the mutations that are known to be the most clinically relevant drivers of relapse than tissue biopsy. Here we focused on TP53 and RAS pathway mutations, given data showing they are associated with ultra-high-risk (UHR) neuroblastoma at diagnosis (19), and also further enriched at relapse (13). 36 out of 72 patients with neuroblastoma had at least 1 UHR gene variant detected in either tissue or plasma at relapse (50 variants detected in total). 7 variants were tissue specific and 17 were cfDNA specific. For all 7 tissue specific variants, ctDNA purity was zero, whereas most tissue biopsies that did not detect cfDNA variants were of high purity (**Figure 5c**). Taken together this shows all

UHR variants can be detected by sequencing of tumour derived cell free DNA, whereas sequencing of a single biopsy does not fully capture all UHR mutations present in the tumour at relapse. The majority of mutations in UHR genes detected in ctDNA were in *TP53* and *NF1*. Three patients had *FGFR1* hotspot mutations detected at relapse, all of which were also detected in cfDNA. **(Figure 5d**).

Two patients with *ALK* mutant neuroblastoma were serially monitored by cfDNA analysis whilst receiving the ALK inhibitor lorlatinib. One patient, with a pathogenic *ALK* F1174L mutation detected in tissue and cfDNA at relapse had an almost complete response to lorlatinib and concordantly the ALK mutant fraction was undetectable after 1 month of therapy (also as seen in (20)). The same pathogenic *ALK* mutation re-merged at low levels (0.4%-2.7% VAF) and remained detectable for a year before the clinical relapse was detected by magnetic resonance imagining (**Figure 5e**). The second patient with metastatic neuroblastoma was commenced on lorlatinib based on the presence of a pathogenic *ALK* F1174L mutation in cfDNA. A partial response to lorlatinib was mirrored by a reduction in mutant ALK levels (**Figure 5f**).

Fragmentomics analysis of cfDNA reveals cancer and tissue-specific signatures

Given that overall mutational frequency in childhood cancers is low, we asked if ctDNA analysis could be used to infer additional epigenetic characteristics relating to tissue of origin and cancer-specific expression signatures. For this purpose, we focused on ctDNA fragmentation patterns derived from analysis of lcWGS.

DNA fragments derived from open chromatin are more sensitive to degradation than nucleosome bound ctDNA fragments, resulting in differences in ctDNA sequencing coverage between open and closed chromatin (21) (**Figure 6a**). To identify fragmentomic patterns in paediatric tumour derived ctDNA we applied the Griffin methodology combined with the GTRD database containing consensus binding sites for 425 transcription factors with loss of coverage relating to assumed sites of transcription factor binding (21). We used Wilcoxon tests and hierarchical clustering to identify transcription factor binding sites that were differentially accessible in one tumour type compared with others. This approach clearly identified cancer specific transcription factor expression signature clusters (**Figure 6b, Supplemental Table 6, Supplementary Figure S9a-d**). Similarity of the TFBSs for TFs within clusters was not explanatory (**Supplementary Figure S10**). Gene ontology enrichment analysis of the 4 main clusters identified gene sets relating to cell of origin including: neural crest derived pathways in neuroblastoma, muscle development in rhabdomyosarcoma, immune pathways in lymphoma and hepatocyte differentiation in the hepatoblastoma cluster (**Figure 6c**).

The largest cluster, associated with non-Hodgkins lymphoma (NHL), contained a large number of transcription factors known to play a role in NHL pathogenesis such as BCL6, IRF4, BATF3, FOXO1 and BCL11A (22-25) (**Figure 6c-e**). In patients with Wilms tumour, the binding sites of only one transcription factor were identified as highly specific and highly accessible: SIX2 - a known poor prognostic biomarker and stem-cell/blastemal subtype associated transcription factor (26,27) (**Figure 6f**). Two out of the three transcription factors known to be hepatocyte specific (28) were identified in the hepatocyte cluster (**Figure 6g, Supplementary Figure S11a-b**).

For the embryonal tumours neuroblastoma and rhabdomyosarcoma, core regulatory circuits (CRC's) related to developmentally derived cell-states have been described (29-31). We therefore evaluated known transcription factors that are activated as part of these defined CRCs. The pan-rhabdomyosarcoma CRC constituents MYOD1 and MYOG (32,33) showed significant loss of coverage in rhabdomyosarcoma samples (**Figure 6h-i**). In neuroblastoma samples, we identified accessibility of the noradrenergic CRC, and of its key constituent transcription factors HAND2, ISL1, TBX2, GATA3 and MYCN (30,34,35) (**Figure 6j-l**). Some of these transcription factor binding sites, such as HAND2, TBX2 and ISL1 were highly specific for neuroblastoma (**Figure 6j, Supplementary Figure S11c-d**), whereas others such as MYCN, although accessible in neuroblastoma, were not specific to neuroblastoma samples only (**Figure 6k**). This is consistent with the major role of MYCN in neuroblastoma oncogenesis, alongside the fact that dysregulation of MYCN has also been associated with other aggressive paediatric malignancies (36). We did not identify differential accessibility for Ewing sarcoma specific CRC transcription factors or the EWS:FLI fusion (**Supplementary Figure S12a-c**).

Taken together, we show that fragmentomics analysis of ctDNA identifies distinct tumour specific clusters related to cell of origin and epigenetic plasticity, with potential application in identifying transcription signatures associated with treatment resistance and relapse.

Discussion

To our knowledge this is the largest pan-paediatric cancer analysis to date of tumour-cfDNA and diagnosis-relapse pairs. Our diagnosis-relapse tumour comparison identifies putative drivers of relapse across the whole cohort. These are predominantly mutations in genes involved in epigenetic regulation in addition to the DNA damage response gene *CHEK2*. Pan-cancer analyses have identified that *CHEK2* mutations occur only at very low frequencies (<2%) in individual poor-outcome paediatric tumours such as medulloblastoma, osteosarcoma, Ewing Sarcoma, rhabdomyosarcoma, and neuroblastoma (37). However, we identify a strong selection for both missense and truncating mutations in *CHEK2* between diagnosis and relapse across our cohort. *CHEK2* mutations have been associated with standard therapy resistance in multiple adult cancers (38-40) but may also associate with sensitivity to novel targeted agents (41). Taken together with our data, this identifies *CHEK2* as a likely driver of therapy resistance in paediatric cancers and also identifies *CHEK2* as an ideal target for arms of tumour agnostic paediatric 'basket' early phase clinical trials such as ESMART (5) and NCI-MATCH (8).

In neuroblastoma, the largest single tumour entity in our cohort we identified an enrichment of both *ALK* and *NF1* mutations at the time of relapse. The enrichment in *ALK* mutations at relapse has previously been described (13,14). Sequencing of primary-relapse neuroblastoma pairs has also identified the acquisition of NF1 mutations in 2/23 relapses (13). This is concordant with our data, identifying relapse specific mutations in 4/59 neuroblastomas. In addition to this, we also identify a further 5 cfDNA unique *NF1* mutations at the time of relapse. Together this identifies clonal selection for *NF1* mutations as a potential driver of therapy resistance in neuroblastoma.

Our paired diagnosis-relapse tumour comparison was performed using an NGS panel which has been designed to detect mutations that are known to be of importance in paediatric cancers (42). This approach was chosen to maximise the chances of successful sequencing in archival diagnostic samples obtained as part of routine clinical care. This is a limitation of our study, where the lack of genome-wide sequencing limits the potential for novel discovery. However, given that our panel is already curated to identify paediatric cancer driver genes, this also increases the probability that the genes identified as being enriched at relapse are likely to be of high biological relevance. Within this context, is is important to note that only a limited number of tumours acquired additional driver mutations at the time of relapse.

Using a high-depth cfDNA panel sequencing approach, we show that we are able to detect tumour mutations in the majority of patients, if enough tumour derived DNA is present in the circulation. In patients with brain tumours where ctDNA levels are limited, there is extremely limited utility of cfDNA analysis from plasma for the detection of tumour associated variants. Here, cfDNA analysis from cerebrospinal fluid is likely to be more informative (43). Conversely, in patients with neuroblastoma there are high amounts of ctDNA and a high concordance between tumour tissue and cfDNA sequencing. In addition, we see a high number of cfDNA unique variants in patients with neuroblastoma demonstrating the heterogeneity of this disease and limitations of performing a single biopsy at relapse. For patients with *ALK* mutant neuroblastoma who receive *ALK* inhibitors our limited data on serial monitoring adds to the increasing body of evidence to support prospective evaluation of cfDNA sequencing as a biomarker of therapy response (44-46).

Further subgroup comparisons are limited by sample size, due to the heterogenous nature of our study population. However, we identify a number of paediatric tumours where cfDNA sequencing for somatic variant detection shows promise, adding to the literature supporting systematic evaluation within disease-specific clinical trials (5,20,47).

Taken together, our data shows that plasma based cfDNA sequencing can provide additional information to tumour sequencing at the time of relapse in extracranial tumours of childhood. This data is supportive

of cfDNA sequencing as a complimentary technique to tissue-based sequencing at the time of relapse to detect tumour associated mutations. However, given the rarity of paediatric cancers, further large-scale analyses will be required to accurately define specificity and sensitivity of cfDNA based approaches in individual tumour types.

Despite the added value of cfDNA sequencing demonstrated in this study, it is important to note that for a significant proportion of patients, no genetic alterations were detected in either tumour tissue or ctDNA at the time of relapse. This is in keeping with other studies (5) and highlights the importance of identifying non-genetic biomarkers of aggressive disease in paediatric tumours. Within this context, the development of clinical grade technologies to accurately profile both epigenetic states and the tumour-immune microenvironment is a defined goal of SMPaeds2. We note also that it is striking how primary and relapse samples are so homogeneous, and no new genomic alteration is strongly enriched at relapse due to treatment. One would have predicted genetic alterations in recurrent driver genes to be the driving force of Darwinian evolution during cancer progression. However, this is not the case in this cohort, and indeed drivers of chemotherapy resistance have not been robustly documented in adult solid tumours so far. Relapse may be driven by a large variety of genetic mechanisms, potentially even multiple ones in the same tumour, with each one being relatively rare and hence difficult to detect without an emormous cohort. A much more parsimonious explanation instead, is that epigenetic alterations may be responsible for relapse. We argue that our study strongly suggest that the pursue of genetic alterations explaining relapse will likely be unfruitful, and that we should concentrate on prospective collection of samples that will allow the epigenetic profiling of matched primary and relapse cancers.

Importantly, here we show that the utility of cfDNA sequencing extends beyond the detection of genetic alterations, and that epigenetic states relating to the developmental origins of paediatric cancers can also be identified by lcWGS of cfDNA. Epigenetic dysregulation is increasingly being recognised as the predominant driver of aggressive childhood cancers (48) and the transcriptional circuits that drive aggressive disease also show specific dependencies that are of high interest for the development of novel therapeutics (35,49). Therefore, new tools are urgently needed that can effectively characterise the epigenetic landscape of childhood tumours. Nucleosome profiling of cfDNA not only has the potential to identify disease subtype in paediatric cancers, but also to identify epigenetic changes associated with aggressive biology. Furthermore, this information is obtained from lcWGS of cfDNA which is minimally invasive, relatively low cost and has the potential to be applied serially and at scale. This offers multiple potential applications including supporting diagnosis, providing prognostic information, identifying epigenetic targets for novel therapeutics and the potential for serial sampling to understand how specific therapies alter epigenetic programming. Furthermore, this approach has potential to be integrated with other data types such as methylation analysis to aid more precise tumour subgroup classification. Further refinement of our methodology, and independent validation in large cohorts will be needed to realise the potential of this tool for paediatric cancers.

In conclusion, we show specific patterns of evolution between diagnosis and relapse in paediatric tumours, including the identification of both tumour specific and pan-cohort drivers of relapse. In addition, we show that cfDNA analysis has potential to identify genetic drivers in aggressive tumours of childhood. Finally, we show that in patients with relapsed and refractory cancers, standard clinical grade lcWGS of cfDNA can be used to infer the epigenetic state of cancer cells, relating to the cell of origin.

Methods

Study design and patient recruitment

The SMPaeds study (IRAS: 246557/264925, doi: 10.1186/ISRCTN21731605) received ethical approval from the UK National Research Ethics Service. It opened in 2019 and recruited patients from 20 paediatric oncology centres across England, Scotland, Wales and Northern Ireland. Children, teenagers and young adults with paediatric type tumours were eligible to enrol at the time of relapse, refractory or progressive disease. Tumour biopsy was performed as part of standard of care and submitted for sequencing.

Alongside this a blood sample was collected at the time of relapse. Blood cell pellet was used for germline sequencing and plasma stored for retrospective cfDNA analysis. Archival tumour tissue from the time of diagnosis was requested from the pathologist at the referring centre. Serial cfDNA sample analysis was performed as part of our local ethically approved protocol (IRAS: 281408).

Sample preparation and sequencing

Either frozen or formalin fixed paraffin embedded (FFPE) tumour samples were accepted in this study. Tumour biopsy specimens were assessed by the study pathologist and analysed for tumour content. Tumour tissue DNA/RNA and buffy coat germline DNA extractions were performed at Great Ormond Street Hospital on Promega Maxwell machines. FFPE tissue was extracted with the Maxwell RSC FFPE Plus DNA Kit (AS1720) and the Maxwell 16 LEV RNA FFPE Kit. Whole blood was collected into cell-free DNA blood collection tubes (Streck). Frozen tissue and buffy-coat samples were extracted with the Maxwell RSC Blood DNA Kit (AS1400) and the Maxwell 16 LEV SimplyRNA Cells Kit.
The Royal Marsden Hospital lab separated plasma from blood within 7 days of venepuncture. Separation was done by double centrifugation: blood was spun at 1600g for 10 min, the plasma fraction was collected and spun again at 1600g for 10 min, aliquoted into cryovials and stored at -80°C. cfDNA was extracted from plasma using QIAamp Circulating Nucleic Acid Kit (55114, Qiagen) and QIAvac 24 Plus Vacuum Manifold system. In cases where the kit was not available, QIAsymphony Circulating DNA Kit (1091063, Qiagen) on the QIAsymphony SP automated extraction system was used instead. Samples with cfDNA above 10ng/µl were analysed using Agilent TapeStation (RRID:SCR_018435) with the genomic DNA (5067-5365) or cfDNA ScreenTape assay (5067-5630) to evaluate the presence and fraction of high molecular weight DNA.
Targeted sequencing panels were constructed to capture driver mutations in paediatric solid tumours (Supplementary Tables S7-9). The paediatric specific panel of 92 genes (42) or a solid tumour panel of 233 genes (Royal Marsden Hospital, UK) was used to construct libraries for tissue samples depending on availability. A marginally smaller paediatric panel of 67 genes (20) was run for cfDNA.
DNA from tissue and buffy-coat were processed by previously established methods (42) using the KAPA HyperPlus Kit and SeqCap EZ adapters (Roche, NimbleGen, Madison WI, USA) with dual-SPRI size selection of the libraries (250-450 bp). 200ng of DNA (samples yielding DNA with median fragment length > 1000 bp) or 400ng of DNA if there was sufficient DNA (for samples with DNA < 1000 bp) was used for library preparation. 1 µg of the pooled library DNA was hybridised to custom panels as described and sequenced to 100X coverage (germline) and 400X coverage (tumour) on Nextseq (RRID:SCR_016381) (100bp paired end reads, mid-output kit or high-output kit depending on sample number). cfDNA sequencing libraries were generated using Cell3™Target (Nonacus) library preparation kit with Dual Index Unique Molecular Identifier (UMI) adapters as per manufacturer protocol (input cfDNA 5-250ng). cfDNA was sequenced using NovaSeq6000 (RRID:SCR_016387) (200 cycle kit, 2x100bp) at ~1500xUMI coverage. For tissue from relapse tumours only, RNA fusion panel sequencing libraries were generated using the TruSight RNA Pan-Cancer Panel (Illumina, San Diego, CA, USA) and sequenced on Nextseq (2x75bp).
For low coverage whole genome sequencing (lcWGS), the pre-capture amplified libraries prepared for panel capture were pooled and sequenced separately on the NextSeq (2x75bp) on a different flow cell at a coverage between 0.5-4X. cfDNA lcWGS were ran on the NovaSeq (200 cycle kit, 100bp paired end reads). There were fewer samples with material available for lcWGS than PanelSeq as PanelSeq was prioritised, all samples used are detailed in Supplementary Tables S10-11.

Panel Sequencing analysis

The initial processing of all NGS data was performed using the Molecular Diagnostics Information Management System v4.0 (50), based on genome build GRCh37, at Clinical Genomics, The Royal Marsden Hospital. In summary raw data was converted to FASTQ files using bcl2fastq2 (RRID:SCR_015058), FASTQs were then aligned to the GRCh37 reference genome using BWA-MEM (RRID:SCR_022192) and quality

metrics were generated using Picard Tools (RRID:SCR_006525). Somatic nucleotide variants for PanelSeq of tumour tissue collected at the time of relapse were called using GATK Mutect2 (51) (RRID:SCR_001876) and were annotated using PCGR (52). Germline NGS panel sequencing results were reported back to the molecular tumour board as part of the main clinical sequencing programme, however for the purposes of this retrospective study, matched germline samples were only used for the purpose of calling somatic variants.

The custom pipeline for the analysis of cfDNA PanelSeq with UMIs was previously described in (43). Following demultiplexing including UMI extraction, fastqs were converted to an unmapped BAM file with annotated UMI sequences with fgbio (https://github.com/fulcrumgenomics/fgbio). Reads are then aligned to the GRCh37 genome with BWA-MEM and grouped by template and UMI to create consensus reads which arise from the same molecule. Reads were filtered using fgbio FilterConsensusReads with the following options: –min-reads 3 –max-read-error-rate 0.05 –max-base-error-rate 0.1 –min-base-quality 30 –max-no-call-fraction 0.1. Filtered reads are once again aligned to the genome with BWA-MEM followed by clipping overlapping reads with fgbio ClipBam. Variant calling was performed with VarDict (RRID:SCR_023658) with the minimum frequency set to 0.0001 and variants were filtered using an in-house blacklist and whitelist. The results of relapse tissue and cfDNA PanelSeq were manually inspected using IGV (RRID:SCR_011793) and reported by the Clinical Genomics team at the Royal Marsden Hospital.

To ensure the greatest sensitivity for longitudinal comparisons, somatic variants occurring at the time of diagnosis were called jointly with relapse tissue and cfDNA using GATK Mutect2 v4.2.3.0 in multi-sample mode. Following calling, variants were annotated using vep101.0 (RRID:SCR_007931) and GATK FilterMutectCalls v4.2.3.0 was used to mark variants for downstream filtering. Variants were imported into R with VcfR v1.13.0 (RRID:SCR_023453) (53) and filtered with R scripts as described. Blacklisted, synonymous and non-coding variants were removed, except for upstream TERT mutations and whitelisted variants. Variants required “MODERATE” or “HIGH” severity, VAF >=5%, AD >=5, DP >=10 and frequency <0.0002 in Gnomad. Ultimately variants that had not previously been reported in relapse were manually inspected, and likely artefacts were removed. FilterMutectCalls flags were used as follows: “normal_artifact” removed if germline allele fraction was greater than 15% of the allele fraction in the tumour; “slippage” removed if AD < 20 and VAF < 0.3%; “strand_bias” up to AD=10, variants with at least 1 read on the opposing strand were kept, for AD > 10 variants were removed if the count difference between strands exceeded 150%. Variants are displayed using the oncoprint function from ComplexHeatmap v2.6.2 (RRID:SCR_017270) (54) and ggplot2 v2.3.4.4 (RRID:SCR_014601) (55). Clinically defined hypermutator SMP0342 was excluded from comparative analysis. For all comparative analyses, genes that were not covered by the panel for all samples were excluded.

RNA fusion panel sequencing was analysed using the BaseSpace Sequencing hub from Illumina (RRID:SCR_011881) using the RNA-Seq Alignment App v2.0.1 with STAR to hg19 and Manta for fusion calling with default parameters. Results were inspected and reported by the Clinical Genomics team at the Royal Marsden Hospital.

Copy number analysis

BAM files were processed as described in the previous section, ichorCNA v0.3.2 (RRID:SCR_024768) was then used to generate normalised relative coverage (log2 ratio = log2 (Tumor / Normal)) in 500kb bins. Data normalisation was performed in accordance with the ichorCNA workflow, using an in-house panel-of-normals, including sex chromosomes. Bins were required to have a minimum mappability of 85 and 95% non-N bases. The log2 ratio was further normalised by subtracting the median log2 ratio of all bins. A selection of bins were removed as they were seen to be recurrently aberrant.

To call absolute copy number, we used a modified version of the ASCATlp script (https://github.com/cresswell-lab/ASCATlp) to fit a range of purities from 0.1 to 1 and ploidies 1.5 to 4.5. Seeking to leverage multiple sampling for optimised segmentation we used the multipcf function from copynumber v1.30.0 (56). For patients with only a single sample, the singular pcf function was used. For both functions a penalty of 20 was used. ASCATlp is based on the approach of the ASCAT tool

(RRID:SCR_016868) (57), in short it calculates the continuous copy number state and the sum of squared differences to the nearest integer for each combination of parameters. A 3x3 search is applied to identify local minima, the solution with the lowest sum of squares is selected. For each sample, the pipeline was run multiple times constraining the search space to find the most feasible solution. For samples in which no fit was produced for the patient ploidy, the purity was set to zero.
The estimated purity value of both tumour and ctDNA content was taken from the lcWGS analysis. Copy number events were defined as follows: Deletion= 0 copies; Loss =1 ; Amplification >= ploidy +2; Gain= ploidy +1. Runs of matching copy number states were extracted from the segments and split by chromosome arm. Chromosomal copy number change was called for runs longer than 75% of the chromosome. Arms were called if the run was under 75% of the chromosome and greater than 75% of the arm. Focal CNVs (genes or chromosome bands) were called when they were shorter than half the length of the chromosome arm. The fraction of genome altered (FGA) was calculated as the fraction of bins with an estimated copy number differing from the median ploidy. Comparative CNV analyses (**Figure 3e and S4**), were performed by subtracting the copy number state of segments at diagnosis from the copy number state at relapse. Amplifications were capped at ploidy=2+ to avoid over-reporting relapse-acquired amplications (Figure 3e, S4a-d). MEDICC2 v1.0.2 (58) was used to generate copy number phylogenies.

dNdS analysis

dNdS analysis was performed using dNdScv v.0.0.1.0 (RRID:SCR_023123) (59) for 278 primary/relapse pairs. dNdS was considered significantly greater than 1 (neutral) when the lower bound of the 95% confidence interval was greater than 1. SMP0342 and SMP0448 were excluded from dNdS analysis as they had more than 3 times as many synonymous mutations as any other sample (272 and 284 synonymous mutations respectively).

Fragmentomics analysis
Transcription factor binding sites were obtained following published methods from the griffin framework (21). The Gene Transcription Regulation Database (GTRD) database v21.12 (60) was filtered for bona fide transcription factors in the CIS-BP database v2.00 (61). The database was then filtered for TFs with at least 10,000 binding sites, leaving 425 TFs. For each TF, the top 10,000 sites with the greatest confidence were picked. BAM files were converted to GRCh38 using UCSC liftOver (RRID:SCR_018160). The Griffin v0.2.0 (21) package was used to retrieve GC content corrected composite coverage for all 425 TFs in 429 cfDNA lcWGS samples. Samples were filtered for ctDNA content >0.2 for further analysis (n=85). For disease groups with more than 3 members, two-sided Wilcoxon rank sum (Mann-Whitney) tests were performed to compare the composite coverage at the centre of TFBSs within a disease with all other samples. TFs with a p-value < 0.05 (n=207) in any comparison, were selected for hierarchical clustering. Clustering was performed using the Heatmap function in ComplexHeatmap v2.6.2 (54). To inspect potential bias in clustering due the similarity of binding sites between TFs, the jaccard coefficient was computed across TFBSs to create a similarity matrix. Gene interaction network and gene ontology enrichment analyses were performed with StringDB (62) for clusters of TFs showing disease specificity.

Data availability Statement
All sequencing data used in this study have been made available on the European Genome-Phenome Archive, accession number: EGAS50000000549. Processed data have been made available on Mendeley (https://doi.org/10.17632/d2gynzkj42.1).

Acknowledgements
This study was funded by Cancer Research UK (CRUK) (A24566), Children with Cancer UK (CWCUK) (17-235). We also received funding from Aoife's Bubbles, Abbie's Fund, and Christopher's Smile charities. We have received support from national networks including the Experimental Cancer Medicine Centre (ECMC) and the VIVO Biobank. We would like to thank Regan Barfoot for her valuable support, and acknowledge

the German INFORM programme for their assistance during initial setup. We thank all staff involved from the University of Birmingham Clinical Trials Unit and all external expert representatives on the study steering committee and the committee lead Professor John Chester. A.S. was supported by the Wellcome Trust (202778/B/16/Z) and Cancer Research UK (A22909). This work was supported by the AIRC/CRUK Accelerator Award (A26815). C.M.S. acknowledges support and funding from the Marie Skłodowska-Curie Actions grant (101106070). We thank all those involved at local sites including oncologists, pathologists, nursing staff and support staff. Thank you to our Patient-Public Involvement (PPI) representatives Sally Hall of Blue Skye Thinking, John Rainsbury and Little Hero for their inspirational and passionate support, and finally but most importantly all patients and parents for their participation.

References

1. Goldstick JE, Cunningham RM, Carter PM. Current Causes of Death in Children and Adolescents in the United States. N Engl J Med **2022**;386(20):1955-6 doi 10.1056/NEJMc2201761.
2. Kyu HH, Stein CE, Boschi Pinto C, Rakovac I, Weber MW, Dannemann Purnat T*, et al.* Causes of death among children aged 5-14 years in the WHO European Region: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Child Adolesc Health **2018**;2(5):321-37 doi 10.1016/S2352-4642(18)30095-6.
3. Perkins SM, Shinohara ET, DeWees T, Frangoul H. Outcome for children with metastatic solid tumors over the last four decades. PLoS One **2014**;9(7):e100396 doi 10.1371/journal.pone.0100396.
4. Ceschel S, Casotto V, Valsecchi MG, Tamaro P, Jankovic M, Hanau G*, et al.* Survival after relapse in children with solid tumors: a follow-up study from the Italian off-therapy registry. Pediatr Blood Cancer **2006**;47(5):560-6 doi 10.1002/pbc.20726.
5. Berlanga P, Pierron G, Lacroix L, Chicard M, Adam de Beaumais T, Marchais A*, et al.* The European MAPPYACTS Trial: Precision Medicine Program in Pediatric and Adolescent Patients with Recurrent Malignancies. Cancer Discov **2022**;12(5):1266-81 doi 10.1158/2159-8290.CD-21-1136.
6. van Tilburg CM, Pfaff E, Pajtler KW, Langenberg KPS, Fiesel P, Jones BC*, et al.* The Pediatric Precision Oncology INFORM Registry: Clinical Outcome and Benefit for Patients with Very High-Evidence Targets. Cancer Discov **2021**;11(11):2764-79 doi 10.1158/2159-8290.CD-21-0094.
7. Wong M, Mayoh C, Lau LMS, Khuong-Quang DA, Pinese M, Kumar A*, et al.* Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. Nat Med **2020**;26(11):1742-53 doi 10.1038/s41591-020-1072-4.
8. Parsons DW, Janeway KA, Patton DR, Winter CL, Coffey B, Williams PM*, et al.* Actionable Tumor Alterations and Treatment Protocol Enrollment of Pediatric and Young Adult Patients With Refractory Cancers in the National Cancer Institute-Children's Oncology Group Pediatric MATCH Trial. J Clin Oncol **2022**;40(20):2224-34 doi 10.1200/JCO.21.02838.
9. Villani A, Davidson S, Kanwar N, Lo WW, Li Y, Cohen-Gogo S*, et al.* The clinical utility of integrative genomics in childhood cancer extends beyond targetable mutations. Nat Cancer **2023**;4(2):203-21 doi 10.1038/s43018-022-00474-y.
10. Grobner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA*, et al.* The landscape of genomic alterations across childhood cancers. Nature **2018**;555(7696):321-7 doi 10.1038/nature25480.
11. Megiorni F. Epigenetics in rhabdomyosarcoma: cues to new biomarkers and targeted therapies. EBioMedicine **2020**;52:102673 doi 10.1016/j.ebiom.2020.102673.
12. Shendy NAM, Zimmerman MW, Abraham BJ, Durbin AD. Intrinsic transcriptional heterogeneity in neuroblastoma guides mechanistic and therapeutic insights. Cell Rep Med **2022**;3(5):100632 doi 10.1016/j.xcrm.2022.100632.
13. Eleveld TF, Oldridge DA, Bernard V, Koster J, Colmet Daage L, Diskin SJ*, et al.* Relapsed neuroblastomas show frequent RAS-MAPK pathway mutations. Nat Genet **2015**;47(8):864-71 doi 10.1038/ng.3333.

14. Rosswog C, Fassunke J, Ernst A, Schomig-Markiefka B, Merkelbach-Bruse S, Bartenhagen C*, et al.* Genomic ALK alterations in primary and relapsed neuroblastoma. Br J Cancer **2023**;128(8):1559-71 doi 10.1038/s41416-023-02208-y.
15. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S*, et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science **2015**;348(6237):880-6 doi 10.1126/science.aaa6806.
16. Sausen M, Leary RJ, Jones S, Wu J, Reynolds CP, Liu X*, et al.* Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. Nat Genet **2013**;45(1):12-7 doi 10.1038/ng.2493.
17. Bellini A, Bessoltane-Bentahar N, Bhalshankar J, Clement N, Raynal V, Baulande S*, et al.* Study of chromatin remodeling genes implicates SMARCA4 as a putative player in oncogenesis in neuroblastoma. Int J Cancer **2019**;145(10):2781-91 doi 10.1002/ijc.32361.
18. van Gerven MR, Bozsaky E, Matser YAH, Vosseberg J, Taschner-Mandl S, Koster J*, et al.* Mutational spectrum of ATRX aberrations in neuroblastoma and associated patient and tumor characteristics. Cancer Sci **2022**;113(6):2167-78 doi 10.1111/cas.15363.
19. Ackermann S, Cartolano M, Hero B, Welte A, Kahlert Y, Roderwieser A*, et al.* A mechanistic classification of clinical phenotypes in neuroblastoma. Science **2018**;362(6419):1165-70 doi 10.1126/science.aat6768.
20. Stankunaite R, George SL, Gallagher L, Jamal S, Shaikh R, Yuan L*, et al.* Circulating tumour DNA sequencing to determine therapeutic response and identify tumour heterogeneity in patients with paediatric solid tumours. Eur J Cancer **2022**;162:209-20 doi 10.1016/j.ejca.2021.09.042.
21. Doebley AL, Ko M, Liao H, Cruikshank AE, Santos K, Kikawa C*, et al.* A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. Nat Commun **2022**;13(1):7475 doi 10.1038/s41467-022-35076-w.
22. Salaverria I, Philipp C, Oschlies I, Kohler CW, Kreuz M, Szczepanowski M*, et al.* Translocations activating IRF4 identify a subtype of germinal center-derived B-cell lymphoma affecting predominantly children and young adults. Blood **2011**;118(1):139-47 doi 10.1182/blood-2011-01-330795.
23. Schleussner N, Merkel O, Costanza M, Liang HC, Hummel F, Romagnani C*, et al.* The AP-1-BATF and -BATF3 module is essential for growth, survival and TH17/ILC3 skewing of anaplastic large cell lymphoma. Leukemia **2018**;32(9):1994-2007 doi 10.1038/s41375-018-0045-9.
24. Roberto MP, Varano G, Vinas-Castells R, Holmes AB, Kumar R, Pasqualucci L*, et al.* Mutations in the transcription factor FOXO1 mimic positive selection signals to promote germinal center B cell expansion and lymphomagenesis. Immunity **2021**;54(8):1807-24 e14 doi 10.1016/j.immuni.2021.07.009.
25. Satterwhite E, Sonoki T, Willis TG, Harder L, Nowak R, Arriola EL*, et al.* The BCL11 gene family: involvement of BCL11A in lymphoid malignancies. Blood **2001**;98(12):3413-20 doi 10.1182/blood.v98.12.3413.
26. Song D, Yue L, Wu G, Ma S, Guo L, Yang H*, et al.* Assessment of promoter methylation and expression of SIX2 as a diagnostic and prognostic biomarker in Wilms' tumor. Tumour Biol **2015**;36(10):7591-8 doi 10.1007/s13277-015-3456-5.
27. Petrosyan A, Villani V, Aguiari P, Thornton ME, Wang Y, Rajewski A*, et al.* Identification and Characterization of the Wilms Tumor Cancer Stem Cell. Adv Sci (Weinh) **2023**;10(20):e2206787 doi 10.1002/advs.202206787.
28. Huang H, Wu L, Lu L, Zhang Z, Qiu B, Mo J*, et al.* Single-cell transcriptomics uncovers cellular architecture and developmental trajectories in hepatoblastoma. Hepatology **2023**;77(6):1911-28 doi 10.1002/hep.32775.
29. Danielli SGW, Y.; Dyer, M.A.; Stewart, E.;Sheppard, H.; Wachtel, M.; Schäfer B.W.; Patel, A.G.; Langenau, D.M. Single cell transcriptomic profiling identifies tumor-acquired and therapyresistant cell states in pediatric rhabdomyosarcoma. Nat Commun **2024**;15(1):6307 doi 10.1038/s41467-024-50527-2.

30. Boeva V, Louis-Brennetot C, Peltier A, Durand S, Pierre-Eugene C, Raynal V*, et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. Nat Genet **2017**;49(9):1408-13 doi 10.1038/ng.3921.
31. van Groningen T, Koster J, Valentijn LJ, Zwijnenburg DA, Akogul N, Hasselt NE*, et al.* Neuroblastoma is composed of two super-enhancer-associated differentiation states. Nat Genet **2017**;49(8):1261-6 doi 10.1038/ng.3899.
32. Gryder BE, Pomella S, Sayers C, Wu XS, Song Y, Chiarella AM*, et al.* Histone hyperacetylation disrupts core gene regulatory architecture in rhabdomyosarcoma. Nat Genet **2019**;51(12):1714-22 doi 10.1038/s41588-019-0534-4.
33. Yohe ME, Gryder BE, Shern JF, Song YK, Chou HC, Sindiri S*, et al.* MEK inhibition induces MYOG and remodels super-enhancers in RAS-driven rhabdomyosarcoma. Sci Transl Med **2018**;10(448):eaan4470 doi 10.1126/scitranslmed.aan4470.
34. Wang L, Tan TK, Durbin AD, Zimmerman MW, Abraham BJ, Tan SH*, et al.* ASCL1 is a MYCN- and LMO1-dependent member of the adrenergic neuroblastoma core regulatory circuitry. Nat Commun **2019**;10(1):5622 doi 10.1038/s41467-019-13515-5.
35. Oldridge DA, Wood AC, Weichert-Leahey N, Crimmins I, Sussman R, Winter C*, et al.* Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. Nature **2015**;528(7582):418-21 doi 10.1038/nature15540.
36. Liu Z, Chen SS, Clarke S, Veschi V, Thiele CJ. Targeting MYCN in Pediatric and Adult Cancers. Front Oncol **2020**;10:623679 doi 10.3389/fonc.2020.623679.
37. Consortium APG. AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer Discov **2017**;7(8):818-31 doi 10.1158/2159-8290.CD-17-0151.
38. Chrisanthar R, Knappskog S, Lokkevik E, Anker G, Ostenstad B, Lundgren S*, et al.* CHEK2 mutations affecting kinase activity together with mutations in TP53 indicate a functional pathway associated with resistance to epirubicin in primary breast cancer. PLoS One **2008**;3(8):e3062 doi 10.1371/journal.pone.0003062.
39. Hansen EB, Karlsson Q, Merson S, Wakerell S, Rageevakumar R, Jensen JB*, et al.* Impact of germline DNA repair gene variants on prognosis and treatment of men with advanced prostate cancer. Sci Rep **2023**;13(1):19135 doi 10.1038/s41598-023-46323-5.
40. Ow GS, Ivshina AV, Fuentes G, Kuznetsov VA. Identification of two poorly prognosed ovarian carcinoma subtypes associated with CHEK2 germ-line mutation and non-CHEK2 somatic mutation gene signatures. Cell Cycle **2014**;13(14):2262-80 doi 10.4161/cc.29271.
41. Tsujino T, Takai T, Hinohara K, Gui F, Tsutsumi T, Bai X*, et al.* CRISPR screens reveal genetic determinants of PARP inhibitor sensitivity and resistance in prostate cancer. Nat Commun **2023**;14(1):252 doi 10.1038/s41467-023-35880-y.
42. George SL, Izquierdo E, Campbell J, Koutroumanidou E, Proszek P, Jamal S*, et al.* A tailored molecular profiling programme for children with cancer to identify clinically actionable genetic alterations. Eur J Cancer **2019**;121:224-35 doi 10.1016/j.ejca.2019.07.027.
43. Stankunaite R, Marshall LV, Carceller F, Chesler L, Hubank M, George SL. Liquid biopsy for children with central nervous system tumours: Clinical integration and technical considerations. Front Pediatr **2022**;10:957944 doi 10.3389/fped.2022.957944.
44. Berko ER, Witek GM, Matkar S, Petrova ZO, Wu MA, Smith CM*, et al.* Circulating tumor DNA reveals mechanisms of lorlatinib resistance in patients with relapsed/refractory ALK-driven neuroblastoma. Nat Commun **2023**;14(1):2601 doi 10.1038/s41467-023-38195-0.
45. Berlak M, Tucker E, Dorel M, Winkler A, McGearey A, Rodriguez-Fos E*, et al.* Mutations in ALK signaling pathways conferring resistance to ALK inhibitor treatment lead to collateral vulnerabilities in neuroblastoma cells. Mol Cancer **2022**;21(1):126 doi 10.1186/s12943-022-01583-z.
46. Bobin C, Iddir Y, Butterworth C, Masliah-Planchon J, Saint-Charles A, Bellini A*, et al.* Sequential analysis of cfDNA reveals clonal evolution in neuroblastoma patients receiving ALK targeted therapy. Clin Cancer Res **2024**;30(15):3316-28 doi 10.1158/1078-0432.CCR-24-0753.

47. Ruhen O, Lak NSM, Stutterheim J, Danielli SG, Chicard M, Iddir Y*, et al.* Molecular Characterization of Circulating Tumor DNA in Pediatric Rhabdomyosarcoma: A Feasibility Study. JCO Precis Oncol **2022**;6:e2100534 doi 10.1200/PO.21.00534.
48. Chen Y, Xu L, Lin RY, Muschen M, Koeffler HP. Core transcriptional regulatory circuitries in cancer. Oncogene **2020**;39(43):6633-46 doi 10.1038/s41388-020-01459-w.
49. Durbin AD, Zimmerman MW, Dharia NV, Abraham BJ, Iniguez AB, Weichert-Leahey N*, et al.* Selective gene dependencies in MYCN-amplified neuroblastoma include the core transcriptional regulatory circuitry. Nat Genet **2018**;50(9):1240-6 doi 10.1038/s41588-018-0191-z.
50. Izquierdo E, Yuan L, George S, Hubank M, Jones C, Proszek P*, et al.* Development of a targeted sequencing approach to identify prognostic, predictive and diagnostic markers in paediatric solid tumours. Oncotarget **2017**;8(67):112036-50 doi 10.18632/oncotarget.23000.
51. Van der Auwera GA, O'Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media; 2020.
52. Nakken S, Fournous G, Vodak D, Aasheim LB, Myklebost O, Hovig E. Personal Cancer Genome Reporter: variant interpretation report for precision oncology. Bioinformatics **2018**;34(10):1778-80 doi 10.1093/bioinformatics/btx817.
53. Knaus BJ, Grunwald NJ. vcfr: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour **2017**;17(1):44-53 doi 10.1111/1755-0998.12549.
54. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics **2016**;32(18):2847-9 doi 10.1093/bioinformatics/btw313.
55. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York **2016** doi https://ggplot2.tidyverse.org.
56. Nilsen G, Liestol K, Van Loo P, Moen Vollan HK, Eide MB, Rueda OM*, et al.* Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. BMC Genomics **2012**;13:591 doi 10.1186/1471-2164-13-591.
57. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W*, et al.* Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A **2010**;107(39):16910-5 doi 10.1073/pnas.1009843107.
58. Kaufmann TL, Petkovic M, Watkins TBK, Colliver EC, Laskina S, Thapa N*, et al.* MEDICC2: whole-genome doubling aware copy-number phylogenies for cancer evolution. Genome Biol **2022**;23(1):241 doi 10.1186/s13059-022-02794-9.
59. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P*, et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. Cell **2017**;171(5):1029-41 e21 doi 10.1016/j.cell.2017.09.042.
60. Yevshin I, Sharipov R, Kolmykov S, Kondrakhin Y, Kolpakov F. GTRD: a database on gene transcription regulation-2019 update. Nucleic Acids Res **2019**;47(D1):D100-D5 doi 10.1093/nar/gky1128.
61. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P*, et al.* Determination and inference of eukaryotic transcription factor sequence specificity. Cell **2014**;158(6):1431-43 doi 10.1016/j.cell.2014.08.009.
62. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R*, et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res **2023**;51(D1):D638-D46 doi 10.1093/nar/gkac1000.

Figure 1: Study Design. a) *Schematic of the study design. Clinical arm, left: rapid turn-around reporting of sequence and copy number variants (CNVs) from Panel-Seq of relapse tissue from 20 centres across the UK. Research arm, right: comparative studies between diagnostic or relapse tissue samples and cfDNA at relapse. Panel-Seq and lcWGS were performed for all analytes if possible, with Panel-Seq taking priority. lcWGS was utilised for CNA, tumour purity/ ctDNA content estimation and fragmentomics of cfDNA.* ***b)*** *Diagnoses of the patients included in the study for which paired analysis was performed. Tumour types with less than 5 samples are grouped together. Totals are indicated at the top of each bar.*

Figure 2: The genomic landscape of matched primary and relapse paediatric tumour samples. a) *Oncoplot of mutations detected in tissue at diagnosis and relapse, excluding clinically defined hypermutator SMP0342 (n= 276). Only cases with at least one SNV or indel are shown (n= 149). Sample purity estimated by lcWGS and patient demographics are annotated (top). Where there is no lcWGS sample the purity is shown as white. Histograms at the top and right indicate the total number of mutations in each patient or gene respectively (purple – detected in both primary and relapse, blue – primary only, orange – relapse only, yellow – different mutations in the same gene between diagnosis and relapse).* ***b)*** *Genes with mutations acquired at relapse, ordered by frequency of relapse specific mutations in the whole cohort (n=276, variant positive n=149), and patients with* ***c)*** *sarcoma (n=78, variant positive n=44) and* ***d****) neuroblastoma (n=58, variant positive n=36).*

Figure 3: Paediatric cancers at relapse show enrichment of mutations in epigenetic modifiers. a) *The number of mutations detected in primary (blue) and relapse (orange) tissue using targeted panel sequencing. n=276 pairs, *** Wilcoxon test, p = 0.000646* ***b)*** *Fraction of genome altered at primary and relapse estimated by lcWGS for pairs with purity >= 0.1, n= 126, ** Wilcoxon test, p = 0.0015. A horizontal bar is drawn at the median FGA.* ***c-e)*** *Mirror plots depicting the frequency of copy number variants observed across the genome in* ***c)*** *primary tumours,* ***d)*** *relapsed tumours and* ***e)*** *comparing primary with relapse tumours. Only cases with purity >0.15 were included, n=119. Amplifications are shown (red) and gains (light red), deletions (blue) and losses (light blue). Key clinically relevant genes with focal copy number variants are labelled at their cytogenetic location. Gene labels above or below the x-axis are amplified or lost respectively. Only genes altered in 5 or more patients are shown.* ***f)*** *Confidence intervals of dN/dS for missense (red) and truncating mutations (blue). For paired cases, n=275, mutations were pooled for primary or relapse samples. "Unique to Relapse" mutations were collected by removing the mutations found in the primary sample in each pair. Samples from patients with an abundance of synonymous mutations (SMP0342 and SMP0448) were excluded from dN/dS analysis.* ***g)*** *Scatterplots comparing dN/dS at primary and relapse for missense (red) and truncating mutations (blue). Genes with positive selection for nonsynonymous mutations at relapse but not at diagnosis are labelled and coloured. The top 2 genes with the greatest selection for nonsynonymous mutations at both diagnosis and relapse are labelled. CDKN2A.p16ink4a and CDKN2A.p14arf are both labelled CDKN2A.*

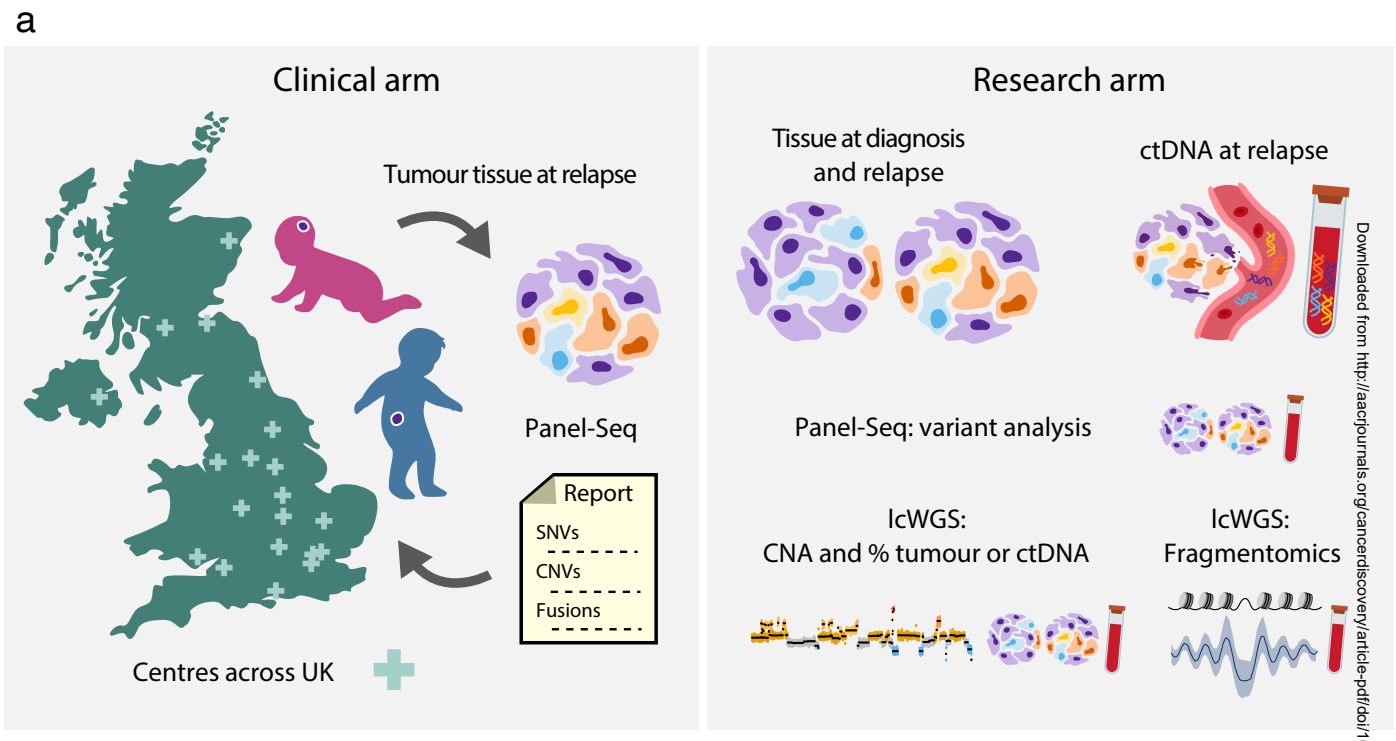
Figure 4: Sequencing of high purity ctDNA at the time of relapse better reflects tumour heterogeneity than tissue sequencing. a) *Fraction of total cell free DNA that is tumour derived, estimated by lcWGS, split according to diagnostic group.* ***b)*** *The proportion of variants detected in cfDNA only, tissue only and both samples across the whole cohort regardless of ctDNA purity, in high purity ctDNA samples (>10% ctDNA by lcWGS) and low purity cfDNA samples (<10% ctDNA by lcWGS).* ***c)*** *Bar chart showing genes with the most ctDNA specific variants detected across the cohort as a whole, ordered by number of cfDNA unique variants.* ***d)*** *Average variant allele frequency of ctDNA specific variants versus variants that were present in both tissue and ctDNA (*** Wilcoxon test p=0.0002)* ***e)*** *Oncoplot showing comparison of tissue and ctDNA sequencing in patients with neuroblastoma.* ***f)*** *Oncoplot showing comparison of tissue and ctDNA sequencing in patients with brain tumours.* ***g)*** *Oncoplot showing comparison of tissue and ctDNA sequencing in patients with lymphoma.*

Figure 5: ctDNA analysis identifies aggressive sub-clones and can predict relapse in neuroblastoma. a) *Copy number phylogeny from patient SMP0083.* ***b)*** *lcWGS copy number profiles of patient SMP0083 with neuroblastoma: primary and relapse tissue sample biopsies at primary thoracic site, cfDNA sample at relapse and additional relapse tissue biopsy sample collected from a metastatic lesion in the liver.* ***c)*** *Tissue and cfDNA purity of samples harbouring cfDNA exclusive ultra-high risk (UHR) mutations or tissue exclusive UHR mutations.* ***d)*** *The distribution of UHR mutations in cfDNA and tissue.* ***e)*** *Levels of ALK mutation detected by ctPC panel and total cfDNA levels (ng/ml of plasma) from patient SMP0409 while on treatment. The % VAF of ALK mutation detected in cfDNA is shown in black circles with the y-axis on the left. At three timepoints, indicated by asterisks, cfDNA sequencing was of suboptimal depth (<250x), therefore VAF should be interpreted with caution. Total cfDNA levels shown in grey with the y-axis on the right. Treatment starting points are indicated by dark red arrows and progression as indicated by MRI scan*

is highlighted with bright red arrows. ***f)*** *Levels of mutations (as indicated by the key) detected by ctPC panel and total cfDNA levels (ng/ml of plasma) in a patient LB001 while on treatment. The % VAF of ALK mutation detected in cfDNA is shown by back circles and the y-axis on the left. Total cfDNA levels shown in grey with the y-axis on the right. Progression as indicated by MRI scan is highlighted with bright red arrows.*

Figure 6: Nucleosome profiling of cfDNA lcWGS reveals signals associated with paediatric cancer subtype. a) *Fragmentomics schematic: cfDNA is enriched for sub-nucleosome fragments (120-180bp). cfDNA preferentially fragments at open chromatin leading to a loss of reads. Therefore, coverage can be used to infer active sites and nucleosome positioning.* ***b)*** *Hierarchically clustered heatmap of the normalised composite coverage for TFBS which show significantly disease dependent variance as determined by Wilcoxon rank sum tests. Disease type, ctDNA content, MYCN amplification and fusion gene presence are annotated at the top. Disease-specific clusters are labelled c1-c4. Genes of interest are labelled. Only cfDNA samples with a ctDNA content of over 20% were used and those belonging to a diagnostic group with less than 3 patients were not included (n=85).* ***c)*** *Gene ontology analysis for clusters c1-c4.* ***d-f)*** *Composite normalised coverage for 1kb +/- TFBSs. The solid lines show the mean and the ribbons indicate the range from the 10^{th} to the 90^{th} percentile. Average coverage for a specific disease group is plotted with all other samples with >20% ctDNA content as indicated by the key. Additionally, a third line for ctDNA content <10% is plotted in* ***(d)*** *and* ***(e)****. The TFBSs plotted are indicated by the x axis title.* ***g-h)*** *Box plots comparing coverage at disease-specific CRC TBFSs in hepatoblastoma* ***(g)*** *and rhabdomyosarcoma* ***(h)****.* ***i-k)*** *Composite normalised coverage for 1kb +/- TFBSs. The solid lines show the mean and the ribbons indicate the range from the 10^{th} to the 90^{th} percentile. Average coverage for a specific disease group is plotted with all other samples with >20% ctDNA content as indicated by the key. The TFBSs plotted are indicated by the x axis title.* ***l)*** *Box plots comparing coverage at disease-specific CRC TBFSs in neuroblastoma. Wilcoxon rank sum test: *=p<0.05, **=p<0.01 , ***=p<0.001, ****=p<0.0001.*

Figure 1: Study Design.



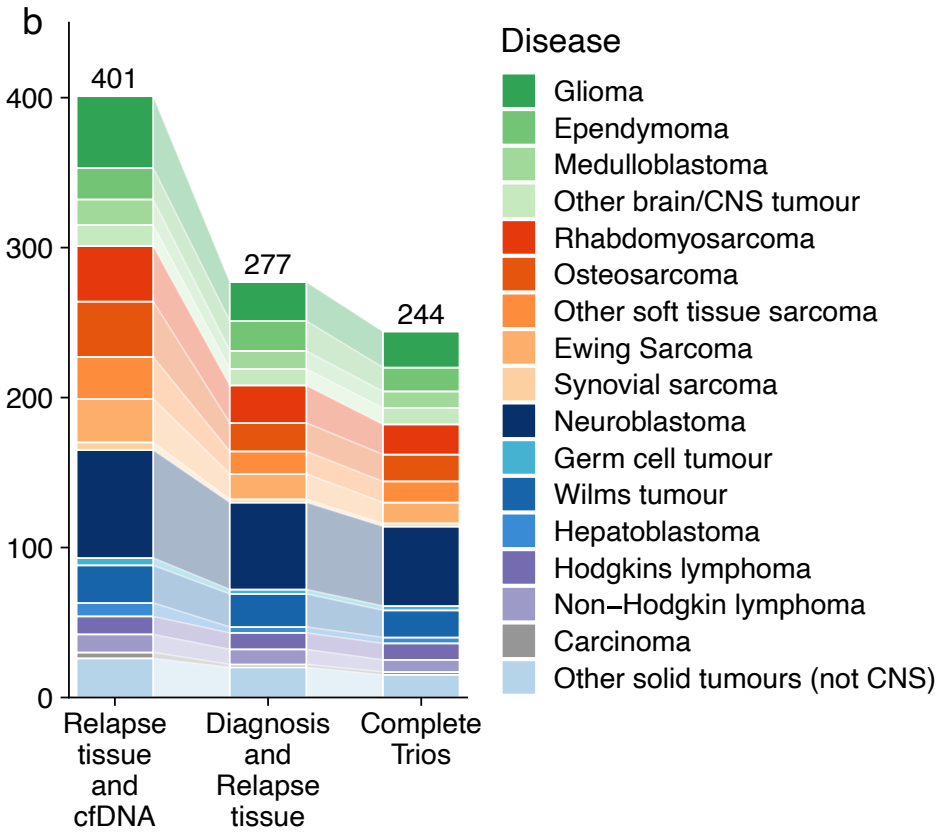


Figure 2: The genomic landscape of matched primary and relapse paediatric tumour samples.

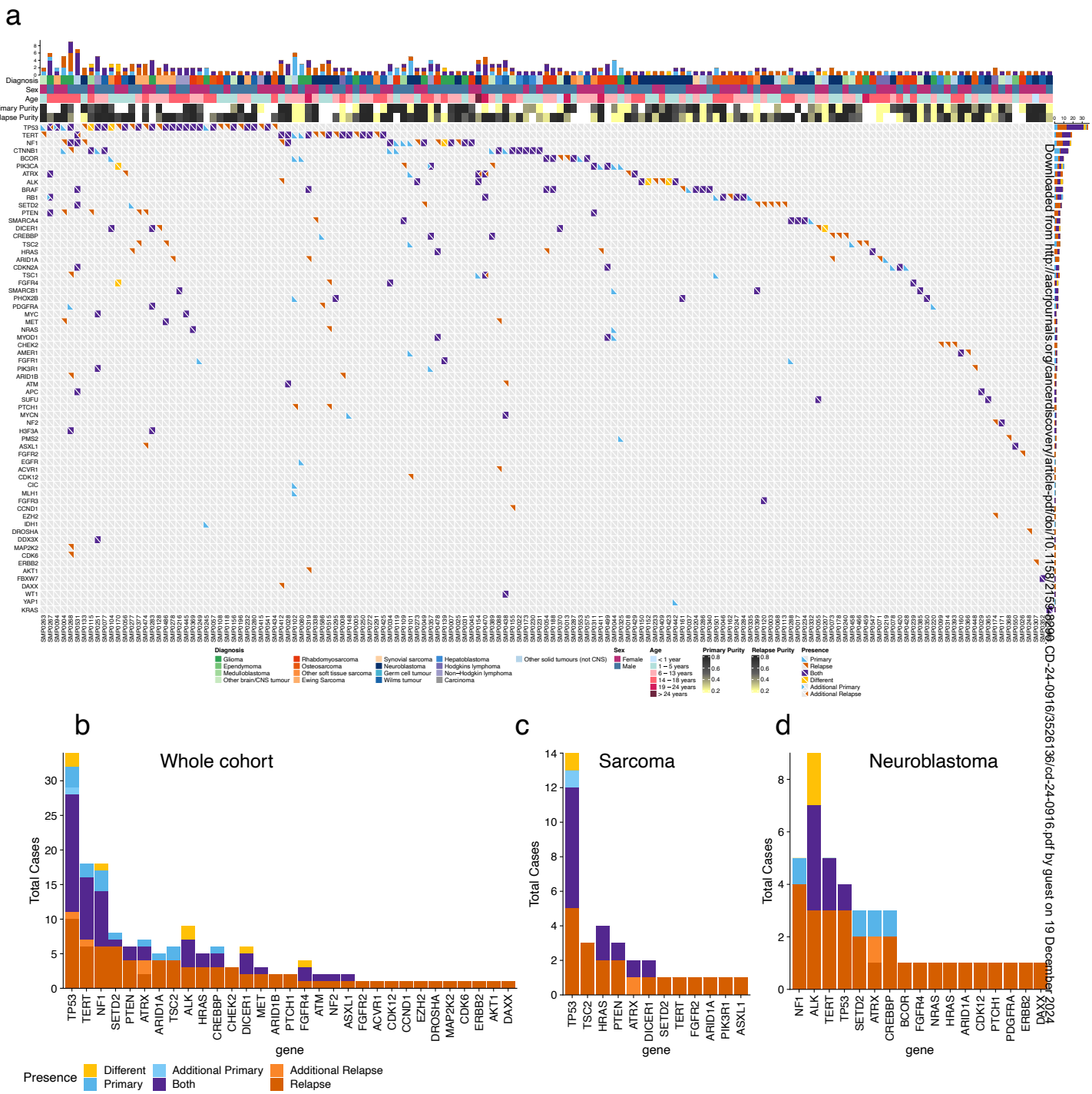


Figure 3: Paediatric cancers at relapse show enrichment of mutations in epigenetic modifiers.

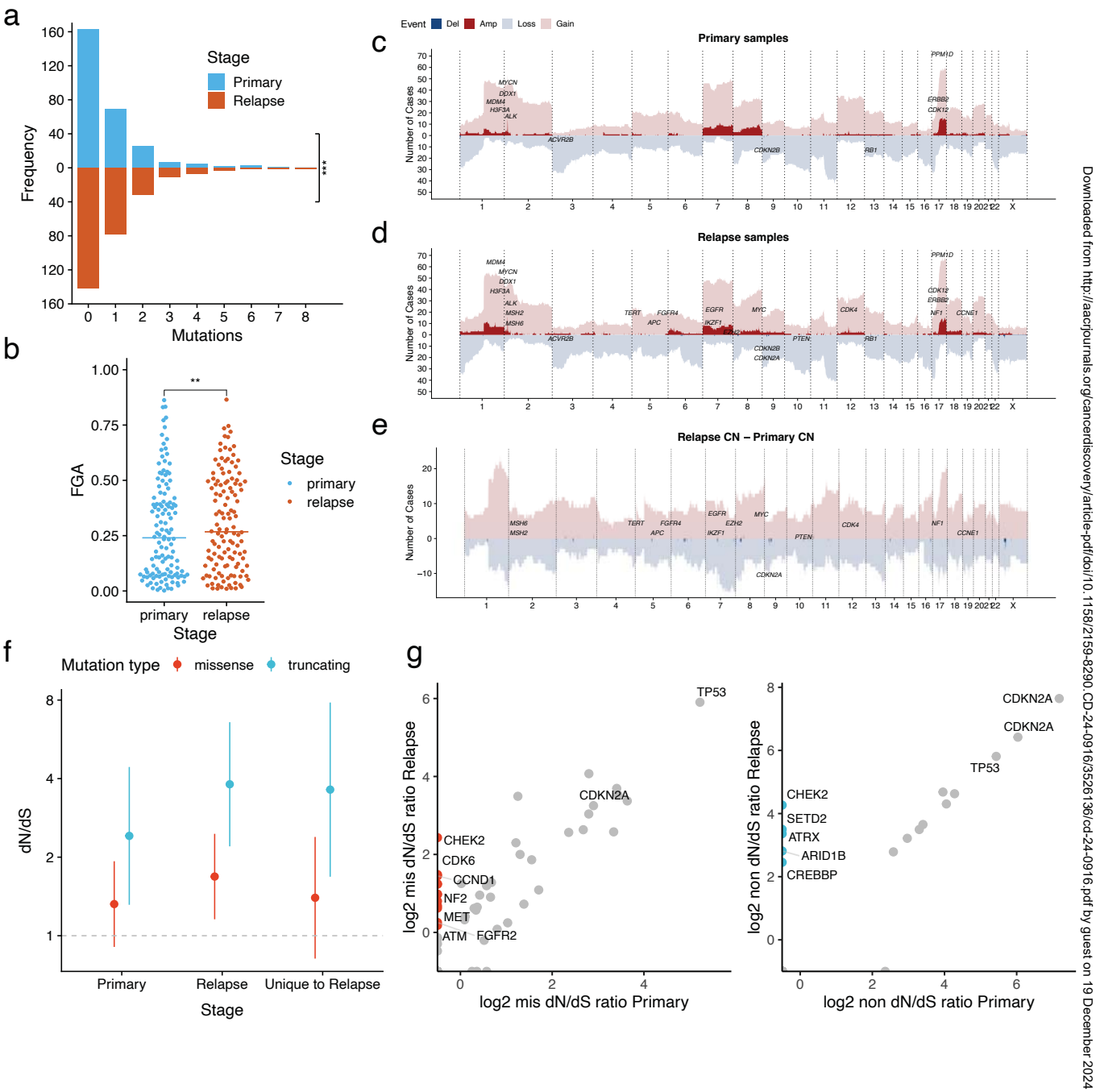


Figure 4: Sequencing of high purity ctDNA at the time of relapse better reflects tumour heterogeneity than tissue sequencing.

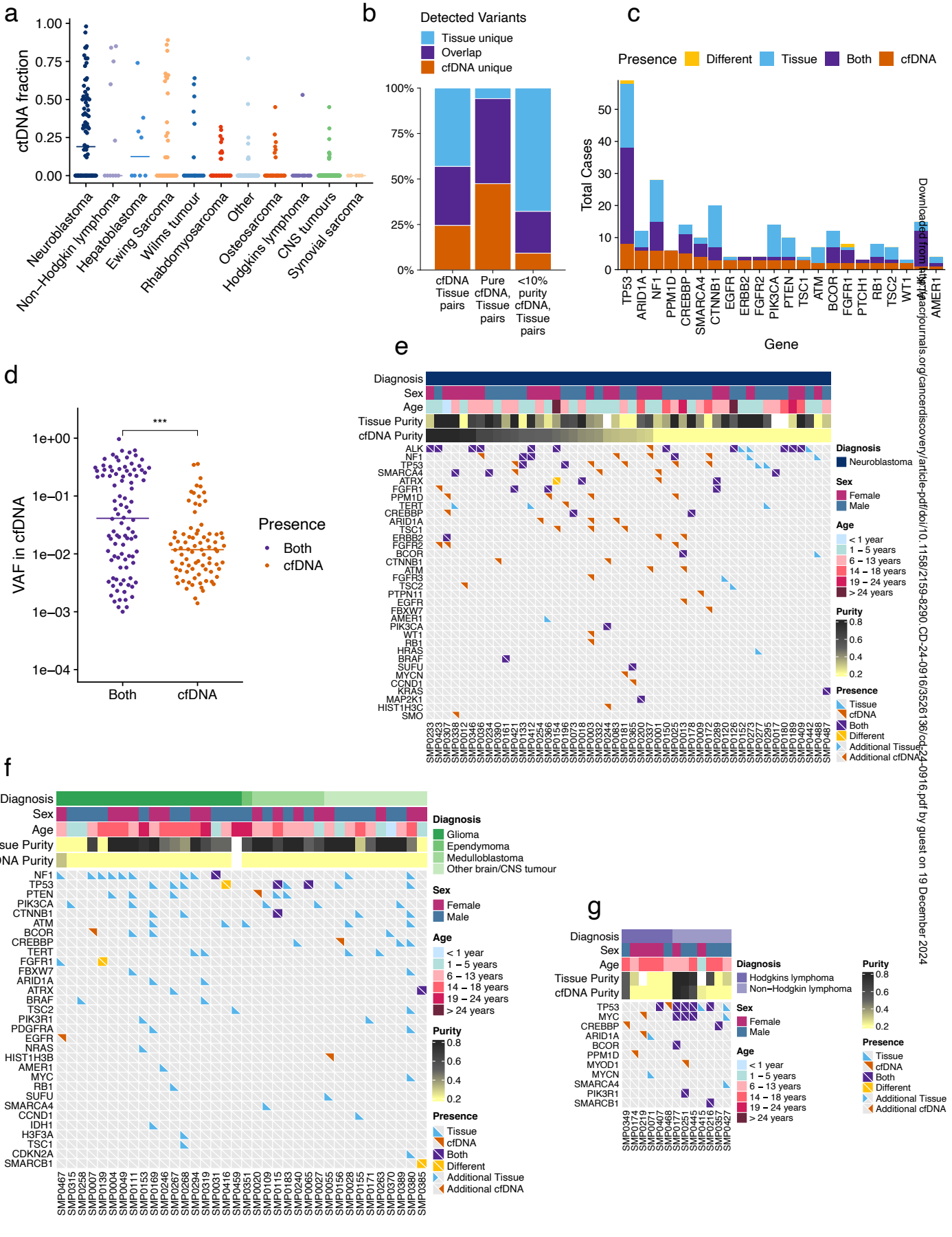


Figure 5: ctDNA analysis identifies aggressive sub-clones and can predict relapse in neuroblastoma.

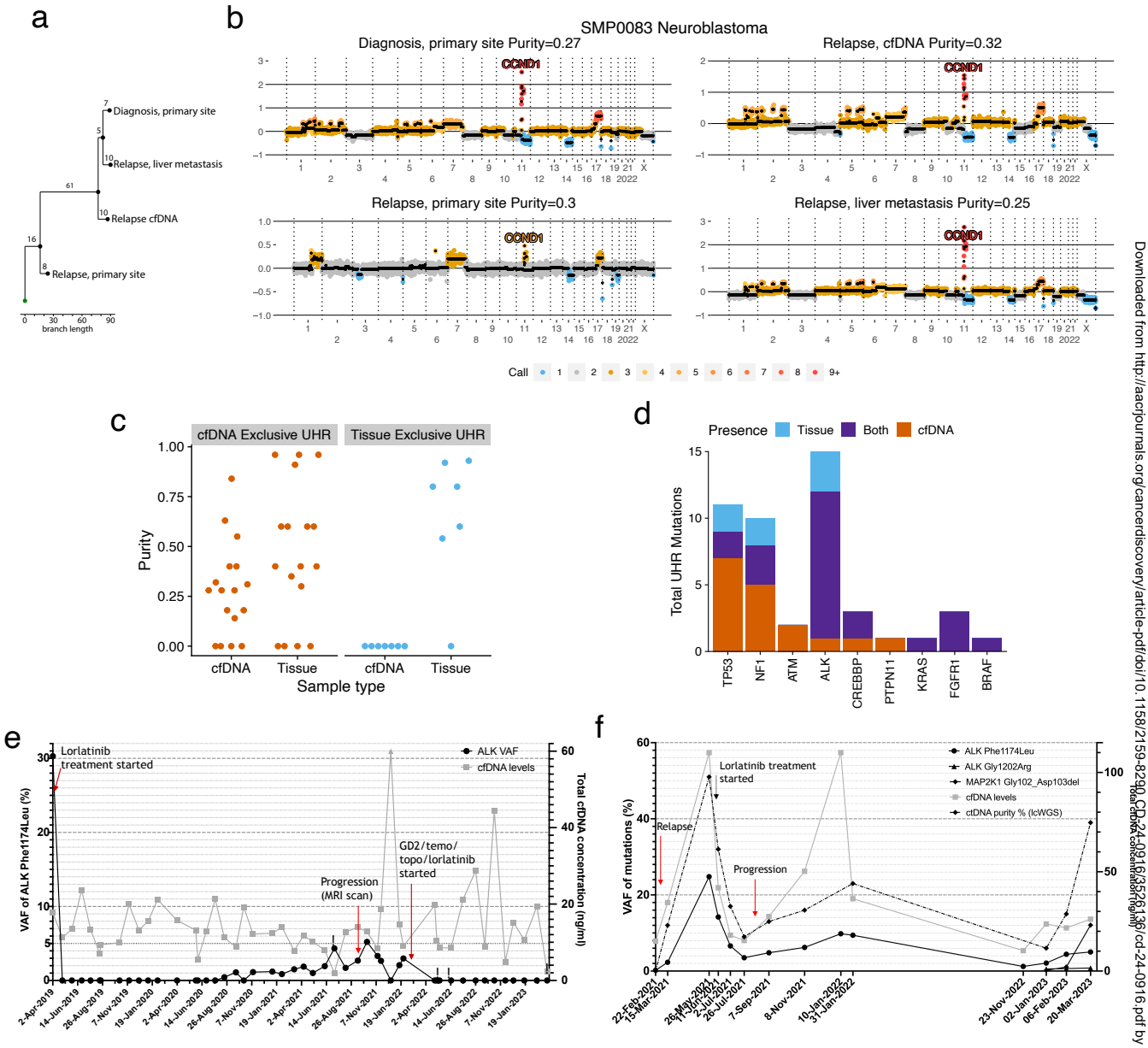


Figure 6: Nucleosome profiling of cfDNA lcWGS reveals signals associated with paediatric cancer subtype.