



# Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer

Marie Wong <sup>1,2,3,25</sup>, Chelsea Mayoh <sup>1,2,25</sup>, Loretta M. S. Lau<sup>1,2,4,25</sup>, Dong-Anh Khuong-Quang <sup>5,6</sup>, Mark Pinese <sup>1,2,3</sup>, Amit Kumar <sup>1,7</sup>, Paulette Barahona<sup>1</sup>, Emilie E. Wilkie <sup>1,2</sup>, Patricia Sullivan <sup>1</sup>, Rachel Bowen-James <sup>1</sup>, Mustafa Syed <sup>1</sup>, Iñigo Martincorena<sup>8</sup>, Federico Abascal<sup>8</sup>, Alexandra Sherstyuk<sup>1</sup>, Noemi A. Bolanos <sup>1,2,4</sup>, Jonathan Baber<sup>9,10</sup>, Peter Priestley <sup>9,10</sup>, M. Emmy M. Dolman<sup>1</sup>, Emmy D. G. Fleuren<sup>1,2</sup>, Marie-Emilie Gauthier<sup>1</sup>, Emily V. A. Mould<sup>1</sup>, Velimir Gayevskiy<sup>3</sup>, Andrew J. Gifford<sup>1,2,11</sup>, Dylan Grebert-Wade<sup>1</sup>, Patrick A. Strong<sup>1</sup>, Elodie Manouvrier<sup>1</sup>, Meera Warby<sup>12</sup>, David M. Thomas <sup>3</sup>, Judy Kirk<sup>13,14</sup>, Katherine Tucker<sup>15,16</sup>, Tracey O'Brien<sup>2,4</sup>, Frank Alvaro<sup>17</sup>, Geoffry B. McCowage<sup>12</sup>, Luciano Dalla-Pozza<sup>12</sup>, Nicholas G. Gottardo<sup>18,19</sup>, Heather Tapp<sup>20</sup>, Paul Wood<sup>21</sup>, Seong-Lin Khaw<sup>5,6</sup>, Jordan R. Hansford<sup>5</sup>, Andrew S. Moore <sup>22,23</sup>, Murray D. Norris<sup>1,2,24</sup>, Toby N. Trahair<sup>1,2,4</sup>, Richard B. Lock<sup>1,2</sup>, Vanessa Tyrrell<sup>1</sup>, Michelle Haber<sup>1,2</sup>, Glenn M. Marshall<sup>1,2,4</sup>, David S. Ziegler <sup>1,2,4,25</sup> , Paul G. Ekert <sup>1,2,6,7,25</sup>  and Mark J. Cowley <sup>1,2,3,25</sup> 

**The Zero Childhood Cancer Program is a precision medicine program to benefit children with poor-outcome, rare, relapsed or refractory cancer. Using tumor and germline whole genome sequencing (WGS) and RNA sequencing (RNAseq) across 252 tumors from high-risk pediatric patients with cancer, we identified 968 reportable molecular aberrations (39.9% in WGS and RNAseq, 35.1% in WGS only and 25.0% in RNAseq only). Of these patients, 93.7% had at least one germline or somatic aberration, 71.4% had therapeutic targets and 5.2% had a change in diagnosis. WGS identified pathogenic cancer-predisposing variants in 16.2% of patients. In 76 central nervous system tumors, methylome analysis confirmed diagnosis in 71.1% of patients and contributed to a change of diagnosis in two patients (2.6%). To date, 43 patients have received a recommended therapy, 38 of whom could be evaluated, with 31% showing objective evidence of clinical benefit. Comprehensive molecular profiling resolved the molecular basis of virtually all high-risk cancers, leading to clinical benefit in some patients.**

Despite advances in therapeutic development and diagnostic technologies, cancer is the leading cause of disease-related death in children in most developed countries<sup>1</sup>. Patients with high-risk pediatric cancer are children, adolescents and young adults ( $\leq 21$  years) with a less than 30% chance of surviving 5 years after their diagnosis. These patients typically have an aggressive tumor with few established treatment options or relapsed or refractory

disease despite standard therapy<sup>2-5</sup>. The potential availability of new targeted therapies and immunotherapies has ushered in the era of precision medicine, where the molecular profile of a tumor helps guide patient management<sup>6,7</sup>. The hypothesis is that matching treatments to molecular changes in the tumor results in more effective cancer control and less long-term treatment-related side effects<sup>8</sup>. There are unique challenges to personalizing pediatric

<sup>1</sup>Children's Cancer Institute, Lowy Cancer Centre, UNSW Sydney, Kensington, NSW, Australia. <sup>2</sup>School of Women's and Children's Health, UNSW Sydney, Kensington, NSW, Australia. <sup>3</sup>Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia. <sup>4</sup>Kids Cancer Centre, Sydney Children's Hospital, Randwick, NSW, Australia. <sup>5</sup>Children's Cancer Centre, Royal Children's Hospital, Parkville, VIC, Australia. <sup>6</sup>Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, VIC, Australia. <sup>7</sup>Peter MacCallum Cancer Centre, Melbourne, VIC, Australia. <sup>8</sup>Wellcome Trust Sanger Institute, Hinxton, UK. <sup>9</sup>Hartwig Medical Foundation, Amsterdam, The Netherlands. <sup>10</sup>Hartwig Medical Foundation Australia, Sydney, NSW, Australia. <sup>11</sup>Department of Anatomical Pathology, Prince of Wales Hospital, Randwick, NSW, Australia. <sup>12</sup>Cancer Centre for Children, The Children's Hospital Westmead, Westmead, NSW, Australia. <sup>13</sup>Familial Cancer Service, Crown Princess Mary Cancer Centre, Westmead Hospital, Westmead, NSW, Australia. <sup>14</sup>Sydney Medical School, University of Sydney Centre for Cancer Research, The Westmead Institute for Medical Research, Westmead, NSW, Australia. <sup>15</sup>Hereditary Cancer Centre, Prince of Wales Hospital, Randwick, NSW, Australia. <sup>16</sup>Prince of Wales Hospital Clinical School, University of New South Wales, Randwick, NSW, Australia. <sup>17</sup>John Hunter Children's Hospital, New Lambton Heights, NSW, Australia. <sup>18</sup>Department of Paediatric and Adolescent Oncology/Haematology, Perth Children's Hospital, Nedlands, WA, Australia. <sup>19</sup>Brain Tumour Research Program, Telethon Kids Institute, Nedlands, WA, Australia. <sup>20</sup>Women's and Children's Hospital, Adelaide, SA, Australia. <sup>21</sup>Monash Children's Hospital, Melbourne, VIC, Australia. <sup>22</sup>Oncology Service, Oncology Service, Queensland Children's Hospital, Brisbane, QLD, Australia. <sup>23</sup>Child Health Research Centre, The University of Queensland, Brisbane, QLD, Australia. <sup>24</sup>University of New South Wales Centre for Childhood Cancer Research, UNSW Sydney, Kensington, NSW, Australia.

<sup>25</sup>These authors contributed equally: Marie Wong, Chelsea Mayoh, Loretta M.S. Lau, David S. Ziegler, Paul G. Ekert, Mark J. Cowley.

✉e-mail: [d.ziegler@unsw.edu.au](mailto:d.ziegler@unsw.edu.au); [pekert@ccia.org.au](mailto:pekert@ccia.org.au); [mcowley@ccia.org.au](mailto:mcowley@ccia.org.au)

cancer treatment: first, only 45% of pediatric cancer driver genes are shared with adult cancers<sup>9,10</sup>, suggesting that new therapeutic agents are required for pediatric cancer; second, pediatric cancers are often driven by structural variants that can be challenging to identify and target; and third, many new targeted drugs lack dosage guidelines and efficacy data in children. Thus, pediatric-specific treatment strategies will be critical to effective personalized medicine.

There are now several large-scale pediatric cancer precision medicine programs that have adopted different combinations of sequencing platforms<sup>7</sup>. Early reports suggested that pathogenic variants were detected in 39–50% of cases using a single platform or targeted approach<sup>11</sup> to 50% or more when multiple sequencing platforms were combined (whole exome and RNAseq)<sup>12–16</sup>. Encouragingly, the Pediatric MATCH trial<sup>17</sup> (NCT03155620) is finding targetable genetic changes in 24–29% of patients using targeted DNA sequencing<sup>18</sup>. In the research setting, deep WGS of tumors (60–100× depth) and matched germline DNA (30× depth) is currently the most accurate and comprehensive platform for the analysis of DNA mutations<sup>19,20</sup>. WGS identifies short protein-coding variants also accessible by targeted sequencing and resolves the noncoding genomic breakpoints of small and large copy number variants (CNVs) and copy number neutral structural variants (SVs). WGS accurately measures hypermutation and mutational signatures<sup>21</sup>, both of which are clinically important<sup>22,23</sup>. RNAseq analysis detects expressed driver fusions<sup>24</sup> and aberrantly expressed genes, which might correlate with drug sensitivity in cell lines<sup>25</sup> and in some patients<sup>26</sup>. Finally, methylation profiling is remarkably accurate in classifying central nervous system (CNS) tumors into specific molecular entities<sup>27</sup>, which is critical for patient management. We hypothesize that combining all these modalities into a comprehensive molecular profiling approach would have higher rates of actionable findings than those in previous reports.

A major challenge of using a comprehensive profiling approach for precision medicine is distilling which of the millions of molecular changes within a tumor are pathogenic and relevant to clinicians<sup>28</sup>. Pathogenicity is determined by a multidisciplinary curation team without the ability to perform additional validation studies, instead relying on guidelines<sup>29</sup>, databases<sup>30</sup>, published literature and clinical experience. Reportable variants are those deemed important drivers of tumor biology, a subset of which are potentially actionable. Actionable variants are a particularly fluid designation and can precipitate changed management, through informing diagnosis, prognosis, familial cancer risk or treatment. For example, actionability depends on drug and trial availability, which differ by jurisdiction and over time. Highly effective neurotrophic receptor tyrosine kinase (NTRK) inhibitors are now highly actionable since receiving US Food and Drug Administration approval<sup>31</sup> but were less so when therapy required compassionate access programs<sup>32</sup>.

The Zero Childhood Cancer Program is Australia's first national pediatric cancer precision medicine program, focused on real-time recruitment and analysis of patients with high-risk pediatric cancer. Here we present the first systematic evaluation of the utility and early clinical effects of a comprehensive molecular profiling platform consisting of germline and tumor WGS, RNAseq and CNS methylome analysis to identify clinically significant variants relevant to the biology of the tumor, diagnosis, clinical management or prognosis. We show that these molecular aberrations can be translated into treatment recommendations that show promising early clinical effects. This study also provides the first comprehensive estimate of the incidence and range of germline cancer susceptibility mutations in Australian patients with high-risk pediatric cancer.

## Results

**The molecular aberration landscape of high-risk pediatric cancers.** From 2015 to June 2019, we recruited 247 patients with high-risk or rare pediatric cancers to the Zero Childhood Cancer

Program and profiled 252 tumors. The first 47 patients were part of the TARGET feasibility study (L.L., C.M., J.Xie., P.B., M.W., et al., submitted manuscript). The next 200 consecutive patients were prospectively recruited onto the national PRISM trial (NCT03336931) from September 2017. The study workflow is described in the Methods and Extended Data Fig. 1. The patient cohort was heterogeneous with respect to cancer type, disease stage, age and treatments (Fig. 1 and Supplementary Table 1). Patients were recruited over the age of 21 years if the tumor type was predominantly a pediatric subtype or was a relapse of a childhood tumor (Fig. 1a). We grouped tumors into five main categories: CNS ( $n=92$ ), sarcoma ( $n=62$ ), non-sarcomatous extracranial solid ( $n=35$ ; hereafter solid tumor), neuroblastoma ( $n=20$ ) and hematological malignancy ( $n=43$ ) (Fig. 1b,c). Patients were enrolled at various disease stages: initial diagnosis (47.2%), relapse (41.3%) or refractory (9.9%). A few (1.6%) were treatment-induced secondary tumors.

For each individual tumor, we prospectively integrated results from WGS ( $n=252$ , 100%) and RNAseq ( $n=228$ , 90.5%) and methylation arrays for CNS tumors ( $n=76$ , 29.3%) (Fig. 1b). We curated single-nucleotide variants (SNVs), short insertions and deletions (indels), SVs, CNVs, gene expression outliers, tumor mutation burden (TMB) and the match between the histopathological diagnosis and the diagnostic subclass suggested by the molecular neuropathology (MNP) classifier in CNS tumors<sup>27</sup> (Methods and Extended Data Fig. 1). Across the cohort, 92.1% (232/252) of tumors had at least one pathogenic or likely pathogenic somatic variant of any type affecting key oncogenic and tumor suppressor pathways (Fig. 2a and Extended Data Figs. 2 and 3a,b). Common SNV mutations affected epigenetic regulation (*H3F3A*, *SMARCB1*, *ATRX* and *SMARCA4*), DNA maintenance (*TP53*), cell cycle (*RB1* and *ATM*) and kinase signaling pathway genes (*PIK3CA*, *PTEN*, *NRAS* and *NF1*), consistent with the cancer subtypes in the cohort and previous studies<sup>9,10</sup>. Frequent CNVs included homozygous deletions of tumor suppressor genes (*CDKN2A/B*, *PTEN*, *RB1* and *TP53*) and epigenetic modifiers (*SMARCB1* and *SMARCA4*) and gene amplifications involving transcriptional regulators (*MYC* and *MYCN*), receptor tyrosine kinases (RTKs) (*PDGFRA*, *ERBB2*, *KIT* and *EGFR*) or the cyclin-dependent kinase *CDK4* (Fig. 2b). WGS resolved precise CNV breakpoints in every case. Chromosome 1q gains were the most common whole chromosome arm gain (centromeric break points) in all tumor types (Fig. 2b and Extended Data Fig. 3b–f). However, whole chromosome arm gains or losses were uncommon compared to segmental CNVs (non-centromeric breakpoints), unlike adult tumors<sup>19</sup>. This likely indicates different mechanisms driving CNVs in childhood cancers.

Aberrant messenger RNA expression levels from RNAseq (Methods), when associated with genomic events such as CNVs, SVs and coding SNVs, provided orthogonal evidence that expression contributed to the biology of the tumor (Fig. 2c). For example, *PDGFRA*, *KIT* or *CDK4* overexpression associated with genomic amplification suggests that RTK signalling pathways or *CDK4*-driven dysregulation of the cell cycle are driver events and potential drug targets. However, 63% of genes curated as reportable due to aberrant expression were not accompanied by explanatory events affecting the coding genome. Some overexpressed genes were reportable because of diagnostic associations (for example, *PHOX2B* in neuroblastoma), others because they suggested activation of targetable pathways, such as JAK-STAT or PI3K-mTOR signaling. We sought to identify noncoding driver variants linked to unexplained expression changes. In *TERT*, a gene for which promoter mutation is a well-established mechanism of activation<sup>33</sup>, we observed canonical *TERT*-activating promoter mutations at *A161T*, *C228T* and *C250T* in eight tumors and three upstream SVs (Supplementary Table 1). An unbiased scan for promoters of other dysregulated genes with a statistical excess of noncoding variation

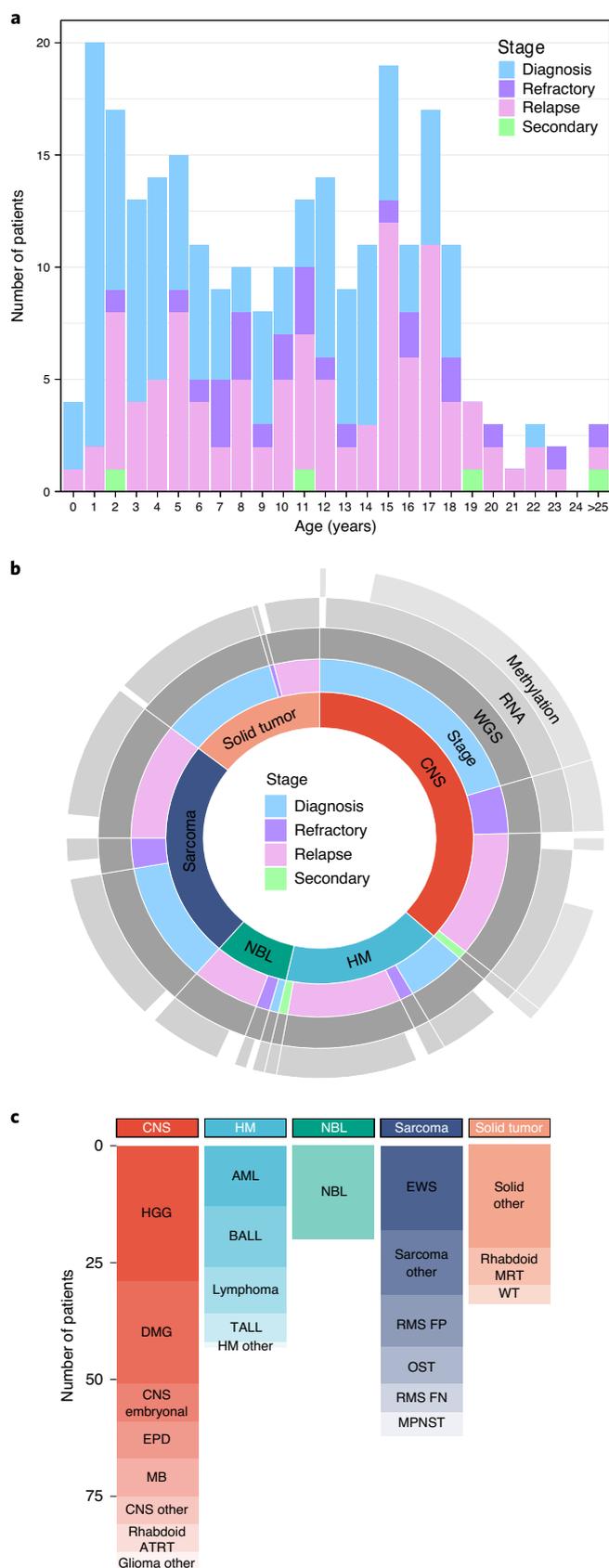
using NBR<sup>34</sup> did not identify any significant loci, likely owing to our heterogeneous and relatively small cohort size that limited power (Supplementary Table 2).

Integrated WGS and RNAseq identified 67 established or novel driver fusions (Fig. 2d). The most frequent were *EWSR1* rearrangements, *PAX3-FOXO1* and *ASPC1-TFE3* in Ewing's sarcoma, alveolar rhabdomyosarcoma and alveolar soft part sarcoma, respectively. There were 15 likely kinase-activating fusions, including six *NTRK* fusions. Other fusions highlighted new tumor biology. A *BRD4-LEUTX* in-frame fusion in an embryonal CNS tumor was identified both by WGS and RNAseq (Extended Data Fig. 4a). This retained the BRD4 bromodomains and the LEUTX homeobox. The homeobox of LEUTX functions as a transcriptional transactivator and normally is expressed during early embryogenesis<sup>35</sup>. LEUTX regulates genes associated with pluripotency and differentiation, similar to other homeobox gene fusions<sup>36</sup>. It is unknown if the BRD4 domains might be targeted by bromodomain and extra-terminal inhibitors<sup>37</sup>.

WGS with RNAseq identified an additional 34 intragenic CNVs or other SVs that either activated oncogenes or inactivated tumor suppressor genes (Fig. 2e). These arose from deletions, duplications, inversions or translocations throughout the genome, frequently expressed as out-of-frame fusions in RNAseq data. WGS detected *IgH-MYC* rearrangements, an established *MYC*-activating lesion<sup>38</sup> associated with elevated RNAseq expression. Interpretation of complex events was enhanced using the novel WGS methods PURPLE, GRIDSS2 and LINX<sup>39</sup>, which resolved clusters of SVs into phased alignment chains. For example, a complex cluster of more than 100 linked SV events involving ten chromosomes created an out-of-frame *TP53-SUZ12* fusion in a malignant peripheral nerve sheath tumor that simultaneously disrupted *TP53* and the histone methyltransferase *SUZ12* (Extended Data Fig. 4b–d). WGS and RNAseq identified three intragenic SVs, including two small deletions and an intragenic inversion affecting exons 1–3 of *IKZF1* in three cases of pre-B acute lymphoblastic leukemia (Extended Data Fig. 4E).

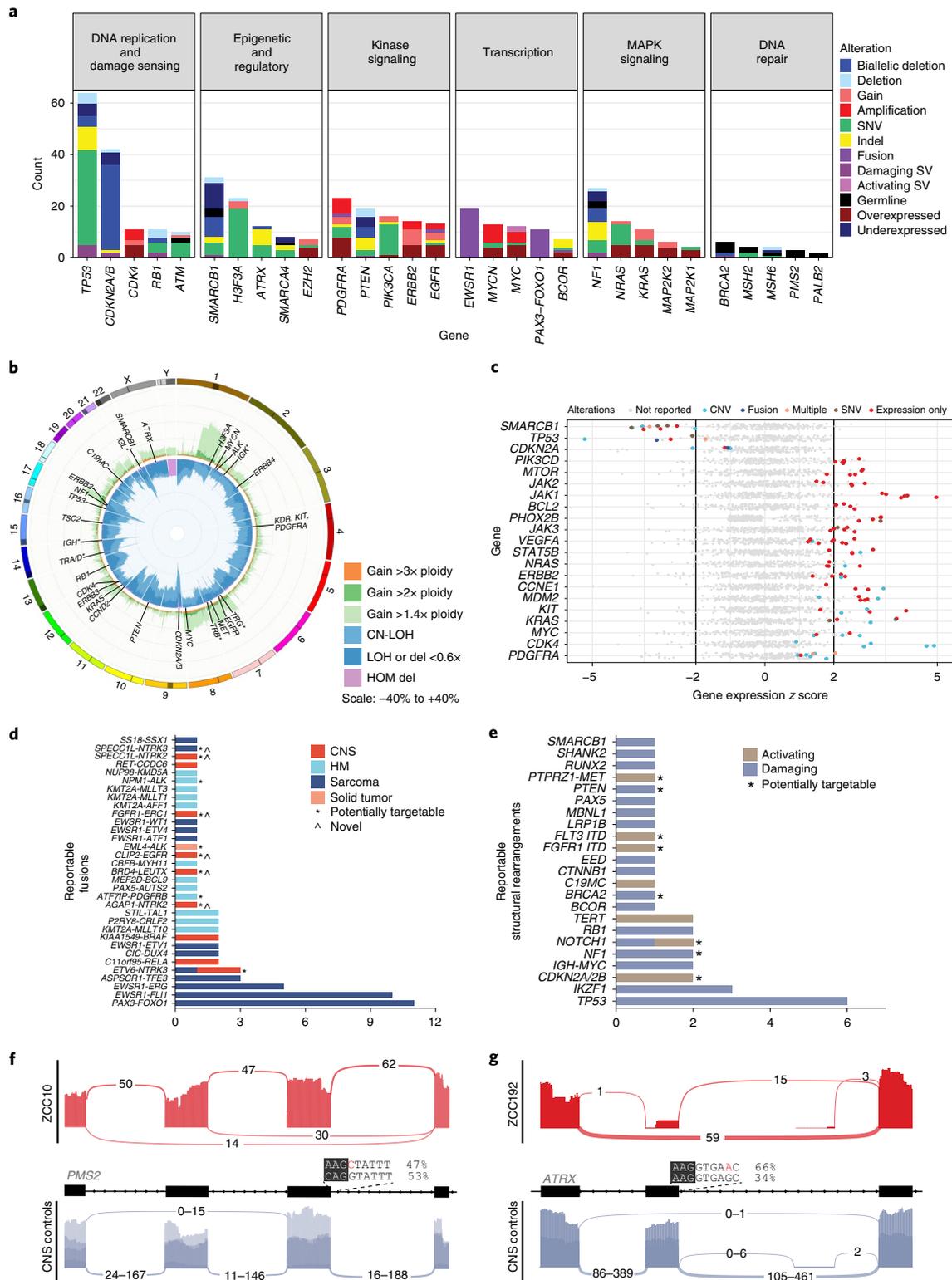
WGS and RNAseq are a powerful combination for interpreting the effect of DNA variants on splicing. Using our in-house algorithm Introme (Methods), we identified six germline and 22 somatic splice-altering variants (Supplementary Table 3). Nineteen affected canonical splice sites, where RNAseq revealed functional effects, including whole or partial exon-skipping and intron readthrough (Fig. 2f and Extended Data Fig. 5). Four intronic variants were at the +3, +5 or –8 residue, including a somatic *ATRX:c.6217+5G>A*

variant causing exon-skipping (Fig. 2g). Five coding variants in the last base of the exon, typically annotated as synonymous ( $n=4$ ) or frameshift ( $n=1$ ), caused exon-skipping and were reported as likely pathogenic or pathogenic.

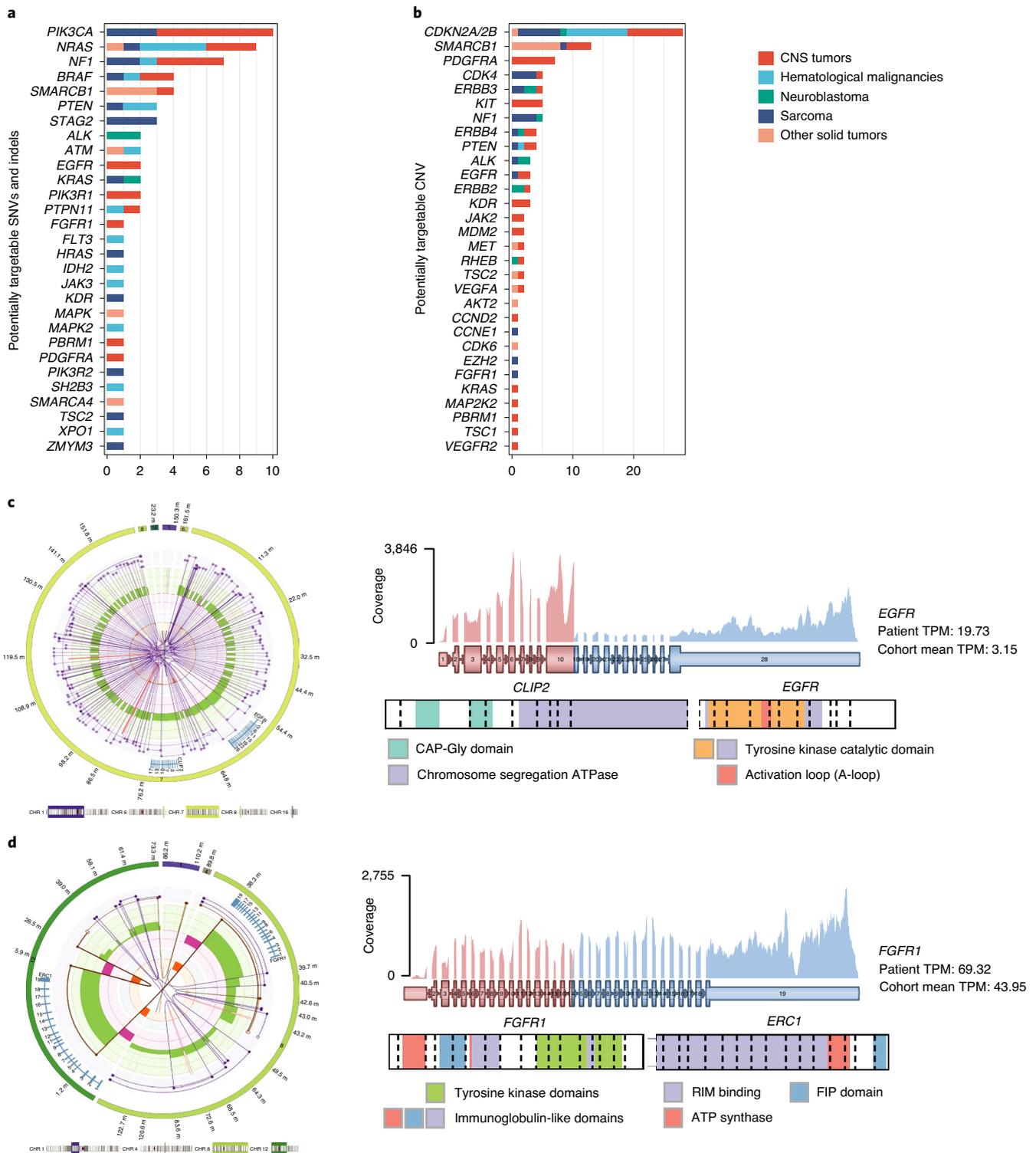


**Fig. 1 | Zero Childhood Cancer Program cohort demographics.**

Representation of the cohort consisting of 247 patients with high-risk pediatric cancer. **a**, Age distribution of the cohort (range, 0–31 years; median, 10 years) highlighting disease stage. **b**, From the innermost ring to the outer ring, the cohort is highlighted by the frequency of samples within each cancer type: CNS cancer (CNS), hematological malignancy (HM), neuroblastoma (NBL), sarcoma and non-sarcomatous extracranial solid tumor (solid tumor), stage of disease (initial diagnosis, refractory disease, relapse and secondary cancer) and sequence platform performed on each (WGS, RNAseq and methylation). **c**, The five main cancer types are broken down into more specific diagnosis with frequency of patients with each diagnosis represented. Cancer diagnosis key: high-grade glioma (HGG), diffuse midline glioma (DMG), ependymoma (EPD), medulloblastoma (MB), atypical teratoid rhabdoid tumor (ATRT), acute myeloid leukemia (AML), B-precursor acute lymphoblastic leukemia (BALL), T cell acute lymphoblastic leukemia (TALL), Ewing's sarcoma (EWS), rhabdomyosarcoma fusion positive (RMS FP), osteosarcoma (OST), rhabdomyosarcoma fusion negative (RMS FN), malignant peripheral nerve sheath tumor (MPNST), malignant rhabdoid tumor (MRT) and Wilms tumor (WT).



**Fig. 2 | Pathogenic aberrations in high-risk pediatric cancers. a**, The top five recurrently aberrated genes within six major pathways, highlighted by specific aberration type. **b**, The genome-wide copy number (CN) profile summarized for all tumors, with amplifications and deletions relative to the tumor's ploidy and loss of heterozygosity (LOH) indicated according to the legend. Both rings are on a scale of 0% to 40%. The most frequently observed amplifications and homozygous (HOM) deletions are shown, as well as rearrangements of immunoglobulin and T cell receptor gene family members (\*). **c**, Genes with reportable expression changes are shown as colored dots, and the associated genomic changes are indicated. **d**, The frequency of patients with a reportable fusion highlighted by cancer type and **(e)** intragenic CNV and other structural rearrangements classified as either activating an oncogene or damaging a tumor suppressor gene (\*, potentially targetable; ^, novel fusions). **f**, A germline canonical splice site variant in *PMS2* (c.903+1G>C) resulted in exon-skipping, confirmed by RNAseq in the tumor (red), not seen in four tissue-matched controls (purple). **g**, A somatic intronic variant in *ATRX* (c.6217+5G>A) resulted in exon-skipping. HM, hematological malignancy.



**Fig. 3 | Clinically actionable aberrations in high-risk pediatric cancers.** **a**, The frequency of somatic SNVs and **(b)** CNVs that were curated as potentially targetable, colored by tumor type. **c**, A novel, targetable *CLIP2-EGFR* fusion arose from a chromothriptic event on chromosome 7, as indicated by a copy number state alternating between 2 and 3 (green segments in the left LINX plot; see explanatory legend in Extended Data Fig. 4b). The fusion gene, the abundance of each transcribed exon, the fusion breakpoints of the *CLIP2-EGFR* fusion (top) and the retained protein domains (below) are shown. The *EGFR* RNAseq abundance in TPM for the tumor bearing the fusion and the cohort mean TPM of *EGFR* are shown. **d**, A novel, potentially targetable *FGFR1-ERC1* fusion presented as for **c**. RIM, Rab3-interacting molecule; FIP, family of Rab11-interacting proteins.

**Therapeutically actionable genomic features in high-risk pediatric cancer.** Therapeutically actionable SNVs, CNVs and SVs are pathogenic variants that directly or indirectly indicate treatment

with targeted anti-cancer drugs. These clustered most frequently in RTK signaling, MAP kinase signaling and PI3K-mTOR signaling pathways (Fig. 3a,b). RTKs were activated by point mutations (*BRAF*,

*ALK*, *EGFR* and *FGFR1*), high copy number gains (*PDGFRA*, *EGFR* and *ERBB2*) or fusions (Fig. 2a). PI3K-mTOR variants included known activating *PIK3CA* mutations or loss-of-function mutations and deletions affecting *PTEN*, *PTPN11*, *PIK3R1*, *PIK3R2*, *MTORC1* and *MTORC2*. Potentially targetable variants were recurrently seen in epigenetic regulation genes, including the SWI/SNF complex genes *SMARCB1*, *SMARCA4* and *EZH2* and in the chromatin-remodeling gene *PBRM1* (Fig. 3a,b). In the cell cycle regulatory pathways, the most common were *CDKN2A/2B* deletions and *CDK4* gene amplifications. Detecting uncommon potentially targetable variants is a potential strength of WGS, such as *STAG2*-inactivating mutations (associated with sensitivity to PARP inhibitors<sup>40</sup>) and inactivating *ZMYM3* (ref. <sup>41</sup>) mutations, required for normal *BRCA1* function. Notably, many common drivers, such as *MYC* or *MYCN* amplification and *TP53* deletion or mutation, remain undruggable and, although pathogenic, would not be considered actionable. In several cases, elevated gene expression was the primary feature, suggesting a possible therapy. For example, unusually high *CTLA4* prompted histological re-evaluation, which identified a significant burden of tumor-infiltrating lymphocytes and potential sensitivity to checkpoint inhibitor immunotherapy. Three tumors with elevated wild-type *BCL2* expression were considered potentially sensitive to potent *BCL2* inhibitors, based on expression of the drug target and lack of resistance mutations.

Reconciling the genomic architecture from WGS, with the reading frame, expression abundance and retained functional domains from RNAseq, provided orthogonal validation and the confidence to assign pathogenicity and targetability to novel driver fusions (Fig. 2d). WGS revealed substantial genomic complexity underlying many in-frame fusions. For example, an in-frame potentially targetable *CLIP2-EGFR* fusion arose from a chromothriptic event on chromosome 7 (Fig. 3c). In another example, WGS identified an out-of-frame *FGFR1-ERC1* fusion arising from a complex derivative chromosome linking fragments of chromosomes 1, 8, 12 and 15. RNAseq indicated exon-skipping adjacent to the breakpoint, resulting in an in-frame transcript, driving *FGFR1* expression (Fig. 3d).

**Molecular profiling platforms can support a change of diagnosis.** Another actionable class of variants were those that supported a change of diagnosis. This occurred in 11 cases through the identification of a specific fusion ( $n=4$ ), CNV ( $n=2$ ) or SNV ( $n=5$ ) not identified by standard testing. In two cases, we identified an *EWSR1* fusion supporting a diagnosis of Ewing's sarcoma and, in another, an *EWSR1-ATF1* fusion supporting a diagnosis of gastrointestinal neuroectodermal tumor. In a high-grade glioma, a *KIAA1549-BRAF* fusion suggested a diagnosis of low-grade glioma<sup>42</sup>. In four CNS tumors, WGS identified histone H3 K27M mutations missed by standard testing (immunohistochemical analysis), resulting in a diagnosis of diffuse midline glioma. In one case, the absence of a K27M variant indicated a high-grade glioma diagnosis. One diagnostic challenge, initially classified as an undifferentiated lymphoma, was altered to rhabdoid tumor of small cell type after WGS identified biallelic *SMARCB1* deletion<sup>43</sup>. A high-grade undifferentiated sarcoma harbored chr17p11.2–13.1 amplification and genome-wide instability with multiple chromothriptic events, supporting a change of diagnosis to osteosarcoma. For 76 CNS tumors, we used the DNA methylation MNP classifier<sup>27</sup> to classify tumors into one of more than 80 distinct tumor classes or subclasses. For 53 (69.7%) tumors, there was a strong (>0.9) or weak (0.5–0.9) match to the original diagnosis (Extended Data Fig. 6). For nine (11.8%) tumors, there was a strong or weak match to a different diagnosis; however, for seven of these, the concordance between the histopathological and genomic features of the original diagnosis remained strong. In two cases, the methylation profile suggested a change of diagnosis: a relapsed medulloblastoma reclassified as a radiation-induced glioblastoma, supported by RNAseq changes<sup>44</sup>,

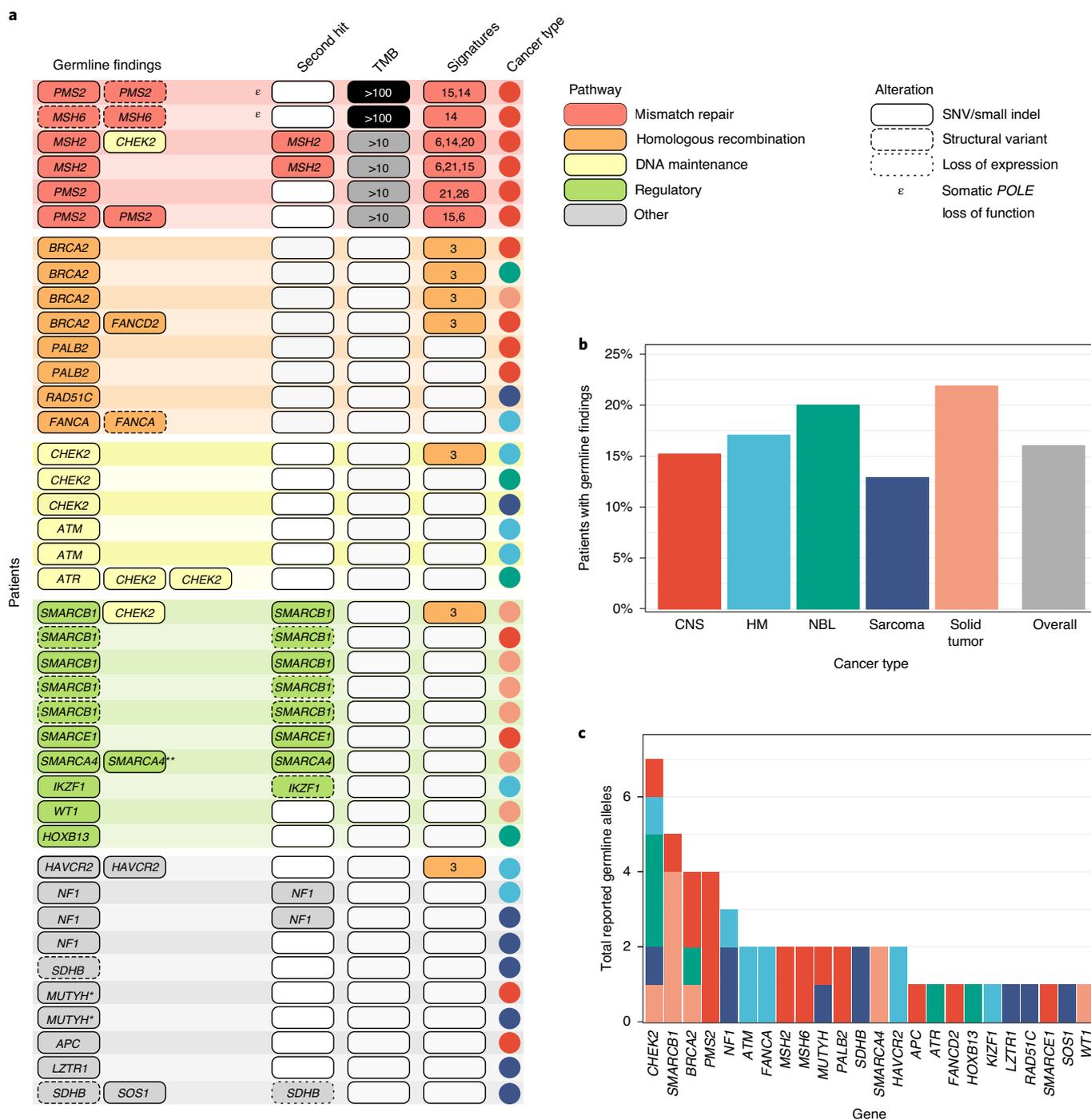
and a recurrent posterior fossa anaplastic ganglioglioma reclassified as an IDH wild-type midline glioblastoma (0.99), despite the presence of a *KIAA1549-BRAF* fusion, normally associated with low-grade gliomas.

**Germline WGS reveals high rates of germline cancer predisposition mutations.** Germline cancer predisposition alleles were identified for 16.2% of patients (40/247), a higher rate than previously described via either whole exome or genome sequencing<sup>45</sup>. These 40 patients harbored 52 pathogenic germline cancer predisposition alleles. In all but two cases, the allele was predicted to result in an increased cancer risk for the patient; the remaining two were heterozygous *MUTYH* loss-of-function variants that were reported owing to their association with an increased lifetime risk of colorectal cancer<sup>46</sup> (Fig. 4). Germline findings were made in all major cancer types, from a rate of 12.9% of patients with sarcoma to 21.9% in solid tumors (Fig. 4b,c). Strikingly, the risk variant was already known to the family in only 14/40 (35%) patients, and the variant was known in that individual in 11/40 (27.5%) patients.

Consistent patterns linked germline pathogenic variants with somatic features of the tumors, such as genome-wide somatic SNV TMB (Extended Data Fig. 7). Five tumors (four CNS and one sarcoma) were hypermutated (10–100 mutations per Mb), and two CNS tumors were ultramutated (>100 mutations per Mb). Six of these seven patients with highly mutated tumors had germline mutations resulting in mismatch repair defects (MMRDs) and a somatic mutation signature linked to MMRD and microsatellite instability (Fig. 4a). Three of these patients were diagnosed with constitutive mismatch repair deficiency based on biallelic germline loss of *MSH6* or *PMS2*, two of which were ultramutated with somatic *POLE* mutation or deletion (Fig. 4a)<sup>47,48</sup>. Review of sequencing data, triggered by the association of *POLE* mutations with signature 14 and the ultramutator phenotype<sup>49</sup>, led to the identification of otherwise missed pathogenic alleles (Extended Data Fig. 8). Across the full cohort, 75% (6/8) of patients with an MMRD somatic signature had a germline pathogenic variant in an MMR gene versus no (0/239) patients lacking an MMRD signature (odds ratio (OR) infinite, 95% confidence interval (CI) > 71, two-sided  $P < 1 \times 10^{-10}$ , Fisher's exact test). Patients with germline pathogenic variants in homologous recombination (HR) pathways often exhibited the associated somatic signature 3 (Fig. 4a and Extended Data Fig. 9). Seventeen percent of patients (4/23) with an HR somatic signature had a germline pathogenic variant in an HR repair gene versus 1.8% of patients (4/224) lacking the signature (OR = 11.3, 95% CI 2.0–66, two-sided  $P = 0.003$ , Fisher's exact test). Somatic second hits were, in certain cases, highly consistent. Germline defects in *SMARC* genes were always associated with a somatic second hit in the same gene (Fig. 4a). Together, these results demonstrate the value of combining germline and somatic WGS for the identification of cancer risk variants.

We investigated why our rate of germline findings was higher than we anticipated<sup>10,45</sup>. Of 40 patients with germline cancer risk findings, four harbored variants in genes not reported in other studies, including *MUTYH*, and five harbored structural and copy number changes that might not be detected by exome-based sequencing. After excluding these, our overall rate of germline risk findings was 11.7% (29/247) of patients. This is not significantly different from the 8.5% previously reported (29/247 versus 95/1120, OR = 1.4, 95% CI 0.89–2.26, two-sided  $P = 0.11$ , Fisher's exact test). This suggests that our enhanced rate of germline findings results from advances in the understanding of pediatric cancer risk, and the enhanced detection of SVs and CNVs by WGS, but does not exclude an effect from the distribution of cancer subtypes and patient ethnicities in our cohort.

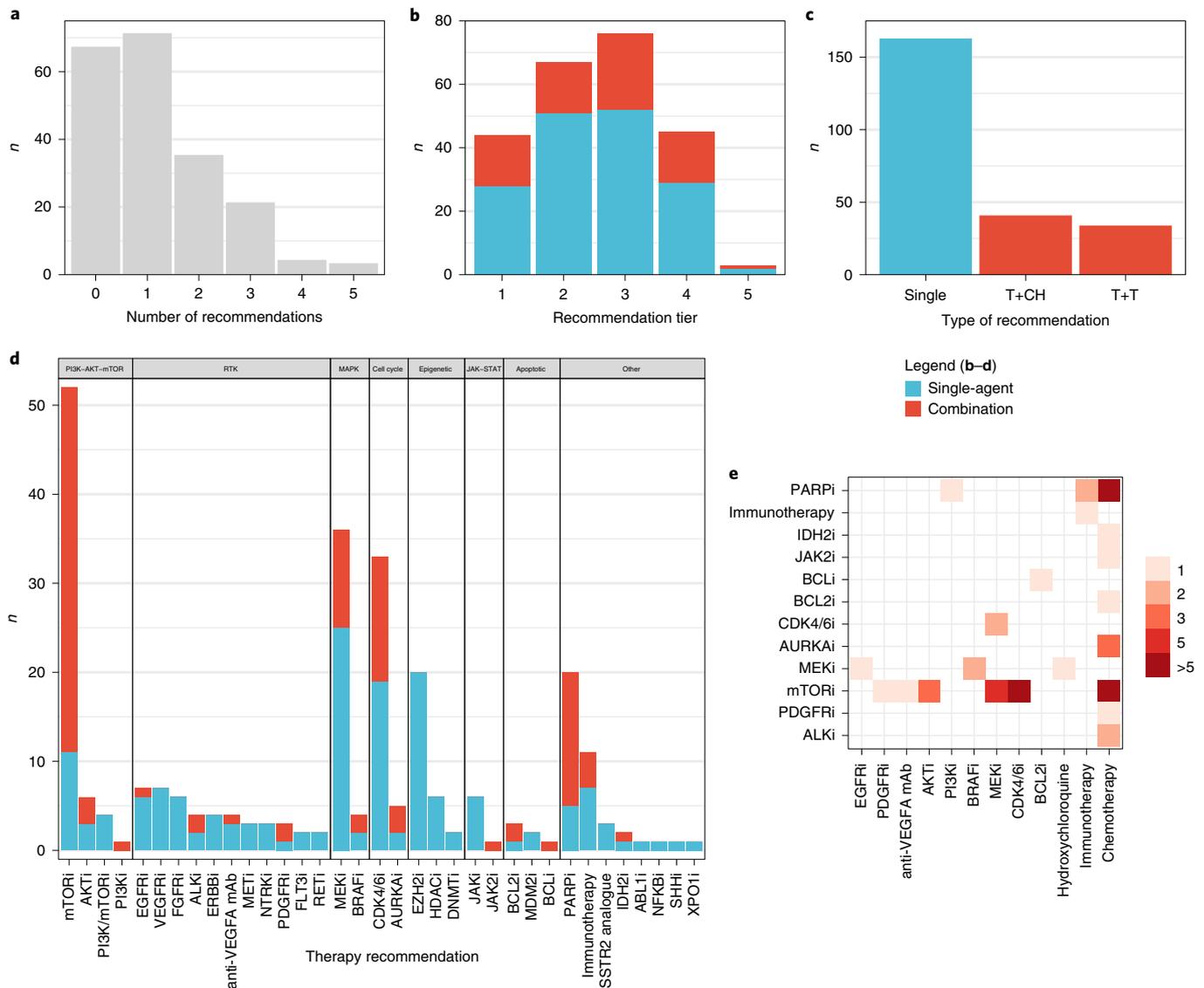
**Overall utility of combined WGS, RNAseq and CNS methylome profiling.** Germline and somatic SNV, indels, CNVs, SVs, TMB and



**Fig. 4 | Integration of WGS, RNAseq and mutation signature analyses reveals a high incidence of germline driver mutations. a**, Reportable cancer risk variants were identified in 40 patients, with each row representing a patient with one or more germline findings, somatic second hits, mutation burden and somatic signatures. Patients are grouped by the predominant cancer risk pathway, colored as per the legend. Cancer type is denoted by colored circles, with colors as in **b**, and the presence of somatic *POLE* loss-of-function mutations denoted by an ε character. **b**, Germline findings were made in all cancer types and **(c)** in a range of cancer predisposition genes. In **c**, allele counts are split by cancer type, following the color scheme in **b**\*, *MUTYH* variants were heterozygous but still reported; \*\*, one reported *SMARCA4* variant was subsequently recategorized to a variant of uncertain significance. HM, hematological malignancy; NBL, neuroblastoma.

signature analysis all contributed to the actionable landscape total of 1,023 individual reportable findings. This included 968 germline and somatic molecular aberrations using WGS and RNAseq and 55 CNS tumor classifications across the cohort (Extended Data Fig. 10). This large number includes instances where the same genomic

event was detected independently by different platforms—compound mutations to different copies of the same gene driving significant effects on gene expression. *TP53* and *SMARCB1* illustrate this, where either gene is affected by combinations of germline and somatic SNVs, deletions and SVs associated with significantly



**Fig. 5 | Comprehensive molecular profiling drives high rates of MTB recommendations.** **a**, The molecular findings from 202 patients were presented at an expert MTB, which returned anywhere from zero to five unique treatment recommendations for each patient. **b**, The frequency of treatment recommendations is shown, from the highest level of supporting evidence to the lowest (Tier 1, clinical evidence in the same cancer; Tier 2, clinical evidence in a different cancer; Tier 3, preclinical evidence in the same cancer; Tier 4, preclinical evidence in a different cancer type; Tier 5, consensus opinion). Single-agent recommendations are shown in blue and combination recommendations in red for **b–d**. **c**, The number of single-agent recommendations, as well as combination recommendations as either two targeted agents (T+T) or a targeted agent with chemotherapy (T+Ch). **d**, The frequency of single-agent and combination drug recommendations made. **e**, For each of the combination treatment recommendations, the frequency of each pair of drugs is shown.

decreased expression. Collectively, 93.7% of patients had at least one reportable aberration (Extended Data Fig. 10a), with a median of three reportable findings per patient (range, 0–13; mean, 4). Overall, 386 (39.9%) of the reportable variants were supported by evidence from both WGS and RNAseq, leaving a striking 582 (60.1%) reportable findings made from either WGS or RNAseq (Extended Data Fig. 10b), reinforcing the complementary nature of these approaches. Notably, outliers of gene expression without associated genomic changes were a major category of reportable findings (242, 25.0%). Ten of 16 samples with no reportable molecular events had low tumor purity estimates (<25%) and CNV profiles consistent with minimal or no tumor in the sample. For the remaining six samples, CNV consistent with tumor was detected, but none was independently actionable.

**Therapy recommendations.** Our national multidisciplinary tumor board (MTB) considered therapeutic recommendations that were recommended only if age-specific safety data were available and if the therapy was accessible via clinical trials, compassionate access or off-label use. Each recommendation was assigned a tier from 1 (highest) to 5 (lowest) as previously described<sup>50</sup> (Fig. 5a, legend).

The MTB was presented 771 molecular aberrations from 201 cases. Of these, 272 variants contributed directly to 237 distinct treatment recommendations. One hundred and thirty-four (67%) patients received at least one therapeutic recommendation, with some receiving up to five (Fig. 5a). Of these, 112 (48%) were Tier 1 or Tier 2 recommendations; 122 (51%) were Tier 3 or Tier 4; and only three (1%) were Tier 5 (Fig. 5b). Most recommendations (70%) were for single-agent therapies. Combination therapies con-

sisted of two targeted agents or a targeted agent and chemotherapy (Fig. 5c). The most common pathway targeted by recommended therapies was the PI3K/AKT/mTOR pathway (20.5%). Therapies targeting RTK (14.6%), the MAPK pathway (13.0%), cell cycle (12.3%) and epigenetic pathways, including the SWI/SNF/PRC2 chromatin remodeling complex (9.1%), were the next most common (Fig. 5d). The most recommended drug combination was an mTOR inhibitor plus chemotherapy (26%), followed by PARP inhibitor plus chemotherapy (16%) and mTOR inhibitor plus CDK4/6 inhibitor (16%) (Fig. 5e).

**Clinical response to treatment.** Of the 134 patients who received an MTB recommendation, 43 (32%) had received the recommended agent(s) at the time of data census. Five patients treated for fewer than 4 weeks were excluded from response assessment. Thus, 38 patients were included in the follow-up of treatment (Fig. 6a), including three patients with no evidence of disease at the start of treatment, two of whom relapsed at 8 and 21 weeks, respectively, and one who remained disease free at 51 weeks. Thirty-five patients were evaluable for response. The best overall responses included four (11.4%) complete responses (CRs), seven (20%) partial responses (PRs), 14 (40%) stable disease (SD) and ten (28.6%) progressive disease (PD). There were similar proportions of Tier 1 and Tier 2 recommendations in patients with a CR/PR (45%), SD (50%) or PD (50%) (Fig. 6b). Some SD responses were remarkably durable, at least 24 weeks in six patients, and three of these patients continued on the recommended treatment for more than 12 months. Of the 28 patients with CNS or solid tumors with measurable disease (Fig. 6c,d), 17 (61%) had a measurable reduction in tumor burden. This was not correlated with the tier of treatment recommendation.

## Discussion

Here we present the molecular findings from the first 247 patients recruited onto Zero Childhood Cancer, a prospective, national pediatric precision medicine program testing the feasibility of WGS, RNAseq and CNS methylome profiling for the clinical management of pediatric patients in real time. We demonstrate the utility of an integrative, comprehensive approach, with each genomic platform, independently and in combination, contributing to the detection of pathogenic variants in 94% of samples. This underpinned the identification of therapeutic targets in 71% of samples, which was higher than anticipated based on earlier experiences<sup>11–16,50–52</sup>. Zero Childhood Cancer employs WGS to identify reportable germline and tumor variants, including single- and multi-exonic CNVs, copy-neutral inversions and other SVs, as well as mutational signatures and TMB. Precise WGS resolution of SV break-points gives increased confidence in identifying short CNV and loss-of-heterozygosity regions that would be challenging to detect by looking only for changes in depth of coverage, particularly in low-purity tumors. In addition to maximizing the number of potential treatment recommendations, multiple platforms corroborating a variant reduces the need for additional validation tests, particularly for novel mutations, which would otherwise be required at least in all instances where a clinical action might result.

We analyzed RNAseq not only to identify fusion genes but also to interrogate abundance, which contributes to the total number of reported aberrations<sup>7</sup>. It remains unclear which gene expression outliers represent legitimate therapeutic targets, particularly in the absence of SNVs, SVs and CNVs. Pathway resources, such as KEGG<sup>53</sup> or Reactome<sup>54,55</sup>, might not be sufficiently specific for pediatric oncogenic pathway activation to be useful in this context. However, data from adult cancer studies suggest that selecting targeted drugs on the basis of expression aberrations can increase response rates<sup>26</sup>. Another study of high-risk pediatric patients showed that gene expression outliers identified potential therapeutic targets in at least 70% of cases<sup>24</sup>. Determining which expression profiles represent true biomarkers of clinical response remains an important objective.

The rate of pathogenic germline variants—16.2% of the cohort—was higher than we anticipated, based on studies in other populations<sup>45</sup>. This is the first estimate of the prevalence of germline mutations driving high-risk pediatric cancer in the Australian population based on tumor–germline WGS. WGS was critical for the assignment of pathogenicity to several variants, notably where somatic features, such as mutation signatures, TMB and second-hit mutations, were instrumental in identifying the underlying germline mutation(s). Targeted sequencing approaches find it challenging to identify copy-neutral structural variants and small CNVs and less accurately detect mutation signatures and TMB. We propose that WGS is the most comprehensive approach to detect hereditary cancer susceptibility.

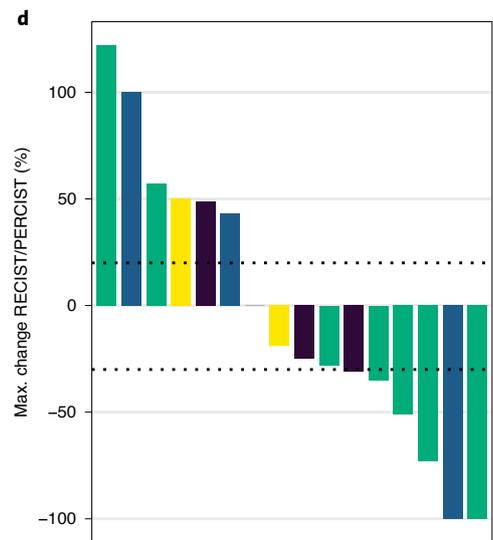
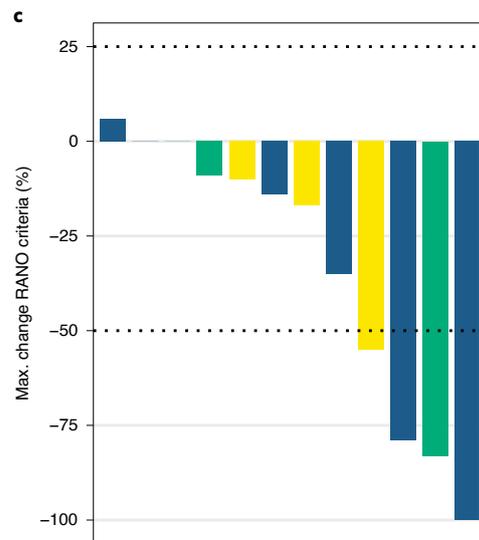
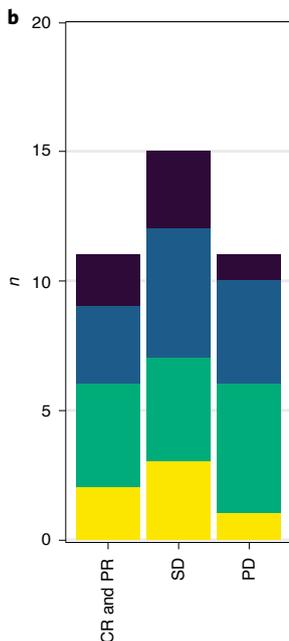
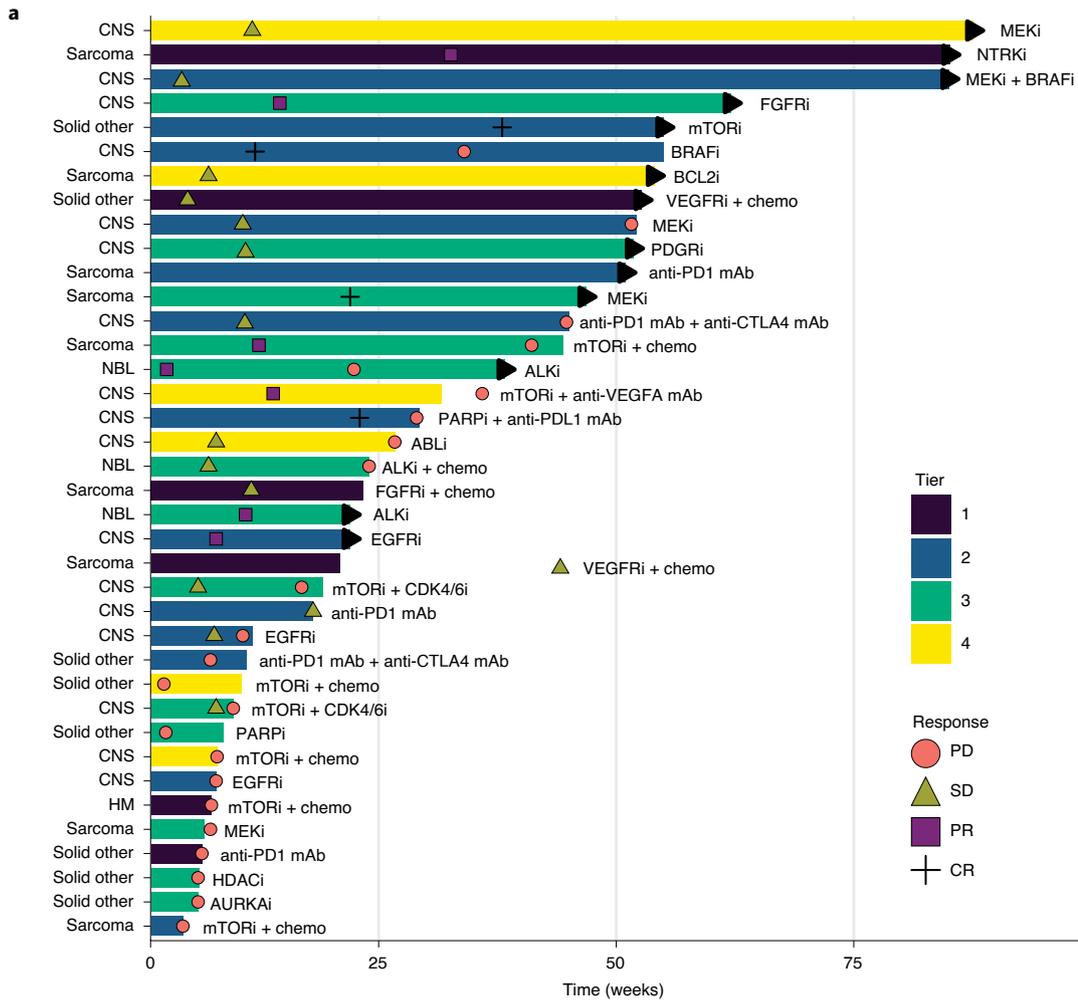
The discovery of all potentially relevant molecular data in an individual tumor is, we contend, essential for the clinical application of such evidence by MTBs and clinicians. The genomic complexity of high-risk cancer is such that there is likely to be more than one genomic driving feature that would require targeting for effective treatment. Appreciating the full targetable landscape permits the MTB and clinicians to consider how targeted therapies might be used, alone or in combination. Just as important is the opportunity for deeper analysis of data from programs such as Zero Childhood Cancer to extend understanding of the biology and therapeutic susceptibilities of high-risk childhood cancer. Although the use of methylation data was limited to CNS tumor subtype classification<sup>27</sup>, increasing the diagnostic confidence in 53 cases and supporting a change of diagnosis in two cases, analyzing the tumor methylome will likely have utility for many other pediatric cancers. Tumour-specific epigenetic changes likely explain features of pediatric tumors, such as the aberrant gene expression profiles.

Although there are gene–drug interactions that can and are being tested now in important trials, such as MATCH<sup>17</sup>, there remain many gaps in the understanding of which molecular biomarkers accurately predict drug responses. For some very promising drugs, such as inhibitors of the anti-apoptotic BCL2 family, the genomic basis on which a drug recommendation might be made is essentially unknown and still relies primarily on the cancer type, potentially specific expression profiles<sup>36,57</sup> and the absence of resistance mechanisms<sup>58,59</sup>. Moreover, the genomic determinants of responses to immunotherapies beyond TMB are still being unraveled. The combined WGS and RNAseq data sets of large-scale programs such

**Fig. 6 | Clinical responses to targeted therapies are observed across all tiers of recommendations.** **a**, Swimmer plot of 38 patients who had received a recommended therapy for at least 4 weeks. x axis is time on treatment (weeks). The color of the bars indicates the MTB tier assigned to the recommendation (see main text). The responses are indicated by the symbols. Where a symbol is shown beyond the end of a bar, this indicates that the assessment occurred after the therapy was ceased. The arrow indicates that the patient continued with treatment. **b**, The stacked bars show the total number of patients for whom the best response fell in the indicated category. The colors of the stacked bars indicate the MTB tier assigned to the recommendation. **c**, Waterfall plot of CNS tumors assessed by RANO criteria in patients receiving recommended therapies. The dotted lines are at –50%, a score below which represents an objective response, and 25%, a score above which represents tumor progression. The color indicates the MTB tier assigned to the recommendation. **d**, Waterfall plot of extracranial solid tumors assessed by RECIST or PRECIST in patients receiving recommended therapies. The dotted lines are at –30%, a score below which represents an objective response, and 20%, a score above which represents tumor progression. The color indicates the MTB tier assigned to the recommendation. HM, hematological malignancy; NBL, neuroblastoma.

as Zero Childhood Cancer will be a most important resource to drive discovery of drugs that target pathways important in childhood cancers and the application of drugs not currently available to children.

We have shown that a comprehensive molecular tumor profile comprising WGS, RNAseq and DNA methylation analysis led to a high rate of treatment recommendations. However, our MTB took a conservative approach, with most potentially targetable



molecular aberrations not being reported. Of those that were, 99% of recommendations fell into tiers requiring at least some degree of relevant preclinical or clinical data to support the recommendation, and only 1% relied on consensus opinion. To date, one third of patients with a treatment recommendation have received the recommended therapy, with a proportion showing evidence of clinical benefit, including 31% with an objective response and others experiencing a period of prolonged SD. Sequencing is a powerful method to identify genomic driver lesions and expand the understanding of cancer biology. It is a substantially greater challenge to translate genomic information into improved patient outcomes. Encouragingly, the responses we have observed, although a small number, compare favorably, for example, with patients treated on pediatric phase 1 trials, unselected by tumor molecular profiling, who have been reported historically to have a response rate of less than 10%<sup>60</sup>. Moreover, responses occurred across all tiers of therapy and across multiple different treatments and pathways targeted, suggesting that there is more to learn about the use of WGS and transcriptome sequencing in the personalized approach to therapy.

In high-risk childhood cancer, WGS and RNAseq both contributed significantly to the identification of the full spectrum of germline and somatic driver variants and to clinical therapeutic recommendations. Whether this holds true in standard-risk pediatric cancers remains to be determined, but it seems likely that the optimal molecular profiling approach might vary by disease type and stage. Our experience demonstrates that WGS and RNAseq offer the best opportunity to identify targetable driver genomic lesions in high-risk pediatric cancer.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-020-1072-4>.

Received: 4 February 2020; Accepted: 20 August 2020;

Published online: 05 October 2020

### References

- Steliarova-Foucher, E. et al. Changing geographical patterns and trends in cancer incidence in children and adolescents in Europe, 1991–2010 (Automated Childhood Cancer Information System): a population-based study. *Lancet Oncol.* **19**, 1159–1169 (2018).
- Gu, Z. et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat. Genet.* **51**, 296–307 (2019).
- Stewart, E. et al. Identification of therapeutic targets in rhabdomyosarcoma through integrated genomic, epigenomic, and proteomic analyses. *Cancer Cell* **34**, e419 (2018).
- Northcott, P. A. et al. The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).
- Pugh, T. J. et al. The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* **45**, 279–284 (2013).
- Berger, M. F. & Mardis, E. R. The emerging clinical relevance of genomics in cancer medicine. *Nat. Rev. Clin. Oncol.* **15**, 353–365 (2018).
- Jones, D. T. W. et al. Molecular characteristics and therapeutic vulnerabilities across paediatric solid tumours. *Nat. Rev. Cancer* **19**, 420–438 (2019).
- Kurzrock, R. & Giles, F. J. Precision oncology for patients with advanced cancer: the challenges of malignant snowflakes. *Cell Cycle* **14**, 2219–2221 (2015).
- Ma, X. et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371–376 (2018).
- Grobner, S. N. et al. The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).
- Parsons, D. W. et al. Diagnostic yield of clinical tumor and germline whole-exome sequencing for children with solid tumors. *JAMA Oncol.* **2**, 616–624 (2016).
- Chang, W. et al. Multidimensional clinomics for precision therapy of children and adolescent young adults with relapsed and refractory cancer: a report from the center for cancer research. *Clin. Cancer Res.* **22**, 3810–3820 (2016).
- Harttrampf, A. C. et al. Molecular screening for cancer treatment optimization (MOSCATO-01) in pediatric patients: a single-institutional prospective molecular stratification trial. *Clin. Cancer Res.* **23**, 6101–6112 (2017).
- Khater, F. et al. Molecular profiling of hard-to-treat childhood and adolescent cancers. *JAMA Netw. Open* **2**, e192906 (2019).
- Mody, R. J. et al. Integrative clinical sequencing in the management of refractory or relapsed cancer in youth. *JAMA* **314**, 913–925 (2015).
- Worst, B. C. et al. Next-generation personalised medicine for high-risk paediatric cancer patients - the INFORM pilot study. *Eur. J. Cancer* **65**, 91–101 (2016).
- Allen, C. E. et al. Target and agent prioritization for the children's oncology group-National Cancer Institute Pediatric MATCH trial. *J. Natl Cancer Inst.* **109**, djw274 (2017).
- Parsons, D. W. et al. Identification of targetable molecular alterations in the NCI-COG Pediatric MATCH trial. *J. Clin. Oncol.* **37**, 10011–10011 (2019).
- Priestley, P. et al. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210–216 (2019).
- Campbell, P. J., Getz, G., Stuart, J. M., Korbel, J. O. & Stein, L. D. -cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
- Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Campbell, B. B. et al. Comprehensive analysis of hypermutation in human cancer. *Cell* **171**, e1010 (2017).
- Davies, H. et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **23**, 517–525 (2017).
- Vaske, O. M. et al. Comparative tumor RNA sequencing analysis for difficult-to-treat pediatric and young adult patients with cancer. *JAMA Netw. Open* **2**, e1913968 (2019).
- McFarland, J. M. et al. Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nat. Commun.* **9**, 4610 (2018).
- Rodon, J. et al. Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial. *Nat. Med.* **25**, 751–758 (2019).
- Capper, D. et al. DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).
- Van Allen, E. M., Wagle, N. & Levy, M. A. Clinical analysis and interpretation of cancer genome data. *J. Clin. Oncol.* **31**, 1825–1833 (2013).
- Hoskinson, D. C., Dubuc, A. M. & Mason-Suares, H. The current state of clinical interpretation of sequence variants. *Curr. Opin. Genet. Dev.* **42**, 33–39 (2017).
- Griffith, M. et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.* **49**, 170–174 (2017).
- Drilon, A. et al. Efficacy of larotrectinib in TRK fusion-positive cancers in adults and children. *N. Engl. J. Med.* **378**, 731–739 (2018).
- Ziegler, D. S. et al. Brief report: potent clinical and radiological response to larotrectinib in TRK fusion-driven high-grade glioma. *Br. J. Cancer* **119**, 693–696 (2018).
- Bell, R. J. et al. Understanding TERT promoter mutations: a common path to immortality. *Mol. Cancer Res.* **14**, 315–323 (2016).
- Rheinbay, E. et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **578**, 102–111 (2020).
- Jouhilahti, E. M. et al. The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. *Development* **143**, 3459–3469 (2016).
- Gough, S. M., Slape, C. I. & Aplan, P. D. NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights. *Blood* **118**, 6247–6257 (2011).
- Stathis, A. & Bertoni, F. BET proteins as targets for anticancer treatment. *Cancer Discov.* **8**, 24 (2018).
- Zimmerman, M. W. et al. MYC drives a subset of high-risk pediatric neuroblastomas and is activated through mechanisms including enhancer hijacking and focal enhancer amplification. *Cancer Discov.* **8**, 320–335 (2018).
- Cameron, D. L., et al. GRIDSS, PURPLE, LINX: unscrambling the tumor genome via integrated analysis of structural variation and copy number. Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/781013v1> (2019).
- Bailey, M. L. et al. Glioblastoma cells containing mutations in the cohesin component STAG2 are sensitive to PARP inhibition. *Mol. Cancer Ther.* **13**, 724–732 (2014).
- Leung, J. W. et al. ZMYM3 regulates BRCA1 localization at damaged chromatin to promote DNA repair. *Genes Dev.* **31**, 260–274 (2017).

42. Jones, D. T. W. et al. Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas. *Cancer Res.* **68**, 8673–8677 (2008).
  43. Kohashi, K. et al. Reclassification of rhabdoid tumor and pediatric undifferentiated/unclassified sarcoma with complete loss of SMARCB1/INI1 protein expression: three subtypes of rhabdoid tumor according to their histological features. *Mod. Pathol.* **29**, 1232–1242 (2016).
  44. Donson, A. M. et al. Unique molecular characteristics of radiation-induced glioblastoma. *J. Neuropathol. Exp. Neurol.* **66**, 740–749 (2007).
  45. Zhang, J., Nichols, K. E. & Downing, J. R. Germline mutations in predisposition genes in pediatric cancer. *N. Engl. J. Med.* **374**, 1391 (2016).
  46. Nielsen, M. et al. Survival of MUTYH-associated polyposis patients with colorectal cancer and matched control colorectal cancer patients. *J. Natl Cancer Inst.* **102**, 1724–1730 (2010).
  47. Shlien, A. et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat. Genet.* **47**, 257–262 (2015).
  48. Wimmer, K. & Eitzler, J. Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg? *Hum. Genet.* **124**, 105–122 (2008).
  49. Haradhvala, N. J. et al. Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. *Nat. Commun.* **9**, 1746 (2018).
  50. Harris, M. H. et al. Multicenter feasibility study of tumor molecular profiling to inform therapeutic decisions in advanced pediatric solid tumors: the individualized cancer therapy (iCat) study. *JAMA Oncol.* **2**, 608–615 (2016).
  51. Marks, L. J. et al. Precision medicine in children and young adults with hematologic malignancies and blood disorders: the Columbia University experience. *Front. Pediatr.* **5**, 265 (2017).
  52. Pincez, T., et al. Feasibility and clinical integration of molecular profiling for target identification in pediatric solid tumors. *Pediatr. Blood Cancer* **64** <https://doi.org/10.1002/pbc.26365> (2017).
  53. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
  54. Fabregat, A. et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* **44**, D481–D487 (2016).
  55. Jassal, B. et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).
  56. Khaw, S. L. et al. Venetoclax responses of pediatric ALL xenografts reveal sensitivity of MLL-rearranged leukemia. *Blood* **128**, 1382–1395 (2016).
  57. Villalobos-Ortiz, M., Ryan, J., Mashaka, T. N., Opferman, J. T. & Letai, A. BH3 profiling discriminates on-target small molecule BH3 mimetics from putative mimetics. *Cell Death Differ.* **27**, 999–1007 (2019).
  58. Blombery, P. et al. Characterization of a novel venetoclax resistance mutation (BCL2 Phe104Ile) observed in follicular lymphoma. *Br. J. Haematol.* **186**, e188–e191 (2019).
  59. Guieze, R. et al. Mitochondrial reprogramming underlies resistance to BCL-2 inhibition in lymphoid malignancies. *Cancer Cell* **36**, e313 (2019).
  60. Lee, D. P., Skolnik, J. M. & Adamson, P. C. Pediatric phase I trials in oncology: an analysis of study conduct efficiency. *J. Clin. Oncol.* **23**, 8431–8441 (2005).
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- © The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

## Methods

Information about the methods used in this study are also available in the Life Sciences Reporting Summary

**Patients and samples.** The pilot study (TARGET) was opened at the two children's hospitals in Sydney, Australia (Sydney Children's Hospital, Randwick and The Children's Hospital at Westmead) from June 2015 to October 2017 and was approved by the Sydney Children's Hospitals Network Human Research Ethics Committee (LNR/14/SCH/497). The PRISM clinical trial (NCT03336931), conducted as part of the Australian Zero Childhood Cancer Precision Medicine Program, was opened at all eight pediatric oncology centers around Australia (Sydney Children's Hospital Randwick, Sydney; The Children's Hospital at Westmead, Sydney; John Hunter Hospital, Newcastle; Queensland Children's Hospital, Brisbane; Royal Children's Hospital, Melbourne; Monash Children's Hospital, Melbourne; Women's & Children's Hospital, Adelaide; and Perth Children's Hospital, Perth) from September 2017 and was approved by the Hunter New England Human Research Ethics Committee of the Hunter New England Local Health District (reference no. 17/02/015/4.06) and the New South Wales Human Research Ethics Committee (reference no. HREC/17/HNE/29). Informed consent for each participant was provided by parents/legal guardian for participants under the age of 18 years and by the participants who were over the age of 18 years. Data sets from 47 patients enrolled on the TARGET pilot study (submitted manuscript) and 200 patients enrolled to June 2019 on the PRISM national clinical trial are included in this manuscript. Five patients had two tumors sequenced. Patients aged 21 years or younger with suspected or confirmed diagnosis of a very rare or high-risk malignancy (at diagnosis, relapse or refractory disease), defined as expected probability of survival of less than 30%, could be consented and registered on the study. Patients older than 21 years with suspected or confirmed high-risk pediatric-type cancers could also be registered with approval from the study chair. After trial registration, patient tumor samples were delivered to the Children's Cancer Institute in Sydney for processing. A patient was deemed eligible for enrolment when all criteria were satisfied: high-risk cancer diagnosis clinically confirmed and when both a tumor and a germline sample were received at the Children's Cancer Institute. Tumor tissue (solid tissue or tumor cells isolated from bone marrow aspirate or peripheral blood) was fresh, fresh frozen or cryopreserved upon receipt. When patients were recruited at relapse or after the onset of refractory disease, their original diagnostic samples were not obtained or sequenced. In those patients who had undergone a bone marrow transplantation during their treatment, both the patient germline (usually a skin punch biopsy or a sample from a previous clinical time point before bone marrow transplantation) and the donor germline (usually peripheral blood from the patient) were sequenced specifically to distinguish tumor-derived somatic variants from donor variants. The average turnaround time from enrollment to return of data to the MTB was 7.5 weeks.

**WGS data analysis.** WGS was conducted at the Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research (Australia), using the Illumina HiSeq X Ten platform with a paired-end read length of 150 bases. Sequencing libraries were prepared from more than 1 µg of DNA using either TruSeq Nano DNA HT Sample Prep Kit (Illumina) or KAPA PCR-Free v2.1 (Roche). All germline samples were sequenced to a minimum mean coverage of 30×, the first five tumor samples to a minimum depth of 60× and the remaining tumor samples to a minimum mean coverage of 90×.

Raw fastq files were aligned to the hs37d5 reference genome using BWA-MEM (v0.17.10-r789)<sup>61</sup>, with resulting BAM files merged and duplicate reads marked using Novosort (v1.03.01; default settings), and read alignment improved using GATK Indel Realignment (v3.3)<sup>62</sup>. Germline SNVs and short (<50-bp) indels were identified using GATK HaplotypeCaller, GenotypeVCFs and VQSR (all v3.3)<sup>63</sup>, annotated with VEP (v87)<sup>64</sup>, converted into a GEMINI (v0.11.0)<sup>65</sup> database and imported into Seave<sup>65</sup> for filtration and prioritization. Somatic SNVs and short indels (<50 bp) were identified using Strelka (v2.0.17)<sup>66</sup> and filtered using these criteria:  $QSS \geq 10$  or  $QSI \geq 10$ ;  $tumourVAF * QS \geq 1.3$ ; from chromosomes 1–22, X, Y, MT; in at most three (of 2,570) individuals in the similarly processed MGRB cohort;<sup>67</sup> if not in the platinum genome's high-confidence region (<https://github.com/Illumina/PlatinumGenomes>) then with  $tumourVAF > 0.1$  and  $QSS > 20$  or  $QSI > 20$  or in a curated hotspot white list;  $NT = ref$ ; and  $tumourVAF > 3 * normalVAF$ . Even after filtering on variants where  $NT = ref$ —that is, with germline genotype '0/0'—some variants have germlineVAF > 0; thus, this final filter retains only variants with excess signal in the tumor, allowing for some tumor in normal contamination. We benchmarked this approach (see below and Supplementary Table 4). Mutation signatures were analyzed using deconstructSigs (v1.8.0)<sup>68</sup> and the Catalogue of Somatic Mutations in Cancer (COSMIC) mutation signatures (v2)<sup>31</sup>. Somatic variants were annotated using SnpEff (v4.3t)<sup>69</sup> and imported into the in-house Glooe platform for filtration and prioritization. Tumor purity, ploidy and somatic copy number alteration were identified using PURPLE (v2.39)<sup>39</sup>, and somatic SVs were identified using GRIDSS (v2.7.2)<sup>70</sup> and then annotated using Ensembl genes. LINX (v1.7) was used to visualize SV clusters and derivative chromosomes. The read evidence supporting all candidate variants was manually inspected using IGV (v2.6.2)<sup>71</sup> before reporting. *Refjyr* (v1.17.8), an

in-house software, was used to create the analysis workflows for running on the DNAnexus platform (<https://www.dnanexus.com/>). Data analysis was performed using R (v3.5.3) (<https://www.R-project.org/>) via RStudio (v1.2.1335) (<http://www.rstudio.com/>) and plotted using ggplot2 (v3.3.2)<sup>72</sup>.

**Strelka somatic SNV and short indel benchmarking.** Using WGS data from the COLO829 matched cancer and normal cell lines data (sequenced by Hartwig Medical Foundation), we ran Strelka (v2.0.17) using default settings and our custom filter (see above) with and without restricting to the GIAB high-confidence regions. We evaluated variant calling performance using bedtools *isec* (v2.28.0)<sup>73</sup> and a truth set of somatic SNVs and short indels<sup>74</sup> from the European Genome-phenome Archive (EGAS00001001385).

**Scan for noncoding promoter variation.** High-confidence promoter regions were defined as the intersection of the PromCore regions previously reported<sup>34</sup>, with the Genome in a Bottle v2.0 high-confidence regions ([ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/release/genome-stratifications/v2.0/GRCh37/union/GRCh37\\_notinalldifficultregions.bed.gz](ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/release/genome-stratifications/v2.0/GRCh37/union/GRCh37_notinalldifficultregions.bed.gz)) and the MGRB tier1+2 region<sup>67</sup>. Overlapping regions were merged without respect to strand using bedtools *merge*<sup>73</sup>, and regions <3 bp were removed. These high-confidence genome regions were then intersected with the promoter regions of our most aberrantly expressed genes (Fig. 2c) to identify eight candidate genes with both a high degree of aberrant expression and well-sequenced promoters.

We performed NBR<sup>34</sup> by training the background mutation model on genome-wide high-confidence promoter regions but testing only the promoters of our eight candidate genes for excess mutation burden. To reduce the influence of false-positive variants and hypermutators on results, variant data were filtered to include only SNVs in patients with fewer than ten mutations per Mb. Multiple testing was controlled by the Benjamini–Hochberg procedure. All noncoding candidate variants were manually inspected using IGV (v2.6.2)<sup>71</sup>.

**Splicing analysis.** To identify the effect of variants on splicing, we developed Introne, which combines prediction scores from SpliceAI (v1.3.1)<sup>75</sup>, MMSplice (v2.1.0)<sup>76</sup>, dbSNV (v1.1)<sup>77</sup>, Branchpointer (v1.3.1)<sup>78</sup> and SPIDEX (v1.0)<sup>79</sup>. Introne (v0.5.1; P.S., R.L. Davis, V.G., M.W., C.M., et al., manuscript in preparation) used machine learning to integrate these individual scores, trained on ~1,000 variants with experimental validation. Rare coding or intronic germline or somatic variants were evaluated for splicing altering potential, with any score greater than 0.5 for any gain or loss of splicing potential manually assessed using IGV (v2.6.2)<sup>71</sup> and Sashimi plots generated from RNAseq data using eggshimi (v0.4.0)<sup>80</sup>.

**Variant filtration.** Germline variants were restricted to a list of cancer predisposition genes curated from literature (Supplementary Table 5) and prioritized by those that were rare in controls (<1% frequency in the Exome Aggregation Consortium (ExAC) database<sup>81</sup>, gnomAD<sup>82</sup>, MGRB<sup>67</sup>, Exome Sequencing Project (ESP)<sup>83</sup> and 1000 Genomes<sup>84</sup>), were previously annotated as likely pathogenic or pathogenic in ClinVar or were novel loss-of-function variants in tumor suppressor genes. In silico prediction tools, including CADD (v1.3)<sup>85</sup>, PolyPhen2 (v2.2.2)<sup>86</sup>, SIFT (v5.0.2)<sup>87</sup>, PROVEAN (v1.1)<sup>88</sup>, FATHMM (via dbNSFP v2.9)<sup>89</sup>, MetaSVM (v1.0)<sup>90</sup> and MetaLR (v1.0)<sup>91</sup>, were used to support the pathogenicity assessment of missense variants using a consensus approach of at least four of seven in silico predictors. Somatic variants were restricted to a curated list of somatic cancer genes and, in addition to the filtering rules defined for germline variants, were prioritized if they were previously reported in somatic cancer databases, including the COSMIC variants<sup>32</sup>, the Cancer Gene Census<sup>93</sup> and the PeCan data portal (<https://pecan.stjude.org/>), with versions updated throughout the project.

**Whole transcriptome data analysis.** Whole transcriptome RNAseq was conducted at the Murdoch Children's Research Institute (Australia) and performed with the TruSeq Stranded mRNA Preparation Kit. RNA sequencing was performed on 228 samples; 24 samples were not attempted because there was insufficient RNA (<400 ng) or the RNA quality was low (RNA integrity number < 5.0). Libraries were pooled, and sequencing runs were performed in paired-end mode using the Illumina HiSeq 4000 platform generating at least 40 million reads per sample ( $n = 47$ ) or the NextSeq 500 platform generating at least 80 million reads per sample ( $n = 181$ ). Initial RNAseq quality control (QC) checks included an evaluation of GC and per-base sequence content using FastQC (v0.11.5). All samples that passed initial QC were aligned to the human genome assembly (build hg19) using the STAR (v2.5)<sup>94</sup> two-pass method with quantMode parameters set to TranscriptomeSAM for alignments translated into transcript coordinates. Alignments were sorted with SAMTools (v1.3.1), duplicates were marked with Picard Tools (v2.4.1), reads were split and trimmed and mapping qualities were reassigned with the Genome Analysis Toolkit (v3.6) using the methods SplitNCigarReads and ReassignOneMappingQuality, respectively. Post-alignment QC required at least 70% of reads to be uniquely aligned, assessed using STAR alignment statistics (100% of samples passed). Raw gene counts, transcripts per kilobase million (TPM), fragments per kilobase million (FPKM) and isoform expression values were calculated using RSEM (v1.2.31)<sup>61</sup>. All RNAseq expression

values are represented as TPM. For identification of outlier expressed genes, the entire cohort was combined, and, using both differential expression (DE) analysis and  $z$  score statistical analysis, each patient's raw counts for DE and TPM for  $z$  score were compared against the cohort. The fold change,  $P$  value and  $z$  score were assessed to determine significance as well as the distribution of the gene in the cohort. If a gene had a  $z$  score greater than 2 and a fold change greater than 2 for overexpressed genes, or a  $z$  score less than  $-2$  and a fold change less than  $-2$  for underexpressed genes, then the gene's distribution in the cohort was assessed to determine if it was a true outlier event. Variants were called on RNA with GATK HaplotypeCaller (v3.6), and ANNOVAR (v20190929)<sup>95</sup> was used for variant annotation. A personalized in-house Python script was developed to extract variant allele frequency, filter out lowly expressed genes and integrate with somatic mutations identified from WGS single-nucleotide polymorphism and indel calls. Fusions were identified using three methods: STAR-Fusion (v1.3.1)<sup>96</sup>, JAFFA (v1.09)<sup>97</sup> and Arriba (v1.1.0; <https://github.com/suhrig/arriba/>). Fusions were analyzed further if identified by a minimum of two algorithms, or by WGS, and one of the gene partners was in our curated list of known fusion genes. All fusions were automatically annotated as high, medium or low confidence and whether they were in-frame or not. High-confidence fusions had evidence of reads spanning the breakpoint (spanning reads) and reads covering both sides of the breakpoint (spanning pairs), with the breakpoint aligning to an exon–exon boundary; medium-confidence fusions had only spanning reads, with the breakpoint aligning to an exon–exon boundary; and low-confidence fusions did not align to an exon–exon boundary. Because Arriba can identify duplications, deletions or inversions, we manually validated these types of SVs identified by WGS to provide an extra level of evidence and support. The read evidence supporting all candidate fusions and SVs were manually inspected using IGV (v2.6.2)<sup>71</sup> before reporting.

**Methylation data analysis.** CNS tumors that had a minimum of 100 ng of DNA were sent for methylation analysis on the EPIC 850K array conducted at the Australian Genome Research Facility (Australia). All samples that passed QC were then processed through the MNP classifier (online; versions might have changed over time)<sup>27</sup> to classify the patient's sample into more than 80 tumor classes and subclasses as well as to obtain the methylation status of the MGMT promoter. CNV profiles from WGS and methylation arrays were compared to check for inconsistencies. The methylation profiling component of Zero Childhood Cancer started after PRISM was launched and was performed on 76 of predominantly the most recent samples.

**Data integration and visualization.** Multi-omic data were integrated and visualized using our in-house interpretation platform Glooe. The platform collates molecular aberrations of all types into a relational database. Each patient is presented as a multi-tab Google spreadsheet highlighting different molecular aberration types, including clinical summaries. Where possible, data types are integrated: in the germline or somatic SNV and indel tables, variants are annotated by their gene-level copy number (including loss of heterozygosity) and gene expression levels; in the germline or somatic CNV tables, gene expression data are integrated. Algorithms reorder aberrations based on secondary mappings listed in the WGS data analysis section. A variant curation interface allows variant-level annotations to be tracked in a database and recalled for future reference and cohort analysis. QC data, plots and data visualization, including genome-wide circos, mutation signatures and CNV profiles, are also displayed in Glooe (versions updated throughout the project).

**Variant curation.** A national variant curation team was established to identify and classify reportable molecular events, including germline and somatic genetic changes, mutation signatures, gene expression, classification results and gene methylation. A pathogenicity score was then assigned to each variant: C1, Benign; C2, Likely Benign; C3, Variant of Unknown Significance; C4, Likely Pathogenic; and C5, Pathogenic, in accordance with published guidelines<sup>29</sup>. Variants that were classified as either C4 or C5 were then deemed reportable to the MTB. Each variant was further curated as a driver mutation and whether it is targetable (linked to a potential change in treatment), diagnostic, prognostic or a known germline cancer predisposition variant. Genes with outlying expression levels (typically  $z$  score  $> 2$  or  $< -2$  as standalone changes or  $> 1.5$  and  $< -1.5$  if supported by expression differences in related genes/pathways) were reported if they helped inform tumor diagnosis or were in genes with known targetable drugs that would have been reported if the gene were amplified or deleted. Reportable variants linked to known anti-cancer drugs with the correct mode of mutation (for example, amplification, in-frame fusions retaining kinase domains or overexpression for RTKs) were established in the curation meetings through extensive literature search and in-house expertise.

**Therapeutic recommendations and response to treatment.** A national MTB meeting occurred every 2 weeks using a Health Insurance Portability and Accountability Act-compliant video-conferencing software and was attended by oncologists, pathologists, geneticists, basic scientists, bioinformaticians and study managers. A templated report with tier-ranked recommendations was issued

thereafter: Tier 1, clinical evidence in the same cancer; Tier 2, clinical evidence in a different cancer; Tier 3, preclinical evidence in the same cancer; Tier 4, preclinical evidence in a different cancer type; and Tier 5, consensus opinion. The treating clinician made the final treatment decisions, including consideration for treatments other than the MTB recommendations. All information distribution was compliant with the Australian Privacy Act of 1988.

Response and progression were evaluated by local board-accredited radiologists at participating centers using the revised Response Evaluation Criteria in Solid Tumors (RECIST) guidelines (version 1.1)<sup>98</sup> and the Positron Emission Tomography Response Criteria in Solid Tumors (PERCIST) guidelines<sup>99</sup> for solid tumors, the Response Assessment in Neuro-Oncology (RANO) criteria<sup>100</sup> for CNS tumors and National Comprehensive Cancer Network guidelines for acute leukemia. To meet criteria for SD, measurements must have met the SD criteria at a minimum interval of 6 weeks after commencing a recommended treatment. All patients with a CR or PR had confirmation of the response with a follow-up scan.

**Statistical methods.** Associations between categorical variables were represented as ORs and tested using two-sided Fisher's exact tests, with group numbers as given. Excess promoter variation was tested by the NBR method, as described in the Supplementary Methods section 'Scan for noncoding promoter variation', with multiple test correction following the Benjamini–Hochberg step-up procedure. All tests were implemented in R (v3.5.3) (<https://www.R-project.org/>).

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

### Data availability

WGS, RNAseq and methylation data generated by this study are available from the European Genome-phenome Archive under accession number EGAS00001004572. Databases used to help filter, prioritize and interpret variants are available online, including COSMIC (<https://cancer.sanger.ac.uk/cosmic>), Cancer Gene Census (<https://cancer.sanger.ac.uk/census>), Pecan (<https://pecan.stjude.cloud/>), dbcsSNV (<http://www.liulab.science/dbcsnv.html>), dbNSFP (<https://sites.google.com/site/jpopgen/dbNSFP>), ExAC (<http://exac.broadinstitute.org/>), gnomAD (<https://gnomad.broadinstitute.org/>), MGRB (<https://sgc.garvan.org.au/>), GIAB (<https://jimb.stanford.edu/giab-resources>), Platinum Genomes (<https://github.com/Illumina/PlatinumGenomes>), ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>), ESP (<https://evs.gs.washington.edu/EVS/>) and 1000 Genomes (<https://www.internationalgenome.org/data>).

### Code availability

Software and scripts related to this publication are available at <https://github.com/CICB/2020-hrPC-landscape>.

### References

- Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv* <https://arxiv.org/abs/1303.3997> (2013).
- Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1–11.10.33 (2013).
- Gayevskiy, V., Roscioli, T., Dinger, M. E. & Cowley, M. J. Seave: a comprehensive web platform for storing and interrogating human genomic variation. *Bioinformatics* **35**, 122–125 (2019).
- McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
- Paila, U., Chapman, B. A., Kirchner, R. & Quinlan, A. R. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput. Biol.* **9**, e1003153 (2013).
- Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
- Pinese, M. et al. The medical genome reference bank contains whole genome and phenotype data of 2570 healthy elderly. *Nat. Commun.* **11**, 435 (2020).
- Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31–31 (2016).
- Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
- Cameron, D. L. et al. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res.* **27**, 2050–2060 (2017).
- Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

72. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
73. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
74. Craig, D. W. et al. A somatic reference standard for cancer genome sequencing. *Sci. Rep.* **6**, 24607 (2016).
75. Jaganathan, K. et al. Predicting splicing from primary sequence with deep learning. *Cell* **176**, e524 (2019).
76. Cheng, J. et al. MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol.* **20**, 48 (2019).
77. Jian, X., Boerwinkle, E. & Liu, X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res.* **42**, 13534–13544 (2014).
78. Signal, B., Gloss, B. S., Dinger, M. E. & Mercer, T. R. Machine learning annotation of human branchpoints. *Bioinformatics* **34**, 920–927 (2018).
79. Xiong, H. Y. et al. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
80. Garrido-Martin, D., Palumbo, E., Guigo, R. & Breschi, A. ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLoS Comput. Biol.* **14**, e1006360 (2018).
81. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
82. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
83. Fu, W. et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
84. Fairley, S., Lowy-Gallego, E., Perry, E. & Flicek, P. The international genome sample resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* **48**, D941–D947 (2019).
85. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
86. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
87. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
88. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).
89. Shihab, H. A. et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65 (2013).
90. Kim, S., Jhong, J. H., Lee, J. & Koo, J. Y. Meta-analytic support vector machine for integrating multiple omics data. *BioData Min.* **10**, 2 (2017).
91. Dong, C. et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).
92. Tate, J. G. et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
93. Sondka, Z. et al. The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).
94. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
95. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
96. Haas, B. J. et al. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol.* **20**, 213 (2019).
97. Davidson, N. M., Majewski, I. J. & Oshlack, A. JAFFA: high sensitivity transcriptome-focused fusion gene detection. *Genome Med.* **7**, 43 (2015).
98. Eisenhauer, E. A. et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**, 228–247 (2009).
99. O, J. H., Lodge, M. A. & Wahl, R. L. Practical PERCIST: a simplified guide to PET response criteria in solid tumors 1.0. *Radiology* **280**, 576–584 (2016).
100. Wen, P. Y. et al. Response assessment in neuro-oncology clinical trials. *J. Clin. Oncol.* **35**, 2439–2449 (2017).

## Acknowledgements

We sincerely thank patients and parents for participating in this study. We thank the many clinicians and health professionals for their time acquiring consenting patients and for collection and coordination of samples and associated clinical data at Sydney Children's Hospital, Randwick; The Children's Hospital at Westmead; the John Hunter Hospital; the Queensland Children's Hospital; the Royal Children's Hospital Melbourne; the Monash Children's Hospital; the Adelaide Women's & Children's Hospital; and the Perth Children's Hospital. We thank M. Weber (Prince of Wales Hospital) for assistance with anatomical pathology expert input. The authors thank the Sydney Children's Tumour Bank Network for providing samples and related clinical information for this study. We thank the staff of the Personalised Medicine Theme of the Children's Cancer Institute for their dedicated

work on Zero Childhood Cancer. We thank M. Dinger, B. Lundie, J. Stockmeyer, A. Statham, J. Coptoy and the staff at the Kinghorn Centre for Clinical Genomics for helpful discussions related to, and rapid processing of, WGS; J. Mattick and M. Dziadek from the Garvan Institute for helpful discussions related to WGS and for supporting fundraising initiatives; S. Eggers and the staff at the Murdoch Children's Research Institute/Victorian Comprehensive Genetics Service for rapid processing of RNAseq; and the staff at the Australian Genome Research Facility for rapid processing of methylation data. We thank the team from the National Computational Infrastructure, which is supported by the Australian Government. We thank G. Asiminos, D. Pollard and the team from DNAnexus for support with genomic data processing and helpful discussions regarding pipeline development. We thank the Australian Federal Government Department of Health, the New South Wales State Government and the Australian Cancer Research Foundation for funding to establish infrastructure to support the Zero Childhood Cancer personalized medicine program. We thank the Kids Cancer Alliance, Cancer Therapeutics Cooperative Research Centre, for supporting the development of a personalized medicine program; Tour de Cure for supporting tumor biobank personnel; The Steven Walter Children's Cancer Foundation and The Hyundai Help 4 Kids Foundation for supporting G.M.M. and P.G.E.; and the Lions Kids Cancer Genome Project, a joint initiative of Lions International Foundation, the Australian Lions Children's Cancer Research Foundation (ALCCRF), the Garvan Institute of Medical Research, the Children's Cancer Institute and the Kids Cancer Centre, Sydney Children's Hospital. Lions International and ALCCRF provided funding to perform WGS and for key personnel, with thanks to J. Collins for project governance and advocacy. We thank the Cure Brain Cancer Foundation for supporting RNA sequencing of patients with brain tumors; the Kids Cancer Project for supporting molecular profiling and molecular and clinical trial personnel; and the University of New South Wales, W. Peters and the Australian Genomics Health Alliance for providing personnel funding support. The New South Wales Ministry of Health-funded Luminesce Alliance provided funding support for computational personnel and infrastructure. The Medical Research Future Fund, the Australian Brain Cancer Mission, the Mindereroo Foundation's Collaborate Against Cancer Initiative and funds raised through the Zero Childhood Cancer Capacity Campaign, a joint initiative of the Children's Cancer Institute and the Sydney Children's Hospital Foundation, supported the national clinical trial and associated clinical and research personnel. We thank the Kinghorn Foundation for personnel support. We thank the National Health and Medical Research Council of Australia (fellowships APP1059804 and APP1157871 to R.B.L.), the Cancer Institute of New South Wales and New South Wales Health (fellowship funding for M.J.C.). We thank the 2018 Priority-Driven Collaborative Cancer Research Scheme, co-funded by Cancer Australia and My Room, for personnel and computational support (grant no. 1165556 awarded to M.J.C.). Zero Childhood Cancer is a joint initiative led by the Children's Cancer Institute and Sydney Children's Hospital, Randwick.

## Author contributions

M.H., G.M.M. and M.D.N. conceived of and designed the Zero Childhood Cancer project. V.T. and E.V.A.M. led and managed the Zero Childhood Cancer Program. D.S.Z., T.O., F. Alvaro, G.B.M., L.D.P., N.G.G., H.T., P.W., S.-L.K., J.R.H. and A.S.M. were clinical leads at recruiting centers and coordinated patient selection. D.G.-W., P.A.S. and E.M. checked sample quality and identity and prepared samples for sequencing. A.J.G. provided pathology analysis. M.W., C.M., M.P., A.K., E.E.W., P.S., R.B.-J., M.S., I.M., F. Abascal, J.B., P.P., M.E.G., V.G. and M.J.C. developed bioinformatic methods and performed the computational genomics analyses. J.K., K.T., M.W., N.A.B. and M.P. interpreted and reported germline findings. L.M.S.L., D.-A.K.-Q., P.B., A.S., M.E.M.D., E.D.G.F., E.V.A.M., P.G.E. and M.J.C. analyzed and curated molecular findings. L.M.S.L., D.-A.K.-Q., T.N.T., G.M.M. and D.S.Z. led the Multidisciplinary Tumor Board discussions and recommendations. L.M.S.L., D.-A.K.-Q. and D.Z. evaluated patient response data. M.D.N., T.N.T. and R.B.L. provided childhood cancer expertise and advice. D.M.T. contributed precision medicine expertise and advice. P.G.E. and M.J.C. conceived of, designed and wrote the manuscript, with comments and contributions from all authors, particularly M.P. who wrote the germline findings section and L.L. and D.S.Z. who wrote the clinical impact sections.

## Competing interests

All authors, except P.G.E., S.-L.K. and D.S.Z., declare that they have no competing interests as defined by Nature Research or other interests that might be perceived to influence the interpretations of the article. D.S.Z. has received funding for travel and consulting from Bayer. P.G.E. and S.-L.K. are recipients of a share in milestone and royalty payments paid to the Walter and Eliza Hall Institute of Medical Research for the development of venetoclax.

## Additional information

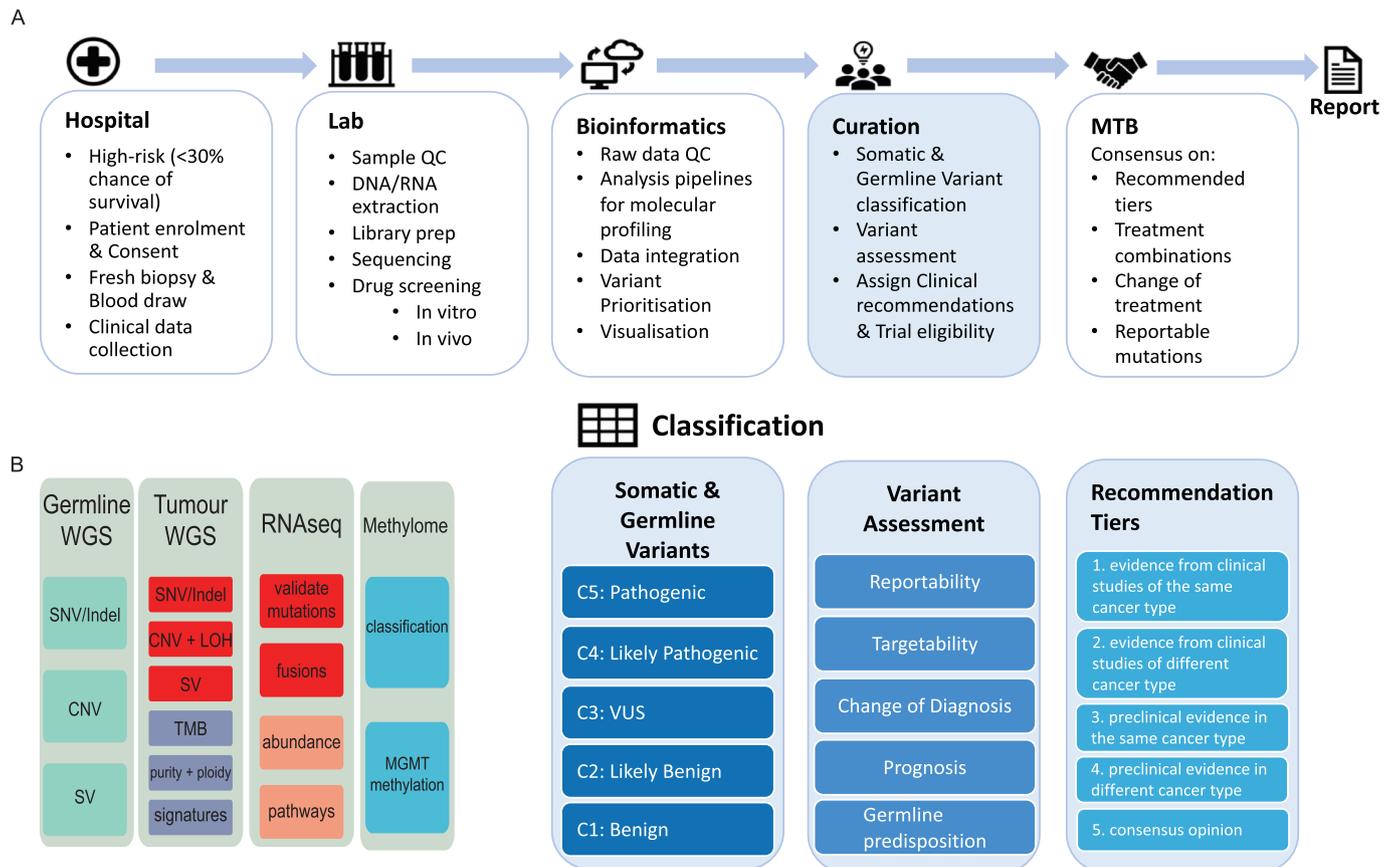
**Extended data** is available for this paper at <https://doi.org/10.1038/s41591-020-1072-4>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41591-020-1072-4>.

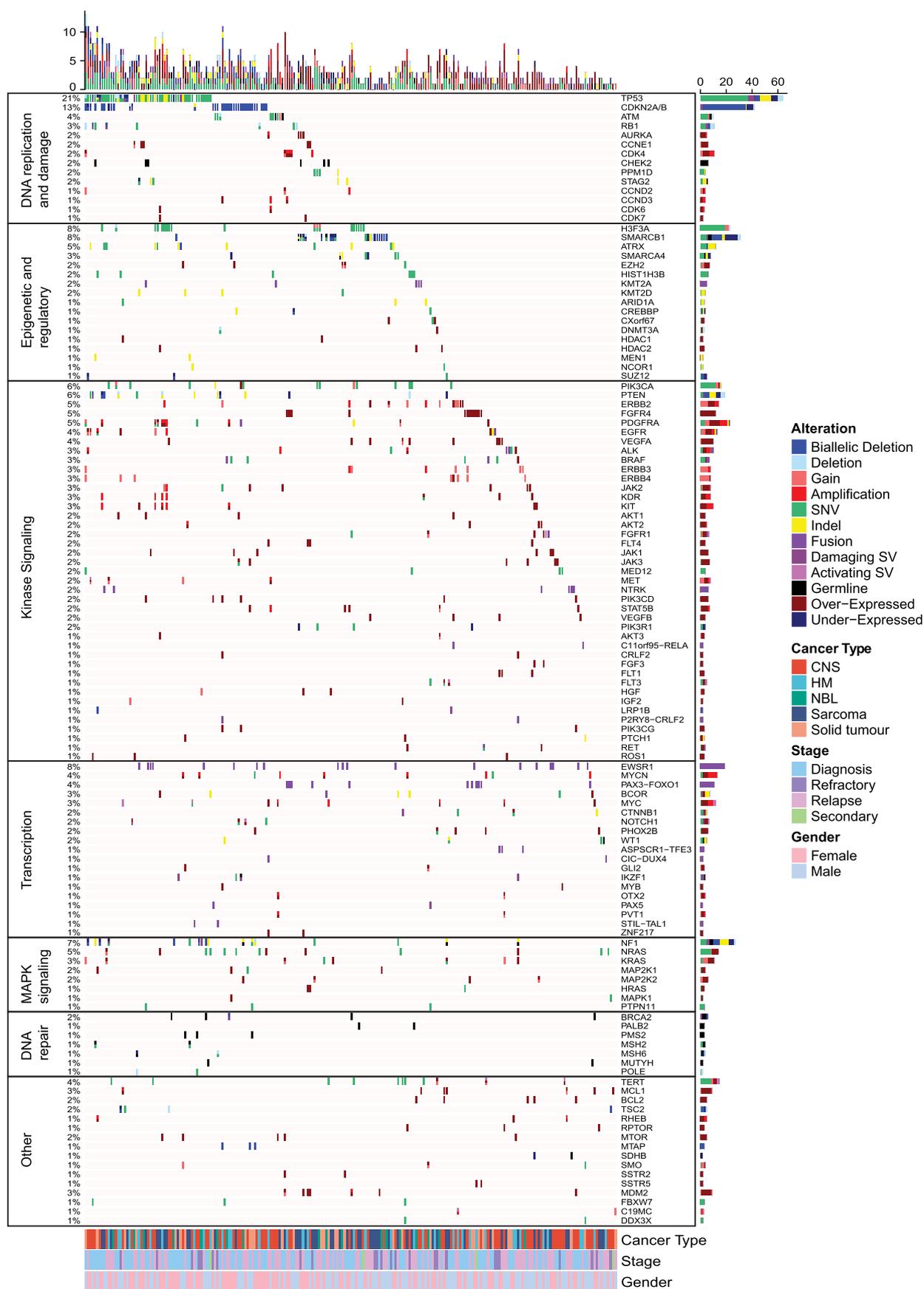
**Correspondence and requests for materials** should be addressed to D.S.Z., P.G.E. or M.J.C.

**Peer review information** Javier Carmona was the primary editor on this article, and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

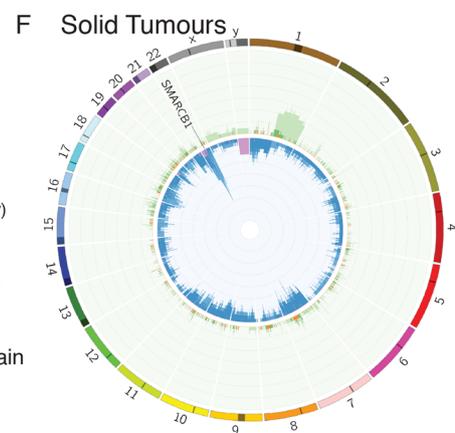
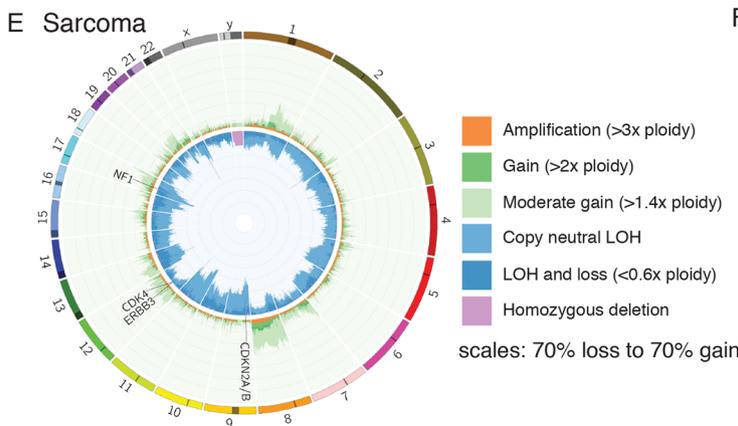
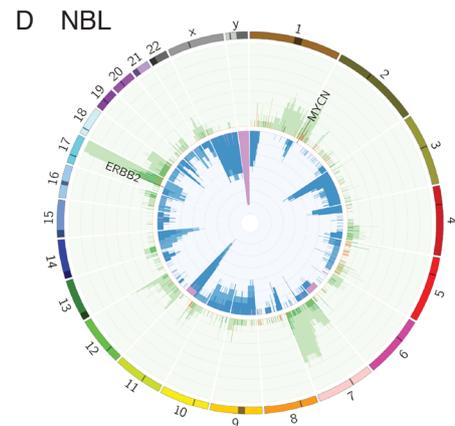
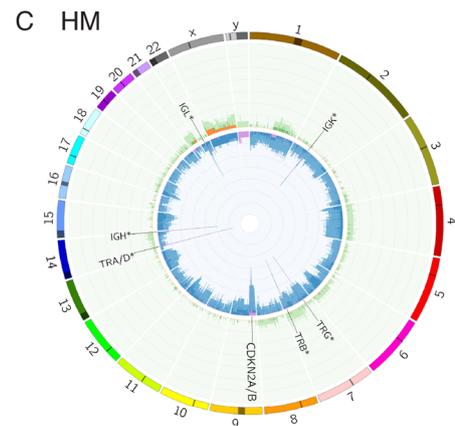
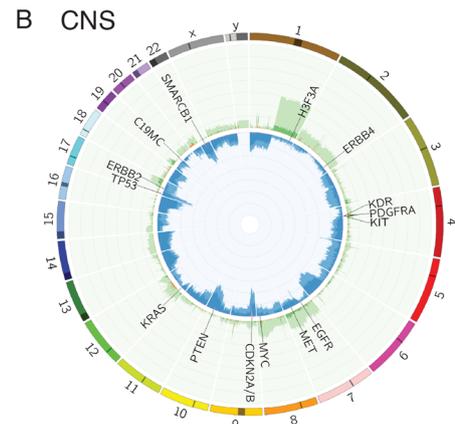
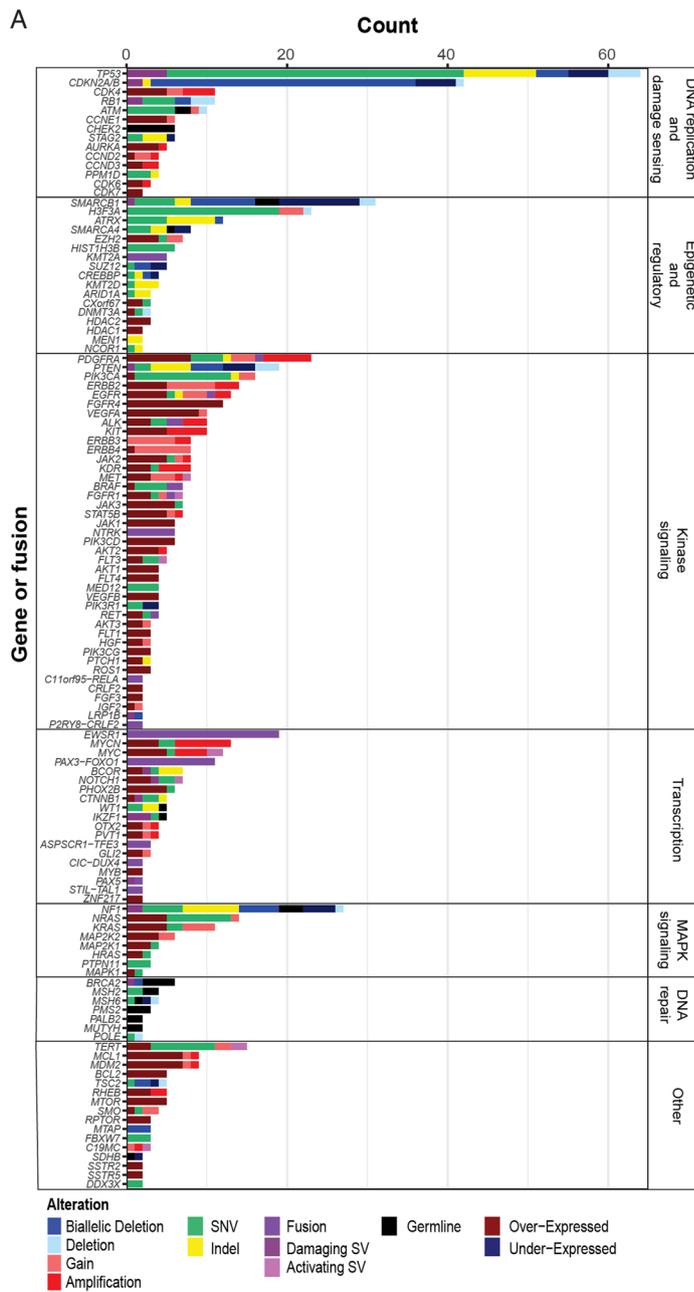


**Extended Data Fig. 1 | ZERO Workflow. a**, The workflow of samples through the ZERO program is shown. Patients were enrolled at one of eight paediatric hospitals, samples processed centrally then profiled at various national centres of excellence for WGS, RNAseq and methylome analysis. The data from each molecular profiling platform are analysed centrally and integrated where possible via dedicated bioinformatics pipelines. The resulting molecular aberrations or classifications from all analysis pipelines are collated and stored in a relational database and integrated by an in-house integration system, Glooe. Here, Glooe applies algorithms that score and rank each variant for prioritisation and generates visualisations to support data quality control assessment and downstream curation. The curation process is critical for classifying molecular aberrations prior to reporting changes to the Molecular Tumour Board (MTB). The multidisciplinary curation team determines the pathogenicity of each variant (C1–C5) consistent with published guidelines<sup>29</sup>. Ultimately the molecular aberrations are assessed for their reportability (that is of potential interest to clinicians), potential targetability with anticancer drugs, potential to support, refine, or propose a change of diagnosis, prognosis, or indicate the presence of a germline cancer predisposition syndrome. The MTB determines the strengths of molecular, preclinical or clinical data supporting potential therapeutic findings. Ultimately, the MTB seeks consensus on what is reported back to the patient. **b**, the variant types for which each sequencing platform was used. The red colouring in Tumour WGS and RNAseq indicates that both platforms were used to corroborate these variant types where possible.



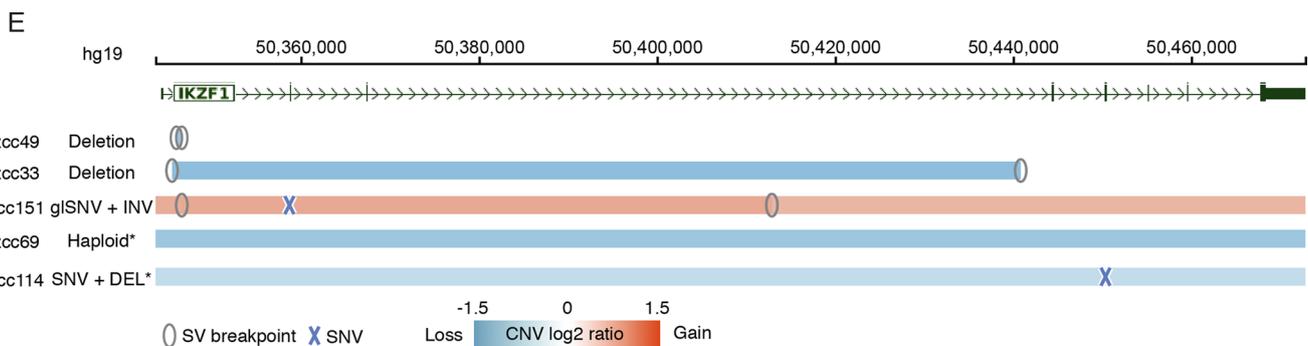
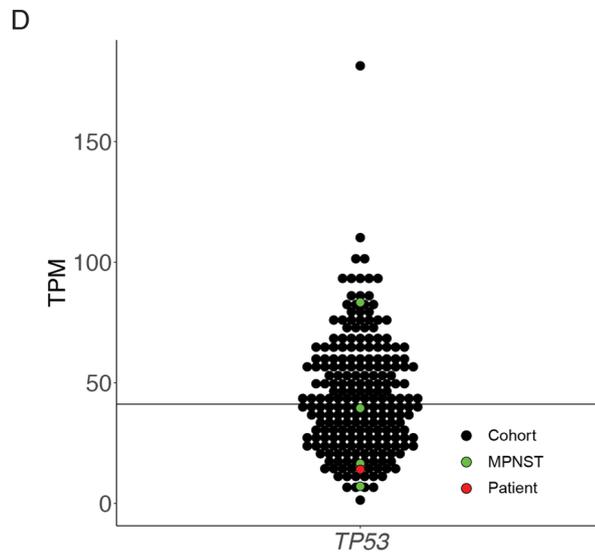
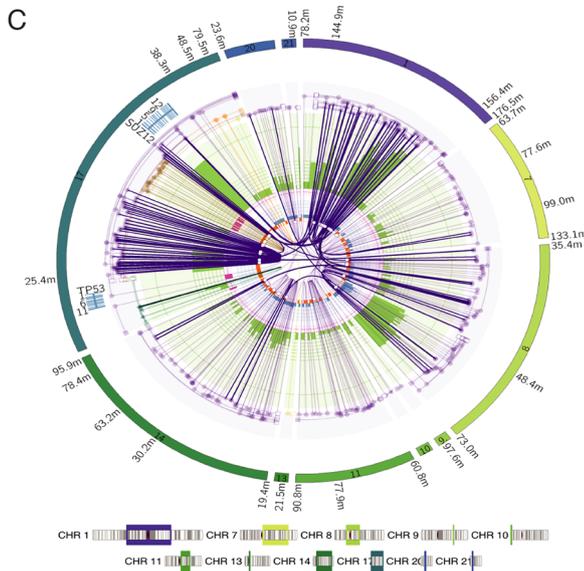
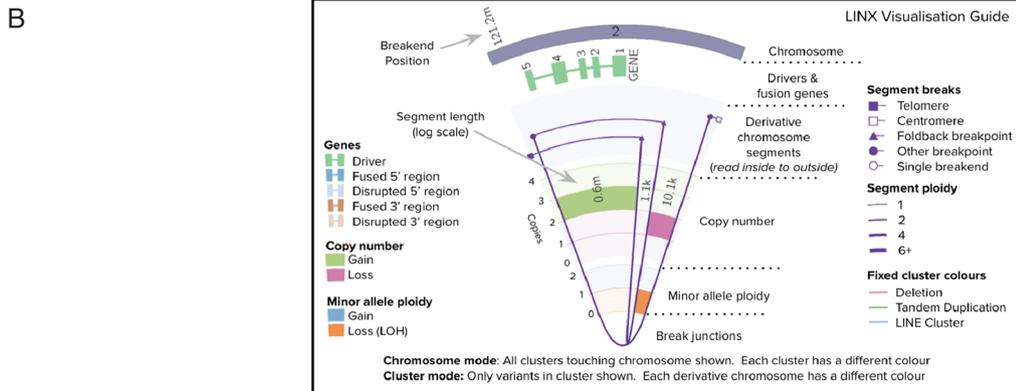
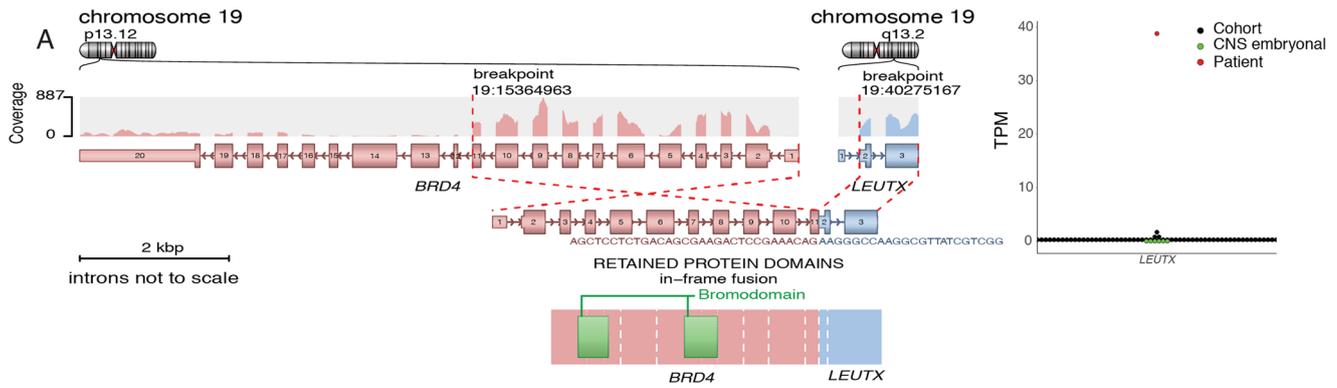
Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Mutational Landscape of High-Risk Paediatric Cancers.** An oncoprint of the reportable germline and somatic events, of any type, observed in more than 2 patients. Genes are classified into broad pathway categories and ordered by the percentage of patients harbouring an alteration in the specified gene. The barplot above shows the number of events in each patient highlighted by aberration type. The barplot on the right shows the total number of events in each indicated gene, highlighted by aberration type and represents genes where more than one aberration, for example CNV gain and over-expression, occurred in the same sample. Where individual genes are shown for each patient, the colour represents the variant type (legend shown on right). Where these bars have multiple colours, it is because the same variants were detected by different techniques, or there are multiple variant types affecting each allele. The heatmaps on the bottom show the distribution of cancer type, stage of disease and sex of patient (legend shown on the right).



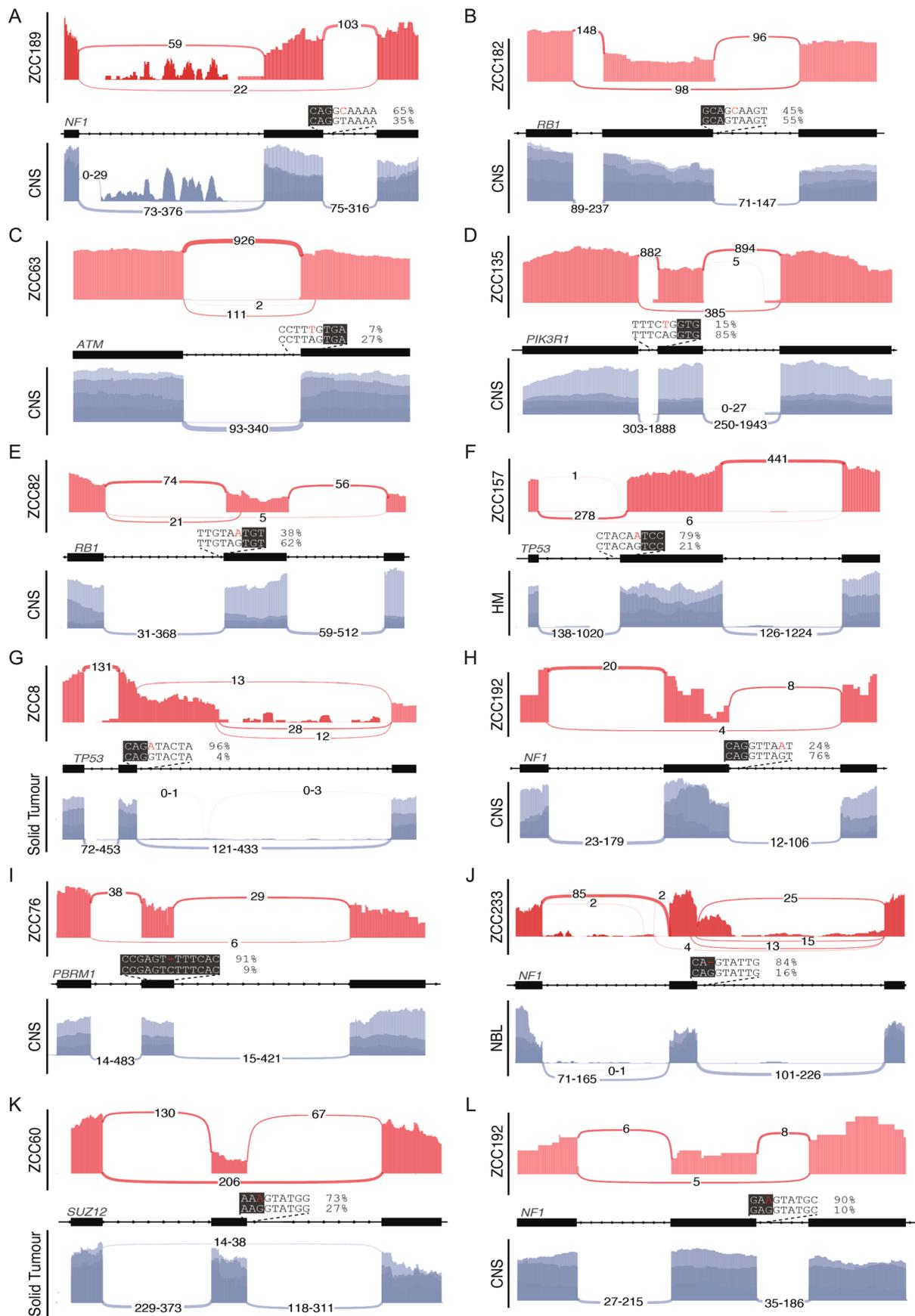
Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | Pathogenic Aberrations in High-Risk Paediatric Cancers.** **a**, Genes recurrently altered by any of the different variant types are shown, grouped by broad pathway categories and ordered by the frequency of the mutation. The plot includes all genes affected in >2 patients. The colouring of the bars indicates the proportion of each variant type. **b-f**, Circos plots of the genome-wide copy number profiles for each major category of cancers. The amplifications and deletions or LOH relative to the tumour ploidy are shown according to the inset legend. The scale on both rings is 0-70% and inverted for the inner ring. The genes most frequently observed with high-level amplifications, homozygous deletions and immunoglobulin or T-cell receptor gene rearrangements (\*) are shown. The number of patients in each subtype are CNS (n=92), HM (n=43), NBL (n=20; n=8 males), Sarcoma (n=61), other solid tumours (n=35).



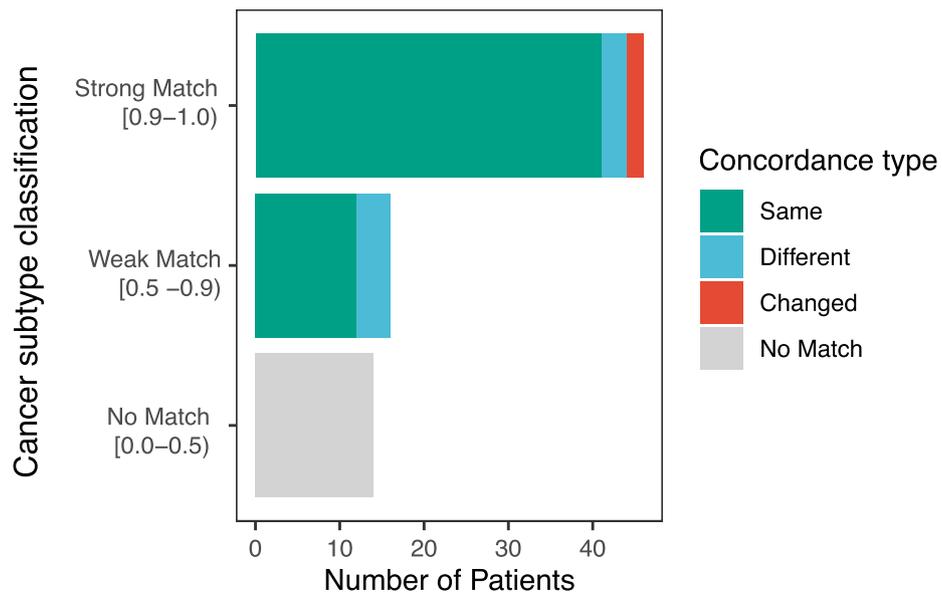
Extended Data Fig. 4 | See next page for caption.

**Extended Data Fig. 4 | Complex structural rearrangements driving novel fusions.** **a**, BRD4-LEUTX fusion identified from RNAseq and WGS from a CNS embryonal tumour at diagnosis in an infant. This figure reveals that at least one SV joined exons 1-11 of *BRD4* to exons 2-3 of *LEUTX*. These exons were highly expressed (coverage histogram above each exon) and the resulting fusion was in-frame and retained the bromodomains in BRD4 and homeobox domain in LEUTX. The expression of LEUTX across cohort is depicted (right), with patient highlighted in red, other CNS embryonal tumours in green and cohort in black. **b**, LINX visualisation guide, highlighting how multiple SV can be chained together into a derivative chromosome, where SV breakpoints flank defined CNV segments. **c**, Complex genome topology underlies an out-of-frame *TP53-SUZ12* fusion identified from an MPNST tumour at initial diagnosis in an adolescent. The plot shows a predicted derivative chromosome, characterised by many SVs (purple lines) and CNV segments (predominantly green segments with ploidy of 2-4) of 10 chromosomes as indicated. **d**, A dot plot representation of expression TPM values of TP53 in the cohort, patient with *TP53-SUZ12* highlighted in red, other MPNST in green and cohort in black, black horizontal line signifies cohort mean TPM. **e**, A pathogenic 607 bp deletion affecting intron 1, a 96 Kb deletion and a 66 Kb inversion, both affecting exons 2-3 of *IKZF1* were identified in three patients with pre-B ALL. The inversion was likely *in trans* with a germline pathogenic splicing variant (*IKZF1*:c.40+1G>A). In two other pre-B ALL cases a tumour with a haploid genome had an *IKZF1* deletion and another had a deletion and a likely pathogenic somatic variant (*IKZF1*:c.544T>C, p.(Cys182Arg)); in both cases the CNV was not deemed to be a driver mutation and was not reported. The plot was made using GenomePaint (X Zhou, *in prep*).

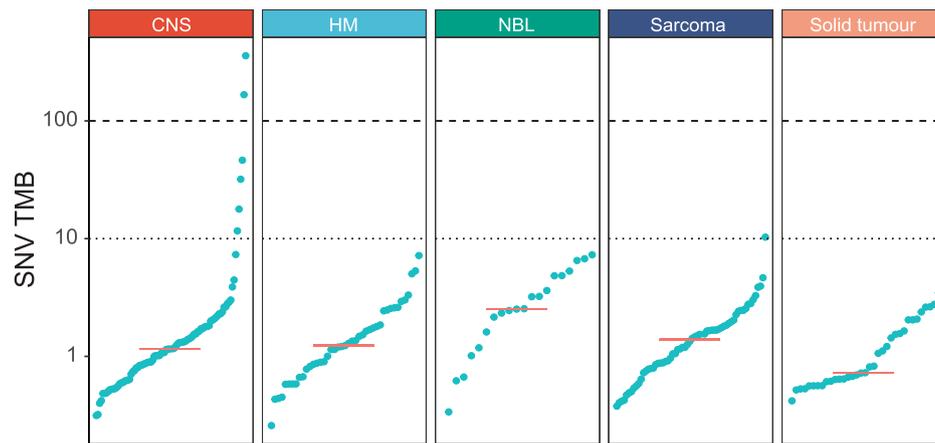


Extended Data Fig. 5 | See next page for caption.

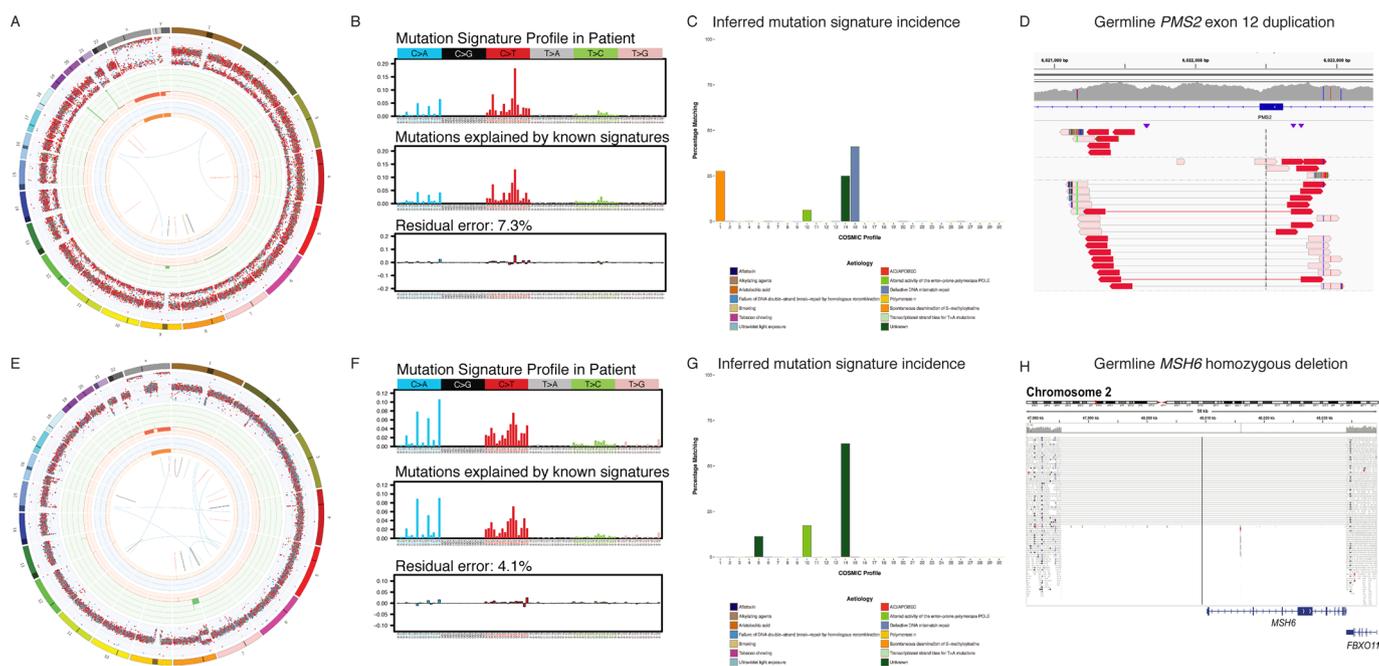
**Extended Data Fig. 5 | Integration of WGS and RNAseq reveals the impact of mutations on splicing. a–l,** The impact of unique somatic mutations on gene splicing, where details about each mutation are in Supplementary Table 3. In each case, a patient's mutation is shown in red, relative to the exon-intron boundary. The patient's RNAseq read coverage is shown above (red) as a histogram, and the number of reads supporting each splice junction, compared to four cancer-type matched controls below (purple).



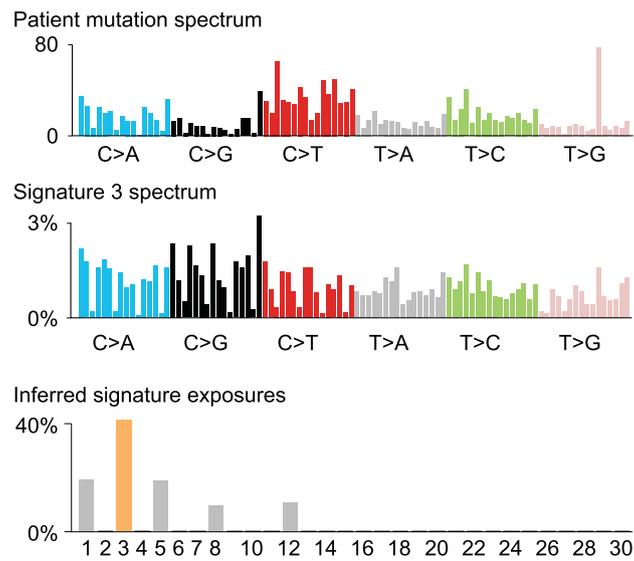
**Extended Data Fig. 6 | Methylation classification results from CNS tumours.** CNS tumour methylation array classification results from 76 patients were binned into those where a strong, weak or no match was made to a known tumour subtype and then by whether that classification matched, differed, or resulted in a change of the initial diagnosis determined from standard of care diagnostic testing.



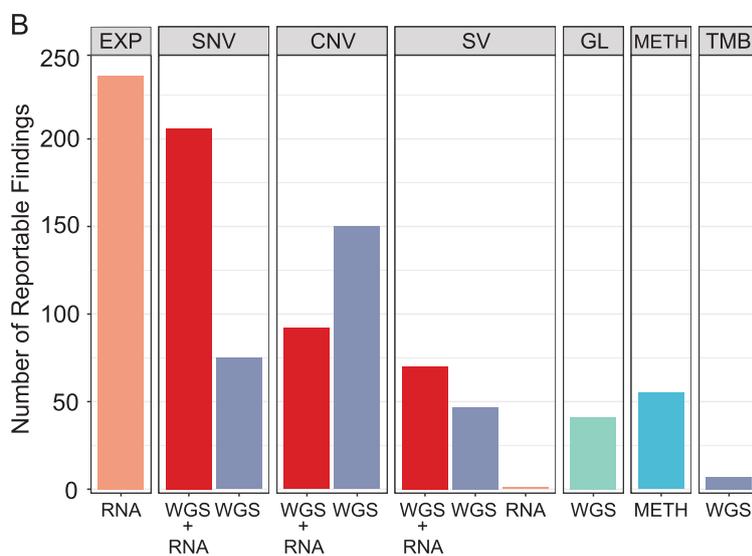
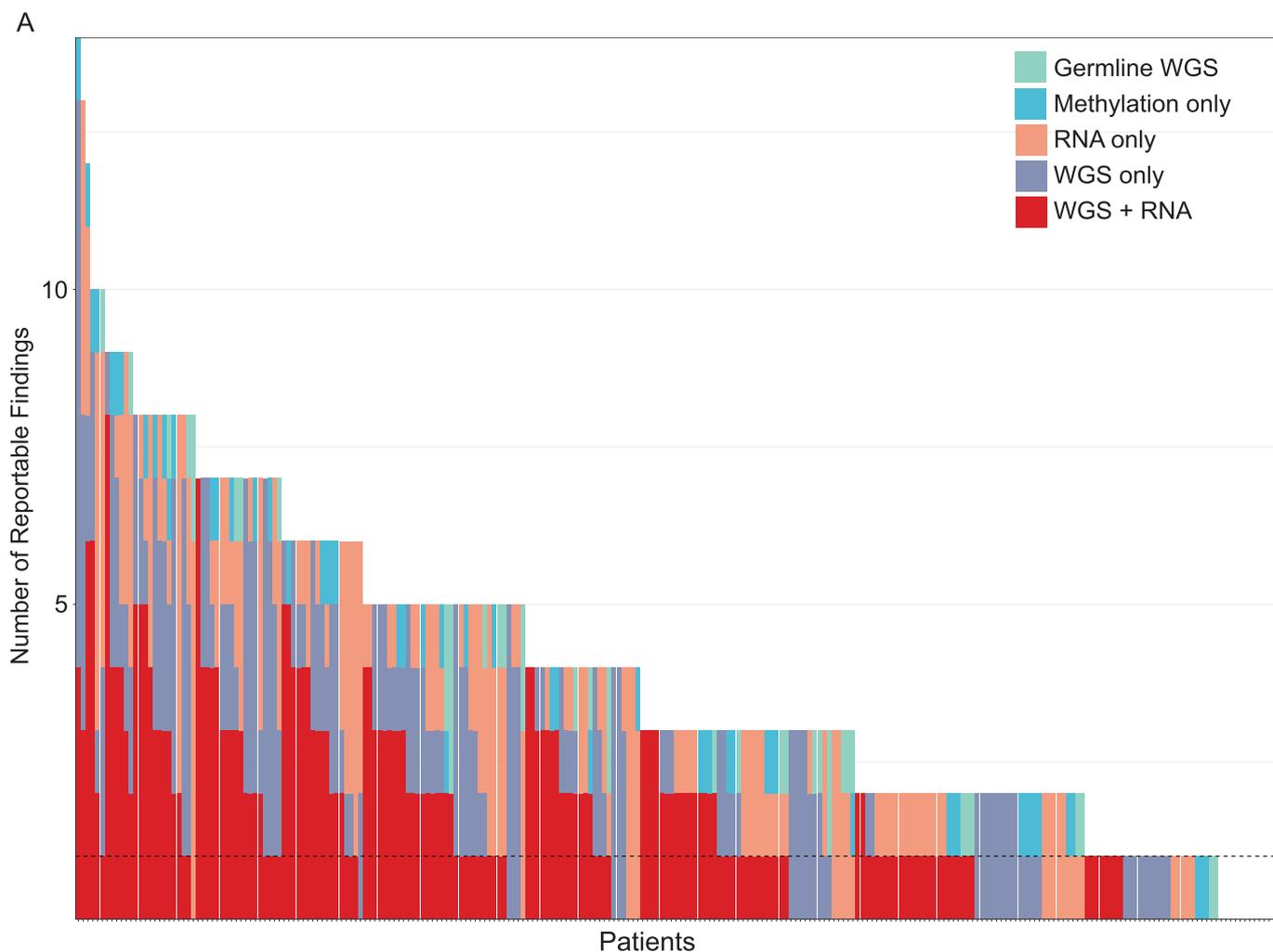
**Extended Data Fig. 7 | Tumour Mutation Burden across the cohort.** Somatic single nucleotide tumour mutation burden (TMB) for each patient and the mean per cancer type (red line), expressed as genome-wide mutations per MB. Ninety-four percent of tumours had low TMB (<10 mutations per MB), consistent with previous reports<sup>10</sup>.



**Extended Data Fig. 8 | Germline and somatic features of ultramutated CNS tumours.** **a**, Circos plot depicting the genome-wide mutation profile of an HGG, from a child, with 165 mutations/MB. The inner area shows SVs, followed by the minor allele ploidy (here highlighting the expected loss of heterozygosity on chrX), then CNV (deletions in red, amplifications in green), followed by somatic SNVs coloured according to the figure legend. **b**, Mutation signature analysis<sup>59</sup> (Methods) showed that most mutations matched previously known mutation signatures. **c**, The burden of each mutation signature is shown, where signature 14 has been associated with CMMRD and somatic *POLE/POLD1* mutations in the exonuclease domain<sup>49</sup>. A somatic *POLE* hotspot mutation (p.Leu424Val) was identified, which was classified as a VUS in ClinVar ([RCV000033144](https://www.ncbi.nlm.nih.gov/clinvar/RCV000033144)) at the time of sequencing. **d**, MLPA analysis revealed a heterozygous exon 12 duplication but could not resolve the breakpoints. We re-inspected the WGS data which revealed the heterozygous exon 12 duplication of *PMS2* from 7:6,021,129–6,023,032 (POLE:p.(Lys670Ala725dup)). None of our SV detection tools from WGS or RNAseq identified this variant. Exons 12–15 of *PMS2* are difficult to sequence due to 99% sequence homology with a pseudogene. WGS easily identified the other pathogenic allele (c.949C>T, p.(Gln317Ter)) and loss of *PMS2* expression confirmed by IHC, resulting in a diagnosis of CMMRD. Panels E–H similarly represent the molecular findings from a hemispheric HGG tumour with 353 mutations/MB. In this case, signature 14 again prompted us to look for *POLE/POLD1* variants, revealing a clear somatic deletion of one *POLE* allele and a subclonal hotspot mutation (p.S459F) with just 4 supporting reads and several VUS on the remaining allele. RNAseq did not support the expression of this hotspot mutation, so it was not reported. Both examples highlight how somatic features of the tumour, in this case TMB and mutation signatures prompted a deeper investigation into the molecular drivers of these tumours.



**Extended Data Fig. 9 | Germline HR mutation driving a tumour with signature 3.** An example of concordant germline and somatic findings supporting a treatment recommendation, showing the tumour's observed mutation spectrum (top panel), the HR-associated signature 3 mutation spectrum (middle panel), and the inferred signature exposure in the tumour (bottom panel), with signature 3 the dominant signal supporting the recommendation of a PARP inhibitor.



**Extended Data Fig. 10 | Multi-platform approach in ZERO increases reportable events.** **a**, The number of reportable events identified in each individual patient, coloured by the source of the data, whether from WGS only, RNAseq only, both WGS and RNA, germline WGS or by methylation array. The x-axis is each patient in the cohort. The y-axis is the number of reportable findings in each patient. **b**, The total number of reportable variants separated by type of event and coloured by the platform identifying the event. SNV: Single Nucleotide Variant, CNV: Copy Number Variant, SV: Structural Variant, EXP: Gene Expression, GL: Germline, METH: Methylation, TMB: Tumour Mutation Burden.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Clinical data collection was performed within LabMatrix by Biofortis, version R7.3.2.0

Data analysis

Code availability

Software and scripts related to this publication are available at <https://github.com/CCICB/2020-hrPC-landscape>.

In addition, these in house, open source and commercial tools were used in this publication:

ANNOVAR (v20190929)  
 Arriba (v1.1.0)  
 BWA-MEM (v0.17.10-r789)  
 Branchpointer (v1.3.1)  
 CADD (v1.3)  
 FATHMM (via dbNSFP v2.9)  
 FastQC (v0.11.5)  
 GATK GenotypeVCFs (v3.3)  
 GATK HaplotypeCaller (v3.3 for WGS and v3.6 for RNAseq)  
 GATK Indel Realignment (v3.3)  
 GATK ReassignOneMappingQuality (v3.6)  
 GATK SplitNCigarReads (v3.6)  
 GATK VQSR (v3.3)  
 GEMINI (v0.11.0)  
 GRIDSS (v2.7.2)  
 IGV (v2.6.2)  
 Introme (v0.5.1)  
 JAFFA (v1.09)  
 LINX (v1.7)

MMSplice (v2.1.0)  
 MNP Classifier (online, versions may have changed over time)  
 MetaLR (v1.0)93  
 MetaSVM (v1.0)  
 NBR (custom scripts available <https://github.com/CCICB/2020-hrPC-landscape/tree/master/NBR>)  
 Novosort (v1.03.01)  
 PROVEAN (v1.1)  
 PURPLE (v2.39)  
 Polyphen2 (v2.2.2)  
 R (v3.5.3)  
 RNA VAF estimator (custom scripts online, [https://github.com/CCICB/2020-hrPC-landscape/blob/master/RNA\\_VAF.py](https://github.com/CCICB/2020-hrPC-landscape/blob/master/RNA_VAF.py))  
 RSEM (v1.2.31)  
 RStudio (v1.2.1335)  
 RefyNr (v1.17.8)  
 SAMTools (v1.3.1)  
 SIFT (v5.0.2)  
 SPIDEX (v1.0)  
 STAR (v2.5)  
 STAR-Fusion (v1.3.1)  
 Seave (<https://seave.bio>; updated throughout the project)  
 SnpEff (v4\_3t)  
 SpliceAI (v1.3.1)  
 Strelka (v2.0.17)  
 Strelka filter (<https://bitbucket.org/cciacb/cci-strelka-filter/src/master/>)  
 Variant Effect Predictor, VEP (v87)  
 bedtools (v2.28.0)  
 dbNSFP (v2.9)  
 dbSNV (v1.1)  
 deconstructSigs (v1.8.0)  
 ggplot2 (v3.3.2)  
 ggsashimi (v0.4.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

### Data availability

WGS, RNAseq and methylation data generated by this study are available from the European Genome Archive, accession number EGAS00001004572.

Databases used to help filter, prioritise and interpret variant are available online, including COSMIC (<https://cancer.sanger.ac.uk/cosmic>), Cancer Gene Census (<https://cancer.sanger.ac.uk/census>), Pecan (<https://pecan.stjude.cloud/>), dbSNV (<http://www.liulab.science/dbscnv.html>), dbNSFP (<https://sites.google.com/site/jpopgen/dbNSFP>), ExAC (<http://exac.broadinstitute.org/>), gnomAD (<https://gnomad.broadinstitute.org/>), MGRB (<https://sgc.garvan.org.au/>), GIAB (<https://jimb.stanford.edu/giab-resources>), Platinum Genomes (<https://github.com/Illumina/PlatinumGenomes>), ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>), ESP (<https://evs.gs.washington.edu/EVS/>), 1000 genomes (<https://www.internationalgenome.org/data>).

### Notes to editor

- 1 - we are still transferring data to EGA. This is being done in 10TB batches and will be completed by 4-6 weeks
- 2 - Please note that a "gemini database" is not available on the web – it is essentially a local SQLite database version of a VCF file, making it easier to query these data.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
  Behavioural & social sciences
  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Sample size

At the time of study design the literature indicated that 90% of the enrolled patients should have adequate tumour tissue for molecular analysis, up to 45% of relapsed/refractory tumour samples would have actionable mutations and up to 10% of paediatric cancer patients may

have a reportable germline mutation.

We therefore hypothesised that:

- 10% of the tumours from enrolled patients cannot be profiled using any of the methods because the quantity and/or quality of the submitted tumour tissue would be inadequate for analysis, i.e. 90% of the patients would have adequate tumour tissue
- 50% of the patients with adequate tumour tissue would have a targetable alteration detected by molecular profiling
- 60% of the above patients would receive a recommendation in at least one of the three categories (drug, change of diagnosis, germline mutations)
- 75% of the recommendations would be made within a clinically relevant timeframe (before terminal progression or death)

Hence the predicted proportion of enrolled patients who would receive a recommendation in at least one of the three categories within a clinically relevant timeframe was 20%. Feasibility is therefore defined as 20% or more of the patients receiving a recommendation for personalised treatment within a clinically relevant timeframe. A sample size of 246 would provide a 95% confidence interval of +/-5% for a 20% recommendation rate.

Data exclusions No data excluded

Replication This was a feasibility study where replication was not applicable.

Randomization This is an observational study testing the feasibility of implementing a precision medicine platform in a high risk paediatric cancer population and therefore randomization was not applicable.

Blinding There is no group allocation to which investigators could be blinded

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

### Methods

- n/a  Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology
- Animals and other organisms
- Human research participants
- Clinical data

- n/a  Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics Age <21 years, any gender, diagnosis of cancer with a 5-year survival probability estimated at 30% or less, at any time of disease course (diagnosis, relapse, progression), any prior treatment with no segregation of treatment categories

Recruitment This is an observational study where paediatric participants, who were eligible as per study criteria, were identified by the treating clinician. While a small number of patients were excluded from tumour analysis because of lack of suitable tumour material, this is unlikely to impact on the results. It is possible that investigators had a bias to enrol patients whose tumours they considered more likely to harbour targetable lesions. However, there was a broad spectrum of tumour types included, and the number and types of patients enrolled were in keeping with the anticipated national annual incidence of high-risk tumour types.

Ethics oversight Ethics approval was provided by Sydney Children's Hospitals Network Human Research Ethics Committee (LNR/14/SCH/497) for the TARGET pilot study and by the Hunter New England Human Research Ethics Committee of Hunter New England Local Health District in New South Wales, Australia (Reference No: 17/02/15/4.06) and New South Wales Human Research Ethics Committee (Reference No: HREC/17/HNE/29) for the PRISM study.

Informed consent for each participant was provided by parents/legal guardian for participants under the age of 18 years and by the participants who were over the age of 18 years.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration NCT03336931

Study protocol	<a href="https://clinicaltrials.gov/ct2/show/NCT03336931">https://clinicaltrials.gov/ct2/show/NCT03336931</a>
Data collection	<p>For TARGET pilot cohort, patients were recruited between June 2015 and October 2017, with data collected prospectively collected during the same time interval.</p> <p>For the PRISM cohort, patients were recruited between September 2017 and May 2019, with data collected prospectively collected between September 2017 and March 2020.</p> <p>All clinical data presented in this manuscript was collected by designated study co-ordinators based at each of the 8 clinical centres where the children with malignancy are managed in Australia; Sydney Children's Hospital, Westmead Hospital for Children, Royal Children's Hospital, Melbourne, Monash Hospital for Children, Melbourne, Queensland Children's Hospital, Brisbane, Women's and Children's Hospital, Adelaide, Perth Children's Hospital, Perth, John Hunter Hospital, NSW.</p>
Outcomes	<p>The primary and secondary clinical outcomes described in the manuscript were predefined in the PRISM protocol as below:</p> <p>Primary endpoint: Proportion of patients receiving a recommendation for (1) personalised therapy within a clinically relevant timeframe (before terminal progression or death), (2) for change of diagnosis or (3) for further action related to germline mutations.</p> <p>Secondary endpoints:</p> <ol style="list-style-type: none"><li>1. Proportion of tumour samples found to have actionable molecular alterations</li><li>2. Proportion of patients who subsequently receive the recommended personalised therapy</li><li>3. Response rate to recommended personalised therapy as measured by RECIST and RANO criteria</li></ol>