The landscape of genomic alterations across childhood cancers

Pan-cancer analyses that examine commonalities and differences among various cancer types have emerged as a powerful way to obtain novel insights into cancer biology. Here we present a comprehensive analysis of genetic alterations in a pan-cancer cohort including 961 tumours from children, adolescents, and young adults, comprising 24 distinct molecular types of cancer. Using a standardized workflow, we identified marked differences in terms of mutation frequency and significantly mutated genes in comparison to previously analysed adult cancers. Genetic alterations in 149 putative cancer driver genes separate the tumours into two classes: small mutation and structural/copy-number variant (correlating with germline variants). Structural variants, hyperdiploidy, and chromothripsis are linked to TP53 mutation status and mutational signatures. Our data suggest that 7–8% of the children in this cohort carry an unambiguous predisposing germline variant and that nearly 50% of paediatric neoplasms harbour a potentially druggable event, which is highly relevant for the design of future clinical trials.

Cure rates for childhood cancers have increased to about 80% in recent decades, but cancer is still the leading cause of death by disease in the developed world among children over one year of age. Furthermore, many children who survive cancer suffer from long-term sequelae of surgery, cytotoxic chemotherapy, and radiotherapy, including mental disabilities, organ toxicities, and secondary cancers. A crucial step in developing more specific and less damaging therapies is the unravelling of the complete genetic repertoire of paediatric malignancies, which differ from adult malignancies in terms of their histopathological entities and molecular subtypes. Over the past few years, many entity-specific sequencing efforts have been launched, but the few paediatric pan-cancer studies thus far have focused only on mutation frequencies, germline predisposition, and alterations in epigenetic regulators.

We have carried out a broad exploration of cancers in children, adolescents, and young adults, by incorporating small mutations and copy-number or structural variants on somatic and germline levels, and by identifying putative cancer genes and comparing them to those previously reported in adult cancers by The Cancer Genome Atlas (TCGA). We have also examined mutational signatures and potential drug targets. The compendium of genetic alterations presented here is available to the scientific community at http://www.pedpancan.com.

This integrative analysis includes 24 types of cancer and covers all major childhood cancer entities, many of which occur exclusively in children (Fig. 1, Supplementary Table 1). Ninety-five per cent of the patients in this study were diagnosed during childhood or adolescence (aged 18 years or younger) and 5% as young adults (up to 25 years) (Extended Data Fig. 1a). This study is biased towards central nervous system tumours, and is complemented by an additional study of a non-overlapping paediatric cohort with mainly leukaemias and extracranial solid tumours.

We compiled paired-end Illumina-based sequencing data for 961 tumours (914 individual patients) from previous cancer-type specific studies (see Methods and Supplementary Note 1) including 547 whole-genome sequences (WGS, median coverage 37×) and 414 whole-exome sequences (WES, 121×) partially complemented by low-coverage whole genomes (Supplementary Tables 1, 2). Tumour and matched germline samples were processed with standardized pipelines to detect single nucleotide variants (SNVs), short insertions and deletions (indels), copy-number variants (CNVs) and other structural variants. Secondary (relapse) tumours (n = 82, including 47 matched to primaries) were analysed separately from the main primary cohort (n = 879).

Mutation frequencies across cancer types

Coding somatic SNV (93%) and indel (7%) counts correlated across all samples (n = 879) (R = 0.27, P = 9.1 × 10−5; Extended Data Fig. 1b, c). Mutation frequencies varied between cancer types (0.02–0.49 mutations per Mb) and were overall 14 times lower than in adult cancers (0.13 versus 1.8 mutations per Mb, TCGA data; Fig. 1, Extended Data Fig. 1c, Supplementary Table 3). Relapse tumours harboured significantly more mutations than primary tumours (P = 0.0015, excluding highly mutated tumours; Extended Data Fig. 1d).

Tumours with more than 10 mutations per Mb have been referred to as ‘hypermutators’, and are often related to deficiencies in mismatch repair (MMR). In this cohort, hypermutation occurred exclusively in H3.3 or H3.1 K27-wildtype (K27wt) high-grade gliomas with biallelic germline mutations in MSH6 or PMS2, with an extremely high mutational burden similar to the highest among adult tumours (in POLE- or POLQ-mutated carcinomas) (Fig. 1). Some paediatric tumours had a mutational burden below this threshold, but markedly above average (2–10 mutations per Mb, referred to as ‘paediatric highly mutated’), including several K27wt high-grade gliomas with monoallelic germline variants in MSH2, MSH6 or PMS2 (Fig. 1). Whether these highly mutated tumours respond to immune checkpoint inhibitors, as described for paediatric glioblastoma, should be of clinical interest.

As in previous reports, the somatic mutation burden increased with patient age (R = 0.39, P = 2.9 × 10−6), except in Burkitt’s lymphoma (immunoglobulin hypermutation) and tumours with ‘kataegis’ events of localized hypermutation at double-stranded breakpoints (Extended Data Fig. 1e, f). Both SNVs (R = 0.37, P = 1.0 × 10−5) and indels (R = 0.27, P = 5.4 × 10−4) correlated with patient age overall, although within some cancers (for example, acute lymphoblastic leukaemia (ALL), Ewing’s sarcoma, and rhabdomyosarcoma), we observed almost random mutational loads (R < 0.2). Rhabdomyosarcomas were largely dominated by embryonal tumours with more mutations than the few alveolar cases (median 0.27 versus 0.12 mutations per Mb, P = 0.002).
Mutational processes in childhood cancers

Most cancer types predominantly harboured C > T transitions (≥30% of SNVs in two-thirds of cancer types) linked to mutational signature 1, whose previously described age-association occurred in some paediatric brain tumours15,16 (P < 0.05; Extended Data Figs 1g, 2a–c). Mutational signatures, possibly reflecting biochemical cellular processes, have previously been investigated for many, mainly adult, cancers15. In this paediatric cohort (WGS, n = 503), we found evidence for major contributions of 16 out of 30 published signatures and also identified one new signature15 (Fig. 2, Extended Data Fig. 2a, Supplementary Table 4). This ‘signature P1’, which is distinct from any previously documented signatures and harbours elevated C > T mutations in a CCG/CCT context, occurred in several atypical rhabdoid tumours (ATRTs) and one ependymoma (Fig. 2, Extended Data Fig. 2d, Supplementary Table 5). Its activity correlated with ‘multiple nucleotide variants’ (MNVs, R = 0.87, P = 1.1 × 10−12), but no particular loci or genes were mutually altered in the affected tumours (Extended Data Fig. 2d). Notably, all ATRTs with signature P1 were in the recently defined subgroup ‘SHH’, and even within one tumour (Extended Data Fig. 2d). Signatures 16 and 18 were heterogeneous-represented within several cancer types, with signature 16 being most prominent in pilocytic astrocytomas, and signature 18, previously proposed to be associated with oxidative DNA damage and related to C > A transversions, in neuroblastomas, rhabdomyosarcomas, and other tumours with multiple structural variants15,18 (Extended Data Figs 1g, 2a, c, 3a).

Signature 3, the ‘canonical’ double-stranded break signature linked to mutations in BRCA1 or BRCA2 or to a ‘BRCAness’ phenotype, and signatures 8 (recently linked to BRCA2 or PALB2 germline mutations in medulloblastomas; S. M. Waszak et al., personal communication) and 13 were linked to chromothripsis and TP53 mutations. This was particularly true for TP53 germline-mutated SHH medulloblastomas, and similarly for adenocortical carcinomas and rhabdomyosarcomas (Extended Data Fig. 3b, c). Overall, signatures 3, 8, and 13 were more pronounced in cancer types with higher genomic instability (that is, structural variants; Extended Data Fig. 2e).

Germline variants in cancer predisposition genes

A recent study of more than 1,000 patients estimated that about 8% of children with cancer harbour a hereditary predisposition5. Accordingly, in our cohort (n = 914 individual patients, about 25% of samples overlapping with the previous study), 7.6% of samples were determined as being likely to be associated with a pathogenic germline variant5,19 (162 genes investigated; Supplementary Tables 6, 7). No general age-of-onset bias was observed in patients with a predisposition; however, onset was later in germline MMR-deficient patients (P = 0.0001), even within the high-grade glioma sub-cohort (P = 0.001).

Hereditary predisposition was most common in adenocortical carcinomas (50%) and hypodiploid B-ALL (28%), followed by K27wt high-grade gliomas, ATRTs, SHH medulloblastomas, and retinoblastomas (15–25% each; Fig. 3a). Compared to the previous study, LZTR1, TSC2, and CHEK2 emerged as new putative predisposition genes, and possible new associations, such as SDHA with medulloblastoma, were detected5 (Fig. 3b).

Most germline variants were related to DNA repair genes from mismatch (MSH2, MSH6, PMS2) and double-stranded break (TP53, BRCA2, CHEK2) repair (Fig. 3b, c). Both groups are clinically relevant: patients with constitutional MMR deficiency could be candidates for immune checkpoint inhibition13 (Figs 1, 3b, c). Carriers of TP53 germline mutations (Li–Fraumeni syndrome), here most common in adenocortical carcinomas, hypodiploid B-ALL, SHH medulloblastomas, and K27wt high-grade gliomas, are at a 50% risk for early-onset cancer compared to 1% overall, and are susceptible to treatment-induced secondary oncogenesis20–22 (Fig. 3b). Correcting the predisposition frequency of 7.6% in this cohort for the relative incidence of

Figure 1 | Somatic mutations in the paediatric pan-cancer cohort.

Somatic coding mutation frequencies in 24 paediatric (n = 879 primary tumours) and 11 adult (n = 3,281) cancer types (TCGA)5. Hypermutated and highly mutated samples are separated by dashed grey lines and highlighted with black squares. Median mutation loads are shown as solid lines (black, cancer types; purple, all paediatric; green, all adult).
cancer types as a whole, we find that approximately 6% of all childhood cancer patients may carry a causative germline variant (Fig. 3d).

**Significance analysis identifies cancer driver genes**

Genome-wide analysis for significant mutation clusters \( (n = 538, \) WGS excluding hypermutators) identified non-coding mutations in the TERT promoter in 2.5% of tumours \( (\text{Extended Data Fig. 4a, b, Supplementary Table 8}) \). Further high-confidence clusters corresponded to coding mutations in frequently mutated genes \((\text{TP53, H3F3A, CTNNB1})\), and to localized hypermutation at the rearranged MYC locus in Burkitt’s lymphoma, while the bulk were classified as likely technical artefacts\( ^{(25)} \) \( (\text{Extended Data Fig. 4b}) \).

MuSiC identified 77 significantly mutated genes \( (\text{SMGs}) \), which were ranked according to their pan-cancer mutation frequency\( ^{(24)} \) \( (\text{Fig. 4, Supplementary Tables 9, 10}) \). Most SMGs were mutually exclusively mutated across cancer types, demonstrating specificity of single putative driver genes in childhood cancers as compared to more frequent co-mutation in adult cancers in the TCGA study\( ^{(7)} \). Without the accompanying paediatric pan-cancer study\( ^{(3)} \), only around 30% of paediatric SMGs overlapped with adult SMGs \( (\text{Extended Data Fig. 5c}) \). On the basis of incidence-normalized mutation frequencies, TP53 is predicted to be the most common somatically mutated gene \((4\% \text{ of childhood tumours})\), followed by TP53 and 76% in multiple SMGs\( ^{(25)} \) \( (\text{Extended Data Fig. 5b}) \). In line with the accompanying paediatric pan-cancer study\( ^{(3)} \), in contrast to multiple DNA repair-related germline mutations, and also in contrast to adult cancers \((9\% \text{ of SMGs, TCGA})^{(7)} \), PI3K-associated SMGs are the most commonly altered \((31\%) \text{ in adult cancers, compared to only 3}\% \text{ in paediatric cancers, which could be related to their often late occurrence in the evolution of multi-hit adult cancers}^{(27)} \)(Extended Data Fig. 5a).

Forty-seven per cent of paediatric tumours harboured at least one SMG mutation, with most tumours \((57\%) \) having only one. SMG mutations were rare \(< (15\%) \text{ in ependymomas, hepatoblastomas, Ewing’s sarcomas (driven by EWSR1 fusions instead of by point mutations)}^{(28)} \), and pilocytic astrocytomas, and common \( (> (90\%) \text{ in K27M high-}

| Cancer Type | Number of Patients with Germline Mutation |
|-------------|------------------------------------------|
| ACC         | 25%                                      |
| MBGR3       | 20%                                      |
| OS          | 15%                                      |
| RMS         | 13%                                      |
| ETMR        | 12%                                      |

**Figure 2 | Mutational processes active in paediatric cancers.** Contributions of thirty known and one novel mutational signature to the somatic mutations for the ten most frequently mutated samples per cancer type; each bar represents one individual tumour.

**Figure 3 | Germline mutations in cancer predisposition genes.**

- **Figure 3a** shows the frequency of patients with a pathogenic germline mutation per cancer type \((n = 914 \text{ tumours})\).
- **Figure 3b** displays mutated genes sorted by number of affected samples \( (\text{del, copy-number alterations; others, SNVs/indels}) \).
- **Figure 3c** demonstrates cellular processes associated with cancer predisposition genes.
- **Figure 3d** illustrates frequency of germline mutations adjusted for incidence and estimated total proportion of childhood cancers likely to be linked to hereditary predisposition.

**Recurrent structural and copy-number variants**

The degree of genomic instability \((\text{that is, the number of structural variants, including insertions, deletions, translocations, and inversions})\), varied substantially \( (\text{median 1–434 structural variants}) \) across regulators and MAP-kinase-associated genes accounted for 12–15% of SMGs. TP53 was the only DNA repair gene among somatic SMGs, in contrast to the multiple DNA repair-related germline mutations, and also in contrast to adult cancers \((9\% \text{ of SMGs, TCGA})^{(7)} \). PI3K-associated SMGs are the most commonly altered \((31\%) \text{ in adult cancers, compared to only 3}\% \text{ in paediatric cancers, which could be related to their often late occurrence in the evolution of multi-hit adult cancers}^{(27)} \)(Extended Data Fig. 5a).

Assessment of high functional impact mutations \( (\text{OncodriveFM})^{(29)} \) revealed well-known tumour suppressor genes \( (\text{TSGs}) \) such as TP53, ATRX, SMARCA4, and RB1, and other putative TSGs, including FMR1 in SHH/WNT medulloblastomas and MALR1 \( (\text{also known as C10orf112})^{(27)} \) in rhabdomyosarcomas \( (\text{Extended Data Fig. 6a}) \). Locally clustered ‘hotspot mutations’ \( (\text{OncodriveClust})^{(29,30)} \) identified known oncogenes, such as CTNNB1, PIK3CA, KRAS, and BRAF, proposed oncogenes \( (\text{ACVR1, KBTBD4, TBR1}) \), and possible new candidates, such as SF3B1, in Group 4 medulloblastomas \( (\text{Extended Data Fig. 6b}) \).
cancer types (WGS, n = 539), with more than 1,000 structural variants in individual samples of adenocortical carcinoma and osteosarcoma (Fig. 5a, Supplementary Table 11). Genomic instability correlated with germline (P = 3 × 10^{-15}) and somatic (P = 2 × 10^{-9}) TP53 mutations across all samples, but differed markedly between cancer types—again suggesting cancer type-specific effects of DNA repair (Fig. 5b, Extended Data Figs 3b, 7a).

Genomically unstable cancers were also more often hyperdiploid^{31} (Supplementary Table 12). Twelve per cent of tumours had a ploidy of four or more, 72% retained a near-diploid state (ploidy 1.5–2.5), and 26% had an aneuploid state (Fig. 5c, Supplementary Table 13). As evidenced by recurrent structural variation outside genes (based on breakpoint clusters in 10-kb windows), rearrangements linked to enhancer hijacking were also found, involving GFI1B and DDX31 in medulloblastomas and TERT in neuroblastomas^{40,41}. Together with genes directly affected by breakpoints, in total 70 structural variant-related putative cancer genes were found, many associated with cell cycle or growth (for example, the tumour suppressor PTPRD) or epigenetic regulators (such as SUZ12)^{42,43} (Extended Data Fig. 8c, Supplementary Tables 18, 19). Cancer type-specific events that occurred together with high expression (data derived from Northcott et al.)^{44} included alterations of RMS2^{45}.

The analysed genomic alterations were combined into 166 ‘likely functional events’ (LFEs) affecting 149 genes, classified as M-(mutation)-type or as SC-(structural/copy-number variant)-type (Extended Data Fig. 10a, Supplementary Table 21). Along the ‘cancer genome hyperbola’, individual tumours (WGS, n = 539) differentiated between an M-class (more M-type LFEs) and an SC-class (more SC-type LFEs)^{46} (Extended Data Fig. 10b, Supplementary Table 21). Fifty-five per cent of tumours were exclusive to one class, 27% were mixed but dominated by one type of LFE, 8% were ambiguous, and 10% had no LFEs (which may be of particular interest in assessing other tumour-driving events at the epigenetic or transcriptomic level).

Germline MMR mutations were enriched in the M-class, and germline TP53 mutations in the SC-class (P = 0.0003 and P = 0.05, respectively, Fisher’s exact test; Extended Data Fig. 10c). Individual cancer types displayed varying relative distributions of mutation classes (Extended Data Fig. 10d).

Drug targets in childhood cancers

To assess the status of druggability of childhood cancers, the cohort (n = 675 with full genomic information; WES-only, n = 39; see Methods) was screened for potentially druggable events^{39} (PDEs, that is, alterations in 179 genes with a directly or indirectly targeted treatment currently available or under development; Supplementary Table 22). This analysis revealed 453 PDEs in 59 genes, including 3% germline events (Supplementary Table 23). Most cancer types had tumours with PDEs related to both M- and SC-type (Fig. 6a). Most commonly, PDEs occurred in Burkitt’s lymphomas and pilocytic astrocytomas, while none were detected in ependymomas or hepatoblastomas (although the latter lacked information regarding CNVs or structural variants). Associated pathways included RTK/MAPK signalling, transcriptional regulation, cell cycle control, and DNA repair (Fig. 6a).

When the data are normalized for relative cancer incidence, 52% of all primary paediatric tumours may harbour a PDE (Fig. 6b); this might be an underestimate, given that some structural variants may not have been detected by this approach (for example, the common MYC translocations in Burkitt’s lymphoma)^{23}. After incidence adjustment, MAPK signalling and cell cycle control were most commonly affected. Notably, the PDEs often varied between primary and relapse tumours from one patient (n = 41): only 37% of primary tumours with PDEs retained these upon progression, while most of them partially or completely gained or lost events. This highlights the need for profiling of the current tumour when considering personalized therapy.

**Figure 4 | Significantly mutated genes in paediatric compared to adult cancer types.** Percentage of tumours with non-silent mutations in 77 SMGs for 24 paediatric tumour types (n = 879 tumours) and the pan-cancer cohort.
mechanisms, and shed light on therapeutic challenges such as tumour heterogeneity.

In summary, this multi-faceted pan-cancer analysis provides a valuable resource for assessing genomic alterations across the spectrum of paediatric tumours. While there are undoubtedly more discoveries to come in terms of expanded cohorts and whole-genome and transcriptome analysis, we believe that this study provides a strong basis for functional follow-up and investigation of potential therapeutic targets in this specific patient population.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 February; accepted 24 December 2017.

Published online 28 February 2018.

Discussion
Our analysis of this pan-cancer compendium outlines the landscape of genomic alterations across multiple childhood cancer types. Although some alteration types and rarer entities are still under-represented and significance analyses are probably limited, this dataset of nearly 1,000 tumours (which can be explored at http://www.pedpancan.com) provides an unprecedented data resource for paediatric cancer research, further complemented by the accompanying pan-cancer study7 (https://pecan.srjc.org/proteinpaint/study/pan-target). The multiple differences found compared to previous studies of adult tumours emphasize the need to consider paediatric cancers separately, further demonstrating a need for mechanism-of-action driven drug development for paediatric indications47.

The predicted frequency of pathogenic germline variants in 6% of patients, together with previous findings, demonstrates the relevance of genetic predisposition in childhood cancer2. Germline TP53 variants, which are clinically highly important, are estimated for 1.5% of children with cancer, and for more than 10% within individual cancer types. Genetic counselling should thus be systematically considered, particularly for patients with indicated high-risk entities.

Although stratified targeted treatment is currently incorporated only rarely into first-line therapy for paediatric cancer patients, our finding that nearly 50% of primary childhood tumours harbour a potentially targetable genetic event is encouraging. It also highlights the need for personalized profiling for each patient, both to increase diagnostic accuracy and to exploit the potential for potentially more effective and less harmful precision therapies. This may also transcend the direct targeting of genes or pathways, for example, through immune checkpoint inhibition in hypermutated tumours13 or through PARP inhibition in genomically unstable (‘BRCAness’) tumours48. It is hoped that ongoing personalized medicine approaches for patients at relapse will give initial information on the use and effectiveness of such targeted drugs (for example, in the clinical trials pedMATCH-NCT03155620; eSMART-NCT02813135; INFORM19). Additional longitudinal monitoring, for example using serial liquid biopsies, may further improve our understanding of tumour biology and the development of resistance

Figure 6 | Potentially druggable events in paediatric cancers.
\( a \), Proportion of primary tumours with potentially druggable events and associated biological pathways, per cancer type (\( n = 675 \) tumours with complete genomic information). NA, not available. \( b \), Proportion of patients with potentially druggable events, projected after normalization for incidence.
support came from the German Ministry for Education and Research (BMBF #01KU1201A) and the German Cancer Aid (#109252) for the ICGC (International Cancer Genome Consortium) PecBraTumor Project and the ICGC MMLM-Seq Project (within Program for Medical Genome Research Grants #01KU1002A–#01KU1002J), and the BioTOP Project (#01EX1502B). This work was also supported by the European Regional Development Fund (ERDF), MMLM-MYCY-SYS (#0316116), ICGC De-Mining (#01KU1505G), the Heidelberg Center for Personalized Oncology (DKFZ-HIPO), and the BMBF-funded Heidelberg Center for Human Bioinformatics (HD-Hub) within the German Network for Biomedical big Data Infrastructures (de.NBI). MSKCC received funding from the Department of Molecular Genomics (DKFZ) and grants from the German Cancer Research Center (DFG) and the Deutsche Forschungsgemeinschaft (DFG). The authors thank the reviewers for their contribution to the peer review of this work.

Author Contributions S.N.G. and B.C.W. performed data analysis and manuscript preparation. S.G.N., J.W., I.B., K.K., V.A.R., G.P.B., M.S.W., B.H., D.H., G.Z., M.H., J.C., and S.L. established workflows and performed data processing. P.D.J., S.B., S.B., D.S., E.P., S.W., U.K., J.M., G.V., C.K., M.K., D.T.W.J., L.C., and M.Z. contributed to design and interpretation of the analyses. P.D.J., D.H., C.B., A.M., K.K., S.F., K.W., A.B., S.B., R.K., D.B., A.E., S.B., R.K.-S., A.E.K., D.L., S.H., C.E., S.B., M.N., C.N., G.H.R., J.S., R.S., F.W., H.W., B.B., U.O., D.W., C.M.V., C.M.K., G.K., F.S.R., M.F.G., J.M.K., S.W., P., E.K., P.A., N.K.W., and M.K. provided data and patient materials. J.K., A.C.R., J.Z., Y.L., X.Z., A.J.W., D.A.Z., and P. curated the databases. S.N.G., B.C.W., D.T.W.J., and S.M.P. prepared the manuscript and figures. B.B., U.D.W., M.Ko., R.M.P., J.O.K., M.S., R.E., D.T.W.J., P.L., C.M.Z., and S.M.P. contributed to project management and provided leadership.

Author Information Reprints and permissions information is available at http://www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer Information Nature thanks S. Chanock and the other anonymous reviewer(s) for their contribution to the peer review of this work.

This research is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, then you will need to obtain permission from the licensor holder to reproduce the material. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Susanne N. Gröblner1,2,3, b, Barbara C. Worst1,2,3, d, Joachim Weischenfeldt2,3, e, Ivo Buchhalter4, a, Kortine Kleinheinz5, f, Vasilis A. Rudnev8, h, Pascal D. Johanni3,2,3,4, g, Gnana Prakash Balasubramanian1,2,9, Maia Segura-Wang10, a, Sebastian Brabetz11,2, b, Sebastian Bender12, c, Barbara Hutter13, d, Dominik Stürmer14,15, e, Elke Pfaff1,2,3,4, f, Daniel Hübschmann4,9,10, g, Gideon Zipprich7, h, Michael Heinold7,10, i, Jürgen Eils1, j, Christian Lawerenz1, k, Serap Erel1,2,3, Sander Lambo1,2,3, l, Sebastian Waszak1, m, Anna-Dorothee von Blomberg1,5, n, Christian Schmidt6, o, Angelaikis Eggert1, p, Simone Fulda3, q, Manfred Gessler1, r, Jenny Wengert5, s, Roland Kappler5, t, Daniel Baumann11, u, Stefan Burdach3,18, v, Reto Zbinden2, w, Hans Christian Eilert4, x, Renate Krüger-Schwenke3, y, Udo Kontry1, z, Andreas E. Kulozik1,5, a, Dietmar Lohmann1, b, Simone Hetterm1, c, Cornelia Eckert2, d, Stefan Bielack1, e, Michael Nathra1, f, Charlotte Niemeyer2, g, Günther H. Richter1, h, Johannes Schulte1, i, Reiner Siebert1, j, Frank Westermann4, k, Jan Molenars6, l, Gilles Vassal6, m, Hendrick Wittt2, n, Olli bigger01, o, CGC MedBRAIN-Seq Project, ICGC MMLM-Seq Project, Birgith Burkhardt27, p, Christian P. Kratz28, q, Olaf Witt3, r, Cornelis M. van Tilburg3, s, Christof M. Kramm3, t, Janet E. Steiner3, u, Stefan Ruckert1, v, Michael Frühwald4, a, Katja von Horn4, t, Stefan Ruckert4, u, Thomas Wolth5, v, Thomas Klingebiel3,8, w, Ewa Koscielniak11, x, Pablo Landgraf37, y, Jan Koster38, z, Anna C. Resnick39, a, Jinghui Zhang40, b, Yanling Liu40, c, Xin Zhou40, d, Angela J. Waanders41, e, Daphne Hubmann34, f, Stefan Ruckert4, g, Michael Frühwald4, h, Katja von Hort4, i, Stefan Ruckert4, j, Thomas Wolth5, k, Thomas Klingebiel3,8, l, Marcel Kool2,3, m, Rosario M. Piro42,43,44, n, Jan O. Korbé5, o, Matthias Schlesner7,45, p, Jan O. Korbé5, q, Matthias Schlesner7,45, r, Roland Els2, s, David T. W. Jones2, t, Peter Lichter3, u, Lukas Chavez2, v, Zwegers5, w, Reto Zbinden2, x, Stefan M. Pfister1, y, z

1Hopp’s Children’s Cancer Center at the NCT Heidelberg (KITZ), Heidelberg, Germany. 2Division of Pediatrische Onkologie, German Cancer Research Center (DKFZ), Heidelberg, Germany. 3Division of Cancer Genomics (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. 4Department of Pediatric Oncology, Hematology & Immunology, Heidelberg University Hospital, Heidelberg, Germany. 5European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany. 6The Finsen Laboratory, Rigshospitalet, Biotech Research and Innovation Centre (BRC), Copenhagen University, Copenhagen, Denmark. 7Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. 8Department of Developmental Neurobiology, St Jude Children’s Research Hospital, Heidelberg, Germany. 9Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. 10Division of Developmental Neurobiology, St Jude Children’s Research Hospital, Heidelberg, Germany.
Memphis, Tennessee, USA. 3Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. 4Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology, Heidelberg University and BioQuant Center, 69120, Heidelberg, Germany. 5Klinikum Stuttgart – Olgahospital, Zentrum für Kinder-, Jugend- und Frauenmedizin, Pädiatrie, Stuttgart, Germany. 6Department of Pediatric Hematology, Oncology, & Clinical Immunology Research Laboratories, Dr von Hauner Children’s Hospital, Ludwig Maximilians University, Munich, Munich, Germany. 7Bone Tumor Reference Center at the Institute of Pathology, University Hospital Basel and University of Basel, Basel, Switzerland. 8Children’s Cancer Research Centre and Department of Pediatrics, Klinikum rechts der Isar, Technische Universität München, Munich, Germany. 9Division of Pediatric Hematology and Oncology, University Medical Center Aachen, Aachen, Germany. 10Department of Human Genetics, University Hospital Essen, Essen, Germany. 11Division of Pediatric Hematology and Oncology, Department of Pediatrics, University Medical Center Freiburg, Freiburg, Germany. 12Department of Pediatric Oncology, Klinikum Kassel, Kassel, Germany. 13Institute of Human Genetics, University of Ulm & University Hospital of Ulm, Ulm, Germany. 14Division of Neuroblastoma Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany. 15Princess Maxima Center for Pediatric Oncology, Utrecht, The Netherlands. 16Innovative Therapies for Children with Cancer Consortium and Department of Clinical Research, Gustave Roussy, Université Paris-Saclay, Villejuif, France. 17Pediatric Hematology and Oncology, University Hospital Münster, Münster, Germany. 18Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany. 19Clinical Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. 20Center for Individualized Pediatric Oncology (ZIPO) and Brain Tumors, University Hospital and German Cancer Research Center (DKFZ), Heidelberg, Germany. 21Division of Pediatric Hematology and Oncology, University Medical Center Göttingen, Göttingen, Germany. 22Pediatric Oncology & Hematology, Pediatrics III, University Hospital of Essen, Essen, Germany. 23Division of Pediatric Hematology and Oncology, University Children’s Hospital Zurich, Zurich, Switzerland. 24Swedish Children’s Cancer Center, Children’s Hospital, Klinikum Augsburg, Augsburg, Germany. 25Genomics and Proteomics Core Facility, High Throughput Sequencing Unit, German Cancer Research Center (DKFZ), Heidelberg, Germany. 26Hospital for Children and Adolescents, University Hospital Frankfurt, Frankfurt, Germany. 27University Hospital Cologne, Klinik und Poliklinik für Kinder- und Jugendmedizin, Köln, Germany. 28Department of Oncogenomics, Academic Medical Center, Amsterdam, The Netherlands. 29Division of Neurosurgery, Center for Childhood Cancer Research, Department of Biomedical and Health Informatics and Center for Data-Driven Discovery in Biomedicine, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. 30Department of Computational Biology, St Jude Children’s Research Hospital, Memphis, Tennessee, USA. 31Division of Oncology, Center for Childhood Cancer Research, Department of Biomedical and Health Informatics and Center for Data-Driven Discovery in Biomedicine, Children’s Hospital of Philadelphia, Philadelphia, USA. 32Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. 33Institute of Computer Science, Free University Berlin, Berlin, Germany. 34Institute of Medical Genetics and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany. 35Bioinformatics and Omics Data Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany. 36EMBL-EBI, European Molecular Biology Laboratory, Hinxton, Cambridge, UK. 37Bioinformatics and Omics Data Analytics, Bioinformatics and Reference Pathology, Christian-Albrechts-Universität, Kiel, Germany. 38Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. 39Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology and Biobank, University of Heidelberg, Heidelberg, Germany. 40Department of Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany. 41Division of Internal Medicine/Hematology, Friedrich-Ebert-Hospital, Neumünster, Germany. 42Department of Medical School Essen - Pediatric Hematology and Oncology, Münster, Germany. 43University Hospital Giessen, Pediatric Hematology and Oncology and Giessen, Germany. 44Department of Medicine III - Campus Grosshadern, University Hospital Munich, Munich, Germany. 45Department of Hematology and Oncology, Georg-August-University of Göttingen, Göttingen, Germany. 46University Hospital Würzburg, Department of Medicine and Poliklinik II, University of Würzburg, Würzburg, Germany. 47Department of Medicine III, Hematology and Oncology, Dr Horst Schmidt-Kliniken of Wiesbaden, Wiesbaden, Germany. 48SMB Bonn-Institute of Pathology, University of Frankfurt Medical School, Frankfurt am Main, Germany. 49Department of Internal Medicine II, Hematology and Oncology, St-Georg Hospital Leipzig, Leipzig, Germany. 50Department of Pathology, Robert-Bosch-Hospital, Stuttgart, Germany. 51Clinic for Hematology and Oncology, St-Antonius-Hospital, Eschweiler, Germany. 52Department of Internal Medicine, University Hospital of Ulm, Ulm, Germany. 53Clinic of Pathology, Charité – Berlin University Medicine Berlin, Germany. 54Institute of Medical Genetics, St-Antonius-Klinik, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany. 55Institute of Human Genetics, University of Munich and University Hospital of Munich, Munich, Germany. 56Institute of Medical Genetics and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany. 57Institute of Computer Science, Free University Berlin, Berlin, Germany. 58Institute of Medical Genetics and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, Germany. 59Institute of Pathology, Charité – Berlin University Medicine Berlin, Germany. 60Clinic of Hematology and Oncology, St-Antonius-Hospital, Eschweiler, Germany. 61Department of Internal Medicine III, University of Ulm and University Hospital of Ulm, Ulm, Germany. 62National Centre for Tumor Disease, Heidelberg, Germany. 63Institute of Cell Biology (Cancer Research), University of Duisburg-Essen, Duisburg-Essen, Medical School Essen, Germany. 64Department of Internal Medicine, University Hospital Ulm, Ulm, Germany. 65Department of Hematology and Oncology, Robert-Bosch-Hospital, Stuttgart, Germany. 66Clinic of Pathology and Oncology, St-Antonius-Hospital, Eschweiler, Germany. 67Department of Pathology, Robert-Bosch-Hospital, Stuttgart, Germany. 68Clinic of Hematology and Oncology, St-Antonius-Hospital, Eschweiler, Germany. 69Institute of Cell Biology (Cancer Research), University of Duisburg-Essen, Duisburg-Essen, Medical School Essen, Germany. 70Institute of Medical Informatics and Epidemiology, University of Leipzig, Leipzig, Germany. 71Clinic of Hematology and Oncology, University Hospital Ulm, Ulm, Germany. 72Clinic of Internal Medicine I, Leibniz Universitaet Hannover, Hannover, Germany. 73Institute of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. 74Bioinformatics Group, Department of Computer, University of Leipzig, Leipzig, Germany. 75EpiCancerBioinformatics, LIFE Research Center for Civilized Diseases, University of Leipzig, Leipzig, Germany. 76Division of Applied Bioinformatics (G200), German Cancer Research Center (DKFZ), Heidelberg, Germany. 77Department of Pediatric Hematology and Oncology, University Hospital, Heidelberg, Germany. 78Institute for Medical Informatics and Epidemiology, University of Leipzig, Leipzig, Germany. 79EMBL-Heidelberg, Genome Biology, Heidelberg, Germany. 80Bioinformatics and Omics Data Analytics (B240), German Cancer Research Center (DKFZ), Heidelberg, Germany. 81RNNomics Group, Fraunhofer Institute for Cell Therapy and Immunology IZI, Leipzig, Germany. 82Santa Fe Institute, Santa Fe, New Mexico, USA. 83Max-Planck-Institute for Mathematics in Sciences, Leipzig, Germany.
METHODS

Samples. The cohort analysed in this study is a compilation of individual sequencing datasets from various sources: the International Cancer Genome Consortium (ICGC) – Pedbrain Tumor and MMML-seq (http://www.icgc.org), the German Cancer Consortium (DKTK) (https://dktk.dkfz.de/en/home), the Pediatric Cancer Genome Project (PCGP) (http://explore.pediatriccancergenomeproject.org/), the Heidelberg Institute for Personalized Oncology (HIPO) (http://www.dkffz.de/en/hipo), the Individualized Therapy For Relapsed Malignancies in Childhood (INFOR, http://www.infor-dkffz.de/eniform), and other previously published datasets (listed below). For all included tumours, matched germline control tissue was available. Ninety-five per cent of the patients were under 18 years of age (or age unspecified but confirmed age group paediatric), but available data were included for patients up to 25 years, as these were considered relevant for cancer types that typically peak at a young age. All centres have approved data access and informed consent had been obtained from all patients.

External data were downloaded from the European Genome-Phenome Archive (EGA; https://www.ebi.ac.ukega/home) using the accessions numbers EGAD00001000085, EGAD00001000135, EGAD00001000159, EGAD00001000160, EGAD00001000161, EGAD00001000162, EGAD00001000163, EGAD00001000164, EGAD00001000165, EGAD00001000259, EGAD00001000260, EGAD00001000261, EGAD00001000268, and EGAD00001000269. v.82, internal datasets are related to previous PMIDs 27747848, 27479119, 26288771, 25570903, 22972766, 24553142, 25135868, 26632267, 26179511, 24615105, 28726821, 23817572, 25962120, 26294725. Supplementary Note 1).

The final cohort included 914 individual patients of no more than 25 years of age including primary tumours for 879 patients with 47 matched relapsed tumours, and an additional 35 independent relapsed tumours (Supplementary Tables 1, 2). Deep-sequencing (~30 ×) whole-genome data (WGS) were available for 547 samples with matched control, whole-exome sequencing (WES) for 414, and low-coverage whole-genome sequencing (lcWGS) for an additional 54 germline and 186 tumour samples. Depending on the requirements of each sub-analysis, we used WES and WGS, WGS only (excluding Ewing’s sarcoma, Wilms tumour, hepatoblastoma, and T-ALL), or WES, WGS and lcWGS (germline excluding Ewing’s sarcoma, Wilms tumour and hepatoblastoma; tumours excluding Ewing’s sarcoma and hepatoblastoma) were used (Supplementary Table 2). Subgroups of cancer types were considered as separate entities if there was significant evidence of differences in terms of clinical and molecular behaviours, if sub-cohort sizes were substantial, and if full annotation of all samples was available. All samples had been sequenced using Illumina technology and 99% of samples were paired-end sequences with 100 bp read length. Ninety-eight per cent of exon sequences are covered with at least 30×, 94% with at least 60×, and the total median exome coverage is 121×. The whole-genome sequenced samples have a median coverage of 37× and 94% of samples are covered with at least 30×. Information on coverage and other metrics for all samples are provided in Supplementary Table 2.

Cancer type incidence. Information on incidence of cancer types in the population was derived from the SEER database (Surveillance, Epidemiology, and End Results program)5, further detailed information on different subgroups of cancer types (central nervous system tumours and subgroups of medulloblastoma, ependymoma, and ALL) was transferred from cancer type-specific publications6–9. Survival data are based on information from the German Childhood Cancer Registry10. Incidence rates of adult cancers were taken from information in the population-based cancer registries (central nervous system tumours and medulloblastoma) and from the SEER database (Surveillance, Epidemiology, and End Results program)6. As proposed by Alexandrov et al.11, the mutational profile of a tumour is expected to reflect a superposition of mutational processes (signatures) acting on its genome, where each mutational process has a different intensity (exposure). For a cohort of tumour genomes, this is modelled as a system of matrices for signatures (P) and exposures (E) defining the observed mutational catalogue (M)5. M ≈ PE. De novo deciphering of signatures was done as described12 based on the mutational characterizations of all samples of the pan-cancer cohort. All resulting signatures were compared to published signatures (available in the COSMIC database, http://cancer.sanger.ac.uk/cosmic/signatures) based on their cosine similarity13. Signatures that did not correspond to any of the previously known signatures (cosine similarity < 0.85) were further analysed to examine their relevance for modelling the cancer genomes. First, linear independence of the known set of signatures was confirmed. Second, for each potentially novel signature, we examined whether the modelling of mutational profiles improved when compared to having used the set of known signatures for each sample, the observed mutational signature was compared to the theoretical profiles calculated using the known set of signatures only, and using the extended set including the new candidate signature. Here, only samples with a total number of mutations over 200 were considered. Reconstruction was calculated as the difference between cosine similarity of the modelled profile and the observed profile. On the basis of the resulting distribution of similarities in both alternatives, a signature was considered to have a relevant contribution to the model, and thus a potential new signature,
if both of the following conditions were fulfilled: the reconstruction (measured as the difference of similarities) of at least one sample increased by 0.02 and that sample had a reconstruction accuracy of c 0.9 based on the known set of signatures only.

This procedure resulted in one new candidate signature, signature P1, which was added to the set of reference signatures. In order to achieve maximum resolution per sample, a sample-wise re-extraction of exposures from the mutational profiles was performed using quadratic programming with the reference signature set used for P and the exposures in E as unknown variables. Samples with a reconstruction accuracy below 0.5 were excluded (resulting in n = 503 tumours with high-quality signature information), as these samples would not be correctly accounted for by the model, which might be due to quality issues or to contributions of unknown signatures that are not present at intensities sufficient to be identified by a de novo approach. The resulting exposures were used for further downstream analyses and visualization. Previously published signatures without validation were first included to model the mutational catalogues as precisely as possible, but then summarized as ‘other’ for representation.

Spearman’s rank correlation and two-sided Kolmogorov–Smirnov tests were used to associate exposure of signatures with numerical and categorical variables, respectively. Exposures to signatures across multiple groups were compared using ANOVA and the post hoc Tukey’s test.

Identifying mutations in genes predisposing to cancers. To identify germline variants with a high likelihood of being implicated in cancer development, we investigated 162 candidate genes adapted from ref. 19 (110 genes regarded as following a dominant inheritance pattern and 52 genes with recessive inheritance) (Supplementary Table 6).

Germline SNVs and indels were subjected to a stepwise filtering approach to eventually classify them into five categories: benign, likely benign, uncertain significance, likely pathogenic, and pathogenic. First, variants reported in both the 1000 Genomes (release November 2010) and dbSNP (v.141) databases were excluded. High-quality variant calls were selected by including only positions with ≥15× coverage, a germline allele frequency of ≥0.02, and a phred-based quality score of ≥10. Variants with a population frequency ≥0.01 reported in additional common databases (esp6500si.v2, XH100g15, and exac03 included in ANNOVAR (http://annovar.openbioinformatics.org)) or with ClinVar (ftp://ftp.ncbi.nlm.nih.gov/clinvar/) annotations of ‘benign’, ‘likely benign’ or ‘uncertain significance’ were removed.

Furthermore,variants with a phred-scaled CADD score ≥15 (http://cadd.gs.washington.edu/info) and with Mutation Assessor (http://mutationassessor.ncbi.nlm.nih.gov/pub/clinvar/) annotations of ‘benign’, ‘likely benign’ or ‘uncertain significance’ were removed.

To identify germline focal deletions (Supplementary Table 6). Every variant was then manually checked and scored as pseudo marker positions 96. All possible marker positions were determined from the ‘SEG’ segmentation format, similar to the output of the Genome instability. Occurrence of chromothripsis was determined by manual inspection of coverage ratio plots (tumour/control) for WGS samples based on previously proposed guidelines 95: at least ten copy-number switches on one chromosome, oscillating copy-number variation (usually with changes of +1 or −1, but also between other levels where additional large-scale copy-number changes interfere), and many more of such copy-number variations in one chromosome or chromosome arm compared to the remaining genome. In samples with an exceptionally high degree of structural variation, several chromosomes could be affected, and some samples showed an ‘amplifier’ type of chromothripsis, which was classified as several high-level focal amplifications on exactly the same copy-number level that are thus likely to be connected to one single event.

Identifying recurrent copy-number/structural variations. GISTIC2.0 (v.2.0.22, gene-gistic default parameter settings) was applied to the segmented copy-number data (per cancer type and pan-cancer) to identify significant copy-number alterations. The resulting peaks were filtered for significance (q ≤0.1) and size (≤10 Mb). Compared to array-based data, which commonly serve as inputs for copy-number significance analysis, sequencing-based copy-number profiles are more prone to artefact copy-number variations, for example, due to repetitive regions leading to ambiguous alignments. Thus, several filtering steps were used to eliminate false-positive GISTIC peak calls and to discover potentially cancer-relevant copy-number alterations: first, peaks overlapping with common fragile genomic sites were excluded, as these are likely to be consequences of genomic instability rather than cancer-driving events 97; next, peaks overlapping within 1 Mb of chromosomal ends were removed, as these regions often overlap with genomic breaks of breakpoints in ≥5 samples in total or ≥2 samples of one cancer type (for samples without chromothripsis). For other samples, genes with breakpoints in ≥5 samples were included as candidates, but these were not used for further downstream analyses. Additionally, recurrent sites of structural variation outside of gene bodies by clustering breakpoints were determined in 10-kb windows.
Scoring of druggable mutations. To identify candidates for targeted therapy, somatic and germline mutations (SNV and indels) were screened for variants in genes that are directly or indirectly involved in pathways with matched drugs either approved or currently being investigated in clinical trials (Supplementary Table 22a, adapted from ref. 19). The mutations were then manually assessed by experts in translational oncology and prioritized according to an internal algorithm taking into account the type of alteration, the mechanism of action of potential drugs within the pathway, the level of evidence for the specific alteration, and its role in the presented cancer types (Supplementary Table 22b, adapted from ref. 19). Only alterations scored ‘intermediate’ or ‘high’ were regarded as being relevant in terms of druggability. A clonality analysis was not performed owing to limited sequencing depth in whole-genome-sequenced tumours.

Additionally, copy-number plots of whole-genome-sequenced data (including low-coverage WGS) were used to manually screen 52 druggable genes for amplifications or deletions (Supplementary Table 2a). Only focal CNVs (<10 Mb) with at least 5 copies (log2 ≥ 1.3) in the case of amplifications or the loss of >1 copy (log2 ≤ −1) for deletions were included and subsequently prioritized as described for the SNVs/indels. The data representation includes all tumours with full genomic information (WES = WGS or WGS; n = 675) and, additionally, tumours analysed by WES only for cancer types without any whole-genome-sequenced tumours (T-ALL, Ewing’s sarcoma, HB; n = 39), but the latter were excluded from downstream analyses.

Data availability. Mutation data have been deposited into commonly used public data portals and are accessible at http://pedpancan.com. They can be explored in and downloaded from the R2 Analysis and Genomics Platform, the PedBio Center for Cancer Visualization, and the TARGET Data Matrix. Sequencing data were obtained from previous studies as listed in Supplementary Note 1 and include the following accession codes: PR018216, PRJEB11430 (European Nucleotide Archive); EGAS00001001139, EGAS00001001953, EGAS00001000607, EGAS00001000381, EGAS00001000906, EGAS00001001297, EGAS00001000443, EGAS00001000213, EGAS00001000263, EGAS0000100192, EGAS00001000255, EGAS0000100254, EGAS0000100253, EGAS0000100246, EGAS0000100379, EGAS0000100380, EGAS00001000346, EGAS00001000349, EGAS0000100347, EGAS0000100192 (European Genome-Phenome Archive).

54. Wu, G. et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat. Genet. 44, 251–253 (2012).
55. Wu, G. et al. The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. Nat. Genet. 46, 444–450 (2014).
56. Cheung, N. K. et al. Association of age at diagnosis and genetic mutations in patients with neuroblastoma. J. Am. Med. Assoc. 307, 1062–1071 (2012).
57. Chen, X. et al. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. Cell Rep. 7, 104–112 (2014).
58. Pinto, E. M. et al. Genomic landscape of paediatric adrenocortical tumours. Nature 495, 602–612 (2013).
59. Ostrom, Q. T. et al. Alex’s Lemonade Stand Foundation infant and childhood primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. Neuro-oncol. 16 (Suppl 10), x1–x36 (2015).
60. Rausch, T. et al. Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups. Cancer Cell 27, 728–743 (2015).
61. Northcott, P. A. et al. Medulloblastomics: the end of the beginning. Nat. Rev. Cancer 12, 819–834 (2012).
62. Stein, L. D., Knoppers, B. M., Campbell, P., Getz, G. & Korbel, J. O. Data analysis: Create a cloud commons. Nature 523, 149–151 (2015).
63. Jones, D. T. et al. Dissecting the genomic complexity underlying medulloblastoma. Nature 488, 100–105 (2012).
64. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010).
65. Rausch, T. et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28, 1333–1339 (2012).
66. Silver, A. et al. A global reference for human genetic variation. Nature 526, 68–74 (2015).
67. Hingorani, D. B. & Donnelly, P. A. Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).
68. Olishansky, A. B. et al. Parent-specific copy number in paired tumor-normal studies using circular binary segmentation. Bioinformatics 27, 2038–2046 (2011).
69. van den Meersche, K., Soetaert, K. & Van Oevelen, D. example() An R function for sampling linear inverse problems. J. Stat. Softw. 30, 1–15 (2009).
70. Roberts, S. A. et al. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol. Cell 46, 424–435 (2012).
71. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. Cell 154, 246–299 (2013).
72. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483, 603–607 (2012).
73. Futreal, P. A. et al. A census of human cancer genes. Nat. Rev. Cancer 4, 177–183 (2004).
74. Vohra, S. & Biggin, P. C. Mutationmapper: a tool to aid the mapping of protein mutation data. PLoS ONE 8, e71711 (2013).
75. Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. Nature 456, 927–932 (2015).
76. Olishansky, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5, 557–572 (2004).
77. LeBelle, B. et al. Creating fragile site profiling in epithelial and erythroid cells reveals that most recent cancer deletions lie in fragile sites hosting large rearrangements. Cell Rep. 4, 420–428 (2013).
78. Sudmant, P. H. et al. An integrated map of structural variation in 2,504 human genomes. Nature 526, 75–81 (2015).
Extended Data Figure 1 | Somatic mutation frequencies and spectra.

a, Patient age at diagnosis (black lines, median); asterisks highlight cancer types with significant correlation of mutation load with age within cancer types (grey, SNVs; black, indels). b, Correlation of SNV and indel loads (left, median per cancer type; right, cross-cohort, n = 876). c, Somatic mutation frequencies (top, indels; bottom, SNVs) in the coding region (n = 879) (black lines, median). d, Mutation loads in primary versus relapse tumours (cross-cohort n = 958, per cancer type, see Supplementary Table 1; two-sided t-test, confidence interval 0.95). n.s., not significant; NA, not applicable. e, Correlation of mutations (SNVs and indels) with age (left, median per cancer type; right, cross-cohort n = 876). f, Proportion of tumours with one or several events of localized hypermutation (WGS samples, n = 540). g, Mutation spectra of SNVs (top, per sample; bottom, average per cancer type; n = 879). Distributions of frequencies per substitution type across cancer types are indicated on the right; outliers are highlighted in the heat map (quartiles, range of whiskers: 1.5 × interquartile range). a, b, e, Linear model, confidence interval 0.95. Hypermutators and ultramutators are considered only in c.
Extended Data Figure 2 | Mutational signatures in paediatric cancer types. a, Summarized contribution of signatures to mutational profiles per cancer type (proportion of mutations per signature and cancer type). Signatures with contributions of ≥5% in at least one cancer type are shown. The colour intensity reflects the relative activity of each signature per cancer type. b, Correlation of signature 1 with patient age per cancer type in this paediatric pan-cancer cohort (left, \( n = 503 \)) compared to results from a global pan-cancer study on 30 cancer types (\( n = 7,042 \))\(^{15} \). c, Relative contributions of mutational signatures to somatic mutations per individual tumour, clustered within cancer types (\( n = 503 \)). d, Correlation of signatures 3, 8, and 13 (somatic mutations) with genome instability (structural variants) per cancer type. e, Substitution type probabilities in trinucleotide context for the newly discovered mutational signature P1; contribution of signature P1 per tumour (\( n = 503 \)); correlation of signature P1 with multiple nucleotide variants (MNVs); activity of signature P1 in ATRT subgroups (Wilcoxon rank-sum test, confidence interval 0.95). b–d. Spearman's correlation, confidence interval 0.95.
Extended Data Figure 3 | Association of mutational signatures with genomic instability. 

(a) Correlation of signatures with the number of structural variants across all tumours and selected cancer types (Spearman’s correlation, confidence interval 0.95).

(b) Association of signatures with chromothripsis across all tumours and within selected cancer types. TP53 mutation status (germline/somatic) is highlighted (Kolmogorov–Smirnov test, confidence interval 0.95, range of whiskers: 1.5 × interquartile range).

(c) Association of signatures with TP53 mutation status (germline/somatic/none) across all tumours and within selected cancer types (ANOVA and post hoc Tukey’s test, confidence interval 0.95, quartiles, range of whiskers: 1.5 × interquartile range).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Extended Data Figure 4 | Characteristics of significantly mutated genomic regions and genes. 

a, Precision-recall curves (mean precision) for various binomial P value cut-offs for the identification of genome-wide mutation clusters. 
b, Manhattan plot for the test statistic of genomic windows. Dashed line indicates the P value cutoff from a. 
c, Significant co-occurrence/mutual exclusivity of SMGs in the pan-cancer dataset (n = 876). 
d, Most frequently mutated genes from c. 
e, Mutations in SMGs selected in d per cancer type. 
f, Allele frequencies of mutations in SMGs compared to mutations in non-SMGs in n = 876 tumours (two-sided t-test, confidence interval 0.95, quartiles, range of whiskers: 1.5 × interquartile range). 
g, SMGs identified from relapse tumours and representation in cancer types.
Extended Data Figure 5 | Significantly mutated genes across age groups. a, Cellular processes associated with paediatric (left) and adult (right) SMGs. b, Frequency of mutations in SMGs in paediatric (n = 879) compared to adult (n = 3,281) cancers. Top, percentage of SMG-mutated samples. Bottom, mutations in SMGs per sample (centre, median; range, minimum to maximum). c, Overlap of SMGs detected in paediatric and adult cancers. d, Projected mutation rates of SMGs based on normalization of the cohort frequencies for cancer type incidence among patients for paediatric and adult cancers. a–d, Information on adult SMGs is based on TCGA data and previous analysis 7.
Extended Data Figure 6 | Mutation needle plots for significantly mutated genes. Mutations in selected significantly mutated genes across pan-cancer cohort: missense (green), truncating (black), in-frame (blue), and other (purple). Hotspot amino acid changes are highlighted. a, Genes with tumour suppressor-like mutation patterns. b, Genes with oncogenic or oncogene-like mutation patterns.
Extended Data Figure 7 | Genomic instability across paediatric cancer types. 

**a**, Structural variant load in relation to TP53 mutation status for individual cancer types (generalized linear model, confidence interval 0.95). 

**b**–**h**, Characteristics of genomic instability (left) and their associations with TP53 mutation status (right) (n.s., not significant). 

**b**, Genome ploidy; density of ploidy across all lineages is summarized on the right. 

**C**, Co-occurrence (Fisher’s exact test) of hyperdiploidy (cross-cohort, \( n = 516 \)) and TP53 mutations (left, somatic; right, germline). 

**d**, Percentage of tumours per cancer type with hyper- (\( \geq 1.5 \)) and hypodiploid (\( \leq 0.5 \)) genomes. 

**e**, Rate of hypodiploidy in relation to TP53 mutation status (left, cross-cohort; right, cancer type-specific (\( n_{SHH} = 38 \)) with co-occurrence highlighted as in b). 

**f**, Rate of chromothripsis (positive/negative). 

**g**, Rate of chromothripsis in relation to TP53 mutation status (left, cross-cohort; right, cancer type-specific (\( n_{SHH} = 38 \)) with co-occurrence highlighted as in b). 

**h**, Cross-cohort (\( n = 516 \)) co-occurrence of samples with chromothripsis and TP53 mutations (top, somatic; bottom, germline).
Extended Data Figure 8 | Recurrent CNVs and structural variations.

a, Genome-wide copy-number profiles normalized for tumour ploidy (n = 516). Cancer types are sorted by genome instability (Fig. 5a). Regions or genes with significant CNVs are indicated (blue, deleted; red, gained or amplified) (Fig. 5b). b, Relative copy-number status (normalized for tumour ploidy to baseline 1) for regions with significant copy-number changes (top, gains or amplifications; bottom, deletions) in n = 516 tumours. Thresholds (amplified: ≥ 1.4, deleted: ≤ 0.6) are based on the overall copy-number distribution indicated on the right. c, Genes affected by breakpoints from structural variants and additional genes associated with clustered breakpoints (in square brackets). Samples are divided into sub-cohorts of tumours with (bottom, n = 73) and without (top, n = 455) chromothripsis. Genes overlapping (direct overlap or within ±200 kb) with genes with significant copy-number changes from a (blue, deletions; red, amplifications).
Extended Data Figure 9 | Averaged copy-number profiles per cancer type. Averaged copy-number profiles for all cancer types ordered by genome instability (Fig. 5a) and significant regions (Fig. 5b). The x-axis represents chromosomal positions in 1-kb windows and the y-axis the log₂ coverage of tumours versus controls. Asterisks indicate in which cancer types a region was called significant (amplifications, red and above copy-number profiles; deletions, blue and below profiles).
Extended Data Figure 10 | Genetic events define mutation classes. a, Genes significantly or recurrently affected by mutations, amplification, deletions, and gene-disrupting structural variants (likely functional events, LFEs). Copy-number and structural variations are summarized as SC-class in contrast to mutations (SNVs or indels) as M-class. b, Number of SC-class (x-axis) and M-class (y-axis) alterations per tumour. c, Proportion of events from M-class and SC-class within each tumour. Tumours with more than 50% (mixed) or 100% (unique) events from one category are considered to be members of the associated class; tumours with equal contributions from both categories are ‘ambiguous’, and tumours without an LFE are assigned class ‘none’ (not shown). Colours indicate germline mutations per tumour. d, Fraction of tumours assigned to different classes per cancer type.
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

 Experimental design

1. Sample size
   Describe how sample size was determined.
   Sample size was determined by tumor/control samples available.

2. Data exclusions
   General: Samples were excluded if the sequencing type was not suitable for a particular analysis, as shown in Fig. 1b.
   Copy-number calling: samples were excluded if no high-quality calls could be made as for example for samples with noisy coverage; regions surrounding centromeres and coverage artifacts were excluded from analyzing significant copy-number changes;
   Mutational signatures: directly adjacent mutations were excluded for calculating signatures; samples with a reconstruction accuracy <0.5 were excluded from any downstream analysis; for evaluating the model samples with <200 mutations were excluded
   Germline analysis: mutations reported in the 1000 genomes release and dbSNPv141 database were excluded

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   No experiments were performed.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   No randomization was done.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   Investigators were not blinded to allocation.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed |
|-----|-----------|
| ☑ | ☑ |

- The exact sample size \( n \) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. \( P \) values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

7. Software

Policy information about availability of computer code

Describe the software used to analyze the data in this study.

All software used is described in the methods section. Publicly available software included: sambamba, SamToFastq, bwa-mem, samtools, platypus, delly, R, ACEseq, impute2, genome music, gistic2.0

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used.

NA

c. Report whether the cell lines were tested for mycoplasma contamination.

NA

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

NA

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used.
12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Data were obtained from previously published sequencing studies and available metadata are provided in Suppl. Table 2.
Author Correction: The landscape of genomic alterations across childhood cancers

Susanne N. Gröbner, Barbara C. Worst, Joachim Weischenfeldt, Ivo Buchhalter, Kortine Kleinheinz, Vasilisa A. Rudneva, Pascal D. Johann, Gnana Prakash Balasubramanian, Maia Segura-Wang, Sebastian Brabetz, Sebastian Bender, Barbara Hutter, Dominik Sturm, Elke Pfaff, Daniel Hübschmann, Gideon Zipprich, Michael Heinold, Jürgen Eils, Christian Lawerenz, Serap Erkek, Sander Lambo, Sebastian Waszak, Claudia Blattmann, Arndt Borkhardt, Michaela Kuhlen, Angelika Eggert, Simone Fulda, Manfred Gessler, Jenny Wegert, Roland Kappler, Daniel Baumhoer, Stefan Burdach, Renate Kirschner-Schwabe, Udo Kontny, Andreas E. Kuolzik, Dietmar Lohmann, Simone Hettner, Cornelia Eckert, Stefan Bielack, Johannes Schulte, Reiner Siebert, Frank Westermann, Jan I. Molenaar, Gilles Vassal, Hendrik Witt, ICGC PedBrain-Seq Project, ICGC MMML-Seq Project, Birgit Burkhardt, Christian P. Kratz, Olaf Witt, Cornelis M. van Tilburg, Christof M. Kramm, Gudrun Fleischhack, Uta Dirksen, Stefan Rutkowski, Michael Frühwald, Katja von Hoff, Stephan Wolf, Thomas Klingebiel, Ewa Koscielniak, Pablo Landgraf, Jan Koster, Adam C. Resnick, Jinghui Zhang, Yanling Liu, Xin Zhou, Angela J. Waanders, Danny A. Zwijnenburg, Pichai Raman, Benedikt Brors, Ursula D. Weber, Paul A. Northcott, Kristian W. Pajtler, Marcel Kool, Rosario M. Piro, Jan O. Korbel, Matthias Schlesner, Roland Eils, David T. W. Jones, Peter Lichter, Lukas Chavez, Marc Zapatka & Stefan M. Pfister

Correction to: Nature 10.1038/nature25480, published online 28 February 2018.

In this Article, author Benedikt Brors was erroneously associated with affiliation number ‘8’ (Department of Developmental Neurobiology, St Jude Children’s Research Hospital, Memphis, Tennessee, USA); the author’s two other affiliations (affiliations ‘3’ and ‘7’, both at the German Cancer Research Center (DKFZ)) were correct. This has been corrected online.