ARTICLE

OPEN

doi:10.1038/nature25480

The landscape of genomic alterations across childhood cancers

A list of authors and affiliations appears at the end of the paper.

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[1,2]. 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[3]. 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[4]. 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[4–6].

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)[7]. 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[8] (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[9].

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 ($n = 879$) ($R = 0.27$, $P = 9.1 \times 10^{-5}$; 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[7] (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)[10,11]. 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)[7,12] (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[13].

As in previous reports, the somatic mutation burden increased with patient age ($R = 0.39$, $P = 2.9 \times 10^{-6}$), except in Burkitt's lymphoma (immunoglobulin hypermutation) and tumours with 'kataegis' events of localized hypermutation at double-stranded breakpoints[14,15] (Extended Data Fig. 1e, f). Both SNVs ($R = 0.37$, $P = 1.0 \times 10^{-5}$) and indels ($R = 0.27$, $P = 5.4 \times 10^{-4}$) correlated with patient age overall, although within some cancers (for example, acute lymphoblastic leukaeumia (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, $P = 0.002$).

A list of authors and affiliations appears at the end of the paper.

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 tumours[15,16] ($P < 0.05$; Extended Data Figs 1g, 2a–c). Mutational signatures, possibly reflecting biochemical cellular processes, have previously been investigated for many, mainly adult, cancers[15]. 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 signature[15] (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 CC̲C/CC̲T context, occurred in several atypical teratoid rhabdoid tumours (ATRTs) and one ependymoma (Fig. 2, Extended Data Fig. 2d, Supplementary Table 5). Its activity correlated with 'multiple nucleotide variants' (MNVs; $R = 0.87$, $P = 1.1 \times 10^{-12}$), 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 proposed methylation subset of these[17] ($P = 0.003$, Wilcoxon rank-sum test; Extended Data Fig. 2d). Signatures 16 and 18 were heterogeneously 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 variants[15,18] (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 adrenocortical 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 predisposition[5]. 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 variant[5,19] (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 adrenocortical 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 detected[5] (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 inhibition[13] (Figs 1, 3b, c). Carriers of *TP53* germline mutations (Li–Fraumeni syndrome), here most common in adrenocortical 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 oncogenesis[2,20–22] (Fig. 3b). Correcting the predisposition frequency of 7.6% in this cohort for the relative incidence of

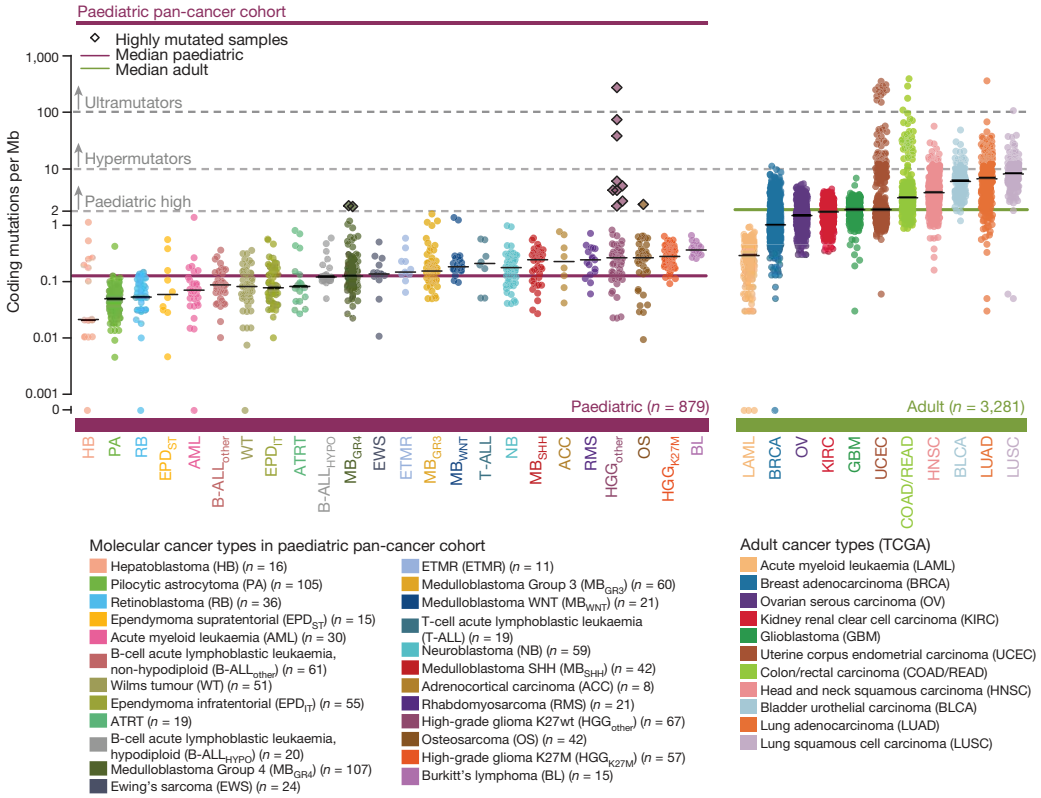


Figure 1 | Somatic mutations in the paediatric pan-cancer cohort. Somatic coding mutation frequencies in 24 paediatric ($n = 879$ primary tumours) and 11 adult ($n = 3,281$) cancer types (TCGA)[7]. Hypermutated and highly mutated samples are separated by dashed grey lines and highlighted with black squares. Median mutation loads are shown as solid lines (black, cancer types; purple, all paediatric; green, all adult).

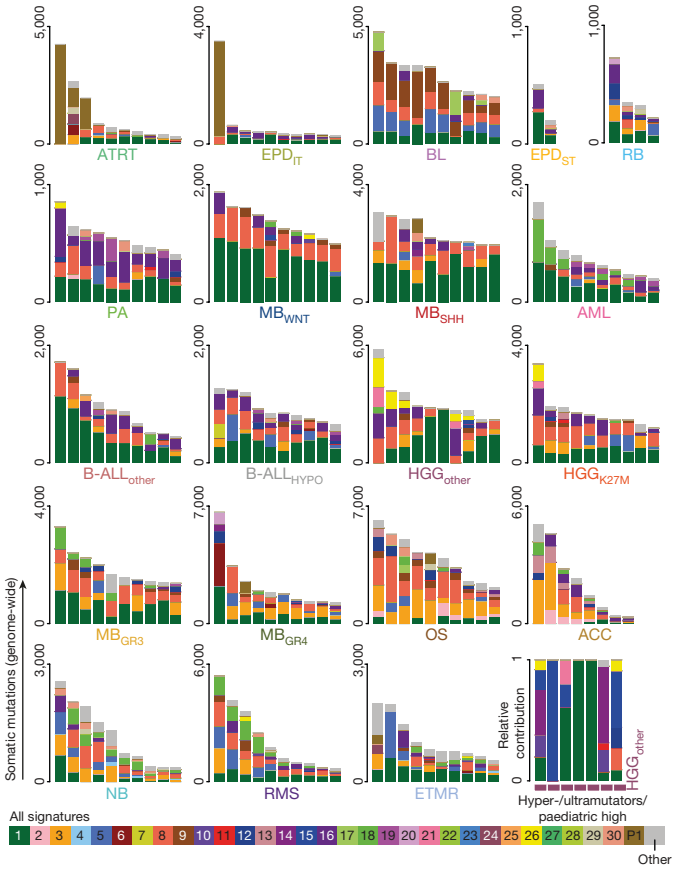


Figure 2 | Mutational processes active in paediatric cancers. Contributions of thirty known and one novel mutational signature to the somatic mutations for the ten most frequently mutated samples per cancer type; each bar represents one individual tumour.

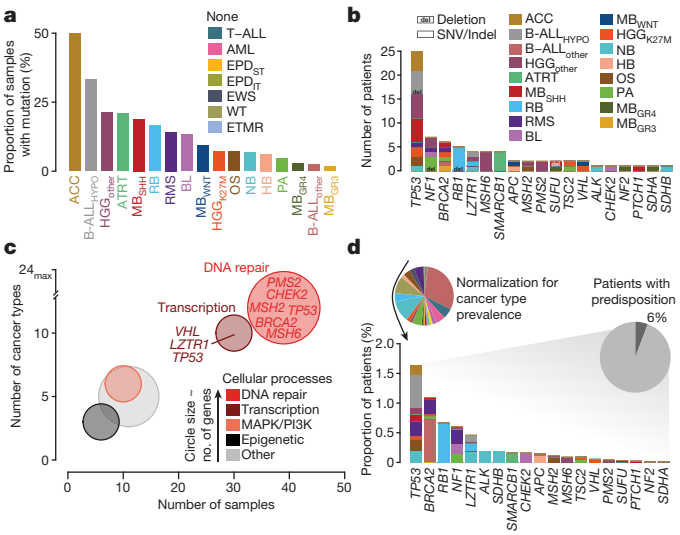


Figure 3 | Germline mutations in cancer predisposition genes. **a**, Frequency of patients with a pathogenic germline mutation per cancer type ($n = 914$ tumours). **b**, Mutated genes sorted by number of affected samples (del, copy-number alterations; others, SNVs/indels). **c**, Cellular processes associated with cancer predisposition genes. **d**, Frequency of germline mutations adjusted for incidence and estimated total proportion of childhood cancers likely to be linked to hereditary predisposition.

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 (Extended Data Fig. 4a, b, Supplementary Table 8). Further high-confidence clusters corresponded to coding mutations in frequently mutated genes (*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[23] (Extended Data Fig. 4b).

MuSiC identified 77 significantly mutated genes (SMGs), which were ranked according to their pan-cancer mutation frequency[24] (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] (Extended Data Fig. 4c–e). None of the SMGs showed a bias towards samples with higher mutation frequencies. The allele frequencies of mutations in SMGs were higher than in non-SMGs, and ranked higher in individual tumours, suggesting an early clonal occurrence of these likely driver events (Extended Data Fig. 4f). Two additional SMGs emerged from analysis of the relapse tumours ($n = 82$): *PRPS1* and *NT5C2*, both of which have been previously implicated in disease progression and chemotherapy resistance[25,26] (Extended Data Fig. 4g).

Genes linked to epigenetic modification emerged as the most common (25% of tumours, 23 of 24 cancer types) and the largest (20%) group of SMGs (Extended Data Fig. 5a). Compared to a previous study[6], for example, we also detected *ARID1A* and *BCOR*. Transcriptional 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% of SMGs, TCGA)[7]. PI3K-associated SMGs are the most commonly altered (31%) genes in adult cancers, compared to only 3% 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%) in ependymomas, hepatoblastomas, Ewing's sarcomas (driven by *EWSR1* fusions instead of by point mutations[28]), and pilocytic astrocytomas, and common (>90%) in K27M high-grade gliomas, WNT medulloblastomas, and Burkitt's lymphomas. By contrast, 93% of adult cancers harbour at least one mutation in an (adult cancer-related) SMG and 76% in multiple SMGs[7] (Extended Data Fig. 5b). In line with the accompanying paediatric pan-cancer study[9], only around 30% of paediatric SMGs overlapped with adult SMGs (Extended Data Fig. 5c). On the basis of incidence-normalized mutation frequencies, *TP53* is predicted to be the most common somatically mutated gene (4% of childhood tumours), followed by *KRAS*, *ATRX*, *NF1,* and *RB1* (1–2% of tumours); in adult cancers, with similarly normalized data, *TP53* is also the most commonly mutated gene, albeit ten times more frequently (Extended Data Fig. 5d).

Assessment of high functional impact mutations (OncodriveFM)[29] revealed well-known tumour suppressor genes (TSGs) such as *TP53*, *ATRX, SMARCA4*, and *RB1*, and further putative TSGs, including *FMR1* in SHH/WNT medulloblastomas and *MALRD1* (also known as *C10orf112*) in rhabdomyosarcomas (Extended Data Fig. 6a). Locally clustered 'hotspot mutations' (OncodriveClust)[29,30] identified known oncogenes, such as *CTNNB1*, *PIK3CA*, *KRAS,* and *BRAF*, proposed oncogenes (*ACVR1, KBTBD4, TBR1*), and possible new candidates, such as *SF3B1*, in Group 4 medulloblastomas (Extended Data Fig. 6b).

Recurrent structural and copy-number variants

The degree of genomic instability (that is, the number of structural variants, including insertions, deletions, translocations, and inversions), varied substantially (median 1–434 structural variants) across

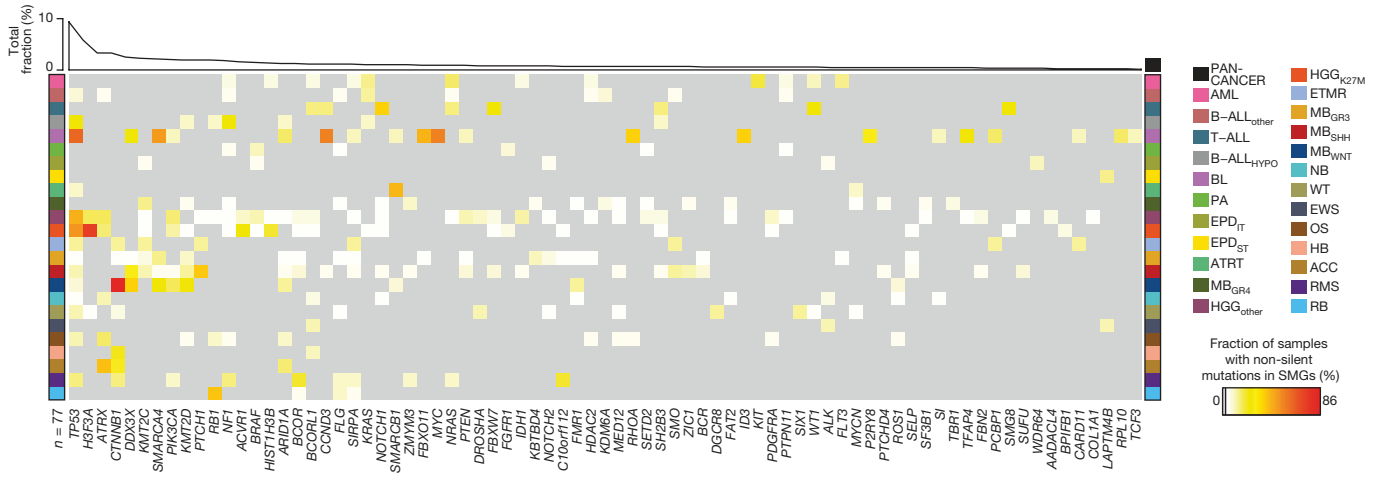


Figure 4 | Significantly mutated genes in paediatric compared to adult cancer types. Percentage of tumours with non-silent mutations in 77 SMGs for 24 paediatric tumour types ($n = 879$ tumours) and the pan-cancer cohort.

cancer types (WGS, $n = 539$), with more than 1,000 structural variants in individual samples of adrenocortical carcinoma and osteosarcoma (Fig. 5a, Supplementary Table 11). Genomic instability correlated with germline ($P = 3 \times 10^{-15}$) and somatic ($P = 2 \times 10^{-4}$) *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 hypodiploidy was observed mainly in hypodiploid B-ALLs (Extended Data Fig. 7b). Hyperdiploidy was associated with somatic ($P = 0.005$) and germline ($P = 0.003$) *TP53* mutations, in line with a role for mutant TP53 in the bypassing of the G1 tetraploidy checkpoint[32] (Extended Data Fig. 7c–e). Chromothripsis was also often observed in hyperdiploid cancers and co-occurred with somatic ($P = 2.3 \times 10^{-10}$) and germline *TP53* ($P = 5 \times 10^{-8}$) mutations in 50% and 66% of these tumours, compared to 8% in *TP53* wild-type tumours[33–35] (Extended Data Fig. 7f–h, Supplementary Table 13).

Thirty-four regions recurrently altered by copy-number changes (17 amplified, 17 deleted) were identified using GISTIC2.0 (WGS, $n = 516$)[36]; candidate driver genes were assigned to each based on known cancer genes and literature review (Fig. 5c, Extended Data Fig. 8a, b, Supplementary Tables 14–17). Alterations per cancer type are summarized in Extended Data Fig. 9.

Recurrently amplified regions contained known oncogenes, including *MYC*, *MYCN*, or *GLI2*, with 11 regions involving high-level amplifications (at least 5-fold gain) (Extended Data Fig. 8b). Further interesting regions included 17q11.2 with 61 genes, containing *NCOR1* as a potential candidate, and a region on 12q24.31 near (~0.1 Mb) the proposed oncogene *KDM2B*[37,38]. Recurrently deleted regions were predominantly associated with epigenetic or cell cycle regulators, most commonly *TP53, PTEN*, *SETD2*, and *CDKN2A* or *CDKN2B*. Further potential tumour suppressors included *RAD51D* on 17q12 and *FOXF1* on 16q24.1, with significant loss across the cohort[39].

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 *RIMS2*[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 20). 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[19] (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.

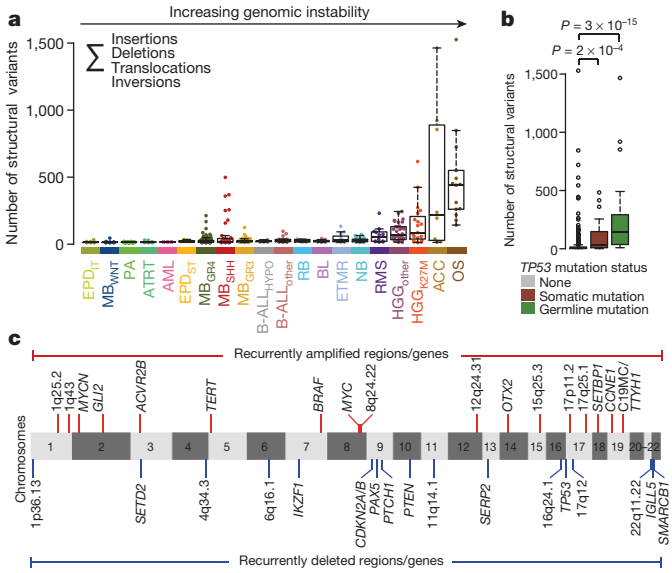


Figure 5 | Genomic instability and recurrent copy-number alterations. **a**, Frequency of structural variants (SVs) across cancer types ($n = 539$ tumours). **b**, Structural variant load from **a** across all tumours in relation to *TP53* mutations (generalized linear model, confidence interval 0.95). **a**, **b**, Quartiles, range of whiskers: 1.5 × interquartile range. **c**, Genomic regions with significant copy-number changes (red, gains or amplifications; blue, deletions; $n = 516$ tumours).

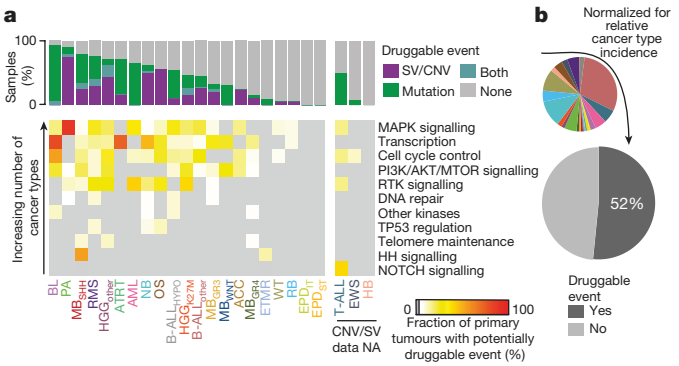


Figure 6 | Potentially druggable events in paediatric cancers. **a**, Proportion of primary tumours with potentially druggable events and associated biological pathways, per cancer type ($n = 675$ tumours with complete genomic information). NA, not available. **b**, Proportion of patients with potentially druggable events, projected after normalization for incidence.

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 study[9] (https://pecan.stjude.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 indications[47].

The predicted frequency of pathogenic germline variants in 6% of patients, together with previous findings, demonstrates the relevance of genetic predisposition in childhood cancer[5]. 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 tumours[13] or through PARP inhibition in genomically unstable ('BRCAness') tumours[48]. 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; INFORM[19]). Additional longitudinal monitoring, for example using serial liquid biopsies, may further improve our understanding of tumour biology and the development of resistance 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.

1. Pui, C. H., Gajjar, A. J., Kane, J. R., Qaddoumi, I. A. & Pappo, A. S. Challenging issues in pediatric oncology. *Nat. Rev. Clin. Oncol.* **8,** 540–549 (2011).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J. Clin.* **66,** 7–30 (2016).
3. Kopp, L. M., Gupta, P., Pelayo-Katsanis, L., Wittman, B. & Katsanis, E. Late effects in adult survivors of pediatric cancer: a guide for the primary care physician. *Am. J. Med.* **125,** 636–641 (2012).
4. Downing, J. R. *et al.* The Pediatric Cancer Genome Project. *Nat. Genet.* **44,** 619–622 (2012).
5. Zhang, J. *et al.* Germline mutations in predisposition genes in pediatric cancer. *N. Engl. J. Med.* **373,** 2336–2346 (2015).
6. Huether, R. *et al.* The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat. Commun.* **5,** 3630 (2014).
7. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502,** 333–339 (2013).
8. Howlader, N. *et al. SEER Cancer Statistics Review, 1975–2012, National Cancer Institute* (National Cancer Institute, SEER Program, NIH, 2014).
9. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* https://doi.org/10.1038/nature25795 (2018).
10. Yuan, Y. *et al.* Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat. Biotechnol.* **32,** 644–652 (2014).
11. Campbell, B. B. *et al.* Comprehensive analysis of hypermutation in human cancer. *Cell* **171,** 1042–1056 (2017).
12. Kandoth, C. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497,** 67–73 (2013).
13. Bouffet, E. *et al.* Immune checkpoint inhibition for hypermutant glioblastoma multiforme resulting from germline biallelic mismatch repair deficiency. *J. Clin. Oncol.* **34,** 2206–2211 (2016).
14. Milholland, B., Auton, A., Suh, Y. & Vijg, J. Age-related somatic mutations in the cancer genome. *Oncotarget* **6,** 24627–24635 (2015).
15. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
16. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47,** 1402–1407 (2015).
17. Johann, P. D. *et al.* Atypical teratoid/rhabdoid tumors are comprised of three epigenetic subgroups with distinct enhancer landscapes. *Cancer Cell* **29,** 379–393 (2016).

18. Pilati, C. *et al.* Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *J. Pathol.* **242,** 10–15 (2017).
19. 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).
20. Sorrell, A. D., Espenschied, C. R., Culver, J. O. & Weitzel, J. N. Tumor protein p53 (TP53) testing and Li–Fraumeni syndrome: current status of clinical applications and future directions. *Mol. Diagn. Ther.* **17,** 31–47 (2013).
21. Kappel, S. *et al.* TP53 germline mutation may affect response to anticancer treatments: analysis of an intensively treated Li–Fraumeni family. *Breast Cancer Res. Treat.* **151,** 671–678 (2015).
22. Heymann, S. *et al.* Radio-induced malignancies after breast cancer postoperative radiotherapy in patients with Li–Fraumeni syndrome. *Radiat. Oncol.* **5,** 104 (2010).
23. Klapproth, K. & Wirth, T. Advances in the understanding of MYC-induced lymphomagenesis. *Br. J. Haematol.* **149,** 484–497 (2010).
24. Dees, N. D. *et al.* MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* **22,** 1589–1598 (2012).
25. Mullighan, C. G. Mutant PRPS1: a new therapeutic target in relapsed acute lymphoblastic leukemia. *Nat. Med.* **21,** 553–554 (2015).
26. Tzoneva, G. *et al.* Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. *Nat. Med.* **19,** 368–371 (2013).
27. Verlaat, W. *et al.* Somatic mutation in *PIK3CA* is a late event in cervical carcinogenesis. *J. Pathol. Clin. Res.* **1,** 207–211 (2015).
28. Delattre, O. *et al.* Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359,** 162–165 (1992).
29. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* **10,** 1081–1082 (2013).
30. Tamborero, D. *et al.* Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* **3,** 2650 (2013).
31. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45,** 1134–1140 (2013).
32. Margolis, R. L., Lohez, O. D. & Andreassen, P. R. G1 tetraploidy checkpoint and the suppression of tumorigenesis. *J. Cell. Biochem.* **88,** 673–683 (2003).
33. Mardin, B. R. *et al.* A cell-based model system links chromothripsis with hyperploidy. *Mol. Syst. Biol.* **11,** 828 (2015).
34. Forment, J. V., Kaidi, A. & Jackson, S. P. Chromothripsis and cancer: causes and consequences of chromosome shattering. *Nat. Rev. Cancer* **12,** 663–670 (2012).
35. Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148,** 59–71 (2012).
36. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12,** R41 (2011).
37. Merve, A. *et al.* Polycomb group gene BMI1 controls invasion of medulloblastoma cells and inhibits BMP-regulated cell adhesion. *Acta Neuropathol. Commun.* **2,** 10 (2014).
38. He, J., Kallin, E. M., Tsukada, Y. & Zhang, Y. The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through $p15^{Ink4b}$. *Nat. Struct. Mol. Biol.* **15,** 1169–1175 (2008).
39. Tamura, M. *et al.* Forkhead transcription factor FOXF1 is a novel target gene of the p53 family and regulates cancer cell migration and invasiveness. *Oncogene* **33,** 4837–4846 (2014).
40. Northcott, P. A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511,** 428–434 (2014).
41. Valentijn, L. J. *et al.* TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. *Nat. Genet.* **47,** 1411–1414 (2015).
42. Veeriah, S. *et al.* The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. *Proc. Natl Acad. Sci. USA* **106,** 9435–9440 (2009).
43. Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED–EZH2 complex. *Mol. Cell* **15,** 57–67 (2004).
44. Northcott, P. A. *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nature* **547,** 311–317 (2017).
45. Kaeser, P. S., Deng, L., Fan, M. & Südhof, T. C. RIM genes differentially contribute to organizing presynaptic release sites. *Proc. Natl Acad. Sci. USA* **109,** 11830–11835 (2012).
46. Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45,** 1127–1133 (2013).
47. Pearson, A. D. *et al.* Implementation of mechanism of action biology-driven early drug development for children with cancer. *Eur. J. Cancer* **62,** 124–131 (2016).
48. Cerrato, A., Morra, F. & Celetti, A. Use of poly ADP-ribose polymerase [PARP] inhibitors in cancer cells bearing DDR defects: the rationale for their inclusion in the clinic. *J. Exp. Clin. Cancer Res.* **35,** 179 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements This project was mainly supported and funded by the German Cancer Consortium (DKTK Pediatric Malignancies Joint Funding Project) and German Cancer Aid (#108128) and Deutsche Kinderkrebsstiftung (German Cancer Childhood Foundation) for the INFORM project. Additional 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) PedBrain Tumor Project and the ICGC MMML-Seq Project (within Program for Medical Genome Research Grants #01KU1002A–#01KU1002J), and the BioTOP Project (#01EK1502B). This work was also supported by an ERC starting grant to J.O.K. (#336045), MMML-MYC-SYS (#0316166I), 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 Bioinformatics Infrastructure (de.NBI) (#031A537A, #031A537C). For technical support and expertise we thank the DKFZ Genomics and Proteomics Core Facility, M. Hain from the Division of Molecular Genetics (DKFZ), N. Jaeger and R. Kabbe from the Department of Pediatric Neurooncology (DKFZ), and S. Oelmez from the Data Management Group (DKFZ). We further thank members and technical staff of the ICGC MMML-Seq (International Cancer Genome Consortium Molecular Mechanisms in Malignant Lymphoma by Sequencing) and the European Renal Tumor Study Group (SIOP-RTSG).

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

Author Information Reprints and permissions information is available at 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. Correspondence and requests for materials should be addressed to S.M.P. (s.pfister@dkfz.de).

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

Susanne N. Gröbner[1,2,3]*, Barbara C. Worst[1,2,3,4]*, Joachim Weischenfeldt[5,6], Ivo Buchhalter[7], Kortine Kleinheinz[7], Vasilisa A. Rudneva[5,8], Pascal D. Johann[1,2,3,4], Gnana Prakash Balasubramanian[1,2,9], Maia Segura-Wang[5], Sebastian Brabetz[1,2,3], Sebastian Bender[1,2], Barbara Hutter[3,7,9], Dominik Sturm[1,2,3,4], Elke Pfaff[1,2,3,4], Daniel Hübschmann[4,9,10], Gideon Zipprich[7], Michael Heinold[7,10], Jürgen Eils[7], Christian Lawerenz[7], Serap Erkek[1,2,3,5], Sander Lambo[1,2,3], Sebastian Waszak[5], Claudia Blattmann[3,11], Arndt Borkhardt[3,12], Michaela Kuhlen[3,12], Angelika Eggert[3,13], Simone Fulda[3,14], Manfred Gessler[15], Jenny Wegert[15], Roland Kappler[3,16], Daniel Baumhoer[17], Stefan Burdach[3,18], Renate Kirschner-Schwabe[3,13], Udo Kontny[3,19], Andreas E. Kulozik[1,3,4], Dietmar Lohmann[3,20], Simone Hettmer[21], Cornelia Eckert[3,13], Stefan Bielack[11], Michaela Nathrath[3,18,22], Charlotte Niemeyer[3,21], Günther H. Richter[3,18], Johannes Schulte[3,13], Reiner Siebert[23], Frank Westermann[3,24], Jan J. Molenaar[25], Gilles Vassal[26], Hendrik Witt[1,2,3,4], ICGC PedBrain-Seq Project, ICGC MMML-Seq Project, Birgit Burkhardt[27], Christian P. Kratz[28], Olaf Witt[1,3,4,29], Cornelis M. van Tilburg[1,3,30], Christof M. Kramm[31], Gudrun Fleischhack[3,32], Uta Dirksen[32], Stefan Rutkowski[33], Michael Frühwald[34], Katja von Hoff[33], Stephan Wolf[35], Thomas Klingebiel[3,36], Ewa Koscielniak[11], Pablo Landgraf[37], Jan Koster[38], Adam C. Resnick[39], Jinghui Zhang[40], Yanling Liu[40], Xin Zhou[40], Angela J. Waanders[41], Danny A. Zwijnenburg[38], Pichai Raman[39], Benedikt Brors[3,7], Ursula D. Weber[3,42], Paul A. Northcott[2,3,8], Kristian W. Pajtler[1,2,3,4], Marcel Kool[1,2,3], Rosario M. Piro[3,42,43,44], Jan O. Korbel[5], Matthias Schlesner[7,45], Roland Eils[7,10], David T. W. Jones[1,2,3], Peter Lichter[3,42], Lukas Chavez[1,2,3]§, Marc Zapatka[42,43]§ & Stefan M. Pfister[1,2,3,4]§

[1]Hopp-Children's Cancer Center at the NCT Heidelberg (KiTZ), Heidelberg, Germany. [2]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [3]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. [4]Department of Pediatric Oncology, Hematology & Immunology, Heidelberg University Hospital, Heidelberg, Germany. [5]European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany. [6]The Finsen Laboratory, Rigshospitalet, Biotech Research and Innovation Centre (BRIC), Copenhagen University, Copenhagen, Denmark. [7]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [8]Department of Developmental Neurobiology, St Jude Children's Research Hospital,

Memphis, Tennessee, USA. [9]Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [10]Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology, Heidelberg University and BioQuant Center, 69120, Heidelberg, Germany. [11]Klinikum Stuttgart - Olgahospital, Zentrum für Kinder-, Jugend- und Frauenmedizin, Pädiatrie, Stuttgart, Germany. [12]Department of Pediatric Oncology, Hematology & Clinical Immunology, University Children's Hospital, Heinrich Heine University, Düsseldorf, Germany. [13]Department of Pediatric Oncology/Hematology, Charité-Universitätsmedizin Berlin, Berlin, Germany. [14]Institute for Experimental Cancer Research in Pediatrics, University Hospital Frankfurt, Frankfurt am Main, Germany. [15]Theodor-Boveri-Institute/Biocenter, Developmental Biochemistry, and Comprehensive Cancer Center Mainfranken, University of Würzburg, Würzburg, Germany. [16]Department of Pediatric Surgery, Research Laboratories, Dr von Hauner Children's Hospital, Ludwig Maximilians University Munich, Munich, Germany. [17]Bone Tumor Reference Center at the Institute of Pathology, University Hospital Basel and University of Basel, Basel, Switzerland. [18]Children's Cancer Research Centre and Department of Pediatrics, Klinikum rechts der Isar, Technische Universität München, Munich, Germany. [19]Division of Pediatric Hematology and Oncology, University Medical Center Aachen, Aachen, Germany. [20]Department of Human Genetics, University Hospital Essen, Essen, Germany. [21]Division of Pediatric Hematology and Oncology, Department of Pediatrics, University Medical Center Freiburg, Freiburg, Germany. [22]Department of Pediatric Oncology, Klinikum Kassel, Kassel, Germany. [23]Institute of Human Genetics, University of Ulm & University Hospital of Ulm, Ulm, Germany. [24]Division of Neuroblastoma Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [25]Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands. [26]Innovative Therapies for Children with Cancer Consortium and Department of Clinical Research, Gustave Roussy, Université Paris-Saclay, Villejuif, France. [27]Pediatric Hematology and Oncology, University Hospital Münster, Münster, Germany. [28]Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany. [29]Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [30]Center for Individualized Pediatric Oncology (ZIPO) and Brain Tumors, University Hospital and German Cancer Research Center (DKFZ), Heidelberg, Germany. [31]Division of Pediatric Hematology and Oncology, University Medical Center Göttingen, Göttingen, Germany. [32]Pediatric Oncology & Hematology, Pediatrics III, University Hospital of Essen, Essen, Germany. [33]Department of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [34]Swabian Children's Cancer Center, Children's Hospital, Klinikum Augsburg, Augsburg, Germany. [35]Genomics and Proteomics Core Facility, High Throughput Sequencing Unit, German Cancer Research Center (DKFZ), Heidelberg, Germany. [36]Hospital for Children and Adolescents, University Hospital Frankfurt, Frankfurt, Germany. [37]University Hospital Cologne, Klinik und Poliklinik für Kinder- und Jugendmedizin, Cologne, Germany. [38]Department of Oncogenomics, Academic Medical Center, Amsterdam, The Netherlands. [39]Division 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. [40]Department of Computational Biology, St Jude Children's Research Hospital, Memphis, Tennessee, USA. [41]Division 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. [42]Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [43]Institute of Computer Science, Freie Universität Berlin, Berlin, Germany. [44]Institute of Medical Genetics and Human Genetics, Charité University Hospital, Berlin, Germany. [45]Bioinformatics and Omics Data Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany.
*These authors contributed equally to this work.
§These authors jointly supervised this work.

International Cancer Genome Consortium PedBrain Project

Coordination (WP1) Peter Lichter[1], Ursula Weber[1], Roland Eils[2]; **Banking (WP2)** Andrey Korshunov[3], Olaf Witt[4,5], Stefan Pfister[4,6]; **Reference Pathology & Quality Control (WP3)** Guido Reifenberger[7], Jörg Felsberg[7]; **Isolation of Analytes (WP4)** Christof von Kalle[8], Manfred Schmidt[9], Cynthia Bartholomä[9], Michael Taylor[10]; **Genomic Sequencing (WP5)** Stefan Pfister[4,6], David Jones[4,6], Peter Lichter[1], Natalie Jäger[4,6], Ivo Buchhalter[2]; **Paired-End Mapping (WP6)** Jan Korbel[11], Adrian Stütz[11], Tobias Rausch[11]; **Methylome Sequencing (WP7)** Bernhard Radlwimmer[1]; **Transcriptome Analysis (WP8)** Marie-Laure Yaspo[12], Hans Lehrach[12], Hans-Jörg Warnatz[12]; **Profiling of small non-coding RNAnome (WP9)** Pablo Landgraf[13], Arndt Borkhardt[13]; **Bioinformatics (WP10)** Benedikt Brors[14], Marc Zapatka[1], Roland Eils[2]; **Bioinformatics (WP11)** Roland Eils[2], Jürgen Eils[15] & Christian Lawerenz[15]

[1]Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [2]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [3]Department of Neuropathology, Heidelberg University Hospital, Heidelberg, Germany. [4]Hopp-Children's Cancer Center at the NCT Heidelberg (KiTZ), Heidelberg, Germany. [5]Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [6]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [7]Department of Neuropathology, Heinrich-Heine-University, Düsseldorf, Germany. [8]Division of Translational Oncology, German Cancer Research Center (DKFZ)/National Center for Tumor Diseases (NCT), Heidelberg, Germany. [9]GeneWerk GmbH, Heidelberg, Germany. [10]Division of Neurosurgery, Hospital for Sick Children, Toronto, Ontario, Canada. [11]Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. [12]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics (MPI-MG), Berlin, Germany. [13]Hematology and Clinical Immunology, University Hospital, Düsseldorf, Germany. [14]Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [15]Data Management Group, German Cancer Research Center (DKFZ), Heidelberg, Germany.

International Cancer Genome Consortium MMML-Seq Project

Coordination (C1) Reiner Siebert[1,2], Susanne Wagner[2], Andrea Haake[2], Julia Richter[2,3], Gesine Richter[2]; **Data Center (C2)** Roland Eils[4,5], Chris Lawerenz[4], Jürgen Eils[4], Jules Kerssemakers[4], Christina Jaeger-Schmidt[4], Ingrid Scholz[4]; **Clinical Centers (WP1)** Anke K. Bergmann[2,6], Christoph Borst[7], Birgit Burkhardt[8,9], Alexander Claviez[6], Martin Dreyling[10], Sonja Eberth[11], Hermann Einsele[12], Norbert Frickhofen[13], Siegfried Haas[7], Martin-Leo Hansmann[14], Dennis Karsch[15], Michael Kneba[15], Jasmin Lisfeld[9], Luisa Mantovani-Löffler[16], Marius Rohde[9], German Ott[17], Christina Stadler[11], Peter Staib[18], Stephan Stilgenbauer[19], Lorenz Trümper[11], Thorsten Zenz[20]; **Normal Cells (WPN)** Martin-Leo Hansmann[14], Dieter Kube[11], Ralf Küppers[21], Marc Weniger[21]; **Pathology & Analyte Preparation (WP2-3)** Michael Hummel[22], Wolfram Klapper[3], Ulrike Kostezka[23], Dido Lenze[22], Peter Möller[24], Andreas Rosenwald[25], German Ott[17], Monika Szczepanowski[3]; **Sequencing & Genomics (WP4-7)** Ole Ammerpohl[1,2], Sietse M. Aukema[2,3], Vera Binder[26], Arndt Borkhardt[26], Andrea Haake[2], Jessica I. Hoell[26], Ellen Leich[25], Peter Lichter[27], Cristina López[1,2], Inga Nagel[2], Jordan Pischimariov[25], Bernhard Radlwimmer[27], Julia Richter[2,3], Philip Rosenstiel[28], Andreas Rosenwald[25], Markus Schilhabel[28], Stefan Schreiber[29], Inga Vater[2], Rabea Wagener[1,2], Reiner Siebert[1,2]; **Bioinformatics (WP8-9)** Stephan H. Bernhart[30,31,32], Hans Binder[30,31], Benedikt Brors[33], Gero Doose[30,31,32], Roland Eils[4,5], Steve Hoffmann[30,31,32], Lydia Hopp[30], Daniel Hübschmann[4,5,34], Kortine Kleinheinz[4,5], Helene Kretzmer[30,31,32], Markus Kreuz[35], Jan Korbel[36], David Langenberger[30,31,32], Markus Loeffler[35], Maciej Rosolowski[35], Matthias Schlesner[4,37], Peter F. Stadler[30,31,32,38,39,40] & Stephanie Sungalee[36]

[1]Institute of Human Genetics, University of Ulm and University Hospital of Ulm, Ulm, Germany. [2]Institute of Human Genetics, Christian-Albrechts-University, Kiel, Germany. [3]Hematopathology Section, Institute of Pathology, Christian-Albrechts-University, Kiel, Germany. [4]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [5]Department for Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology and Bioquant, University of Heidelberg, Heidelberg, Germany. [6]Department of Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany. [7]Department of Internal Medicine/Hematology, Friedrich-Ebert-Hospital, Neumünster, Germany. [8]University Hospital Muenster - Pediatric Hematology and Oncology, Münster, Germany. [9]University Hospital Giessen, Pediatric Hematology and Oncology, Giessen, Germany. [10]Department of Medicine III - Campus Grosshadern, University Hospital Munich, Munich, Germany. [11]Department of Hematology and Oncology, Georg-August-University of Göttingen, Göttingen, Germany. [12]University Hospital Würzburg, Department of Medicine and Poliklinik II, University of Würzburg, Würzburg. [13]Department of Medicine III, Hematology and Oncology, Dr Horst-Schmidt-Kliniken of Wiesbaden, Wiesbaden, Germany. [14]Senckenberg Institute of Pathology, University of Frankfurt Medical School, Frankfurt am Main, Germany. [15]Department of Internal Medicine II: Hematology and Oncology, University Medical Centre, Campus Kiel, Kiel, Germany. [16]Hospital of Internal Medicine II, Hematology and Oncology, St-Georg Hospital Leipzig, Leipzig, Germany. [17]Department of Pathology, Robert-Bosch-Hospital, Stuttgart, Germany. [18]Clinic for Hematology and Oncology, St-Antonius-Hospital, Eschweiler, Germany. [19]Department for Internal Medicine III, University of Ulm and University Hospital of Ulm, Ulm, Germany. [20]National Centre for Tumor Disease, Heidelberg, Germany. [21]Institute of Cell Biology (Cancer Research), University of Duisburg-Essen, Duisburg-Essen, Medical School, Essen, Germany. [22]Institute of Pathology, Charité – University Medicine Berlin, Berlin, Germany. [23]Comprehensive Cancer Center Ulm (CCCU), University Hospital Ulm, Ulm, Germany. [24]Institute of Pathology, University of Ulm and University Hospital of Ulm, Ulm, Germany. [25]Institute of Pathology, University of Würzburg, Würzburg, Germany. [26]Department of Pediatric Oncology, Hematology and Clinical Immunology, Heinrich-Heine-University, Düsseldorf, Germany. [27]German Cancer Research Center (DKFZ), Division of Molecular Genetics, Heidelberg, Germany. [28]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany. [29]Department of General Internal Medicine, University Kiel, Kiel, Germany. [30]Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig, Germany. [31]Bioinformatics Group, Department of Computer, University of Leipzig, Leipzig, Germany. [32]Transcriptome Bioinformatics, LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany. [33]Division of Applied Bioinformatics (G200), German Cancer Research Center (DKFZ), Heidelberg, Germany. [34]Department of Pediatric Immunology, Hematology and Oncology, University Hospital, Heidelberg, Germany. [35]Institute for Medical Informatics Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. [36]EMBL Heidelberg, Genome Biology, Heidelberg, Germany. [37]Bioinformatics and Omics Data Analytics (B240), German Cancer Research Center (DKFZ), Heidelberg, Germany. [38]RNomics Group, Fraunhofer Institute for Cell Therapy and Immunology IZI, Leipzig, Germany. [39]Santa Fe Institute, Santa Fe, New Mexico, USA. [40]Max-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.dkfz.de/en/hipo), the Individualized Therapy For Relapsed Malignancies in Childhood (INFORM) registry (www.dkfz.de/en/inform), 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.uk/ega/home) using the accession numbers EGAD00001000085, EGAD00001000135, EGAD00001000159, EGAD00001000160, EGAD00001000161, EGAD00001000162, EGAD00001000163, EGAD00001000164, EGAD00001000165, EGAD00001000259, EGAD00001000260, EGAD00001000261, EGAD00001000268, and EGAD00001000269[49–62]; internal datasets are related to previous PMIDs 27748748, 27479119, 26923874, 25670083, 25253770, 24972766, 24553142, 25135868, 26632267, 26179511, 24651015, 28726821, 23817572, 25962120, 26294725[17,19,44,63–74] (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 24). 'Subgroups' of cancer types were considered as separate entities if there was considerable 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 exome 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)[8]; 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 publications[75–79]. Survival data are based on information from the German Childhood Cancer Registry[80]. Incidence rates of adult cancers were taken from information in the German GEKID database (http://www.gekid.de/, 2003–2012).

Data preprocessing. All data were processed using a standardized alignment and variant calling pipeline, which was developed in the context of the ICGC Pan-Cancer project (https://dockstore.org/containers/quay.io/pancancer/pcawg-dkfz-workflow)[81].

Alignments. Datasets were available in either raw FASTQ or aligned BAM format. To allow standardized processing for all included samples, BAM files were sorted by read name using sambamba (v.0.4.6) and converted to a raw-like FASTQ format using SamToFastq (v.1.61). Reads were then aligned to the phase II reference human genome assembly of the 1000 Genomes Project including decoy sequences (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz) using BWA-MEM (v.0.7.8 using default settings except '-T 0'). Matching genotypes of tumour and control samples were confirmed by calculating pairwise DNA sequence similarities at 1,000 reference SNPs (dbSNP v.138)[82].

Mutation calling. SNVs were called with the previously described samtools-based DKFZ pipeline adjusted for ICGC Pan-Cancer settings, and short indels were called using Platypus (v.0.7.4)[74,83]. Variants were first identified in the tumour sample and germline or somatic origin was determined based on their presence or absence in the matched control tissue. Functional effects were annotated using ANNOVAR and GENCODE19 (http://www.gencodegenes.org/releases/19.html)[84].

Somatic structural variant discovery. Somatic structural variant discovery was pursued across all whole-genome sequenced samples (high-quality structural variants available for $n = 539$ primary tumours) using the DELLY ICGC Pan-Cancer analysis workflow (https://github.com/ICGC-TCGA-PanCancer/pcawg_delly_workflow)[85]. A high-stringency structural variant set was obtained by additionally filtering somatic structural variants detected in 1% or more of a set of 1,105 germline samples from healthy individuals belonging to phase I of the 1000 Genomes Project and by removing somatic structural variants present in any of the paediatric germline samples of this study[86]. High-stringency structural variants were further required to have at least four supporting read pairs with a minimum mapping quality of 20 and were restricted to somatic structural variant sizes from 300 bp to 500 Mb.

Copy-number calling. Copy numbers were estimated using ACEseq (allele-specific copy-number estimation from sequencing) (K. Kleinheinz *et al.*, unpublished data), using a binned tumour–control coverage ratio and B-allele frequency (BAF). Allele frequencies were obtained for all single nucleotide polymorphism (SNP) positions recorded in dbSNP version 135[82]. To improve sensitivity with regard to imbalanced and balanced regions, SNP positions in the control were phased with impute2[87]. Additionally, the coverage for 10-kb windows with sufficient mapping quality and read density was recorded and subsequently corrected for GC content and replication timing.

The genome was segmented using the PSCBS package incorporating structural variant breakpoints defined by DELLY[88,89]. Segments were clustered based on coverage ratio and BAF using *k*-means and neighbouring segments in the same cluster were joined; focal segments (<9 Mb) were stitched to the more similar neighbour. Tumour cell content and ploidy were estimated by testing how well different combinations of both explain the data. Segments with balanced BAF were assigned to even-numbered copy-number states, whereas unbalanced segments were allowed to match with uneven numbers as well. Finally, estimated tumour cell content and ploidy were used to compute the total and allele-specific copy-number for each segment. High-quality copy-number calls were available for $n = 516$ of the WGS samples.

Mutation statistics. The frequency of somatic mutations in coding regions was determined for each sample individually by normalizing the total number of coding mutations for the number of sufficiently covered (≥6×) coding bases to account (determined using MuSiC-bmr) for different data types (WGS/WES) and for different exome target enrichment kits[24]. Mutation spectra were obtained by categorizing observed SNVs into base substitution types in pyrimidine context. Spearman's rank correlation test was applied to infer correlations between different types of mutation counts or between mutation counts and age. Generalized linear models were used to fit regression lines. Clusters of localized hypermutation were identified using a previously presented approach adjusted for mutation rates in human paediatric cancers[90].

Deciphering mutation signatures. Exome-sequenced tumours, except for hypermutator cases, were excluded from signature analysis owing to their low numbers of mutations. In brief, signatures are represented as probability distributions of substitution types of SNVs in pyrimidine context. Considering the immediate sequence context of each SNV, this results in 96 possible mutation types with directly adjacent mutations (multiple nucleotide variants, MNVs) being excluded, which are counted per tumour to compile its mutational profile.

As proposed by Alexandrov *et al.*[91], 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)[91]: $M \approx P \times E$.

De novo deciphering of signatures was done as described[91] based on the mutational catalogues of all cancer types and 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 similarity[15]. 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 from the known set of signatures was confirmed. Second, for each potentially novel signature, we examined whether the modelling of mutation profiles improved when compared to having used the set of known signatures: for each sample, the observed mutational profile was compared to the theoretical profiles calculated using the set of known 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 <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 ≥0.2, and a phred-based quality score of ≥10. Variants with a population frequency ≥0.01 reported in additional common databases (esp6500siv2, X1000g2015, and exac03 included in ANNOVAR (http://annovar.openbioinformatics.org)) or with ClinVar (ftp://ftp.ncbi.nlm.nih.gov/pub/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.org/r3/) categories 'medium' and 'high', or no available annotation, were included. Variants with a dbSNP classification of 'precious' were not subject to these two filtering steps. As indel calling is more prone to alignment and calling errors, potentially deleterious indels were manually investigated for artefacts. For recessive tumour genes, variants were included only with an allele frequency of one or with two compound heterozygous mutations of the same gene in the same patient. In total, the filtering steps narrowed down the number of potentially pathogenic mutations to $n = 433$. Every variant was then manually checked and scored by the use of varied, mainly gene-specific online databases (http://p53.iarc.fr/, http://www.lovd.nl/3.0/home, https://www.ncbi.nlm.nih.gov/clinvar/, and others). Only likely pathogenic and pathogenic mutations were considered as cancer-relevant and used for representation in Fig. 3. Additionally, whole-genome sequenced samples were manually screened for copy-number losses in 13 tumour suppressor genes of the candidate list, which are known to occasionally harbour germline focal deletions (*MLH1*, *MSH2*, *MSH6*, *NF1*, *PMS2*, *PRKAR1A*, *PTCH1*, *PTEN*, *RB1*, *SMARCA4*, *SMARCB1*, *SUFU*, *TP53*).

Detecting genome-wide mutation clusters. To identify genomic regions with single or clusters of recurrent mutations, the human genome was binned into non-overlapping windows of various sizes (50–500 bp) and compared the observed mutations to a background model (V. A. Rudneva *et al.*, unpublished data) which was estimated using the 'global' model: the genome was stratified into 25 evenly sized groups of genomic windows based on the combined vector of five genetic and epigenetic features (replication timing, gene expression level, GC content, H3K9me3, and open versus closed chromatin conformation). For each region an enrichment score, binomial P value, and negative binomial test P value were computed.

Cross-validations were used to determine the significance cut-off that would provide reproducible results (with samples segregated by subgroup). A combination of the window size (500 bp), test statistics (enrichment score, mutational recurrence, binomial test P value, and gamma Poisson test P value), and a cut-off value that ensured high precision and recall values based on the precision-recall analysis ($P = 10^{-20}$) were chosen (Extended Data Fig. 4a). Recall was calculated as the number of regions that satisfied the cut-off in results obtained on both halves of the dataset; precision was calculated as a fraction of the recalled regions to the total number of regions that satisfied the cut-off in each of the datasets. The chosen parameters were then used to run the pipeline on the complete dataset and then the mutations in the resulting regions were further examined manually for potential false positives in order to identify high-confidence candidate regions (Extended Data Fig. 4b).

Significantly mutated genes. Significantly mutated genes based on somatic SNVs and indels were identified with the SMG module of the MuSiC tools suite[24] separately from all cancer types and from the pan-cancer cohort, and then merged.

This kind of significance analysis often produces false positive hits (for example, very large genes), despite normalization procedures, and thus several filters were applied to the raw output[30]. First, all genes of >30,000 bp exonic length or >10,000 bp with additional replication timing >800 were excluded (Cancer Cell Line Encyclopedia; CCLE)[92]. Genes that scored significant in three or more cancer types, or that were recurrently mutated at the same position, were manually inspected for artefacts from ambiguous alignments (for example, repetitive sequence regions). Also, genes that are probably not associated with tumour development but rather represent non-neoplastic somatic hypermutation processes in the context of immune function were removed. Furthermore, genes mutated in <2% of the cohort were included only if they had a secondary signal from either functional impact or from localized clustering bias (Intogen modules OncodriveFM and OncodriveClust v. 3.0 beta) or from being among known cancer genes[29,93]. Mutation needle plots were generated using MutationMapper[94]. Biological processes were assigned to the significantly mutated genes mostly exclusively, except for a few genes with high relevance for multiple processes, as specified in Supplementary Table 9.

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.

Generation of copy-number profiles. Copy-number calls reported by ACEseq were converted to the 'SEG' segmentation format, similar to the output of the circular binary segmentation algorithm based on chromosomal segment borders as pseudo marker positions[96]. All possible marker positions were determined from the whole cohort before assessing sample-wise copy-number profiles per marker in order to achieve identical resolution for all samples. Owing to sparse and highly oscillating sequencing coverage at centromeres, centromeric coordinates (±3 Mb around the centre of annotated centromeres) were excluded from whole-genome segmentation, as were two likely artefact regions on chromosomes 7 and 14 with nonspecific occurrences of relative copy-number gains and losses in 28% and 30% of all analysed samples in 17 of 19 entities (14q11.2, 7p14.1), which were identified using GISTIC2.0 (as described below) with ±1 Mb.

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[36]. The resulting peaks were filtered for significance ($q \leq 0.1$) 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 here sequencing coverage tends to vary frequently; and last, peaks overlapping with copy-number variable regions[98] (regions ranked 1–100) were excluded. Additionally, some of the resulting peaks were classified as 'passengers' of variable regions that were called as separated peaks from most likely one event, for example, a peak with *MYCNOS* as passenger peak of *MYCN* amplification. For overlapping peaks called in multiple entities and/or pan-cancer, the final region was determined based on the analysis with highest significance for each peak, respectively.

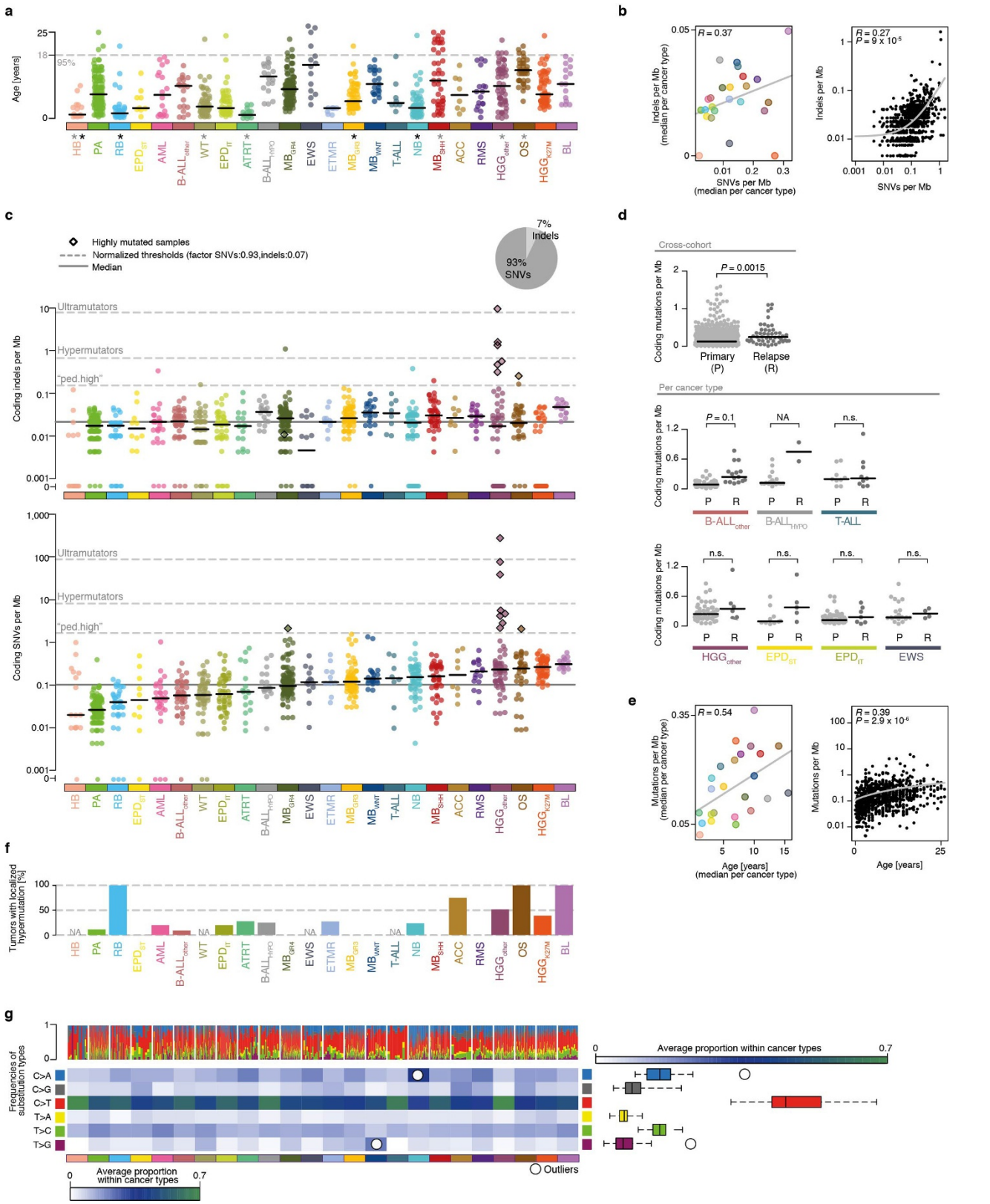
Genes with a breakpoint inside the gene borders were assumed to be altered by structural variation and considered as recurrently altered if they had breakpoints in ≥5 samples in total or in ≥2 samples of one cancer type (for samples without chromothripsis). For other samples, genes with breakpoints in ≥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 present cancer type (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.

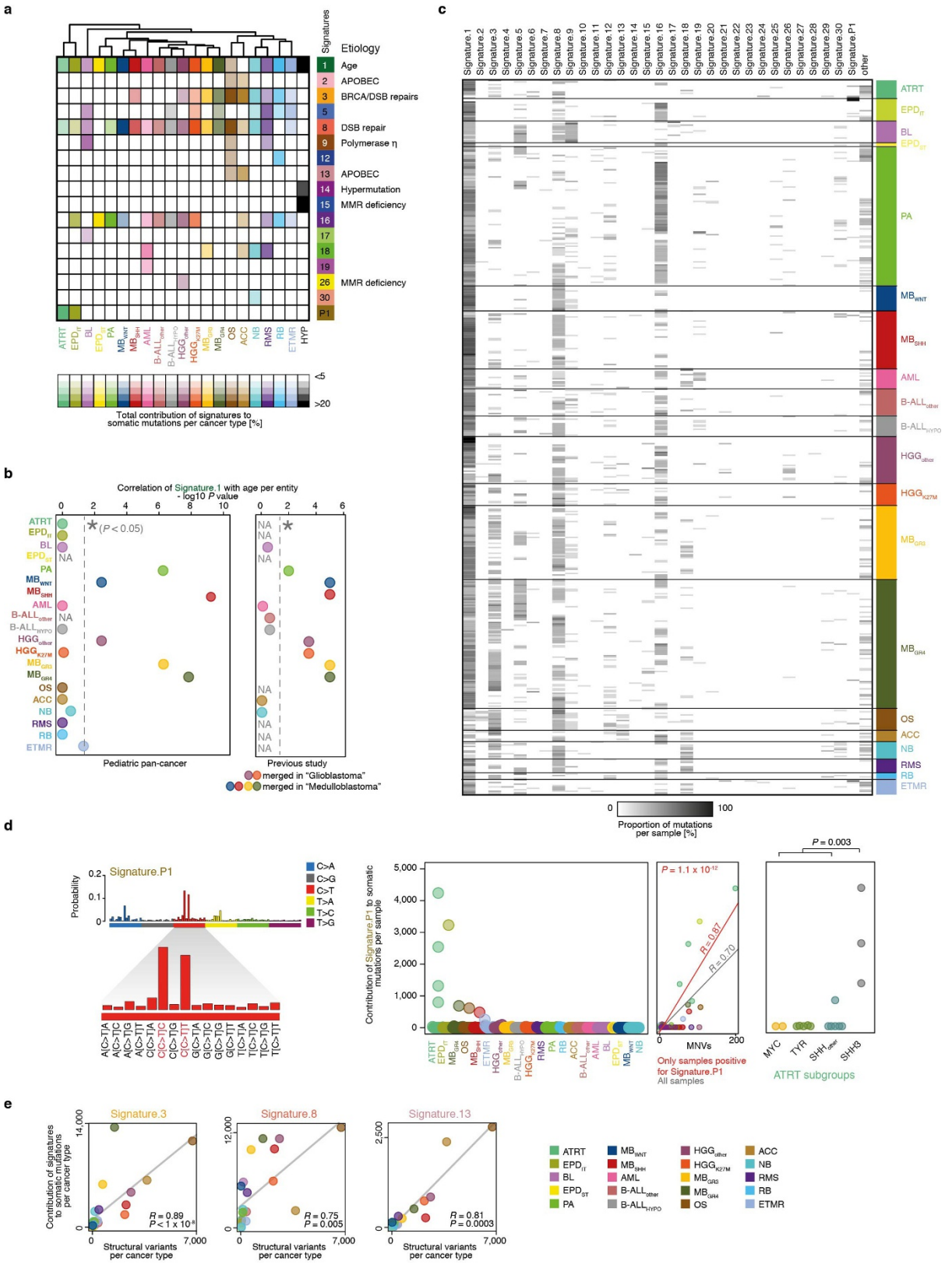
Additionally, copy-number plots of whole-genome-sequenced data (including low-coverage WGS) were used to manually screen 52 druggable genes for amplifications or deletions (Supplementary Table 22a). Only focal CNVs (<10 Mb) with at least 5 copies ($\log_2 \geq 1.3$) in the case of amplifications or the loss of ≥ 1 copy ($\log_2 \leq -1$) for deletions were included and subsequently prioritized as described for the SNVs/indels. The data representation includes all tumours with full genomic information (WES + lcWGS 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 PedcBio Portal 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: RP012816, PRJEB11430 (European Nucleotide Archive); EGAS00001001139, EGAS00001001953, EGAS00001000607, EGAS00001000381, EGAS00001000906, EGAS00001001297, EGAS00001000443, EGAS00001000213, EGAS00001000263, EGAS00001000192, EGAS00001000255, EGAS00001000254, EGAS00001000253, EGAS00001000256, EGAS00001000246, EGAS00001000379, EGAS00001000380, EGAS00001000346, EGAS00001000349, EGAS00001000347, EGAS00001000192 (European Genome-Phenome Archive).

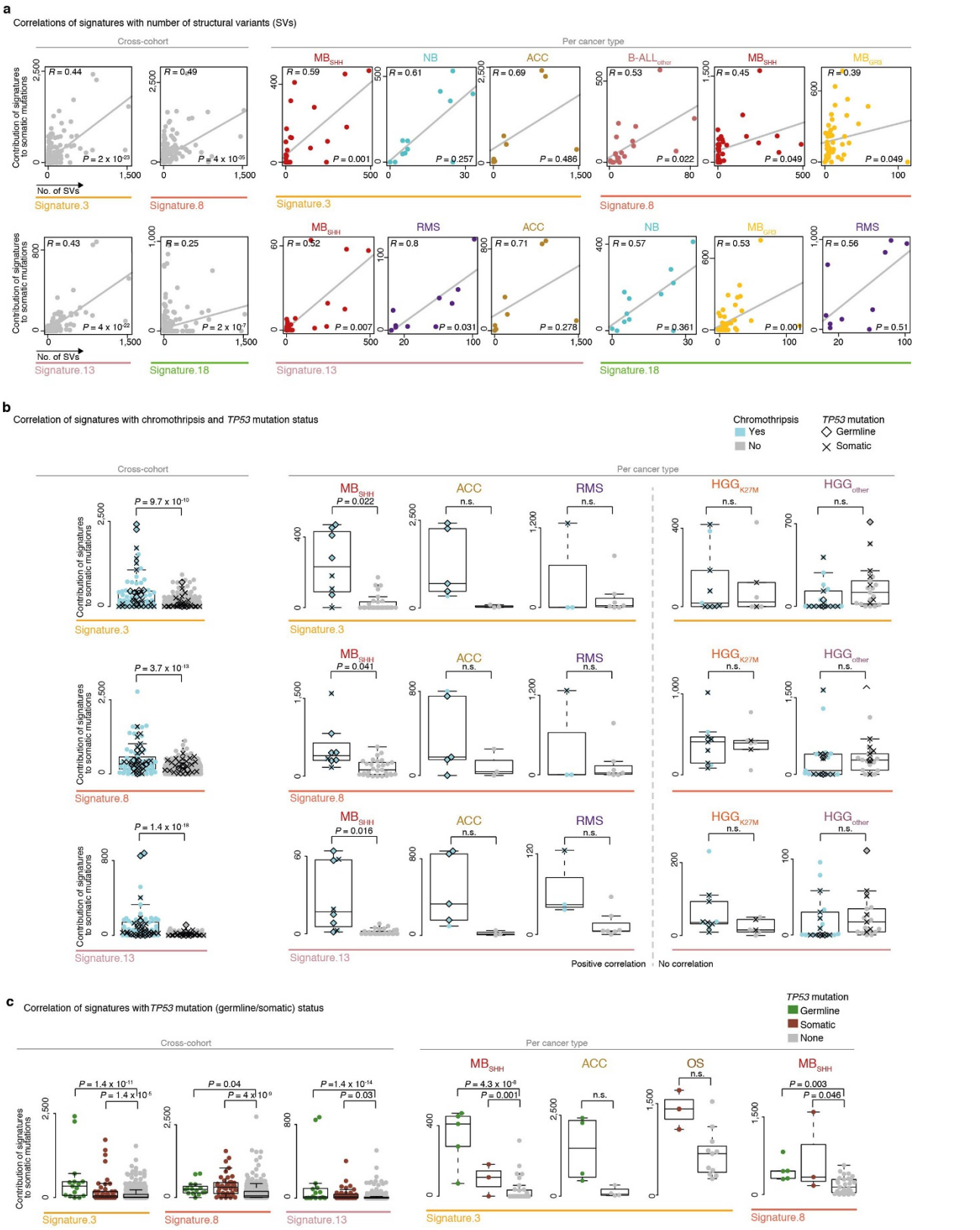
49. Wu, G. *et al.* Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat. Genet.* **44,** 251–253 (2012).
50. Wu, G. *et al.* The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nat. Genet.* **46,** 444–450 (2014).
51. 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).
52. Chen, X. *et al.* Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep.* **7,** 104–112 (2014).
53. Pinto, E. M. *et al.* Genomic landscape of paediatric adrenocortical tumours. *Nat. Commun.* **6,** 6302 (2015).
54. Zhang, J. *et al.* Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nat. Genet.* **45,** 602–612 (2013).
55. Parker, M. *et al.* C11orf95-RELA fusions drive oncogenic NF-κB signalling in ependymoma. *Nature* **506,** 451–455 (2014).
56. Chen, X. *et al.* Targeting oxidative stress in embryonal rhabdomyosarcoma. *Cancer Cell* **24,** 710–724 (2013).
57. Andersson, A. K. *et al.* The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat. Genet.* **47,** 330–337 (2015).
58. Gruber, T. A. *et al.* An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. *Cancer Cell* **22,** 683–697 (2012).
59. Holmfeldt, L. *et al.* The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat. Genet.* **45,** 242–252 (2013).
60. Zhang, J. *et al.* A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* **481,** 329–334 (2012).
61. Faber, Z. J. *et al.* The genomic landscape of core-binding factor acute myeloid leukemias. *Nat. Genet.* **48,** 1551–1556 (2016).
62. Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488,** 43–48 (2012).
63. International Cancer Genome Consortium PedBrain Tumor Project. Recurrent MET fusion genes represent a drug target in pediatric glioblastoma. *Nat. Med.* **22,** 1314–1320 (2016).
64. Wegert, J. *et al.* Mutations in the SIX1/2 pathway and the DROSHA/DGCR8 miRNA microprocessor complex underlie high-risk blastemal type Wilms tumors. *Cancer Cell* **27,** 298–311 (2015).
65. Irving, J. *et al.* Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. *Blood* **124,** 3420–3430 (2014).
66. Bandapalli, O. R. *et al.* The activating STAT5B N642H mutation is a common abnormality in pediatric T-cell acute lymphoblastic leukemia and confers a higher risk of relapse. *Haematologica* **99,** e188–e192 (2014).
67. Mack, S. C. *et al.* Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506,** 445–450 (2014).
68. Eichenmüller, M. *et al.* The genomic landscape of hepatoblastoma and their progenies with HCC-like features. *J. Hepatol.* **61,** 1312–1320 (2014).
69. Kovac, M. *et al.* Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency. *Nat. Commun.* **6,** 8940 (2015).
70. Agelopoulos, K. *et al.* Deep sequencing in conjunction with expression and functional analyses reveals activation of FGFR1 in Ewing sarcoma. *Clin. Cancer Res.* **21,** 4935–4946 (2015).
71. Kool, M. *et al.* Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer Cell* **25,** 393–405 (2014).
72. Li, B. *et al.* Negative feedback-defective PRPS1 mutants drive thiopurine resistance in relapsed childhood ALL. *Nat. Med.* **21,** 563–571 (2015).
73. Kunz, J. B. *et al.* Pediatric T-cell lymphoblastic leukemia evolves into relapse by clonal selection, acquisition of mutations and promoter hypomethylation. *Haematologica* **100,** 1442–1450 (2015).
74. Jones, D. T. *et al.* Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nat. Genet.* **45,** 927–932 (2013).
75. 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).
76. Pajtler, K. W. *et al.* Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups. *Cancer Cell* **27,** 728–743 (2015).
77. Northcott, P. A. *et al.* Medulloblastomics: the end of the beginning. *Nat. Rev. Cancer* **12,** 818–834 (2012).
78. Harrison, C. J. *et al.* Three distinct subgroups of hypodiploidy in acute lymphoblastic leukaemia. *Br. J. Haematol.* **125,** 552–559 (2004).
79. Pui, C. H., Relling, M. V. & Downing, J. R. Acute lymphoblastic leukemia. *N. Engl. J. Med.* **350,** 1535–1548 (2004).
80. Kaatsch, P. S. C. *German Childhood CancerRegistry - Report 2013/14 (1980-2013)* (Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), Univ. Medical Center of Johannes Gutenberg Univ., 2014).
81. Stein, L. D., Knoppers, B. M., Campbell, P., Getz, G. & Korbel, J. O. Data analysis: Create a cloud commons. *Nature* **523,** 149–151 (2015).
82. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–311 (2001).
83. Jones, D. T. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488,** 100–105 (2012).
84. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).
85. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28,** i333–i339 (2012).
86. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
87. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** e1000529 (2009).
88. Olshen, A. B. *et al.* Parent-specific copy number in paired tumor–normal studies using circular binary segmentation. *Bioinformatics* **27,** 2038–2046 (2011).
89. Van den Meersche, K., Soetaert, K. & Van Oevelen, D. xsample(): An R function for sampling linear inverse problems. *J. Stat. Softw.* **30,** 1–15 (2009).
90. 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).
91. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3,** 246–259 (2013).
92. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–607 (2012).
93. Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4,** 177–183 (2004).
94. Vohra, S. & Biggin, P. C. Mutationmapper: a tool to aid the mapping of protein mutation data. *PLoS ONE* **8,** e71711 (2013).
95. Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152,** 1226–1236 (2013).
96. Olshen, 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).
97. Le Tallec, B. *et al.* Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep.* **4,** 420–428 (2013).
98. Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526,** 75–81 (2015).



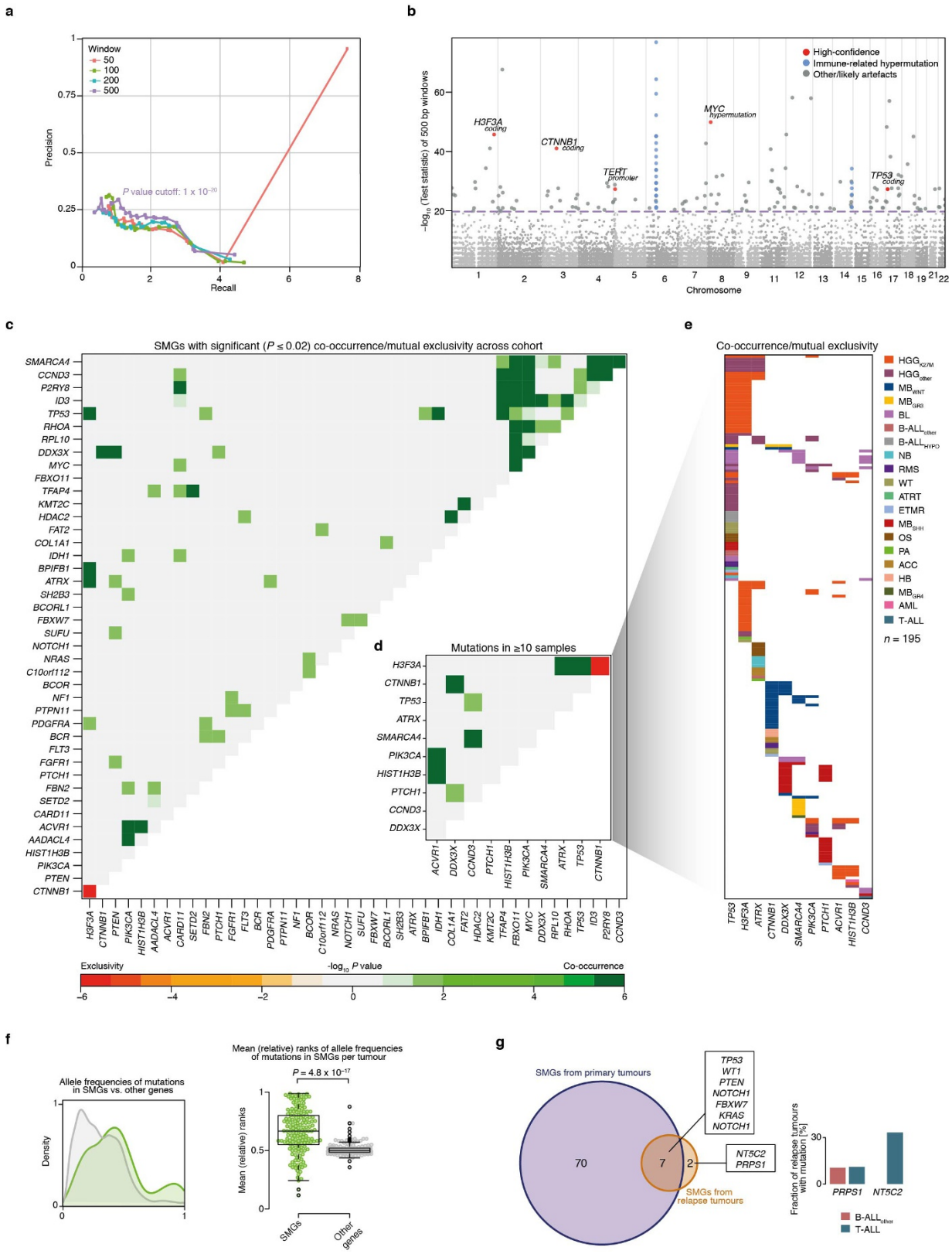
Extended Data Figure 1 | Somatic mutation frequencies and spectra. **a**, Patient age at diagnosis (black lines, median); asterisks highlight cancer types with significant correlation of mutation load with age within cancer types (grey, SNVs; black, indels). **b**, Correlation of SNV and indel loads (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**.



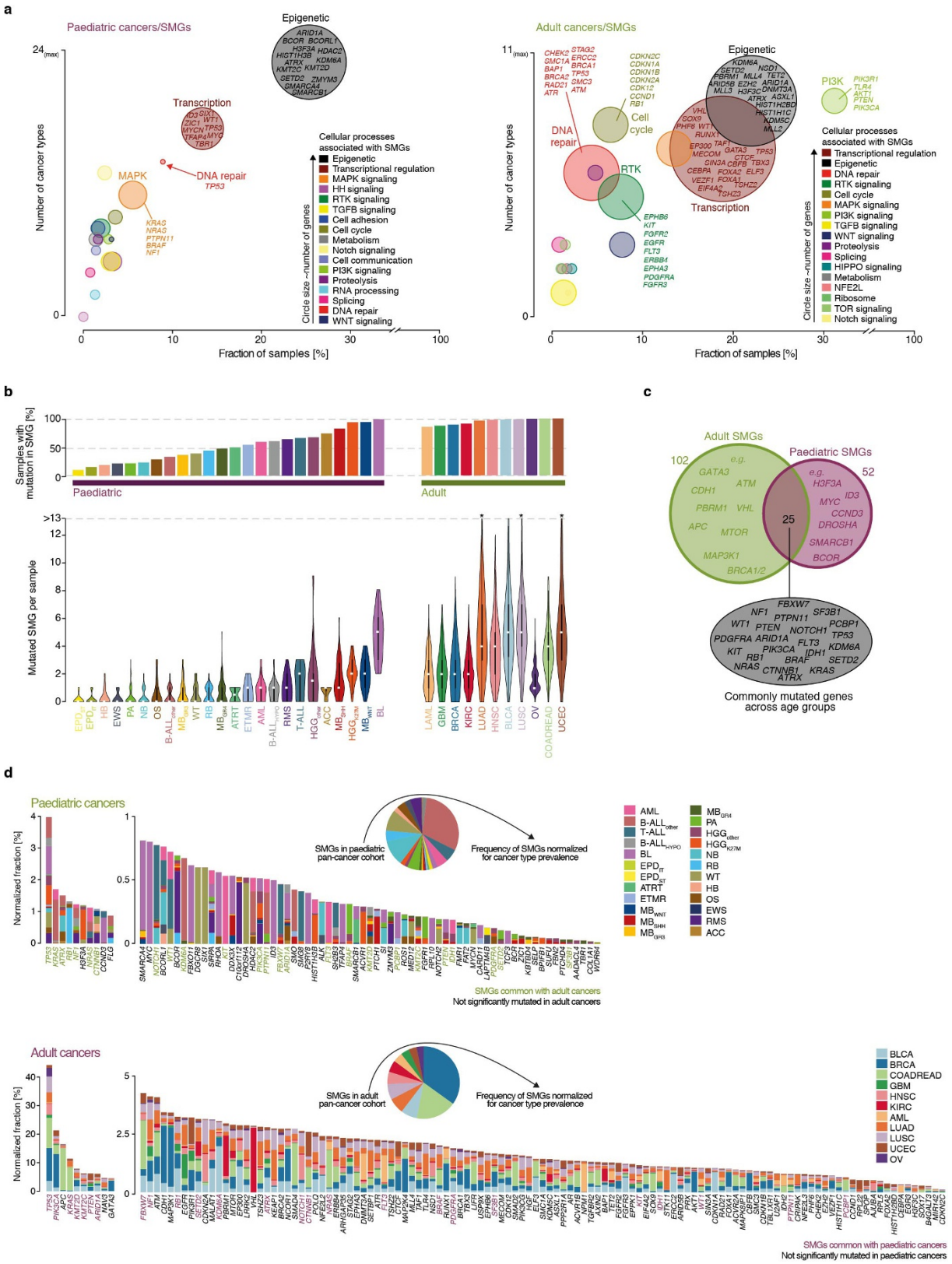
Extended Data Figure 2 | Mutational signatures in paediatric cancer types. a, Summarized contribution of signatures to mutational profiles per cancer type (proportion of mutations per signature and cancer type). Signatures with contributions of ≥5% in at least one cancer type are shown. The colour intensity 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.



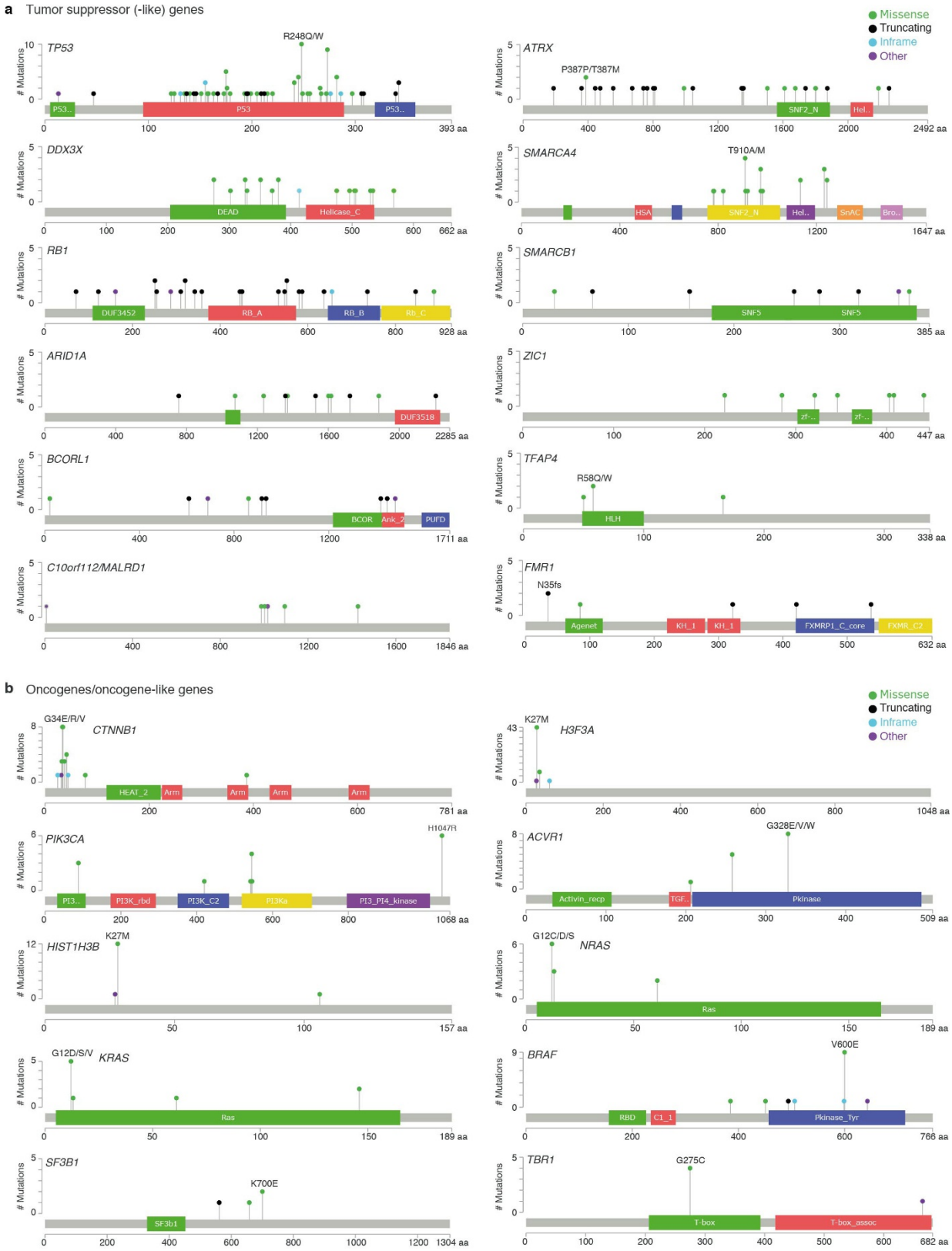
Extended Data Figure 3 | Association of mutational signatures with genomic instability. a, Correlation of signatures with the number of structural variants across all tumours and selected cancer types (Spearman's correlation, confidence interval 0.95). **b**, Association of signatures with chromothripsis 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). **a**–**c**, Cross-cohort $n = 503$, cancer types see Supplementary Tables 1, 4.



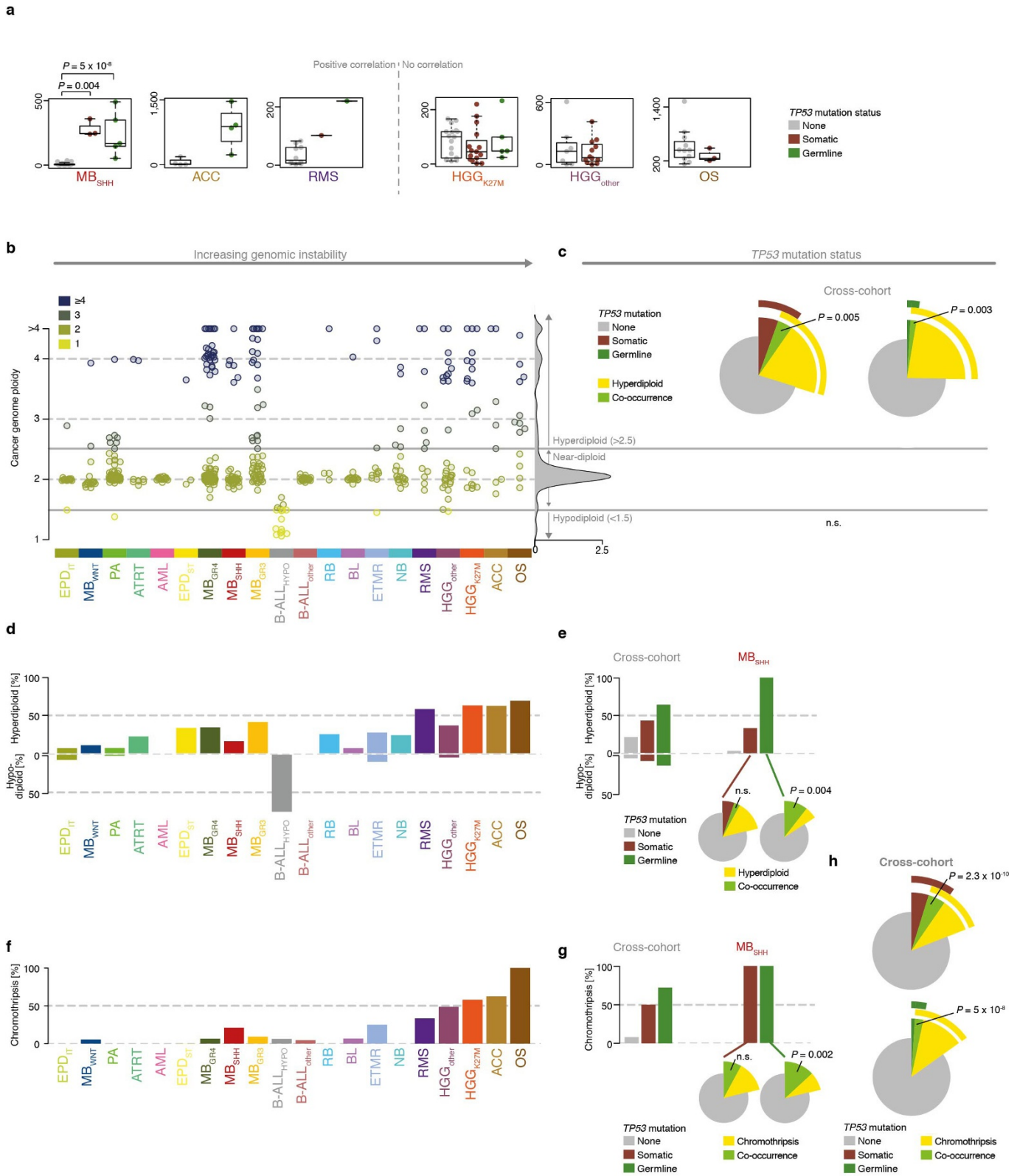
Extended Data Figure 4 | Characteristics of significantly mutated genomic regions and genes. a, Precision-recall curves (mean precision) for various binomial *P* value cut-offs for the identification of genome-wide mutation clusters. **b**, Manhattan plot for the test statistic of genomic windows. Dashed 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.



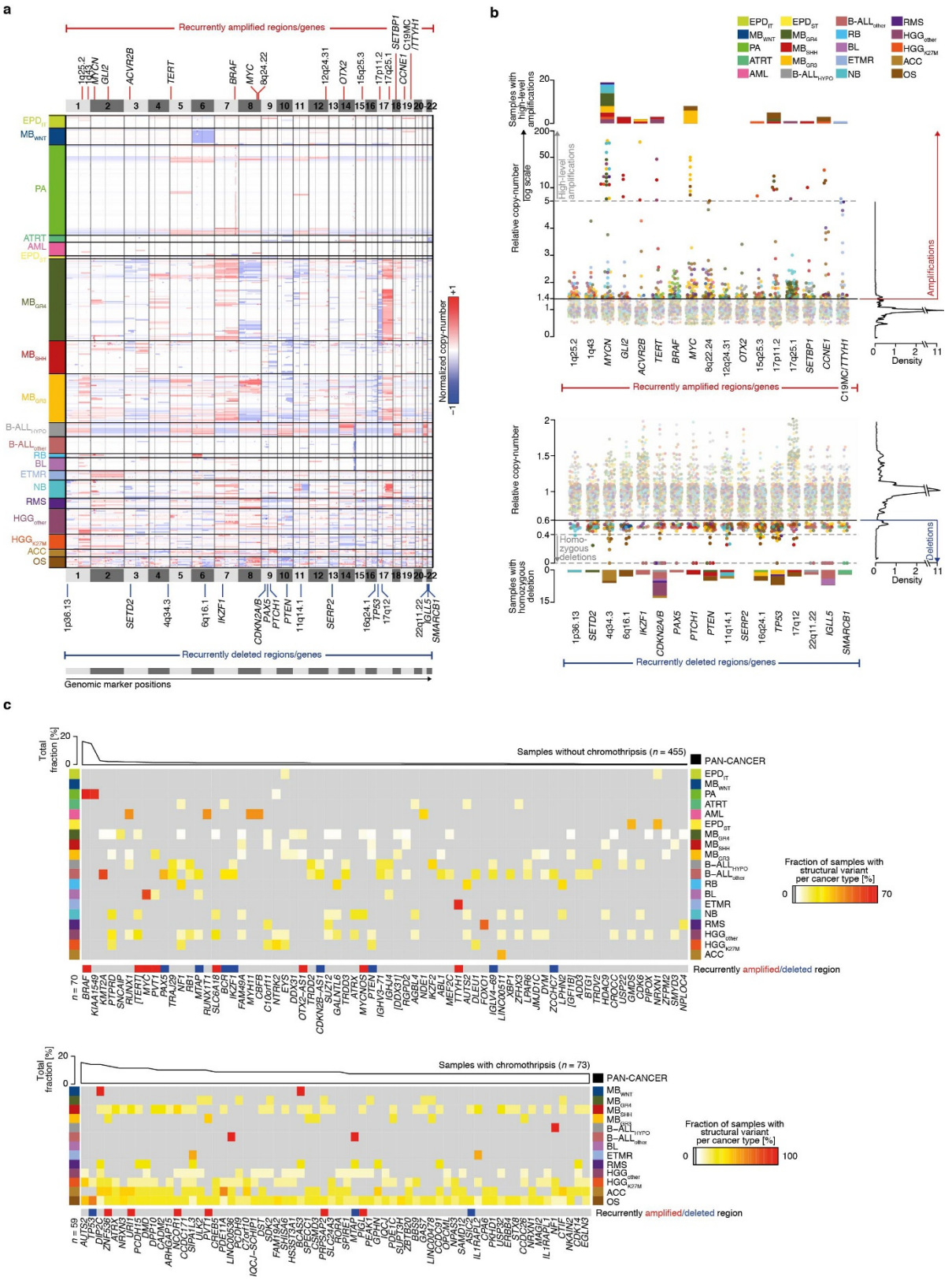
Extended Data Figure 5 | Significantly mutated genes across age groups. a, Cellular processes associated with paediatric (left) and adult (right) SMGs. **b**, Frequency of mutations in SMGs in paediatric ($n = 879$) compared to adult ($n = 3{,}281$) cancers. Top, percentage of SMG-mutated samples. Bottom, 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].



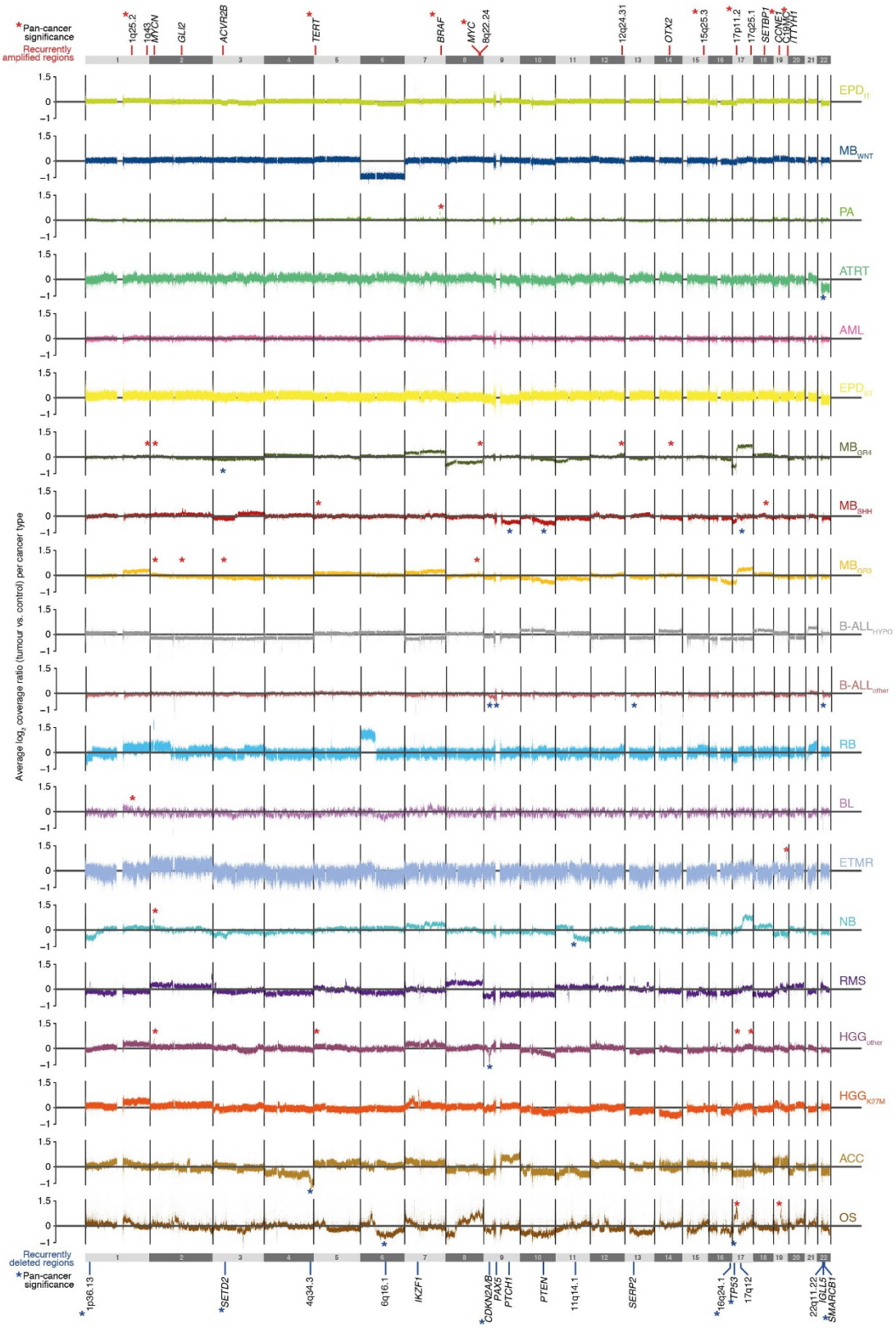
Extended Data Figure 6 | Mutation needle plots for significantly mutated genes. Mutations in selected significantly mutated genes across pan-cancer cohort: missense (green), truncating (black), in-frame (blue), and other (purple). Hotspot amino acid changes are highlighted. **a**, Genes with tumour suppressor-like mutation patterns. **b**, Genes with oncogenic or oncogene-like mutation patterns.



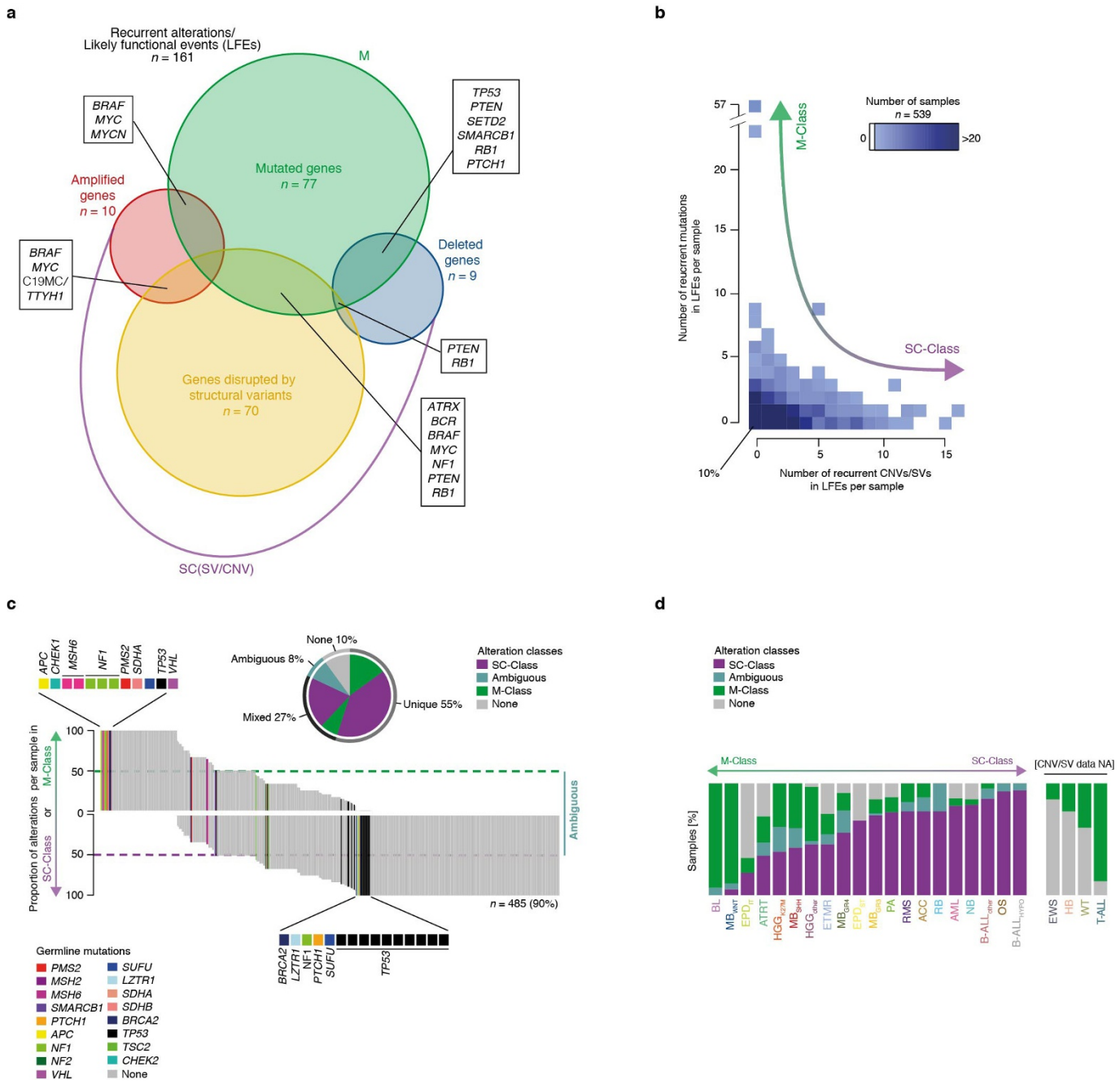
Extended Data Figure 7 | Genomic instability across paediatric cancer types. a, Structural variant load in relation to *TP53* mutation status for individual cancer types (generalized linear model, confidence interval 0.95). **b**–**h**, Characteristics of genomic instability (left) and their associations 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- (≥1.5) and hypodiploid (≤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 $\log_2$ 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.

nature research

Corresponding author(s): Stefan M. Pfister

☐ Initial submission ☐ Revised version ☒ Final submission

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

   Describe any 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.

Software

Policy information about availability of computer code

7. Software

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.

Materials and reagents

Policy information about availability of materials

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

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used.

Policy information about studies involving human research participants

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.

CORRECTION

https://doi.org/10.1038/s41586-018-0167-2

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. Kulozik, Dietmar Lohmann, Simone Hettmer, Cornelia Eckert, Stefan Bielack, Michaela Nathrath, Charlotte Niemeyer, Günther H. Richter, Johannes Schulte, Reiner Siebert, Frank Westermann, Jan J. 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.