Genomes for Kids: The Scope of Pathogenic Mutations in Pediatric Cancer Revealed by Comprehensive DNA and RNA Sequencing

Scott Newman[1], Joy Nakitandwe[2], Chimene A. Kesserwan[3], Elizabeth M. Azzato[2], David A. Wheeler[1], Michael Rusch[1], Sheila Shurtleff[2], Dale J. Hedges[1], Kayla V. Hamilton[3], Scott G. Foy[1], Michael N. Edmonson[1], Andrew Thrasher[1], Armita Bahrami[2], Brent A. Orr[2], Jeffery M. Klco[2], Jiali Gu[2], Lynn W. Harrison[3], Lu Wang[2], Michael R. Clay[2], Annastasia Ouma[3], Antonina Silkov[1], Yanling Liu[1], Zhaojie Zhang[1], Yu Liu[1], Samuel W. Brady[1], Xin Zhou[1], Ti-Cheng Chang[1], Manjusha Pande[1], Eric Davis[1], Jared Becksfort[1], Aman Patel[1], Mark R. Wilkinson[1], Delaram Rahbarinia[1], Manish Kubal[3], Jamie L. Maciaszek[2], Victor Pastor[1], Jay Knight[1], Alexander M. Gout[1], Jian Wang[1], Zhaohui Gu[2], Charles G. Mullighan[2], Rose B. McGee[3], Emily A. Quinn[3], Regina Nuccio[3], Roya Mostafavi[3], Elsie L. Gerhardt[3], Leslie M. Taylor[3], Jessica M. Valdez[3], Stacy J. Hines-Dowell[3], Alberto S. Pappo[3], Giles Robinson[3], Liza-Marie Johnson[3], Ching-Hon Pui[3], David W. Ellison[2], James R. Downing[2], Jinghui Zhang[1], and Kim E. Nichols[3]

ABSTRACT

Genomic studies of pediatric cancer have primarily focused on specific tumor types or high-risk disease. Here, we used a three-platform sequencing approach, including whole-genome sequencing (WGS), whole-exome sequencing (WES), and RNA sequencing (RNA-seq), to examine tumor and germline genomes from 309 prospectively identified children with newly diagnosed (85%) or relapsed/refractory (15%) cancers, unselected for tumor type. Eighty-six percent of patients harbored diagnostic (53%), prognostic (57%), therapeutically relevant (25%), and/or cancer-predisposing (18%) variants. Inclusion of WGS enabled detection of activating gene fusions and enhancer hijacks (36% and 8% of tumors, respectively), small intragenic deletions (15% of tumors), and mutational signatures revealing of pathogenic variant effects. Evaluation of paired tumor–normal data revealed relevance to tumor development for 55% of pathogenic germline variants. This study demonstrates the power of a three-platform approach that incorporates WGS to interrogate and interpret the full range of genomic variants across newly diagnosed as well as relapsed/refractory pediatric cancers.

SIGNIFICANCE: Pediatric cancers are driven by diverse genomic lesions, and sequencing has proven useful in evaluating high-risk and relapsed/refractory cases. We show that combined WGS, WES, and RNA-seq of tumor and paired normal tissues enables identification and characterization of genetic drivers across the full spectrum of pediatric cancers.

INTRODUCTION

High-throughput next-generation sequencing (NGS) of pediatric cohorts has provided seminal insights into the genomic landscapes of the major subtypes of childhood cancer. It is now well established that these NGS approaches add significant value by refining or changing cancer diagnoses (1–5), providing prognostic information (3, 6), identifying therapeutic targets or markers of therapy resistance (2, 3, 5, 7, 8), detecting variants of pharmacogenetic significance (3), and uncovering underlying genetic predisposition (2–5, 8, 9).

To date, most pediatric NGS studies have focused on patients with high-risk disease, including difficult-to-treat or relapsed/refractory cancers. In many of these studies, patients with newly diagnosed or standard-risk cancers are absent or underrepresented. However, many standard-risk cancers do not respond or recur following treatment with current best available therapies. Indeed, such is the case for 15% to 20% of acute lymphoblastic leukemia (ALL; ref. 10) or Wilms tumor (11) cases and more than 30% of rhabdomyosarcomas (12) or other non–central nervous system (CNS) solid tumors. To improve upon these outcomes, it is essential that we comprehensively examine and elucidate the molecular underpinnings of childhood cancer across the full spectrum of presentations.

Each childhood cancer harbors a unique combination of somatic alterations on a background of inherited and *de novo* germline variants. Moreover, novel diagnostic and prognostic subgroups and the full constellation of genetic drivers have yet to be defined for many rare pediatric cancers. As such, an individual pediatric cancer genome could be described as an "*N* of 1" case study for which genome-wide analysis can uncover unique molecular drivers and elucidate how individual combinations of somatic and germline variants influence clinical presentation and response to cancer therapy. Therefore, it is essential that comprehensive genomic data are generated and systematically investigated if we are to fully capitalize on the potential of precision medicine across pediatric cancers, including those that are common as well as those that are rare.

Prior studies using distinct sequencing platforms have demonstrated that a genome-wide approach is necessary to enable full discovery of novel driver variants in pediatric cancers (1, 5, 8, 13–15). Toward this end, we developed a three-platform sequencing pipeline that includes whole-genome sequencing (WGS), whole-exome sequencing (WES), and RNA sequencing (RNA-seq) of paired tumor and normal samples. This three-platform pipeline has been validated on a retrospective cohort of highly selected high-risk pediatric cancer cases harboring known genomic alterations as shown by classical molecular assays. Furthermore, it greatly improved the accuracy of detection of genomic alterations, obviated the need for validation testing of somatic and germline variants through orthogonal approaches, and facilitated discovery of novel oncogenic processes (14).

[1]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, Tennessee. [2]Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee. [3]Department of Oncology, St. Jude Children's Research Hospital, Memphis, Tennessee.

S. Newman, J. Nakitandwe, C.A. Kesserwan, E.M. Azzato, D.A. Wheeler, D.W. Ellison, J.R. Downing, J. Zhang, and K.E. Nichols contributed equally to this article.

Current address for C.A. Kesserwan: Center for Cancer Research, Cancer Genetics Branch, NIH/NCI, Bethesda, Maryland.

Corresponding Authors: Kim E. Nichols, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105. Phone: 901-595-8385; Fax: 901-595-6086; E-mail: kim.nichols@stjude.org; and Jinghui Zhang, jinghui.zhang@stjude.org

Here we present data from Genomes for Kids (G4K), a prospective nontherapeutic three-platform sequencing study of 309 patients with pediatric cancer, unselected for tumor type or stage, treated at St. Jude Children's Research Hospital (Memphis, TN). The aims of this study were to (i) demonstrate the utility of comprehensive whole genome, exome, and transcriptome sequencing of paired tumor and normal samples for patients across the spectrum of standard-risk to high-risk cancers; (ii) show how integrating data from multiple sequencing platforms can elucidate the functional impacts of difficult-to-interpret variants; (iii) analyze rare genomic findings in *N* of 1 cases and show how these findings inform understanding of tumor biology and, when possible, patient care; and (iv) discover novel mechanisms driving a diverse array of childhood cancers.

RESULTS

Patient Enrollment

From August 2015 to March 2017, 365 patients with pediatric cancer were approached for enrollment in G4K (NCT02530658), a nontherapeutic study of three-platform sequencing of paired tumor and normal samples. Three hundred nine (85%) agreed to participate, 53 (15%) declined, and 3 were later removed from the study (Fig. 1A). Race/ethnicity was the only variable that significantly correlated with declining enrollment, with families of black children more likely to decline compared with families of non-Hispanic or Hispanic white children (*P* < 0.001) as we have reported elsewhere (16).

Forty-seven patients did not undergo tumor biopsy for safety reasons or because biopsy was not considered clinically indicated. Nine patients did not have sufficient tumor DNA or RNA to complete three-platform sequencing. Thus, 253 of 309 (82%) enrolled patients had their tumors examined using WGS, WES, and RNA-seq. Nine of 309 patients' families (3%) declined the return of germline results, leaving 300 who underwent analysis of germline samples using WES and WGS, followed by in-depth evaluation and reporting of 156 cancer predisposition genes. Study participants included 166 males and 143 females, with an average age at cancer diagnosis of 7.4 years (range: 4 days to 25.7 years; Fig. 1B; Supplementary Table S1).

At the time of G4K enrollment, 262 patients (85%) had newly diagnosed cancers and 47 (15%) had relapsed or refractory disease (Fig. 1C). The spectrum of cancers in the G4K cohort was similar to that observed in the NCI Surveillance, Epidemiology, and End Results (SEER) registry (Fig. 1D) except for Hodgkin and non-Hodgkin lymphoma, which were underrepresented, and leukemia and retinoblastoma, which were overrepresented. A total of 128 patients (41%) had hematologic malignancies of 28 subtypes, 97 (31%) had brain tumors of 27 subtypes, and 84 (27%) had non-CNS solid tumors of 26 subtypes. Forty-five patients (15%) had 18 very rare tumor types, defined here as tumors present in fewer than 2 cases per million annually in the United States (Fig. 1E). Among these 18 tumor types, only craniopharyngioma (17) and mixed phenotype acute leukemia (18) have been studied in enough detail to provide an initial understanding of associated somatic alterations. Thus, the ability to examine the impact of genomic lesions

in an *N* of 1 context is of paramount importance to understanding the biologic basis and therapeutic targets in these very rare tumors.

Overview of Somatic Alterations

Tumor samples were evaluated using 45× PCR-free WGS and 150× WES from DNA and 100 million RNA-seq reads from total RNA. Paired germline samples were evaluated using both WGS and WES; 1,200 genes known to play a role in the pathogenesis of cancer were interrogated using tumor genomic data. Genes were considered in the context of each patient's tumor type and prioritized for review as described previously (19). Given the potential effects of structural events and gene fusions on protein expression and function, we reported novel events if there was strong sequence support and the structural event or gene fusion involved genes of potential relevance to cancer. We detected somatic single-nucleotide variants (SNV), small insertions/deletions (indels), loss of heterozygosity (LOH), structural variants (SV) including fusions, enhancer hijacks, internal tandem duplications (ITD), and copy-number alterations (CNA) using automated computational pipelines followed by analyst curation (14). Among the SNVs and indels observed on both WGS and WES platforms, 41% had a variant allele fraction (VAF) below 0.2 and 15% had a VAF below 0.1, revealing a large set of subclonal variants, including pathogenic or likely pathogenic (P/LP) variants in both SNVs and indels, using this approach (Supplementary Fig. S1A and S1B).

All variants were reviewed by a committee of computational biologists, pathologists, oncologists, geneticists, and genetic counselors. Reportable P/LP alterations included SNVs (22%); indels (10%); gross chromosomal losses, gains, and LOH (33%); subarm copy (focal)-number abnormalities (20%); and SVs/gene fusions (44%). Overall, we reported a mean of three P/LP somatically acquired sequence variants per case (range 0–14) in addition to ploidy alterations, such as gross chromosomal gains and losses and chromosome arm–level regions of copy-neutral LOH (CNLOH). Two brain tumors had no reportable P/LP findings despite pathology review indicating adequate tumor tissue. The prevalence of variants deemed P or LP in hematologic malignances, brain tumors, and non-CNS solid tumors, affected by diverse mutational mechanisms, were consistent with other recent pediatric cancer genomic studies (refs. 8, 14, 15, 20–22; Fig. 2A; Supplementary Table S2).

Identification of Gene Fusions, Enhancer Hijacks, and Microdeletions

SVs causing gene fusions or enhancer hijacking are important drivers of pediatric cancers (23–25), yet they are challenging to detect by WES, as SV breakpoints are most often located in noncoding regions of the genome. Consistent with this finding, we showed previously that inclusion of WGS in the study of pediatric cancer genomes significantly improves the detection of driver gene alterations when compared with WES alone (14). Using the three-platform sequencing approach, we identified in-frame gene fusions in 90 tumors (36%) representing 44 distinct gene pairings and 34 distinct fusion driver genes, of which 23 were diagnostic for a specific cancer type or subgroup. Thirty of the 34 conferred a clear or potential diagnostic, prognostic, or therapeutic utility
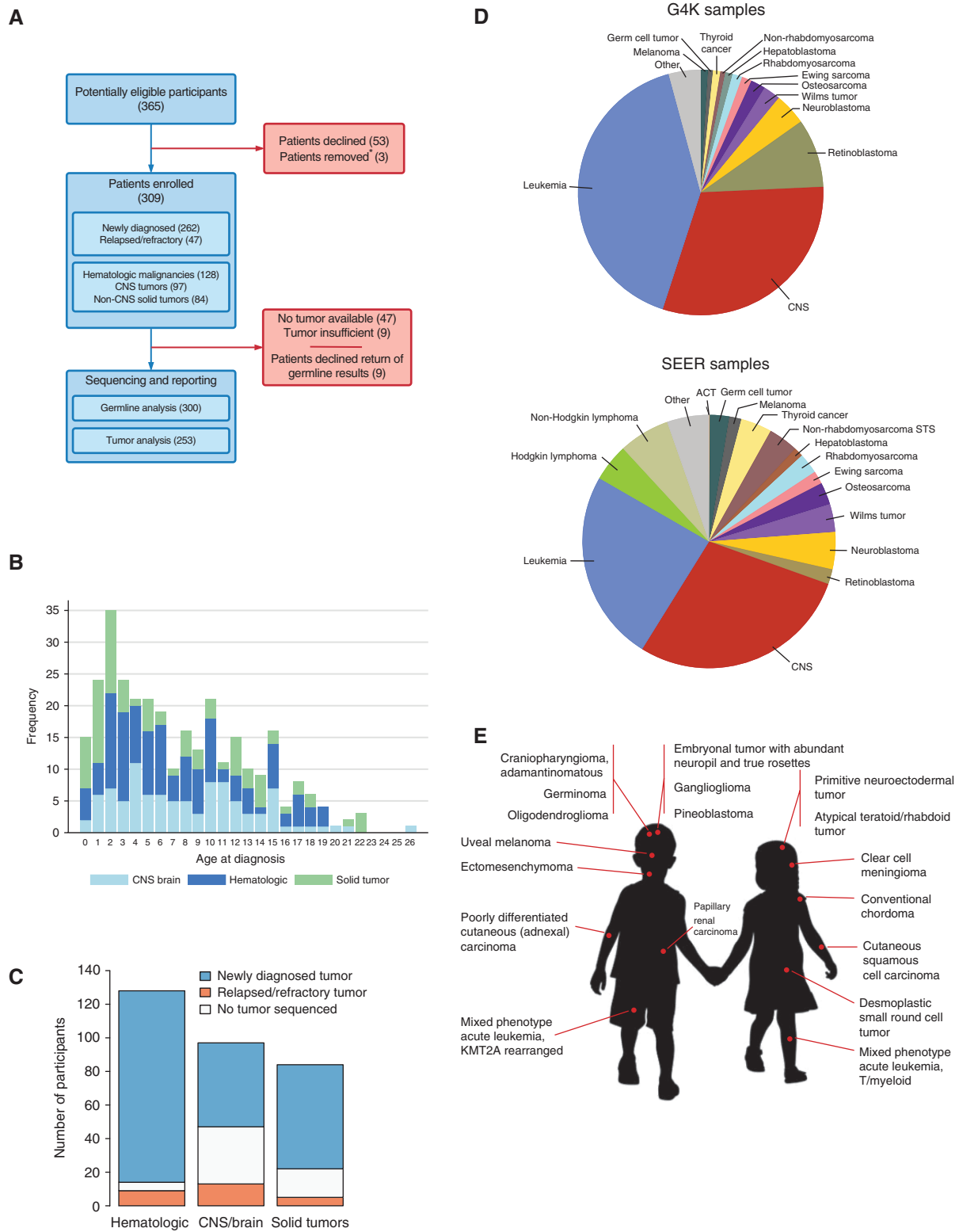
Figure 1. Patient accrual and demographic data. **A,** Consort diagram depicting patient accrual into G4K. *Three patients were removed from the study when pathology revealed that 1 patient had no cancer, another patient died before a germline sample could be collected, and another patient declined return of germline results and there was insufficient tumor for sequencing. **B,** Age distribution of patients. **C,** Number of patients with newly diagnosed or relapsed or refractory tumors, or no tumor available for sequencing, broken down by major tumor type. **D,** The distribution of cancer types represented in the G4K cohort (top) compared to the distribution of pediatric cancers in the NCI SEER program (bottom). ACT, adrenocortical tumor; STS, soft-tissue sarcoma. **E,** Eighteen rare tumor types found in the G4K cohort. Rare tumor types were defined as those present at less than 2 per million children annually in the United States.
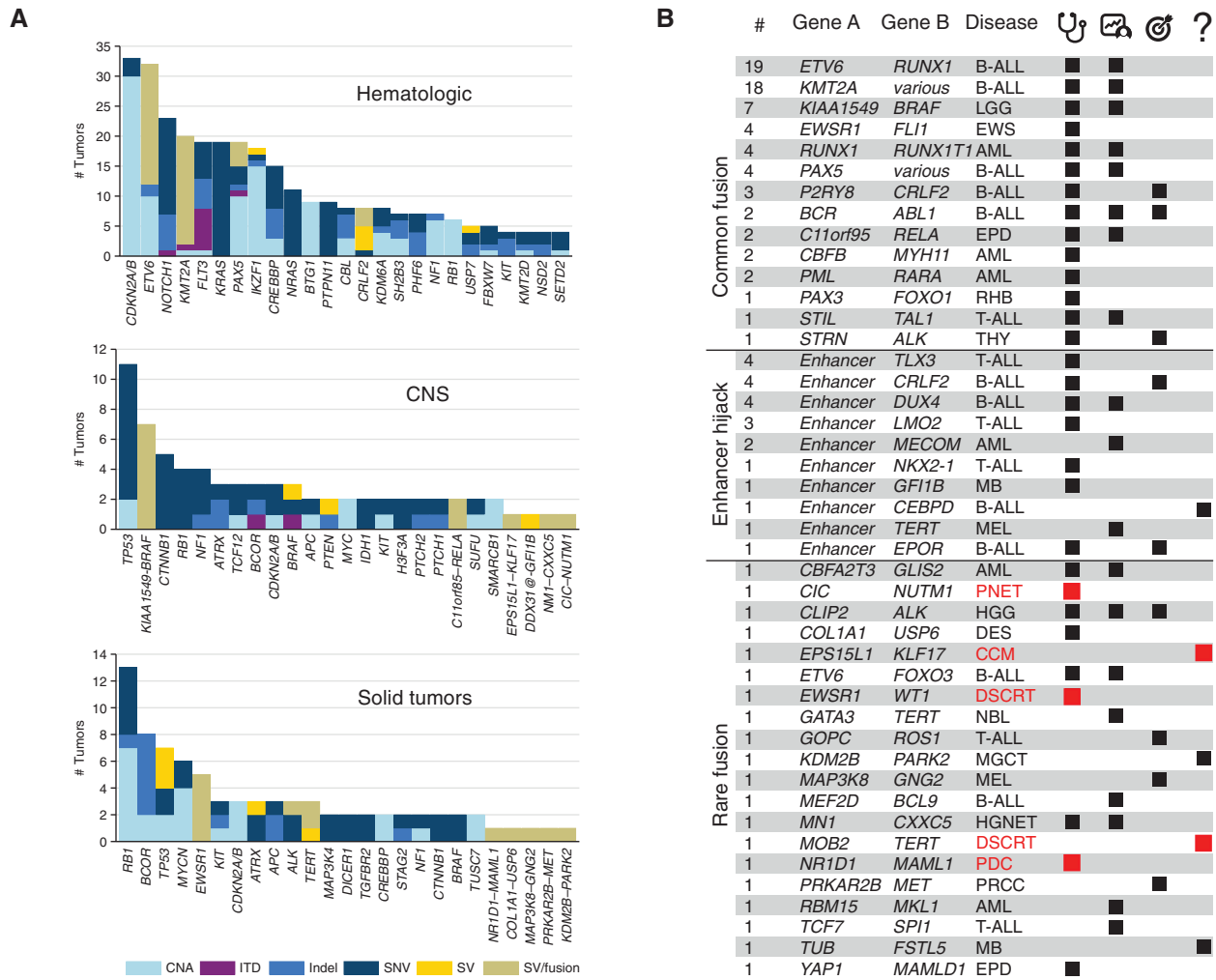
Figure 2. Somatic findings in the 253 analyzed tumors. **A,** Bar charts showing the numbers and relative contributions of mutational mechanisms affecting cancer genes in the tumors analyzed through G4K. The top 25 mutated genes for hematologic, CNS, and solid tumors are shown as are gene fusions or enhancer hijack events for singletons in CNS and solid tumors. **B,** Gene fusions and enhancer hijacks detected in G4K samples. Number of samples with a given fusion are indicated in the leftmost column, followed by the genes/loci involved and the diseases in which they were detected (see Supplementary Fig. S3 for schematics depicting the 20 rare fusions). Black or red tiles indicate whether the identified gene fusions or enhancer hijacks have a clear or likely clinical utility, arranged into three columns indicating diagnostic (stethoscope), prognostic (patient chart), and therapeutically relevant (target) categories. In the rightmost column (question mark), tiles indicate lesions with an unknown clinical utility, but considered biologically relevant to the tumor. Red disease names and tiles were identified in the rare tumors, as shown in Fig. 1E. Disease abbreviations and additional details regarding SV classifications and literature citations can be found in Supplementary Table S2.

(Fig. 2B; Supplementary Figs. S2 and S3; Supplementary Table S2). For example, brain tumor SJBT030081, originally classified as a high-grade glioma, was found to harbor an *MN1–CXCC5* gene fusion and thus reclassified as a CNS high-grade neuroepithelial tumor with *MN1* alteration, an entity only recently described (26). Five of the fusions shown in Fig. 2B (shown in red) were in patients with very rare tumor types (as shown in Fig. 1E).

In an additional 21 tumors (8%), combined WGS and RNA-seq analysis enabled detection of 10 distinct enhancer hijacking translocations, wherein an SV brings a transcriptionally active locus into close proximity to an oncogene, thereby driving its expression. We correlated outlier oncogene expression with SV breakpoints from WGS data to facilitate the accurate

detection of enhancer hijacking events, which ranged from 0 to 716 kb up or downstream of the target oncogene with outlier expression (see "Methods"; Supplementary Fig. S4; Supplementary Table S3). Enhancer hijacks included well-characterized events such as *IGH@-CRLF2*, *TCR@-LMO2*, *IGH@-DUX4*, *IGH@-EPOR*, and *DDX31@-GFI1B*. In addition, there were less well-characterized enhancer hijacks such as two instances of *CDK6@-MECOM* in acute myeloid leukemia (AML), which may have a negative prognostic impact (27), and a novel *TLX3*-activating translocation in a T-lineage ALL (T-ALL) sample as described below (see "Disease Relevance of Novel Somatic Variants").

Acute leukemias are susceptible to small microdeletions, resulting from off-target recombinase activating gene

(RAG)–dependent effects (28, 29). Detection of these deletions can be of prognostic significance, as is the case with *IKZF1* in B-lineage ALL (B-ALL), where intragenic deletions are associated with a poorer outcome (30). However, intragenic microdeletions can be difficult to detect using WES, because exons rarely capture deletion breakpoints and their identification must rely on subtle changes in depth of coverage (31, 32). Consistent with this notion, we previously observed that single-nucleotide polymorphism (SNP) array or WES has limited power for detecting focal and exon-poor CNAs due to insufficient physical space for robust detection of read-depth changes (14, 33). Using WGS data, we detected small intragenic deletions involving as few as one or two exons in cancer-relevant genes in 38 tumors (15%), with all but six being leukemias (Supplementary Table S4). Twenty-six genes were affected, including those expected (28), such as *BTG1*, *CDKN2A*, *ETV6*, *RAG1*, *TCF4*, *IKZF1*, and *TP53* among the most commonly involved. Several clinically relevant genes that are less commonly observed as targets of microdeletion, including *CREBBP*, *SH2B3*, *USP9X*, *FBXW7*, and *NR3C2*, were also impacted. The majority of these microdeletions (38 of 53 total events; 72%) involved loss of a single gene copy and thus would have been difficult to detect using WES.

Mutation Signatures and Etiology of Pediatric Tumors

The whole-genome mutational landscape of a tumor records its natural history and reflects the environmental and endogenous exposures that have contributed to tumor initiation and progression (34, 35). To elucidate the mutational landscape of pediatric cancers, we evaluated WGS tumor data for the relative proportions of mutation signatures reported by Alexandrov (ref. 36; Supplementary Table S5). In addition to well-known mutational processes such as spontaneous deamination (signature 1) evident in nearly every patient, signatures 2 and 13 reflecting activation induced cytidine deaminase (AID), and apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) activity were present in several of B-ALL cases (37, 38); signatures of ultraviolet radiation exposure were present in four B-ALL samples and a cutaneous melanoma (see also ref. 15), and three tumors exhibited signatures 6 and 15, indicating mismatch repair (MMR) deficiency, as well as signature 10, which is associated with mutated DNA polymerase epsilon (*POLE*). Six tumors harbored high levels of signature 18, thought to arise from DNA damage induced by reactive oxygen species (ROS), with one tumor exhibiting biallelic loss of *MUTYH*, the glycosylase that executes the first step in repair of ROS-induced mutations. The roles of the MMR deficiency and *POLE* and ROS signatures in disease pathogenicity are described further in the next section.

Disease Relevance of Novel Somatic Variants

Most pediatric cancer genomes harbor a limited number of somatic alterations, many of which are nonrecurrent, making it difficult to ascertain pathogenicity. Visualization of these variants in the context of large-scale public data was essential to classify some of these challenging *N* of 1 events. We illustrate this process with two examples. The first case, SJTALL030071, is a T-ALL that harbored a translocation t(5;8)(q35q24) linking an intergenic region on 8q24, 1 Mb

distal of *MYC*, to an intergenic region 29 kb distal to *TLX3*, a known driver of T-ALL (39). The copy-number profile of this tumor showed an 18-Mb duplication of the 8q24→8qter region that did not include *MYC* (Fig. 3A) but did include the *NOTCH MYC* enhancer (N-Me; ref. 40) that is known to activate *MYC* in T-ALL. Indeed, N-Me activity was confirmed by the high level of enhancer RNA and *MYC* expression in this tumor. The translocation juxtaposed N-Me 29 kb upstream of *TLX3*, leading to elevated TLX3 expression (Fig. 3B). Moreover, using SNP markers in the DNA at the *TLX3* locus, we demonstrated allele-specific expression (ASE) in the RNA, consistent with activation by a *cis*-acting regulator (Fig. 3C). Consistent with these findings, SJTALL030071 clustered by t-distributed stochastic neighbor embedding (t-SNE) analysis among other *TLX3*-driven T-ALLs from the NCI-Therapeutically Applicable Research to Generate Effective Treatments (NCI-TARGET) cohort (Fig. 3D). Altogether, these data suggested mechanistic similarity to known enhancer hijack events linking *TLX3* to T-cell receptor enhancer loci (39, 41), and accordingly we classified the translocation as LP.

The second case, SJBALL030052, is a B-ALL with a complex SV that exhibited DNA and RNA evidence of a novel fusion including exon 2 of *ETV6* with exon 2 of *FOXO3* (Fig. 3E). The impact of this fusion was not initially clear as it could conceivably result in *ETV6* loss of function (LOF) or in a gene fusion functionally similar to *ETV6–RUNX1*. To gain further insights, we compared the RNA-seq gene expression profiles between our B-ALL sample and the publicly available RNA-seq profiles from 1,988 other childhood B-ALL cases (42). SJBALL030052 clustered with the *ETV6–RUNX1* subgroup (Fig. 3F), suggesting that the fusion impacted gene expression similarly to *ETV6–RUNX1*. We thus classified the novel fusion as LP, which supported this patient's clinically determined standard-risk classification.

Across all 253 tumors, 65 (25%) harbored a total of 89 genetic alterations that were considered rare or unreported in the tumor type being investigated (Supplementary Table S6). For example, we found *AKT1*-activating variants in two patients with T-ALL, SJTALL030064 and SJTALL030134. *AKT1* represents a targetable oncogene in epithelial cancers, but it is only very rarely altered in T-ALL, and its role and efficacy as a therapeutic target have yet to be investigated in this cancer type. These rare events were often suggestive of unusual mechanisms of activation of an oncogene or signaling pathway. For example, *USP9X* LOF was found in two cases and potentially activates JAK/STAT signaling (43). Of these 90 rare or novel events, 23 (26%) were SVs predicted to generate fusions or enhancer hijacks identified by WGS and RNA-seq analysis. These observations demonstrate that the spectrum of cryptic intergenic events continues to expand through use of these combined nonbiased genomic approaches.

Germline Cancer-Predisposing Variants

We analyzed 156 cancer predisposition genes using germline WES and WGS data for the 300 patients who consented to return of germline results. This list included 63 genes associated with moderate to high cancer penetrance (9), as well as 93 additional genes, approximately half of which are associated with autosomal recessive cancer-predisposing conditions (Supplementary Table S7). We classified variants according

A

chr8:128,500,000–130,600,000 [hg19]

B

C

D

E

F

to the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) 2015 germline variant interpretation guidelines (44). Germline P, LP, or variants of uncertain significance (VUS) among the 63 moderate/highly penetrant genes were reported back to providers and patients. For the remaining 93 genes, only P or LP variants were included in the final clinical reports.

We identified 58 germline P or LP variants affecting 29 genes in 55 (18%) of 300 patients (Fig. 4A; Supplementary Table S8) and 420 VUS affecting 111 genes in 230 (77%) patients (Supplementary Table S9). The prevalence of P/LP variants ranged from 10% for patients with hematologic malignancies to 40% to 50% for those with retinoblastoma and other solid tumors (Fig. 4B). Thirty-two (55%) of the germline P/LP variants affected genes not generally associated with the patient's tumor type, such as a germline mutation in *WT1* in a patient with B-ALL (SJBALL030057) or a germline mutation in *BRCA2* in a patient with a diffuse intrinsic pontine glioma (SJHGG030328; Supplementary Table S8). Therefore, if targeted, germline gene panels relevant to the patient's tumor type had guided testing, these variants might have escaped detection, as many of the panels would not have included the affected genes.

Disease Relevance of Novel Germline Variants

An integral component of our germline variant pathogenicity classification involved simultaneous review of paired tumor–normal data to determine the molecular impacts of germline variants on RNA splicing and expression, tumor mutation signatures, and tumor mutation burden (TMB). For example, transcriptome data from Ewing sarcoma, SJEWS030332, harboring a novel *BAP1* intronic variant at the -3 position of exon 5 (NM_004656.3: c.256–3C>A), revealed evidence of intron retention in the tumor relative to other Ewing sarcoma samples in the study. Fisher exact test showed a significant difference in number of variant-bearing RNA-seq intronic reads relative to PCR-free tumor WGS ($P = 0.047$; Fig. 4C). Intron retention, unveiled by the use of tumor RNA data, provided sufficient evidence to classify this novel germline variant as LP.

A patient with B-ALL, SJBALL030144, presented with café au lait macules but no coding mutations to explain this clinical phenotype. Analysis of tumor RNA data revealed that a germline *NF1* variant at the +3 position of exon 45 (NM_000267.3: c.6858+3A>G) caused skipping of exon 45,

which is predicted to lead to out-of-frame translation of the NF1 protein (Fig. 4D). This variant was classified as LP. The opposite conclusion was obtained for a germline *APC* variant (NM_000038.5: c.449A>G) in patient SJST030310. Analysis at an external clinical laboratory predicted creation of a *de novo* splice site with subsequent LOF; however, review of tumor transcriptome data provided no evidence of altered splicing, and we thus classified this variant as a VUS (Supplementary Fig. S5).

TMB and mutation signatures derived from WGS were used to establish the roles of germline variants in generating the molecular phenotypes observed in some tumors. Tumors were classified as hypermutators based on a TMB >10 mutations per Mb and ultramutators with >100 mutations per Mb (45). Two cases (SJHGG030335, SJHM030291) harbored germline biallelic *PMS2* pathogenic variants leading to tumors that exhibited TMB and mutational signatures consistent with MMR deficiency (Supplementary Table S5). SJHGG030335 harbored >100 mutations per Mb due to an acquired *POLE* S459F pathogenic somatic variant. The *POLE*-related mutational signature was also exhibited by this tumor.

Finally, SJST030211, a squamous carcinoma of the lip, exhibited a hypermutator phenotype corresponding to Catalogue of Somatic Mutations in Cancer (COSMIC) signature 11, which is linked to temozolomide treatment (Supplementary Fig. S6A and S6B). Consistent with this finding, this patient had received prior therapy with temozolomide for a low-grade glioma and had no detectable mutations in MMR genes in the germline or the squamous carcinoma. Thus, the mutator phenotype was caused by therapy, with this information obviating the need for follow-up germline testing for the patient and family.

As more children with cancer undergo gene panel or multiplatform sequencing, an increasing number of pathogenic germline variants are being identified in genes not generally associated with the patient's tumor type, and in some cases, these variants are associated with adult-onset conditions or autosomal-recessive cancer predisposition syndromes (4, 5, 9, 22, 46). Nevertheless, it remains poorly understood whether or how these germline variants contribute to the pathogenesis of pediatric cancers. Therefore, we reviewed all germline P/LP variants in the context of each patient's tumor type to determine whether the germline variant might

Figure 3. Using multi-omics data to interpret pathogenicity of SVs. **A,** GenomePaint plots showing two regions, chr8:128500000–130600000 (top) and chr5: 170710000–170790000 (bottom) [hg19], from SJTALL030071. Both panels consist of the RefSeq gene model in green, with *MYC* and *TLX3* highlighted by red boxes; orange bars show regions of copy-number gain supported by increased WGS coverage plotted as the blue histogram immediately below. Gray lollipops marked t(5;8) indicate the position of the translocation breakpoint. RNA-seq coverage is shown below the whole-genome coverage histogram. Additional data from NCI-TARGET are also shown, with narrow red bars representing regions of copy-number gain and gray lollipops representing SV breakpoints surrounding the *TLX3* locus. A region or recurrent copy-number gain in TARGET samples is adjacent to the chromosome 8 breakpoint in SJTALL030071. Generally high but nonspecific RNA-seq coverage at this locus suggests a region of high transcriptional activity (distal *MYC* enhancer) is brought into proximity of *TLX3* by the translocation. CNV, copy-number variation. **B,** A rank-order plot of T-ALL from TARGET showing expression levels of *TLX3* mRNA in a set of *TLX3*-activated tumors compared with tumors in which *TLX3* was not activated. SJTALL030071 *TLX3* expression (red dot) groups with the activated set. **C,** Allele-specific expression (ASE) of the *TLX3* locus in SJTALL030071. The tumor DNA (top row) shows a series of heterozygous alleles in the *TLX3* locus (blue and red stacked bars show relative VAF from WGS data). In the RNA-seq data (second row), expression of only one allele is observed. RNA coverage, read counts at each allele, are shown numerically (third row). Beneath the read counts, black lines map the locations of the alleles to the chromosome 5 coordinates surrounding the *TLX3* locus. Beneath the coordinate line, the location of the SJTALL030071 translocation breakpoint is indicated. **D,** Two-dimensional t-SNE plot of RNA-seq–derived gene expression data from 264 T-ALL samples (41). Major T-ALL subgroups are indicated on the plot, with SJTALL030071 localizing among the *TLX3* cluster, as shown by the black arrow. **E,** Schematic representation of an *ETV6–FOXO3* fusion found in SJBALL030052 joining the N-terminal region to the *ETV6* sterile alpha motif domain (green) with oligomerization interfaces (red) and ETS domain (purple), with the C-terminal *FOXO3* forkhead binding (green), KIX binding (purple), and transactivation domains (red). **F,** Two-dimensional t-SNE plot of RNA-seq–derived gene expression data from 1,988 B-ALL samples (42). Major subgroups are indicated on the plot, with SJBALL030052 localizing to the periphery of the *ETV6–RUNX1* subgroup, as shown by the black arrow.
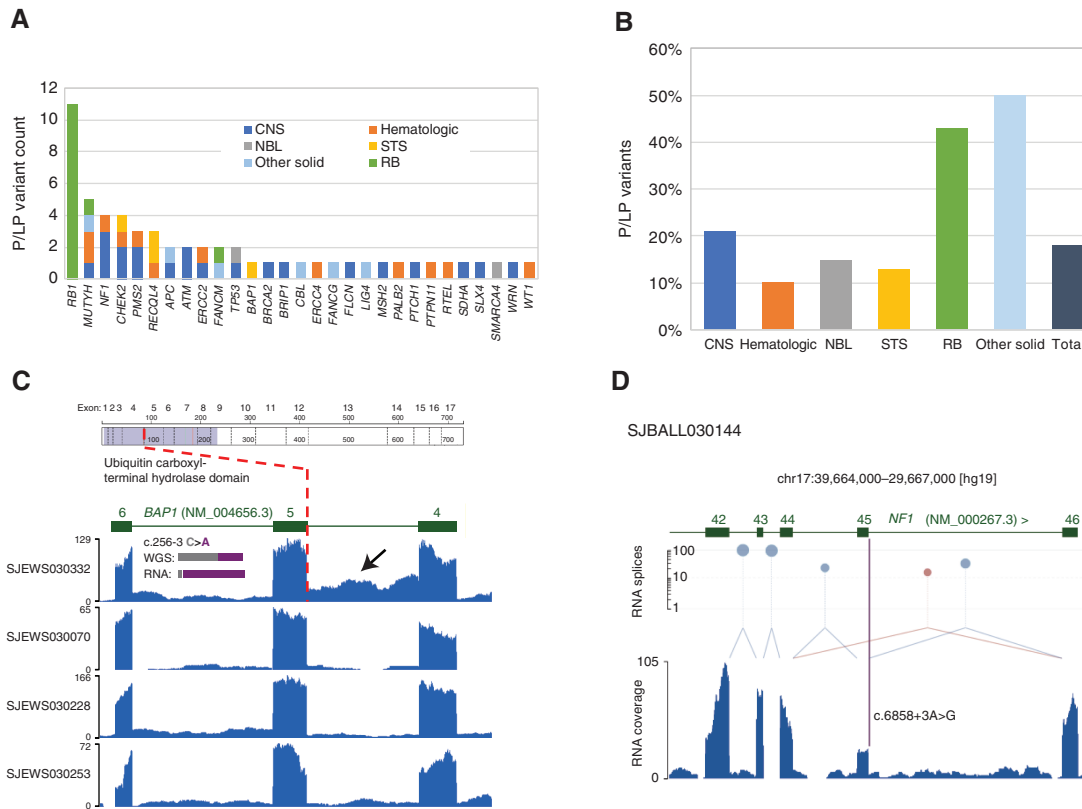
Figure 4. Germline variants and assessment of variant pathogenicity based on RNA data. **A,** Numbers of germline P/LP variants, broken down by gene and tumor type. NBL, neuroblastoma; RB, retinoblastoma. **B,** Proportions of germline P/LP variants, broken down by tumor type. **C,** *BAP1* intron 4 retention in SJEWS030332 compared to other G4K Ewing sarcoma cases. Each blue histogram shows hg19-aligned RNA-seq coverage relative to the *BAP1* gene model in green (note that *BAP1* is on the negative strand). The position of the exon 5 splice acceptor mutation is indicated by the red dotted line. Increased read coverage in the SJEWS030332 (bearing a mutation at the -3 position of exon 5) intron relative to the three other samples indicates intron 4 retention (black arrow). Inset histograms show the relative proportion of reference and variant alleles in tumor-derived WGS and RNA-seq in gray and purple, respectively. Corresponding read counts are WGS: 32G/21T (40% variant allele) and RNA 2G/28T (93% variant allele). Above the RNA coverage plots is a schematic of the BAP1 protein with the location of the splice variant leading to protein truncation marked. **D,** *NF1* exon 45 skipping in SJBALL030144. The blue histogram shows RNA-seq coverage relative to the *NF1* gene model in green. Canonical splices are shown as light blue links between exons, and a noncanonical splice is shown in mauve. The height of mauve and blue lollipops is proportional to the number of splice junction reads detected plotted on a log scale on the *y*-axis. The purple bar indicates the position of the *NF1* exon 45 splice acceptor mutation. Exon 45 expression is diminished relative to flanking exons, and a noncanonical splice linking exons 44 and 46 is observed, indicating an exon skipping event.

have played a causal role based on the molecular phenotypes observed in the tumor. We considered a germline variant relevant to development of the child's tumor if the gene had a known association with the child's tumor type or if there was specific molecular evidence supporting a functional consequence of the mutation in the tumor. If neither of these criteria were met, the relevance of the variant in the tumor was considered to be unknown (Supplementary Fig. S7).

Using this scheme, 32 of 58 (55%) germline P/LP variants affecting 15 genes were characterized as relevant to tumor formation ("Disease related," Supplementary Table S8; Fig. 5). Most of these genes have known relationships with pediatric cancer, such as *RB1* in retinoblastoma, *NF1* and *PMS2* in glioblastoma, and *PTCH1* in medulloblastoma. Examination of tumor WGS and RNA-seq data, protein expression, and the literature enabled classification of disease relevance in several *N* of 1 cases, illustrated in the following examples.

First, we observed a missense germline *TP53* variant, (A161T), in patient SJNBL030203 with neuroblastoma. Although prior studies have suggested that this variant is

damaging, with a dominant negative LOF effect (47–49), neuroblastoma occurs only very rarely (~1%) in individuals with pathogenic germline *TP53* variants (50). Furthermore, like most neuroblastoma (51), this patient's tumor retained the wild-type *TP53* gene copy. However, examination of tumor RNA-seq data revealed expression of only the variant allele (Fig. 6A). Together these data suggest that there is transcriptional silencing of the wild-type gene copy, a mechanism not yet described in neuroblastoma, and we considered this germline variant relevant to the child's tumor.

Patient SJRB030050 with retinoblastoma harbored a germline *MUTYH* founder variant (G396D; Fig. 6B) that was rendered homozygous in the tumor through CNLOH. The TMB was in the upper quartile among all retinoblastoma tumors in the G4K study, as well as a larger set of retinoblastoma tumors evaluated through the Pediatric Cancer Genome Project (PCGP; Fig. 6C). Notably, over half of this tumor's mutations were attributable to the ROS signature 18 (Fig. 6D). These features indicate that the loss of *MUTYH* function contributed to the mutational processes in this tumor. As a result, the germline
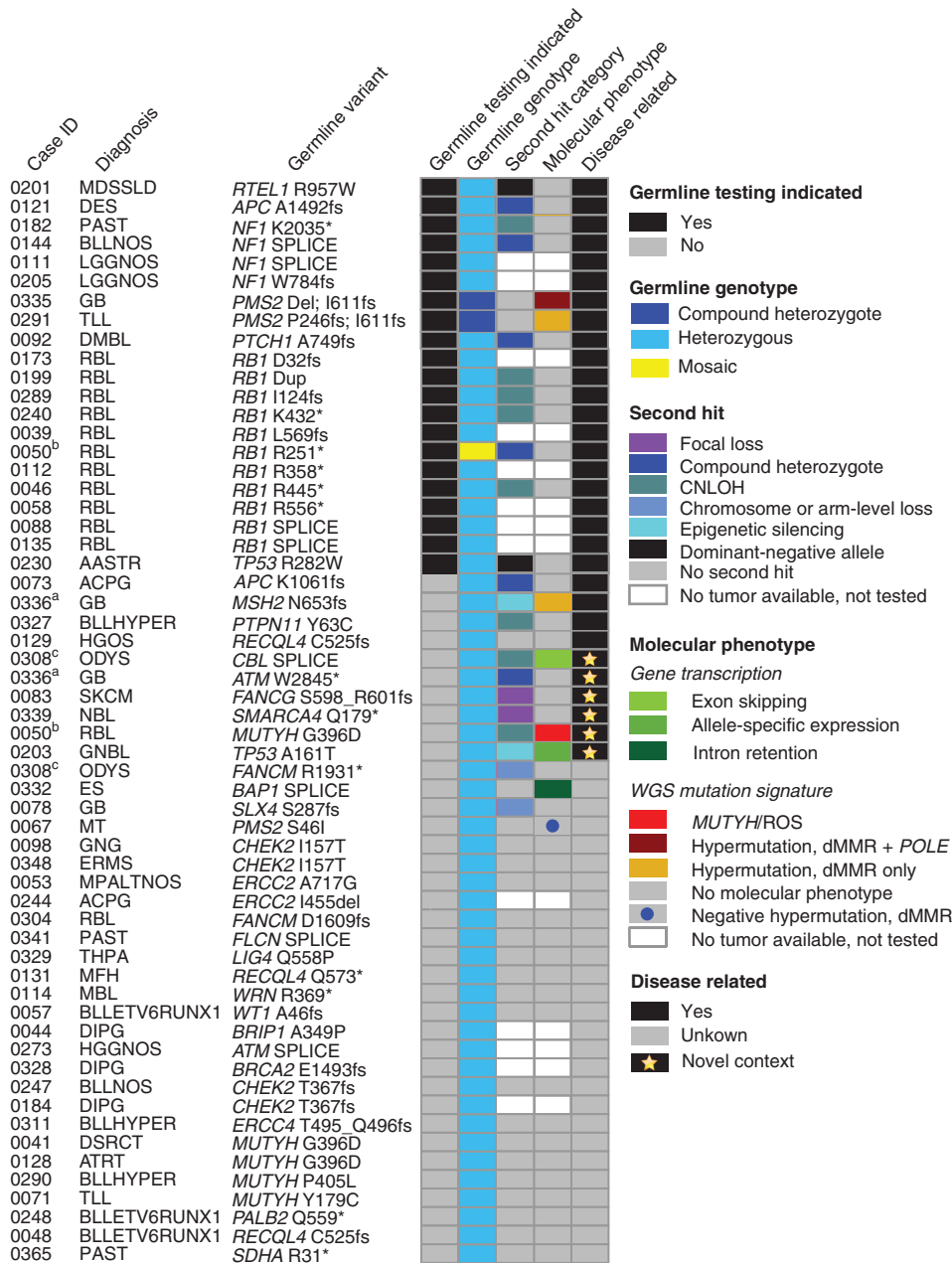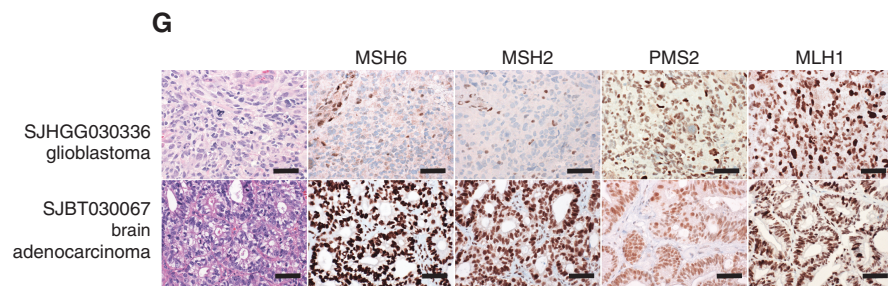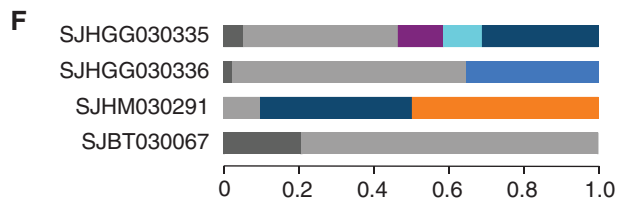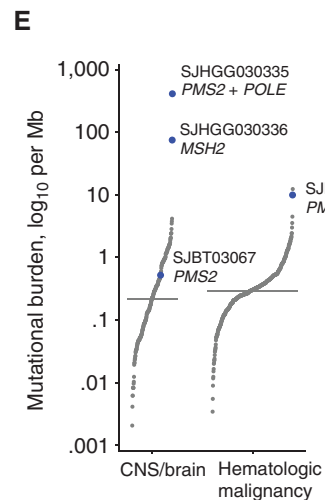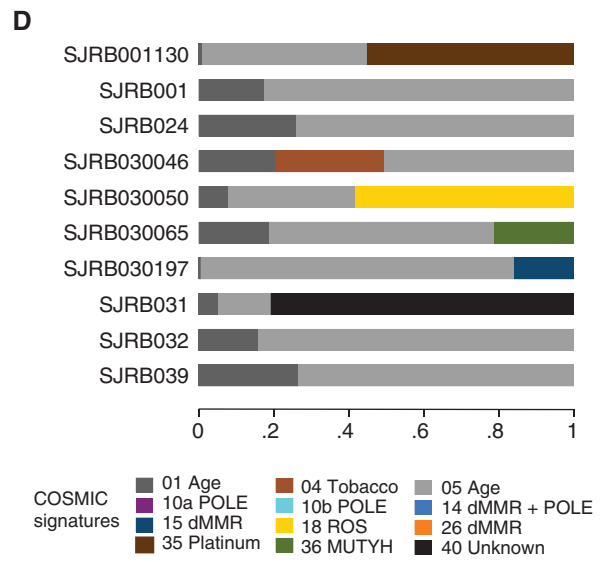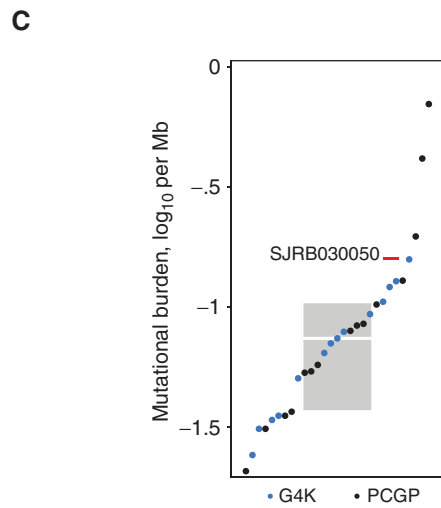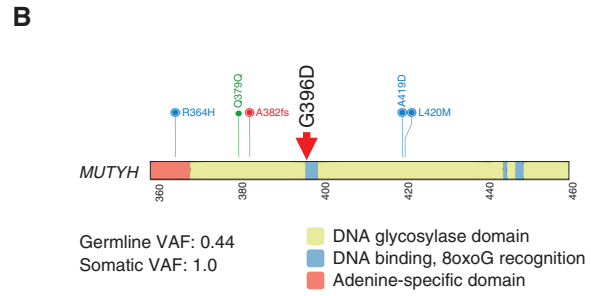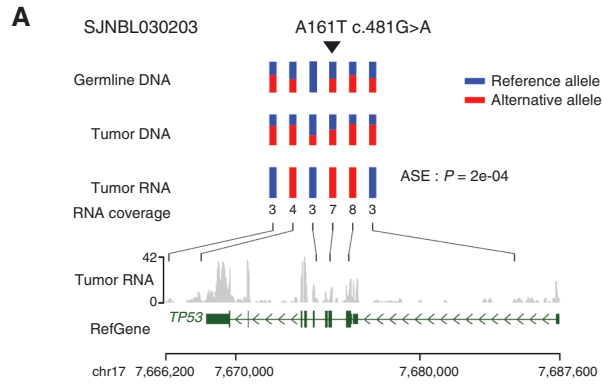
Figure 5. The impact of somatic variation in establishing disease relevance of deleterious germline variants. Each row represents a unique patient. From left to right, the columns are as follows: "Case ID" gives the last 4 digits of the patients' IDs (compare Supplementary Table S8, column A). "Diagnosis" gives the disease code used internally (compare Supplementary Table S8, column B). Note: Matched superscripts indicate mutations in the same patient. "Germline variant" lists the gene and amino acid change. "Germline Testing Indicated" signifies patients whose cancer or other phenotypic characteristics suggested the patient and possibly the family should undergo germline testing (black tiles). "Germline genotype" gives the genotype of the variant indicated in the "Germline variant" column. "Second hit category" gives the genetic configuration of any genetic or epigenetic alterations affecting the remaining wild-type gene copy in the tumor. "Molecular phenotype" pertains to evidence in the DNA sequence of the tumor as to the activity of the germline variant. "Molecular phenotypes" includes features such as splice aberrations visible in the RNA-seq data and mutation signatures. Tumor second hits and molecular phenotype were not available for some patients due to absence or inadequate tumor for testing (see Methods). The 5 columns of tiled cells are sorted to group germline mutations by "Disease related" and then secondarily by "Germline Testing Indicated." dMMR, deficient MMR.

MUTYH variant was classified as relevant to retinoblastoma development in this patient. In 2 other patients harboring the same germline *MUTYH* founder mutation (Supplementary Table S8), the variant was considered of uncertain relevance to disease because there was no loss of the wild-type allele and no ROS mutation signature in their tumors.

Two patients harbored a single heterozygous mutation in one of the MMR genes characteristic of Lynch syndrome (LS; also known as hereditary nonpolyposis colon cancer). Patient SJHGG030336 with glioblastoma harbored a germline *MSH2* variant (N653fs) and patient SJBT030067 with an unusual adenocarcinoma of the pineal region of the brain harbored

A

SJNBL030203

A161T c.481G>A

Germline DNA

Reference allele
Alternative allele

Tumor DNA

Tumor RNA

ASE : *P* = 2e-04

RNA coverage   3 4 3 7 8 3

Tumor RNA

RefGene   *TP53*

chr17 7,666,200   7,670,000   7,680,000   7,687,600

B

G396D

R364H Q379Q A382fs   A419D L420M

MUTYH
360   380   400   420   440   460

Germline VAF: 0.44
Somatic VAF: 1.0

DNA glycosylase domain
DNA binding, 8oxoG recognition
Adenine-specific domain

C

Mutational burden, log$_{10}$ per Mb

SJRB030050

• G4K   • PCGP

D

SJRB001130
SJRB001
SJRB024
SJRB030046
SJRB030050
SJRB030065
SJRB030197
SJRB031
SJRB032
SJRB039

0   .2   .4   .6   .8   1

COSMIC
signatures

01 Age
10a POLE
15 dMMR
35 Platinum

04 Tobacco
10b POLE
18 ROS
36 MUTYH

05 Age
14 dMMR + POLE
26 dMMR
40 Unknown

E

Mutational burden, log$_{10}$ per Mb

SJHGG030335
PMS2 + POLE

SJHGG030336
MSH2

SJHM030291
PMS2

SJBT03067
PMS2

CNS/brain   Hematologic
malignancy

F

SJHGG030335
SJHGG030336
SJHM030291
SJBT030067

0   0.2   0.4   0.6   0.8   1.0

G

MSH6   MSH2   PMS2   MLH1

SJHGG030336
glioblastoma

SJBT030067
brain
adenocarcinoma

a *PMS2* variant (S46I). Neither tumor exhibited mutation or loss of the respective wild-type gene copy. Nevertheless, the *MSH2*-mutant glioblastoma (SJHGG030336) was hypermutated (Fig. 6E) and exhibited a mutation signature associated with microsatellite instability (MSI; Fig. 6F), consistent with the LS phenotype. On the basis of the high TMB and mutation signatures, IHC staining of this child's tumor was performed, revealing absent MSH2 and MSH6 expression (Fig. 6G, top), and confirming a deficiency in MMR. The germline *MSH2* N653fs variant was thus deemed relevant to this child's glioblastoma. In contrast, the pineal adenocarcinoma in patient SJBT030067 did not exhibit hypermutation (Fig. 6E) or the mutation signature of MSI (Fig. 6F), and it retained PMS2 expression (Fig. 6G, bottom). Therefore, the disease relevance of the germline *PMS2* variant to development of this child's adenocarcinoma remained uncertain.

Finally, patient SJNBL030339 with neuroblastoma harbored a pathogenic germline variant in *SMARCA4* with focal deletion of the wild-type allele observed in the tumor (Fig. 5). The oncogenic potential of germline *SMARCA4* variants is known to predispose to small-cell carcinoma of the ovary and less commonly to rhabdoid tumor (52), but it has not been linked to neuroblastoma. However, our finding raises the possibility that *SMARCA4* represents a target for mutational inactivation in neuroblastoma. Consistent with this notion, SMARCA4 is expressed in neural tissues and somatic biallelic *SMARCA4* variants are present in 1% of neuroblastoma tumors (53). Furthermore, the literature includes two cases of neuroblastoma with pathogenic germline *SMARCA4* variants (4, 54). Taken together, these data strengthen the association between germline *SMARCA4* variants and predisposition to neuroblastoma.

Management Implications Resulting from Genomic Findings

Although the G4K study was nontherapeutic by design and thus not intended to match patients to a targeted therapy, we sought to investigate whether genomic findings obtained from three-platform sequencing would inform diagnosis, risk stratification, cancer treatment, and/or genetic counseling of the patient and family (6, 55, 56). We deemed variants as therapeutically actionable or potentially actionable if they were classified as P/LP from a somatic perspective and they or their downstream pathways could be targeted (6). To make

these determinations, we used available evidence (46, 57, 58), including FDA-approved targeted therapies, eligibility for treatment on a clinical trial based on NGS results, and other professional guidelines or associated information (https://www.oncokb.org/; https://civicdb.org/home).

In total, 218 of 253 patients (86%) who underwent sequencing of paired tumor and normal tissues had at least one finding that was diagnostic (53%), prognostic (57%), therapeutically targetable (25%), involved in cancer predisposition (18%), or some combination of these features. (Fig. 7A and B; Supplementary Table S10). Among the somatic alterations, the preponderance of targetable abnormalities included gene fusions or hotspot mutations in kinases including *BRAF*, *ALK*, *MET*, and *ABL1*. We also detected several indirectly targetable mutations/fusions affecting genes upstream of the JAK–STAT pathway involving *EPOR*, *CRLF2*, *SH2B3*, and *USP9X*. Fusions of *EPOR* and *CRLF2* and truncations of *SH2B3* and *USP9X* are associated with Philadelphia chromosome–like B-ALL (57, 59), which is under investigation for treatment with JAK1/2 inhibitors (NCT03117751, NCT02723994).

At the time of diagnosis, the majority of pediatric patients with cancer, including those in the G4K study, are placed on Institutional Review Board (IRB)–approved clinical trials. However, 78 G4K patients with sequenced tumors presented with relapsed or metastatic tumors for which there was no clinical trial available, or they developed metastatic, relapsed, or refractory disease during the course of the study. Thus, approximately one third of the patients were eligible for a change in therapy, including an NGS-directed agent. Thirty-two (41%) of the tumors in this relapsed/refractory group harbored one or more targetable or potentially targetable lesions or a targetable mutation signature. Twelve patients (38%) received a targeted agent matched to their tumor's genetic lesion or mutational signature (Supplementary Table S11). Five of these 12 patients responded to the genomics-directed therapy [1 patient with AML (complete response), 1 with melanoma (partial response), 2 with glioblastoma (stable disease, SD), and 1 with craniopharyngioma (SD); Fig. 7C]. The 2 patients whose glioblastomas exhibited mutational signatures consistent with MMR deficiency were treated with checkpoint inhibitors with 1 of these 2 surviving for 2 years under this therapy while exhibiting no significant side effects. Two patients remain alive at the time of writing: SJAML030286, whose AML harbored a somatic *FLT3* ITD

Figure 6. Evaluation of germline and tumor data to establish disease relevance. **A,** P53 A161T is a weakly functional likely pathogenic p53 variant in SJNBL030203. Six markers, including the germline variant C>T at chr17:7578449 (demarked by upside down triangle), are represented by three rows of vertical bars. The bars are colored red and blue to show the allele fractions in WGS data. The rows indicating germline and tumor DNA show most positions as heterozygous, with both red and blue portions. In the RNA, all markers are monocolored, indicating that only one allele is expressed. At the location of the germline variant (C>T), the bar is completely red, indicating that only the variant allele is expressed, suggesting that the wild-type allele is transcriptionally silenced. ASE, allele-specific expression. **B,** A pathogenic *MUTYH* germline founder mutation, G396D, in retinoblastoma patient SHRB030050. **C,** The TMB for SHRB030050, from WGS data, is in the upper quartile compared with other retinoblastoma patients, including those from St. Jude Cloud. Box includes the second to third quartiles; horizontal bar within box is median. **D,** Mutation signatures from WGS data for retinoblastoma patients available in G4K and from the PCGP in the St. Jude Cloud resource (https://www.stjude.cloud/). Sixty percent of tumor mutations in SHRB030050 are attributable to damage by ROS. **E,** TMB plots of brain and hematologic tumors. The three tumors that are hypermutated (those with >10 mutations/Mb) are labeled with patient ID and the genes that were mutated to cause hypermutation. A fourth brain tumor with a TMB close to median was heterozygous for mutation of *PMS2*. The two hypermutated patients with *PMS2* carried compound heterozygous mutation of the gene. The patient with the highest TMB also carried an S459F mutation of *POLE*. **F,** Mutation signatures of the patients in **E**. **G,** IHC staining for the indicated MMR proteins in the brain tumors of patients SJHGG030336 and SJBT030067. Top, infiltrative astrocytoma with severe cytologic atypia, mitotic activity, and necrosis, diagnostic of glioblastoma. IHC for MMR proteins exhibited loss of *MSH2* and *MSH6* but retained staining in stromal elements. Bottom, gastrointestinal-type adenocarcinoma arising in malignant mixed germ cell tumor. Staining for MMR proteins demonstrated retained expression of *MSH2*, *MSH6*, *PMS2*, and *MLH1*. All images are 40× magnification; scale bars, 40 μm.
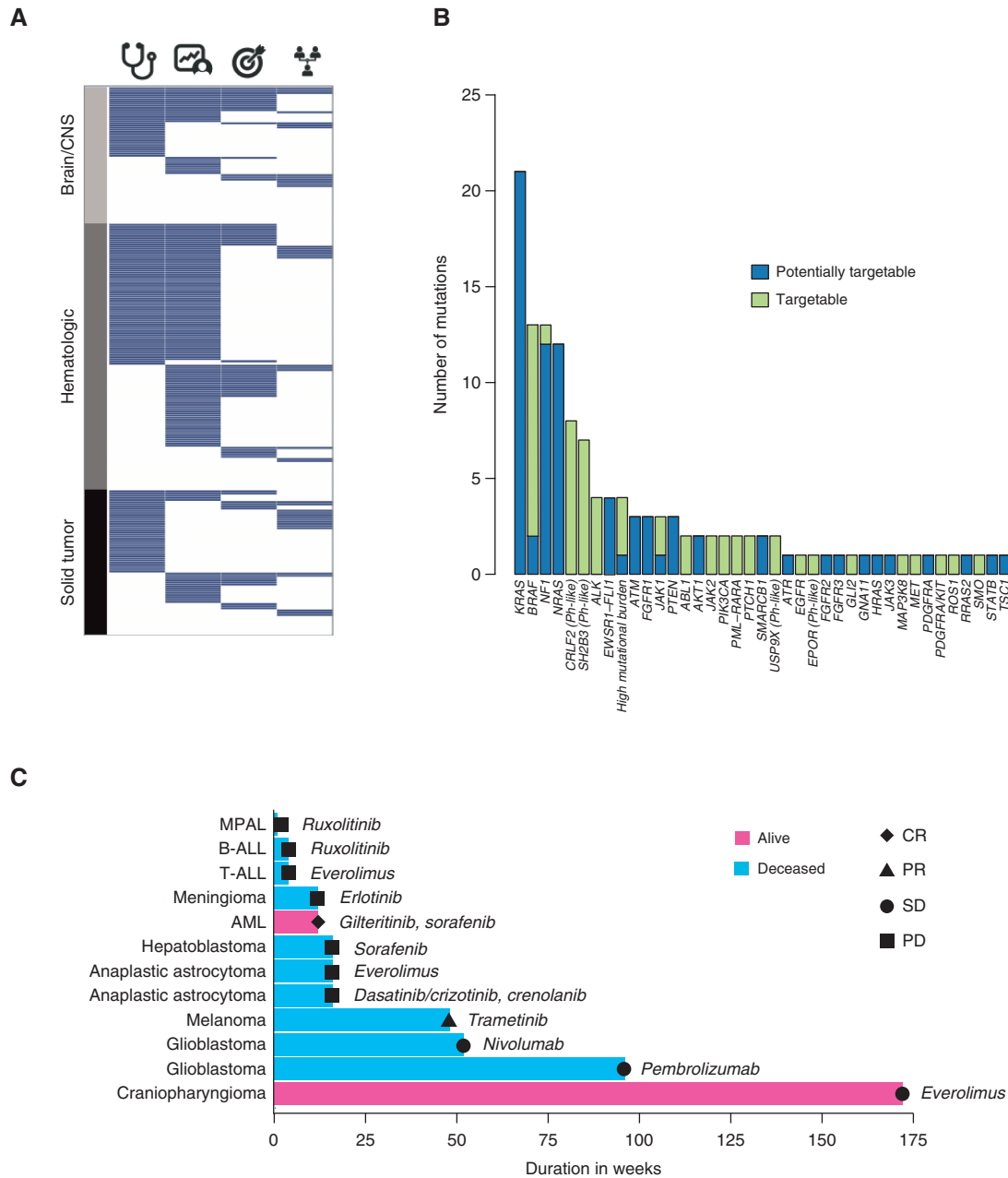
Figure 7. Clinically actionable findings. **A,** Tile plot summarizing clinically actionable findings in the 253 patients who had both tumor and normal tissues sequenced. Each tumor is represented as a row and is grouped according to major tumor type. Columns represent the presence (blue) or absence (white) of a diagnostic (stethoscope), prognostic (patient chart), therapeutically relevant (target), or cancer-predisposing mutation (pedigree). **B,** Tumors with targetable (tiers 1 and 2, green) or potentially targetable (blue) lesions, categorized by the affected gene, identified by the three-platform sequencing approach. Additional information can be found in Supplementary Tables S2 and S10. **C,** Swimmer plot depicting patients receiving a targeted therapy matched to their tumor genetic lesion. Each bar is 1 patient, with the disease as labeled. Pink bars, patient is alive; blue bars, patient is deceased. Best response on the targeted therapy is as labeled: CR, complete response; PD, progressive disease; PR, partial response; SD stable disease. The drugs used are labeled adjacent to each bar. See Supplementary Table S11 for further details. MPAL, mixed phenotype acute leukemia.

and was treated with sorafenib and gilteritinib-containing regimens for a total of 12 weeks followed by allogeneic stem cell transplantation, and SJBT030073, whose multiply recurrent craniopharyngioma harbored a somatic *PIK3CA* mutation and was treated with everolimus for 172 weeks. This patient's tumor progressed and was subsequently partially resected with sequence analysis of the resected tumor showing loss of the *PIK3CA* variant.

All participants with P/LP germline variants have undergone genetic counseling, and first-degree relatives have been offered clinical evaluation and targeted genetic testing; more distant family members have been encouraged to pursue

counseling in their home communities. Although cascade testing is still ongoing, 27 of 77 (35%) tested individuals from 31 families have been found to harbor a cancer-predisposing germline variant (Supplementary Table S8), and all have been provided with recommendations for cancer surveillance and risk-reducing measures when appropriate. To date, 19 of 58 germline P/LP variants (33%) have been confirmed as inherited, while 12 (17%) are *de novo* in origin. The remaining 27 are of unknown inheritance.

Comparison to Gene Panels

We next sought to determine what proportion of the events found by the three-platform sequencing approach would also be detected by commercially available gene panels. For this analysis, we selected four commonly used somatic panels: FoundationOne CDx, FoundationOne Heme, Oncomine v3, and OncoKids Hotspot. All four panels detect coding SNVs and indels, including exonic ITDs and focal CNAs. For a defined set of genes, FoundationOne CDx detects common intronic translocation breakpoints in DNA and the other panels detect fusion transcripts in cDNA. For comparison of novel gene fusions, we required that the panel include only one partner gene of the fusion pair discovered in our data set (see "Methods," "Comparison to Gene Panels," and Supplementary Tables S12–S14 for further details).

We observed that 42% to 84% of the evaluable P/LP SNVs/indels/ITDs/fusions/SVs and focal CNAs could have been detected by one or more of the panel designs (Supplementary Tables S12 and S13). Oncomine, applied to pediatric hematologic malignancy cases, covered the G4K variants least efficiently. This is not surprising because the panel is optimized for evaluation of adult solid tumors. The pediatric-specific OncoKids panel fared better, covering 65% and 74% of reported abnormalities for brain/CNS tumors and hematologic malignancies, respectively. The best performing panel was FoundationOne Heme when applied to pediatric hematologic malignancy cases, covering 84% of reported abnormalities. Confining our analysis to include only clearly targetable mutations (i.e., omitting potentially targetable mutations such as $KRAS^{G12D}$, $EWSR1$–$FLI1$, and others as defined in Supplementary Table S2), panel coverage ranged from 39% to 85% with Oncomine applied to pediatric hematologic malignancies detecting the fewest alterations and FoundationOne CDx applied to solid and brain tumors detecting the most (Supplementary Table S14).

On a per-patient basis, between 36% and 69% of patients had at least one reported mutation through G4K that was not covered by one of these panel designs (Supplementary Tables S13 and S14). This number includes abnormalities with a functional impact on the protein but of no known clinical significance. Considering only those variants with clear clinical impact, between 18% and 30% of patients had at least one diagnostic, prognostic, therapeutically relevant, or cancer predisposing mutation that was not covered. Interestingly, 2 of the 12 patients with relapsed or refractory cancers who received a genomics-specified therapy harbored variants that would not have been detected by these panels. These include patient SJMEL030083 with a metastatic melanoma harboring a *MAP3K8–GNG2* fusion (treated with the MEK inhibitor trametinib) and patient SJBT030076 whose clear

cell meningioma harbored an *EPS15L1–KLF17* fusion (treated with the EGFR inhibitor erlotinib; Supplementary Table S11).

DISCUSSION

Through the G4K study, we performed comprehensive WGS, WES, and transcriptome sequencing of tumor and/or paired normal samples from 309 prospective unselected children with cancer to quantify the prevalence and spectrum of genomic variants and evaluate the potential benefits of a three-platform sequencing approach. Several recent reports have focused on cohorts enriched in children with high-risk, relapsed, and refractory cancers (1, 3, 8, 57, 58, 60, 61). The G4K study intentionally differed from these others in that it was designed to investigate an unbiased cancer cohort, including patients with newly diagnosed standard-risk tumors, as well as those with more aggressive forms of disease, to specifically ask whether novel insights into the biology of cancer can emerge from comprehensive sequence analysis of all pediatric cancers.

Incorporation of whole-genome DNA sequencing enhanced the identification of clinically relevant lesions that are detected inefficiently or not at all by other genomic platforms. WGS detected inter- and intragenic SV breakpoints, which demarked the location of translocations and large deletions on whole chromosome scales. Translocations that activated oncogenes through enhancer hijacking with chromosomal breakpoints up to more than 750 kb away from the target oncogene were present in 8% of cases. On a smaller scale, WGS detected intragenic microdeletions affecting one or two exons in 15% of tumors, some of which may not be detectable by exome-based copy-number analysis, as shown in our previous study (14). Moreover, the passenger mutations in WGS identified the mutational processes supporting the role of germline lesions in MMR (*MSH2*, *PMS2*) and base excision repair genes (*MUTYH*) that contributed to the pathogenesis of several tumors. Taken together, up to one third of tumors in the G4K study were impacted by oncogenic events detected most sensitively by WGS. Recent results demonstrate that topologically associated chromatin domains (62) can be disrupted by focal copy-number variation (CNV) leading to dysregulated expression of oncogenes and tumor suppressors located in *cis*, often at long distances from the genomic event (63, 64). These data point toward an increasingly important role in the near future for WGS combined with transcriptome sequencing in the clinical evaluation of patient tumor genomes.

Eighty-six percent of patients had at least one finding that was diagnostic, prognostic, targetable, or indicating an underlying germline predisposition. One quarter of the tumors analyzed harbored a possible therapeutic target, excluding cases wherein MEK inhibitors would be an option. Benefit from MEK inhibitors, used downstream of mutant *KRAS*, *NRAS* and *NF1*, is more difficult to predict, but including MEK inhibitors, our potentially targetable group would comprise 43% of all tumors tested. While this proportion falls within the range of recent studies (4, 5, 8), it is notable given that the majority of patients on the G4K study had newly diagnosed standard-risk cancers, which might be expected to harbor fewer variants when compared to high-risk or relapsed/refractory cases. Potentially targetable abnormalities consisted mostly of JAK/STAT pathway mutations in B-ALL that can be treated with ruxolitinib or other JAK

inhibitors; high TMB samples that can be treated with pembroli-zumab, ipilimumab, or additional checkpoint inhibitors; and kinase fusions including *BRAF* in low-grade glioma, which may be treated with tyrosine kinase or MEK inhibitors. The propor-tion of patients who respond to these or other therapies remains to be determined and is an active area of investigation.

During the course of the G4K study, 78 patients whose tumors were sequenced presented with or developed meta-static, relapsed, or refractory disease with almost half of the tumors harboring potentially targetable lesions. Therapy was changed for 12 patients based on tumor genomic data, with 1 patient moving on to curative allogeneic stem cell transplanta-tion and 4 patients exhibiting prolonged disease stabilization. Recently the Zero Childhood Cancer Program reported on 38 patients receiving a genomics-specified targeted therapy, with 31% demonstrating clinical benefit (22), a proportion similar to ours. Nevertheless, despite favorable clinical outcomes in a limited number of patients, overall responses were modest across both of these studies. Together, these findings highlight that we still do not understand the full spectrum of genomic lesions that drive therapy response in most children with can-cer. These findings support our premise that the collection and interrogation of comprehensive genomic information are critical if we are to further push the boundaries of cure.

Revealing an inherited cancer predisposition can impact patient management by informing genetic counseling, ena-bling familial genetic testing, facilitating implementation of cancer surveillance and cancer risk–reducing measures, and directing cancer treatment. Our screening through G4K uncov-ered germline P/LP variants in 55 patients of whom almost two thirds would not have been detected on the basis of routine clinical indications for genetic testing. Importantly, simultane-ous assessment of tumor and germline genomes, including features such as TMB and mutation spectrum as well as allele-specific expression, improved variant classification in several cases. Indeed, we found corroborating evidence in the tumor that at least half of the germline P/LP variants we reported were likely to be relevant to the development of the child's tumor.

Gene panels designed to match patients with currently avail-able targeted therapies offer a rapid turnaround time, low expense, deep sequencing coverage, and sparing of tissue when limited biopsy material is available—all valuable characteristics in the clinical setting. Moreover, the trend toward adapting panels to capture cDNA addresses the need to detect fusion transcripts, allowing for the discovery of novel fusion partners of the more commonly involved oncogenes (65). Thus, for rou-tine standard of care, comprehensive gene panels address the medical need of the majority of patients. Nevertheless, there is much yet to be learned about childhood cancer. Among our set of 12 patients with relapsed or refractory cancers whose thera-pies were changed based on tumor genomic data, 2 received therapy as the result of novel alterations that would not have been detected by the any of the panels examined. Further-more, the relevance of a comprehensive view of the mutational landscape in informing understanding of tumorigenesis and management, even in the diagnostic setting, should not be underestimated. When evaluating the yield of gene panels in the G4K cohort, approximately 1 in 5 patients harbored a clinically relevant mutation that we conservatively estimated would have remained undetected using one of the panels examined. For the

G4K patients sequenced at diagnosis, it remains to be deter-mined what impact these events will have on overall outcomes.

Despite its potential benefits, comprehensive genomic pro-filing, such as that reported herein, is currently beyond the scope of most cancer clinical services for a variety of reasons. First, obtaining fresh-frozen tissue for three-platform sequenc-ing is not feasible for all patients. For the cancers where only limited biopsy is possible, alternative workflows involving formalin-fixed, paraffin-embedded samples may need to weigh the benefits of the breadth of WES and/or WGS against the coverage depth of a robust panel design. Second, the infra-structure and computational resources necessary to perform three-platform sequencing are substantial. Although recent innovations in cloud computing and reductions in the costs of NGS instruments and reagents are rapidly making whole-genome approaches more affordable, comprehensive genomic profiling will not be standard of care for some time to come. Third, turn-around time is a major challenge to a comprehen-sive genomics approach in the clinical setting. Throughout the G4K study, the entire workflow—from analyte preparation to data analysis, classification, and reporting—was under 7 weeks for 95% of the cases. Currently, turnaround times are under 6 weeks, and we are investigating improved laboratory and computational methods to reduce this further. Finally, as we recently reported (16), there was a significant overrepresenta-tion of African American patients who declined participation in the G4K study. While all 53 patients were later offered more focused germline testing, only 11 chose to pursue such test-ing. Thus, disparities exist in the context of cancer genomics and additional efforts are warranted to understand patients' perspectives and decision-making surrounding such testing to enhance future enrollment of diverse populations.

In summary, the G4K study provides evidence that three-platform sequencing of tumor and paired normal tissues gener-ates a more detailed picture of the genetic landscape of a tumor, at times revealing clinically relevant information that would go undetected if one used more targeted NGS approaches. As genomic sequencing technologies become less expensive and more widely available, their use will be an important adjunct to gene panels in the evaluation and management of children with newly diagnosed as well as relapsed or refractory cancers.

METHODS

Patient Eligibility and Accrual

This study was approved by an IRB (IRB# Pro00005011) and con-ducted in accordance with institutional and ethical guidelines. Over 20 months, 918 patients at St. Jude Children's Research Hospital were assessed for enrollment in the G4K study (NCT02530658). Enrollment criteria consisted of availability of a fresh-frozen tumor sample and a paired normal (i.e., germline) sample. Tumor purity was determined by a pathologist using visual inspection of a hematoxylin and eosin (H&E)–stained section of the tumor just adjacent to the portion sent for DNA and RNA extraction. For leukemia samples, tumor purity was determined based on a blast count determined by visual inspection of an H&E-stained bone marrow section or by flow cytometric analysis. A tumor purity >40% tumor was preferred for sequencing. Tumor purity was >40% in all but 16 cases, where it ranged from 23% to 37%. Never-theless, sequencing was successful in each of these cases.

A total of 365 patients were eligible to enroll. Patients and their par-ents met with a research nurse, trained by certified genetic counselors,

who introduced the study and answered questions. Patients were provided with written materials describing the study and a copy of the consent form to review. Interested patients were scheduled for an informed consent visit during which the research nurse reviewed key concepts, assessed parent and patient understanding, and obtained written informed consent. When possible, patients also met with an oncologist, clinical geneticist, or nurse practitioner to undergo collection of a full medical history and completion of a physical examination, and with a genetic counselor who obtained the family history and constructed a three-generation pedigree. A positive family history was defined as the presence of at least one first- or second-degree relative with cancer diagnosed before age 50, excluding cervical and non-melanoma skin cancers (see Supplementary Table S15). This same definition was used in our prior report of germline findings from the PCGP (9), and it is very similar to those in other recent reports (66–68).

Nucleic Acid Extraction and Sequencing

For solid tumors, sample adequacy and tumor cell percentage were assessed using an H&E-stained section from a block of tumor tissue adjacent to the one from which DNA and RNA were extracted. For leukemias, adequacy was based on bone marrow blast count. Tumor tissue was not available for many patients with retinoblastoma, craniopharyngioma, optic pathway glioma, and diffuse intrinsic pontine glioma due to safety concerns around biopsy (see Supplementary Table S1 for more details).

Nucleic acid extraction, library preparation, and sequencing on the Illumina HiSeq 2500/4000 were as described (14) but with a single modification. Specifically, in our previous clinical genomics pilot study, WES was performed using the TruSeq DNA LT Sample Prep kit (Illumina), whereas in the present study, it was completed using the Nextera Rapid Capture Exome Kit and the TruSeq Exome kits (both from Illumina). Samples were named according to the following convention: SJ (St. Jude), disease code (for example, RB; disease codes are listed in Supplementary Table S1), patient number, and sample type (D1 is first diagnostic sample, D2 is second diagnostic sample, R1 is first relapsed sample, G1 is first germline sample, etc.). Sequencing coverage and other statistics are listed in Supplementary Table S16.

Postsequencing quality control (QC) thresholds established during the pilot phase of our clinical service (14) were as follows: For WGS, ≥40% coverage of coding exons at 45×; however, approximately 80% coverage at 30× was acceptable upon the medical director's review. For WES, coverage of coding exons is ideally ≥65% at 45×, but approximately 80% coverage at 30× was acceptable upon the medical director's review. For RNA-seq, ideally ≥15% of coding exons were covered at 45×, but 20% at 30× was acceptable upon the medical director's review. Additional sequencing was performed in some samples to meet acceptable coverage QC thresholds.

Variant Classification and Reporting

WGS, WES, and RNA-seq data were aligned and variants called and annotated using previously published algorithms (19, 33, 69, 70) and automated pipeline infrastructure (14). Genome analysts reviewed variant calls and presented each case to a multidisciplinary tumor board consisting of representatives from Pathology, Computational Biology, and Oncology. Because our objective was to report SNVs/indels and CNAs from the somatic and germline context, as well as somatically acquired SVs, we adopted a unified reporting nomenclature that encompassed every variant type. As a convention, in referring to regional or focal copy-number changes, we use CNA in the somatic context and CNV in the germline context. We classified germline SNVs and indels according to the ACMG 2015 guidelines (44) and subsequently adopted the same nomenclature for other variant types. We reported germline P, LP, and VUS found among 63 high-risk cancer predisposition genes, including those recommended by ACMG v2.0 (71) and those for which clinical management recommendations exist

or are under development (44), as well as P/LP variants in 93 additional cancer predisposition genes. Germline CNVs were assessed for their functional impact on the gene of interest. If a CNV caused an LOF, we applied the PVS1 tag (72), and if it was rare or absent from databases such as the Database of Genomic Variants (DGV; http://dgv.tcag.ca/dgv/app/home), we also attached the PM2 tag, allowing us to classify these variants as P/LP in keeping with germline SNV/indel calls.

Somatic variants with a clear or likely impact on the function of a cancer-relevant gene were classified as P or LP, respectively, and those with an uncertain impact were classified as VUS. Somatic variants classified as VUS were not included in clinical tumor reports. When assessing the pathogenicity of somatic SNVs/indels, we considered multiple lines of evidence obtained using functional prediction algorithms and literature mining, as well as recurrence in the PCGP (20) and NCI-TARGET (15) databases, other pediatric cancer mutation data in the St. Jude PeCAN portal (73), and adult cancer data from COSMIC (74). Whenever possible, we used tumor RNA data to establish the functional impact for novel SVs by seeking evidence of truncation or in-frame gene fusion, and for gene amplification or enhancer hijacks, we used these data to examine gene expression.

We based our assessment criteria for clinical actionability on the system described by ACMG/American Society of Clinical Oncology (ASCO)/AMP (56, 75). For SNVs/indels, we called any variant with tier IA/B or tier IIC as actionable, as these comprised diagnostic, prognostic, and therapeutically relevant lesions with a high level of supporting evidence that an oncologist might reasonably act upon. Several lesions that did not meet this stringent threshold were mostly placed in tier IID and labeled as "potentially actionable" (Supplementary Table S2). For CNAs, we considered tier 1A/B and 2 variants as actionable (75) and applied these same assessment criteria to SVs.

Bioinformatics Analysis

Bioinformatics analysis and variant review were as described (14). Briefly, sequence reads were aligned using Burrows–Wheeler alignment (BWA; ref. 76) for DNA or BWA and Spliced Transcripts Alignment to a Reference (STAR) via our Strongarm pipeline (77) for RNA. Sequence variants were called using Bambino (69); DNA SVs using CREST (70); RNA SVs including ITD using CICERO (78), a local assembly-based algorithm that integrates RNA-seq read support with extensive annotation for candidate ranking (https://platform.stjude.cloud/workflows/rapid_rna-seq); and CNVs and allelic imbalance with CONSERTING (33). Variants were annotated using the Medal Ceremony/PeCanPIE pipeline (ref. 19; https://pecan.stjude.cloud/pie). We estimated TMB by counting all exonic and untranslated region (UTR) variants that passed manual review by a genome analyst in addition to high-quality intronic and noncoding somatic variants falling outside of repeat regions. For genome-wide calculations, we used only somatically acquired high-quality variants whose Fisher exact test P value for somatic origin was <0.05. To calculate mutations per Mb of DNA, we took all high-quality somatic variants from WGS and divided by 1445 Mb—the total amount of genome capable of generating a mutation call consisting of tier 1 (coding exons, splice regions and UTRs): 108,721,345 bp, tier 2 (potential regulatory regions): 163,138,648 bp, and tier 3 (intergenic and intronic regions): 1,173,180,850 bp. Tier 4 (repeat regions: 1,448,043,873) regions were omitted from the calculation, since our mutation calling pipeline masks these regions. For mutational burden in exome-only regions, we used 108.7 Mb as the denominator—the size of tier 1 regions listed above. B-ALL and T-ALL two-dimensional t-SNE distributions used data and methods described previously by Gu and colleagues (2019; ref. 42) and Liu and colleagues (2017; ref. 41). All additional analyses used standard parameters unless stated otherwise. Gene expression was quantified using the Gencode 75-gene model and HTSeq (version 0.11.1; ref. 79) and normalized between samples using the DESeq2 (version 1.26) variance stabilizing transformation function (80). SigProfilerSingleSample was used

to test for the presence and abundance of the COSMIC v3.1 Mutational Signatures (36) as described (45). Briefly, samples with 400 or more mutations (485 samples) were analyzed for 46 of the COSMIC signatures (excluding 23 COSMIC signatures that were rarely detected and manually found to be spurious in each positive case). Samples with fewer than 400 mutations (583 samples) were analyzed for a core set of 13 signatures (1, 2, 3, 5, 7a, 7b, 7c, 7d, 8, 13, 18, 36, and 40) that could be reliably detected in low mutation burden samples and are common in pediatric cancers (Supplementary Table S5).

Analysis of Putative Enhancer Hijacks

SJTALL030071 harbored a noncanonical interchromosomal translocation juxtaposing N-Me to the vicinity of *TLX3*, causing upregulation of the gene. To investigate further the possibility of activation of *TLX3* by somatic regulatory noncoding variants, we analyzed WGS and RNA-seq of this T-ALL to identify *cis*-activated genes that have outlier expression using cis-X (version 1.4.0), a computational method for discovering regulatory noncoding variants in cancer by integrating whole-genome and transcriptome sequencing data from a single cancer sample (81). cis-X first finds aberrantly *cis*-activated genes that exhibit allele-specific expression accompanied by an elevated outlier expression. For each gene, outlier high expression of a cancer sample of interest was determined by comparing its expression level to those of reference samples with the same tissue type. A null distribution of "leave-one-out (LOO)" t-statistic score was established using the reference samples. This was then used to determine the false discovery rate (FDR) of the LOO t-statistic score of a cancer sample of interest; those with an FDR<0.05 were retained as having significant outlier high expression. A minimum of 20 cases is required for this analysis.

Data used for reference expression matrix were obtained from publicly available data on St. Jude Cloud or NCI -TARGET to meet the minimum sample size criteria. For T-ALL samples such as SJTALL030071, RNA-seq data from 264 T-ALLs profiled by NCI-TARGET (41) were used as the reference expression matrix for evaluation of outlier expression status. For the remaining cancer subtypes that harbor candidate enhancer hijacking events (e.g., medulloblastoma, B-ALL, and AML), reference expression matrices were prepared by querying the relevant cancer diagnosis and selecting file type "Feature Counts" (precalculated using Gencode V31) using the Data Browser of Genomic Platform of the St. Jude Cloud (82). Feature counts were subsequently converted to reference FPKM matrix using DEseq2 for calculating outlier high expression status of samples presented in Supplementary Table S3. The AML reference expression matrix did not include therapy-related myeloid neoplasms. The number of samples used to construct the reference expression matrix for each cancer subtype is recorded in Supplementary Table S3.

For case SJTALL030071, we identified 18 genes that exhibit both allele-specific expression and outlier high expression surrounding *TLX3* (Supplementary Table S17). None of the genes other the *TLX3* play a known role in T-cell development or cancer. There were 29 consecutive SNPs with heterozygous genotypes in tumor DNA around the *TLX3* locus exhibiting monoallelic expression in tumor RNA (see Fig. 3C).

Comparison to Gene Panels

We selected four commercially available gene panels: FoundationOne CDx, FoundationOne Heme, Thermo Fisher Oncomine v3, and Children's Hospital Los Angeles OncoKids to represent the breadth and diversity of clinical gene panels. FoundationOne CDx is a large, general-use DNA-based panel; FoundationOne Heme is a large combined DNA/RNA panel focused on hematologic malignancies; Thermo Fisher Oncomine v3 is a smaller DNA/RNA panel focused on adult cancers; and OncoKids is a pediatric-specific version of the Oncomine platform used in support of NCI's pediatric MATCH trial (see Supplementary Table S12 footnotes for further details).

All four panels are capable of detecting SNVs/indels, gene fusions, and focal CNAs at the gene level. However, gross chromosomal changes

cannot be officially reported by these assays. In addition, complex copy-number abnormalities such as chromothripsis and regions of LOH are not reported. Of 253 G4K patients, 75 patients (29.6%), predominantly with hematologic malignancies, had diagnostic or prognostic gross chromosomal abnormalities including hyperdiploidy, hypodiploidy, and iAMP21. However, an additional karyotype, FISH, or microarray test could be run to detect these abnormalities. To make a direct comparison of reportable gene-level alterations, we omitted gross chromosomal gains and losses from the analysis.

We arrived at a conservative estimate of how many reported and clinically actionable genes are not represented on the exemplar panels as follows:

(i) We included both somatic and germline mutations in the comparison because our combined test reported both.
(ii) We used the coding sequence DNA portion of the panel assay for SNVs/indels/ITDs and focal CNAs. We assumed the observed alteration would have been captured and reported at the tumor VAF observed in the G4K patient. This is especially relevant for CNVs, as panels generally only report homozygous deletions or multiple copy gains.
(iii) We used the RNA panel gene lists combined with intronic sequences designed to capture recurrent translocation breakpoints for fusions, enhancer hijacks, and intragenic SVs.
(iv) We required only one of the partner genes to be present on the panel for gene fusions. This approach was permissive and enabled a conservative estimate of how many events might be detected by these technologies; we noted in our previous study that detection of gene fusions with low expression, such as *KMT2A* rearrangement and *KIAA1549–BRAF*, benefits from a multiplatform approach (14).
(v) For enhancer hijacks, we were unable to determine if the translocation breakpoints that we observed by WGS were featured on current panel designs. Furthermore, we did not assume outlier expression would be required on the panel data. As such, we took the same approach as for gene fusions, requiring only one of the partner loci being present on the panel design.

If a G4K variant was covered by a given panel, it was tallied as detected (Supplementary Table S12). We evaluated all variants within each patient, and if one variant was missed by a given panel, the patient was tallied as a miss by that panel (Supplementary Tables S13 and S14).

Visualization of Multi-omics Data

To examine the impact of somatic variants within tumors, we used visual exploration of aggregated multi-omics data from PCGP, NCI-TARGET, and other studies (15, 20, 41, 42) using GenomePaint (https://genomepaint.stjude.cloud/).

Data Availability

All raw and intermediate data are available as a free community resource hosted in St. Jude Cloud (https://www.stjude.cloud) under accession number SJC-1004. Study-level visualizations are available at https://pecan.stjude.cloud. Germline variants returned to patients during the study have been deposited in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/submitters/506672/).

Software/Code Availability

All software used in this study has been previously published; references are listed throughout the article or are available in St. Jude Cloud (https://www.stjude.cloud). Furthermore, our gene fusion, variant annotation, and mutational signature pipelines are available on St. Jude Cloud.

Authors' Disclosures

C.G. Mulligan reports personal fees from Illumina during the conduct of the study, as well as grants from Pfizer and AbbVie, and

Disclaimer

The opinions expressed in this article are the authors' own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Authors' Contributions

S. Newman: Resources, data curation, software, formal analysis, supervision, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing. **J. Nakitandwe:** Resources, data curation, formal analysis, supervision, validation, investigation, methodology, project administration, writing–review and editing. **C.A. Kesserwan:** Data curation, formal analysis, investigation, visualization, writing–original draft, writing–review and editing. **E.M. Azzato:** Data curation, formal analysis, supervision, validation, investigation. **D.A. Wheeler:** Data curation, software, formal analysis, supervision, validation, investigation, visualization, methodology, writing–original draft, writing–review and editing. **M. Rusch:** Conceptualization, resources, software, supervision, validation, methodology. **S. Shurtleff:** Conceptualization, resources, data curation, formal analysis, supervision, validation, investigation, methodology, project administration. **D.J. Hedges:** Data curation, validation. **K.V. Hamilton:** Resources, data curation, investigation. **S.G. Foy:** Data curation. **M.N. Edmonson:** Resources, data curation, software, validation. **A. Thrasher:** Resources, data curation, software, validation. **A. Bahrami:** Resources, data curation, formal analysis, validation, investigation. **B.A. Orr:** Resources, data curation, formal analysis, validation, investigation. **J.M. Klco:** Resources, data curation, formal analysis, validation, investigation. **J. Gu:** Resources, data curation, software. **L.W. Harrison:** Resources, data curation, project administration, writing–review and editing. **L. Wang:** Data curation, formal analysis, validation, investigation. **M.R. Clay:** Resources, data curation. **A. Ouma:** Resources, data curation, project administration. **A. Silkov:** Data curation. **Yanling Liu:** Data curation. **Z. Zhang:** Data curation, software, validation. **Yu Liu:** Data curation. **S.W. Brady:** Data curation, validation, visualization. **X. Zhou:** Resources, data curation, software, validation, investigation, visualization. **T.-C. Chang:** Data curation, software. **M. Pande:** Resources, data curation, software. **E. Davis:** Data curation, software. **J. Becksfort:** Data curation, software. **A. Patel:** Resources, data curation, software. **M.R. Wilkinson:** Resources, data curation, software. **D. Rahbarinia:** Resources, data curation, software. **M. Kubal:** Resources, data curation, project administration. **J.L. Maciaszek:** Data curation. **V. Pastor:** Software, visualization. **J. Knight:** Resources, data curation, software. **A.M. Gout:** Data curation, writing–review and editing. **J. Wang:** Resources, data curation, software, visualization. **Z. Gu:** Data curation, software, visualization. **C.G. Mullighan:** Resources, data curation, software, visualization. **R.B. McGee:** Resources, data curation, validation, investigation. **E.A. Quinn:** Resources, data curation, validation, investigation. **R. Nuccio:** Resources, data curation, validation, investigation. **R. Mostafavi:** Resources, data curation, validation, investigation. **E.L. Gerhardt:** Resources, data curation, validation, investigation. **L.M. Taylor:** Data curation, project administration. **J.M. Valdez:** Resources, data curation, validation, investigation. **S.J. Hines-Dowell:** Resources, data curation, validation, investigation, project administration. **A.S. Pappo:** Data curation, validation, investigation. **G. Robinson:** Data curation, validation, investigation. **L.-M. Johnson:** Conceptualization, resources, project administration. **C.-H. Pui:** Conceptualization, project administration. **D.W. Ellison:** Conceptualization, resources, data curation, supervision, funding acquisition, validation, investigation, methodology, project administration. **J.R. Downing:** Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, validation, investigation, methodology, project administration. **J. Zhang:** Conceptualization, resources, data curation, software, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing. **K.E. Nichols:** Conceptualization, resources, data curation, software, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing.

Acknowledgments

Note

Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

REFERENCES

1. Chang W, Brohl AS, Patidar R, Sindiri S, Shern JF, Wei JS, et al. Multi-Dimensional ClinOmics for precision therapy of children and adolescent young adults with relapsed and refractory cancer: a report from the center for cancer research. Clin Cancer Res 2016;22:3810–20.
2. Kline CN, Joseph NM, Grenert JP, van Ziffle J, Talevich E, Onodera C, et al. Targeted next-generation sequencing of pediatric neuro-oncology patients improves diagnosis, identifies pathogenic germline mutations, and directs targeted therapy. Neuro Oncol 2017;19:699–709.
3. Oberg JA, Glade Bender JL, Sulis ML, Pendrick D, Sireci AN, Hsiao SJ, et al. Implementation of next generation sequencing into pediatric hematology-oncology practice: moving beyond actionable alterations. Genome Med 2016;8:133.
4. Parsons DW, Roy A, Yang Y, Wang T, Scollon S, Bergstrom K, et al. Diagnostic yield of clinical tumor and germline whole-exome sequencing for children with solid tumors. JAMA Oncol 2016;2:616–24.
5. Surrey LF, MacFarland SP, Chang F, Cao K, Rathi KS, Akgumus GT, et al. Clinical utility of custom-designed NGS panel testing in pediatric tumors. Genome Med 2019;11:32.
6. Marks LJ, Oberg JA, Pendrick D, Sireci AN, Glasser C, Coval C, et al. Precision medicine in children and young adults with hematologic malignancies and blood disorders: The Columbia University Experience. Front Pediatr 2017;5:265.
7. Lin JJ, Riely GJ, Shaw AT. Targeting ALK: precision medicine takes on drug resistance. Cancer Discov 2017;7:137–55.
8. Mody RJ, Wu YM, Lonigro RJ, Cao X, Roychowdhury S, Vats P, et al. Integrative clinical sequencing in the management of refractory or relapsed cancer in youth. JAMA 2015;314:913–25.
9. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, et al. Germline mutations in predisposition genes in pediatric cancer. N Engl J Med 2015;373:2336–46.

10. Tran TH, Hunger SP. The genomic landscape of pediatric acute lymphoblastic leukemia and precision medicine opportunities. Semin Cancer Biol 2020 Nov 13 [Epub ahead of print].

11. Brok J, Lopez-Yurda M, Tinteren HV, Treger TD, Furtwangler R, Graf N, et al. Relapse of Wilms' tumour and detection methods: a retrospective analysis of the 2001 Renal Tumour Study Group-International Society of Paediatric Oncology Wilms' tumour protocol database. Lancet Oncol 2018;19:1072–81.

12. Pappo AS, Anderson JR, Crist WM, Wharam MD, Breitfeld PP, Hawkins D, et al. Survival after relapse in children and adolescents with rhabdomyosarcoma: a report from the Intergroup Rhabdomyosarcoma Study Group. J Clin Oncol 1999;17:3487–93.

13. Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. Nature 2017;548: 297–303.

14. Rusch M, Nakitandwe J, Shurtleff S, Newman S, Zhang Z, Edmonson MN, et al. Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome. Nat Commun 2018;9:3962.

15. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature 2018;555:371–6.

16. Howard Sharp KM, Jurbergs N, Ouma A, Harrison L, Gerhardt E, Taylor L, et al. Factors associated with declining to participate in a pediatric oncology next generation sequencing study. JCO Precis Oncol 2020;4:202–11.

17. Goschzik T, Gessi M, Dreschmann V, Gebhardt U, Wang L, Yamaguchi S, et al. Genomic alterations of adamantinomatous and papillary craniopharyngioma. J Neuropathol Exp Neurol 2017;76:126–34.

18. Alexander TB, Gu Z, Iacobucci I, Dickerson K, Choi JK, Xu B, et al. The genetic basis and cell of origin of mixed phenotype acute leukaemia. Nature 2018;562:373–9.

19. Edmonson MN, Patel AN, Hedges DJ, Wang Z, Rampersaud E, Kesserwan CA, et al. Pediatric Cancer Variant Pathogenicity Information Exchange (PeCanPIE): a cloud-based platform for curating and classifying germline variants. Genome Res 2019;29:1555–65.

20. Downing JR, Wilson RK, Zhang J, Mardis ER, Pui CH, Ding L, et al. The Pediatric Cancer Genome Project. Nat Genet 2012;44:619–22.

21. Roberts KG, Li Y, Payne-Turner D, Harvey RC, Yang YL, Pei D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. N Engl J Med 2014;371:1005–15.

22. Wong M, Mayoh C, Lau LMS, Khuong-Quang DA, Pinese M, Kumar A, et al. Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. Nat Med 2020;26:1742–53.

23. He B, Gao P, Ding YY, Chen CH, Chen G, Chen C, et al. Diverse noncoding mutations contribute to deregulation of cis-regulatory landscape in pediatric cancers. Sci Adv 2020;6:eaba3064.

24. Northcott PA, Lee C, Zichner T, Stutz AM, Erkek S, Kawauchi D, et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature 2014;511:428–34.

25. Zimmerman MW, Liu Y, He S, Durbin AD, Abraham BJ, Easton J, et al. MYC drives a subset of high-risk pediatric neuroblastomas and is activated through mechanisms including enhancer hijacking and focal enhancer amplification. Cancer Discov 2018;8:320–35.

26. Lehman NL, Usubalieva A, Lin T, Allen SJ, Tran QT, Mobley BC, et al. Genomic analysis demonstrates that histologically-defined astroblastomas are molecularly heterogeneous and that tumors with MN1 rearrangement exhibit the most favorable prognosis. Acta Neuropathol Commun 2019;7:42.

27. Haferlach C, Bacher U, Grossmann V, Schindela S, Zenger M, Kohlmann A, et al. Three novel cytogenetically cryptic EVI1 rearrangements associated with increased EVI1 expression and poor prognosis identified in 27 acute myeloid leukemia cases. Genes Chromosomes Cancer 2012;51:1079–85.

28. Paulsson K, Cazier JB, Macdougall F, Stevens J, Stasevich I, Vrcelj N, et al. Microdeletions are a general feature of adult and adolescent acute lymphoblastic leukemia: unexpected similarities with pediatric disease. Proc Natl Acad Sci U S A 2008;105:6708–13.

29. Waanders E, Scheijen B, van der Meer LT, van Reijmersdal SV, van Emst L, Kroeze Y, et al. The origin and nature of tightly clustered BTG1 deletions in precursor B-cell acute lymphoblastic leukemia support a model of multiclonal evolution. PLoS Genet 2012;8:e1002533.

30. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LA, Miller CB, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med 2009;360:470–80.

31. Engelhardt KR, Xu Y, Grainger A, Germani Batacchi MG, Swan DJ, Willet JD, et al. Identification of heterozygous single- and multi-exon deletions in IL7R by whole exome sequencing. J Clin Immunol 2017;37:42–50.

32. Marchuk DS, Crooks K, Strande N, Kaiser-Rogers K, Milko LV, Brandt A, et al. Increasing the diagnostic yield of exome sequencing by copy number variant analysis. PLoS One 2018;13:e0209185.

33. Chen X, Gupta P, Wang J, Nakitandwe J, Roberts K, Dalton JD, et al. CONSERTING: integrating copy-number analysis with structural-variation detection. Nat Methods 2015;12:527–30.

34. Phillips DH. Mutational spectra and mutational signatures: insights into cancer aetiology and mechanisms of DNA damage and repair. DNA Repair 2018;71:6–11.

35. Van Hoeck A, Tjoonk NH, van Boxtel R, Cuppen E. Portrait of a cancer: mutational signature analyses for cancer diagnostics. BMC Cancer 2019;19:457.

36. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578:94–101.

37. Li B, Brady SW, Ma X, Shen S, Zhang Y, Li Y, et al. Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. Blood 2020;135:41–55.

38. Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat Genet 2014;46:116–25.

39. Bernard OA, Busson-LeConiat M, Ballerini P, Mauchauffe M, Della Valle V, Monni R, et al. A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. Leukemia 2001;15:1495–504.

40. Herranz D, Ambesi-Impiombato A, Palomero T, Schnell SA, Belver L, Wendorff AA, et al. A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. Nat Med 2014;20:1130–7.

41. Liu Y, Easton J, Shao Y, Maciaszek J, Wang Z, Wilkinson MR, et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. Nat Genet 2017;49:1211–8.

42. Gu Z, Churchman ML, Roberts KG, Moore I, Zhou X, Nakitandwe J, et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. Nat Genet 2019;51:296–307.

43. Schwartzman O, Savino AM, Gombert M, Palmi C, Cario G, Schrappe M, et al. Suppressors and activators of JAK-STAT signaling at diagnosis and relapse of acute lymphoblastic leukemia in Down syndrome. Proc Natl Acad Sci U S A 2017;114:E4030–E9.

44. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–24.

45. Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, et al. Comprehensive analysis of hypermutation in human cancer. Cell 2017;171:1042–56.

46. Grobner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. Nature 2018;555:321–7.

47. Freitas AC, Opiniao A, Fragoso S, Nunes H, Santos M, Clara A, et al. Men seeking counselling in a Breast Cancer Risk Evaluation Clinic. Ecancermedicalscience 2018;12:804.

48. Giacomelli AO, Yang X, Lintner RE, McFarland JM, Duby M, Kim J, et al. Mutational processes shape the landscape of TP53 mutations in human cancer. Nat Genet 2018;50:1381–7.

49. Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, et al. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proc Natl Acad Sci U S A 2003;100:8424–9.

50. Bougeard G, Renaux-Petel M, Flaman JM, Charbonnier C, Fermey P, Belotti M, et al. Revisiting Li-Fraumeni syndrome from TP53 mutation carriers. J Clin Oncol 2015;33:2345–52.

51. Van Maerken T, Vandesompele J, Rihani A, De Paepe A, Speleman F. Escape from p53-mediated tumor surveillance in neuroblastoma: switching off the p14(ARF)-MDM2-p53 axis. Cell Death Differ 2009;16:1563–72.

52. Agaimy A, Foulkes WD. Hereditary SWI/SNF complex deficiency syndromes. Semin Diagn Pathol 2018;35:193–8.

53. Bellini A, Bessoltane-Bentahar N, Bhalshankar J, Clement N, Raynal V, Baulande S, et al. Study of chromatin remodeling genes implicates SMARCA4 as a putative player in oncogenesis in neuroblastoma. Int J Cancer 2019;145:2781–91.

54. Wang Z, Wilson CL, Easton J, Thrasher A, Mulder H, Liu Q, et al. Genetic risk for subsequent neoplasms among long-term survivors of childhood cancer. J Clin Oncol 2018;36:2078–87.

55. Joseph L, Cankovic M, Caughron S, Chandra P, Emmadi R, Hagenkord J, et al. The spectrum of clinical utilities in molecular pathology testing procedures for inherited conditions and cancer: a report of the Association for Molecular Pathology. J Mol Diagn 2016; 18:605–19.

56. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. J Mol Diagn 2017;19:4–23.

57. Pikman Y, Tasian SK, Sulis ML, Stevenson K, Blonquist TM, Apsel Winger B, et al. Matched targeted therapy for pediatric patients with relapsed, refractory, or high-risk leukemias: a report from the LEAP Consortium. Cancer Discov 2021;11:1424–39.

58. Worst BC, van Tilburg CM, Balasubramanian GP, Fiesel P, Witt R, Freitag A, et al. Next-generation personalised medicine for high-risk paediatric cancer patients - the INFORM pilot study. Eur J Cancer 2016;65:91–101.

59. Tasian SK, Loh ML, Hunger SP. Philadelphia chromosome-like acute lymphoblastic leukemia. Blood 2017;130:2064–72.

60. Allen CE, Laetsch TW, Mody R, Irwin MS, Lim MS, Adamson PC, et al. Target and agent prioritization for the Children's Oncology Group-National Cancer Institute Pediatric MATCH Trial. J Natl Cancer Inst 2017;109:djw274.

61. Harttrampf AC, Lacroix L, Deloger M, Deschamps F, Puget S, Auger N, et al. Molecular screening for cancer treatment optimization (MOSCATO-01) in pediatric patients: a single-institutional prospective molecular stratification trial. Clin Cancer Res 2017;23:6101–12.

62. Szabo Q, Bantignies F, Cavalli G. Principles of genome folding into topologically associating domains. Sci Adv 2019;5:eaaw1668.

63. Akdemir KC, Le VT, Chandran S, Li Y, Verhaak RG, Beroukhim R, et al. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. Nat Genet 2020;52:294–305.

64. Li L, Barth NKH, Pilarsky C, Taher L. Cancer is associated with alterations in the three-dimensional organization of the genome. Cancers 2019;11:1886.

65. Chmielecki J, Bailey M, He J, Elvin J, Vergilio JA, Ramkissoon S, et al. Genomic profiling of a large set of diverse pediatric cancers identifies known and novel mutations across tumor spectra. Cancer Res 2017;77:509–19.

66. Goudie C, Coltin H, Witkowski L, Mourad S, Malkin D, Foulkes WD. The McGill Interactive Pediatric OncoGenetic Guidelines: an approach to identifying pediatric oncology patients most likely to benefit from a genetic evaluation. Pediatr Blood Cancer 2017;64.

67. Hampel H, Bennett RL, Buchanan A, Pearlman R, Wiesner GL; Guideline Development Group, American College of Medical Genetics and Genomics Professional Practice and Guidelines Committee and National Society of Genetic Counselors Practice Guidelines Committee. A practice guideline from the American College of Medical Genetics and Genomics and the National Society of Genetic Counselors: referral indications for cancer predisposition assessment. Genet Med 2015;17:70–87.

68. Ripperger T, Bielack SS, Borkhardt A, Brecht IB, Burkhardt B, Calaminus G, et al. Childhood cancer predisposition syndromes-a concise review and recommendations by the Cancer Predisposition Working Group of the Society for Pediatric Oncology and Hematology. Am J Med Genet A 2017;173:1017–37.

69. Edmonson MN, Zhang J, Yan C, Finney RP, Meerzaman DM, Buetow KH. Bambino: a variant detector and alignment viewer for next-generation sequencing data in the SAM/BAM format. Bioinformatics 2011;27:865–6.

70. Wang J, Mullighan CG, Easton J, Roberts S, Heatley SL, Ma J, et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat Methods 2011;8:652–4.

71. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med 2017;19:249–55.

72. Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. Hum Mutat 2018;39:1517–24.

73. Zhou X, Edmonson MN, Wilkinson MR, Patel A, Wu G, Liu Y, et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. Nat Genet 2016;48:4–6.

74. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2011;39:D945–50.

75. Mikhail FM, Biegel JA, Cooley LD, Dubuc AM, Hirsch B, Horner VL, et al. Technical laboratory standards for interpretation and reporting of acquired copy-number abnormalities and copy-neutral loss of heterozygosity in neoplastic disorders: a joint consensus recommendation from the American College of Medical Genetics and Genomics (ACMG) and the Cancer Genomics Consortium (CGC). Genet Med 2019;21:1903–16.

76. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60.

77. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29: 15–21.

78. Tian L, Li Y, Edmonson MN, Zhou X, Newman S, McLeod C, et al. CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. Genome Biol 2020;21:126.

79. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166–9.

80. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106.

81. Liu Y, Li C, Shen S, Chen X, Szlachta K, Edmonson MN, et al. Discovery of regulatory noncoding variants in individual cancer genomes by using cis-X. Nat Genet 2020;52:811–8.

82. McLeod C, Gout AM, Zhou X, Thrasher A, Rahbarinia D, Brady SW, et al. St. Jude Cloud: a pediatric cancer genomic data-sharing ecosystem. Cancer Discov 2021;11:1082–99.