Data-driven approaches to advance research and clinical care for pediatric cancer

Alexander M. Gout, Sasi Arunachalam, David B. Finkelstein, Jinghui Zhang

PII: S0304-419X(21)00068-8

DOI: https://doi.org/10.1016/j.bbcan.2021.188571

Reference: BBACAN 188571

To appear in: *BBA - Reviews on Cancer*

Received date: 26 February 2021

Revised date: 7 May 2021

Accepted date: 22 May 2021

Please cite this article as: A.M. Gout, S. Arunachalam, D.B. Finkelstein, et al., Data-driven approaches to advance research and clinical care for pediatric cancer, *BBA - Reviews on Cancer* (2021), https://doi.org/10.1016/j.bbcan.2021.188571

Data-driven approaches to advance research and clinical care for pediatric cancer

Alexander M. Gout[1], Sasi Arunachalam[1], David B. Finkelstein[1], and Jinghui Zhang[1,1*]
jinghui.zhang@stjude.org
[1]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, 38105, Tennessee, USA
[*]Corresponding author.

Abstract

Pediatric cancer is a rare disease with a distinct etiology and mutational landscape compared with adult cancer. Multi-omics profiling of retrospective and prospective cohorts coupled with innovative computational analysis have been instrumental in uncovering mechanisms of tumorigenesis and drug resistance that are now informing pediatric cancer clinical therapy. In this review we present the major data resources of pediatric cancer and actionable insights into pediatric cancer etiology stemming from the identification of oncogenic gene fusions, mutational signature analysis, systems biology, cancer predisposition and survivorship studies - that have led to improved clinical diagnosis, discovery of new drug-targets, pharmacological therapy, and screening for genetic predisposition. Ultimately, integration of large-scale omics datasets generated through the international collaboration is required to maximize the power of data-driven approach to advance pediatric cancer research informing clinical therapy.

Keywords: Pediatric cancer, genomics, target discovery, data integration, data portals.

1. Introduction

At an annual incidence of ~16,000 in the US (https://www.acco.org/us-childhood-cancer-statistic), pediatric cancer is a disease comprised of many subtypes—more than 50% of which are rare cancers with an annual incidence of <200 cases based on the annual cancer diagnoses collected from NCI's Surveillance, Epidemiology and End Results (SEER) program (https://seer.cancer.gov). To improve the clinical outcome of pediatric cancer, a concerted effort across the world is required to share data generated from this disease. Historically, many

[1] Mailing Address: Department of Computational Biology, St. Jude Children's Research Hospital, 262 Danny Thomas Pl., Memphis, TN 38105, USA.

therapeutic agents used for treating childhood cancers were a re-application of adult cancer therapies, which may not be optimal given the physiologic and developmental differences between adults and children. Indeed, multi-omics landscape mapping efforts in recent years have unveiled distinct drivers for pediatric cancers (1,2), re-affirming that the etiology and tumorigenic mechanisms involved do not always align with those of adult cancers. It should also be noted that the current cure rate (80%) of childhood cancer is built on the use of cytotoxic chemotherapy and radiotherapies that are often associated with major side effects that can reduce the quality of life for survivors (3). Thus, there is a critical need for advancing the treatment of childhood cancer based on our evolving knowledge of molecular targets identified through multi-omics profiling of patient samples.

Pediatric cancers are thought to arise from developing tissues that undergo substantial expansion during early organ formation. This developmental origin has a profound effect on the mutational landscape which may require specialized computational tools to decode. For example, the paucity of observed somatic mutations and the critical role of fusion oncoproteins requires software tools designed for detecting variants at high sensitivity and for assessing variant pathogenicity given limited sample size. Therefore, establishing computational and data resources is essential to enable the iterative cycle of discovery on a research cohort, implementation of precision treatments prospectively in the clinic, study of the molecular features associated with clinical outcomes, and the investigation of new therapies to improve the quality of life for survivors (Figure 1). As genomics research leading to new insights on the etiology and treatment of pediatric cancer have already been reviewed thoroughly (4,5), we will discuss key resources involved in this iterative data-driven approach and highlight the impact of data-driven discovery on patient clinical care below.

2. Publicly available pediatric cancer genomic data resources

At the center of recent advances in our understanding of pediatric cancer tumorigenesis are a series of large-scale omics profiling studies using retrospective and prospective pediatric cancer patient samples (see summary in Table 1). The notable retrospective studies include the St. Jude Children's Research Hospital/Washington University- Pediatric Cancer Genome Project (PCGP)(6), the National Cancer Institute (NCI) Therapeutically Applicable Research to Generate Effective Treatments (TARGET) program, the International Cancer Genome

Consortium (ICGC), and the Gabriella Miller Kids First Data Resource Center (KFDRC) Initiative. Each of these retrospective studies involved whole-genome (WGS) and/or whole-exome sequencing (WES) and often transcriptome-sequencing (RNA-Seq). Some initiatives also involved micro-RNA sequencing (miRNA-Seq), bisulfite-sequencing or methylation array data.

PCGP generated comprehensive omics profiling for 15 pediatric cancer subtypes in leukemia (3 subtypes) (6-11), solid tumor (7 subtypes) (12-16) and brain tumor (5 subtypes) (17-20). TARGET (https://ocg.cancer.gov/programs/target) focuses on leukemia (3 subtypes) (21-30) and solid tumors (5 subtypes) (31-35) while the primary focus of ICGC is on brain tumors (36-44). Gabriella Miller KFDRC (https://commonfund.nih.gov/kidsfirst) focuses on familial leukemia, 3 subtypes of solid tumors and the major histological subtypes of brain tumors (45).The resulting data sets are publicly available on a variety of data portals: St Jude Cloud (https://stjude.cloud) for PCGP, NCI Genome Data Commons (https://gdc.cancer.gov) for TARGET, ICGC portal (https://dcc.icgc.org), and Gabriella Miller KFDRC portal (https://kidsfirstdrc.org). St. Jude Cloud also hosts germline genomes of long-term survivors from the St Jude Life (SJLIFE) cohort and Childhood Cancer Survivorship Study (CCSS) profiled by WGS and WES. A retrospective cohort used for a benchmark analysis of a 3-platform WGS, WES and RNA-Seq clinical sequencing pipeline is also accessible on St. Jude Cloud (46).

Several prospective clinical research studies have incorporated genome-wide sequencing (e.g. WGS, WES plus RNA-seq). The resulting data can initiate new iterations of research investigations in addition to their clinical utility providing molecular diagnostics, stratification for clinical trials, and/or providing avenues for personalized cancer therapy. Currently, publicly available genome-wide sequencing data sets include: 1) St Jude Genomes for Kids (G4K), a feasibility study assessing the utility of multiplatform genomic testing for precision oncology (https://clinicaltrials.gov/ct2/show/NCT02530658); 2) St Jude Real-time Clinical Genomics (47), an initiative where upon performing clinical sequencing on every eligible patient at St. Jude (to assist diagnostic and personalized therapy), data is publicly released through a comprehensive workflow involving verification of patient consent, de-identification, data harmonization, and quality checking – being the first instance of an institutional deposition of prospective clinical genomics data to the scientific research community as soon as possible; 3) Zero Childhood Cancer (ZERO) – a precision medicine program profiling children with poor-outcome, rare,

relapsed or refractory cancer from Children's Cancer Institute and Kids Cancer Centre, Sydney Children's Hospital Randwick, Australia (48); and 4) The Pediatric Brain Tumor Atlas (PBTA) representing the world's largest collection of childhood brain tumor data, available to assess in real-time via the Gabriella Miller Kids First Data Resource Portal.

Resources for accessing these large data sets generated from retrospective and prospective studies are summarized in Table 1 and further detail presented in Supplementary Tables 1 and 2. Furthermore, somatic variants from published pediatric genomic studies can be accessed on several data portals such as PeCan (https://pecan.stjude.cloud) (49), COSMIC (https://cancer.sanger.ac.uk) (50) and PedcBioPortal (https://pedcbioportal.kidsfirstdrc.org) (51,52), which are also valuable resources to the community. Further, the National Cancer Institute has developed the Pediatric Genomic Data Inventory (https://datascience.cancer.gov/resources/nci-data-catalog/pediatric-genomic-data-inventory), which documents genomic data sets generated from 50 studies. In addition to data derived from primary tumor samples, genomic data and associated compound or genetic screening data pertaining to pediatric cancer cell lines are also available. For instance, a total of 113 pediatric cancer cell lines are included as part of the Cancer Cell Line Encyclopedia (CCLE) (53) and a recent functional genomics screen of 82 cell lines performed by Dharia et al. (54) also represent valuable pediatric cancer genomic data resources.

3. Genomics analysis discover new therapeutic targets for pediatric cancer

Driver genetic aberrations identified by analyzing pediatric cancer genomic data have helped guide clinical trial design. An early example is neuroblastoma, the most common extracranial solid pediatric cancer originating from neural crest cells of the sympathetic nervous system. Anaplastic lymphoma kinase (*ALK*) genomic aberrations in neuroblastoma were reported in 2008 in a series of studies via genome-wide comparative genomic hybridization analysis on a large series of neuroblastomas (55), genome-wide scans using high-density SNP arrays on primary neuroblastoma samples (56) or familial neuroblastoma pedigrees (57,58). In each case, activating *ALK* hotspot mutations were identified, providing a novel therapeutic opportunity – leading to the Next Generation Personalized Neuroblastoma Therapy (NEPENTHE) clinical trial (https://clinicaltrials.gov/ct2/show/NCT02780128) for patients whose tumors involve mutations in

ALK. Here patients receive combination therapy of ribociclib, a dual inhibitor of cyclin-dependent kinase (CDK) and the ALK-inhibitor ceritinib – having demonstrated synergy against neuroblastoma (59). Additional studies by Maris and colleagues employing computational approaches have led to the identification of candidate immunotherapeutic targets such as CAMKV in *MYCN* amplified neuroblastoma (60) or GPC2 in high-risk neuroblastoma (61) from the analysis of NCI-TARGET neuroblastoma RNA-Seq data – where the potential of the latter as a candidate for CART immunotherapy being validated upon observed anti-tumor activity of CAR T-cells that target *GPC2* in vitro and mouse models (62).

The promise of immunotherapeutic approaches relies on harnessing the specific memory of the immune system to target malignant cell neoepitopes, enabling durable cures with minimal toxicity. However, given the paucity of somatic mutations in pediatric cancers it was initially unclear whether immunotherapy would represent a viable approach for treatment. To ascertain the neoepitope landscape in pediatric cancer, following the development of an analytical workflow, Chang et al. (63) defined the neoepitope landscape of somatic alterations comprised of missense mutations and oncogenic gene fusions among 540 childhood cancer genomes and transcriptomes revealing at least one predicted neoepitope in 88% of leukemias, 78% of central nervous system, and 90% of solid tumors – where a high proportion (69.6%) of neoepitopes were identified within *ETV6-RUNX1* in leukemias containing this fusion. Importantly, a subsequent study by Zamora et al. (64) reported the majority of predicted peptide neoepitopes in pediatric acute lymphoblastic leukemia (ALL) were recognized by patient T cells and induced functional responses *in vitro*. Further, *CBFB-MYH11* fusion neoepitopes, found in 12% of pediatric acute myeloid leukemia (AML) (29), enable T cell recognition and killing *in vitro* and *in vivo* in a patient-derived murine xenograft (65). Collectively, these studies demonstrate pediatric ALL and AML are not necessarily immunologically silent and necessitate further exploration of immunotherapy for targeting fusion positive leukemias.

Gene fusions are a major class of drug targets and are important biomarkers for defining subgroups of pediatric leukemia (e.g. *ETV6-RUX1* and *BCR-ABL1*) for risk stratification. Historically, gene fusions were characterized by cytogenetics or RT-PCR in a molecular pathology laboratory. However, since profiling of the entire transcriptome by RNA-Seq has become a standard assay for research and clinical applications, novel gene fusions can be discovered from RNA-Seq or a combination of RNA-Seq and WGS through the use of innovative computational methods such as CICERO (66), FusionCatcher (67), CREST (68), and

DELLY (69). For example, targetable fusions involving a diverse number of kinases (e.g. *ABL1, ABL2, CRLF2, CSF1R, EPOR, JAK2, PDGFRB, PTK2B, TSLP, or TYK2*) were detected by CICERO in high-risk acute lymphoblastic leukemia (ALL) samples exhibiting a gene expression signature reminiscent of BCR-ABL1 (Ph-like) (70,71), suggesting tyrosine kinase inhibitor therapy may be effective for these patients, as had previously been demonstrated in *BCR-ABL*-positive ALL patients (72). Indeed, this notion was further supported by leukemic cells and cell lines harboring these fusions, to exhibit respective sensitivity to ruxolitinib and dasatinib *in vitro* (70); and induction of remission of refractory *EBF1-PDGFRB* positive ALL following tyrosine kinase inhibitor therapy (73,74). Importantly, these findings contributed to the initiation of precision-medicine testing and treatment for Ph-like ALL in Children's Oncology Group ALL (https://childrensoncologygroup.org/aall1521) and St. Jude Total17 (https://clinicaltrials.gov/ct2/show/NCT03117751) clinical trials.

Targetable gene fusions have also been detected in multiple pediatric cancer types. For example, gene fusions involving neurotrophin receptor kinases (*NTRK1*, *NTRK2*, and *NTRK3*) have been found by PCGP and ICGC in low grade gliomas (17,44) and non-brainstem high-grade glioma (HGG) infant patients (18). Importantly, the subsequent report of the successful treatment of an *ETV6-NTRK3* positive HGG with Larotrectinib (75) paved the way for FDA approval of this pan-TRK inhibitor for the treatment of solid tumors with *NTRK* gene fusions (https://www.fda.gov/drugs/fda-approves-larotrectinib-solid-tumors-ntrk-gene-fusions). Further, upon the FDA approval of entrectinib (https://www.cancer.gov/news-events/cancer-currents-blog/2019/fda-entrectinib-ntrk-fusion) for the treatment of patients with solid tumors harboring a *NTRK1*, *2*, and *3* fusions, the results of the abovementioned studies enabled development of, in addition to the stratification of pediatric cancer patients into appropriate clinical trials such as the St. Jude STARTRK phase I/II clinical trial study (https://www.stjude.org/research/clinical-trials/startrk-study-of-entrectinib-in-children-with-brain-or-solid-tumors.html) of entrectinib. Given a *NTRK* fusion was also detected in leukemias by PCGP and other studies by RNA-seq (70,76), with demonstrated high sensitivity to TRK inhibition in mouse models (76,77), similar TRK-inhibitor therapeutic opportunities may exist for *NTRK*-fusion positive leukemias.

4. Mutational signature analysis unveils etiology of pediatric cancer initiation and relapse

In addition to driver gene discovery, analysis of mutational signatures has provided insight into the unique etiology of pediatric cancers. The mutational signatures, first reported by Alexandrov and colleagues (78-80), unveiled underlying mutational processes involved in tumorigenesis. More recently, analysis of WGS- and WES-identified somatic mutations in 31 adult and 1 pediatric cancers revealed a diverse set of over 100 mutational signatures (81), involving single-base-substitutions, doublet-base-substitutions, clustered-base-substitutions, and small insertion-and-deletion signatures. These were identified using nonnegative matrix factorization (NMF) approaches such as SigProfiler (81) and SignatureAnalyzer (a Bayesian variant of NMF) (82-84).

Mutational signature analysis employing these approaches identified a subset of the signatures in pediatric cancers, suggesting new avenues for cancer prevention or therapy. For example, COSMIC signature 18, associated with reactive oxygen species (ROS) exposure, was first discovered exclusively in neuroblastoma by Alexandrov et al. (78) in a study of 26 adult and 4 pediatric cancers. In a recent pan-neuroblastoma study Brady et al. (85) showed that signature 18 arises early in tumor initiation and is associated with increased expression of mitochondrial ribosome and electron transport chain (ETC)-associated genes, which may explain the connection of signature 18 and ROS. A similar upregulated mitochondrial gene expression pattern was observed in signature 18 positive pediatric rhabdomyosarcomas and points to future therapeutic avenues targeting altered mitochondrial function in these cancers. The ultraviolet-light (UV) exposure associated mutational signature was another exposure-related signature identified in pediatric cancers, in melanoma, as expected, and, intriguingly, in a subset of B-cell acute lymphoblastic leukemias (1,47,86). Here, further investigation may associate a novel pathway with presence of UV signature in B-ALL and thus potential therapeutic avenue.

Therapy-induced signatures in pediatric cancer were first reported in osteosarcoma and several brain and solid tumors treated with cisplatin (47,87). A recent study on relapsed pediatric ALLs identified two novel therapy-related signatures (COSMIC mutational signatures 86 and 87), one a result of thiopurine treatment which is employed during ALL maintenance therapy. These were present in 27% of patients and accounted for 46% of the acquired resistance mutations in *TP53*, *NR3C1*, *PRPS1*, and *NT5C2* (86), which may explain the reported increase in the reported risk of relapse with intensification of thiopurine maintenance therapy (88). Furthermore, therapy-induced mutations likely induce secondary malignancies in children (89) as demonstrated in a recent study which shows cisplatin- (signature 31 and 35) and thiopurine-signatures likely

caused *TP53* and Ras-pathway driver variants in pediatric therapy-related myeloid neoplasms following exposure to cytotoxic therapy (89). This suggests altering the dosage or timing of thiopurine or other cytotoxic treatments should be carefully considered to circumvent relapse or secondary cancers in these children.

5. Machine learning and systems biology approaches for improved clinical diagnosis and pharmacological therapy.

Machine learning and systems biology approaches have proved useful in overcoming some of the limitations common to rare diseases. Here, combining genomic, methylation, histological, or pharmacological data has led to refined tumor classification, identified target genes, and highlighted the drug-repurposing avenue for pediatric cancer therapy. For example, the use of a molecular classification approach incorporating a support vector machine classifier trained on DNA methylation array data (90) was used by Northcott et al. (36) for the verification of medulloblastoma diagnosis and subtype status. DNA methylation data has also been crucial for diagnosing other epigenetically modified rare central nervous tumors (CNS) (91) along with pediatric sarcomas (92). In a study by Capper et al. (91) in order to reduce histological variability and enhance the precision of CNS tumor diagnoses, random forest (RF) classifiers utilizing genomic scale DNA methylation array data were developed across all ages and tumor subtypes, computationally cross-validated and clinically assessed. Here, the successful application of RF and other machine learning workflows on DNA methylation array data (93) provides an attractive avenue for precision cancer diagnostics for pediatric CNS tumors. Importantly, availability of the resultant classifiers online (https://www.molecularneuropathology.org/mnp), significantly broadens the clinical impact of these methodologies for pediatric cancer patients worldwide.

Innovative systems biology methods have proved successful in identifying new avenues for cancer therapy as exemplified by a study concerning pediatric ALL (94). By combining pharmacological data and gene expression within a Bayesian network, they discovered an existing drug (dasatinib) could benefit 41% of children with T-cell acute lymphoblastic leukemia (T-ALL), providing a new therapeutic avenue for this hematologic cancer (94). A similarly advanced systems biology network approach has been applied to another significant clinical problem in T-ALL, namely glucocorticoid resistance. Here, a computationally inferred network of

master regulators was used to identify the AKT kinase activity as the major driver of glucocorticoid resistance in T-ALL (95). *In vitro* and *in vivo* inhibition of this gene was seen to effectively reverse glucocorticoid resistance (95), thus providing a new therapeutic avenue for these at-risk patients.

Systems biology network approaches have also shown utility in the context of pediatric solid tumors such as neuroblastoma. Novel targets for high-risk MYCN amplified neuroblastomas have been historically difficult to identify due to their lack of frequent somatic mutations and rarity. However, integration of data from a whole-genome shRNA library screen with a computational model of master regulator proteins revealed the transcription factor activating protein 4 (TFAP4) as a master regulator of MYCN-amplified neuroblastoma and is synthetically lethal, providing a novel target for this cancer (96). Additionally, regulatory network analysis identified subtype-specific master regulator proteins that were conserved across independent neuroblastoma cohorts. Using this approach, TEAD4 was uncovered and demonstrated to robustly predict poor survival, thereby suggesting a further novel therapeutic avenue for these tumors (97). Importantly, these studies demonstrate advanced machine learning approaches, when combined with genomic and other datasets, are able to overcome the limitations imposed by the rarity of pediatric cancer and are able to provide insight into clinical phenomenon otherwise intractable using conventional analytical approaches.

6. Pediatric cancer high throughput screening data providing avenues for clinical therapy

As is the case for the above mentioned genomic sequencing data, analysis of high-throughput screening data also provides a unique opportunity for novel therapeutic target identification. Functional genomics approaches such as the genome-scale CRISPR-Cas9 loss-of-function screen reported by Dharia et al. (54) provide critical insight into the genes required for pediatric cancer survival. Here, analysis of 82 pediatric cancer cell lines, representing 13 pediatric solid and brain tumor types, revealed a similar complexity of genetic dependencies to adult cancers. However, the vulnerabilities observed were often distinct from those in adult cancer suggesting adult oncology drug repurposing to unlikely be successful. Visualization tools for the data and associated analyses are available at the Cancer Dependency Map Portal (https://depmap.org and https://depmap.org/peddep).

High-throughput drug screening of pediatric cell line models also present an opportunity for unique therapeutic target identification. As mentioned above, the CCLE, comprising a large collection of gene expression, chromosomal copy number and high-throughput sequencing data for over 1000 human cancer cell lines, includes data for 113 pediatric cancer lines, providing opportunities for the identification of actionable drugs. This is highlighted in a CCLE study by Barretina et al. (53) which analyzed anticancer drug screening and sequencing data using computational approaches to correlate drug activity with genetic changes. In addition to reporting known interactions between gene mutations and drug sensitivities (e.g. *BRAF* mutations and *RAF* inhibitors), novel chemotherapeutic sensitivities were found including within Ewing sarcoma (tumor of the bone and soft tissue, primarily affecting adolescents and young adults) cell lines. Barretina et al. found Schlafen family member 11 (*SLFN11*) expression, a cell cycle control protein, was the top predictor of drug response across all cell lines where, interestingly, Ewing sarcoma lines exhibited the highest *SLFN11* expression. This suggests *SLFN11* expression could stratify Ewing sarcoma patients in clinical trials that use any conventional chemotherapy. The results of a second large-scale CCLE drug screening study by Garnett et al. (98), also highlighted Ewing sarcoma, reporting a high sensitivity to poly(ADP-ribose) polymerase (PARP) inhibitors - of consequence given 30% of Ewing sarcoma patients experience recurrent or metastatic disease and exhibit poor survival. Subsequently, Dyer and colleagues reported PARP inhibitor treatment of Ewing sarcoma cell lines induced 10-to 1,000-fold greater cytotoxicity following treatment with the DNA-damaging agents temozolomide or irinotecan, likely due in part to the observed defective DNA break repair processes (99). The study also shown an orthotopic Ewing sarcoma mouse model exhibiting a complete and durable response to combination therapy comprising the PARP inhibitor irinotecan with temozolomide. This therapeutic combination approach was later incorporated into the NCI-phase II trial studying how well irinotecan hydrochloride, temozolomide, and combination chemotherapy treats newly diagnosed Ewing sarcoma patients (https://www.cancer.gov/about-cancer/treatment/clinical-trials/search/v?id=NCI-2013-01094&r=1). Consequently, careful analysis of pediatric cell line drug screening data in conjunction with subsequent *in vitro* and *in vivo* studies has paved the way for exploration of combinatoric therapy for Ewing sarcoma patients. A further example of high-throughput screening of cell lines and xenografts leading to the identification of new therapeutic opportunities is presented by Vernooij et al. (100) who report the identification of idasanutlin as a resensitizing drug for venetoclax-resistant neuroblastoma.

7. Germline cancer predisposition for childhood cancer patients and long-term survivors

The study of germline cancer predisposition is critical for understanding the etiology of pediatric cancer. This is exemplified in the case of Wilms tumor, the most common renal malignancy in childhood where studies of familial cases via unbiased DNA sequencing of tumor genomes has identified a series of genes harboring germline mutations in this cancer such as *WT1* (101) *DICER1* (102), *PALB2 and CHEK2* (32). A copy number variation (CNV) genome-wide analysis by Egolf et al. (103) identified enrichment of a germline 550kb deletion at 16p11.2 within two independent neuroblastoma cohorts (total n=5,585). Interestingly, this CNV has been linked to other neurodevelopmental disorders and Egolf et al. suggest the inherent dysregulation of neurodevelopmental pathways likely result in neublastoma in addition to influencing other neurological phenotypes.

More broadly, the prevalence of germline mutations in cancer predisposition genes has been investigated by Zhang et al. (104) in a comprehensive study involving analyzing WGS and WES of 1,120 patients enrolled in PCGP. Here a comprehensive analysis pipeline involving a germline variant classifier, MedalCeremony (105), was developed, and they reported 8.5% germline mutational prevalence in cancer predisposition genes (104). More importantly, <50% of the patients had known family history, suggesting the need for screening germline cancer predisposition mutations for all pediatric cancer patients. Similarly, a pan-cancer analysis estimated a 6% prevalence in all childhood cancer patients after correction for cohort bias (2). Higher rates of pathogenic germline variants were reported within the: BASIC3 study of 150 children with CNS or non-CNS solid tumors (10% germline mutation prevalence) (106); Zero Childhood Cancer Program's cohort (n=252) of poor-outcome, rare, relapsed or refractory cancers (pathogenic cancer-predisposing variants identified in 16.2%) (48) and Memorial Sloan Kettering Cancer Center-integrated mutation profiling of actionable cancer targets (MSK-IMPACT) analysis of 751 pediatric solid tumor patients (pathogenic/likely pathogenic variants identified in 18% patients) (107). The differences in mutational prevalence reflect the differences in patient cohorts, gene lists as well as criteria for inclusion of mutations affecting genes associated with autosomal recessive cancer-predisposition syndromes.

The impact of germline mutations in cancer predisposition genes has been analyzed in long-term survivors of pediatric cancer. Wang et al reported 5.8% germline mutation prevalence in the study of 3,006 survivors enrolled in SJLIFE, a retrospective cohort with prospective clinical follow-up of childhood cancer survivors (108). Mutations were associated with secondary neoplasms such as breast cancer and sarcoma among irradiated survivors. A subsequent study which expands the cohort to include SJLIFE as well as CCSS (Childhood Cancer Survivor Study) identified *BRCA2* as a predisposition gene for pediatric or adolescent non–Hodgkin lymphoma (108). These findings highlight the importance of the knowledge on germline cancer predisposition in the clinical management of pediatric cancer patients, which may enable developing potentially lifesaving measures for cancer surveillance and prevention among survivors

8. Concluding Remarks

When compared to adults, pediatric cancer has an impressive cure rate of ~80% (109) in developed countries. However, as this rate is largely due to the success of therapies for acute lymphoblastic leukemias, much effort is still required for the identification of treatments for many subtypes of pediatric cancer, often having poor outcomes. Further, successful treatment of childhood cancers is largely dependent on the use of cytotoxic chemotherapy and radiotherapies, often associated with side effects that reduce the quality of life of survivors (3). It is widely believed targeting genetic alterations underpinning childhood cancer will facilitate the development of less toxic treatments. Central to this endeavor are large-scale pediatric data resources and innovative computational tools, when combined, have led to the discovery of new drug-targets, new clinical trials, a greater understanding of cancer predisposition, and insights into the genetic risk factors associated with secondary neoplasms in survivors. In addition to the abovementioned datasets, single-cell and circulating tumor DNA (ctDNA) data are becoming increasingly available and present new opportunities. Single-cell tumor gene expression data may enable deeper insight into the developmental origin of pediatric cancer (110,111) and clonal evolution/relapse, especially when compared to single-cell gene expression in normal tissues as represented within the Pediatric Cell Atlas (https://humancellatlas.org/pca) (112). Further, ctDNA screening data and development of associated analytical tools holds promise for the early detection and/or non-invasive diagnosis of new and relapsed pediatric cancer.

While adoption of a data-driven approach has shed light on underlying genomic aberrations and guided therapeutic development, challenges surrounding data sharing and integration remain. Data sharing ecosystems such as St. Jude Cloud (47) and Gabriella Miller KFDRC and initiatives such as the Childhood Cancer Data Initiative (CCDI) aim to address these challenges – the former by creating a data sharing and analysis ecosystem – and the latter by enabling a coordinated effort for the collection, analysis, and sharing of data via the establishment of a data ecosystem of connected repositories, registries, and tools for the research community. Data federation of large scale pediatric omic data resources (TARGET, ICGC, KFDRC, St. Jude Cloud, ZERO) in addition to integration with associated clinical phenotypic and patient information represent an additional exciting opportunity to leverage these large-scale datasets for further research and clinical benefit.

In summary, the activities surrounding the data-driven approach for pediatric cancer research and clinical care requires the concerted effort of dedicated teams of computational biologists, genomic experts, molecular and cellular biologists, pharmacologists, pathologists and oncologists to effectively translate these omic-derived findings into clinical care. This broad-based scientific collaboration offers an unprecedented opportunity to overcome the unique limitations associated with pediatric cancer research and improve clinical outcomes for these children.

9. Display Items

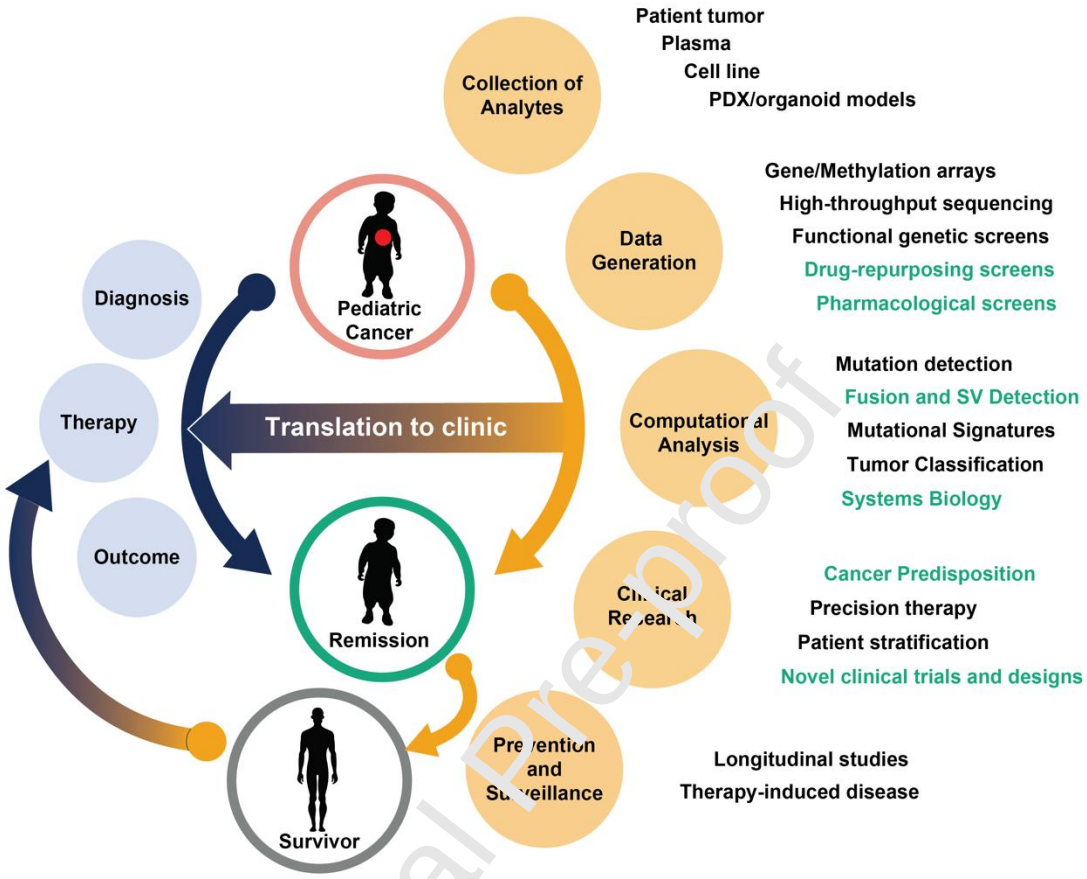

Figure 1. **Data-driven Approach for Pediatric Cancer Research and Clinical Care**. Following the collection and generation of patient tumor analytes, computational analysis of various data types in conjunction with clinical research activities directly impact the clinical management and care of pediatric cancer patients. 'Prevention and surveillance' of childhood cancer survivors: refers to activities concerned with the surveillance and identification of secondary neoplasms arising in survivors following 5+ years of remission; and also provides retrospective insights directly refining current treatment protocols. Features of this data-driven approach to cancer patient care specific to pediatrics are indicated in green font. Abbreviation: structural variant (SV).

Table 1. **Publicly available datasets generated from major pediatric cancer genomic studies.**

Data Access Portal	Pediatric Cohort	#Subtypes	WGS	WES	RNA-Seq	miRNA-Seq
NCI Genome Data Commons (https://gdc.cancer.gov)	NCI-TARGET[1]	8	1171 (2199)	1457 (2955)	1342 (1561)	2409 (2529)

ICGC data portal (https://dcc.icgc.org)	ICGC[1,*]	5	769	23	303	20
Kids First Data Resource Center Resource portal (https://kidsfirstdrc.org)	KFDRC[1,2,*]	13	1817 (2987)	170 (314)	1050 (1995)	21 (42)
St. Jude Cloud (https://www.stjude.cloud)	Cancer Patients[1,2]	27	1962 (3811)	2446 (4714)	2300 (2476)	
	Survivors[1]	26	7743 (7746)	3317 (3322)	0	
EGA (https://ega-archive.org)	ZERO[2]	23	252 (504)		228 (228)	

Patient cohorts are denoted as retrospective[1] or prospective[2].

*For ICGC and KFDRC, we only include data from pediatric cancer patients (aged 20 years or younger). Genomic profiling data is summarized as: #subjects (#samples).

Declaration of Competing Interest

None

Acknowledgments

We would like to thank Dr. Sam Brady, Ms. Delaram Rahbarinia, and Ms. Stephanie Wiggins for their help in data preparation, proof-reading and design of figure 1.

Supplementary data

Supplementary material

References

1. 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(7696):371-6 doi 10.1038/nature25795.
2. Grobner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA*, et al.* The landscape of genomic alterations across childhood cancers. Nature **2018**;555(7696):321-7 doi 10.1038/nature25480.
3. Hudson MM, Neglia JP, Woods WG, Sandlund JT, Pui CH, Kun LE*, et al.* Lessons from the past: opportunities to improve childhood cancer survivor care through outcomes investigations of historical therapeutic approaches for pediatric hematological malignancies. Pediatr Blood Cancer **2012**;58(3):334-43 doi 10.1002/pbc.23385.
4. Sweet-Cordero EA, Biegel JA. The genomic landscape of pediatric cancers: Implications for diagnosis and treatment. Science **2019**;363(6432):1170-5 doi 10.1126/science.aaw3535.

5. Jones DTW, Banito A, Grunewald TGP, Haber M, Jager N, Kool M*, et al.* Molecular characteristics and therapeutic vulnerabilities across paediatric solid tumours. Nat Rev Cancer **2019**;19(8):420-38 doi 10.1038/s41568-019-0169-x.
6. Downing JR, Wilson RK, Zhang J, Mardis ER, Pui CH, Ding L*, et al.* The Pediatric Cancer Genome Project. Nat Genet **2012**;44(6):619-22 doi 10.1038/ng.2287.
7. Andersson AK, Ma J, Wang J, Chen X, Gedman AL, Dang J*, et al.* The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. Nat Genet **2015**;47(4):330-7 doi 10.1038/ng.3230.
8. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D*, et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature **2012**;481(7380):157-63 doi 10.1038/nature10725.
9. Gruber TA, Larson Gedman A, Zhang J, Koss CS, Marada S, Ta HQ*, et al.* An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. Cancer Cell **2012**;22(5):683-97 doi 10.1016/j.ccr.2012.10.007.
10. Holmfeldt L, Wei L, Diaz-Flores E, Walsh M, Zhang J, Ding L*, et al.* The genomic landscape of hypodiploid acute lymphoblastic leukemia. Nat Genet **2013**;45(3):242-52 doi 10.1038/ng.2532.
11. Faber ZJ, Chen X, Gedman AL, Boggs K, Cheng J, Ma J*, et al.* The genomic landscape of core-binding factor acute myeloid leukemias. Nat Genet **2016**;48(12):1551-6 doi 10.1038/ng.3709.
12. Zhang J, Benavente CA, McEvoy J, Flores-Otero J, Ding L, Chen X*, et al.* A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature **2012**;481(7381):329-34 doi 10.1038/nature10733.
13. Chen X, Bahrami A, Pappo A, Easton J, Dalton J, Hedlund E*, et al.* Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. Cell Rep **2014**;7(1):104-12 doi 10.1016/j.celrep.2014.03.003.
14. Chen X, Stewart E, Shelat AA, Qu C, Bahrami A, Hatley M*, et al.* Targeting oxidative stress in embryonal rhabdomyosarcoma. Cancer Cell **2013**;24(6):710-24 doi 10.1016/j.ccr.2013.11.002.
15. Pinto EM, Chen X, Easton J, Finkelstein D, Liu Z, Pounds S*, et al.* Genomic landscape of paediatric adrenocortical tumours. Nat Commun **2015**;6:6302 doi 10.1038/ncomms7302.
16. Lu C, Zhang J, Nagahawatte P, Easton J, Lee S, Liu Z*, et al.* The genomic landscape of childhood and adolescent melanoma. J Invest Dermatol **2015**;135(3):816-23 doi 10.1038/jid.2014.425.
17. Zhang J, Wu G, Miller CP, Tatevossian RG, Dalton JD, Tang B*, et al.* Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. Nat Genet **2013**;45(6):602-12 doi 10.1038/ng.2611.
18. Wu G, Diaz AK, Paugh BS, Rankin SL, Ju B, Li Y*, et al.* The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. Nat Genet **2014**;46(5):444-50 doi 10.1038/ng.2938.

19. Robinson G, Parker M, Kranenburg TA, Lu C, Chen X, Ding L*, et al.* Novel mutations target distinct subgroups of medulloblastoma. Nature **2012**;488(7409):43-8 doi 10.1038/nature11213.
20. Parker M, Mohankumar KM, Punchihewa C, Weinlich R, Dalton JD, Li Y*, et al.* C11orf95-RELA fusions drive oncogenic NF-kappaB signalling in ependymoma. Nature **2014**;506(7489):451-5 doi 10.1038/nature13109.
21. 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(5):470-80 doi 10.1056/NEJMoa0808253.
22. Roberts KG, Morin RD, Zhang J, Hirst M, Zhao Y, Su X*, et al.* Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. Cancer Cell **2012**;22(2):153-66 doi 10.1016/j.ccr.2012.06.005.
23. Mullighan CG, Zhang J, Harvey RC, Collins-Underwood JR, Schulman BA, Phillips LA*, et al.* JAK mutations in high-risk childhood acute lymphoblastic leukemia. Proc Natl Acad Sci U S A **2009**;106(23):9414-8 doi 10.1073/pnas.0811761106.
24. Harvey RC, Mullighan CG, Wang X, Dobbin KK, Davidson GS, Bedrick EJ*, et al.* Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. Blood **2010**;116(23):4874-84 doi 10.1182/blood-2009-08-239681.
25. Harvey RC, Mullighan CG, Chen IM, Wharton W, Mikhail FM, Carroll AJ*, et al.* Rearrangement of CRLF2 is associated with mutation of JAK kinases, alteration of IKZF1, Hispanic/Latino ethnicity, and a poor outcome in pediatric B-progenitor acute lymphoblastic leukemia. Blood **2010**;115(26):5312-21 doi 10.1182/blood-2009-09-245944.
26. Zhang J, Mullighan CG, Harvey RC, Wu G, Chen X, Edmonson M*, et al.* Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. Blood **2011**;118(11):3080-7 doi 10.1182/blood-2011-03-341412.
27. Kang H, Chen IM, Wilson CS, Bedrick EJ, Harvey RC, Atlas SR*, et al.* Gene expression classifiers for relapse-free survival and minimal residual disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia. Blood **2010**;115(7):1394-405 doi 10.1182/blood-2009-05-218560.
28. Farrar JE, Schuback HL, Ries RE, Wai D, Hampton OA, Trevino LR*, et al.* Genomic Profiling of Pediatric Acute Myeloid Leukemia Reveals a Changing Mutational Landscape from Disease Diagnosis to Relapse. Cancer Res **2016**;76(8):2197-205 doi 10.1158/0008-5472.CAN-15-1015.
29. Bolouri H, Farrar JE, Triche T, Jr., Ries RE, Lim EL, Alonzo TA*, et al.* The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med **2018**;24(1):103-12 doi 10.1038/nm.4439.
30. Brunner AM, Graubert TA. Genomics in childhood acute myeloid leukemia comes of age. Nat Med **2018**;24(1):7-9 doi 10.1038/nm.4469.

31. Walz AL, Ooms A, Gadd S, Gerhard DS, Smith MA, Guidry Auvil JM*, et al.* Recurrent DGCR8, DROSHA, and SIX homeodomain mutations in favorable histology Wilms tumors. Cancer Cell **2015**;27(2):286-97 doi 10.1016/j.ccell.2015.01.003.
32. Gadd S, Huff V, Walz AL, Ooms A, Armstrong AE, Gerhard DS*, et al.* A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. Nat Genet **2017**;49(10):1487-94 doi 10.1038/ng.3940.
33. Gooskens SL, Gadd S, Guidry Auvil JM, Gerhard DS, Khan J, Patidar R*, et al.* TCF21 hypermethylation in genetically quiescent clear cell sarcoma of the kidney. Oncotarget **2015**;6(18):15828-41 doi 10.18632/oncotarget.4682.
34. Oldridge DA, Wood AC, Weichert-Leahey N, Crimmins I, Sussman R, Winter C*, et al.* Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. Nature **2015**;528(7582):418-21 doi 10.1038/nature15540.
35. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D*, et al.* The genetic landscape of high-risk neuroblastoma. Nat Genet **2013**;45(3):279-84 doi 10.1038/ng.2529.
36. Northcott PA, Buchhalter I, Morrissy AS, Hovestadt V, Weischenfeldt J, Ehrenberger T*, et al.* The whole-genome landscape of medulloblastoma subtypes. Nature **2017**;547(7663):311-7 doi 10.1038/nature22973.
37. Hovestadt V, Jones DT, Picelli S, Wang W, Kool M, Northcott PA*, et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. Nature **2014**;510(7506):537-41 doi 10.1038/nature13268.
38. Kool M, Jones DT, Jager N, Northcott PA, Pugh TJ, Hovestadt V*, et al.* Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. Cancer Cell **2014**;25(3):393-405 doi 10.1016/j.ccr.2014.02.004.
39. 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(7510):428-34 doi 10.1038/nature13379.
40. Kool M, Korshunov A, Remke M, Jones DT, Schlanstein M, Northcott PA*, et al.* Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. Acta Neuropathol **2012**;123(4):473-84 doi 10.1007/s00401-012-0958-8.
41. Jones DT, Jager N, Kool M, Zichner T, Hutter B, Sultan M*, et al.* Dissecting the genomic complexity underlying medulloblastoma. Nature **2012**;488(7409):100-5 doi 10.1038/nature11284.
42. Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J*, et al.* Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature **2012**;488(7409):106-10 doi 10.1038/nature11329.
43. International Cancer Genome Consortium PedBrain Tumor P. Recurrent MET fusion genes represent a drug target in pediatric glioblastoma. Nat Med **2016**;22(11):1314-20 doi 10.1038/nm.4204.
44. Jones DT, Hutter B, Jager N, Korshunov A, Kool M, Warnatz HJ*, et al.* Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. Nat Genet **2013**;45(8):927-32 doi 10.1038/ng.2682.

45. Petralia F, Tignor N, Reva B, Koptyra M, Chowdhury S, Rykunov D*, et al.* Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer. Cell **2020**;183(7):1962-85 e31 doi 10.1016/j.cell.2020.10.044.
46. 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(1):3962 doi 10.1038/s41467-018-06485-7.
47. 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** doi 10.1158/2159-8290.CD-20-1230.
48. Wong M, Mayoh C, Lau LMS, Khuong-Quang DA, Pinese M, Kumar A*, et al.* Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. Nat Med **2020**;26(11):1742-53 doi 10.1038/s41591-020-1072-4.
49. 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(1):4-6 doi 10.1038/ng.3466.
50. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N*, et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res **2019**;47(D1):D941-D7 doi 10.1093/nar/gky1015.
51. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA*, et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov **2012**;2(5):401-4 doi 10.1158/2159-8290.CD-12-0095.
52. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO*, et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal **2013**;6(269):pl1 doi 10.1126/scisignal.2004088.
53. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S*, et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature **2012**;483(7391):603-7 doi 10.1038/nature11003.
54. Dharia NV, Kugener G, Guenther LM, Malone CF, Durbin AD, Hong AL*, et al.* A first-generation pediatric cancer dependency map. Nat Genet **2021**;53(4):529-38 doi 10.1038/s41588-021-00819-w.
55. Janoueix-Lerosey I, Lequin D, Brugieres L, Ribeiro A, de Pontual L, Combaret V*, et al.* Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. Nature **2008**;455(7215):967-70 doi 10.1038/nature07398.
56. Chen Y, Takita J, Choi YL, Kato M, Ohira M, Sanada M*, et al.* Oncogenic mutations of ALK kinase in neuroblastoma. Nature **2008**;455(7215):971-4 doi 10.1038/nature07399.
57. Mosse YP, Laudenslager M, Longo L, Cole KA, Wood A, Attiyeh EF*, et al.* Identification of ALK as a major familial neuroblastoma predisposition gene. Nature **2008**;455(7215):930-5 doi 10.1038/nature07261.
58. George RE, Sanda T, Hanna M, Frohling S, Luther W, 2nd, Zhang J*, et al.* Activating mutations in ALK provide a therapeutic target in neuroblastoma. Nature **2008**;455(7215):975-8 doi 10.1038/nature07397.

59. Wood AC, Krytska K, Ryles HT, Infarinato NR, Sano R, Hansel TD*, et al.* Dual ALK and CDK4/6 Inhibition Demonstrates Synergy against Neuroblastoma. Clin Cancer Res **2017**;23(11):2856-68 doi 10.1158/1078-0432.CCR-16-1114.
60. Sussman RT, Rokita JL, Huang K, Raman P, Rathi KS, Martinez D*, et al.* CAMKV Is a Candidate Immunotherapeutic Target in MYCN Amplified Neuroblastoma. Front Oncol **2020**;10:302 doi 10.3389/fonc.2020.00302.
61. Bosse KR, Raman P, Zhu Z, Lane M, Martinez D, Heitzeneder S*, et al.* Identification of GPC2 as an Oncoprotein and Candidate Immunotherapeutic Target in High-Risk Neuroblastoma. Cancer Cell **2017**;32(3):295-309 e12 doi 10.1016/j.ccell.2017.08.003.
62. Li N, Fu H, Hewitt SM, Dimitrov DS, Ho M. Therapeutically targeting glypican-2 via single-domain antibody-based chimeric antigen receptors and immunotoxins in neuroblastoma. Proc Natl Acad Sci U S A **2017**;114(32):E6623-E31 doi 10.1073/pnas.1706055114.
63. Chang TC, Carter RA, Li Y, Li Y, Wang H, Edmonson MN*, et al.* The neoepitope landscape in pediatric cancers. Genome Med **2017**;9(1):78 doi 10.1186/s13073-017-0468-3.
64. Zamora AE, Crawford JC, Allen EK, Guo XJ, Bakke J, Carter RA*, et al.* Pediatric patients with acute lymphoblastic leukemia generate abundant and functional neoantigen-specific CD8(+) T cell responses. Sci Transl Med **2019**;11(498) doi 10.1126/scitranslmed.aat8549.
65. Biernacki MA, Foster KA, Woodward KB, Coon ME, Cummings C, Cunningham TM*, et al.* CBFB-MYH11 fusion neoantigen enables T cell recognition and killing of acute myeloid leukemia. J Clin Invest **2020**;130(10):5127-41 doi 10.1172/JCI137723.
66. 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(1):126 doi 10.1186/s13059-020-02043-x.
67. Daniel Nicorici MŞ, Henrik Edgren, Sara Kangaspeska, Astrid Murumägi, Olli Kallioniemi, Sami Virtanen, Olavi Kilkku. FusionCatcher – a tool for finding somatic fusion genes in paired-end RNA-sequencing data. bioRxiv doi https://doi.org/10.1101/011650.
68. 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(8):652-4 doi 10.1038/nmeth.1628.
69. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics **2012**;28(18):i333-i9 doi 10.1093/bioinformatics/bts378.
70. 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(11):1005-15 doi 10.1056/NEJMoa1403088.
71. Reshmi SC, Harvey RC, Roberts KG, Stonerock E, Smith A, Jenkins H*, et al.* Targetable kinase gene fusions in high-risk B-ALL: a study from the Children's Oncology Group. Blood **2017**;129(25):3352-61 doi 10.1182/blood-2016-12-758979.
72. Schultz KR, Bowman WP, Aledo A, Slayton WB, Sather H, Devidas M*, et al.* Improved early event-free survival with imatinib in Philadelphia chromosome-positive acute lymphoblastic leukemia: a children's oncology group study. J Clin Oncol **2009**;27(31):5175-81 doi 10.1200/JCO.2008.21.2514.

73. Weston BW, Hayden MA, Roberts KG, Bowyer S, Hsu J, Fedoriw G*, et al.* Tyrosine kinase inhibitor therapy induces remission in a patient with refractory EBF1-PDGFRB-positive acute lymphoblastic leukemia. J Clin Oncol **2013**;31(25):e413-6 doi 10.1200/JCO.2012.47.6770.
74. Lengline E, Beldjord K, Dombret H, Soulier J, Boissel N, Clappier E. Successful tyrosine kinase inhibitor therapy in a refractory B-cell precursor acute lymphoblastic leukemia with EBF1-PDGFRB fusion. Haematologica **2013**;98(11):e146-8 doi 10.3324/haematol.2013.095372.
75. Ziegler DS, Wong M, Mayoh C, Kumar A, Tsoli M, Mould E*, et al.* Brief Report: Potent clinical and radiological response to larotrectinib in TRK fusion-driven high-grade glioma. Br J Cancer **2018**;119(6):693-6 doi 10.1038/s41416-018-0251-2.
76. Taylor J, Pavlick D, Yoshimi A, Marcelus C, Chung SS, Hechtman JF*, et al.* Oncogenic TRK fusions are amenable to inhibition in hematologic malignancies. J Clin Invest **2018**;128(9):3819-25 doi 10.1172/JCI120787.
77. Roberts KG, Janke LJ, Zhao Y, Seth A, Ma J, Finkelstein D*, et al.* ETV6-NTRK3 induces aggressive acute lymphoblastic leukemia highly sensitive to selective TRK inhibition. Blood **2018**;132(8):861-5 doi 10.1182/blood-2018-05-849554.
78. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV*, et al.* Signatures of mutational processes in human cancer. Nature **2013**;500(7463):415-21 doi 10.1038/nature12477.
79. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep **2013**;3(1):246-59 doi 10.1016/j.celrep.2012.12.008.
80. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K*, et al.* Mutational processes molding the genomes of 21 breast cancers. Cell **2012**;149(5):979-93 doi 10.1016/j.cell.2012.04.024.
81. 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(7793):94-101 doi 10.1038/s41586-020-1943-3.
82. Kasar S, Kim J, Improgo R, Tiao G, Polak P, Haradhvala N*, et al.* Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. Nat Commun **2015**;6:8866 doi 10.1038/ncomms9866.
83. Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Kwiatkowski DJ*, et al.* Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. Nat Genet **2016**;48(6):600-6 doi 10.1038/ng.3557.
84. Haradhvala NJ, Kim J, Maruvka YE, Polak P, Rosebrock D, Livitz D*, et al.* Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. Nat Commun **2018**;9(1):1746 doi 10.1038/s41467-018-04002-4.
85. Brady SW, Liu Y, Ma X, Gout AM, Hagiwara K, Zhou X*, et al.* Pan-neuroblastoma analysis reveals age- and signature-associated driver alterations. Nat Commun **2020**;11(1):5183 doi 10.1038/s41467-020-18987-4.

86. 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(1):41-55 doi 10.1182/blood.2019002220.
87. Brady SW, Ma X, Bahrami A, Satas G, Wu G, Newman S*, et al.* The Clonal Evolution of Metastatic Osteosarcoma as Shaped by Cisplatin Treatment. Mol Cancer Res **2019**;17(4):895-906 doi 10.1158/1541-7786.MCR-18-0620.
88. Schmiegelow K, Bjork O, Glomstein A, Gustafsson G, Keiding N, Kristinsson J*, et al.* Intensification of mercaptopurine/methotrexate maintenance chemotherapy may increase the risk of relapse for some children with acute lymphoblastic leukemia. J Clin Oncol **2003**;21(7):1332-9 doi 10.1200/JCO.2003.04.039.
89. Schwartz JR, Ma J, Kamens J, Westover T, Walsh MP, Brady SW*, et al.* The acquisition of molecular drivers in pediatric therapy-related myeloid neoplasms. Nat Commun **2021**;12(1):985 doi 10.1038/s41467-021-21255-8.
90. Hovestadt V, Remke M, Kool M, Pietsch T, Northcott PA, Fischer R*, et al.* Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. Acta Neuropathol **2013**;125(6):913-6 doi 10.1007/s00401-013-1126-5.
91. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D*, et al.* DNA methylation-based classification of central nervous system tumours. Nature **2018**;555(7697):469-74 doi 10.1038/nature26000.
92. Koelsche C, Schrimpf D, Stichel D, Sill M, Sahm F, Reuss DE*, et al.* Sarcoma classification by DNA methylation profiling. Nat Commun **2021**;12(1):498 doi 10.1038/s41467-020-20603-4.
93. Maros ME, Capper D, Jones DTW, Hovestadt V, von Deimling A, Pfister SM*, et al.* Machine learning workflows to estimate class probabilities for precision cancer diagnostics on DNA methylation microarray data. Nat Protoc **2020**;15(2):479-512 doi 10.1038/s41596-019-0251-6.
94. Y Gocho LJ, Hu J, Yang JJ, Dharia NV, Zhang J, et. al. Network-based systems pharmacology reveals heterogeneity in LCK and BCL2 signaling and therapeutic sensitivity of T-cell acute lymphoblastic leukemia. Nature Cancer **2021** doi 10.1038/s43018-020-00167-4.
95. Piovan E, Yu J, Tosello V, Herranz D, Ambesi-Impiombato A, Da Silva AC*, et al.* Direct reversal of glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. Cancer Cell **2013**;24(6):766-76 doi 10.1016/j.ccr.2013.10.022.
96. Boboila S, Lopez G, Yu J, Banerjee D, Kadenhe-Chiweshe A, Connolly EP*, et al.* Transcription factor activating protein 4 is synthetically lethal and a master regulator of MYCN-amplified neuroblastoma. Oncogene **2018**;37(40):5451-65 doi 10.1038/s41388-018-0326-9.
97. Rajbhandari P, Lopez G, Capdevila C, Salvatori B, Yu J, Rodriguez-Barrueco R*, et al.* Cross-Cohort Analysis Identifies a TEAD4-MYCN Positive Feedback Loop as the Core Regulatory Element of High-Risk Neuroblastoma. Cancer Discov **2018**;8(5):582-99 doi 10.1158/2159-8290.CD-16-0861.

98. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW*, et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature **2012**;483(7391):570-5 doi 10.1038/nature11005.
99. Stewart E, Goshorn R, Bradley C, Griffiths LM, Benavente C, Twarog NR*, et al.* Targeting the DNA repair pathway in Ewing sarcoma. Cell Rep **2014**;9(3):829-41 doi 10.1016/j.celrep.2014.09.028.
100. Vernooij L, Bate-Eya LT, Alles LK, Lee JY, Koopmans B, Jonus HC*, et al.* High-throughput screening identifies idasanutlin as a resensitizing drug for venetoclax-resistant neuroblastoma cells. Mol Cancer Ther **2021** doi 10.1158/1535-7163.MCT-20-0666.
101. Huff V. Wilms tumor genetics. Am J Med Genet **1998**;79(4):260-7 doi 10.1002/(sici)1096-8628(19981002)79:4<260::aid-ajmg6>3.0.co;2-q.
102. Palculict TB, Ruteshouser EC, Fan Y, Wang W, Strong L, Huff V. Identification of germline DICER1 mutations and loss of heterozygosity in familial Wilms tumour. J Med Genet **2016**;53(6):385-8 doi 10.1136/jmedgenet-2015-103311.
103. Egolf LE, Vaksman Z, Lopez G, Rokita JL, Modi A, Basta PV*, et al.* Germline 16p11.2 Microdeletion Predisposes to Neuroblastoma. Am J Hum Genet **2019**;105(3):658-68 doi 10.1016/j.ajhg.2019.07.020.
104. 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(24):2336-46 doi 10.1056/NEJMoa1508054.
105. 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(9):1555-65 doi 10.1101/gr.250357.119.
106. 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(5):616-24 doi 10.1001/jamaoncol.2015.5699.
107. Fiala EM, Jayakumaran, G., Mauguen, A. et al. . Prospective pan-cancer germline testing using MSK-IMPACT informs clinical translation in 751 patients with pediatric solid tumors. Nature Cancer **2021**;2():357–65 doi https://doi.org/10.1038/s43018-021-00172-1.
108. 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(20):2078-87 doi 10.1200/JCO.2018.77.8589.
109. Pui CH, Gajjar AJ, Kane JR, Qaddoumi IA, Pappo AS. Challenging issues in pediatric oncology. Nat Rev Clin Oncol **2011**;8(9):540-9 doi 10.1038/nrclinonc.2011.95.
110. Hovestadt V, Smith KS, Bihannic L, Filbin MG, Shaw ML, Baumgartner A*, et al.* Resolving medulloblastoma cellular architecture by single-cell genomics. Nature **2019**;572(7767):74-9 doi 10.1038/s41586-019-1434-6.
111. Filbin MG, Tirosh I, Hovestadt V, Shaw ML, Escalante LE, Mathewson ND*, et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. Science **2018**;360(6386):331-5 doi 10.1126/science.aao4750.

112. Taylor DM, Aronow BJ, Tan K, Bernt K, Salomonis N, Greene CS*, et al.* The Pediatric Cell Atlas: Defining the Growth Phase of Human Development at Single-Cell Resolution. Dev Cell **2019**;49(1):10-29 doi 10.1016/j.devcel.2019.03.001.