special articles

Cancer Informatics for Cancer Centers: Scientific Drivers for Informatics, Data Science, and Care in Pediatric, Adolescent, and Young Adult Cancer

Anthony R. Kerlavage, PhD[1]; Anne C. Kirchhoff, PhD, MPH[2]; Jaime M. Guidry Auvil, PhD[1]; Norman E. Sharpless, MD[3]; Kara L. Davis, DO[4]; Karlyne Reilly, PhD[5]; Gregory Reaman, MD[6]; Lynne Penberthy, MD, MPH[7]; Dennis Deapen, DrPh[8]; Amie Hwang, PhD[8]; Eric B. Durbin, DrPH, MS[9]; Sara L. Gallotto, MS[10]; Richard Aplenc, MD, PhD, MSCE[11]; Samuel L. Volchenboum, MD, PhD[12]; Allison P. Heath, PhD[11]; Bruce J. Aronow, PhD[13]; Jinghui Zhang, PhD[14]; Olena Vaske, PhD[15]; Todd A. Alonzo, PhD[16]; Paul C. Nathan, MD, MSc[17]; Jenny N. Poynter, PhD, MPH[18]; Greg Armstrong, MD, MSCE[14]; Erin E. Hahn, PhD, MPH[19]; Karen J. Wernli, MS, PhD[20]; Casey Greene, PhD[21]; Jack DiGiovanna, PhD[22]; Adam C. Resnick, PhD[11]; Eve R. Shalley, BA[23]; Sorena Nadaf, MS, MMI[24]; and Warren A. Kibbe, PhD[25]

abstract

Cancer Informatics for Cancer Centers (CI4CC) is a grassroots, nonprofit 501c3 organization intended to provide a focused national forum for engagement of senior cancer informatics leaders, primarily aimed at academic cancer centers anywhere in the world but with a special emphasis on the 70 National Cancer Institute–funded cancer centers. This consortium has regularly held topic-focused biannual face-to-face symposiums. These meetings are a place to review cancer informatics and data science priorities and initiatives, providing a forum for discussion of the strategic and pragmatic issues that we faced at our respective institutions and cancer centers. Here, we provide meeting highlights from the latest CI4CC Symposium, which was delayed from its original April 2020 schedule because of the COVID-19 pandemic and held virtually over three days (September 24, October 1, and October 8) in the fall of 2020. In addition to the content presented, we found that holding this event virtually once a week for 6 hours was a great way to keep the kind of deep engagement that a face-to-face meeting engenders. This is the second such publication of CI4CC Symposium highlights, the first covering the meeting that took place in Napa, California, from October 14-16, 2019. We conclude with some thoughts about using data science to learn from every child with cancer, focusing on emerging activities of the National Cancer Institute's Childhood Cancer Data Initiative.

JCO Clin Cancer Inform 5:881-896.

INTRODUCTION TO THE FALL 2020 CI4CC MEETING

The Fall 2020 Symposium was the 15th organized National Cancer Institute (NCI) and Community Cancer Center Informatics Symposium (see review of Fall 2019 Symposium[1]). This conference focused on the evolving and increasing dependence upon data science and informatics in pediatric and Adolescent and Young Adult (AYA) cancers to provide insights into biology and potential therapies. Participants discussed scientific advances in the understanding of pediatric and AYA cancers and how lessons learned in these domains can affect cancer research and care in adult populations.

Pediatric and AYA cancers are collectively considered rare diseases because there are so few cases diagnosed each year, particularly when compared with cancers diagnosed in older patients. This fact does not lessen the devastating effects of cancer, particularly the prolonged and highly toxic treatments, on every individual patient and family. Fortunately, the number of survivors of childhood and AYA cancers is steadily increasing; however, they are at great risk for secondary cancers later in life and often suffer sustained adverse effects. There is a growing focus on data collection within the cancer research community alongside increasing utilization of cloud-based repositories and digital tools for analysis. Opportunities exist to improve survivorship and reduce the burden of cancer on patients and their families through (1) focus on data generation to fill key knowledge gaps, (2) creating linkages among existing data repositories, and (3) development of new, nationwide registries and cohorts to provide the power to discover novel treatments for pediatric and AYA cancers. The incorporation of data science and informatics approaches into pediatric and AYA research is key to advancing our understanding of cancer etiology, identifying new approaches for cancer prevention and early detection,

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on June 10, 2021 and published at ascopubs.org/journal/cci on August 24, 2021: DOI https://doi.org/10.1200/CCI.21.00040

CONTEXT

Key Objective

How do we incorporate informatics and data science capabilities to learn from every child with cancer? The Cancer Informatics for Cancer Centers (CI4CC) Fall 2020 Symposium provided an opportunity for leaders in childhood cancer research to share their latest efforts with the cancer informatics community.

Knowledge Generated

The CI4CC Fall 2020 symposium focused on innovation in childhood cancer research and how leaders in the field could contribute to the National Cancer Institute's Childhood Cancer Data Initiative's goal of learning from every child with cancer. Topics included the following: discoveries in immunotherapy, precision medicine, and rare tumors; childhood cancer registries; data repositories and data interoperability; recent health services research revelations about childhood cancers; and innovation in data science for pediatric cancer.

Relevance

The 15th CI4CC Symposium highlighted critical initiatives in childhood cancer research that rely on data science and informatics to further our understanding of disease and facilitate progress toward novel treatments. The National Cancer Institute, Cancer Centers, and the broader childhood cancer community have demonstrated commitment to working together toward these goals.

improving cancer patient outcomes, and enhancing cancer care delivery throughout the community. A focused, unified approach exploiting the power of data science to provide mechanistic insights and clinical predictive value in understanding these cancers is imperative. Ensuring data from every patient are available for analysis as an integrated whole is crucial to increasing this understanding.

HIGHLIGHTS FROM THE MEETING

Day 1

Drs Tony Kerlavage (NCI) and Anne Kirchhoff (Huntsman Cancer Institute and University of Utah, School of Medicine), the conference cochairs, set the stage for the meeting by outlining the topics to be covered, including

- Updates on progress in NCI's Childhood Cancer Data Initiative (CCDI)
- Emerging scientific discoveries in immunotherapy, precision medicine, and rare tumors
- Registries specifically focusing on childhood cancers and how to effectively capture data
- Data repositories and how to facilitate interoperability through the use of standards and applying data harmonization
- Recent health services research revelations about childhood and AYA cancers and strategies to address these findings
- How foundations and collaborations are fueling innovation in data science for pediatric cancer.

The Childhood Cancer Data Initiative. NCI Director, Dr Ned Sharpless, and Dr Jaime Guidry Auvil (NCI) provided an overview of the NCI's CCDI,[2] which laid a foundation for the remaining talks in the conference. Dr Sharpless outlined CCDI's vision to learn from every child diagnosed with cancer through building a community centered around childhood cancer research and clinical care data. CCDI received its initial $50 million US dollars (USD) federal investment in fiscal year 2020, with an additional $50 million USD proposed each year for a total of 10 years. These funds allow NCI to bring the nation's childhood and AYA cancer research, advocacy, and care communities together in this ambitious effort in data collection, sharing, analysis, and access. Together with the broader community, NCI can bring together childhood cancer care and research data from across the nation to allow researchers to make new discoveries that improve treatments, outcomes, quality of life, and survivorship for patients.

Dr Guidry Auvil updated on foundational activities supported through CCDI in its first year, highlighted recommendations of the NCI Board of Scientific Advisors ad hoc Working Group[3] in support of CCDI, and described goals that the initiative will strive to meet over 10 years through enhanced data sharing. In the first year of the initiative, NCI focused largely on strengthening existing childhood cancer research programs, developing systems, and building on the working group report to support future CCDI activities. NCI is working to achieve these goals through (1) building an infrastructure of federated, searchable data resources and tools; (2) developing comprehensive data sets that include all types of data for broader discovery; (3) creating resources that make data sets easier to use and understand by multiple types of users; and (4) supporting research to improve treatments and meet the needs of childhood and AYA patients, survivors, and their families. Specifically, CCDI has supported enhanced data sharing within molecular repositories and clinical registries across NCI and sequencing and data collection to complete critical programs like My Pediatric and Adult Rare Tumor Network (MyPART),[4] Pediatric Molecular Analysis for Therapy

Choice (MATCH),[5] and the Childhood Cancer Survivor Study (CCSS).[6] Furthermore, CCDI has provided funding for development and characterization of cancer models within the Human Cancer Models Initiative[7] and the Pediatric Preclinical Testing Consortium (PPTC).[8] In the coming years of the initiative, CCDI will continue on its quest to learn from every patient through identifying and engaging all patients diagnosed with childhood or AYA cancers and developing a national strategy to minimally characterize and treat each patient affected by childhood and AYA cancers and to further collect, manage, and share the data obtained through these efforts to accelerate understanding of disease and improve therapeutics and quality of life for all patients and survivors.

Scientific innovation in childhood cancer programs. This session highlighted several innovative clinical and preclinical research approaches to understanding childhood cancers.

Dr Kara Davis (Stanford University) began by describing the work of the NCI-sponsored Pediatric Immunotherapy Discovery and Development Network (PIDDN).[9] Immunotherapies have garnered well-deserved excitement for the treatment of cancer. Immune checkpoint inhibition has been approved for many cancer histologies; yet, the impressive treatment responses to immune checkpoint inhibition in adult cancers have been rarely seen in pediatric tumors outside of lymphoma.[10] By contrast, cellular immunotherapies and monoclonal antibody therapies have demonstrated clinically meaningful responses in pediatric tumors (relapsed lymphoblastic leukemia and neuroblastoma).[11,12] The PIDDN represents a focused investment in expanding the reach of immunotherapies to childhood tumors. The work of Dr Davis and her colleagues focuses on determining the tumor immune architecture in childhood solid and brain tumors and identifying coexpression of candidate immunotherapy targets using a single-cell, high-dimensional imaging platform (Multiplexed Ion Beam Imaging [MIBI]).[13] Starting with neuroblastoma, a tumor of the peripheral sympathetic nervous system that accounts for 10% of pediatric cancer-associated deaths, they have optimized a 40-antibody panel to capture infiltrating immune cells, tissue architecture, and neuroblastoma proteins including those associated with adrenergic or mesenchymal identity.[14] The team optimized their antibody staining and determined tissue specificity using a unique tissue microarray, which included 32 neuroblastoma cores, additional childhood solid tumors (osteosarcoma, Ewing's sarcoma, rhabdomyosarcoma, Wilms tumor, and others), and healthy pediatric tissues from various organs (kidneys, liver, and tonsil). They uncovered divergent neuroblastoma cell populations differentiated by coexpression of CD57, Ki67, and GPC2, as well as mesenchymal (FN1 and SNAI2/SLUG) and adrenergic (TH, GATA3, and PHOX2B) proteins, demonstrating promising early results of their MIBI application to neuroblastoma. Dr Davis' work demonstrates the feasibility of using MIBI to make translatable discoveries in neuroblastoma, and they are now developing the approach for application to childhood sarcomas. Working within the PIDDN enables the opportunity to leverage the expertise and discoveries of the network to improve immunotherapies for childhood cancers.

Dr Kim Stegmaier (Dana-Farber Cancer Institute) described her laboratory's work on a Cancer Moonshot–funded program to identify fusion oncoprotein drivers in pediatric cancers. Fusion-driven cancers are common, particularly in young populations, and occur across the spectrum of diseases from solid tumors to hematologic malignancies. Dr Stegmaier highlighted several themes: (1) the fact that fusions occur in the context of few recurrent point mutations, (2) the fusion is the initiating cancer event, (3) they occur in specific development states, (4) tumors are immunologically cold and nonresponsive to checkpoint inhibitors, and (5) fusions often involve transcription factors of chromatin regulators. She suggested that now is an excellent time for this research, as massively parallel sequencing allows for the identification of novel fusions, and new technologies (eg, CRISPR/Cas9 and Cryo-EM) and degrader chemistry present promising approaches to treatment. The team adopted a strategy to identify cancer vulnerabilities using a cancer dependency map. This involved profiling cancer cell lines for molecular features, introducing chemical or genetic perturbations, creating algorithms to analyze the resultant data generated, and developing new therapeutic hypotheses. Using the AVANA library, which contains more than 70,000 CRISPR guides, 18,000 targeted genes, and more than 700 cancer cell lines screened for dropouts, the team screened 100 pediatric cancer cell lines, which were integrated into the larger cell line map. These maps are released quarterly to the public for research. The team found that pediatric cancers have as many selective gene dependencies (ie, a specific cancer is dependent upon a particular gene) as adult cancers, despite having fewer mutations, and many are unique dependencies in these childhood cancers. Dr Stegmaier described a particular example of her work in Ewing Sarcoma, which is driven by an EWS/ETS fusion. Taking a CRISPR screening approach to identify proteins that regulate the stability of the EWS/FLI protein, her laboratory engineered a model system and identified TRIM8 as a novel regulator in Ewing Sarcoma. It also scored as the top dependency in the pediatric dependency map, and multiple independent screens confirmed the same target to have a direct effect on the fusion itself.

Dr Nita Seibel (NCI) presented early lessons learned from the Pediatric MATCH trial, a collaboration between the NCI and the Children's Oncology Group (COG). This signal-finding trial is designed to determine whether matching certain targeted drugs in children and adolescents whose tumors have specific gene abnormalities will effectively treat their cancer, regardless of the cancer type. Enrollment on this trial has greatly exceeded expectations, accruing more than 1,000 patients in the first 3 years from a diverse

and growing collection of COG institutions. Targeted Oncomine DNA and RNA mutation or fusion panels covering more than 400 genes and 4,000 mutations of interest were used for tumor analysis of each sample. Actionable mutations were detected in 29% of all tumors tested, and a significantly higher percentage (46%) was detected in CNS tumors. Approximately 25% of study patients were assigned to a treatment arm, and 10% of patients screened were enrolled as of September 2020.

Dr Karlyne Reilly (NCI) described her work with NCI colleagues through the MyPART Network to integrate longitudinal clinical, molecular, and patient-reported data from rare tumors in young patients. MyPART is funded as part of the Cancer Moonshot through the Center for Cancer Research in the NCI Intramural Research Program and is a patient engagement network focused on pediatric and young adult rare solid tumors. Dr Reilly discussed how development of new therapies to treat rare cancers is hampered by difficulties in designing effective clinical trials because of (1) incomplete understanding of causes and behavior of rare cancers, (2) lack of appropriate trial end points, and (3) low trial enrollment because of rarity of the condition. Draft guidance from the US Food and Drug Administration (FDA)[15] and experience from development of selumetinib for treatment of rare plexiform neurofibromas[16] have shown the importance of well-designed natural history studies to collect observations critical for further therapeutic developments. Patients and patient advocates provide an important perspective in cancer research[17] and can help galvanize the rare cancer patient community to participate in clinical trials. The goals of MyPART are to (1) develop a network of clinical or research sites and accompanying patient and advocacy interface, (2) provide state-of-the-art expertise and personalized health care and data to children and young adults with rare tumors, and (3) build databases and tools to advance research on novel treatments. Dr Reilly et al have developed a natural history and biospecimen acquisition study for children and young adults with rare solid tumors,[18] enrolling patients with rare tumors of interest to work with MyPART investigators. Standardized medical and family history, patient-reported outcomes, and biospecimens are obtained directly from patients, who can participate remotely. In cases where patients may benefit, they are brought to the National Institutes of Health (NIH) Clinical Center for more extensive tests and collection of additional data and tissue samples. Patients will be followed yearly to collect longitudinal data on how they respond to treatments through standard of care and treatment trials that they may join at other institutions.

Dr John Maris (Children's Hospital of Philadelphia) described his work in the development of mouse models for childhood cancers and the importance of these models as tools for pediatric drug development. Because pediatric cancers are rare diseases, with biologic characteristics and oncogenic drivers distinctive from adult cancers, prioritization of potential targets and treatments for clinical evaluation in children is essential. Recent initiatives, such as the NCI CCDI, the Research to Accelerate Cures and Equity for Children Act, and the Foundation for the NIH ACT4PEDS preclinical testing program, present a confluence of opportunities to make progress in these areas. ACT4PEDS seeks to build upon the NCI-supported PPTC, which, in turn, builds upon 10 years of experience with the Pediatric Preclinical Testing Program. The Pediatric Preclinical Testing Program and PPTC collaborated with scores of pharmaceutical companies to test more than 170 novel agents against the programs' pediatric preclinical models of sarcoma and renal tumors, neuroblastoma, brain tumors, osteosarcoma, and leukemias. Multiple drugs were moved to clinical evaluations as a result of the testing results. This work confirmed the importance of murine-human pharmacodynamic and pharmacokinetic comparisons and the importance of genomic characterization for interpreting preclinical testing results. A diverse set of 261 pediatric cancer PDX models were genomically characterized, and the preclinical models recapitulated the genomic alterations associated with those observed in the clinical setting, with enrichment of genomic alterations observed at relapse. In conclusion, the preclinical research consortium is planning for Research to Accelerate Cures and Equity Act activities including (1) making the FDA Relevant Molecular Targets List computable, (2) developing the infrastructure to adjudicate the Relevant Molecular Targets List with proteogenomic and drug trial data from both humans and mice, and (3) scaling a preclinical testing program to allow for increased throughput and for expansion to include models that can be used in evaluating the growing arsenal of immuno-oncology agents.

Dr Gregory Reaman (Oncology Center of Excellence, US FDA) closed out the session with a talk on molecular targets and cancer drug development to facilitate precision oncology for children. Recent changes in the regulatory environment surrounding pediatric cancer drug development provide unprecedented opportunity to advance the concept of precision oncology for children. Innovative drugs directed at molecular targets associated with adult cancers may provide therapeutic options for children with cancer despite their etiologic and biologic differences.[19,20] New requirements for early evaluation of appropriate drugs in children, on the basis of their molecular mechanism of action rather than clinical indication, will shorten the unacceptable time lag between first in human and first in children studies. Specifically, the FDA Reauthorization Act of 2017, Sec. 504[21] amended the Pediatric Research Equity Act Sec 505B of the Food, Drug, and Cosmetic Act[22] to require the sponsor of an original new drug application or biologic license application submitted on or after August 18, 2020, to conduct a pediatric study (to yield meaningful data regarding dosing, safety, and preliminary efficacy to

inform potential pediatric labeling) of any new drug intended for the treatment of an adult cancer and directed at a molecular target that the FDA determines to be substantially relevant to the growth or progression of a pediatric cancer. The FDA and NCI were required to create and update a list of relevant molecular targets (informed by peer-reviewed literature and publicly available databases and now published on the FDA website[23]), to which this new pediatric study requirement applies. The framework for constructing the list of relevant molecular targets included the following: (1) genomic data indicating an association of specific gene aberrations, protein overexpression, and pathway dysregulation in one or more pediatric cancers; (2) functional evidence of a target's role in synthetic lethality; (3) evidence that target modulation affects tumor cell sustainability; and (4) existence of predictive and/or response biomarkers. Evidence of target actionability and the consequence of target inhibition in specific pediatric cancers are incomplete. Extending the evidence base for determination of relevance of purported targets requires broad genomic sequencing of childhood cancers and rational decisions for preclinical testing in appropriate models. Data sharing of these findings is essential and mandated for meaningful transformation of regulatory change to improve therapeutics for childhood cancer.

Day 2

Pediatric registries. Dr Lynne Penberthy (NCI) opened the second day of the conference by describing efforts toward establishing a National Childhood Cancer Registry (NCCR).[24] Drs Amie Hwang and Dennis Deapen (both from University of Southern California, Norris Comprehensive Cancer Center) followed with a deeper dive into critical aspects of the NCCR. In 2020, the NCI Board of Scientific Advisors Working Group in support of CCDI reported a set of recommendations[3] including Developing the National Childhood Cancer Registry, as part of the CCDI data ecosystem, to enhance access to patient-linked childhood and AYA cancer and survivorship data. The purpose of the NCCR is to leverage and link disparate health data from multiple existing sources to create an infrastructure that can better support research on childhood cancer. The core data are derived from population-based cancer registries including patients with cancer under age 20 years diagnosed since 1995. The NCCR will securely maintain personally identifiable information to facilitate frequent, efficient addition of data on all childhood cancer cases. The NCCR is exempt from Health Insurance Portability and Accountability Act of 1996 regulations and will comply with regulatory requirements in each state. The NCCR will be expanded to include dozens of sources of relevant, individually linked health information including pharmacy, treatment, outcomes, genetics, genomics, birth records, and residential history. Initial registry participation includes approximately 70% of US childhood cancer cases and is coordinated by the North American Association of Central Cancer Registries,[25] supported by NCI and advised by a variety of clinical, epidemiologics, and genomics experts. Four working groups on Metadata, Genomics and Biospecimens, Data Access or Data Release, and Data Products have been established to inform NCCR development. The Data Products workgroup, led by Drs Hwang and Deapen, is tasked with the development of output products and utilization tools targeted to benefit researchers, clinicians, cancer registries, policy makers, educators, and patients and their families. Six preliminary data products are proposed: (1) a public-facing interface providing counts and indexing tools allowing sample size determination of registry, (2) precalculated public health statistics similar to SEER*Explorer, (3) selectable descriptive analytics similar to SEER*Stat, (4) secure cloud-based patient-level analysis platform, (5) individual-level data sets requiring protocol approval, and (6) public information including facts and infographics.

Dr Eric Durbin (University of Kentucky, Markey Cancer Center) discussed the Virtual Tissue Repository (VTR) as a key component of the NCCR approach. Childhood cancer biorepositories are constrained by two key limitations that include insufficient cases to study rare tumors and inherent bias. Representing $< 1\%$ of all cancer diagnoses, childhood cancers are extremely rare by histologic and molecular subtypes.[26] Cases are not randomly selected from the underlying population, rendering them unrepresentative of all children diagnosed.[27] To address these limitations, the SEER[28] Program is developing VTRs that enable the acquisition of cohort-based pathology specimens for cancer research. Mandated by state laws, central cancer registries serve as public health authorities that are explicitly exempt from Health Insurance Portability and Accountability Act privacy regulations. In this capacity, registries may serve as honest brokers to provide deidentified biospecimens and data for nonhuman subjects' research, while protecting patient confidentiality. To support institutional review board–approved studies, VTRs leverage registry data and electronic pathology reports to identify population-based cohorts of cases and negotiate with pathology laboratories to obtain the surgical and biopsy archival tissue specimens. Significant resources are required by central registries to support VTR staff and infrastructure. Pathology laboratories also charge to retrieve tissue blocks and slides, as well as for pathologists' time to screen cases. In addition to costs, investigators should be aware of the effort and time that may be required to collect the specimens and to conduct custom annotations when needed. Appalachian children in Kentucky and other states experience significantly high rates of brain and CNS tumors.[29,30] Dr Durbin presented a real-world example of using the Kentucky Cancer Registry VTR to obtain specimens needed to explore associated molecular risk factors. The VTR succeeded in negotiating with 11 pathology

laboratories to obtain the complete census of available specimens (N = 258). Specimens are being sequenced and analyzed in collaboration with the Center for Data Driven Discovery in Biomedicine at the Children's Hospital of Philadelphia and the Kids First Data Resource Center (KFDRC).

Sara Gallotto (Massachusetts General Hospital [MGH]) described the use of natural language processing (NLP) and other automated data entry in the Pediatric Proton/Photon Consortium Registry (PPCR).[31] The PPCR was established in 2012 to create a comprehensive description of the pediatric cancer patient cohort being treated with proton radiotherapy. The main goal has since evolved to expediting outcomes research in all pediatric radiation oncology patients, including any modality.[32] The PPCR has enrolled more than 3,300 patients at 17 centers. All data are manually entered into a centralized database containing 1,600 potential data fields. Using NLP to automate data entry will improve data completeness within the PPCR and save time and resources. Previous attempts to extract tumor site from pathology reports were piloted at MGH. Existing rules-based NLP models from the MGH breast cancer clinic were used to test 679 pediatric reports to determine if the tumor was in the brain or another organ in the body. These attempts failed because of small patient numbers and lack of standardization in clinic notes. Successful NLP models in the pediatric cancer space will require thousands of free-text clinic notes to improve model sensitivity and specificity. The PPCR is currently working with MIM Software Inc to use their artificial intelligence–assisted contouring software, Contour Protégé AI, to fine tune pediatric models and allow for automated contouring of radiotherapy plans in future PPCR research and into the clinical realm. Standardization of data is critical to the success of NLP efforts.[33] The PPCR team will continue to work with the NCI under the CCDI to partner with researchers at the Massachusetts Institute of Technology to develop rules-based and machine learning tools to extract and categorize key radiation treatment data from clinic notes at the various PPCR centers.

Dr Kelly Getz (Perelman School of Medicine, University of Pennsylvania) presented an overview of the Pediatric Health Information System (PHIS),[34] a comparative pediatric database that includes claims data from 50 children's hospitals across the United States. The PHIS system covers more than 9.2 million inpatient stays and 40.9 million emergency department (ED) encounters, including charge data on the basis of nearly 255 million total International Classification of Disease, 9th Revision and Tenth Revision codes collected since 2004. The strengths of this data system include detailed information at the patient level on both inpatient and affiliated outpatient stays, daily resources use, and costs from a broad and heterogeneous set of hospitals representing variability in practice and outcomes. Because of the large sample size, PHIS provides an opportunity to study rare diseases and events. Limitations of this system are that the data are administrative (thus not designed for research), largely represent inpatient-based utilization, and lack detailed laboratory data important as covariates. Furthermore, outcomes collected are incomplete because of limited follow-up and focus largely on inpatient utilization. Nevertheless, exploiting strengths of multiple data resources through linkages can overcome limitations inherent in the use of any one resource alone.

Dr Richard Aplenc (Children's Hospital of Philadelphia) provided a plenary talk on Understanding the Differences Between Childhood and Adult Cancer: Outcomes, Genetics, and Impact on Research and Data Science. He described how pediatric oncology has a long-standing history of collaborative clinical trials and cooperative research efforts aimed at improving outcomes for children with cancer, particularly as compared with their adult counterparts. As is well-known, Dr Sidney Farber published the first report detailing an effective therapy for pediatric leukemia in 1948.[35] Less well-known is the first multi-institutional (11 centers and 27 participating investigators) clinical trial in pediatric oncology to randomly assign treatments for 125 patients in 1955.[36] Although the trial did not identify a meaningful difference between the two treatment arms, it did establish the first pediatric cooperative oncology group. This establishment was followed by four other pediatric cancer cooperative groups and the subsequent coalescence of these groups in 2000 into the current COG.[37] These cooperative groups play a critical role in pediatric cancer research because of the relatively small number of children diagnosed with cancer on an annual basis in the United States. As a result, definitive efficacy testing or biomarker validation must occur within a coordinated, collaborative setting. Cooperative groups have been very effective in creating the clinical and biology studies needed to achieve these aims. As an example, the largest currently open trial in acute myeloid leukemia (AML) in the United States is the COG trial with a target enrollment of nearly twice that of the second largest open US trial in adults. In addition, previous COG AML studies have provided critical data used to support the approval for gemtuzumab (targeted antibody-drug conjugate for AML) and to define the molecular differences between adult and pediatric AML.[38] However, COG studies, like all cooperative group trials, have important constraints, including limited data collection and the lack of a fully represented source population.[39] Multiple groups are working to address these issues through automated data collection from the electronic health record.

Pediatric data resources and integration. The topic for the remainder of day 2 pivoted to pediatric data resources. Dr Sam Volchenboum (University of Chicago) started the session by discussing how the proper infrastructure and data harmonization can transform the way that researchers share and use data. Congressional approval of the first year

of funding the $500 million USD CCDI in December 2019 represents an opportunity to transform how pediatric cancer data are collected, aggregated, and shared. Data commons allow researchers to achieve these goals through the creation of a cloud-based repository of aggregated and harmonized data with tools for cohort discovery, visualization, and data request and fulfillment. The development of data commons presents challenges in infrastructure development and deployment, data governance, and data life cycle management. The University of Chicago's Pediatric Cancer Data Commons (PCDC)[40] has created a platform for aggregating and sharing pediatric oncology clinical trials data from around the world. Such a platform enables research into the causes of cancer, better methods of risk stratification, novel therapies, and more sensitive ways to detect disease and monitor response to therapy. The PCDC team has identified features critical for successful data commons design and implementation. First, contributors and core research aims must be identified to establish the foundation for the data commons. Second, long-term funding must be identified to support commons development and sustainability. In addition, the selection of infrastructure should include ease of adding data, tools for data model development, standard processes for change management, and ease of connecting to external data sources. Next, a strong governance program helps disease consortium members work toward shared goals through consensus. Importantly, the development of a common data dictionary through a balloting process is critical for harmonizing and aggregating clinical data. Additionally, connecting to external data sources via a common identifier enhances the usefulness of the data. The interface to engage with a commons must be well-designed and easy to use to promote discovery and hypothesis generation. Finally, education and training help ensure widespread proper use of the commons and its tools. The PCDC contains clinical data from more than 25,000 children treated on cancer clinical trials. Soon, 10 different disease areas will be represented in the PCDC with more and updated data being added continually. Data are being connected to other data sources and nodes in the NCI's Cancer Research Data Commons[41] ecosystem through a common identifier available for US patients. The PCDC data model is being integrated into the harmonized Cancer Research Data Commons model designed by the Center for Cancer Data Harmonization. Ultimately, the PCDC is a paradigm system for how data commons can lower barriers to research and facilitate access to high-quality harmonized international data aggregated across clinical trials.

Dr Allison Heath (Children's Hospital of Philadelphia) provided an overview of the Gabriella Miller KFDRC.[42] Since launching to the public in October 2018, KFDRC has made an increasing number of pediatric genomic studies, including family-focused cohorts, available to the research community. By the end of 2020, genomic and clinical data from almost 15,000 participant samples will be available across a variety of structural birth defect and pediatric cancer cohorts. The KFDRC has architected a secure, cloud-based platform that supports the ability of researchers to not only find, access, and reuse data but also integrate, collaborate, and analyze data quickly at scale. A best-of-breed approach has been taken to develop a portal, with reusable components from Overture, as the entry point for Kids First data. From there, users can integrate with platforms such as Cavatica[43] for bioinformatics workflows and PedcBioPortal[44] for cancer genomic visualizations. Additionally, a set of framework services, powered by Gen3,[45] provide a foundation for interoperability with other large-scale data sources, platforms, and a growing ecosystem of analytic and visualization applications that provide capabilities for rapid in-place analyses without download. In early 2021, these services included a new functionality in a variant workbench, which provides authorized researchers with cloud-based, high performance access to the billions of germline variants that have been called on Kids First data. Recognizing the value in both the source data collected by the original studies and the power in having harmonized genomic and clinical data for cross-study analysis, the KFDRC makes both available in the platform. The explore data feature on the portal allows users to search and browse in real-time across all Kids First studies to identify virtual cohorts of interest for further study. Within the portal, these cohorts can be saved and shared with collaborators for iterative refinement and analysis. With appropriate approvals, the associated data can be accessed and analyzed seamlessly in Cavatica or other platforms with interoperable framework services. More recently, clinical data have been identified as a key priority to better empower the ability for discovery and translational research. Toward this end, the KFDRC has piloted representing research data in the Fast Healthcare Interoperability Resources (FHIR) standard and plans to provide the ability to access clinical data via FHIR end point in the first half of 2021.

Dr Bruce Aronow (Cincinnati Children's Hospital) showed how a consistent approach to the analysis of single-cell genomic profiles of primary tumors and patient-derived xenografts versus background tissues (with or without comparisons with those tissues from development and organoids) has the potential to identify driver gene regulatory networks that distinguish replicative from post-replicative tumor cells. Importantly, these signatures and modular representations of tumor cell program can be easily saved and shared for cross-comparisons between different tumor types and subtypes that can be used for large-scale database generation, a preview of which can be seen at ToppCell.[46].

Dr Jinghui Zhang (St Jude Children's Research Hospital) spoke about BIG Pediatric Cancer Genomic Data: Discovery, Precision Medicine, and Data Sharing. High-

throughput next-generation sequencing has enabled tremendous advances in our understanding of the genomic landscape of pediatric cancer, leading to discoveries of new mechanisms of tumor initiation and progression, novel targets, and diagnostic and prognostic markers. By focusing on the development of innovative computational analysis tools, researchers at St Jude have investigated the genomic variants in pediatric cancer in the following areas: (1) genomic landscapes of > 20 subtypes of pediatric cancer; (2) a pan-cancer study of genomes and transcriptomes of pediatric cancer, which unveiled that > 50% of the driver genes are absent in adult cancer; (3) clonal evolution of relapsed pediatric leukemia driven by therapy-induced variants bearing novel mutational signature; and (4) pathogenic germline mutations in cancer predisposition genes. These insights led to implementation of clinical cancer genomic profiling by three-platform analyses of whole-genome, whole-exome, and transcriptome sequencing for all eligible patients with pediatric cancer at St Jude. Since 2015, they have analyzed > 1,200 pediatric oncology patients providing critical data that may affect patient care. The omics data generated from their research and clinical programs can be accessed on St Jude Cloud,[47] a cloud-based data sharing ecosystem for accessing, analyzing, and visualizing genomic data generated from > 10,000 patients with pediatric cancer, long-term survivors of pediatric cancer, and > 800 pediatric sickle cell patients. Access to three interconnected Apps on St Jude Cloud, that is, Genomics Platform, Pediatric Cancer (PeCan) Knowledgebase, and Visualization Community, provides a unique experience for simultaneously performing advanced data analysis and enhancing the knowledgebase for pediatric cancer.

Dr Olena M. Vaske (University of California [UC], Santa Cruz) closed out the day discussing leveraging large genomic data sets for clinical impact on individual children with cancer. Pediatric cancers have significantly lower DNA mutation rates than adult cancers.[19] However, activated and targetable oncogenic driver pathways can be identified in RNA sequencing (RNA-Seq) data from cancer biopsies or resections.[48,49] Therefore, the UC Santa Cruz Treehouse Childhood Cancer Initiative[50] leverages genomic data to perform comparative RNA-Seq analysis between a single pediatric tumor and a background compendium of cancers to identify unusually expressed oncogenes and oncogenic pathways. This analysis identifies genes with outlier expression in a single tumor sample as compared with the background cohort.[48] Treehouse partners with existing clinical genomics trials and provides a report of outlier genes and pathways, and the top most correlated tumor samples to the focus sample. For example, the Treehouse team demonstrated that the identification of similarly profiled tumor samples aided in refining the diagnosis of a 10-year-old female patient originally diagnosed with immature teratoma.[51] After finding that the top six most correlated samples to her tumor were all glioma, additional histopathologic analysis was performed and the diagnosis refined to gliomatosis peritonei. This tumor consists of mature glial tissue in the peritoneum, which can co-occur with teratoma, and is difficult to identify via histopathologic analysis alone.[52] They also showed that targeting overexpressed genes can have therapeutic impact, as seen in the case of a 1-year-old male patient diagnosed with myoepithelial carcinoma of the liver. The Treehouse investigators identified overexpression of several receptor tyrosine kinases and pathways (targetable by pazopanib), and cell cycle genes and pathways (targetable by ribociclib), in the tumor sample. The patient was treated with pazopanib for 3 months until growth of lung nodules was observed and then received 12 cycles of ribociclib over 1 year. All lung nodules remained stable and were removed surgically, and the patient now has no evidence of disease. Their work shows the critical value of data sharing and comparative RNA-Seq analysis in precision medicine trials and pediatric cancer.

Day 3

Pediatric and AYA cohorts. The third day opened with a series of talks discussing Pediatric, AYA cancer cohorts. Dr Todd Alonzo (University of Southern California, Keck School of Medicine) started the session with an overview of the COG Project: EveryChild (PEC).[53] He highlighted the challenges in pediatric cancer research, including the need to improve cure rates for refractory cancers, diminish acute toxicity for patients with cancer enduring harsh treatments, and minimize risk for late effects in survivors. Future progress in childhood cancer research hinges on being able to link patient-level omic data (genomics, proteomics, etc) with rich patient-level clinical data. The ability to provide this link was a huge motivation for the development of PEC, also referred to as COG study APEC14B1. The COG PEC was activated in October 2015 and has enrolled more than 27,000 infants, children, adolescents, and young adults up to age 25 years at the time of diagnosis. Enrollment can occur at the time of diagnosis, relapse or progression, second or secondary malignancy, and postmortem. The COG PEC comprises five major components for which patients can provide consent:

(1) Eligibility screening—Clinical, pathologic, imaging, surgical, and biologic data are used to determine study eligibility or risk stratification for enrollment onto therapeutic trials. (2) Biobanking for future research—Well-annotated biospecimens from pediatric cancers are collected at initial diagnosis, progression, relapse, and postmortem. As of June 2020, nearly 200,000 specimens have been received from roughly 12,000 patients. This provides researchers an outstanding resource for future biology studies. (3) Tracking outcome—PEC tracks therapeutic approach and outcome of treatment. (4) Childhood Cancer Registry—PEC aims to establish a near population-based pediatric cancer registry. (5) Contact for future research—PEC allows patients to consent to be contacted

for future research, which enhances the conduct of COG biology, epidemiology, and survivorship studies. In summary, PEC links data from the laboratory to information on each child's cancer, provides well-annotated biospecimens to research laboratories around the world, and aims to capture biology and outcome data for children diagnosed with cancer—no matter how common or rare.

Dr Paul Nathan (Hospital for Sick Children, University of Toronto) focused his talk on research efforts to link Ontario, Canada's administrative health care data to provincial pediatric and AYA cancer registries. The Pediatric Oncology Group of Ontario (POGO) maintains a registry of all patients treated before age 18 years at one of Ontario's five pediatric cancer centers since 1985.[54] Data holdings include demographic, disease, and detailed treatment information. Although these data are primarily used to obtain estimates of Ontario's current and future childhood cancer incidence to inform the distribution of oncology resources in the province, it is also used for research. POGO's registry is complemented by the Initiative to Maximize Progress in AYA Cancer Therapy,[55] a research database of patients 15-21 years old and diagnosed with one of the five common AYA cancers between 1992 and 2011. The POGO Networked Information System[56] holds data of more than 14,000 children with cancer, whereas Initiative to Maximize Progress in AYA Cancer Therapy has data of 5,349 AYA cases. These two databases have been linked in multiple research studies to administrative health data held at the Institute for Clinical Evaluative Sciences, a nonprofit research institute that applies the study of health informatics[57] for and population-wide health outcomes research using data collected through the routine administration of Ontario's publicly funded health care system.[58] Children and AYAs with cancer can be linked to determine their health care utilization records using an encrypted identification number derived from their provincial health card. These linkages have been used to study outcomes across the cancer spectrum, including cancer diagnoses, disparities in care, cancer therapy, the impact of locus of care on outcomes,[59] late effects of cancer therapy,[60] palliative care, health economics, and cost-effectiveness,[61] and to develop methodology for the use of these databases for cancer research.[62]

Dr Jenny Poynter (University of Minnesota) discussed the COG's registry protocols, ACCRN07 and APEC14B1, including an overview of the available data and an example of funded work conducted using the registry. The Childhood Cancer Research Network (ACCRN07) was the initial registry protocol within the COG and included a tiered consent including participation in the registry and a separate consent for future research contact.[63] More than 56,000 patients with childhood cancer were enrolled in the registry during the 10 years of its existence. On the basis of a comparison with SEER data, this represents approximately 40% of all patients with childhood cancer diagnosed during this time period in the United States.[64] The CCRN was replaced by COG's Project: EveryChild (APEC14B1), which opened in 2015 and combined the registry with disease-specific biobanking protocols. A brief demographic and epidemiologic questionnaire is also completed at the time of enrollment. The COG registries have facilitated epidemiologic studies of multiple childhood cancers, including acute lymphoblastic leukemia, rhabdomyosarcoma, neuroblastoma, osteosarcoma, Ewing sarcoma, Wilms tumor, germ cell tumors, and histiocytosis. Dr Poynter provided highlights from her work studying the etiology of germ cell tumor[65,66] and plans for an ongoing survivorship study involving cases recruited through the registry protocols. The COG registry protocols will support current and future etiologic studies of childhood cancer and currently serve as the largest pediatric cancer biobank in the nation.

Dr Greg Armstrong (St Jude Children's Research Hospital) discussed the lifetime impact of childhood cancer and often harsh cancer therapies on survivors and how the CCSS[6] can be used as a resource for survivorship research. The CCSS is a multi-institutional, multidisciplinary collaborative research resource comprising a cohort of 38,036 five-year survivors of childhood cancer and a comparison cohort of 5,059 siblings.[67,68] The CCSS resource permits investigators to conduct high-quality research addressing long-term morbidity, mortality, and health-related quality of life. With the successful recruitment and longitudinal follow-up of the cohort, which includes survivors diagnosed and treated over three decades (1970-1999), the CCSS is a resource for conducting observational and intervention-based research on a broad spectrum of long-term outcomes across the life span of aging survivors. The resource includes comprehensive annotation of treatment exposures, ongoing longitudinal follow-up, and an established biorepository, with germline genetic information on 8,380 survivors (combinations of single nucleotide polymorphism array, whole-exome, and whole-genome sequencing) for identification of genetic susceptibility for disease- and treatment-related late effects.[69] Enhanced access to CCSS data by the scientific community has been achieved through posting of genetic and phenotype data to the Database of Genotypes and Phenotypes at NIH and the St Jude Cloud.[47] Development of a web-based and mobile health (mHealth) platform with enrollment of 10,700 participants (enrollment ongoing) available for rapid contact for recruitment to intervention trials has expanded the availability of this population of cancer survivors to researchers. The CCSS will continue to be a primary resource for research that informs exposure-based screening and health surveillance recommendations for the growing population of childhood cancer survivors as well as intervention trials vital to changing the trajectory of health outcomes in this population.[70,71] CCSS is an open resource for investigators actively seeking to leverage the current successful organizational and research infrastructure to address the long-term mortality, morbidity, and quality of life of

survivors of childhood cancer. Understanding the risk for late effects of childhood cancer and its therapy provides the basis for health screening recommendations and interventions that can mitigate long-term health problems in this high-risk population.

Dr Anne Kirchhoff (Huntsman Cancer Institute, University of Utah) provided an overview of the Utah Population Database (UPDB)[72] resource, which is the only population-based resource of its kind in the United States. UPDB links personally identifiable information with residential, demographic, and cancer registry data from the Utah Cancer Registry,[73] family history, and medical records. There are more than 4,200 patients with pediatric cancer diagnosed between age 0 and 14 years and more than 20,700 AYA patients diagnosed between age 15 and 39 years, with records starting in 1986. The UPDB allows for novel questions to be asked in pediatric and AYA cancer because of the depth of available data. The UPDB contains family pedigree information spanning up to 20 generations that identify an individual's family history of cancer.[74] Additionally, UPDB records enable research on coaggregation of health conditions, such as infertility and cancer, as well as longitudinal information from health care records, enabling late effects surveillance. Because of the available data, UPDB can also be used to generate cancer-free comparison cohorts for research studies. Examples of pediatric and AYA cancer studies that have been conducted using the UPDB include high health care utilization among survivors compared with populations without cancer,[75,76] the effects of air pollution on health outcomes, and the association of fine particulate matter air pollution with greater mortality among patients with pediatric and AYA cancer.[77] UPDB data have been linked to environmental, geographic, and Census data, can be linked to other data sources at the individual or spatial level, and provide an important population-based resource for epidemiologic research in pediatric and AYA cancer.

Dr Erin Hahn (Kaiser Permanente Southern California [KPSC]) provided an overview of the current KPSC research on AYA cancer survivors. Health services and epidemiologic research incorporating data from integrated health systems such as KPSC have several advantages, including a very large and stable patient population, long-term electronic records, access to SEER-affiliated cancer registries, and strong partnerships between clinicians and researchers. AYA-focused studies from KPSC research scientists, led by Dr Chun R. Chao, showed that AYA cancer type distribution and 5-year survival are very similar to published AYA SEER statistics,[78] allowing for valid external generalizability from KPSC studies. Recent KPSC studies demonstrated that AYA survivors have an overall two-fold increased risk for cardiovascular disease compared with age- and sex-matched noncancer controls[79] and that the risk of comorbidities was significantly elevated in survivors compared with noncancer controls, with incidence rate ratios ranging from 1.3 for dyslipidemia to 8.3 for avascular necrosis.[80] Another KPSC study found that less than half of AYA Hodgkin lymphoma survivors received the guideline-recommended combination of office visits, imaging, and laboratory testing within the first 2 years after treatment.[81] This body of work helps to lay the foundation for future interventions to improve care delivery and patient outcomes for this unique population.

Dr Karen Wernli (Kaiser Permanente Washington Health Research Institute) discussed her research leveraging national claims data from a large national commercial insurer to understand end-of-life quality in patients with AYA cancer. Provenance of claims data was from an AYA cohort enrolled from 2001 to 2016 and used to describe receipt of end-of-life care. A subset of the study cohort was linked to the National Death Index[82] to validate date and cause of death. Proportions of receipt of ED visits, hospitalizations, intensive care unit (ICU) stays, and surgery within 30 days and receipt of chemotherapy within 14 days of death were calculated. Differences in proportions were calculated in unadjusted analysis by geographic region and temporality, using Chi-square tests. The study population was approximately 33 years old and evenly split between men and women. End-of-life measures within the last 30 days were similar to other AYA populations for ED visits (16.3%) and ICU stays (40.1%).[83-85] However, a higher proportion of AYAs in the cohort were hospitalized (78.4%). Hospitalizations were highest in the Northeast (83.4%) and lowest in the West (74.3%). ICU stays were highest in the South (42.6%) and lowest in the Midwest (36.2%). Visiting ED more than once in the last 30 days increased from 2001 to 2016 ($P = .01$). Furthermore, an increasing proportion of AYA is in the ICU by 2016 than in 2001 ($P = .004$). There are no differences in other end-of-life quality measures. National claims data from large US insurers are valuable assets in health services in AYA oncology. Continuous linkage of such data to other publicly available data sets will enhance their usability in future research.

Foundations, collaborations, and advocacy: fueling innovation in data science. The final session of the symposium focused on how advocacy groups and foundations partner with researchers to further innovation in data science approaches to pediatric cancer research. Often, these innovations are sparked by team science approaches, whether in workshops, hackathons, or collaborations with nationwide or international partners. Dr Casey Greene (Alex's Lemonade Stand Foundation's Childhood Cancer Data Lab [CCDL][86] and the University of Pennsylvania) discussed how the CCDL has developed workshops that aim to empower pediatric cancer researchers to analyze their own data. These workshops are inspired by Data Carpentry[87] and others.[88,89] The workshops developed to date have a modular architecture so that they can be rearranged to focus on topics of interest to participants at a specific institution or meeting.[90] The CCDL was launched by Alex's

Lemonade Stand Foundation to empower pediatric cancer researchers poised for the next big discovery with the knowledge, data, and tools to reach their goals, and the workshops are a critical part of the program. In the years ahead, it will be important to implement, evaluate, and improve training efforts as we aim to maximize the value returned on openly shared data.[91]

Dr Jack DiGiovanna (Seven Bridges) described an international academic-commercial partnership to create a valuable pediatric cancer resource. The ZERO Childhood Cancer Program[92] is a world-leading personalized medicine program that supports approximately 200 new cases of high-risk pediatric cancer in Australia annually. ZERO's diagnosis and treatment recommendations are partially based on outlying gene expression. Aggregating Australian data with global data would help better localize new patients on the transcriptome landscape and develop strategies to more effectively treat high-risk childhood cancers. Here, Seven Bridges, in a collaborative partnership between ZERO, Data Driven Discovery in Biomedicine at Children's Hospital of Philadelphia, and the Australian Bioinformatics Commons, created an internationally federated computational infrastructure and increased the number of available research samples by approximately 8×. This consortium used Cavatica—a mature and widely used genomics analysis platform underpinning the Gabriella Miller Kids First Data Resource to enable efficient harmonized analyses across geographically separated and jurisdictionally protected data resources. Specifically, multicloud functionality allowed computation in the same region where the data were stored, in Sydney (ZERO) or Northern Virginia (Children's Brain Tumor Tissue Consortium [CBTTC]).[93] Common Workflow Language workflows were used interchangeably across both locations, aggregating both data sets into a single, virtual, pan-continental data set accessible through the Cavatica analysis platform. Gene expression data were clustered to identify brain cancer subtypes. Gene expression was calculated with official Kids First RNA-Seq workflow on ZERO samples. By combining these ZERO gene expression data with 8× more CBTTC samples, the clustering precision should be improved. A future step, in the case of previously unidentified brain cancer subtypes, would be to probe molecular drivers. Toward that end, the ZERO somatic structural variant and copy number variation caller were optimized in Common Workflow Language 1.0 such that it is ready to process both ZERO and CBTTC whole genome sequencing data. Overall, this has created access to the largest pediatric cancer data resource ever generated in Australia, immediately increasing samples available to Australian researchers over eight-fold, without additional recruitment or sequencing costs.

Dr Adam Resnick (Children's Hospital of Philadelphia) closed the symposium with a presentation on powering science in pediatric cancer. Pediatric cancers comprise a heterogeneous and often complex rare disease landscape of data-driven discovery and clinical translation with diverse, asynchronous sources of data generation, various loci of analysis and robust clinical trial efforts spanning multiple stakeholder communities, data modalities, and infrastructures. To address emerging data coordination and data sharing challenges, as well as harness what might be unique opportunities for convergence research underpinned by collaborative and interconnected rare disease communities, a number of recently launched initiatives have piloted new modes of federating resources and data commoning through cloud-based platforms, repositories, portals, and real-time analytics.[94] Initiatives like the NIH-supported Gabriella Miller KFDRC have piloted both data and infrastructure interoperability across the pediatric landscape through platform-based connectivity that empowers cross-disease discovery across the developmental biology context of pediatric cancer and structural birth defects, focusing on the altered germline disease setting as informed by whole-genome sequencing. Along with efforts that intersected with the NCI's Genomic Data Commons and Therapeutically Applicable Research To Generate Effective Treatments data sets, KFDRC piloted consortia-based efforts (including the Children's Brain Tumor Network[95] and Pacific Pediatric Neuro-Oncology Consortium,[96] both focused on pediatric brain tumors) that leveraged the KFDRC discovery platform coordination, Cavatica cloud compute environments, and PedcBioportal's knowledgebase to support the more than 40,000 biospecimens and associated clinical, genomic, and assay-based resources spanning $>$ 25 institutions on behalf of advancing coordinated data commons practices and workflows that seek to meet the translational gap for what is the leading cause of disease-related death in children in the United States.[97] Such efforts have preliminarily defined the potential value of a shared, federated data-driven infrastructure that spans the NIH, consortia, and institutional ecosystem through joint initiatives and publications.[98] Key to the success of these efforts is the need to address both the unique disease context of any one specific pediatric cancer histology or subtype and its domain expert community, while also harnessing cross-disease and cross-age continuums of research on behalf of both vertical and horizontal convergence pediatric research. Together, these combined previous NIH-funded initiatives and existing community and institutional resources set the stage for a shovel-ready ecosystem poised for clinical transformation and patient impact through programs like the NCI CCDI that focuses on data coordination and federation efforts that harness the full potential of real-time data flows and clinical translation while building tomorrow's pediatric cancer translational ecosystem on behalf of accelerated impact for children.

PLENARY AND CORPORATE KEYNOTES

In addition to the presentations outlined above, Dr Michael Teitell, Cancer Center Director and Chief of Pediatric and Neonatal Pathology at UC Los Angeles Jonsson

Comprehensive Cancer Center (JCCC), presented a Keynote Lecture on Data Science Opportunities at a Matrix Cancer Center. Dr Teitell began by describing the setting of the JCCC in the most populous and diverse county in the United States and discussing the landscape of health disparities in the context of the largest number of cancer cases in any US county. He provided an overview of JCCC's six research programs across basic, clinical, translational, and population science. Describing work from his own research laboratory on tumor heterogeneity and pathology sampling, Dr Teitell presented innovations in live cell interference microscopy, which can calculate the biomass of a cell to identify dynamics of tumor-killing cytotoxic T-lymphocytes and examine drug response profiles. He discussed the JCCC oncology clinical trials network, which has global reach, generating large data sets with repeated success in gaining FDA approval for treatments in breast, prostate, and other cancers. Dr Teitell provided an overview of a UC-wide strategic priority to create a UC Cancer Consortium among the five UC campuses having medical centers. They have made significant progress in cancer genomics data integration, having performed more than 30,000 genomic analyses across the five campuses. A data sharing pilot is underway with the goal of creating a UC Cancer Center–wide database linked with multiple data types from a very diverse population for decision support and research purposes. These efforts will allow the consortium to emphasize precision medicine approaches to treatment, enable quality assessment for driving best practices, and create opportunities to collaborate with industry.

The symposium also featured three corporate keynotes. David Fenstermacher (DNAnexus)[99] presented Creating an Oncology-Focused Molecular Precision Medicine Hub. Ian Maurer (GenomOncology)[100] presented Extracting Clinical Information from Unstructured Documents at Scale. Finally, Cassandra Wesselman and Jean Lozach (OnRamp Bioinformatics)[101] presented Collaboration in the Cloud: Empowering the Oncology Research Community.

CHILDHOOD CANCER DATA SCIENCE: LEARNING FROM EVERY CHILD

The 15th Cancer Informatics for Cancer Centers Symposium highlighted critical initiatives in pediatric and AYA cancer research that rely on data science and informatics to further our understanding of disease and facilitate progress toward novel treatments. With lower overall rate of incidence and a wide spectrum of representative disease subtypes, this particular population of young patients with cancer embodies an exemplary model to optimize data sharing policies and practices that inform and guide precision oncology. With an aim to learn from every child with cancer, the NCI is leading efforts to develop a set of strategies to (1) gather data and comprehensively report on every patient, survivor, and their families diagnosed with childhood cancers in the United States; (2) establish a platform to federate, aggregate, and integrate data from all relevant NCI-supported and community-based childhood and AYA data resources, and (3) offer appropriate clinical and molecular characterization to optimally treat every child with cancer.

The initiatives described throughout the conference showcase the trajectory of the pediatric and AYA cancer community through all facets of research and clinical care. It will be important to consider how ongoing and new efforts can extend the important work being done, from basic research to population studies to clinical treatment, to reach each patient and fill gaps in knowledge. To facilitate discovery, NCI is establishing a framework for an ecosystem of data repositories and registries, both within the government and the larger cancer research and care community, that can bring together a variety of data collected and make them available through an interoperable and sustainable network. An initial phase of a Childhood Molecular Characterization Protocol is being defined to determine how clinical sequencing in DNA and RNA, along with epigenomic characterization and rich phenomic annotation, should be optimized for both treatment and subsequent disease research. Data science and informatics will play an increasingly vital role in the utility and understanding of these data to meet and adapt to the needs of pediatric and AYA cancer community.

The culture of research is evolving, and data are being used increasingly to generate novel hypotheses in addition to validating published results. Rare cancers with few cases are particularly reliant on high-quality data to allow maximal discovery with precious tissue reserves. Collaborative team science and broad open access to research and clinical care data are crucial to building enough power to drive innovation and find answers to improve treatments for young patients and survivors. The NCI is committed to working closely with Cancer Centers and the broader pediatric and AYA cancer community, both nationally and internationally, to maximize the success of its data initiatives. The CCDI will leverage these partnerships to benefit all pediatric and AYA activities supported through NCI and to lay a foundation of data-driven progress for all types of cancer research and care.

AFFILIATIONS

[1]Center for Biomedical Informatics and Information Technology, National Cancer Institute, Rockville, MD
[2]Huntsman Cancer Institute and University of Utah, School of Medicine, Salt Lake City, UT
[3]National Cancer Institute, Bethesda, MD
[4]Maternal and Child Health Research Institute, Stanford School of Medicine, Stanford, CA
[5]Center for Cancer Research, National Cancer Institute, Bethesda, MD
[6]Center for Drug Evaluation and Research, Food and Drug Administration, Bethesda, MD

[7]Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD
[8]Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA
[9]University of Kentucky, Markey Cancer Center, Lexington, KY
[10]Massachusetts General Hospital, Boston, MA
[11]Children's Hospital of Philadelphia, Philadelphia, PA
[12]University of Chicago, Chicago, IL
[13]Cincinnati Children's Hospital, Cincinnati, OH
[14]St Jude Children's Research Hospital, Memphis, TN
[15]University of California, Santa Cruz, Santa Cruz, CA
[16]University of Southern California, Keck School of Medicine, Los Angeles, CA
[17]The Hospital for Sick Children, Toronto, Canada
[18]University of Minnesota, Masonic Cancer Center, Minneapolis, MN
[19]Kaiser Permanente Southern California, Los Angeles, CA
[20]Kaiser Permanente Washington Health Research Institute, Seattle, WA
[21]University of Colorado, Aurora, CO
[22]Seven Bridges, Boston, MA
[23]Essex Management, Rockville, MD
[24]City of Hope, Duarte, CA
[25]Duke University, Durham, NC

CORRESPONDING AUTHOR

Anthony R. Kerlavage, PhD, Center for Biomedical Informatics and Information Technology, National Cancer Institute, 9609 Medical Center Dr, Room 1W420, Rockville, MD 20850; e-mail: kerlavagear@nih.gov.

SUPPORT

Supported by the National Cancer Institute (R01CA151284: J.N.P., ZIABC011754: K.R., ZIABC011852: B. Widemann, R21CA205309: K.J.W., R01CA216354 and R01CA216391: J.Z., U10CA180899: T.A.A., U10CA180886: D. Hawkins, P30CA2014: University of Utah Huntsman Cancer Institute, 5U24CA220457-05: N. Schultz, P30CA177558: University of Kentucky Markey Cancer Center, and HHSN261201800013I/HHSN26100001: University of Kentucky—Kentucky SEER Program); the National Heart, Lung, and Blood Institute (5U2CHL138346-05: A.C.R.); Alex's Lemonade Stand Foundation (CCDL17-01293: C.G.); Huntsman Cancer Foundation (A.C.K.); Commonwealth of Australia (GFA-244: Smith); St Baldrick's Foundation Consortium Award (427053: O.V.); and AACR NextGen Award for Transformative Cancer Research (O.V.).

AUTHOR CONTRIBUTIONS

Conception and design: Anthony R. Kerlavage, Anne C. Kirchhoff, Norman E. Sharpless, Dennis Deapen, Eric B. Durbin, Samuel L. Volchenboum, Bruce J. Aronow, Olena Vaske, Todd A. Alonzo, Jenny N. Poynter, Erin E. Hahn, Karen J. Wernli, Casey Greene, Jack DiGiovanna, Adam C. Resnick, Sorena Nadaf, Warren A. Kibbe
Provision of study materials or patients: Anne C. Kirchhoff, Karlyne Reilly, Dennis Deapen, Allison P. Heath, Sorena Nadaf
Collection and assembly of data: Anthony R. Kerlavage, Jaime M. Guidry Auvil, Kara L. Davis, Karlyne Reilly, Gregory Reaman, Dennis Deapen, Amie Hwang, Eric B. Durbin, Sara L. Gallotto, Richard Aplenc, Allison P. Heath, Jinghui Zhang, Paul C. Nathan, Greg Armstrong, Karen J. Wernli, Adam C. Resnick, Eve R. Shalley, Sorena Nadaf
Data analysis and interpretation: Anthony R. Kerlavage, Kara L. Davis, Lynne Penberthy, Dennis Deapen, Amie Hwang, Eric B. Durbin, Richard Aplenc, Bruce J. Aronow, Jinghui Zhang, Karen J. Wernli, Sorena Nadaf
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by the authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.
Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Anne C. Kirchhoff
Stock and Other Ownership Interests: Medtronic

Kara L. Davis
Honoraria: Novartis
Research Funding: Jazz Pharmaceuticals

Karlyne Reilly
Consulting or Advisory Role: Saul Ewing LLC
Patents, Royalties, Other Intellectual Property: I hold a patent to a potential therapeutic: Reilly KM, Beutler JA, Turbyville T, Wiemer DF: The Natural Product Schweinfurthin A and Synthetic Schweinfurthin Analogs Specifically Inhibit Nf1-Null Cells and may be Useful as a Therapy for Neurofibromatosis Type 1. US patent 61/174,338, April 19, 2014; International patent PCT/US10/33153. There are no royalties or licensing fees from this patent at the time

Richard Aplenc
Expert Testimony: Vorys

Samuel L. Volchenboum
Stock and Other Ownership Interests: Litmus Health
Consulting or Advisory Role: Accordant
Travel, Accommodations, Expenses: Sanford Health

Bruce J. Aronow
Patents, Royalties, Other Intellectual Property: Patents issued for some data analysis algorithms related to data mining and discovery approaches to drug repositioning for new clinical indications

Olena Vaske
Employment: NantWorks
Stock and Other Ownership Interests: NantHealth

Casey Greene
Other Relationship: Alex's Lemonade Stand Foundation

Jack DiGiovanna
Employment: Biogen (I)
Stock and Other Ownership Interests: Biogen

Sorena Nadaf
This author is a member of the *JCO Clinical Cancer Informatics* Editorial Board. Journal policy recused the author from having any role in the peer review of this manuscript.

Warren A. Kibbe
This author is a member of the *JCO Clinical Cancer Informatics* Editorial Board. Journal policy recused the author from having any role in the peer review of this manuscript.

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

We thank all participants in this symposium for their presentations and engaging discussion. We also thank Eve Shalley and Tim Adamich for their excellent work in organization and execution of the meeting. We acknowledge the contribution of Lauren Sanders, PhD (University of California Santa Cruz), to the Treehouse Initiative summary.

REFERENCES

1. Barnholtz-Sloan J, Rollison D, Basu A, et al: Cancer Informatics for Cancer Centers (CI4CC): Building a community focused on sharing ideas and best practices to improve cancer care and patient outcomes. JCO Clin Cancer Inform 4:108-111, 2020
2. Childhood Cancer Data Initiative (CCDI). https://www.cancer.gov/research/areas/childhood/childhood-cancer-data-initiative
3. National Cancer Institute Board of Scientific Advisors: Data sharing opportunities in childhood, adolescent and young adult (AYA) cancer research for the National Cancer Institute. 2020. https://deainfo.nci.nih.gov/advisory/bsa/sub-cmte/CCDI/CCDI%20BSA%20WG%20Report_Final%20061620.pdf
4. National Cancer institute: My Pediatric and Adult Rare Tumor Network. https://www.cancer.gov/pediatric-adult-rare-tumor/
5. National Cancer institute: NCI-COG Pediatric MATCH. https://www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/pediatric-match
6. St. Jude Children's Research Hospital: The Childhood Cancer Survivor Study. https://ccss.stjude.org/
7. National Cancer Institute: Human Cancer Models Initiative. https://ocg.cancer.gov/programs/HCMI
8. RTI International: Pediatric Preclinical Testing Consortium. http://www.ncipptc.org/
9. National Cancer Institute: Pediatric Immunotherapy Network. https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/implementation/pediatric-immunotherapy-network
10. Davis KL, Fox E, Merchant MS, et al: Nivolumab in children and young adults with relapsed or refractory solid tumours or lymphoma (ADVL1412): A multicentre, open-label, single-arm, phase 1-2 trial. Lancet Oncol 21:541-550, 2020
11. Maude SL, Laetsch TW, Buechner J, et al: Tisagenlecleucel in children and young adults with B-cell lymphoblastic leukemia. N Engl J Med 378:439-448, 2018
12. Yu AL, Gilman AL, Ozkaynak MF, et al: Anti-GD2 antibody with GM-CSF, interleukin-2, and isotretinoin for neuroblastoma. N Engl J Med 363:1324-1334, 2010
13. Angelo M, Bendall SC, Finck R, et al: Multiplexed ion beam imaging of human breast tumors. Nat Med 20:436-442, 2014
14. van Groningen T, Koster J, Valentijn LJ, et al: Neuroblastoma is composed of two super-enhancer-associated differentiation states. Nat Genet 49:1261-1266, 2017
15. US Food and Drug Administration: Rare diseases: Natural history studies for drug development. 2019. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/rare-diseases-natural-history-studies-drug-development
16. Gross AM, Wolters PL, Dombi E, et al: Selumetinib in children with inoperable plexiform neurofibromas. N Engl J Med 382:1430-1442, 2020
17. Perlmutter J, Roach N, Smith ML: Involving advocates in cancer research. Semin Oncol 42:681-685, 2015
18. Del Rivero J: Natural history and biospecimen acquisition for children and adults with rare solid tumors. 2018. https://clinicaltrials.gov/ct2/show/NCT03739827
19. Gröbner S, Worst B, Weischenfeldt J, et al: The landscape of genomic alterations across childhood cancers. Nature 555:321-327, 2018
20. Ma X, Liu Y, Liu Y, et al: Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature 555:371-376, 2018
21. US Food and Drug Administration: Food and Drug Administration Reauthorization Act of 2017 (FDARA). 2017. https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/fda-reauthorization-act-2017-fdara
22. US Food and Drug Administration: Pediatric research equity ActIPREA. 2003. https://www.fda.gov/drugs/development-resources/pediatric-research-equity-act-prea
23. US Food and Drug Administration: Relevant pediatric molecular target list. 2020. https://www.fda.gov/media/120331/download
24. National Cancer Institute: National Childhood Cancer Registry. https://cancercontrol.cancer.gov/research-emphasis/childhood-cancer-registry
25. North American Association of Central Cancer Registries. https://www.naaccr.org/
26. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2020. CA Cancer J Clin 70:7-30, 2020
27. Tucker TC, Durbin EB, McDowell JK, et al: Unlocking the potential of population-based cancer registries. Cancer 125:3729-3737, 2019
28. National Cancer Institute: Surveillance, Epidemiology, and End Results Program. https://seer.cancer.gov/
29. Huang B, Luo A, Durbin EB, et al: Incidence of CNS tumors in Appalachian children. J Neurooncol 132:507-512, 2017
30. Durbin EB, Christian J, Badgett TC, et al: Informatics methods and infrastructure needed to study factors associated with high incidence of pediatric brain and central nervous system tumors in Kentucky. J Registry Manag 47:127-134, 2020
31. Pediatric Proton/Photon Consortium Registry. https://www.pediatricradiationregistry.org/
32. Lawell MP, Indelicato DJ, Paulino AC, et al: An open invitation to join the Pediatric Proton/Photon Consortium Registry to standardize data collection in pediatric radiation oncology. Br J Radiol 93:20190673, 2020
33. Mayo CS, Kessler ML, Eisbruch A, et al: The big data effort in radiation oncology: Data mining or data farming? Adv Radiat Oncol 1:260-271, 2016
34. Children's Hospital Association: Pediatric Health Information System. https://www.childrenshospitals.org/phis
35. Farber S, Diamond LK, Mercer RD, et al: Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid (aminopterin). N Engl J Med 238:787-793, 1948
36. Heyn RM, Brubaker CA, Burchenal JH, et al: The comparison of 6-mercaptopurine with the combination of 6-mercaptopurine and azaserine in the treatment of acute leukemia in children: Results of a cooperative study. Blood 15:350-359, 1960
37. O'Leary M, Krailo M, Anderson JR, et al: Progress in childhood cancer: 50 Years of research collaboration, a report from the Children's Oncology Group. Semin Oncol 35:484-493, 2008
38. Gamis AS, Alonzo TA, Meshinchi S, et al: Gemtuzumab ozogamicin in children and adolescents with de novo acute myeloid leukemia improves event-free survival by reducing relapse risk: Results from the randomized phase III Children's Oncology Group trial AAML0531. J Clin Oncol 32:3021-3032, 2014
39. Miller TP, Li Y, Kavcic M, et al: Accuracy of adverse event ascertainment in clinical trials for pediatric acute myeloid leukemia. J Clin Oncol 34:1537-1543, 2016
40. University of Chicago: Pediatric Cancer Data Commons. http://commons.cri.uchicago.edu/

41. National Cancer Institute: Cancer Research Data Commons. https://datacommons.cancer.gov/
42. Kids First Data Resource Center. https://kidsfirstdrc.org/
43. The Children's Hospital of Philadelphia: Cavatica. https://d3b.center/our-research/cavatica/
44. The Children's Hospital of Philadelphia: PedcBioPortal. https://d3b.center/our-research/pedcbioportal/
45. University of Chicago: Gen3 Data Commons. https://gen3.org/
46. Cincinnati Children's Hospital Medical Center: ToppCell. http://toppcell.cchmc.org
47. St. Jude Children's Research Hospital: St. Jude Cloud. https://www.stjude.cloud/
48. Vaske OM, Bjork I, Salama SR, et al: Comparative tumor RNA sequencing analysis for difficult-to-treat pediatric and young adult patients with cancer. JAMA Netw Open 2:e1913968, 2019
49. Newton Y, Rassekh SR, Deyell RJ, et al: Comparative RNA-sequencing analysis benefits a pediatric patient with relapsed cancer. JCO Precis Oncol 2, 2018 doi:10.1200/PO.17.00198
50. University of California Santa Cruz: Treehouse Childhood Cancer Initiative. https://treehousegenomics.soe.ucsc.edu/
51. Sanders LM, Rangaswami A, Bjork I, et al: Comparative RNA-seq analysis aids in diagnosis of a rare pediatric tumor. Cold Spring Harb Mol Case Stud 5:a004317, 2019
52. Liang L, Zhang Y, Malpica A, et al: Gliomatosis peritonei: A clinicopathologic and immunohistochemical study of 21 cases. Mod Pathol 28:1613-1620, 2015
53. Children's Oncology Group: Project:EveryChild. http://projecteverychild.org/
54. Greenberg ML, Barr RD, DiMonte B, et al: Childhood cancer registries in Ontario, Canada: Lessons learned from a comparison of two registries. Int J Cancer 105:88-91, 2003
55. Baxter NN, Daly C, Gupta S, et al: The initiative to maximize progress in adolescent and young adult cancer therapy (IMPACT) cohort study: A population-based cohort of young Canadians with cancer. BMC Cancer 14:805, 2014
56. Pediatric Oncology Group of Ontario: POGONIS – Childhood Cancer Database. https://www.pogo.ca/research-data/pogonis-childhood-cancer-database/
57. Wikipedia: Health informatics. https://en.wikipedia.org/wiki/Health_informatics
58. Wikipedia: Publicly funded health care. https://en.wikipedia.org/wiki/Publicly_funded_health_care
59. Gupta S, Pole JD, Baxter NN, et al: The effect of adopting pediatric protocols in adolescents and young adults with acute lymphoblastic leukemia in pediatric vs adult centers: An IMPACT cohort study. Cancer Med 8:2095-2103, 2019
60. Khanna A, Pequeno P, Gupta S, et al: Increased risk of all cardiovascular disease subtypes among childhood cancer survivors: Population-based matched cohort study. Circulation 140:1041-1043, 2019
61. Furzer J, Gupta S, Nathan PC, et al: Cost-effectiveness of tisagenlecleucel vs standard care in high-risk relapsed pediatric acute lymphoblastic leukemia in Canada. JAMA Oncol 6:393-401, 2020
62. Gupta S, Nathan PC, Baxter NN, et al: Validity of administrative data in identifying cancer-related events in adolescents and young adults: A population-based study using the IMPACT cohort. Med Care 56:e32-e38, 2018
63. Steele JR, Wellemeyer AS, Hansen MJ, et al: Childhood cancer research network: A North American Pediatric Cancer Registry. Cancer Epidemiol Biomarkers Prev 15:1241-1242, 2006
64. Musselman JR, Spector LG, Krailo MD, et al: The Children's Oncology Group Childhood Cancer Research Network (CCRN): Case catchment in the United States. Cancer 120:3007-3015, 2014
65. Williams LA, Pankratz N, Lane J, et al: Klinefelter syndrome in males with germ cell tumors: A report from the children's oncology group. Cancer 124:3900-3908, 2018
66. Williams LA, Mills L, Hooten AJ, et al: Differences in DNA methylation profiles by histologic subtype of paediatric germ cell tumours: A report from the Children's Oncology Group. Br J Cancer 119:864-872, 2018
67. Robison LL, Armstrong GT, Boice JD, et al: The Childhood Cancer Survivor study: A National Cancer Institute-supported resource for outcome and intervention research. J Clin Oncol 27:2308-2318, 2009
68. Leisenring WM, Mertens AC, Armstrong GT, et al: Pediatric cancer survivorship research: Experience of the Childhood Cancer Survivor Study. J Clin Oncol 27:2319-2327, 2009
69. Morton LM, Sampson JN, Armstrong GT, et al: Genome-wide association study to identify susceptibility loci that modify radiation-related risk for breast cancer after childhood cancer. J Natl Cancer Inst 109:djx058, 2017
70. Armstrong GT, Chen Y, Yasui Y, et al: Reduction in late mortality among 5-year survivors of childhood cancer. N Engl J Med 374:833-842, 2016
71. Oeffinger KC, Ford JS, Moskowitz CS, et al: Promoting breast cancer surveillance: The EMPOWER study, a randomized clinical trial in the Childhood Cancer Survivor study. J Clin Oncol 37:2131-2140, 2019
72. Huntsman Cancer Institute: Utah Population Database. https://uofuhealth.utah.edu/huntsman/utah-population-database/
73. Huntsman Cancer Institute: Utah Cancer Registry. https://uofuhealth.utah.edu/utah-cancer-registry/
74. Curtin K, Smith KR, Fraser A, et al: Familial risk of childhood cancer and tumors in the li-fraumeni spectrum in the Utah Population Database: Implications for genetic evaluation in pediatric practice. Int J Cancer 133:2444-2453, 2013
75. Kirchhoff AC, Fluchel MN, Wright J, et al: Risk of hospitalization for survivors of childhood and adolescent cancer. Cancer Epidemiol Biomarkers Prev 23:1280-1289, 2014
76. Anderson C, Kaddas HK, Ou JY, et al: Hospitalization after adolescent and young adult (AYA) cancer: A population-based study in Utah. Cancer Epidemiol Biomarkers Prev 29:336-342, 2020
77. Ou JY, Hanson HA, Ramsay JM, et al: Fine particulate matter air pollution and mortality among pediatric, adolescent, and young adult cancer patients. Cancer Epidemiol Biomarkers Prev 29:1929-1939, 2020
78. Chao C, Chiu V, Mueller LA, et al: Exploring the feasibility of establishing a retrospective cohort of survivors of adolescent and young adult cancer to study long-term health outcomes in an integrated managed care environment. J Adolesc Young Adult Oncol 2:59-65, 2013
79. Chao C, Xu L, Bhatia S, et al: Cardiovascular disease risk profiles in survivors of adolescent and young adult (AYA) cancer: The Kaiser Permanente AYA Cancer Survivors study. J Clin Oncol 34:1626-1633, 2016
80. Chao C, Bhatia S, Xu L, et al: Chronic comorbidities among survivors of adolescent and young adult cancer. J Clin Oncol 38:3161-3174, 2020
81. Hahn EE, Wu YL, Munoz-Plaza CE, et al: Use of recommended post-treatment services for adolescent and young adult survivors of Hodgkin lymphoma. Cancer 125:1558-1587, 2019
82. Centers for Disease Control and Prevention: National Death Index. https://www.cdc.gov/nchs/ndi/index.htm

83. Mack JW, Cannavale K, Sattayapiwat O, et al: Care in the final month of life among adolescent and young adult cancer patients in Kaiser Permanente Southern California. J Palliat Med 19:1136-1141, 2016
84. Mack JW, Chen K, Boscoe FP, et al: High intensity end-of-life care among adolescent and young adult cancer patients in New York state Medicaid program. Med Care 53:1018-1026, 2015
85. Mack JW, Chen LH, Cannavale K, et al: End-of-life care intensity among adolescent and young adult patients with cancer in Kaiser Permanente Southern California. JAMA Oncol 1:592-600, 2015
86. Alex's Lemonade Stand Foundation: Childhood Cancer Data Lab. https://www.ccdatalab.org/
87. Teal TK, Cranston KA, Lapp H, et al: Data carpentry: Workshops to increase data literacy for researchers. Int J Digit Curat 10, 2015
88. Taroni JN: Making workshops work: Insights from EDAMAME. mSystems 4:e00467-19, 2019
89. Shade A, Dunivin TK, Choi J, et al: Strategies for building computing skills to support microbiome analysis: A five-year perspective from the EDAMAME workshop. mSystems 4:e00297-19, 2019
90. AlexsLemonade/training-modules. https://github.com/alexslemonade/training-modules
91. Byrd JB, Greene AC, Prasad DV, et al: Responsible, practical genomic data sharing that accelerates research. Nat Rev Genet 21:615-629, 2020
92. Children's Cancer Institute and Kids Cancer Centre at Sydney Children's Hospital: Zero Childhood Cancer. https://www.zerochildhoodcancer.org.au/
93. The Children's Hospital of Philadelphia: Children Brain Tumor Tissue Consortium. https://www.chop.edu/clinical-trial/cbttc-collection-protocol
94. Hess C, Ostrom E: Understanding Knowledge as a Commons: From Theory to Practice. Cambridge, MA, MIT Press, 2007
95. Children's Brain Tumor Network. https://cbtn.org/
96. Pacific Pediatric Neuro-Oncology Consortium. https://pnoc.us/
97. Surveillance, Epidemiology, and End Results Program: Annual Report to the Nation 2020: Overall Cancer Statistics. National Cancer Institute, 2020. https://seer.cancer.gov/report_to_nation/statistics.html
98. Petralia F, Tignor N, Reva B, et al: Integrated proteogenomic characterization across major histological types of pediatric brain cancer. Cell 183:1962-1985, 2020
99. DNAnexus. https://www.dnanexus.com/
100. GenomOncology. https://www.genomoncology.com/
101. ROSALIND, Inc. https://www.onramp.bio/

