

**Data Sharing Opportunities in Childhood, Adolescent and Young Adult (AYA) Cancer  
Research for the National Cancer Institute**

Report of the Board of Scientific Advisors on the Childhood Cancer Data Initiative (CCDI)

June 15, 2020

**NATIONAL INSTITUTES OF HEALTH  
National Cancer Institute  
Board of Scientific Advisors**

***Ad Hoc* Working Group in Support of the Childhood Cancer Data Initiative**

**CO-CHAIR**

**Kevin M. Shannon, M.D.**

American Cancer Society Research Professor,  
Roma and Marvin Auerback Distinguished Professor of Molecular Oncology  
Professor, Department of Pediatrics  
School of Medicine  
University of California, San Francisco  
San Francisco, California

**CO-CHAIR**

**Otis W. Brawley, M.D., M.A.C.P., F.A.S.C.O., F.A.C.E.**

Bloomberg Distinguished Professor of Oncology and Epidemiology  
The Sidney Kimmel Comprehensive Cancer Center  
Johns Hopkins University  
Baltimore, Maryland

**MEMBERS**

**Peter C. Adamson, M.D.**

Global Head  
Oncology Development and Pediatric Innovation  
Sanofi  
Cambridge, Massachusetts

**Tom Curran, Ph.D.**

Senior Vice President, Executive Director,  
and Chief Scientific Officer  
Children's Research Institute  
Donald J. Hall Eminent Scholar in  
Pediatric Research  
Children's Mercy Kansas City  
Kansas City, Missouri

**James R. Downing, M.D.**

President and Chief Executive Officer  
St. Jude Children's Research Hospital  
Memphis, Tennessee

**Julie Guillot**

Volunteer  
Leukemia & Lymphoma Society  
Chair, Pediatric Partnerships & Outreach  
Founder, Target Pediatric AML Project  
Former Tech Executive  
Cisco/NetSolve, Dell, Accenture  
Austin, Texas

**Amanda Haddock**

President  
Dragon Master Foundation  
Kechi, Kansas

**Douglas S. Hawkins, M.D.**

Chair, Children's Oncology Group  
Professor of Pediatrics  
Division of Hematology/Oncology  
Center for Clinical and Translational Research  
Seattle Children's Hospital  
Seattle, Washington

**Katherine A. Janeway, M.D.**  
Director of Clinical Genomics  
**Dana-Farber Cancer Institute**  
Assistant Professor of Pediatrics  
Harvard Medical School  
Boston, Massachusetts

**Andrea Hayes-Jordan, M.D., F.A.C.S, F.A.A.P.**  
Byah Thompson Doxey Distinguished Professor  
of Surgery  
Division Chief, Pediatric Surgery  
Surgeon-in-Chief  
University of North Carolina Children's Hospital

**Warren Kibbe, Ph.D.**  
Chief  
Translational Biomedical Informatics  
Department of Biostatistics and Bioinformatics  
Chief Data Officer  
Duke Cancer Institute  
Duke University School of Medicine  
Durham, North Carolina

**Andrew L. Kung, M.D., Ph.D.**  
Chairman and Professor  
Department of Pediatrics  
Memorial Sloan Kettering Cancer Center  
New York, New York

**John M. Maris, M.D.**  
Giulio D'Angio Chair in Neuroblastoma Research  
Professor of Pediatrics  
Department of Pediatrics  
Children's Hospital of Philadelphia  
University of Pennsylvania  
Philadelphia, Pennsylvania

**Samuel L. Volchenboum, M.D., Ph.D.**  
Associate Professor of Pediatrics  
Dean, Master's Education  
Associate Chief Research Informatics Officer  
Associate Director, Institute for Translational  
Medicine  
University of Chicago  
Chicago, Illinois

### **Ex Officio Members**

**Stephn J. Chanock, M.D.**  
Director  
Division of Cancer Epidemiology & Genetics  
National Cancer Institute  
National Institutes of Health  
Bethesda, Maryland

**Anthony Kerlavage, Ph.D.**  
Director  
Center for Biomedical Informatics and  
Information Technology  
National Cancer Institute  
National Institutes of Health  
Bethesda, Maryland

**Lynne Penberthy, M.D., M.P.H.**  
Associate Director  
Surveillance Research Program  
Division of Cancer Control and Population  
Sciences  
Office of the Director  
National Cancer Institute  
National Institutes of Health  
Bethesda, Maryland

**Gregory H. Reaman, M.D.**  
Associate Director for Oncology Sciences  
Office of Hematology and Oncology Products  
Center for Drug Evaluation and Research  
U.S. Food & Drug Administration  
Silver Spring, Maryland

**Malcolm Smith, M.D.**  
Associate Branch Chief for Pediatric Oncology  
Clinical Investigations Branch  
Cancer Therapy Evaluation Program  
Division of Cancer Treatment and Diagnosis  
National Cancer Institute  
National Institutes of Health  
Bethesda, Maryland

**Brigitte V. Widemann, M.D.**  
Chief  
Pediatric Oncology Branch  
Center for Cancer Research  
National Cancer Institute  
National Institutes of Health  
Bethesda, Maryland

### **Executive Secretary**

**Jaime M. Guidry Auvil, Ph.D.**  
Director  
Office of Data Sharing  
Center for Biomedical Informatics and  
Information Technology  
National Cancer Institute  
National Institutes of Health  
Bethesda, Maryland

## Table of Contents

<b>Executive Summary .....</b>	<b>2</b>
<b>Introduction.....</b>	<b>4</b>
<b>Types of Data for Collection and Aggregation.....</b>	<b>6</b>
<b>Current Landscape of Childhood/AYA Cancer Research &amp; Needs Analysis.....</b>	<b>9</b>
<b>Potential Barriers to Progress.....</b>	<b>14</b>
<b>Generating New Data .....</b>	<b>16</b>
<b>Distinction Between Research and Clinical Data.....</b>	<b>18</b>
<b>Engaging Diverse Array of Stakeholders for Input.....</b>	<b>20</b>
<b>Potential Opportunities for Transformative Discoveries .....</b>	<b>22</b>

## Executive Summary

In November 2019, Dr. Norman Sharpless charged the Board of Scientific Advisors (BSA) *Ad Hoc* Working Group (WG) in Support of the Childhood Cancer Data Initiative (CCDI) to provide general guidance to the National Cancer Institute (NCI) related to the development and implementation of the CCDI to establish more efficient ways to share and use childhood cancer data to improve outcomes for children who develop cancer and to better understand the biology of childhood cancers. Because childhood cancers are rare diseases, strategies for increasing scientific understanding of childhood cancers can become models for the entire cancer research community. Data aggregation will lead to answers that can positively affect all children with cancer; therefore, the NCI envisions that the CCDI will develop specific and robust goals to facilitate optimal data sharing. Efforts within the CCDI will include support for the identification of novel therapeutic targets and approaches to advance new drug development. The WG was further charged with how to best advance CCDI's areas of focus and might recommend other research activities to enhance the Initiative's efforts.

The overall purpose of this WG report is to advise the NCI in implementing the CCDI so that this new resource will enable broad, rapid data sharing in ways that will optimally facilitate childhood, adolescent and young adult (AYA) cancer research and accelerate the development of better and less toxic therapies for the benefit of pediatric and AYA cancer patients and their families. This overriding goal is foremost to the members of the WG and we believe it should underpin all administrative and organizational decisions.

This WG report addresses the following broad thematic areas in pediatric and AYA cancer research in individual sections: (1) Types of Data for Collection/Aggregation; (2) Landscape of Pediatric/AYA Cancer Research Data and Needs Analysis; (3) Potential Barriers to Progress; (4) Generating New Data; (5) Research versus Clinical Data; (6) Patients, Families, and Advocates; and (7) Potential Opportunities for Transformative Discoveries. Specific recommendations relative to each subject area are presented in the individual sections of this report.

We highlight here some key over-arching recommendations drawn from the more complete list presented in the report for consideration by the NCI.

1. We recommend that the NCI commit to support ongoing data sharing over time through policy and funding. The WG identified a lack of dedicated financial support to personnel, time and other resources needed to implement effective data sharing as a substantial barrier to achieving the overall goals of CCDI.
2. We recommend that the NCI perform a comprehensive inventory of data, databases and shared research infrastructures with each assessed for quality, relevance and integration feasibility to better understand the landscape of pediatric and AYA cancer and survivorship data.
3. We recommend that the CCDI aggregate six broad categories of data: (1) clinical, treatment, and outcome data from clinical trials and the electronic health record (EHR); (2) molecular data including research sequencing (i.e. genomic, epigenetic and proteomic data) and clinical molecular profiling; (3) information regarding the availability and location of archived biospecimens, including germline and tumor DNA; (4) longitudinal population data from patients and survivors of pediatric and AYA cancers; (5) characteristics of cell line, patient derived xenograft (PDX), and genetically engineered mouse (GEM) models of pediatric and AYA cancers; and (6) any existing preclinical data generated from studies performed in these models.

4. We recommend that the NCI implement strategies to make all CCDI data as representative as possible of the full spectrum of pediatric/AYA cancer patients in the United States (U.S.). Recruitment targets should reflect the percentage of minorities in the U.S. population and to achieve this, we recommend the NCI include resources for persons obtaining consent to communicate and educate in a racially and ethnically sensitive manner.
5. We recommend convening experts to develop consensus guidelines addressing what clinical/molecular characterization and sample archiving should be performed and when they should be performed for each pediatric and AYA cancer diagnosis. The WG suggests that this group consider both clinically actionable sequencing to inform diagnosis and therapy and discovery analysis for research purposes. Finally, the NCI should consider potential ways of paying for molecular testing that is not currently funded by insurance or other sources. Addressing these challenging issues is central to the achieving the CCDI's mission of learning from every child with cancer.
6. We recommend that the NCI develop a long-term strategy for tracking pediatric and AYA cancer patients over many decades across their lifespan. During this time, they will become independent adults and transition their lives (and care) many times. Assigning a universal, anonymized and privacy-protected unique patient identifier to link clinical information, biospecimens, and any molecular data over time is a particularly appealing solution to this problem that should be strongly considered.

The report will conclude with some potential opportunities for the NCI to consider implementing through the CCDI that the WG feels could lead to transformative discoveries in pediatric and AYA cancer research and survivorship.

## Introduction

The Childhood Cancer Data Initiative (CCDI) was announced last year during the President's State of the Union address as a proposed federal investment of \$50 million per year, for the next 10 years, to support cancer research in patients and survivors of childhood, adolescent and young adult (AYA) malignancies. Given the amount of appropriated annual funds, it was decided CCDI would focus on the critical need to collect, analyze, and share data better in order to maximize NCI's ongoing investment in pediatric and AYA cancers and survivorship. Tissue samples from patients with cancer in this age group are critically limited and a valuable resource. The overall goal of the CCDI is not simply to generate more data, but to build processes that transform data into knowledge that moves the field forward in meaningful ways. The CCDI supports the wider pediatric cancer community's goal of maximizing the stated goal of "*learning from every patient*" so that ultimately those patients, survivors and their families can materially benefit in terms of higher cure rates and improved long-term health outcomes.

Approximately 16,000 children and adolescents from birth to age 15 years are diagnosed with cancer annually in the United States. This relatively small number of patients does not account for the large impact of pediatric and AYA cancers on the expected lifespan, particularly given the substantial proportion of adult cancers arising in individuals who are >70 years old. While 84% of children with cancer survive 5 years or more and are cured of their primary malignancy, a substantial proportion of patients and survivors experience significant long-term health problems as a result of the cancer itself or the treatment. Indeed, many studies have shown that the life expectancy of childhood cancer survivors is substantially lower than the general population, due to treatment-related co-morbidities. These findings underscore the need to develop more effective as well as less toxic therapies for pediatric and AYA cancers. If we could reduce the damage to the immature nervous and cardiovascular systems, and to other organs, we could reduce the frequency and severity of adverse late health effects. It should also be noted that many pioneering cancer discoveries such as curative combination chemotherapy, bone marrow transplantation, molecular targeted therapies and cellular immunotherapies were first developed in the small pediatric/AYA population, but these discoveries go on to benefit all cancer patients. Thus, investments in pediatric/AYA cancer research and learnings from those patient communities can speed progress for cancer patients of all ages.

Recent advances in molecular medicine have identified key driver mutations in virtually every pediatric/AYA cancer and have also identified unique biologic subtypes within different malignancies. This is perhaps most evident in acute lymphoblastic leukemia (ALL), which is the most common pediatric cancer. While the overall cure rate in pediatric ALL is now ~90%, deep genomic analyses have further characterized subtypes of ALL with a high rate of relapse and poor prognosis. Importantly, genomic profiling of those ALL subtypes also uncovered actionable drug targets in some patients who were otherwise lacking potentially curative treatment options. Despite the high cure rates for many pediatric/AYA malignancies, cancer remains the leading cause of disease-related death in children. Furthermore, the outcomes for children with certain "high risk" cancers (e.g., most sarcomas and some brain tumors) have not improved substantially over the past two decades, despite intensification in chemoradiotherapy. Indeed, the outcomes for some of these cancers have not changed since they were first identified and classified accurately.

In comparison to adult cancers, where lifestyle and other environmental factors often contribute significantly, pediatric/AYA cancer is far less common on a population-wide basis and inherited factors play a larger role. In addition, pediatric and AYA patients are treated in diverse clinical settings across the nation, including many institutions that see few cases each year. With this said, there is a remarkable history of successfully conducting national and international collaborative clinical trials through the NCI-funded Children's Oncology Group (COG) and other consortia. The relatively small number of relapsed pediatric cancers is an additional challenge that is being addressed, in part, through networks that are conducting phase I/II trials

across different tissue histologies. These efforts include the NCI Pediatric Molecular Analysis for Therapy Choice (MATCH) precision trial and the development of focused collaborative efforts for testing new therapies in leukemia, neuroblastoma, brain tumors, and some other pediatric/AYA cancers.

Although genome-wide analyses have elucidated the broad molecular landscape of most pediatric and AYA cancers, there is no national “standard of care” for molecular profiling at diagnosis or relapse and no consensus regarding the optimal type(s) of analyses that should be performed. Although it is estimated that some form of molecular testing is currently performed on approximately half of all pediatric cancers, there is a need for more data in this area. In addition to representing a substantial barrier to achieving the CCDI’s stated goal of “learning from every child with cancer”, uneven access to modern molecular analyses raises key questions about health equity and access to care.

In addressing the mission of the CCDI, the WG identified other distinctive aspects of pediatric/AYA cancers including: (1) lack of substantial industry support for data collection and aggregation; (2) the potential value of performing extensive preclinical testing of new agents across different models given the small number of patients available for clinical trials; (3) the positive role of engaged and effective advocates in promoting key initiatives such as The Childhood Cancer Survivorship, Treatment, Access & Research (STAR) Act of 2018 and Research Acceleration for Cure and Equity (RACE) for Children Act of 2017; (4) the importance of longitudinal follow-up and the many challenges in tracking survivors of pediatric/AYA cancers; (5) the need to educate patients and families (especially the underserved and minority) regarding the importance of participating in the CCDI and related tissue and blood acquisition studies; and, (6) key issues around privacy, particularly given the implications of discovering germline mutations in cancer predisposition genes for patients, siblings, and other family members.

The WG considered the following general questions when preparing this report:

- 1) What specific types of data should ideally be brought together within the CCDI?
- 2) What data currently exists, where is it housed, and what are the best strategies for bringing it together (landscape and needs analysis)?
- 3) What are the key barriers to aggregating these existing data?
- 4) What new data will enhance the CCDI and how might this be generated?
- 5) Are there specific practical considerations for encouraging or implementing robust data sharing?
- 6) What tools and systems need to be created and updated to utilize shared data to improve childhood cancer outcomes?
- 7) What is the perspective of the families of pediatric cancer patients and how can they best contribute to advancing the goals of the CCDI?
- 8) What key scientific questions can the CCDI address over the next 2-3 years?

This report is organized into seven sections that outline the consensus opinions of the WG and conclude with one or more specific recommendations for consideration by the NCI. The intent is to assist the NCI in implementing the CCDI so that it will rapidly advance discovery in order to accelerate the development of better and less toxic therapies for the benefit of pediatric/AYA cancer patients and their families. This overriding goal is foremost to the members of the WG and should underpin all administrative and organizational decisions.

## I. Types of Data for Collection or Aggregation

The WG first discussed the general types of data that would ideally be aggregated by the CCDI and accessible through this new NCI resource. More specific topics such as the need to systematically inventory what data currently exist and in what form(s), operational strategies for coordinating with other NCI and non-NCI resources, and potential barriers are discussed later in this report. The WG identified three broad categories of data for possible inclusion in the CCDI: (1) clinical, treatment, and outcome data from individual patients; (2) the availability of biospecimens, including germline and tumor DNA and the results of any molecular profiling; and, (3) resources for preclinical drug testing and data generated from studies performed in cell line, patient derived xenograft (PDX), and genetically engineered mouse (GEM) models of pediatric and AYA cancers. Integrating these categories of data will result in a robust, multi-modal dataset that will empower both traditional research and additional artificial intelligence efforts. The linking of molecular data to clinical outcomes data, as well as aggregation and generation of data to support synthetic control arms, is of critical importance and is a key enabler to advancing precision medicine. A key point of emphasis for the WG is the perceived need to implement policies and practices for updating existing CCDI data on a regular basis.

Clinical Data. Clinical includes data collected from both patients enrolled on clinical trials, as well as non-clinical trials data found in the electronic health record (EHR) and other systems. Given the substantial impact of inherited cancer predisposition mutations/syndromes on the pathogenesis of pediatric and AYA cancers, the WG suggests that the CCDI collect information regarding a history of early-onset cancers in other family members and/or the presence of a known cancer predisposition syndrome (e.g. neurofibromatosis type 1 (NF1) or Down syndrome) from all pediatric and AYA cancer patients. Some germline polymorphisms and mutations interact with the genotoxic therapies used to cure a primary cancer to cause histologically distinct second cancers. Given this, it would also be useful for researchers to have access to a resource that aggregates data from individuals who develop therapy-induced cancers. The WG also agreed that rich, standardized data regarding the primary cancer diagnosis, such as tumor type/sub-type, stage, histologic grade, treatment given, response (complete remission/partial remission/primary refractory), and outcome (cure/relapse/death from disease, death from a non-cancer cause, morbidities), and adverse events should be available. Information regarding any salvage treatment that relapsed patients received, and their responses, would be of great value. Given the increasing availability and ease of collection of real-world data/evidence, the NCI might consider strategies to leverage data from wearables and sensors as well as information collected from patients and families through standardized patient-reported outcome measures. Finally, the NCI should consider implementing strategies for collecting data regarding social determinants of health on pediatric/AYA cancer patients to facilitate research into how this influences treatment decisions and outcomes. Efforts to include all minority races and cultures also must be prioritized. Relevant data collected should ideally address all aspects of social determinants of health, including but not limited to education, literacy, poverty race, ethnicity and socioeconomic status.

Molecular Profiling. These types of data include: (1) cytogenetic and FISH (Fluorescence In Situ Hybridization) analysis of tumor cells; (2) immunostaining to measure the expression of individual proteins, such as MYC in neuroblastoma; (3) transcriptome (RNA-seq) analysis and/or any targeted or genome-wide DNA sequencing of biospecimens; (4) epigenetic and proteomic profiling of biospecimens; and (5) data derived from the tumor microenvironment. Ideally, paired germline/tumor sequencing data will be housed in a data commons or similar resource through the CCDI and made available to the research community by providing access to both curated/summated and primary data (e.g. FAST-Q sequencing files). Finally, as there will likely be scientific value in retrospectively performing additional molecular analyses on specimens from some patients where the clinical outcome is known, the NCI should both consider how to include

information on the existence and location(s) of archived biospecimens from pediatric/AYA cancer patients in the CCDI. In addition, this WG suggests that the NCI review current practices for collecting and cataloging biospecimens from pediatric/AYA cancer patients and consider implementing strategies for enhancing these key activities, particularly in the context of clinical trials.

Data from Preclinical Drug Testing. Testing new therapies for pediatric cancer is challenging due to the relatively small number of patients with progressive disease who become eligible for clinical trials of experimental drugs. This restricts the number of novel investigational agents that can be evaluated in children with specific cancers and underscores the need for robust models of pediatric cancer that can be used to identify the most promising agents and drug combinations. Furthermore, new anticancer agents are almost invariably developed based on activity against adult tumors, so that target mutant proteins uniquely expressed in childhood malignancies are rarely pursued. On the other hand, the initiating “driver” mutations are now known for many childhood cancers, and the overall mutational burden is relatively low in most childhood/AYA malignancies. These characteristics both facilitate generating accurate GEM models and suggest that panels of characterized PDX models will capture the key molecular features of many pediatric/AYA cancers. In addition, many drugs are being developed that target proteins encoded by genes that are mutated in both adult and pediatric cancers such as: *ALK*, *BRAF*, *FLT3*, *KTM2A/MLL*, *KRAS/NRAS*, and *NTRK*. Indeed, the RACE for Children Act mandates that pharmaceutical and biotechnology implement a pediatric development plan for these shared biochemical targets. A robust infrastructure for performing drug testing in pediatric cancer models will facilitate access to the most promising new agents.

The NCI currently supports the Pediatric Preclinical Testing Consortium (PPTC) through a cooperative agreement mechanism to test new agents primarily in PDX models of pediatric cancers. This panel of 261 PDX models across 27 distinct childhood cancer histologies was recently fully characterized with DNA and RNA sequencing in work that was funded by the Alex’s Lemonade Stand Foundation. The WG views the CCDI as a key resource for aggregating existing information about the PPTC panel and other pediatric cancer cell line, PDX and GEM models, and for providing access to the current and future results of drug testing. In anticipation of implementing the RACE Act, the NCI is actively developing a Pediatric Preclinical Public-Private Partnership (PPP3) as a strategy for continuing high throughput *in vitro* drug testing in cell line models and to support *in vivo* studies in PDX and GEM models of pediatric/AYA cancers in academia and with industry partners. The WG envisions the PPP3 as playing a key role in collecting data generated from this effort and for making it available to the scientific community. The WG is not aware of any centralized resource for investigators who are interested in reviewing published and unpublished preclinical data from studies performed in GEM and PDX models of pediatric/AYA cancers and believes that the CCDI could potentially address this important gap.

## **Recommendations**

1. We recommend that the CCDI aggregate six broad categories of data: (1) clinical, treatment, and outcome data from clinical trials and the electronic health record (EHR); (2) molecular data including research sequencing (i.e. genomic, epigenetic and proteomic data) and clinical molecular profiling; (3) information regarding the availability and location of archived biospecimens, including germline and tumor DNA; (4) longitudinal population data from patients and survivors of pediatric and AYA cancers; (5) characteristics of cell line, patient derived xenograft (PDX), and genetically engineered mouse (GEM) models of pediatric and AYA cancers; and (6) any existing preclinical data generated from studies performed in these models.

## **Recommendations (continued)**

2. We recommend that the CCDI aggregate data from previous molecular analyses of pediatric/AYA cancers irrespective of whether these studies were performed for diagnostic or research purposes and/or as part of a clinical trial.
3. We recommend that the NCI implement strategies to make all CCDI data as representative as possible of the full spectrum of pediatric/AYA cancer patients in the United States (U.S.). Recruitment targets should reflect the percentage of minorities in the U.S. population and to achieve this, we recommend the NCI include resources for persons obtaining consent to communicate and educate in a racially and ethnically sensitive manner.
4. We recommend that the 10-year CCDI effort includes plans to include new data and adapt treatments, diagnostics and prevention strategies to the knowledge gained as methods advance.

## II. Current Landscape of Pediatric/AYA Cancer Research Data and Needs Analysis

Childhood cancers are fundamentally different than malignancies arising in adults. Pediatric and AYA cancers are typically initiated by gain- or loss-of-function mutations in a key gene that alters a single lineage-restricted protein that plays a fundamental role in the development of their cell of origin. In addition, germline genetic variation potently influences tumor initiation, and epigenetic somatic alterations are critical mediators in childhood cancer phenotypes. At diagnosis, targetable mutated kinase alterations are rare in pediatric/AYA cancers, unlike tumors acquired later in life. However, genotoxic stress from chemotherapy and/or radiation therapy dramatically influences mutation burden at relapse and can sometimes create new therapeutic vulnerabilities. Consequently, childhood/AYA cancers must be investigated and treated as biologically distinct from their adult counterparts.

The RACE for Children Act has the potential to significantly and favorably alter the landscape for childhood cancer drug development by shortening the timeline between “first-in-human” (almost always adult) and “first-in-children” clinical trials. For drugs that show some activity in adults, it takes, on average, over 6 years before they are tested in pediatric cancers. The RACE Act is designed to address this problem by requiring that targeted drugs developed for adult cancer indications are rapidly tested in children when the molecular target is relevant to the growth or progression of one or more pediatric cancers. Expanding the evidence base of the relevance of specific genetic aberrations, pathway alterations, and membrane antigen expression to cancer survival and proliferation in children is hampered by the diminished access to clinically actionable tumor DNA sequencing, gene and protein expression analyses, and methylome sequencing. Furthermore, the lack of a central source for aggregation and integration of extant data developed at individual institutions for use by clinical investigators, industry sponsors, and regulatory agencies is a substantial impediment to translation. The CCDI can support the RACE for Children Act by addressing these needs. Expanding the evidence base for individual targets within individual pediatric cancer types and across the spectrum of cancer diagnoses will inform clinical research strategies for investigators, alert sponsors of the probable requirement for early consideration of pediatric development plans for new molecularly targeted agents, and engage the FDA and other international regulatory agencies in decision-making related to their authority in implementing the RACE ACT and other legislative initiatives.

There are several databases and repositories that include pediatric/AYA cancer data. Some of these contain only pediatric data, while others contain data from cancer patients of all ages. Some databases primarily host research data, while others are for clinical care. Major efforts generating data relevant for patients and survivors of childhood/AYA cancers within NCI and associated Cancer Centers include the trans-NIH (National Institutes of Health) Gabriella Miller Kids First Pediatric Research Program; NCI Pediatric MATCH trial; NCI Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative; the Childhood Cancer Survivor Study (CCSS), the Pediatric Immunotherapy Discovery and Development Network (PI-DDN); and the St. Jude LIFE study. In addition, most clinical trials data in the U.S. are collected by the NCI-funded Children’s Oncology Group (COG). The data from these efforts are accessible through multiple, disparate data repositories within and outside of NCI such as the Genomics Data Commons (GDC), Cancer Research Data Commons (CRDC), Kids First Data Resource Center (DRC), Pediatric Cancer Data Commons (PCDC), St. Jude Cloud, cBioPortal, and AACR Project Genomics Evidence Neoplasia Information Exchange (GENIE). In addition, there are several private for-profit organizations such as Tempus, Archer, and Foundation Medicine with relevant cancer data and databases. This list is only a partial representation of data available on childhood/AYA cancer patients and survivors. It will be critical for CCDI to consider how to aggregate or use data from each available pediatric/AYA or survivorship data source to meet its goals.

Unfortunately, there is no registry or catalog of existing childhood/AYA cancer data or data repositories, and there is little to no standardization among the existing databases. Among those with molecular profiling data, there is also high variability as to what types of information from such testing is available – from targeted somatic gene panel sequencing, to whole exome sequencing, to whole genome and transcriptome (RNA-seq) sequencing, as well as rich epigenetic (methylation profiling, ChIP Seq, etc.) and proteomic data sets that can inform childhood and AYA cancer research. Additionally, the critical clinical and demographic information associated with these molecular data are sometimes insufficient to support rich analyses.

In addition to existing data resources listed, the NCI has begun to develop initial phases of a National Childhood Cancer Registry (NCCR) through the CCDI that will provide access to clinically relevant patient data from childhood cancers. The various databases and registries contain very different, yet complimentary, types of data; integrating these resources has potential significant value in helping to answer key scientific and treatment questions in childhood/AYA cancers and can also avoid some redundancies. Ideally the CCDI will facilitate meaningful research through connection or federation of these data and repositories to enable interoperability.

One issue with making many different data types interoperable, regardless of the repository, is the general lack of standards employed for each dataset. The WG identified defining and promoting the use of standards for future data collection as a key potential role of the CCDI. One model for this activity could be the CRDC-H (common data model for data in the NCI Cancer Research Data Commons) from the NCI's Center for Cancer Data Harmonization. Having a clear, defined standard will help researchers in preparing data for submission to the various data platforms and repositories and further ensure interoperability for future data collection.

The WG discussed two ongoing initiatives with goals aligning to the CCDI. The AACR Project GENIE and Children's Oncology Group's (COG) Project:EveryChild both focus on comprehensive data collection and open sharing of pediatric/AYA cancer sequencing data. GENIE is a major international registry of real-world data from patients of all ages formed in 2015. It involves data sharing among 19 academic institutions.<sup>1</sup> Today it has over 80,000 clinical-grade sequencing (CLIA/ISO) records and an established legal and technical infrastructure. The consortium has safely shared data with a 7,500-person registered user base. A unique value of sequencing data generated in clinical laboratories at academic institutions is the ability to perform deep clinical annotation because the sequencing data remain linked to the medical record. The primary goal of the registry is to improve clinical decision making, particularly in patients with rare cancers and patients who have rare variants in common cancers, by linking genotypes with patient outcomes; secondary goals include catalyzing clinical and translational research.

Later in 2020, there is a planned release of nearly 3,600 sequenced pediatric tumors (age at sequencing  $\leq 18$ ) from 16 participating Institutions in GENIE. To enhance the utility of deposited pediatric data, in 2018 Drs. Katie Janeway and Andrew Kung initiated a NCI-funded project to annotate pediatric cancer records within GENIE from the Dana-Farber Cancer Institute (DFCI) and Memorial Sloan Kettering Cancer Center (MSKCC), with disease specific staging, risk factors and longitudinal clinical treatment and outcome data. PRISSMM is an artificial intelligence tool. It is a phenomics data standard developed Dr. Deborah Schrag at DFCI. It takes unstructured data from text reports in electronic health records and structures them so that they can be readily analyzed. PRISSMM uses a standardized framework for determining outcomes from real-world data. In lung and other adult cancers, pathology, imaging, signs and symptoms, tumor markers, and medical provider assessments collected into the PRISSMM model have been transformed into cancer

---

<sup>1</sup> <https://www.aacr.org/professionals/blog/aacr-project-genie-the-sharing-economy-comes-to-the-clinic/>

outcome data and found to be predictive of expected disease outcomes<sup>1</sup>. The PRISSMM framework has been process hardened and deployed for largescale clinical annotation within GENIE, including the Biopharma Collaborative Project (BPC), a five-year, \$36 million, pre-competitive collaboration between GENIE and nine leading pharma companies to annotate 50,000 GENIE records<sup>2</sup>. The GENIE pediatric clinical data annotation project has developed a pediatric version of PRISSMM adding data elements appropriate for pediatric cancers that are, when possible, harmonized with existing or emerging national standards such as the Toronto staging guidelines.

Taken together, Project GENIE and the PRISSMM framework represents an extant infrastructure and an emerging standard for aggregating and sharing clinically annotated cancer genomic data. With existing pilot-scale pediatric projects establishing the feasibility of leveraging this shovel-ready infrastructure as a standard for pediatric cancers, the overall goals of the CCDI and AACR Project GENIE are aligned.

As previously mentioned, the Children's Oncology Group provides the largest resource of U.S. clinical trial data in childhood and adolescent cancers, and a majority of the COG biobanking is done through a registry called Project:EveryChild. It was launched in 2015 to replace its former registry of histology-specific biobanking studies. Based upon 2017-2019 data, COG enrolls approximately 6,000 patients annually on Project:EveryChild, with >70,000 specimens banked since inception. COG enrolls 2,200 patients per year on phase III and pilot studies and 190 patients annually on phase II studies. Enrollment is dependent upon broad institutional participation across a network of more than 220 COG sites. The top 50 enrolling COG institutions account for 52% of all COG enrollments, with the remainder coming from the smaller 170 institutions. The COG Biorepository in Columbus, Ohio contains tissues from more than 32,000 children with cancer and related disease and is supported by an NCI U24 grant. However, institutional reimbursement for tissue and data submission through Project:EveryChild is supported by philanthropic funds for patients who do not enroll on clinical trials. The COG Biorepository is further responsible for the critical function of assigning the Unique Specimen Identifier (USI), the unique and privacy-preserving identifier that connects data and specimens for COG patients. The WG considers childhood and AYA biobanking to be underfunded, which restricts the potential impact of this resource on research and developing new therapies. It is important for NCI to consider how CCDI can support, align with or utilize tools and best practices of initiatives like GENIE and Project:EveryChild.

Aggregating sequencing data already being generated in commercial and academic medical center laboratories and linking it to harmonized clinical data from trials and/or the medical record will be a major step forward. However, its utility will be limited by the lack of systematic implementation of precision oncology in pediatrics in the United States. Large cohorts of children with cancer in the U.S. have had genomic characterization of their tumors by programs such as TARGET, Gabriella Miller Kids First Pediatric Research Program, and the St. Jude—Washington University Pediatric Cancer Genome Project. Yet, the United States has not had any coordinated effort to comprehensively characterize every childhood cancer on a national level, unlike many European countries. Consequently, many childhood/AYA cancers are not profiled at all. When genomic evaluation is performed for clinical or research purposes the approaches vary significantly. As a result, the overlap in genomic events captured across different tests can be small. A further issue is many platforms used to characterize pediatric and AYA cancers are designed with a focus on adult malignancies missing key genes or variants in pediatric cancers. For example, private (not previously described) gene fusions are very important molecular targets in pediatric cancers. They are best identified with RNA sequencing approaches designed specifically for pediatric cancers that are not

---

<sup>2</sup> <https://www.aacr.org/about-the-aacr/newsroom/news-releases/aacr-project-genie-begins-five-year-collaborative-research-project-with-36-million-in-new-funding/>

widely available. Implementing guidelines for the optimal molecular data set that appropriately characterizes pediatric and AYA cancers would: 1) allow the CCDI to evaluate the utility of genomic datasets to be aggregated; and 2) highlight areas where existing efforts are failing to generate adequate genomic data and are therefore priorities for CCDI investment in data generation. There are substantial challenges inherent in organizing and funding a national effort to sequence pediatric/AYA cancers. With this in mind, members of the WG recommended prioritizing deep molecular analysis (whole exome or genome + transcriptome) of germline/diagnosis/relapse “trios” and on rare pediatric cancers and on tumors where no effective front-line therapies currently exist.

From a technology standpoint, CCDI has enormous potential to speed discovery, but it’s also a significant “paradigm shifting” undertaking that presents a variety of challenges and complexities, such as:

- Numerous, geographically dispersed data owners with varying needs, priorities and degrees of ability (technical capabilities and resources) to share data. Some have invested millions in data generation and technology platforms and may be hesitant to embrace new data sharing and generation models.
- Ten-year funding window, potentially with year-to-year budget uncertainty, that will require carefully phasing of work and integration considerations.
- Cultural and legal challenges, like academic publishing concerns and privacy issues, introduce added complexity.

In addition to adequate financial resourcing, investments in “time and team” for CCDI solution design and “master” planning/phasing will be critical to maximize benefits and ensure that all projects align with the big picture. Inclusion of a broad spectrum of subject matter experts and stakeholders in CCDI visioning and decision-making (across diseases, organizations/institutions, and roles) – led by technology design experts with related experience – can help mitigate risks and ensure success.

## **Recommendations**

1. We recommend that the NCI perform a comprehensive inventory of data, databases and shared research infrastructures with each assessed for quality, relevance and integration feasibility to better understand the landscape of pediatric and AYA cancer and survivorship data.
2. We recommend convening experts to develop consensus guidelines addressing what clinical/molecular characterization and sample archiving should be performed and when they should be performed for each pediatric and AYA cancer diagnosis. The WG suggests that this group consider both clinically actionable sequencing to inform diagnosis and therapy and discovery analysis for research purposes. Finally, the NCI should consider potential ways of paying for molecular testing that is not currently funded by insurance or other sources. Addressing these challenging issues is central to the achieving the CCDI’s mission of learning from every child with cancer.
3. We recommend that CCDI efforts include collection and integration of new “ideal” data types prospectively, on every newly diagnosed patient -- offering standardized biomarker/omics testing (biology panel) to every child and seamlessly linking their biology, clinical, and other critical data.

## **Recommendations (continued)**

4. We recommend that the NCI consider how to collect off-trial patient clinical data, as well as data to enable synthetic control arms to support precision match trials.
5. We recommend that the NCI review current practices for collecting and cataloging biospecimens from pediatric/AYA cancer patients and assigning unique and privacy-preserving identifiers and consider implementing strategies for enhancing these key activities, particularly in the context of clinical trials.
6. We recommend developing the National Childhood Cancer Registry (NCCR) as a national resource for aggregating and accessing high-quality curated clinical information, molecular data, and other associated critical patient information and make every effort to avoid costly redundancies.
7. We recommend that the NCI convene a broad spectrum of subject matter experts and stakeholders with expertise in technology, data science and disease biology and clinical care for patients and survivors of pediatric and AYA cancers to assist with architecture design and road-mapping of the NCCR and overall CCDI data infrastructure.
8. We recommend that the CCDI include activities focused on generating and aggregating preclinical data and the development of infrastructure and tools that will enhance clinical translation. These efforts would ideally focus on the FDA's Relevant Molecular Targets List and on fulfilling the mandates of the RACE for Children Act. The NCI should also consider provide resources to validate potential targets in pediatric and AYA cancers to expand the reach of precision oncology.

### III. Potential Barriers to Progress

While addressing the key systemic (“landscape”) questions outlined in the previous section will be key to the success of the CCDI, members of the WG identified other barriers to data collection, curation, and aggregation. A lack of dedicated financial support is a substantial impediment. Specifically, the NCI, other funding agencies, institutions, and investigators invariably underestimate the personnel, time, and other resources needed to implement effective data sharing. Furthermore, members of the WG articulated the need for a commitment to support ongoing data sharing over time. For example, some NCI-funded Cancer Centers decided not to apply for the P30 supplements announced earlier this year because they are for one year only.

As resources for data aggregation and sharing are finite, the WG also discussed whether the most efficient strategy involves supporting major centers with large existing datasets and ongoing molecular profiling programs or reaching out more broadly to collect data from as many patients as possible from across the country. Of the first 2,300 pediatric cancer patients enrolled in the GENIE database, over 80% were from four institutions (MSKCC, DFCI, University of California, San Francisco (UCSF), and Children’s Hospital of Philadelphia (CHOP)). While the GENIE leadership is enthusiastic about collecting additional data from pediatric/AYA cancers, there is no funding to support data aggregation at additional institutions. Contributions from biopharmaceutical companies provides resources in support of GENIE’s activities to aggregate and curate data in common adult cancers, but there is no real incentive to do this in rare cancers, including pediatric malignancies. However, it may be possible to implement creative public-private partnerships with other entities.

Pediatric and AYA cancer patients who receive care at the largest pediatric oncology research centers appear to differ in important ways from the overall pediatric/AYA cancer population. For example, Dr. Douglas Hawkins, chair of the Children’s Oncology Group noted that patients cared for at smaller institutions are more likely to enroll on COG Phase III national clinical trials and less likely to have access to state-of-the-art molecular profiling until they relapse and can participate in an initiative like the NCI Pediatric MATCH trial.

Finally, enabling the CCDI to access and “federate” (*i.e. form into a connected infrastructure that functions together*) pediatric/AYA cancer data generated by foundations (e.g. Foundation Medicine), other funding or commercial entities, academic centers, and health systems (e.g. Kaiser Permanente) is highly desirable but may also prove challenging for various reasons. The WG suggests that the NCI address these issues proactively by providing incentives and resources to individual stakeholders. On a more positive note, if the CCDI is successful in effectively aggregating pediatric/AYA data from multiple sources, this would establish a model for data sharing that might also be applicable to adult cancers.

#### **Recommendations**

1. We recommend that the NCI commit to support ongoing data sharing over time through policy and funding. The WG identified a lack of dedicated financial support to personnel, time and other resources needed to implement effective data sharing as a substantial barrier to achieving the overall goals of CCDI.

## **Recommendations** *(continued)*

2. We recommend that the NCI consider multiple, creative ways to engage with and utilize expertise from a wide variety of stakeholders in the pediatric and AYA cancer and survivor communities. NCI should explore public-private partnerships with other entities including large pediatric care centers, foundations, and industry for interoperability and sustainability models for data and infrastructure. NCI could also consider, in addition to a panel of scientific advisors, creating a panel of technology advisors to bring valuable outside perspectives to the CCDI.
3. We recommend that the NCI consider committing resources to consent, sequence samples from, and retain diverse populations in the CCDI database to proactively ensure that this resource accurately reflects the diversity of the population.

#### IV. Generating New Data

Generating new cancer data aligns with the recommendations of the National Cancer Advisory Board (NCAB) *ad Hoc* Data Science Working Group, which highlighted the need for NCI to play a leadership role in: (1) funding the creation and sharing of cancer research data; (2) ensuring sustainability of its investment in the creation of such data; (3) creating mechanisms for these data to be made broadly available to the research community; (4) defining responsible data use policies and processes; and, (5) supporting the training of the next generation of cancer data scientists. Nowhere in the NCI portfolio are these recommendations more profound or transformative than in pediatric and AYA cancers. Likewise, three recommendations in the Data Science WG report focused on key areas of high relevance for the CCDI: (1) further investments in existing high-value data sets; (2) harmonization of terminology between cancer research data and clinical care data; (3) increase the application of artificial intelligence (AI), including machine learning and other approaches in cancer research and implement these methods in cancer care.

To achieve these goals, it is essential that datasets, methodologies, analytic pipelines, and results be rapidly and openly shared, with consistent metadata and annotation information also openly shared. As with the NCI Cancer Moonshot, it will be very important for the CCDI to have a clear, open data and open science policy. The WG suggests that NCI and NIH work together to consider and implement methods for rewarding and reinforcing good data sharing practices. Finally, to bring in the machine learning and AI communities, it is important to have clear approval processes and diminished barriers to data access. Most computationally trained researchers outside of medicine do not understand the regulatory space that currently limits access to pediatric and young adult cancer datasets. Innovative IRBs (Institutional Review Boards), data platforms, and access policies that are designed to work together to eliminate these barriers have the potential to revolutionize the application of these techniques to childhood cancer. The CCDI is a historic opportunity to create a true learning health system for pediatric cancer.

For the CCDI to generate the evidence needed to change outcomes for pediatric cancer, there will need to be investments in generating new data either for new or existing projects, aligned with the NCAB *ad Hoc* Data Science recommendations. This recommendation is even more compelling for children and young adult cancers than it is for adults, as described below:

1. Pediatric cancer is infrequent compared to adult cancer. Consequently, it is very challenging to develop data collections with sufficient numbers of subjects for comprehensive analyses, particularly for the rare pediatric cancers.
2. Recent studies have uncovered a high frequency of germline predisposition mutations in pediatric cancer. In contrast to adult cancer, this implies there is a significant familial risk that is important to document and manage. These data have implications for the entire family.
3. There is a risk that databases will not reflect the diversity of the population. Underserved populations are underrepresented in the current genomic studies of pediatric and AYA cancers. By ensuring that every child has access to cancer genomics we can redress this imbalance. Increasing the diversity of databases is a fundamental principle of precision medicine. Committing to resources to consent and retain diverse populations should be considered. Specifically, since some minority cultures are reticent to be included in “experiments” on their children, persons obtaining consent should include minority populations to engender trust. For all potential participants, communication at the education level of the participant and family should be highlighted.
4. Comprehensive genomic profiles identify variants that affect outcomes in a myriad of ways. For example, pharmacogenomic analysis can change the selection of therapies or the dose of drugs given

to individuals. Pediatric cancer survivors encounter many long-term effects, including second malignancies, that could be ameliorated with sophisticated incorporation of genomic data.

5. Many pediatric cancers - though they may share the same name with adult cancers - are biologically distinct from the corresponding adult cancers. For example, acute myeloid leukemia (AML) in children differs significantly from AML in older individuals, particularly in patients >60 years old. Generation of pediatric/AYA-specific biology data, is critical to the discovery of new therapeutic targets and treatment strategies.

For all of these reasons and as outlined above in Recommendation #2 in the Landscapes section, we recommend that the NCI develop a plan to generate new data and adapt treatments, diagnostics and prevention strategies, to the knowledge gained as methods advance in accordance with the advice of an expert consensus panel convened to address this broad subject area.

Examples of new data that could fill existing gaps include:

- Expanded cases (e.g. current data may focus on high-risk cases, but sequencing of intermediate cases or biological extremes [exceptional responders] may be beneficial)
- Expanded data types (e.g. single sequencing cell data, germline/tumor whole genome, whole exome, methylome, proteome, phenomics including social determinants and other patient/family data, etc.) as warranted by disease

## **Recommendations**

1. We recommend that the NCI seek ways to support the generation of evidence needed to change outcomes for pediatric cancer. This will require an investment in generating data through new or existing NCI initiatives, aligned with the NCAB *ad Hoc* Data Science WG report.
2. We recommend that the CCDI harmonize terminologies and coding between cancer research and clinical care data as recommended in the NCAB *ad Hoc* Data Science WG report.
3. We concur with applying of AI and machine learning approaches to analyze pediatric/AYA cancer datasets as recommended in the NCAB *ad Hoc* Data Science WG report.
4. We recommend that the CCDI remove barriers to data access for the broader community by supporting the creation of data resources that simplify data access, such as limited data sets, safe harbor data sets, and synthetic data sets that can be broadly accessed.

## V. Distinctions Between Research and Clinical Data

WG members articulated key differences between data collected through a clinical trial or for other research purposes and data obtained for routine clinical care. As summarized below, access to each type will likely require different approaches and governance.

Research and Registry Data. Some of the major childhood/AYA cancer databases are not open to most investigators. The WG considered how to encourage institutions to participate fully in data sharing. Incentives are likely more effective than penalties, but data sharing should be required, with clearly defined penalties for non-compliance. Some felt the NCI should make data sharing a clear expectation of NCI Designated Cancer Centers and that non-competing and competitive renewal funding should be contingent on tangible support of the CCDI. In rare circumstances the NCI could also decide to withhold other institutional funding such as T32 awards to ensure active participation in the CCDI. This would be regrettable if it were to become necessary. It is essential that the NIH/NCI make it clear that the expectation - from the start - will be for broad data sharing. While embargo periods on data sharing are reasonable, the WG suggests establishing clear expectations that any data collected using NCI funds will be made available to the research community within 6-12 months of collection. While this “stick” approach is essential, the “carrot” in this scenario will be the systems and processes set up through the CCDI and other programs that make it as easy as possible for researchers to “do the right thing.” The easiest course of action should be for researchers to share their data with the broader community. While these tactics might be effective for NIH/NCI funded databases, other strategies may be required for privately held databases and registries.

There must also be a clear set of tools and resources for researchers to leverage for data curation and submission. This includes educational and training resources and expert “conciierge” to support researchers preparing and submitting data. The NCI’s Center for Cancer Data Harmonization (CCDH) will be an excellent source of tools and educational and training resources to aid researchers in data harmonization and submission. Some additional resources are emerging through data resources like the Genomic Data Commons (GDC) and the Gabriella Miller Kids First Data Resource Center, and we support a concerted NCI-wide effort to make processes standardized and available to all.

Clinical Data. Leveraging data from the electronic health record (EHR) will be critical for obtaining useful data for studying pediatric cancer. Currently, most clinical data available for further research are from completed clinical trials and are limited by the manual abstraction and curation of data and entry into prescribed case report forms. The NCI is advised to support efforts to standardize the extraction and reporting of data directly from the EHR. The CCDI-funded NCCR is one project that is planning on leveraging data from the EHR, including diagnostic data and treatment and outcome information. Since EHR vendors have not productively participated in widespread standardized data sharing, it is incumbent on organizations like the NCI to incentivize academic medical centers to collect, standardize, and share the rich granular and free-text data found in the EHR.

In addition to EHR data, there may be opportunities to collect real-world data/evidence, including patient/parent-reported outcomes, data from wearables and sensors, and from other non-traditional sources. For instance, information from a wearable could be used to monitor patient activity, sleep, and heart rate after receiving chemotherapy. Standardized surveys can be administered directly to patients and families through portals such as Epic’s MyChart, with the resulting data flowing directly back into the EHR for subsequent extraction. These surveys could leverage PROMIS (Patient-Reported Outcomes Measurement Information System) quality measures or even gather data from families about therapies delivered across multiple institutions; information often best reported by parents. Data about financial toxicity could be

collected by leveraging geocoded information to infer patient socioeconomic status. Technology to leverage these data is maturing, and the most common way to move data in and out of EHRs is via HL7-FHIR (Health Level 7 - Fast Healthcare Interoperability Resources). Most healthcare institutions now have the facility to use FHIR to import and export data. The NCI's Center for Cancer Data Harmonization, part of the Cancer Research Data Commons (CRDC) ecosystem, is evaluating FHIR resources and tooling as a solution for sharing data. Further, the NCCR has asked NCI-funded cancer centers to augment their cancer registry reporting by leveraging additional EHR data. While these pilot projects are just starting, it is likely that many successful implementations will utilize FHIR messaging for pushing data from the EHR to the NCI data resources (CCDH, NCCR). The CCDI should build upon the initial efforts of these NCI resources (CCDH, NCCR) and potentially others to leverage the current work on using FHIR to standardize the movement of data out of EHRs.

The usefulness of any data is partially contingent on being able to connect those data to other sources of information. For instance, genomic data in TARGET is enriched by connecting those data to the clinical information in the Pediatric Cancer Data Commons (derived from COG clinical trials). Connecting these data relies on a privacy-preserving linked patient identifier. The COG has solved this issue by having the Biopathology Center at Nationwide Children's Hospital (Columbus, OH) deidentify each patient by assigning a publicly available Unique Specimen Identifier (USI), a six-letter code unique to each patient, at the time of biospecimen accessioning. This code can usually be publicly displayed with any data related to that patient. Since COG acts as the honest broker (the only group knowing the association between COG IDs and USIs), this process relies on COG for any disambiguation activities when updating patient data from separate sources. While potentially useful across all childhood cancer, the current process is unlikely to be scalable without additional resources available for both COG and the Biopathology Center, which administers the USI system on behalf of COG. In addition to the USI, there are several other privacy-protecting patient identifier systems in production. The team from the NCCR has had a formal consultant report prepared detailing the most impactful systems (e.g., Datavant, EUPID), and we suggest that the NCI review this report. Importantly, the WG believes that whatever system is chosen should be utilized across different NCI programs (i.e., CCDI, NCCR, CRDC), and non-NCI efforts should be incentivized to use the same system.

## **Recommendations**

1. We recommend that the NIH/NCI set clear expectations regarding the need for broad data sharing policies governing all CCDI activities. An embargo period on data sharing is reasonable, but the expectation should be that data collected using NCI funds will be made available to the research community within 6 – 12 months of collection.
2. We recommend that the NCI explore barcoding approaches to data and metadata from individual patient data so that uniqueness can be assured, and further evaluate this by funding one or more pilot projects using existing data sets.
3. We recommend that the NCI develop a long-term strategy for tracking pediatric and AYA cancer patients over many decades across their lifespan. During this time, they will become independent adults and transition their lives (and care) many times. Assigning a universal, anonymized and privacy-protected unique patient identifier to link clinical information, biospecimens, and any molecular data over time is a particularly appealing solution to this problem that should be strongly considered.

## VI. Engaging Diverse Array of Stakeholders for Input

Patient and Family Perspectives: Engaged and informed advocacy is a cornerstone of the pediatric/AYA cancer community and played a key role in passage of critical legislation, including the STAR and RACE for Children Acts. In addition, multiple philanthropic foundations are dedicated to curing childhood cancer and fund essential career development and research activities across the nation.

Pediatric/AYA cancer advocates contributed valuable perspectives and suggestions to the WG. First and foremost, the families of children with cancer strongly support the overall goal of the CCDI to create a national data ecosystem that will advance research and treatment. In this spirit, they expressed the view that molecular and other data generated from pediatric/AYA cancer patients belongs to the patients and their families and not to a given institution, lab, or organization. Accordingly, the WG suggests that the NCI explore creative ways for parents to “opt-in” to share (de-identified) data collected on their child (versus institutional or organizational ownership of data), especially genomic data. Platforms already exist that allow parents and patients to control use of their data in a dynamic way, and the WG suggests that the NCI review these models. The NCI should consider developing the CCDI’s infrastructure in a way that could allow patients and families to see their data and allow them to share it with other families, clinicians, and/or researchers. There should also ideally be a mechanism for switching control of that consenting process from parent to child when the child comes of age.

Traditionally, longitudinal data has been difficult to collect and even harder for families to access. As this information is both important for research and critical for decision making around care, the WG strongly encourages the NCI to develop the CCDI with the capacity to aggregate longitudinal follow-up data regarding outcomes and quality of life in pediatric/AYA cancer patients. We also recommend that authoritative, data-driven health care recommendations be collected in an accessible format for pediatric/AYA cancer patients and their families through the CCDI. Effective linkage to the CCSS, St. Jude Life and other cancer survivorship efforts will be essential for achieving this goal. To ensure that CCDI is delivering desired benefits, the NCI should consider developing CCDI-specific key performance indicators (KPIs; e.g. childhood cancer performance metrics/dashboard) and regular progress reports, like tracking number of new targets discovered.

In addition to their involvement as parents, family members, friends, and patients, pediatric/AYA cancer advocates possess valuable knowledge, experience and perspectives due to their various professional backgrounds, expertise, and cancer journey. Given this, the WG suggests that the NCI explore ways of harnessing these relevant skills to advance specific aspects of the CCDI in domains such as data architecture, information technology planning, and legal affairs through a rigorous nomination and selection process. In addition, the WG encourages that the NCI to consider involving pediatric/AYA advocates and industry stakeholders as participants in the CCDI governance structure, oversight or review processes, and change management activities (e.g. communications, process or policy changes).

And finally, siloes are a problem not unique to the research community -- it exists in the “advocacy/funding community” as well. First, foundations inspired to support childhood cancer research have difficulty finding each other -- a barrier to forming partnerships and collaboratively funding large-scale, strategic efforts. Second, families and social workers experience the continued challenge of locating financial and other support resources amid the sea of charities. Projects to enable funder visibility and partnering, as well as enabling patients to easily locate and centrally apply for support resources could be small-scale but visible “wins” for CCDI.

## **Recommendations:**

1. We recommend that the NCI consider multiple, creative ways to engage with and utilize expertise from a wide variety of stakeholders in the pediatric and AYA cancer and survivor communities. NCI should explore public-private partnerships with other entities including large pediatric care centers, foundations and industry for interoperability and sustainability models for data and infrastructure. NCI could also consider, in addition to a panel of scientific advisors, creating a panel of technology advisors to bring valuable outside perspectives to the CCDI.
2. We recommend that the NCI explore creative ways for parents to “opt-in” to share (de-identified) data collected on their child (versus institutional or organizational ownership of data), especially genomic data.
3. We recommend that the NCI consider developing the CCDI’s infrastructure in a way that could allow patients and families to see their data and allow them to share it with other families, clinicians, and/or researchers. There should also ideally be a mechanism for switching control of that consenting process from parent to child when the child comes of age.

## VII. Potential Opportunities for Transformative Discoveries

Given the unpredictable nature of scientific and clinical discovery, the WG acknowledged the inherent challenges of identifying specific areas where the CCDI might serve a platform for catalyzing “game changing” advances. With this said, the WG identified the following specific key areas where the NCI might consider undertaking new initiatives:

- Developing a national strategy to offer appropriate biospecimen collection and genomic testing to every pediatric/AYA patient with cancer within two years. While this is a large and complex topic, specific priorities include: (1) developing a comprehensive plan for collecting and archiving diagnostic germline and tumor samples from all pediatric/AYA cancer patients irrespective of whether they are enrolled on a clinical trial and for linking this to treatment data and outcome; (2) implementing a national standard for performing deep sequencing of the ~15% of pediatric/AYA cancers that relapse after an initial response to front-line therapies (germline, diagnostic, relapse triads); and, (3) consider what type(s) of molecular profiling should be performed at diagnosis in different pediatric/AYA cancers to guide care and enhance discovery.
- Aggregate data from existing cell line, patient derived xenografts (PDX), and genetically engineered mouse (GEM) models of pediatric or AYA cancers) that might inform rapid pediatric clinical translation in accordance with the RACE for Children Act. Consider funding focused efforts to test promising new agents in the most relevant pediatric cancer models.
- Identify patients who have a remarkable initial response to conventional chemotherapy and/or targeted therapies. Deep molecular profiling of these “outliers” have provided mechanistic insights and identified key dependencies in adult cancers. Systematically pursuing this approach will likely also be informative in pediatric/AYA cancers.
- The COG Biorepository includes multiple rare childhood cancer specimens. A focused biologic effort to delineating the molecular landscapes (and potential therapeutic vulnerabilities) of specific rare cancers should leverage the COG Biorepository and other archived specimens. Strategies to increase submission of high value tissues could increase the number of cases available.
- The COG Biorepository has a large number of germline samples that could be used for predisposition gene discovery, especially for histologies for which large germline analysis has not been done (e.g. liver tumors, germ cell tumors, etc.).
- Improve biobanking efforts at intake by adding quality control, digital image generation, and nucleic acid extraction for all children diagnosed with specific solid tumors.