



## American Society of Hematology Recommendations to the NCI's Childhood Cancer Data Initiative

The National Cancer Institute's (NCI) [Childhood Cancer Data Initiative](#) aims to improve treatments and outcomes for children, adolescents, and young adults with cancer. Below are ASH's recommendations in response to the NCI's request for input on this initiative:

### RESEARCH QUESTIONS

Research questions that can be addressed through improved childhood cancer data sharing including common data needed to answer such questions:

- a) ***Understanding the biology of pediatric hematologic cancer development through:***
  - *Insights from epigenetics:* In all hematologic malignancies, including acute and chronic leukemias, and lymphomas, there are both inherited and somatic genetic alterations that contribute to predisposition, transformation, and disease progression. While genomics plays a significant role in the onset and progression of such malignancies, it is becoming increasingly evident that epigenetic mechanisms are also heavily involved in oncogenesis. For instance, mutations in epigenetic modifiers are the most common alteration found in pediatric acute lymphocytic leukemia (ALL) at relapse, and while there are high upfront cure rates of pediatric ALL, the cure rates of relapsed pediatric ALL patients remain dismal. Insights from epigenetics could help investigators further understand the molecular mechanisms involved in normal and malignant cell development. Through the sharing of epigenetic data, investigators could begin to understand how epigenetic dysregulation in children and young adults contributes to hematologic cancer initiation, transformation, and evolution.
  - *Correlation of cancers with other hematologic diseases:* There is a need for data to understand the pathophysiologic mechanisms that increase the risk of coagulopathies in pediatric patients with cancers. Given that the prevalence of venous thromboembolism (VTE) continues to increase in the pediatric oncology population (with adolescents and young adults being the population at high risk of developing VTE) and because there is limited information regarding the impact of VTE on pediatric cancer outcome, leveraging data from existing adult cancer studies investigating this issue could elucidate the correlation between cancer and thrombosis in this patient population.

Understanding the biological underpinnings of tumor development, and correlation to other diseases would be vital in enhancing diagnosis and therapy development for the pediatric population. To ensure that such research questions are answered, the National Cancer Institute should allocate resources to fund studies aimed at: (1) identifying proteins that may impact the epigenome thus leading to oncogenesis and determining how to reverse malignant histone modifications; (2) developing locus-specific epigenetic reprogramming tools that could enhance current and future epigenetic therapies; (3) understanding epigenetic factors that affect stem cell biology and influence response to immunotherapeutic approaches; and (4) exploring the interplay between the coagulation system and tumor onset and progression. Furthermore, by implementing and supporting a federated data sharing

system, the institute could facilitate access to data generated from such studies including tools to assist with data analytics.

b) ***Refining risk stratification protocols and enhancing risk prediction in the pediatric population to identify children at risk of developing hematologic cancers:***

While the treatment outcome of children with acute leukemias (e.g., acute myeloid leukemia and ALL) has improved over the years, there remains a population of these patients who are unable to respond favorably to available therapies. Predicting poor response remains to be a challenge. In addition to supporting sequencing studies that determine risk factors for hematologic cancers in the pediatric population, this initiative could facilitate the development of a data eco-system that connects genomic datasets with biospecimen data. By interconnecting various datasets with biospecimen repositories, the field may begin to determine effective predictive biomarkers that could identify non-responders or pediatric patients at risk of developing severe adverse events to therapies. The recognition of such predictive biomarkers could help define optimal therapeutic strategies for children with acute leukemias.

c) ***Prioritizing novel therapeutic agents identified in pre-clinical research for testing in the clinical setting:***

Predictive disease models of pediatric hematologic cancers can inform efficient testing of possible novel therapeutic agents in the clinical setting. However, to ensure that these models are useful in prioritizing novel therapeutic targets, pre-clinical assessments must be conducted. Conducting such analysis efficiently and consistently will require investigators to share pre-clinical data generated from models (e.g., genetically engineered mouse models, patient-derived xenografts, etc.) that depict hematologic cancers likely to occur in the pediatric patient population as well as pharmacokinetic data. Through this initiative, the National Cancer Institute, could foster the development of a data eco-system that allows for such pre-clinical data to be shared in a comprehensive and harmonized fashion. Such an eco-system must allow for easy data deposition and include bioinformatic and interpretation tools that would efficiently predict and/or prioritize novel therapeutic targets to be tested in the clinical setting. Establishing such data-sharing infrastructure could also be instrumental in defining targets that might benefit from being targeted with combination therapies.

## DATA ANALYSIS TOOLS

Existing tools that can be adapted to make it easier to develop, maintain, and/or use childhood cancer data in a common data infrastructure. What new tools need to be developed?

a) ***Current pediatric immune-oncology data analysis tools that can be leveraged:***

Through this initiative, the National Cancer Institute could leverage existing national infrastructure such as the Center for International Blood and Marrow Transplant Research (CIBMTR), [Pediatric Health Information System](#), and electronic health records to define optimal therapeutic strategies for pediatric patients with hematologic cancers. Such platforms could be useful because they contain comprehensive information on patient demographics, treatment information, response and outcome data, and there is growing experience in the field on how to analyze such data in a meaningful way. A thoughtful, minimally cumbersome process for granting investigators access to such data should also be put in place.

The American Society of Hematology has a vested interest in immune-oncology specifically as it relates to adverse events resulting from these therapies as well as collecting and analyzing data to inform future pre-clinical and clinical studies in the immune-oncology area. To that end, the Society continues to work collaboratively with CIBMTR and other professional societies to collect cellular immunotherapy data in a unified manner and to create tools for efficient data analysis.

As for new tools, the identification of new targets and therapies for pediatric cancers would benefit from a pediatric cancer data ecosystem that captures genomic, epigenetic and non-genomic data (e.g., proteomic, epidemiologic, biospecimens, and phenotypic data) in a manner that allows for interconnectivity and interoperability and includes effective analytic tools.

b) ***Commitment to democratizing access to visualization and analytical tools:***

The National Cancer Institute (NCI) has made a significant investment in the Genomic Data Commons (GDC) which already possesses robust analytical and visualization tools. There is a sizeable amount of genomic and phenotypic information relevant to pediatric malignancies currently residing in the database of Genotypes and Phenotypes (dbGaP). The NCI could play a role in ensuring that data in dbGaP is made available through the GDC. Subsequently, data analysis would not be limited to researchers who possess and maintain in-house analytical and visualization tools. NCI's efforts to expedite the process of requesting access to dbGaP data would incentivize investigators' use of the GDC and in turn foster research.

## DATA COLLECTION

What would we need to do to collect the most informative datasets possible? What data types do we need to consider? What are the minimal molecular and clinical data elements needed for a broadly used dataset? What tissue sources should be considered?

a) ***Need to collect comprehensive data on pediatric patients with hematologic cancers:***

In addition to collecting genomic data on such patients, the following types of datasets and samples would be quite informative in disease diagnosis, therapy development, and treatment of pediatric patients with hematologic cancers:

- Proteomic data with clinical correlation
- Safety and toxicity data profiles at different time points
- Treatment regimen
- Outcomes data
- Socioeconomic and demographic data
- Biospecimens (e.g., bone marrow and blood cells)

b) ***Commitment to depositing data from curation efforts to a common repository:***

The American Society of Hematology is also committed to the understanding of inherited genomic profiles that predispose individuals to hematologic malignancies. As such, the Society has partnered with the Clinical Genome Resource to curate variants in genes (such as *RUNX1* and *GATA2*) that are involved in predisposition to myeloid malignancies. Predisposition

syndromes often manifest in children or young adults. Data collection as a result of this effort will lead to the deposition of high-quality annotated variants to a public resource (ClinVar) where other researchers will be able to benefit from these findings. More variants known to affect the pediatric population should be submitted to this common resource. The National Cancer Institute should encourage efforts from organizations and individual laboratories to deposit their variant curation results to this repository. Bulk, rapid submission of variants to ClinVar with tissue information and clinical diagnosis should be supported beyond current processes.

## DATA SHARING

What challenges are there to broadly sharing childhood cancer data? How are data successfully being shared now? How can we improve upon it?

a) ***Facilitating interoperability across various data repositories as well as novel and harmonized data analytic tools that can be used across different systems:***

Many hematologic diseases, including those that impact the pediatric population are categorized as rare diseases. The rare, yet deleterious nature of these diseases (e.g., chronic myelomonocytic leukemia) make it challenging to build a critical mass of data, that can be leveraged to inform research questions, therapy development and/or practice patterns related to these diseases. The American Society of Hematology (ASH) has recognized such limitations and has invested in the creation of the ASH Research Collaborative (ASH RC) Data Hub to facilitate the sharing of data on hematologic malignancies such as multiple myeloma in support of scientific inquiry and discovery. Through this initiative, the National Cancer Institute could help facilitate interoperability across various data repositories as well as novel and harmonized data analytic tools that can be used across different systems. This will facilitate crosstalk between various data repositories like the ASH Data Hub and ensure that all repositories are learning and benefiting from each other.

Sincerely,



Roy Silverstein, MD  
President