Executive Summary: Future Directions in Childhood Cancer Genetics Research

On May 26, 2016, researchers and clinicians with expertise relevant to pediatric cancer genetics gathered at NCI to discuss the current state of the field and high priority research opportunities for childhood cancer genetics. The workshop was co-sponsored by NCI and the NIH Common Fund Gabriella Miller Kids First Pediatric Research Program. Also participating in the workshop were pediatric cancer research advocates, NCI researchers, and NIH staff from the Kids First Program.

The workshop focused on four primary areas: the current status of childhood cancer genetic sequencing programs, analytical and technical challenges/opportunities for large scale sequencing programs, clinical implications of cancer genetics research advances, and computational challenges relating to establishing harmonized childhood cancer genetics data resources. Key discussion points and priority research areas identified are summarized below.

Workshop participants noted that the germline DNA of at least 12,000 pediatric cancer cases have or are undergoing extensive sequencing, which provides a strong foundation for studying germline susceptibility and the relationship between the germline and tumor genomes. This work has been accomplished through projects such as the Pediatric Cancer Genome Project, the Childhood Cancer Survivor Study (CCSS), the NCI TARGET project, BASIC\(^3\), and others. An important initial finding from the existing sequencing datasets is that pathogenic or likely pathogenic germline mutations in known cancer predisposition genes are observed in approximately 10% of children with cancer. This number may change as more cases are investigated and as methods for studying non-coding pathogenic variants are enhanced. The existing data has limitations, including a large number of cases (~2,500 cases) studied only by cancer gene panel sequencing, only preliminary integration of germline and tumor genomic data to infer mechanisms of tumor etiology, and almost complete lack of paired parental DNA analyses to more completely understand heritability and mosaicism. To promote more rapid understanding of the functional and clinical relevance of inherited rare variants, it will be important to focus on 1) child-parent trios, and 2) paired analyses of germline and somatic genomes from each patient. In many instances, corroborative laboratory work will be required to shed light on the functional importance of rare variants, especially when it is unlikely that a sufficient number of cases will be collected to establish adequate power to define which variants have relevant implications in the clinic and for genetic counseling. Collaborative efforts are required to achieve consensus on interpretation of variants, which, in turn, can have important implications for clinical decisions and genetic counseling.

A number of large-scale sequencing projects focused on tumor/normal pairs are ongoing or planned. Examples discussed at the workshop included the Terry Fox PROFYLE Precision Medicine Program and several projects led by St. Jude Children’s Research Hospital (SJCRH). Together with other projects ongoing around the globe, it is likely that another 10,000 pediatric cancer cases will be sequenced within three years, advancing the capacity of researchers to define clinically important variants for both susceptibility and outcomes, particularly if clinical annotation is well captured. With larger data sets and tumors sequenced at higher depth, it will be critical to characterize subclonal heterogeneity (both for insights into drivers of cancer and into relationships with clinical outcome) as well as to identify mutational signatures (for new insights into etiology).

Workshop participants noted that tumor sequencing of DNA and RNA can add important information in assessing the significance of germline sequencing results, but further work is needed to optimize the algorithms to investigate germline variation in somatic sequence data. For example, the importance of a germline variant allele may be informed by the corresponding tumor, which could harbor an alteration.
in the second allele. As well, the tumor genomic profiles provide insight into the biological role of germline events in the pathogenesis of the cancer. Hence, when possible, sequencing both tumor and germline specimens is highly encouraged.

Childhood cancer survivor cohorts are essential in providing insights into both cancer susceptibility and into the genetics of therapy-related toxicity and long-term adverse events. This line of research has added importance as 1 in 750 Americans is a survivor of pediatric cancer. Survivor cohorts undergoing large-scale sequencing (e.g., exome or whole genome) include the Childhood Cancer Survivor Study and the St. Jude Life survivor study. Together, germline sequencing results will be available for nearly 10,000 survivors of childhood cancer within the next year. The sequencing results, when combined with the rich annotation of the treatment modalities and the health status of the survivors, will allow the identification of genetic variants associated with second cancers after childhood cancer and other late adverse effects after childhood cancer. Better understanding of these genetic associations can lead to modified risk stratification for future therapeutic interventions to reduce rates of long-term toxicities.

There was agreement at the workshop that continued whole genome sequencing of childhood cancer cohorts is essential for the discovery of the catalog of important variants as well as for clinical translation of this information. To capture additional cancer predisposition genes and alleles, further sequencing studies are required to increase the number of patients analyzed and to more precisely estimate effect sizes and penetrance, as well as to identify new associations. Moreover, molecular and clinical data for large numbers of patients with a specific type of cancer are needed to have statistical power to detect new genetic insights into both susceptibility and tumor drivers. Efforts should continue to identify control sets that should be useful for improving both the discovery and characterization of cancer predisposition genes and alleles.

While early studies have shown that pediatric cancers possess strong heritability components, the current estimates are unstable due to lack of parental and family history information as well as small numbers. The pursuit of heritability of pediatric cancer requires access to family members for sequence analysis as well as generation of sufficiently high coverage of sequencing in order to discriminate between (a) heritable mutations, (b) de novo alleles, and (c) somatic mosaicism; this distinction has important implications for counseling and addressing parental concerns. Already, de novo mutations and somatic mosaicism have emerged as important genetic mechanisms in many pediatric developmental and overgrowth disorders. It is also critical to recognize that the underlying genetic architecture of susceptibility to pediatric cancer is complex and involves more than Mendelian disorders (e.g., oligogenic or epistatic models), and that future research projects need to adequately assess for these mechanisms.

The clinical implications of germline testing were discussed at the workshop. The field of returning genetic results to families is rapidly changing and new approaches are evolving, ones that must respond to the needs and requests of the family constellation. There has already been a dramatic rise in requests for genetic counseling, which creates the need for increasing the numbers of persons trained to convey this information and the need to identify ways in which health care professionals can more efficiently convey the probabilistic information (e.g., information delivery at a distance, investing resources in pre-testing counseling, etc.). Research is also needed to identify novel methods of presenting genetic information and the uncertainty provided by genetic diagnosis. As well, there is a need to better assess what patients/families truly understand about their genetic condition and how they respond based on their understanding. Finally, research is needed to address cancer prevention in the context of predisposition and to evaluate the assumptions on which prevention methods are based.
Questions such as the extent to which a clinical surveillance protocol can detect presymptomatic tumors in cancer predisposition syndrome gene mutation carriers need to be addressed, as does the extent to which presymptomatic tumors are biologically less aggressive and more effectively treated compared to symptomatic tumors.

Because of the need to study large numbers of patients to reliably investigate the above questions, workshop participants identified developing methods of combining data across sequencing projects for analysis as a very high priority. A pediatric cancer-specific commons for sharing data through a federated system was envisioned and will require extensive effort to harmonize sequence, clinical and family history annotation. To create such a pediatric cancer commons, incorporating the largest studies is a priority both for efficiency and to stabilize the harmonization of genomic and phenotypic data. The NCI Genomic Data Commons (GDC) can play a key role in establishing methods for sharing and combining childhood cancer datasets for analysis, including cross-cancer studies that have already uncovered shared mechanisms between distinct types of cancers. The mission of the NCI GDC is to provide the cancer research community with a unified and harmonized data repository that enables data sharing across cancer genomic studies in support of precision medicine. The NCI GDC currently includes both TCGA adult cancer data and TARGET pediatric cancer data. Childhood cancer data from the Kids First cancer projects will also reside in the NCI GDC. The NCI GDC has the capability to be interoperative with other cloud-based solutions that will respect the data collectors and provide attribution. The pediatric data commons that is envisioned can be the nidus for the future integration of additional pediatric data sets, privately and publicly funded, into a federated cloud system. Additionally, there needs to be a portal to the Gabriela Miller Kids First Fund Sequencing Data Resource to facilitate and accelerate the study of pleiotropy, namely improving the understanding of shared genes and pathways that play critical roles in susceptibility to cancer and structural birth defects in children.

In summary, workshop participants identified a number of high priority childhood cancer genetics research opportunities, including:

- Accelerated large-scale germline sequencing of well-annotated pediatric cancer patient and parent trios, with an increasing emphasis on including paired diagnostic and/or relapse tumor specimens when available.
- Application of germline sequencing to highly phenotyped cohorts of long term childhood cancer survivors to identify associations between genetic variants and specific patient-related outcomes such as second malignancies, treatment-related morbidities and survival.
- Development of high throughput in vitro and in vivo methods to test the functional consequences of both germline and somatic rare variants discovered through these sequencing efforts. This is of particular importance because in many cases it is unlikely that a sufficient number of cases will be collected to establish a stable genetic effect.
- Research to enhance the clinical utility of germline sequencing data (e.g., through more effective communication of results, improved screening protocols, prevention/early detection, etc.).
- Establishment of a federated or unified inter-operable pediatric cancer data commons, such as the NCI Genomic Data Commons, for the deposition and sharing uniform analysis of sequencing data. The pediatric cancer data commons needs to have harmonized data pipelines as well as harmonization of clinical and family history annotation, and should provide the ability to compute in the cloud environment with standardized analytic pipelines.
- Interoperative access between the pediatric cancer data commons and the Gabriella Miller Kids First Data Resource as well as other available pediatric and adult cancer resources.
- Establishment of standards/guidelines and methods for collecting robust phenotype information, including clinical outcomes.