Data Commons to Support Pediatric Cancer Research

Samuel L. Volchenboum, MD, PhD, Suzanne M. Cox, PhD, MPH, Allison Heath, PhD, Adam Resnick, PhD, Susan L. Cohn, MD, and Robert Grossman, PhD

OVERVIEW

The falling costs and increasing fidelity of high-throughput biomedical research data have led to a renaissance in cancer surveillance and treatment. Yet, the amount, velocity, and complexity of these data have overcome the capacity of the increasing number of researchers collecting and analyzing this information. By centralizing the data, processing power, and tools, there is a valuable opportunity to share resources and thus increase the efficiency, power, and impact of research. Herein, we describe current data commons and how they operate in the oncology landscape, including an overview of the International Neuroblastoma Risk Group data commons as a paradigm case. We outline the practical steps and considerations in building data commons. Finally, we discuss the unique opportunities and benefits of creating a data commons within the context of pediatric cancer research, highlighting the particular advantages for clinical oncology and suggested next steps.

Simply, a data commons consists of cloud-based infrastructure that includes storage for data and the computational resources and tools for analysis.[1] The research community interacts with a data commons in several ways: (1) through the submission of data, (2) by requesting and downloading data, or (3) by collecting and analyzing data on the data commons infrastructure. By facilitating these tasks, the presence of a data commons relieves the need for the researcher to purchase and manage local storage, compute, or processing tools.

There are many sources of international genomic data, both public and private, and notably some specific to pediatric cancer.[2] The examples here focus on the data available through formal mechanisms at U.S. federal agencies such as the National Cancer Institute (NCI) and the National Center for Biotechnology Information (NCBI). The NCBI hosts the Database of Genotypes and Phenotypes (dbGaP), which contains a collection of studies focused on the interaction of genotype and phenotype in humans.[3] The NCI has funded the Therapeutically Applicable Research To Generate Effective Treatments (TARGET) consortium, which is a collaboration of investigators comprised mainly of members of the Children's Oncology Group (COG), a clinical trials group devoted exclusively to childhood and adolescent cancer research.[4] TARGET researchers collaborate with the COG to access clinical expertise and biospecimens across the network, with the goal of producing genomics data that will facilitate molecular discoveries and aid translation of those findings into effective therapies. The NCI also funds the Genomic Data Commons (GDC) that centralizes, standardizes, and makes accessible data from large-scale NCI programs

such as TARGET and its adult equivalent, The Cancer Genome Atlas (TCGA).[5]

Data commons share four requirements. First, storage and compute resources must be available for cloud-based analysis of data. Until recently, the most efficient and cost-effective storage was local, on premises, managed arrays of hard drives. The availability of cheap, reliable cloud-based storage obviates the need for researchers to purchase and manage their own storage or compute clusters. Second, a data commons must access publicly available data sets. This includes data from NCI-controlled resources, such as dbGaP and TARGET, but also could include data from any source. Note that "publicly available" does not mean "freely available." In many cases, access to the data will be controlled and only available through an application process. For genomic data, whole-exome or whole-genome data deposited in dbGaP or TARGET is only available after the investigator completes an application and once their sponsoring institution completes the necessary materials transfer agreement. Third, a data commons must contain software services and tools to enable cloud-based analysis of the data. Increasingly, researchers do not have the means or expertise to provision their own servers and tools or keep up with updates and new versions. Centralized access to tools facilitates a common platform for analysis across data sets. Fourth, data must conform to the FAIR digital compliance model (Findable, Accessible, Interoperable, Reusable).[6,7]

To be findable, data must contain sufficient metadata to be persistently identifiable and distinguishable from other objects. For example, a genomic sequence should contain machine-readable metadata that ties it back to the origi-

nal source in a unique and persistent way. Accessible data are defined as requiring appropriate authorization via a well-defined protocol. Individual-level genomic data stored in the genomic data commons is accessed once the user receives authorization via the dbGaP system. The GDC authenticates users to access these data based on information passed through from dbGaP. Interoperability is key to building a successful data commons, and key components are shared vocabularies and ontologies. Furthermore, the data must be machine actionable. By being both syntactically parseable and semantically machine accessible, data can be shared between systems, facilitating a degree of sharing and interoperability that is currently a rare commodity. If these three criteria are met—findability, accessibility, and interoperability—the data can be reusable, if the metadata and other descriptors are sufficiently rich that the data can be linked to other sources. Furthermore, these descriptors should allow linkage back to the original source. Increasingly, documentation of data lineage is being required for publication, and observation of these FAIR principles will help to ensure sufficient data provenance.[8]

CURRENT OVERVIEW OF U.S. DATA COMMONS

Across biomedical disciplines, communities have acknowledged the need to "eliminate data silos and promote data sharing."[9,10] The concept of collective research that led to the success of the human genome project exemplifies how open, rapid sharing of data can accelerate discovery. In both oncology and pediatric medicine, data commons projects have been created and are already being used in research.

In pediatrics, the development of a collaborative chronic care network for children with Crohn's disease and ulcerative colitis called ImproveCareNow has allowed researchers and clinicians to collaborate to improve care and outcomes.[11,12] As of March 2015, the ImproveCareNow network had data from 73 centers, including 8,205 patients.[12] Additionally, pa-

tients play a key role in the ImproveCareNow collaborative chronic care network through input from a patient advisory council and also by providing patients the opportunity to share their stories, test new ideas, or contribute their data.[13]

In October 2016, a new collaboration, Cavatica, was announced between the Children's Brain Tumor Tissue Consortium Consortium and the Pacific Pediatric Neuro-Oncology Consortium. Cavatica also works in partnership with Seven Bridges, which provides a cloud-based environment for analyzing genomic data.[14] The goal of the collaboration is to create a data analysis platform that will help researchers collaboratively access and share data about pediatric cancers, congenital disorders, and rare diseases, such as epilepsy and autism. In building the Cavatica platform, one of the goals is to ensure that adult and pediatric data can intersect in meaningful ways. Therefore, Cavatica will interoperate with the GDC, existing NIH data repositories, and other emerging data commons.[14] Cavatica provides data access and shared use via a combination of models depending on the data source. Access to all NIH-managed data requires dbGAP approval, whereas other datasets and projects are managed by project-specific administration or a data depositor's Data Use Committee. One aspect of interoperability is accomplished by harmonizing key data concepts with those provided by the GDC and providing an application program interface (API) layer that allows querying of these concepts across multiple commons. At the same time, Cavatica allows for defining additional fields for projects, as important disease and pediatric-specific concepts may not be defined as part of the GDC. As described in the general data commons model above, by leveraging cloud technology, researchers can set up collaborative projects in Cavatica that link together data sets for coanalysis and sharable results. Additionally, all of the data, as well as the workflows run on the data, have unique identifiers that allow for reuse and reproducibility of analyses.

In addition to funding the GDC, the NCI also funds the Cancer Genomics Cloud Pilots. These projects are all designed to make cancer genomics data broadly accessible, computable, and usable by researchers worldwide, with the goal of fostering the molecular diagnosis and treatment of cancer. The GDC launched in June 2016 and serves as an integral part of the National Cancer Moonshot and the President's Precision Medicine Initiative.[5] Within the GDC, genomics data and associated clinical data can be stored and analyzed, allowing researchers to compare finding across studies. One of the key GDC initiatives was to harmonize the NCI's cancer genomics data. This included both processing the genomic data with uniform pipelines as well as developing a data model with uniform terms and definitions for biospecimen and clinical data. The GDC currently contains approximately 5 petabytes (a petabyte is 1,000,000 gigabytes) of data, which includes legacy data that were transferred from previous NCI projects as well as the newly stored harmonized data. There are approximately 42 types of cancer represented across 14,200 patients with a total of over 578,000 files. The data include 10 major types, ranging from raw sequencing

data and raw microarray data to copy number variation, simple nucleotide variation, and gene expression. The data are derived from 17 different experimental strategies, with the major ones being RNA expression (RNA-Seq), whole-exome sequencing (WXS), whole-genome sequencing (WGS), micro-RNA sequencing (miRNA-Seq), and genotyping array.

The current GDC system is a hybrid cloud that is based upon OpenStack running at an on-premise University of Chicago data center and on Amazon Web Services. In its current operations, the GDC uses these cloud-based services for internal operations, for processing data submitted via GDC's bioinformatics pipelines, and for responding to user requests through the GDC API, but the cloud-based services are not directly exposed to GDC users. The GDC provides APIs and a set of core tools built over the API, including the data portal and a data download tool. Key to the concept of the commons is the ability to seed an ecosystem in which people can create custom tools over the data and APIs to meet their needs. This is beginning to emerge around the GDC with projects such as TCGABiolinks and the related TCGABiolinksGUI.[15,16] These projects use the GDC API and provide a model for how data commons can foster applications that further improve the accessibility of data.

The NCI Cloud Pilots provide GDC users with the ability to use public clouds to execute their own bioinformatics pipelines on GDC data that have already been imported to the Cloud Pilot or are accessed directly through GDC API. For other custom analyses over NCI data, researchers currently have access to the following NCI cloud pilots: (1) FireCloud, developed by the Broad Institute, (2) the Cancer Genomics Cloud developed by Seven Bridges Genomics, and (3) the Institute for Systems Biology Cancer Genomics Cloud. Via these pilots, different ways of leveraging cloud capabilities on platforms such as Amazon Web Services and Google Cloud Platform are being investigated to inform future cloud and commons architectures. Each of the pilots has suites of tools, including APIs and user interfaces, to facilitate access and use of large-scale data sets relevant to the cancer research community.[7]

These examples of current developments in both pediatrics and oncology offer insight into the possible ways in which data commons can be organized and used across research communities. The potential to foster significant collaborative efforts and results holds great promise for the pediatric oncology community as well.

NEED FOR PEDIATRIC ONCOLOGY DATA COMMONS

Pediatric cancer is rare, and advances in diagnosis and treatment have been made through large consortium-driven trials. The total number of new pediatric cancer diagnoses per year in the United States is around 16,000.[17] There are about 3,000 new cases per year of the most common pediatric cancer, acute lymphoblastic leukemia, whereas the most common solid tumor, neuroblastoma, affects 800 new children each year. In contrast, there were 1.7 million new cases of adult cancer in the United States predicted for

2016.[18] For perspective, the total number of new pediatric cancer cases per year in the entire United States is about equivalent to the number of new breast cancer cases in Florida alone.[19] The paucity of childhood cancer cases makes it particularly challenging to study, and the emergence of data commons for pediatric cancer could be a transformative innovation.

PARADIGM CASE: THE INRG DATA COMMONS

In 2004, an International Neuroblastoma Risk Group (INRG) Task Force formed with representation from cooperative children's cancer groups in North America, Europe, Australia, and Japan. The initial charge of this multiconsortium group was to statistically analyze prognostic markers on a combined patient data set to establish an international risk group classification system. Working with statisticians from each geographic area, a standard data dictionary was created to map all of the data elements into this framework. This initiative led to development of a database that contained information on 32 clinical elements from 8,800 patients with neuroblastoma diagnosed around the world between 1990 and 2002. During the following decade, data were added to the database under supervision of cooperative group statisticians, and over a dozen high-impact, peer-reviewed papers were published.[20-23]

In 2012, the University of Chicago Center for Research Informatics (CRI) set out to remedy three severe limitations inherent in the INRG database. First, at the time, the data remain sequestered in a single flat spreadsheet, not readily available except through an onerous and lengthy data request process. Furthermore, until the request was processed, the researcher would have little insight into the feasibility of the study. The second limitation was that biospecimen availability presented a particularly difficult challenge. In the United States, most biologic samples for pediatric subjects with cancer are collected through COG clinical trials or tumor registries. These specimens are stored in a common repository at Children's National Medical Center in Columbus, Ohio. To request samples, the researchers must first query sample availability for their cohort of interest and await a response. So, even before an application for specimens can be submitted, the researchers can waste precious time waiting to document specimen availability. A third challenge facing neuroblastoma researchers was that associated genomic data were not linked to the clinical phenotype information in the INRG database, and furthermore, there were no associated resources such as storage and compute available for analysis.

Realizing these challenges, University of Chicago faculty Dr. Susan Cohn, co-chair of the INRG, and Dr. Samuel Volchenboum, director of the CRI, leveraged philanthropic funding to finance a partnership between the CRI and the University of Chicago Center for Data Intensive Science, directed by Dr. Robert Grossman. Given their experience in building and managing the NCI GDC, the Center for Data Intensive Science was the ideal partner for CRI in setting up a neuroblastoma data commons to address the above

challenges.

Key to this development was the establishment of an international governance process. Led by Dr. Volchenboum, the INRG Data Governance Committee met via phone conference several times and presented twice to the international neuroblastoma community at meetings in Cologne, Germany (2014), and Cairns, Australia (2016). This group consisted of an international team of attorneys, ethicists, and neuroblastoma subject matter experts. The work product was (1) a set of operating principles for the INRG, (2) a data-use agreement for researchers wanting data from the INRG, and (3) a data-contributor agreement, for those wanting to deposit data into the INRG.

Because the University of Chicago was acting as a service provider for the INRG, a memorandum of understanding was established between the INRG and the University of Chicago that covered the responsibilities of both groups. The CRI designed and built a database to house the INRG phenotype data, and a front-end interface was created to allow anyone to query the entire cohort of neuroblastoma patients (Fig. 1). Currently, over 18,000 neuroblastoma patients are represented in the database, spanning a period from 1980 to present.[24] Data are updated at regular intervals. In addition to phenotype information, the database can filter on biospecimen availability via an API at the COG Biospecimen Repository. As a result, a search process that took weeks or months previously can now be accomplished in a few minutes.

The process of linking phenotype and genotype information is often made difficult by the lack of a common associated identifier. Fortunately, COG assigns a Universal Specimen Identifier (USI) to each sample collected. The USI is associated with any subsequently generated samples, data, or other information. In the case of the genomic data, all samples in the TARGET data set have USIs that link back to the phenotype data in the INRG database, permitting easy association of data sets.

As illustrated in Fig. 1, central to the data commons is an application and approval workflow that requires users to formally request clinical data through a project approval portal. Access to the genomic data are governed separately through the NCBI's own mechanisms, and this approval is passed through to the GDC. Once the necessary permissions have been secured, including from the INRG Executive Committee for clinical data, from the NCBI for genomic data, and from the University of Chicago for establishment of the virtual infrastructure, the system deposits the clinical data into an object store, and a virtual machine is launched in which the user can use command line tools for data analysis. Genomic data can be pulled from the GDC through the GDC API and lined up against the clinical data for subsequent analysis.

In summary, the Neuroblastoma Data Commons enables research over clinical data in the INRG database and associated genomic information pulled from the GDC or the NCBI's Gene Expression Omnibus in a cloud-based infrastructure. The process is governed by an international data governance committee, and the University of Chicago acts as a service provider with a formal agreement with the INRG on behalf of the COG, the German Gesellschaft für Pädiatrische Onkologie und Hämatologie, the Japanese Advanced Neuroblastoma Study Group, the Japanese Infantile Neuroblastoma Cooperative Study Group, and the Society of Pediatric Oncology Europe Neuroblastoma Group.

FIGURE 1. INRG Data Commons Design and Workflow
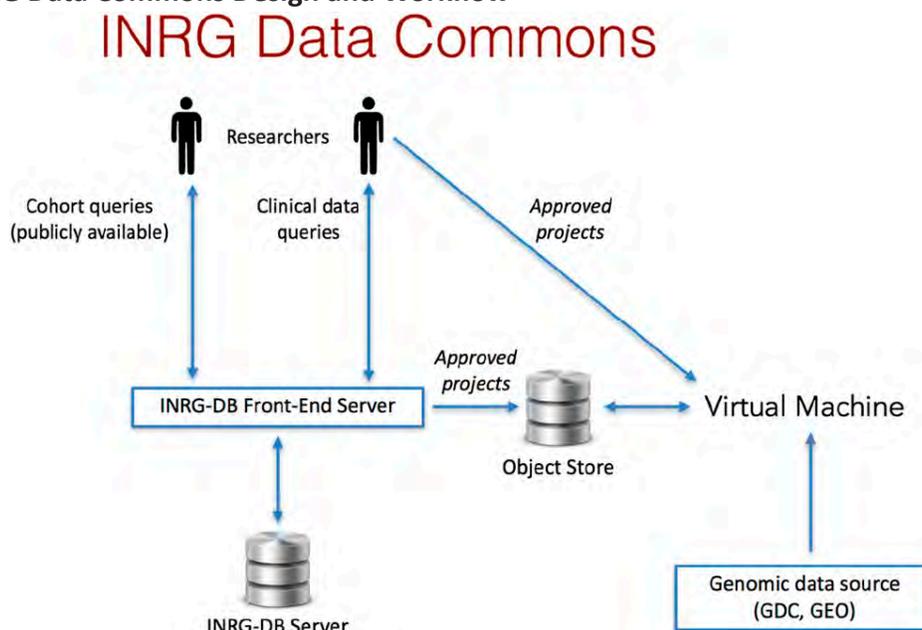
This figure illustrates the relationship between the INRG database and the permissions and technology that allow researchers to access the data.

PRACTICAL ISSUES IN BUILDING DATA COMMONS

With the establishment of the INRG Data Commons, other pediatric solid tumor groups have expressed interest in building similar pediatric cancer data commons. Efforts are underway with groups representing pediatric sarcoma, germ cell tumors, and brain tumors to collect and standardize data and build data commons. In considering development, it is useful to use a paradigm, which is emerging for building cancer data commons. Key elements have been drawn from data science, data governance, and engineering.

Figure 2 outlines this paradigm and some of the important issues and steps in building a pediatric cancer data commons. Because data commons are ultimately a shared community resource, engaging stakeholders from the outset of the design is critical to defining the data model and achieving standardization. Furthermore, cooperation from consortium members is critical in defining scope and being able to collect data from multiple sources. Finally, cooperative group members will serve as a resource throughout the process for advice and governance, maximizing the chance for an accepted, usable, and successful data commons. Other key considerations that must be considered are: identifying a funding source and infrastructure; engaging a project team; and identifying relevant data sources.

Once the basic purpose and structure is defined, then establishing governance policies and procedures such as contributor and user agreements is integral to the functioning of the system. Standards for data definition, format, and ontologies must be agreed upon across the group, and engaging external experts in this area such as the Clinical Data Interchange Standards Consortium can help guide the process. The Clinical Data Interchange Standards Consortium mission is "to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of health care."[25] Once standards are established, the project team can create a data dictionary and map various

FIGURE 2. Paradigm for Building Pediatric Cancer Data Commons

Paradigm for pediatric cancer data commons
1. Engage cooperative groups
2. Define scope
3. Identify funding
4. Choose infrastructure
5. Engage team
6. Identify data sources
7. Establish governance
8. Create agreements
9. Establish standards
10. Create database
11. Build front-end tools
12. Establish back-end connections to data sources and infrastructure
13. Create and execute communication and education plans
14. Create sustainability model

This figure offers a model, drawn from the fields of data science, data governance, and engineering, of the critical steps in building a successful and sustainable pediatric cancer data commons.

elements that can be used to create database. A front-end query engine allows users to easily access information about the data commons and increases impact by lowering the threshold for determining key questions about sample size and statistical power.

The final steps to ensure continued success of a data commons includes the creation and execution of a robust communication and education plan that will inform potential contributors and users about the data commons on a regular basis. In conjunction with ongoing stakeholder engagement, this will help to build a sustainable data commons for the future. An essential component of the consortium is real patient and family engagement. Only in building such a model can we authentically empower patients and foundations to take an active role as data producers in their participation and contribution to research.

NEXT STEPS: FUTURE OF PEDIATRIC CANCER RESEARCH

Like data-driven progress in adult cancers, "big data" genomics in pediatric cancers requires a shared computation/harmonization infrastructure and, more importantly, the necessary associated availability of well-curated genomic and phenotypic data from large cohorts of pediatric patients. But because of the paucity of pediatric cancer cases, children with cancer are failing to benefit from the technological revolution driving the precision-medicine age. Although 5-year survival rates for children diagnosed with cancer have improved in recent decades, survival rates remain very low for some types of cancer, especially for metastatic tumors and those with high-risk genetic features. Despite steady advances in diagnosis, treatment, and follow-up, cancer remains the leading cause of death from disease in children in the United States.[26,27]

We propose that the next important step in addressing pediatric cancer is the creation of a shared and well-curated pediatric cancer data commons or an ecosystem of connected data commons. This fully empowered data commons would accelerate discovery through existing cohorts and would include a consortium-led approach to developing robust processes for data contribution, data attribution, data sharing, collaborative discovery, and shared analysis and further provide for interoperability and access of data. This can be accomplished by utilizing the paradigm offered here (Fig. 2) to ensure comprehensive and collaborative development of the data commons infrastructure, policies, and processes.

One of the most promising benefits of developing a shared and well-curated pediatric cancer data commons or ecosystem is the ability to develop novel personalized medicine approaches to treating pediatric cancer. In personalized medicine, interventions are tailored to individual variation in risk and treatment response.[28] Genomics elevates personalized medicine by allowing for precise classification of an individual's disease and potentially expected responsiveness to treatment. A combination of the characterization of cancers and known responsiveness to treatment allows for

targeted interventions. These targeted therapies could be especially beneficial for children who may live many years after treatment and are at risk for treatment-related late effects. Currently, one in 1,000 individuals in the United States is a childhood cancer survivor, which means there are numerous survivors who are at increased risk for long-term health consequences resulting from their treatments.[29] Personalized medicine offers the opportunity to identify cohorts of children most in need of aggressive treatment while reducing exposure to ineffective therapies.[27]

The future of pediatric cancer research depends on our ability to develop collaborative data commons and to bring together data, research expertise, and clinical practice to translate the benefits of these collaborations into real change for children. Advances in genomic profiling, along with the democratization of data storage and compute resources, have resulted in a computational landscape ideally positioned for studying pediatric cancer. What remains is a focused and systematic effort to collect, standardize, and combine multiple disparate phenotypic, genomic, and other data sets from children with cancer into one or more connected data commons. If built with robust data governance and a well-conceived sustainability model, these resources could have a transformative effect on pediatric cancer research, resulting in novel and better ways to diagnose and treat children with oncologic diseases.

ACKNOWLEDGMENT

References

1. Grossman RL, Heath A, Murphy M, et al. A case for data commons: toward data science as a service. *Comput Sci Eng*. 2016;18:10-20.

2. St. Jude Children's Research Hospital. PeCan Data Portal. https://pecan.stjude.org. Accessed March 27, 2017.

3. dbGaP. www.ncbi.nlm.nih.gov/gap. Accessed March 4, 2017.

4. National Cancer Institute Office of Genomics. Therapeutically Applicable Research to Generate Effective Treatments (TARGET). https://ocg.cancer.gov/programs/target. Accessed March 4, 2017.

5. National Institutes of Health. Newly Launched Genomic Data Commons to Facilitate Data and Clinical Information Sharing. www.nih.gov/news-events/news-releases/newly-launched-genomic-data-commons-facilitate-data-clinical-information-sharing. Accessed February 26, 2017.

6. Force11. Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0. www.force11.org/fairprinciples. Accessed March 4, 2017.

7. National Institutes of Health. The NIH Commons. https://datascience.nih.gov/commons. Accessed February 21, 2017.

8. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018.

9. Delaney SK, Hultner ML, Jacob HJ, et al. Toward clinical genomics in everyday medicine: perspectives and recommendations. *Expert Rev Mol Diagn*. 2016;16:521-532.

10. Rehm HL, Berg JS, Brooks LD, et al; ClinGen. ClinGen: the Clinical Genome Resource. *N Engl J Med*. 2015;372:2235-2242.

11. Margolis PA, Peterson LE, Seid M. Collaborative Chronic Care Networks (C3Ns) to transform chronic illness care. *Pediatrics*. 2013;131(Suppl 4):S219-S223.

12. Marsolo K, Margolis PA, Forrest CB, et al. A digital architecture for a network-based learning health system: integrating chronic care management, quality improvement, and research. *EGEMS (Wash DC)*. 2015;3:1168.

13. Collaborative Chronic Care Network. Patients. http://c3nproject.org/patients. Accessed February 22, 2017.

14. "Pediatric Research Consortia Unveil CAVATICA Data Platform." Innovation District, October 20, 2016. http://innovationdistrict.childrensnational.org/pediatric-research-consortia-unveil-cavatica-platform/. Accessed February 26, 2017.

15. Bioconductor. TCGAbiolinks: An R/Bioconductor Package for Integrative Analysis With TCGA Data. http://bioconductor.org/packages/TCGAbiolinks/. Accessed February 28, 2017.

16. TCGAbiolinksGUI. http://bioinformaticsfmrp.github.io/TCGAbiolinksGUI/. Accessed February 28, 2017.

17. CureSearch. Number of Diagnoses. https://curesearch.org/Number-of-Diagnoses. Accessed March 4, 2017.

18. Surveillance, Epidemiology, and End Results Program. https://seer.cancer.gov/. Accessed March 4, 2017.

19. American Cancer Society. Cancer Facts and Figures 2016. www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2016.html. Accessed March 4, 2017.

20. Dubois SG, London WB, Zhang Y, et al. Lung metastases in neuroblastoma at initial diagnosis: a report from the International Neuroblastoma Risk Group (INRG) project. *Pediatr Blood Cancer*. 2008;51:589-592.

21. Meany HJ, London WB, Ambros PF, et al. Significance of clinical and biologic features in stage 3 neuroblastoma: a report from the International Neuroblastoma Risk Group project. *Pediatr Blood Cancer*. 2014;61:1932-1939.

22. Bagatell R, Beck-Popovic M, London WB, et al; International Neuroblastoma Risk Group. Significance of MYCN amplification in international neuroblastoma staging system stage 1 and 2 neuroblastoma: a report from the International Neuroblastoma Risk Group database. *J Clin Oncol*. 2009;27:365-370.

23. Thompson D, Vo KT, London WB, et al. Identification of patient subgroups with markedly disparate rates of MYCN amplification in neuroblastoma: a report from the International Neuroblastoma Risk Group project. *Cancer*. 2016;122:935-945.

24. Cohn SL, Pearson ADJ, London WB, et al; INRG Task Force. The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report. *J Clin Oncol*. 2009;27:289-297.

25. CDISC. CDISC Vision and Mission. www.cdisc.org/system/files/all/CDISC-4-Pager_pages_web.pdf. Accessed March 4, 2017.

26. National Cancer Institute. Cancer in Children and Adolescents. www.cancer.gov/types/childhood-cancers/child-adolescent-cancers-fact-sheet. Accessed February 28, 2017.

27. Adamson PC. Improving the outcome for children with cancer: development of targeted new agents. *CA Cancer J Clin*. 2015;65:212-220.

28. Conti R, Veenstra DL, Armstrong K, et al. Personalized medicine and genomics: challenges and opportunities in assessing effectiveness, cost-effectiveness, and future research priorities. *Med Decis Making*. 2010;30:328-340.

29. Landier W, Bhatia S, Eshelman DA, et al. Development of risk-based guidelines for pediatric cancer survivors: the Children's Oncology Group Long-Term Follow-Up Guidelines from the Children's Oncology Group Late Effects Committee and Nursing Discipline. *J Clin Oncol*. 2004;22:4979-4990.