Gabriella Miller Kids First Pediatric Research Program

First Annual Meeting

DRC Executive Summary

September 6-7 2017

Introduction

The Gabriella Miller Kids First Pediatric Research Program's (Kids First) Annual Meeting was intended to bring together for the first time the newly-awarded Kids First Data Resource Center, representatives from the Kids First sequencing centers and X01 cohort projects, and other members of the research and advocacy communities to discuss efforts needed to build a centralized data resource that will be widely used by the pediatric research community.

An additional focus of the meeting was on partnership and engagement of patients and families. Foundations and patient representatives were invited to the annual meeting to receive an overview of the Kids First program and an outline of the DRC goals for the development of the Kids First Data Resource. There were opportunities for the patient/foundation groups to ask questions during a designated session of the meeting. Lastly, there was a "**Meet and Greet**" which provided the opportunity for patient/foundations to interact with the researchers involved in analyzing Kids First genetic data and building the data resource as well as Kids First Program staff from the NIH.

Meeting Goals

- Update the Public on the Kids First Program's Progress
- Introduce the Vision for Kids First Data Resource Center
- Review the Current State of the Art of Data Portals
- Propose Models for Data Sharing and Access
- Discuss Genomic Data Processing and Integration
- Pursue Phenotype and Clinical Data Harmonization
- Brainstorm Portal Development Process
- Foster Cross Analyses, Joint Discovery, and Collaboration

Wednesday, September 6, 2017

Overview of the Kids First Program & Kids First Progress Update

Lorette Javois Ph.D. (NICHD)

The Gabriella Miller Kids First Pediatric Research Program was initiated in response to the 2014 Gabriella Miller Kids First Research Act. The vision of the program is to alleviate suffering from childhood cancer and structural birth defects by fostering collaborative research to uncover the etiology of these diseases and supporting data sharing within the pediatric research community. The program is overseen by the Kids First Working Group and Leadership Team with input from external scientific advisors. The program has accomplished many major milestones to date including:

- Identifying cohorts of children with childhood cancer and/or structural birth defects;
- Funding the Kids First Sequencing Centers and providing whole genome sequencing for over 18,000 genomes;

- Awarding the Kids First Data Resource grant to develop the infrastructure that will provide researchers with necessary computational infrastructure and analysis tools to analyze large and complex data sets.

Future initiatives of the program will support data analysis, data mining, and demonstration projects. Discussion points and audience questions noted the need to provide support for analysis and additional support for clinical and phenotypic data collection.

The Kids First Data Resource Center Team Composition and Vision

The core Children's Hospital of Philadelphia (CHOP) team and their partner sites from University of Chicago (UChicago), Ontario Institute for Cancer Research (OICR), Oregon Health Sciences University (OHSU), Children's National Medical Center (CNMC) and Seven Bridges Genomics (SBG) were introduced to the X01 teams. An overview of the proposed roles and responsibilities of the DRC across its functional units was provided, describing the Data Coordinating Center (DCC), Data Resource Portal (DRP), and Administration and Outreach (AOC) components of the Data Resource Center (DRC). Due to the recent nature of the award to CHOP, a key component of the meeting was for the DRC to learn more about the current cohorts, scientific objectives, investigators and to define current and future needs and requirements.

Engaging the Advocacy Community

There was a total of **14 patient advocates/foundations** represented during the Kids First Annual Meeting either in-person or via WebEx. The breakdown was as follows:

Group	Count
Cancer	12
Structural Birth Defects	2

Following the annual meeting, the DRC further engaged patient/foundation/advocacy groups. A post-meeting evaluation survey, composed of 10 questions, was distributed to all the patient/foundation attendees soliciting feedback that will ultimately enable the DRC to improve future meetings and better meet the needs of our supporters and patient family advocates. There were 8 respondents to the total survey with 7 completing the question regarding satisfaction.

Very Dissatisfied	Somewhat Dissatisfied	Neither Satisfied Nor Dissatisfied	Somewhat Satisfied	Very Satisfied
0% 0	0% 0	0% 0	43% 3	57% 4

When asked what could have made their experience better, patient/foundation/advocacy groups responded that, 1. they would have liked more notice so they could attend in-person, 2. for those that attended in-person, they would have liked more time allotted to interact with researchers and other advocates, 3. they would have appreciated more materials available to take back with them and 4. lastly, they were disappointed by the misperception that NIH was not requiring data sharing.

Lastly, per the questions that were posed during the "Session 1: Questions from the Advocacy Community", the AOC team will create a Patient/Foundation FAQ sheet that will be easily accessible on the Kids First website and available as a handout for future advocacy partners.

Data Resource Portal Background and Initial Steps

Prior to the meeting, attendees were engaged via an online survey to inform the DRC of their current needs, practices, and expectations. A review of the prior portals that the DRC teams have been involved in developing was given to help provide context for what could be reused as well as new requirements and opportunities for the Kids First DRP recognizing that many of these portals are cancer-centric and that additional efforts will be needed to inform cross-disease analysis and workflows. The two prior portals presented included:

- **GDC/ICGC Data Portal by** Vincent Ferretti Ph.D. (OICR). Dr. Ferretti gave a live demonstration of the main functionality of the International Cancer Genome Consortium (ICGC) Data Portal. This includes sophisticated, faceted search capabilities making data exploration extremely fast and easy, a suite of interactive JavaScript components for in-depth analysis and visualization of specific genomic features, embedded genome and pathway browsers, synthetic cohort comparisons, and a streaming data download service. The portal integrates a large variety of annotations such as variant consequences of variants and variant frequencies, functional impact factors and druggability. The portal also offers cloud-based tools for searching a catalog of raw ICGC data files stored in worldwide repositories and cloud-based computing environments.t

- **PedcBioPortal** by Pichai Raman Ph.D. (CHOP) PedcBioPortal is a web-based application geared at allowing basic and translational pediatric cancer researchers ways to explore and interrogate high-dimensional genomic data. The application is an instance of the original cBioPortal with certain functionality expansions and a focus on pediatric cancer data sets. We have implemented this application at the Children's Hospital of Philadelphia to support research efforts and transitively through many collaborations with other institutes. This was presented during the first annual meeting to demonstrate the capability for implementing and helping to design enterprise level highly used genomic applications. Additionally, we wanted to use this as a springboard and framework to spur discussion among X01 investigators as to the best ways to visualize their data.

To wrap up the session, Dr. Ferretti shared an initial vision for the Kids First Data Resource Portal. He described the four main user types the portal aims to serve (biomedical researchers, clinicians, computational biologists, and patients) and presented the overall portal design strategy. This was followed by a description of the main portal components and their interconnections. The DRP team

is very interested in further engaging the X01 investigators and the Kids First community to further shape this vision and prioritize the development roadmap.

Kids First Data Coordinating Center

An overview of the proposed Kids First Data Coordinating Center roles and responsibilities was given by Drs. Allison Heath and Deanne Taylor. This included ingesting, harmonizing, and shepherding the data through the various distribution methods available to the Data Resource Center to both the X01 investigators and their direct collaborators as well as the larger research community. For the former, the DRC will set up cloud-based workspaces that will allow X01 investigators to both work directly on the cloud as well as download the data locally if desired. For the secondary users and larger research community, the DRC is working to attain NIH Trusted Partner status via a partnership with University of Chicago. This will allow the DRC to distribute data under authorization of dbGaP. An initial presentation of genomic harmonization strategies was made, but the main consensus was that the DRC should gather more information about the specific genomic data needs of the X01 project teams before any decisions are made. General agreement was that harmonization is a key component of the DRC to facilitate cross-cohort analysis. Immediate next steps for the DCC are to transfer genomic data from WashU and BCM and set up ongoing transfer processes with Broad and Hudson Alpha for the ongoing sequencing efforts.

Prior projects that the DRC team has been involved in relevant to the DCC component of the DRC were presented:

- **GDC: Towards a Data Commons Environment** by Robert Grossman Ph.D. (UChicago). The University of Chicago will: 1) develop commons services for the Kids First Data Resource Center (KFDRC) by using, and adapting as required, the Bionimbus and Genomic Data Commons software stack; 2) operate data commons services provided by the Common Services Operations Center (CSOC) at the University of Chicago; and 3) develop and operate bioinformatics tools and services for understanding pediatric cancer and birth defects. This will provide technical resources for integration with commons services, data management tools, and biomedical informatics tools and services.

- **Cancer genomics cloud pilots and Cavatica** presentations by Brandi Davis-Dusenbery Ph.D. (SBG) and Adam Resnick Ph.D. (CHOP) provided an overview of the various cloud-based models of performing analysis next to large scale data, models for PI and consortium-driven collaboration and sharing in the context of the pediatric genomic data.

Kids First DRC Collaborations, Partners and Commitment Makers

As part of the DRC's efforts for engaging the wider landscape of consortia-based initiatives involving Kids First X01 investigators and the DRC, several presentations were given as examples to determine the integrative potential for the Kids First Program:

- **Children's Brain Tumor Tissue Consortium (CBTTC)** Adam Resnick Ph.D. (CHOP) CBTTC is generating data including whole genome and RNA-seq on a cohort of about 2,000

patients with pediatric brain tumors.

- **International Neuroblastoma Risk Group (INRG) Data Commons** Samuel Volchenboum MD, Ph.D. (UChicago) The neuroblastoma data commons as a paradigm system in which the data elements have been harmonized across four international consortia. The neuroblastoma data commons now houses over 19,000 subjects and links out to available genomic data and tissue availability. Dr. Volchenboum discussed ongoing work in building a soft-tissue sarcoma data commons and a germ cell tumor data commons.

- **OPEN DIPG** Javad Nazarian Ph.D. (CNMC) The objectives of the initiative are to 1) Generate DIPG'Omics for primary analysis 2) Centralize all DIPG'Omics for secondary analysis 3) Integrate the new DIPG data 4) Unify DIPG expertise (bioinformaticians, data scientists, clinicians, researchers and new talent) to analyze the DIPG genomic data. Dr. Nazarian indicated that the initiative aligns with the aims of the Kids First program, and he looks forward to donating the data for cross analysis.

- **Alex's Lemonade Stand Foundation (ALSF) Childhood Cancer Data Lab** Casey Greene Ph.D. (UPENN) Dr. Greene provided an introduction to the Alex's Lemonade Stand Foundation and the mission of the Childhood Cancer Data Lab. The mission is to accelerate the pace of finding cures for childhood cancer by empowering scientists and doctors to harnessing the power of Big Data. The first informatics lab of its kind, CCDL will construct tools that make data and analysis widely available, easily mineable, and broadly reusable. Finally, for training, Dr. Greene focused on providing a "carpentry" for childhood cancer genomics (like Data Carpentry).

Thursday, September 7, 2017

Phenotype and Clinical Data Harmonization

The aim of this session was to gather a better understanding of how phenotype harmonization will support the scientific aims of the Kids First cohorts. This understanding would be used to help prioritize the tasks of the DRC as phenotype harmonization in general is a very challenging problem. To provide background for the current state of the art, several existing initiatives for harmonizing phenotypic and clinical data were presented:

- **Retrospective Data Harmonization: Maelstrom Guidelines and Tools** by Vincent Ferretti Ph.D (OICR) Dr. Ferretti presented the Maelstrom Research (www.maelstrom-research.org) work on retrospective data harmonization of population-based cohorts. He gave an overview of the Maelstrom Research guidelines for rigorous retrospective data harmonization published in the International Journal of Epidemiology (Fortier et al., IJE, 2016) and explained how these guidelines have been implemented and applied in large-scale international data harmonization projects using the OBiBa software suite (Doiron et al. IJE, 2017) developed by Maelstrom Research.

- **Deep Phenotyping for Improved Diagnostics and Analysis** by Melissa Haendel Ph.D (OHSU) Dr. Haendel presented several consortium efforts that have helped to improve diagnostics across a wide disease landscape. This included the Monarch Initiative, Human

Phenotype Ontology, and the Undiagnosed Disease Network. She presented several platforms and tools that have made data collection and harmonization easier, resulting in better scientific and translational results.

Following the presentations, an interactive session was held to help gather an understanding of the current cohorts, their status, and needs for phenotype/clinical data harmonization. A general theme that emerged was that one of the focuses of the DRC should be leveraging existing standards and ontologies to harmonize across the cohorts. One unique challenge identified was providing cross analysis capabilities between the structural birth defect cohorts and the cancer cohorts. Additionally, incorporation of familial aspects along the genomic analysis will be needed. There was high variability across the cohorts in terms of the breadth and depth of the phenotypic and clinical data that has been gathered. There may also be potential to work with foundations and patient groups to augment the X01 efforts. Similar to genomic harmonization, one key recommendation was to gather further details by talking directly with each one of the X01 investigators and their team.

Portal Development Process: Interactive Session

This session was developed to identify the data elements and functionalities that X01 investigators were most interested in. While previous sessions focused on what was done and what was possible (e.g., presentation of portals such as PedcBioPortal and ICGC) this discussion was geared toward understanding their current pain points with respect to informatics research. Essentially, the DRC was trying to compile the wish-list of features requested by X01 investigators to make data querying, analysis, and reporting easier. The sessions were structured as a set of open questions to the audience who were then grouped into teams to discuss and report back. The results of this session demonstrated that the wish-list was finite and there were a few very common struggles that researchers face. However, there were many more features that were specific to the disease and investigator. All features were gathered in a document and are being prioritized (based on scope) for development in the portal.

Data Access, Programmatic and Data Set Integrations

Updates were given by the Kids First Genome Sequencing Centers:

- **Current processes, workflows, and collaborations at the Broad by** Daniel MacArthur Ph.D. (Broad). Dr. MacArthur provided the Broad sequencing center update. This included:
    - Dr. Engle - X01 data is now loaded
    - Dr. Vilain (disorders of sex development) – in processing
    - Dr. Shen (hearing loss) – finishing sequencing at Baylor
    - Drs. Seidman, Chung, Marazita – pending processing but need to understand how to work DRC on long-term integration with their analysis solutions

- Dr. MacArthur called attention to one problem regarding the general incompatibility of data aligned and processed at different sites. Varying choices in reference genome, aligner, or data processing steps produce incompatible data resulting in different variants and genotypes. Also, batch effects encumber analyses that seek to combine data. His proposed solution was to create "functionally equivalent" alignments.

- **Kids First Genome Sequencing Center: Current processes, workflows, and collaborations at St. Jude/Hudson Alpha** Xiaotu Ma Ph.D. (St. Jude), Shawn Levy Ph.D. (Hudson Alpha) This presentation provided the current state of workflows and LIMS process as well as highlighted the importance of sample QC and data harmonization to inform downstream analysis.

Data Access Practices and Models Review: Consortia, dbGaP, and Others
The genomic data storage and distribution landscape is shifting at the NIH. The DRC will need to work closely with the Program staff to determine the best strategies moving forward and communicating any changes to the X01 and other Kids First data community stakeholders. The DRC can technically support several different models, so it is primarily policy decisions that are needed to guide the appropriate technical solutions.

X01 Cross Analysis and Joint Discovery

To close out the annual meeting, there were several presentations from investigators on their scientific analysis with a focus on cross analysis and joint discovery:

- **GRIN** by Deanne Taylor Ph.D. (CHOP) This presentation described the Genomic Research & Innovation Network, a consortium between CHOP, Boston Children's Hospital, and the Cincinnati Children's Hospital Medical Center that was formed in 2016. The talk discussed how the GRIN collaboration enhances data sharing and collaboration through reducing barriers by having common broad consent forms, shared and centralized architecture and data structures, shared genomic pipelines, shared data use agreements, and legal and intellectual property agreements between the three institutions. The accomplishments of the GRIN consortium so far in analysis and discovery of genetic variants that may contribute to rare disease were discussed. GRIN's website can be accessed at http://www.grinnetwork.org
- **Bench to Bassinet (PCGC)** Christine Seidman MD (HMS): This presentation provided key insights into the PCGC cohort and the scientific questions investigators are seeking to engage via the Kids First X01 program.
- **TARGET/Neuroblastoma by** Pichai Raman Ph.D. (CHOP) This presentation was delivered on behalf of Dr. John Maris, one of the 2016 X01 investigators leading a project on Neuroblastoma titled "Genetic Basis of Neuroblastoma initiation and progression". The talk was geared to provide a summary of the research done to date, specifically the state of sequencing and processing. Additionally, an ongoing analysis plan with the X01 data and other analytics involving integration with more datasets wwas explored. Finally, a vignette regarding the discovery of immunotherapy targets in neuroblastoma supported by data integration efforts was detailed.
- **PPTC** Malcolm Smith MD, Ph.D. (NCI), Pichai Raman Ph.D. (CHOP) This presentation gave an overview of the PPTC program and its scientific goals. It also introduced the idea of preclinical testing data alongside patient data in the Kids First DRC.

Recap, Actions, and Meeting Closing:

The meeting concluded with a brief overview of the next steps and a commitment to engage the X01 investigators directly via one-on-one meetings in the context of a DRC "listening tour" in which the DRC staff would learn more about the individual needs and priorities of X01 grantees.