National Cancer Institute (NCI)
Integrated Canine Data Commons (ICDC) Steering Committee (SC)

Teleconference
Wednesday, August 21, 2019

Participants (*Present)
External Committee Members
Matthew Breen*
Renee Chambers*
Dawn Duval*
Allison Heath*
Will Hendricks*
Warren Kibbe*
Debbie Knapp, ICDC-SC Chair*
Cheryl London*
Jeff Trent*
Roel Verhaak
Shaying Zhao*

Internal Committee Members (NCI, National Institutes of Health, and Frederick National Laboratory for Cancer Research [FNLCR])
Matthew Beyers*
Allen Dearry
Toby Hecht*
Paula Jacobs*
Tony Kerlavage
Erika Kim*
Amy LeBlanc*
Christina Mazcko*
Philip Musk
Elaine Ostrander*
John Otridge
Ralph Parchment, ICDC-SC Managing Secretary*
Connie Sommers*
Greg Tawa*

Others
Lori Lydard
Tara Whipp
Mary Cerny (writer)*

Opening and Welcome

Drs. Knapp and Parchment opened the meeting at 11:35 a.m. ET and welcomed those in attendance.

Minutes of the June Meeting

The minutes of the June 26, 2019, ICDC-SC meeting were accepted with the following corrections: (1) The date of the Data Governance Advisory Board (DGAB) meeting was June 24, not June 17, and (2) reference to "collaborating with Vox" should be changed to "collaborating using the Box application."

Follow-up: Childhood Cancer Data Initiative (CCDI) Symposium and Poster Presentation

The CCDI focuses on the critical need to collect, analyze, and share data to address the burden of cancer in children, adolescents, and young adults. The CCDI symposium was held in Washington, D.C., on July 29–31, 2019, and constituted a scientific planning session to gain a common understanding of the current issues and opportunities in childhood cancer research that can be addressed through enhanced data collection and maximum utilization of that data.

The ICDC-SC submitted an abstract for the poster session to present the ICDC and identify ways the ICDC and CCDI can work together (Abstract title: "Synergy Between the CCDI and the ICDC to Uniquely Propel Advances in Childhood Cancer"). Dr. Trent had a brief presentation in support of the ICDC on the second day of the symposium. There was a lot of traffic surrounding the poster, and Dr. LeBlanc, who presented the poster, was invited to speak about the ICDC at an upcoming Board meeting. Other ICDC-SC members also attended the conference. Many symposium participants expressed interest in the ICDC, and further follow-up from the conference is expected.

NCI has fully funded the ICDC for the next 5 years. By participating in the symposium and engaging in other activities, the SC hopes to increase awareness about the ICDC and to keep the canine component of the larger data commons a high priority for NCI funding.

Dr. Hecht will continue to follow up as a central person on inquiries about the ICDC and will direct questions to the appropriate individuals. She also will track requests for proposals or other solicitations for the ICDC (e.g., through supplements or the Cancer Center Core grant mechanism), and she will report back to the committee with updates.

Links to the symposium agenda, poster abstracts, and podcasts can be found at https://www.cancer.gov/research/areas/childhood/childhood-cancer-data-initiative/symposium.

Reports from the Working Groups

Data Governance Advisory Board (DGAB) Chair's Report

The DGAB has met regularly over the past few months; the most recent meeting was August 20. The Chair reported that the discussions have been very productive.

The *initial* mission of the DGAB is to help prioritize studies for submission of data to the ICDC. The board has developed draft guidance for prioritization on the type of data for the ICDC. This guidance includes minimum data requirements and will evolve over time with ongoing feedback and revisions. The guidance document has been uploaded to the DGAB folder under the ICDC-SC folder on the Box site, which can be logged into using one's NIH account or individual log-in information. Mr. Beyers will forward the Box website link, the link to the guidance document/folder, and a copy of the draft guidance via email. It was noted that the easiest way to access the online information is via the folder link.

Further clarification regarding the format of the data to be submitted to the ICDC is needed. It was noted, for example, that the U24 holder, which is the coordinating center for the U01s, is collecting data from the canine clinical trials and related studies and has asked what data should be submitted to the ICDC (e.g., sequences, clinical trials and outcomes data, correlative data), how to submit data to the ICDC, and in what form(s) these data should be submitted (e.g., raw, annotated, analyzed).

The expectation is that the ICDC will be able to capture all types of data and the full spectrum of data, from raw data to clinical/outcomes data, to annotated and analyzed data. Data collection should be completed before being submitted to the ICDC, including for canine clinical studies. Because current trials are ongoing under the U01s, complete data sets should be available in about a year. The coordinating center is collecting clinical trials data, which is under their purview, but has conveyed that additional funding is needed to be able to collect and submit correlative data. Dr. Hecht will follow up in the coming year regarding possible options to provide these additional funds. The National Cancer Advisory Board (NCAB) Data Working Group has recommended submitting enhanced data sets to the data commons. It would be helpful to tie any request for funding for the coordinating center to the NCAB's recommendation.

The person heading the coordinating center, Dr. Qi Long, has contacted Dr. Otridge and requested a contact person for the DGAB. Per this request, the coordinating center can follow up with Dr. Kibbe or Mr. Beyers with any questions or concerns.

The Best Practices Subcommittee (BPS) guidance for standards for baseline data collection is forthcoming. The BPS will work with the DGAB (in coordination with the DGAB's initial and subsequent guidance) and the ICDC-SC in defining the types of data to be submitted to the ICDC. There was general agreement that raw data, clinical trial data, and correlative data should be uploaded to the canine commons. Whether "analyzed" data should also be submitted is not as clear, however, and depends on the level of analysis or how "analysis" is defined. A key aspect of the commons is to have users access raw and correlative data for their own analyses, within their own team and/or collaboratively with others. Meta-analyses and secondary analyses may need to be addressed separately.

Committee members discussed possible parameters and guidance for data submitted to the ICDC:

- The ICDC should include a broad range of data and data sets: sequencing and proteomic data, imaging data, mass spectrometry data, immune system data, and clinical data including outcomes with and without interventions.
- It is unlikely that any one study will cover all data for a complete/comprehensive analysis.
- Data from unique studies that address a limited or specific issue should not be excluded.
- Finished analyses should be published and referenced in the ICDC.
- The form in which raw data are submitted should not be overly proscriptive, but guidance should include how to curate data prior to submission to the commons. The standards developed by the BPS will define the format in which the various types of data will received by the ICDC.
- Although some flexibility is necessary, standardization of formats and nomenclature (including definitions and semantics) needs to be part of the guidance to assure as much uniformity across data as possible and to improve the ability to co-analyze data sets.
- Annotation of imaging data in particular is important for non-imagers to understand as it involves a highly specialized file format for computer-extracted characterizations. Thus, how to annotate imaging data should be clearly delineated in the guidance for data submission. Characterization of other types of data will similarly involve different levels of analysis and annotation (e.g., alignment and variant calling for sequence data).
- Alignment with guidelines of other repositories and data commons needs to be taken into consideration and built into the ICDC where applicable and appropriate.
- Customization of data collection and submission for the ICDC was discussed. An example of types of data that might be collected for normal vs. tumor tissue was provided. In this example, three types of data would be considered: sequencing data, variants and re-arrangements data, and neoantigen production. Within these data, different stages of disease progression and treatment response could be assessed and compared by different investigators to see what conclusions can be made. The ICDC should be able to accommodate this type of scenario, including secondary analyses and meta-analyses, but how to assure that the ICDC has the rich data sets to achieve this outcome still needs to be worked out. The canine clinical trials conducted under the U01s are required to be put in public databases, which would be ideal for the ICDC.
- In other repositories, it is common for annotations to be done on submitted data sets, but then the annotated data sets cannot be tied to the original data. How would this be addressed for the ICDC if, for example, variant calls differ and it is not possible to align variants with the initial data set?
    - In imaging repositories, when annotation results in a different outcome, a new DOI connected to the original data set is assigned to each annotation. This process is a workflow issue, not a technical capability, but it needs to be set up up front so that multiple analyses on the same primary data set remain tied to each other. The annotations, in turn, should be identified accordingly.
- Harmonizing data as they are submitted is an intensive process. Having clear guidance and standards for submission of data, including how data and the process for submission are defined, should reduce the extent to which the extra step of harmonization is needed.
- How the ICDC will tie into cloud resources, from data submission and access to analytical tools, also needs to be configured.

- Annotation and harmonization are workflow issues. These processes need to be built into the system up front.

The ICDC-SC recognized that developing and then implementing guidance for collection, curation, submission, and analysis of data for the ICDC will evolve over time to reach the goal of a successful process and workflow.

Best Practices Subcommittee (BPS) Chair's Update

Members and Chairs have been assigned to each of the four BPS working groups (WGs). An effort has been made to have a balance between NIH and non-NIH team leaders and members within the WGs. The Chairs and Co-Chairs for each WG include the following:

- Clinical and pathology standards: Dr. Chambers (Chair), Dr. London (Co-Chair)
- Genomics/proteomics/multi-omics standards: Dr. Heath (Chair), Dr. Zhao (Co-Chair)
- Immunology standards: Dr. London (Chair), Dr. Knapp (Co-Chair)
- Imaging standards: Dr. Jacobs (Chair), Dr. LeBlanc (Co-Chair)

Each WG will develop a data framework for their respective data types for the ICDC. One of the challenges is how to enable an analysis pipeline via the cloud, and what can be accomplished regarding this goal with the DGAB. The Imaging WG team leaders met since the last ICDC-SC meeting; the committee reviewed a summary of the objectives and initial plans for this group. The other three WGs had not yet formally convened as of the August ICDC-SC meeting, but there has been discussion regarding immunology standards. These initial efforts represent the first steps of an iterative process to bring together best practices that can be successfully applied and implemented for the ICDC.

Imaging WG Update

The overarching objective of the Imaging WG is to provide the comparative cancer research community with guidelines and best practices for the collection, formatting, and reporting of imaging data collected from canine companion animals with naturally-occurring cancers for the ICDC. The aims include harmonizing standards for imaging data to be deposited into the ICDC.

The initial discussions of the Imaging WG have focused on the curation of clinical images and imaging data for the ICDC. Key issues to address include identifying the types of clinical imaging and pathology data to be included in the ICDC and recognizing the unique aspects of having imaging data beyond clinical imaging data (e.g., CT, MRI, radiography), in particular scanned pathology imaging data. Other key issues include whether a unique patient-specific identifier should be established for every dog (e.g., implanted chip), use of a specific list of patient descriptors for each patient, and development of a de-identification plan/policy.

An existing process for the curation and sharing of imaging data gathered from human patients--The Cancer Imaging Archive (TCIA)--is available to serve as a model for these efforts. The TCIA has an on-line application process that captures the key features of individual imaging data sets and provides customized scripts for the submission process. The TCIA supports both clinical imaging data and pathology data (scanned tissue images) as well as multiple species (see https://www.cancerimagingarchive.net/primary-data/). The TCIA also has a process in place for curating data from outside contributors, which can be used as a prototype for the ICDC. Some

support to develop the processes for collection and curation of canine clinical trial imaging data into the TCIA is included in an established task order at FNLCR that is also for the collection and curation of human clinical trial imaging data.

The Imaging WG plans to explore these issues in a model (pilot) project using an existing high-quality data set that includes clinical data (no outcomes, just demographics), glioma diagnosis, MRI data, and pathology data (hematoxylin and eosin and IHCx5 stains) from 200 dogs with malignant glioma from about a dozen veterinary schools across the country. Most of these cases also are likely to have MRI data that were used to make a diagnosis, after which the dogs were either euthanized or put in a clinical trial or given standard care treatment. For a subset of this cohort (80–90 dogs), there are multi-omic genomic profiling data available through funding of a P30 supplement to the Jackson Laboratory (reviewed in Amin et al., *Cancer Cell*). Data were collected at necropsy and via biopsy for about 80% and 20% of these dogs, respectively. There should be some treatment and follow-up data for animals who have been biopsied. In addition, Dr. LeBlanc has access to all of the scanned pathology images from these dogs that she can share for this test project. The WG can follow up to see what clinical trial data and outcomes are available for this cohort.

The first steps would involve assessing how many dogs from this group have MRI data, then adopting a process to execute collection and curation of the MRI data into TCIA from one of the submitting institutions that has case material in the canine cohort. A small contract to this site would support beta testing of the process of conforming to submission standards and transferring the image data first into the TCIA and later into the ICDC portal whenever that system is configured to manage Digital Imaging and Communications in Medicine (DICOM) and pathology files. The process for use of pathology and clinical imaging is well established for humans and has also been tested in murine and other animal models. The process must be customized for each site, however, given differences in equipment and instrumentation and how sites use DICOM files and information. Some funding is available for acquiring and curating canine imaging data as part of a task order awarded with the archiving program. This pilot project would leverage the extensive clinical research experience with human data and permit necessary modifications for canine patients.

The group also should consider other data sets and tumor types as well, including models with deep interrogation that lead to clinical trials, whether for the pilot or further development of the ICDC prototype, to see whether the same data collection methods can be applied across studies and cohorts. For example, data from approximately 100 cases of canine pulmonary adenocarcinoma were recently published in *Clinical Cancer Research.*

Immunology WG Update
Immunology standards will be applied in three settings. Once the ICDC is set up and immunology data are being accrued, the standards for data and sample collection and analysis can be prospectively applied. In the early stages of the ICDC, the expectation is that data collected for other purposes can be mined for immune response. For data that cannot be repeated (e.g., cytokine assays), it will be important to record the methods used to generate these data.

A series of validated tools to allow for more quantitative analysis of immunology markers in canines is expected in the near future. No single approach is available, however, and in the interim, a set of immunology assays and data collection methods should be identified for use with the ICDC. For example, if major histocompatibility complex assays have captured canine data, they should be brought into the ICDC. On the other hand, although many assays for different tumor types, mutations, and markers are available for dogs, as for humans, most tests are not standardized. In the absence and limitations of methods and markers of immunohistochemistry and flow cytometry for dogs, preference should be given to testing using RNA sequencing and peripheral blood mononuclear cells.

A key resource for this component of the ICDC is the Pre-medical Cancer Immunotherapy Network for Canine Trials (PRECINCT), which provides infrastructure to a collaborative and interactive network of researchers and clinician scientists working at participating U01 sites to investigate immunotherapeutic strategies in dogs with cancer and identify correlates of immunological and therapeutic responses. PRECINCT has developed SOPs that can be applied across all PRECINCT-associated U01 projects to ensure high quality, repeatable data sets.

The WG leaders have been working with researchers and clinicians in the field and with PRECINCT, which has already generated a list of reagents and assays used in their canine studies.

The WG has also discussed using curated data based on established templated REDCap libraries and standardized collections. Because what is considered "normal" bloodwork can vary by study and institution, data can be identified as "within range" or "not within range." These curated libraries and a list of what should be collected will be available as templates for the ICDC.

Other Resources
There was a question as to whether The Cancer Genome Atlas has a best practices group and best practices standards and, if so, how that information could be incorporated into the ICDC. The group agreed that the ICDC should take advantage of existing models and best practices studies and perhaps consult with other groups for added value but also recognized that there is no one set of standards that can be used for the canine commons at this point.

WG Participation
The working groups are open to feedback and input from all ICDC-SC members. Those who are interested in participating or observing in any of the subgroups should contact the BPS Chair (Dr. Trent) or team leaders for additional information.

Other Issues

Attendance at Working Group/Advisory Board Meetings
ICDC-SC members who are not formal members of the BPS or DGAB are welcome to attend and participate in the subcommittee and board meetings or join the meetings as observers, if preferred.

Administrative Items

September ICDC-SC Meeting
The next meeting of the ICDC-SC will be held via teleconference on Wednesday, September 18, 2019, from 11:30 a.m. to 1:00 p.m. ET. Dr. Parchment will forward meeting information and materials ahead of the September teleconference.

DGAB and BSP Meetings
Dates and times of upcoming meetings are TBD. Details will be distributed to committee members when available.

Honorarium
A copy of the honorarium reimbursement form for ICDC-SC meeting participation was distributed ahead of the teleconference. External members were reminded to forward paperwork for their honorarium to Ms. Lydard.

Action Items

- Dr. Hecht will continue to follow up on inquiries about the ICDC, requests to share data, and funding opportunities for the canine commons
- Mr. Beyers will forward/email the link for the ICDC folders on the Box site, as well as the DGAB's draft guidance for prioritization and the link to the guidance document/folder on the Box site.
- Mr. Beyers will continue to post materials for the ICDC-SC, including meeting minutes, under the Box application.
- Topics for future meetings should be forwarded to Dr. Knapp and Dr. Parchment.

Adjournment

The meeting was adjourned at 12:35 p.m. ET.