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

Teleconference
Wednesday, April 10, 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*
Phillip Musk*
Jeff Trent*
Roel Verhaak*
Shaying Zhao*

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

Others
Lori Lydard*
Tara Whipp*
Mary Cerny (writer)*

Opening, Welcome, and New Member Introduction

Dr. Parchment opened the meeting at 11:30 a.m. EDT and welcomed those in attendance. He then introduced the newest member of the committee, Dr. Shaying Zhao, Professor, Department of Biochemistry and Molecular Biology, University of Georgia. Dr. Zhao is also a member of the Institute of Bioinformatics.

ICDC-SC Chair Introduction

Dr. Hecht announced the appointment of Debbie Knapp, DVM, MS, as Chair of the ICDC-SC. As Chair, Dr. Knapp will provide leadership for the ICDC-SC, the End User Working Group, and the task forces and will ensure that activities and plans that fall under those ICDC entities continue to advance.

Dr. Knapp is a veterinary medical oncologist at the Purdue University College of Veterinary Medicine. She also has served as Director of the Comparative Oncology Program at the College since 1992, and she is a program co-leader at the school's NCI-designated cancer center. Dr. Knapp's career has focused on naturally occurring canine cancer models and on raising awareness about this field of veterinary medicine and translation to human clinical research and practice. She is looking forward to working with the Committee members on this important mission.

The ICDC-SC congratulated Dr. Knapp on her appointment.

Minutes of the February Meeting

The minutes of the February 20, 2019, ICDC-SC meeting were accepted as written.

Data Governance Advisory Board (DGAB)

A DGAB has been established for the ICDC. The DGAB will be involved in the prioritization and processing of data requests for the ICDC. The DGAB's primary activity will be to evaluate requests for data submission, based on criteria that the Board will define, publish, and report back to the ICDC-SC. The DGAB will also report review metrics to the ICDC. Prioritized data requests will be submitted to NCI, and the NCI Executive Team will make final prioritization decisions. In addition, the Board will consider implications of the data set with respect to governance process for the overall NCI Cancer Research Data Commons (CRDC).

The DGAB will comprise six members, all of whom will be selected from the members of the ICDC-SC:

- Four external (non-NIH) ICDC-SC members
- Two internal (NIH) ICDC-SC members, including one member from the Center for Biomedical Informatics and Information Technology (CBIIT)

The DGAB will be supported by the FNL staff.

A draft process for how data submission requests will be reviewed and submitted to the ICDC has been proposed:

- Data requests will be submitted to FNL through a data submission package. FNL staff will review the requests to make sure they meet all initial submission criteria. Those people submitting the requests will have the opportunity to address questions about any missing items or information.
- Accepted requests will be forward to the DGAB, which will review and prioritize the requests based on its published evaluation criteria. Any questions and concerns identified by the DGAB will be provided to the data request submitter.
- The DGAB will forward its prioritized requests to the NCI Executive Team to finalize prioritization. As in prior steps in the process, questions raised by the Executive Team will be forwarded to the data request submitter.
- Upon approval by the Executive Team, the decision is forwarded/communicated to the FNL staff, who will inform the submitter of that decision and direct the submitter to work with the ICDC data submission team. Submitters will also be informed of requests that were not approved.
- Data for approved requests will be entered into the ICDC system.

ICDC-SC members will be contacted in the future regarding their interest in serving on the DGAB.

Questions/Discussion
It was noted that priorities will likely need to be re-aligned as the number of requests submitted increases. In addition, some requests might need to be designated as very high priorities at the time of submission, such as when data are needed for imminent publication, and thus might need to be moved up in the prioritization queue. The process for review and approval of data requests will include prioritization as well as a mechanism for re-prioritizing as the size of the commons and the number of requests increase and as otherwise needed.

Scientific Use Cases and Data

In creating and implementing the ICDC, the overarching question for NCI is whether pet dogs with spontaneous cancers can serve as close models of human disease so that new drugs, immunotherapeutic agents, and combinations can be evaluated for further development for human cancer patients. One approach to answering this question is to establish a publicly accessible canine database as an interoperable node in the larger CRDC.

The ICDC is one of several data-driven nodes that constitute the larger cloud-based CRDC. Canine use cases and data and supporting tools are the foundation of the ICDC. Plans are in place to set up a system to support the collection of data, but the data collected need to focus on answering specific scientific questions that the ICDC needs to ask and will be able to answer. The purpose of this discussion is to start the process of understanding and then delineating those questions. The discussion will be ongoing, and this process and questions will be used to build the ICDC structure in terms of the types of data and tools that the ICDC will contain.
The ICDC-SC's role will be to contribute to this process and to understand these questions over time.

As the features of the data and tools to answer these questions are better defined, they will be documented and prioritized.

Discussion

The Committee noted that while the ultimate goal of the ICDC is interoperability between the ICDC and the NCI data commons, initial steps should focus on identifying what works within canine data sets and the ICDC per se, including data analysis and tools. Committee members generally agreed with focusing on the ICDC first, with some members pointing out that one advantage of this project is that its goals and components do not have to relate back to humans as the first and only framework.

The Committee members explored the types of scientific questions ("use cases") for the ICDC and the types of data and tools to address those questions as follows:

Data collection, quality, and accessibility

Standardized data collection methods and data quality and accessibility are foundational to the scientific questions and operability of the ICDC node and the ICDC's interfacing with the CRDC.

Receipt of genomic, sequence, and proteomic data from the National Center for Advancing Translational Sciences (NCATS) has started, establishing import mechanisms. The ICDC will initially draw from existing data but, in the long term (i.e., within 2 years), will be populated with data and specimens collected prospectively from U01 trials. Other researchers have relevant data and may be willing to contribute to the ICDC; examples include data sets from individual labs that have focused on sequencing one or two canine tumors and their normal equivalents.

A preliminary data model with a longitudinal clinical trials arm has been chosen and is based on a Comparative Oncology Trials Consortium (COTC) trial. Steps to map the model onto the Cancer Genomics Cloud (CGC) have already started. Development of a uniform data collection model for the ICDC is discussed below, under "Annotation and data collection platforms."

Having data in an accessible form is another key issue. The form of the data uploaded to the ICDC (raw data/sequences, final analyzed data, or both) and the tools to research and access the data also need to be considered. Presuming that data are easily accessed, it will be important to delineate what questions will be asked regarding the data or data set. Operational questions are whether and how the data can be downloaded. Committee members noted, for example, that access to the cloud may be a barrier for some groups, which will limit their ability to use data and tools if the ICDC system is cloud-based only. Alternative mechanisms to assist these researchers and teams who lack access to the cloud or who cannot convert their existing pipeline to a cloud-compatible pipeline will need to be considered. One suggestion was to have a data coordinating center with a small bioinformatics core to run an offline harmonization pipeline. Another component is how the canine data sets will be linked to the human data sets and how cross-linking to all data sets and nodes will be achieved. As a starting point to envision the interface within and across data sets and nodes in the ICDC and CRDC, the Committee might

want to draw up paper-based cartoon mock-ups that show all aspects of the data use/access process.

The DGAB will evaluate the quality of the data, including assessment of the protocols used to collect data and specimens to ensure that the ICDC is populated with high-quality data.

The Committee also discussed having a collaborative group set up pre-submission/pre-uploading standards for genomic data for the ICDC.

Data management and processing
Experience with the CGC shows that the process of managing incoming data on an ongoing basis is very time-consuming and requires substantial budgetary support. The plan for the ICDC is to provide a high level of support for data collection, processing, and submission in the early stages of the project, with users eventually submitting their own data as the ICDC expands. Per this model, there will be a core critical mass of harmonized data to serve as the foundation at the outset of the ICDC. An SOP or guideline can be developed once the data collection and submission process is clearly defined. Whether the same level of management in place for the CGC will be applied to the ICDC will be based in part on available funding. Having a data coordinating center that includes offline operation of a data harmonization pipeline, as discussed above, could be a cost-effective alternative to a more comprehensive data management plan.

Comparative analyses
Comparative analyses would focus on advancing the understanding of relationship between canine tumors and human cancers. The types of cancers to be studied (e.g., mammary, bladder) and the types of comparisons that would be analyzed in dogs versus humans would need to be identified for the ICDC.

A key question for comparative analyses is the ability to understand how mutations in a canine tumor translate to the human genome. From the ICDC perspective, the scientific questions at the mutation level would be to compare mutations in canine tumors with those in human tumors and to compare gene expression and amino acid profiles between the two species. Assessing correlations between the mutation/mutational status/tumor microenvironment and clinical outcomes (e.g., pathology, resistance, long-term outcomes, response to treatment, relapse) will be critical to comparative analyses and is consistent with the broader aims of the CRDC. The available data on canine tumor mutations would need to be taken into consideration. It was noted that some canine data sets may have as many as half a million mutations that could be mined.

Collecting data that can be tested in high-profile comparative use cases is an end point for the ICDC. A more immediate goal would be to determine whether it is possible to query the data in the shorter term to get meaningful results to the scientific question(s) being asked.

Predictive outcomes
The ICDC could go beyond outcomes such as treatment response and question how to survey cases to determine the predictive capacity of the data for specific cancers (e.g., disease progression, survival). For example, approximately half of bladder cancers in both people and

dogs progress to metastatic disease; such progression appears to be related to molecular subtypes.

Annotation and data collection platforms
There currently are no standards in the field for clinical annotation with respect to outcomes, which presents a challenge for analyses that involve more than genomics. The ICDC therefore will need to develop a standardized process for collecting data in a meaningful and relatively easy way in the clinical setting. It would be helpful to look at other structures for clinical outcome data and then beta-test them to see how they work within community veterinary practice and as possible prototypes for the ICDC.

An underlying concern with canine genomic data is the adaptation of pathogenic variants, including somatic substitutions and translocations. Developing pipelines and tools that can identify and annotate these variants and their putative pathogenicity remains a fundamental question and would be a good starting point to explore using existing data sets from various labs. Being able to easily move between human and canine genomic data sets to see, for example, whether mutations are in the same regions in the two species or pathogenic in only one species is a very useful tool. Conservation is used to annotate sequence variance, and conservation metrics, while not considered robust comparative genomics, can be built into the process to understand the pathogenicity of individual variants.

One team is looking at the Tufts University Cummings School of Veterinary Medicine common data platform, the Observational Medical Outcomes Partnership (OMOP), to extract data from pets' medical records. Initial efforts indicate that this process is somewhat challenging. Unlike with human patient records, reimbursement is not tied to diagnostic information; thus, the coding is not always consistent. The framework based on the Tufts platform could be distributed through the Clinical and Translational Science Award One Health Alliance (COHA). Another model that uses a platform similar to that at Tufts but for human clinical data is the Oncology Research Information Exchange Network (ORIEN) consortium, composed of 19 comprehensive cancer centers, all of which have adopted a single, uniform set of data. Canines-N-Kids Foundation, which promotes research that integrates efforts for the benefit of both kids and dogs with cancer, is an additional resource that should be considered.

Committee members working with these or similar platforms and other members who can share data or additional relevant examples could work together in a subcommittee to identify best practices as a starting point. Uploading information from different data sets can also present challenges. One member suggested bringing in a programmer with experience in extracting data from medical records and moving the data into a single platform. The subcommittee could choose an existing data set (or sets) to closely review as prototypes for the ICDC. As the characteristics of data sets vary, this task will assist the ICDC-SC in designing and shaping the platform for the canine commons.

It was noted that the Common Terminology Criteria for Adverse Events (CTCAE) guidelines for canine clinical trials are being updated. This update goes hand in hand with plans to revisit the practices for reporting and standardization of clinical data writ large. The revised CTCAE guidelines can be incorporated into the ICDC platform.

Among the key issues identified during this discussion will be explored further are:

- Annotation and standardization of the annotation of clinical outcomes
- Identification of the types of data to be accessed and the questions to ask of the data
- Setup of a subcommittee to identify best practices for data collection and uploading and platform development to unify the data fields to align clinical information and genomic data
- Promotion of collaborations to share/donate data
- Establishment of linkages across all data types between humans and dogs (aspirational)
- Development of a cloud-based model for access, standardization, and analysis of data across groups (aspirational)

The ICDC-SC will continue to discuss these and other ideas at future meetings.

Administrative Items

May meeting
The next meeting of the ICDC-SC will convene via teleconference on Wednesday, May 22, 2019, from 11 a.m. to 1 p.m. EDT. Topics for discussion should be sent to Dr. Knapp, Dr. Parchment, Dr. Hecht, or Mr. Beyers.

COI and honoraria
Committee members were reminded to complete and submit their COI disclosure form (if they have not already done so) and their honorarium paperwork so that the honorarium for their participation and meeting attendance can be processed. Dr. Parchment will distribute the required forms with the May meeting notice.

ICDC-SC member profiles
Committee members were asked to put together a short profile (two or three sentences) of their research interests and how they align with the ICDC. The profiles should be forwarded to Dr. Parchment or Mr. Beyers.

Action Items

- Mr. Beyers will contact ICDC-SC members, asking them to forward a brief bio/profile on how their research interests and background fit with the ICDC.
- The ICDC-SC will set up the best practices subcommittee. Members volunteering for the subcommittee include Drs. LeBlanc, London, Ostrander, Trent, and Zhao.
- External ICDC-SC members were reminded to submit their COI disclosure form (if they have not already) to Dr. Parchment and to forward paperwork for their honorarium to Ms. Lydard.
- Topics for future meetings should be forwarded to Dr. Knapp, Dr. Parchment, Dr. Hecht, or Mr. Beyers.

Adjournment

The meeting was adjourned at 12:38 p.m. EDT.