

E.5 Quality Standards of Samples and Clinical Data

Introduction

The ICGC in its first phase overcame the major challenge of generating collections of high-quality tumor samples and committed partners and funding agencies invested substantial effort and funds to ensure genomic data were generated from the highest quality samples. In this next generation of the ICGC, clinical data is deemed to be the major challenge: obtaining, curating and harmonising large sets of detailed clinical data from a multitude of tumour types and programs globally. As there is no current standardized clinical trials data set that spans across all cancer types, data fields and values will be adopted that explicitly suit ARGO requirements. While it was intended in the first phase of ICGC to require such clinical data, in practice this has only been accomplished within a few projects. The high standards for clinical annotation are known from the outset, and projects will be required to allow for significant investment, resources and efforts to build the necessary processes to curate and submit comprehensive clinical data sets.

The Tissue and Clinical Annotation Working Group (2018-2021) has developed quality metrics for clinical data and the more recently formed Clinical and Metadata Working group, will harmonize standardisations and implementation of clinical nomenclature and data values.

POLICY: Every project will adhere to the following recommendations regarding Quality of samples:

1. Tumor types should be defined using the existing international standards of the WHO (including ICD-10 and ICD-O). If novel molecular subtypes are studied, these should be defined with sufficient detail
2. All samples will have to be reviewed by two or more reference pathologists. This assessment will need to be performed on stained sections of the very same tissue piece from which biomolecules will be purified. Histological examination should be documented and respective high-resolution digital images have to be stored and made available i) to those studying the given samples and ii) on a dedicated web-page for open access. The Pathology Working Group will provide guidance
3. Patient-matched control samples, representative for the germline genome, are mandatory to discern “somatic” from “inherited” mutations. For solid tumors, the mononuclear cell fraction from peripheral blood is the ideal source, while for hematological malignancies skin biopsies or (lymphocytes from patients in remission) are recommended.

POLICY: Every project will adhere to the following recommendations regarding quality and submission of clinical data:

1. Member programs and projects commit to submitting the Mandatory Clinical Data set for each participant. The mandatory data elements are required to address clinically relevant analyses within as well as across entities. These data points constitute the critical elements

of clinical correlation to allow harmonization of diverse ICGC ARGO projects, and will be required as a minimum. All of these data points are commonly acquired in cohort-based studies (patients studied outside of clinical trials such as observational and longitudinal studies, retrospective or prospective) and clinical trials and, therefore, are in principle available. Project leads are required to ensure that projects can meet the standard for the Core data elements and where not able to, submit an exception for review by the Clinical and Metadata Working Group.

2. Further information regarding data submission is contained in the Data Management Policy, and through the [ICGC ARGO Documentation site](#). Status of completion of Clinical Data Submission will be available to users through the individual member program dashboard.

Guidelines regarding the quality and submission of clinical data:

1. Acquisition of follow-up information is highly recommended on an annual basis for collection of updated treatment and outcome information. This will inform subsequent interpretation of ICGC data and clinical correlations
2. Clinical data will be submitted to ICGC ARGO using controlled vocabulary as detailed in the [Data Dictionary](#), which has been developed with consultation from programs, and wherever possible on international standards, such as ICD (WHO), AJCC, or from widely used matrices (in particular those used the Genomic Data Commons, IARC and others) to allow co-aggregation with data from these sources. The Data Dictionary defines the clinical data model and includes rigorous validation performed as quality control steps at the time of submission.
3. Generation of an Extended Data Set is under way consisting of additional variables that are recommended for the analysis of biological processes that are considered hallmarks of cancer etiology and progression. These data points will encompass detailed lifestyle, predictive and prognostic factors, family history information and additional treatment and response data along the trajectory of individual therapies. Data sets will likely be developed within specific tumour groups, and this extended data is encouraged to be completed by regulated clinical trials or where deeper clinical data is available
4. All Core data must have a valid value submitted for all fields for a clinical data submission to be classified as **complete**. Unless an exception has been granted formally (see below section: Clinical Data Exceptions Process). A donor must be clinically complete before any molecular analysis files are released to program members. Specifically;
 - i. A donor must have a donor file submitted with all core fields provided (unless an exception is granted, see below).
 - ii. A donor must have at least one primary diagnosis with all core fields provided.
 - iii. A donor must have at least one tumour and one normal specimen submitted.
 - iv. For each registered specimen, a donor must have all specimen core fields provided.

- v. A donor must have at least one treatment and a corresponding treatment detail file (if applicable, e.g. for chemotherapy, hormonal therapy or radiation) with all core fields provided.
- vi. A donor must have at least one followup with all core fields provided.
- 5. Exemptions may exist where a data element is not applicable to a particular tumour type. These cases must be documented appropriately through the submission process following the guidelines of the Data Dictionary.
- 6. Ensure, where appropriate, the sustainability of the data submitted through both archiving and using appropriate identification and retrieval systems
- 7. Member projects and leads should facilitate a process for the demonstration of traceability of data, including Good Documentation Practices, and these be documented in the program or institutional Standard Operating Procedures (SOP).

Guidelines regarding the quality standards of samples:

- 1. Histological examination will have to be documented and respective digital images be stored and made available to those studying the given tumor entity. Specifically the degree of 1) necrosis; 2) debris; 3) inflammatory tissue; and 4) fibrosis are to be assessed
- 2. Standard Operating Procedures (SOPs) for freezing samples will be those established by WHO/IARC ([“Common Minimum Technical Standards and Protocols for Biological Resource Centres dedicated to Cancer Research”](#) by the World Health Organization - International Agency for Research on Cancer (WHO-IARC, working group reports Vol.2, 2007)
- 3. As a basis for the exchange of tissue specimens between countries with different national regulations that need to be respected, a coordinating rule has been formulated on the basis of the ‘home-country principle’
- 4. Although many types of macromolecules should be isolated, priority should be given to the isolation of high quality DNA (which is also valid for some epigenomic analyses)
- 5. The quality of the isolated classes of macromolecules needs to be controlled by standardized procedures used by all members of the ICGC. The choice of these tests will be defined by an ICGC working group
- 6. Controls for transcriptomic and epigenomic analyses may require site-matched tissue control samples. This aspect must be dealt with in the recommendations of the tumor-specific expert panel.

Clinical Data Exceptions Policy

As clinical data forms a central part of the ICGC ARGO mission; management and governance of it is critical to ensure the balance between maximum engagement and program requirements. Due to the comprehensive nature of the ARGO clinical data model, it is accepted that some groups will need margin for exceptions, particularly in cases involving retrospective data, disease specific circumstances, availability or inaccessibility of data. ICGC has a standard set of criteria and a consistent approach to assessing applications for exceptions as outlined below.

Key Points:

- The Core clinical data elements are mandatory, but within this Core Data set there are critical elements that are not subject to exceptions. These include key donor attributes and clinical endpoints such as treatment response and survival data. As these elements are vital to answering ARGO's research questions, cases missing this information would be excluded.
- Exceptions are rare and granted on a case by case basis, thresholds may exist due to technical capability.
- Programs submit a form containing a standard set of detail surrounding the rationale for the request and numbers involved- requests are then reviewed and discussed centrally with clinical expertise involved.
- Exceptions that are related to inherent tumour type conditions- ie tumour grade in blood cancers, will be built into the validation rules and these will not be required to be submitted as exceptions.
- Projects which are prospective in nature or are regulatory grade clinical trials are expected to meet the requirements for all Core clinical data elements and are discouraged from submitting exceptions.

Process

1. **Programs submit an [exceptions request](#)**
2. **Requests reviewed by the Clinical and Metadata working group**, discussed and decision reached. Review will consider the type of program (questions being asked, retrospective vs prospective data etc), value of dataset, rationale for exemption (if legitimate for tumour type, country, etc) and the potential impact to overall data set if field not provided. If a request spans a considerable percentage of the data sets donors an exception may be granted to the entire data element for that program.
3. **Outcomes communicated to applicants** with a full justification for the decision.
4. **Approval forwarded on to the DCC** where technical edits are put in place to allow the exemption. This is logged and documented.
5. **DCC provides confirmation** to the program/applicant to allow for data submission.