MARATHON OF HOPE: GOLD COHORT DATA
Guideline for Institutions

OVERVIEW

The Marathon of Hope Canadian Cancer Network (MOHCCN) initiative funded by Health Canada is in the process of creating a "gold standard" cohort of clinical cancer specimens with a well-annotated, uniformly generated, and consistently quality-controlled dataset (including clinical and genomic) from 15,000 (15k) cases collected from across Canada over 5 years.

Each MOHCCN case is defined as a unique patient and includes whole transcriptome, whole genome, and clinical data (over 190 data fields including specimen, primary diagnosis, biomarker, surgery and treatment fields). This description outlines the minimum data set however, imaging, health technology assessment and additional disease-specific data points can also be included (Figure 1). The gold cohort includes various cancer types from breast, prostate, GI and brain cancers to melanoma, leukemia and pediatric cancers.
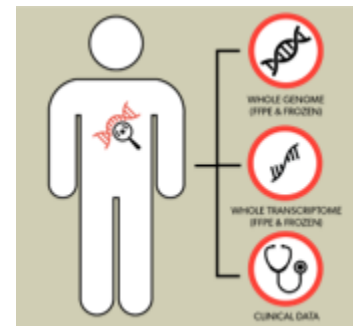

Figure 1: MOHCCN Case

MOHCCN DATA PATHWAY

Supporting the development of the MOHCCN Gold Cohort, four main components of the MOHCCN data pathway have been identified (Figure 2). Network committees, working groups and technical partners are engaged in activities across the pathway to support the generation, curation, ingestion, and utilization of the MOHCCN Gold Cohort.
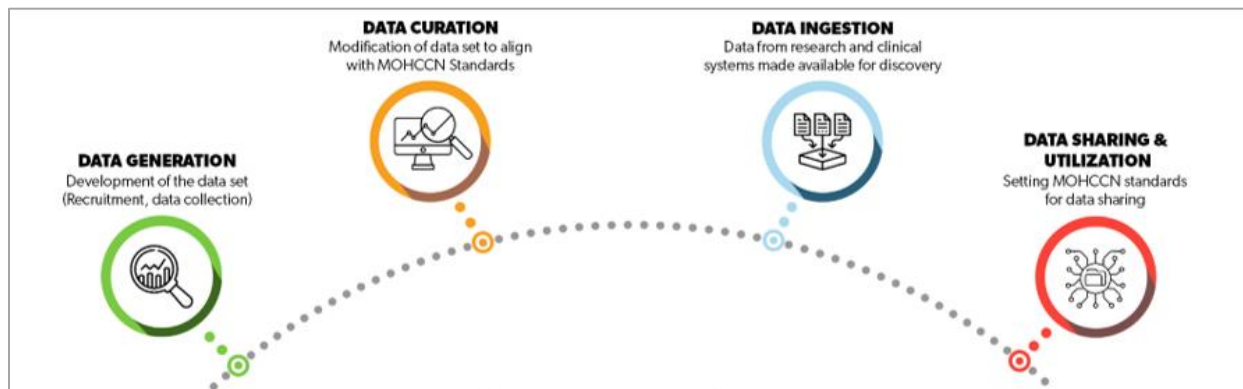

Figure 2: MOHCCN Data Pathway

DATA STANDARDS, POLICIES AND GUIDELINES

Supported by MOHCCN Committees and Working Groups, the development of [MOHCCN policies and guidelines](#) are key components required for the development of the Gold Cohort and responsible sharing of Network data. These include:

DATA CURATION

MOHCCN Gold Cohort Policy: The Technology Working Group (TWG) has developed a policy to define the data requirements to complete a MOHCCN Gold Cohort Case. MOHCCN data standards are available for all required data points. A complete list can be found in the Gold Cohort Data Standards Policy.

MOHCCN Clinical Data Model v2: The Data Standards Sub-Committee (DSC) is a sub-committee of the Data Policies and Standards Committee. The DSC has developed the MOHCC clinical data model and corresponding guidelines to support data collection.

DATA INGESTION

Technology infrastructure: As a Network, MOHCCN is collaborating across top hospitals and cancer research facilities in Canada to share genomic and clinical data from thousands of patients across the country. This data will then be analyzed using artificial intelligence and other state-of-the-art technologies and methods, allowing researchers to pinpoint exactly how cancers are formed.

Together with the MOHCCN, The Canadian Distributed Infrastructure for Genomics platform (CanDIG) is helping to close the gap between research in the lab and patient care in the clinic by providing a Canadian-developed, open-source solution that takes into account the multi-jurisdictional nature of Canadian privacy and security policies. Through CanDIG (Figure 3), data will be made available via the Digital Health Discovery Platform (DHDP).

1. CanDIG provides discovery of national-scale health genomics and associate clinical and phenotypic data.
2. DHDP provides the platform for health researchers, AI and data scientists and industry

Figure 3: The bolded icons represent locations where CanDIG is deployed to ingest Network data. The additional icons show the locations identified for future deployment.

**DATA
SHARING &
UTILIZATION**
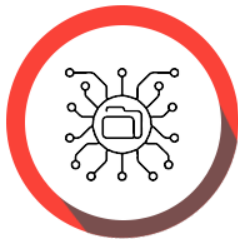
Network Agreements: Binds Network Member (institutions) to share coded genomic and health data across the network in a timely manner, according to Network policies. *For more information about Network Agreements please refer to the MOHCCN Network Agreement Guideline.*
MOHCCN Data Access and Use Policy: The Data Policy and Standards Committee (DPSC) has developed a policy establishing timelines for sharing Network Data with Network members and External users, including commercial "for profit" companies. Timelines apply to each patient case from the time when the sequence data is generated.
MOHCCN Code of Conduct: In alignment with the Data Access and Use policy, the aim of the code of conduct is to promote research integrity and collaboration.
MOHCCN Publication and Authorship Policy: This policy defines Network publication types and authorship/ acknowledgement guidelines.

The policies and guidelines listed above are endorsed by the MOHCCN Network Council. Additional recommendations, policies and guidelines that support Network activities are in development by the MOHCCN committees and working groups. Members of the Network are invited to participate through MOHCCN committees and working groups in the formation and ongoing refinement of Network recommendations, policies and guidelines.