Informatics

ABOUT INFORMATICS

Our mission is to advance the knowledge and treatment of cancer through computational biology.

The research objectives of Informatics are to:

- Gain new and deeper understanding of cancer biology through the application of computational and data-intensive techniques;
- Develop information systems that apply this understanding to cancer clinical care;
- Harmonize and disseminate large cancer-related data sets;
- Train the next generation of computational biologists to work on cancer-related problems;
- Foster efficiency, communication and collaboration within and among Informatics, OICR and the wider community.

Informatics has a broad set of research interests ranging from open research and reproducibility, to software development for big data, long-read sequence algorithm development, biomarker discovery, viral detection, and population-based genomics approaches to cancer mutations, as well as pathway and network analysis. While our research activities and expertise focus on cancer, they also have broader application in genomic research.

Informatics' research expertise in cancer informatics is internationally recognized and sought-after. Our seven principal investigators and senior scientists bring extensive expertise to the representation, evaluation and interpretation of cancer genomes, particularly in the following areas:

- Flexible, scalable databases
- Public data web portal development
- Algorithm and software development (for long-read sequence alignment, methylation analysis, viral detection, etc.)
- Big Data analytics
- Population genomics
- Pathway and network analysis
- Biomarker discovery
- Open, reproducible research

TECHNOLOGIES AND INFRASTRUCTURE

To facilitate Informatics' ongoing large-scale, complex data analyses and software development projects, OICR has built and maintains an in-house state-of-the-art compute infrastructure. This shared resource consists of 1) a High Performance Computing (HPC) environment, with over 8,000 cores, an aggregate of 30+TB of memory and 6PB of network-attached storage; and 2) an onsite Cloud Computing environment (OICR Stratus OpenStack and OICR mcVMM). OICR's Research IT Team expertly maintains these services using procedures and architectures that provide availability, performance, and security. Research IT provides user support, infrastructure services and business information systems for all MaRS-based OICR staff, as well as a secure IT infrastructure to house provincial, national and global portals such as the International Cancer Genome Consortium (ICGC), Ontario Health Study, and Canadian Partnership Against Cancer among others.

Internally, our Informatics researchers use the HPC environment to process sequence generated by the Genomics Program, as well as other algorithm development tasks. Stratus is an OpenStack Cloud Computing infrastructure, which aims to be a stable, scalable and high-performing platform to enable OICR and colleagues to self-provision deployments and software on a fee-based model.

SERVICES

Informatics offers the following services at a competitive rate to all OICR-affiliated researchers.

1) The **Genome Sequence Informatics (GSI)** Team within Informatics offers tools and workflows for sequence analysis and quality control. The team has extensive expertise in a wide variety of cancer sequence data types generated on various next-generation sequencing platforms. Using highly automated and reproducible workflows, GSI provides both primary and secondary analyses, and can also perform bespoke analysis using specific cutting-edge software. Our analysis team provides assistance with experimental design, as well as the interpretation and the publication of experimental results. Data storage and secure transfer or sharing options can also be provided.

An example of some of the analysis services we provide include:

- Whole genome and exome;
- RNA-seq;
- Targeted sequencing for standard and custom gene panels;
- Difficult sample analysis (low cellularity; degraded samples such as FFPE; molecular barcoding);
- Variant calling, including SNVs, copy number and structural variants;
- ATAC-seq;
- Methylation seq;
- Integration of public data sets; and
- Pathway and other functional analyses.

2) OICR's **Research IT** enables the strategic objectives of the Institute by providing investigators with compute infrastructure access for rapid and intelligent data analysis, modeling, storage and transfer. The following subset of OICR Research IT services are available to OICR-affiliated researchers:

- OICR Research IT managed virtual machine (VM) hosting with or without Internet access;
- OpenStack-based tenancies for user managed VMs with or without Internet access;
- High Performance Compute (HPC): OICR can provide offsite remote access to OICR's Grid Engine-based HPC compute infrastructure to approved collaborative researchers. Queue and job resources (number of CPU cores, amount of RAM, job runtime, etc.) are assessed and allocated on a case-by-case basis;
- Scratch/temporary storage;
- Mid-term storage for a maximum of four years (may be extended or renewed);
- Vulnerability assessments and threat risk assessments; and
- DNS, Directory, SSL Certificate management.

3) The **Web Development Team (WEBDEV)** designs and builds websites and web applications for OICR's researchers and their programs. Through its website products, the Web Development team facilitates awareness of OICR's research activities and promotes OICR's translational research initiatives. Its products also enable bioinformatics training, international research online engagement and provide tools and data services to harmonize and disseminate large cancer-related data sets. Exemplar websites include OICR's corporate sites, Global Alliance for Genomics and Health, ICGC sites, Genomic Data Commons, Bioinformatics.ca and Cancer Collaboratory, among numerous other internal and external sites. Development rates are calculated hourly per team member's contribution to the final product.

The following services from the WEBDEV team are available to OICR-affiliated researchers:

- Custom web application design and development services;
- Custom design and production services; and
- Creative and technical analysis - user interactivity, usability and content production.

4) **Cancer Genome Collaboratory (cancercollaboratory.org)** is a unique compute cloud-based biomedical research resource designed to make the extensive genomic holdings of the ICGC project available in convenient form to the cancer research community. Researchers can create a project and request an account on the compute cloud.

5) **Canadian Data Integration Centre (CDIC, genome-cdic.ca)**, which supports large scale genomics projects on population-wide and clinical cohorts, and provides analytical and bioinformatics support through access to the software and analytic systems needed to collect and harmonize diverse health and lifestyle data, analyze it and electronically publish the results. Researchers can request access and services for their project needs.

FEATURED PROJECTS

One of the strengths of Informatics has been its leadership and scientific roles in many highly recognized, successful international research projects.

We have played a very visible and collaboratory role in many cancer projects, including the following initiated at OICR:

- The **International Cancer Genome Consortium (icgc.org)**, where Informatics housed the ICGC Secretariat and designed, built and maintained the **ICGC Data Coordination Centre (ICGC DCC, dcc.icgc.org)**, a portal providing tools for visualizing, querying and downloading the data released by ICGC consortium members;
- **Canadian Prostate Cancer Genome Network (CPC-GENE)**, a project aimed at understanding the prostate cancer genome to better predict treatment failure for intermediate risk prostate cancers;
- **PRONTO**, a pan-Canada research project to rapidly develop novel diagnostic markers for early prostate cancer; and
- **Genome Data Commons (gdc.nci.nih.gov)**, a unified data repository supporting the data sharing of several cancer genome programs at the NCI, including The Cancer Genome Atlas.

As well, we have produced many seminal, open-source, open-access community resources including:

- **Reactome (reactome.org)**, an open, curated knowledgebase of biological pathways in humans;
- **WormBase (wormbase.org)**, an online biological database about the biology and genome of the nematode model organism Caenorhabditis elegans; and
- **Dockstore (dockstore.org)**, developed by the Cancer Genome Collaboratory, is an open platform for sharing Docker-based tools described with the Common Workflow Language used by the Global Alliance for Genomics and Health.

CONTACT INFORMATION

TECHNOLOGY PROGRAM:	Informatics
INSTITUTION:	Ontario Institute for Cancer Research (OICR) MaRS Centre, 661 University Avenue, Suite 510, Toronto, Ontario, Canada, M5G 0A3
PROGRAM DIRECTOR:	**Lincoln Stein** Lincoln.Stein@oicr.on.ca | 416-673-8541
PROGRAM MANAGER:	**Michelle Brazas** Michelle.Brazas@oicr.on.ca | 416-673-8502

Funding for OICR is provided by the Government of Ontario