BACKGROUNDER

Results of Genome Canada 2012 Bioinformatics and Computational Biology Competition

Genome Canada, in partnership with the Canadian Institutes of Health Research (CIHR), is pleased to announce the 17 successful projects resulting from the Genome Canada 2012 Bioinformatics and Computational Biology Competition. The Competition supports the next generation of tools that are required to deal with the large amounts of data produced by modern genomics technologies and to provide the research community broad access to these tools. Listed by region from West to East, the successful projects are as follows:

BRITISH COLUMBIA

Project Title:	**Next generation bioinformatics for clinical genomics: using de novo assembly in personalized medicine**
Project Type:	**Large-Scale Applied**
Project Leaders:	**Drs. Inanc Birol, Steven Jones and Aly Karsan**
Institution:	**BC Cancer Agency Genome Sciences Centre**

Genomics technologies that detect variations and mutations in DNA and RNA can advance cancer care and reduce health care costs by improving preventive care, diagnosis and treatment. Advanced high throughput DNA and RNA sequencing technologies can help realize this vision by generating large amounts of sequence data rapidly and at low cost.

However, solid analysis of the generated data is essential to reach its full potential and will provide the backbone to application. Drs. Inanc Birol, Steven Jones, Aly Karsan and team are developing an analytical approach to detect variations and mutations in DNA and RNA related to cancer diagnosis and care. This approach could lead to more efficient and effective clinical testing for various types of cancers across Canada.

Project Title: **A federated bioinformatics platform for public health microbial genomics**
Project Type: **Large-Scale Applied**
Project Leaders: **Drs. Fiona Brinkman, Gary Van Domselaar and William Hsiao**
Institution: **Simon Fraser University**

Despite recent advances in medical, sanitary and public health practices, diseases communicable via bacteria still remain a serious threat. Whole genome sequencing of disease-causing bacteria can provide comprehensive data to identify and track these organisms. However, the current complexity of analytic tools makes it difficult for public health investigators to use and share genomics information in a timely and effective manner.

Drs. Fiona Brinkman, Gary Van Domselaar, William Hsiao and team aim to address these gaps by providing public health workers with easy-to-use bioinformatics and genomics analytic tools to allow them to better manage communicable diseases and provide quicker responses to infectious disease outbreaks.

Project Title: **Computational interpretation of cancer genomes: defining mutational landscapes for translational genomics**
Project Type: **Large-Scale Applied**
Project Leaders: **Drs. Sohrab Shah and Paul C. Butros**
Institution: **BC Cancer Agency/University of British Columbia**

Tumours develop through the accumulation of mutations in DNA. Recent advances in high-throughput (next-generation) DNA sequencing allow researchers rapid identification of mutations in a genome. This has increased understanding of the biology of cancer cells, and has led to more effective drugs and better predictions of patient outcomes.

However, maximizing the clinical use of next-generation sequencing data requires sophisticated software to improve the analysis of genomes and identification of mutant sequences related to tumours.

Drs. Sohrab Shah, Paul C. Boutros and team are developing innovative software that will improve patient care by identifying and analyzing the mutations involved in cancer progression.

Project Title: **Applied bioinformatics of Cis-regulation for disease exploration (ABC4DE)**
Project Type: **Large-Scale Applied**
Project Leader: **Dr. Wyeth Wasserman**
Institution: **University of British Columbia Centre for Molecular Medicine & Therapeutics**

The goal of personalized medicine is to treat patients in the manner that is most appropriate for each individual. Before this can happen, doctors will need software that can perform detailed high-speed and low-cost analyses of patients' specific genetic mutations.

Dr. Wyeth Wasserman and team are developing software that will understand and categorize the pieces of DNA that help turn genes on or off. These small sequences of DNA are spread throughout the human genome and serve a critical role in controlling when and where genes are turned on. Mutations in these on/off switches can cause birth defects, disease risk and adverse drug reactions.

The software will help physicians analyze patients' genetic mutations and pave the way for personalized medicine.

Project Title: **Tool for proteome-wide identification of regulatory switches**
Project Type: **Small-Scale Innovative**
Project Leader: **Dr. Joerg Gsponer**
Institution: **University of British Columbia**

Despite significant progress in diagnosis and treatment, cancer remains Canada's leading cause of death. Although scientists have made major efforts in identifying mutations in some cancers, it is still not known how these mutations cause cancer. Cancer is often related to the disruption of regulatory mechanisms in the cell, including auto-inhibition, a process that allows proteins to switch their functions on and off. Mutations can alter these protein switches, which can lead to changes in cell behaviour and ultimately cancer.

However, there is no easy way to determine when cancer-causing mutations affect auto-inhibitory switches. Dr. Joerg Gsponer and team are developing a new method to identify auto-inhibitory switches through the use of genomic and proteomic information.

Project Title: **A compressed sensing framework for identifying differentially expressed isoforms and transcriptomic aberrations in cancer samples**
Project Type: **Small-Scale Innovative**
Project Leaders: **Drs. Cenk Sahinalp and Colin Collins**
Institution: **Simon Fraser University**

New technology provides fast and accurate ways to analyze RNAs (which act as messengers carrying instructions from DNA) that code protein in a tissue. While there are many potential RNA products in a gene, it is believed that it is only necessary to identify and quantify a small number of RNA products in order to get information about a sample's RNA content.

Drs. Cenk Sahinalp, Colin Collins and team are using a computing technique called *compressed sensing* to find the smallest possible number of RNA products from each gene. Their approach will also help pinpoint the essential differences between samples of RNA products from the same gene. Sahinalp and Collins will also use *compressed sensing* to identify tissues from prostate cancer patients that develop into more aggressive forms of the disease.

Project Title: **Measuring and modeling tumour evolution from next generation sequencing data: enabling clinical study of clonal diversity in cancer patients**
Project Type: **Small-Scale Innovative**
Project Leader: **Dr. Sohrab Shah**
Institution: **BC Cancer Agency**

Breast and ovarian cancers are significant causes of disease and death among North American women. Tumours in these cancers can acquire different mutations, resulting in cells that may respond differently to therapy. However, this genetic diversity within tumours is rarely considered when it comes to treatment, even though it is believed to contribute to drug resistance and disease progression.

While new sequencing technologies have provided some insight into the nature of tumour evolution, it is still unclear how evolutionary processes contribute to cancer.

Dr. Sohrab Shah and team are using sequencing data gathered from breast and ovarian cancer patient samples to create software that will improve understanding of tumour evolution and help predict clinical results.

ONTARIO

Project Title: **MedSavant: An integrative framework for clinical and research analysis of human genomes**
Project Type: **Large-Scale Applied**
Project Leaders: **Drs. Michael Brudno and Gary Bader**
Institution: **University of Toronto, Centre for Computational Medicine, The Hospital for Sick Children**

Physicians will soon be able to use patients' whole genome sequence to search for information about the person's risk of developing a disease, thereby improving clinical decision-making. This promises significant medical and economic benefits, including early detection and treatment of high-risk patients and eliminating multiple genetic tests.

Integrating whole genome sequencing into clinical practice requires software that will allow clinicians to identify relevant genetic variants in patients. Drs. Michael Brudno, Gary Bader and team aim to improve health care by developing broadly shared software that will prioritize the genetic variants in patients who may require medical attention.

Project Title: **ProHits Next Generation: A flexible system for tracking, analyzing and reporting functional proteomics data**
Project Type: **Large-Scale Applied**
Project Leaders: **Drs. Anne-Claude Gingras and Mike Tyers**
Institution: **Samuel Lunenfeld Research Institute at Mount Sinai Hospital**
Institut de Recherche en Immunologie et cancérologie, Université de Montréal

Human cells are built from tens of thousands of different proteins that perform most of the activities necessary for life. To gain insight into the cause of a disease and to develop new approaches to treat disease, it is necessary to understand how proteins interact with and modify each other. Mass spectrometry is now being used to identify proteins and their modifications and interactions.

Drs. Anne-Claude Gingras, Mike Tyers and team aim to develop innovative ways to analyze the data generated by mass spectrometry and to increase the amount of information about protein interactions and modifications

Their research will improve the analysis of protein interactions and increase understanding of the effects of disease states and drug treatments.

Project Title: **Development of a unified Canadian clinical genomic database as a community resource for standardizing and sharing genetic interpretations**
Project Type: **Large-Scale Applied**
Project Leaders: **Drs. Jordan Lerner-Ellis and Matthew Lebo**
Institution: **Mount Sinai Hospital**

Canadian scientists have made exciting discoveries about the complex relationship between genetic mutations and disease. However, much of this information is spread across dozens of databases in widely differing formats. In order to use this information to improve patient outcomes, researchers and clinicians need a more widely-accessible resource designed for sharing and collaboration.

Drs. Jordan Lerner-Ellis, Matthew Lebo and their team aim to address this issue by creating a shared, open-source genetic database that will amalgamate the work of participating clinical and research laboratories across Canada and internationally. This resource will provide sophisticated new tools for the diagnosis and management of common and rare diseases, while improving the effectiveness of healthcare delivery.

Project Title: **Large data sets and novel tools for plant biology for use in international consolidation-tier data repositories and portals**
Project Type: **Large-Scale Applied**
Project Leaders: **Drs. Nicholas Provart and Stephen Wright**
Institution: **University of Toronto**

New technologies allow plant biologists to identify important DNA sequences in an organism's genome. Among other things, these advances have helped gain insight into the expression level of genes in many different parts of plants under different conditions, the interactions between the proteins present in organism and the 3D structures of these proteins. However, researchers still find it difficult to draw meaningful conclusions from the huge amounts of data that confront them.

The research team led by Drs. Nicholas Provart, Stephen Wright aims to develop data visualization tools and applications to accelerate advances in plant biology. Their contribution to making vast amounts of data easier to interpret will increase our understanding of plant biology, which is important for feeding, housing, clothing and providing energy to the world's growing population.

Project Title: **Applying genomic signal processing methods to accelerate crop breeding**
Project Type: **Small-Scale Innovative**
Project Leaders: **Drs. Lewis Lukens and Cortland Griswold**
Institution: **University of Guelph**

Selective breeding improves plant and animal products by identifying desirable traits such as quality, yield, and ability to grow in difficult conditions, ensuring that that there is sufficient production for food, fuel and raw materials. Factors like climate change and population growth are making selective breeding more important than ever. One of the largest challenges facing the plant research community is identifying the suite of genes that make organisms well adapted to their environment and using this information in breeding programs.

Drs. Lewis Lukens, Cortland Griswold and their team are using bioinformatics tools to understand how organisms that adapt well to their environments can be selected to accelerate the development of new plant varieties.

Project Title: **Leveraging meta-transcriptomics for functional interrogation of microbiomes**
Project Type: **Small-Scale Innovative**
Project Leaders: **Dr. John Parkinson**
Institution: **The Hospital for Sick Children**

Bacteria do not live in isolation but tend to form parts of microbial communities (called "microbiomes"), displaying complex inter-dependencies between themselves and their

environments. The composition of these communities is increasingly viewed as having a significant impact on human health and disease.

To understand more about how bacteria function within their communities, whole-microbiome gene expression profiling has emerged as a powerful tool to study their influence on their environment. However, few methods and tools to fully understand the data resulting from this profiling have been developed.

Dr. John Parkinson and team aim to bridge this gap by developing new software that enables the identification of genes and pathways that have critical roles within the microbiome. Such genes and pathways represent potential targets for new treatments that help maintain healthy microbiomes and reduce the risk of diseases such as Type 1 diabetes, irritable bowel disease and rheumatoid arthritis.

Project Title: **Pathway and network visualization for personal genomes**
Project Type: **Small-Scale Innovative**
Project Leader: **Dr. Lincoln Stein**
Institution: **Ontario Institute for Cancer Research**

Cancer is a disease caused by the accumulation of multiple genetic mutations. Highly specific drugs that target mutated proteins in cancer cells are currently being used to treat the disease. However, since cancer patients have different mutation profiles, a drug that is effective in one may not have the same result in another. Personalized medicine based on genomic data would allow doctors to determine the best targeted therapy for each patient.

Dr. Lincoln Stein and team aim to develop software that will improve the treatment of cancer patients by enabling physicians to study and visualize the genomic aberrations of individual patients. It will help identify genes related to cancers and other disease.

QUÉBEC

Project Title: **PIATEA: A portal for integrative approaches to transposable element annotation**
Project Type: **Small-Scale Innovative**
Project Leaders: **Drs. Mathieu Blanchette and Thomas Bureau**
Institution: **McGill University**

Recent advances in genomics have revealed that most of the DNA in complex organisms is comprised of transposable elements (DNA sequences that can change their positions within the genome). Transposable elements have major impacts on how organisms function and evolve. They can cause disease or adapt to new or changing environments.

Researchers need to better understand the role of transposable elements, but current tools that identify them accurately are difficult to use.

Drs. Mathieu Blanchette, Thomas Bureau and team aim to develop a user-friendly system to identify transposable elements more accurately and quickly. Doing so will also help speed up the development of genomic applications such as disease treatments and breeding strategies for agriculture.

Project Title: **A development and deployment platform for citizen science games in genomics**
Project Type: **Small-Scale Innovative**
Project Leaders: **Drs. Jerome Waldispuhl and Mathieu Blanchette**
Institution: **McGill University**

Computer games, which are tremendously popular worldwide, can bring massive resources together through *crowdsourcing (a practice of obtaining contributions from a large group of people, especially from an online community)*. Applied to genomics, scientific games have excellent potential to solve complex problems arising in molecular biology.

Drs. Jerome Waldispuhl, Mathieu Blanchette and team are designing a platform that will allow scientists to build video games for studying, mining and processing molecular biology data.

The games will be assembled on an online gaming website that will be open to contributions from researchers and game companies to enhance the development of techniques in personalized genomics and health.

ATLANTIC

Project Title: **Exploiting the full potential of next generation DNA sequencing for crop improvement**
Project Type: **Small-Scale Innovative**
Project Leader: **Dr. Sean Myles**
Institution: **Dalhousie University**

A future with safe and secure food requires the breeding of new disease-resistant crops that provide higher yields and require fewer chemicals to grow. Current breeding practices have given us most of the food we enjoy today, but many of the current breeding methods are laborious, time-consuming and expensive.

New genomics technologies can help make the traditional breeding process more efficient, cost-effective and accurate.

Dr. Sean Myles and team are making sense of data generated from modern DNA sequencing technologies to develop user-friendly genomics-assisted breeding software that will quicken the development of better-tasting and healthier food. The software will also be able to be used in other areas of genomics research, such as forestry, bioenergy, conservation biology and aquaculture.