Project title:
Genomics in the CARTaGENE cohort

Collaborators: Simon Gravel, Guillaume Lettre, Daniel Taliun, Anne-Marie Laberge, Julie Hussin, Claude Bhérer, Sarah Gagliano-Taliun, Emmanuel Milot, Simon Girard, Martine Tétreault, Sirui Zhou, Yue Li, Vincent Ferretti

Project summary:
This project aims to highlight the usefulness of genetic data in CARTaGENE by performing a wide-range of analyses to be published in a high-impact peer-reviewed research paper. The aim of these efforts is to showcase the CARTaGENE dataset, increase its visibility and promote future collaborations. The manuscript will provide updates about the latest genetic and phenotypic data available through the CARTaGENE biobank to researchers in Canada and worldwide. It will showcase the potential value of the CARTaGENE biobank data for genetic and population health research using multiple application examples from population genetics, clinical variant interpretation, genome-wide and phenome-wide association studies (GWAS/PheWAS), and polygenic risk scores (PRS). In terms of genetic data, the main analyses will use the newly generated whole-genome sequencing (WGS) data and the most recent version of genotype data imputed to the TOPMed reference panel. The project is a collaborative effort of multiple research groups with vast experiences in genetics and population health and which were/are actively using the CARTaGENE biobank in their research programmes.

The proposed main manuscript will include the following:

1) Overview of the biobank's structure, participants, and past/future developments.
2) Population genetics insights: patterns of genetic variations unique to the region and local populations, enrichment of clinically relevant variations, unique identity-by-descent (IBD) patterns, combining genetic and genealogical data.
3) Complex variations: characterization of structural variation and HLA alleles in different sub-populations.
4) Genotype imputation: optimal strategies for genotype imputation using heterogeneous arrays, and genotype imputation panel for local populations.
5) Genome-wide associations with disease and disease-related traits: multi-ancestry GWAS/PheWAS with available phenotypes, comparison of GWAS/PheWAS results against other biobanks, application of GWAS/PheWAS results for PRS, gene-based association tests with rare genetic variation.
6) Analytical web tools: interactive variant browser, interactive browser for GWAS/PheWAS - PheWeb.

We expect that some of these topics will also be studied in more depth in “companion” papers.

Due to its broad set of topics, this project will involve the collaboration of many researchers who already have CARTaGENE projects. All analyses already fit within an approved project. We are writing this proposal as a separate joint project to allow researchers to work together by sharing limited amounts of data about participants.