Evaluating a new method for Hepatitis C Virus resistance screening

Hepatitis C is a disease present globally, with 130-150 million chronic carriers of the virus and an estimated 500,000 deaths related to it annually. Although the treatment has improved drastically in recent years, it is unfortunately associated with resistance development. Resistance in HCV is caused by mutations in specific genes of the virus, resulting in failed drug therapy. These are called resistance-associated variants (RAVs), and the mutations occur in the NS3, NS5A and NS5B genes of HCV – which are also the proteins targeted by drugs. In clinical practice, resistance screening is required in cases of treatment failure and also to detect pre-existing RAVs in treatment naive patients. The RAVs may be present at different frequencies in a patient, since large quantities of viruses are produced daily. RAVs are detected by sequencing the HCV genome and analysing the mutation sites. Today this is performed mainly by using Sanger sequencing – a method with the ability to detect resistant variants present at a 15-20% frequency or more in a sample. We have previously developed a new method for detecting lower frequency RAVs, based on long-read sequencing using Pacific Biosciences' (PacBio's) technology, that we believe will ensure a more reliant way of investigating resistance in Hepatitis C patients [1].

In a new pilot project, we have generated Sanger sequencing as well as PacBio data for over 124 HCV infected patients. The aim of this master's project will be to perform comparative analyses of this data and to evaluate the results. For each individual, the RAVs will be identified both in the Sanger sequencing data as well as in the PacBio data, after which statistical methods will be applied to compare and summarize the results. Most of the analysis tools are already available but some own programming code (for example in R) may be required. The ideal candidate should have an interest for infection biology, be able to work in a Unix/Linux environment and have some basic programming skills. To our knowledge this is the largest study conducted so far using long-read PacBio sequencing of HCV. Therefore, we believe the results of this work will be of interest for the research community and may even result in a scientific publication.

- Available data:
    - Sanger sequencing data (for 124 HCV infected patients)
    - PacBio sequencing data (for the same 124 individuals)

- Methods:
    - Programming (in R or other language)
    - Analysis of HCV variants in sequencing data
    - Statistical analysis

Location: National Genomics Infrastructure, SciLifeLab, Uppsala University (BMC building)

Main supervisor: Adam Ameur (adam.ameur@igp.uu.se)

References:

[1] Bergfors A, et al. Analysis of Hepatitis C NS5A resistance associated polymorphisms using ultra deep single molecule real time (SMRT) sequencing. Antiviral Res. 2016 Feb;126:81-9.