Characterising Swedish-specific genetic variation using next generation sequencing data

Description of the project:
In light of recent developments in sequencing technologies, we can gain unprecedented insights into underlying genetic contribution to human complex traits, such as body mass, height, and cardiovascular diseases. Whole genome sequencing (WGS) is to become a golden standard in large-scale genetic studies, due to declining costs and improved data quality. WGS projects has been performed in many human populations from different geographical regions, including the SweGen project in Sweden [1]. These projects provide a wealth of information about genetic factors that shape our genome and potentially influence our risk for disease.

In parallel to the SweGen project, our group (PI: Åsa Johansson) has conducted short-read Illumina WGS in 1000 individuals from the Northern Sweden Population Health Study (NSPHS), aiming to further understand the effects of rare and population specific genetic variation to complex traits. One of the main objectives of this study is to identify large genetic variations, e.g. copy number variation, insertions and deletions. We have already aligned short-read WGS data to the current reference genome assembly, an assembly that has been constructed from multiple individuals to present a pan-human genome. However, it has been shown previously, that individual populations might include genetic material and rearrangements that are not possible to identify using the current reference genome[2]. Therefore, in this project, we aim to further harness the power of NGS technology to characterise Swedish-specific genetic variation using a local reference genome that has been constructed using a combination of methods for short- and long-read sequencing including SMRT PacBio, BioNano, Chromium and Illumina WGS.

Objectives:
- Realigning the Illumina WGS data form the NSPHS to the local reference genome.
- Identifying population specific genetic variations across the genome.

Requirements:
We are looking for one/two enthusiastic student(s) to do $\geq$ 30 ECTs project and work as a member of a team and independently after some instructions. Students with experience and training in NGS analyses are very welcome to apply.
Please send us your CV, motivation letter, and a short description of your relevant experience. Also, please indicate the period which you will be available.

Åsa Johansson
Åsa.johansson@igp.uu.se, nima.rafati@igp.uu.se

Reference:
1. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. Eur J Hum Genet. Nature Publishing Group; 2017;25: 1253–1260. doi:10.1038/ejhg.2017.130
2. Seo J-S, Rhie A, Kim J, Lee S, Sohn M-H, Kim C-U, et al. De novo assembly and phasing of a Korean human genome. Nature. Nature Publishing Group; 2016;538: 243–247. doi:10.1038/nature20098