Investigation of functional common and rare variant in cardiovascular disease

Currently genome wide association studies are targeted to common variants that are present in higher frequency in a population, which had lead into hypothesis of common variant common disease hypothesis. However most of common variants identified so far are within the non-coding regions that are difficult to link with disease. On the other end of spectrum, rare variants: less common in population, are less studied due to large sample requirement and higher cost of sequencing. However, recent studies have shown these genetic variants have higher effect on common diseases as well.

One of such method to study these variant using HiCap (variant of Hi-C ) methods, that identifies promoter-anchored interaction between variants that are thousand of bases apart from each other. Thus providing connectivity information from GWAS variants with potential genes.

In this current project, we will investigate role of both common and rare variant in one of common disease, cardiovascular diseases. Our group already carried out Hicap on large sample cohort thus having interaction data between putative promoters and enhancers.

The aim of this project would:

* Identify these variants as rare and common variants based on external databases such as (1000G project, SwedGen Variant frequency database).
* Annotate variants based on functional marks from ENCODE database. Basically target command and rare variants with TFBSS (Tranacription factor binding sites sequences) marks.
* From these common variant annotate variants with eQTL effect from GTEx project on relevant cell types
* On the rare variants side, identify variants with functional effect on candidate genes

From this project,

I should be able to aptly use and manipulate all the large-scale databases from all the projects. Additionally, I should be programmatically manipulating all the analysis.