Investigation of SNPs and sequence segments of low compositional complexity in genomic DNA and protein sequences.

Vidhya Gomathi Krishnan and David Robert Westhead

This project is investigating several aspects of SNPs (Single Nucleotide Polymorphisms) in *C. elegans* and the human genome. SNPs are important because they make a large contribution to genetic differences between individuals. It is expected that SNPs data will make a major contribution to pharmacogenomics. This is the new science aiming to tailor drug treatments to individual genotypes to produce more efficacious therapy with fewer side effects.

Statistical analyses have revealed interesting tendencies of SNPs to lie in regions of low sequence complexity and outside exons, reflecting parts of the genome that are evolving more quickly and are more susceptible to change. However a significant number of SNPs occur in protein coding regions, and these are more likely to have functional effects. We are now moving on to create machine learning methods to predict the functional significance of SNPs based protein structure predictions.

In parallel we are investigating links between low sequence complexity, domain boundaries and protein function. An overview of the project is shown in Figure 1

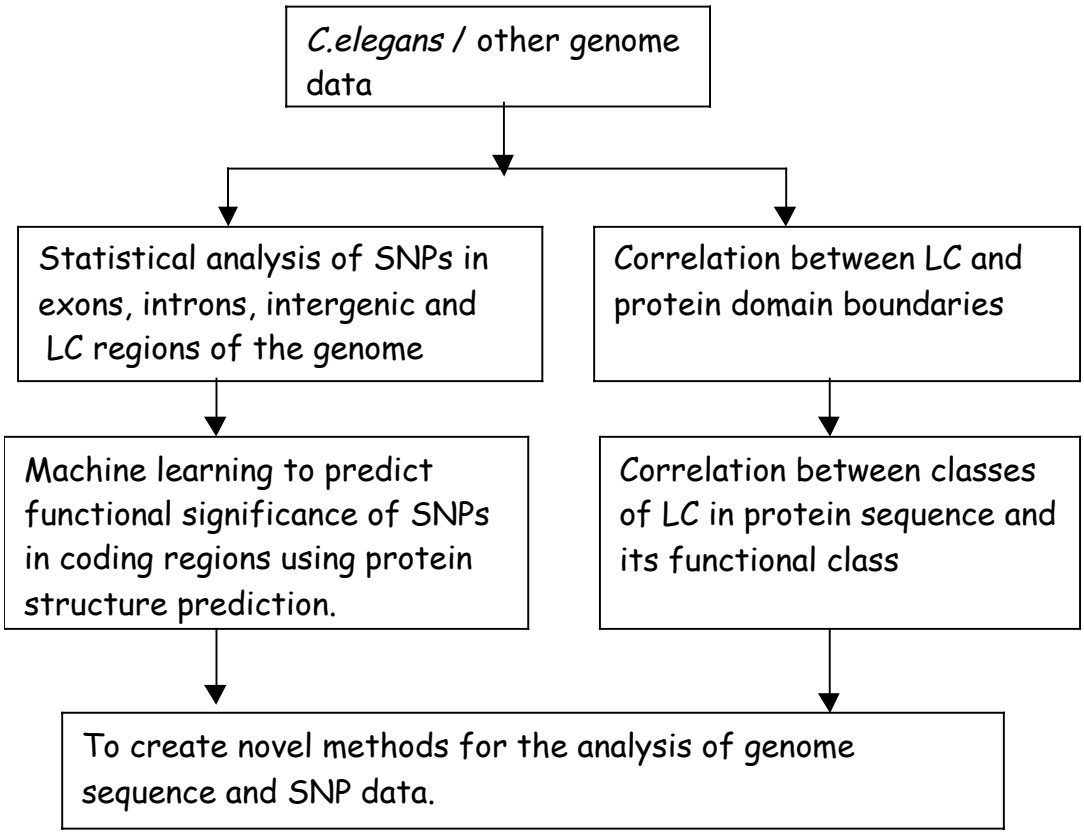

Figure 1: The flowchart of the overview of the project.

Funding
We thank AstraZeneca plc and the ORS award scheme for funding this work.