**Project Summary**

In recent years, the numerous large-scale sequencing projects combined with fast sequencing techniques have generated enormous amounts of sequence data. This has led to the identification of thousands of previously unseen genes whose function awaits to be characterized. A fundamental goal is therefore to identify the function of uncharacterized genes on a genomic scale. It is difficult to design functional assays for uncharacterized genes so a major current challenge in bioinformatics is to devise algorithmic methods that, given a gene, can predict a hypothesis for its function that can then be validated experimentally.

In this project, we shall focus on three different aspects of gene function: molecular function, cellular role and organismal phenotype. Our aim will be to build a general system that, given a gene, can predict its function at these three different levels. This multi-scale prediction will be carried out exploiting the structure of biological networks.

**Intellectual Merit**

Protein Function Prediction through networks

**Broader Impacts**

The proposed work will lead to impacts beyond the protein function community – in particular, in the fields of personal genomics and network science. In recent years, there has been an avalanche of personal genomics sequencing data, providing new needs for the large-scale interpretation of protein function.

The proposal to integrate function prediction with networks will have direct relevance to the network scientific community. It will help transform biological networks from static node-and-edge representations into dynamic entities that better reflect molecular reality. Insights gained from such network analyses may reveal relationships between network centrality (e.g. hubiness) and mobility, or the roles that motions play in precluding or facilitating interactions.

The results of Aims 1 and 2 will be integrated into a series of web tools that will serve as a bridge for investigators coming from networks and protein function communities. Finally, we outline a comprehensive set of plans to use the developed tools as devices for education in bioinformatics, as well as designing, organizing and running various workshop on the topic.