Project Acronym	Mut2Dis
Project Code	PIOF-GA-2009-237225
Project Title	New methods to evaluate the impact of single point protein mutation on human health.
Periodic Report	Returning Phase, Sep 2011 – Aug 2012 (12 months)

RESEARCH SUMMARY

1. Summary of the project objectives

In this report we summarize the research activity performed by Dr. Emidio Capriotti during the returning phase of the Marie-Curie IOF at the Department of Mathematics and Computer Science, University of Balearic Islands under the supervision of Dr. Jairo Rocha.
The main aims of our proposal are the following:

i. Study and characterization of the rate of evolution of Single Nucleotide Polymorphisms and their effect in human disease.
ii. Study and characterization of the structural determinants of human disease.
iii. Development of new general machine learning methods for disease prediction.
iv. Development of disease-specific predictors.
v. Development of a World Wide Web server for predicting the likelihood of a SNP variant to be associated with human disease.

These 5 aims correspond to 6 different tasks that have to be accomplished in 36 months. In the proposal's timeline, during the returning phase (12 months), we planned to perform the two final 2 tasks. According to this, we mainly accomplished the 4th and 5th. In conclusion during the whole period of the project (36 months) we achieved all the five objectives described in our proposal.

2. Description of the work performed since the beginning of the project

During the last 12 months of the project Dr. Emidio Capriotti developed a new methods for the prediction of disease-specific mutations focusing on cancer. In addition, he implemented two different web servers for the predictions of disease-related variants.
In details EC selected a manually curated set cancer driver missense Single Nucleotide Variants (mSNVs). This dataset previously used to train another method (Carter at al, Cancer Research 2009) and analyzed it performing sequence analysis of the protein under mutation. For each protein the sequence profile has been calculated using similar protein retrieved using the BLAST algorithm. Using all sequence information previously calculated EC developed a machine learning approach to discriminate between cancer causing and neutral variants. For this particular task only sequence information has been used because the number of cancer mutations for which protein three-dimensional

structure was available were not enough abundant to train a machine learning method. Finally, EC implemented two web servers: the first one more general for the prediction of disease-related mSNVs and the second one more specific for the detection of cancer causing mSNVs.

3. Description of the achieved results

The research activity performed during the returning phase reached all the aims described in our proposal. In particular, it has been demonstrated that for diseases for which a good number of annotated mutations are available it is possible to build disease-specific predictors. In particular we tested this hypothesis in the case of cancer causing mSNVs showing that the disease-specific methods reaches performs better than the general method. In addition we implemented web available version of the method that can be used by the scientific community to evaluate possible deleterious mutations in human.

4. Expected final results and their potential impact

At the end of the returning phase we have developed a user-friendly web server interface for the prediction of the effect of mSNVs. The implemented web tools include a general method for the detection of disease-related variants that uses both protein sequence and structure information and a cancer specific algorithm that takes in to account only sequence information. In conclusion we demonstrated that structural information is important to improvement the prediction of deleterious variants. When structural information is not available but a good set of mutations have been annotated, the function information are important to improve the performance of the predictors on a specific class of diseases. We believe that in the near future, when more mSNVs data will be available, the development of disease-specific methods will be key strategy for the development of more accurate algorithms and for the understanding of the disease mechanism.

More details about the project are available at: http://snps.uib.es/mut2dis/