

From comparative genomics to 'molecular bioinformatics': an integrated approach to functional genomics

Francesco Filippini^{1*}, Valeria Rossi¹, Raffaella Picco¹, Massimo Floriduz¹, Tiziana Naccari¹, Andrea Carpi¹, Alberta Budillon², Marcella Vacca², Fernando Gianfrancesco², Alfredo Ciccodicola², Michele DiUrso²

¹ Dipartimento di Biolog

Despite several genomes have been largely or completely sequenced, we are still far from a 'whole proteome' functional characterization. Thousands of 'hypothetical proteins', 'unique' (i.e. not homologous to any other) gene products, and even partially characterized proteins hide unknown, possibly important functions. In the absence of experimental confirmation, sequence analysis alone never demonstrates a function, and simple homology search is even unable to infer function from sequence for about a third of the gene products in any of the sequenced genomes. On the other hand, a 'brute-force' experimental characterization of whole proteomes would be too expensive and time-consuming. In order to find an alternative to shotgun characterization of gene products with unknown function, we are following a "molecular bioinformatics" approach, combining step-by-step in silico analyses with in vitro and in vivo prediction-driven experiments. This allowed to unravel molecular mechanisms underlying function of rolB plant oncogene (F. Filippini et al., Nature 379:499-500, 1996; V. Rossi et al., ms. in preparation) and of bifunctional phospholipase-chitinase EP3 (R. Picco et al., ms. in preparation); further, work is in progress about characterization of the unique N-terminal domain of SYBL1 (M. D'Esposito et al., Nature Genetics, 227-230, 1996), the molecular significance of MeCP2 mutations in Rett syndrome patients, to attempt a genotype-phenotype correlation (Vacca et al et al. , ms. in preparation) and about comparative analysis of mammalian and plant adaptins and protein tyrosine kinases. Using comparative analysis of mammalian and plant genomes we are searching for evolutionarily conserved "sequence function tags" - such as motifs and even weak homology regions - in order to get suggestions for a 'targetted' experimental approach to the demonstration of function or for the identification of homologous elements. This approach is finally aimed to unravel possibly the deregulation of important functions underlying a genetic diseases or fundamental genes involved in plant and mammalian developmental control.