

Poster C-1

Elucidating the complex nature of gene fusion events for moving towards automatic quantification of protein function



Authors:

Ashwin Sivakumar (*Bioinformatics group, Division of Genetics, Institute of Biotechnology, University of Helsinki PO B*)

Christopher Wilton (*Bioinformatics group, Division of Genetics, Institute of Biotechnology, University of Helsinki PO Bo*)

Swapn Mallick (*Bioinformatics group, Division of Genetics, Institute of Biotechnology, University of Helsinki PO Bo*)

Liisa Holm (*Bioinformatics group, Division of Genetics, Institute of Biotechnology, University of Helsinki PO Bo*)

Short Abstract: A complete functional unit of a protein can be multi-domain, and it is the co-occurrence and interaction of these multiple domains that determines the function and functional diversity of their gene products. We automatically mine these “functional units” (modules) using only sequence information for introducing a biologically novel automatic functional classification system of contemporary proteins.

Long Abstract:

Diversity in contemporary genomes is elucidated by the functional diversities of proteins encoded by them. During the course of evolution, proteins recruit new functions, which can generally be predicted through specific sequence patterns encoding a functional unit.

A complete functional unit of a protein can be multi-domain, and it is the co-occurrence and interaction of these multiple domains that determines the function and functional diversity of their gene products. Most of the eukaryotic genomes are made of multi-domain proteins. They offer numerous case studies where functional diversities are elucidated by interaction between constituent domains of a protein. With periodic sequencing of new eukaryotic genomes, functional annotation and characterization of multi-domain proteins present a major challenge.

We automatically mine these “functional units” (modules) using only sequence information for introducing an automatic functional classification system of contemporary proteins. The main biological theme behind the concept of modules lies in systems scale identification what we loosely term ‘gene fusion events’.

Our progressive study [1, 4] while developing the concept of modules suggests that at least 10% of the proteins from contemporary genomes exist as fused genes. The study also remarkably revealed that function of protein complexes or proteins resulting from a fusion/fission event may not be just a binary addition of functions of their fusing component proteins. We argue that gene fusion and fission events have a pivotal role in evolution of contemporary functional diversity and can be complex in nature.

Here we discuss how these results and the concept of modules have special repercussions on the following key problems in functional genomics:

a) Computational quantification of biological function and studying diversity in protein

functions at the gene product level.

b) How the modular classification of the protein universe compares and improves upon the concept of domain classifications [2, 3].

c) Complex nature of gene fusions and their role in maintaining continuous diversity in functions of the protein universe. Roles of gene fusions in complex diseases like Cancer.

d) Quantification of horizontal gene transfers- Old questions for new answers.

References:

1. From sequence to a functional unit. Sivakumar A, Wilton C, Holm L. Jan 3, 2006. Physiological genomics, 2006.

2. ADDA: A domain database with global coverage of the protein universe. Heger A, Wilton C, Sivakumar A, Holm L. Nucleic Acids Research. Jan 1, 33 Database Issue: D188-91., 2005

3. The Pfam protein families database. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. Nucleic Acids Res. 2000 Jan 1;28(1):263-6

4. Approaches to define biological function-dissecting the diverse crotonase superfamily Manuscript under preparation, Sivakumar A, Holm L, Mallick S, Wierenga R