Contribution to the ontology and system biology of muscle genes and application to microarray expression studies

Mittempergher L, Picelli S, Feltrin E, Colluto L, Nofrate V, Caldara F, Millino C, Campanaro S, Valle G

CRIBI Biotechnology Centre, Department of Biology, University of Padova, Padova

Motivation

Over the past ten years, our group has been working on the identification and characterization of skeletal muscle genes. A problem that remains open raises from the observation that vertebrate skeletal muscle consists of fibers having different contractile properties: fast and slow. Although several fast- and slow-specific genes have been known for many years, an overall picture of gene expression in these two types of muscle is still unclear. Therefore, we performed some gene expression experiments using microarrays, taking the mouse as a model organism because, differently from human, many of its muscles are mainly composed by a single type of fiber, making easier to delineate a "fast" or "slow" expression profile. A very effective approach to analyse microarray data is based on Gene Ontology (GO) (http://www.geneontology.org/), but there are two main problems: firstly, the list of probes needs to be continuously revised as the genome annotation and GO terms become better defined. Secondly, we found that the GO terms related to muscle were very scanty, making worthless their application to our studies. This gave us a good motivation to contribute to the GO project and to start a collaboration with the GO Consortium to enrich our domain of interest (neuro-muscular genes) with new terms. At the same time, the System Biology Markup Language (SBML, http://sbml.org/) can be used to annotate protein interactions.

Methods

Re-annotation of microarray probes. Oligo sets used for microarray experiments must be reannotated because: 1) new genes and splice variants are continuously identified and frequently an oligo designed specifically for a gene cannot discriminate between splice variants or gene isoforms; 2) gene annotation in public databases is continuously updated. We have implemented a procedure that make use of GoMiner (http://discover.nci.nih.gov/gominer/) and other programs in order to obtain a faster annotation of our microarray platforms. GO and sub-ontology of neuro-muscular genes. The methodological approach is based on a deep understanding of specific problems by expert people. Every problem (for instance a group of proteins involved in a given process) is then "translated" into a series of GO terms that are submitted to the GO Consortium. In this respect we are collaborating with the GO curators at the EBI and we must acknowledge their help and availability. Once the new GO terms are defined, they can be associated to the corresponding genes. Since GO terms must be for general use, some very specific terminology cannot be included. Therefore we have started the development of a specific sub-ontology that could better explain some biological aspects such as muscle contraction or nervous system functions. This sub-ontology will complement the general GO terms. Annotation of protein interaction with SBML. GO terms are intended to describe proteins, not the network of their interaction and their functional regulation. Therefore, we annotate this information by SBML (level 2), using the CellDesigner software (Kitano et al. Nature Biotechnology 23, 961-966, 2005).

Results

Although this work of annotation is still in progress and will probably continue for some time, we have applied the methods described above for a whole-genome microarray expression analysis of mouse skeletal muscles using oligo microarray technology. We considered three muscles composed respectively of slow (soleus), fast (tibialis) and mixed fibers (gastrocnemius). Our preliminary results are shown in the table enclosed to the abstract, where the number of genes having an associated GO term rises from 6571 to 9933.

Contact email: lorenza.mittempergher@unipd.it

	old (OPERON) annotation	new (updated) annotation
Genes annotated with a GO term	**6571**	**9933**
Biological Process	**5430**	**8020**
muscle_contraction	**44**	**52**
muscle_development	**93**	**107**
heart_development	**65**	**74**
Cellular Component	**5575**	**8300**
cytoskeleton	**370**	**522**
actin_cytoskeleton	**91**	**135**
myosin (protein_complex)	**25**	**34**
Molecular Function	**5835**	**8739**
cytoskeletal_protein_binding	**147**	**215**
actin_binding	**115**	**159**
motor_activity	**61**	**88**