Poster G-22

From peptides - identified by mass spec - to biomolecular pathways

Authors:
Claus A. Andersen (*Sienabiotech SpA*)
Stefano Gotta (*Sienabiotech SpA*)
Roberto Raggiaschi (*Sienabiotech SpA*)
Andreas Kremer (*Sienabiotech SpA*)
Eduardo Gonzalez Couto (*Sienabiotech SpA*)
Rainer Kern (*Sienabiotech SpA*)
Georg C. Terstappen (*Sienabiotech SpA*)

Short Abstract: Today several tools support protein identification by mass spec, but bringing the acquired experimental results into an improved understanding of the biomolecular pathways is missing. Grouping the MS identified proteins can support this step by connecting splice variants, paralogs, orthologs, homologs, GO functions etc. as presented here.

Long Abstract:
Experimental setup: Rat cortical neurons treated with amyloid-beta; were compared to untreated controls in order to find differentially modulated phospho-proteins. The phospho-proteins were first extracted using an IMAC resin, then separated on a 2D-gel from which differentially modulated spots were excised. The proteins were trypsinated and finally run through a nanoLC prior to peptide determination by MS/MS using an ESI-QTOF-MS.
Problem setting: It is well known that the number of proteins identified by mass spec is too high; both from the point of view of 2D-gel electrophoresis where it is reasonable to expect ~5 proteins in a spot as well as from the perspective of biological interpretation where resources are invested to understand which biomolecular causes could have produced the differential modulation observed. The high number of proteins is partially due to the genomic origins of proteins where a set of peptides match several splice variants and paralogs. To a large extent it is also due to the large list of redundant orthologs also matching the same list of peptides and the desire to avoid false negatives. To limit the latter one could simply exclude proteins from other organisms than the one studied, but in the present case of rat many proteins are still not described. This is evident from the fact that presently only 17967 genes are known in rat, while 20306 genes are known in mouse (13% increase between ENSEMBL assemblies RGSC 3.4 and NCBI m35, respectively) and that 20-30% of the observed peptides are not found in rat proteins. One could also argue that the MS/MS spectra to peptide matching could be done searching the genome in 6 frame translation, but this is in most cases prohibitively computationally intensive, introduces a large degree of noise, opens up further problems such as full length protein determination and finally does not cover peptides which cross intron-exon boundaries. In summary this means that at present one is forced to search databases containing proteins from other species for protein identification by MS/MS and handle the increased complexity downstream in the analysis workflow.
Analysis methodology and results: The spectra-to-peptide matching is performed using the

Mascot waterfall model on ENSEMBL rat (known, novel and ab-initio) followed by UniRef100. This is done to maximize the yield and minimize the computation time required, which was demonstrated by an analysis of the unique tryptic peptides present in various databases. A threshold is applied to exclude proteins which are unlikely to have been present based on the observed peptide match quality analogous to the ‘Trans Proteomic Pipeline’. Various grouping methodologies are presented to handle the large number of resulting proteins observed. Representative proteins are furthermore selected for each group, where the objective is to reduce the complexity presented to the user while retaining the possibility to dig into the data and retrieve any underlying sub-parts. There are primarily three types of groupings: one which addresses complexity introduced by the genomic origins of proteins, a second which focuses on the supporting experimental data and a third on the proteins’ functional aspects. Examples are also shown of how the groupings support the biological interpretation of the data and its mapping onto pathways.