

Open Access

Recent Trends in Data Mining in Proteomics and Various Applications of Mass Spectrometry in Proteomic Studies

Siva Kishore Nandikolla^{1*}, Satya Varali M² and Mahaboobbi Shaik³

¹Department of Biochemistry and Bioinformatics, GITAM University, Visakhapatnam, India ²Department of Human Genetics, Andhra University, Visakhapatnam, India ³Department of Biotechnology, Andhra University, Visakhapatnam, India

Abstract

Data mining approaches are collected at various levels of proteomic analysis. Various recent studies in proteomics and its collaborated works with mass spectrometry are data mined. Applications like Global Proteomics, Leu-d₃ labelling, Clinical proteomics, In Mascot scores, Serum proteome analysis, In breast cancer, In Genetic studies, Cancer biomarker, in the recurrence of pterygia, Identifying proteins in the cells injured by alcohol, Identification of host-derived biomarkers, proteome expression database, PDAC. Tools like MODELLER 9v2, PIANA, MUTATOR, MALDI-TOF, PDAC. Algorithms like VBA- code, unique-*m* mining algorithm are studied and concatenated in this article.

Keywords: Data mining; Proteomics; Mass spectrometry; Biomarkers; Algorithms

Introduction

Data mining in proteomics

Data mining is the search for hidden trends within large data. Data mining is needed at all levels of genomics and proteomics analyses. These studies can provide a lot of information and generate large quantities of data from the analysis of biological specimens. The amount of data generated from these studies will require the development of improved bioinformatics and computational biology tools for efficient and accurate data analyses [1].

Proteomics is a branch of functional genomics to study protein properties, such as protein expression, post translational modifications and protein-protein interactions to obtain a global view of cellular process. The proteome is a dynamic feature of an organism, it's tissue location and its state changes constantly in response to stimuli both internally and externally. Unlike genes, proteins vary widely in their chemical behaviors, making it difficult to be sort out with one technique that works well on all proteins. Proteomic analysis requires sampling, separation and concentration, identification, structure and proteinprotein interaction [2] network determination and accurate analysis.

Functional analysis and interpretation of large-scale proteomics and gene expression data needs a variety bioinformatics tools and huge public knowledge resources. The intertwined bioinformatics approach gave way to pathway analysis, hypothesis generation and target protein identification [3].

Precise identification of peptides and proteins in biological samples from proteomic mass-spectra is a challenging issue in bioinformatics. The sensitivity of identification algorithms depend on the prevailing scoring methods, some being more sensitive, and others more specific. Combinations of algorithms, called 'consensus methods provide more accurate results than individual algorithms. In-depth analysis of various approaches to consensus scoring using known protein mixtures, and evaluation generated from consensus of three different search algorithms: Mascot [4], Sequest [5], and X!Tandem [6] is considered. The findings indicate that the union of Mascot, Sequest, and X!Tandem performed well (considering overall accuracy), and methods using 80-99.9% protein probability and/or minimum 2 peptides and/or 0-50% minimum peptide probability for protein identification performed better (on average) among all consensus methods tested in terms of overall accuracy [7]. Strategies for optimizing sensitivity or specificity of peptide identification in MS/MS spectra for different user-specific conditions are provided [8].

To predict the three-dimensional structure of a protein. Secondary structure prediction of a protein from its amino acid sequence is an important step. Many of the existing algorithms use the similarity and homology [9,10] to proteins with known secondary structures in the Protein Data Bank, other proteins with low similarity measures need single sequence approach to the discovery of their secondary structure. An algorithm based on the deterministic sequential sampling method and hidden Markov model for the single-sequence protein secondary structure prediction are deduced. The predictions are based on windowed observations and by the obtained average over possible conformations within the observation window [11].

The automated calculation of unique peptide sequences was done in two steps: Step-1: a SQL-based database of theoretically digested peptides from a given FASTA file formatted protein database is generated on choosing protease. Step-2: *in silico* generated peptides from a pre-defined protein sequence are compared to the peptide database in order to identify unique peptides [12].

Several projects were initiated by the Human Proteome Organization (HUPO) aiming for the proteome analysis of distinct human organs. The vertical regarding brain, its development and correlated diseases is the HUPO Brain Proteome Project (HUPO BPP). The relevant bioinformatics data is drawn from the inter-laboratory comparisons as well as from the rechecking of all data sets submitted

*Corresponding author: Siva Kishore Nandikolla, Department of Biochemistry and Bioinformatics, GITAM University, India, E-mail: sivakishoren@gmail.com

Received May 15, 2011; Accepted June 18, 2011; Published July 25, 2011

Citation: Siva Kishore N, Satya Varali M, Mahaboobbi S (2011) Recent Trends in Data Mining in Proteomics and Various Applications of Mass Spectrometry in Proteomic Studies. J Proteomics Bioinform S11:001. doi:10.4172/jpb.S11-001

Copyright: © 2011 Siva Kishore N, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

by the different groups [13]. Due to the large variety of Proteomics workflows, as well as the large variety of instruments and data-analysis software available Human Proteome Organization (HUPO) [14,15] is facing a major challenge and expects data management of greater accuracy, reproducibility and comparability where a new generation of the ProteinScape[™] bioinformatics platform, now enabling researchers to manage Proteomics data from the warehouse to a central data repository with a strong focus on the improved accuracy, reproducibility and comparability of protein data.

Self organizing Maps (SOM) were employed for the classifying and clustering similar sequences and visualization of high dimensional data spaces as they are known for their capability to maintain the essence of topological relationships between the features. SOM effectively yielded 4 clusters which were distinct from each other and marked by characteristic features [16].

Recent implementations of proteomics

Nowadays the knowledge of proteomic studies is very helpful in various research fields, assessments of various biological functions. Some of the recent implementations are:

Global proteomics

Global Proteomics is an alternate methodology, where all blood proteins modulated by disease or drug are used to resolve pharmacodynamic questions in developing an immunoassay. The Global Proteomic approach was applied to one Alzheimer study [17] where it demonstrated a large panel of plasma proteins predictive of disease severity (as measured by the Mini Mental State Examination). This methodology can readily distinguish patients who are responsive and non-responsive to hypertension therapies. Global Proteomics approach is based on bioinformatics analysis which combines samples by proteomic similarity and then uses a geometric representation of sample similarity to answer common pharmacodynamic questions [18].

Leu-d, labelling

The deuterated-leucine (Leu-d₃) labeling is a kind of stable isotope labeling by amino acids in cell culture (SILAC), which is widely used to compare and quantify protein relativeness. Expansion of integrated immune precipitation (IP) coupled with SILAC approach (SILAC-IP) to differentiate the specific binding partners associated with a bait protein in two populations of cells. the accuracy of the SILAC-based quantitative approach not only depends on the integrity of labeled cells, equal mixture of two group of proteins, but also the abundance and signal-to-noise ratio of the peptide pair [19]. By this SILAC-IP strategy, the identified specific-binding proteins were quantified by tracking pairs of Leu-d₃ labeled and unlabeled peptides from the mass spectra, which can differentiate specific-binding proteins from nonspecific partners with great accuracy [20].

In MODELLER 9v2 software

By using the MODELLER 9v2 software homology modeling was done. BLAST [21] search was made against Protein Data Bank (PDB) with the default parameters to find suitable templates for homology modeling. After sequence alignment the one that showed the maximum identity with high score and lower e-value. Final model was obtained by molecular mechanics and molecular dynamics methods and was tested by PROCHECK and VERIFY 3D graph, which showed that the final refined model is reliable. The model could be further studied for the protein characterizing [22].

PIANA

Pathway modeling is one of the most interesting as well as new aspects of systems biology to design and analyze pathway for various diagnostic as well as other purposes [23,24]. Integration and prediction of protein protein interactions with the help of a PPI software framework "PIANA" solves many of the nomenclature issues common to systems dealing with biological data [25,26].

Clinical proteomics

Clinical proteomics are unimaginable without the help of metaanalysis. Due to development of the Proteomics Identification Database Engine (PRIDE) extensive markup language (XML) data format in the year 2005 a standardized data format was created for proteomics data publishing. The Proteomics Identifications Database Engine (PRIDE) has revolutionized proteomics publishing Compared to other fields of science proteomics data can hardly be condensed [27]. GPDE, (http://www.griss.co.at/?GPDE) (Figure 1), an open source web-based bioinformatic tool which uses the possibilities of this standardized format to create biologically-based meta-analysis over number of proteomics experiments. There the GPDE is used in three different ways: 1) in *in silico* alternative to multiple Western analyses; 2) to access easily to cellular proteome reference maps; 3) reliable support to the assessment of the specificity of biomarkers [28].

Visualization

DataBiNS-Viz – a visualization and exploration environment for non-synonymous coding single nucleotide polymorphisms (nsSNPs) data collected by the BioMoby-based DataBiNS workflow. *In silico* [29] analysis of the potential biological impact of nsSNPs requires integration of data and knowledge from various Web-based resources, both databases and analytical tools. Manual retrieval and integration of this information is error-prone and tedious. This provided the motivation for the original DataBiNS - data-mining workflow [30] for the BioMOBY [31] and Taverna [32] environments which retrieved and integrated data relating to nsSNPs and the biological pathways affected by them. DataBiNSViz enables the workflow of DataBiNS on proteins described by KEGG, PubMed, or OMIM identifiers, followed by manual exploration of structure/function and pathway data for those proteins, with focus on nsSNP data [33].

In mascot scores

Proteomics has large amount of proteomics data which are archived for documentation purposes. Since proteomics search engines like Mascot [4,34] or Sequest, are used for peptide sequencing, which resulted in peptide hits that are ranked by giving scores. Ranking algorithms [35] are applied to combine archived search results into predictive models. In this way peptide sequences are identified which made to achieve high scores. All peptide sequences and Mascot scores from four years, proteomics experiments on *Homo sapiens* of the Proteome Center Tuebingen for training are taken. MacroModel and DragonX were used to encode the peptides for molecular descriptor computation. By using the Greedy Search Algorithm ranking-specific feature selection is done to significantly improve the performance of RankNet and FRank. Therefore ranking algorithms can be used for analysing long term proteomics data to identify frequently top scoring peptides [36].



Serum proteome analysis

Analysis of Serum-Containing Conditioned Medium from Primary Astrocyte Cultures [37] is a study to identify secreted proteins in serum containing medium using stable isotope labeling [38]. Serum proteome analysis is a prominent approach in disease diagnosis and therapeutic monitoring. Large dynamic range of proteins, high-abundant proteins, excess of salt and lipid in serum makes the analysis very difficult. This analysis increased sample loading capacity and improved the detection sensitivity of low abundant proteins [39]. Hence, it is important to improve the dissolution of proteins in 2D gel electrophoresis (2-DE), and increase the analysis of serum under a wide variety of physiological conditions [38].

MUTATOR: Tool for proteomics

MUTATER is a computer based tool that is designed to create custom mutations in Nucleotide sequences and Protein. It has the ability to work on both nucleotide sequences and protein, and in both cases, inputs must be given by the user in RAW format. The user has to submit the position of mutation and the amino acid/nucleotide change that is required at that specific position. The output is given in the box. Therefore it serves as a Basic but completely Novel and unique tool when it comes to create Mutations and changes in Sequences. MUTATER (Figure 2) has been developed in C # and is compatible with Windows PC [40].

In breast cancer

The huge progress in proteomics, enabled by the advancements in mass spectrometry, [41] has brought protein analysis back into the limelight of breast cancer research, reviving old areas as well as opening new fields of study like early detection, prognosis, diagnosis, and therapy. Several proteomics technologies like Quantitative proteomic analyses [42], Biomarkers [43,44] are used to uncover molecular mechanisms associated with breast carcinoma at the global level to discover protein patterns that distinguish disease and disease-free states with high sensitivity and specificity. In this review, the basic features of proteomic technologies, including MS, and the main current applications and challenges of proteomics in breast cancer research, including (i) protein expression profiling of breast tumours, tumour cells, tumor fluids and the auto-immune response of the breast cancer cells is considered. All of these applications continue to benefit from further technological advances, such as the development of proteomics methods, high-resolution, high sensitivity MS, SERPA approach, and advanced bioinformatics for data handling and interpretation [45].

In genetic studies

Unique substrings in genomes may indicate high level of specificity

Page	4	of	9
------	---	----	---

			M	UTATE	R		
Protein Mutate Nuc	cleotide Mutate	Gene to Pro	tein				
Browse File	C:\Users\Zan	nmy\Desktop\F	^p rotein.bt				
Original Sequ	uence						
MEDGKPVWAPHF ANILIAVNPYFDIPI EANPLLEAFGNAK NRGCTRYFANKE LKNKSAQSLEYCA VLDIAGFEYFEHN AVHQKHKDHFRL KLSFISVGNKFKTQ ARI DPRI FCKAI E	PTDGFQMGNI KIYSSEAIKSYO (TVRNNNSSR TDKQILQNRK VELLGLDQDDI SFEQFCINYCM TIPRKSKLAVH QLNLLLDKLRS KAI GI NFNDY	VDIGPDSLTIE QGKSLGTRPP FGKFVEIHFNE SPEYLKAGSM IRVSLTTRVM IEKLQQFFNEF IRNIRDDEGFI TGASFIRCIKF YEGI TKVFFR	PLNQKGKTI HVFAIADKA EKSSVVGGF KDPLLDDH LTTAGGTKO RILKEEQELY IRHFAGAVC PNLKMTSHF PGKFAFEDC	FLALINQVFPAEE FRDMKVLKMSQ VSHYLLEKSRIC GDFIRMCTAMKI GTVIKVPLKVEQ/ QKEGLGVNEVH YETTQFVEKNN IFEGAQILSQLQQ DIMKSDPDHLAF	DSKKDVEDNCSL SIIVSGESGAGKT VQGKEERNYHIFY KIGLDDEEKLDLFF NNNARDALAKTVY IYVDNQDCIDLIEA DALHMSLESLICE CSGMVSVLDLMQC I VKRVNHWI TCS	MYLNEATLLHNIKVR ENTKFVLRYLTESYG 'RLCAGASEDIREKLH 'VVAGVLHLGNIDFEE 'SHLFDHVVNRVNQC KLVGILDILDEENRLF SRDKFIRELFESSTNI GGYPSRASFHELYNN RWKKVQWCSL SVIK	AYSKDRIYTYV ITGQDIDDRIV ILSSPDNFRYL EAGSTSGGCN CFPFETSSYFIG PQPSDQHFTS NNKDTKQKAG AYKKYMPDKL I KNKIKYRAF
Mutation Box	IN CALCULATION I	RIGEIRVIIIA	FURFALFDG	AIMING OF DITLAL	LVKNVINNVETCS	NWKKVQWC5L5VIK	
Enter The M	Autation						
Position	10 A	mino Acid	D	ОК	Cancel	1	
Mutated Se	quence					Amino Cour	nt 1285
MEDGKPVWAD TYVANILIAVNPI DDRIVEANPLLE PDNFRYLNRGC AGSTSGGCNLK QCFPFETSSYFIC ENRLPQPSDQH FESSTNNNKDT ASFHELYNMYK	HPTDGFQMG DFDIPKIYSSE, AFGNAKTVRI :TRYFANKETU INKSAQSLEYC GVLDIAGFEYF IFTSAVHQKH KQKAGKLSFI: KYMPDKLARL	NIVDIGPDSLT AIKSYQGKSLG NNNSSRFGKF DKQILQNRKSF AELLGLDQDI EHNSFEQFCI KDHFRLTIPRF SVGNKFKTQLI DPRLFCKALF	IEPLNQKGK ITRPPHVFA VEIHFNEKS 'PYLKAGSM DLRVSLTTR NYCNEKLQC (SKLAVHRN NLLLDKLRS KALGLNENI	TFLALDNQVFP, IADKAFRDMKVI ISVVGGFVSHYLL KUPLLDDHGDF VMLTTAGGTKG QFFNERILKEQI IRDDEGFIIRHF, TGASFIRCIKPNI DYKFGLTKVFFR	AEEDSKKDVEDNO KMSQSIIVSGESG LEKSRICVQGKEEI IRMCTAMKKIGLD TVIKVPLKVEQAN ELYQKEGLGVNEV GAVCYETTQFVE LKMTSHHFEGAQI PGKFAEFDQIMKS	SLMYLNEATLLHNIK AGKTENTKFVLRYLT RNYHIFYRLCAGASE DEEKLDLFRVVAGVL NARDALAKTVYSHLF HYVDNQDCIDLIEAK KNNDALHMSLESLIC LSQLQCSGMVSVLDI DPDHLAELVKRVNH	VRYSKDRIY TESYGTGQDI DIREKLHLSS HLGNIDFEE DHVVNRVN LVGILDILDE ESRDKFIREL MQGGYPSR WLTCSRWK
Change Box Normal Amino	x Acid At Po	osition= 10 i	s Proline	AND After	Mutation New	Amino Acid At F	Position=
10 is Aspartic	Acid						

Figure 2: MUTATER GUI. Amino acid sequence is showing along with mutated sequence with other details including sequence length, Position and mutated amino acid [40].

which is crucial and fundamental to many genetics studies, such as PCR, microarray hybridization, Southern and Northern blotting, RNA interference (RNAi), and genome (re)sequencing [46]. In this study the concept of unique-m substrings of genomes for controlling specificity in genome-wide assays is proposed. A unique-*m* substring is defined if it only has a single perfect match on one strand of the entire genome while all other approximate matches must have more than m mismatches. A pattern growth approaches to systematically mine such unique-m substrings from a given genome. The algorithm does not need a pre-processing step to extract sequential information which is required by most of other rival methods. The search for unique-msubstrings from genomes is performed as a single task of regular data mining so that the similarities among queries are utilized to achieve tremendous speedup. In addition, the unique-*m* mining algorithm has been parallelized to facilitate genome-wide computation on a cluster or a single machine of multiple CPUs with shared memory [47].

The algorithm is as follows. Let *a* be a substring and S_a be the socalled projected database that contains all sequences that contain the substring *a*, where the last element of each occurrence of *a* is marked (the so-called *a*-locations) in both strands of the sequences and the number of mismatches for each occurrence is also recorded. The computation of S_a ' from S_a requires, for each *a*-location, the check of whether or not the base on its right side (if any) equals the newly appended item b. In case of equality this gives an *a*'-location in S_a ' with the same number of mismatches; otherwise one more mismatch will be logged. If the number of mismatches exceeds *m*, this *a*'-location is not stored.

unique-m (a, S_a) if $|S_a| \ge 1$ then if *min_l* \le *length* (*a*) \le *max_l* and $|S_a| = 1$ then report *a* and S_a if *length* (*a*) < *max_l* for each base b do

a' := a with b appended to it

unique-m (a', S_a')

The main call is unique-m (L, SL), where L is the empty substring. Note that each base in database SL is a L-location, including the position before each sequence (with f mismatch). This call creates a projected database that marks all occurrences of the first base b, reports all unique-m substrings that begin with this b, and then proceeds to the next base [47].

Proteomics in tool designing

A 3 dimensional model (3D) was developed for the Paralytic insecticidal toxin of the (ITX-1) of Tegenaria agrestis (Hobo spider). A homology modeling method was used for the prediction of the structure. For the modeling, a template protein was obtained by mGenTHERADER [48], namely the high-resolution X-ray crystallography structure of a FERREDOXIN (1FCA) of Clostridium acidurici. By comparing the template protein a rough model was constructed for the target protein using MODELLER [49], a program for comparative modelling. The model was validated using protein structure checking tools such as PROCHECK [50,51] and WHAT IF for reliability. The final model obtained by molecular mechanics and dynamics method and was assessed by PROCHECK and VERIFY 3D graph, which showed that the final refined model is reliable [22]. The information thus discussed provides insight to the molecular understanding of Paralytic insecticidal toxin (Tegenaria agrestis). The predicted 3-D model may be further used in characterizing the protein in wet laboratory [52].

Predicting the antigenic regions of a protein is of prime importance in assessing the states of a polypeptide chain as exposed or buried regions. This can be achieved by calculating and plotting the hydrophilicity of a protein using values of Hoop-Woods scale. Hence, hydrophilicity plot of Hoop-Woods scale amino acid sequence of a protein on its x-axis, and degree of hydrophobicity or hydrophilicity on its y-axis using python language as architecture by utilizing various functional attributes such as scipy, matplot and numpy modules is given [53].

Cancer biomarker

In cancer biomarker research the development of statistical methods to identify expression signatures showing the heterogeneity of cancer across affected individuals is an active area. [54-56]. This is collaborated by analysis of proteomics data from a melanoma study, in which the differential expression is most often present throughout the distribution, rather than being accumulated in the tails, albeit with a few proteins showing expression patterns consistent with outer expression [57].

VBA- code

For refining large extensible markup language, (XML) – based proteomics dataset with a high stringent and simple approach using VBA- coded plug-in is developed. It is termed as (All and None) methodology. By testing the reliability and efficiency of this method, All and None was confirmed to be an applicable process for initial screening of biological biomarkers in complex specimens and tissue extract [58]. Confirmation of high confidence, high quality of a given protein when its corresponding peptides mostly over 2 and its scores were above identity and homology threshold of MOWSE algorithm [58,59] used in Mascot search [4].

VB code for All and None [58]

Sub find_Matches ()

Dim CompareRange As Variant, x As Variant, y As Variant

Set CompareRange equal to the range to which you will

Compare the selection.

Set CompareRange – Range ("C1:C1000000")

Note: If the compare range is located on another workbook

Or worksheet, use the following syntax.

Set CompareRange = Workbook ("Book2").

Worksheets ("Sheet2"). Range ("C1:C1000000")

Loop through each cell in the selection and compare it to

Each cell in CompareRange.

For Each x In Selection

For Each y in CompareRange

If x = y Then X.Offset (0, 1) = x

Next y

Next x

End Sub

Visual basic codes (VBA) [58]

Following are VB codes used for "All and None comparison.

General instructions for using visual basic codes as macro embedded files in Microsoft spread sheet.

1. Start excel (spread sheet)

2. Enable macro in your excel sheet

3. Press ALT+F11 to start activating visual basic editor

4. On the Insert menu, click Module

5. Enter the following code for all and none in a module sheet. You can also copy and paste the codes

6. Save the module and press AlT+F11 for returning to spreadsheet

7. Enter you data to be analyzed or drawn in excel sheet; for all and non comparison; compare each duplicate of your group. Insert one list (IPI, gene symbol, other formats) in A and C columns

8. Select the range from A1:Ax where Ax is the end of you data

9. In excel 2003 and earlier, point to **Macro** on the **tools** menu, and then click **Macro**

10. In Excel 2007, click the **developer** tab, and then click **Macro** in the code group

11. Select the range from A1:Ax where Ax is the end of you data

12. Click Find_Matches and then click run

13. Duplicates will appear in B column and Unique will be protein Ids besides to blank boxes in column B

14. Press Alt+F8 (or go to tools then macros)

A proteomic approach to identify proteins which may be involved in the recurrence of pterygia. Tissues from a recurrent pterygium and from a primary pterygium were surgically resected and were analyzed by proteomics to identify proteins that were significantly up- or downregulated. Eleven proteins were differentially expressed; seven proteins were up-regulated and four proteins were down-regulated in the recurrent pterygium. The identified proteins are known to regulate cell cycle, cell organization, extracellular matrix, cholesterol metabolism, and cell signaling. Up- and down-regulation of proteins for cellular signaling are most likely involved in the recurrence of pterygia [60].

A proteomic-based approach was applied to characterize cellular responses of neuronal cells to Pyridostigmine Bromide exposure. Protein extracts from cultured neuroblastoma cells treated with 700nM PB for 10 days, as well as extracts from control cells were separated using two-dimensional gel electrophoresis. Twenty two differentiallyexpressed proteins were identified by MALDI-TOF mass spectrometry (MS) [61]. Similarly Maldi –TOF MS was applied to identify the affected proteins when exposed to 1800 MHz GSM mobile phone [62].

Mass spectrometry in proteomics

In the past decade, various mass spectrometry-based approaches have been applied to investigate the proteomes of diseased and normal samples from pancreatic tissues, juice, cell lines, and serum, with the goals of dissecting the abnormal signaling pathways underlying oncogenesis and identifying new biomarkers [63-73]. Several techniques are available in proteomics but LC -MS based analysis of complex protein mixtures turned to a main stream analytical technique in quantitative proteomics [74].

A mass spectrometry-based proteomics strategy to examine protein-protein interactions using anti-Green Fluoroscent Protein single-chain antibody V(H)H in a combination with a novel stable isotopic labeling reagent, isotope tag on amino groups (iTAG) [75]. Classification of the identified proteins into their functional categories indicated that Side Population cells over express stress proteins, cytoskeletal proteins and enzymes of the glycolytic metabolism [76].

The application of novel methods for identifying S-nitrosylated proteins, especially when combined with mass-spectrometry based proteomics to provide site-specific identification of the modified cysteine residues, promises to deliver critical clues for the regulatory role of this dynamic posttranslational modification in cellular processes [77].

Mass spectrometry coupled with protein separation using 2D-PAGE or multidimensional liquid chromatography is the currently technologies for proteomics. This technology can generate huge amount of raw mass spectra and/or tandem mass spectra. These MS data would be analyzed by bioinformatics tools for the rapid retrieval of known proteins from protein databases, and the identification of novel proteins whose functions are hitherto unknown. Large availability of high resolution and accuracy MS instruments, shotgun quantitative proteomics has obtained great reputation in recent years due to its capacity of comparing a large number of samples without resource intensive and potentially biased labeling steps [78]. Many computational methods have been developed in recent years to help these processes [79-84]. The results of 2D-DIGE and protein identification, data are freely accessible in (GeMDBJProteomics, http://gemdbj.nibio.go.jp/dgdb/DigeTop.do) [85].

Different ion-activation methods for MS/MS, such as collisioninduced dissociation (including postsource decay) and surface-induced dissociation, electron capture and electron-transfer dissociation, infrared multi photon and blackbody infrared radiative dissociation have been discussed as they are used in proteomic research [86].

Phosphopeptide/protein identification using tandem mass spectrometry

Phosphopeptide/protein identification using tandem mass spectrometry (MS/MS) is a challenging issue in proteomics research. The existing available algorithms for protein and peptide identification generates high false discovery rate when comparing with phosphopeptide spectra. As a result, huge increase in signalto-noise ratio is provided and the rate of detecting important peaks are significantly high. Experiments which used MASCOT [4] and SEQUEST [5] with Peptide/ProteinProphet [87,88] and a decoy database approach gained a good amount of improvement in the sensitivity of phosphopeptide identification without compromising specificity, explaining the new strategy for MS spectra preprocessing is a powerful proteomics tool for improvising phosphopeptide identifications [89].

Imaging mass spectrometry

Imaging mass spectrometry (IMS) is an emerging technology, pioneered by Prof. Richard Caprioli's group started more than a decade ago. The initial set up for IMS experiments with available automated matrix deposition, MALDI-TOF mass spectrometry instrumentation and data handling software for image generation [90].

MALDI-TOF mass-spectrometry

Mass spectrometry has enormous potential in biomedical research. MALDI-TOF [91] is the new statistical method for analyzing massspectrometry data in proteomic research is proposed. The peak detection method directly affects the process, like possible biomarker identification of a protein [92]. Location-shifted Poisson distribution is imbibed to the deamidated isotopic distribution of a peptide molecule by a specific location. To estimate the parameters of the distribution maximum likelihood estimation by the expectation-maximization (EM) technique is used [93].

Identifying proteins in the cells injured by alcohol

When alcohol induces systemic injures to the cell, it releases organ/ tissue-speific proteins into the blood. These proteins are detected using proteomic approach. Serum proteomic profiles using MALDI –OTOF Mass Spectrometry were compared between before and after treated samples. Potential markers like fragment of alpha fibrinogen, isoform 1 [94,95] were detected in mass spectral profiles which are useful clinically for the determination of alcohol drinking status by MALDI –OTOF mass spectrometry.

Identification of host-derived biomarkers

Host derived biomarkers in the circulating low molecular mass fraction (<25kDa) of blood proteome tested in murine models by using novel mass spectrometry reveals the flexibility of their occurrence in the sera of doxycycline-treated mice. Treated murine responded to the therapeutic intervention and thus making them a useful tool for monitoring efficacies of existing and novel treatment regimens [96].

It is widely accepted that discovery of specific, reliable and sensitive tumor biomarkers can improve the treatment of cancer. Compared

with non-TNBC samples, stem cell markers were over expressed in triple negative breast cancer (TNBC) [97].

In open-access proteome expression database

Ewing sarcoma is the second most common primary malignant bone tumor in children and adolescents worldwide.

An open-access proteome expression database of eight Ewing sarcoma cases using proteome data obtained by two dimensional difference gel electrophoresis (2D-DIGE) and mass spectrometry.. The results of 2D-DIGE and protein identification by mass spectrometry, and part of the corresponding clinico-pathological data such as prognosis after treatments are freely accessible in the public proteome database Genome Medicine Database of Japan Proteomics (GeMDBJ Proteomics, https://gemdbj.nibio.go.jp/dgdb/DigeTop.do) [98].

PDAC

Phosphoproteomes of pancreatic ductal adenocarcinoma (PDAC) cells and normal pancreatic duct cells are characterized by mass spectrometry using LTQ-Orbitrap. More than 700 phosphoproteins from each sample, and revealed differential phosphorylation of many proteins involved in cell adhesion, cell junction, and cytoskeleton are identified. Since post-translational phosphorylation [99] is a common and important mechanism of acute and reversible regulation of protein function in mammalian cells, an understanding of differential phosphorylation of these proteins and resulting signal transduction changes in PDAC will help in comprehending the complex dynamics of tumor invasion and metastasis in pancreatic cancer [100]. The "false discovery rate (FDR)" was estimated by searching a combined forward-reversed database as described by Elias [101].

Current projects in proteomics and MS:

- The HUPO Human Plasma Proteome Project
- Proteomic and metabonomic expression analysis of blood proteins and metabolites following cerebral ischaemia
- · Identification of novel neuroprotective pathways
- Proteome analysis of hematopoietic stem cells from CML patients after treatment with BMS-214662
- A proteomic comparison of dexamethasone-induced nuclear protein in GC-sensitive and GC-resistant acute lymphoblastic leukaemia cell lines
- Proteomic Analysis of the Dicer ribonuclease function in CD4 T cells
- Organelle Proteomics
- Circadian Proteomics
- Statistical considerations within quantitative proteomics
- Asperger Syndrome Mass Spectrometry

Conclusion

These various implementations and recent advancements that are mentioned upon give us a deeper view and knowledge in data mining in proteomics and mass spectrometry studies. Development of new tools can reduce the amount of wet-lab research time for the researchers. The refining of the algorithms can produce more accurate results and uplift the level of research. Advancements in Mass spectrometry, MALDI

References

- 1. Halima Bensmail, Abdelali Haoudi (2005) Data Mining in Genomics and Proteomics. J Biomed Biotechnol 2: 63-64.
- Gupta AK, Goel A, Seneviratne JM, Joshi GK, Kumar A (2011) Molecular Cloning of MAP Kinase Genes and In silico Identification of their Downstream Transcription Factors Involved in Pathogenesis of Karnal bunt (Tilletia indica) of Wheat. J Proteomics Bioinform 4: 160-169.
- Hu ZZ, Huang H, Cheema A, Jung M, Dritschilo A, et al. (2008) Integrated Bioinformatics for Radiation-Induced Pathway Analysis from Proteomics and Microarray Data. J Proteomics Bioinform 1: 47-60.
- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20: 3551-3567.
- Eng JK, McCormack AL, Yates JR III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5: 976-989.
- Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20: 1466-1467.
- Sultana T, Jordan R, Lyons-Weiler J (2009) Optimization of the Use of Consensus Methods for the Detection and Putative Identification of Peptides via Mass-spectrometry Using Protein Standard Mixtures. J Proteomics Bioinform 2: 262-273.
- Dagda RK, Sultana T, Lyons-Weiler J (2010) Evaluation of the Consensus of Four Peptide Identification Algorithms for Tandem Mass Spectrometry Based Proteomics. J Proteomics Bioinform 3: 39-47.
- Zhang Y, Li T, Yang C, Li D, Cui Y, et al. (2011) Prelocabc: A Novel Predictor of Protein Sub-cellular Localization Using a Bayesian Classifier. J Proteomics Bioinform 4: 44-52.
- Vaseeharan B, Valli SJ (2011) In silico Homology Modeling of Prophenoloxidase activating factor Serine Proteinase Gene from the Haemocytes of Fenneropenaeus indicus. J Proteomics Bioinform 4: 53-57.
- Liang K, Wang X (2011) Protein Secondary Structure Prediction using Deterministic Sequential Sampling. J Data Mining in Genom Proteomics 2:107.
- Michael K, Gorden R, Martin E, Anke S, Helmut EM, et al. (2008) Automated Calculation of Unique Peptide Sequences for Unambiguous Identification of Highly Homologous Proteins by Mass Spectrometry. J Proteomics Bioinform 1: 006-010.
- Hamacher M, Gröttrup B, Eisenacher M, Marcus K, Park YM, et al. (2011) Interlab proteomics: data mining in collaborative projects on the basis of the HUPO brain proteome project's pilot studies. Methods Mol Biol 696: 235-246.
- Thiele H, Jörg G, Peter H, Gerhard K, Martin B (2008) Managing Proteomics Data: From Generation and Data Warehousing to Central Data Repository. J Proteomics Bioinform 1: 485-507.
- Nahalka J (2011) Quantification of Peptide Bond Types in Human Proteome Indicates How DNA Codons were Assembled at Prebiotic Conditions. J Proteomics Bioinform 4: 153-159.
- Murty USN, Amit KB, Neelima A (2009) An In Silico Approach to Cluster CAM Kinase Protein Sequences. J Proteomics Bioinform 2: 97-107.
- 17. Takasaki S (2009) Mitochondrial Haplogroups Associated with Japanese Centenarians, Alzheimer's Patients, Parkinson's Patients, Type 2 Diabetes Patients, Healthy Non-Obese Young Males, and Obese Young Males . J Genet Genomics 36: 425-434.
- Paul K, Nathan LC, Daniel C, Clarissa D, Patrice H, et al. (2008) Global Proteomics: Pharmacodynamic DecisionMaking via Geometric Interpretations of Proteomic Analyses. J Proteomics Bioinform 1: 315-328.
- Ong SE, Mann M (2005) Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol 1: 252-262.

Page 8 of 9

- Shufang L, Xuejiao X, Haojie L, Pengyuan Y (2008) Development of Deuterated-Leucine Labeling with Immunoprecipitationto Analyze Cellular Protein Complex. J Proteomics Bioinform 1: 293-301.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.
- Sunil K, Priya RD, Prakash CS (2008) Prediction of 3-Dimensional Structure of Cathepsin L Protein of *RattusNorvegicus*. J Proteomics Bioinform 1: 307-314.
- Somnath T, Virendra VG, Rajat KD (2008). Pathway Modeling: New face of Graphical Probabilistic Analysis. J Proteomics Bioinform 1: 281-286.
- 24. Nanjappa V, Raju R, Muthusamy B, Sharma J, Thomas JK, et al. (2011) A Comprehensive Curated Reaction Map of Leptin Signaling Pathway. J Proteomics Bioinform 4: 184-189.
- Ramón A, Javier GG, Baldo O (2008) Integration and Prediction of PPI Using Multiple Resources from Public Databases. J Proteomics Bioinform 1: 166-187.
- 26. Mohammed A, Guda C (2011) Computational Approaches for Automated Classification of Enzyme Sequences. J Proteomics Bioinform 4: 147 152.
- 27. Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, et al. (2005) PRIDE: the proteomics identifications database. Proteomics 5: 3537-3545.
- Griss J, Gerner C (2009) GPDE : A Biological View on PRIDE. J Proteomics Bioinform 2: 167-174.
- Asthana M, Singh VK, Kumar R, Chandra R (2011) Isolation, Cloning and *In silico* Study of Hexon Gene of Fowl Adenovirus 4 (FAV4) Isolates Associated with Hydro Pericardium Syndrome in Domestic Fowl. J Proteomics Bioinform 4: 190-195.
- Song YC, Kawas E, Good BM, Wilkinson MD, Tebbutt, SJ (2007) DataBiNS: a BioMoby-based data-mining workflow for biological pathways and nonsynonymous SNPs. Bioinformatics 23: 780-782.
- Wilkinson MD, Links M (2002) BioMOBY: an open source biological web services proposal. Brief Bioinform 3: 331-341.
- Oinn T, Addis M, Ferris J, Marvin D, Senger M, et al. (2004) Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics 20: 3045-3054.
- Fong CC, Edward AK, Mark DW, Scott JT (2008) DataBiNS-Viz: A Web-Based Tool for Visualization of Non-Synonymous SNP Data. J Proteomics Bioinform 1: 233-236.
- 34. Yates JR 3rd, Eng J, McCormack A, Schieltz D (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem 67: 1426-1436.
- Burges C, Shaked T, Renshaw E, Lazier A, Deeds M, et al. (2005) Learning to rank using gradient descent ICML '05: Proceedings of the 22nd international conference on Machine learning. 89-96.
- Henneges C, Hinselmann G, Jung S, Madlung J, Schütz W, et al. (2009) Ranking Methods for the Prediction of Frequent Top Scoring Peptides from Proteomics Data. J Proteomics Bioinform 2: 226-235.
- Annika T, Jonas F, Fredrik B, Michael N, Kaj B, et al., (2008) Proteome Analysis of Serum-Containing Conditioned Medium from Primary Astrocyte Cultures. J Proteomics Bioinform 1: 128-142.
- Fengming G, Shufang L, Chunmei H, Guobo S, Yuhuan X, et al. (2008) The Optimized Conditions of Two Dimensional Polyacrylamide Gel Electrophoresis for Serum Proteomics. J Proteomics Bioinform 1: 250- 257.
- Bjöhall K, Miliotis T, Davidsson P (2005) Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. Proteomics 5: 307-317.
- Butt AM, Ahmed A (2009) MUTATER: Tool for the Introduction of Custom Position Based Mutations in Protein and Nucleotide Sequences. J Proteomics Bioinform 2: 344-348.
- Hondermarck H, Vercoutter-Edouart AS, Révillion F, Lemoine J, El-Yazidi-Belkoura I, et al. (2001) Proteomics of breast cancer for marker discovery and signal pathway profiling. Proteomics 1: 1216-1232.
- 42. Anderson NG, Anderson NL (1996) Twenty years of two-dimensional Electrophoresis: past, present and future. Electrophoresis 17: 443-53.

- Vercoutter-Edouart AS, Lemoine J, Bourhis X, Louis H, Boilly B, et al. (2001) Proteomic analysis reveals that 14-3-3 sigma is down-regulated in human breast cancer cells. Cancer Res 61: 76-80.
- 44. Kumar GSS, Venugopal AK, Selvan LDN, MarimuthuA, Keerthikumar S, et al. (2011) Gene Expression Profiling of Tuberculous Meningitis. J Proteomics Bioinform 4: 098-105.
- 45. Hamrita B, Nasr HB, Chahed K, Kabbage M, Chouchane L (2010) Proteomic Analysis of Human Breast Cancer: New Technologies and Clinical Applications for Biomarker Profiling. J Proteomics Bioinform 3: 091-098.
- 46. Cooley P, Clark RF, Page G (2011) The Influence of Errors Inherent in Genome Wide Association Studies (GWAS) in Relation To Single Gene Models. J Proteomics Bioinform 4: 138-144.
- Ye K, Jia Z, Wang Y, Flicek P, Apweiler R (2010) Mining Unique-*m* Substrings from Genomes. J Proteomics Bioinform 3: 099-100.
- McGuffi n LJ, Jones DT (2003) Improvement of the GenTHREADER method for genomic fold recognition. Bioinformatics 19: 874-881.
- TranQui D, Jesior JC (1995) Structure of the ferredoxin from *Clostridium acidurici*: Model at 1.8 Å resolutions. Acta Cryst D 51:155-159.
- 50. Vriend G (1990) WHAT IF: A molecular modeling and drug design program. J Mol Graph 8: 52-56.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26: 283-291.
- Ingale AG, Chikhale NJ (2010) Prediction of 3D Structure of Paralytic Insecticidal Toxin (ITX-1) of *Tegenaria agrestis* (Hobo Spider). J Data Mining in Genom Proteomics 1:102.
- 53. Pandarinath P, Rao AA, Devi GL (2011) A Python Based Hydrophilicity Plot to Assess the Exposed and Buried Regions of a Protein. J Proteomics Bioinform 4: 145-146.
- 54. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, et al. (2005) Recurrent Fusion of TMPRSS2 and ETS Transcription Factor Genes in Prostate Cancer. Science 310: 644-648.
- 55. Pierre M, De Hertogh B, De Meulder B, Bareke E, Depiereux S, et al. (2011) Enhanced Meta-analysis Highlights Genes Involved in Metastasis from Several Microarray Datasets. J Proteomics Bioinform 4: 036-043.
- Marimuthu A, Jacob HKC, Jakharia A, Subbannayya Y, Keerthikumar S, et al. (2011) Gene Expression Profiling of Gastric Cancer. J Proteomics Bioinform 4: 074-082.
- Vuong H, Shedden K, Liu Y, Lubman DM (2011) Outlier-Based Differential Expression Analysis in Proteomics Studies. J Proteomics Bioinform 4: 116-122.
- Magdeldin S, Yoshida Y, Zhang Y, Xu B, Yaoita E, et al. (2011) "All and None" Refining Strategy; Fishing Your Correct Protein from Proteomics Ocean. J Proteomics Bioinform 4: 123-124.
- 59. TM4 Software Suite's MultiExperiment Viewer.
- Pappin DJ, Hojrup P, Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol 3: 327-332.
- 61. Kanamoto T, Souchelnytskyi N, Toda R, Rimayanti U, Kiuch Y (2011) Proteomic Analyses of Proteins Differentially Expressed in Recurrent and Primary Pterygia. J Proteomics Bioinform 4: 058-061.
- 62. Abdullah L, Reed J, Kayihan G, Mathura V, Mouzon B, et al. (2009) Proteomic Analysis of Human Neuronal Cells Treated with the Gulf War Agent Pyridostigmine Bromide. J Proteomics Bioinform 2: 439-444.
- Nylund R, Tammio H, Kuster N, Leszczynski D (2009) Proteomic Analysis of the Response of Human Endothelial Cell Line EA.hy926 to 1800 GSM Mobile Phone Radiation. J Proteomics Bioinform 2: 455-462.
- 64. Shekouh AR, Thompson CC, Prime W, Campbell F, Hamlett J et al. (2003) Application of laser capture microdissection combined with two-dimensional electrophoresis for the discovery of differentially regulated proteins in pancreatic ductal adenocarcinoma. Proteomics 3: 1988-2001.
- 65. Chen R, Yi EC, Donohoe S, Pan S, Eng J et al. (2005) Pancreatic cancer proteome: the proteins that underlie invasion, metastasis, and immunologic escape. Gastroenterology 129: 1187-1197.

- Gronborg M, Bunkenborg J, Kristiansen TZ, Jensen ON, Yen CJ et al. (2004) Comprehensive proteomic analysis of human pancreatic juice. J Proteome Res 3: 1042-1055.
- 67. Chen R, Pan S, Yi EC, Donohoe S, Bronner MP et al. (2006) Quantitative proteomic profiling of pancreatic cancer juice. Proteomics 6: 3871-3879.
- Gronborg M, Kristiansen TZ, Iwahori A, Chang R, Reddy R et al. (2006) Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. Mol Cell Proteomics 5: 157-171.
- Bloomston M, Zhou JX, Rosemurgy AS, Frankel W, Muro-Cacho CA et al. (2006) Fibrinogen gamma overexpression in pancreatic cancer identified by large-scale proteomic analysis of serum samples. Cancer Res 66: 2592-2599.
- Chen R, Pan S, Brentnall TA, Aebersold R (2005) Proteomic profiling of pancreatic cancer for biomarker discovery. Mol Cell Proteomics 4: 523-533.
- Aspinall-O'Dea M, Costello E (2007) The pancreatic cancer proteome recent advances and future promise. Proteomics Clin Appl 1: 1066-1079.
- Chen R, Pan S, Aebersold R, Brentnall TA (2007) Proteomics studies of pancreatic cancer. Proteomics Clin Appl 1: 1582-1591.
- Tonack S, Aspinall-O'Dea M, Neoptolemos JP, Costello E (2009) Pancreatic cancer: proteomic approaches to a challenging disease. Pancreatology 9: 567-576.
- 74. Zhou W, Capello M, Fredolini C, Piemonti L, Liotta LA et al. (2011) Proteomic analysis of pancreatic ductal adenocarcinoma cells reveals metabolic alterations. J Proteome Res 10: 1944-1952.
- 75. Tuli L, Ressom HW (2009) LC–MS Based Detection of Differential Protein Expression. J Proteomics Bioinform 2: 416-438.
- Galan JA, Paris LL, Zhang HJ, Adler J, Geahlen RL, et al. (2011) Proteomic studies of Syk-interacting proteins using a novel amine-specific isotope tag and GFP nanotrap. J Am Soc Mass Spectrom. 22:319-328.
- Satyavani R, Fatima A, Sundaram CS, Anabalagan C, Saritha CV, et al. (2009) Proteomic Analysis Of The "Side Population" (SP) Cells From Murine Bone Marrow. J Proteomics Bioinform 2: 398-407.
- Raju K, Doulias PT, Tenopoulou M, Greene JL, Ischiropoulos H (2011) Strategies and tools to explore protein S-nitrosylation. Biochim Biophys Acta.
- Zhang R, Barton A, Brittenden J, Huang JT, Crowther D (2010) Evaluation for Computational Platforms of LC-MS Based Label-Free Quantitative Proteomics: A Global View. J Proteomics Bioinform 3: 260-265.
- Jaffe JD, Mani DR, Leptos KC, Church GM, Gillette MA, et al. (2006) PEPPeR, a platform for experimental proteomic pattern recognition. Mol Cell Proteomics 5: 1927-1941.
- 81. Li XJ, Yi EC, Kemp CJ, Zhang H, Aebersold R (2005) A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry. Mol Cell Proteomics 4: 1328-1340.
- May D, Fitzgibbon M, Liu Y, Holzman T, Eng J, et al. (2007) A platform for accurate mass and time analyses of mass spectrometry data. J Proteome Res 6: 2685-2694.
- Mueller LN, Brusniak MY, Mani DR, Aebersold R (2008) An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. J Proteome Res 7: 51-61.
- 84. Sturm M, Bertsch A, Gropl C, Hildebrandt A, Hussong R, et al. (2008) OpenMS - an open-source software framework for mass spectrometry. BMC Bioinformatics 9: 163.
- 85. Kosaihira S, Tsunehiro Y, Tsuta K, Tochigi N, Gemma A, et al. (2009) Proteome Expression Database of Lung Adenocarcinoma: a segment of the Genome Medicine Database of Japan Proteomics. J Proteomics Bioinform 2: 463-465.
- Ahmed FE (2009) Utility of mass spectrometry for proteome analysis: part II. Ion-activation methods, statistics, bioinformatics and annotation. Expert Rev Proteomics 6: 171-191.
- Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 74: 5383-5392.
- Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 75: 46460-4658.

 Cerqueira FR, Morandell S, Ascher S, Mechtler K, Huber LA, et al. (2009) Improving Phosphopeptide/protein Identification using a New Data Mining Framework for MS/MS Spectra Preprocessing. J Proteomics Bioinform 2: 150-164.

Page 9 of 9

- Gustafsson JOR, McColl SR, Hoffmann P (2008) Imaging Mass Spectrometry and Its Methodological Application to Murine Tissue. J Proteomics Bioinform 1: 458-463.
- 91. Li X (2001) PROcess R Library by Xiaochun Li / R 2:1-1.
- Noy K, Fasulo D (2007) Improved model-based, platform- independent feature extraction for mass spectrometry. Bioinformatics 23: 2528-2535
- 93. Atlas M, Datta S (2009) A Statistical Technique for Monoisotopic Peak Detection in a Mass Spectrum. J Proteomics Bioinform 2: 202-216.
- 94. Ezzat MO, AL- Obaidi OHS, Mordi MN (2011) Theoretical Study of Benzylic Oxidation and Effect of Para- Substituents by Using Hyperchem Program. J Proteomics Bioinform 4: 113-115.
- 95. Liangpunsakul S, Lai X, Ringham HN, Crabb DW, Witzmann FA (2009) Serum Proteomic Profiles in Subjects with Heavy Alcohol Abuse. J Proteomics Bioinform 2: 236-243.
- 96. Narayanan A, Zhou W, Ross M, Tang J, Liotta L, et al. (2009) Discovery of Infectious Disease Biomarkers in Murine Anthrax Model Using Mass Spectrometry of the Low-Molecular-Mass Serum Proteome. J Proteomics Bioinform 2: 408-415.
- 97. Lu M, Whitelegge JP, Whelan SA, He J, Saxton RE (2010) Hydrophobic Fractionation Enhances Novel Protein Detection by Mass Spectrometry in Triple Negative Breast Cancer. J Proteomics Bioinform 3: 01-10.
- Kikuta K, Tsunehiro Y, Yoshida A, Tochigi N, Hirohahsi S, et al. (2009) Proteome Expression Database of Ewing sarcoma: a segment of the Genome Medicine Database of Japan Proteomics. J Proteomics Bioinform 2: 500-504.
- Frahm JL, Li LO, Grevengoed TJ, Coleman RA (2011) Phosphorylation and Acetylation of Acyl-CoA Synthetase- I. J Proteomics Bioinform 4: 129-137.
- 100. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4: 207-214.
- 101.Zhou W, Capello M, Fredolini C, Racanicchi L, Piemonti L, et al. (2011) Phosphoproteomic Analysis of Pancreatic Ductal Adenocarcinoma Cells Reveals Differential Phosphorylation of Cell Adhesion, Cell Junction and Structural Proteins. J Proteomics Bioinform 4: 170-178.