Journal of Computational Intelligence in Bioinformatics
ISSN 0973-385X Volume 6, Number 2 (2013) pp. 75-79

http://www.ripublication.com/jcib.htm

# Analysis of Mass Spectra Resources for Identifying the Novel Marker(s) for Non Obstructive Azoospermia

**Varsha Ravikumar, Tanushree Chaudhuri and Kusum Paul**

*Department of Biotechnology, The Oxford college of Engineering, Bommanahalli, Hosur road, Bangalore – 560 068, Karnataka state, India.*
*Corresponding author: varshar304@gmail.com*

**Abstract**

The identification of new biomarker is one of the most important task in the field of clinical and pharmaceutical research. The testis produce male gametes in the germinal epithelium. About 10 to 15 % of cases of male infertility are due to azoosperima. We for the first time identifying the novel marker by doing genome wide analysis of mass spectra data for infertility (Non obstructive Azoospermia). Proteins can particularly be good biomarkers. One of the key approach is to first establish which proteins are expressed in normal and disease conditions of any specific tissue of interest and then identify the proteins that are differentially produced or not-produced in the disease condition. Mass spectrometry has become one of the most informative methods for studying proteins.After extracting proteins of significance, they were compared with previously established microarray data to identify potential biomarkers for normal mouse testis. 2374 proteins are identified in NOA condition by Mass spectrometry experiments. 583 proteins were found only in NOA condition and not found in normal condition. COL6A2,ANG, GMPPA,MAN2B2 andCLSTN3 proteins can be novel markers for NOA condition.

## Introduction:

Proteins play an important role in disease initiation, progress and even identification of disease-stages. Identified potential marker can be used as a drug targets. Proteins are the most abundant element present in living organisms. Mass spectrometry had

become one of the most informative methods for studying proteins and standard technique for protein identification, Quantification and Characterization in proteomics data. Mass spectrometry can be used as a large-scale technique and can be used for identifying the biomarker in proteomics data with more accuracy. Proteomic studies may identify potential markers for infertility and other diseases of the genito-urinary tract [1].Mass spectrometry-based search for biomarker patterns is widely recognized as a valuable research tool for predictive medicine and pharmacological monitoring [2].

Normally identification of markers is done based on single or multiple samples, but we tried to identify the markers by genome wide analysis of mass spectra data. That is we collected all the mass spectra data which are available in the mass spectra resources for NOA condition.

Human infertility affects ~15% of couples, with the male contributing to the infertility in 50% of all cases [3, 4]. One of the most severe forms of male infertility is azoospermia, which is characterized by an absence of sperm in the semen [5].

## Materials and methods:

There are several mass spectrometry based proteomics databases are publicly available. The main resources are: PRoteomics IDEntifications database(PRIDE), Global Proteome Machine database(GPMD), Peptide Atlas, Tranche and NCBI Peptidome [6].Initially searched all databases for the NOA condition and we found datasets only in GPM database. In the available resources Tranche and NCBI Peptidome were not working.

We found 12 datasets for NOA condition from GPM database. Mgex tdb is the database which gives the information (list) of genes expressed in testis in different conditions. This database list the specific genes present in testis tissue for different conditions [7]. The proteins identified by mass spectrometry are then compared with the list of specific genes for NOA condition obtained from Mgex tdb. Compared that specific list of proteins with the normal testis specific genes obtained from Mgex tdb. The proteins which are present in NOA condition and absent in normal testis were identified.

Protein score is calculated for identified proteins by using the data obtained from the GPM database.

Score is calculated by using the formula:

Score=.1(A) +.2(B)

Where

A ➔ The number of distinct peptide identified for that particular protein.

B➔ The number of repeated peptide identified for that particular protein.

High score proteins are identified and then based on protein- protein interaction we identified novel protein biomarker for NOA condition.

## Result and Discussions:

In GPM database we found 12 experiments for NOA condition by mass spectra analysis of seminal plasma sample taken from the people suffering from NOA Table1. 2374 proteins were identified in the seminal plasma sample of NOA affected peoples by mass spectrometry analysis. Among 2374 proteins 1067 were specific to NOA condition. Around 700 proteins were identified in all 12 experiments fig1.so more confidently we can say that expression of that proteins in NOA condition. 583 proteins were found only in NOA condition and absent in normal condition. Protein scores are calculated for 583 proteins. The proteins which have less number of interaction can become a good marker.Based on the protein score and protein-protein interactions and number of experiments identified that particular protein COL6A2,ANG, GMPPA,MAN2B2 andCLSTN3 Table 2.

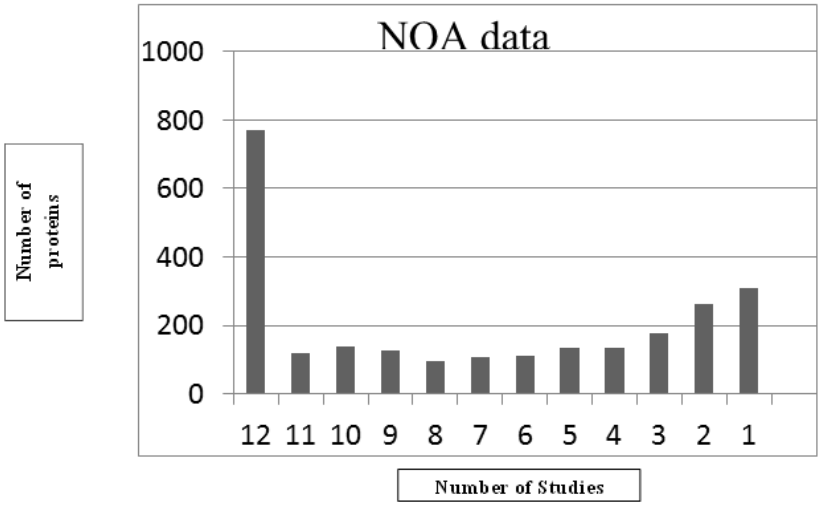


**Fig1.** The graph shows around 700 proteins were identified by all the studies, 120 proteins identified by 11 studies.

**Table1:** NOA dataset details:

| ACC No | Identified Protein | Identified Peptide |
|---|---|---|
| GPM32010005808 | 1500 | total selected peptides = 33433 |
| | | unique selected peptides = 7170 |
| GPM32010005809 | 1526 | total selected peptides = 32802 |
| | | unique selected peptides = 7213 |
| GPM32010005810 | 1513 | total selected peptides = 34845 |
| | | unique selected peptides = 7243 |
| GPM32010005812 | 1418 | total selected peptides = 38745 |
| | | unique selected peptides = 6969 |
| GPM32010005813 | 1468 | total selected peptides = 42796 |

| | | |
|---|---|---|
| | | unique selected peptides = 7392 |
| GPM32010005814 | 1482 | total selected peptides = 44138 |
| | | unique selected peptides = 7301 |
| GPM32010005815 | 1500 | total selected peptides = 33115 |
| | | unique selected peptides = 7170 |
| GPM32010005816 | 1526 | total selected peptides = 32802 |
| | | unique selected peptides = 7213 |
| GPM32010005817 | 1540 | total selected peptides = 37043 |
| | | unique selected peptides = 7397 |
| GPM32010005818 | 1300 | total selected peptides = 41063 |
| | | unique selected peptides = 6379 |
| GPM32010005819 | 1464 | total selected peptides = 39678 |
| | | unique selected peptides = 7322 |
| GPM32010005820 | 1651 | total selected peptides = 44837 |
| | | unique selected peptides = 8461 |

Table1. This table shows the number of studies available for NOA condition and the number of proteins, peptide identified in each studies.

**Table2:** Details of top 10 proteins:

| Protein | Protein score | Protein –protein interaction | Number of studies identifies that protein |
|---|---|---|---|
| CLSTN3 | 15.23333 | 0 | 9 |
| MAN2B2 | 19.85833 | 1 | 10 |
| COL6A2 | 78.875 | 3 | 12 |
| AMPD3 | 8.633333 | 3 | 2 |
| PDE9A | 8.025 | 3 | 8 |
| SRGAP1 | 49.5875 | 4 | 1 |
| BCAM | 11.66667 | 4 | 7 |
| GMPPA | 9.366667 | 5 | 11 |
| ANG | 15.5 | 7 | 12 |
| NUDT5 | 33.2 | 11 | 5 |

Table 2. This table shows the details of top 10 proteins, which contains the name of protein, protein score, protein-protein interaction, number of studies identifies that proteins.

## Acknowledgments:

We extend our sincere thanks to the management of The Oxford College of Engineering for their support. We also thank our principal Dr. Nagaraj and our HOD Dr. Kusum Paul, Department of Biotechnology for providing necessary resources.

## References:

[1] Jimenez CR., 2008,"Mass spectrometry-based proteomics: trends in tools and strategies", J Proteome Res.

[2] Batruch, Lecker, Kagedan D, Smith CR, Mullen BJ, Grober E, Lo KC, Diamandis EP, Jarvi KA., 2011, " Proteomic analysis of seminal plasma from normal volunteers and post-vasectomy patients identifies over 2000 proteins and candidate biomarkers of the urogenital system",J Proteome Res.

[3] Cooke H. J., Saunders P. T., 2002, "Mouse models of male infertility",Nat. Rev. Genet. 3, 790–801.

[4] De Kretser D. M., Baker H. W., 1999, "Infertility in men: Recent advances and continuing controversies", J. Clin. Endocrinol. Metab. 84, 3443–3450.

[5] Kolettis P. N., 2002,"The evaluation and management of the azoospermic patient",J. Androl. 23, 293–305.

[6] Juan Antonio Vizcaíno, Joseph M. Foster and Lennart Martens, 2010,"Proteomics data repositories: Providing a safe haven for your data and acting as a springboard for further research", J Proteomics. 73(11), 2136–2146.

[7] Kshitish K Acharya, Darshan S Chandrashekar, Neelima Chitturi, Hardik Shah, Varun Malhotra, Sreelakshmi KS, Deepti H, Akhilesh Bajpai, Sravanthi Davuluri, Pranami Bora and Leena Rao, 2010,"A novel tissue-specific meta-analysis approach for gene expression predictions, initiated with a mammalian gene expression testis database", BMC Genomics, 11:467.