INNSPUB

**International Journal of Biosciences | IJB |**
**ISSN: 2220-6655 (Print), 2222-5234 (Online)**
**http://www.innspub.net**
**Vol. 16, No. 1, p. 107-119, 2020**

RESEARCH PAPER OPEN ACCESS

# Identification of Biomarker Signatures and Candidate Drugs in Non-Small Cell Lung Cancer

**Md. Parvez Mosharaf[1], Amina Rownaq[2], S.M. Shahinul Islam[2], Md. Nurul Haque Mollah[1*]**

[1]*Bioinformatics Lab., Department of Statistics, University of Rajshahi, Bangladesh*

[2]*Institute of Biological Sciences, University of Rajshahi, Bangladesh*

**Key words:** Non-small cell lung cancer, Molecular signature, Molecular pathway, differentially expressed genes, Protein-protein interaction.

http://dx.doi.org/10.12692/ijb/16.1.107-119 Article published on January 15, 2020

## Abstract

Lung cancer is the most important health risk for human in worldwide. Non-small cell lung cancer (NSCLC) is the most common cause of premature death from malignant disease. The aim of the study was to determine the pathways and expression profile of the genes to discover molecular signature at RNA and protein levels which could serve as potential drug targets for therapeutics innovation and the identification of novel targets. Eight proteins, six TFs and seven miRNAs came into prominence as potential drug targets. The differential expression profiles of these reporter biomolecules were cross-validated by independent RNA-Seq and miRNA-Seq. Risk discrimination performance of the reporter biomolecules NPR3, JUN, PPARG, TP53, CKMT1A, SP3 and TFAP2A were also evaluated. Total 213 drugs and 7 proteins was found for non-small cell lung cancer through dgidb. Among these identified drugs seven drugs such as- Gemcitabine, Carboplatin, paclitaxel, Docetaxel, Crizotinib, Bevacizumab and Gemcitabine is used for NSCLC which is approved by National Cancer Institute. The molecular signatures and repurposed drugs presented here permit further attention for experimental studies which are offer significant potential as biomarkers and candidate therapeutics for precision medicine approaches to clinical management of NSCLC.

[*] **Corresponding Author:** Md. Nurul Haque Mollah ✉ mollah.stat.bio@ru.ac.bd

## Introduction

Lung cancer is the second leading cause of cancer death worldwide among the human cancer. In Bangladesh, Lung cancers deaths reached 9,660 or 1.33% of total deaths according to the latest WHO data published in May, 2014. International Agency for Research on Cancer calculated cancer-related death rates in Bangladesh was 7.5% in 2005 and it will be 13% in 2030. Smoking, alcohol and high air pollutions are the main risk factor of lung cancer (Alberg *et al.*, 2013). Non-small cell lung cancer (NSCLC) account for approximately 75% of all of lung cancers, which is the most common type of bronchial tumor in the world (Jemal *et al.*, 2007). Conventional diagnosis of lung cancer is currently based on tumor histology. The main histological types of NSCLC include adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Despite the well-defined histological types of NSCLC, patient survival and treatment outcomes can differ substantially, even among NSCLCs of the same histological type and stage. Both sub-types adenocarcinoma and squamous cell carcinoma are very different in regard to copy number, DNA methylation, genetic mutations, transcriptome, proteome and biomarkers. Defining different lung cancer entities based on clinical, pathological and molecular alterations also determines disease evolution and treatment options. In spite of significant progress in the development of targeted-therapy, the high mortality rate in lung cancer firmly emphasizes the need for prevention and efficient detection of lung cancer, as well as a better classification, enabling patients to benefit from more specific therapy. Microarrays have been Lung cancer morphology exhibits a broad spectrum and many tumors are atypical or lack the morphologic features necessary for improved differential diagnoses. Lung cancer diagnoses based solely on morphological features are in many cases insufficient (D'Amico, 2008).

In recent years, there has been an expanding interest in the molecular genetic classification of cancers. Gene expression profiling is one of the methods used for this purpose. The data obtained from gene expression analyses may allow the differentiation of cancer subgroups based on molecular phenotypes. Subgroups determined by gene expression in combination with clinical features like spread reflect the cancer's biology and progression better than histology alone (Golub, 2001).

The aim to reduce the high morbidity rate in lung cancer is related to growing efficiency of early detection and prevention of cancer risk factor strategies of lung cancer while still in a curable stage. Depend on the site of the primary tumor, these patient are treated with surgery and/or radiotherapy. Despite advances in early detection and standard treatment, NSCLC is often diagnosed at an advanced stage and carries a poor prognosis. Greater knowledge of the molecular origins and progression of lung cancer may lead to improvements in the treatment and prevention of the disease. Gene expression profiling offers a more functional molecular understanding of lung cancer that may grant insights into its pathophysiology and yield relevant information for staging, prognosis and therapeutic decision-making.

Microarrays have been used for more than two decades in pre-clinical research to determine gene-expression profiling associated with different pathological sub-groups or diverse clinical evolution and to evaluate biological and cellular functions determined by these gene panels. Improved outcomes for NSCLC are therefore clearly needed. There are a number of different strategies currently under evaluation. However, a better understanding of the molecular mechanisms that determine clinical outcomes is likely to provide the basis for more effective therapeutic intervention. Questions have been raised regarding the reproducibility of microarray result and there are examples of gene expression signature that even though validated in independent patient cohorts, show relative little overlap of genes. It has also been stated that differential gene expression on transcript level matches protein abundance to only about 40% (Tian *et al.*, 2004).

In the present study, system biomedicine framework was used to identify molecular biomarker signature and inform systemic drug targeting in non-small cell lung cancer (Fig. 1).

For this purpose, gene expression profiles from lung cancer were subjected to integrative analyses with genome-scale biomolecular networks to reveal molecular signature at mRNA, miRNA and protein levels, which serve as potential drug targets for therapeutics innovation. The differential expression profiles and the risk discrimination performance of the reporter biomolecules were cross-validated in independent transcriptome datasets.

## Materials and methods

*Gene Expression Profiling in Non-Small Cell Lung Cancer*

For non-small cell lung cancer, the transcriptome data (GSE54712) was obtained from the previous study (Lopez *et al.*, 2014) through the publicly available Gene Expression Omnibus (NCBI-GEO) dataset (Barrett *et al.*, 2013), mRNA profiling was hybridized to in-situ oligonucleotide microarrays (Agilent-014850 Whole Human Genome Microarray 4x44K).

*Differentially expressed gene*

The gene expression data set was normalized for identifying DEGs through the Robust Multi-Array Average (RMA) expression measure and it was implemented in the "LIMMA" package of

Bioconductor platform. To control the false discovery rate in multiple-testing the p-values were adjusted by Benjamini Hochberg's method and by using adjusted p-value >0.01 criterion was cutoff to designate statistical significance.

The hierarchical clustering analysis was applied to categorize patients and gene into groups using Heatmap3 package. All analyses packages were implemented in R (version 3.2.3), analysis were performed through the Bioconductor platform (Gentleman *et al.*, 2004).

*Gene function enrichment*

To find out molecular function, biological process and molecular pathway annotations of the identified DEGs through Database for Annotation, Visualization and Integrated Discovery (DAVID) (version v6.8) bioinformatics resources by using gene overrepresentation analyses. Whole set of genome annotation for human genome was used as the background reference set. Analyses were carried out by using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database as the annotation sources. DAVID bioinformatics resources were used to find out molecular pathway annotations of the identified DEGs (Huang *et al.*, 2009). P-values were determined by using Fisher Exact test and Benjamini-Hochberg's correction was used the multiple testing correction technique. Results with adjusted-p <0.01 were considered as statistically significant.

*Analysis and Reconstruction of Protein-Protein Interaction Sub-network*

The previously reconstructed protein-protein interaction (PPI) network of *Homo sapiens* (Karagoz *et al.*, 2016), which consist of 288,033 physical interactions between 21,052 proteins, was recruited in the present study to construct a PPI subnetwork around the protein encoded by the identified DEGs.

The subnetwork was shown as an undirected graph, where nodes represent the proteins and the edges represent the interactions between the proteins. The sub-network was visualized and analyzed via Cytoscape (v3.6.1).

To determine highly connected proteins (i,e., hub proteins) a topological analysis was applied through Cyto-Hubba plugin (Chin *et al.*, 2009) and the dual-metric approach considering degree and between's centrality metrics simultaneously were employed (Calimlioglu *et al.*, 2015).

*Determination of Reporter Transcription Factors and miRNAs*

Transcriptional regulatory biomolecules (i.e. TFs) was obtained through human transcriptional regulation interaction data bases TRRUST v2 (www.grnpedia.org/trrust) which significant changes has occurred at transcriptional level. To identify miRNAs around which significant changes occur at transcriptional level, the experimentally verified miRNA-target gene interactions was involved in this study which were obtained through human transcriptional regulation interaction (Bovolenta *et al.*, 2012) and miRTarbase (Release 6.0) (Chou *et al.*, 2016) as well as previous study (Gov *et al.,* 2017). To obtained z-score and corresponding p-values of the regulatory molecules, the reporter features algorithms (Patil and Nielsen, 2005) was used and implemented as described previously (Kori *et al.*, 2016). P-values were corrected via Benjamini-Hochberg's methods and statistically significant (adjusted p <0.01) results were considered as reporter regulatory elements.

*Cross-validation and Evaluation of Performance of Reporter Biomolecules*

According to their prognostic index and survival multivariate the patients were divided into low and high-risk group and risk assessments were performed through oncolnc (Anaya, 2016).

The differences between the risk groups in gene expression levels were represented via box plots and statistical significance of the differences was estimated through t-test. Survival signatures of reporter biomolecules were evaluated through Kaplan-Meier plots and a log –rank p-value <0.01 was considered as a cut-off to describe statistical significance in all survival analyses.

*Identification of Candidate Drugs through Drug Repositioning*

To identify novel therapeutics in NSCLC through drug repositioning the transcriptome guided drug repositioning tool and Drug Gene Interaction Database (dgidb) (Cotto *et al.*, 2017) was used to designate the candidate drugs targeting the hub protein and reporter TFs. Total 213 drugs and 7 proteins was found for non-small cell lung cancer through dgidb. Drugs associated with each disease dataset were expected by employing the hypergeometric probability test. For exploring more accurate drug candidates for specific disease, it is suggested to search based on the particular condition (Turanli *et al.*, 2017).

## Results

*Genome Reprogramming in Non-Small Cell Lung Cancer*

In this present study, statistical test LIMMA was used to identify DEGs from noise and outliers in the transcriptome dataset. The parametric method LIMMA was identified 380 DEGs with statistical significance of adjusted p<0.01. Among those, 169 genes were up regulated, whereas 211 genes were down regulated (Fig. 2A). Through hierarchical clustering analysis based on the profiles of DEGs, it was shown that the tumor and normal sample were distinguished with 100% specificity and 92% sensitivity (Fig. 2B). These DEGs were considered in further analysis.

To clarify the molecular functions, biological processes, cellular components and KEGG pathways functional overrepresentation were carried out, which associated with proteins encoded by the DEGs. GO analysis was performed by DAVID results represent that both up regulated and down regulated DEGs are significantly enriched in biological processes (BP), where up regulated genes include regulation of small GTPase mediated signal transduction, regulation of Rho protein signal transduction, positive regulation of GTPase activity, small GTPase mediated signal transduction, positive regulation of GTPase activity, cell adhesion, regulation of heart rate by cardiac conduction (Table 1) and down regulated genes include negative regulation of endopeptidase activity, oxidation-reduction process, daunorubicin metabolic process, doxorubicin metabolic process, progesterone metabolic process (Table 1).

In case of cellular component (CC), up regulated genes are involved in plasma membrane extracellular region, integral component of plasma membrane, intercalated disc, cell junction, myosin complex (Table 1) and down regulated genes are involved in integral component of plasma membrane, integral component of membrane, proteinaceous extracellular matrix (Table 1). In addition, up regulated genes were enriched in molecular function (MF) including guanyl-nucleotide exchange factor activity, Rho guanyl-nucleotide exchange factor activity, serine-type endopeptidase inhibitor activity, calmodulin binding, actin filament binding (Table 1) and down regulared enriched in serine-type endopeptidase inhibitor activity, oxidoreductase activity, ketosteroid monooxygenase activity, trans-1,2-dihydrobenzene-1,2-diol dehydrogenase activity, phenanthrene 9,10-monooxygenase activity, chemorepellent activity, symporter activity, metallopeptidase activity (Table 1).

**Table 1.** Functional overrepresentation of differentially expressed genes in non-small cell lung cancer.

| Gene Cluster | Category | Pathway | No. of gene | P-value |
|---|---|---|---|---|
| Cluster 1 (Up Regulated Genes) | Gene Ontology Biological Process | Regulation of small GTPase mediated signal transduction | 5 | 2.0E-2 |
| | | regulation of Rho protein signal transduction | 4 | 2.5E-2 |
| | | positive regulation of GTPase activity | 9 | 7.2E-2 |
| | | small GTPase mediated signal transduction | 6 | 4.3E-2 |
| | | positive regulation of GTPase activity | 9 | 7.2E-2 |
| | | regulation of heart rate by cardiac conduction | 3 | 3.0E-2 |
| | | extracellular matrix organization | 7 | 4.2E-3 |
| | | cell adhesion | 11 | 3.0E-3 |
| | Gene Ontology Cellular Component | plasma membrane | 45 | 3.6E-3 |
| | | extracellular region | 19 | 4.5E-2 |
| | | integral component of plasma membrane | 18 | 2.9E-2 |
| | | intercalated disc | 4 | 4.3E-3 |
| | | cell junction | 9 | 1.9E-2 |
| | | myosin complex | 3 | 5.2E-2 |
| | Gene Ontology Molecular Function | guanyl-nucleotide exchange factor activity | 6 | 2.1E-3 |
| | | Rho guanyl-nucleotide exchange factor activity | 4 | 2.1E-2 |
| | | serine-type endopeptidase inhibitor activity | 4 | 3.8E-2 |
| | | calmodulin binding | 5 | 5.7E-2 |
| | | actin filament binding | 4 | 8.1E-2 |
| | | motor activity | 3 | 9.1E-2 |
| | KEGG Pathways | ECM-receptor interaction | 3 | 1.6E-1 |
| | | Focal adhesion | 4 | 2.3E-1 |
| | | PI3K-Akt signaling pathway | 4 | 5.4E-1 |
| Cluster 2 (Down Regulated Genes) | Gene Ontology Biological Process | negative regulation of endopeptidase activity | 8 | 3.3E-4 |
| | | oxidation-reduction process | 15 | 4.7E-3 |
| | | cellular response to jasmonic acid stimulus | 3 | 6.8E-4 |
| | | daunorubicin metabolic process | 3 | 3.1E-3 |
| | | doxorubicin metabolic process | 3 | 3.1E-3 |
| | | progesterone metabolic process | 3 | 3.9E-3 |
| | Gene Ontology Cellular Component | plasma membrane | 63 | 4.3E-4 |
| | | integral component of plasma membrane | 26 | 4.7E-3 |
| | | integral component of membrane | 70 | 4.9E-3 |
| | | proteinaceous extracellular matrix | 7 | 5.6E-2 |
| | Gene Ontology Molecular Function | serine-type endopeptidase inhibitor activity | 7 | 5.2E-4 |
| | | oxidoreductase activity | 10 | 2.4E-4 |
| | | ketosteroid monooxygenase activity | 3 | 3.2E-4 |
| | | trans-1,2-dihydrobenzene-1,2-diol dehydrogenase activity | 3 | 6.3E-4 |
| | | phenanthrene 9,10-monooxygenase activity | 3 | 6.3E-4 |
| | | chemorepellent activity | 3 | 3.2E-2 |
| | | symporter activity | 4 | 1.6E-2 |
| | | metallopeptidase activity | 4 | 5.2E-2 |
| | KEGG Pathways | Phosphatidylinositol signaling system | 6 | 8.0E-3 |
| | | Pathways in cancer | 9 | 1.3E-1 |
| | | PI3K-Akt signaling pathway | 6 | 4.5E-1 |
| | | Regulation of actin cytoskeleton | 4 | 5.1E-1 |

The up regulated DEGs and down regulated DEGs both are significantly enriched which are analyzed in KEGG pathway analysis.

The results are represented up regulation of pathways in ECM-receptor interaction, focal adhesion and PI3K-Akt signaling pathway. On the other hand, down regulated DEGs were enriched in Phosphatidylinositol signaling system, Pathways in cancer, PI3K-Akt signaling pathway and regulation of actin cytoskeleton pathways which were represent in non-small cell lung cancer (Table 1).

**Table 2.** Reporter biomolecules are in non-small cell lung cancer. Statistically significant (p<0.01) values are given in bold and italic.

| Symbol | Description | Feature | Hazard ratio (TCGA) |
|---|---|---|---|
| PTPN20 | Protein Tyrosine Phosphatase, Non-Receptor Type 20 | Hub Protein | NA |
| ST6GALNAC2 | ST6 N-Acetylgalactosaminide Alpha-2,6-Sialyltransferase 2 | Hub Protein | 1.12 |
| ELMO1 | Engulfment And Cell Motility 1 | Hub Protein | *1.59* |
| TXLNGY | Taxilin Gamma Pseudogene, Y-Linked | Hub Protein | |
| NPR3 | Natriuretic Peptide Receptor 3 | Hub Protein | 1.03 |
| CFH | Complement Factor H | Hub Protein | 1.05 |
| CNTNAP3B | Contactin Associated Protein Like 3B | Hub Protein | NA |
| CKMT1A | Creatine Kinase, Mitochondrial 1A | Hub Protein | 1.21 |
| TP53 | tumor protein p53 | Reporter Transcription Factor | 1.02 |
| SP3 | SP3 transcription factor | Reporter Transcription Factor | 1.08 |
| JUN | JUN proto-oncogene | Reporter Transcription Factor | 1.08 |
| PPARG | Peroxisome proliferator- activated receptor gamma | Reporter Transcription Factor | 1.23 |
| TFAP2A | Transcription factor AP-2 alpha | Reporter Transcription Factor | *1.68* |
| TWIST1 | Twist basic helix-loop-helix transcription factor 1 | Reporter Transcription Factor | 1.25 |
| miR-146a-5p | MicoRNA-146a | Receptor mirRNA | NA |
| miR-210-3p | MIcroRNA-210 | Receptor mirRNA | NA |
| miR-374a-5p | MIcroRNA-374a | Receptor mirRNA | NA |
| miR-548b-3p | MIcroRNA-548b | Receptor mirRNA | NA |
| miR-544a | MIcroRNA-544a | Receptor mirRNA | NA |
| miR-152-3p | MIcroRNA-152 | Receptor mirRNA | NA |
| miR-204-5p | MIcroRNA-204 | Receptor mirRNA | NA |
| All Hubs | PTPN20, ST6GALNAC2, ELMO1, TXLNGY, NPR3, CFH, CNTNAP3B, CKMT1A | All Hub Proteins | 1.51 |
| All TFs | TP53, SP3, JUN, PPARG, TFAP2A, TWIST1 | All receptor transcription factors | *1.65* |

*Proteomic Signatures in Non-Small Cell Lung Cancer*

Central proteins, so called "hub proteins" have a significant role in signal transduction events during disease progression.

To identify these hub proteins a PPI sub-network was constructed around proteins encoded by the DEGs via their physical interactions and topological analysis was formed.

As a result, eight hub proteins were found (Table 2) and those are Protein Tyrosine Phosphatase, Non-Receptor Type 20 (PTPN20), ST6 N-Acetylgalactosaminide Alpha-2,6-Sialyltransferase 2 (ST6GALNAC2), Engulfment And Cell Motility 1 (ELMO1), Taxilin Gamma Pseudogene, Y-Linked (TXLNGY), Natriuretic Peptide Receptor 3 (NPR3), Complement Factor H (CFH), Contactin Associated Protein Like 3B (CNTNAP3B),Creatine Kinase U-type and mitochondria (CKMT1A).

The transcriptional and post-transcriptional regulatory components (i.e., TFs and miRNAs) occurs significant changes at transcriptional level, which were also identified. The results have found six TFs and seven miRNAs as reporter transcriptional regulatory components (Table 2).

**Table 3.** Selected repositioned drugs in non-small cell lung cancer.

| Repositioning drug | Drug Class/ Status/ Description |
|---|---|
| Docetaxel | Approved, Investigational. Docetaxel is a clinically well-established anti-mitotic chemotherapy medication used mainly for the treatment of breast, ovarian, and non-small cell lung cancer. Docetaxel reversibly binds to tubulin with high affinity in a 1:1 stoichiometric ratio. |
| Gemcitabine | Approved, Gemcitabine is used in various carcinomas: non-small cell lung cancer, pancreatic cancer, bladder cancer and breast cancer. It is being investigated for use in oesophageal cancer, and is used experimentally in lymphomas and various other tumor types. |
| Carboplatin | Approved an organoplatinum compound that possesses antineoplastic activity. |
| Paclitaxel | An organoplatinum compound that possesses antineoplastic activity. It is available as an intravenous solution for injection and the newer formulation contains albumin-bound paclitaxel marketed under the brand name Abraxane. |
| Crizotinib | Crizotinib an inhibitor of receptor tyrosine kinase for the treatment of non-small cell lung cancer (NSCLC). Verification of the presence of ALK fusion gene is done by Abbott Molecular's Vysis ALK Break Apart FISH Probe Kit. This verification is used to select for patients suitable for treatment. FDA approved in August 26, 2011. |
| Bevacizumab | A recombinant humanized monoclonal IgG1 antibody that binds to and inhibits the biologic activity of human vascular endothelial growth factor (VEGF). Bevacizumab contains human framework regions and the complementarity-determining regions of a murine antibody that binds to VEGF. |
| Gemcitabine | Gemcitabine is used in various carcinomas: non-small cell lung cancer, pancreatic cancer, bladder cancer and breast cancer. It is being investigated for use in oesophageal cancer, and is used experimentally in lymphomas and various other tumor types. |

### *Cross-validation and Risk Discrimination Performance of Reporter Biomolecules*

The differential expression signatures was validated and the analysis of the risk discrimination performance of reporter biomolecules (i.e., 8 hub proteins, 6 TFs and 7miRNAs) were performed using independent RNA-seq and miRNA-seq dataset obtained from TCGA. According to the expression levels of the reporter biomolecules, the samples were divided into two groups and groups are entitled as low-risk and high-risk considering their risk discrimination performance. The differences in the gene expression levels (encoding hub proteins or reporter TFs) between the risk groups were presented via box-plots and prognostic capabilities based on survival data were analyzed Kaplan-Meier plots, log-rank test and hazard ratio (Figs 3 and 4). Furthermore, the discrimination power of the all hub proteins and all reporter TFs were also analyzed and statistically significant performance in terms of survival probabilities in all datasets were found in both groups.

### *Identification of Candidate Drugs through Drug Repositioning*

Considering hub proteins and transcription factors as potential drug targets and the transcriptomic signatures of genes within the disease-gene-drug triad we identified potential drug for each target protein using the transcriptome guided drug repositioning tool and the Drug Gene Interaction Database (dgidb). As a result 213 drugs and 7 proteins was found (Fig. 5).

Among these identified drugs seven drugs is used for non-small cell lung cancer which is approved by National Cancer Institute. Gemcitabine, Carboplatin, paclitaxel, Docetaxel, Crizotinib, Bevacizumab and

Gemcitabine drugs are used for the treatment of NSCLC (Table 3).

## Discussion

Although the prevalence and mortality of NSCLC are among the highest worldwide and the disease has been studied extensively, there is still a need for more accurate prognostic and diagnostic markers as well as for information about the underlying molecular events in the disease's development. In the present study, we followed a multi-omics data integration framework to reveal molecular signatures at mRNA, miRNA and protein levels, which offers the promise as biomarkers and potential drug targets for efficacious treatments. In the gene expression study, we performed several analyses of NSCLC, including GO, clustering, new biomarker discovery, a comparison of survival based on histology and gene expression profiles, and analysis of patient survival.

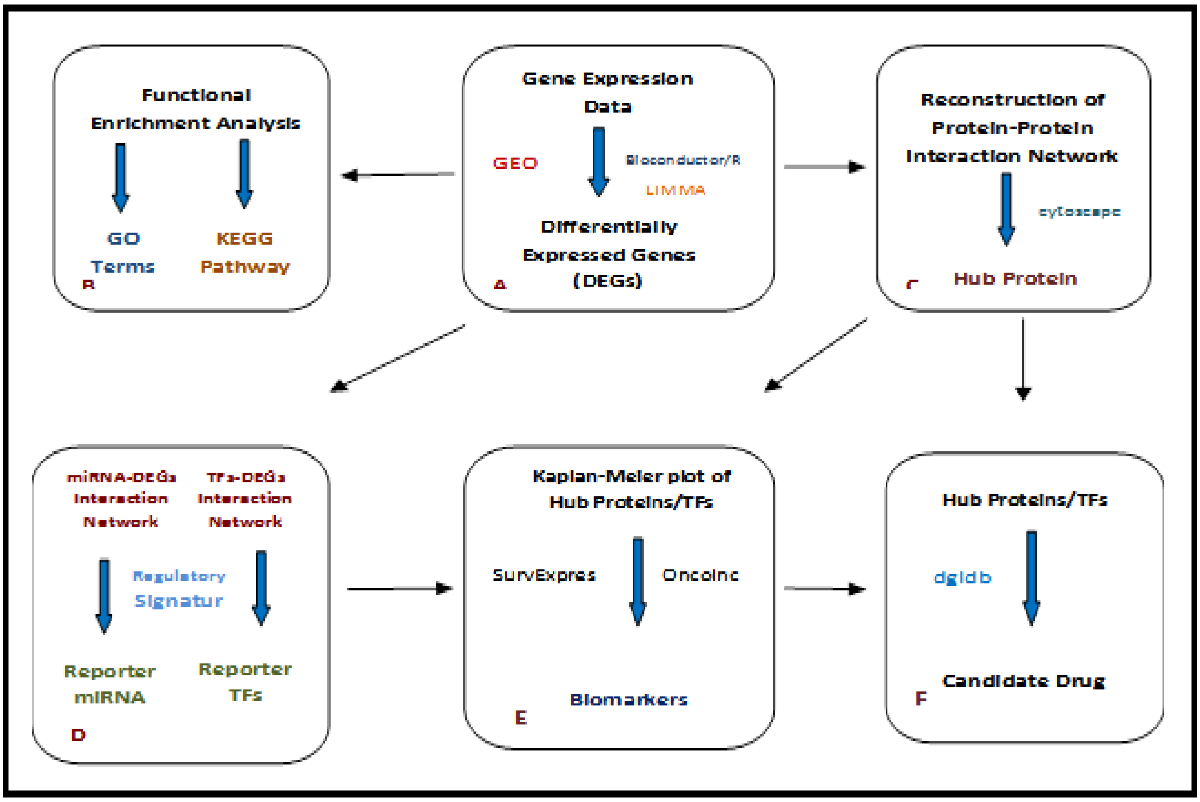

**Fig. 1.** The methodology of multi-stage analysis which is employed in the present study. (A) Gene expression data were obtained from the Gene Expression Omnibus (GEO) database. The dataset was statistically analyzed to identify differentially expressed genes (DEGs). (B) Functional enrichment of DEGs was performed to identify significantly enriched Gene Ontology (GO) terms and Pathways. (C) Protein-protein interaction network was reconstructed around DEGs and the topological analyses were performed via Cytoscape to identify hub proteins. (D) To identify reporter micro-RNA (miRNA) and transcription factors (TFs), DEGs were integrated with miRNA-target gene and TF-target interactions and the statistically significant miRNAs and TFs were considered as the reporter transcriptional regulatory elements. (E) The survival analysis of the hub biomolecules was done through The Cancer Genome Atlus (TCGA) NSCLC datasets via SurvExpress and Oncolnc. (F) The candidate drug molecules were identified by Drug Gene Interaction Database (dgidb).

The analysis of gene expression profiles in lung cancer samples resulted with a total of 380 DEGs with statistically significant changes in their expression profiles. Overrepresentation analyses for DEGs indicated the prominence of ECM organization and associated cell communication/signaling pathways (such as focal adhesion, ECM-receptor interaction pathways, and PI3K-Akt signaling pathway); hence, emphasized the critical roles of tumor microenvironment in NSCLC. Recent studies indicated the importance of tumor microenvironment as a decisive factor in tumorigenesis in various cancers (Dzobo *et al.*, 2016, Gollapalli *et al.*, 2017, Gov *et al.*, 2017, Hu *et al.*, 2017, Miskolczi *et al.*, 2018). GO analysis of statistically significant up regulated genes in NSCLC uncovered biological processes involved in cell adhesion and positive regulation of GTPase activity , whereas GO analysis of statistically significant down regulated genes in NSCLC uncovered genes involved in much more diverse processes, such as anatomical structure development, oxidation-reduction process and negative regulation of endopeptidase activity.

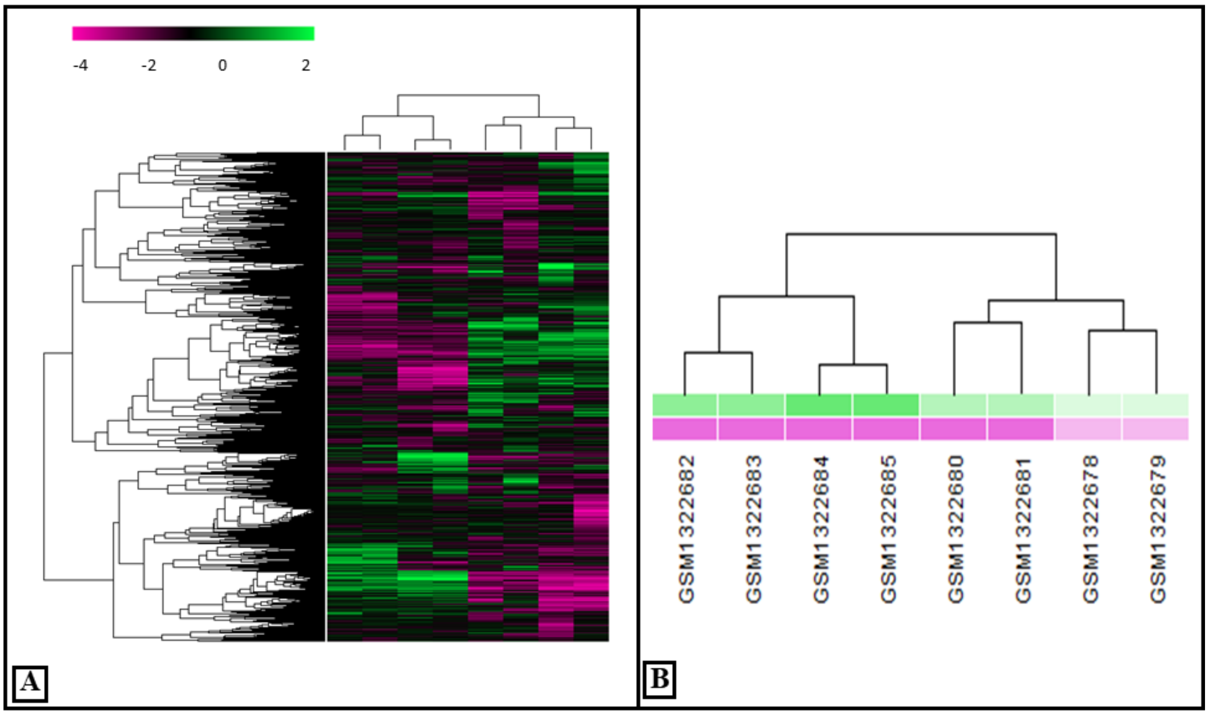

**Fig. 2.** (A & B): Heat map of differentially expressed genes and clustering analysis of DEGs. The above of the heat map presents clustering of samples and the left of the heat map presents clustering of the DEGs. Purple: high expression level and Green: low expression level. (A) Showed the hierarchical clustering (B) represent the clustering of samples/ patients.

Therefore, in the last decades, the reconstruction of accurate PPI sub-networks gained importance with the emergence of systems medicine and systems pharmacology concepts, which aim to offer solutions to many emerging problems associated with diseases such as identifying efficacious biomarkers and therapeutic targets for accurate diagnosis, prognosis, and therapeutics (Sevimoglu and Arga, 2014, Turanli and Arga, 2017). We reconstructed a PPI sub-network around the DEGs in NSCLC, and analyzed the topology of the reconstructed network to identify central proteins, i.e. hub proteins, that have the potential to contribute to the initiation and/or progression of cancer development. The regulation of gene expression is controlled by TFs and miRNAs at transcriptional and/or post-transcriptional levels; thus changes in these molecules may provide crucial information on dysregulation of gene expression in NSCLC. Therefore, reporter TFs and miRNAs have been identified.

The role of systems outlook gradually increases through improving existent approaches and developing novel concepts to provide possible solutions to the grand challenge in pharmacology, i.e., the development of efficacious treatment strategies for tackling the complexities of the diseases (Turanli and Arga, 2017).

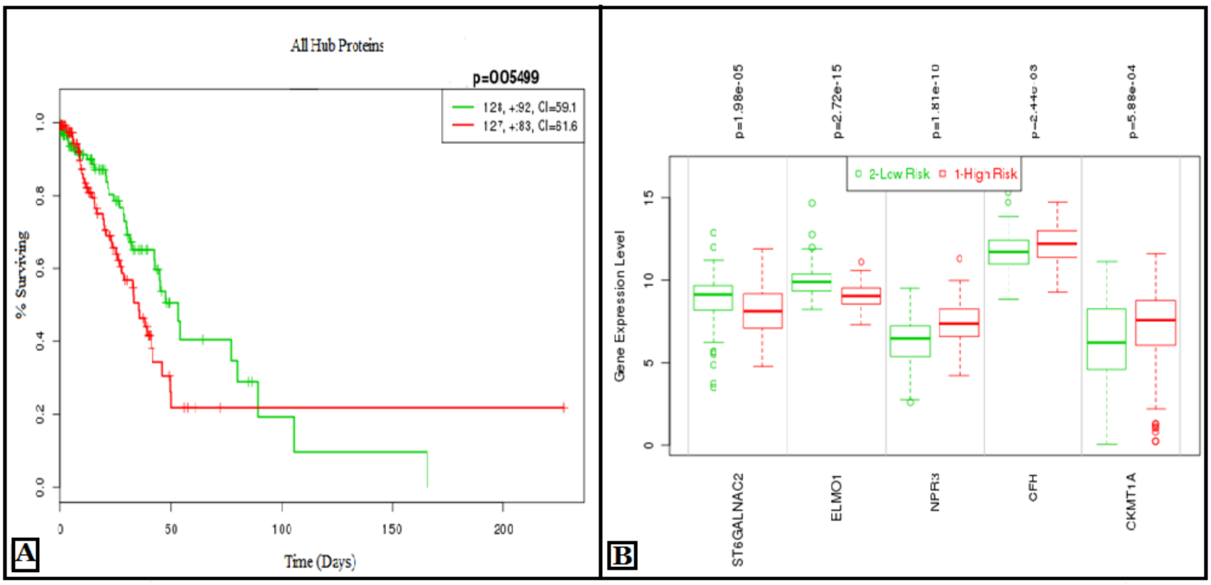

**Fig. 3.** Cross validation and prognostic performance analyses of hub proteins. (A) Kaplan-Meier plots representing the prognostic power of all hub proteins in NSCLC. (B) Box-plots present differential expression of genes encoding hub proteins in two risk groups.

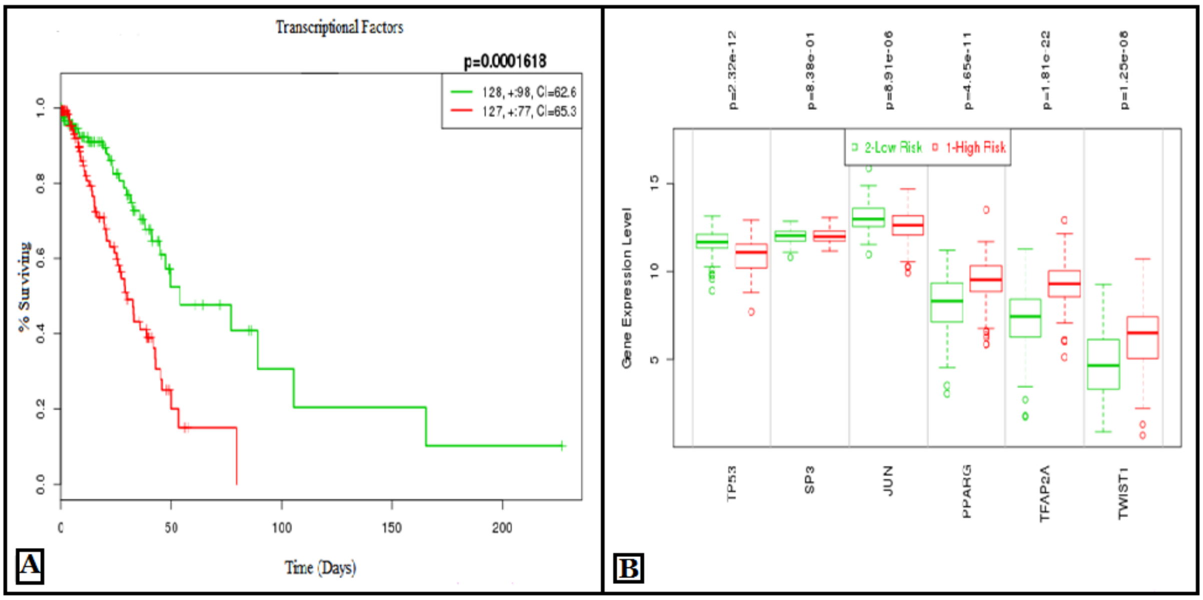

**Fig. 4.** Cross validation and prognostic performance analyses of Transcriptional Factors. (A) Kaplan-Meier plots representing the prognostic power of all TFs in NSCLC. (B) Box-plots present differential expression of genes encoding TFs in two risk groups.

The development of novel and effective treatment strategies for complex pathophysiology's, such as cancers, requires an understanding of the generation and progression mechanisms. Besides, systems biology research generates new information that can help in understanding the mechanisms behind disease pathogenesis to identify new biomarkers and therapeutic targets and to enable drug discovery. By using molecular signatures 213 drugs and seven proteins was identified. Among these drugs seven drugs is used for NSCLC which is approved by national cancer institute.

Integration of transcriptome datasets with biological network models provides molecular signatures, which will be beneficial for understanding etiopathogenesis and biological mechanisms of the diseases, and for developing effective and safe medications.

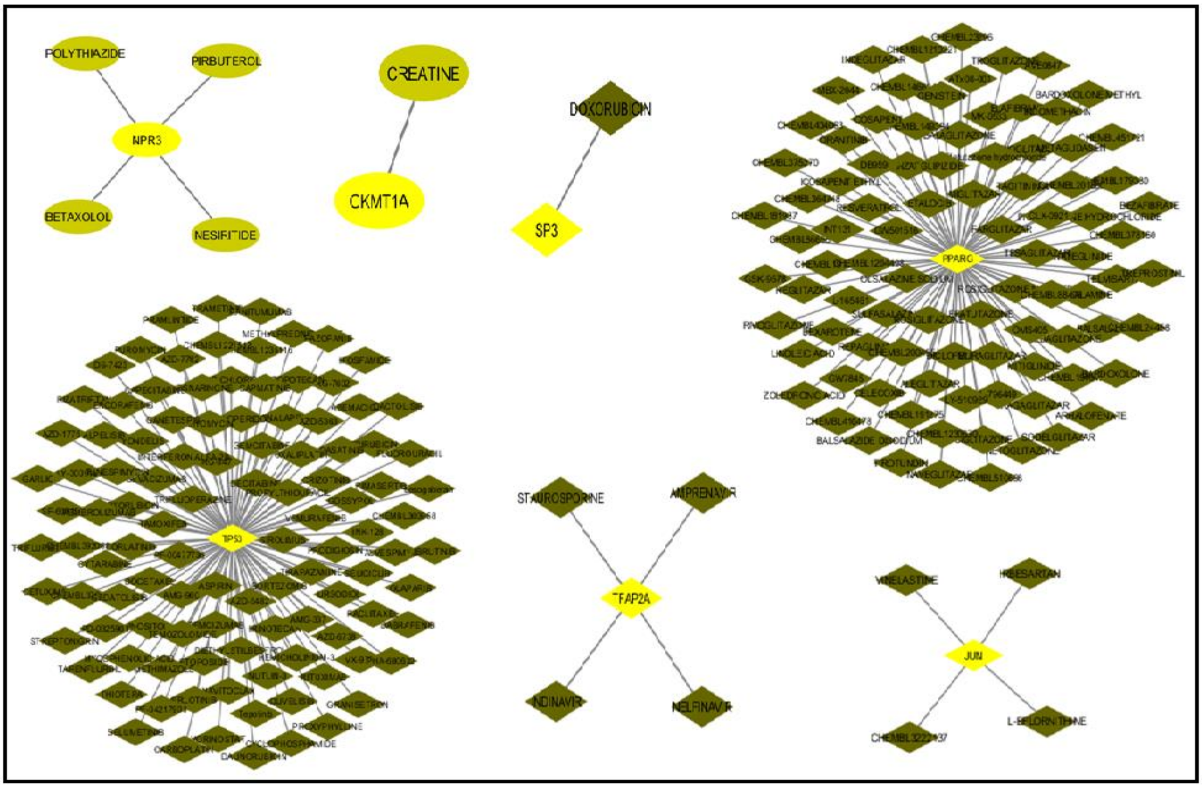

**Fig. 5.** Identified potential drugs for molecular signature by using the transcriptome guided drug repositioning tool and dgidb.

## Conclusion

In this present study, transcriptome data were integrated with genome-scale biological networks to reveal molecular signatures at RNA and protein levels.

Analysis of the genome reprogramming in NSCLC emphasized the decisive role of tumor microenvironment in tumorigenesis. Eight protein, six TFs and seven miRNA came into prominence as potential drug targets. The differential expression profile of these reporter biomolecules were cross validated in independent RNA-seq and miRNA-seq datasets. Moreover, seven novel candidate drugs were identified for NSCLC which is approved by national cancer institute.

## Acknowledges

We are grateful to Institute of Biological Sciences and Lab of Bioinformatics, University of Rajshahi, Rajshahi, Bangladesh for providing research facility. Also thanks to Ministry of Science and Technology (MOST) for providing scholarship of this study.

## Declaration

The authors declare no conflict of interest.

## References

**Alberg AJ, Brock MV, Ford JG, Samet JM, Spivack SD.** 2013. Epidemiology of lung cancer: Diagnosis and management of lung cancer. Third Edition: American College of Chest Physicians evidence-based clinical practice guidelines. Chest **143,** e1S, e29S.

**Anaya J.** 2016. OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. PeerJ Computer Science **2,** e67.

**Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall 1 KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Alexandra Soboleva A.** 2013. NCBI GEO: Archive for functional genomics data sets – Update. Nucleic Acids Research **41,** 991-995.

**Bovolenta LA, Acencio ML, Lemke N.** 2012. HTRIdb: An open-access database for experimentally verified human transcriptional regulation interactions. BMC Genomics **13,** 405.

**Calimlioglu B, Karagoz K, Sevimoglu T, Kilic E, Gov E, Arga KY.** 2015. Tissue-Specific Molecular Biomarker Signatures of Type 2 Diabetes: An Integrative Analysis of Transcriptomics and Protein-Protein Interaction Data. OMICS **19,** 563-573.

**Chin C, Chen S, Wu H.** 2009. Cyto-Hubba: A Cytoscape Plug-in for Hub Object Analysis in Network Biology. Genome Informatics **5,** 2-3.

**Chou CH, Chang NW, Shrestha S, Hsu SD, Lim YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ, Tsai TR, Ho SY, Jian TY, Wu HY, Chen PR, Lin NC, Huang HT, Yang TL, Pai CY, Tai CS, Chen WL, Huang CY, Liu CC, Weng SL, Liao KW, Hsu WL, Huang HD.** 2016. miRTarBase 2016: Updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Research **44,** D239-D247.

**Cotto KC, Wagner AH, Feng Y, Kiwala S, Coffman AC, Spies G, Wollam A, Spies NC, Griffith OL, Griffith M.** 2017. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. Nucleic Acids Research **17.**

**D'Amico TA.** 2008. Molecular biologic staging of lung cancer. The Annals of Thoracic Surgery **85,** S737- S742.

**Dzobo K, Senthebane DA, Rowe A, Thomford NE, Mwapagha LM, Al-Awwad N, Dandara C, Parker MI**. 2016. Cancer Stem Cell Hypothesis for Therapeutic Innovation in Clinical Oncology? Taking the Root Out, Not Chopping the Leaf. OMICS **20(12),** 681-691.

**Gentleman RC, Carey VJ and Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacas S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith G, Tierney L, Yang JY, Zhang J.** 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biology **5,** R80.

**Gollapalli K, Ghantasala S, Atak A, Rapole S, Moiyadi A, Epari S, Srivastava S.** 2017. Tissue Proteome Analysis of Different Grades of Human Gliomas Provides Major Cues for Glioma Pathogenesis. OMICS **21(5),** 275-284.

**Golub TR.** 2001. Genome-wide views of cancer. The New England Journal of Medicine **344,** 601–602.

**Gov E, Kori M, Arga KY.** 2017. Multiomics Analysis of Tumor Microenvironment Reveals Gata2 and miRNA-124-3p as Potential Novel Biomarkers in Ovarian Cancer. OMICS **21,** 603-615.

**Hu M, Qian C, Hu Z, Fei B, Zhou H.** 2017. Biomarkers in Tumor Microenvironment? Upregulation of Fibroblast Activation Protein-α Correlates with Gastric Cancer Progression and Poor Prognosis. OMICS **21(1),** 38-44.

**Huang Z, Cheng Y, Chiu PM, Cheung FM, Nicholls JM, Kwong DL, Lee AW, Zabarovsky ER, Stanbridge EJ, Lung HL, Lung ML.** 2012. Tumor suppressor Alpha B-crystallin (CRYAB) associates with the cadherin/catenin adherens junction and impairs NPC progression-associated properties. Oncogene **31,** 3709–3720.

**Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ.** 2007. Cancer statistics, CA: A Cancer Journal for Clinicians **57,** 43-66.

**Karagoz K, Sevimoglu T, Arga KY.** 2016. Integration of multiple biological features yields high confidence human protein interactome. Journal of Theoretical Biology **403,** 85-96.

**Kori M, Gov E, Arga KY.** 2016. Molecular signatures of ovarian diseases: Insights from network medicine perspective. Systems Biology in Reproductive Medicine **62,** 266–282.

**Lopez-Ayllon BD, Moncho-Amor V, Abarrategi A, Ibañez-de-Cáceres I, Castro-Carpeño J, Belda-Iniesta C, Perona R, Sastre L.** 2014. Cancer stem cells and cisplatin-resistant cells isolated from non-small-lung cancer cell lines constitute related cell populations. Cancer Medicine **3(5),** 1099-111.

**Miskolczi Z, Smith MP, Rowling EJ, Ferguson J, Barriuso J, Wellbrock C.** 2018. Collagen abundance controls phenotypes through lineage-specific microenvironment sensing. Oncogene. http://dx.doi.org/10.1038/s41388018-0209-0.

**Patil KR, Nielsen J.** 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proceedings of the National Academy of Sciences USA **102,** 2685-2689.

**Sevimoglu T, Arga KY.** 2014. The role of protein interaction networks in systems biomedicine. Computational and Structural Biotechnology Journal **11,** 22-27.

**Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, Yi EC, Dai H, Thorsson V, Eng J, Goodlett D, Berger JP, Gunter B, Linseley PS, Stoughton RB, Aebersold R, Collins SJ, Hanlon WA, Hood LE.** 2004. Integrated genomic and proteomic analyses of gene expression in Mammalian cells. Molecular & Cellular Proteomics **3(10),** 960-969.

**Turanli B, Arga KY.** 2017. Systems biomedicine acts as a driver for the evolution of pharmacology. Annals of Pharmacology and Pharmaceutics **2,** 1087.