Integrated Bioinformatics Analysis Deciphering the microRNA Regulation in Protein-Protein Interaction Network in Lung Adenocarcinoma

Pratyay Sengupta[a], Sayoni Saha[a], Moumita Maji[a], Monidipa Ghosh[a,*]

[a]Department of Biotechnology, National Institute of Technology Durgapur, Mahatma Gandhi Avenue, Durgapur, West Bengal 713209, India

*Author to whom correspondence to be addressed

Email: monidipa.ghosh@bt.nitdgp.ac.in (M.G.)

Abstract

Background: The architecture of the protein-protein interaction (PPI) network in any organism relies on their gene expression signature. microRNAs (miRNAs) have recently emerged as major post transcriptional regulators that control PPI by targeting mainly untranslated regions of the gene encoding proteins. Here, we aimed to unveil the role of miRNAs in the PPI network for identifying potential molecular targets for lung adenocarcinoma (LUAD).

Materials and methods: The expression profiles of miRNAs and mRNAs were collected from the NCBI Gene Expression Omnibus (GEO) database (GSE74190 and GSE116959). Abnormally expressed mRNAs from the data were appointed to construct a PPI network and hence incorporated with the miRNA-mRNA regulatory network. The miRNAs and mRNAs in this network were subjected to functional enrichment. Through the network analysis, hubs were identified and their mutation rate and probability of cooccurrence were calculated.

Results:

We identified 17 miRNAs and 429 mRNAs signature for differentially altered transcriptome in LUAD. The combined miRNA–mRNA regulatory network exhibited scale-free characteristics. Network analysis showed 5 miRNA (including hsa-miR-486-5p, hsa-miR-200b-5p, and hsa-miR-130b-5p) and 10 mRNA (including ASPM, CCNB1, TTN, TPX2, and BIRC5) which expressively contribute in the LUAD. We further investigated the hub genes and noticed that ASPM and TTN had the maximum rate of mutation and possessed a high tendency of cooccurrence in LUAD.

Conclusion:

This study provides a unique network approach to the exploration of the underlying molecular mechanism in LUAD. Identified mRNAs and miRNAs may therefore, serve as significant prognostic predictors and therapeutic targets.

Keywords: Lung adenocarcinoma, miRNA regulation, Protein-protein interaction, Network analysis

1. Introduction

Lung adenocarcinoma (LUAD), the most common primary lung cancer, mainly manifests in the form of non-small cell lung cancer (NSCLC) (Myers and Wallen 2020). As with other tumors, lung adenocarcinoma shows heterogeneity by high rates of genetic mutation (Dong et al. 2020). Despite the availability of diverse technologies, such as stereotactic radiotherapy, targeted therapy, minimal invasion methods and immunotherapy, the long-term survival is still poor in LUAD (Ferrara et al. 2018). One of the main reasons for this could be a delayed diagnosis. Thus, it is necessary to understand the molecular mechanisms behind lung adenocarcinoma genesis and metastasis, and identify effective biomarkers of this disease. MicroRNAs (miRNAs) are the 19-25 nucleotides long single-stranded non-coding RNAs, that regulate genes by mainly binding to the 3' untranslated regions (UTRs) of their target genes (Berindan-Neagoe et al. 2014). In recent, miRNAs were demonstrated to be significantly dysregulated in LUAD tissues and could be a potential target for this disease (Han and Li 2018).

In the post-genomic era, biological networks offer exciting possibilities in understanding gene regulation and pattern recognition for autonomous disease identification. Protein-protein interaction (PPI) network is conventionally a static network, where proteins are represented by nodes and interactions by edges (Barabasi and Oltvai 2004). However, in reality, PPI network has dynamic property, which is intrinsically regulated by complex cellular mechanisms through time and space (Liang and Li 2007). Recent studies showed the significance of the miRNA regulation on protein-protein interaction networks (Du et al. 2019; Lei et al. 2019; Liang and Li 2007). Moreover, different methods have been developed to predict miRNA targets accurately at the genomic level (John et al. 2004; Lewis et al. 2003). But much effort had not been given to understand the relationship between miRNA regulations on PPI in LUAD from a regulatory network approach.

Gene Expression Omnibus (GEO) is a public data repository of array- and sequence-based functional genomics data (Edgar et al. 2002). Owing to the advancement of high-throughput bioinformatics technologies, we aimed to re-analyze the mRNA (GSE116959) and miRNA (GSE74190) expression profile data sets in LUAD from GEO to unveil more specific molecular targets associated with the tumorigenesis of the disease. We sought to construct a regulatory network combining miRNA-mRNA and protein-protein interactions in LUAD. Our study presented an integrative bioinformatics data analysis that could be appointed for the improvement in the clinical consequences aided through endorsement of the potential biomarkers in LUAD.

2. Materials and Methods

2.1 Data

The transcriptomic data used in this publication were collected from the National Center of Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) (Edgar et al. 2002). This gene expression profile was obtained from the GSE116959 dataset (Moreno Leon et al. 2019). The platform used for this data is the GPL17077 Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381 (Probe Name version). The GSE74190 dataset contains microRNA expression profiles of LUAD patients. GPL19622 Agilent-019118 Human miRNA Microarray V1 G4470A [miRBase release 9.1 miRNA ID version] platform used for the miRNA profile data.

2.2 Screening of Dysregulated miRNA and mRNA

In our study, GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/), an R based interactive web tool, was used to identify differentially expressed miRNAs and mRNAs between tumor and adjacent normal samples separately within each dataset (Barrett et al. 2013). A log fold change (logFC) >2 was set as a cut off for differentially expressed miRNAs and mRNAs. Probes without a corresponding gene symbol were filtered. The data were processed through the paired samples t-test and corrected using the Benjamini – Hochberg method. The adjusted P-value <0.05 cut off was applied. The log2 normalized expression values of differentially expressed miRNAs and mRNAs were employed for the complete unsupervised clustering using online tool MORPHEUS (https://software.broadinstitute.org/morpheus/). Further, Principal component analysis (PCA) and volcano plot of the same were performed using the R based online tool ClustVis(https://biit.cs.ut.ee/clustvis/) and the MATLAB v.R2018a respectively (Metsalu and Vilo 2015).

2.3 Protein-Protein Interaction Network Construction

The Protein-protein interaction (PPI) network was constructed and analyzed by using the STRING: functional protein association network database (https://string-db.org). The dysregulated mRNA signature in LUAD was set as the input and the highest confidence (0.90) of active interaction was chosen. The isolated nodes were hidden and supervised clustering (K-means) was performed for recognizing three clusters in the provided data set.

2.4 Development of miRNA-mRNA Expression Network

Interactions between dysregulated miRNAs and mRNAs were anticipated using miRWalk 3.0 (http://mirwalk.umm.uni-heidelberg.de/) (Sticht et al. 2018), which compiled the prediction results of both TargetScan(Agarwal et al. 2015) and miRDB(Chen and Wang 2020), and a maximum score (= 1) was considered for the prediction analysis in miRWalk. The miRNA-mRNA interaction network was incorporated with the PPI network for a better understanding of the underlying mechanism in LUAD. The Cytoscape v.3.8.0 (https://cytoscape.org/) was used for visualizing the regulatory network between miRNA-mRNA**.** Cytohubba, a cytoscape plug-in, was employed to predict top 5 miRNAs and top 10 mRNAs based on their degree. A network with these 5 miRNAs and their common target mRNAs was constructed and analyzed. BoxPlotR (http://shiny.chemgrid.org/boxplotr/), an online tool was used to create violin plots for the degree centrality (DC), closeness centrality (CC), and betweenness centrality (BC) data of the nodes present in the developed interactive network.

2.5 Functional Enrichment Analysis

Functional enrichment analysis was done with the significantly dysregulated miRNAs and mRNAs in LUAD development using FUNRICH tool v.3.1.3 (http://www.funrich.org/) taking the human proteome as background from the Gene Ontology database (http://geneontology.org/) (Pathan et al. 2015; Pathan et al. 2017). Top 5 significant cellular components (CC), molecular functions (MF) and biological process (BP) for miRNAs and mRNAs were plotted.

2.6 Pathway Analysis

Pathway analysis was carried out for differentially expressed mRNA signatures in LUAD using the open access, manually curated, peer-reviewed pathway database Reactome (https://reactome.org/). A cut-off of p value < 0.05 was set to recognize the altered signaling pathways.

2.7 Screening and Overall Survival Analysis of the Hub Genes

The mutation rates and the tendency of cooccurrence of the top 10 hub genes were examined with the cBioPortal platform (http://www.cbioportal.org). Further, the overall survival plots for the hub miRNAs and mRNAs were plotted using the Kaplan–Meier plotter (https://kmplot.com/analysis/) (Gyorffy et al. 2013) to determine their prognostic characteristics.

3. Results

3.1 Genetic Alteration in miRNAs and mRNAs Expression in LUAD

Our analysis revealed 17 miRNAs and 429 mRNAs to be differentially expressed in LUAD tissues (**Supplementary Table S1 and S2**). Dysregulated 17 miRNAs from the GSE74190 dataset includes 6 over-expressed and 11 under-expressed miRNAs. Analysis of the GSE116959 dataset showed 122 up-regulated and 307 down-regulated mRNAs. Unsupervised clustering of dysregulated mRNAs and miRNAs exhibited a distinct difference between tumor and normal samples (**Figure. 1A-B**). To identify the expression levels of differentially expressed mRNAs and miRNAs the volcano plots were made (**Figure. 1C-D**). The principal component analysis was plotted by using first and second components with 69.5% variability in miRNA level and 59.4% variability in mRNA level, represents the percentage of variation in their respective levels (**Figure. 1E-F**). This results further substantiates the clustering pattern of the dysregulated mRNA and miRNA signatures. The particulars of the data dimension employed in this report are provided in **Table 1**.

GSE Accession Numbers	RNA type	Tumor Sample Count	Adjacent Normal Sample Count	Identified RNA
GSE116959	messenger RNA	57	11	33985
GSE74190	micro RNA	36	44	821

Table 1. Details of the data sets of lung adenocarcinoma downloaded from the NCBI GEO data portal.

3.2 Interaction of Dysregulated Proteins in LUAD

The 429 differentially expressed mRNAs were assigned for the PPI network formation. The network exposed two major clusters with GNG11 (Guanine nucleotide-binding protein subunit gamma) and CCNB2 (G2/mitotic-specific cyclin-B2).

GNG11 was identified to be a key hub protein that interacted with 15 other proteins including GRK5 (G protein-coupled receptor kinase 5), CALCRL (Calcitonin gene-related peptide type 1 receptor), VIPR1 (Vasoactive intestinal polypeptide receptor 1), NMUR1 (Neuromedin-U receptor 1), ADRB1 (Beta-1 adrenergic receptor), SSTR1 (Beta-1 adrenergic receptor), ADRB2 (Beta-2 adrenergic receptor), CX3CR1 (CX3C chemokine receptor 1), EDN1 (Endothelin-1), CHRM1 (Muscarinic acetylcholine receptor M1), NMU (Muscarinic acetylcholine receptor M1), CACNA2D2 (Voltage-dependent calcium channel subunit alpha-2/delta-2), CXCL13 (C-X-C motif chemokine 13), GPER1 (C-X-C motif chemokine 13), and RAMP3 (Receptor activity-modifying protein 3) (**Supplementary Figure S1A**).

CCNB2, another key hub protein interacted with KIF2C (Kinesin-like protein KIF2C), BIRC5 (Baculoviral IAP repeat-containing protein 5), CENPF (Centromere protein F), UBE2C (Ubiquitin-conjugating enzyme E2 C), TPX2 (Targeting protein for Xklp2), SPAG5 (Sperm-associated antigen 5), ASPM (Abnormal spindle-like microcephaly-associated protein), KIAA0101 (PCNA Clamp Associated Factor), CDC45 (Cell division control protein 45 homolog),CDT1 (DNA replication factor Cdt1),HJURP (Holliday junction recognitionprotein), CEP55 (Centrosomal protein of 55 kDa), CCNB1 (G2/mitotic-specific cyclin-B1), PKMYT1 (Membrane-associated tyrosine- and threonine-specific cdc2-inhibitory kinase), and CDCA5 (Sororin) (**Supplementary Figure S1B**). The details of the PPI network were provided in **Supplementary Table S3.**

3.3 miRNA-mRNA Interacting Network Analysis

The role of miRNA on PPI network in LUAD was investigated by constructing a miRNA-mRNA regulatory network and visualized using Cytoscape (**Figure. 2A**). Interaction analysis showed that 17 dysregulated miRNAs aimed 321 dysregulated mRNAs for up-regulating or down-regulating them. In details, 6 over-expressed miRNAs down-regulated 178 mRNAs, whereas 11 under-expressed miRNAs up-regulated 69 mRNAs. Eventually, we incorporate the PPI interactions which include 287 interactions. In the combined interaction network, total number of nodes is 363 with the 1346 connections (**Supplementary Table S4**).

Degree of a node in a network implies the number of edges connected with the given node. Moreover, the degree distribution can be defined as the probability distribution of degrees over the network. In this study, the degree of distribution of the nodes in the regulatory network was studied, and the power-law distribution was observed (**Figure. 2B**). It indicated that the network exhibited the scale-free characteristics alike typical biological networks. We further analyze the indicators of the network centrality such as DC (degree centrality), CC (closeness centrality), and, BC (betweenness centrality) which recognizes the important single nodes possessing significant activity in a complex network (**Supplementary Table S5**). Our result showed significant alteration in the DC, CC and BC among miRNAs and mRNAs (p value < 0.01 for DC, CC, and BC using Mann Whitney U test). Moreover, the miRNA nodes having higher centrality indicating that the miRNA nodes tended to be critical in this context (**Figure. 2C-E**). However, the scale-free distribution of the miRNA-mRNA network indicate the existance of functionally relevant nodes in the development of LUAD.

3.4 Functional Enrichment Analysis of the Dysregulated miRNAs and mRNAs

Gene ontology (GO) enrichment analyses for differentially expressed miRNAs and mRNAs were performed. We identified the top 5 most significant GO terms of each group. Differentially expressed miRNAs were enriched in signal transduction, regulation of nucleic acid metabolism, on the BP level. On CC level, they were enriched in nucleus, cytoplasm and lysosome. Transcription factor activity, GTPase activity, protein serine/ threonine kinase activity and phosphatase activity were significant on MF level for dysregulated miRNAs (**Figure. 3A**).

However, differentially expressed mRNAs were enriched in extracellular matrix (ECM) organization, caveola assembly, nitric oxide transport, on the BP level. Cell surface and extracellular regions were enriched on CC level for dysregulated mRNAs. On MF level, calcium ion binding, lipoprotein particle binding, heparin binding were noteworthy (**Figure. 3B**).

3.5 Extracellular Matrix Organisation Was the Most Enriched Pathway in LUAD

Considering the signature 429 dysregulated mRNAs, a pathway enrichment analysis was done by using the Reactome pathway analysis database. Top 10 most enhanced pathways were shown (**Figure. 4A**). Extracellular matrix organization pathway was one of the most enriched pathways in the LUAD ($p = 1.96 \times 10^{-5}$). Twenty-eight proteins were enriched in the ECM organization pathway. These proteins were further appointed for the network analysis (**Figure. 4B**). Collagen degradation ($p = 4.37 \times 10^{-4}$) was among the other key dysregulated pathways in LUAD. The dysregulated proteins involved in the ECM organization pathways were listed in **Supplementary Table S6.**

3.6 Hub Selection and Analysis

To identify the hub nodes in the network we obtained Cytohubba plug-in of Cytoscape. hsa-miR-486-5p, hsa-miR-200b-5p, hsa-miR-130b-5p, hsa-miR-183-5p, and hsa-miR-139-5p were the top 5 hub miRNAs with the degree 119, 110, 102, 97, and 85 respectively, in the network. On the other hand, ASPM (20), CCNB1 (19), TTN (Titin) (17), CENPF (16), GNG11 (15), CCNB2 (15), BIRC5 (15), SPAG5 (15), KIF2C (14) and TPX2 (14) were the top 10 hub genes (with their degree) in the network. The interaction between these hub miRNAs and their common targets were identified and represented (**Figure. 5**).

Oncopoint analysis of the hub genes with the cBioPortal showed that TTN (38%) and ASPM (14%) have the exceptionally high genetic mutation rates in LUAD (**Supplementary Figure S2**). We also analyzed the tendency for the cooccurrence of these hub genes (**Supplementary**

Table S7). The result showed TTN and ASPM also had a high tendency of cooccurrence in LUAD with a log2 odd ratio 1.368 and $p < 0.001$.

4. Discussion

Understanding the elementary molecular mechanisms of LUAD development to the identification of the therapeutic targets could considerably improve patient prognosis. Large expression profiling data sets have demonstrated to be beneficial in determining the regimen for cancer treatment (Deb et al. 2020; Dong et al. 2020; Klaeger et al. 2017; Zhang et al. 2020). In the post-genomic era, biological networks become a useful tool to understand gene regulation critically. A few studies have been performed to understand the role of miRNAs in LUAD (Jin et al. 2020; Wang et al. 2019). Since the PPI network can easily explain the functions of proteins involved in any disease, it is crucial to understand the effect of miRNA on PPI. In this present report, we have re-investigate and combined the miRNA and mRNA expression data from LUAD patients. We identified 17 differentially expressed miRNAs and 429 differentially expressed mRNAs between lung adenocarcinoma and adjacent normal tissue samples. Biological roles of the dysregulated miRNAs and mRNAs were studied using functional enrichment analyses. The differentially expressed miRNAs were mostly overrepresented in cell communication, signal transduction, regulation of gene expression processes. Earlier studies showed strong evidence suggesting a powerful role of miRNAs in the communication between tumor microenvironment and cancer cells via releasing exosomes, cytokines, growth factors, altering various signaling pathways (Eichelmann et al. 2018). Significantly altered mRNAs were mostly overrepresented in the extracellular matrix organization, caveola assembly. The ECM provides a biomechanical anchorage to the cells. Cancer metastasis is closely related to the detachment of abled cells from the ECM of the primary tumor site. Malignant cells traverse ECM barrier to gain the accessibility of circulation using an anchor independent manner and reattached to the secondary tumor site (Poltavets et al. 2018; Ranganathan et al. 2020). Lim et al. also reported previously the role of ECM organization as a prognostic and predictive indicator for early-stage NSCLC (Lim et al. 2017). Also, the higher CAV1 (Caveolin 1) expression, which leads to caveola assembly, induce filopodia formation, cell migration and metastasis in LUAD cells (Ho et al. 2002). In summary, our results are consistent with these previous reports.

We further analyzed the miRNA-mRNA regulatory network for identifying specific key genes and miRNAs related to LUAD. Among identified top 10 hub genes, ASPM had the highest degree at 20. ASPM gene, which involved in mitotic spindle regulation, is responsible for the

coordination of cell cycle. A recent study by Deb et al. showed alteration of phosphorylation pattern in S425 position in ASPM targeted by Serine/threonine-protein kinase Nek2 (NEK2), is a common signature pattern for five cancer types including LUAD (Deb et al. 2020).Wang et al. assessed ASPM expression and the functional significance in LUAD and reported that increasing level of expression of ASPM was positively correlated with poor disease prognosis (Wang et al. 2020). However, ASPM was also reported to be a potential biomarker candidate for bladder cancer(Saleh et al. 2020; Xu et al. 2019) breast cancer (Tang et al. 2019), and prostate cancer (Pai et al. 2019). CCNB1, another hub gene with second highest degree (19), an important mitosis initiator and is a key member of the cyclin family. This gene was previously reported as candidates for the development and prognosis of lung cancer (Arinaga et al. 2003; Soria et al. 2000; Wang et al. 2020), hepatocellular carcinoma (Huang et al. 2018; Long et al. 2019), breast cancer (Ding et al. 2014).In the present study, we found that ASPM and CCNB1 were significantly upregulated in LUAD samples compared to the paired adjacent normal samples. Furthermore, overall survival analysis showed that the LUAD patients having higher expression of these two genes (ASPM and CCNB1) related with the poor overall survival (**Supplementary Figure S3A**).

Moreover, we also identified top 5 miRNA hubs. Among them hsa-miR-200b-5p, hsa-miR-130b-5p and hsa-miR-183-5p were significantly over-expressed with degree 110, 102 and 97. In other hand, hsa-miR-486-5p and hsa-miR-139-5p were under-expressed with degree 119 and 85 respectively. hsa-miR-486-5p, having highest degree in the network, could serve as a potential biomarker for LUAD. Tian et al. validated the significant downregulation of hsa-miR-486-5p in NSCLC tissue, serum and cell samples (Tian et al. 2019). Our data showed hsa-miR-486-5p regulated both ASPM and CCNB1 during the alteration of cellular activities in LUAD patients. In addition, Rui et al. reported a significant involvement of hsa-miR-200b-5p expression in SPC-A1 LUAD cells (Rui et al. 2010). High expression of hsa-miR-130b-5p was associated with the poor overall survival and act as an oncogenic miRNA in NSCLC patients (Hirono et al. 2019). In our study, hsa-miR-130b-5p regulated hub gene SPAG5, CENPF and TTN. The miRNA, hsa-miR-183-5p is involved in cellular migration and invasion (Zaporozhchenko et al. 2016). In contrast to many other studies, we found hsa-miR-183-5p to be upregulated in LUAD patients. However, it plays a significant oncogenic role in other cancer types including breast (Macedo et al. 2017), gastric (Cao et al. 2014), and pancreatic cancer (Lu et al. 2016). Another downregulated miRNA, hsa-miR-139-5p was reported to be associated with occurrence along with development of NSCLC and may serve as a tumor

suppressor gene for LUAD(Yong-Hao et al. 2019). The overall survival plots of the hub miRNAs in LUAD were plotted using KM plotter (**Supplementary Figure S3B**).

We further screened the mutation rates of these 10 LUAD hub genes with cBioPortal platform. TTN (38%), ASPM (14%), CENPF (9%), TPX2 (4%) were the four genes with the highest mutation rates, which implies that these genes may play a significant role in tumorigenesis. TTN which is a key component in the vertebrate striated muscles assembly and functioning, has correlation with lung cancer upon missense mutation (Cheng et al. 2019).We observed that TTN was regulated by all five hub miRNAs, worth mentioning. A cooccurrence study among genes showed ASPM had significantly higher cooccurrence log odd ratios with CENPF, TTN, SPAG5, TPX2, and KIF2C. Hub gene KIF2C was reported as a regulator of cell signaling pathway, significantly contribute in LUAD progression (Bai et al. 2019).Survival analysis showed that a higher expression of TPX2, KIF2C is significantly related to the poor overall survival in LUAD patients(Ma et al. 2019) (**Supplementary Figure S3A**). CENPF, another critical hub gene in our network, is associated with the centromere-kinetochore complex and plays a vital role in tumor development in lung (Li et al. 2020), breast (O'Brien et al. 2007; Sun et al. 2019), prostate (Shahid et al. 2018) and nasopharynx (Cao et al. 2010). BIRC5, which is overexpressed in our data, performed a role in apoptosis and cell proliferation (Zhang et al. 2020).CCNB2 which was essential for the regulation of the cell cycle at the G2/M (mitosis) transition, played a significant role in LUAD development (Chen et al. 2020; Zhang et al. 2020). A current study by He et al. reported SPAG5 as “As emerging oncogene” for its overexpression in cancer types (He et al. 2020). SPAG5 was also reported to be upregulated in most of the LUAD cell lines (Kim et al. 2019), which further confirmed our analysis. Nevertheless, these predictions ought to be validated exhaustively through the clinical experiments and accomplished execution in LUAD therapy and treatment.

5. Conclusion

From the above discussion it can be concluded that, it is a network-based comprehensive re-investigation of the transcriptomic data-sets has been obtained from NCBI GEO. Our study aimed to unveil the role of miRNAs in PPI network of LUAD. miRNA-mRNA regulatory network has identified 5 hub miRNAs and 10 hub mRNAs which may serve as potential biomarkers and can be combined for the treatment of LUAD. However, provided the signaling complexity in cells and tumor heterogeneity, these discoveries require further investigation and validation in larger cohorts of patients.

Supplementary Materials:

Figure S1: Protein–protein interaction network showing the top 2 protein clusters with (**A**) GNG11 and (**B**) CCNB2 as hub, involved in the LUAD with the highest confidence (0.90) acquired using the online tool STRING; **Figure S2:** Mutational rate of the hub genes in lung adenocarcinoma were screened using cBioPortal platform; **Figure S3:**Kaplan-Meier plotsshowing overall survival according to high and low expression of mentioned mRNAs (**A**) and miRNAs (**B**) in LUAD;

Table S1: Dysregulated (logFC>2) miRNA signature in LUAD; **Table S2:** Dysregulated (logFC>2) mRNA signature in LUAD; **Table S3:** Details of protein-protein interaction network in LUAD; **Table S4:** Interactions of miRNA-mRNA regulatory network in LUAD; **Table S5:** Node details of miRNA-mRNA regulatory network in LUAD; **Table S6:** Proteins enriched in the extracellular matrix organization pathway; **Table S7:** Tendency of cooccurrence of the hub genes in LUAD;

Author Contributions:

Conceptualization, P.S., and M.G.; Methodology, P.S., S.S., and M.M; Software, P.S.; Formal analysis, P.S., M.M. and M.G.; Investigation, P.S., and M.G.; Illustration, P.S.; Resources, M.G.; Data curation, P.S., S.S., and M.M.; Writing—original draft preparation, P.S., and S.S.; Writing—review and editing, M.G.; Visualization, P.S., S.S., M.M., and M.G.; Supervision, M.G.; Project administration, M.G.; All authors have read and agreed to the published version of the manuscript.

Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments:

We thank the Ministry of Human Resource Development, Government of India and Department of Biotechnology, National Institute of Technology Durgapur for the research support.

Conflicts of Interest:

The authors declare no conflict of interest.

Code availability:

Not Applicable

Figure Legends:

Figure 1: Differential expression pattern analysis of miRNA and mRNA obtained from lung adenocarcinoma tumor and normal samples.A-B. Unsupervised clustering of dysregulated mRNAs (A) and miRNAs (B) along with their expression pattern in LUAD is illustrated in the heatmap developed using MORPHEUS (https://software.broadinstitute.org/morpheus/). **C-D**. Relative expression of differentially expressed 465 mRNAs (C) and 17 miRNAs (D) can be observed in the Volcano Plot. These plots were made by using MATLAB v.R2018a. Green, Red and Black dots on the plot display upregulated, downregulated and dysregulated expression respectively. **E-F**. Principal component analysis (PCA) of the dysregulated mRNAs (E) and miRNAs (F) was performed by ClustVis (https://biit.cs.ut.ee/clustvis/) showing their corresponding variations in the clusters of tumor and normal samples.

Figure 2: miRNA-mRNA regulatory network construction and characterization. A. This regulatory network analysis was visualized in Cytoscape v.3.8.0 (https://cytoscape.org/) showing that 17 dysregulated miRNAs targeting 321 dysregulated mRNAs. Yellow nodes in the regulatory network represent the mRNAs and blue nodes represent the miRNAs.**B**. Degree distribution of the regulatory network. **C**. The difference of degree centrality between miRNAs and mRNAs. The miRNA nodes showed an aberrantly higher degree centrality than mRNA nodes. **D**. The difference of closeness centrality between miRNAs and mRNAs. The miRNA nodes showed a significantly higher closeness centrality than mRNA nodes. **E**. The difference of betweenness centrality between miRNAs and mRNAs. The miRNA nodes had a significantly higher betweenness centrality than mRNAs. P-values were calculated based on Mann – Whitney U test.BoxPlotR (http://shiny.chemgrid.org/boxplotr/) was used to create violin plots for the degree centrality (DC), closeness centrality (CC), and betweenness centrality (BC) data of the nodes present in the developed interactive network.

Figure 3: Functional enrichment analysis of the dysregulated miRNA and mRNA signatures.A-B. Top five enriched gene ontology terms for each of the biological process, cellular components and molecular functions were shown for miRNA (A) and mRNA (B) signatures using a double y-axis plot. Y-axis in left indicates the percentage of identified dysregulated genes involve in that particular process (represented using blue bars) and y-axis in right indicates the -log10(P- values) (marked using the orange line).

Figure 4: Altered pathways and interaction clusters enriched in LUAD. A. Bar graph of showing top ten enriched pathways in lung adenocarcinoma identified using the online pathway analysis tool Reactome (https://reactome.org/). **B**. Protein–protein interaction network comprising the protein clusters associated with the ECM organization pathway with the highest confidence (0.90) obtained using the STRING database (https://string-db.org).

Figure 5: **Schematic diagram of interactions between hub miRNA and their common targets.** This diagram was visualized in Cytoscape v.3.8.0 (https://cytoscape.org/) and was modified by Adobe Illustrator (Adobe Illustrator CC 2018, Adobe Inc., retrieved from https://adobe.com/products/illustrator).The dysregulated top 5 miRNAs with maximum degree and their common mRNA targets were depicted by a schematic diagram. Their respective expression pattern was also highlighted.

References:

Agarwal V, Bell GW, Nam JW, Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs Elife 4 doi:10.7554/eLife.05005

Arinaga M, Noguchi T, Takeno S, Chujo M, Miura T, Kimura Y, Uchida Y (2003) Clinical implication of cyclin B1 in non-small cell lung cancer Oncol Rep 10:1381-1386

Bai Y, Xiong L, Zhu M, Yang Z, Zhao J, Tang H (2019) Co-expression network analysis identified KIF2C in association with progression and prognosis in lung adenocarcinoma Cancer Biomark 24:371-382 doi:10.3233/CBM-181512

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization Nat Rev Genet 5:101-113 doi:10.1038/nrg1272

Barrett T et al. (2013) NCBI GEO: archive for functional genomics data sets--update Nucleic Acids Res 41:D991-995 doi:10.1093/nar/gks1193

Berindan-Neagoe I, Monroig Pdel C, Pasculli B, Calin GA (2014) MicroRNAome genome: a treasure for cancer diagnosis and therapy CA Cancer J Clin 64:311-336 doi:10.3322/caac.21244

Cao JY et al. (2010) Prognostic significance and therapeutic implications of centromere protein F expression in human nasopharyngeal carcinoma Mol Cancer 9:237 doi:10.1186/1476-4598-9-237

Cao LL et al. (2014) miR-183 inhibits invasion of gastric cancer by targeting Ezrin Int J Clin Exp Pathol 7:5582-5594

Chen C, Guo Q, Song Y, Xu G, Liu L (2020) SKA1/2/3 serves as a biomarker for poor prognosis in human lung adenocarcinoma Transl Lung Cancer Res 9:218-231 doi:10.21037/tlcr.2020.01.20

Chen Y, Wang X (2020) miRDB: an online database for prediction of functional microRNA targets Nucleic Acids Res 48:D127-D131 doi:10.1093/nar/gkz757

Cheng X et al. (2019) Aggregate analysis based on TCGA: TTN missense mutation correlates with favorable prognosis in lung squamous cell carcinoma J Cancer Res Clin Oncol 145:1027-1035 doi:10.1007/s00432-019-02861-y

Deb B, Sengupta P, Sambath J, Kumar P (2020) Bioinformatics Analysis of Global Proteomic and Phosphoproteomic Data Sets Revealed Activation of NEK2 and AURKA in Cancers Biomolecules 10 doi:10.3390/biom10020237

Ding K, Li W, Zou Z, Zou X, Wang C (2014) CCNB1 is a prognostic biomarker for ER+ breast cancer Med Hypotheses 83:359-364 doi:10.1016/j.mehy.2014.06.013

Dong S, Men W, Yang S, Xu S (2020) Identification of lung adenocarcinoma biomarkers based on bioinformatic analysis and human samples Oncol Rep 43:1437-1450 doi:10.3892/or.2020.7526

Du X, Li Q, Cao Q, Wang S, Liu H, Li Q (2019) Integrated Analysis of miRNA-mRNA Interaction Network in Porcine Granulosa Cells Undergoing Oxidative Stress Oxid Med Cell Longev 2019:1041583 doi:10.1155/2019/1041583

Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository Nucleic Acids Res 30:207-210 doi:10.1093/nar/30.1.207

Eichelmann AK, Matuszcak C, Hummel R, Haier J (2018) Role of miRNAs in cell signaling of cancer associated fibroblasts Int J Biochem Cell Biol 101:94-102 doi:10.1016/j.biocel.2018.05.015

Ferrara R et al. (2018) Hyperprogressive Disease in Patients With Advanced Non-Small Cell Lung Cancer Treated With PD-1/PD-L1 Inhibitors or With Single-Agent Chemotherapy JAMA Oncol 4:1543-1552 doi:10.1001/jamaoncol.2018.3676

Gyorffy B, Surowiak P, Budczies J, Lanczky A (2013) Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer PLoS One 8:e82241 doi:10.1371/journal.pone.0082241

Han Y, Li H (2018) miRNAs as biomarkers and for the early detection of non-small cell lung cancer (NSCLC) J Thorac Dis 10:3119-3131 doi:10.21037/jtd.2018.05.32

He J, Green AR, Li Y, Chan SYT, Liu DX (2020) SPAG5: An Emerging Oncogene Trends Cancer doi:10.1016/j.trecan.2020.03.006

Hirono T et al. (2019) MicroRNA-130b functions as an oncomiRNA in non-small cell lung cancer by targeting tissue inhibitor of metalloproteinase-2 Sci Rep 9:6956 doi:10.1038/s41598-019-43355-8

Ho CC, Huang PH, Huang HY, Chen YH, Yang PC, Hsu SM (2002) Up-regulated caveolin-1 accentuates the metastasis capability of lung adenocarcinoma by inducing filopodia formation Am J Pathol 161:1647-1656 doi:10.1016/S0002-9440(10)64442-2

Huang Q, Li J, Wei AJTCR (2018) Identification of potential therapeutic targets in hepatocellular carcinoma using an integrated bioinformatics approach 2018 7:849-858

Jin X, Guan Y, Zhang Z, Wang H (2020) Microarray data analysis on gene and miRNA expression to identify biomarkers in non-small cell lung cancer BMC Cancer 20:329 doi:10.1186/s12885-020-06829-x

John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS (2004) Human MicroRNA targets PLoS Biol 2:e363 doi:10.1371/journal.pbio.0020363

Kim J, Hyun SW, Lee IS, Jo K, Kim YS, Kim JS, Kim CS (2019) Aster koraiensis extract lowers postprandial glucose in normoglycemic and high-fat-diet-induced obese mice Food Sci Biotechnol 28:563-568 doi:10.1007/s10068-018-0497-8

Klaeger S et al. (2017) The target landscape of clinical kinase drugs Science 358 doi:10.1126/science.aan4368

Lei H, Liu W, Si J, Wang J, Zhang T (2019) Analyzing the regulation of miRNAs on protein-protein interaction network in Hodgkin lymphoma BMC Bioinformatics 20:449 doi:10.1186/s12859-019-3041-9

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets Cell 115:787-798 doi:10.1016/s0092-8674(03)01018-3

Li R, Wang X, Zhao X, Zhang X, Chen H, Ma Y, Liu Y (2020) Centromere protein F and Forkhead box M1 correlation with prognosis of non-small cell lung cancer Oncol Lett 19:1368-1374 doi:10.3892/ol.2019.11232

Liang H, Li WH (2007) MicroRNA regulation of human protein protein interaction network RNA 13:1402-1408 doi:10.1261/rna.634607

Lim SB, Tan SJ, Lim WT, Lim CT (2017) An extracellular matrix-related prognostic and predictive indicator for early-stage non-small cell lung cancer Nat Commun 8:1734 doi:10.1038/s41467-017-01430-6

Long J et al. (2019) Construction and comprehensive analysis of a ceRNA network to reveal potential prognostic biomarkers for hepatocellular carcinoma Cancer Cell Int 19:90 doi:10.1186/s12935-019-0817-y

Lu YY, Zeng Y, Zheng JY, Huang CL, Wang XP (2016) [MiR-183 Regulates Proliferation of SW1990 Pancreatic Cancer Cell Line by Targeting PDCD4] Sichuan Da Xue Xue Bao Yi Xue Ban 47:691-696

Ma W et al. (2019) Prognostic significance of TOP2A in non-small cell lung cancer revealed by bioinformatic analysis Cancer Cell Int 19:239 doi:10.1186/s12935-019-0956-1

Macedo T, Silva-Oliveira RJ, Silva VAO, Vidal DO, Evangelista AF, Marques MMC (2017) Overexpression of mir-183 and mir-494 promotes proliferation and migration in human breast cancer cell lines Oncol Lett 14:1054-1060 doi:10.3892/ol.2017.6265

Metsalu T, Vilo J (2015) ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap Nucleic Acids Res 43:W566-570 doi:10.1093/nar/gkv468

Moreno Leon L et al. (2019) The nuclear hypoxia-regulated NLUCAT1 long non-coding RNA contributes to an aggressive phenotype in lung adenocarcinoma through regulation of oxidative stress Oncogene 38:7146-7165 doi:10.1038/s41388-019-0935-y

Myers DJ, Wallen JM (2020) Cancer, Lung Adenocarcinoma. In: StatPearls. Treasure Island (FL),

O'Brien SL et al. (2007) CENP-F expression is associated with poor prognosis and chromosomal instability in patients with primary breast cancer Int J Cancer 120:1434-1443 doi:10.1002/ijc.22413

Pai VC et al. (2019) ASPM promotes prostate cancer stemness and progression by augmenting Wnt-Dvl-3-beta-catenin signaling Oncogene 38:1340-1353 doi:10.1038/s41388-018-0497-4

Pathan M et al. (2015) FunRich: An open access standalone functional enrichment and interaction network analysis tool Proteomics 15:2597-2601 doi:10.1002/pmic.201400515

Pathan M et al. (2017) A novel community driven software for functional enrichment analysis of extracellular vesicles data J Extracell Vesicles 6:1321455 doi:10.1080/20013078.2017.1321455

Poltavets V, Kochetkova M, Pitson SM, Samuel MS (2018) The Role of the Extracellular Matrix and Its Molecular and Cellular Regulators in Cancer Cell Plasticity Front Oncol 8:431 doi:10.3389/fonc.2018.00431

Ranganathan S, Kumar S, Mohanty SS, Jolly MK, Rangarajan A (2020) Cellular Plasticity in Matrix-attached and -Detached Cells: Implications in Metastasis Journal of the Indian Institute of Science doi:10.1007/s41745-020-00179-0

Rui W, Bing F, Hai-Zhu S, Wei D, Long-Bang C (2010) Identification of microRNA profiles in docetaxel-resistant human non-small cell lung carcinoma cells (SPC-A1) J Cell Mol Med 14:206-214 doi:10.1111/j.1582-4934.2009.00964.x

Saleh AA, Gohar SF, Hemida AS, Elgharbawy M, Soliman SE (2020) Evaluation of ASPM and TEF Gene Expressions as Potential Biomarkers for Bladder Cancer Biochem Genet 58:490-507 doi:10.1007/s10528-020-09962-1

Shahid M et al. (2018) Centromere protein F (CENPF), a microtubule binding protein, modulates cancer metabolism by regulating pyruvate kinase M2 phosphorylation signaling Cell Cycle 17:2802-2818 doi:10.1080/15384101.2018.1557496

Soria JC, Jang SJ, Khuri FR, Hassan K, Liu D, Hong WK, Mao L (2000) Overexpression of cyclin B1 in early-stage non-small cell lung cancer and its clinical implication Cancer Res 60:4000-4004

Sticht C, De La Torre C, Parveen A, Gretz N (2018) miRWalk: An online resource for prediction of microRNA binding sites PLoS One 13:e0206239 doi:10.1371/journal.pone.0206239

Sun J et al. (2019) Overexpression of CENPF correlates with poor prognosis and tumor bone metastasis in breast cancer Cancer Cell Int 19:264 doi:10.1186/s12935-019-0986-8

Tang J et al. (2019) Overexpression of ASPM, CDC20, and TTK Confer a Poorer Prognosis in Breast Cancer Identified by Gene Co-expression Network Analysis Front Oncol 9:310 doi:10.3389/fonc.2019.00310

Tian F et al. (2019) MiR-486-5p Serves as a Good Biomarker in Nonsmall Cell Lung Cancer and Suppresses Cell Growth With the Involvement of a Target PIK3R1 Front Genet 10:688 doi:10.3389/fgene.2019.00688

Wang J, He Z, Duan R (2020) [Expression of ASPM in Lung Adenocarcinoma and Its Relationship with Development and Prognosis] Zhongguo Fei Ai Za Zhi 23:29-35 doi:10.3779/j.issn.1009-3419.2020.01.05

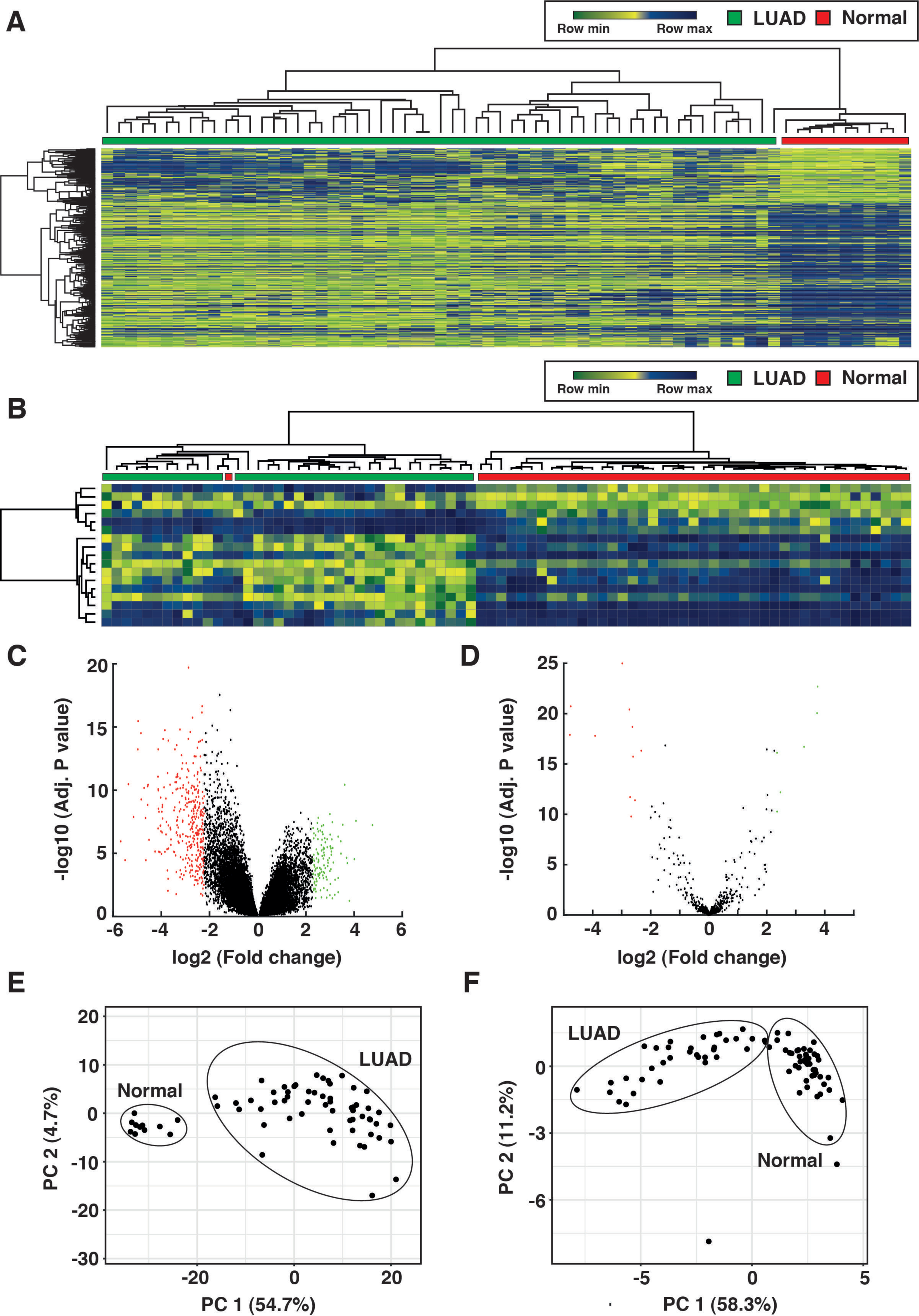
Wang SS, Fang YY, Huang JC, Liang YY, Guo YN, Pan LJ, Chen G (2019) Clinical value of microRNA-198-5p downregulation in lung adenocarcinoma and its potential pathways Oncol Lett 18:2939-2954 doi:10.3892/ol.2019.10610

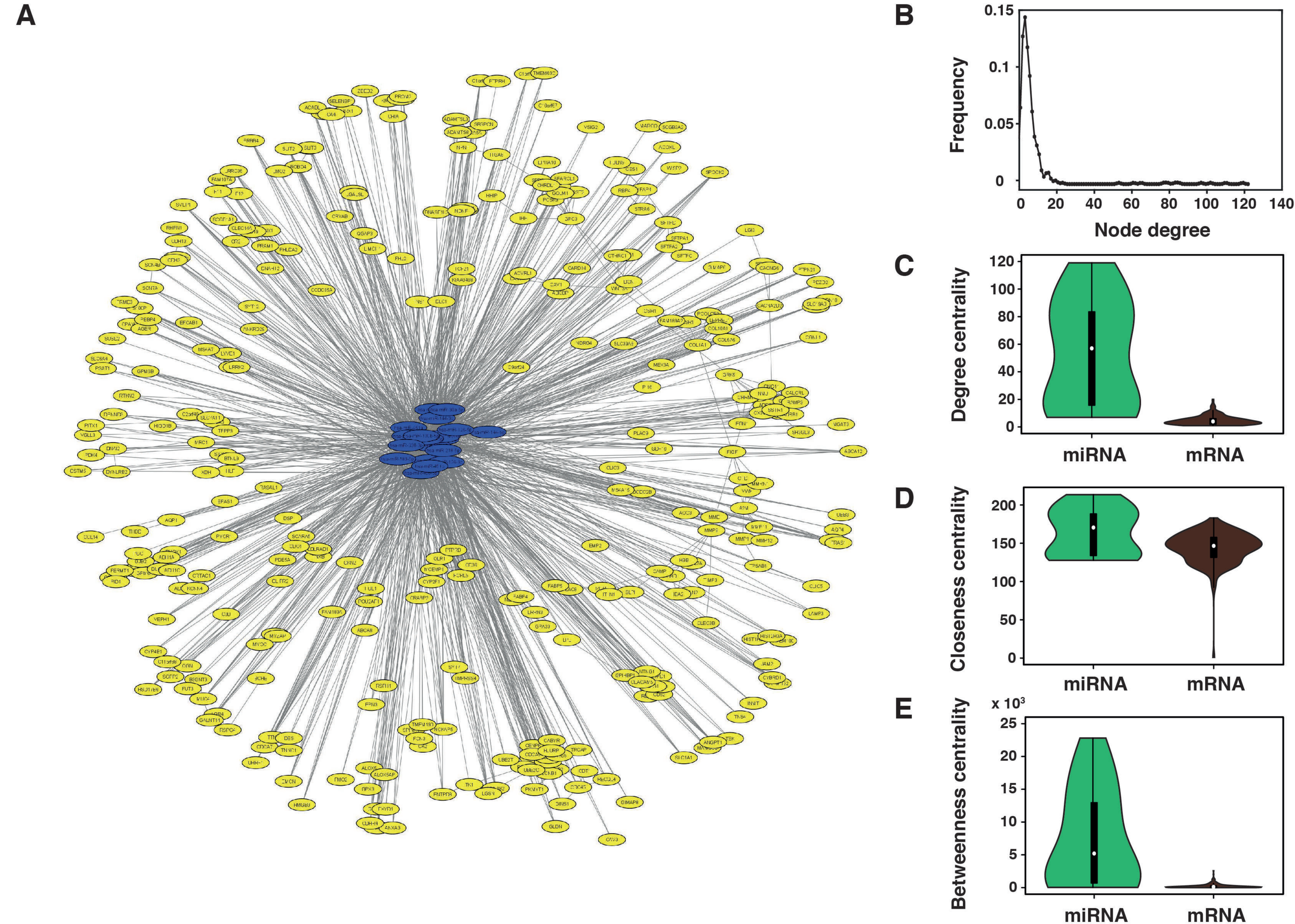
Xu Z, Zhang Q, Luh F, Jin B, Liu X (2019) Overexpression of the ASPM gene is associated with aggressiveness and poor outcome in bladder cancer Oncol Lett 17:1865-1876 doi:10.3892/ol.2018.9762

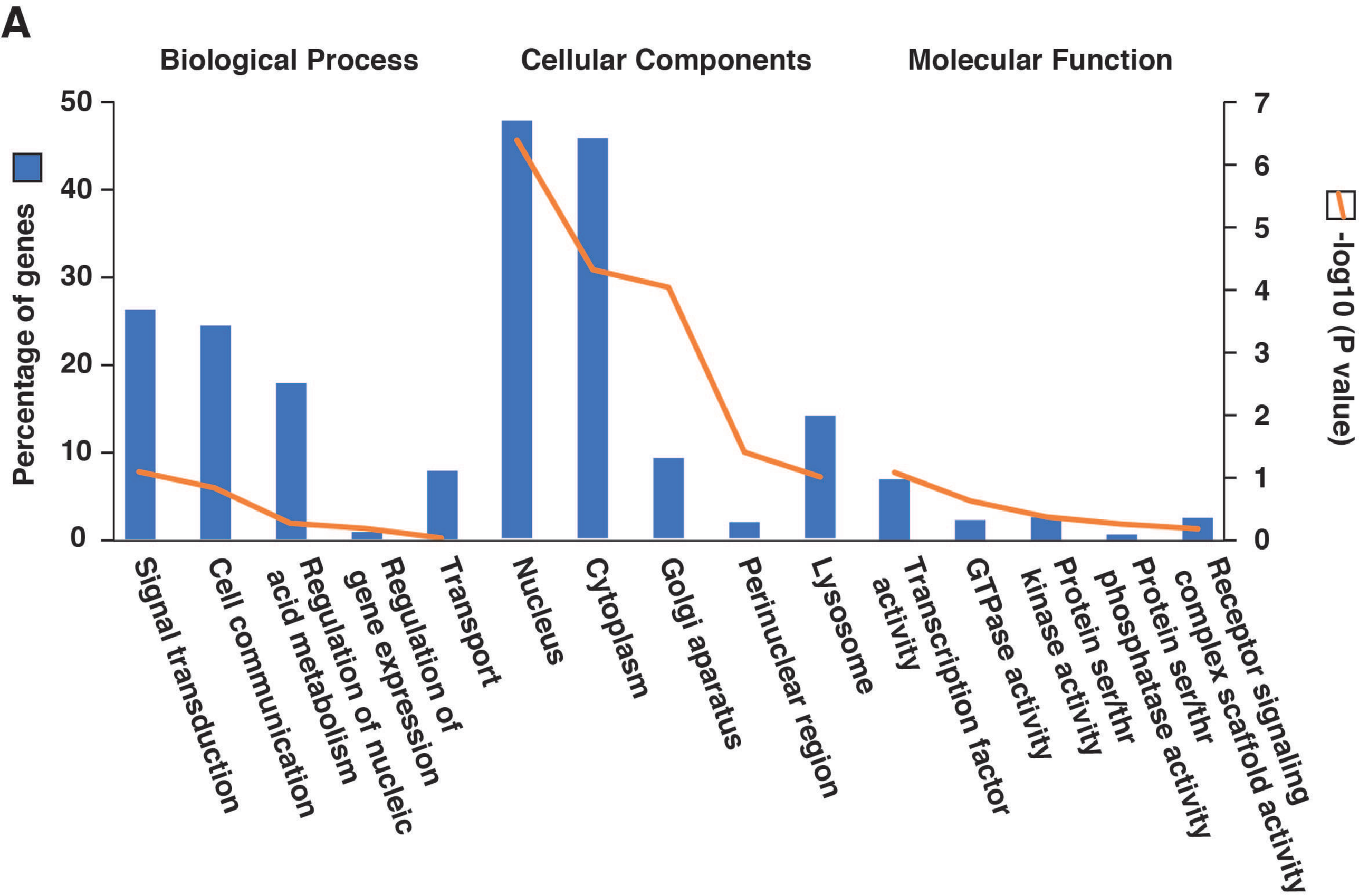
Yong-Hao Y, Xian-Guo W, Ming X, Jin-Ping Z (2019) Expression and clinical significance of miR-139-5p in non-small cell lung cancer J Int Med Res 47:867-874 doi:10.1177/0300060518815379

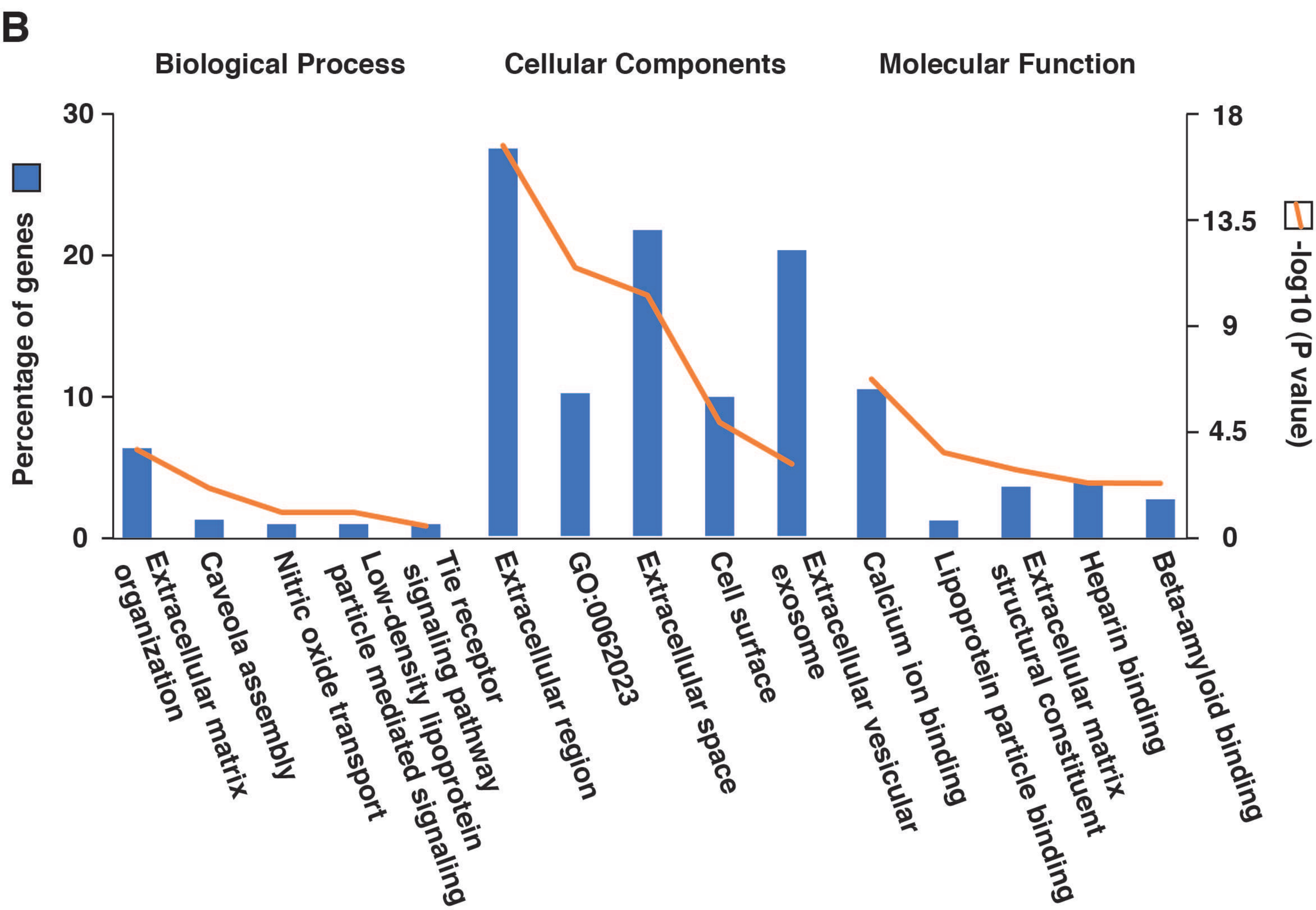
Zaporozhchenko IA et al. (2016) Plasma miR-19b and miR-183 as Potential Biomarkers of Lung Cancer PLoS One 11:e0165261 doi:10.1371/journal.pone.0165261

Zhang J, Li D, Zhang Y, Ding Z, Zheng Y, Chen S, Wan Y (2020) Integrative analysis of mRNA and miRNA expression profiles reveals seven potential diagnostic biomarkers for nonsmall cell lung cancer Oncol Rep 43:99-112 doi:10.3892/or.2019.7407

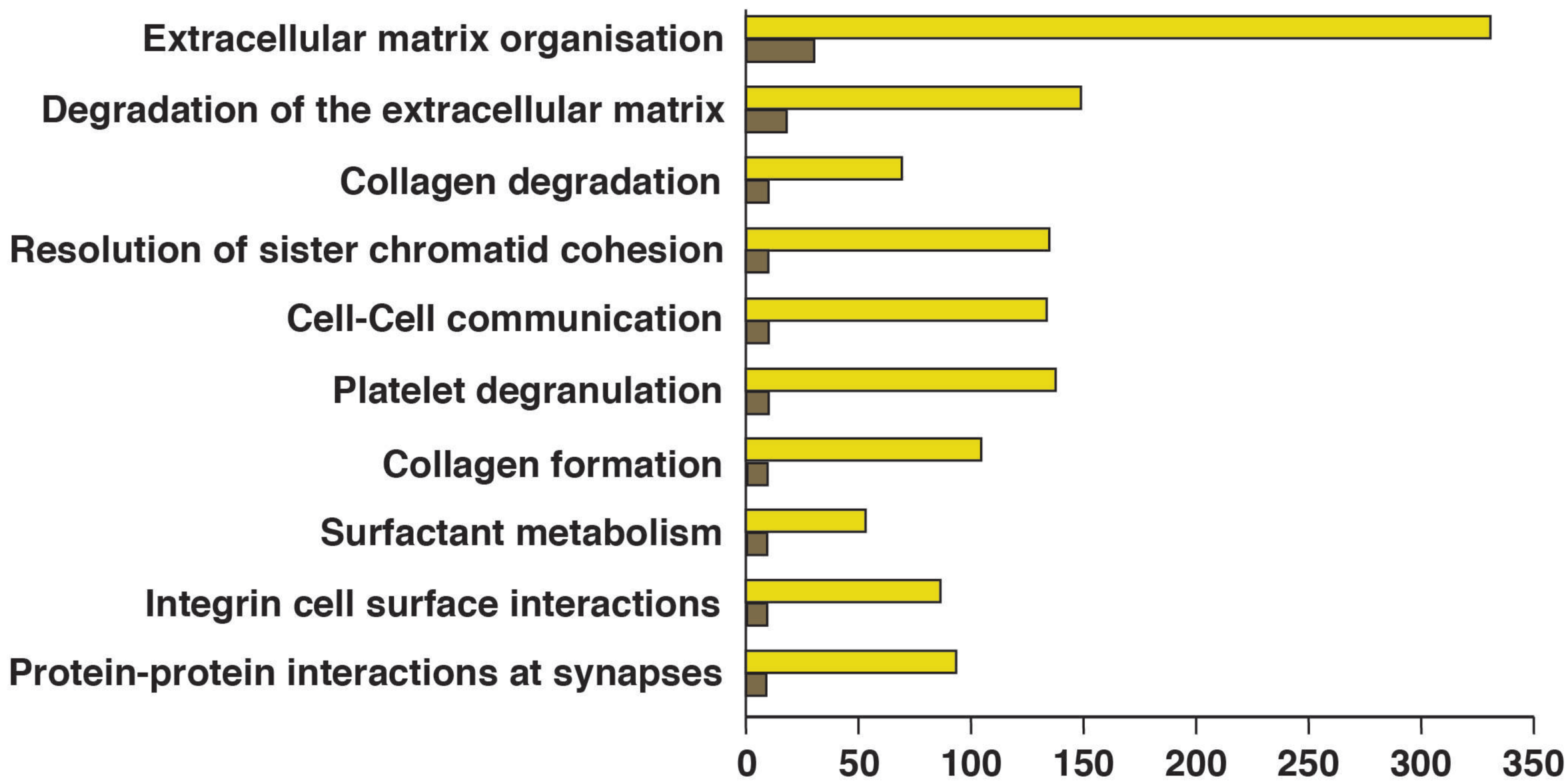

A
Row min
Row max
LUAD
Normal
B
Row min
Row max
LUAD
Normal
C
-log10 (Adj. P value)
log2 (Fold change)
D
-log10 (Adj. P value)
log2 (Fold change)
E
Normal
LUAD
PC 2 (4.7%)
PC 1 (54.7%)
F
LUAD
Normal
PC 2 (11.2%)
PC 1 (58.3%)


A
B
Frequency
Node degree
C
Degree centrality
miRNA
mRNA
D
Closeness centrality
miRNA
mRNA
E
Betweenness centrality
x 10^3
miRNA
mRNA

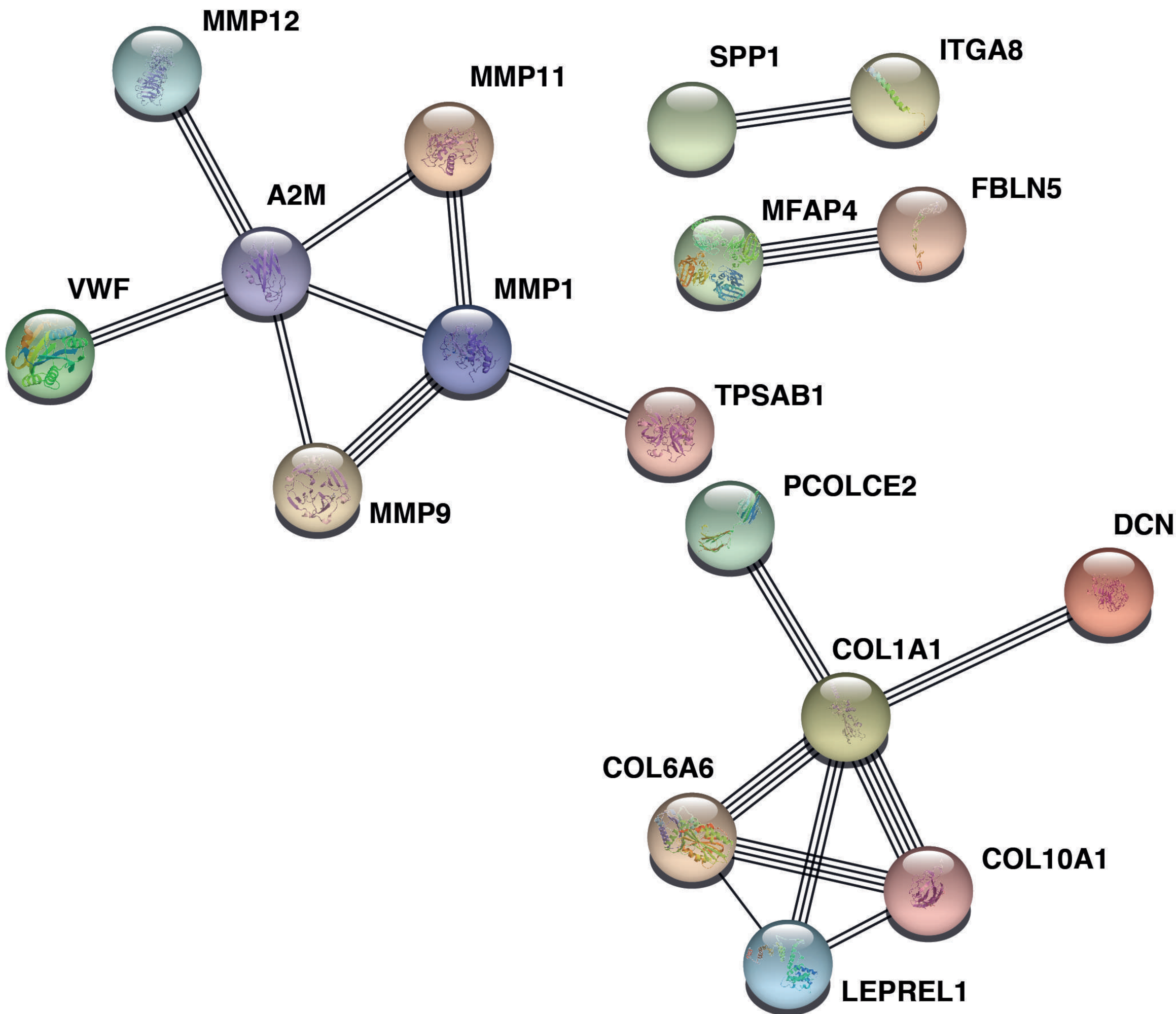

A
Biological Process
Cellular Components
Molecular Function
Percentage of genes
-log10 (P value)
50
40
30
20
10
0
7
6
5
4
3
2
1
0
Signal transduction
Cell communication
Regulation of acid metabolism
Regulation of gene expression
Transport
Nucleus
Cytoplasm
Golgi aparatus
Perinuclear region
Lysosome
Transcription factor activity
GTPase activity
kinase activity
Protein ser/thr phosphatase activity
Protein ser/thr
Receptor signaling complex scaffold activity

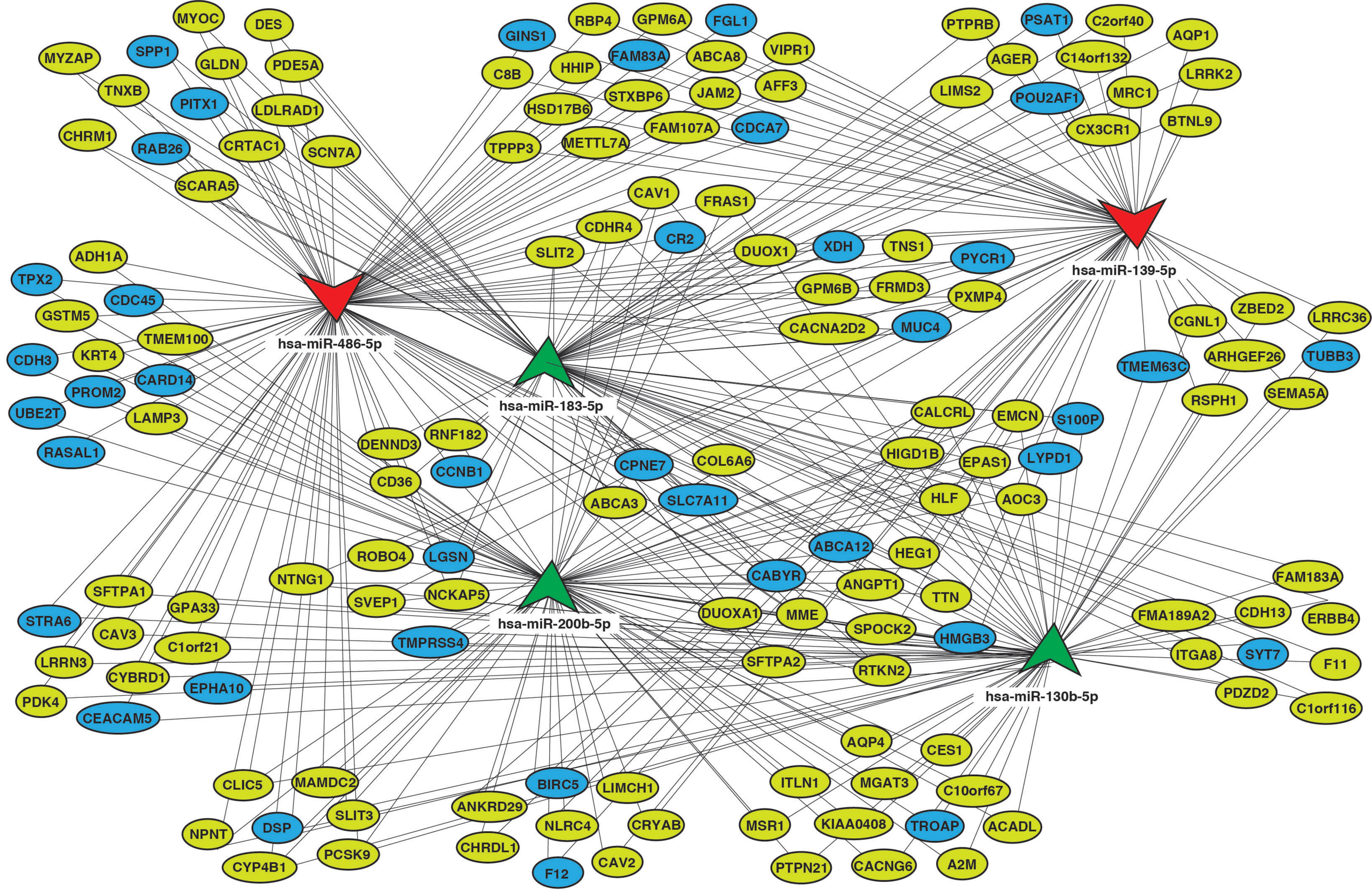

B
Biological Process
Cellular Components
Molecular Function
Percentage of genes
-log10 (P value)
30
20
10
0
18
13.5
9
4.5
0
Extracellular matrix organization
Caveola assembly
Nitric oxide transport
Low-density lipoprotein particle mediated signaling
Tie receptor signaling pathway
Extracellular region
GO:0062023
Extracellular space
Cell surface
Extracellular vesicular exosome
Calcium ion binding
Lipoprotein particle binding
Extracellular matrix structural constituent
Heparin binding
Beta-amyloid binding

A


Entities found
Entities total
Extracellular matrix organisation
Degradation of the extracellular matrix
Collagen degradation
Resolution of sister chromatid cohesion
Cell-Cell communication
Platelet degranulation
Collagen formation
Surfactant metabolism
Integrin cell surface interactions
Protein-protein interactions at synapses
0
50
100
150
200
250
300
350

B


MMP12
MMP11
SPP1
ITGA8
A2M
FBLN5
MFAP4
VWF
MMP1
TPSAB1
PCOLCE2
MMP9
DCN
COL1A1
COL6A6
COL10A1
LEPREL1


hsa-miR-486-5p
hsa-miR-183-5p
hsa-miR-139-5p
hsa-miR-200b-5p
hsa-miR-130b-5p
MYOC
DES
MYZAP
SPP1
GLDN
PDE5A
TNXB
PITX1
LDLRAD1
CHRM1
RAB26
CRTAC1
SCN7A
SCARA5
GINS1
RBP4
GPM6A
FGL1
FAM83A
ABCA8
VIPR1
C8B
HHIP
STXBP6
JAM2
AFF3
HSD17B6
FAM107A
CDCA7
TPPP3
METTL7A
PTPRB
PSAT1
C2orf40
AQP1
AGER
C14orf132
LRRK2
LIMS2
POU2AF1
MRC1
CX3CR1
BTNL9
CAV1
FRAS1
CDHR4
CR2
SLIT2
DUOX1
XDH
TNS1
PYCR1
GPM6B
FRMD3
PXMP4
CACNA2D2
MUC4
ADH1A
TPX2
CDC45
GSTM5
TMEM100
CDH3
KRT4
CARD14
PROM2
UBE2T
LAMP3
RASAL1
DENND3
RNF182
CD36
CCNB1
CPNE7
COL6A6
ABCA3
SLC7A11
CALCRL
EMCN
S100P
HIGD1B
EPAS1
LYPD1
HLF
AOC3
ZBED2
LRRC36
CGNL1
ARHGEF26
TUBB3
TMEM63C
RSPH1
SEMA5A
ROBO4
LGSN
NTNG1
SVEP1
NCKAP5
TMPRSS4
SFTPA1
GPA33
STRA6
CAV3
C1orf21
LRRN3
CYBRD1
EPHA10
PDK4
CEACAM5
ABCA12
HEG1
CABYR
ANGPT1
TTN
DUOXA1
MME
SPOCK2
HMGB3
SFTPA2
RTKN2
FAM183A
FMA189A2
CDH13
ERBB4
ITGA8
SYT7
F11
PDZD2
C1orf116
AQP4
CES1
ITLN1
MGAT3
C10orf67
MSR1
KIAA0408
TROAP
ACADL
PTPN21
CACNG6
A2M
CLIC5
MAMDC2
SLIT3
NPNT
DSP
PCSK9
CYP4B1
BIRC5
LIMCH1
ANKRD29
NLRC4
CRYAB
CHRDL1
F12
CAV2
Over-expressed miRNA
Under-expressed miRNA
Up-regulated mRNA
Down-regulated mRNA