

# Central Nodes in Protein Interaction Networks Drive Critical Functions in Transforming Growth Factor Beta-1 Stimulated Kidney Cells

Reyhaneh Rabeian, M.Sc.<sup>1#</sup>, Maryam Abedi, M.Sc.<sup>1#</sup>, Yousof Gheisari, M.D., Ph.D.<sup>1, 2\*</sup>

1. Department of Genetics and Molecular Biology, Isfahan University of Medical Sciences, Isfahan, Iran  
2. Regenerative Medicine Lab, Isfahan Kidney Diseases Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

\*Corresponding Address: P.O.Box: 8174673461, Department of Genetics and Molecular Biology, Isfahan University of Medical Sciences, Isfahan, Iran  
Email: ygheisari@med.mui.aci.ir

#The first two authors equally contributed to this manuscript.

Received: 27/Sep/2015, Accepted: 17/Mar/2016

## Abstract

**Objective:** Despite the huge efforts, chronic kidney disease (CKD) remains as an unsolved problem in medicine. Many studies have shown a central role for transforming growth factor beta-1 (TGF $\beta$ -1) and its downstream signaling cascades in the pathogenesis of CKD. In this study, we have reanalyzed a microarray dataset to recognize critical signaling pathways controlled by TGF $\beta$ -1.

**Materials and Methods:** This study is a bioinformatics reanalysis for a microarray data. The GSE23338 dataset was downloaded from the gene expression omnibus (GEO) database which assesses the mRNA expression profile of TGF $\beta$ -1 treated human kidney cells after 24 and 48 hours incubation. The protein interaction networks for differentially expressed (DE) genes in both time points were constructed and enriched. In addition, by network topology analysis, genes with high centrality were identified and then pathway enrichment analysis was performed with either the total network genes or with the central nodes.

**Results:** We found 110 and 170 genes differentially expressed in the time points 24 and 48 hours, respectively. As the genes in each time point had few interactions, the networks were enriched by adding previously known genes interacting with the differentially expressed ones. In terms of degree, betweenness, and closeness centrality parameters 62 and 60 nodes were considered to be central in the enriched networks of 24 hours and 48 hours treatment, respectively. Pathway enrichment analysis with the central nodes was more informative than those with all network nodes or even initial DE genes, revealing key signaling pathways.

**Conclusion:** We introduced a method for the analysis of microarray data that integrates the expression pattern of genes with their topological properties in protein interaction networks. This holistic novel approach allows extracting knowledge from raw bulk *omics* data.

**Keywords:** Chronic Kidney Disease, Microarray Analysis, Protein Interaction Maps, Systems Biology, Transforming Growth Factor Beta-1

Cell Journal (Yakhteh), Vol 18, No 4, Jan-Mar (Winter) 2017, Pages: 514-531

**Citation:** Rabeian R, Abedi M, Gheisari Y. Central nodes in protein interaction networks drive critical functions in transforming growth factor beta-1 stimulated kidney cells. Cell J. 2017; 18(4): 514-531.

## Introduction

Chronic kidney disease (CKD) is a public health problem and a leading cause of death. Despite using current therapies to slow progression of CKD, respective patients are still reaching the end stage renal disease (ESRD) at alarming proportions (1). The histological feature of this debilitating disorder

is excessive deposition of extra-cellular matrix (ECM) defined as renal fibrosis. Recent studies declared that transforming growth factor beta-1 (TGF $\beta$ -1) is the major driver of fibrosis in kidney, stimulating a variety of signaling pathways related to deposition of ECM components (2). In spite of enormous researches on the role of TGF $\beta$ -1 and

downstream elements in the progression of CKD (3, 4), few studies have employed holistic and computational methods for investigation of kidney disorders. Among these studies, there is an elegant report presented by Jin et al. (5) who employed gene regulatory network concepts to analyze high-throughput gene expression data. They could predict and experimentally validate HIPK2 as a potential drug target in HIV-associated nephropathy.

In this study, we propose a holistic approach to investigate the molecular interactions and signaling pathways in response to TGF $\beta$ -1 stimulation in human kidney cells. A microarray dataset has been generated by Walsh et al. (6) that examines the expression profile of human tubular epithelial cells before and after treatment with TGF $\beta$ -1 for 24 and 48 hours. However, they only focused on the few top differentially expressed (DE) genes including *GREM1*, *JAG1* and *HES1*. They identified Notch signaling as a critical pathway in diabetic nephropathy. In the current study, we introduced a new method for the analysis of the same microarray dataset that integrated the expression pattern of genes with their topological location in the gene interaction network. Using this strategy, we could infer more informative signaling pathways related to TGF $\beta$ -1 stimulation. This approach could also be employed for other large data to improve our understanding of biological processes by extracting remarkable concepts from bulk *omics* data.

## Materials and Methods

### Microarray data

This study is a bioinformatics analysis of GSE23338 dataset, originally generated by Walsh et al. (6). mRNA expression profile was downloaded from the Gene Expression Omnibus (GEO) database (7). In this microarray experiment, transcriptional response of human proximal tubule epithelial cells (HK-2) to TGF $\beta$ -1 stimulation after 24 and 48 hours was assessed. Using GEO2R tool of GEO, the TGF $\beta$ -1 treated cells (24 or 48 hours) were compared to untreated HK-2 cells. Benjamini-Hochberg false discovery rate method was applied for P value

adjustment. Genes with adjusted  $P \leq 0.05$  were considered as differentially expressed.

### Protein-protein interaction network

Using CluePedia plugin (8) of the Cytoscape software version 3.1.0 (9), a protein-protein interaction (PPI) network was constructed for the DE genes in time point of 24 hours or 48 hours. Topology of networks was analyzed by the NetworkAnalyzer tool of Cytoscape software.

### Pathway enrichment analysis

Pathway enrichment analysis for DE genes was carried out using ClueGO plugin (10) of Cytoscape. In this analysis, KEGG and Reactome databases were chosen for retrieving data and network specificity was adjusted to medium. Bonferroni step down was applied for P value adjustment and pathways with adjusted  $P \leq 0.05$  were chosen.

## Results

In this study, we reanalyzed the GSE23338 microarray dataset assessing mRNA expression profile of HK-2 cells after 24 and 48 hours of treatment with TGF $\beta$ -1. Analysis by GEO2R revealed that 110 genes after 24 hours and 170 genes after 48 hours were differentially expressed with adjusted  $P \leq 0.05$  (Table 1). To investigate the interaction between variably expressed genes, a network was constructed for each time point. Although different kind of interactions (activation, post-translational modification, expression and binding) were allowed to be shown, unexpectedly, few interactions were appeared in both networks (Fig.1A, B). To infer pathways related to the DE genes and understand the down-stream processes controlled by TGF $\beta$ -1, pathway enrichment analysis was performed, showing only 12 pathways for 24 hours (Fig.1C) and 10 pathways for 48 hours treatments (Fig.1D), with few connections between the signaling pathways.

The scarcity of interactions in PPI and pathway networks was not unexpected, as they were derived from mRNA microarray data which can only detect genes with altered mRNA level, thus regulated genes at other levels were

missed. Hence, to predict other role players, we enriched both PPI networks by adding one interacting node for each gene. This resulted in expansion from 110 to 199 nodes for 24 hours (Fig.2A) and from 170 to 301 nodes for 48 hours treatment (Fig.2B). PPI networks were reconstructed with the same parameters applied initially. To determine the most central genes in these enriched networks, their topology was assessed by graph theory measures such as degree, betweenness centrality, and closeness centrality. In each network, the genes were sorted based on each of these features. Then, the top 20% genes in 24 hours treatment and 15% genes with higher rank in 48 hours were chosen. Because of overlapping nodes between the above three centrality parameters, a total of 62 genes in time point of 24 hours (Table 2) and 60 genes in time point of 48 hours (Table 3) were finally selected. Again, pathway enrichment

analysis was performed with either the central genes or the total genes in these two enriched networks. The central genes in time points 24 and 48 hours networks were related to 29 and 49 pathways, respectively (Fig.3). These pathways were strongly related to CKD and formed a deeply connected network in both time points. Interestingly, pathway enrichment analysis with the total enriched networks genes, only determined 16 and 18 pathways for time points of 24 and 48 hours, respectively. These pathways were less inter-connected compared to those derived from the central genes (Fig.4).

Pathway enrichment analysis with the central genes predicted Notch, TNF, P53, Activin and TGF $\beta$  signaling as well as platelet-related pathways, affected after TGF $\beta$ -1 treatment in both 24 and 48 hours. However, Hippo, PDGF and FGFR signaling pathways were enriched only in the second time point.

**Table 1:** Differentially expressed genes in time 24 hours and 48 hours with adjusted  $P \leq 0.05$ . The genes are sorted by log2 of fold change (LogFC)

Genes	Time 24		Genes	Time 48	
	adj.P.Val	logFC		adj.P.Val	logFC
<i>GDF15</i>	0.012817	-4.03492	<i>GDF15</i>	0.004294	-3.77276
<i>CRYM</i>	0.046546	-3.35307	<i>CRYM</i>	0.020195	-3.74094
<i>SCNN1A</i>	0.012817	-3.19552	<i>CD9</i>	0.000557	-3.3273
<i>CD9</i>	0.003455	-2.96886	<i>SCNN1A</i>	0.006484	-2.86473
<i>RBM47</i>	0.012817	-2.96538	<i>RBM47</i>	0.010066	-2.73215
<i>MAL</i>	0.012817	-2.6579	<i>MAL</i>	0.007941	-2.72193
<i>HLF</i>	0.033274	-2.44538	<i>AREG</i>	0.014332	-2.71598
<i>DEPTOR</i>	0.011983	-2.38064	<i>HLF</i>	0.021497	-2.52256
<i>IMPA2</i>	0.002857	-2.22728	<i>PLA1A</i>	0.007423	-2.46499
<i>RTEL1</i>	0.003588	-2.11992	<i>PDZK1IP1</i>	0.026161	-2.45799
<i>MEGF9</i>	0.03429	-2.04315	<i>DUSP5</i>	0.005251	-2.37922
<i>GSE1</i>	0.011894	-2.04015	<i>ACSL1</i>	0.003583	-2.36964
<i>ELOVL6</i>	0.004534	-2.02884	<i>DEPTOR</i>	0.014818	-2.23285

Table 1: Continued

Genes	Time 24		Genes	Time 48	
	adj.P.Val	logFC		adj.P.Val	logFC
<i>BIRC3</i>	0.012817	-1.98537	<i>DEFB1</i>	0.001178	-2.1258
<i>SLC17A3</i>	0.006063	-1.96502	<i>IMPA2</i>	0.001964	-2.11942
<i>SULT1C2</i>	0.045879	-1.93073	<i>HLA-DMB</i>	0.036113	-2.11004
<i>DUSP6</i>	0.018789	-1.93001	<i>FXSD2</i>	0.002471	-2.09686
<i>CEBPD</i>	0.015951	-1.89181	<i>RTEL1</i>	0.003148	-1.99502
<i>DEFB1</i>	0.003455	-1.87388	<i>CLDN1</i>	0.002102	-1.9428
<i>ACSL1</i>	0.003455	-1.84878	<i>BIRC3</i>	0.008587	-1.93307
<i>PLA1A</i>	0.030906	-1.79724	<i>SULT1C2</i>	0.028474	-1.89456
<i>DUSP5</i>	0.011894	-1.78577	<i>FAS</i>	0.040775	-1.84699
<i>CA12</i>	0.011983	-1.70822	<i>CEBPD</i>	0.014469	-1.81201
<i>CLDN1</i>	0.006732	-1.69617	<i>SLC17A3</i>	0.010066	-1.78837
<i>PDZK1IP1</i>	0.031449	-1.66723	<i>LY6E</i>	0.003332	-1.70064
<i>ADAMTS3</i>	0.009793	-1.64873	<i>SERPINA1</i>	0.021497	-1.68148
<i>CDKN2AIP</i>	0.047829	-1.62696	<i>SLCO4A1</i>	0.03808	-1.67053
<i>GULP1</i>	0.049153	-1.55674	<i>SOD2</i>	0.003686	-1.65771
<i>ACVR1B</i>	0.019538	-1.47953	<i>TSPAN1</i>	0.011747	-1.65484
<i>ID2</i>	0.018571	-1.45204	<i>PLIN2</i>	0.026161	-1.62099
<i>EPAS1</i>	0.049153	-1.42294	<i>MEGF9</i>	0.024224	-1.61932
<i>SOD2</i>	0.016073	-1.41158	<i>RAB20</i>	0.026161	-1.59433
<i>ANXA4</i>	0.047613	-1.37096	<i>CLU</i>	0.002471	-1.54936
<i>RAB20</i>	0.015265	-1.34593	<i>SLC4A4</i>	0.03487	-1.50061

Table 1: Continued

Genes	Time 24		Genes	Time 48	
	adj.P.Val	logFC		adj.P.Val	logFC
<i>MMD</i>	0.030004	-1.33753	<i>GULP1</i>	0.047026	-1.46306
<i>CLU</i>	0.01997	-1.32415	<i>EPAS1</i>	0.038561	-1.42677
<i>BDNF</i>	0.018571	-1.26903	<i>ACVR1B</i>	0.013621	-1.3911
<i>EPCAM</i>	0.015265	-1.26628	<i>GPRC5C</i>	0.026161	-1.34555
<i>NR2F2</i>	0.044918	-1.26334	<i>GSE1</i>	0.041643	-1.32532
<i>TMEM159</i>	0.047829	-1.25784	<i>LRRC61</i>	0.020785	-1.32277
<i>FAS</i>	0.019538	-1.23999	<i>ANXA4</i>	0.038789	-1.31199
<i>LY6E</i>	0.014942	-1.20673	<i>CDKN2AIP</i>	0.03949	-1.30584
<i>LRRC61</i>	0.033972	-1.17462	<i>MMD</i>	0.021485	-1.29784
<i>PPP2R5A</i>	0.023781	-1.16917	<i>PPP2R5A</i>	0.019989	-1.25554
<i>SERPINA1</i>	0.039821	-1.09323	<i>NR2F2</i>	0.012081	-1.22902
<i>IL24</i>	0.011983	-1.09102	<i>GLRX</i>	0.035692	-1.22902
<i>HGD</i>	0.019538	-1.08015	<i>SERPINA6</i>	0.00653	-1.22661
<i>ELF3</i>	0.026977	-1.07437	<i>EMP1</i>	0.030041	-1.22491
<i>GCH1</i>	0.032261	-1.0672	<i>MAPKAPK3</i>	0.037211	-1.20559
<i>ALDH5A1</i>	0.030004	-1.05748	<i>IFI30</i>	0.039032	-1.1775
<i>FXD2</i>	0.020961	-1.02587	<i>EPCAM</i>	0.014332	-1.17347
<i>TRIM38</i>	0.043165	-0.92721	<i>SYS1-DBNDD2</i>	0.039499	-1.16256
<i>NHLRC2</i>	0.018571	-0.92091	<i>ADAMTS3</i>	0.014586	-1.12871
<i>TBL1X</i>	0.040887	-0.88595	<i>SHMT1</i>	0.036579	-1.12397
<i>LAD1</i>	0.04193	-0.87726	<i>GGT2</i>	0.007492	-1.10696
<i>GLRX</i>	0.035216	-0.87251	<i>LAD1</i>	0.014332	-1.09515
<i>TPM1</i>	0.030916	0.782848	<i>FOSL1</i>	0.023626	-1.08872
<i>AMIGO2</i>	0.032261	0.803279	<i>ELF3</i>	0.022045	-1.078
<i>MISP</i>	0.030916	0.808838	<i>ID2</i>	0.03219	-1.07757
<i>ACLY</i>	0.030778	0.809032	<i>SMAD3</i>	0.042933	-1.05481
<i>FN1</i>	0.03429	0.860918	<i>IL24</i>	0.030041	-1.03178
<i>LYPD1</i>	0.046955	0.922314	<i>SH2B2</i>	0.020195	-1.00971
<i>RALA</i>	0.030004	0.95394	<i>DUSP6</i>	0.038561	-0.98235
<i>EFNB2</i>	0.030004	0.9589	<i>ITPR3</i>	0.021485	-0.9804
<i>SMURF2</i>	0.044772	1.000129	<i>PDLIM1</i>	0.044481	-0.96321
<i>TFPI2</i>	0.019538	1.042945	<i>ALDH5A1</i>	0.019989	-0.95377
<i>MARCH3</i>	0.026013	1.048251	<i>FAM3C</i>	0.039499	-0.93464
<i>NREP</i>	0.031449	1.121914	<i>REPIN1</i>	0.038561	-0.9095

Table 1: Continued

Genes	Time 24		Genes	Time 48	
	adj.P.Val	logFC		adj.P.Val	logFC
<i>LTBP2</i>	0.015265	1.133197	<i>GGT1</i>	0.036579	-0.893
<i>PLEK2</i>	0.025143	1.137329	<i>ANXA1</i>	0.03141	-0.8635
<i>RFTN1</i>	0.014768	1.141252	<i>UXS1</i>	0.037211	-0.78881
<i>PRPS1</i>	0.021243	1.212761	<i>HGD</i>	0.039499	-0.77866
<i>ADA</i>	0.012817	1.214286	<i>TBL1X</i>	0.028029	-0.76181
<i>TNS1</i>	0.027064	1.276677	<i>MGLL</i>	0.039499	-0.75719
<i>COL1A1</i>	0.044918	1.349036	<i>GNPDA1</i>	0.028029	-0.75096
<i>LAMC2</i>	0.015265	1.448205	<i>PAX8</i>	0.031546	-0.73263
<i>CREB3L1</i>	0.004425	1.453935	<i>TRIM38</i>	0.026161	-0.69388
<i>TSPAN13</i>	0.030916	1.468138	<i>PROSC</i>	0.047627	-0.68991
<i>F3</i>	0.049854	1.537792	<i>TPM1</i>	0.045542	0.614444
<i>AKAP12</i>	0.030004	1.541307	<i>ARL4C</i>	0.038561	0.67729
<i>HES1</i>	0.015265	1.549119	<i>IFNGR2</i>	0.045542	0.695846
<i>SGK1</i>	0.006063	1.584326	<i>RFTN1</i>	0.037889	0.727815
<i>PAX6</i>	0.014768	1.602106	<i>ACLY</i>	0.021485	0.74412
<i>GREM1</i>	0.004818	1.607941	<i>EFNB2</i>	0.026486	0.789092
<i>PTHLH</i>	0.018571	1.651867	<i>CLTCL1</i>	0.043748	0.805174
<i>SLN</i>	0.030916	1.66995	<i>SMURF2</i>	0.019989	0.813175
<i>ADAM19</i>	0.046955	1.673182	<i>FAM208B</i>	0.038561	0.815648
<i>TUFT1</i>	0.01997	1.708363	<i>TPM4</i>	0.036579	0.816674
<i>PPP1R13L</i>	0.044622	1.715701	<i>PLEK2</i>	0.040607	0.838742
<i>VEGFC</i>	0.006732	1.731189	<i>FHOD3</i>	0.043748	0.840283
<i>GPR56</i>	0.005222	1.757315	<i>CADM1</i>	0.014818	0.842324
<i>LRP4</i>	0.006732	1.839036	<i>DLC1</i>	0.035692	0.861077
<i>SIK1</i>	0.028431	1.847404	<i>ELK3</i>	0.037211	0.866603
<i>C1orf106</i>	0.014768	1.852771	<i>AMIGO2</i>	0.013633	0.891177
<i>KCNK3</i>	0.019891	1.928548	<i>PGRMC2</i>	0.038561	0.892116
<i>WNT5B</i>	0.015265	1.950651	<i>RAB32</i>	0.039499	0.911187
<i>SNAI2</i>	0.021356	1.996987	<i>UAP1</i>	0.02966	0.914231
<i>GALNT10</i>	0.022735	2.016561	<i>SKIL</i>	0.037889	0.927445
<i>GADD45B</i>	0.005222	2.081882	<i>MAGED2</i>	0.047466	0.933606
<i>FSTL3</i>	0.006871	2.18737	<i>DYRK2</i>	0.045542	0.941228
<i>WNT5A</i>	0.015265	2.199978	<i>PALLD</i>	0.039499	0.960395
<i>SCG5</i>	0.006063	2.421762	<i>MKL1</i>	0.012081	0.986708

Table 1: Continued

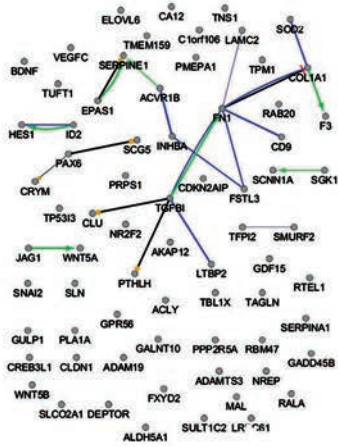
Genes	Time 24		Genes	Time 48	
	adj.P.Val	logFC		adj.P.Val	logFC
<i>TGFBI</i>	0.010068	2.585222	<i>MARCH3</i>	0.039989	1.008954
<i>TP53I3</i>	0.018571	2.591672	<i>LTBP2</i>	0.007423	1.013795
<i>IL11</i>	0.006063	2.680544	<i>GABARAPL1</i>	0.026161	1.018263
<i>PMEPA1</i>	0.002821	2.69133	<i>TFPI2</i>	0.045542	1.023787
<i>TAGLN</i>	0.015265	2.807473	<i>NOV</i>	0.03219	1.037359
<i>SLCO2A1</i>	0.002821	2.969782	<i>NUAK1</i>	0.010962	1.041704
<i>INHBA</i>	0.006732	3.742935	<i>SLC22A4</i>	0.021701	1.057375
<i>JAG1</i>	0.012993	4.819474	<i>PDLIM7</i>	0.036579	1.075928
			<i>SEMA3C</i>	0.040214	1.084533
			<i>PRPS1</i>	0.018933	1.090259
			<i>COL4A1</i>	0.014469	1.103866
			<i>NREP</i>	0.013884	1.110733
			<i>LYPD1</i>	0.028474	1.112816
			<i>TCF4</i>	0.044016	1.140686
			<i>GADD45B</i>	0.047627	1.201497
			<i>INPP4B</i>	0.003583	1.212552
			<i>SGK1</i>	0.010594	1.225169
			<i>IL15</i>	0.036579	1.22672
			<i>MAP3K4</i>	0.028944	1.263727
			<i>TUFT1</i>	0.037211	1.284833
			<i>SPARC</i>	0.019989	1.288601
			<i>COL7A1</i>	0.00653	1.297757
			<i>ADAM12</i>	0.008731	1.356895
			<i>CREB3L1</i>	0.003148	1.386727
			<i>PTHLH</i>	0.013101	1.415775
			<i>ADAM19</i>	0.026161	1.427201
			<i>IGF1R</i>	0.047026	1.47119
			<i>ARHGEF40</i>	0.01087	1.471459
			<i>WNT5B</i>	0.037889	1.474394
			<i>C1orf106</i>	0.018696	1.482021
			<i>FSTL3</i>	0.010621	1.530293
			<i>LRP4</i>	0.019989	1.533742
			<i>NEDD9</i>	0.040607	1.541275
			<i>HES1</i>	0.019989	1.573046

Table 1: Continued

Genes	Time 24		Genes	Time 48	
	adj.P.Val	logFC		adj.P.Val	logFC
			<i>SPOCK1</i>	0.014586	1.577949
			<i>TSPAN13</i>	0.014818	1.599124
			<i>SPHK1</i>	0.024113	1.599544
			<i>THBS1</i>	0.047872	1.633499
			<i>BCAT1</i>	0.003332	1.666823
			<i>AKAP12</i>	0.010066	1.677861
			<i>SLN</i>	0.015745	1.68979
			<i>DSP</i>	0.048348	1.726407
			<i>FN1</i>	0.004908	1.832317
			<i>SCG5</i>	0.000815	1.864837
			<i>GPR56</i>	0.010962	1.900793
			<i>GALNT10</i>	0.028612	1.917703
			<i>PAX6</i>	0.005251	1.918114
			<i>GREM1</i>	0.001644	1.934697
			<i>SIK1</i>	0.012081	1.972459
			<i>TP53I3</i>	0.035787	1.979721
			<i>VEGFC</i>	0.010621	1.991006
			<i>EFEMP1</i>	0.007516	2.118009
			<i>SLC26A2</i>	0.026357	2.161277
			<i>FBN1</i>	0.019989	2.339046
			<i>WNT5A</i>	0.001354	2.385963
			<i>MMP13</i>	0.024732	2.392553
			<i>TAGLN</i>	0.010066	2.43914
			<i>SNAI2</i>	0.002471	2.45019
			<i>PMEPA1</i>	0.000815	2.475717
			<i>TNS1</i>	0.012081	2.514322
			<i>TGFBI</i>	0.002471	2.622322
			<i>IL11</i>	0.001964	2.698275
			<i>SLCO2A1</i>	0.001145	2.774921
			<i>SLC7A11</i>	0.00226	3.076526
			<i>MMP1</i>	0.021973	3.220728
			<i>SERPINE1</i>	0.000815	3.452134
			<i>INHBA</i>	0.000815	3.757617
			<i>JAG1</i>	0.007516	4.928316

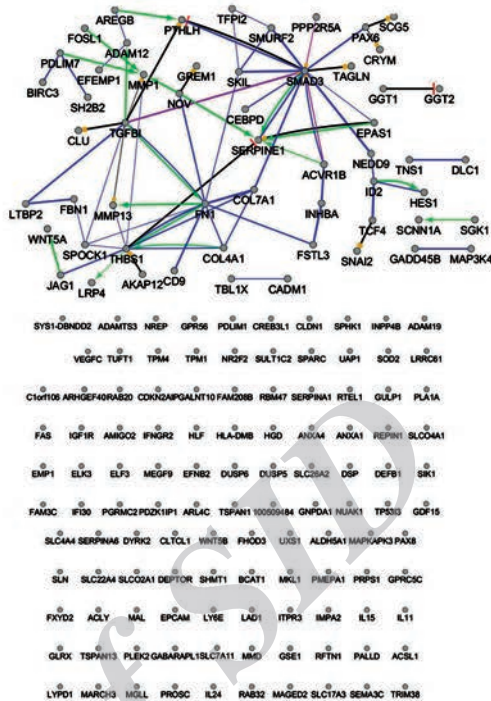


A

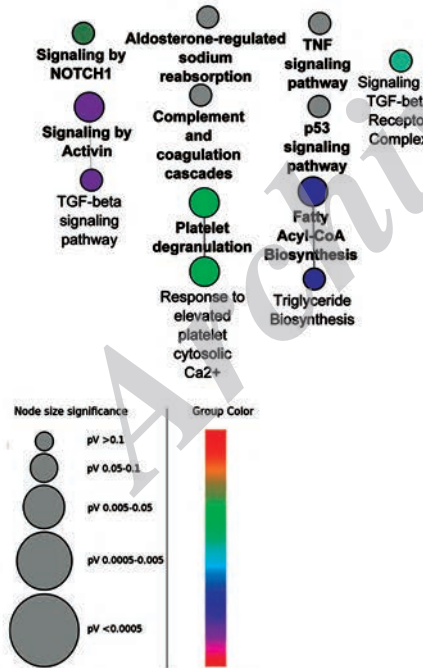


■ Activation  
■ Binding  
■ Expression  
■ Post-Translational Modification

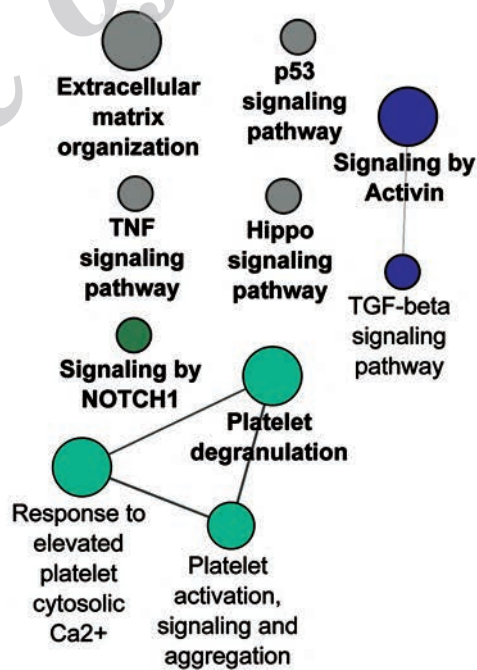
B



C

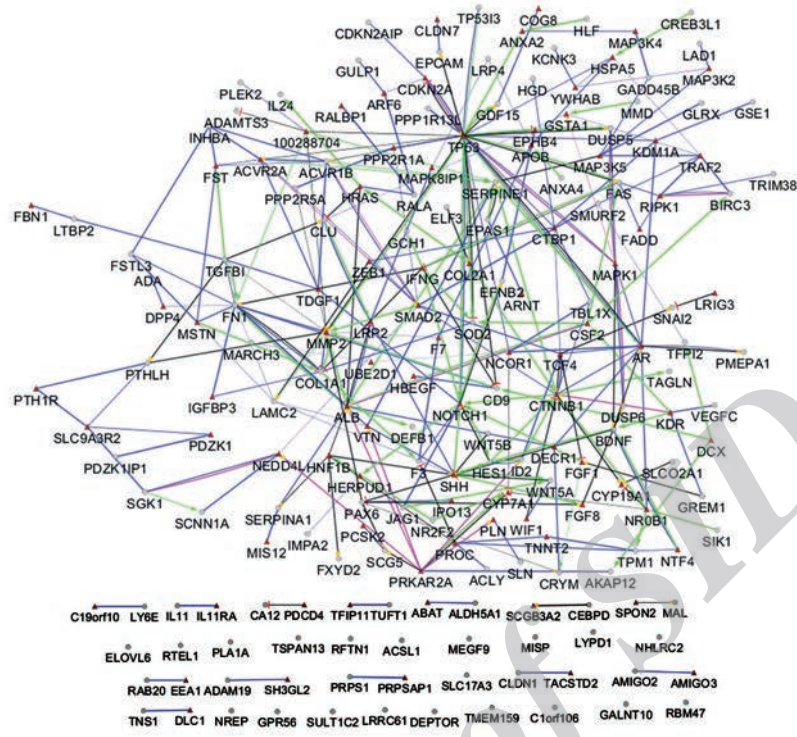


D

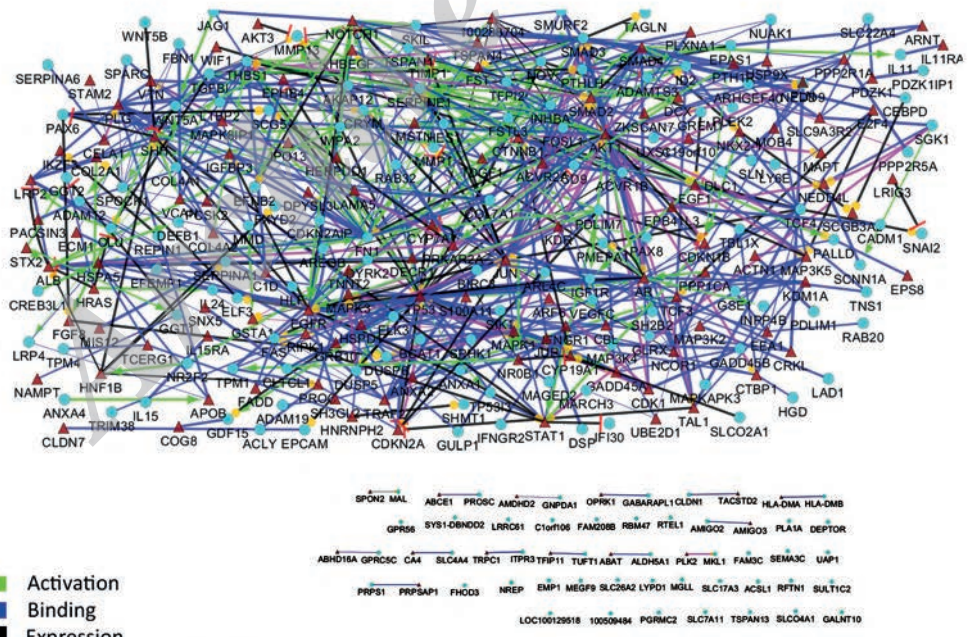


**Fig.1:** Interaction networks of the DE genes in the microarray dataset were poor and few signaling pathways were enriched. The expression profiles of human kidney cells treated with TGFβ-1 for 24 or 48 hours were compared to untreated cells. The interaction networks of the differentially expressed genes in the time points of **A**, 24 hours and **B**, 48 hours have few edges. In addition, pathway enrichment analysis of these genes in **C**, 24 hours and **D**, 48 hours could not detect key signaling pathways. Pathways with adjusted  $P \leq 0.05$  are shown. Color represents the gene ontology (GO) term level. TGFβ-1; Transforming growth factor Beta-1 and DE; Differentially expressed.

A



B



**Fig.2:** Enrichment of the protein-protein interaction (PPI) network is an efficient method to predict the missed interacting nodes. The networks of **A.** 24 hours and **B.** 48 hours treatment were enriched. The selected nodes from microarray experiment are depicted with ellipse and enriched nodes with triangle.

**Table 2:** The top 20% genes with the best rank in degree, betweenness centrality, and closeness centrality parameters in the enriched protein-protein interaction (PPI) network of time 24 hours

Genes	Degree	Genes	Betweenness	Genes	Closeness
<i>TP53</i>	35	<i>TP53</i>	0.338181	<i>TP53</i>	0.397906
<i>FN1</i>	16	<i>MMP2</i>	0.171374	<i>MMP2</i>	0.38191
<i>CTNNB1</i>	15	<i>ALB</i>	0.145517	<i>NOTCH1</i>	0.361045
<i>MMP2</i>	15	<i>CTNNB1</i>	0.125341	<i>ALB</i>	0.356808
<i>ALB</i>	14	<i>NOTCH1</i>	0.11946	<i>AR</i>	0.35023
<i>AR</i>	14	<i>SERPINE1</i>	0.100643	<i>CTNNB1</i>	0.347032
<i>NOTCH1</i>	14	<i>AR</i>	0.085536	<i>SERPINE1</i>	0.344671
<i>SHH</i>	13	<i>FN1</i>	0.075905	<i>SMAD2</i>	0.334802
<i>SMAD2</i>	11	<i>SHH</i>	0.069747	<i>FN1</i>	0.333333
<i>SERPINE1</i>	10	<i>SMAD2</i>	0.069443	<i>ACVR1B</i>	0.326882
<i>COL1A1</i>	9	<i>PRKAR2A</i>	0.067309	<i>ACVR2A</i>	0.326882
<i>PRKAR2A</i>	9	<i>HSPA5</i>	0.067049	<i>SHH</i>	0.325482
<i>MAPK1</i>	9	<i>MAP3K5</i>	0.049224	<i>CD9</i>	0.324094
<i>TGFBI</i>	8	<i>PTHLH</i>	0.047677	<i>MAPK1</i>	0.319328
<i>ACVR1B</i>	8	<i>HRAS</i>	0.045799	<i>NCOR1</i>	0.316667
<i>IFNG</i>	8	<i>TGFBI</i>	0.044403	<i>LAMC2</i>	0.31405
<i>TCF4</i>	8	<i>HNF1B</i>	0.040631	<i>VTN</i>	0.312115
<i>ACVR2A</i>	8	<i>CDKN2A</i>	0.039982	<i>FAS</i>	0.310838
<i>FAS</i>	7	<i>NCOR1</i>	0.039374	<i>TCF4</i>	0.310204
<i>BDNF</i>	7	<i>PAX6</i>	0.038418	<i>SOD2</i>	0.308943
<i>CD9</i>	6	<i>CD9</i>	0.038417	<i>CTBP1</i>	0.307692

Table 2: Continued

Genes	Degree	Genes	Betweenness	Genes	Closeness
<i>TP53</i>	35	<i>TP53</i>	0.338181	<i>TP53</i>	0.397906
<i>LAMC2</i>	6	<i>TCF4</i>	0.037912	<i>PAX6</i>	0.306452
<i>CTBP1</i>	6	<i>NROB1</i>	0.035918	<i>HES1</i>	0.306452
<i>PAX6</i>	6	<i>FAS</i>	0.035234	<i>HSPA5</i>	0.305835
<i>HES1</i>	6	<i>MAPK1</i>	0.031996	<i>IFNG</i>	0.305221
<i>CSF2</i>	6	<i>NEDD4L</i>	0.030261	<i>KDM1A</i>	0.305221
<i>NROB1</i>	6	<i>SLC9A3R2</i>	0.028821	<i>TGFB1</i>	0.304609
<i>HNF1B</i>	6	<i>IFNG</i>	0.028734	<i>CSF2</i>	0.304609
<i>LRP2</i>	6	<i>CSF2</i>	0.027151	<i>PRKAR2A</i>	0.304
<i>TRAF2</i>	6	<i>ANXA2</i>	0.026874	<i>DECR1</i>	0.303393
<i>RIPK1</i>	6	<i>PROC</i>	0.026277	<i>PPP2R1A</i>	0.302187
<i>NCOR1</i>	5	<i>KDR</i>	0.024832	<i>DECR1</i>	0.303393
<i>VTN</i>	5	<i>CTBP1</i>	0.024626	<i>PPP2R1A</i>	0.302187
<i>SOD2</i>	5	<i>APOB</i>	0.024534	<i>COL1A1</i>	0.30099
<i>HSPA5</i>	5	<i>TRAF2</i>	0.024347	<i>BDNF</i>	0.298625
<i>CDKN2A</i>	5	<i>F3</i>	0.022625	<i>TDGF1</i>	0.295146
<i>HRAS</i>	5	<i>BDNF</i>	0.022597	<i>F7</i>	0.294574
<i>CYP7A1</i>	5	<i>LRP2</i>	0.022317	<i>NROB1</i>	0.293436
<i>KDR</i>	5	<i>COL2A1</i>	0.022231	<i>HNF1B</i>	0.292308
<i>ID2</i>	5	<i>GSTA1</i>	0.021794	<i>CDKN2A</i>	0.291747
<i>MAP3K5</i>	5	<i>VTN</i>	0.021145	<i>DUSP5</i>	0.290631
<i>CLU</i>	5	<i>ARF6</i>	0.020175	<i>LRP2</i>	0.290076
<i>NEDD4L</i>	5	<i>YWHAB</i>	0.01996	<i>ANXA2</i>	0.289524
<i>FST</i>	5	<i>ACVR1B</i>	0.018667	<i>F3</i>	0.288425
<i>MSTN</i>	5	<i>ACVR2A</i>	0.018667	<i>PTH1H</i>	0.287335
<i>PROC</i>	5	<i>RALA</i>	0.018621	<i>HRAS</i>	0.286792

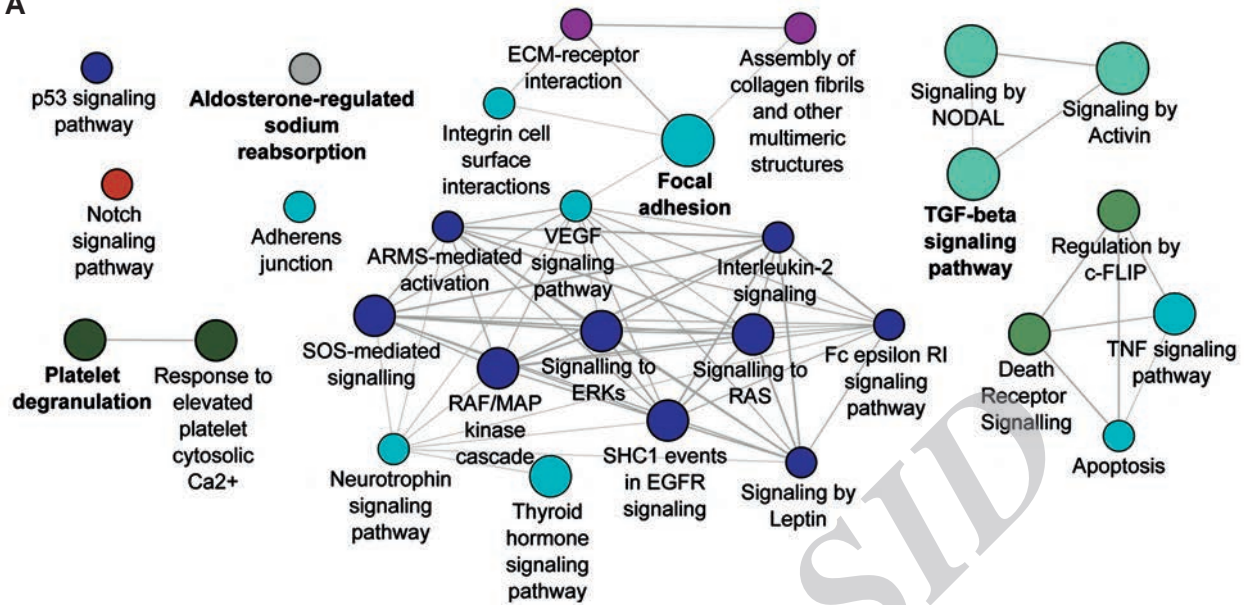
**Table 3:** The top 15% genes with the best rank in degree, betweenness centrality, and closeness centrality parameters in the enriched protein-protein interaction (PPI) network of time 48 hours

Genes	Degree	Genes	Betweenness	Genes	Closeness
<i>TP53</i>	55	<i>TP53</i>	0.218425	<i>JUN</i>	0.419966
<i>AKT1</i>	49	<i>AKT1</i>	0.180618	<i>TP53</i>	0.419244
<i>EGFR</i>	33	<i>EGFR</i>	0.129406	<i>AKT1</i>	0.415673
<i>SMAD3</i>	32	<i>JUN</i>	0.121849	<i>EGFR</i>	0.403974
<i>JUN</i>	32	<i>SMAD3</i>	0.091028	<i>AR</i>	0.403306
<i>AR</i>	28	<i>ALB</i>	0.08664	<i>SMAD3</i>	0.394184
<i>FN1</i>	25	<i>CTNNB1</i>	0.077284	<i>CTNNB1</i>	0.3904
<i>THBS1</i>	24	<i>AR</i>	0.063727	<i>SMAD4</i>	0.387917
<i>CTNNB1</i>	23	<i>SMAD4</i>	0.059499	<i>SERPINE1</i>	0.387917
<i>SMAD2</i>	23	<i>FN1</i>	0.056896	<i>NOTCH1</i>	0.380655
<i>SERPINE1</i>	20	<i>THBS1</i>	0.049411	<i>THBS1</i>	0.377709
<i>SMAD4</i>	20	<i>SHH</i>	0.0474	<i>SMAD2</i>	0.375963
<i>NOTCH1</i>	18	<i>NOTCH1</i>	0.044617	<i>FN1</i>	0.371951
<i>ALB</i>	16	<i>SERPINE1</i>	0.041249	<i>MMP1</i>	0.369138
<i>SHH</i>	16	<i>STAT1</i>	0.039746	<i>MAPK1</i>	0.365269
<i>PLG</i>	15	<i>HSPA5</i>	0.037599	<i>STAT1</i>	0.364179
<i>MMP1</i>	14	<i>PLG</i>	0.035914	<i>MMP13</i>	0.359882
<i>TCF4</i>	13	<i>TRAF2</i>	0.035805	<i>ALB</i>	0.357247
<i>TGFBI</i>	12	<i>PRKAR2A</i>	0.03185	<i>IGF1R</i>	0.357247
<i>MAPK1</i>	12	<i>SMAD2</i>	0.028572	<i>ACVR1B</i>	0.350575
<i>ACVR1B</i>	11	<i>SLC9A3R2</i>	0.028142	<i>ACVR2A</i>	0.350575
<i>CSF2</i>	11	<i>HSPD1</i>	0.027604	<i>CSF2</i>	0.34907

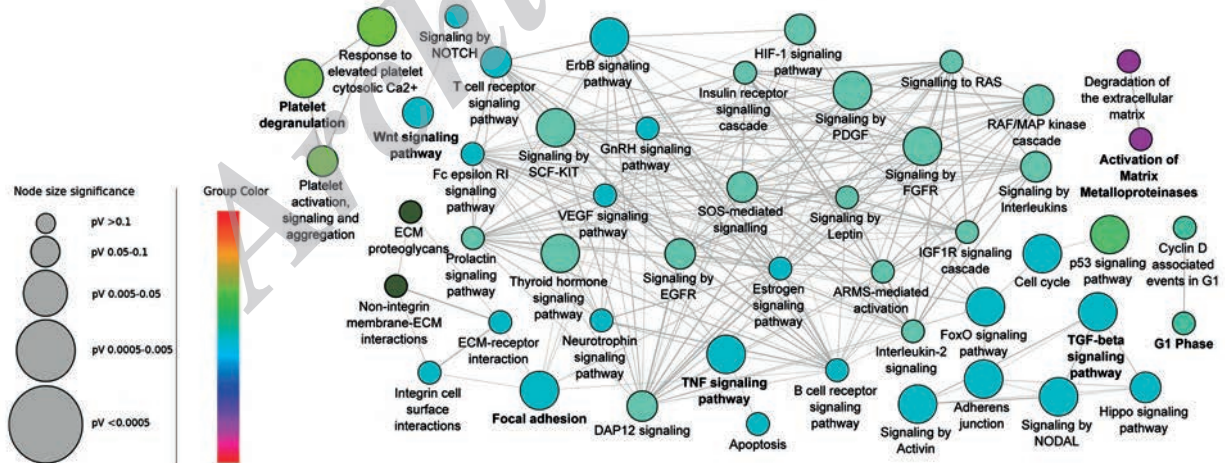
Table 3: Continued

Genes	Degree	Genes	Betweenness	Genes	Closeness
<i>PRKAR2A</i>	11	<i>HRAS</i>	0.025499	<i>KDR</i>	0.348074
<i>STAT1</i>	11	<i>TCF4</i>	0.024472	<i>CDK1</i>	0.348074
<i>TRAF2</i>	11	<i>PALLD</i>	0.024455	<i>CTBP1</i>	0.347578
<i>IGF1R</i>	10	<i>TGFBI</i>	0.024427	<i>PPP2R1A</i>	0.346591
<i>CDKN2A</i>	10	<i>STX2</i>	0.024388	<i>SHH</i>	0.343662
<i>MAP3K5</i>	10	<i>CDKN2A</i>	0.022895	<i>SPOCK1</i>	0.343662
<i>ACVR2A</i>	10	<i>CD9</i>	0.021863	<i>GRB10</i>	0.343179
<i>ID2</i>	9	<i>NCOR1</i>	0.021568	<i>NOV</i>	0.342216
<i>MMP13</i>	9	<i>MAP3K5</i>	0.020279	<i>GSTA1</i>	0.33936
<i>SKIL</i>	9	<i>HNF1B</i>	0.020242	<i>TCF4</i>	0.338889
<i>SPOCK1</i>	9	<i>SPOCK1</i>	0.018718	<i>FAS</i>	0.336088
<i>PDLIM7</i>	9	<i>CTBP1</i>	0.018114	<i>CDKN2A</i>	0.335626
<i>KDR</i>	9	<i>TPM1</i>	0.018056	<i>NCOR1</i>	0.335626
<i>LRP2</i>	9	<i>PTHLH</i>	0.017657	<i>TGFBI</i>	0.335165
<i>TCF3</i>	9	<i>TBL1X</i>	0.016927	<i>VCAN</i>	0.334705
<i>NOV</i>	8	<i>CSF2</i>	0.015788	<i>HSPA5</i>	0.334247
<i>PTHLH</i>	8	<i>GSTA1</i>	0.015304	<i>CLTCL1</i>	0.333333
<i>CDKN1B</i>	8	<i>KDR</i>	0.015185	<i>SKIL</i>	0.333333
<i>GADD45A</i>	8	<i>MMP13</i>	0.014604	<i>PLG</i>	0.332879
<i>GRB10</i>	8	<i>ANXA2</i>	0.01411	<i>PTHLH</i>	0.332879
<i>LAMA5</i>	8	<i>CLTCL1</i>	0.013602	<i>PRKAR2A</i>	0.332425
<i>VTN</i>	8	<i>MAPK1</i>	0.012772	<i>MAP3K5</i>	0.330623
<i>CBL</i>	8	<i>TCF3</i>	0.012502	<i>LAMA5</i>	0.330176

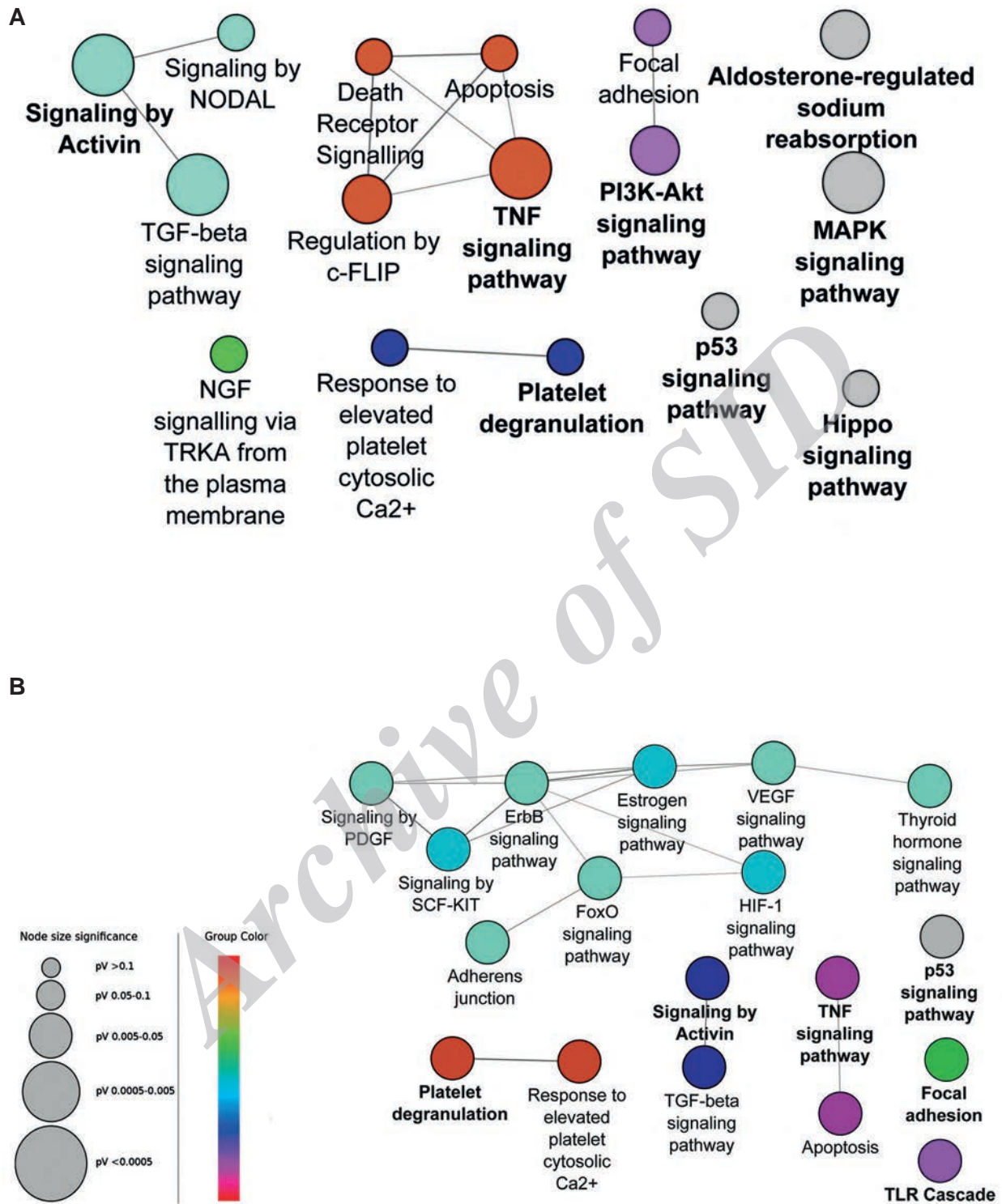
A



B



**Fig.3:** Selection of central nodes for pathway enrichment analysis can detect critical signaling pathways. In the enriched protein-protein interaction (PPI) networks, 62 genes for 24 hours treatment network and 60 genes for 48 hours treatment network were chosen as nodes with high centrality. These central nodes are related to 29 and 49 highly connected pathways in A. 24 hours and B. 48 hours, respectively. Pathways with adjusted  $P \leq 0.05$  are shown. Color represents the gene ontology (GO) term level.



**Fig.4:** Pathway enrichment analysis with total genes in the enriched network is not informative. Pathway enrichment analysis with all 199 genes in 24 hours, or 301 genes in 48 hours treatment in enriched PPI networks only demonstrated **A.** 16 or **B.** 18 poorly inter-connected pathways, respectively. Pathways with adjusted  $P \leq 0.05$  are shown. Color represents the gene ontology (GO) term level.



## Discussion

In this study, we reanalyzed a microarray dataset to determine gene expression alteration in response to TGF $\beta$ -1 in a human kidney cell line. The investigators who originally generated this data emphasized the involvement of Notch signaling pathway based on a few DE genes (6). In contrast, we have constructed PPI networks for DE genes in the time points of 24 and 48 hours treatment. We found that expansion of these networks followed by selection of central nodes for pathway enrichment analysis is an efficient method to recognize key signaling pathways in response to TGF $\beta$ -1 stimulation. Our analysis also predicted the potential role of some novel pathways in this *in vitro* model and also pointed out time-dependent activation of particular pathways. Interestingly, the same investigators later repeated the experiment and assessed the mRNA expression profile by RNA-Seq and found that this technique is superior to microarray in identification of the DE genes and altered signaling pathways (11). Noteworthy, the signaling pathways determined by our analysis on the original microarray dataset is similar to the pathways identified with RNA-Seq data.

An interesting finding in this study was that pathway enrichment analysis with the DE genes in the microarray experiment was not efficient for prediction of key signaling pathways. However, it was expected that all important genes were not regulated at the mRNA level and so they were not detectable by mRNA microarrays. Therefore, to compensate for this limitation, we constructed a PPI network of DE genes and then enriched this network by adding genes that were previously known to be interacting with the initial network nodes. This expanded gene set was more informative for detecting signaling pathways. Indeed, it is perfect to perform multi-level assessments in biological experiments, but for practical reasons it is not commonly feasible. In this case, it is possible to measure changes at one level and then make bioinformatics predictions to fill the gaps at other levels.

Several previous studies have shown that highly connected nodes (hubs) in the networks, determined by degree parameter, are vital for the organism survival (12). Next studies revealed that essential genes in the network can be determined

not only by degree but also by other centrality parameters, such as betweenness or closeness centrality (13, 14). Here, we have used a combination of these three network topology parameters to determine the central nodes. Interestingly, pathway enrichment with these central genes was more informative than enrichment with the initial genes or even with the total genes in the expanded PPI networks. This observation is in line with our recent study on diabetic nephropathy showing the central network nodes tend to be present in signaling pathways and cross talks (15).

In pathway enrichment analysis, Hippo, PDGF, and FGFR signaling pathways were detected only in the second time point, 48 hours treatment. Actually, the initial activation of upstream signaling pathways detected in 24 hours treatment may lead to the expression of genes, related to these three pathways after 48 hours. This finding on time-specific expression of genes underscores the importance of time-course designs for gene expression analysis experiments.

Most of the predicted pathways in our analysis such as Notch, TNF, P53, and TGF $\beta$  signaling have been previously known to be involved in the pathogenesis of CKD (16-19), whereas, for some others, such as platelet degranulation pathway, there is not currently direct experimental proof for participation in renal fibrosis. However, previous experiments have shown megakaryocytes as mediators of fibrosis in a subset of hematologic malignancies, idiopathic pulmonary fibrosis, as well as bone marrow (20-22). The role of megakaryocytes in kidney fibrosis is an interesting topic for future studies.

## Conclusion

We have here employed a holistic approach to assess the consequences of TGF $\beta$ -1 stimulation in kidney cells. Although, high-throughput techniques are frequently applied in biological investigations, data interpretation is yet commonly limited to the assessment of most up or down-regulated factors missing the huge effect of interactions for genes with subtle expression change. Systems biology provides novel concepts and methods to infer the underlying mechanisms of biological phenomena from *omics* raw data and hopefully will bring a higher quality of life to those suffering from chronic diseases.

## Acknowledgments

This study was financially supported by Iran National Science Foundation (INSF) and Isfahan University of Medical Sciences (393088). There is no conflict of interest in this study.

## References

1. Thomas R, Kanso A, Sedor JR. Chronic kidney disease and its complications. *Prim Care*. 2008; 35(2): 329-344.
2. Lan HY. Diverse roles of TGF-beta/Smads in renal fibrosis and inflammation. *Int J Biol Sci*. 2011; 7(7): 1056-1067.
3. Bottinger EP. TGF-beta in renal injury and disease. *Semin Nephrol*. 2007; 27(3): 309-320.
4. Munoz-Felix JM, Gonzalez-Nunez M, Martinez-Salgado C, Lopez-Novoa JM. TGF-beta/BMP proteins as therapeutic targets in renal fibrosis. Where have we arrived after 25 years of trials and tribulations? *Pharmacol Ther*. 2015; 156: 44-58.
5. Jin Y, Ratnam K, Chuang PY, Fan Y, Zhong Y, Dai Y, et al. A systems approach identifies HIPK2 as a key regulator of kidney fibrosis. *Nat Med*. 2012; 18(4): 580-588.
6. Walsh DW, Roxburgh SA, McGettigan P, Berthier CC, Higgins DG, Kretzler M, et al. Co-regulation of Gremlin and Notch signalling in diabetic nephropathy. *Biochim Biophys Acta*. 2008; 1782(1): 10-21.
7. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, et al. NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Res*. 2005; 33(Database issue): D562-566.
8. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics*. 2013; 29(5): 661-663.
9. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13(11): 2498-2504.
10. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009; 25(8): 1091-1093.
11. Brennan EP, Morine MJ, Walsh DW, Roxburgh SA, Lindenmeyer MT, Brazil DP, et al. Next-generation sequencing identifies TGF-beta1-associated gene expression profiles in renal epithelial cells reiterated in human diabetic nephropathy. *Biochim Biophys Acta*. 2012; 1822(4): 589-599.
12. He X, Zhang J. Why do hubs tend to be essential in protein networks? *PLoS Genet*. 2006; 2(6): e88.
13. Raman K, Damaraju N, Joshi GK. The organisational structure of protein networks: revisiting the centrality-lethality hypothesis. *Syst Synth Biol*. 2014; 8(1): 73-81.
14. Ozgur A, Vu T, Erkan G, Radev DR. Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics*. 2008; 24(13): i277-285.
15. Abedi M, Gheisari Y. Nodes with high centrality in protein interaction networks are responsible for driving signaling pathways in diabetic nephropathy. *PeerJ*. 2015; 3: e1284.
16. Wang W, Huang XR, Li AG, Liu F, Li JH, Truong LD, et al. Signaling mechanism of TGF-beta1 in prevention of renal inflammation: role of Smad7. *J Am Soc Nephrol*. 2005; 16(5): 1371-1383.
17. Sharma S, Sirin Y, Susztak K. The story of Notch and chronic kidney disease. *Curr Opin Nephrol Hypertens*. 2011; 20(1): 56-61.
18. Misseri R, Meldrum DR, Dinarello CA, Dagher P, Hile KL, Rink RC, et al. TNF-alpha mediates obstruction-induced renal tubular cell apoptosis and proapoptotic signaling. *Am J Physiol Renal Physiol*. 2005; 288(2): F406-411.
19. Samarakoon R, Overstreet JM, Higgins SP, Higgins PJ. TGF-beta1 --> SMAD/p53/USF2 --> PAI-1 transcriptional axis in ureteral obstruction-induced renal fibrosis. *Cell Tissue Res*. 2012; 347(1): 117-128.
20. Papadantonakis N, Matsuura S, Ravid K. Megakaryocyte pathology and bone marrow fibrosis: the lysyl oxidase connection. *Blood*. 2012; 120(9): 1774-1781.
21. Thachil J. The lung megakaryocytes and pulmonary fibrosis in systemic sclerosis. *Med Hypotheses*. 2009; 72(3): 291-293.
22. Cripe LD, Hromas R. Malignant disorders of megakaryocytes. *Semin Hematol*. 1998; 35(3): 200-209.