Network-based genetic profiling, and therapeutic target identification of Thyroid Cancer

1st Md. Ali Hossain
Dept. of CSE
MIU, Jahangirnagar University
Dhaka, Bangladesh
ali.cse.bd@gmail.com

2nd Tania Akter Asa
Dept. of EEE
Islamic University
Kushtia, Bangladesh
tania.eee.iu@gmail.com

3rd Julian Quinn
Bone biology divisions
Garvan Institute of Medical Research
Sydney, Australia
j.quinn@garvan.org.au

4rdMd. Mijanur Rahman
Dept. of CSE
JKKNIU
Bangladesh
mijan_cse@yahoo.com

5th Fazlul Huq
Biomedical Sciences, Faculty of Medicine and Health
The University of Sydney
Sydney, Australia
fazlul.huq@sydney.edu.au

6th Mohammad Ali Moni
Biomedical Sciences, Faculty of Medicine and Health
The University of Sydney
Sydney, Australia
mohammad.moni@sydney.edu.au

***Abstract*—Molecular mechanisms that underlie the pathogenesis and progression of malignant thyroid cancer (TC) are poorly understood. In this study, we employed network-based integrative analyses of TC lesions to identify key molecules that are possible hub genes and proteins in molecular pathways active in TC. We thus studied a microarray gene expression dataset (GSE82208, n=52) that compared TC to benign thyroid tumours (follicular thyroid adenomas) in order to identify differentially expressed genes (DEGs) in TC. We used Gene Ontology (GO) and KEGG pathway analyses to identify potential gene and pathways roles as well as candidate hub genes identified by protein-protein interaction (PPI) analysis.**

TC showed altered levels of 598 gene transcripts, with 133 genes up- and 465 genes down-regulated. We observed four significant pathways (One carbon pool by folate, p53 signaling, Progesterone-mediated oocyte maturation signaling and Cell cycle pathways) connected with the significantly up-regulated DEGs; eight pathways were connected with the significantly down-regulated DEGs in TC. We observed ten significant GO groups connected with the significantly up-regulated DEGs and eighty connected with the significantly down-regulated DEGs in TC. The PPI analysis identified 12 potential hub genes based on degree and betweenness centrality. These included TOP2A, JUN, EGFR, CDK1, FOS, CDKN3, EZH2, TYMS, PBK, CDH1, UBE2C and CCNB2. Moreover, transcription factors (TFs) that may influence the observed TC gene expression were identified, including FOXC1, GATA2, YY1, FOXL1, E2F1, NFIC, SRF, TFAP2A, HINFP, and CREB1. We also identified potential microRNA (miRNAs) that may also affect transcript levels of DEGs, including hsa-mir-335-5p, -26b-5p, -124-3p, -16-5p, -192-5p, -1-3p, -17-5p, -92a-3p, -215-5p, and -20a-5p.

Thus, our study identified DEGs, molecular pathways, TFs and miRNAs that identify pathways and molecular mechanisms that may underly malignant progression in TC. These include novel candidates for further studies to identify new TC biomarkers and therapeutic targets.

***Index Terms*—thyroid cancer, protein-protein interaction, reporter transcription factors, reporter microRNAs, molecular pathways**

I. Introduction

Thyroid cancer (TC) is the most common endocrine malignant disease [24], and although it has a relatively low mortality rate compared to most other common metastatic diseases its incidence is rising and is now the fifth most common cancer disease in women, notably affecting those in the 20 to 34 year age range. Global incidence of thyroid cancer is increasing about 5% per year [1], although some of this increase may be due to better detection technologies. TC prevalence in the United States in 2012 was 56,460 new diagnoses of thyroid cancer and 1,780 TC-related deaths. The disease has a particularly high incidence in Korea [13]. While TC is three times more common in women than men, the death rate is similar between the sexes due to a much higher mortality rate in men. Where the disease is localized only in the thyroid, the 5-year survival rate is over 99%, but presence of distant metastasis is a associated with a survival rate below 55%. Four major types of thyroid cancer are defined, namely medullary thyroid carcinoma (the most common, arising from parafollicaular cells), follicular thyroid carcinoma, papillary thyroid carcinoma, and anaplastic thyroid carcinoma [3].

The causes and development processes that give rise to TC are poorly understood, and these cancers can be difficult to treat. Treatments vary by stage and type, with surgery (thyroidectomy) regarded as first line, but also including radioactive iodine, tyrosine kinase inhibitors (TKIs), as well as external beam radiation. However, these can be of limited utility with advanced disease. There is therefore a great need to understand the mechanisms that underlie TC development and progression. For this we can employ molecular analysis approaches, for example using differential expression of genes (DEG) to discover factors involved in dysregulation of cell functions and metabolic pathways that may underlie morbidity and comorbidity [20], [22]. The repertoire of candidate factors can be extended using approaches such as gene ontology anal-

ysis and protein-protein interactions (PPI) studies. Diseases and disease processes can be functionally connected if they involve with the same dysregulated genes or gene pathways.

For identifying DEGs in pathological and normal tissues, microarray gene expression profiling is widely used [9] and several such studies have been performed in TC [30]. For example, Huang et al. [12] studied gene expression profiles of TC patients but addressed only transcript data, while functional interactions among the gene products were not considered. To better understand the molecular mechanisms behind diseases and to identify critical biomolecules, integrative analysis within the network medicine context is needed [21]. Candidate TC biomarkers obtained from such studies may be potential therapeutic targets. Such a systems biology approach integrating network statistical and topological analyses of experimental datasets can clarify disease mechanisms [17]. For this reason, in our study, we used such a systems approach to identify TC molecular signatures at miRNA, mRNA, and protein levels, that differ from that of follicular thyroid adenomas, a type of benign thyroid tumour. Such a comparison is much more meaningful than a comparison of TC to normal thyroid tissue as the tumours are similar while differing markedly in their malignant behaviour. In the present work, the advances in these signaling pathways, ontology enrichment analysis, molecular signatures and targeted treatments of thyroid cancer were reviewed (Figure 1).

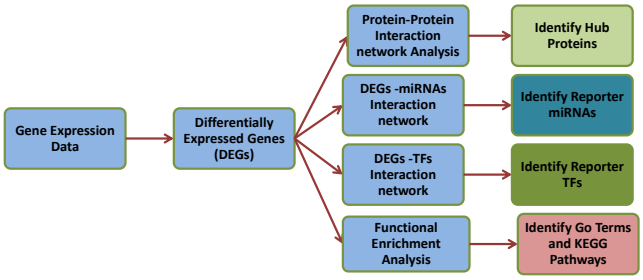


Fig. 1: The multi-stage analysis methodology was employed in this study. Gene expression datasets related to TC were collected from the NCBI Gene Expression Omnibus (NCBI-GEO) database. The dataset was statistically analyzed using GEO2R to identify DEGs. Four types of functional enrichment analyses of DEGs were then performed to identify significantly enriched pathways. Thus, we constructed protein-protein interaction networks around DEGs topological analyses to identify putative pathways hub proteins, identified possible microRNA (miRNA) and transcriptions factor (TF) interactors, and used Gene Ontology annotation terms to provide pathways enrichment. TF and miRNA studies employed JASPAR and miRTarbase databases, respectively. DEGs were integrated with those networks and higher degree and betweenness centrality was used to designate TFs and miRNAs as the reporter transcriptional regulatory elements. The target DEGs of reporter miRNAs and TFs were subjected to pathway enrichment analyses.

II. MATERIALS AND METHODS

In this study, the multi-step analysis method we developed and applied is shown in Fig.1. We statistically analyzed Gene expression datasets to identify the DEGs and their regulatory patterns. We employed DEGs to identify enriched pathways, biological processes and annotation terms (i.e., Gene Ontology terms) by using functional enrichment methods. Then, to identify reporter biomolecules, We integrated the intermediate analysis results with biomolecular networks.

Dataset Employed and Statistical Methods Used

We obtained the gene expression data of TC (GSE82208) for our study from the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) [4]. The data included analyses of patients of different age and sex. We performed several rounds of filtering, normalization and statistical analysis. To identify genes that was differentially expressed in patients, $Students$ unpaired t-test was performed. To identify a gene as DEG a threshold of at least a 2-fold change in expression levels, and a p-value < 0.05 for the t-tests were chosen.

Functional Enrichment of Gene Sets

We performed gene ontology and pathway analysis on identified DEGs using the David bioinformatics resources (https://david-d.ncifcrf.gov/) (version v6.8) [34] to get further insight into the molecular pathways of TC. In these analyses, Gene Ontology (GO) terminology and KEGG pathway databases were used as annotation sources. Enrichment results showing an adjusted $p - value < 0.05$ were considered significant.

Construction and Analysis of Protein-Protein Interaction (PPI) Sub-networks

The PPI network was first constructed with the DEGs and analyzed using STRING [35] a web-based visualization software resource. The constructed PPI network was represented as an undirected graph, where nodes represent the proteins and the edges represent the interactions between the proteins. The PPI network was then visualized and analyzed using Cytoscape (v3.5.1) [33]. Topological analysis was applied to identify highly connected proteins (i.e., hub proteins) through the Cyto-Hubba plugin [7]. The top three modules (i.e., the three most highly interconnected protein clusters) in the PPI subnetwork were identified using the MCODE plug-in [7]. These modules were further analyzed and characterized using enrichment analyses by NetworkAnalyst [38]. The KEGG pathway enrichment analysis of the PPI networks involved DEGs were performed by NetworkAnalyst [38].

Identifying TFs and miRNAs that Influence Expression of Candidate Genes

To identify TFs and miRNAs that affect transcript levels around which significant changes occur at the transcriptional level, we obtained experimentally verified TF-target genes from the JASPAR database [15] and miRNA-target gene

interactions from TarBase [32] and miRTarBase [11] by using NetworkAnalyst tools [38].

III. Results and Discussion

A. Results of DEG analyses

Transcriptomic signatures: Differentially expressed genes

Compared to normal subjects, the TC patient gene expression patterns were analysed using oligonucleotide microarrays from the NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE82208) [37]. Gene expression profiling was performed in 27 malignant follicular thyroid cancers and 25 follicular thyroid adenomas. In this analysis, it was found that 598 genes ($p < 0.05, > 1.0$ log2 fold change) were differentially expressed as compared to controls (follicular thyroid adenomas), amongst which 465 genes were significantly down-regulated and 133 genes were up-regulated (see Additional file 1:Table S1).

B. Pathway and functional correlation analysis

Combining large scale, state of the art transcriptom and proteome analysis, we have performed a regulatory analysis to gain further insight into the molecular pathways associated with the TC and predicted links to the affected pathways. DEGs and pathways were analysed using KEGG pathway database $(http://www.genome.jp/kegg/pathway.html)$ and functional annotation analysis tool DAVID v 6.8 (http://niaid.abcc.ncifcrf.gov) to identify overrepresented pathway groups amongst DEGs and to group them into functional categories. It was observed that four significant pathways, including the "One carbon pool by folate" pathway, p53 signaling pathway, cell cycle and progesterone-mediated oocyte maturation signaling pathway. These are associated with the significantly up-regulated genes of TC. Fold enrichment, adjusted p-values and genes associated with these pathways are presented in the table Table I (a) A. It was also observed that eight pathways were significantly over-represented, including mineral absorption, thyroid hormone synthesis, pathways in cancer, estrogen signaling, sphingolipid metabolism, protein processing in endoplasmic reticulum, axon guidance and focal adhesion pathways (Table I (a) B) were identified. These are associated with the significantly down-regulated DEGs in TC. We then performed gene ontology analysis using DAVID to obtain further insight into the molecular roles and biological function of DEGs in TC. From this analysis, nine gene ontology groups were associated with the significantly up-regulated DEG (see Table II (a)) and eighty GO groups were associated with the DEG significantly down-regulated in our TC dataset, the ten most significant gene ontology groups are presented in Table II (a). Fold enrichment, adjusted p-values and genes associated with these ontologies are also indicated.

Proteomic signatures: Hub target proteins from PPI analysis

The protein-protein interaction network was constructed with the DEGs (Fig. 2 A.) by using STRING. Topological analysis using STRING and further anlysis by Cytoscape's Cyto-Hubba plugin identified the most significant hub proteins, which were identified as gene products of TOP2A, JUN, EGFR, CDK1, FOS, CDKN3, EZH2, TYMS, PBK, CDH1, UBE2C and CCNB2. The simplified protein-protein interaction network for most significant hub genes were constructed by using Cyto-Hubba plugin and shown in fig. 2B.

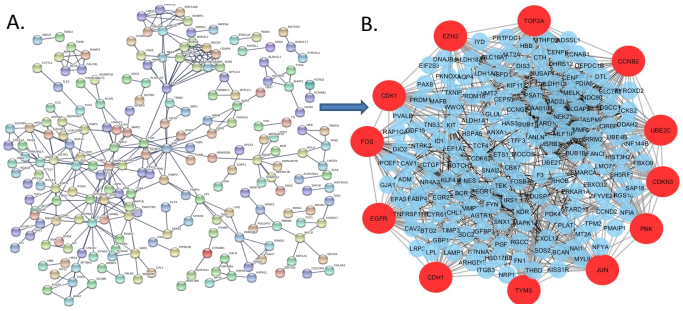


Fig. 2: A. The protein-protein interaction network found in the DEGs of TC. B. The protein-protein interaction(PPI) network identified in TC dataset DEGs showing the most significant hub genes on the periphery (red)

Enrichment Analysis of the Modules found in PPI of DEG.

The PPI network was analyzed by using MCODE plug-in in Cytoscape (version 3.5.1), and top three modules were selected (Fig. 3). The enrichment analysis of the top three modules

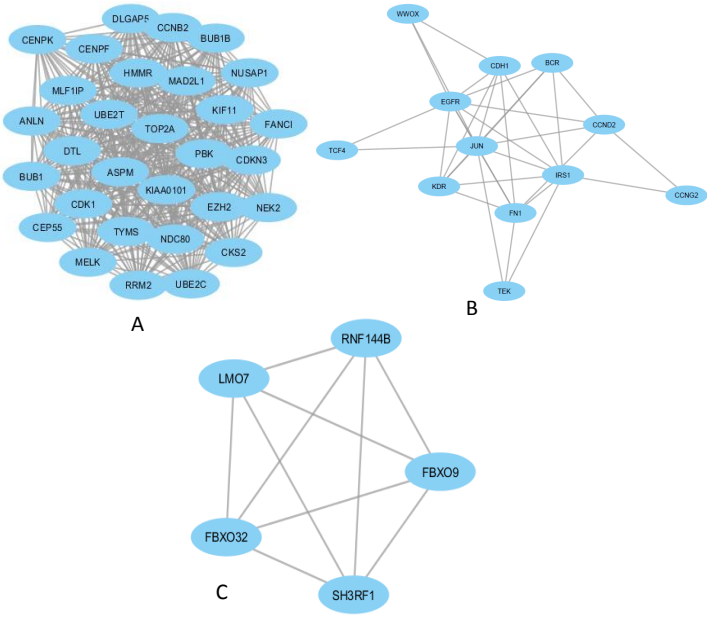


Fig. 3: Top three modules in the protein-protein interaction network of the DEGs in TC. The nodes indicate the DEGs and the edges indicate the interactions between two genes.

was analyzed by using DAVID. Module 1 represented a set of biological processs significantly enriched for ATP binding, cell division, cell proliferation; associated enriched cellular components were kinetochore, condensed chromosome kinetochore,

TABLE I: (a) A. Pathways associated with the significantly genes showing up-regulated expression in TC. The set of significantly up-regulated genes for TC showed an enrichment of four KEGG pathways using DAVID. (a) B. Pathways associated with significantly down-regulated genes of TC. The set of significantly down-regulated genes for TC showed in an enrichment of eight KEGG pathways using DAVID. Fold enrichment, adjusted p-values and genes associated with these pathways are presented in the table Tables. (b) Summary of reporter biomolecules (Hub genes, TFs, and miRNAs) in TC

A.

KEGG ID	Pathway	Genes in the pathway	Fold Enrichment	Adjusted p-value
hsa00670	One carbon pool by folate	MTHFD2, TYMS, SHMT2, ALDH1L1	24	5.21E-04
hsa04115	p53 signaling pathway	CDK1, CCNB2, RRM2, PMAIP1, IGFBP3	9	2.03E-03
hsa04110	Cell cycle	CDK1, MAD2L1, CCNB2, BUB1, BUB1B	5	1.77E-02
hsa04914	Progesterone-mediated oocyte maturation	CDK1, MAD2L1, CCNB2, BUB1	6	3.31E-02

B.

KEGG ID	Pathway	Genes in the pathway	Fold Enrichment	Adjusted p-value
hsa04978	Mineral absorption	MT1M, MT2A, CYBRD1, MT1E, MT1H, MT1X, MT1G, MT1F	6	3.44E-04
hsa04918	Thyroid hormone synthesis	SLC26A4, TG, HSP90B1, PAX8, TPO, HSPA5, LRP2, TSHR, IYD	4	9.55E-04
hsa05200	Pathways in cancer	FGFR2, BCR, EPAS1, PGF, RXRB, GNAI1, RUNX1T1, FZD1, CDH1, CDK6, KIT, CTNNA1, MMP2, CXCL12, TCF7L1, AGTR1, FOS, HSP90B1, JUN, PAX8, SOS2, LAMB1	2	5.48E-03
hsa04915	Estrogen signaling pathway	HSPA1L, FOS, HSP90B1, GNAI1, JUN, SOS2, HSPA1A, HSPA1B, MMP2	3	8.41E-03
hsa00600	Sphingolipid metabolism	SPTLC1, SGMS2, SPTLC3, GALC, PLPP3, ASAH1	4	1.17E-02
hsa04360	Axon guidance	NRP1, LIMK2, SEMA6D, FYN, GNAI1, PLXNA2, SEMA3D, UNC5C, CXCL12, EPHA3	3	1.22E-02
hsa04141	Protein processing in endoplasmic reticulum	HSPA1L, HSP90B1, ERO1B, UBE4B, PDIA6, HSPA1A, HSPA1B, HSPA5, CALR, MAN1C1, SEC62	2	2.68E-02
hsa04510	Focal adhesion	ITGA9, CAV2, CAV1, CCND2, FYN, PGF, JUN, SOS2, ITGB3, LAMB1, KDR, MYL9	2	3.95E-02

(a)

Symbol	Description	Feature
UBA52	DNA topoisomerase 2-alpha	Hub protein
JUN	c-Jun in combination with c-Fos	Hub protein
EGFR	The epidermal growth factor receptor	Hub protein
CDK1	Cyclin Dependent Kinase 1	Hub protein
FOS	proto-oncogene	Hub protein
CDKN3	Cyclin Dependent Kinase Inhibitor 3	Hub protein
EZH2	Enhancer of zeste homolog 2	Hub protein
TYMS	Thymidylate Synthetase	Hub protein
PBK	PDZ binding kinase	Hub protein
CDH1	Cadherin-1	Hub protein
UBE2C	Ubiquitin Conjugating Enzyme E2 C	Hub protein
CCNB2	G2/Mitotic-Specific Cyclin-B2	Hub protein
FOXC1	Forkhead Box C1	Reporter Transcription Factor
GATA2	GATA Binding Protein 2	Reporter Transcription Factor
YY1	YY1 Transcription Factor	Reporter Transcription Factor
FOXL1	Forkhead Box L1	Reporter Transcription Factor
NFIC	Nuclear Factor I C	Reporter Transcription Factor
E2F1	E2F Transcription Factor 1	Reporter Transcription Factor
TFAP2A	Activating enhancer binding Protein 2 alpha	Reporter Transcription Factor
SRF	Serum Response Factor	Reporter Transcription Factor
HINFP	Histone H4 transcription factor	Reporter Transcription Factor
CREB1	Cyclic AMP-responsive element-binding protein 1, CREB transcription factor	Reporter Transcription Factor
hsa-mir-335-5p	MicroRNA 355	Reporter microRNA
hsa-mir-26b-5p	MicroRNA 26b	Reporter microRNA
hsa-mir-124-3p	MicroRNA 124	Reporter microRNA
hsa-mir-16-5p	MicroRNA 16	Reporter microRNA
hsa-mir-192-5p	MicroRNA 192	Reporter microRNA
hsa-mir-1-3p	MicroRNA 11	Reporter microRNA
hsa-mir-17-5p	MicroRNA 17	Reporter microRNA
hsa-mir-92a-3p	MicroRNA 92a	Reporter microRNA
hsa-mir-215-5p	MicroRNA 215	Reporter microRNA
hsa-mir-20a-5p	MicroRNA 20a	Reporter microRNA

(b)

nucleoplasm, cytoplasm. Molecular functions significantly enriched in module 1 included chromatin binding, protein kinase activity (Table III (a)). Module-2 showed significantly enriched biological process positive regulation of protein phosphorylation, protein autophosphorylation; the significantly enriched cellular components included membrane raft,endosome,apical plasma membrane, focal adhesion. Molecular functions of module-2 were enriched in protein tyrosine kinase activity, enzyme binding, protein binding, integrin binding (Table III (b)). Module-3 gene ontology annotation found significant enrichment for protein ubiquitination, with significantly enriched cellular components being ubiquitin ligase complex, SCF ubiquitin ligase complex. Module-3 molecular functions showed significantly enrichment in ubiquitin-protein transferase activity and zinc ion binding (Table III (b)). It was also notable that pathway enrichment analyses found modules 1 and 2 were significantly enriched in protein binding pathway which was not seen in was obtained in module-3 (Table III (b)).

C. Regulatory signatures: TFs and miRNAs affecting DEG Transcript Levels.

1) Transcriptional regulatory network construction and TF enrichment analysis: The construction of reporter TFs and DEGs interaction network constructed by NetworkAnalyst revealed a number of potentially important TFs selected by the topological analysis using dual metric approach involving degree and betweenness (Fig. 4). The top 10 ranked TFs with the highest degree and betweenness centrality included FOXC1, GATA2, YY1, FOXL1, E2F1, NFIC, SRF, TFAP2A, HINFP, and CREB1. Pathway enrichment analysis of the TFs of the TC DEGs mainly identified as statistically significant a number of relevant pathways that included Transcriptional misregulation in cancer, Acute myeloid leukemia, Pathways in cancer, Thyroid cancer, Cell cycle, Dorso-ventral axis formation, Arrhythmogenic right ventricular cardiomyopathy (ARVC), Cocaine addiction, Prostate cancer pathways (Table IV (b)).

2) miRNA regulatory network construction and enrichment analysis of miRNA: We constructed TC DEG-miRNA interactions networks using NetworkAnalyst (Fig. 5). By topological analysis, the top 10 ranked miRNAs by highest degree and betweenness centrality were selected which includes hsa-mir-335-5p, hsa-mir-26b-5p, hsa-mir-124-3p, hsa-mir-16-5p, hsa-mir-192-5p, hsa-mir-1-3p, hsa-mir-17-5p, hsa-mir-92a-3p, hsa-mir-215-5p, and hsa-mir-20a-5p. The analysis indicates that these are miRNAs having the highest potential to regulate levels of the DEG transcripts. The pathway enrichment analysis of the miRNA associated DEGs networks identified the p53 signaling pathway, Dorso-ventral axis formation, Pertussis, Adherens junction, Thyroid cancer, Pathways in cancer, Focal

TABLE II: (a) Gene ontologies associated with the up-regulated significant genes of TC. (b) Gene ontologies associated with the down-regulated significant genes of TC. Ten most significant biological functional ontologies are identified that are associated with these significant genes. Fold enrichment, adjusted p-values and genes associated with these ontologies are presented in these tables.

Go Term	Pathway	Genes in the pathway	Fold Enric-hment	Adjusted p-value
GO:0006636	unsaturated fatty acid biosynthetic process	ELOVL4, SCD, ELOVL7	45	1.88E-03
GO:0007094	mitotic spindle assembly checkpoint	MAD2L1, BUB1, BUB1B	33	3.39E-03
GO:0045773	positive regulation of axon extension	TNFRSF12A, MAP1B, FN1	24	6.83E-03
GO:0044772	mitotic cell cycle phase transition	CDK1, CKS2	67	2.92E-02
GO:0046602	regulation of mitotic centrosome separation	KIF11, NEK2	67	2.92E-02
GO:0021930	cerebellar granule cell precursor proliferation	ATF5, RORA	45	4.35E-02
GO:0035999	tetrahydrofolate interconversion	TYMS, SHMT2	45	4.35E-02
GO:0035338	long-chain fatty-acyl-CoA biosynthetic process	ELOVL4, ELOVL7	45	4.35E-02
GO:0046777	protein autophosphorylation	EGFR, NEK2, STK26, MELK	5	4.78E-02

(a)

Go Term	Pathway	Genes in the pathway	Fold Enrichment	Adjusted p-value
GO:0045926	negative regulation of growth	ING5, MT1M, MT2A, MT1E, MT1H, MT1X, MT1G, MT1F	18	1.37E-07
GO:0071294	cellular response to zinc ion	MT1M, MT2A, MT1E, MT1H, MT1X, MT1G, MT1F	16	3.17E-06
GO:0006590	thyroid hormone generation	DIO2, CPQ, FOXE1, TPO, DIO1, IYD	21	4.56E-06
GO:0001525	angiogenesis	FGFR2, EMCN, CAV1, NRP1, ACVRL1, EPAS1, PGF, MMP2, KDR, PKNOX1, CTGF, ID1, JUN, MAPK14, TEK, ECSCR, RHOB, CALCRL	3	1.89E-05
GO:0001938	positive regulation of endothelial cell proliferation	CAV2, NRP1, ACVRL1, PGF, F3, JUN, TEK, ITGB3, CXCL12, KDR	6	2.97E-05
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	FGFR2, GLIS3, TSHZ3, ZMYND11, CAV1, TFCP2L1, ZNF366, PRDM16, CALR, CBX7, CITED2, MINA, TCF4, BHLHE41, NFIL3, NR2F1, TXNIP, EGR1, NR4A2, TLE1, FOSB, SNAI2, FOXP2, CD36, BTG2, ID1, SALL1, FOXE1, LRP8, ID3, PRDM1, SMARCA2, KLF4, NFIA, RERE, PEG3	2	3.16E-05
GO:0010628	positive regulation of gene expression	CAV1, EPHX2, GJA1, TLE1, CDK6, KIT, CALR, CITED2, ANK2, GSN, CTGF, ID1, MAPK14, CYP26B1, NTRK2, RGCC, PRDM1, NFIL3, KLF4	3	4.33E-05
GO:0030198	extracellular matrix organization	B4GALT1, FBN1, OLFML2A, CDH1, ITGB3, KDR, CSGALNACT1, SMOC2, ITGA9, TNFRSF11B, COL9A3, ERO1B, FBLN5, JAM2, LAMB1, CYR61	4	5.55E-05
GO:0043434	response to peptide hormone	BTG2, GNAI1, CTGF, TEK, ANXA1, GJA1, CXCL12, IRS1	8	6.31E-05
GO:0071276	cellular response to cadmium ion	MT1E, MT1H, MT1X, MT1G, MT1F	13	5.36E-04

(b)

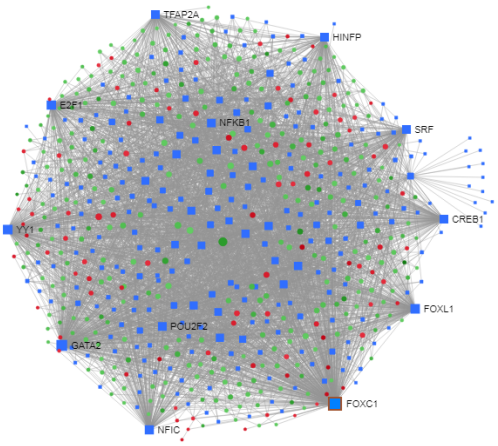


Fig. 4: Construction of regulatory networks of TFs-DEGs interaction. Red nodes indicate up-regulation and green node represents down-regulated DEGs.

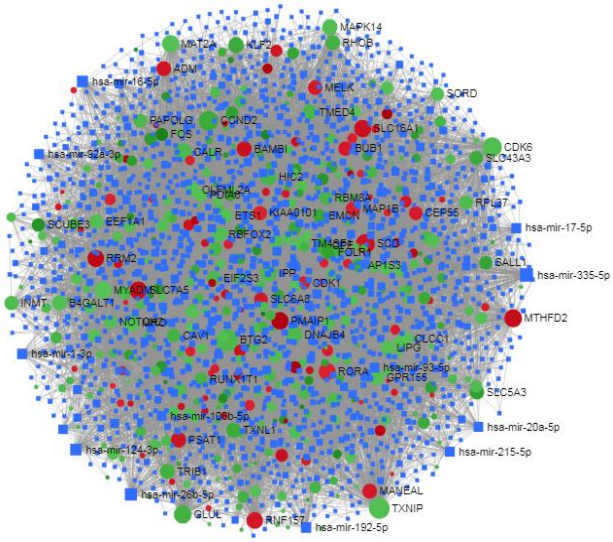


Fig. 5: Construction of regulatory networks of DEGs-miRNAs interaction. Red nodes indicate up-regulation and green node represents down-regulated DEGs.

adhesion, Sphingolipid metabolism, PPAR signaling pathway, and ECM-receptor interaction signaling system statistically significant (Table IV (a))

IV. Discussion:

To clarify the molecular mechanisms that underlie the pathological features TC, differential gene expression between TC lesions and benign lesions were examined. Note that a strength of this dataset is that comparing malignant follicular TC with benign follicular adenoma is more meaningful than comparing TC to the more complex normal thyroid tissue where follicular cells form only a minority of cells. The DEGs identified were used to identify possible regulatory patterns, molecular key pathways, and protein-protein interactions. This approach can be used to find molecular signatures that may serve either as potential therapeutic targets or as biomarkers to differentially diagnose or identify TC. We identified a number of significantly altered genes that are candidate signature genes for TC. However, another important use of such data is to identify and characterize biological functions associated with these genes that may give insights into the biology and

TABLE III: (a) The enrichment analysis of the subnetwork module-1 (b) The enrichment analysis of the subnetworks module-2 and module-3

Module	Category	Term	P-value	Genes
Module-1	Biological Process	ATP binding	1.00E-13	CDK1, KIF11, CCNB2, NEK2, BUB1, CENPF, BUB1B, NDC80, ANLN, PBK, CEP55, ASPM
		Cell division	1.28E-10	CDK1, KIF11, MAD2L1, CCNB2, NEK2, CKS2, BUB1, CENPF, BUB1B, NDC80, UBE2C
		Sister chromatid cohesion	6.92E-07	MAD2L1, BUB1, CENPF, BUB1B, NDC80, CENPK
		Cell proliferation	1.76E-06	CDK1, TYMS, DLGAP5, CKS2, BUB1, CENPF, BUB1B, MELK
		mitotic spindle assembly checkpoint	4.64E-06	MAD2L1, BUB1, CENPF, BUB1B
		chromosome segregation	4.68E-06	KIF11, NEK2, CENPF, NDC80, TOP2A
		mitotic sister chromatid segregation	9.32E-06	MAD2L1, NEK2, NUSAP1, NDC80
		G2/M transition of mitotic cell cycle	7.46E-05	CDK1, CCNB2, NEK2, MELK, HMMR
		negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	2.20E-04	CDK1, MAD2L1, BUB1B, UBE2C
		positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition	2.69E-04	CDK1, MAD2L1, BUB1B, UBE2C
	Cellular Component	Kinetochore	1.39E-07	MAD2L1, NEK2, BUB1, CENPF, BUB1B, NDC80
		Condensed chromosome kinetochore	1.99E-07	MAD2L1, NEK2, BUB1, BUB1B, NDC80, CENPK
		Nucleoplasm	1.74E-05	CDK1, DTL, EZH2, KIAA0101, CENPF, ANLN, UBE2C, CENPK, TYMS, CCNB2, FANCI, RRM2, BUB1, TOP2A, UBE2T
		Nucleus	3.38E-05	CDK1, DTL, NEK2, DLGAP5, EZH2, KIAA0101, NUSAP1, CENPF, NDC80, PBK, CDKN3, CENPK, TYMS, CCNB2, MAD2L1, RRM2, TOP2A, UBE2T, ASPM, MELK
		Midbody	4.30E-05	CDK1, NEK2, CENPF, CEP55, ASPM
		Cytoplasm	9.43E-05	CDK1, KIF11, DTL, NEK2, DLGAP5, EZH2, KIAA0101, NUSAP1, CENPF, UBE2C, CDKN3, TYMS, FANCI, RRM2, BUB1, BUB1B, TOP2A, UBE2T, ASPM
		Centrosome	4.30E-04	CDK1, CCNB2, DTL, NEK2, CENPF, CEP55
		Cytosol	5.65E-04	CDK1, KIF11, NEK2, CENPF, NDC80, UBE2C, CENPK, HMMR, TYMS, CCNB2, MAD2L1, RRM2, BUB1, BUB1B
		Spindle pole	6.11E-04	KIF11, MAD2L1, NEK2, CENPF
		Spindle microtubule	0.0020693	CDK1, KIF11, NUSAP1
	Molecular Function	Chromatin binding	3.63E-05	CDK1, EZH2, CKS2, KIAA0101, CENPF, TOP2A, UBE2T
		Protein kinase activity	2.78E-04	CDK1, NEK2, BUB1, BUB1B, PBK, MELK
		protein serine/threonine kinase activity	3.44E-04	CDK1, NEK2, BUB1, BUB1B, PBK, MELK
		ATP binding	4.78E-04	CDK1, KIF11, NEK2, BUB1, BUB1B, PBK, UBE2C, TOP2A, UBE2T, MELK
		Protein binding	9.40E-04	CDK1, DTL, NEK2, DLGAP5, EZH2, KIAA0101, NUSAP1, CENPF, NDC80, PBK, CEP55, UBE2C, CENPK, CDKN3, HMMR, MAD2L1, CCNB2, FANCI, RRM2, BUB1, CKS2, BUB1B, TOP2A, MELK
		Ubiquitin conjugating enzyme activity	4.70E-2	UBE2C, UBE2T

(a)

Module	Category	Term	Pvalue	Genes
Module-2	Biological Process	positive regulation of protein phosphorylation	6.67E-05	EGFR, CCND2, TEK, KDR
		peptidyl-tyrosine phosphorylation	1.16E-04	EGFR, BCR, TEK, KDR
		protein autophosphorylation	1.64E-04	EGFR, BCR, TEK, KDR
		positive regulation of cell proliferation	1.65E-04	EGFR, CCND2, IRS1, KDR, FN1
		positive regulation of ERK1 and ERK2 cascade	1.73E-04	EGFR, JUN, TEK, KDR
		positive regulation of GTPase activity	3.47E-04	EGFR, BCR, JUN, TEK, IRS1
		Angiogenesis	3.52E-04	JUN, TEK, KDR, FN1
		positive regulation of fibroblast proliferation	5.48E-04	EGFR, JUN, FN1
		positive regulation of endothelial cell proliferation	8.94E-04	JUN, TEK, KDR
		transmembrane receptor protein tyrosine kinase signaling pathway	1.72E-03	EGFR, TEK, KDR
	Molecular Function	protein tyrosine kinase activity	5.51E-05	EGFR, BCR, TEK, KDR
		transmembrane receptor protein tyrosine kinase activity	2.20E-04	EGFR, TEK, KDR
		enzyme binding	8.24E-04	EGFR, BCR, JUN, WWOX
		protein binding	1.45E-03	EGFR, BCR, CCND2, JUN, TEK, CDH1, TCF4, IRS1, WWOX, KDR, FN1
		integrin binding	1.67E-03	EGFR, KDR, FN1
		Ras guanyl-nucleotide exchange factor activity	2.00E-03	EGFR, TEK, IRS1
		receptor signaling protein tyrosine kinase activity	5.91E-03	EGFR, KDR
		identical protein binding	8.26E-03	EGFR, JUN, TCF4, FN1
		growth factor binding	1.59E-02	TEK, KDR
		chromatin binding	2.13E-02	EGFR, JUN, TCF4
	Cellular Component	membrane raft	6.54E-03	EGFR, TEK, KDR
		endosome	7.76E-03	EGFR, CDH1, KDR
		apical plasma membrane	1.27E-02	EGFR, TEK, FN1
		focal adhesion	2.22E-02	EGFR, TEK, CDH1
		cell junction	3.00E-02	BCR, CDH1, KDR
		microvillus	3.39E-02	TEK, WWOX
Module-3	Biological Process	protein ubiquitination	3.82E-05	SH3RF1, LMO7, FBXO32, FBXO9
	Cellular Component	ubiquitin ligase complex	2.07E-04	RNF144B, LMO7, FBXO9
		SCF ubiquitin ligase complex	1.09E-02	FBXO32, FBXO9
	Molecular Function	ubiquitin-protein transferase activity	2.89E-05	RNF144B, LMO7, FBXO32, FBXO9
		zinc ion binding	2.62E-02	SH3RF1, RNF144B, LMO7

(b)

behavior of TC lesions themselves. Disease genes play a central role in the human interactomes through the pathways. In this study, We found four important pathways that are associated with the thyroid cancer up-regulated genes and eight with the significantly down-regulated genes in TC. The p53 signaling pathway plays an important role in cancer cell apoptosis and DNA repair by causing cycle arrest [41]. Cell cycle progression is facilitated by cyclin-dependent kinases that are activated by cyclins including cyclin D1. The cyclin D1 expression may play a role in tumor progression and may have prognostic significance in TC [36]. One published study confirmed that thyroid cancer is actively associated with the p53 signaling pathway [10].Another study indicated that abnormal p53 activation is closely related to the development of TC [25]. It was seen that one-carbon metabolism pathways is actively involved in cancer [23].

We also found that nine important gene ontology groups are associated with the significantly up-regulated genes and eighty important gene ontology groups are associated with the significantly down-regulated genes for TC where top ten are shown in table II. To understand the central mechanisms behind the disease mechanisms, PPI network reconstruction and analysis of reconstructed PPI network represent a powerful approach. For this reason, to construct a PPI network around DEGs for our study, we have combined results of statistical analyses with the protein interactome network. For finding hub proteins, topological analysis strategies were employed. From the PPI network analysis, it was observed that 12 hub genes (Table I (b)) are involved in the TC and relevant diseases. Among the hub genes TOP2A is involved in the thyroid cancer, breast cancer, andon-small cell lung cancers [28]. The hub gene JUN is linked to different types of mitral valve disease (MVD), including mitral regurgitation (MR) and mitral stenosis (MS) [6]. From gene ontology, we found that JUN gene actively in Angiogenesis, positive regulation of endothelial cell proliferation (Table-II (b)). The hub protein EGFR is implicated in cardiovascular diseases and thyroid cancer [31] and is linked to growth in many types of cancers. The hub gene CDK1 involved in the breast Cancer and ovary carcinoma [18]. CDK1 overexpression has been documented in lung cancer, lymphoma and advanced melanoma and loss of cytoplasmic CDK1 predicts poor survival and confers chemotherapeutic resistance in the latter [27]. Cyclin B1 (in the same family as CDK1) overexpression and/or mislocalization has been described in several primary cancers including thyroid carcinoma, colon, gastric, prostate, breast, and non small-cell lung cancer [27]. It is also actively involved in mitotic cell cycle phase transition (table-II (a)). The hub gene FOS family

TABLE IV: (a) Top 10 KEGG pathway in genes of micro-RNAs and differentially expressed genes interaction networks (b) Top 10 KEGG pathway in genes of TFs and differentially expressed genes interaction networks

Pathway	Total	Expected	Hits	P.Value
p53 signaling pathway	68	2.25	8	1.65E-03
Dorso-ventral axis formation	12	0.397	3	6.29E-03
Pertussis	52	1.72	6	6.92E-03
Adherens junction	70	2.32	7	7.95E-03
Thyroid cancer	28	0.927	4	1.27E-02
Pathways in cancer	310	10.3	18	1.32E-02
Focal adhesion	200	6.62	13	1.45E-02
Sphingolipid metabolism	46	1.52	5	1.71E-02
PPAR signaling pathway	64	2.12	6	1.84E-02
ECM-receptor interaction	84	2.78	7	2.05E-02

(a)

Pathway	Total	Expected	Hits	P.Value
Transcriptional misregulation in cancer	19	0.942	11	1.87E-10
Acute myeloid leukemia	57	2.83	16	7.48E-09
Pathways in cancer	310	15.4	40	9.50E-09
Thyroid cancer	28	1.39	10	4.41E-07
HTLV-I infection	199	9.86	26	3.86E-06
Cell cycle	124	6.15	18	3.12E-05
Dorso-ventral axis formation	12	0.595	5	1.71E-04
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	12	0.595	5	1.71E-04
Cocaine addiction	43	2.13	9	1.99E-04
Prostate cancer	87	4.31	13	3.01E-04

(b)

is made up four members: c-FOS, FOSB, FOS-like antigen (FOSL) -1 and -2. Generally, the FOSL-2 is connected with the motility, cell adhesion, invasiveness and metastasis than cell growth. This gene is associated with the tumorigenesis of breast cancer, colorectal, and also associated with TC [16]. The hub gene CDKN3 was over expressed in lung adenocarcinoma (ADC), squamous cell carcinoma (SCC), hepatocellular carcinoma, cervical cancer and epithelial ovarian cancer [40]. Over expression of EZH2 is frequently observed in multiple cancer types, including gastric, esophageal, bladder, prostate, breast, ovarian, pancreatic, kidney, liver, brain, and lung cancer and melanoma [39]. The gene TYMS is associated with risk of development of breast cancer, colorectal, lung, pancreatic, gastric, and lymphoma cancer [14]. The gene PBK was found to be over-expressed in haematological and breast tumours and in poorly differentiated thyroid carcinomas (PDTC) cell lines [29]. Besides Hereditary diffuse gastric cancer (HDGC), The hub gene CDH1 mutations are connected with an increased incidence of lobular carcinoma of the breast and, possibly, prostate cancer and colorectal carcinoma [26]. This gene is also involved in extracellular matrix organization (Table-II (b)). Previous studies have reported overexpression of UBE2C in some cancers such as colorectal carcinoma, pancreatic carcinoma, cervical cancer, bladder carcinoma, esophageal squamous cell carcinoma and lung cancer, etc. [19]. Hub gene CCNB2 is known to be underexpressed in thyroid cell tumors [8]. We found that CDK1 and CCNB2 hubs genes actively involved in p53 signaling pathway, Progesterone-mediated oocyte maturation and Cell cycle pathways. The hub gene TYMS involved in One carbon pool by folate pathway, FOS, JUN and CDH1 genes involved in Pathways in cancer and JUN, FOS genes in Estrogen signaling pathway. By considering the possible roles of these hub proteins in pathogenesis of TC and other related diseases, we suggest further work may identify new roles for these proteins.

As we know that the regulation of gene expression is controlled by TFs and miRNAs at post-transcriptional and/or transcriptional levels, so changes in these molecules may provide crucial information on dysregulation of gene expression in the TC. Therefore, reporter miRNAs and TFs, and their target DEGs have been found out.

Reporter TFs and their relationship to DEGs were analysed ans the top ten TFs selected (Table I (b)). One previous study evaluated that FOXC1 strongly associated with TC [5]. One study found the overexpression of YY1 in differentiated thyroid cancer [2]. The pathway analysis of this TFs regulated DEGs were statistically significant with TC and relevant diseases. From the DEGs- miRNA interaction network top ten miRNA (Table I(b)) were selected and analysed. The pathway analysis of this miRNA regulated DEGs were statistically significant with the thyroid cancer and relevant diseases. It was obtained that the target DEGs of the reporter miRNAs and TFs were involved in p53 signaling pathway, PPAR signaling pathway, Dorso-ventral axis formation, Focal adhesion (Table-IV(a)) and Pathways in cancer, Transcriptional misregulation in cancer,Thyroid cancer, respectively.

V. CONCLUSIONS

Our data-driven approaches are useful for the identification of disease mechanism in a quantitative way. In this study, using integrated bioinformatics analyses, we analyzed gene expression profiles in the TC. Candidate hub proteins identified in this study may play significant roles in the pathogenesis of TC, and indicate focuses for future experimental work to clarify the roles of these proteins. This study provides networks between the genes for understanding the pathogenic properties of TC and maps significant pathways for the future development of novel therapeutic strategies. Our study thus identifies new perspectives in the molecular mechanism for TC that may lead to new potential therapeutic development.

ACKNOWLEDGMENTS

The work was supported by the KARTE ($https://www.karte-health.com/home$) and the University of Sydney Fellowship award.

REFERENCES

[1] Thyroid cancer: Statistics— Cancer. net, 2017. [Online; accessed 10-06-2018].

[2] J. Arribas, J. Castellví, R. Marcos, C. Zafón, and A. Velázquez. Expression of yy1 in differentiated thyroid cancer. *Endocrine pathology*, 26(2):111–118, 2015.

[3] A. Bansal and J. Ramana. Tcgdb: a compendium of molecular signatures of thyroid cancer and disorders. *J. Cancer Sci. Ther*, 7:198–201, 2015.

[4] T. Barrett, D. B. Troup, S. E. Wilhite, P. Ledoux, D. Rudnev, C. Evangelista, I. F. Kim, A. Soboleva, M. Tomashevsky, and R. Edgar. Ncbi geo: mining tens of millions of expression profilesdatabase and tools update. *Nucleic acids research*, 35(suppl_1):D760–D765, 2006.

[5] M. Bullock, G. Lim, C. Li, I. H. Choi, S. Kochhar, C. Liddle, L. Zhang, and R. J. Clifton-Bligh. Thyroid transcription factor foxe1 interacts with ets factor elk1 to co-regulate tert. *Oncotarget*, 7(52):85948, 2016.

[6] C. Chang, C. Zhang, X. Zhao, X. Kuang, H. Tang, and X. Xiao. Differential regulation of mitogen-activated protein kinase signaling pathways in human with different types of mitral valvular disease. *Journal of Surgical Research*, 181(1):49–59, 2013.

[7] S.-H. Chen, C.-H. Chin, H.-H. Wu, C.-W. Ho, M.-T. Ko, and C.-Y. Lin. cyto-hubba: A cytoscape plug-in for hub object analysis in network biology. In *20th International Conference on Genome Informatics*, 2009.

[8] C. González, E. Mato, O. Bell, E. Lerma, A. Moral, J. Perez, et al. Predictive value of molecular markers that involved in a genetic signature in thyroid tissue and their involvement in the tumour dedifferentiation process.

[9] Y. Guo, Y. Bao, M. Ma, and W. Yang. Identification of key candidate genes and pathways in colorectal cancer by integrated bioinformatical analysis. *International journal of molecular sciences*, 18(4):722, 2017.

[10] W. He, B. Qi, Q. Zhou, C. Lu, Q. Huang, L. Xian, and M. Chen. Key genes and pathways in thyroid cancer based on gene set enrichment analysis. *Oncology reports*, 30(3):1391–1397, 2013.

[11] S.-D. Hsu, F.-M. Lin, W.-Y. Wu, C. Liang, W.-C. Huang, W.-L. Chan, W.-T. Tsai, G.-Z. Chen, C.-J. Lee, C.-M. Chiu, et al. mirtarbase: a database curates experimentally validated microrna–target interactions. *Nucleic acids research*, 39(suppl_1):D163–D169, 2010.

[12] Y. Huang, M. Prasad, W. J. Lemon, H. Hampel, F. A. Wright, K. Kornacker, V. LiVolsi, W. Frankel, R. T. Kloos, C. Eng, et al. Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. *Proceedings of the National Academy of Sciences*, 98(26):15044–15049, 2001.

[13] K.-W. Jung, Y.-J. Won, H.-J. Kong, C.-M. Oh, H. G. Seo, and J.-S. Lee. Cancer statistics in korea: incidence, mortality, survival and prevalence in 2010. *Cancer research and treatment: official journal of Korean Cancer Association*, 45(1):1, 2013.

[14] J. d. S. N. Junior, F. A. de Lima Marson, and C. S. Bertuzzo. Thymidylate synthase gene (tyms) polymorphisms in sporadic and hereditary breast cancer. *BMC research notes*, 5(1):676, 2012.

[15] A. Khan, O. Fornes, A. Stigliani, M. Gheorghe, J. A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Cheneby, S. R. Kulkarni, G. Tan, et al. Jaspar 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic acids research*, 46(D1):D260–D266, 2017.

[16] S. W. Kim, C. D. Kim, J. H. Chung, and K. H. Kwon. Association study of fos-like antigen-2 promoter polymorphisms with papillary thyroid cancer in korean population. *Clinical and experimental otorhinolaryngology*, 7(1):42, 2014.

[17] M. Kori, E. Gov, and K. Y. Arga. Molecular signatures of ovarian diseases: Insights from network medicine perspective. *Systems biology in reproductive medicine*, 62(4):266–282, 2016.

[18] Z. Liu, G. Liang, L. Tan, A. Su, W. Jiang, and C. Gong. High-efficient screening method for identification of key genes in breast cancer through microarray and bioinformatics. *Anticancer research*, 37(8):4329–4335, 2017.

[19] C.-h. Mo, L. Gao, X.-f. Zhu, K.-l. Wei, J.-j. Zeng, G. Chen, and Z.-b. Feng. The clinicopathological significance of ube2c in breast cancer: a study based on immunohistochemistry, microarray and rna-sequencing data. *Cancer cell international*, 17(1):83, 2017.

[20] M. A. Moni and P. Liò. comor: a software for disease comorbidity risk assessment. *Journal of clinical bioinformatics*, 4(1):1, 2014.

[21] M. A. Moni and P. Liò. Network-based analysis of comorbidities risk during an infection: Sars and hiv case studies. *BMC bioinformatics*, 15(1):1, 2014.

[22] M. A. Moni and P. Liò. How to build personalized multi-omics comorbidity profiles. *Frontiers in cell and developmental biology*, 3, 2015.

[23] A. C. Newman and O. D. Maddocks. One-carbon metabolism in cancer. *British journal of cancer*, 116(12):1499, 2017.

[24] Q. T. Nguyen, E. J. Lee, M. G. Huang, Y. I. Park, A. Khullar, and R. A. Plodkowski. Diagnosis and treatment of patients with thyroid cancer. *American health & drug benefits*, 8(1):30, 2015.

[25] Y. Ni, X. He, J. Chen, J. Moline, J. Mester, M. S. Orloff, M. D. Ringel, and C. Eng. Germline sdhx variants modify breast and thyroid cancer risks in cowden and cowden-like syndrome via fad/nad-dependant destabilization of p53. *Human molecular genetics*, 21(2):300–310, 2011.

[26] J. A. Norton, C. M. Ham, R. Jacques Van Dam, T. A. Longacre, D. G. Huntsman, N. Chun, A. W. Kurian, and J. M. Ford. Cdh1 truncating mutations in the e-cadherin gene: an indication for total gastrectomy to treat hereditary diffuse gastric cancer. *Annals of surgery*, 245(6):873, 2007.

[27] M. Peyressatre, C. Prével, M. Pellerano, and M. C. Morris. Targeting cyclin-dependent kinases in human cancers: from small molecules to peptide inhibitors. *Cancers*, 7(1):179–237, 2015.

[28] J. Piao, J. Sun, Y. Yang, T. Jin, L. Chen, and Z. Lin. Target gene screening and evaluation of prognostic values in non-small cell lung cancers by bioinformatics analysis. *Gene*, 2018.

[29] J. Pita, A. Banito, B. Cavaco, and V. Leite. Gene expression profiling associated with the progression to poorly differentiated thyroid carcinomas. *British journal of cancer*, 101(10):1782, 2009.

[30] R. Ria, V. Simeon, A. Melaccio, G. Di Meo, S. Trino, C. Mazzoccoli, I. Saltarella, A. Lamanuzzi, A. Morano, A. Gurrado, et al. Gene expression profiling of normal thyroid tissue from patients with thyroid carcinoma. *Oncotarget*, 7(20):29677, 2016.

[31] A. Sarajlić, V. Janjić, N. Stojković, D. Radak, and N. Pržulj. Network topology reveals key cardiovascular disease genes. *PloS one*, 8(8):e71537, 2013.

[32] P. Sethupathy, B. Corda, and A. G. Hatzigeorgiou. Tarbase: A comprehensive database of experimentally supported animal microrna targets. *Rna*, 12(2):192–197, 2006.

[33] M. E. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang, and T. Ideker. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, 27(3):431–432, 2010.

[34] G. Smyth, M. Ritchie, and N. Thorne. Linear models for microarray data users guide. *Bioinformatics*, 20:3705–3706, 2011.

[35] D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, et al. The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*, page gkw937, 2016.

[36] S. Wang, R. V. Lloyd, M. J. Hutzler, M. S. Safran, N. A. Patwardhan, and A. Khan. The role of cell cycle regulatory protein, cyclin d1, in the progression of thyroid cancer. *Modern Pathology*, 13(8):882, 2000.

[37] B. Wojtas, A. Pfeifer, M. Oczko-Wojciechowska, J. Krajewska, A. Czarniecka, A. Kukulska, M. Eszlinger, T. Musholt, T. Stokowy, M. Swierniak, et al. Gene expression (mrna) markers for differentiating between malignant and benign follicular thyroid tumours. *International journal of molecular sciences*, 18(6):1184, 2017.

[38] J. Xia, E. E. Gill, and R. E. Hancock. Networkanalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nature protocols*, 10(6):823, 2015.

[39] H. Yamaguchi and M.-C. Hung. Regulation and role of ezh2 in cancer. *Cancer research and treatment: official journal of Korean Cancer Association*, 46(3):209, 2014.

[40] X. Zang, M. Chen, Y. Zhou, G. Xiao, Y. Xie, and X. Wang. Identifying cdkn3 gene expression as a prognostic biomarker in lung adenocarcinoma via meta-analysis. *Cancer informatics*, 14:CIN–S17287, 2015.

[41] W. Zhu, H. Hu, P. Qiu, and G. Yan. Triptolide induces apoptosis in human anaplastic thyroid carcinoma cells by a p53-independent but nf-κb-related mechanism. *Oncology reports*, 22(6):1397–1401, 2009.

A.

KEGG ID	Pathway	Genes in the pathway	Fold Enrichment	Adjusted p-value
hsa00670	One carbon pool by folate	MTHFD2, TYMS, SHMT2, ALDH1L1	24	5.21E-04
hsa04115	p53 signaling pathway	CDK1, CCNB2, RRM2, PMAIP1, IGFBP3	9	2.03E-03
hsa04110	Cell cycle	CDK1, MAD2L1, CCNB2, BUB1, BUB1B	5	1.77E-02
hsa04914	Progesterone-mediated oocyte maturation	CDK1, MAD2L1, CCNB2, BUB1	6	3.31E-02

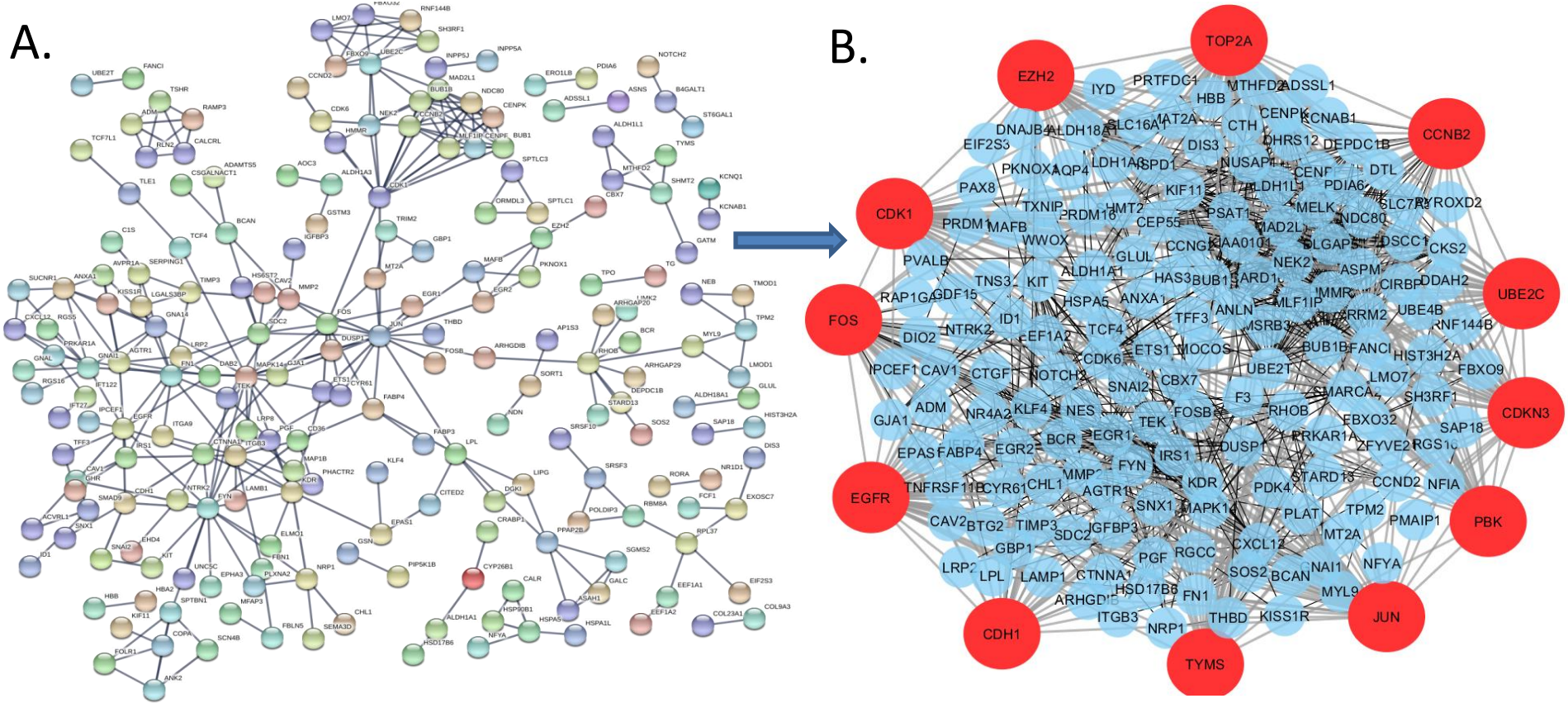
B.

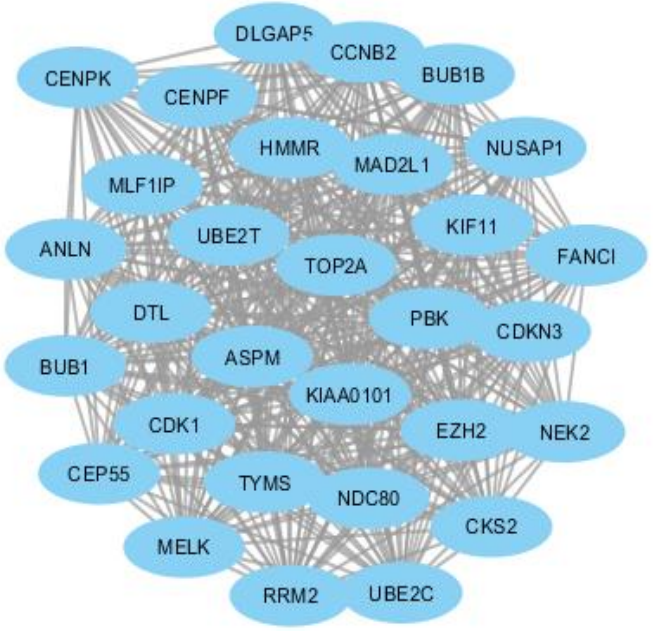
KEGG ID	Pathway	Genes in the pathway	Fold Enrichment	Adjusted p-value
hsa04978	Mineral absorption	MT1M, MT2A, CYBRD1, MT1E, MT1H, MT1X, MT1G, MT1F	6	3.44E-04
hsa04918	Thyroid hormone synthesis	SLC26A4, TG, HSP90B1, PAX8, TPO, HSPA5, LRP2, TSHR, IYD	4	9.55E-04
hsa05200	Pathways in cancer	FGFR2, BCR, EPAS1, PGF, RXRB, GNAI1, RUNX1T1, FZD1, CDH1, CDK6, KIT, CTNNA1, MMP2, CXCL12, TCF7L1, AGTR1, FOS, HSP90B1, JUN, PAX8, SOS2, LAMB1	2	5.48E-03
hsa04915	Estrogen signaling pathway	HSPA1L, FOS, HSP90B1, GNAI1, JUN, SOS2, HSPA1A, HSPA1B, MMP2	3	8.41E-03
hsa00600	Sphingolipid metabolism	SPTLC1, SGMS2, SPTLC3, GALC, PLPP3, ASAH1	4	1.17E-02
hsa04360	Axon guidance	NRP1, LIMK2, SEMA6D, FYN, GNAI1, PLXNA2, SEMA3D, UNC5C, CXCL12, EPHA3	3	1.22E-02
hsa04141	Protein processing in endoplasmic reticulum	HSPA1L, HSP90B1, ERO1B, UBE4B, PDIA6, HSPA1A, HSPA1B, HSPA5, CALR, MAN1C1, SEC62	2	2.68E-02
hsa04510	Focal adhesion	ITGA9, CAV2, CAV1, CCND2, FYN, PGF, JUN, SOS2, ITGB3, LAMB1, KDR, MYL9	2	3.95E-02

Symbol	Description	Feature
UBA52	DNA topoisomerase 2-alpha	Hub protein
JUN	c-Jun in combination with c-Fos	Hub protein
EGFR	The epidermal growth factor receptor	Hub protein
CDK1	Cyclin Dependent Kinase 1	Hub protein
FOS	proto-oncogene	Hub protein
CDKN3	Cyclin Dependent Kinase Inhibitor 3	Hub protein
EZH2	Enhancer of zeste homolog 2	Hub protein
TYMS	Thymidylate Synthetase	Hub protein
PBK	PDZ binding kinase	Hub protein
CDH1	Cadherin-1	Hub protein
UBE2C	Ubiquitin Conjugating Enzyme E2 C	Hub protein
CCNB2	G2/Mitotic-Specific Cyclin-B2	Hub protein
FOXC1	Forkhead Box C1	Reporter Transcription Factor
GATA2	GATA Binding Protein 2	Reporter Transcription Factor
YY1	YY1 Transcription Factor	Reporter Transcription Factor
FOXL1	Forkhead Box L1	Reporter Transcription Factor
NFIC	Nuclear Factor I C	Reporter Transcription Factor
E2F1	E2F Transcription Factor 1	Reporter Transcription Factor
TFAP2A	Activating enhancer binding Protein 2 alpha	Reporter Transcription Factor
SRF	Serum Response Factor	Reporter Transcription Factor
HINFP	Histone H4 transcription factor	Reporter Transcription Factor
CREB1	Cyclic AMP-responsive element-binding protein 1, CREB transcription factor	Reporter Transcription Factor
hsa-mir-335-5p	MicroRNA 355	Reporter microRNA
hsa-mir-26b-5p	MicroRNA 26b	Reporter microRNA
hsa-mir-124-3p	MicroRNA 124	Reporter microRNA
hsa-mir-16-5p	MicroRNA 16	Reporter microRNA
hsa-mir-192-5p	MicroRNA 192	Reporter microRNA
hsa-mir-1-3p	MicroRNA 11	Reporter microRNA
hsa-mir-17-5p	MicroRNA 17	Reporter microRNA
hsa-mir-92a-3p	MicroRNA 92a	Reporter microRNA
hsa-mir-215-5p	MicroRNA 215	Reporter microRNA
hsa-mir-20a-5p	MicroRNA 20a	Reporter microRNA

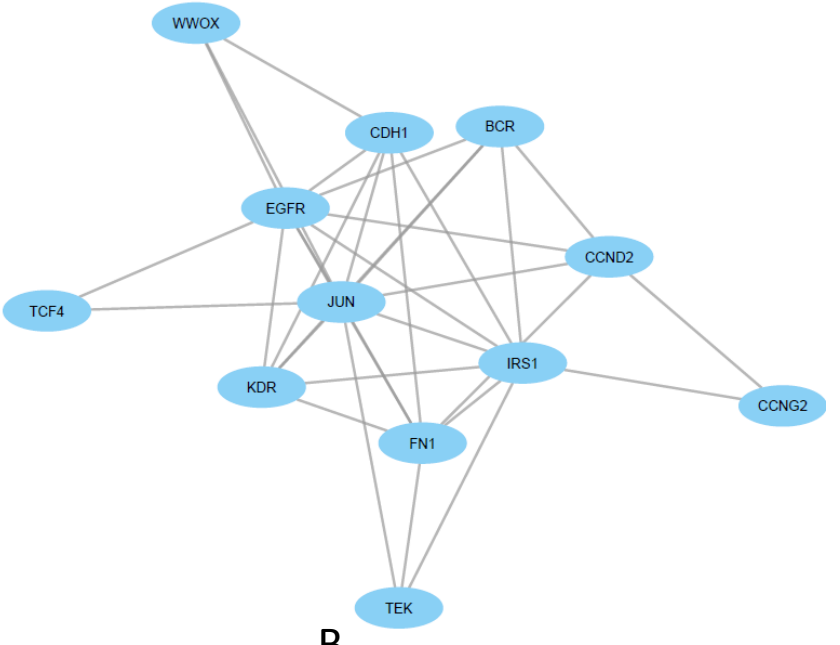
Go Term	Pathway	Genes in the pathway	Fold Enrichment	Adjusted p-value
GO:0045926	negative regulation of growth	ING5, MT1M, MT2A, MT1E, MT1H, MT1X, MT1G, MT1F	18	1.37E-07
GO:0071294	cellular response to zinc ion	MT1M, MT2A, MT1E, MT1H, MT1X, MT1G, MT1F	16	3.17E-06
GO:0006590	thyroid hormone generation	DIO2, CPQ, FOXE1, TPO, DIO1, IYD	21	4.56E-06
GO:0001525	angiogenesis	FGFR2, EMCN, CAV1, NRP1, ACVRL1, EPAS1, PGF, MMP2, KDR, PKNOX1, CTGF, ID1, JUN, MAPK14, TEK, ECSCR, RHOB, CALCRL	3	1.89E-05
GO:0001938	positive regulation of endothelial cell proliferation	CAV2, NRP1, ACVRL1, PGF, F3, JUN, TEK, ITGB3, CXCL12, KDR	6	2.97E-05
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	FGFR2, GLIS3, TSHZ3, ZMYND11, CAV1, TFCP2L1, ZNF366, PRDM16, CALR, CBX7, CITED2, MINA, TCF4, BHLHE41, NFIL3, NR2F1, TXNIP, EGR1, NR4A2, TLE1, FOSB, SNAI2, FOXP2, CD36, BTG2, ID1, SALL1, FOXE1, LRP8, ID3, PRDM1, SMARCA2, KLF4, NFIA, RERE, PEG3	2	3.16E-05
GO:0010628	positive regulation of gene expression	CAV1, EPHX2, GJA1, TLE1, CDK6, KIT, CALR, CITED2, ANK2, GSN, CTGF, ID1, MAPK14, CYP26B1, NTRK2, RGCC, PRDM1, NFIL3, KLF4	3	4.33E-05
GO:0030198	extracellular matrix organization	B4GALT1, FBN1, OLFML2A, CDH1, ITGB3, KDR, CSGALNACT1, SMOC2, ITGA9, TNFRSF11B, COL9A3, ERO1B, FBLN5, JAM2, LAMB1, CYR61	4	5.55E-05
GO:0043434	response to peptide hormone	BTG2, GNAI1, CTGF, TEK, ANXA1, GJA1, CXCL12, IRS1	8	6.31E-05
GO:0071276	cellular response to cadmium ion	MT1E, MT1H, MT1X, MT1G, MT1F	13	5.36E-04

Go Term	Pathway	Genes in the pathway	Fold Enric-hment	Adjusted p-value
GO:0006636	unsaturated fatty acid biosynthetic process	ELOVL4, SCD, ELOVL7	45	1.88E-03
GO:0007094	mitotic spindle assembly checkpoint	MAD2L1, BUB1, BUB1B	33	3.39E-03
GO:0045773	positive regulation of axon extension	TNFRSF12A, MAP1B, FN1	24	6.83E-03
GO:0044772	mitotic cell cycle phase transition	CDK1, CKS2	67	2.92E-02
GO:0046602	regulation of mitotic centrosome separation	KIF11, NEK2	67	2.92E-02
GO:0021930	cerebellar granule cell precursor proliferation	ATF5, RORA	45	4.35E-02
GO:0035999	tetrahydrofolate interconversion	TYMS, SHMT2	45	4.35E-02
GO:0035338	long-chain fatty-acyl-CoA biosynthetic process	ELOVL4, ELOVL7	45	4.35E-02
GO:0046777	protein autophosphorylation	EGFR, NEK2, STK26, MELK	5	4.78E-02

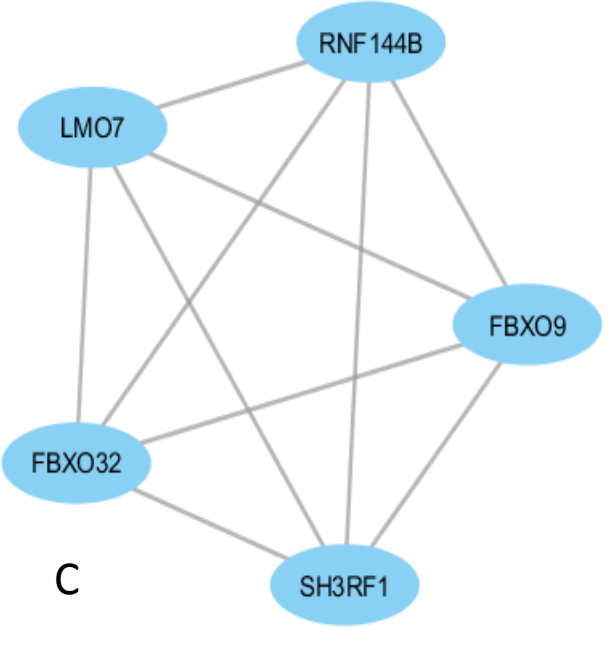

A.
B.
TOP2A
EZH2
CCNB2
CDK1
UBE2C
FOS
CDKN3
EGFR
PBK
CDH1
JUN
TYMS


DLGAP5
CCNB2
CENPK
BUB1B
CENPF
HMMR
NUSAP1
MAD2L1
MLF1IP
KIF11
UBE2T
ANLN
FANCI
TOP2A
DTL
PBK
CDKN3
BUB1
ASPM
KIAA0101
CDK1
EZH2
NEK2
CEP55
TYMS
NDC80
CKS2
MELK
RRM2
UBE2C

A


WWOX
CDH1
BCR
EGFR
CCND2
TCF4
JUN
IRS1
KDR
CCNG2
FN1
TEK

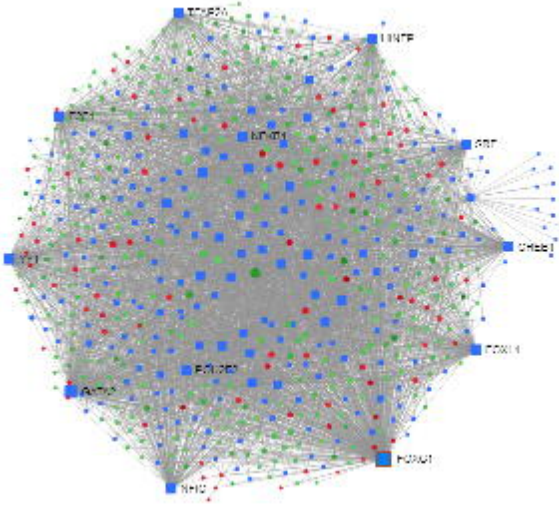
B

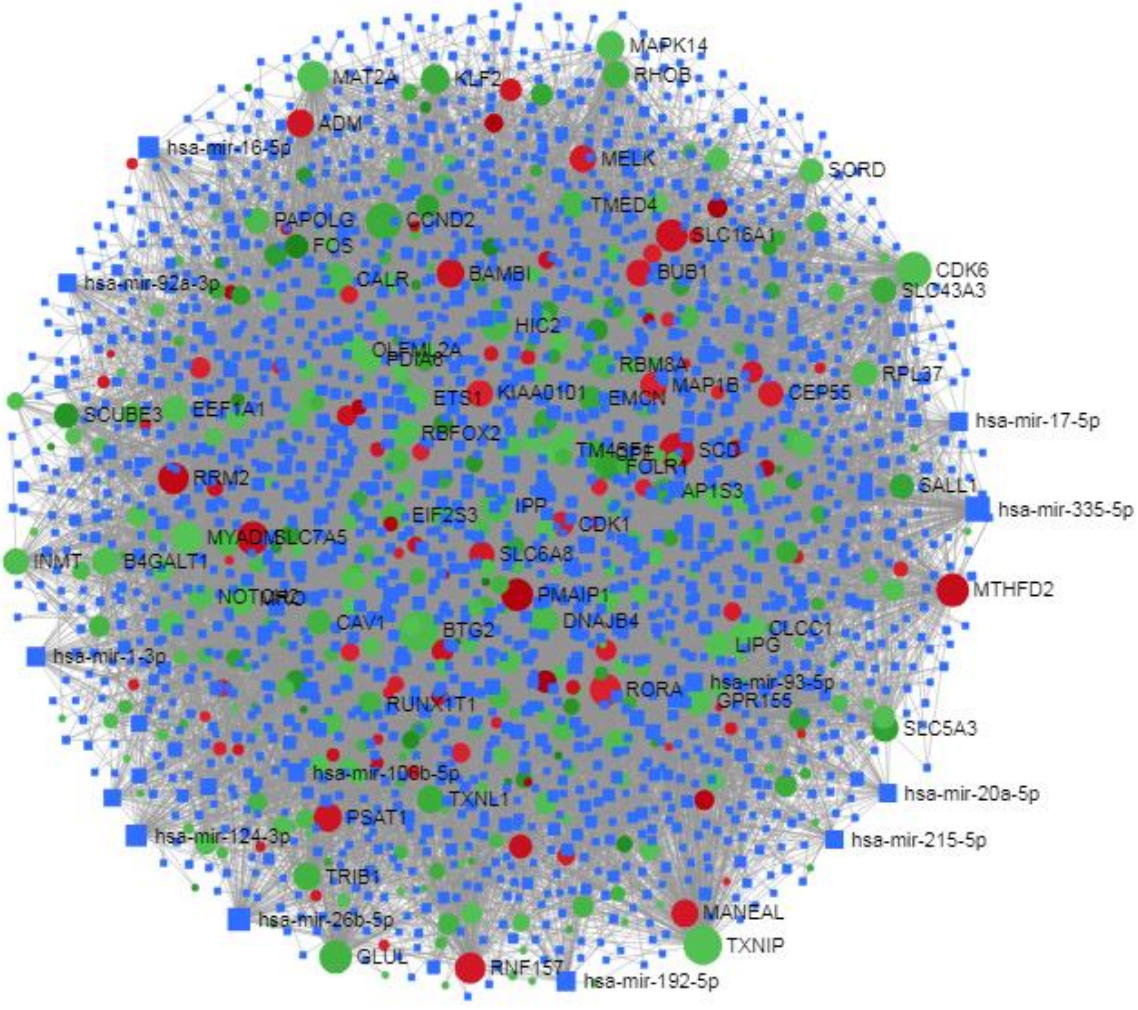

RNF144B
LMO7
FBXO9
FBXO32
SH3RF1

C

Module	Category	Term	P-value	Genes
Module-1	Biological Process	ATP binding	1.00E-13	CDK1, KIF11, CCNB2, NEK2, BUB1, CENPF, BUB1B, NDC80, ANLN, PBK, CEP55, ASPM
		Cell division	1.28E-10	CDK1, KIF11, MAD2L1, CCNB2, NEK2, CKS2, BUB1, CENPF, BUB1B, NDC80, UBE2C
		Sister chromatid cohesion	6.92E-07	MAD2L1, BUB1, CENPF, BUB1B, NDC80, CENPK
		Cell proliferation	1.76E-06	CDK1, TYMS, DLGAP5, CKS2, BUB1, CENPF, BUB1B, MELK
		mitotic spindle assembly checkpoint	4.64E-06	MAD2L1, BUB1, CENPF, BUB1B
		chromosome segregation	4.68E-06	KIF11, NEK2, CENPF, NDC80, TOP2A
		mitotic sister chromatid segregation	9.32E-06	MAD2L1, NEK2, NUSAP1, NDC80
		G2/M transition of mitotic cell cycle	7.46E-05	CDK1, CCNB2, NEK2, MELK, HMMR
		negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	2.20E-04	CDK1, MAD2L1, BUB1B, UBE2C
		positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition	2.69E-04	CDK1, MAD2L1, BUB1B, UBE2C
	Cellular Component	Kinetochore	1.39E-07	MAD2L1, NEK2, BUB1, CENPF, BUB1B, NDC80
		Condensed chromosome kinetochore	1.99E-07	MAD2L1, NEK2, BUB1, BUB1B, NDC80, CENPK
		Nucleoplasm	1.74E-05	CDK1, DTL, EZH2, KIAA0101, CENPF, ANLN, UBE2C, CENPK, TYMS, CCNB2, FANCI, RRM2, BUB1, TOP2A, UBE2T
		Nucleus	3.38E-05	CDK1, DTL, NEK2, DLGAP5, EZH2, KIAA0101, NUSAP1, CENPF, NDC80, PBK, CDKN3, CENPK, TYMS, CCNB2, MAD2L1, RRM2, TOP2A, UBE2T, ASPM, MELK
		Midbody	4.30E-05	CDK1, NEK2, CENPF, CEP55, ASPM
		Cytoplasm	9.43E-05	CDK1, KIF11, DTL, NEK2, DLGAP5, EZH2, KIAA0101, NUSAP1, CENPF, UBE2C, CDKN3, TYMS, FANCI, RRM2, BUB1, BUB1B, TOP2A, UBE2T, ASPM
		Centrosome	4.30E-04	CDK1, CCNB2, DTL, NEK2, CENPF, CEP55
		Cytosol	5.65E-04	CDK1, KIF11, NEK2, CENPF, NDC80, UBE2C, CENPK, HMMR, TYMS, CCNB2, MAD2L1, RRM2, BUB1, BUB1B
		Spindle pole	6.11E-04	KIF11, MAD2L1, NEK2, CENPF
		Spindle microtubule	0.0020693	CDK1, KIF11, NUSAP1
	Molecular Function	Chromatin binding	3.63E-05	CDK1, EZH2, CKS2, KIAA0101, CENPF, TOP2A, UBE2T
		Protein kinase activity	2.78E-04	CDK1, NEK2, BUB1, BUB1B, PBK, MELK
		protein serine/threonine kinase activity	3.44E-04	CDK1, NEK2, BUB1, BUB1B, PBK, MELK
		ATP binding	4.78E-04	CDK1, KIF11, NEK2, BUB1, BUB1B, PBK, UBE2C, TOP2A, UBE2T, MELK
		Protein binding	9.40E-04	CDK1, DTL, NEK2, DLGAP5, EZH2, KIAA0101, NUSAP1, CENPF, NDC80, PBK, CEP55, UBE2C, CENPK, CDKN3, HMMR, MAD2L1, CCNB2, FANCI, RRM2, BUB1, CKS2, BUB1B, TOP2A, MELK
		Ubiquitin conjugating enzyme activity	4.70E-2	UBE2C, UBE2T

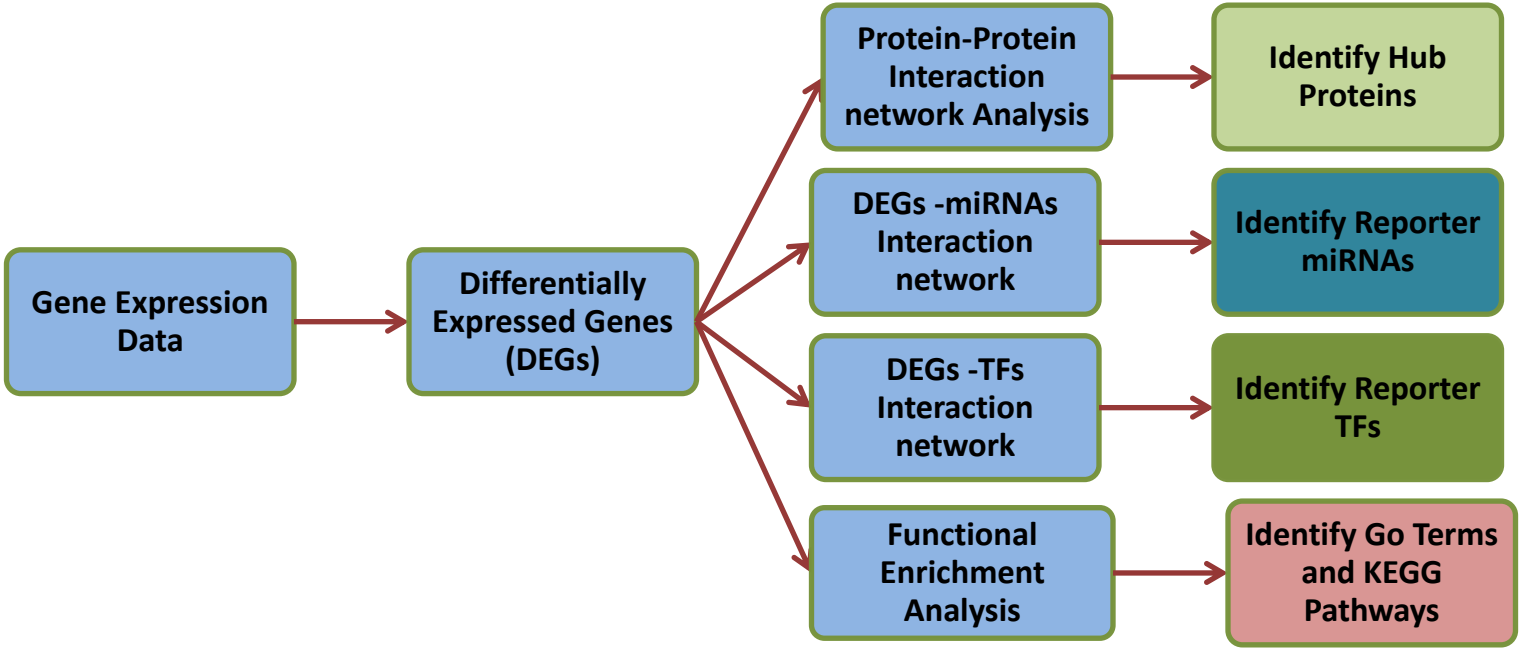
Module	Category	Term	Pvalue	Genes
Module-2	Biological Process	positive regulation of protein phosphorylation	6.67E-05	EGFR, CCND2, TEK, KDR
		peptidyl-tyrosine phosphorylation	1.16E-04	EGFR, BCR, TEK, KDR
		protein autophosphorylation	1.64E-04	EGFR, BCR, TEK, KDR
		positive regulation of cell proliferation	1.65E-04	EGFR, CCND2, IRS1, KDR, FN1
		positive regulation of ERK1 and ERK2 cascade	1.73E-04	EGFR, JUN, TEK, KDR
		positive regulation of GTPase activity	3.47E-04	EGFR, BCR, JUN, TEK, IRS1
		Angiogenesis	3.52E-04	JUN, TEK, KDR, FN1
		positive regulation of fibroblast proliferation	5.48E-04	EGFR, JUN, FN1
		positive regulation of endothelial cell proliferation	8.94E-04	JUN, TEK, KDR
		transmembrane receptor protein tyrosine kinase signaling pathway	1.72E-03	EGFR, TEK, KDR
	Molecular Function	protein tyrosine kinase activity	5.51E-05	EGFR, BCR, TEK, KDR
		transmembrane receptor protein tyrosine kinase activity	2.20E-04	EGFR, TEK, KDR
		enzyme binding	8.24E-04	EGFR, BCR, JUN, WWOX
		protein binding	1.45E-03	EGFR, BCR, CCND2, JUN, TEK, CDH1, TCF4, IRS1, WWOX, KDR, FN1
		integrin binding	1.67E-03	EGFR, KDR, FN1
		Ras guanyl-nucleotide exchange factor activity	2.00E-03	EGFR, TEK, IRS1
		receptor signaling protein tyrosine kinase activity	5.91E-03	EGFR, KDR
		identical protein binding	8.26E-03	EGFR, JUN, TCF4, FN1
		growth factor binding	1.59E-02	TEK, KDR
		chromatin binding	2.13E-02	EGFR, JUN, TCF4
	Cellular Component	membrane raft	6.54E-03	EGFR, TEK, KDR
		endosome	7.76E-03	EGFR, CDH1, KDR
		apical plasma membrane	1.27E-02	EGFR, TEK, FN1
		focal adhesion	2.22E-02	EGFR, TEK, CDH1
		cell junction	3.00E-02	BCR, CDH1, KDR
		microvillus	3.39E-02	TEK, WWOX
Module-3	Biological Process	protein ubiquitination	3.82E-05	SH3RF1, LMO7, FBXO32, FBXO9
	Cellular Component	ubiquitin ligase complex	2.07E-04	RNF144B, LMO7, FBXO9
		SCF ubiquitin ligase complex	1.09E-02	FBXO32, FBXO9
	Molecular Function	ubiquitin-protein transferase activity	2.89E-05	RNF144B, LMO7, FBXO32, FBXO9
		zinc ion binding	2.62E-02	SH3RF1, RNF144B, LMO7




MAPK14
MAT2A
KLF2
RHOB
ADM
hsa-mir-16-5p
MELK
SORD
TMED4
PAPOLG
CCND2
FOS
SLC16A1
CDK6
hsa-mir-92a-3p
CALR
BAMBI
BUB1
SLC43A3
HIC2
OLFML2A
PDIA6
RBM8A
RPL37
MAP1B
CEP55
ETS1
KIAA0101
EMCN
SCUBE3
EEF1A1
hsa-mir-17-5p
RBFOX2
TM4SF1
SCD
FOLR1
RRM2
AP1S3
GALL1
hsa-mir-335-5p
EIF2S3
IPP
CDK1
MYADM
SLC7A5
INMT
B4GALT1
SLC6A8
MTHFD2
NOTCH2
PMAIP1
CAV1
BTG2
DNAJB4
CLCC1
LIPG
hsa-mir-1-3p
hsa-mir-93-5p
RORA
GPR155
RUNX1T1
SLC5A3
hsa-mir-106b-5p
TXNL1
hsa-mir-20a-5p
PSAT1
hsa-mir-124-3p
hsa-mir-215-5p
TRIB1
hsa-mir-26b-5p
MANEAL
TXNIP
GLUL
RNF157
hsa-mir-192-5p

Pathway	Total	Expected	Hits	P.Value
p53 signaling pathway	68	2.25	8	1.65E-03
Dorso-ventral axis formation	12	0.397	3	6.29E-03
Pertussis	52	1.72	6	6.92E-03
Adherens junction	70	2.32	7	7.95E-03
Thyroid cancer	28	0.927	4	1.27E-02
Pathways in cancer	310	10.3	18	1.32E-02
Focal adhesion	200	6.62	13	1.45E-02
Sphingolipid metabolism	46	1.52	5	1.71E-02
PPAR signaling pathway	64	2.12	6	1.84E-02
ECM-receptor interaction	84	2.78	7	2.05E-02

Pathway	Total	Expected	Hits	P.Value
Transcriptional misregulation in cancer	19	0.942	11	1.87E-10
Acute myeloid leukemia	57	2.83	16	7.48E-09
Pathways in cancer	310	15.4	40	9.50E-09
Thyroid cancer	28	1.39	10	4.41E-07
HTLV-I infection	199	9.86	26	3.86E-06
Cell cycle	124	6.15	18	3.12E-05
Dorso-ventral axis formation	12	0.595	5	1.71E-04
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	12	0.595	5	1.71E-04
Cocaine addiction	43	2.13	9	1.99E-04
Prostate cancer	87	4.31	13	3.01E-04


Gene Expression Data
Differentially Expressed Genes (DEGs)
Protein-Protein Interaction network Analysis
Identify Hub Proteins
DEGs -miRNAs Interaction network
Identify Reporter miRNAs
DEGs -TFs Interaction network
Identify Reporter TFs
Functional Enrichment Analysis
Identify Go Terms and KEGG Pathways