A system biology approach reveals cellular pathway differences between follicular thyroid carcinoma and follicular thyroid adenoma

Md. Ali Hossain[a,b,*], Tania Akter Asa[c], Md. Mijanur Rahman[d], Julian M.W. Quinn[e], Fazlul Huq[f], Mohammad Ali Moni[e,f,*]

[a]*Dept of CSE, Manarat International University, Dhaka-1212, Bangladesh*
[b]*Dept of CSE, Jahangirnagar University, Savar, Dhaka, Bangladesh*
[c]*Dept. of EEE, Islamic University, Kushtia, Bangladesh*
[d]*Dept. of CSE, Jatiya Kabi Kazi Nazrul Islam University, Bangladesh*
[e]*Bone biology divisions, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia*
[f]*The University of Sydney, Sydney Medical School, School of Medical Sciences, Discipline of Pathology, NSW 1825, Australia*

Abstract

Pathogenic mechanisms that underlie malignant follicular thyroid carcinoma (FTC) development are poorly understood. To identify key genes and pathways driving malignant behaviour we employed a system biology-based integrative analyses comparing FTC transcriptomes with a similar but benign lesion, follicular thyroid adenoma (FTA). We identified differentially expressed genes (DEGs) in microarray gene expression datasets (n=52) of FTCs and FTA tissues. Pathway analyses of DEGs using gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) resources revealed significant pathways, and pathway hub genes using protein-protein interactions (PPI). We identified 598 DEGs (relative to FTAs) in FTCs and 12 significant pathways with altered expression in FTC. 10 GO groups were significantly connected with FTC-high expression DEGs and 80 with low-FTC expression. PPI analysis identified 12 potential hub genes based on degree and betweenness centrality. Moreover, 10 transcription factors (TFs) were iden-

*Corresponding Author
Email address: ali.hossain@manarat.ac.bd (Md. Ali Hossain), *mohammad.moni@sydney.edu.au* (Mohammad Ali Moni)

tified that may underlie DEG expression as well as a number of microRNA (miRNAs). Thus, we identified DEGs, pathways, TFs and miRNAs that reflect molecular mechanisms differing between FTC and benign FTA. These may constitute biomarkers that distinguish these lesions and, given the similarities and common origin of the lesions, they may also be indicators of malignant progression potential.

Keywords: Thyroid cancer carcinoma, protein-protein interaction, reporter transcription factors, reporter microRNAs, molecular pathways

---

1. Introduction

Thyroid cancers are the most common type of endocrine malignancy, although they have a relatively low mortality rate compared to most other common metastatic diseases. The United States had 56,460 new diagnoses of thyroid cancer and 1,780 related deaths reported in 2012 [18]. Its incidence is also rising globally at about 5% per year, although some of this increase may be due to improved detection, and it notably affects those in the 20 to 34 year age range [1]. Thyroid cancers include several major types including papillary thyroid carcinomas, medullary thyroid carcinoma, anaplastic thyroid carcinoma and follicular thyroid carcinomas (FTCs) [3]; FTC is one of the more aggressive types, although it accounts for a minority (14%) of total thyroid cancers [34].

The causes and cellular processes that give rise to FTC and control these tumors behaviour are poorly understood; because of this, these cancers have few effective treatment options [37]. There is, therefore, a great need to understand the mechanisms that drive development and progression in FTC to identify new approaches to detection, estimate risk of progression and find new therapies. In addition, differential diagnosis of FTC is problematic as it can be difficult to distinguish from follicular thyroid adenoma (FTA), a benign and non-invasive lesion. For this reason, molecular markers that distinguish FTC and FTA (and other types of thyroid lesions) are much sought. Thus, Wojtas *et al.* conducted a gene expression comparison of FTC and FTA lesions which identified potential markers that can distinguish FTC from FTA with a sensitivity and specificity of 78% and 80%, respectively. [51]. We can also use this dataset to identify pathways with different levels of activity in benign and aggressive thyroid tumours (here, FTA and FTC) that may reflect important molecular mechanisms that underlie their behaviour.

For such pathway studies our starting point is the differential expression of genes (DEGs) between these lesions. We can then use analytic tools to study these DEGs and discover pathways, and pathway hub genes that may affect cell functions and thereby underlie their tumour morbidity, growth and invasion [28, 14, 30]. The repertoire of candidate pathway factors can be extended using approaches such as gene ontology analysis and protein-protein interactions (PPI) studies.

To identify such DEGs, microarray gene expression profiling is widely used [11] and several such studies have been performed for thyroid cancer subtypes [40]. For example, Huang et al. [16] studied gene expression profiles of thyroid cancers, although not functional interactions between gene products. To understand better the underlying molecular mechanisms and identify critical biomolecules, integrative analysis within a gene network context is needed [38, 29, 39]. These may lead to candidate FTC biomarkers (the focus of Wojtas *et al.*) but our main interest is to obtain key or hub genes distinguishing malignant FTC from benign FTA as these may be potential therapeutic targets. Such a systems biology approach integrating network statistical and topological analyses of experimental datasets can thus clarify disease mechanisms [22]. Thus, we used such a systems approach (Figure 1) to identify FTC molecular signatures at the miRNA, mRNA, and protein levels that differ from that of FTAs. Such a comparison should be more meaningful than a comparison of FTC to normal thyroid tissue as the tumours are similar, yet differ markedly in their potential for invasion metastasis.

2. 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 these 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 FTC and FTA (GSE82208) for our study from the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/)

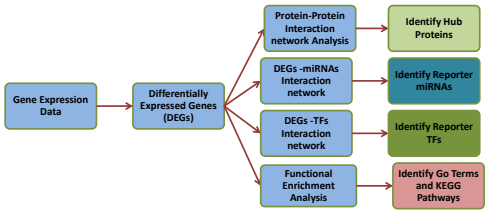


Figure 1: The multi-stage analysis methodologies employed in this study. Gene expression datasets related to FTC and FTA tissue were collected from the NCBI Gene Expression Omnibus (NCBI-GEO) database and statistically analyzed using GEO2R to identify DEGs. Four types of functional enrichment analyses 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 micro-RNA (miRNA) and transcription 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 were 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.

[4, 51]. This dataset contained analyses of RNA from frozen tumour tissue specimens from 27 FTC and 25 FTA lesions using Affymetrix human genome U133 (Plus 2.0) arrays. The lesions were diagnosed histologically, many with a second diagnosis to confirm, when paraffin embedded material was available [51]. Gene expression analysis using microarrays is a widely used method to develop and refine the molecular determinants of human disorders. We used data using these technologies to analyze the gene expression profiles of FTC and FTA. To identify the DEGs between FTC samples and FTA, t-test method was used by using the Limma package [47]. To identify the up-regulated genes, we used the condition of $p - value < .05$ and $logFC > 2$ (FC, fold change) and for identifying the down-regulated genes, $p - value < .05$ and $logFC < -2$ were used. All identified up-regulation genes and down-regulation genes were considered as DEGs. We applied the topological and neighborhood based benchmark methods to find gene-gene associations. A gene-gene network was constructed by using the gene-gene associations, where the nodes in the network represent gene [53, 29]. This network can also be characterized as a bipartite graph. These topological and neighborhood based benchmark methods were adopted from our previous studies [30].

The common neighbours are the based on the Jaccard Coefficient method,

where the edge prediction score for the node pair is as [31]:

$$E(i,j) = \frac{N(G_i \cap G_j)}{N(G_i \cup G_j)} \tag{1}$$

where G is the set of nodes and E is the set of all edges. We used R software packages "comoR" [28] and "POGO" [30] to cross check the genes-diseases associations.

Functional Enrichment of Gene Sets

We performed gene ontology and pathway analysis on identified up-regulation genes and down-regulation genes using DAVID bioinformatics resources (https://david-d.ncifcrf.gov/) (version v6.8) [46] to get further insight into the molecular pathways that differ between FTC and FTA. In these analyses, GO 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 [48] 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. To construct the PPI network from the STRING database (http://string-db.org) [48], we used database data, data mined from PubMed abstract text, Co-expression, gene fusion and Neighborhood as active interaction sources and a combined score that is greater than 0.4 was set as the level of significance. The PPI network was then visualized and analyzed using Cytoscape (v3.5.1) [45, 32]. Then, topological analysis was applied to identify highly connected proteins (i.e., hub proteins) through the Cyto-Hubba plugin [7] where betweenness centrality and higher degree were employed. After then, 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]. Finally, these modules were further analyzed and characterized using enrichment analyses by NetworkAnalyst [52, 13]. The KEGG pathway enrichment analysis of the PPI networks involved DEGs were performed by NetworkAnalyst [52, 13].

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 experimen-

tally verified TF-target genes from the JASPAR database [20] and miRNA-target gene interactions from TarBase [43] and miRTarBase [15] by using NetworkAnalyst tools [52] where betweenness centrality and higher degree filters were used. Currently, there were many techniques to measure the topological properties. We used the Degree Centrality (DC) and Betweenness Centrality (BC) to find out networks topological properties. We can define the DC of a node v in a network as the total number of nodes which are directly connected to node v in that network. The definition can also be written as follows:

$$DC(v) = \sum_{j \in G} \frac{a_{vj}}{n-1} \tag{2}$$

Whereas, n represents total number of nodes in the network and avj represents that node v and node j are directly connected. In the case of Betweenness Centrality (BC), the total number of times of node v appearing in the shortest path between other nodes are quantified. It is also defined as follows:

$$BC(v) = \sum_{i \neq j \neq v \in V} \frac{\sigma_{ivj}}{\sigma_{ij}} \tag{3}$$

Where σ_{i_j} = total number of shortest paths from node i to node j, and σ_{ivj} = total number of paths through node v.

3. Results

3.1. Results of DEG analyses

Transcriptomic signatures: Differentially expressed genes

FTA and FTC tissue gene expression patterns were analysed using oligonucleotide microarrays from the NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=G [51]. Gene expression profiling was performed in 27 malignant FTCs and 25 FTAs. 598 genes were differentially expressed ($p < 0.05, > 1.0$ log2 fold change) relative to FTAs, of which 465 genes were significantly lower expression and 133 genes were higher expression levels in FTC lesions (see Additional file 1:Table S1).

3.2. Pathway and functional correlation analysis

Combining large scale, state of the art transcriptome and proteome analysis, we performed a regulatory analysis to gain further insight into the molecular pathways associated with the FTC and predicted links to pathways that

differ relative to the benign FTAs. DEGs and pathways were analysed using KEGG pathway database ($http://www.genome.jp/kegg/pathway.html$) and functional annotation tool DAVID v 6.8 (http://niaid.abcc.ncifcrf.gov) to identify overrepresented pathway groups and it was observed that 4 pathways were associated with DEGs with higher expression in FTC compared to FTA, namely the "One carbon pool by folate" pathway, p53 signaling pathway, Cell cycle and Progesterone-mediated oocyte maturation signaling pathway. Fold enrichment, adjusted p-values and genes associated with these pathways are presented in Table 1(a) A.

Table 1: (a) Part A shows pathways associated with the genes showing higher expression in FTC than FTA. This set showed an enrichment of four KEGG pathways, annotated using DAVID. Part B shows pathways associated with genes showing lower expression levels in FTC. This set showed in an enrichment of eight KEGG pathways using DAVID. Fold enrichment, adjusted p-values and genes associated with these pathways are indicated. (b) Summary of reporter biomolecules (including putative hub genes, TFs, and miRNAs) in that differ between FTC and FTA.

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)

It was also observed that 8 significantly over-represented pathways 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 1(a) B) were identified. These are associated with the DEGs expressed at lower lev-

els in FTC lesions compared to FTA. We then performed GO analysis using DAVID to obtain further insight into the molecular roles and biological function of DEGs identified in this study. From this analysis, 9 GO groups were associated with DEGs that were more highly expressed in FTC (see Table 2(a)). Reflecting their greater number, the genes with lower expression in FTC compared to FTA were associated with 80 GO groups; the 10 most significant GO groups are presented in Table 2(b). Fold enrichment, adjusted p-values and genes associated with these ontologies are also indicated.

Table 2: (a) Gene ontologies associated with genes with higher expression in FTC. (b) Gene ontologies associated with genes with lower expression in FTC compared to FTA. The 10 most significant biological functional ontologies are indicated and are those 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 Enrichment	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)

3.3. Proteomic signatures: Hub target proteins from PPI analysis

A PPI network was constructed using the DEGs identified in this study (Fig. 2A) using the STRING package. Topological analyses using STRING as well as further analysis 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 PPI network for most significant hub genes were constructed by using Cyto-Hubba plugin and are shown in fig. 2B.

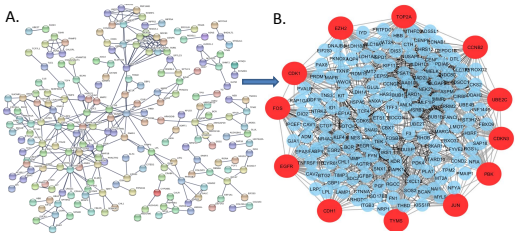


Figure 2: A. The protein-protein interaction network constructed using the DEGs identified in the FTC/FTA comparison. B. The protein-protein interaction(PPI) network identified in FTC/FTA dataset DEGs showing the most significant hub genes on the periphery (red)

3.4. 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

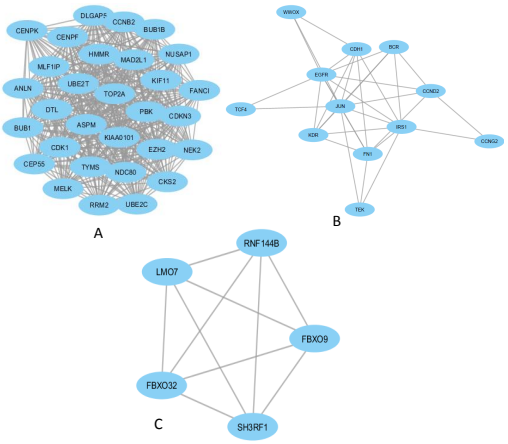


Figure 3: Top three modules in the protein-protein interaction network of the DEGs in TC. The nodes indicate the DEGs and the edges indicate the interactions between two genes.

analysis of the top three modules was analyzed by using DAVID. Module 1 represented a set of biological process significantly enriched for ATP binding, Cell division, Cell proliferation; associated enriched cellular components were Kinetochore, Condensed chromosome kinetochore (both of which relate to cell division) and Nucleoplasm and Cytoplasm which are very general categories. Molecular functions significantly enriched in module 1 included chromatin binding and protein kinase activity (Table 3 (a)). Module 2 showed

significantly enriched biological process positive regulation of Protein phosphorylation and Protein autophosphorylation; these significantly enriched cellular components included membrane raft, endosome, apical plasma membrane and focal adhesion. Molecular functions of module 2 were enriched in Protein tyrosine kinase activity, Enzyme binding, Protein binding and Integrin binding (Table 3(b)).

Table 3: (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)

Module 3 gene ontology annotation found significant enrichment for Protein ubiquitination, with significantly enriched cellular components being ubiquitin ligase complex and SCF ubiquitin ligase complex, which have involvement in proteosomal and autophagy functions. Module 3 molecular functions showed significant enrichment in ubiquitin-protein transferase activity and zinc ion binding (Table 3 (b)). It was also notable that pathway enrichment analyses found modules 1 and 2 were significantly enriched in protein binding pathways, which was not seen in was obtained in module 3 (Table 3 (b)).

3.5. Regulatory signatures: TFs and miRNAs affecting DEG transcript Levels.

3.5.1. 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 that differ between FTC and FTA and 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 4(b)).

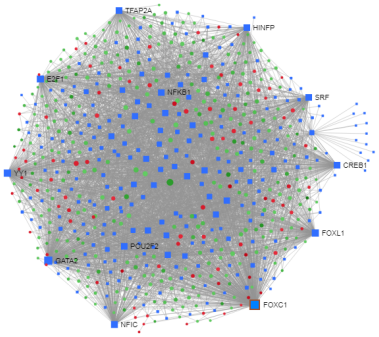


Figure 4: Construction of regulatory networks of TF-DEG interactions. Red nodes indicate up-regulation and green node represents down-regulated DEGs.

3.5.2. 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 po-

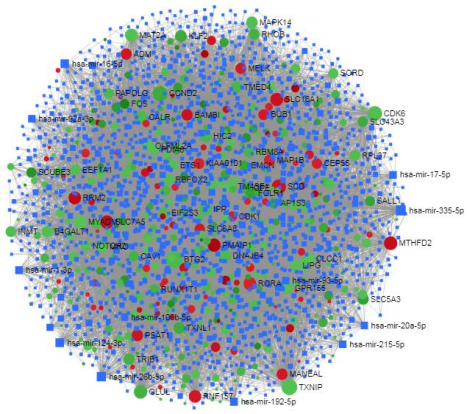


Figure 5: Construction of regulatory networks of DEGs-miRNAs interaction. Red nodes indicate up-regulation and green node represents down-regulated DEGs.

tential to regulate levels of the DEG transcripts differently in FTC and FTA. 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 adhesion, Sphingolipid metabolism, PPAR signaling pathway, and ECM-receptor interaction signaling system statistically significant (Table 4(a))

Table 4: (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)

4. Discussion:

To clarify the molecular mechanisms that underlie the pathological features (especially malignancy potential) that distinguish malignant and benign thyroid lesions we examined differential gene expression between FTC from FTA. Such a comparison is likely to be more informative than making such a comparison of malignant and normal tissues which will contain many more differences related to neoplastic cell development. In addition, the lesions contain significant non-neoplastic cell components (such as mesenchymally-derived and endothelial cells) that are also likely to differ markedly from those of normal tissue.

The DEGs identified from the FTC/FTA comparison were used to identify possible regulatory patterns, key molecular pathways, and protein-protein interactions among these pathway gene products. Such an analytical approach can be used to find molecular signatures that may serve either as potential therapeutic targets or as biomarkers to differentially diagnose or identify FTC. As did the original study by Wojtas *et al.* [51], we identified significantly altered genes that may be candidate biomarkers that distinguish FTC. However, another important use of such data (and our main focus here) is to identify and characterize biological functions associated with these genes that may give insights into the biology and behaviour of FTC lesions themselves.

Our study thus identified a number of important pathways that were differentially expressed in FTC, some of which would be expected by the malignant profile of FTC compared to FTA. For example, the p53 signaling pathway plays an important role in cancer cell apoptosis and DNA repair by causing cycle arrest in response to DNA damage [57] and previously implicated in thyroid cancers [12]. Cell cycle progression is regulated and facilitated by cyclin-dependent kinases that are activated by cyclins such as cyclin D1. Indeed, cyclin D1 levels can influence tumour progression, and may have prognostic significance in FTC [50]. Related to this, it has also been reported that the One-carbon metabolism pathway is actively involved in cancer progression, probably due to its involvement in nucleotide synthesis [33].

We also found a number of significant GO groups are associated with the DEGs as shown in table 2. PPI network reconstruction and the analysis of reconstructed PPI network represent a powerful approach for understanding disease mechanisms, so to construct a PPI network for the DEGs in our study

we combined results of statistical analyses with the protein interactome network. To identify potential hub proteins, topological analysis strategies were employed. This identified 12 hub genes (Table 1(b)). These include a number of genes commonly associated with cancers (including thyroid cancers), such as EGRF, JUN, FOS, CDK1 and other cyclin pathway genes, E-cadherin (CDH1) and E2F1. The hub protein EGFR is linked to growth in many types of cancer as well as cardiovascular diseases, and has been previously linked to thyroid cancers [42]. JUN and FOS are best characterised as subunits of the AP-1 transcript factor that has a crucial role in many types of cell differentiation and inflammation processes [6]. GO databases also delineated particular roles for JUN in Angiogenesis and regulation of Endothelial cell proliferation, both of which have central roles in metastasis and invasion (Table-2 (b)). CDK1 is an important cell cycle regulator and is involved in the breast, lung and ovary carcinomas [26]. CDK1 overexpression has been documented in lung cancer, lymphoma and advanced melanoma while loss of cytoplasmic CDK1 predicts poor patient survival and may confer chemotherapeutic resistance in the latter [36]. 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 [36]. It is also actively involved in mitotic cell cycle phase transition (Table 2 (a)). CDKN3 encodes a cyclin inhibitor (regulating cell cycle) and has been described as overexpressed in lung adenocarcinoma (ADC), squamous cell carcinoma (SCC), hepatocellular carcinoma, cervical cancer and epithelial ovarian cancer [56]. Over-expression of EZH2 is frequently observed in many cancer types [54].

TYMS expression is also associated with the risk of development of epithelial cancers and lymphoma cancer[19]. Besides, in 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 [35]. This gene is also involved in extracellular matrix organization (Table 2(b)). Hub gene CCNB2 is known to be underexpressed in thyroid cell tumors [10]. We found that CDK1 and CCNB2 hubs genes were actively associated with p53 signaling pathway, Progesterone-mediated oocyte maturation and Cell cycle pathways. The hub gene TYMS is also involved in One carbon pool by folate pathway, while FOS, JUN and CDH1 genes are involved in Pathways in cancer and JUN, FOS genes in Estrogen signaling pathway. A number of other hub genes have been described as associated with malignant thyroid cancers including

EXH2, UBCH10,TFAP2A, TOP2A and SRF [27, 23, 8, 41, 21]. By considering the possible roles of these hub proteins in pathogenesis of FTC and other related diseases new roles for these proteins may be identified. In addition to the above hub genes, we also identified a number of factors not previously noted as having a role in FTC or thyroid cancers, including UBA52, GATA2, FOXL1 and NFIC.

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 regarding dysregulation of gene expression in FTC. Thus, we investigated TFs and their relationship to DEGs in these tumours. One of the TFs, we identified was FOXC1 which a previous study found that FOXC1 was strongly associated with thyroid cancers [5]. Overexpression of YY1 in differentiated thyroid cancers has also been noted [2]. The pathway analysis of these TFs showed statistical significance in FTC. From the DEGs a miRNA interaction network of miRNAs (Table 1(b)) was also identified and analysed. Our pathway analysis of these miRNA also showed statistically significant relationships with FTC. Notably, the target DEGs of these miRNAs and TFs included the p53 signaling pathway, PPAR signaling pathway, Dorso-ventral axis formation, Focal adhesion (Table 4(a)) and Pathways in cancer, Transcriptional misregulation in cancer, and Thyroid cancer, respectively. The miRNA species we identified that have previously been shown to have a role in thyroid cancer included hsa-mir-335-5p, hsa-124-3p, hsa-mir-17-5p and hsa-mir-20a-5p [9, 55, 25, 17]. However, 6 other miRNAs we identified have not previously been associated with thyroid lesions. miRNA species have wide ranging effects on gene expression which may be indirect, so further analysis for these are needed.

We analysed the dataset originally generated by Wojtas *et al.*, who sought markers distinguishing FTC and FTA and combined microarray methods and literature meta-analysis to achieve this. The gene markers they identified included TFF3 and CPQ (previously characterised FTC markers), PLVAP and ACVRL1, all genes expressed in thyroid although ACVR1 is an activin receptor with wide expression. These latter genes were not identified by our analytical approach, but it should be noted that our study used very different analytic tools, focussing on identifying pathways and on finding evidence for gene functions using PPI and other resources. Our analytical approach was somewhat more similar to that of Wang *et al.* [49] who examined an older thyroid cancer dataset (GSE27155) containing 10 FTA, 14 FTC, 4 normal thyroid tissue samples plus other types of thyroid lesion. While we employed

a number of additional analytical tools (some developed previously by us) there were a number of significant genes and pathways we found in common in the FTC/FTA analyses Wang *et al.* [49]. These notably include the AP-1 TF and its components (e.g., JUN and FOS), BCL2, factors relating to DNA damage and cyclin-related factors that regulate cell cycle. We identified no genes that Wang *et al.* identified in comparing FTC and normal tissues. Recent studies by Shang *et al.* and by Liang and Sun [24] examined datasets generated from papillary thyroid cancer lesions that were compared with and normal tissues. DEGs identified in these studies were analysed by methods related to our approach [44]. However, unlike the Wang *et al.* study cited above, very few of the hub genes these studies identified were found in our analysis; BCL2 was the main exception. This lack of concordance with the FTC studies may well reflect the different tumour type (and lower malignancy rates) represented by papillary thyroid cancer, and it might be useful to compare all these datasets in detail to determine how these types of tumour differ.

5. CONCLUSIONS

Our data-driven approach has uncovered a number of significant molecular mechanisms that may underlie the difference between FTC and FTA. In this study, using integrated bioinformatics analyses, we analyzed gene expression profiles in FTC and FTA and used this information to identify candidate hub proteins that could play significant roles in FTC development. Our results indicate the directions for future experimental work needed to clarify the roles of these cellular proteins. In sum, the study identified gene networks that advance our understanding of the pathogenesis of FTC, and indicates new avenues to develop therapies for FTC.

Acknowledgments

The work was supported by the iHealthOmics
(*http : //www.ihealthomics.com/*) and a University of Sydney Fellowship award.

Author Contribution

conceptualization, MAH and MAM and TAA; methodology, MAH and TAA.; software, MAH; validation, MAH, TAA and MAM; formal analysis,

MAH; investigation, MAH; resources, MAH; data curation, MAH and TAA; writingoriginal draft preparation, MAH; writingreview and editing, MAH, MMR, JQ and MAM; visualization, MAH and TAA; supervision, MAM; project administration, FH and MAM; funding acquisition, FH and MAM

This research received no external funding

We have No conflict of interest.

The following abbreviations are used in this manuscript:

ADC	Adenocarcinoma
DEG	Differencially Expressed Genes
FC	fold change
FTC	Follicular Thyroid Carcinomas
FTA	Follicular Thyroid Adenoma
GO	Gene Ontology
GEO	Gene Expression Omnibus
HDGC	Hereditary diffuse gastric cancer
KEGG	Kyoto encyclopedia of genes and genomes
miRNA	Micro-RNA
PPI	Protein-protein Interactions
SCC	Squamous Cell Carcinoma
TF	Transcription Factor

[1] Thyroid cancer: Statistics— Cancer. net. https://www.cancer.net/cancer-types/thyroid-cancer/statistics, 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] J. P. De Andrade, A. W. Lorenzen, V. T. Wu, M. V. Bogachek, J. M. Park, V. W. Gu, C. M. Sevenich, V. C. Cassady, A. C. Beck, M. V. Kulak, et al. Targeting the sumo pathway as a novel treatment for anaplastic thyroid cancer. *Oncotarget*, 8(70):114801, 2017.

[9] M. Geraldo, A. Yamashita, and E. Kimura. Microrna mir-146b-5p regulates signal transduction of tgf-β by repressing smad4 in thyroid cancer. *Oncogene*, 31(15):1910, 2012.

[10] 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.

[11] 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.

[12] 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.

[13] M. A. Hossain, T. A. Asa, F. Huq, J. M. Quinn, and M. A. Moni. A network-based approach to identify molecular signatures and comorbidities of thyroid cancer. *Proceedings of International Joint Conference on Computational Intelligence(IJCCI)*, pages 235–246, 2018.

[14] M. A. Hossain, S. M. S. Islam, J. Quinn, F. Huq, and M. A. Moni. Identification of ovarian cancer gene expression patterns associated with disease progression and mortality. *bioRxiv*, page 473165, 2018.

[15] 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.

[16] 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.

[17] I. Jahanbani, A. Al-Abdallah, R. H. Ali, N. Al-Brahim, and O. Mojiminiyi. Discriminatory mirnas for the management of papillary thyroid carcinoma and noninvasive follicular thyroid neoplasms with papillary-like nuclear features. *Thyroid*, 28(3):319–327, 2018.

[18] 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.

[19] 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.

[20] 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.

[21] H. J. Kim, K. R. Kim, H. S. Park, K. Y. Jang, M. J. Chung, M. Shong, and W. S. Moon. The expression and role of serum response factor in papillary carcinoma of the thyroid. *International journal of oncology*, 35(1):49–55, 2009.

[22] 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.

[23] J.-J. Lee, A. Y. Au, T. Foukakis, M. Barbaro, N. Kiss, R. Clifton-Bligh, J. Staaf, Å. Borg, L. Delbridge, B. G. Robinson, et al. Array-cgh identifies cyclin d1 and ubch10 amplicons in anaplastic thyroid carcinoma. *Endocrine-related cancer*, 15(3):801–815, 2008.

[24] W. Liang and F. Sun. Identification of key genes of papillary thyroid cancer using integrated bioinformatics analysis. *Journal of endocrinological investigation*, 41(10):1237–1245, 2018.

[25] L. Liu, J. Yang, X. Zhu, D. Li, Z. Lv, and X. Zhang. Long noncoding rna h19 competitively binds mir-17-5p to regulate yes 1 expression in thyroid cancer. *The FEBS journal*, 283(12):2326–2339, 2016.

[26] 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.

[27] K. Masudo, N. Suganuma, H. Nakayama, T. Oshima, Y. Rino, H. Iwasaki, K. Matsuzu, K. Sugino, K. Ito, T. Kondo, et al. Ezh2 overexpression as a useful prognostic marker for aggressive behaviour in thyroid cancer. *in vivo*, 32(1):25–31, 2018.

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

[29] 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.

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

[31] M. A. Moni and P. Lio. Genetic profiling and comorbidities of zika infection. *The Journal of infectious diseases*, 216(6):703–712, 2017.

[32] M. A. Moni, H. Xu, and P. Lio. Cytocom: a cytoscape app to visualize, query and analyse disease comorbidity networks. *Bioinformatics*, 31(6):969–971, 2014.

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

[34] 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.

[35] 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.

[36] 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.

[37] M. Rahman, T. Islam, E. Gov, B. Turanli, G. Gulfidan, M. Shahjaman, N. A. Banu, M. Mollah, N. Haque, K. Y. Arga, and M. A. Moni. Identification of prognostic biomarker signatures and candidate drugs in colorectal cancer: Insights from systems biology analysis. *Medicina*, 55(1):20, 2019.

[38] M. R. Rahman, T. Islam, B. Turanli, T. Zaman, H. M. Faruquee, M. M. Rahman, M. N. H. Mollah, R. K. Nanda, K. Y. Arga, E. Gov, et al. Network-based approach to identify molecular signatures and therapeutic agents in alzheimers disease. *Computational biology and chemistry*, 2018.

[39] H. K. Rana, M. R. Akhtar, M. B. Ahmed, P. Lio, J. M. Quinn, F. Huq, and M. A. Moni. Genetic effects of welding fumes on the progression of neurodegenerative diseases. *Neurotoxicology*, 71:93–101, 2019.

[40] 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.

[41] M. Rossing. Classification of follicular cell-derived thyroid cancer by global rna profiling. *Journal of molecular endocrinology*, 50(2):R39–R51, 2013.

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

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

[44] J. Shang, Q. Ding, S. Yuan, J.-X. Liu, F. Li, and H. Zhang. Network analyses of integrated differentially expressed genes in papillary thyroid carcinoma to identify characteristic genes. *Genes*, 10(1):45, 2019.

[45] 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.

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

[47] G. K. Smyth, J. Michaud, and H. S. Scott. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*, 21(9):2067–2075, 2005.

[48] 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.

[49] Q. Wang, Y. Shen, B. Ye, H. Hu, C. Fan, T. Wang, Y. Zheng, J. Lv, Y. Ma, and M. Xiang. Gene expression differences between thyroid carcinoma, thyroid adenoma and normal thyroid tissue. *Oncology Reports*, 40(6):3359–3369, 2018.

[50] 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.

[51] 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.

[52] 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.

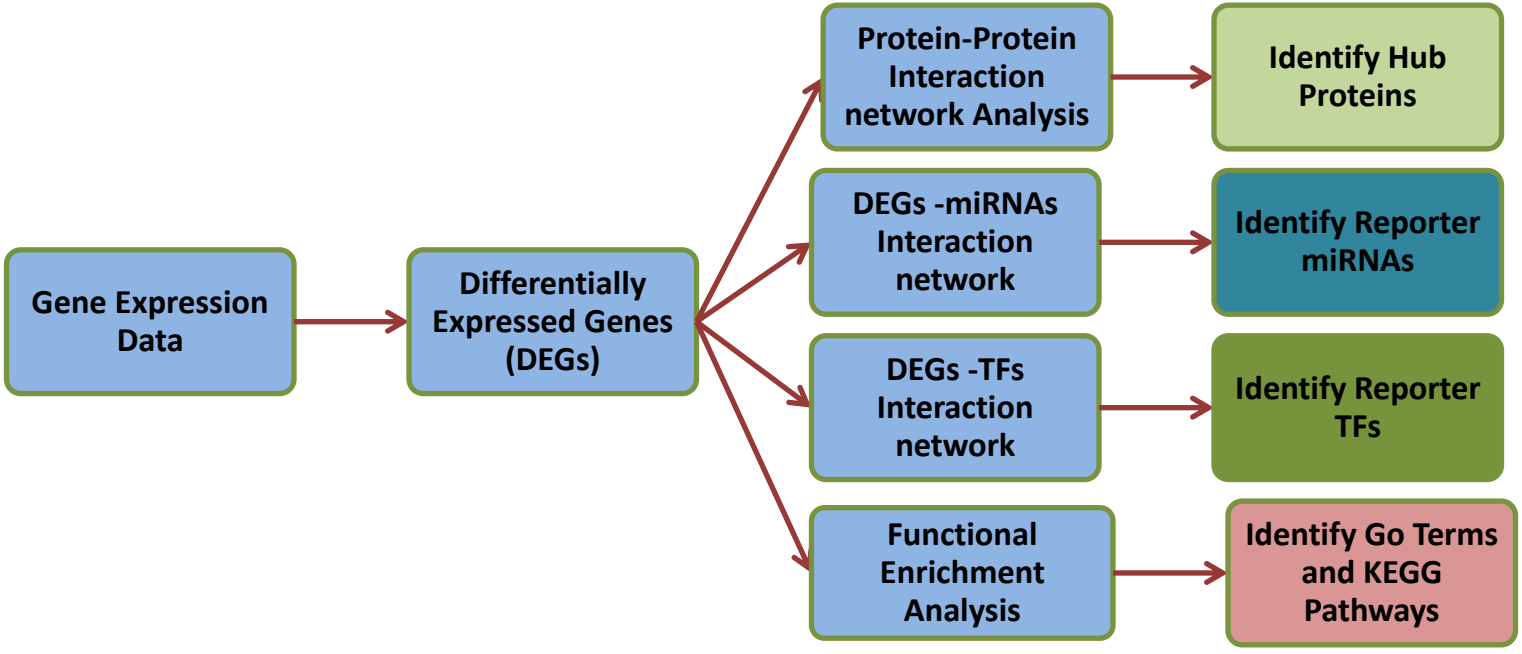
[53] H. Xu, M. A. Moni, and P. Liò. Network regularised cox regression and multiplex network models to predict disease comorbidities and survival of cancer. *Computational biology and chemistry*, 59:15–31, 2015.

[54] 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.

[55] S. Yu, X. Liu, Y. Zhang, J. Li, S. Chen, H. Zheng, R. Reng, C. Zhang, J. Chen, and L. Chen. Circulating microrna124-3p, microrna9-3p and microrna196b-5p may be potential signatures for differential diagnosis of thyroid nodules. *Oncotarget*, 7(51):84165, 2016.

[56] 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.

[57] 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.


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