A Comparative Analysis of Gene Expression Profile in Liver and Esophageal Cancer using Expressed Sequence Tags

Mahboobeh Ramezani [1], Fatemeh Shamsabadi[2], Morteza Oladnabi*[3,4,5]

1. Student Research Committee, Golestan University of Medical Sciences, Gorgan, Iran
2. Department of Medical Biotechnology, Golestan University of Medical Sciences, Gorgan, Iran
3. Cancer Research Center, Golestan University of Medical Sciences, Gorgan, Iran
4. Ischemic Disorders Research Center, Golestan University of Medical Sciences, Gorgan, Iran
5. Gorgan Congenital Malformations Research Center, Golestan University of Medical Sciences, Gorgan, Iran

ABSTRACT

Background and objectives: Liver and esophageal cancers are common among the Iranian population. This study aims to explore the common up-regulated genes in liver and esophageal cancer tissues using expressed sequence tags (ESTs) and to identify the role of key genes in cancer development.

Methods: EST profiles of protein-coding genes in normal and cancerous hepatic and esophageal tissues were extracted from the UniGene database. Genes with > 1500 transcripts per million were selected as highly expressed. The cancer to normal ratio of up-regulated genes was calculated. The shared overexpressed genes between liver and esophageal cancer tissues were determined. Finally, functional classification and pathway analysis were performed on the genes using the STRING and Enrichr databases.

Results: Of 17,242 genes, 53 and 26 genes were overexpressed in the liver and esophageal cancer tissues, respectively. Nine up-regulated genes (APLP2, EEF1G, ENO1, HSP90AA1, HSP90AB1, HSPA8, KRT18, RPL4 and UBC) were shared between the two cancer tissues, which were involved in cell cycle progression through G2/M checkpoint, G1/S transition and DNA replication. They were also involved in the vascular endothelial growth factor, hypoxia-inducible factor 1 and estrogen signal pathways as well as the Toll-like receptor cascade.

Conclusion: Based on the results, the identically up-regulated genes and underlying molecular mechanisms implicated in both cancers could be valuable targets for diagnosis and treatment of cancer.

Keywords: Liver cancer, Esophagus cancer, ESTs, Up-regulated genes, Molecular mechanism

INTRODUCTION

The cause of many sporadic and highly heterogeneous cancers remains unknown. Compressive total RNA analysis of diversely expressed genes in tumor tissues compared to normal counterparts can help discover candidate genes that could play an important role in human malignancies. A number of techniques, ranging from differential display and differential nucleic acid subtraction to sequential analysis of gene expression, expression microarrays and gene chips have been used to identify such aberrantly expressed cancer-related genes (1-3).

Differential screening is the most efficient and widely used technique for comparison of gene expression between two different tissue samples under normal and pathological conditions. The computer-based differential display methodology could identify transcripts preferentially expressed or repressed in the tumor context by comparing cancerous libraries (present in public databases) against the remaining libraries (4-7).

The expressed sequences tags (ESTs) are partial sequences of cDNA segments generated from different tissues. ESTs from normal and cancer tissues have been accumulated in public sequence databases, such as the UniGene. In this database, data are presented as transcripts per million (TPM), which is used to normalize the expression level of genes. The value of EST in a certain tissue is calculated based on the following formula (8):

$$TPM = \frac{(Number\ of\ EST\ in\ the\ cluster)}{(Number\ of\ Total\ EST\ of\ the\ tissue)} \times 1000000.$$

The UniGene database (https://www.ncbi.nlm.nih.gov/unigene) computationally determines the transcripts from similar loci, analyzes the expression of genes based on tissue types, age and health status, and reports relevant proteins and clone resources. The information and content of this database is gathered from GenBank, a large collection of cDNA and ESTs resulting from

decades of universal work (9-10). On the other hand, ESTs represent the set of all transcribed genes in genome that provide valuable information regarding gene mapping and gene annotation. Importantly, the abundance of a specific mRNA can be estimated from the number of corresponding ESTs found in a specific tissue library (11).

The promoter region of the genes is a critical element for gene expression regulation and directly affects the gene transcription level. In eukaryotic systems, these fragments are classified into TATA-containing and TATA-less promoters. TATA box is located 25 to 30 nucleotides upstream of the transcriptional start site and directs RNA polymerase II for transcription process. Earlier studies have identified regulatory elements in 76% of human promoters with high GC content enriched with SP1 binding sites (12, 13). To date, few studies have focused on the nucleotide composition of the immediate upstream sequences to the transcription start sites. Currently available evidence has failed to demonstrate the impact of GC content and the purine/pyrimidine ratio on gene expression. Previous research on short tandem repeats (STRs) in the 5′ untranslated region (UTR), 3′UTR, coding and intronic regions have revealed the functional roles of STRs (14, 15). However, the amount of STRs in the human core promoter region remains unknown. Here, we assess up-regulated genes in liver and esophageal cancer tissues by determining EST profiles and STR analysis of the core promoter region of these genes.

MATERIALS AND METHODS

In this study, all human protein-coding genes were selected from the GeneCards database (http://www.genecards.org/List). EST profile of genes was extracted from the UniGene database (https://www.ncbi.nlm.nih.gov/unigene), which contains EST profile of only 17,242 protein-coding genes. The EST profiles were examined in normal and cancerous hepatic and esophageal tissues. The TPM value of

17,242 genes was categorized into 5 cut off points: no, low, moderate, high and very high gene expression, and the maximum rank was selected (>1500 TPM). Following the assessment of the up-regulated genes (>1 fold change), the genes shared between the two tissues were identified.

To evaluate presence of STRs in the core promoter region of the shared up-regulated genes, the 120-bp sequence flanking region upstream of the TSS (+1) of these genes (−120 to +1) were retrieved from the Ensembl database (http://www. ensembl.org/index.html; accessed on May 2018). Subsequently, these sequences were evaluated for STR, type, copy number and nucleotide percentage in the Insilico (http://insilico.ehu.es/mini_tools/microsatellit es) and ALGGEN (http://alggen.lsi.upc.es/cgibin/promo_v3/pro mo/promoinit.cgi?driDB=TF_8.3) databases. Finally, functional classification and pathway analysis were carried out for the shared up-regulated genes using the STRING (v 11.0) and Enrichr bioinformatics databases. Protein network interacting with the shared up-regulated genes was retrieved based on experimental, co-expression, textmining and database data with at least medium confidence of 0.4.

Statistical analysis
Statistical analysis was performed using the SPSS 20.0 software. The Student's t-test was used to evaluate intergroup differences. P-values less ≤ 0.05 were considered statistically significant.

RESULTS
Up-regulated genes in esophageal cancer tissues
Of the total 17,242 genes, 4,758 (27.6%) were expressed (>1 fold change) in the normal esophageal tissue and 4,465 (25.9%) were expressed in the esophageal cancer tissues. The second set of analyses demonstrated that 53 genes are highly transcribed in esophageal tumor tissues (Table 1). A high proportion of these genes were expressed with ratio of 1.17.

Table 1. Up-regulated genes in the esophageal cancer tissue

Gene Name	Ensembl ID	Ratio TPM esophageal Cancer/Normal	Gene Name	Ensembl ID	Ratio TPM esophageal Cancer/Normal
PSAP	ENST00000394936	1.17	SLC2A1	ENST00000372500	1.13
KRT18	ENST00000388835	1.17	CCT8	ENST00000286788	1.13
OLR1	ENST00000309539	1.17	XRCC6	ENST00000359308	1.13
HSP90AA1	ENST00000216281	1.17	P4HB	ENST00000331483	1.11
HSP90B1	ENST00000299767	1.17	KRT16	ENST00000301653	1.1
SERPINH1	ENST00000358171	1.17	CCT3	ENST00000295688	1.1
THBS1	ENST00000260356	1.17	RPN1	ENST00000296255	1.1
RRM2	ENST00000304567	1.17	CCT5	ENST00000280326	1.1
MMP1	ENST00000315274	1.17	BSG	ENST00000333511	1.1
MMADHC	ENST00000303319	1.17	CCT7	ENST00000258091	1.09
CD44	ENST00000263398	1.17	TRIM16L	ENST00000395671	1.09
EIF4G2	ENST00000339995	1.17	PERP	ENST00000421351	1.08
KRT17	ENST00000311208	1.17	ZNF207	ENST00000321233	1.08
ATF4	ENST00000337304	1.17	EEF1G	ENST00000329251	1.08
MMP10	ENST00000279441	1.17	RPL4	ENST00000307961	1.06
ADH7	ENST00000209665	1.17	PGK1	ENST00000373316	1.06
RRM1	ENST00000300738	1.17	KRT6C	ENST00000252250	1.03
FOS	ENST00000303562	1.17	HSP90AB1	ENST00000353801	1.03
B2M	ENST00000544417	1.17	ENO1	ENST00000234590	1.02
SLC3A2	ENST00000338663	1.16	HSPA8	ENST00000227378	1.02

NOP56	ENST00000329276	1.15	SRSF2	ENST00000359995	1.01
GPI	ENST00000356487	1.15	EIF3L	ENST00000381683	1.01
APLP2	ENST00000263574	1.14	UBC	ENST00000339647	1.01
GPNMB	ENST00000258733	1.14	ATP5A1	ENST00000282050	1.17
SCPEP1	ENST00000262288	1.14	DDX5	ENST00000225792	1.13
HSPA1B	ENST00000375650	1.14	SYPL1	ENST00000011473	1.13
RBMX	ENST00000320676	1.13			

Up-regulated genes in liver cancer tissues
Analysis of the EST data revealed that 11,345 genes (65.8 %) were transcribed in the normal liver tissue, while 9,189 genes (53.3%) were transcribed in liver cancer tissues. As shown in table 2, 26 genes were highly expressed in liver cancer tissues compared to normal liver tissue. Approximate two-fold overexpression of *APLP2* and *PKM* were observed in liver cancer tissues.

Table 2. Up-regulated genes in the liver tumor tissues

Gene Name	Ensembl ID	Ratio TPM liver Cancer/Normal	Gene Name	Ensembl ID	Ratio
KRT18	ENST00000388835	1.79	PKM	ENST00000335181	1.87
HSP90AA1	ENST00000216281	1.4	TUBA1B	ENST00000336023	1.74
APLP2	ENST00000263574	1.98	RPL3	ENST00000216146	1.51
EEF1G	ENST00000329251	1.67	AKR1C1	ENST00000380872	1.54
RPL4	ENST00000307961	1.7	ANXA2	ENST00000332680	1.7
HSP90AB1	ENST00000353801	1.49	EIF4A1	ENST00000293831	1.52
ENO1	ENST00000234590	1.63	ANXA1	ENST00000257497	1.7
HSPA8	ENST00000227378	1.46	TUBB	ENST00000327892	1.61
UBC	ENST00000339647	1.66	VIM	ENST00000224237	1.57
LDHA	ENST00000227157	1.65	RPS3A	ENST00000274065	1.65
ACTB	ENST00000331789	1.5	RPL10	ENST00000344746	1.55
RPL13A	ENST00000391857	1.19	ALDH1A1	ENST00000297785	1.5
APOB	ENST00000233242	1.27	AGT	ENST00000366667	1.19

TPM: Transcripts per million

Common up-regulated genes in esophageal and liver cancers
Table 3 presents characteristics of identical highly transcribed genes in both cancers. Of the nine up-regulated genes, three are located on the long arm of chromosome 11. No significant difference was observed in the recurrence of STR sequences between tumor and normal tissues.
In order to examine the functional classification of proteins coded by these genes, network analysis was performed using the STRING database. The functional mode of these proteins is illustrated in figure 1A. According to this result, these proteins were mostly implicated in amino acid and nucleic acid metabolism, suggesting their significant function in cancer progression. Figure 1B shows the physical interaction of these proteins. Results illustrated a complex network between these proteins based on empirical and textmining criteria. Interestingly, there was no interaction found with APLP2 and KRT18.

Table 3. List of up-regulated genes identically overexpressed in liver and esophageal cancer tissues

| Gene Name	Ratio TPM esophageal Cancer/Normal	Ratio TPM Liver Cancer/Normal	Gene Ontology	GC content (%)	Chromosome location	STR Formula
APLP2	1.14	1.98	Amino acid metabolism, small molecule biochemistry, dermatological diseases & conditions	43.97	11q24.3	-
EEF1G	1.08	1.67	Protein synthesis, gene expression, RNA post-transcriptional modification	47.78	11q12.3	CCCA/3
ENO1	1.02	1.63	Nucleic acid metabolism, small molecule biochemistry, cellular movement	50.23	1p36.23	GC/3
HSP90AA1	1.17	1.4	Protein refolding , signal transduction, protein tyrosine kinase activity	47.25	14q32.31	-
HSP90AB1	1.03	1.49	ATP-dependent protein binding, protein kinase regulator activity,	49.1	6p21.1	-
HSPA8	1.02	1.46	Cancer, gastrointestinal disease	49.25	11q24.1	CG/4
KRT18	1.17	1.79	Gene expression, protein synthesis, cellular assembly and organization	56.19	12q13.13	GG/3
RPL4	1.06	1.7	Molecular transport, RNA trafficking, DNA damage response, detection of DNA damage, Cell cycle (G1/S)	42.06	15q22.31	AG/3
UBC	1.01	1.66		50.7	12q24.31	-

TPM: Transcripts per million, STR: Short tandem repeat
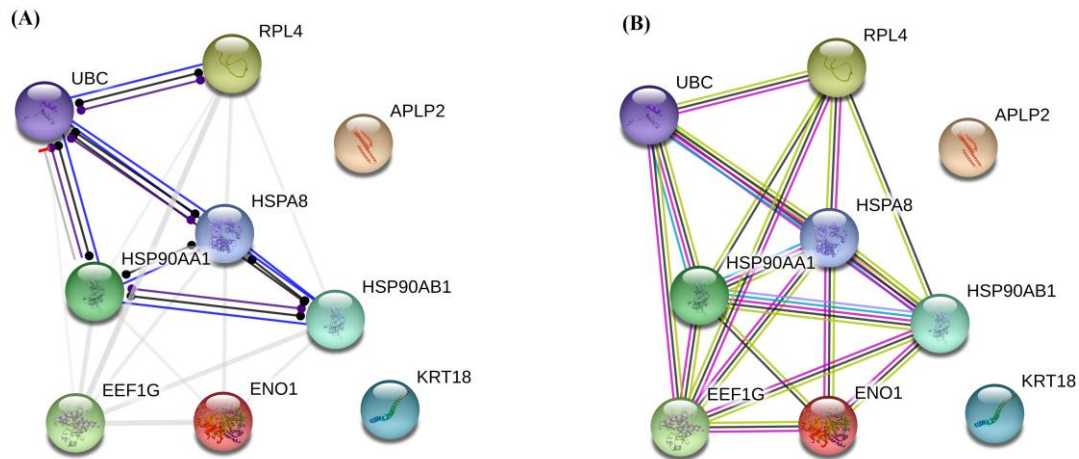
Figure 1. The functional protein association networks of up-regulated genes shared between esophageal and liver cancer tissues.(A) The predicted mode of action of the shared up-regulated proteins is demonstrated. Color of lines show binding (blue), reaction (black), catalysis (violet) and inhibition (red). Line shapes represent the unspecified effect (dot) and negative effect (bar). The thickness of grey lines indicates the validation status. (B) The physical interactions of the shared up-regulated proteins are selected with medium score (0.400) according to the experimental (pink), co-expression (grey), textmining (green) and database (cyan) data.

Moreover, we explored the most significant biological pathways attributed to these genes using the WikiPathways tool in Enrichr (Figure 2). The results indicated absence of RPL4 in these pathways. The heat shock protein (HSPs) HSP90AA1 and HSP90AB1 are involved in PI3K-Akt and mTOR pathways, suggesting their possible involvement in cancer development. As a

modulator of glucose and insulin homeostasis, APLP2 is activated by MAPK9 in the NOVA-regulated splicing of synaptic proteins. Moreover, *KRT18* participates in pathogenic *Escherichia coli* infection pathway that alters the structure of intermediate filament complex.
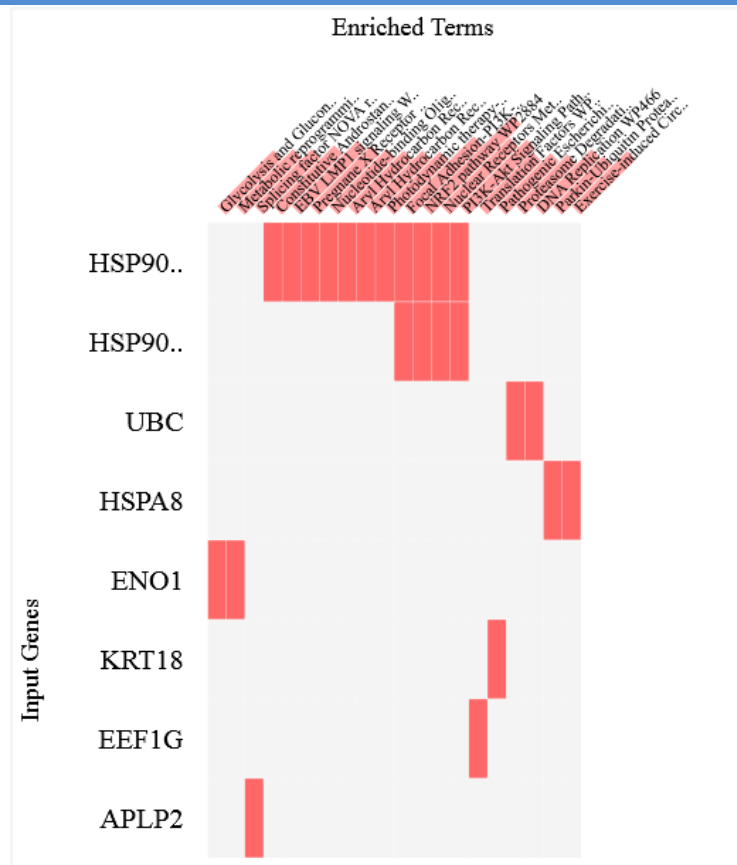
Figure 2. Associated pathways in the genes shared between esophageal cancer and liver cancer. Enriched terms and input genes are presented in the column and the row, respectively. There is no report for the *RPL4* gene.

DISCUSSION

Different genome and proteome high-throughput analysis methods have been used to study gene expression in normal and cancer specimens. These methods improve our knowledge on genes and their variations that might be directly related to cancer phenotype. In addition, by the help of integrative bioinformatics analysis, it is possible to distinguish the differential expressed genes across normal and tumor samples (16, 17). Tumor-associated molecules could be identified using multiple digital differential display methods and experimental gene expression data (16).

To our knowledge, this is the first in silico EST study on the expression variation of protein-coding genes in normal and cancerous esophageal and liver tissues. We found that 26 and 53 genes were highly transcribed in esophageal and liver cancer tissues, respectively. We identified nine genes including APLP2, EEF1G, ENO1, HSP90AA1, HSP90AB1, HSPA8, KRT18, RPL4 and UBC which are overexpressed in both cancer tissues. This finding suggests that these candidate genes may be used as biomarkers for diagnosis and treatment of these cancers. Moreover, they might be considered as targets for developing novel therapeutic approaches. For instance, the expression of enolase 1, cytokeratin 18 and HSPs in tumors may be valuable diagnostic and therapeutic targets. Overexpression of ubiquitin C in this study corroborates a previous research that found down-regulation of ubiquitin C suppresses growth and increases the radio-sensitivity of human non-small cell lung carcinoma cells (18).

Interestingly, three of the shared genes are related to the HSP family, suggesting their

possible involvement in cancer progression. It has been reported that these molecules are involved in the post-translational regulation and the protein folding processes. Moreover, they can cause induce apoptosis under stressful conditions (19). These genes were highly expressed in several types of cancer including breast, endometrial, ovarian, colon, lung and prostate cancer affecting the survival of tumor cells (20).

The expression of HSP90AA1 regulates mitochondrial apoptosis, signal transduction, stress signals and growth factors during tumor growth. It has been shown that HSPA8 may be involved in tumor cell proliferation and apoptosis regulation in hepatocellular carcinoma (21). Overexpression of HSP90AA1/HSPA8 has been detected in patients with hepatocellular carcinoma, indicating their involvement in the VEGF-associated recurrence (22).

The most important clinically relevant finding was the up-regulation of alpha-enolase, which has been previously described as a promising candidate molecular target for immunotherapy (23).

The enrichment analysis revealed the involvement of the shared genes in cell cycle through the G2/M checkpoint, G1/S transition and DNA replication. Similarly, they participate in VEGF, hypoxia-inducible factor 1 and estrogen signaling pathways as well as the Toll-like receptor cascade.

A limitation of our study was lack of in vitro expression analysis of the shared genes. The fact that the EST profile of all protein-coding genes in tissues is yet to be completed can be considered another limitation of the present study. This may increase risk of systematic errors and bias.

CONCLUSION

We identified nine up-regulated genes shared between esophageal and liver cancers and their underlying molecular mechanisms. These genes could be valuable targets for diagnosis and treatment of cancer via gene therapy. More studies should be carried out to explore the potential usefulness of these genes as biomarkers for the two cancers. An *in silico* analysis of RNA-sequence data for the detection of the shared genes could confirm our results.

Conflict of interest
The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES
1. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer. 2010; 1805(1):105-17. [DOI:10.1016/j.bbcan.2009.11.002]

2. Xie H, Zhu WY, Wasserman A, Grebinskiy V, Olson A, Mintz L. Computational analysis of alternative splicing using EST tissue information. Genomics. 2002; 80(3):326-30. [DOI:10.1006/geno.2002.6841]

3. Aouacheria A, Navratil V, Barthelaix A, Mouchiroud D, Gautier C. Bioinformatic screening of human ESTs for differentially expressed genes in normal and tumor tissues. BMC genomics. 2006; 7(1):94. [DOI:10.1186/1471-2164-7-94]

4. Gray JW, Collins C. Genome changes and gene expression in human solid tumors. Carcinogenesis. 2000; 21(3):443-52. [DOI:10.1093/carcin/21.3.443]

5. Wang J, Liang P. DigiNorthern, digital expression analysis of query genes based on ESTs. Bioinformatics. 2003; 19(5):653-4. [DOI:10.1093/bioinformatics/btg044]

6. Lal A, Lash AE, Altschul SF, Velculescu V, Zhang L, McLendon RE, Marra MA, Prange C, Morin PJ, Polyak K, Papadopoulos N. A public database for gene expression in human cancers. Cancer Research. 1999; 59(21):5403-7.

7. Baranova AV, Lobashev AV, Ivanov DV, Krukovskaya LL, Yankovsky NK, Kozlov AP. In silico screening for tumour-specific expressed sequences in human genome. FEBS letters. 2001; 508(1):143-8. [DOI:10.1016/S0014-5793(01)03028-9]

8. Chen YC, Hsiao CD, Lin WD, Hu CM, Hwang PP, Ho JM. ZooDDD: a cross-species database for digital differential display analysis. Bioinformatics. 2006; 22(17):2180-2. [DOI:10.1093/bioinformatics/btl358]

9. Jiang J, Jacob HJ. EbEST: an automated tool using expressed sequence tags to delineate gene structure. Genome research. 1998; 8(3):268-75. [DOI:10.1101/gr.8.3.268]

10. Pontius JU, Wagner L, Schuler GD. 21. UniGene: A unified view of the transcriptome. The NCBI Handbook. Bethesda, MD: National Library of Medicine (US), NCBI. 2003.

11. Zhang J, Zhang L, Coombes KR. Gene sequence signatures revealed by mining the UniGene affiliation network. Bioinformatics. 2005; 22(4):385-91. [DOI:10.1093/bioinformatics/bti796]

12. Darvish H, Nabi MO, Firouzabadi SG, Karimlou M, Heidari A, Najmabadi H, Ohadi M. Exceptional human core promoter nucleotide compositions. Gene. 2011; 475(2):79-86. [DOI:10.1016/j.gene.2010.12.013]

13. Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene. 2007; 389(1):52-65. [DOI:10.1016/j.gene.2006.09.029]

14. Li YC, Korol AB, Fahima T, Beiles A, Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular ecology. 2002; 11(12):2453-65. [DOI:10.1046/j.1365-294X.2002.01643.x]

15. Li YC, Korol AB, Fahima T, Nevo E. Microsatellites within genes: structure, function, and evolution. Molecular biology and evolution. 2004; 21(6):991-1007. [DOI:10.1093/molbev/msh073]

16. Aouacheria A, Navratil V, Barthelaix A, Mouchiroud D, Gautier C. Bioinformatic screening of human ESTs for differentially expressed genes in normal and tumor tissues. BMC genomics. 2006; 7(1):94. [DOI:10.1186/1471-2164-7-94]

17. Mochida K, Kawaura K, Shimosaka E, Kawakami N, Shin T, Kohara Y, Yamazaki Y, Ogihara Y. Tissue expression map of a large number of expressed sequence tags and its application to in silico screening of stress response genes in common wheat. Molecular Genetics and Genomics. 2006; 276(3):304-12. [DOI:10.1007/s00438-006-0120-1]

18. Tang Y, Geng Y, Luo J, Shen W, Zhu W, Meng C, Li M, Zhou X, Zhang S, Cao J. Downregulation of ubiquitin inhibits the proliferation and radioresistance of non-small cell lung cancer cells in vitro and in vivo. Scientific reports. 2015; 5:9476. [DOI:10.1038/srep09476]

19. Kennedy D, Jäger R, Mosser DD, Samali A. Regulation of apoptosis by heat shock proteins. IUBMB life. 2014; 66(5):327-38. [DOI:10.1002/iub.1274]

20. Lianos GD, Alexiou GA, Mangano A, Mangano A, Rausei S, Boni L, Dionigi G, Roukos DH. The role of heat shock proteins in cancer. Cancer letters. 2015; 360(2):114-8. [DOI:10.1016/j.canlet.2015.02.026]

21. Xiang X, You XM, Li LQ. Expression of HSP90AA1/HSPA8 in hepatocellular carcinoma patients with depression. OncoTargets and therapy. 2018; 11:3013. [DOI:10.2147/OTT.S159432]

22. Xiang X, You XM, Li LQ. Expression of HSP90AA1/HSPA8 in hepatocellular carcinoma patients with depression. OncoTargets and

therapy.   2018;   11:3013. [DOI:10.2147/OTT.S159432]

23. Capello M, Ferri-Borgogno S, Riganti C, Chattaragada MS, Principe M, Roux C, Zhou W, Petricoin EF, Cappello P, Novelli F. Targeting the Warburg effect in cancer cells through ENO1 knockdown rescues oxidative phosphorylation and induces growth arrest. Oncotarget. 2016; 7(5):5598. [DOI:10.18632/oncotarget.6798]