



## Analysis of differentially expressed genes in malignant biliary strictures

D.C. Qi<sup>1,2\*</sup>, B. Wu<sup>1\*</sup>, S.L. Tao<sup>3</sup>, J. Zhou<sup>3</sup>, H.X. Qian<sup>2</sup> and D. Wang<sup>4</sup>

<sup>1</sup>Department of Surgery, Affiliated Sixth People's Hospital, Shanghai Jiaotong University, Shanghai, China

<sup>2</sup>Department of Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China

<sup>3</sup>Department of Surgery, Jinghong People's Hospital, Xishuangbanna, Yunnan, China

<sup>4</sup>Department of Gastroenterology, Shanghai Changhai Hospital, Shanghai, China

\*These authors contributed equally to this study.

Corresponding authors: H.X. Qian / D. Wang

E-mail: QianHaixin1@hotmail.com / dongwang@ailyun.com

Genet. Mol. Res. 13 (2): 2674-2682 (2014)

Received December 13, 2012

Accepted June 14, 2013

Published April 8, 2014

DOI <http://dx.doi.org/10.4238/2014.April.8.10>

**ABSTRACT.** Microarray data were collected from bile duct samples from subjects with malignant biliary strictures by endoscopic retrograde cholangiopancreatography to screen for key genes associated with this disease. A predicted interaction network was constructed for these genes to interpret their functions. The gene expression dataset GSE34166 (10 samples: 6 malignant and 4 benign control samples) was downloaded from the Gene Expression Omnibus database. R package scripts were used to process the data and screen for differentially expressed genes. Genes identified were uploaded to the analysis tool String 8.3 to generate a gene interaction network. A hub gene was identified by calculating the node degree. The interaction network of the hub gene with other genes in the human genome was constructed and screened (score >0.9), and pathway-enrichment analysis was performed to elucidate the hub gene

function. In total, 377 differentially expressed genes were identified and a network comprising 209 pairs of interactions was constructed. The most critical hub gene was identified as *GSTA1*, and a *GSTA1*-based interaction network was constructed consisting of 25 genes (containing the differentially expressed gene *GSTA3*). The cytochrome P450 (CYP450)-metabolic pathway displayed the most significant enrichment. Additionally, 4 transcription factors and their binding sites were also identified. In conclusion, we have identified the differentially expressed genes *GSTA1* (a hub gene) and *GSTA3*, which may cause abnormal gene expression and tumorigenesis through CYP450-metabolic pathways. The transcription factors and their binding sites in the promoter of the hub gene provide potential directions for future drug design.

**Key words:** Endoscopic retrograde cholangiopancreatography; Malignant biliary strictures; Protein interaction network; Hub gene; Transcription factor; *GSTA1*

## INTRODUCTION

Biliary strictures consist of abnormal narrowing of bile duct and are the result of different causes, the most common of which is surgical injury to the bile duct. Other risk factors include cancer of the bile duct (Gamblin et al., 2009), pancreatitis, and gallstones (Dooley, 2011). The condition can be divided into benign and malignant biliary strictures, the latter of which results from primary or metastatic cancers.

Endoscopic retrograde cholangiopancreatography (ERCP) is a minimally invasive surgery method to examine the pancreatic duct and was developed by McCune in 1968. Since then, endoscopic drainage, biliary lithotomy, endoscopic pancreatic duct drainage, and embolization have also been developed (Nourallah et al., 1999; Lehman, 2002). These new methods are used with additional techniques, such as high-frequency electric knife, laser, microwave, and electrohydraulic shock wave methods, which have further broadened the therapeutic range of ERCP. Today, ERCP is an important tool for diagnosing and treating pancreaticobiliary disease. Nevertheless, several risks remain especially serious postoperative complications that require vigilance in clinical practice.

With the recent advance of life science technologies for diagnosis and treatment of various tumor diseases, such as lung cancer, thyroid tumor, the use of genetics approaches has increasingly attracted attention (Bernig and Chanock, 2006; Garcia and Folpe, 2010). Progress has also been made in detecting malignant biliary strictures. For example, previous studies have identified a point mutation in the *KRAS* gene that can be used for diagnosing malignant and benign biliary strictures. Nischalke et al. (2012) have shown that, in addition of cytology, the use of RT-PCR for measuring the expression of insulin-like growth factor 2 mRNA binding protein 3 (*IGF2BP3*), *HOXB7*, and NIMA (never in mitosis gene A)-related kinase 2 (*NEK2*) genes improves the detection of malignant biliary disorders. Hass et al. (2008) have shown that osteopontin was the most consistently overexpressed gene in intrahepatic cholangiocarcinoma.

However, few studies have used high-throughput screening for and functional analysis of genes differentially expressed in malignant biliary strictures. To this end, we ana-

lyzed microarray data obtained from benign and malignant biliary stricture samples obtained by ERCP to identify hub and related genes through interaction network analysis by bioinformatic methods. These genes may be potential targets for future drug design or detection of cholangiocarcinoma.

## **MATERIAL AND METHODS**

### **Microarray data**

The data set GSE34166 (Chapman et al., 2012), including raw files and probe annotations, was downloaded from the Gene Expression Omnibus (GEO) database, which contains 10 biliary stricture samples, representing 4 benign and 6 malignant tissues (cholangiocarcinoma). Chip data were acquired by using the GPL570 [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array. Probe annotation was obtained from Affymetrix that included information for all the probes on the Affymetrix ATH1 (25K) chip.

### **Data processing and identification of differentially expressed genes**

Raw data were processed using the package Affy of R, including filling in of missing data with the KNN method (Marques et al., 2011) and data normalization with the median method (Fujita et al., 2006). The LIMMA package (Smyth, 2005) was used to identify differentially expressed gene, followed by multiple-testing corrections with Bayesian methods (Benjamini and Hochberg, 1995). A P value of  $<0.01$  and fold-changes in expression of  $>2$  or  $<-2$  were used as cutoffs to identify differentially expressed genes.

### **Screening for hub genes**

All genes identified as differentially expressed were uploaded to String (Szklarczyk et al., 2011), to identify interactions between gene products using database information as well as characteristics of the input sequence. According to known protein interaction networks, most of them possess a scale-free property (Albert, 2005), i.e., most nodes have only a few connections, and a few nodes in the network have a large number of connections. These few nodes are called critical network nodes (hub nodes). After the interaction network was generated, a node analysis was performed to screen of hub protein by using the scale-free property.

### **Construction of the interaction network for the hub gene**

The hub gene was analyzed again by String to establish its predicted interaction network. Interaction pairs with score  $>0.9$  were considered hub gene candidates.

### **Pathway-enrichment analysis**

All of the genes in the predicted interaction network were uploaded to the DAVID software (Huang et al., 2009), to calculate enriched pathways using hypergeometric distribution algorithm (false-discovery rate [FDR]  $<0.05$  and count  $>2$ ).

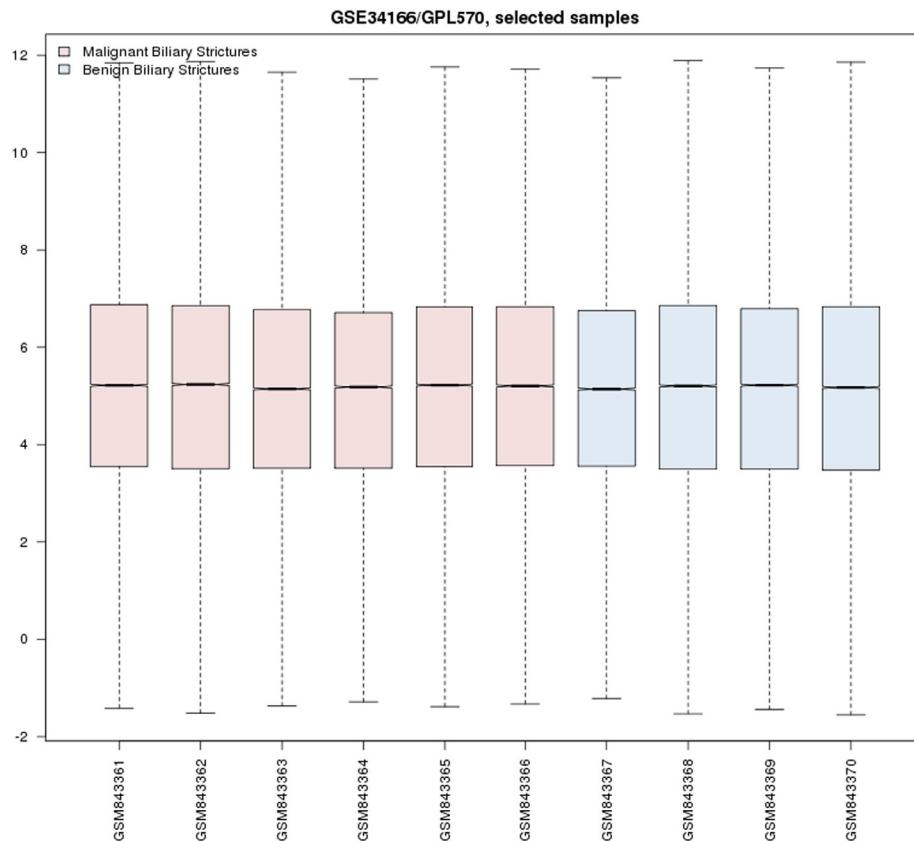
## Hub gene-related transcription factors (TFs)

The DNA-binding domain in transcription factors binds to cis-acting elements in DNA sequences (Stower, 2011), to inhibit or enhance gene expression. The text mining-based function of the free online tool EpiTect ChIP qPCR Primers (Han et al., 2010) was used to extract TFs of the hub gene from the published literature and other public resources.

## RESULTS

### Differentially expressed genes

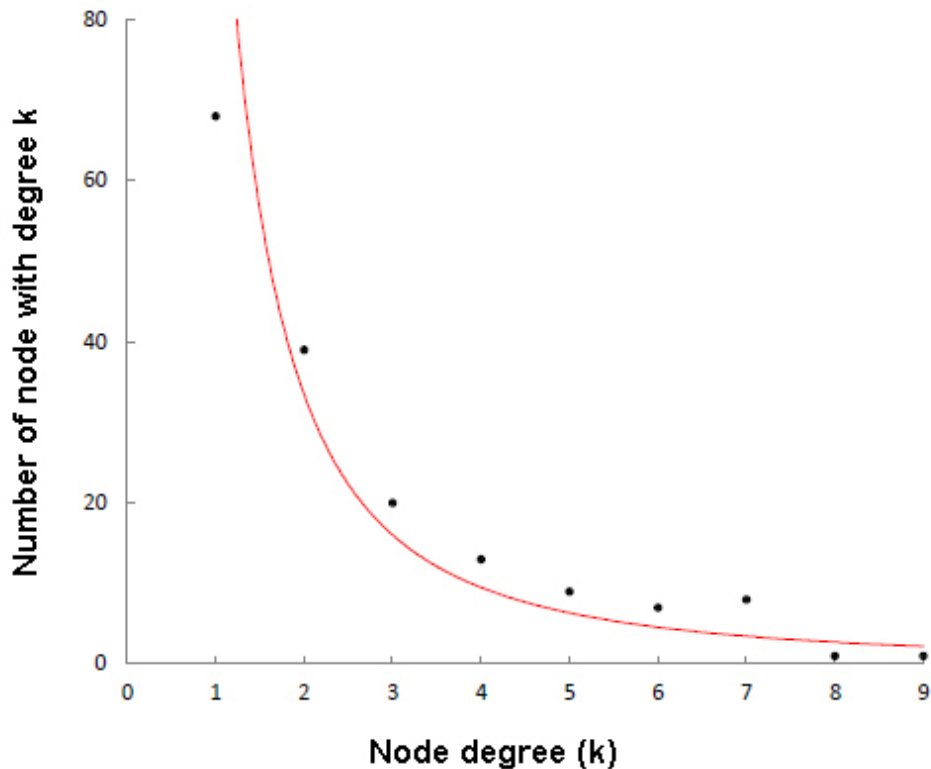
Raw data were normalized (see Figure 1) and then screened for differentially expressed genes. In total, 377 differentially expressed genes were identified according to the pre-set criteria ( $P < 0.01$ ,  $|\log_{2}FC| > 1$ ): 212 of these genes displayed down-regulated and 165 up-regulated expression.



**Figure 1.** Box plots for the normalized chip data, 4 blue boxes for benign biliary stricture samples and 6 pink boxes for malignant samples. Black line in the box represents the median of data set. The black lines are almost on the same line, indicating good effectiveness of normalization.

### Screening for hub genes

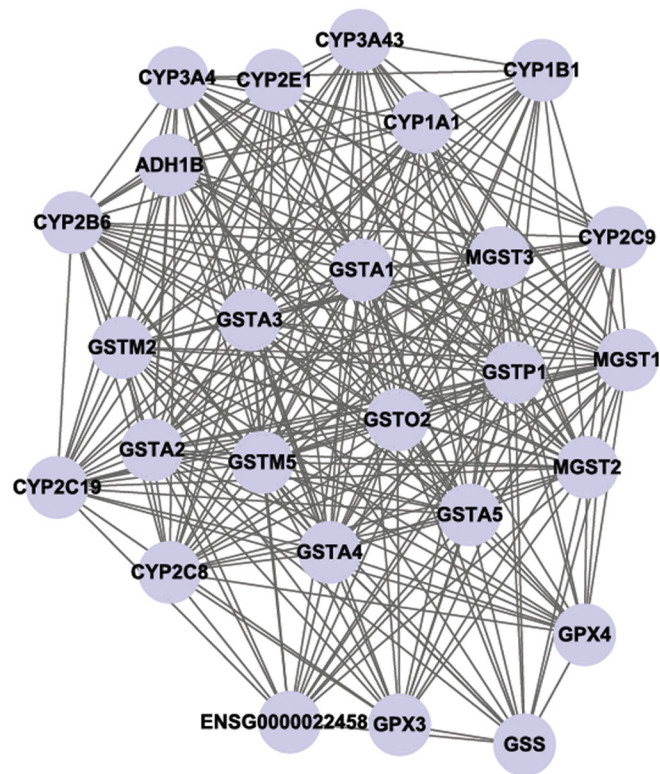
The protein interaction network for the differentially expressed genes was generated using the String software and 209 pairs of interaction were identified. Topological analysis was performed on the network and showed a scale-free property (Figure 2). Fitting of the power law generated the following equation:  $y = 117.8x^{(-1.81)}$  (see curve in Figure 2), wherein the x-axis is the node degree (k) and the y-axis is the number of nodes with degree k. As shown in figure, *GSTA1* (9) had the highest number of nodes, indicating its important role in the network and suggesting that *GSTA1* is a hub gene.



**Figure 2.** Distribution of node degree for the interaction network. The x-axis is the node degree (k), i.e., number of node directly connected to the specific node, and the y-axis is the number of node with degree k. The red line is a power function curve fitting the data points.

### Determining the functions for the hub gene

The newly identified hub gene *GSTA1* was set at the center to construct an interaction network consisting of 25 nodes against the background of the entire human genome (Figure 3). The differentially expressed gene *GSTA3* was included in the network, whereas most of the other gene members belonged to the cytochrome P450 (CYP450) family.



**Figure 3.** Predicted interaction network for the hub gene *GSTA1*.

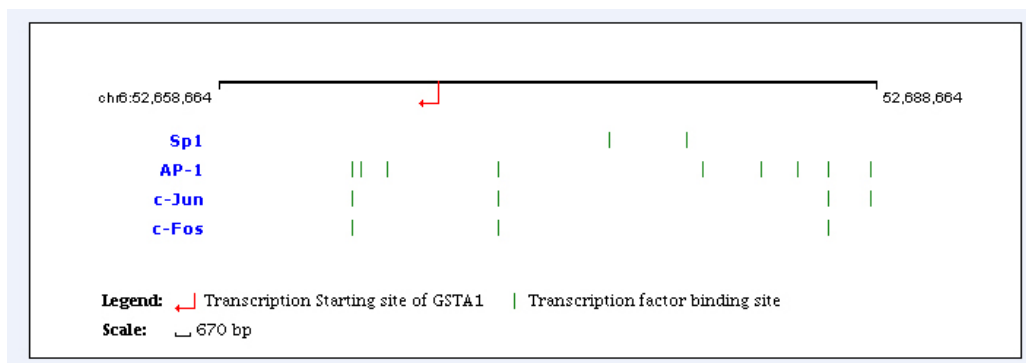
The 25 nodes were analyzed with the DAVID software, which uncovered 3 significantly enriched pathways (FDR < 0.05) (Table 1), in which metabolism of exogenous chemicals by CYP450 was the most significant pathway. As shown in Figure 3, most nodes interacting with *GSTA1* and *GSTA3* belonged to the CYP450 family. Thus, we speculate that the hub gene *GSTA1* and the differentially expressed gene *GSTA3* interact with CYP450 members and cause tumorigenesis through the pathways of exogenous metabolism by CYP450.

**Table 1.** Pathway-enrichment analysis result for all the genes in the predicted interaction network.

ID	Term	Count	P value	FDR
hsa00980	Metabolism of xenobiotics by cytochrome P450	22	1.24E-39	8.41E-37
hsa00982	Drug metabolism	20	8.35E-34	5.67E-31
hsa00480	Glutathione metabolism	15	1.97E-23	1.34E-20

FDR = false-discovery rate.

In addition, 4 transcription factors including Sp-1, Ap-1, c-Jun, and c-Fos that bind to the promoter of the key gene *GSTA1* and their corresponding binding sites were identified (Figure 4).



**Figure 4.** Transcription factors and binding sites for the hub gene *GSTA1*. Red arrow indicates the starting site of the transcription, and green bars indicate the binding sites for each transcription factor.

## DISCUSSION

Our analysis of the chip data from cholangiocarcinoma tissues (Chapman et al., 2012) identified several differentially expressed genes, including *GSTA1* and *GSTA3*. According to the network analysis, these two genes interacted with CYP450 members and thus participate in the CYP450 exogenous chemical metabolism.

Both *GSTA1* and *GSTA3* encode glutathione S-transferases (GST), which are involved in detoxification of electrophilic compounds, including drugs, toxins, and carcinogens, by conjugating these compounds with glutathione (Khan et al., 2011). Kobayashi et al. (2000) reported a GST as a new sensitive marker of hepatocellular damage in biliary atresia, and Nakajima et al. (2003) reported that GST-pi is involved in resistance to anti-cancer drugs used for treating cholangiocarcinoma. Daorueang et al. (2012) uncovered the underlying mechanism for GST-promoted cell proliferation that accelerates the formation of cholangiocarcinoma.

CYP450 is a protein family widely distributed in human tissues and requires an iron porphyrin as prosthetic group. Most members are monooxygenases that catalyze a variety of metabolic reactions (Bruno and Njar, 2007). The CYP450 superfamily can be broadly divided into two categories according to the metabolic substrates used: one contains P450s for the metabolism of endogenous compounds (such as steroids), and the other for the metabolism of exogenous substances (such as drugs, poisons, and carcinogens or mutagens). The pathway interacting with the hub gene is involved in oxidation, reduction of exogenous substances, as well as oxidative metabolism, and thus plays an important role in activating prodrugs or carcinogens or degrading drugs or poisons (Karlgrén et al., 2006). CYP450 enzymes frequently produce intermediates or final products that have strong affinities for electrons and these intermediates may interact with nucleophilic groups in intracellular macromolecules, such as DNAs, RNAs, or proteins. This may cause destruction of cell structures, enzyme inactivation, gene mutations, or suppression of gene expression, and ultimately result in cell damage, programmed death, or tumorigenesis (Iizasa et al., 2005). Changes in CYP450 genes can lead to increased or decreased P450 activity or complete inactivation of the affected enzyme (Klotz, 2006), and increased expression of CYP450 has been associated with cancer (Gharavi and El-

Kadi, 2004; Bandala et al., 2012). Several studies have reviewed the association of CYP450 polymorphisms with susceptibility to cancer (Han and Zhou, 2000; Sergentanis et al., 2012), and because CYP450 enzymes are involved in the metabolism of many drugs, CYP450 activities have to be considered in the design of anticancer therapies (van Schaik, 2008).

Because GSTA1, GSTA3, and CYP450 have similar functions, it is not unexpected that these proteins strongly interact with each other. Several studies have indicated that these proteins are involved in drug metabolism and are therefore associated with susceptibility to cancer (Forrester et al., 1990; Lin et al., 1998; Tan et al., 2000) and other diseases (Burim et al., 2004). We believe that further studies are needed to investigate the associations between these proteins to improve our understanding of the mechanisms and processes underlying the formation of malignant biliary strictures.

In addition, TFs have attracted increasing attention because they are involved in regulating gene expression and thus play an important role in the onset, development, invasion, and metastasis of tumors. Because TFs are potential targets of new anticancer drugs (Darnell Jr., 2002), we screened for the TFs that control expression of the *GSTA1* gene and for their binding sites in the *GSTA1* promoter. This analysis yielded identification of the 4 TFs including Sp1 (Ebert et al., 2003), AP-1 (Romero et al., 2006), c-Jun (Hayes and McMahon, 2001), and c-Fos (Tang et al., 2008) and of their binding sites in the *GSTA1* promoter. The differential expression in *GSTA1* genotypes defined by the C-69T substitution has been attributed to an alteration of the Sp1 transcription factor (Tijhuis et al., 2005). c-Jun participates in transducing chemical stresses like chemopreventive blocking agents (Hayes and McMahon, 2001). Therefore, these TFs may be potential drug targets worthy of further characterization.

## REFERENCES

- Albert R (2005). Scale-free networks in cell biology. *J. Cell Sci.* 118: 4947-4957.
- Bandala C, Floriano-Sánchez E, Cárdenas-Rodríguez N, López-Cruz J, et al. (2012). RNA expression of cytochrome P450 in Mexican women with breast cancer. *Asian Pac. J. Cancer Prev.* 13: 2647-2653.
- Benjamini Y and Hochberg Y (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B* 57: 289-300.
- Bernig T and Chanock SJ (2006). Challenges of SNP genotyping and genetic variation: its future role in diagnosis and treatment of cancer. *Expert Rev. Mol. Diagn.* 6: 319-331.
- Bruno RD and Njar VC (2007). Targeting cytochrome P450 enzymes: a new approach in anti-cancer drug development. *Bioorg. Med. Chem.* 15: 5047-5060.
- Burim RV, Canalle R, Martinelli AL and Takahashi CS (2004). Polymorphisms in glutathione S-transferases GSTM1, GSTT1 and GSTP1 and cytochromes P450 CYP2E1 and CYP1A1 and susceptibility to cirrhosis or pancreatitis in alcoholics. *Mutagenesis* 19: 291-298.
- Chapman MH, Tidswell R, Dooley JS, Sandanayake NS, et al. (2012). Whole genome RNA expression profiling of endoscopic biliary brushings provides data suitable for biomarker discovery in cholangiocarcinoma. *J. Hepatol.* 56: 877-885.
- Daorueang D, Thuwajit P, Roittrakul S, Laha T, et al. (2012). Secreted *Opisthorchis viverrini* glutathione S-transferase regulates cell proliferation through AKT and ERK pathways in cholangiocarcinoma. *Parasitol. Int.* 61: 155-161.
- Darnell JE Jr (2002). Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer* 2: 740-749.
- Dooley JS (2011). Gallstones and Benign Biliary Diseases. In: Sherlock's Diseases of the Liver and Biliary System (Dooley JS, Lok ASF, Burroughs AK and Heathcote EJ, eds.). 12th edn. Wiley-Blackwell, Oxford, 257-293.
- Ebert MN, Klinder A, Peters WH, Schaferhenrich A, et al. (2003). Expression of glutathione S-transferases (GSTs) in human colon cells and inducibility of GSTM2 by butyrate. *Carcinogenesis* 24: 1637-1644.
- Forrester LM, Hayes JD, Millis R, Barnes D, et al. (1990). Expression of glutathione S-transferases and cytochrome P450 in normal and tumor breast tissue. *Carcinogenesis* 11: 2163-2170.
- Fujita A, Sato JR, Rodrigues Lde O, Ferreira CE, et al. (2006). Evaluating different methods of microarray data normalization. *BMC Bioinformatics* 7: 469.



- Gamblin TC, Krasinskas AM, Slivka AS, Tublin ME, et al. (2009). Fibroinflammatory biliary stricture: a rare bile duct lesion masquerading as cholangiocarcinoma. *J. Gastrointest. Surg.* 13: 713-721.
- Garcia JJ and Folpe AL (2010). The impact of advances in molecular genetic pathology on the classification, diagnosis and treatment of selected soft tissue tumors of the head and neck. *Head Neck Pathol.* 4: 70-76.
- Gharavi N and El-Kadi AO (2004). Expression of cytochrome P450 in lung tumor. *Curr. Drug Metab.* 5: 203-210.
- Han L, Suzek TO, Wang Y and Bryant SH (2010). The text-mining based PubChem bioassay neighboring analysis. *BMC Bioinformatics* 11: 549.
- Han XM and Zhou HH (2000). Polymorphism of CYP450 and cancer susceptibility. *Acta Pharmacol. Sin.* 21: 673-679.
- Hass HG, Nehls O, Jobst J, Friling A, et al. (2008). Identification of osteopontin as the most consistently over-expressed gene in intrahepatic cholangiocarcinoma: Detection by oligonucleotide microarray and real-time PCR analysis. *World J. Gastroenterol.* 14: 2501-2510.
- Hayes JD and McMahon M (2001). Molecular basis for the contribution of the antioxidant responsive element to cancer chemoprevention. *Cancer Lett.* 174: 103-113.
- Huang W, Sherman BT and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57.
- Iizasa T, Baba M, Saitoh Y, Suzuki M, et al. (2005). A polymorphism in the 5'-flanking region of the CYP2E1 gene and elevated lung adenocarcinoma risk in a Japanese population. *Oncol. Rep.* 14: 919-923.
- Karlgren M, Gomez A, Stark K, Svard J, et al. (2006). Tumor-specific expression of the novel cytochrome P450 enzyme, CYP2W1. *Biochem. Biophys. Res. Commun.* 341: 451-458.
- Khan MS, Khan MK, Siddiqui MH and Arif JM (2011). An *in vivo* and *in silico* approach to elucidate the tocotrienol-mediated fortification against infection and inflammation induced alterations in antioxidant defense system. *Eur. Rev. Med. Pharmacol. Sci.* 15: 916-930.
- Klotz U (2006). Clinical impact of CYP2C19 polymorphism on the action of proton pump inhibitors: a review of a special problem. *Int. J. Clin. Pharmacol. Ther.* 44: 297-302.
- Kobayashi H, Horikoshi K, Yamataka A, Lane GJ, et al. (2000). Alpha-glutathione-S-transferase as a new sensitive marker of hepatocellular damage in biliary atresia. *Pediatr. Surg. Int.* 16: 302-305.
- Lehman GA (2002). What are the determinants of success in utilization of ERCP in the setting of pancreatic and biliary diseases? *Gastrointest. Endosc.* 56: S291-S293.
- Lin DX, Tang YM, Peng Q, Lu SX, et al. (1998). Susceptibility to esophageal cancer and genetic polymorphisms in glutathione S-transferases T1, P1, and M1 and cytochrome P450 2E1. *Cancer Epidemiol. Biomarkers Prev.* 7: 1013-1018.
- Marques FZ, Campain AE, Tomaszewski M, Zukowska-Szczechowska E, et al. (2011). Gene expression profiling reveals renin mRNA overexpression in human hypertensive kidneys and a role for microRNAs. *Hypertension* 58: 1093-1098.
- Nakajima T, Takayama T, Miyaniishi K, Nobuoka A, et al. (2003). Reversal of multiple drug resistance in cholangiocarcinoma by the glutathione S-transferase-pi-specific inhibitor O1-hexadecyl-gamma-glutamyl-S-benzylcysteinyl-D-phenylglycine ethylester. *J. Pharmacol. Exp. Ther.* 306: 861-869.
- Nischalke HD, Schmitz V, Luda C, Aldenhoff K, et al. (2012). Detection of IGF2BP3, HOXB7, and NEK2 mRNA expression in brush cytology specimens as a new diagnostic tool in patients with biliary strictures. *PLoS One* 7: e42141.
- Nourallah H, Issa H and Al-Salem AH (1999). The role of ERCP in the evaluation, diagnosis, and therapy of biliary and pancreatic diseases in children. *Ann. Saudi Med.* 19: 163-166.
- Romero L, Ng L and Kirby GM (2006). Chemical inducers of rodent glutathione S-transferases down-regulate human GSTA1 transcription through a mechanism involving variant hepatic nuclear factor 1-C. *Mol. Pharmacol.* 70: 277-286.
- Sergentanis TN, Economopoulos KP, Choussein S and Vlahos NF (2012). Cytochrome P450 1A1 (CYP1A1) gene polymorphisms and ovarian cancer risk: a meta-analysis. *Mol. Biol. Rep.* 39: 9921-9930.
- Smyth GK (2005). Limma: Linear Models for Microarray Data. In: Bioinformatics and Computational Biology Solutions Using R and Bioconductor (Gentleman R, Carey V, Dudoit S, Irizarry R, et al., eds.). Springer, New York, 397-420.
- Stower H (2011). Gene regulation: Resolving transcription factor binding. *Nat. Rev. Genet.* 13: 71.
- Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39: D561-D568.
- Tan W, Song N, Wang GQ, Liu Q, et al. (2000). Impact of genetic polymorphisms in cytochrome P450 2E1 and glutathione S-transferases M1, T1, and P1 on susceptibility to esophageal cancer among high-risk individuals in China. *Cancer Epidemiol. Biomarkers Prev.* 9: 551-556.
- Tang NJ, Liu J, Coenraads PJ, Dong L, et al. (2008). Expression of AhR, CYP1A1, GSTA1, c-fos and TGF-alpha in skin lesions from dioxin-exposed humans with chloracne. *Toxicol. Lett.* 177: 182-187.
- Tijhuis MJ, Wark PA, Aarts JM, Visker MH, et al. (2005). GSTP1 and GSTA1 polymorphisms interact with cruciferous vegetable intake in colorectal adenoma risk. *Cancer Epidemiol. Biomarkers Prev.* 14: 2943-2951.
- van Schaik RH (2008). CYP450 pharmacogenetics for personalizing cancer therapy. *Drug Resist. Updat.* 11: 77-98.