Serial analysis of gene expression in adenocarcinoma samples and normal colonic mucosa in a Chinese population

X.L. Zhang[1], F. Gao[1], W. Li[2], W.Z. Tang[1] and S. Zhang[1]

[1]Department of Colorectum Anus Surgery,
The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, China
[2]Center of Laboratory, Guangxi Medical University, Nanning, Guangxi, China

Corresponding author: F. Gao
E-mail: gaofenggx@163.com

ABSTRACT. The goal of this study was to identify the differences in gene expression between adenocarcinoma samples and colonic normal mucosa of a Chinese population. Gene expression libraries of adenocarcinoma and normal colonic mucosa were generated by serial analysis of gene expression (SAGE). Expression tags were obtained from sequencing results using the SAGE2000 software. Representative genes of tags were determined based on the SAGEmap from NCBI. Expressed genes were selected and their expression in different libraries was compared using the SPSS 17.0 software. The expression of 78 tags representing 88 types of genes showed significant differences. Compared with normal mucosa, 38 genes showed decreased expression in cancer, whereas the expression of the remaining 50 genes showed significant up-regulation. Expressed tags in SAGE libraries of normal and cancerous tissues were significantly different. Further studies examining these genes, which showed different

expression levels between the 2 tissues, may be used as tumor markers and offer clues for studying the etiopathogenesis of colon cancer.

Key words: Colon normal mucosa; Colonic adenocarcinoma; Difference; Serial analysis of gene expression

INTRODUCTION

Human colon cancer occurs worldwide with a variable geographic distribution and ranks third in order of occurrence and fifth as the leading cause of cancer mortality, affecting more men than women (Willaert and Ceelen, 2015). Colonic adenocarcinoma is the most frequent subtype of colon cancer, the incidence of which is increasingly quickly in China. Despite the use of modern surgical techniques combined with various therapy modalities, such as chemotherapy and radiotherapy, the overall 5-year survival rate of colon cancer remains at 50% (Missiaglia et al., 2014). Therefore, it is very important to identify new diagnostic markers and therapeutic targets for colonic adenocarcinoma.

Research on molecular level is important for improving tumor diagnosis and treatment. After the human genome project was completed, research aims regarding tumors entered the post-genomic era. An increased understanding of changes in gene expression that occur during carcinogenesis may lead to improvements in the prevention of colonic adenocarcinoma. Genes encoding secretory/transmembrane proteins presented specifically in cancers may be used as diagnostic biomarkers (Buckhaults et al., 2001). Genes can not only act as biomarkers, but also may be therapeutic targets (Yasui et al., 2004). Comprehensive gene expression analysis might be useful for identifying potential markers for early detection and therapeutic targets for colonic adenocarcinoma. The construction of gene expression library is a useful method for studying gene expression. By analyzing a gene expression library, the expression information of nearly all genes can be obtained and their alterations can be compared with other related libraries. Additionally, in the expression library, we can analyze the interaction of gene expression, identify new tumor markers, gene expression, and even new genes, and explore the etiopathogenesis and treatment methods for cancer. Serial analysis of gene expression (SAGE) is a common and powerful approach used for constructing expression libraries (Velculescu et al., 1995). This technique has been widely used in cancer studies because new genes can be identified. Two SAGE libraries of colon cancer were constructed in the US in 1996. However, no SAGE libraries for colon cancer in Chinese populations have been established.

In the present study, we generated a SAGE library from one colonic adenocarcinoma sample from a Chinese patient. By comparing SAGE tags from colonic adenocarcinoma with those from the normal human colonic mucosa, we identified 78 tags and 88 genes that were differentially expressed between these 2 tissues.

MATERIAL AND METHODS

Tissue samples

For SAGE analysis, one primary colonic adenocarcinoma (tubular adenocarcinoma Class

II, T4N0M0) sample from a Chinese patient was used. It was confirmed microscopically that the tumor specimens consisted mostly of carcinoma tissue. The samples were obtained following surgery of one male patient diagnosed sigmoid colon carcinoma. Colonic carcinoma tissue samples and corresponding normal colonic mucosa samples (>12 cm above the tumor segment) were used. Samples were frozen immediately in liquid nitrogen and stored until use. The patient signed informed consent for the use of his biopsy material.

RNA isolation

Total RNA was extracted from biopsies using Trizol Reagent (Invitrogen, Carlsbad, CA, USA). Briefly, 200 µL Trizol was added to lyse the tissues. After phenol/chloroform extraction, isopropanol was used to precipitate RNA. Next, RNA was washed with 70% ethanol and air-dried. The sample was dissolved in RNase-free H_2O and stored at -80°C. The concentration of total RNA was determined by spectrophotometry.

SAGE procedure

Two SAGE libraries were obtained according the SAGE protocol as described by Velculescu et al. (1995). Briefly, 5-10 µg polyA + RNA was isolated from 1 mg total RNA using MagneSphere Magnetic Separation Products (Promega, Madison, WI, USA). First- and second-strand cDNA was synthesized using biotinylated poly A primers, DNA polymerase, and M-MLV reverse transcriptase. Cleavage of cDNA was carried out by the anchoring enzyme HinIII. The biotinylated cDNA was connected with magnetic beads. Bound cDNA was ligated with linkers. Next, blunt-end cDNA tags were released using tagging enzyme. The tags were ligated to form 102-bp ditag concatemers. The polymerase chain reaction (PCR) amplification of ditags is shown in Figure 1. The ditags were isolated, purified, and digested (Figure 2).
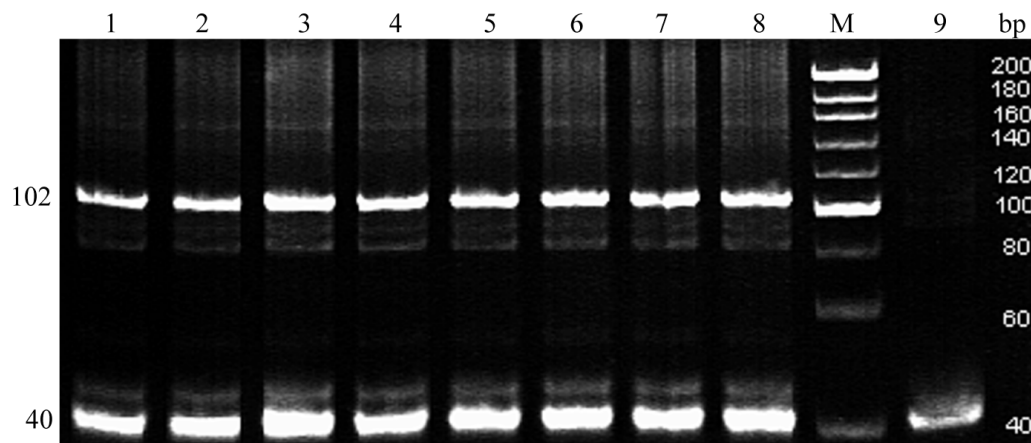

Figure 1. PCR analysis of ditags M: DNA marker; 1-8: ditags; 9: negative control.
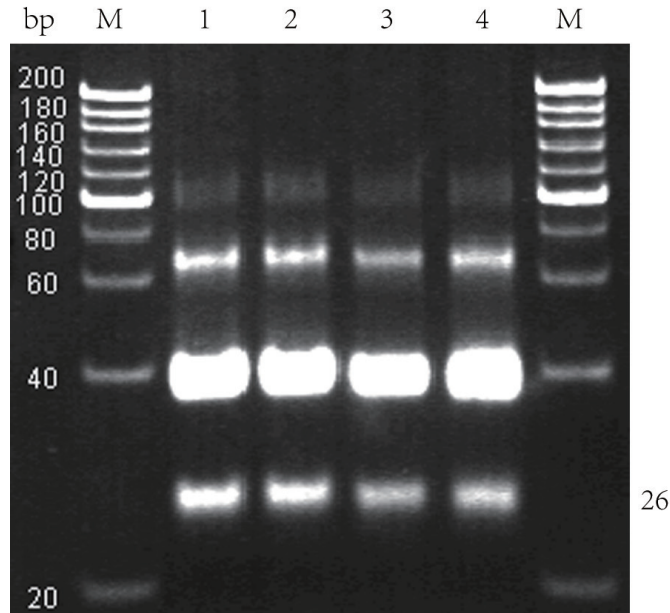
Figure 2. Isolation and purification of 26 bp ditags. M: DNA marker; 1-4: ditags.

The 26-bp ditags were then ligated to form concatemers. The isolated concatemers, consisting of serially ligated tags, were ligated into the pUCL-8 vector and transformed in *Escherichia coli* cells. PCR was performed on positive colonies using the specific primer M13, and PCR products were analyzed by agarose gel electrophoresis. The colonies were sequenced when the length of their PCR products was more than 400 bp. The results of sequencing revealed 4 bases (CATG) repeated at an interval of 26 bp, as shown in Figure 3.
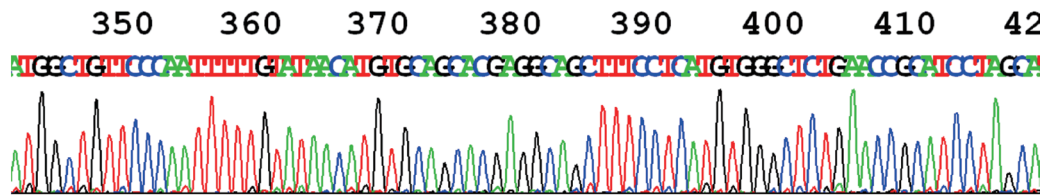

Figure 3. Sequence analysis of clones showing 4-base repetition (CATG) with an interval of 26 bp.

SAGE and statistical analysis

The expression tags were obtained from the sequencing results using SAGE2000 software. Representative genes of tags were identified using SAGEmap from NCBI. Highly expressed genes were selected and their expression in different libraries was compared using SPSS 17.0 software (SPSS, Inc., Chicago, IL, USA) with the Chi-squared test. $P < 0.01$ was regarded as statistically significant.

RESULTS

SAGE library characteristics

Based on the SAGE libraries of normal mucosa and adenocarcinoma, 250 and 233 clones, respectively, were selected and sequenced. Respective totals of 7926 and 7672 tags were obtained from the sequencing results using of SAGE2000 software. In the libraries of colon cancer and normal mucosa, 123 tags expressed 16 times. Among these, 45 tags showed no significant difference between 2 libraries, while the other 78 tags showed obvious differences. Five tags differently expressed were not identified in the SAGEmap. Because one tag may represent several different genes or several tags may suggest one gene, the remaining 73 tags represented a total of 88 different genes.

Comparison of expression profiles of adenocarcinoma and normal mucosa

We identified 38 genes with decreased expression in tumor tissues. Genes corresponding to tags showing more than a 5-fold down-regulation in colon cancer are shown in Table 1. Some genes, including *FABP1*, *SF3B2*, *SELENBP1*, *AASS*, *LGALS3*, *WDR36*, *CCNB1IP1*, *ZNF394*, *FXYD3*, *FLJ31401*, *FTH1*, and *CKB*, among others, decreased by less than 5-fold in tumor tissues. Among these, 3 tags represented the *MGC27165* gene.

Table 1. Five folds or more down-regulated tags in colorectal adenocarcinoma (CA) compared to normal colonic mucosa (normal).

Tag sequence	Tag per mill		Symbol and Description	P value
	CA	Normal		
GGAAAACAGA	11	56	SBZF3 Zinc finger protein 695	<0.01
GGAAAACAGA	11	56	LGALS4 Lectin, galactoside-binding, soluble, 4 (galectin 4)	<0.01
CTCCCCCAAG	2	56	MGC27165 Hypothetical protein MGC27165	<0.01
GCCCAGGTCA	3	41	FLJ10815 Amino acid transporter	<0.01
CTCCCCCAAA	1	36	MGC27165 Hypothetical protein MGC27165	<0.01
GCGGAGGTGG	2	31	MGC27165 Hypothetical protein MGC27165	<0.01
GTGCACTGAG	3	30	HLA-A Major histocompatibility complex, class I, A	<0.01
GTGCACTGAG	3	30	HLA-C Major histocompatibility complex, class I, C	<0.01
ATACTCCACT	0	30	GUCA2A Guanylate cyclase activator 2A (guanylin)	<0.01
TGCTCCTACC	0	29	FCGBP Fc fragment of IgG binding protein	<0.01
GACCCAAGAT	1	28	PIGR Polymeric immunoglobulin receptor	<0.01
GAAATACAGT	4	27	CTSD Cathepsin D (lysosomal aspartyl peptidase)	<0.01
GAAATACAGT	4	27	NT5C 5', 3'-nucleotidase, cytosolic	<0.01
GACATCAAGT	3	26	KRT19 Keratin 19	<0.01
GCGAAACCCT	4	25	C10orf95 Chromosome 10 open reading frame 95	<0.01
GCGAAACCCT	4	25	DKFZp761E198 Hypothetical protein DKFZp761E198	<0.01
CTCATAAGGA	4	23	ZF HCF-binding transcription factor Zhangfei	<0.01
CCAGGGGAGA	2	22	SDHA Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	<0.01
CCAGGGGAGA	2	22	IFI27 Interferon, alpha-inducible protein 27	<0.01
GAAATAAAGC	1	22	IGHG1 Immunoglobulin heavy constant gamma 1 (G1m marker)	<0.01
GAAATAAAGC	1	22	IGHG3 Immunoglobulin heavy constant mu	<0.01
TATAGTCCTC	2	19	CLDN7 Claudin 7	<0.01
AGGTGGCAAG	3	18	ANGPTL4 Angiopoietin-like 4	<0.01
TACATAATTA	2	18	TncRNA Trophoblast-derived noncoding RNA	<0.01
ACACAGCAAG	3	17	EMID2 EMI domain containing 2	<0.01
CATAGGTTTA	1	16	SLC26A3 Solute carrier family 26, member 3	<0.01

Fifty genes, including 28 ribosomal protein genes, were expressed highly in colon cancer. Table 2 shows examples of genes corresponding to tags up-regulated by more than 5-fold. Other genes with less than 5-fold up-regulation in tumor tissues included *RBM9*, *KIAA1068*, *BRCA1*, *CEBPA*, *NET-5*, *LTBP3*, *LOC388524*, and *KIAA1602* and ribosomal protein genes such as *RPL41*, *L10*, *LP2*, *L21*, *L13*, *L18a*, *S27*, *S3A*, *L8*, *L13a*, *L11*, *SA*, *L23*, *L39*, *L30*, and *S12*, among others.

Many tags showed no representative genes in SAGEmap. These included 9 genes that expressed 16 times or more. Among these, the expression of 4 tags was not significant between the 2 libraries, while CTAACTAGTT, CAAGCATCCC, CCCATCGTTC, CTAAGACTTC, and CACCTAAT TG were significantly down-regulated in cancer.

Table 2. Five folds or more up-regulated tags in colorectal adenocarcinoma (CA) compared to normal colonic mucosa (normal).

Tag sequence	Tag per mill		Symbol and Description	P value
	CA	Normal		
AATCCTGTGG	54	10	RPL8 Ribosomal protein L8	<0.01
CAATAAATGT	50	4	RPL37 Ribosomal protein L37	<0.01
GAGGGAGTTT	48	7	RPL27A Ribosomal protein L27a	<0.01
CGCCGCCGGC	45	9	RPL35 Ribosomal protein L35	<0.01
TAGGTTGTCT	45	5	TPT1 Tumor protein, translationally-controlled 1	<0.01
GGCAAGCCCC	37	3	PTPRG Protein tyrosine phosphatase, receptor type, G	<0.01
GGCAAGCCCC	37	3	RPL10A Ribosomal protein L10a	<0.01
TTCAATAAAA	31	1	RPLP1 Ribosomal protein, large, P1	<0.01
CTGGGTTAAT	30	5	RPS19 Ribosomal protein S19	<0.01
ATTATTTTTC	30	1	-Transcribed locus, moderately similar to XP_519807.1 PREDICTED: similar to 60S ribosomal protein L7 [Pan troglodytes]	<0.01
ATTATTTTTC	30	1	-Full-length cDNA clone CS0CAP007YH10 of Thymus of Homo sapiens (human)	<0.01
ACCTGTATCC	23	1	IFITM3 Interferon induced transmembrane protein 3 (1-8U)	<0.01
TGTGCTAAAT	20	3	-Similar to Zinc finger protein 492	<0.01
TGTGCTAAAT	20	3	USP36 Ubiquitin specific peptidase 36	<0.01
TGTGCTAAAT	20	3	RPL34 Ribosomal protein L34	<0.01
GCTTTTAAGG	20	1	RPS20 Ribosomal protein S20	<0.01
AAGACAGTGG	19	3	HDAC2 Histone deacetylase 2	<0.01
CCTAGCTGGA	19	2	LOC388817 Peptidylprolyl isomerase A-like	<0.01
CCTAGCTGGA	19	2	KNSL7 Kinesin family member 15	<0.01
CTGTTGGTGA	17	1	KIAA1036 Vasohibin 1	<0.01
CTGTTGGTGA	17	1	-Full-length cDNA clone CS0DJ002YF04 of T cells (Jurkat cell line) Cot 10-normalized of Homo sapiens (human)	<0.01
AATAGGTCCA	17	1	RPS25 Ribosomal protein S25	<0.01
CACTTGCCCT	16	3	NDUFB9 NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9, 22kDa	<0.01
CACTTGCCCT	16	3	ACAS2 Acyl-CoA synthetase short-chain family member 2	<0.01
ATCAAGGGTG	16	2	RPL9 Ribosomal protein L9	<0.01
AAGAAGATAG	16	1	- Luminal binding protein 1 (BiP-1) (BP1)	<0.01
AAGAAGATAG	16	1	RPL23A Ribosomal protein L23a	<0.01

DISCUSSION

In the study, SAGE technology was used to identify the entire transcription profile of colonic adenocarcinoma and compared to the profile of the normal colon mucosa. The specific information obtained from the study will enable identification of factors involved in the process of colonic carcinogenesis and identification of uniquely expressed tissue-specific genes. Moreover, this information can be applied to obtain insight into the biology of neoplastic lesions and may ultimately lead to better management of colonic adenocarcinoma.

Compared with expression microarrays and other gene chip technologies, the principal advantage of SAGE is that it generates a library of thousands of expressed genes without the requirement of previous knowledge of the cell's repertoire. The transcriptome gained from SAGE technology conveys

not only the identity of each expressed gene but also quantifies its level of expression.

We used tissue samples from one male individual diagnosed with colonic adenocarcinoma to prepare the SAGE library and to compare this information with corresponding normal colonic mucosa. We used the full tumor tissue, which provided more accurate insight into the gene expression profiles of important malignant conditions. In this study, more than 15,598 tags were analyzed. Comparison of the SAGE-generated tag expression profiles of tumor and normal tissue revealed that hundreds of transcripts were expressed differentially. Because of sequencing errors, singleton tags should be carefully considered. A malignant biopsy may contain a larger fraction of tumor than stromal cells. The cancers may be significantly heterogeneous, which should be taken into account when interpreting these expression libraries. It was necessary to verify the RNA expression levels of genes of interest on a larger panel of samples to confirm the SAGE results. In addition, some genes may not have been expressed. In this study, approximately 25% of tags in each library were singleton tags; this is in accordance with previously reported SAGE libraries (Ibrahim et al., 2005).

Previous gene expression profiles analyses have been performed to compare colon cancer and normal tissues to obtain an overview of genes involved in colon cancer. Seventy-two colon cancer-specific genes were identified using cDNA microarrays by comparing expression profiles in human colon biopsies of tumor tissue and normal tissue (van Erk et al., 2005). Huang et al. (2006) reported 136 down-regulated genes and 54 up-regulated genes in colon cancer using a combination of CGAP analysis and RNA interference.

In this study, mapping the SAGE tags to known genes and mRNAs in the SAGE Genie database suggested that a large number of genes were known to be expressed in colon cancer and many genes were not previously recognized in colonic adenocarcinoma. For instance, the tumor SAGE library confirmed low expression of *FLJ10815*, *GUCA2A*, *SDHA*, *FXYD3*, *ANGPTL4*, *SLC26A3*, *FABP1*, *SELENBP1*, *AASS*, *FLJ31401*, and *CKB*, which are known to be related to material transport and metabolism. Previous studies showed that these genes were involved in grade malignancy and tumor prognosis. Particularly, *FLJ10815* over-expression may promote cancer cell proliferation, while *ANGPTL4* was related to the migration and invasion of cancer. Moreover, some genes associated with cell cycle control, nucleic acid metabolism, immunologic function, gene expression and regulation, and cell-cell and cell-matrix interaction are down-regulated in tumor tissues. For example, *CCNB1IP1*, known to be expressed at low levels in many types of tumors such as non-small cell lung cancer, may degrade cyclin B1 and regulate mitosis (Confalonieri et al., 2009). Immunologic inadequacy is associated with tumor development. *PIGR*, which is down-regulated in nasopharynx cancer and colon cancer, was found to play an important role in immunologic defense (Chang et al., 2005). The expression of *CTSD*, known to be associated with the cancer invasion and metastasis through digestion of the extracellular matrix and destruction of the basement membrane (Minarowska et al., 2008), was shown to be significantly decreased in this colon cancer. However, further studies are needed to confirm these results. In the comparison of SAGE-generated gene expression profiles of colon cancer and normal tissues, 50 genes were significantly highly expressed and were identified in tumor tissues. Most were ribosomal protein genes. The expression levels of some ribosomal protein genes in colonic adenocarcinoma were similar to those observed in previous studies. However, others were not, indicating the diverse functions of ribosomal proteins. The expression of genes related to cell apoptosis regulation and gene expression and regulation, such as *TPT1*, *KIAA1068*, *BRCA1*, *NET-5*, *LTBP3*, *KIAA1602*, *PTPRG*, *USP36*, *KNSL7*, *BP1*, *RBM9*, *CEBPA*, and *HDAC2*, was increased in colon cancer. *TPT1*, an anti-apoptosis gene reported to be correlated with drug resistance and cancer prognosis, and *HDAC2*, known to be associated with drug resistance of

tumors, were up-regulated in many tumor types (Graidist et al., 2004; Fritsche et al., 2009). *BP1* may lead to tumor development by regulating Cyclin D1 (Moghaddam et al., 2007). The *CEBPA* gene has been associated with acute leukemia (Pabst and Mueller, 2009). The relationship between the expression of these genes and the development of colon cancer has not been widely examined. In this study, we found that increased expression of these genes may play an important role in the pathogenesis of colonic adenocarcinoma.

Compared with these 2 SAGE libraries of primary colon cancer, Tul02 and Tu98, the SAGE-generated high tag expression profiles of tumor in this study were significantly different, indicating that the SAGE library for colon cancer in a Chinese population has specific characteristics. However, additional studies are needed to confirm this. Some SAGE-generated tags in this study were not found for the corresponding genes, which may have been because of base mutation and mispairing, or may represent new genes. The significance of these tags requires further study.

In summary, we compared the gene expression profiles of colonic adenocarcinoma and normal colonic mucosa and found that colonic cancer showed specific characteristics. The unique gene expression profiles described in this study provide information and identified novel genes that may be involved in several cell signaling pathways important in colon tumors, the process of metaplasia, or can be used as tissue identification markers. Our results may improve the control and prevention of tumors, thus leading to better disease management.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Buckhaults P, Rago C, St Croix B, Romans KE, et al. (2001). Secreted and cell surface genes expressed in benign and malignant colorectal tumors. *Cancer Res*. 61: 6996-7001.

Chang Y, Lee TC, Li JC, Lai TL, et al. (2005). Differential expression of osteoblast-specific factor 2 and polymeric immunoglobulin receptor genes in nasopharyngeal carcinoma. *Head Neck* 27: 873-882.

Confalonieri S, Quarto M, Goisis G, Nuciforo P, et al. (2009). Alterations of ubiquitin ligases in human cancer and their association with the natural history of the tumor. *Oncogene* 28: 2959-2968.

Fritsche P, Seidler B, Schüler S, Schnieke A, et al. (2009). HDAC2 mediates therapeutic resistance of pancreatic cancer cells via the BH3-only protein NOXA. *Gut* 58: 1399-1409.

Graidist P, Phongdara A and Fujise K (2004). Antiapoptotic protein partners fortilin and MCL1 independently protect cells from 5-fluorouracil-induced cytotoxicity. *J. Biol. Chem*. 279: 40868-40875.

Huang ZG, Ran ZH, Lu W and Xiao SD (2006). Analysis of gene expression profile in colon cancer using the Cancer Genome Anatomy Project and RNA interference. *Chin. J. Dig. Dis*. 7: 97-102.

Ibrahim AF, Hedley PE, Cardle L, Kruger W, et al. (2005). A comparative analysis of transcript abundance using SAGE and Affymetrix arrays. *Funct. Integr. Genomics* 5: 163-74.

Minarowska A, Gacko M, Karwowska A and Minarowski Ł (2008). Human cathepsin D. *Folia. Histochem. Cytobiol*. 46: 23-38.

Missiaglia E, Jacobs B, D'Ario G, Di Narzo AF, et al. (2014). Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features. *Ann. Oncol*. 25: 1995-2001.

Moghaddam SJ, Haghighi EN, Samiee S, Shahid N, et al. (2007). Immunohistochemical analysis of p53, cyclinD1, RB1, c-fos and N-ras gene expression in hepatocellular carcinoma in Iran. *World J. Gastroenterol*. 13: 588-593.

Pabst T and Mueller BU (2009). Complexity of CEBPA dysregulation in human acute myeloid leukemia. *Clin. Cancer Res*. 15: 5303-5307.

van Erk MJ, Krul CA, Caldenhoven E, Stierum RH, et al. (2005). Expression profiling of colon cancer cell lines and colon biopsies: towards a screening system for potential cancer-preventive compounds. *Eur. J. Cancer Prev*. 14: 439-457.

Velculescu VE, Zhang L, Vogelstein B and Kinzler KW (1995). Serial analysis of gene expression. *Science* 270: 484-487.

Willaert W and Ceelen W (2015). Extent of surgery in cancer of the colon: Is more better? *World J. Gastroenterol*. 21: 132-138.

Yasui W, Oue N, Ito R, Kuraoka K, et al. (2004). Search for new biomarkers of gastric cancer through serial analysis of gene expression and its clinical implications. *Cancer Sci*. 95: 385-392.