Jurnal Biosains, 17(1), 19–37, 2006

A PRELIMINARY STUDY OF DIFFERENTIALLY EXPRESSED GENES IN MALAYSIAN COLORECTAL CARCINOMA CASES

[1]Sim E U H*, [2]Bong I P N, [2]Balraj P, [2]Tan S K, [3]Jamal R, [3]Sagap I, [3]Nadeson S, [3]Rose I M and [4]Lim P K M

[1]Human Molecular Genetics Laboratory, Resource Biotechnology Programme, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
[2]Institute for Medical Research (IMR), 50588 Kuala Lumpur, Malaysia
[3]Faculty of Medicine, Universiti Kebangsaan Malaysia, 56000 Kuala Lumpur, Malaysia
[4]Malaysian Bio-Diagnostics Research Sdn. Bhd. (MBDr), Block Intron-Ekson UKM-MTDC Smart Technology Centre, 43650 Bangi, Selangor Darul Ehsan, Malaysia

Abstrak: Setakat ini, mekanisme genetik yang lengkap tentang kemajuan adenoma kepada karsinoma dalam karsinoma usus besar masih belum diketahui. Demi memperoleh maklumat genetik tentang lintasan karsinogenesis ini, kami menggunakan kaedah pencarian gen yang diekspres secara berbeza dalam tumor karsinoma usus besar. Profil pengekspresan gen daripada kes karsinoma usus besar dikaji menggunakan sistem DNA Microarray. Kami melaporkan pengawalaturan tinggi dan rendah untuk 819 dan 98 gen masing-masing dalam tumor relatif kepada kawalan normal berkenaan. Corak pengekspresan berbeza dalam 121 gen adalah kekal dalam semua tumor yang dikaji. Tiga puluh tiga daripada gen-gen ini adalah gen protein ribosom (RP). Perbandingan data dengan pangkalan data pengekspresan gen domain awam (*Cancer Gene Expression Database, CGED*) menunjukkan 47 gen pengekspresan berbeza adalah konsisten. Dua puluh dua daripada gen-gen ini adalah gen RP. Antara semua gen RP yang dikenal pasti dalam kajian ini, corak pengekspresan untuk enam gen adalah selaras dengan sorotan kajian. Pengekspresan tinggi gen RP *L32* dilaporkan buat pertama kali dalam kajian ini dan telah disahkan melalui penganalisaan *RT-PCR*. Gen-gen yang tidak berkaitan dengan RP tetapi penting untuk dipertimbang adalah gen *tumour susceptibility* (*TSG101*) dan *20-kDa myosin light chain* (*MLC-2*). Walaupun penganalisaan *microarray* kami berdasarkan saiz sampel kecil (n = 2), kajian awal ini mengenal pasti banyak gen yang dikaitkan dengan konteks pembentukan dan kemajuan karsinoma usus besar buat pertama kali. Maka, penemuan kami membekalkan maklumat baru untuk kejadian genetik dalam proses karsinoma usus besar.

Abstract: Presently, the complete genetic mechanisms for the progression of adenoma to carcinoma for colorectal carcinoma (CRC) remain largely unclear. In order to obtain genetic information of this cancer pathway, we searched for differentially expressed genes in tumours of CRC. Gene expression profiles from CRC cases were assessed via the DNA microarray system. We report up-regulation and down-regulation of 819 and 98 genes respectively, in the tumours relative to their normal controls. The differential expression patterns of 121 genes were persistent in all tumours. Thirty three of these are ribosomal proteins (RPs) genes. Comparison of the 121 genes with a public domain gene expression database, the Cancer Gene Expression Database (CGEP), revealed 47 genes to be consistently differentially expressed in colorectal tumours. Among these, 22 are RP genes. Among all RP genes identified in this study the over-expression pattern for six of them is consistent with literature. The up-regulation of RP *L32* in CRC tumours was demonstrated for the first time in this study and it was also verified via reverse transcription-polymerase

*Corresponding author: uhsim@frst.unimas.my

chain reaction (RT-PCR) analysis. Non-RP genes worth noting are the tumour susceptibility gene (*TSG101*) and the 20-kDa myosin light chain (*MLC-2*). Albeit small sample size (n = 2 for microarray analysis), our preliminary studies revealed many genes that are brought into the context of CRC tumourigenesis for the first time, thus providing new clues to the genetic events during colorectal carcinogenesis.

Keywords: Colorectal Carcinoma, DNA Microarray, Ribosomal Proteins (RP), *TSG101*, *MLC-2*.

INTRODUCTION

Colorectal carcinoma (CRC) is one of the human cancer models where the progressive histopathological stages correlate with gradual but sequential perturbation of specific genes. These basic and sequential events that correlate with the adenoma-carcinoma progression have been reviewed by Fodde *et al.* (2001). Generally, in the case of a type of inheritable CRC – the familial adenomatous polyposis (FAP), initiation of tumour formation and clonal evolution of tumourigenic cells can be triggered by germline inactivating mutations in the adenomatous polyposis coli (*APC*) gene. In the context of Wnt-signaling pathway, the improper function of mutant *APC* leads to the stabilization of *β-catenin*, and hence the formation of the *β-catenin/TCF-LEF* (T-cell factor-lymphoid enchancer factor) complex. This results in the ectopic activation of oncogenes, one of which is *k-ras*. This in turn leads to a proportion of precancerous colorectal cells becoming adenomatous polyps. Further genetic disruption specifically in the form of Loss of Heterozygosity (LOH) at chromosomes 18q and 17p involving *SMAD2/SMAD4* and *TP53* genes respectively, causes the malignant transformation of benign adenomas to invasive carcinoma of the colorectum. This basic sequence of genetic events in colorectal cancer evolution may also be true for sporadic CRC, as *APC* mutations have been verified in early stages of sporadic CRC tumours (Powell *et al.* 1992). Contrary to the phenomenon of *APC*-initiated tumourigenesis, genetic defects (causing tumour progression) in the second familial forms of CRC, the Hereditary Nonpolyposis Colorectal Cancer (HNPCC), occur in the DNA mismatch repair (MMR) genes (Kinzler & Vogelstein 1996). In the HNPCC cases, MMR deficiencies concomitantly cause adenomas to acquire higher rate of mutation relative to normal colorectal cells. The resultant accumulation of mutations in oncogenes and tumor suppressor genes will lead to malignant transformation of adenomas (Kinzler & Vogelstein 1996).

The variety of genetic pathways that pertain to the adenoma-carcinoma sequence between FAP (and some sporadic CRC) and HNPCC reflects the complexities of the molecular mechanisms underlying tumor initiation to malignant progression of colorectal carcinoma. Presently, the understanding of these mechanisms remains basic if not partially elucidated. Of late, several groups have attempted to study the molecular events of colorectal carcinoma using approaches of gene expression analysis. These include the Serial Analysis of Gene Expression (SAGE) analysis (Zhang *et al.* 1997), the Affymetrix Human GeneChip™ (6500 and 6800) Set oligonucleotide arrays (Notterman *et al.* 2001), the 19,200-Element Complementary DNA microarray (Hedge *et al.* 2001), the

Suppressive Subtractive Hybridization (SSH) techniques (Hufton *et al.* 1999; Luo & Lai 2001) and a combination of SSH with cDNA library array technology (Swearingen *et al.* 2003). The findings of these studies produced a large repertoire of differentially expressed genes that are yet to be comprehensively and accurately incorporated into the existing molecular pathway.

In this study we report the assessment of gene expression profile of CRC cases via DNA microarray strategy. This has allowed us to identify and verify a number (121) of persistently differentially expressed genes. Our findings will inevitably provide important information for a comprehensive delineation of molecular pathway of CRC.

MATERIALS AND METHODS

Total RNA

Commercially available normal colon (cat. no. 64065-1) and colon tumour total RNAs (cat. no. 64014-1) were purchased from BD BioSciences (Clontech Laboratories, Inc.). For our study, these normal colon and colon tumour total RNA samples are designated as CN and CT respectively. The total RNAs from colorectal carcinoma and their paired normal tissue biopsies from local patients are designated as 019T and 056T, and 019N and 056N respectively. These were extracted using the Trizol method. Basically, frozen CRC and paired normal tissues were cut into smaller pieces with sterile surgical blade and then homogenized in 1 ml Trizol reagent (Invitrogen) using a polytron homogenizer. Following incubation at room temperature for 10 minutes, the homogenate was mixed with 200 μl chloroform, incubated at room temperature for 3 minutes and then centrifuged at 12,000 g for 15 minutes at 4°C. Total RNA present in the colourless aqueous layer (top layer) was precipitated using 0.5 ml isopropanol. The resulting total RNA pellet was washed with 75% ethanol (1 ml ethanol/ 1 ml Trizol ratio) and rinsed with 1 ml absolute ethanol. The final and purified dry total RNA pellet was then dissolved with 70 μl warm elution buffer. Estimation of concentration and purity of total RNA was performed via spectrophotometric assay at A_{260} and A_{280}. For this study, colorectal tumour samples are designated as 019T and 056T, and their paired normal tissues as 019N and 056N, respectively.

DNA Microarray

MICROMAX™ Human cDNA I Array slide (cat. no. MPS621) were from AlphaGene Inc. and PerkinElmer Life Science Inc. (USA). Each microarray slide contains 2382 elements comprising known human genes, control genes and housekeeping genes. All genes were spotted in duplicate.

Probe Preparation and Hybridization of Array

Labelling of probes with fluorescent dyes was performed using MICROMAX™ Direct Labeling Kit (AlphaGene Inc. & PerkinElmer Life Science Inc.; cat no. MPS502) and according to the manufacturer's protocol. Tumour samples were labeled with Cyanine 5 (Cy5) and normal samples were labeled with Cyanine 3

(Cy3) in this study. Purification of labeled cDNA probes was conducted using the isopropanol precipitation method. An equal concentration of Cy3 and Cy5-labeled cDNA probes were mixed, dried and dissolved in 20 μl hybridizaton buffer, prior to hybridization on the array slides for overnight at 65°C in a hybridization incubator. Washing of microarray slides was carried out in 0.5X SSC, 0.01% SDS; 0.06X SSC, 0.01% SDS; and 0.06X SSC – each for 15 minutes at room temperature with gentle agitation.

Microarray Data Analysis

Hybridization signals were detected using a flourescence scanner (Typhoon 8600 variable mode imager, Amersham Pharmacia Biotech) and documented using the ImageQuant software. The data was then processed and analyzed using GenePix Array Ver. 4.1 program (Axon Instruments Inc., Canada).

Reverse Transcriptase - PCR (RT-PCR)

One to two microlitres of total RNA was used as template for RT-PCR assay in a 20 μl reaction volume. For each assay, generation of first strand cDNAs was catalyzed by 200U Moloney murine leukemia virus (MMLV-RT), using oligo-dT_{18} primers. PCR amplification of the first strand cDNAs was carried out using the parameters of 95°C for 15 minutes (hot start); and then 21 cycles of 95°C × 30 seconds (denature), 59–64°C × 2 minutes (annealing) and 72°C × 2 minutes (extension); and a final extension step of 72°C for 10 minutes. Primer pairs of selected genes were designed to amplify regions within the coding region of each gene. Results of RT-PCR assays were assessed on agarose via gel electrophoretic analysis.

RESULTS

cDNA Microarray Analysis

Genes that are prominently differentially expressed in both sets are listed in Table 1 and 2. Comparative assessment of CT versus CN revealed 625 and 94 genes up-regulated and down-regulated in the tumour case, respectively. The highest level of up-regulation in CT was demonstrated by the osteoblast specific factor 2 (*OSF-2p1*) gene (82 folds), whereas the highest level of down-regulation in CT are found in the glutathione peroxidase (12 folds) and *TYL* genes (12 folds). In the case of the colorectal tumour sample (019T) and its paired normal control tissue (019N), 315 genes showed up-regulation in tumour whereas 5 genes were found to be down-regulated. The *MPV17* gene showed the highest level of up-regulation (98 folds), and the *MT-1I* gene showed the highest level of down-regulation (5 folds).

In total, of the 2382 genes (MICROMAX™ Human cDNA I Array system) analyzed, 917 genes appeared to have differential expression patterns between normal and tumour cases. Of these, mRNA levels for 819 genes are up-regulated and 98 genes are down-regulated in tumour cases. Relative expression difference ranges from 2 to 92 folds, with 12.3% (113/917) of the genes showing 10 folds difference or higher. Generally, the CT/CN set exhibited higher number of differentially expressed genes (719) compared to the 019T/019N set (320

genes). The ratio of up-regulated to down-regulated genes in the CT/CN set is approximately 6.67:1 (625:94), in contrast to 63:1 (315:5) for the 019T/019N set.

Table 1: Prominently differentially expressed genes/proteins from microarray analysis of CN versus CT via the MICROMAX™ Human cDNA I Array system.

Acc. No.	Transcripts up-regulated in CT	Fold	Acc. No.	Transcripts down-regulated in CT	Fold
D13665	*OSF-2p1*	82	D00632	Glutathione peroxidase (*Glu-Ox*)	12
L22587	Immunoglobulin heavy chain, V region (*IGH@*)	64	X99688	*TYL*	12
Y14737	Immunoglobulin lambda heavy chain (*Igλ-H)*	48	X68485	A1 adenosine receptor	10
X83703	Cytokine inducible nuclear protein	32	L10335	Neuroendocrine-specific protein C (*NSP*)	8
J04765	*Osteopontin*	32	D29808	T-cell acute lymphoblastic	7
M80927	Glycoprotein	32	J03483	Leukemia associated antigen 1 (*TALLA-1*)	7
M77844	*Oculorhombin*	20	L33404	Chromogranin A	7
X13694	*Osteopontin*	20	U05598	Stratum corneum chymotryptic enzyme	6
L42611	Keratin 6 isoform K6e (*KRT6E*)	16	Y11588	Dihydrodiol dehydrogenase	6
U62962	*Int-6*	16	X52426	Apoptosis specific protein	6

Table 2: Prominently differentially expressed genes from microarray analysis of 019N versus 019T via the MICROMAX™ Human cDNA I Array system.

Acc. No.	Transcripts up-regulated in 019T	Fold	Acc. No.	Transcripts down-regulated in 019T	Fold
X76538	*MPV17*	98	X76717	*MT-1l*	5
M86917	Oxysterol-binding protein (OSBP)	72	M95787	22-kDa smooth muscle protein (*SM22*)	4
Z68204	Succinyl CoA synthethase	56	J02854	20-kDa myosin light chain (*MLC-2*)	2
M58018	Beta-myosin heavy chain (*MYH7*)	39	X13839	Vascular smooth muscle alpha-actin	2
M11354	H3.3 histone, class B	36	L25798	3-hydroxy-3-methylglutaryl coenzyme A synthase	2
S82470	*BB1*, tumor progression-enhanced factor gene	36			
D88378	Proteasome inhibitor hPl31 subunit	33			
M36340	ADP-ribosylation factor 1 (*ARF1*)	32			
U54558	Translation initiation factor eIF3p66 subunit	32			
U75283	Sigma receptor	29			

Persistently Differentially Expressed Genes

Comparison of results of differentially expressed genes results between the two colorectal carcinoma sets revealed 121 genes that are persistently differentially expressed in both sets (Table 3). Persistent differential expression refers to expression behaviour of genes that showed consistent differential expression in both sets of sample analyzed. Except for *MLC-2*, all of these genes are up-regulated in tumours. Only five genes (RP L26, chondroitin sulfate proteoglycan, vasopressin activated calcium mobilizing receptor-like protein, *KIAA0428* and vimentin) showed consistency in the differential expression pattern, where equal relative fold-difference in the two sets (019N/019T and CN/CT) was observed. A majority of persistently differentially expressed genes (107 of 121) in the 019N/019T set showed higher level of fold-difference when compared to that of the CN/CT set. Comparing our results with the Cancer Gene Expression Database (CGED) (Kato *et al*. 2005), available at http://cged.hgc.jp, reveals 47 genes that are consistently differentially expressed (Table 3). These are genes that are differentially expressed in both tumour sets studied, and also listed in the Cancer Gene Expression Database (CGED) as differentially expressed in colorectal tumours. Of the 47 genes, 11 show down-regulation in tumours in contrast to our results of their up-regulation in tumours studied. As such, 29.8% (36/121) of our microarray results on persistently differentially expressed genes is validated against an independent set of data (CGED).

Table 3: Persistent differentially expressed genes obtained from comparison among differentially expressed genes of CN/CT, 019N/019T and CGEP (Kato *et al.* 2005) sets.

Acc. No.	Protein/Gene	Approximate fold difference		
		CT/CN	019T/019N	CGED
X13584	mRNA for gamma-aminobutyric-acidreceptor alpha-subunit (*GABA-A* receptor alpha subunit)	2	18	-
U46751	Phosphotyrosine independent ligand p62 for the *Lck SH2* domain	3	18	-
U83857	*Aac11* (*aac11*)	3	17	-
D42054	mRNA for KIAA0092 gene	3	15	-
X14420	mRNA for pro-alpha-1 type 3 collagen	4	15	-
U96915	sin3 associated polypeptide p18 (*SAP18*)	2	14	3.4
U10248	Ribosomal protein L29 (*humrpl29*)	3	14	1.4
M81757	S19 ribosomal protein	3	14	1.5
L11566	Ribosomal protein L18 (*RPL18*)	2	13	-
M14648	Cell adhesion protein (vitronectin) receptor alpha subunit	4	13	-
D45887	mRNA for calmodulin	2	12	-1.6
Z47087	mRNA for RNA polymerase II elongation factor-like protein	2	12	9

(continue on next page)

Table 3 *(continued)*

Acc. No.	Protein/Gene	Approximate fold difference		
		CT/CN	019T/019N	CGED
J02939	Membrane glycoprotein 4F2 antigen heavy chain mRNA	2	12	-
AF010187	*FGF-1* intracellular binding protein (*FIBP*)	3	12	-
L25899	Ribosomal protein L10	2	12	1.5
U14968	Ribosomal protein L27a	3	12	1.4
X76534	*NMB* mRNA	2	11	-
Y00387	mRNA for glutamine synthetase (E.C. 6.3.1.2)	2	11	-2.4
X06323	*MRL3* mRNA for ribosomal protein L3 homologue (*MRL3*, mammalian ribosome L3)	2	11	3
X56999	*UbA52* placental mRNA for ubiquitin-52 amino acid fusion protein	3	11	1
J03068	*DNF1552* (lung)	2	11	1.4
U51678	Small acidic protein	5	11	-
L38961	Putative transmembrane protein precursor (B5)	2	11	-
D37991	*SSR2* mRNA for beta-signal sequence receptor	2	11	-
S42658	S3 ribosomal protein	2	11	1.3
D00759	mRNA for proteasome subunit HC2	2	10	NI
U14971	Ribosomal protein S9	2	10	-
D50419	mRNA for *OTK18*	3	9	-
M15661	Ribosomal protein	5	9	-
D21239	mRNA for *C3G* protein	3	9	-
X15998	mRNA for the chondroitin sulphate proteoglycan versican, V1 splice-variant; precursor peptide	6	9	-
U32944	Cytoplasmic dynein light chain 1 (*hdlc1*)	2	9	2.6
J05500	Beta-spectrin (*SPTB*)	3	9	-
M55067	47-kD autosomal chronic granulomatous disease protein	4	9	-
J04543	Synexin	2	9	-
L06499	Ribosomal protein L37a (*RPL37A*)	3	9	1.3
J03040	*SPARC*/osteonectin	10	8	-5.1
X02152	mRNA for lactate dehydrogenase-A (*LDH-A, EC 1.1.1.27*)	3	8	-
S54761	Beta 2- mu, beta 2-microglobulin	2	8	-3.2
L06505	Ribosomal protein L12	2	8	9.0
M36072	Ribosomal protein L7a (*surf 3*) large subunit	4	8	-
L36645	Receptor protein-tyrosine kinase (*HEK8*)	3	8	-
M58458	Ribosomal protein S4 (*RPS4X*) isoform	3	8	1.9

(continue on next page)

Table 3 *(continued)*

Acc. No.	Protein/Gene	Approximate fold difference		
		CT/CN	019T/019N	CGED
M17733	Thymosin beta-4	3	8	-
U69645	Zinc finger protein	2	8	-
M17886	Acidic ribosomal phosphoprotein P1	2	8	1.5
D14530	Homolog of yeast ribosomal protein S28	5	8	1.5
D89729	mRNA for *CRM1* protein	6	7	-1.0
D87735	mRNA for ribosomal protein L14	3	7	2.6
Y00711	mRNA for lactate dehydrogenase B (*LDH-B*)	2	7	1.9
X16064	mRNA for translationally controlled tumor protein	3	7	-
L13740	*TR3* orphan receptor	2	7	-
X03342	mRNA for ribosomal protein L32	2	7	-
M77804	Tryptophanyl tRNA synthetase (*IFNWRS*)	2	7	-
X91257	mRNA for seryl-tRNA synthetase	2	7	-
D89289	mRNA for N-Acetyl-beta-D-glucosaminide	3	7	-
M21300	Small proline rich protein (*sprl*) mRNA, clone 15B	4	7	-
U09510	glycyl-tRNA synthetase	3	7	-
U87460	Putative endothelin receptor type B-like protein	2	7	-
L25085	*Sec61*-complex beta-subunit	2	6	1.8
M74002	Arginine-rich nuclear protein	2	7	-
X52966	mRNA for ribosomal protein L35a	4	7	-
Y10275	mRNA for L-3-phosphoserine phosphatase	3	7	-
X80909	Alpha *NAC* mRNA	2	7	-1.3
X63237	*Uba80* mRNA for ubiquitin	3	7	1.4
X53505	mRNA for ribosomal protein S12	2	6	-
D88674	mRNA for antizyme inhibitor	2	6	-
M20020	Ribosomal protein S6	8	6	-
U35622	*EWS-E1A-F* chimeric protein	2	6	-
AF026844	Ribosomal protein L41	2	6	1.7
AF083255	RNA helicase-related protein	2	6	-
M81635	Erythrocyte membrane protein	3	6	-
Z26876	Gene for ribosomal protein L38	4	6	-
AF013168	Hamartin (*TSC1*)	3	6	-
U26173	bZIP protein NF-IL3A (*IL3BP1*)	3	6	-
L06498	Ribosomal protein S20 (*RPS20*)	3	6	-5.1
U14973	Ribosomal protein S29	2	6	2.1
M61866	Krueppel-related DNA-binding protein (*PF4*) mRNA, 5' end	3	6	-
U10550	Gem GTPase (gem)	2	6	-
U51432	Nuclear protein Skip	4	6	-
L10413	Farnesyltransferase alpha-subunit	3	6	-
X69392	mRNA for ribosomal protein L26	6	6	-

(continue on next page)

Table 3 *(continued)*

Acc. No.	Protein/Gene	Approximate fold difference		
		CT/CN	019T/019N	CGED
X55954	mRNA for HL23 ribosomal protei homologue	4	6	1.4
X64707	*BBC1* mRNA	2	5	2.4
L28010	*HnRNP F* protein	2	5	-
X62691	mRNA for ribosomal protein (homologuous to yeast S24)	4	5	1.2
S75725	Prostacyclin-stimulating factor	4	5	-
D50372	mRNA for myosin regulatory light chain	4	5	1.6
D80009	mRNA for KIAA0187 gene	5	5	-
M13932	Ribosomal protein S17	2	5	2.2
M76979	Pigment epithelium-differentiation factor (*PEDF*)	2	5	-
M62831	Transcription factor *ETR101*	2	5	-
L19527	Ribosomal protein L27 (*RPL27*)	3	5	-
M14219	Chondroitin/dermatan sulfate proteoglycan (*PG40*) core protein	4	4	-
M34671	Lymphocytic antigen *CD59/MEM43*	3	4	3.6
L39060	Transcription factor SL1	2	4	-
X81882	mRNA for for vasopressin activated calcium mobilizing receptor-like protein	4	4	-
U12404	*Csa-19*	2	3	1.5
M99701	*pp21*	3	4	-
X12451	mRNA for pro-cathepsin L (major excreted protein *MEP*)	3	4	-
D13665	mRNA for osteoblast specific factor 2 (*OSF-2p1*)	82	4	-
D90402	mRNA for endothelin receptor (*ETR*)	3	4	-
AB007888	KIAA0428	4	4	4.6
M64716	Ribosomal protein S25	3	4	-
J02947	Extracellular-superoxide dismutase (*SOD3*)	3	4	2.3
U14967	Ribosomal protein L21	5	4	-1.0
U47742	Monocytic leukaemia zinc finger protein (*MOZ*)	3	4	-
U37230	Ribosomal protein L23a	3	4	-
M94314	Ribosomal protein L30	3	4	-1.3
U82130	Tumor susceptiblity protein (*TSG101*)	2	4	9
X89401	mRNA for large subunit of ribosomal protein L21	6	4	-1.0
X52022	RNA for type VI collagen alpha3 chain	4	3	-1.9
AF010313	Pig8 (*PIG8*)	2	3	-
L03555	Ig rearranged kappa-chain mRNA variable region, joining region, constant region	2	3	-
M28212	GTP-binding protein (RAB6	4	3	-
U09953	Ribosomal protein L9	2	3	9.0

(continue on next page)

Table 3 *(continued)*

Acc. No.	Protein/Gene	Approximate fold difference		
		CT/CN	019T/019N	CGED
X16478	mRNA 5'-fragment for vimentin N-terminal fragment	3	3	-
U27655	*RGP3*	2	3	-
L12535	*RSU-1/RSP-1*	2	3	-
U76992	*Tat-SF1*	3	19	-
U75283	Sigma receptor (*hSigmaR1*)	2	29	-
J02854	20-kDa myosin light chain (*MLC-2*)	-3	-2	NI

Note: Fold difference of differential expression from CGED is expressed as the ratio of frequency of ESTs in cancer libraries to frequency of all expressed seqnencetag (ESTs), or vice versa. Positive and negative values indicate up-regulation and down-regulation in tumours, respectively. The abbreviation, NI, means no information for frequency of EST is available although expression in cancer is reported.

RT-PCR Verification

For the purpose of verifying our microarray findings, RT-PCR assays were performed for a selected number of differentially expressed genes reported. Genes targeted for analysis were either randomly selected (for 019T/019N set and persistently expressed genes) or based on prominence in their fold-difference of differential expression (for the CT/CN set).

In the CT/CN set – *OSF-2p1*, RT L7a (*surf-3*), immunoglobulin lambda heavy chain (*Igλ-H*) and immunoglobulin heavy chain V region (*IGH@*) genes were demonstrated to be up-regulated in the tumour, while the glutathione peroxidase (*Glu-Ox*), *TYL* and the 20-kDa myosin light chain (*MLC-2*) genes were shown to be down-regulated in the tumour [Fig. 1 (A)]. Analysis of the 019N/019T set revealed up-regulation of *BB1*, *MPV17* and RP *L32* genes in the tumour [(Fig. 1 (B)]. The up-regulation of *TSG101* gene in tumour cases was verified from the analysis of a second local CRC case, the 056N/056T set [(Fig. 1 (C)]. Analysis of the 056N/056T set [Fig. 1 (C)] also revealed reproducibility of results for the RP *L7a (surf-3)* gene [comparison with CN/CT set; [Fig. 1(A)], and the RP *L32* gene [comparison with 019N/019T set; [Fig. 1 (B)]. The *GAPDH* control [Fig.1 (A & C)] affirmed the equimolar concentration of starting total RNA used in the study. On the whole, the RT-PCR assays confirmed the authenticity of the differentially expressed genes procured from our microarray analysis.

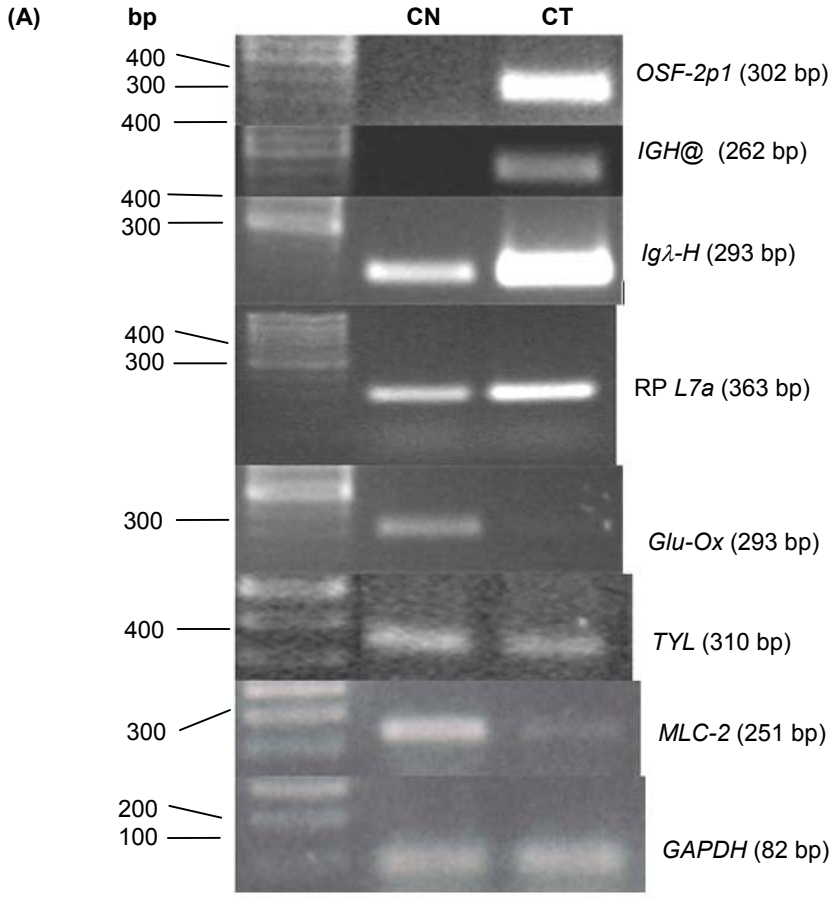

Figure 1: RT-PCR verification of a subset of differentially expressed genes identified in the microarray analysis. (A) are results from analysis of CN/CT (commercially-available) system; and (B) and (C) are results from analysis of local CRC tumour and its paired normal samples – 019N/019T and 056N/056T, respectively. Fragment sizes indicated in parentheses are the expected RT-PCR product sizes. The DNA size reference used was the GeneRuler™ 100 bp DNA ladder (MBI Fermentas). *(continue on next page)*

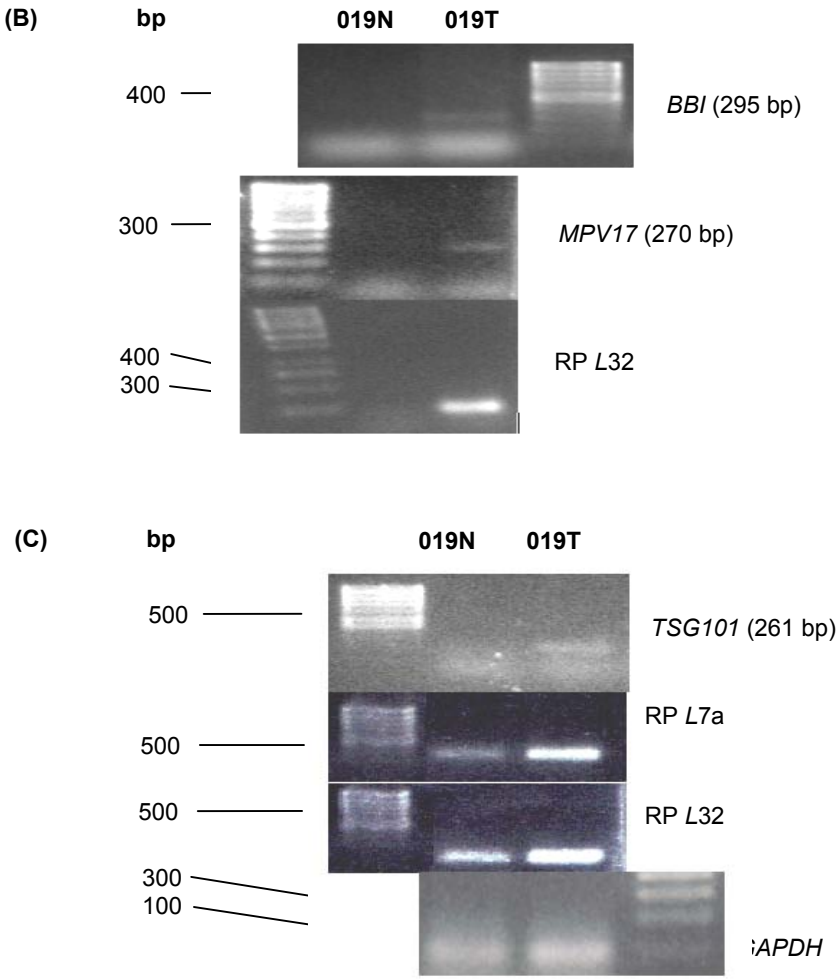

Figure 1: *(continued)*

DISCUSSION

The results of this study, like several of others using similar approaches, yielded a large repertoire of differentially expressed markers in tumours of CRC. The comparison of their *in vivo* expression profiles, and hence identification of common differentially expressed genes has allowed greater accuracy in targeting genes associated with the genetic events of CRC development. In the CN/CT system, the tumour sample was not compared to a paired normal. Although this may suggests incompatible comparison to a certain extent, such approach is still considered valid, as similar strategy was employed in the published work by

Swearingen *et al.* (2003). In their case, primary tumours from different individuals were compared to the normal colon tissue from another set of different individuals that were neither afflicted with CRC nor related to those where the tumour specimens were procured. Furthermore, the occurrence of many persistently differentially expressed markers in both sets studied (CN/CT and 019N/019T) provides validation to our approach. To this end, we have identified 121 persistently differentially expressed genes – 47 of which were also listed as differentially expressed in colorectal tumours according to the public domain gene expression database, CGED (Kato *et al*. 2005; http://cged.hgc.jp).

Of the 47 consistent differentially expressed genes, 22 encode RP or ribosomal protein-related genes. Four of these have been reported by others to be over-expressed in CRC cases. These are the RP genes of *S3* (Pogue-Geile *et al.* 1991), *S19* (Kondoh *et al.* 1992), *L7a* (Wang *et al*. 2000) and human homologue of yeast RP S28 (Otsuka *et al.* 2001). Besides these, two other RP genes not listed in CGED but have been reported by others to be differentially expressed are the *S6* and *S12* RP genes. The *S3*, *S6* and *S12* RP genes have been demonstrated to be over-expressed in adenocarcinoma of the colon relative to normal colonic mucosa, and also more abundant in adenomatous polyps (Pogue-Geile *et al.* 1991). In fact, we showed up-regulation of *L26* and *L35* RP genes in our tumour sets, despite failure by Pogue-Geile's group (1991) to detect their transcripts in either normal or malignant colon using Northern analysis. The *S19* RP genes has been shown to be up-regulated in colon carcinoma tissue and has increased level of expression that correlates with tumour progression in colon cancer cell lines (Kondoh *et al.* 1992). In the case of *L7a* (*surf 3*), its up-regulation in colorectal cancer was demonstrated by Wang *et al.* (2000). The human homologue of yeast RP S28 was shown by Otsuka's group (2001) to be up-regulated in metastatic-tumour-derived cells of colorectal carcinoma compared with primary-tumour-derived cells.

The fact that a majority (22 of 47) of consistent differentially expressed genes represent RP genes suggests strong correlation between high levels of RP mRNAs with neoplasia of the colorectum. It would seem logical that the increased level of RP mRNAs can be simply due to the presence of higher percentage of proliferating cells in neoplasia situation. However, Pogue-Geile *et al.* (1991) explained that the proliferation rate of colorectal carcinomas cells is not significantly higher than that of normal colonic mucosa – hence a possibility that higher mRNAs levels may be due to a decreased mRNA degradation rather than an increased transcription. This remains to be proven. It is also unclear whether the increased level of RP transcripts can be causative of carcinogenesis. Indeed this seems plausible through several findings that include the suggestion of S6 RP's tumour suppressive function in *Drosophila* hematopoietic system (Watson *et al.* 1992), formation of protein complex between human L5 RP with Mdm2 and p53 (Marechal *et al.* 1994), mutation of human S19 RP genes in sporadic and familial cases of Diamond-Blackfan anemia – a syndrome with increased risk of developing leukemia (Draptchinskaia *et al.* 1999) and heterozygous mutations in 11 different RP genes that predispose zebrafish to cancer (Amsterdam *et al.* 2004). Although the general indication from these findings favours the notion that a reduced level of RP genes would lead to carcinogenesis, studies by others and

ours on human CRC cases suggested otherwise. Thus, the role(s) of RPs in cancer and the mechanism(s) by which increased RP gene transcripts leads to carcinogenesis in CRC remains to be studied.

The tumour susceptibility gene, *TSG101* was consistently up-regulated in all tumour samples studied (CT, 019T and 056T) and also listed in the CGED as up-regulated in colorectal tumours. The *TSG101* gene is located at chromosome 11p15.1–15.2 – a region known to contain tumour suppressor genes and was initially shown to have large intragenic deletions in human breast cancers (Li *et al.* 1997). Abnormal/truncated or aberrantly spliced *TSG101* transcripts have been commonly reported in the human breast cancers – amongst other forms of cancer (Lee & Feinberg 1997; Carney *et al.* 1998; Turpin *et al.* 1999; Balz *et al.* 2002). Interestingly, for cases of CRC, the presence of aberrant transcripts were considered to be consequence of PCR artifacts as they were also found in the normal controls to the colon tumours studied (Hampl *et al.* 1998; Lin *et al.* 1998). In its suspected involvement with oncoprotein network, specifically the p53-*MDM2* circuitry, Li *et al.* (2001) explained that interaction between *TSG101* and *MDM2* elevate the level of *MDM2*, and consequently promote decay of p53. Coupled with the findings that the over-expression of *MDM2* promote loss of *TSG101* (Li *et al.* 2001), it appears that *TSG101* may play the role of regulator and target of the p53-*MDM2* pathway during the control of cellular apoptosis and proliferation. The correlation between *TSG101* and p53 decay was further strengthened through the findings of homozygous *TSG101* -/- mouse embryos that showed significant accumulation of p53 protein (Ruland *et al.* 2001). However, unlike a typical tumour suppressor, studies have shown that under-expression of *TSG101* did not cause cellular over-proliferation. Despite the suggestive role(s) of the *TSG101-MDM2-P53* complex in mediating carcinogenesis, there is still no data or findings that support predispositional inactivating intragenic *TSG101* mutations in CRC or any other forms of cancers. One thing remains consistent is the up-regulation of *TSG101* in cancers. Our results from all CRC tumours samples studied (CT, 019T and 056T), the records from CGED (Kato *et al.* 2005) and findings by Koon *et al.* (2004) in malignant gastrointestinal stromal tumours are evidence to this consistency. Perhaps this suggests that *TSG101* has greater roles in tumour progression of CRC rather than neoplastic predisposition or tumour suppression. Indeed the deficiency of *TSG101* expression has been directly correlated to cellular growth arrest, specifically at the growth 1/synthesis (G1/S) phase of the cell cycle checkpoint, leading to prohibition of cellular growth and proliferation. Evidence to this were revealed in studies that include the quenching of *TSG101* protein in cells via the introduction of *TSG101*-specific antibodies (Zhong *et al.* 1998), observation of primary mouse embryonic fibroblast derived from *TSG101* conditional knockout mice (Krempler *et al.* 2002) and RNA interference experiment on prostate cancer *(PC3)* and breast cancer *(MDA-MB-231)* cell lines (Zhu *et al.* 2004). The actual influence of *TSG101* expression on the control of proliferation of colorectal or colorectal-derived cells will require further investigation.

In this study, persistent down-regulation in CRC was demonstrated for the 20-kDa regulatory myosin light chain gene, *MLC-2*. Although expression of human *MLC-2* has been reported in normal human colon tissue (Kumar *et al.*

1989), its down-regulated expression in colorectal tumours is reported for the first time in the study. Products of the *MLC-2* gene is important for regulation of smooth muscle and non-muscle cell contractile activity via phosphorylated *MLC-2*-mediated increase of actin-activated myosin ATPase activity (Kumar *et al.* 1989). Despite using a small number of samples, we suspect that the expression behaviour observed attributes to the malfunctioning of colorectal smooth muscle tissues. In fact, when neonatal human prostate epithelial cells were subjected to multiple X-ray exposures, the derived cell lines showed reduced cell size with poorly organized actin stress fibres and demonstrated progressive loss of *MLC-2* expression (Prasad *et al.* 1997). Similar expression behaviour of *MLC-2* was observed in human osteosarcoma derived cell clonal cells that have undergone transformation following infection by Kirsten murine sarcoma virus or by chemical carcinogen (Kumar & Chang 1992). Interestingly, the oncoprotein-mediated signally pathway affecting *MLC-2* repression was proven by an *in vitro* study – where the repression of chicken *MLC-2* promoter by the proto-oncogene, *fos*, was via a *fos*-responsive element (FRE) at –1130 to – 1200 bp upstream of the *MLC*-2 transcription initiation site (Goswami *et al.* 1992). In addition, the complete repression of *MLC-2* expression was reported in human osteosarcoma derived clonal cells transformed by *Ha-ras* oncogene (Kumar & Chang 1992). Taken together, these reports and our results suggest that oncogene-mediated repression of *MLC-2* in CRC probably caused both neoplastic development and malfunction of affected colorectal smooth muscle cells.

Besides differentially expressed genes common for both sets analyzed, a few of those that are differentially expressed in either set have been shown in literature to be associated with CRC, or other cancers. Amongst these, are the oncogenes of *K-ras, dek* and *set*. As reviewed in Fodde *et al.* (2001), expression of *K-ras* and other oncogenes is crucial for transformation of aberrant crypt foci to adenoma. In studies by Hedge *et al.* (2001), the *set* and *dek* oncogenes have been demonstrated to be over-expressed in liver metastatic cell line derived from CRC. Similarly, we found these three genes to be up-regulated in one of the cancer sets studied (CN/CT). The level of up-regulation of the three oncogenes (*K-ras*, *dek* and *set*) between studies of Hedge's group and ours is comparable (Table 4). This consistency among findings suggests that the *set* and *dek* oncogenes work in association with *K-ras* in tumorigenic (adenomatous) transformation of colorectal mucosal cells, and that the maintenance of their high expression beyond adenomatous tissue (metastatic and advanced stages) suggests that their roles may persist throughout early intermediate stages of carcinogenesis into invasive carcinoma malignancy.

Table 4: Consistent differentially expressed genes from this study and in comparison to findings from literature.

Genes	Fold of differential expression in tumour cases		
	Hedge *et al.* (2001)	Notterman *et al.* (2001)	This study (CT/CN)
K-ras	2.4 – 3.7	–	2
dek	2.5	–	3
set	2 – 4	–	3
CAS	–	4.7	6

Finally, despite the large database of differentially expressed genes from our findings that could be implicated as CRC-associated genes, a majority of these genes have never been brought into the context of CRC tumourigenesis. Therefore, much has yet to be done to fully understand their physiological role(s) in colorectal organogenesis and carcinogenesis. More importantly, the consistent expression behaviours of these genes in most forms of CRC have to be established, as well as their place in the existing gene-gene network and pathway in CRC development. Two areas of investigation will be crucial for future work. Firstly, since our findings (DNA microarray data) are based on small sample size (n = 2), there is a need to establish reproducibility of the differential expression signatures of these genes via expression profile analysis of more CRC cases. Secondly, the increasing number of newly identified differentially expressed genes in CRC must be fully utilize for more comprehensive functional and gene-gene interaction studies. Ultimately these efforts will provide accurate information on molecular events occurring during onset of tumourigenesis to progression of malignancy in cancer of the colorectum.

ACKNOWLEDGEMENTS

The facilities for scanning of microarray slides and software for microarray data analysis (GenePix Array Ver 4.1) were provided by the Institute of Health and Community Medicine (IHCM), Universiti Malaysia Sarawak. This work was supported by a grant from the Malaysian National Biotechnology Directorate – Medical Biotechnology Cooperative Centre research initiatives (Project No.: 06-05-01-003 BTK/ER/018).

REFERENCES

Amsterdam A, Sadler K C, Lai K, Farrington S, Bronson R T, Lees J A and Hopkins N. (2004). Many ribosomal protein genes are cancer genes in zebrafish. *PLoS Biology* 2(5): 690–698.

Balz V, Prisack H B, Bier H and Bojar H. (2002). Analysis of *BRCA1*, *TP53* and *TSG101* germline mutations in German breast and/or ovarian cancer families. *Cancer Genet and Cytogenet* 138(2): 120–127.

Carney M E, Maxwell G L, Lancaster J M, Gumbs C, Marks J, Berchuck A and Futreal P A. (1998). Aberrant splicing of the *TSG101* tumor suppressor gene in human breast and ovarian cancers. *J. Soc. Gynecol. Investig.* 5(5): 28–85.

Draptchinskaia N, Gustavsson P, Anderson B, Pettersson M, Willig T N, Dianzani I, Ball S, Tchernia G, Klar J, Matsson H, Tentler D, Mohandas N, Carlsson B and Dahl N. (1999). The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anemia. *Nat. Genet.* 21: 169–175.

Fearnhead N S, Britton M P and Bodmer W F. (2001). The ABC of APC. *Human Molecular Genetics* 10(7): 721–733.

Fodde R, Smits R and Clevers H. (2001). *APC*, signal transduction and genetic instability in colorectal cancer. *Nature Reviews* 1: 55–67.

Goswami S K, Zarraga A M, Martin M E, Morgenstern D and Siddiqui M A. (1992). *fos*-mediated repression of cardiac myosin light chain-2 gene transcription. *Cell. Mol. Biol.* 38(1): 49–58.

Hampl M, Hampl J, Plaschke J, Fitze G, Schackert G, Saeger H D and Schackert H K. (1998). Evidence that *TSG101* aberrant transcripts are PCR artifacts. *Biochem. Biophys. Res. Commun.* 248(3): 753–760.

Hedge P, Rong Q, Gaspard R, Abernathy K, Dharap S, Earle-Hughes J, Gay C, Nwokekeh N U, Chen T, Saeed A, Sharov V, Lee N H, Yeatman T J and Quackenbush J. (2001). Identification of tumour markers in sets of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer Res.* 61: 7792–7797.

Hufton S E, Moerkerk P T, Brandwijk R, de Bruine A P, Arends J W and Hoogenboom H R. (1999). A profile of differentially expressed genes in primary colorectal cancer using suppression subtractive hybridization. *FEBS Lett.* 463(1–2): 77–82.

Kato K, Yamashita R, Matoba R, Monden M, Shinzaburo N, Takagi T and Nakai K. (2005). Cancer gene expression database (CGED): A database for gene expression profiling with accompanying clinical information of human cancer tissues. *Nucleic Acids Research* 33: Database issue D533–D536.

Katoh M. (2001). Frequent up-regulation of *WNT2* in primary gastric and colorectal cancer. *International Journal of Oncology* 19: 1003–1007.

Kinzler K W and Vogelstein B. (1996). Lessons from hereditary colorectal cancer. *Cell* 87: 159–170.

Kondoh N, Schweinfest C W, Henderson K W and Papas T S. (1992). Differential expression of S19 ribosomal protein laminin-binding protein and human lymphocyte antigen class-I messenger RNAs associated with colon carcinoma progression and differentiation. *Cancer Res.* 52(4): 791–796.

Koon N, Schneider-Stock R, Sarlomo-Rikala M, Lasota J, Smolkin M, Petroni G, Zaika A, Boltze C, Meyer F, Andersson L, Knuutila S, Miettinen M and El-Rifai W. (2004). Molecular targets for tumour progression in gastrointestinal stromal tumours. *Gut* 53(2): 235–240.

Krempler A, Henry M D, Triplett A A and Wagner K U. (2002). Targeted deletion of the *TSG101* gene results in cell cycle arrest at G1/S and p53-independent cell death. *J. Biol. Chem.* 277(45): 43216–43223.

Kumar C C, Mohan S R, Zavodny P J, Narula S K and Leibowitz P J. (1989). Characterization and differential expression of human vascular smooth muscle myosin light chain-2 isoform in non-muscle cells. *Biochemistry* 28(9): 4027–4035.

Kumar C C and Chang C. (1992). Human smooth muscle myosin light chain-2 gene expression is repressed in *ras* transformed fibroblast cells. *Cell Growth Differ.* 3(1): 1–10.

Lee M P and Feinberg A P. (1997). Aberrant splicing but not mutations of *TSG101* in human breast cancer. *Cancer Res.* 57(15): 3131–3134.

Li L, Li X, Francke U and Cohen S N. (1997). The *TSG101* tumor susceptibility gene is located in chromosome 11 band p15 and is mutated in human breast cancer. *Cell* 88(1): 143–154.

Li L, Liao J, Ruland J, Mak T W and Cohen S N. (2001). A *TSG101/MDM2* regulatory loop modulates *MDM2* degradation and *MDM2*/p53 feedback control. *Proc. Natl. Acad. Sci. USA* 98(4): 1619–1624.

Lin P M, Liu TC, Chang J G, Chen T P and Lin S F. (1998). Aberrant *TSG101* transcripts in acute myeloid leukaemia. *Br. J. Haematol*. 102(3): 753–758.

Luo M J and Lai M D. (2001). Identification of differentially expressed genes in normal mucosa, adenoma and adenocarcinoma of colon by SSH. *World J. Gastroenterol.* 7(5): 726–731.

Marechal V, Elenbaas B, Piette J, Nicolas J C and Levine A J. (1994). The ribosomal L5 protein is associated with mdm-2 and mdm-2-p53 complexes. *Mol. Cell Biol.* 14: 7414–7420.

Notterman D A, Alon U, Sierk A J and Levine A J. (2001). Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma and normal tissue examined by oligonucleotide arrays. *Cancer Res.* 61: 3124–3130.

Otsuka M, Kato M, Yoshikawa T, Chen H, Brown E J, Masuho Y, Omata M and Seki N. (2001). Differential expression of the L-plastin gene in human colorectal cancer progression and metastasis. *Biochem. Biophys. Res. Commun.* 289(4): 876–881.

Pogue-Geile K, Geiser J R, Shu M, Miller C, Wool I G, Meisler A I and Pipas J M. (1991). Ribosomal protein genes are overexpressed in colorectal cancer: Isolation of a cDNA clone encoding the human S3 ribosomal protein. *Mol. Cell Biol.* 11(8): 3842–3849.

Powell S M, Zilz N, Beazer-Barclay Y, Bryan T M, Hamilton S R, Thibodeau S N, Vogelstein B and Kinzler K W. (1992). APC mutations occur early during colorectal tumorigenesis. *Nature* 339: 235–237.

Prasad S C, Thraves P J, Dritschilo A and Kuettel M R. (1997). Protein expression changes associated with radiation-induced neoplastic progression of human prostate epithelial cells. *Electrophoresis* 18(3–4): 629–637.

Ruland J, Sirard C, Elia A, MacPherson D, Wakeham A, Li L, de la Pompa J L, Cohen S N and Mak T W. (2001). *p53* accumulation, defective cell proliferation and early embryonic lethality in mice lacking *TSG101*. *Proc. Natl. Acad. Sci. USA* 98(4): 1859–1864.

Sparks A B, Morin P J, Vogelstein B and Kinzler K W. (1998). Mutational analysis of the *APC*/beta-catenin/Tcf pathway in colorectal cancer. *Cancer Res.* 58: 1130–1134.

Swearingen M L, Sun D, Bourner M and Weinstein E J. (2003). Detection of differentially expressed HES-6 gene in metastatic colon carcinoma by combination of suppression subtractive hybridization and cDNA library array. *Cancer Letters* 198: 229–239.

Turpin E, Dalle B, de Roquancourt A, Plassa L F, Marty M, Janin A, Beuzard Y and de The H. (1999). Stress-induced aberrant splicing of *TSG101*: Association to high tumor grade and p53 status in breast cancers. *Oncogene* 18(54): 7834–7837.

Wang Y, Cheong D, Chan S and Hooi S C. (2000). Ribosomal protein L7 a gene is up-regulated but not fused to the tyrosine kinase receptor as chimeric trk oncogene in human colorectal carcinoma. *Int. J. Oncol.* 16(4): 757–762.

Watson K L, Konrad K D, Woods D F and Bryant P J. (1992). *Drosophila* homolog of the human S6 ribosomal protein is required for tumor suppression in the hematopoietic system. *Proc. Natl. Acad. Sci. USA* 89: 11302–11306.

Zhang L, Zhou W, Velculescu V E, Kern S E, Hruban R H, Hamilton S R, Vogelstein B and Kinzler K W. (1997). Gene expression profiles in normal and cancer cells. *Science* 276: 1268–1272.

Zhong Q, Chen Y, Jones D and Lee W H. (1998). Perturbation of *TSG101* protein affects cell cycle progression. *Cancer Res.* 58(13): 2699–2702.

Zhu G, Gilchrist R, Borley N, Chng H W, Morgan M, Marshall J F, Camplejohn R S, Muir G H and Hart I R. (2004). Reduction of *TSG101* protein has a negative impact on tumor cell growth. *Int. J. Cancer* 109(4): 541–547.