



**Distinctive Gene Expression Signatures For Early And  
Advanced Stage Sporadic Colorectal Adenocarcinomas In A  
Malaysian Patient Cohort**

Journal:	<i>BMJ Open</i>
Manuscript ID:	bmjopen-2014-004930
Article Type:	Research
Date Submitted by the Author:	14-Feb-2014
Complete List of Authors:	Lau, Tze Pheng; University of Malaya, Biomedical Science Roslani, April Camilla; University of Malaya, Surgery Lian, Lay Hoong Lee, Ping Chin; Universiti Sabah Malaysia, School of Science and Technology Hilmi, Ida; University of Malaya, Medicine Goh, Khean Lee; University of Malaya, Medicine Chua, Kek Heng; University of Malaya, Biomedical Science
<b>Primary Subject Heading</b>:	Gastroenterology and hepatology
Secondary Subject Heading:	Gastroenterology and hepatology
Keywords:	Gastrointestinal tumours < GASTROENTEROLOGY, Cancer genetics < GENETICS, PUBLIC HEALTH

SCHOLARONE™  
Manuscripts

Only

1  
2  
3  
4 **Distinctive Gene Expression Signatures For Early And Advanced Stage Sporadic**  
5 **Colorectal Adenocarcinomas In A Malaysian Patient Cohort**  
6  
7  
8

9  
10  
11 Tze Pheng Lau<sup>1</sup>, April Camilla Roslani<sup>2</sup>, Lay Hoong Lian<sup>3</sup>, Ping Chin Lee<sup>4</sup>, Ida Hilmi<sup>5</sup>,  
12  
13  
14 Khean Lee Goh<sup>5</sup>, Kek Heng Chua<sup>1\*</sup>  
15  
16  
17  
18

19 <sup>1</sup>*Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala*  
20 *Lumpur, Malaysia.*  
21

22 <sup>2</sup>*Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur,*  
23 *Malaysia.*  
24

25 <sup>3</sup>*Casa Vista, Jalan Penaga, 59100 Kuala Lumpur, Malaysia.*  
26

27 <sup>4</sup>*School of Science and Technology, Universiti Sabah Malaysia, 88400 Kota Kinabalu, Sabah,*  
28 *Malaysia.*  
29

30 <sup>5</sup>*Division of Gastroenterology and Hepatology, Department of Medicine, Faculty of Medicine,*  
31 *University of Malaya, 50603 Kuala Lumpur, Malaysia.*  
32  
33  
34

35  
36  
37  
38 *\*Correspondence author: Prof. Chua Kek Heng, Department of Biomedical Science, Faculty*  
39 *of Medicine, University Malaya, 50603 Kuala Lumpur, Malaysia.*  
40

41  
42 *e-mail: khchua@um.edu.my*  
43

44  
45 *Phone no: 603-79676607*  
46

47  
48 *Fax no: 603-79676600*  
49

50 **Key words:** Gene expression, non-hereditary, colorectal cancer, Malaysian  
51

52  
53 **Word count:** 3056 words  
54  
55  
56  
57  
58  
59  
60

## ABSTRACT

**Objectives:** To characterise the mRNA expression patterns of early- and advanced stage colorectal adenocarcinomas of Malaysian patients.

**Design:** Comparative expression analysis.

**Setting and participants:** We performed a combination of ACP-based PCR and RT-qPCR for the identification of differentially expressed genes (DEGs) associated with early- and advanced stage primary CRC tumours. We recruited four paired samples from CRC patients of Dukes' A and B for the preliminary differential expression study, and a total of 27 paired samples, ranging from CRC Stages I – IV, for subsequent confirmatory test. The tumouric samples were obtained from CRC patients undergoing curative surgical resection without pre-operative chemoradiotherapy. The recruited CRC patients were newly-diagnosed with CRC, and were not associated with any hereditary syndromes, previously diagnosed cancer, or positive family history of CRC. The paired non-cancerous tissue specimens were excised from macroscopically normal colonic mucosa distally located from the CRC tumours.

**Primary and secondary outcome measures:** The differential mRNA expression patterns of early- and advanced stage colorectal adenocarcinomas compared to macroscopically normal colonic mucosa were characterised by ACP-based PCR and RT-qPCR.

**Results:** The *RPL35*, *RPS23* and *TIMP1* genes were found to be over-expressed in both early- and advanced stage colorectal adenocarcinomas ( $p < 0.05$ ). On the other hand, the *ARPC2* gene was significantly under-expressed in early colorectal adenocarcinomas, while the advanced stage primary CRC tumours exhibited an additional over-expression of the *C6orf173* gene ( $p < 0.05$ ).

**Conclusions:** We characterised two distinctive gene expression patterns to aid in the stratification of primary CRC neoplasms among Malaysian CRC patients. Further work can be done to assess and compare the mRNA expression levels of these identified DEGs between each CRC stage group, Stages I – IV.

## ARTICLE SUMMARY

### Article focus

- The latest staging system of CRC tumours, which relies mainly on the clinicopathological assessment of primary tumours, is insufficient to address the complexity and heterogeneity of this disease. Hence, a non-anatomical, molecular-oriented staging system which can effectively predict the patients' outcome(s) and direct targeted treatment to different subgroups of CRC patients, is needed.
- Despite continuous revisions in the CRC classification criteria and expansion in transcriptomic and proteomic studies, there is yet to be any molecular marker(s) incorporated for clinical purposes.
- Article focus of this study is to characterise mRNA expression patterns of early- and advanced stage CRC tumours of Malaysian patients.

### Key messages

- The under-expression of *ARPC2* and over-expression of *C6orf173* gene were distinctive for the early- and advanced stage sporadic colorectal adenocarcinomas, respectively.

### Strengths and limitations of this study

This regional-based study has a relatively small sample size due to the strict sample recruitment criteria where all subjects were newly-diagnosed with CRC, and were not associated with any hereditary syndromes, previously diagnosed cancer, or positive family history of CRC. However, the findings of this study are still reliable in view of our stringent sample selection criteria, high specificity primers and probes, as well as reliable statistical analysis.

**Abbreviations:** CRC: colorectal cancer; TNM: tumour-node-metastasis; AJCC: American Joint Committee on Cancer; SAGE: serial analysis of gene expression; CEA: carcinoembryonic antigen; ACP: Annealing Control Primer; RT-qPCR:

reverse transcription-quantitative real-time PCR; RIN: RNA integrity number; DEG: differentially expressed gene;  $\Delta\Delta C_T$ : comparative  $C_T$ ; MAD: median absolute deviation

## INTRODUCTION

Cancer staging is vital for patient management, especially in prognosis prediction and planning of treatment intervention[1]. This is especially in the CRC staging system. As such, there have been many noteworthy improvements since the introduction of the classical Dukes' staging system, followed by the modified Astler-Coller staging system; to the latest 7<sup>th</sup> edition of TNM staging system published by the AJCC[2-4]. The TNM staging system allows the incorporation of various clinical information (which are obtained through histopathological examination, radiologic imaging and surgical findings), for accurate CRC stratification[5]. However, these clinical assessments are greatly dependent on the expertise of pathologists, radiologists and clinicians.

The TNM classification is applicable for both clinical (cTNM) and pathological (pTNM) staging of primary CRC tumours. Typically, it involves the assessment on the depth of bowel wall invasion at the time of diagnosis and the presence of regional lymph nodes metastases, as well as the presence of distant organ metastasis[4]. As a potentially worse patient outcome with more advanced disease stage is the core concept in cancer staging, AJCC revises the TNM classification system every few years with an attempt to formulate it for more accurate patient prognostication[5]. The latest 7<sup>th</sup> edition has further detailed the subclassification of the pN category and the assessment of discontinuous/satellite tumour foci. However, these revisions have increased the complexity and subjectivity during evaluation, and thus might lead to inter-observer variability and hamper its efficiency in routine clinical practise[5,6]. In addition, current clinicopathological parameters are insufficient to address the great biologic and genetic heterogeneity of CRC, and thus are inadequate for accurate individual prognostic prediction. From the perspective of clinical oncology, a new classification scheme based on molecular biomarkers would be relatively more efficient and accurate.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Over the past decades, many researchers have attempted to establish gene expression signatures specifically for the diagnosis, prognostication and recurrence prediction of sporadic CRC. Transcriptional profiling promises a fairly dynamic view on the cellular functions, regulatory mechanisms and biochemical pathways involved in the disease pathogenesis and progression[7]. Various gene expression profiling techniques ranging from differential display, SAGE to microarrays have been utilised. Despite its wide application in gene expression profiling, microarray experiments have been subjected to various sources of variability, false-positives, as well as statistical and bioinformatic challenges. To date, none of the molecular markers described has been validated and employed in routine clinical practise owing to the poor reproducibility of the identified differentially expressed genes (DEGs) between different profiling platforms[8]. Although the KRAS mutation and mismatch repair status have showed promising prognostic and predictive values, they have yet to be incorporated into either routine pathological reporting systems or TNM staging systems[5].

Since most of the molecular studies on CRC were based in Western populations and different molecular changes were thought to underlie the development of sporadic CRC in populations with different genetic backgrounds, we aimed to investigate the changes in mRNA expression patterns in primary sporadic CRC tumours with regards to our Malaysian patients. In our study, we have employed a combined approach of a two-step ACP-based PCR and real-time reverse transcription PCR to characterise the gene expression patterns for both early- and advanced stage sporadic colorectal adenocarcinomas.

## MATERIALS AND METHODS

### Patient selection and specimen collection

All patients presented with histologically confirmed colorectal adenocarcinomas and were staged accordingly to the AJCC TNM staging system (Table 1). The staging of cancer was performed by taking into consideration their histopathological reports, computed tomography images, morphological evaluations during surgery and serum CEA levels. Initially, four CRC patients of Stages I - III were recruited for the preliminary ACP-based PCR analysis, while another 27 patients with CRC Stages I – IV were recruited for subsequent RT-qPCR analysis. All subjects were newly-diagnosed with CRC, and were not associated with any hereditary syndromes, previously diagnosed cancer, or positive family history of CRC.

The subjects were admitted to the University Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia, and underwent curative surgical resection between 2010 and 2011. None had received pre-operative chemoradiotherapy. The study protocol was approved by the Ethics Committee Board of UMMC (Ref. No.: 654.1), and written informed consent was obtained from all study subjects. Both CRC tumour and paired non-cancerous tissue specimens were immersed in *RNAlater* RNA Stabilization Reagent (Qiagen) immediately after excision and stored at  $-80^{\circ}\text{C}$ .

Table 1 Cancer staging of recruited subjects.

Subject	Cancer Stage
T1	Stage I / pT1N0M0
T2	Stage II / pT3N0M0
T3	Stage II / pT2N0M0
T4	Stage II / pT3N0M0
T5	Stage II / pT3N0M0
T6	Stage II / pT4N0M0
T7	Stage II / pT4N0M0
T8	Stage II / pT4N0M0
T9	Stage II / pT3N0M0
T10	Stage II / pT3N0M0
T11	Stage IV / pT3N2M1
T12	Stage IV
T13	Stage III / pT3N1M0
T14	Stage IV / pT3N1M1
T15	Stage III / pT3N1M0
T16	Stage III / pT3N2M0
T17	Stage IV / pT4N1M1
T18	Stage III / pT3N1M0
T19	Stage IV / pT3N0M1
T20	Stage III / pT4N1M0
T21	Stage III
T22	Stage II
T23	Stage III / pT3N1M0
T24	Stage II / pT3-4N0M0
T25	Stage IV / pT4N1M1
T26	Stage II / pT3N0M0
T27	Stage III / pT3N1M0



## Total RNA extraction

Total RNA was extracted from homogenised colonic tissues with the RNeasy Plus Mini Kit (Qiagen) according to manufacturer's instructions. Subsequently, the RNA yield and integrity were ascertained via Agilent 2100 Bioanalyzer in conjunction with Agilent RNA 6000 Nano Kits (Agilent Technologies). The values of RIN were then determined in order to assess the integrity of the isolated total RNA. In this study, only RNA samples with RIN values of 8.0 – 10.0 and rRNA ratios [28S/18S] of 1.5 – 2.5 were selected for successive applications.

## ACP-based PCR analysis

### a) First-strand cDNA synthesis

The synthesis of first-strand cDNA was performed according to the manufacturer's protocol for the GeneFishing DEG Premix Kit (Seegene), as follows: 3 µg of total RNA was added with 2 µl of 10 µM dT-ACP1 (5'-CTGTGAATGCTGCGACTACGA TXXXXX(T)<sub>18</sub>-3') and RNase-free water to a final volume of 9.5 µl. The mixture was then incubated at 80 °C for 3 min, followed by chilling on ice for another 2 min. Subsequently, 4 µl of 5X RT buffer (Mbiotech), 5 µl of 2mM dNTP (Fermentas), 0.5 µl of 40 U/µl RNase inhibitor (Mbiotech) and 1 µl of 200 U/µl M-MLV reverse transcriptase (Mbiotech) were added. This mixture was then incubated at 42 °C for 90 min, heated at 94 °C for another 2 min and chilled on ice for 2 min. Finally, 80 µl of DNase-free water was added to dilute the synthesised cDNA. The first-strand cDNA was stored under -20 °C until further analysis.

### b) ACP-based GeneFishing PCR

First, all four cDNA samples within each CRC and control group samples were pooled together in equal amounts. The characterisation of DEGs was then conducted via ACP-based PCR based on 20 arbitrary ACP primers (Cat. No.: K1021) in a thermal cycler (Mastercycler Gradient, Eppendorf) according to the manufacturer's protocol (GeneFishing DEG Premix Kit, Seegene). Initially, the synthesis of second-strand cDNA was commenced in a one-cycle first-stage PCR: 94 °C for 5 min, 50 °C for 3 min and 72 °C for 1 min. Next, the constructed second-strand cDNA was subjected to second-stage PCR with 40 cycles of a denaturing step

1  
2  
3  
4 at 94 °C for 40 sec, annealing step at 65 °C for 40 sec and extension step at 72 °C for 40 sec.  
5  
6 Lastly, a final extension step at 72 °C for 5 min was carried out. The amplified products were  
7  
8 then separated on 3 % (w/v) agarose gels stained with ethidium bromide.  
9

#### 10 11 c) Cloning and sequencing

12  
13 The identified differentially expressed bands were extracted from the agarose gel by using the  
14  
15 PureLink Quick Gel Extraction Kit (Invitrogen). Each of these extracted DNA fragments was  
16  
17 then individually cloned with the use of the TOPO TA Cloning Kit for Sequencing  
18  
19 (Invitrogen). Subsequently, the plasmid containing the inserted DNA fragment was extracted  
20  
21 from clones of interest via PureLink Quick Plasmid Miniprep Kit (Invitrogen). The isolated  
22  
23 cloned plasmids were then sequenced with the ABI 3730xl DNA Analyzer (Applied  
24  
25 Biosystems). Finally, all the sequences obtained were analysed and matched for similarities  
26  
27 with reference to the BLAST programme under the NCBI database.

#### 28 29 **RT-qPCR analysis**

##### 30 31 a) Reverse transcription

32  
33 The total RNA isolated from 27 paired samples was reverse transcribed to first-strand cDNA,  
34  
35 with the following protocol: 3 µg of total RNA was added with 2 µl of 0.5 µg/µl oligo(dT)<sub>12-18</sub>  
36  
37 (Invitrogen) and RNase-free water to a final volume of 9.5 µl. The reaction mixture was then  
38  
39 incubated at 80°C for 3 min, followed by chilling on ice for another 2 min. Next, 4 µl of 5X  
40  
41 first strand buffer (Invitrogen), 5 µl of 2mM dNTP (Fermentas), 0.5 µl of 40 U/µl RNaseOUT  
42  
43 recombinant RNase inhibitor (Invitrogen) and 1 µl of 200 U/µl M-MLV reverse transcriptase  
44  
45 (Invitrogen) were added to the mixture. Finally, the reaction mixture was incubated at 42 °C  
46  
47 for 90 min, heated at 94 °C for another 2 min and chilled on ice for 2 min. The synthesised  
48  
49 first-strand cDNA was stored under -20 °C until further usage.

##### 50 51 b) $\Delta\Delta C_T$ analysis

52  
53 The relative expression of identified DEGs in all paired CRC tumours and control samples  
54  
55 was determined via  $\Delta\Delta C_T$  method. The RT-qPCR was performed in a singleplex reaction  
56  
57

1  
2  
3  
4 containing 50 ng first-strand cDNA under universal thermal cycling conditions with the ABI  
5 7500 Fast Real-Time PCR System (Applied Biosystems). Both *ACTB* (Assay ID:  
6 Hs99999903\_m1) and *GAPDH* (Assay ID: Hs99999905\_m1) were used as reference genes  
7 and are commercially available as TaqMan Pre-designed Assays (Applied Biosystems). Prior  
8 to the analysis of gene expression, the amplification efficiency for all target and reference  
9 genes assays was measured by using the standard curve method with 2-log measurements.  
10 The amplification efficiency value of 90 – 110 % was acceptable (Applied Biosystems). In  
11 this relative quantification method, the  $2^{-\Delta\Delta C_t}$  values obtained represented the fold change in  
12 gene expression of the CRC tumours, which was normalised with both reference genes, in  
13 relative to the calibrator (control sample)[9].  
14  
15  
16  
17  
18  
19  
20  
21  
22

### 23 c) Statistical analysis

24  
25 The difference in the expression level between CRC tumour and paired non-cancerous tissues  
26 was analysed by using Real-Time StatMiner software (Integromics). The paired t-test was  
27 then performed to assess the statistical significance of the observed differential expression  
28 patterns.  
29  
30  
31  
32

## 33 RESULTS

### 34 DEGs between CRC tumours and non-cancerous colonic tissues

35  
36 This preliminary study was conducted on paired samples pooled from four patients with CRC  
37 Stages I - III. In ACP-based GeneFishing PCR, 20 sets of arbitrary ACP primers were used to  
38 randomly amplify gene products in both CRC tumours and normal colonic samples. Upon  
39 visualisation on agarose gels, a total of 13 differentially expressed bands were observed by  
40 means of comparing bands intensity between the tumouric and non-cancerous samples, as  
41 shown in Figure 1. These bands were further sequenced for gene identification, and 16 DEGs  
42 were successfully reported. Among them, 13 were over-expressed in CRC tumours, whilst  
43 three were under-expressed, as listed in Table 2.  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 2 Sequence similarities and identification of DEGs.

Differentially Expressed Band	DEG	Identity	Sequence Homology (%)	Accession Number	UniGene Number	Description
<i>Over-expressed</i>						
A4.1	DEG1	<i>Homo sapiens</i> proteasome (prosome, macropain) 26S subunit, ATPase, 5 (PSMC5), mRNA	502/506 (99%)	NM_002805.4	Hs.79387	Involves in the ATP-dependent degradation of ubiquitinated proteins.
	DEG2	<i>Homo sapiens</i> ubiquinol-cytochrome c reductase hinge protein (UQCRH), mRNA	514/521 (98%)	NM_006004.2	Hs.481571	A component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex, which is part of the mitochondrial respiratory chain.
A4.2	DEG3	<i>Homo sapiens</i> ribosomal protein S23 (RPS23), mRNA	551/551 (100%)	NM_001025.4	Hs.527193	A component of the 40S subunit of human ribosomes.
A6.1	DEG4	<i>Homo sapiens</i> ribosomal protein L10 (RPL10), transcript variant 1, mRNA	554/557 (99%)	NM_006013.3	Hs.534404	A component of the 60S subunit of human ribosomes.
A9.2	DEG6	<i>Homo sapiens</i> actin related protein 2/3 complex, subunit 2, 34kDa (ARPC2), transcript variant 2, mRNA	473/473 (100%)	NM_005731.2	Hs.529303	Involves in the regulation of actin polymerization as an actin-binding component of the Arp2/3 complex, and mediates the formation of branched actin networks together with an activating nucleation-promoting factor (NPF).
	DEG7	<i>Homo sapiens</i> TIMP metalloproteinase inhibitor 1 (TIMP1), mRNA	503/511 (98%)	NM_003254.2	Hs.522632	Irreversibly inactivates the metalloproteinases by binding to their catalytic zinc cofactor.
A10.1	DEG8	<i>Homo sapiens</i> ATP synthase, H <sup>+</sup> -transporting, mitochondrial F1 complex, beta polypeptide (ATP5B), nuclear gene encoding mitochondrial protein, mRNA	917/919 (99%)	NM_001686.3	Hs.406510	A subunit of mitochondrial ATP synthase that catalyzes the synthesis of ATP by utilizing an electrochemical gradient of protons across the inner membrane during oxidative

						phosphorylation.
A13.2	DEG11	<i>Homo sapiens</i> chromosome 11 open reading frame 10 (C11orf10), mRNA	273/273 (100%)	NM_014206.3	Hs.437779	Unknown.
A13.3	DEG12	<i>Homo sapiens</i> mitochondrial ribosomal protein L24 (MRPL24), nuclear gene encoding mitochondrial protein, transcript variant 2, mRNA	408/411 (99%)	NM_024540.3	Hs.418233	Involves in protein synthesis within the mitochondrion.
A13.4	DEG13	<i>Homo sapiens</i> similar to OK/SW-CL.16 (LOC100288418)	635/644 (98%)	XM_002342023.1	-	Unknown.
A18.1	DEG14	<i>Homo sapiens</i> family with sequence similarity 96, member B (FAM96B), transcript variant 2, transcribed RNA	486/487 (99%)	NR_024525.1	Hs.9825	Involves in chromosome segregation as part of the mitotic spindle-associated MMXD complex.
A20.1	DEG15	<i>Homo sapiens</i> ribosomal protein L35 (RPL35), mRNA	440/446 (99%)	NM_007209.3	Hs.182825	A component of the 60S subunit of human ribosomes.
A20.2	DEG16	<i>Homo sapiens</i> chromosome 6 open reading frame173 (C6orf173), mRNA	551/554 (99%)	NM_001012507.2	Hs.486401	May be required for proper chromosome segregation during mitosis and involved with CENPT in the establishment of centromere chromatin structure.
<b><i>Under-expressed</i></b>						
A9.1	DEG5	<i>Homo sapiens</i> ribosomal protein L37 (RPL37), mRNA	284/284 (100%)	NM_000997.4	Hs.731513	A component of the 60S subunit of human ribosomes, and can bind to the 23S rRNA.
A13.1	DEG9	<i>Homo sapiens</i> solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1 (SLC25A1), nuclear gene encoding mitochondrial protein, mRNA	165/165 (100%)	NM_005984.2	Hs.111024	A mitochondrial tricarboxylate transporter which is responsible for the movement of citrate across the mitochondrial inner membrane.
	DEG10	<i>Homo sapiens</i> similar to cytochrome c oxidase subunit II (LOC100288578), miscRNA	141/146 (97%)	XR_078216.1	-	Unknown.

### Differential ability of the identified DEGs on early and advanced colorectal neoplasia

Following the identification of DEGs, the gene sequences obtained were then used to design primers and TaqMan probes for RT-qPCR analysis by Applied Biosystems, as listed in Table 3. In an attempt to assess the differential ability of identified DEGs on early and advanced colorectal adenocarcinoma, the recruited paired samples were further stratified into two groups according to the cancer stage. Among them, 13 patients with Stages I and II were grouped as early stage CRC, whilst the advanced stage CRC group comprised of 14 patients with Stages III and IV.

Table 3 Primers and TaqMan probes for relative quantification with Comparative C<sub>T</sub> method.

DEG	Primers Sequence		TaqMan Probe Sequence
DEG1	Forward:	5'-GGGCGTGTGCACAGAAG-3'	5'-CTCGCAGGGCATAACAT-3'
	Reverse:	5'-AAGTCCTCCTGAGTGACATGGA-3'	
DEG2	Forward:	5'-GATGCTTACCGAATCCGGAGATC-3'	5'-CCTCTTCTCTCCTCCTCC-3'
	Reverse:	5'-GCATTGCTCTCTCACTGTTGTTAG-3'	
DEG3	Forward:	5'-CAACCGTCATTGGGTACAAAGG-3'	5'-ATGGCAAGAAAATCAC-3'
	Reverse:	5'-TGTAAGGGTCCAGCTGATCAAGA-3'	
DEG4	Forward:	5'-CGGCCAGGAACTTGAACCTG-3'	5'-CAGGGCCTCAATCACA-3'
	Reverse:	5'-CCGAGCTGCAGAACAAGGA-3'	
DEG5	Forward:	5'-CTGGTTCGAATGAGGCACCTAAAA-3'	5'-CATGCCTGAATCTGC-3'
	Reverse:	5'-TGGGTTTAGGTGTTTCCTCAC-3'	
DEG6	Forward:	5'-AGATTAGCGGGATGAAAACGTCTT-3'	5'-CCCCGTGATTGTTTTTC-3'
	Reverse:	5'-CGCCCAGATGCCGAGAAAA-3'	
DEG7	Forward:	5'-GGTAGTGATGTGCAAGAGTCCAT-3'	5'-CATTGCTGGAAAACCTG-3'
	Reverse:	5'-CCGCAGCGAGGAGTTTCT-3'	
DEG8	Forward:	5'-GAAGGAGACCATCAAAGGATTCCA-3'	5'-ATTCACCTGCCAAAATC-3'
	Reverse:	5'-GAAGGCCTGTTCTGGGAGATG-3'	
DEG9	Forward:	5'-GGCAGGGTGGTCTGAGA-3'	5'-CCTCTCTCCGCCCCGGACA-3'
	Reverse:	5'-CCGCCATTGGCCTTAACTG-3'	
DEG11	Forward:	5'-CAGGTTTCAGTGAAGCCATCTG-3'	5'-CACCCAAGGGTAACAAC-3'
	Reverse:	5'-GGGTTGGCATCTACGTGTGA-3'	
DEG12	Forward:	5'-CCAGGTCAAACCTGTGGATCCT-3'	5'-ATGGACAGGAAACCCAC-3'
	Reverse:	5'-GCTTCAGTAAATCTCCACTCGATCT-3'	
DEG14	Forward:	5'-CCCGCTCCTTATCTGCAAGTT-3'	5'-CATGCAGTGAACAAGC-3'
	Reverse:	5'-TCAAGATGGACGTGCACATTACTC-3'	
DEG15	Forward:	5'-CGGCCTCCAAGCTCTCT-3'	5'-CCGGACGACTCGGATCT-3'
	Reverse:	5'-TGAGAACACGGGCAATGGATTT-3'	
DEG16	Forward:	5'-GGACTCTTCTGCTAATCGATGAACA-3'	5'-CAGATGGACCAATAAGTCA-3'
	Reverse:	5'-GCCTCAACTTCGTCTGGAGAAAA-3'	

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

The analysis of RT-qPCR results was performed via Real-Time StatMiner software by importing the raw Ct data. The within-group correlation then determined by calculating the MAD for all the samples within the same experimental group. The biological samples which do not correlate well with other samples in the same group, were detected as group outliers and excluded from subsequent analysis. Both *ACTB* and *GAPDH* were used for normalisation in computing the  $\Delta C_T$  (Figure 2) and  $2^{-\Delta\Delta C_T}$  values (Table 4). The statistical significance of the observed fold change in expression was determined by paired t-test for all the DEGs. A  $p$  value of less than 0.05 is considered as statistically significant (Table 4).

In both early and advanced stage CRC groups, the expression of four out of 16 DEGs was reported to be significantly differed between tumouric and non-cancerous tissues. Remarkably, the combination of this panel of four genes is different among two groups. The *RPL35*, *RPS23* and *TIMP1* genes were found to be over-expressed in both early- and advanced colorectal neoplasms ( $p < 0.05$ ) (Figures 3 and 4). It is interesting to note that, the under-expression of *ARPC2* gene ( $p < 0.05$ ) was only observed in early stage CRC tumours (Figure 3). On the other hand, the *C6orf173* gene was found to be over-expressed ( $p < 0.05$ ) in advanced colorectal adenocarcinomas, but not in early stage CRC tumours (Figure 4).

Table 4  $\Delta C_T$  mean,  $\Delta\Delta C_T$ ,  $2^{-\Delta\Delta C_T}$  and  $p$  values for all the DEGs in both early- and advanced stage CRC groups.

DEG	Early Stage CRC					Advanced Stage CRC				
	$\Delta C_T$ Mean (CRC)	$\Delta C_T$ Mean (Normal)	$\Delta\Delta C_T$	$2^{-\Delta\Delta C_T}$	$p$ value	$\Delta C_T$ Mean (CRC)	$\Delta C_T$ Mean (Normal)	$\Delta\Delta C_T$	$2^{-\Delta\Delta C_T}$	$p$ value
<i>ARPC2</i>	2.6854	2.0664	0.6190	0.6511	0.0282*	2.7240	2.3300	0.3940	0.7610	0.2424
<i>ATP5B</i>	1.5846	1.2702	0.3144	0.8042	0.3524	1.9558	1.3838	0.5720	0.6727	0.1484
<i>C11orf10</i>	3.2897	3.3639	-0.0742	1.0528	0.8333	3.3281	3.6709	-0.3428	1.2682	0.3710
<i>C6orf173</i>	6.1083	7.1943	-1.0860	2.1228	0.0905	5.9949	7.9087	-1.9138	3.7680	0.0013*
<i>FAM96B</i>	3.5602	3.8955	-0.3353	1.2616	0.2935	3.5276	3.9920	-0.4644	1.3797	0.2113
<i>MRPL24</i>	4.9171	5.0839	-0.1668	1.1226	0.3564	4.9728	5.1467	-0.1739	1.1281	0.7001
<i>PSMC5</i>	3.8232	3.9617	-0.1385	1.1008	0.6812	3.7705	3.8455	-0.0750	1.0534	0.8048
<i>RPL10</i>	-0.7462	-0.4853	-0.2609	1.1982	0.4001	-1.1576	-0.5196	-0.6380	1.5562	0.0950
<i>RPL35</i>	-0.1926	0.6222	-0.8148	1.7591	0.0024*	0.1748	0.8769	-0.7021	1.6269	0.0372*
<i>RPL37</i>	-0.0059	-0.1539	0.1480	0.9025	0.8645	0.2184	0.7143	-0.4959	1.4102	0.1537
<i>RPS23</i>	0.2176	0.7739	-0.5563	1.4705	0.0310*	0.0676	0.9431	-0.8755	1.8346	0.0250*
<i>SLC25A1</i>	3.7514	3.5430	0.2084	0.8655	0.5721	3.5565	3.4428	0.1137	0.9242	0.7991
<i>TIMP1</i>	2.9096	4.3059	-1.3963	2.6323	0.0440*	2.3330	3.8547	-1.5217	2.8713	0.0062*
<i>UQCRH</i>	2.0087	2.2216	-0.2129	1.1590	0.4108	2.3375	2.4459	-0.1084	1.0780	0.7808

\* $p < 0.05$  = statistically significant



## DISCUSSION

Our current study has revealed two distinctive 4-gene signatures for both early- and advanced stage colorectal adenocarcinomas. The early stage sporadic CRC was characterised by the over-expression of *RPL35*, *RPS23* and *TIMP1* genes, as well as under-expression of *ARPC2* gene. On the other hand, the advanced primary CRC tumours were reported with over-expression of *C6orf173*, *RPL35*, *RPS23* and *TIMP1* genes. Although the relative fold change for *ARPC2*, *RPL35* and *RPS23* genes is below 2, the individual result does not affect the analysis since gene expression patterns of all four genes in combination were proposed to distinguish between the early- and advanced stage CRC neoplasms. The potential involvement of these DEGs and their altered expression levels in CRC were further supported by previous researches.

In fact, several proto-oncogenes and tumour suppressors are previously reported to regulate the ribosome production, i.e., the *RB*[10], *TP53*[11], *PTEN* genes[12], as well as the *MYC* gene family[13]. It is suggested that the alterations in ribosome biogenesis might affect the translation of genes that are involved in neoplastic transformation. In addition, the additional extra-ribosomal functions of the ribosomal proteins (r-proteins) in cellular apoptosis, cellular proliferation, cellular transformation, genes transcription, mRNA translation, DNA repair and inflammation, might also trigger and support the neoplastic development[14]. Hence, the over-expression of r-proteins-encoding genes observed in colorectal adenocarcinomas is not unexpected[15-17]. Our current study has revealed the significant over-expression of two r-proteins that were not previously described in CRC tumours, i.e., the *RPL35* and *RPS23*. The observed fold changes for the *RPL35* and *RPS23* mRNA levels were comparable between the early- and advanced stage CRC tumours in our sample cohort. This was in agreement with previous reports by Barnard et al. and Frigerio et al., where the changes in the mRNA expression levels of the r-proteins were irrespective of the cancer stage[18,19]. The hypothesis that the same ribosomal protein may contribute in different stages of cancer progression with their hitherto unknown extra-ribosomal roles might provide an explanation to these observations[20].

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

On the other hand, our present study also demonstrated an over-expression of the *TIMP1* gene in both early- and advanced stage primary CRC tumours. This finding is supported by Zeng et al., where the over-expression of TIMP1 was reported in all stages of primary CRC tumours[21]. Under normal physiological conditions, the proteolytic activities of MMPs are kept at bay by their natural inhibitors, the TIMPs[22]. Previous studies have reported the over-expression of MMPs in both early- and advanced stage CRC tumours, as well as other cancer types[23-25], which is in accordance to their biological roles. Hence, a similar scenario is expected for TIMPs and indeed, their suppressive role in tumour invasion and metastasis has been demonstrated in various cancer models[26]. However, more recent studies have revealed a direct correlation between TIMP1 expression and tumour aggressiveness in cancer, including CRC[21,27]. These findings, which are contradictory to its protease-inhibiting function, have suggested a possible tumour-promoting role of TIMP1 in tumorigenesis. It is postulated that the TIMP1 exhibits the abilities to inhibit tumour cell apoptosis and promote tumour angiogenesis, as well as other growth-factor-like effects[28]. In our present study, the observed comparable over-expression of TIMP1 in both early- and advanced stage sporadic CRC neoplasms was in line with its MMP inhibitory and MMP-independent tumour-promoting activities.

In cancer biology, the expression of mRNAs and proteins of the ARP2/3 complex is often studied due to its role in cell migration, which contributes to cancer invasion and metastasis if aberrantly regulated[29]. We have detected a significant under-expression of ARPC2 in our cohort of early stage primary CRC tumours. Surprisingly, this finding is contradictory with the role played by ARPC2 in cancer invasion and metastasis theoretically. Previously, Kaneda et al. has reported the decreased expression of all the seven genes encoding the subunits of ARP2/3 complex in human gastric cancers. Among them, the Arp2, ARPC2 and ARPC3 showed the most prominent reduction in their expression levels[30]. The exact mechanism underlying this observation still remains unknown, but the epigenetic alteration might potentially provide an explanation for it. For instance, promoter hypermethylation that causes gene silencing is responsible for the reduced expression of ARPC1 in human gastric

1  
2  
3  
4 cancer[31]. Similarly, the epigenetic study might also offer a clue for the under-expression of  
5 ARPC2 in CRC neoplasms.  
6  
7

8  
9 *C6orf173*, which is also known as *CUG2* or *CENP-W*, is a novel oncogene that has been  
10 found to be up-regulated in many human cancer tissues. Its high expression level is  
11 profoundly reported in tumours of the ovary, liver, lung, pancreas, breast, colon, rectum and  
12 stomach. The CENP-W is a new member of the constitutive centromere-associated network,  
13 which specifically interacts with the CENP-T and plays an important role in mitosis[32]. In  
14 our current study, the CENP-W is over-expressed in advanced CRC adenocarcinoma. This  
15 finding correlates to its function in kinetochore assembly, where its aberrant expression might  
16 lead to abnormal cell division and aneuploidy in cancer[32]. In our study, the over-expression  
17 of CENP-W was observed in both early- and advanced cohort of CRC neoplasms but only  
18 statistically significant in the latter group. Given the fact that aneuploidy is constantly  
19 associated with a greater proportion of advanced CRC cases, the aberrant expression of  
20 CENP-W might potentially relate to a poorer prognosis of CRC[33].  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

31 In conclusion, we have characterised two distinctive gene expression patterns, which comprise  
32 of the *ARPC2*, *C6orf173*, *RPL35*, *RPS23* and *TIMP1* genes, for the stratification of primary  
33 colorectal adenocarcinomas among Malaysian CRC patients. Our current sample size was  
34 relatively small owing to the lack of a designated Tissue Bank in our institution. There were  
35 also not many CRC patient volunteers. Moreover, our stringent criteria for patient selection  
36 have also limited the availability of suitable specimens within the short sample collection  
37 period. Nevertheless, our identified mRNA expression patterns specific for early- and  
38 advanced stage CRC tumours are still convincing with our stringent sample selection criteria,  
39 high specificity primers and probes, as well as reliable statistical analysis. In future, the  
40 validation of these DEGs should be performed on a larger set of clinical samples, and  
41 extensive inter-laboratory testing of their differential abilities on each CRC stage is also  
42 desired. In addition, we should also integrate other imaging and histological information to  
43 complement our identified gene expression patterns, which then hold promises for better  
44 stratification of CRC tumours.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 **Funding** – This study was supported by FS176/2007C, PS172/2008C and Research  
5 Collaborative Grant, CG041-2013 from the University of Malaya.  
6  
7  
8

9  
10 **Contributorship:** TPL, LHL, PCL and KHC were responsible for the design of the study and  
11 analysis. ACR, IH and KLG were involved in samples collection. TPL performed the  
12 experiment. All authors were involved in drafting the manuscript.  
13  
14

15  
16  
17 **Competing interests** – None.  
18  
19

20 **Data sharing statement** – There is no additional data available.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**REFERENCES**

1. Greene FL, Page D, Fleming ID, et al., eds. AJCC cancer staging manual (6th ed.). New York: Springer 2002.
2. Dukes CE. The classification of cancer of the rectum. *J Pathol Bacteriol* 1932; 35: 323-32.
3. Astler VB, Coller FA. The prognostic significance of direct extension of carcinoma of the colon and rectum. *Ann Surg* 1954; 139: 846-52.
4. Edge SB, Byrd DR, Compton CC, et al. AJCC cancer staging manual (7th ed.). New York: Springer 2010.
5. Hu HK, Krasinskas A, Willis J. Perspectives on current tumour-node-metastasis (TNM) staging of cancers of the colon and rectum. *Semin Oncol* 2011; 38: 500-10.
6. Doyle VJ, Bateman AC. Colorectal cancer staging using TNM 7: is it time to use this new staging system? *J Clin Pathol* 2012; 65: 372-4.
7. Russo G, Zegar C, Giordano A. Advantages and limitations of microarray technology in human cancer. *Oncogene* 2003; 22: 6497-507.
8. Puppa G, Sonzogni A, Colombari R, et al. TNM staging system of colorectal carcinoma: a critical appraisal of challenging issues. *Arch Pathol Lab Med* 2010; 134: 837-52.

- 1  
2  
3  
4  
5 9. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time  
6  
7 quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 2001; 25: 402-8.  
8  
9
- 10  
11 10. Voit R, Schafer K, Grummt I. Mechanism of repression of RNA polymerase I  
12  
13 transcription by the retinoblastoma protein. *Mol Cell Biol* 1997; 17: 4230-7.  
14
- 15  
16 11. Zhai W, Cornai L. Repression of RNA polymerase I transcription by the tumour  
17  
18 suppressor p53. *Mol Cell Biol* 2000; 20: 5930-8.  
19  
20
- 21  
22 12. Backman S, Stambolic V, Mak T. PTEN function in mammalian cell size regulation.  
23  
24 *Curr Opin Neurobiol* 2002; 12: 516-22.  
25  
26
- 27  
28 13. Greasley PJ, Bonnard C, Amati B. Myc induces the nucleolin and BN51 genes:  
29  
30 possible implications in ribosome biogenesis. *Nucleic Acids Res* 2000; 28: 446-53.  
31  
32
- 33  
34 14. Montanaro L, Treré D, Derenzini M. Nucleolus, ribosomes, and cancer. *Am J Pathol*  
35  
36 2008; 173: 301-10.  
37  
38
- 39  
40 15. Sharp MG, Adams SM, Elvin P, et al. A sequence previously identified as metastasis-  
41  
42 related encodes an acidic ribosomal phosphoprotein, P2. *Br J Cancer* 1990; 61: 83-8.  
43  
44
- 45  
46 16. Chester KA, Robson L, Begent RH, et al. Identification of a human ribosomal protein  
47  
48 mRNA with increased expression in colorectal tumours. *Biochim Biophys Acta* 1989;  
49  
50 1009: 297-300.  
51  
52
- 53  
54 17. Pogue-Geile K, Geiser JR, Shu M, et al. Ribosomal protein genes are overexpressed in  
55  
56 colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein.  
57  
58 *Mol Cell Biol* 1991; 11: 3842-9.  
59  
60

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
18. Barnard GF, Staniunas RJ, Mori M, et al. Gastric and hepatocellular carcinomas do not overexpress the same ribosomal protein messenger RNAs as colonic carcinoma. *Cancer Res* 1993; 53: 4048-52.
19. Frigerio JM, Dagorn JC, Iovanna JL. Cloning, sequencing and expression of the L5, L21, L27a, L28, S5, S9, S10 and S29 human ribosomal protein mRNAs. *Biochim Biophys Acta* 1995; 1262: 64-8.
20. Lai MD, Xu J. Ribosomal proteins and colorectal cancer. *Curr Genomics* 2007; 8: 43-9.
21. Zeng ZS, Cohen AM, Zhang ZF, et al. Elevated tissue inhibitor of metalloproteinase 1 RNA in colorectal cancer stroma correlates with lymph node and distant metastases. *Clin Cancer Res* 1995; 1: 899-906.
22. Ennis BW, Matrisian LM. Matrix degrading metalloproteinases. *J Neurooncol* 1994; 18: 105-9.
23. Urbanski SJ, Edwards DR, Maitland A, et al. Expression of metalloproteinases and their inhibitors in primary pulmonary carcinomas. *Br J Cancer* 1992; 66: 1188-94.
24. Boag AH, Young ID. Immunohistochemical analysis of type IV collagenase expression in prostatic hyperplasia and adenocarcinoma. *Mod Pathol* 1993; 6: 65-8.
25. Newell KJ, Witty JP, Rodgers WH, et al. Expression and localisation of matrix-degrading metalloproteinases during colorectal tumourigenesis. *Mol Carcinogen* 1994; 10: 199-206.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
26. Khokha R, Waterhouse P. The role of tissue inhibitor of metalloproteinase-1 in specific aspects of cancer progression and reproduction. *J Neurooncol* 1994; 18: 123-7.
  27. Lu XQ, Levy M, Weinstein IB, et al. Immunological quantitation of levels of tissue inhibitor of metalloproteinase-1 in human colon cancer. *Cancer Res* 1991; 51: 6231-5.
  28. Egeblad M, Werb Z. New functions for the matrix metalloproteinases in cancer progression. *Nat Rev Cancer* 2002; 2: 161-74.
  29. Yamaguchi H, Wyckoff J, Condeelis J. Cell migration in tumours. *Curr Opin Cell Biol* 2005; 17: 559-64.
  30. Kaneda A, Kaminishi M, Sugimura T, et al. Decreased expression of the seven ARP2/3 complex genes in human gastric cancers. *Cancer Lett* 2004; 212: 203-10.
  31. Kaneda A, Kaminishi M, Nakanishi Y, et al. Reduced expression of the insulin-induced protein 1 and p41 ARP2/3 complex genes in human gastric cancers. *Int J Cancer* 2002; 100: 57-62.
  32. Hori T, Amano M, Suzuki A, et al. CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* 2008; 135: 1039-52.
  33. Chen HS, Sheen-Chen SM, Lu CC. DNA index and S-phase fraction in curative resection of colorectal adenocarcinoma: analysis of prognosis and current trends. *World J Surg* 2002; 26: 626-30.



For peer review only

### Figure Legends

Figure 1 Differential banding patterns on 3 % agarose gel post ACP-based PCR amplification between normal colon and CRC tumour samples (N: normal sample; C: CRC sample)

- 1  
2  
3  
4 Figure 2 Box-plots showing  $\Delta C_T$  values of all CRC tumours and normal colonic tissues  
5 in each early- (a) and advanced (b) stage CRC group.  
6  
7  
8  
9 Figure 3 Differential expression patterns of all the identified DEGs in early stage CRC  
10 group.  
11  
12  
13 Figure 4 Differential expression patterns of all the identified DEGs in advanced stage  
14 CRC group.  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

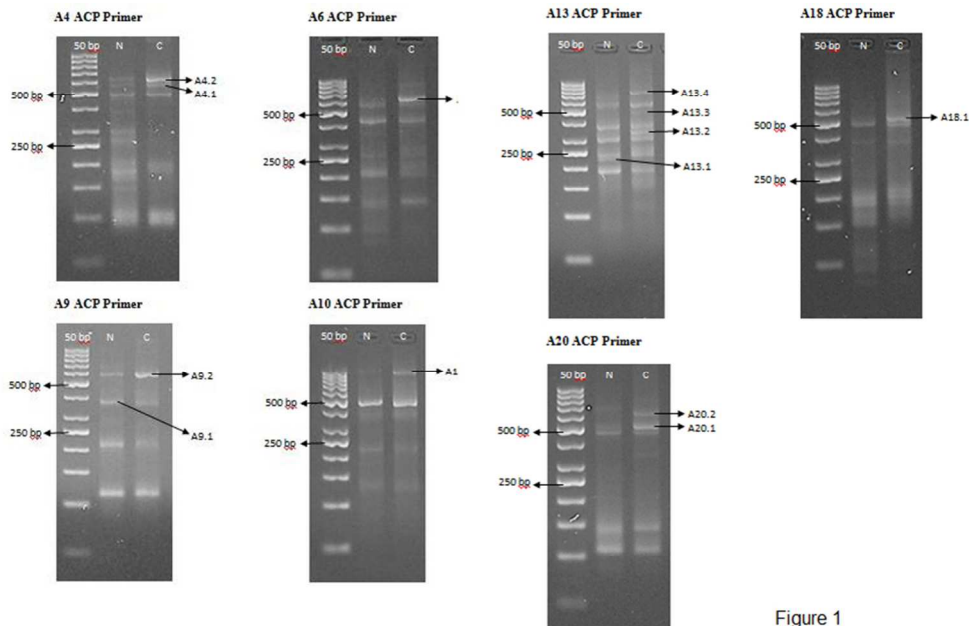


Figure 1

Figure 1  
73x47mm (300 x 300 DPI)

review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

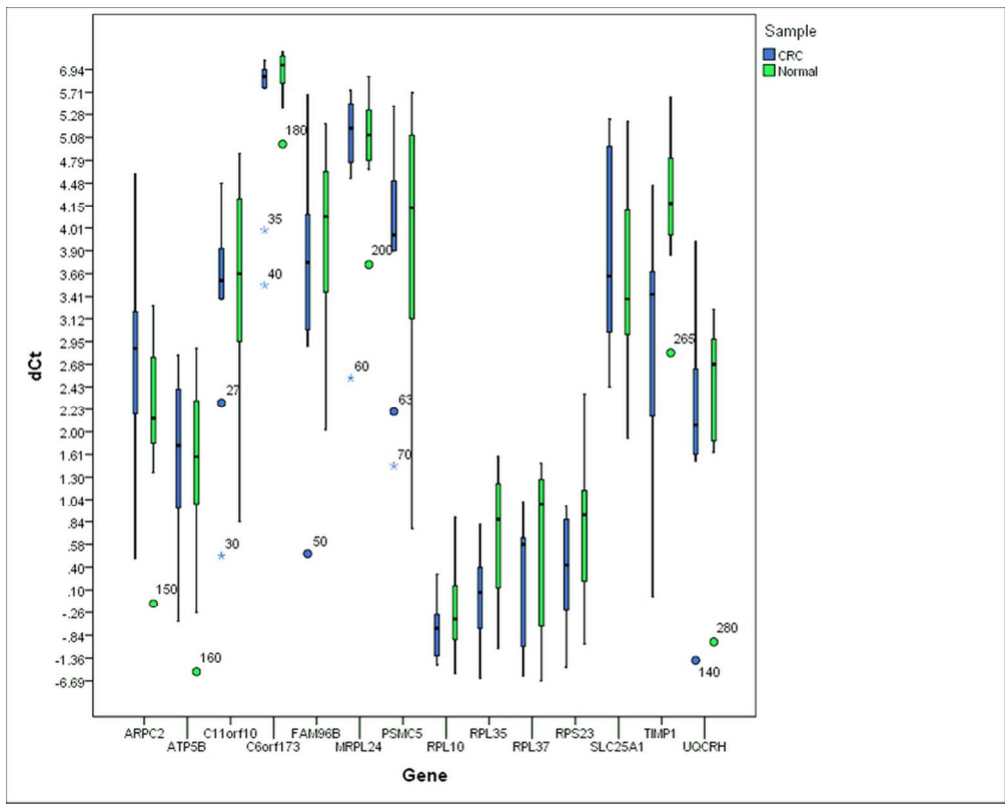


Figure 2(a)  
71x56mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

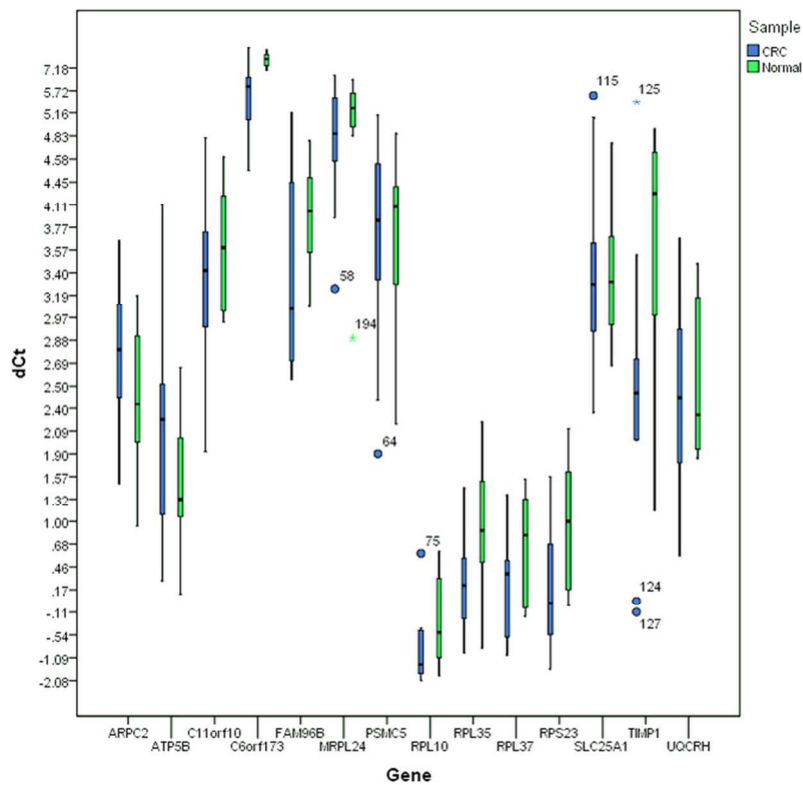


Figure 2(b)  
71x57mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

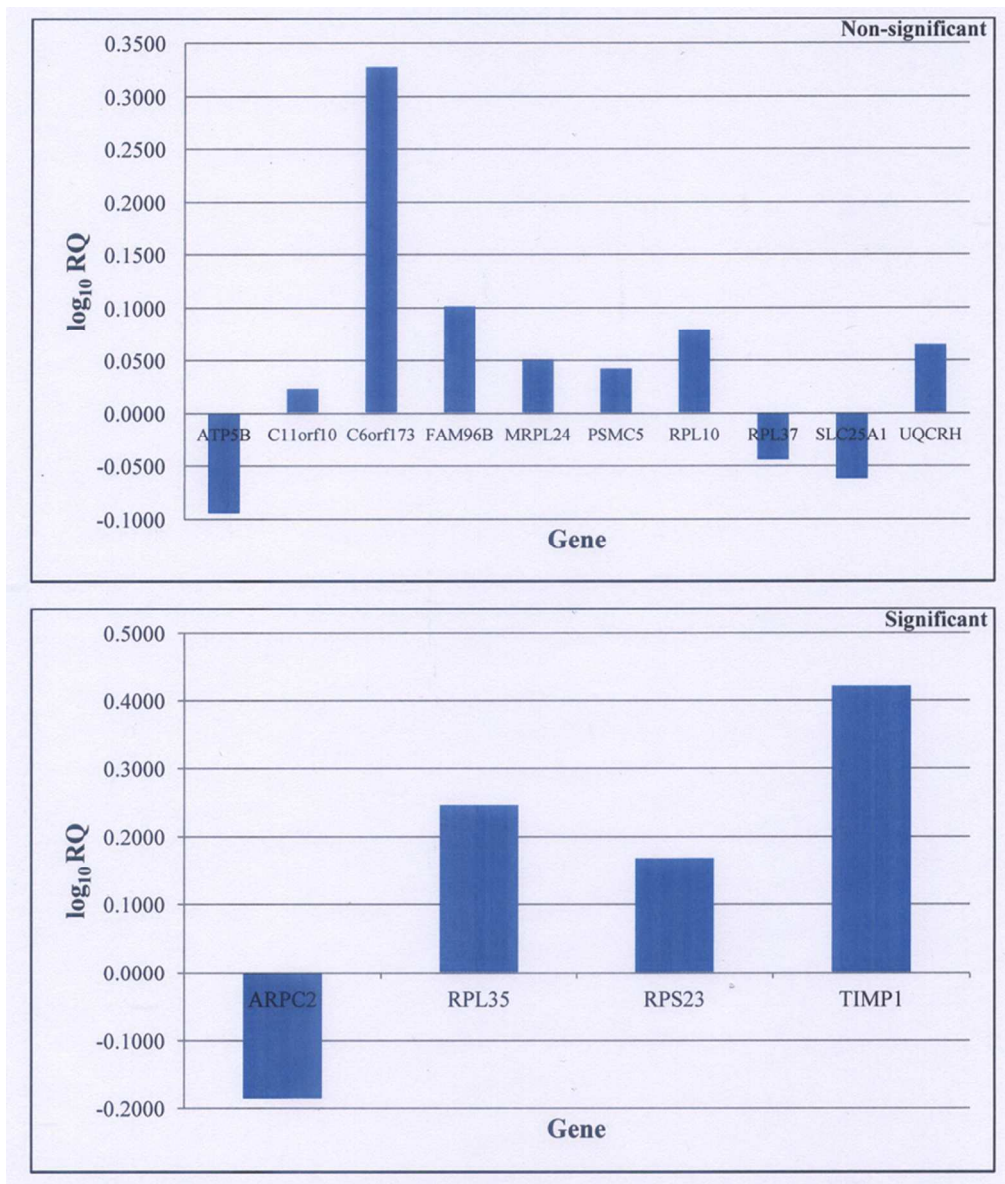


Figure 3  
238x280mm (96 x 96 DPI)

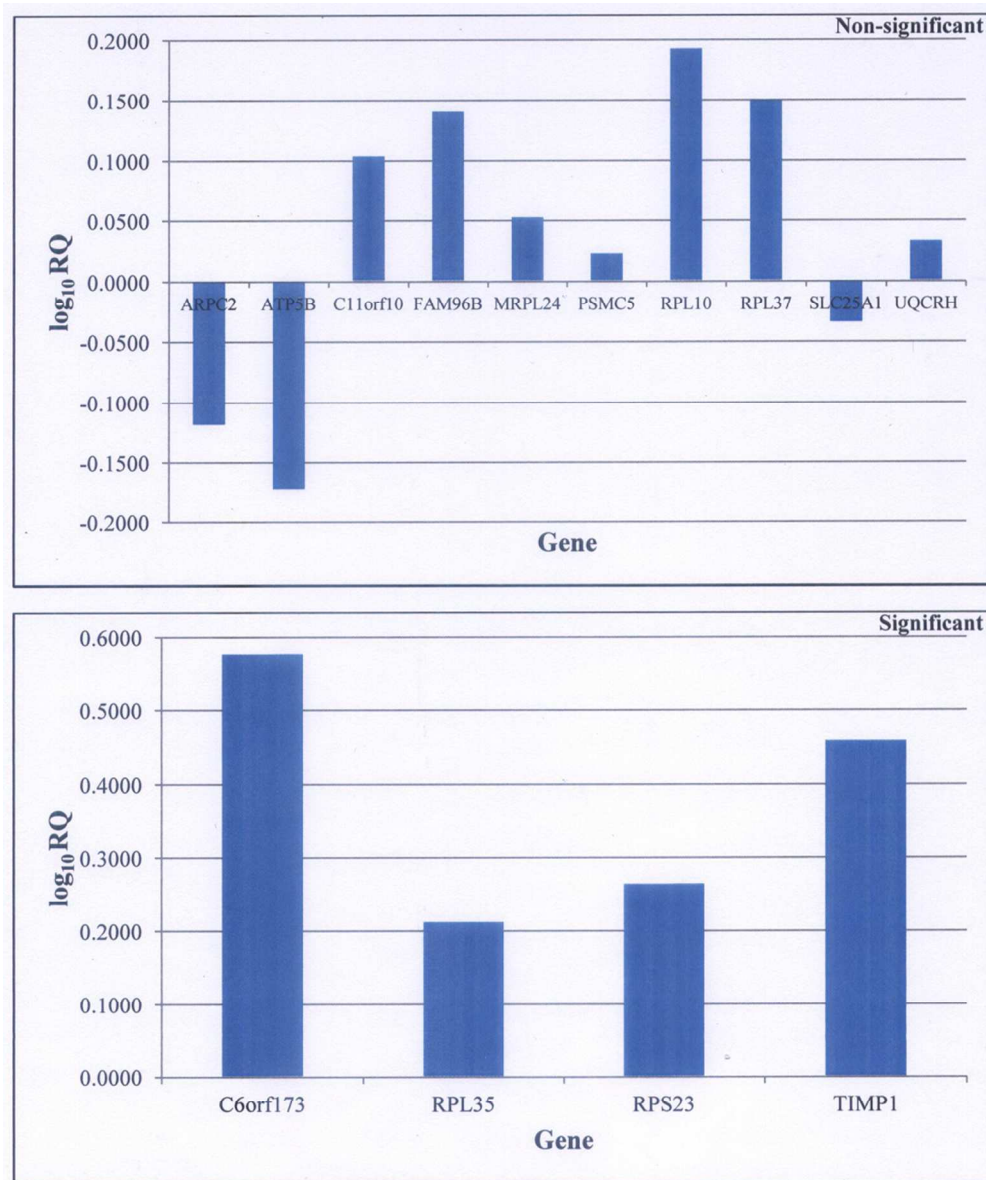


Figure 4  
293x350mm (96 x 96 DPI)

## STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
<b>Title and abstract</b>	1	<p>✓ (a) Indicate the study's design with a commonly used term in the title or the abstract</p> <p>✓ (b) Provide in the abstract an informative and balanced summary of what was done and what was found</p>
<b>Introduction</b>		
Background/rationale	2	✓ Explain the scientific background and rationale for the investigation being reported
Objectives	3	✓ State specific objectives, including any prespecified hypotheses
<b>Methods</b>		
Study design	4	✓ Present key elements of study design early in the paper
Setting	5	✓ Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	<p>✓ (a) <i>Cohort study</i>—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up</p> <p>✓ <i>Case-control study</i>—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls</p> <p><i>Cross-sectional study</i>—Give the eligibility criteria, and the sources and methods of selection of participants</p> <p>✓ (b) <i>Cohort study</i>—For matched studies, give matching criteria and number of exposed and unexposed</p> <p>✓ <i>Case-control study</i>—For matched studies, give matching criteria and the number of controls per case</p>
Variables	7	✓ Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	✓ For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	✓ Describe any efforts to address potential sources of bias
Study size	10	✓ Explain how the study size was arrived at
Quantitative variables	11	✓ Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	<p>✓ (a) Describe all statistical methods, including those used to control for confounding</p> <p>✓ (b) Describe any methods used to examine subgroups and interactions</p> <p>✓ (c) Explain how missing data were addressed</p> <p>✓ (d) <i>Cohort study</i>—If applicable, explain how loss to follow-up was addressed</p> <p>✓ <i>Case-control study</i>—If applicable, explain how matching of cases and controls was addressed</p> <p><i>Cross-sectional study</i>—If applicable, describe analytical methods taking account of sampling strategy</p> <p>✓ (e) Describe any sensitivity analyses</p>

Continued on next page



**Results**

Participants	13*	<p>√ (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed</p> <p>√ (b) Give reasons for non-participation at each stage</p> <p>(c) Consider use of a flow diagram</p>
Descriptive data	14*	<p>√ (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders</p> <p>(b) Indicate number of participants with missing data for each variable of interest</p> <p>(c) <i>Cohort study</i>—Summarise follow-up time (eg, average and total amount)</p>
Outcome data	15*	<p><i>Cohort study</i>—Report numbers of outcome events or summary measures over time</p> <p>√ <i>Case-control study</i>—Report numbers in each exposure category, or summary measures of exposure</p> <p><i>Cross-sectional study</i>—Report numbers of outcome events or summary measures</p>
Main results	16	<p>√ (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included</p> <p>(b) Report category boundaries when continuous variables were categorized</p> <p>(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period</p>
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses

**Discussion**

Key results	18	√ Summarise key results with reference to study objectives
Limitations	19	√ Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	√ Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results

**Other information**

Funding	22	√ Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based
---------	----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------

\*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at [www.strobe-statement.org](http://www.strobe-statement.org).

# BMJ Open

## Characterisation Of Differential mRNA Expression Profiles For Early And Advanced Stage Sporadic Colorectal Adenocarcinomas In A Malaysian Patient Cohort

Journal:	<i>BMJ Open</i>
Manuscript ID:	bmjopen-2014-004930.R1
Article Type:	Research
Date Submitted by the Author:	16-Jun-2014
Complete List of Authors:	Lau, Tze Pheng; University of Malaya, Biomedical Science Roslan, April Camilla; University of Malaya, Surgery Lian, Lay Hoong; University of Malaya, Biomedical Science Lee, Ping Chin; Universiti Sabah Malaysia, School of Science and Technology Hilmi, Ida; University of Malaya, Medicine Goh, Khean Lee; University of Malaya, Medicine Chua, Kek Heng; University of Malaya, Biomedical Science
<b>Primary Subject Heading</b>:	Gastroenterology and hepatology
Secondary Subject Heading:	Gastroenterology and hepatology
Keywords:	Gastrointestinal tumours < GASTROENTEROLOGY, Cancer genetics < GENETICS, PUBLIC HEALTH

SCHOLARONE™  
Manuscripts

Only

1  
2  
3  
4 **Characterisation Of Differential mRNA Expression Profiles For Early And Advanced**  
5 **Stage Sporadic Colorectal Adenocarcinomas In A Malaysian Patient Cohort**  
6  
7  
8

9  
10  
11 Tze Pheng Lau<sup>1</sup>, April Camilla Roslani<sup>2</sup>, Lay Hoong Lian<sup>1</sup>, Ping Chin Lee<sup>3</sup>, Ida Hilmi<sup>4</sup>,  
12  
13 Khean Lee Goh<sup>4</sup>, Kek Heng Chua<sup>1\*</sup>  
14  
15

16  
17  
18  
19 <sup>1</sup>*Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala*  
20 *Lumpur, Malaysia.*  
21

22 <sup>2</sup>*Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur,*  
23 *Malaysia.*  
24

25 <sup>3</sup>*School of Science and Technology, Universiti Sabah Malaysia, 88400 Kota Kinabalu, Sabah,*  
26 *Malaysia.*  
27

28 <sup>4</sup>*Division of Gastroenterology and Hepatology, Department of Medicine, Faculty of Medicine,*  
29 *University of Malaya, 50603 Kuala Lumpur, Malaysia.*  
30  
31

32  
33  
34  
35  
36 *\*Correspondence author: Prof. Chua Kek Heng, Department of Biomedical Science, Faculty*  
37 *of Medicine, University Malaya, 50603 Kuala Lumpur, Malaysia.*  
38

39  
40 *e-mail: khchua@um.edu.my*  
41

42  
43 *Phone no: 603-79676607*  
44

45  
46 *Fax no: 603-79676600*  
47

48 **Key words:** Gene expression, non-hereditary, colorectal cancer, Malaysian  
49

50 **Word count:** 3056 words  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## ABSTRACT

**Objectives:** To characterise the mRNA expression patterns of early- and advanced stage colorectal adenocarcinomas of Malaysian patients.

**Design:** Comparative expression analysis.

**Setting and participants:** We performed a combination of ACP-based PCR and RT-qPCR for the identification of differentially expressed genes (DEGs) associated with early- and advanced stage primary colorectal tumours. We recruited four paired samples from CRC patients of Dukes' A and B for the preliminary differential expression study, and a total of 27 paired samples, ranging from CRC Stages I – IV, for subsequent confirmatory test. The tumouric samples were obtained from CRC patients undergoing curative surgical resection without pre-operative chemoradiotherapy. The recruited CRC patients were newly-diagnosed with CRC, and were not associated with any hereditary syndromes, previously diagnosed cancer, or positive family history of CRC. The paired non-cancerous tissue specimens were excised from macroscopically normal colonic mucosa distally located from the colorectal tumours.

**Primary and secondary outcome measures:** The differential mRNA expression patterns of early- and advanced stage colorectal adenocarcinomas compared to macroscopically normal colonic mucosa were characterised by ACP-based PCR and RT-qPCR.

**Results:** The *RPL35*, *RPS23* and *TIMP1* genes were found to be over-expressed in both early- and advanced stage colorectal adenocarcinomas ( $p < 0.05$ ). On the other hand, the *ARPC2* gene was significantly under-expressed in early colorectal adenocarcinomas, while the advanced stage primary colorectal tumours exhibited an additional over-expression of the *C6orf173* gene ( $p < 0.05$ ).

**Conclusions:** We characterised two distinctive gene expression patterns to aid in the stratification of primary colorectal neoplasms among Malaysian CRC patients. Further work can be done to assess and compare the mRNA expression levels of these identified DEGs between each CRC stage group, Stages I – IV.

## ARTICLE SUMMARY

### Article focus

- The latest staging system of colorectal tumours, which relies mainly on the clinicopathological assessment of primary tumours, is insufficient to address the complexity and heterogeneity of this disease. Hence, a non-anatomical, molecular-oriented staging system which can effectively predict the patients' outcome(s) and direct targeted treatment to different subgroups of CRC patients, is needed.
- Despite continuous revisions in the CRC classification criteria and expansion in transcriptomic and proteomic studies, there is yet to be any molecular marker(s) incorporated for clinical purposes.
- Article focus of this study is to characterise mRNA expression patterns of early- and advanced stage colorectal tumours of Malaysian patients.

### Key messages

- The under-expression of *ARPC2* and over-expression of *C6orf173* gene were distinctive for the early- and advanced stage sporadic colorectal adenocarcinomas, respectively.

### Strengths and limitations of this study

This regional-based study has a relatively small sample size due to the strict sample recruitment criteria where all subjects were newly-diagnosed with CRC, and were not associated with any hereditary syndromes, previously diagnosed cancer, or positive family history of CRC. However, the findings of this study are still reliable in view of our stringent sample selection criteria, high specificity primers and probes, as well as reliable statistical analysis.

**Abbreviations:** CRC: colorectal cancer; TNM: tumour-node-metastasis; AJCC: American Joint Committee on Cancer; SAGE: serial analysis of gene expression; CEA: carcinoembryonic antigen; ACP: Annealing Control Primer; RT-qPCR:

reverse transcription-quantitative real-time PCR; RIN: RNA integrity number; DEG: differentially expressed gene;  $\Delta\Delta C_T$ : comparative  $C_T$ ; MAD: median absolute deviation

## INTRODUCTION

Cancer staging is vital for patient management, especially in prognosis prediction and planning of treatment intervention[1]. This is especially in the CRC staging system. As such, there have been many noteworthy improvements since the introduction of the classical Dukes' staging system, followed by the modified Astler-Coller staging system; to the latest 7<sup>th</sup> edition of TNM staging system published by the AJCC[2-4]. The TNM staging system allows the incorporation of various clinical information (which are obtained through histopathological examination, radiologic imaging and surgical findings), for accurate CRC stratification[5]. However, these clinical assessments are greatly dependent on the expertise of pathologists, radiologists and clinicians.

The TNM classification is applicable for both clinical (cTNM) and pathological (pTNM) staging of primary colorectal tumours. Typically, it involves the assessment on the depth of bowel wall invasion at the time of diagnosis and the presence of regional lymph nodes metastases, as well as the presence of distant organ metastasis[4]. As a potentially worse patient outcome with more advanced disease stage is the core concept in cancer staging, AJCC revises the TNM classification system every few years with an attempt to formulate it for more accurate patient prognostication[5]. The latest 7<sup>th</sup> edition has further detailed the subclassification of the pN category and the assessment of discontinuous/satellite tumour foci. However, these revisions have increased the complexity and subjectivity during evaluation, and thus might lead to inter-observer variability and hamper its efficiency in routine clinical practise[5,6]. In addition, current clinicopathological parameters are insufficient to address the great biologic and genetic heterogeneity of CRC in patients' outcome and treatment response prediction. From the perspective of clinical oncology, the integration of molecular

1  
2  
3  
4 biomarkers into existing clinicopathological assessment will further refine the cancer  
5 management in future.  
6  
7

8  
9 Over the past decades, many researchers have attempted to establish gene expression  
10 signatures specifically for the diagnosis, prognostication and recurrence prediction of sporadic  
11 CRC. Transcriptional profiling promises a fairly dynamic view on the cellular functions,  
12 regulatory mechanisms and biochemical pathways involved in the disease pathogenesis and  
13 progression[7]. Various gene expression profiling techniques ranging from differential  
14 display, SAGE to microarrays have been utilised. Despite its wide application in gene  
15 expression profiling, microarray experiments have been subjected to various sources of  
16 variability, false-positives, as well as statistical and bioinformatic challenges. To date, none  
17 of the molecular markers described has been validated and employed in routine clinical  
18 practise owing to the poor reproducibility of the identified differentially expressed genes  
19 (DEGs) between different profiling platforms[8]. Although the KRAS mutation and  
20 mismatch repair status have showed promising prognostic and predictive values, they have yet  
21 to be incorporated into either routine pathological reporting systems or TNM staging  
22 systems[5].  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

34 Since most of the molecular studies on CRC were based in Western populations and different  
35 molecular changes were thought to underlie the development of sporadic CRC in populations  
36 with different genetic backgrounds, we aimed to investigate the changes in mRNA expression  
37 patterns in primary sporadic colorectal tumours with regards to our Malaysian patients. In our  
38 study, we have employed a combined approach of a two-step ACP-based PCR and real-time  
39 reverse transcription PCR to characterise the gene expression patterns for both early- and  
40 advanced stage sporadic colorectal adenocarcinomas.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## MATERIALS AND METHODS

### Patient selection and specimen collection

All patients presented with histologically confirmed colorectal adenocarcinomas and were staged accordingly to the AJCC TNM staging system (Table 1). The staging of cancer was performed by taking into consideration their histopathological reports, computed tomography images, morphological evaluations during surgery and serum CEA levels. Initially, four CRC patients of Stages I - III were recruited for the preliminary ACP-based PCR analysis, while another 27 patients with CRC Stages I – IV were recruited for subsequent RT-qPCR analysis. All subjects were newly-diagnosed with CRC, and were not associated with any hereditary syndromes, previously diagnosed cancer, or positive family history of CRC.

The subjects were admitted to the University Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia, and underwent curative surgical resection between 2010 and 2011. None had received pre-operative chemoradiotherapy. The study protocol was approved by the Ethics Committee Board of UMMC (Ref. No.: 654.1), and written informed consent was obtained from all study subjects. Both colorectal tumour and paired non-cancerous tissue specimens were immersed in *RNAlater* RNA Stabilization Reagent (Qiagen) immediately after excision and stored at  $-80^{\circ}\text{C}$ .



Table 1 Cancer staging of recruited subjects.

Subject	Cancer Stage
T1	Stage I / pT1N0M0
T2	Stage II / pT3N0M0
T3	Stage II / pT2N0M0
T4	Stage II / pT3N0M0
T5	Stage II / pT3N0M0
T6	Stage II / pT4N0M0
T7	Stage II / pT4N0M0
T8	Stage II / pT4N0M0
T9	Stage II / pT3N0M0
T10	Stage II / pT3N0M0
T11	Stage IV / pT3N2M1
T12	Stage IV
T13	Stage III / pT3N1M0
T14	Stage IV / pT3N1M1
T15	Stage III / pT3N1M0
T16	Stage III / pT3N2M0
T17	Stage IV / pT4N1M1
T18	Stage III / pT3N1M0
T19	Stage IV / pT3N0M1
T20	Stage III / pT4N1M0
T21	Stage III
T22	Stage II
T23	Stage III / pT3N1M0
T24	Stage II / pT3-4N0M0
T25	Stage IV / pT4N1M1
T26	Stage II / pT3N0M0
T27	Stage III / pT3N1M0

## Total RNA extraction

Total RNA was extracted from homogenised colonic tissues with the RNeasy Plus Mini Kit (Qiagen) according to manufacturer's instructions. Subsequently, the RNA yield and integrity were ascertained via Agilent 2100 Bioanalyzer in conjunction with Agilent RNA 6000 Nano Kits (Agilent Technologies). The values of RIN were then determined in order to assess the integrity of the isolated total RNA. In this study, only RNA samples with RIN values of 8.0 – 10.0 and rRNA ratios [28S/18S] of 1.5 – 2.5 were selected for successive applications.

## ACP-based PCR analysis

### a) First-strand cDNA synthesis

The synthesis of first-strand cDNA was performed according to the manufacturer's protocol for the GeneFishing DEG Premix Kit (Seegene), as follows: 3 µg of total RNA was added with 2 µl of 10 µM dT-ACP1 (5'-CTGTGAATGCTGCGACTACGA TXXXXX(T)<sub>18</sub>-3') and RNase-free water to a final volume of 9.5 µl. The mixture was then incubated at 80 °C for 3 min, followed by chilling on ice for another 2 min. Subsequently, 4 µl of 5X RT buffer (Mbiotech), 5 µl of 2mM dNTP (Fermentas), 0.5 µl of 40 U/µl RNase inhibitor (Mbiotech) and 1 µl of 200 U/µl M-MLV reverse transcriptase (Mbiotech) were added. This mixture was then incubated at 42 °C for 90 min, heated at 94 °C for another 2 min and chilled on ice for 2 min. Finally, 80 µl of DNase-free water was added to dilute the synthesised cDNA. The first-strand cDNA was stored under -20 °C until further analysis.

### b) ACP-based GeneFishing PCR

First, all four cDNA samples within each CRC and control group samples were pooled together in equal amounts. The characterisation of DEGs was then conducted via ACP-based PCR based on 20 arbitrary ACP primers (Cat. No.: K1021) in a thermal cycler (Mastercycler Gradient, Eppendorf) according to the manufacturer's protocol (GeneFishing DEG Premix Kit, Seegene). Initially, the synthesis of second-strand cDNA was commenced in a one-cycle first-stage PCR: 94 °C for 5 min, 50 °C for 3 min and 72 °C for 1 min. Next, the constructed second-strand cDNA was subjected to second-stage PCR with 40 cycles of a denaturing step

1  
2  
3  
4 at 94 °C for 40 sec, annealing step at 65 °C for 40 sec and extension step at 72 °C for 40 sec.  
5  
6 Lastly, a final extension step at 72 °C for 5 min was carried out. The amplified products were  
7  
8 then separated on 3 % (w/v) agarose gels stained with ethidium bromide.  
9

#### 10 11 c) Cloning and sequencing

12  
13 The identified differentially expressed bands were extracted from the agarose gel by using the  
14  
15 PureLink Quick Gel Extraction Kit (Invitrogen). Each of these extracted DNA fragments was  
16  
17 then individually cloned with the use of the TOPO TA Cloning Kit for Sequencing  
18  
19 (Invitrogen). Subsequently, the plasmid containing the inserted DNA fragment was extracted  
20  
21 from clones of interest via PureLink Quick Plasmid Miniprep Kit (Invitrogen). The isolated  
22  
23 cloned plasmids were then sequenced with the ABI 3730xl DNA Analyzer (Applied  
24  
25 Biosystems). Finally, all the sequences obtained were analysed and matched for similarities  
26  
27 with reference to the BLAST programme under the NCBI database.

#### 28 29 **RT-qPCR analysis**

##### 30 31 a) Reverse transcription

32  
33 The total RNA isolated from 27 paired samples was reverse transcribed to first-strand cDNA,  
34  
35 with the following protocol: 3 µg of total RNA was added with 2 µl of 0.5 µg/µl oligo(dT)<sub>12-18</sub>  
36  
37 (Invitrogen) and RNase-free water to a final volume of 9.5 µl. The reaction mixture was then  
38  
39 incubated at 80°C for 3 min, followed by chilling on ice for another 2 min. Next, 4 µl of 5X  
40  
41 first strand buffer (Invitrogen), 5 µl of 2mM dNTP (Fermentas), 0.5 µl of 40 U/µl RNaseOUT  
42  
43 recombinant RNase inhibitor (Invitrogen) and 1 µl of 200 U/µl M-MLV reverse transcriptase  
44  
45 (Invitrogen) were added to the mixture. Finally, the reaction mixture was incubated at 42 °C  
46  
47 for 90 min, heated at 94 °C for another 2 min and chilled on ice for 2 min. The synthesised  
48  
49 first-strand cDNA was stored under -20 °C until further usage.

##### 50 51 b) $\Delta\Delta C_T$ analysis

52  
53 The relative expression of identified DEGs in all paired colorectal tumours and control  
54  
55 samples was determined via  $\Delta\Delta C_T$  method. The RT-qPCR was performed in a singleplex  
56  
57

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

reaction containing 50 ng first-strand cDNA under universal thermal cycling conditions with the ABI 7500 Fast Real-Time PCR System (Applied Biosystems). Both *ACTB* (Assay ID: Hs99999903\_m1) and *GAPDH* (Assay ID: Hs99999905\_m1) were used as reference genes and are commercially available as TaqMan Pre-designed Assays (Applied Biosystems). Prior to the analysis of gene expression, the amplification efficiency for all target and reference genes assays was measured by using the standard curve method with 2-log measurements. The amplification efficiency value of 90 – 110 % was acceptable (Applied Biosystems). In this relative quantification method, the  $2^{-\Delta\Delta C_t}$  values obtained represented the fold change in gene expression of the colorectal tumours, which was normalised with both reference genes, in relative to the calibrator (control sample)[9].

### c) Statistical analysis

The difference in the expression level between colorectal tumour and paired non-cancerous tissues was analysed by using Real-Time StatMiner software (Integromics). The paired t-test was then performed to assess the statistical significance of the observed differential expression patterns.

## RESULTS

### DEGs between colorectal tumours and non-cancerous colonic tissues

This preliminary study was conducted on paired samples pooled from four patients with CRC Stages I - III. In ACP-based GeneFishing PCR, 20 sets of arbitrary ACP primers were used to randomly amplify gene products in both colorectal tumours and normal colonic samples. Upon visualisation on agarose gels, a total of 13 differentially expressed bands were observed by means of comparing bands intensity between the tumouric and non-cancerous samples, as shown in Figure 1. These bands were further sequenced for gene identification, and 16 DEGs were successfully reported. Among them, 13 were over-expressed in colorectal tumours, whilst three were under-expressed, as listed in Table 2.

Table 2 Sequence similarities and identification of DEGs.

Differentially Expressed Band	DEG	Identity	Sequence Homology (%)	Accession Number	UniGene Number	Description
<i>Over-expressed</i>						
A4.1	DEG1	<i>Homo sapiens</i> proteasome (prosome, macropain) 26S subunit, ATPase, 5 (PSMC5), mRNA	502/506 (99%)	NM_002805.4	Hs.79387	Involves in the ATP-dependent degradation of ubiquitinated proteins.
	DEG2	<i>Homo sapiens</i> ubiquinol-cytochrome c reductase hinge protein (UQCRH), mRNA	514/521 (98%)	NM_006004.2	Hs.481571	A component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex, which is part of the mitochondrial respiratory chain.
A4.2	DEG3	<i>Homo sapiens</i> ribosomal protein S23 (RPS23), mRNA	551/551 (100%)	NM_001025.4	Hs.527193	A component of the 40S subunit of human ribosomes.
A6.1	DEG4	<i>Homo sapiens</i> ribosomal protein L10 (RPL10), transcript variant 1, mRNA	554/557 (99%)	NM_006013.3	Hs.534404	A component of the 60S subunit of human ribosomes.
A9.2	DEG6	<i>Homo sapiens</i> actin related protein 2/3 complex, subunit 2, 34kDa (ARPC2), transcript variant 2, mRNA	473/473 (100%)	NM_005731.2	Hs.529303	Involves in the regulation of actin polymerization as an actin-binding component of the Arp2/3 complex, and mediates the formation of branched actin networks together with an activating nucleation-promoting factor (NPF).
	DEG7	<i>Homo sapiens</i> TIMP metalloproteinase inhibitor 1 (TIMP1), mRNA	503/511 (98%)	NM_003254.2	Hs.522632	Irreversibly inactivates the metalloproteinases by binding to their catalytic zinc cofactor.
A10.1	DEG8	<i>Homo sapiens</i> ATP synthase, H <sup>+</sup> -transporting, mitochondrial F1 complex, beta polypeptide (ATP5B), nuclear gene encoding mitochondrial protein, mRNA	917/919 (99%)	NM_001686.3	Hs.406510	A subunit of mitochondrial ATP synthase that catalyzes the synthesis of ATP by utilizing an electrochemical gradient of protons across the inner membrane during oxidative

						phosphorylation.
A13.2	DEG11	<i>Homo sapiens</i> chromosome 11 open reading frame 10 (C11orf10), mRNA	273/273 (100%)	NM_014206.3	Hs.437779	Unknown.
A13.3	DEG12	<i>Homo sapiens</i> mitochondrial ribosomal protein L24 (MRPL24), nuclear gene encoding mitochondrial protein, transcript variant 2, mRNA	408/411 (99%)	NM_024540.3	Hs.418233	Involves in protein synthesis within the mitochondrion.
A13.4	DEG13	<i>Homo sapiens</i> similar to OK/SW-CL.16 (LOC100288418)	635/644 (98%)	XM_002342023.1	-	Unknown.
A18.1	DEG14	<i>Homo sapiens</i> family with sequence similarity 96, member B (FAM96B), transcript variant 2, transcribed RNA	486/487 (99%)	NR_024525.1	Hs.9825	Involves in chromosome segregation as part of the mitotic spindle-associated MMXD complex.
A20.1	DEG15	<i>Homo sapiens</i> ribosomal protein L35 (RPL35), mRNA	440/446 (99%)	NM_007209.3	Hs.182825	A component of the 60S subunit of human ribosomes.
A20.2	DEG16	<i>Homo sapiens</i> chromosome 6 open reading frame173 (C6orf173), mRNA	551/554 (99%)	NM_001012507.2	Hs.486401	May be required for proper chromosome segregation during mitosis and involved with CENPT in the establishment of centromere chromatin structure.
<b><i>Under-expressed</i></b>						
A9.1	DEG5	<i>Homo sapiens</i> ribosomal protein L37 (RPL37), mRNA	284/284 (100%)	NM_000997.4	Hs.731513	A component of the 60S subunit of human ribosomes, and can bind to the 23S rRNA.
A13.1	DEG9	<i>Homo sapiens</i> solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1 (SLC25A1), nuclear gene encoding mitochondrial protein, mRNA	165/165 (100%)	NM_005984.2	Hs.111024	A mitochondrial tricarboxylate transporter which is responsible for the movement of citrate across the mitochondrial inner membrane.
	DEG10	<i>Homo sapiens</i> similar to cytochrome c oxidase subunit II (LOC100288578), miscRNA	141/146 (97%)	XR_078216.1	-	Unknown.

### Differential ability of the identified DEGs on early and advanced colorectal neoplasia

Following the identification of DEGs, the gene sequences obtained were then used to design primers and TaqMan probes for RT-qPCR analysis by Applied Biosystems, as listed in Table 3. In an attempt to assess the differential ability of identified DEGs on early and advanced colorectal adenocarcinoma, the recruited paired samples were further stratified into two groups according to the cancer stage. Among them, 13 patients with Stages I and II were grouped as early stage CRC, whilst the advanced stage CRC group comprised of 14 patients with Stages III and IV.

Table 3 Primers and TaqMan probes for relative quantification with Comparative C<sub>T</sub> method.

DEG	Primers Sequence		TaqMan Probe Sequence
DEG1	Forward:	5'-GGGCGTGTGCACAGAAG-3'	5'-CTCGCAGGGCATAACAT-3'
	Reverse:	5'-AAGTCCTCCTGAGTGACATGGA-3'	
DEG2	Forward:	5'-GATGCTTACCGAATCCGGAGATC-3'	5'-CCTCTTCTCTCCTCCTCC-3'
	Reverse:	5'-GCATTGCTCTCTCACTGTTGTTAG-3'	
DEG3	Forward:	5'-CAACCGTCATTGGGTACAAAGG-3'	5'-ATGGCAAGAAAATCAC-3'
	Reverse:	5'-TGTAAGGGTCCAGCTGATCAAGA-3'	
DEG4	Forward:	5'-CGGCCAGGAACTTGAACCTG-3'	5'-CAGGGCCTCAATCACA-3'
	Reverse:	5'-CCGAGCTGCAGAACAAGGA-3'	
DEG5	Forward:	5'-CTGGTTCGAATGAGGCACCTAAAA-3'	5'-CATGCCTGAATCTGC-3'
	Reverse:	5'-TGGGTTTAGGTGTTTCCTCAC-3'	
DEG6	Forward:	5'-AGATTAGCGGGATGAAAACGTCTT-3'	5'-CCCCGTGATTGTTTTTC-3'
	Reverse:	5'-CGCCCAGATGCCGAGAAAA-3'	
DEG7	Forward:	5'-GGTAGTGATGTGCAAGAGTCCAT-3'	5'-CATTGCTGGAAAACCTG-3'
	Reverse:	5'-CCGCAGCGAGGAGTTTCT-3'	
DEG8	Forward:	5'-GAAGGAGACCATCAAAGGATTCCA-3'	5'-ATTCACCTGCCAAAATC-3'
	Reverse:	5'-GAAGGCCTGTTCTGGGAGATG-3'	
DEG9	Forward:	5'-GGCAGGGTGGTCCTGAGA-3'	5'-CCTCTCTCCGCCCCGGACA-3'
	Reverse:	5'-CCGCCATTGGCCTTAACTG-3'	
DEG11	Forward:	5'-CAGGTTTCAGTGAAGCCATCTG-3'	5'-CACCCAAGGGTAACAAC-3'
	Reverse:	5'-GGGTTGGCATCTACGTGTGA-3'	
DEG12	Forward:	5'-CCAGGTCAAACCTGTGGATCCT-3'	5'-ATGGACAGGAAACCCAC-3'
	Reverse:	5'-GCTTCAGTAAATCTCCACTCGATCT-3'	
DEG14	Forward:	5'-CCCGCTCCTTATCTGCAAGTT-3'	5'-CATGCAGTGAACAAGC-3'
	Reverse:	5'-TCAAGATGGACGTGCACATTACTC-3'	
DEG15	Forward:	5'-CGGCCTCCAAGCTCTCT-3'	5'-CCGGACGACTCGGATCT-3'
	Reverse:	5'-TGAGAACACGGGCAATGGATTT-3'	
DEG16	Forward:	5'-GGACTCTTCTGCTAATCGATGAACA-3'	5'-CAGATGGACCAATAAGTCA-3'
	Reverse:	5'-GCCTCAACTTCGTCTGGAGAAAA-3'	

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

The analysis of RT-qPCR results was performed via Real-Time StatMiner software by importing the raw Ct data. The within-group correlation then determined by calculating the MAD for all the samples within the same experimental group. The biological samples which do not correlate well with other samples in the same group, were detected as group outliers and excluded from subsequent analysis. Both *ACTB* and *GAPDH* were used for normalisation in computing the  $\Delta C_T$  (Figure 2) and  $2^{-\Delta\Delta C_T}$  values by using the following formulas (Table 4).

$$C_{T(\text{Target gene})} - C_{T(\text{Reference gene})} = \Delta C_T$$

$$\Delta C_{T(\text{Sample})} - \Delta C_{T(\text{Calibrator})} = \Delta\Delta C_T$$

$$\text{Relative Fold Change in Expression (RQ)} = 2^{-\Delta\Delta C_T}$$

The relative fold change in the mRNA expression level between the colorectal tumours and adjacent normal colonic mucosa were shown as the  $2^{-\Delta\Delta C_T}$  values. The statistical significance of the observed fold change in expression was determined by paired t-test for all the DEGs. A *p* value of less than 0.05 is considered as statistically significant (Table 4).

In both early and advanced stage CRC groups, the expression of four out of 16 DEGs was reported to be significantly differed between tumouric and non-cancerous tissues. Remarkably, the combination of this panel of four genes is different among two groups. The *RPL35*, *RPS23* and *TIMP1* genes were found to be over-expressed in both early- and advanced colorectal neoplasms ( $p < 0.05$ ) (Figures 3 and 4). It is interesting to note that, the under-expression of *ARPC2* gene ( $p < 0.05$ ) was only observed in early stage colorectal tumours (Figure 3). On the other hand, the *C6orf173* gene was found to be over-expressed ( $p < 0.05$ ) in advanced colorectal adenocarcinomas, but not in early stage colorectal tumours (Figure 4).



Table 4  $\Delta C_T$  mean,  $\Delta\Delta C_T$ ,  $2^{-\Delta\Delta C_T}$  and  $p$  values for all the DEGs in both early- and advanced stage CRC groups.

DEG	Early Stage CRC					Advanced Stage CRC				
	$\Delta C_T$ Mean (CRC)	$\Delta C_T$ Mean (Normal)	$\Delta\Delta C_T$	$2^{-\Delta\Delta C_T}$	$p$ value	$\Delta C_T$ Mean (CRC)	$\Delta C_T$ Mean (Normal)	$\Delta\Delta C_T$	$2^{-\Delta\Delta C_T}$	$p$ value
<i>ARPC2</i>	2.6854	2.0664	0.6190	0.6511	0.0282*	2.7240	2.3300	0.3940	0.7610	0.2424
<i>ATP5B</i>	1.5846	1.2702	0.3144	0.8042	0.3524	1.9558	1.3838	0.5720	0.6727	0.1484
<i>C11orf10</i>	3.2897	3.3639	-0.0742	1.0528	0.8333	3.3281	3.6709	-0.3428	1.2682	0.3710
<i>C6orf173</i>	6.1083	7.1943	-1.0860	2.1228	0.0905	5.9949	7.9087	-1.9138	3.7680	0.0013*
<i>FAM96B</i>	3.5602	3.8955	-0.3353	1.2616	0.2935	3.5276	3.9920	-0.4644	1.3797	0.2113
<i>MRPL24</i>	4.9171	5.0839	-0.1668	1.1226	0.3564	4.9728	5.1467	-0.1739	1.1281	0.7001
<i>PSMC5</i>	3.8232	3.9617	-0.1385	1.1008	0.6812	3.7705	3.8455	-0.0750	1.0534	0.8048
<i>RPL10</i>	-0.7462	-0.4853	-0.2609	1.1982	0.4001	-1.1576	-0.5196	-0.6380	1.5562	0.0950
<i>RPL35</i>	-0.1926	0.6222	-0.8148	1.7591	0.0024*	0.1748	0.8769	-0.7021	1.6269	0.0372*
<i>RPL37</i>	-0.0059	-0.1539	0.1480	0.9025	0.8645	0.2184	0.7143	-0.4959	1.4102	0.1537
<i>RPS23</i>	0.2176	0.7739	-0.5563	1.4705	0.0310*	0.0676	0.9431	-0.8755	1.8346	0.0250*
<i>SLC25A1</i>	3.7514	3.5430	0.2084	0.8655	0.5721	3.5565	3.4428	0.1137	0.9242	0.7991
<i>TIMP1</i>	2.9096	4.3059	-1.3963	2.6323	0.0440*	2.3330	3.8547	-1.5217	2.8713	0.0062*
<i>UQCRH</i>	2.0087	2.2216	-0.2129	1.1590	0.4108	2.3375	2.4459	-0.1084	1.0780	0.7808

\* $p < 0.05$  = statistically significant

## DISCUSSION

Our current study has revealed two distinctive 4-gene signatures for both early- and advanced stage colorectal adenocarcinomas. The early stage sporadic CRC was characterised by the over-expression of *RPL35*, *RPS23* and *TIMP1* genes, as well as under-expression of *ARPC2* gene. On the other hand, the advanced primary colorectal tumours were reported with over-expression of *C6orf173*, *RPL35*, *RPS23* and *TIMP1* genes. Although the relative fold change for *ARPC2*, *RPL35* and *RPS23* genes is below 2, the individual result does not affect the analysis since gene expression patterns of all four genes in combination were proposed to distinguish between the early- and advanced stage colorectal neoplasms. The potential involvement of these DEGs and their altered expression levels in CRC were further supported by previous researches.

In fact, several proto-oncogenes and tumour suppressors are previously reported to regulate the ribosome production, i.e., the *RB*[10], *TP53*[11], *PTEN* genes[12], as well as the *MYC* gene family[13]. It is suggested that the alterations in ribosome biogenesis might affect the translation of genes that are involved in neoplastic transformation. In addition, the additional extra-ribosomal functions of the ribosomal proteins (r-proteins) in cellular apoptosis, cellular proliferation, cellular transformation, genes transcription, mRNA translation, DNA repair and inflammation, might also trigger and support the neoplastic development[14]. Hence, the over-expression of r-proteins-encoding genes observed in colorectal adenocarcinomas is not unexpected[15-17]. Our current study has revealed the significant over-expression of two r-proteins that were not previously described in colorectal tumours, i.e., the *RPL35* and *RPS23*. The observed fold changes for the *RPL35* and *RPS23* mRNA levels were comparable between the early- and advanced stage colorectal tumours in our sample cohort. This was in agreement with previous reports by Barnard et al. and Frigerio et al., where the changes in the mRNA expression levels of the r-proteins were irrespective of the cancer stage[18,19]. The hypothesis that the same ribosomal protein may contribute in different stages of cancer progression with their hitherto unknown extra-ribosomal roles might provide an explanation to these observations[20].

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

On the other hand, our present study also demonstrated an over-expression of the *TIMP1* gene in both early- and advanced stage primary colorectal tumours. This finding is supported by Zeng et al., where the over-expression of TIMP1 was reported in all stages of primary colorectal tumours[21]. Under normal physiological conditions, the proteolytic activities of MMPs are kept at bay by their natural inhibitors, the TIMPs[22]. Previous studies have reported the over-expression of MMPs in both early- and advanced stage colorectal tumours, as well as other cancer types[23-25], which is in accordance to their biological roles. Hence, a similar scenario is expected for TIMPs and indeed, their suppressive role in tumour invasion and metastasis has been demonstrated in various cancer models[26]. However, more recent studies have revealed a direct correlation between TIMP1 expression and tumour aggressiveness in cancer, including CRC[21,27]. These findings, which are contradictory to its protease-inhibiting function, have suggested a possible tumour-promoting role of TIMP1 in tumorigenesis. It is postulated that the TIMP1 exhibits the abilities to inhibit tumour cell apoptosis and promote tumour angiogenesis, as well as other growth-factor-like effects[28]. In our present study, the observed comparable over-expression of TIMP1 in both early- and advanced stage sporadic colorectal neoplasms was in line with its MMP inhibitory and MMP-independent tumour-promoting activities.

In cancer biology, the expression of mRNAs and proteins of the ARP2/3 complex is often studied due to its role in cell migration, which contributes to cancer invasion and metastasis if aberrantly regulated[29]. We have detected a significant under-expression of ARPC2 in our cohort of early stage primary colorectal tumours. Surprisingly, this finding is contradictory with the role played by ARPC2 in cancer invasion and metastasis theoretically. Previously, Kaneda et al. has reported the decreased expression of all the seven genes encoding the subunits of ARP2/3 complex in human gastric cancers. Among them, the Arp2, ARPC2 and ARPC3 showed the most prominent reduction in their expression levels[30]. The exact mechanism underlying this observation still remains unknown, but the epigenetic alteration might potentially provide an explanation for it. For instance, promoter hypermethylation that causes gene silencing is responsible for the reduced expression of ARPC1 in human gastric

1  
2  
3  
4 cancer[31]. Similarly, the epigenetic study might also offer a clue for the under-expression of  
5 ARPC2 in colorectal neoplasms.  
6  
7

8  
9 *C6orf173*, which is also known as *CUG2* or *CENP-W*, is a novel oncogene that has been  
10 found to be up-regulated in many human cancer tissues. Its high expression level is  
11 profoundly reported in tumours of the ovary, liver, lung, pancreas, breast, colon, rectum and  
12 stomach. The CENP-W is a new member of the constitutive centromere-associated network,  
13 which specifically interacts with the CENP-T and plays an important role in mitosis[32]. In  
14 our current study, the CENP-W is over-expressed in advanced colorectal adenocarcinoma.  
15 This finding correlates to its function in kinetochore assembly, where its aberrant expression  
16 might lead to abnormal cell division and aneuploidy in cancer[32]. In our study, the over-  
17 expression of CENP-W was observed in both early- and advanced cohort of colorectal  
18 neoplasms but only statistically significant in the latter group. Given the fact that aneuploidy  
19 is constantly associated with a greater proportion of advanced CRC cases, the aberrant  
20 expression of CENP-W might potentially relate to a poorer prognosis of CRC[33].  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

31 In conclusion, we have characterised two distinctive gene expression patterns, which comprise  
32 of the *ARPC2*, *C6orf173*, *RPL35*, *RPS23* and *TIMP1* genes, for the stratification of primary  
33 colorectal adenocarcinomas among Malaysian CRC patients. Our current sample size was  
34 relatively small owing to the lack of a designated Tissue Bank in our institution. There were  
35 also not many CRC patient volunteers. Moreover, our stringent criteria for patient selection  
36 have also limited the availability of suitable specimens within the short sample collection  
37 period. Nevertheless, our identified mRNA expression patterns specific for early- and  
38 advanced stage colorectal tumours are still convincing with our stringent sample selection  
39 criteria, high specificity primers and probes, as well as reliable statistical analysis. In future,  
40 the validation of these DEGs should be performed on a larger set of clinical samples, and  
41 extensive inter-laboratory testing of their differential abilities on each CRC stage is also  
42 desired. In addition, we should also integrate other imaging and histological information to  
43 complement our identified gene expression patterns, which then hold promises for better  
44 stratification of colorectal tumours.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 **Funding** – This study was supported by FS176/2007C, PS172/2008C and Research  
5 Collaborative Grant, CG041-2013 from the University of Malaya.  
6  
7  
8

9  
10 **Competing interests** – None.  
11

12  
13 **Data sharing statement** – There is no additional data available.  
14  
15

16  
17 **Contributorship statement** – All authors were responsible for the design and analysis of the  
18 study. All authors were involved in drafting the manuscript, providing the intellectual input  
19 and approving the final version of the manuscript.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**REFERENCES**

1. Greene FL, Page D, Fleming ID, et al., eds. AJCC cancer staging manual (6th ed.). New York: Springer 2002.
2. Dukes CE. The classification of cancer of the rectum. *J Pathol Bacteriol* 1932; 35: 323-32.
3. Astler VB, Coller FA. The prognostic significance of direct extension of carcinoma of the colon and rectum. *Ann Surg* 1954; 139: 846-52.
4. Edge SB, Byrd DR, Compton CC, et al. AJCC cancer staging manual (7th ed.). New York: Springer 2010.
5. Hu HK, Krasinskas A, Willis J. Perspectives on current tumour-node-metastasis (TNM) staging of cancers of the colon and rectum. *Semin Oncol* 2011; 38: 500-10.
6. Doyle VJ, Bateman AC. Colorectal cancer staging using TNM 7: is it time to use this new staging system? *J Clin Pathol* 2012; 65: 372-4.
7. Russo G, Zegar C, Giordano A. Advantages and limitations of microarray technology in human cancer. *Oncogene* 2003; 22: 6497-507.
8. Puppa G, Sonzogni A, Colombari R, et al. TNM staging system of colorectal carcinoma: a critical appraisal of challenging issues. *Arch Pathol Lab Med* 2010; 134: 837-52.

- 1  
2  
3  
4  
5 9. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time  
6  
7 quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 2001; 25: 402-8.  
8  
9
- 10  
11 10. Voit R, Schafer K, Grummt I. Mechanism of repression of RNA polymerase I  
12  
13 transcription by the retinoblastoma protein. *Mol Cell Biol* 1997; 17: 4230-7.  
14
- 15  
16 11. Zhai W, Cornai L. Repression of RNA polymerase I transcription by the tumour  
17  
18 suppressor p53. *Mol Cell Biol* 2000; 20: 5930-8.  
19  
20
- 21  
22 12. Backman S, Stambolic V, Mak T. PTEN function in mammalian cell size regulation.  
23  
24 *Curr Opin Neurobiol* 2002; 12: 516-22.  
25  
26
- 27  
28 13. Greasley PJ, Bonnard C, Amati B. Myc induces the nucleolin and BN51 genes:  
29  
30 possible implications in ribosome biogenesis. *Nucleic Acids Res* 2000; 28: 446-53.  
31  
32
- 33  
34 14. Montanaro L, Treré D, Derenzini M. Nucleolus, ribosomes, and cancer. *Am J Pathol*  
35  
36 2008; 173: 301-10.  
37  
38
- 39  
40 15. Sharp MG, Adams SM, Elvin P, et al. A sequence previously identified as metastasis-  
41  
42 related encodes an acidic ribosomal phosphoprotein, P2. *Br J Cancer* 1990; 61: 83-8.  
43  
44
- 45  
46 16. Chester KA, Robson L, Begent RH, et al. Identification of a human ribosomal protein  
47  
48 mRNA with increased expression in colorectal tumours. *Biochim Biophys Acta* 1989;  
49  
50 1009: 297-300.  
51  
52
- 53  
54 17. Pogue-Geile K, Geiser JR, Shu M, et al. Ribosomal protein genes are overexpressed in  
55  
56 colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein.  
57  
58 *Mol Cell Biol* 1991; 11: 3842-9.  
59  
60

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
18. Barnard GF, Staniunas RJ, Mori M, et al. Gastric and hepatocellular carcinomas do not overexpress the same ribosomal protein messenger RNAs as colonic carcinoma. *Cancer Res* 1993; 53: 4048-52.
19. Frigerio JM, Dagorn JC, Iovanna JL. Cloning, sequencing and expression of the L5, L21, L27a, L28, S5, S9, S10 and S29 human ribosomal protein mRNAs. *Biochim Biophys Acta* 1995; 1262: 64-8.
20. Lai MD, Xu J. Ribosomal proteins and colorectal cancer. *Curr Genomics* 2007; 8: 43-9.
21. Zeng ZS, Cohen AM, Zhang ZF, et al. Elevated tissue inhibitor of metalloproteinase 1 RNA in colorectal cancer stroma correlates with lymph node and distant metastases. *Clin Cancer Res* 1995; 1: 899-906.
22. Ennis BW, Matrisian LM. Matrix degrading metalloproteinases. *J Neurooncol* 1994; 18: 105-9.
23. Urbanski SJ, Edwards DR, Maitland A, et al. Expression of metalloproteinases and their inhibitors in primary pulmonary carcinomas. *Br J Cancer* 1992; 66: 1188-94.
24. Boag AH, Young ID. Immunohistochemical analysis of type IV collagenase expression in prostatic hyperplasia and adenocarcinoma. *Mod Pathol* 1993; 6: 65-8.
25. Newell KJ, Witty JP, Rodgers WH, et al. Expression and localisation of matrix-degrading metalloproteinases during colorectal tumourigenesis. *Mol Carcinogen* 1994; 10: 199-206.



- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
26. Khokha R, Waterhouse P. The role of tissue inhibitor of metalloproteinase-1 in specific aspects of cancer progression and reproduction. *J Neurooncol* 1994; 18: 123-7.
  27. Lu XQ, Levy M, Weinstein IB, et al. Immunological quantitation of levels of tissue inhibitor of metalloproteinase-1 in human colon cancer. *Cancer Res* 1991; 51: 6231-5.
  28. Egeblad M, Werb Z. New functions for the matrix metalloproteinases in cancer progression. *Nat Rev Cancer* 2002; 2: 161-74.
  29. Yamaguchi H, Wyckoff J, Condeelis J. Cell migration in tumours. *Curr Opin Cell Biol* 2005; 17: 559-64.
  30. Kaneda A, Kaminishi M, Sugimura T, et al. Decreased expression of the seven ARP2/3 complex genes in human gastric cancers. *Cancer Lett* 2004; 212: 203-10.
  31. Kaneda A, Kaminishi M, Nakanishi Y, et al. Reduced expression of the insulin-induced protein 1 and p41 ARP2/3 complex genes in human gastric cancers. *Int J Cancer* 2002; 100: 57-62.
  32. Hori T, Amano M, Suzuki A, et al. CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* 2008; 135: 1039-52.
  33. Chen HS, Sheen-Chen SM, Lu CC. DNA index and S-phase fraction in curative resection of colorectal adenocarcinoma: analysis of prognosis and current trends. *World J Surg* 2002; 26: 626-30

## Figure Legends

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- Figure 1 Differential banding patterns on 3 % agarose gel post ACP-based PCR amplification between normal colon and colorectal tumour samples (N: normal sample; C: CRC sample)
- Figure 2 Box-plots showing  $\Delta C_T$  values of all colorectal tumours and normal colonic tissues in each early- (a) and advanced (b) stage CRC group.
- Figure 3 Differential expression patterns of all the identified DEGs in early stage CRC group.
- Figure 4 Differential expression patterns of all the identified DEGs in advanced stage CRC group.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 | **Characterisation Of Differential mRNA Distinctive Gene Expression Profiles Signatures**  
11 | **For Early And Advanced Stage Sporadic Colorectal Adenocarcinomas In A Malaysian**  
12 | **Patient Cohort**  
13  
14

Formatted

15  
16  
17 5 Tze Pheng Lau<sup>1</sup>, April Camilla Roslani<sup>2</sup>, Lay Hoong Lian<sup>1</sup>, Ping Chin Lee<sup>3</sup>, Ida Hilmi<sup>4</sup>,  
18  
19 6 Khean Lee Goh<sup>4</sup>, Kek Heng Chua<sup>1\*</sup>  
20  
21 7

22  
23 8 <sup>1</sup>*Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala*  
24 *Lumpur, Malaysia.*

25  
26 10 <sup>2</sup>*Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur,*  
27 *Malaysia.*

28  
29 12 <sup>3</sup>*School of Science and Technology, Universiti Sabah Malaysia, 88400 Kota Kinabalu, Sabah,*  
30 *Malaysia.*

31  
32 14 <sup>4</sup>*Division of Gastroenterology and Hepatology, Department of Medicine, Faculty of Medicine,*  
33 *University of Malaya, 50603 Kuala Lumpur, Malaysia.*  
34  
35

36  
37 17 \*Correspondence author: Prof. Chua Kek Heng, Department of Biomedical Science, Faculty  
38 of Medicine, University Malaya, 50603 Kuala Lumpur, Malaysia.

39  
40 19 e-mail: khchua@um.edu.my

41  
42 20 Phone no: 603-79676607

43  
44 21 Fax no: 603-79676600  
45

46 22 **Key words:** Gene expression, non-hereditary, colorectal cancer, Malaysian

47  
48 23 **Word count:** 3056 words  
49  
50  
51

52 1  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1  
11  
12 2 **ABSTRACT**

13  
14 3 **Objectives:** To characterise the mRNA expression patterns of early- and advanced stage  
15 4 colorectal adenocarcinomas of Malaysian patients.

16  
17 5 **Design:** Comparative expression analysis.

18  
19 6 **Setting and participants:** We performed a combination of ACP-based PCR and RT-qPCR  
20 7 for the identification of differentially expressed genes (DEGs) associated with early- and  
21 8 advanced stage primary colorectalCRC tumours. We recruited four paired samples from CRC  
22 9 patients of Dukes' A and B for the preliminary differential expression study, and a total of 27  
23 10 paired samples, ranging from CRC Stages I – IV, for subsequent confirmatory test. The  
24 11 tumouric samples were obtained from CRC patients undergoing curative surgical resection  
25 12 without pre-operative chemoradiotherapy. The recruited CRC patients were newly-diagnosed  
26 13 with CRC, and were not associated with any hereditary syndromes, previously diagnosed  
27 14 cancer, or positive family history of CRC. The paired non-cancerous tissue specimens were  
28 15 excised from macroscopically normal colonic mucosa distally located from the colorectalCRC  
29 16 tumours.

30  
31 17 **Primary and secondary outcome measures:** The differential mRNA expression patterns of  
32 18 early- and advanced stage colorectal adenocarcinomas compared to macroscopically normal  
33 19 colonic mucosa were characterised by ACP-based PCR and RT-qPCR.

34  
35 20 **Results:** The *RPL35*, *RPS23* and *TIMP1* genes were found to be over-expressed in both early-  
36 21 and advanced stage colorectal adenocarcinomas ( $p < 0.05$ ). On the other hand, the *ARPC2*  
37 22 gene was significantly under-expressed in early colorectal adenocarcinomas, while the  
38 23 advanced stage primary colorectalCRC tumours exhibited an additional over-expression of the  
39 24 *C6orf173* gene ( $p < 0.05$ ).

40  
41 25 **Conclusions:** We characterised two distinctive gene expression patterns to aid in the  
42 26 stratification of primary colorectalCRC neoplasms among Malaysian CRC patients. Further

1 work can be done to assess and compare the mRNA expression levels of these identified  
2 DEGs between each CRC stage group, Stages I – IV.

### 3 ARTICLE SUMMARY

#### 4 Article focus

- 5 |     ▪ The latest staging system of colorectalCRC tumours, which relies mainly on the  
6     clinicopathological assessment of primary tumours, is insufficient to address the  
7     complexity and heterogeneity of this disease. Hence, a non-anatomical, molecular-  
8     oriented staging system which can effectively predict the patients' outcome(s) and  
9     direct targeted treatment to different subgroups of CRC patients, is needed.
- 10 |    ▪ Despite continuous revisions in the CRC classification criteria and expansion in  
11 |    transcriptomic and proteomic studies, there is yet to be any molecular marker(s)  
12 |    incorporated for clinical purposes.
- 13 |    ▪ Article focus of this study is to characterise mRNA expression patterns of early- and  
14 |    advanced stage colorectalCRC tumours of Malaysian patients.

#### 15 Key messages

- 16 |     ▪ The under-expression of *ARPC2* and over-expression of *C6orf173* gene were  
17 |     distinctive for the early- and advanced stage sporadic colorectal adenocarcinomas,  
18 |     respectively.

#### 19 Strengths and limitations of this study

20 This regional-based study has a relatively small sample size due to the strict sample  
21 recruitment criteria where all subjects were newly-diagnosed with CRC, and were not  
22 associated with any hereditary syndromes, previously diagnosed cancer, or positive family  
23 history of CRC. However, the findings of this study are still reliable in view of our stringent  
24 sample selection criteria, high specificity primers and probes, as well as reliable statistical  
25 analysis.

1 **Abbreviations:** CRC: colorectal cancer; TNM: tumour-node-metastasis; AJCC: American  
2 Joint Committee on Cancer; SAGE: serial analysis of gene expression; CEA:  
3 carcinoembryonic antigen; ACP: Annealing Control Primer; RT-qPCR:  
4 reverse transcription-quantitative real-time PCR; RIN: RNA integrity  
5 number; DEG: differentially expressed gene;  $\Delta\Delta C_T$ : comparative  $C_T$ ; MAD:  
6 median absolute deviation  
7

## 8 INTRODUCTION

9 Cancer staging is vital for patient management, especially in prognosis prediction and  
10 planning of treatment intervention[1]. This is especially in the CRC staging system. As such,  
11 there have been many noteworthy improvements since the introduction of the classical Dukes'  
12 staging system, followed by the modified Astler-Coller staging system; to the latest 7<sup>th</sup> edition  
13 of TNM staging system published by the AJCC[2-4]. The TNM staging system allows the  
14 incorporation of various clinical information (which are obtained through histopathological  
15 examination, radiologic imaging and surgical findings), for accurate CRC stratification[5].  
16 However, these clinical assessments are greatly dependent on the expertise of pathologists,  
17 radiologists and clinicians.

18 The TNM classification is applicable for both clinical (cTNM) and pathological (pTNM)  
19 staging of primary colorectalCRC tumours. Typically, it involves the assessment on the depth  
20 of bowel wall invasion at the time of diagnosis and the presence of regional lymph nodes  
21 metastases, as well as the presence of distant organ metastasis[4]. As a potentially worse  
22 patient outcome with more advanced disease stage is the core concept in cancer staging, AJCC  
23 revises the TNM classification system every few years with an attempt to formulate it for  
24 more accurate patient prognostication[5]. The latest 7<sup>th</sup> edition has further detailed the  
25 subclassification of the pN category and the assessment of discontinuous/satellite tumour foci.  
26 However, these revisions have increased the complexity and subjectivity during evaluation,  
27 and thus might lead to inter-observer variability and hamper its efficiency in routine clinical  
28 practise[5,6]. In addition, current clinicopathological parameters are insufficient to address

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 the great biologic and genetic heterogeneity of CRC in patients' outcome and treatment  
11 2 response prediction, and thus are inadequate for accurate individual prognostic prediction.  
12 3 From the perspective of clinical oncology, the integration of molecular biomarkers into  
13 4 existing clinicopathological assessment will further refine the cancer management in future, a  
14 5 new classification scheme based on molecular biomarkers would be relatively more efficient  
15 6 and accurate.

17  
18  
19 7 Over the past decades, many researchers have attempted to establish gene expression  
20 8 signatures specifically for the diagnosis, prognostication and recurrence prediction of sporadic  
21 9 CRC. Transcriptional profiling promises a fairly dynamic view on the cellular functions,  
22 10 regulatory mechanisms and biochemical pathways involved in the disease pathogenesis and  
23 11 progression[7]. Various gene expression profiling techniques ranging from differential  
24 12 display, SAGE to microarrays have been utilised. Despite its wide application in gene  
25 13 expression profiling, microarray experiments have been subjected to various sources of  
26 14 variability, false-positives, as well as statistical and bioinformatic challenges. To date, none  
27 15 of the molecular markers described has been validated and employed in routine clinical  
28 16 practise owing to the poor reproducibility of the identified differentially expressed genes  
29 17 (DEGs) between different profiling platforms[8]. Although the KRAS mutation and  
30 18 mismatch repair status have showed promising prognostic and predictive values, they have yet  
31 19 to be incorporated into either routine pathological reporting systems or TNM staging  
32 20 systems[5].

33  
34  
35  
36  
37  
38  
39 21 Since most of the molecular studies on CRC were based in Western populations and different  
40 22 molecular changes were thought to underlie the development of sporadic CRC in populations  
41 23 with different genetic backgrounds, we aimed to investigate the changes in mRNA expression  
42 24 patterns in primary sporadic colorectalCRC tumours with regards to our Malaysian patients.  
43 25 In our study, we have employed a combined approach of a two-step ACP-based PCR and real-  
44 26 time reverse transcription PCR to characterise the gene expression patterns for both early- and  
45 27 advanced stage sporadic colorectal adenocarcinomas.

## MATERIALS AND METHODS

### Patient selection and specimen collection

All patients presented with histologically confirmed colorectal adenocarcinomas and were staged accordingly to the AJCC TNM staging system (Table 1). The staging of cancer was performed by taking into consideration their histopathological reports, computed tomography images, morphological evaluations during surgery and serum CEA levels. Initially, four CRC patients of Stages I - III were recruited for the preliminary ACP-based PCR analysis, while another 27 patients with CRC Stages I – IV were recruited for subsequent RT-qPCR analysis. All subjects were newly-diagnosed with CRC, and were not associated with any hereditary syndromes, previously diagnosed cancer, or positive family history of CRC.

The subjects were admitted to the University Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia, and underwent curative surgical resection between 2010 and 2011. None had received pre-operative chemoradiotherapy. The study protocol was approved by the Ethics Committee Board of UMMC (Ref. No.: 654.1), and written informed consent was obtained from all study subjects. Both colorectal CRC tumour and paired non-cancerous tissue specimens were immersed in RNAlater RNA Stabilization Reagent (Qiagen) immediately after excision and stored at  $-80^{\circ}\text{C}$ .



1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1  
11 2  
12 3  
13 4  
14 5  
15 6  
16 7  
17  
18  
19  
20 Table 1 Cancer staging of recruited subjects.

Subject	Cancer Stage
T1	Stage I / pT1N0M0
T2	Stage II / pT3N0M0
T3	Stage II / pT2N0M0
T4	Stage II / pT3N0M0
T5	Stage II / pT3N0M0
T6	Stage II / pT4N0M0
T7	Stage II / pT4N0M0
T8	Stage II / pT4N0M0
T9	Stage II / pT3N0M0
T10	Stage II / pT3N0M0
T11	Stage IV / pT3N2M1
T12	Stage IV
T13	Stage III / pT3N1M0
T14	Stage IV / pT3N1M1
T15	Stage III / pT3N1M0
T16	Stage III / pT3N2M0
T17	Stage IV / pT4N1M1
T18	Stage III / pT3N1M0
T19	Stage IV / pT3N0M1
T20	Stage III / pT4N1M0
T21	Stage III

7

T22	Stage II
T23	Stage III / pT3N1M0
T24	Stage II / pT3-4N0M0
T25	Stage IV / pT4N1M1
T26	Stage II / pT3N0M0
T27	Stage III / pT3N1M0

## 2 Total RNA extraction

3 Total RNA was extracted from homogenised colonic tissues with the RNeasy Plus Mini Kit  
 4 (Qiagen) according to manufacturer's instructions. Subsequently, the RNA yield and integrity  
 5 were ascertained via Agilent 2100 Bioanalyzer in conjunction with Agilent RNA 6000 Nano  
 6 Kits (Agilent Technologies). The values of RIN were then determined in order to assess the  
 7 integrity of the isolated total RNA. In this study, only RNA samples with RIN values of 8.0 –  
 8 10.0 and rRNA ratios [28S/18S] of 1.5 – 2.5 were selected for successive applications.

## 9 ACP-based PCR analysis

### 10 a) First-strand cDNA synthesis

11 The synthesis of first-strand cDNA was performed according to the manufacturer's protocol  
 12 for the GeneFishing DEG Premix Kit (Seegene), as follows: 3 µg of total RNA was added  
 13 with 2 µl of 10 µM dT-ACP1 (5'-CTGTGAATGCTGCGACTACGA TXXXXX(T)<sub>18</sub>-3') and  
 14 RNase-free water to a final volume of 9.5 µl. The mixture was then incubated at 80 °C for 3  
 15 min, followed by chilling on ice for another 2 min. Subsequently, 4 µl of 5X RT buffer  
 16 (Mbiotech), 5 µl of 2mM dNTP (Fermentas), 0.5 µl of 40 U/µl RNase inhibitor (Mbiotech)  
 17 and 1 µl of 200 U/µl M-MLV reverse transcriptase (Mbiotech) were added. This mixture was  
 18 then incubated at 42 °C for 90 min, heated at 94 °C for another 2 min and chilled on ice for 2  
 19 min. Finally, 80 µl of DNase-free water was added to dilute the synthesised cDNA. The first-  
 20 strand cDNA was stored under -20 °C until further analysis.

### 21 b) ACP-based GeneFishing PCR

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 First, all four cDNA samples within each CRC and control group samples were pooled  
11 2 together in equal amounts. The characterisation of DEGs was then conducted via ACP-based  
12 3 PCR based on 20 arbitrary ACP primers (Cat. No.: K1021) in a thermal cycler (Mastercycler  
13 4 Gradient, Eppendorf) according to the manufacturer's protocol (GeneFishing DEG Premix Kit,  
14 5 Seegene). Initially, the synthesis of second-strand cDNA was commenced in a one-cycle first-  
15 6 stage PCR: 94 °C for 5 min, 50 °C for 3 min and 72 °C for 1 min. Next, the constructed  
16 7 second-strand cDNA was subjected to second-stage PCR with 40 cycles of a denaturing step  
17 8 at 94 °C for 40 sec, annealing step at 65 °C for 40 sec and extension step at 72 °C for 40 sec.  
18 9 Lastly, a final extension step at 72 °C for 5 min was carried out. The amplified products were  
19 10 then separated on 3 % (w/v) agarose gels stained with ethidium bromide.

#### 24 11 c) Cloning and sequencing

25 12 The identified differentially expressed bands were extracted from the agarose gel by using the  
26 13 PureLink Quick Gel Extraction Kit (Invitrogen). Each of these extracted DNA fragments was  
27 14 then individually cloned with the use of the TOPO TA Cloning Kit for Sequencing  
28 15 (Invitrogen). Subsequently, the plasmid containing the inserted DNA fragment was extracted  
29 16 from clones of interest via PureLink Quick Plasmid Miniprep Kit (Invitrogen). The isolated  
30 17 cloned plasmids were then sequenced with the ABI 3730xl DNA Analyzer (Applied  
31 18 Biosystems). Finally, all the sequences obtained were analysed and matched for similarities  
32 19 with reference to the BLAST programme under the NCBI database.

#### 33 20 RT-qPCR analysis

##### 34 21 a) Reverse transcription

35 22 The total RNA isolated from 27 paired samples was reverse transcribed to first-strand cDNA,  
36 23 with the following protocol: 3 µg of total RNA was added with 2 µl of 0.5 µg/µl oligo(dT)<sub>12-18</sub>  
37 24 (Invitrogen) and RNase-free water to a final volume of 9.5 µl. The reaction mixture was then  
38 25 incubated at 80 °C for 3 min, followed by chilling on ice for another 2 min. Next, 4 µl of 5X  
39 26 first strand buffer (Invitrogen), 5 µl of 2mM dNTP (Fermentas), 0.5 µl of 40 U/µl RNaseOUT  
40 27 recombinant RNase inhibitor (Invitrogen) and 1 µl of 200 U/µl M-MLV reverse transcriptase

1 (Invitrogen) were added to the mixture. Finally, the reaction mixture was incubated at 42 °C  
2 for 90 min, heated at 94 °C for another 2 min and chilled on ice for 2 min. The synthesised  
3 first-strand cDNA was stored under -20 °C until further usage.

#### 4 b) $\Delta\Delta C_T$ analysis

5 | The relative expression of identified DEGs in all paired colorectalCRC tumours and control  
6 samples was determined via  $\Delta\Delta C_T$  method. The RT-qPCR was performed in a singleplex  
7 reaction containing 50 ng first-strand cDNA under universal thermal cycling conditions with  
8 the ABI 7500 Fast Real-Time PCR System (Applied Biosystems). Both *ACTB* (Assay ID:  
9 Hs99999903\_m1) and *GAPDH* (Assay ID: Hs99999905\_m1) were used as reference genes  
10 and are commercially available as TaqMan Pre-designed Assays (Applied Biosystems). Prior  
11 to the analysis of gene expression, the amplification efficiency for all target and reference  
12 genes assays was measured by using the standard curve method with 2-log measurements.  
13 The amplification efficiency value of 90 – 110 % was acceptable (Applied Biosystems). In  
14 this relative quantification method, the  $2^{-\Delta\Delta C_T}$  values obtained represented the fold change in  
15 | gene expression of the colorectalCRC tumours, which was normalised with both reference  
16 genes, in relative to the calibrator (control sample)[9].

#### 17 c) Statistical analysis

18 | The difference in the expression level between colorectalCRC tumour and paired non-  
19 cancerous tissues was analysed by using Real-Time StatMiner software (Integromics). The  
20 paired t-test was then performed to assess the statistical significance of the observed  
21 differential expression patterns.

## 22 | RESULTS

### 23 | DEGs between colorectalCRC tumours and non-cancerous colonic tissues

24 | This preliminary study was conducted on paired samples pooled from four patients with CRC  
25 Stages I - III. In ACP-based GeneFishing PCR, 20 sets of arbitrary ACP primers were used to  
26

1 randomly amplify gene products in both colorectalCRC tumours and normal colonic samples.  
2 Upon visualisation on agarose gels, a total of 13 differentially expressed bands were observed  
3 by means of comparing bands intensity between the tumouric and non-cancerous samples, as  
4 shown in Figure 1. These bands were further sequenced for gene identification, and 16 DEGs  
5 were successfully reported. Among them, 13 were over-expressed in colorectalCRC tumours,  
6 whilst three were under-expressed, as listed in Table 2.

Formatted: Highlight

Table 2 Sequence similarities and identification of DEGs.

Differentially Expressed Band	DEG	Identity	Sequence Homology (%)	Accession Number	UniGene Number	Description
<i>Over-expressed</i>						
A4.1	DEG1	<i>Homo sapiens</i> proteasome (prosome, macropain) 26S subunit, ATPase, 5 (PSMC5), mRNA	502/506 (99%)	NM_002805.4	Hs.79387	Involves in the ATP-dependent degradation of ubiquitinated proteins.
	DEG2	<i>Homo sapiens</i> ubiquinol-cytochrome c reductase hinge protein (UQCRH), mRNA	514/521 (98%)	NM_006004.2	Hs.481571	A component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex, which is part of the mitochondrial respiratory chain.
A4.2	DEG3	<i>Homo sapiens</i> ribosomal protein S23 (RPS23), mRNA	551/551 (100%)	NM_001025.4	Hs.527193	A component of the 40S subunit of human ribosomes.
A6.1	DEG4	<i>Homo sapiens</i> ribosomal protein L10 (RPL10), transcript variant 1, mRNA	554/557 (99%)	NM_006013.3	Hs.534404	A component of the 60S subunit of human ribosomes.
A9.2	DEG6	<i>Homo sapiens</i> actin related protein 2/3 complex, subunit 2, 34kDa (ARPC2), transcript variant 2, mRNA	473/473 (100%)	NM_005731.2	Hs.529303	Involves in the regulation of actin polymerization as an actin-binding component of the Arp2/3 complex, and mediates the formation of branched actin networks together with an activating nucleation-promoting factor (NPF).
	DEG7	<i>Homo sapiens</i> TIMP metalloproteinase inhibitor 1 (TIMP1), mRNA	503/511 (98%)	NM_003254.2	Hs.522632	Irreversibly inactivates the metalloproteinases by binding to their catalytic zinc cofactor.
A10.1	DEG8	<i>Homo sapiens</i> ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, beta polypeptide (ATP5B), nuclear gene encoding mitochondrial protein, mRNA	917/919 (99%)	NM_001686.3	Hs.406510	A subunit of mitochondrial ATP synthase that catalyzes the synthesis of ATP by utilizing an electrochemical gradient of protons across the inner membrane during oxidative

12

						phosphorylation.
A13.2	DEG11	<i>Homo sapiens</i> chromosome 11 open reading frame 10 (C11orf10), mRNA	273/273 (100%)	NM_014206.3	Hs.437779	Unknown.
A13.3	DEG12	<i>Homo sapiens</i> mitochondrial ribosomal protein L24 (MRPL24), nuclear gene encoding mitochondrial protein, transcript variant 2, mRNA	408/411 (99%)	NM_024540.3	Hs.418233	Involves in protein synthesis within the mitochondrion.
A13.4	DEG13	<i>Homo sapiens</i> similar to OK/SW-CL.16 (LOC100288418)	635/644 (98%)	XM_002342023.1	-	Unknown.
A18.1	DEG14	<i>Homo sapiens</i> family with sequence similarity 96, member B (FAM96B), transcript variant 2, transcribed RNA	486/487 (99%)	NR_024525.1	Hs.9825	Involves in chromosome segregation as part of the mitotic spindle-associated MMXD complex.
A20.1	DEG15	<i>Homo sapiens</i> ribosomal protein L35 (RPL35), mRNA	440/446 (99%)	NM_007209.3	Hs.182825	A component of the 60S subunit of human ribosomes.
A20.2	DEG16	<i>Homo sapiens</i> chromosome 6 open reading frame173 (C6orf173), mRNA	551/554 (99%)	NM_001012507.2	Hs.486401	May be required for proper chromosome segregation during mitosis and involved with CENPT in the establishment of centromere chromatin structure.
<b><i>Under-expressed</i></b>						
A9.1	DEG5	<i>Homo sapiens</i> ribosomal protein L37 (RPL37), mRNA	284/284 (100%)	NM_000997.4	Hs.731513	A component of the 60S subunit of human ribosomes, and can bind to the 23S rRNA.
A13.1	DEG9	<i>Homo sapiens</i> solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1 (SLC25A1), nuclear gene encoding mitochondrial protein, mRNA	165/165 (100%)	NM_005984.2	Hs.111024	A mitochondrial tricarboxylate transporter which is responsible for the movement of citrate across the mitochondrial inner membrane.
	DEG10	<i>Homo sapiens</i> similar to cytochrome c oxidase subunit II (LOC100288578), miscRNA	141/146 (97%)	XR_078216.1	-	Unknown.

### Differential ability of the identified DEGs on early and advanced colorectal neoplasia

Following the identification of DEGs, the gene sequences obtained were then used to design primers and TaqMan probes for RT-qPCR analysis by Applied Biosystems, as listed in Table 3. In an attempt to assess the differential ability of identified DEGs on early and advanced colorectal adenocarcinoma, the recruited paired samples were further stratified into two groups according to the cancer stage. Among them, 13 patients with Stages I and II were grouped as early stage CRC, whilst the advanced stage CRC group comprised of 14 patients with Stages III and IV.

Table 3 Primers and TaqMan probes for relative quantification with Comparative  $C_T$  method.

DEG	Primers Sequence	TaqMan Probe Sequence
DEG1	Forward: 5'-GGGCGTGTGCACAGAAG-3' Reverse: 5'-AAGTCCTCCTGAGTGACATGGA-3'	5'-CTCGCAGGGCATAACAT-3'
DEG2	Forward: 5'-GATGCTTACCGAATCCGGAGATC-3' Reverse: 5'-GCATTGCTCTCTCACTGTTGTAG-3'	5'-CCTCTCCTCTCTCCTCCTCC-3'
DEG3	Forward: 5'-CAACCGTCATTGGGTACAAAGG-3' Reverse: 5'-TGTAAGGGTCCAGCTGATCAAGA-3'	5'-ATGGCAAGAAAATCAC-3'
DEG4	Forward: 5'-CGGCCAGGAACTTGAACCTG-3' Reverse: 5'-CCGAGCTGCAGAACAAGGA-3'	5'-CAGGGCCTCAATCACA-3'
DEG5	Forward: 5'-CTGGTCGAATGAGGCACCTAAAA-3' Reverse: 5'-TGGGTTTAGGTGTTGTTCCCTCAC-3'	5'-CATGCCTGAATCTGC-3'
DEG6	Forward: 5'-AGATTAGCGGGATGAAAACGTCTT-3' Reverse: 5'-CGCCAGATGCCGAGAAAA-3'	5'-CCCCGTGATTGTTTC-3'
DEG7	Forward: 5'-GGTAGTGATGTGCAAGAGTCCAT-3' Reverse: 5'-CCGCAGCGAGGAGTTTCT-3'	5'-CATTGCTGGAAAACCTG-3'
DEG8	Forward: 5'-GAAGGAGACCATCAAAGGATTCCA-3' Reverse: 5'-GAAGGCCTGTTCTGGGAGATG-3'	5'-ATTCACCTGCCAAAATC-3'
DEG9	Forward: 5'-GGCAGGGTGGTCCTGAGA-3' Reverse: 5'-CCGCCATTGGCCTTAACCTG-3'	5'-CCTCTCTCCGCCCGGACA-3'
DEG11	Forward: 5'-CAGGTTTCAGTGAAGCCATCTG-3' Reverse: 5'-GGGTTGGCATCTACGTGGA-3'	5'-CACCCAAGGGTAACAAC-3'
DEG12	Forward: 5'-CCAGGTCAAACCTGTGGATCCT-3' Reverse: 5'-GCTTCAGTAAATCTCCACTCGATCT-3'	5'-ATGGACAGGAAACCCAC-3'
DEG14	Forward: 5'-CCCGTCTTATCTGCAAGTT-3' Reverse: 5'-TCAAGATGGACGTGCACATTACTC-3'	5'-CATGCAGTGAACAAGC-3'
DEG15	Forward: 5'-CGGCCTCCAAGCTCTCT-3' Reverse: 5'-TGAGAACACGGGCAATGGATT-3'	5'-CCGGACGACTCGGATCT-3'
DEG16	Forward: 5'-GGACTCTTCTGCTAATCGATGAACA-3' Reverse: 5'-GCCTCAACTTCGCTGGAGAAAA-3'	5'-CAGATGGACCAATAAGTCA-3'



1 The analysis of RT-qPCR results was performed via Real-Time StatMiner software by  
 2 importing the raw Ct data. The within-group correlation then determined by calculating the  
 3 MAD for all the samples within the same experimental group. The biological samples which  
 4 do not correlate well with other samples in the same group, were detected as group outliers  
 5 and excluded from subsequent analysis. Both *ACTB* and *GAPDH* were used for normalisation  
 6 in computing the  $\Delta C_T$  (Figure 2) and  $2^{-\Delta\Delta C_T}$  values **by using the following formulas** (Table 4).

$$C_T(\text{Target gene}) - C_T(\text{Reference gene}) = \Delta C_T$$

$$\Delta C_T(\text{Sample}) - \Delta C_T(\text{Calibrator}) = \Delta\Delta C_T$$

$$\text{Relative Fold Change in Expression (RQ)} = 2^{-\Delta\Delta C_T}$$

10 **The relative fold change in the mRNA expression level between the colorectal tumours and**  
 11 **adjacent normal colonic mucosa were shown as the  $2^{-\Delta\Delta C_T}$  values.** —The statistical  
 12 significance of the observed fold change in expression was determined by paired t-test for all  
 13 the DEGs. A *p* value of less than 0.05 is considered as statistically significant (Table 4).

14 In both early and advanced stage CRC groups, the expression of four out of 16 DEGs was  
 15 reported to be significantly differed between tumouric and non-cancerous tissues.  
 16 Remarkably, the combination of this panel of four genes is different among two groups. The  
 17 *RPL35*, *RPS23* and *TIMP1* genes were found to be over-expressed in both early- and  
 18 advanced colorectal neoplasms (*p* < 0.05) (Figures 3 and 4). It is interesting to note that, the  
 19 under-expression of *ARPC2* gene (*p* < 0.05) was only observed in early stage **colorectalCRC**  
 20 tumours (Figure 3). On the other hand, the *C6orf173* gene was found to be over-expressed (*p*  
 21 < 0.05) in advanced colorectal adenocarcinomas, but not in early stage **colorectalCRC**  
 22 tumours (Figure 4).

Formatted: Highlight

Formatted: Highlight

Formatted: Centered

Formatted: Highlight

Formatted: Highlight

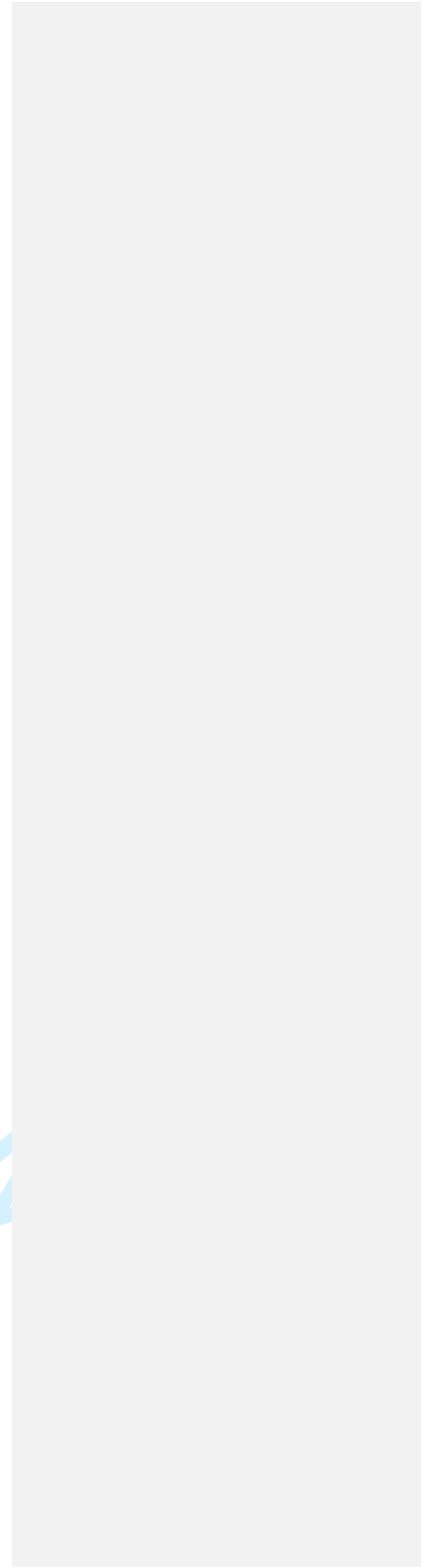
Field Code Changed

Formatted: Highlight

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6

For peer review only



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

Table 4  $\Delta C_T$  mean,  $\Delta\Delta C_T$ ,  $2^{-\Delta\Delta C_T}$  and  $p$  values for all the DEGs in both early- and advanced stage CRC groups.

DEG	Early Stage CRC					Advanced Stage CRC				
	$\Delta C_T$ Mean (CRC)	$\Delta C_T$ Mean (Normal)	$\Delta\Delta C_T$	$2^{-\Delta\Delta C_T}$	$p$ value	$\Delta C_T$ Mean (CRC)	$\Delta C_T$ Mean (Normal)	$\Delta\Delta C_T$	$2^{-\Delta\Delta C_T}$	$p$ value
<i>ARPC2</i>	2.6854	2.0664	0.6190	0.6511	0.0282*	2.7240	2.3300	0.3940	0.7610	0.2424
<i>ATP5B</i>	1.5846	1.2702	0.3144	0.8042	0.3524	1.9558	1.3838	0.5720	0.6727	0.1484
<i>C11orf10</i>	3.2897	3.3639	-0.0742	1.0528	0.8333	3.3281	3.6709	-0.3428	1.2682	0.3710
<i>C6orf173</i>	6.1083	7.1943	-1.0860	2.1228	0.0905	5.9949	7.9087	-1.9138	3.7680	0.0013*
<i>FAM96B</i>	3.5602	3.8955	-0.3353	1.2616	0.2935	3.5276	3.9920	-0.4644	1.3797	0.2113
<i>MRPL24</i>	4.9171	5.0839	-0.1668	1.1226	0.3564	4.9728	5.1467	-0.1739	1.1281	0.7001
<i>PSMC5</i>	3.8232	3.9617	-0.1385	1.1008	0.6812	3.7705	3.8455	-0.0750	1.0534	0.8048
<i>RPL10</i>	-0.7462	-0.4853	-0.2609	1.1982	0.4001	-1.1576	-0.5196	-0.6380	1.5562	0.0950
<i>RPL35</i>	-0.1926	0.6222	-0.8148	1.7591	0.0024*	0.1748	0.8769	-0.7021	1.6269	0.0372*
<i>RPL37</i>	-0.0059	-0.1539	0.1480	0.9025	0.8645	0.2184	0.7143	-0.4959	1.4102	0.1537
<i>RPS23</i>	0.2176	0.7739	-0.5563	1.4705	0.0310*	0.0676	0.9431	-0.8755	1.8346	0.0250*
<i>SLC25A1</i>	3.7514	3.5430	0.2084	0.8655	0.5721	3.5565	3.4428	0.1137	0.9242	0.7991
<i>TIMP1</i>	2.9096	4.3059	-1.3963	2.6323	0.0440*	2.3330	3.8547	-1.5217	2.8713	0.0062*
<i>UQCRH</i>	2.0087	2.2216	-0.2129	1.1590	0.4108	2.3375	2.4459	-0.1084	1.0780	0.7808

\* $p < 0.05$  = statistically significant

## DISCUSSION

Our current study has revealed two distinctive 4-gene signatures for both early- and advanced stage colorectal adenocarcinomas. The early stage sporadic CRC was characterised by the over-expression of *RPL35*, *RPS23* and *TIMP1* genes, as well as under-expression of *ARPC2* gene. On the other hand, the advanced primary colorectal CRC tumours were reported with over-expression of *C6orf173*, *RPL35*, *RPS23* and *TIMP1* genes. Although the relative fold change for *ARPC2*, *RPL35* and *RPS23* genes is below 2, the individual result does not affect the analysis since gene expression patterns of all four genes in combination were proposed to distinguish between the early- and advanced stage colorectal CRC neoplasms. The potential involvement of these DEGs and their altered expression levels in CRC were further supported by previous researches.

In fact, several proto-oncogenes and tumour suppressors are previously reported to regulate the ribosome production, i.e., the *RB*[10], *TP53*[11], *PTEN* genes[12], as well as the *MYC* gene family[13]. It is suggested that the alterations in ribosome biogenesis might affect the translation of genes that are involved in neoplastic transformation. In addition, the additional extra-ribosomal functions of the ribosomal proteins (r-proteins) in cellular apoptosis, cellular proliferation, cellular transformation, genes transcription, mRNA translation, DNA repair and inflammation, might also trigger and support the neoplastic development[14]. Hence, the over-expression of r-proteins-encoding genes observed in colorectal adenocarcinomas is not unexpected[15-17]. Our current study has revealed the significant over-expression of two r-proteins that were not previously described in colorectal CRC tumours, i.e., the *RPL35* and *RPS23*. The observed fold changes for the *RPL35* and *RPS23* mRNA levels were comparable between the early- and advanced stage colorectal CRC tumours in our sample cohort. This was in agreement with previous reports by Barnard et al. and Frigerio et al., where the changes in the mRNA expression levels of the r-proteins were irrespective of the cancer stage[18,19]. The hypothesis that the same ribosomal protein may contribute in different stages of cancer progression with their hitherto unknown extra-ribosomal roles might provide an explanation to these observations[20].

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 On the other hand, our present study also demonstrated an over-expression of the *TIMP1* gene  
11 2 | in both early- and advanced stage primary colorectalCRC tumours. This finding is supported  
12 3 | by Zeng et al., where the over-expression of TIMP1 was reported in all stages of primary  
13 4 | colorectalCRC tumours[21]. Under normal physiological conditions, the proteolytic activities  
14 5 | of MMPs are kept at bay by their natural inhibitors, the TIMPs[22]. Previous studies have  
15 6 | reported the over-expression of MMPs in both early- and advanced stage colorectalCRC  
16 7 | tumours, as well as other cancer types[23-25], which is in accordance to their biological roles.  
17 8 Hence, a similar scenario is expected for TIMPs and indeed, their suppressive role in tumour  
18 9 invasion and metastasis has been demonstrated in various cancer models[26]. However, more  
19 10 recent studies have revealed a direct correlation between TIMP1 expression and tumour  
20 11 aggressiveness in cancer, including CRC[21,27]. These findings, which are contradictory to  
21 12 its protease-inhibiting function, have suggested a possible tumour-promoting role of TIMP1 in  
22 13 tumorigenesis. It is postulated that the TIMP1 exhibits the abilities to inhibit tumour cell  
23 14 apoptosis and promote tumour angiogenesis, as well as other growth-factor-like effects[28].  
24 15 In our present study, the observed comparable over-expression of TIMP1 in both early- and  
25 16 | advanced stage sporadic colorectalCRC neoplasms was in line with its MMP inhibitory and  
26 17 | MMP-independent tumour-promoting activities.

18 In cancer biology, the expression of mRNAs and proteins of the ARP2/3 complex is often  
19 studied due to its role in cell migration, which contributes to cancer invasion and metastasis if  
20 aberrantly regulated[29]. We have detected a significant under-expression of ARPC2 in our  
21 | cohort of early stage primary colorectalCRC tumours. Surprisingly, this finding is  
22 | contradictory with the role played by ARPC2 in cancer invasion and metastasis theoretically.  
23 Previously, Kaneda et al. has reported the decreased expression of all the seven genes  
24 encoding the subunits of ARP2/3 complex in human gastric cancers. Among them, the Arp2,  
25 ARPC2 and ARPC3 showed the most prominent reduction in their expression levels[30]. The  
26 exact mechanism underlying this observation still remains unknown, but the epigenetic  
27 alteration might potentially provide an explanation for it. For instance, promoter  
28 hypermethylation that causes gene silencing is responsible for the reduced expression of

1 ARPC1 in human gastric cancer[31]. Similarly, the epigenetic study might also offer a clue  
2 for the under-expression of ARPC2 in colorectalCRC neoplasms.

3 *C6orf173*, which is also known as *CUG2* or *CENP-W*, is a novel oncogene that has been  
4 found to be up-regulated in many human cancer tissues. Its high expression level is  
5 profoundly reported in tumours of the ovary, liver, lung, pancreas, breast, colon, rectum and  
6 stomach. The CENP-W is a new member of the constitutive centromere-associated network,  
7 which specifically interacts with the CENP-T and plays an important role in mitosis[32]. In  
8 our current study, the CENP-W is over-expressed in advanced colorectalCRC adenocarcinoma.  
9 This finding correlates to its function in kinetochore assembly, where its aberrant expression  
10 might lead to abnormal cell division and aneuploidy in cancer[32]. In our study, the over-  
11 expression of CENP-W was observed in both early- and advanced cohort of colorectalCRC  
12 neoplasms but only statistically significant in the latter group. Given the fact that aneuploidy  
13 is constantly associated with a greater proportion of advanced CRC cases, the aberrant  
14 expression of CENP-W might potentially relate to a poorer prognosis of CRC[33].

15 In conclusion, we have characterised two distinctive gene expression patterns, which comprise  
16 of the *ARPC2*, *C6orf173*, *RPL35*, *RPS23* and *TIMPI* genes, for the stratification of primary  
17 colorectal adenocarcinomas among Malaysian CRC patients. Our current sample size was  
18 relatively small owing to the lack of a designated Tissue Bank in our institution. There were  
19 also not many CRC patient volunteers. Moreover, our stringent criteria for patient selection  
20 have also limited the availability of suitable specimens within the short sample collection  
21 period. Nevertheless, our identified mRNA expression patterns specific for early- and  
22 advanced stage colorectalCRC tumours are still convincing with our stringent sample  
23 selection criteria, high specificity primers and probes, as well as reliable statistical analysis.  
24 In future, the validation of these DEGs should be performed on a larger set of clinical samples,  
25 and extensive inter-laboratory testing of their differential abilities on each CRC stage is also  
26 desired. In addition, we should also integrate other imaging and histological information to  
27 complement our identified gene expression patterns, which then hold promises for better  
28 stratification of colorectalCRC tumours.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 **Funding** – This study was supported by FS176/2007C, PS172/2008C and Research  
11 2 Collaborative Grant, CG041-2013 from the University of Malaya.

12 3  
13 4 **Competing interests** – None.

14 5  
15 6 **Data sharing statement** – There is no additional data available.

16 7  
17 8 **Contributorship statement** – All authors were responsible for the design and analysis of the  
18 9 study. All authors were involved in drafting the manuscript, providing the intellectual input  
19 10 and approving the final version of the manuscript.

20 11  
21 12  
22 13 **REFERENCES**

- 23 14 1 Greene FL, Page D, Fleming ID, et al., eds. AJCC cancer staging manual (6th ed.).  
24 15 New York: Springer 2002.
- 25 16 2. Dukes CE. The classification of cancer of the rectum. J Pathol Bacteriol 1932; 35:  
26 17 323-32.
- 27 18 3. Astler VB, Collier FA. The prognostic significance of direct extension of carcinoma of  
28 19 the colon and rectum. Ann Surg 1954; 139: 846-52.
- 29 20 4. Edge SB, Byrd DR, Compton CC, et al. AJCC cancer staging manual (7th ed.). New  
30 21 York: Springer 2010.
- 31 22 5. Hu HK, Krasinskas A, Willis J. Perspectives on current tumour-node-metastasis  
32 23 (TNM) staging of cancers of the colon and rectum. Semin Oncol 2011; 38: 500-10.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 6. Doyle VJ, Bateman AC. Colorectal cancer staging using TNM 7: is it time to use this  
11 new staging system? *J Clin Pathol* 2012; 65: 372-4.  
12  
13  
14 3 7. Russo G, Zegar C, Giordano A. Advantages and limitations of microarray technology  
15 in human cancer. *Oncogene* 2003; 22: 6497-507.  
16  
17  
18 5 8. Puppa G, Sonzogni A, Colombari R, et al. TNM staging system of colorectal  
19 carcinoma: a critical appraisal of challenging issues. *Arch Pathol Lab Med* 2010; 134:  
20 837-52.  
21  
22  
23  
24  
25 8 9. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time  
26 quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 2001; 25: 402-8.  
27  
28  
29 10 10. Voit R, Schafer K, Grummt I. Mechanism of repression of RNA polymerase I  
30 transcription by the retinoblastoma protein. *Mol Cell Biol* 1997; 17: 4230-7.  
31  
32  
33  
34 12 11. Zhai W, Cornai L. Repression of RNA polymerase I transcription by the tumour  
35 suppressor p53. *Mol Cell Biol* 2000; 20: 5930-8.  
36  
37  
38 14 12. Backman S, Stambolic V, Mak T. PTEN function in mammalian cell size regulation.  
39 *Curr Opin Neurobiol* 2002; 12: 516-22.  
40  
41  
42 16 13. Greasley PJ, Bonnard C, Amati B. Myc induces the nucleolin and BN51 genes:  
43 possible implications in ribosome biogenesis. *Nucleic Acids Res* 2000; 28: 446-53.  
44  
45  
46  
47 18 14. Montanaro L, Treré D, Derenzini M. Nucleolus, ribosomes, and cancer. *Am J Pathol*  
48 2008; 173: 301-10.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 15. Sharp MG, Adams SM, Elvin P, et al. A sequence previously identified as metastasis-  
11 2 related encodes an acidic ribosomal phosphoprotein, P2. *Br J Cancer* 1990; 61: 83-8.  
12  
13  
14 3 16. Chester KA, Robson L, Begent RH, et al. Identification of a human ribosomal protein  
15 4 mRNA with increased expression in colorectal tumours. *Biochim Biophys Acta* 1989;  
16 5 1009: 297-300.  
17  
18  
19  
20 6 17. Pogue-Geile K, Geiser JR, Shu M, et al. Ribosomal protein genes are overexpressed in  
21 7 colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein.  
22 8 *Mol Cell Biol* 1991; 11: 3842-9.  
23  
24  
25  
26  
27 9 18. Barnard GF, Staniunas RJ, Mori M, et al. Gastric and hepatocellular carcinomas do not  
28 10 overexpress the same ribosomal protein messenger RNAs as colonic carcinoma.  
29 11 *Cancer Res* 1993; 53: 4048-52.  
30  
31  
32  
33 12 19. Frigerio JM, Dagorn JC, Iovanna JL. Cloning, sequencing and expression of the L5,  
34 13 L21, L27a, L28, S5, S9, S10 and S29 human ribosomal protein mRNAs. *Biochim*  
35 14 *Biophys Acta* 1995; 1262: 64-8.  
36  
37  
38  
39 15 20. Lai MD, Xu J. Ribosomal proteins and colorectal cancer. *Curr Genomics* 2007; 8: 43-9.  
40  
41  
42 16 21. Zeng ZS, Cohen AM, Zhang ZF, et al. Elevated tissue inhibitor of metalloproteinase 1  
43 17 RNA in colorectal cancer stroma correlates with lymph node and distant metastases.  
44 18 *Clin Cancer Res* 1995; 1: 899-906.  
45  
46  
47  
48 19 22. Ennis BW, Matrisian LM. Matrix degrading metalloproteinases. *J Neurooncol* 1994;  
49 20 18: 105-9.  
50  
51

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 23. Urbanski SJ, Edwards DR, Maitland A, et al. Expression of metalloproteinases and  
11 their inhibitors in primary pulmonary carcinomas. *Br J Cancer* 1992; 66: 1188-94.  
12 2  
13  
14 3 24. Boag AH, Young ID. Immunohistochemical analysis of type IV collagenase  
15 expression in prostatic hyperplasia and adenocarcinoma. *Mod Pathol* 1993; 6: 65-8.  
16 4  
17  
18 5 25. Newell KJ, Witty JP, Rodgers WH, et al. Expression and localisation of matrix-  
19 degrading metalloproteinases during colorectal tumourigenesis. *Mol Carcinogen* 1994;  
20 6  
21 10: 199-206.  
22 7  
23  
24 8 26. Khokha R, Waterhouse P. The role of tissue inhibitor of metalloproteinase-1 in  
25 specific aspects of cancer progression and reproduction. *J Neurooncol* 1994; 18: 123-7.  
26 9  
27  
28 10 27. Lu XQ, Levy M, Weinstein IB, et al. Immunological quantitation of levels of tissue  
29 inhibitor of metalloproteinase-1 in human colon cancer. *Cancer Res* 1991; 51: 6231-5.  
30 11  
31  
32 12 28. Egeblad M, Werb Z. New functions for the matrix metalloproteinases in cancer  
33 progression. *Nat Rev Cancer* 2002; 2: 161-74.  
34 13  
35  
36 14 29. Yamaguchi H, Wyckoff J, Condeelis J. Cell migration in tumours. *Curr Opin Cell Biol*  
37 2005; 17: 559-64.  
38 15  
39  
40 16 30. Kaneda A, Kaminishi M, Sugimura T, et al. Decreased expression of the seven  
41 ARP2/3 complex genes in human gastric cancers. *Cancer Lett* 2004; 212: 203-10.  
42 17  
43  
44 18 31. Kaneda A, Kaminishi M, Nakanishi Y, et al. Reduced expression of the insulin-  
45 induced protein 1 and p41 ARP2/3 complex genes in human gastric cancers. *Int J*  
46 19  
47 20  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 32. Hori T, Amano M, Suzuki A, et al. CCAN makes multiple contacts with centromeric  
2 DNA to provide distinct pathways to the outer kinetochore. Cell 2008; 135: 1039-52.

3 33. Chen HS, Sheen-Chen SM, Lu CC. DNA index and S-phase fraction in curative  
4 resection of colorectal adenocarcinoma: analysis of prognosis and current trends.  
5 World J Surg 2002; 26: 626-30.

6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16

For peer review only

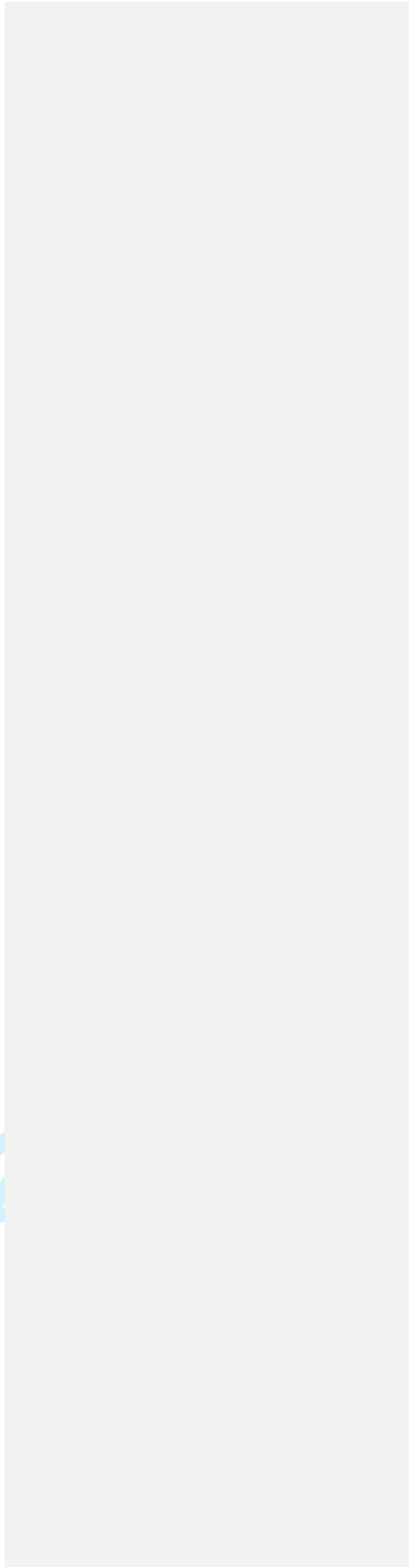
17 **Figure Legends**

18 Figure 1 Differential banding patterns on 3 % agarose gel post ACP-based PCR  
19 amplification between normal colon and colorectalCRC tumour samples (N:  
20 normal sample; C: CRC sample)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1 | Figure 2      Box-plots showing  $\Delta C_T$  values of all colorectalCRC tumours and normal
- 2                    colonic tissues in each early- (a) and advanced (b) stage CRC group.
  
- 3    Figure 3      Differential expression patterns of all the identified DEGs in early stage CRC
- 4                    group.
  
- 5    Figure 4      Differential expression patterns of all the identified DEGs in advanced stage
- 6                    CRC group.

For peer review only



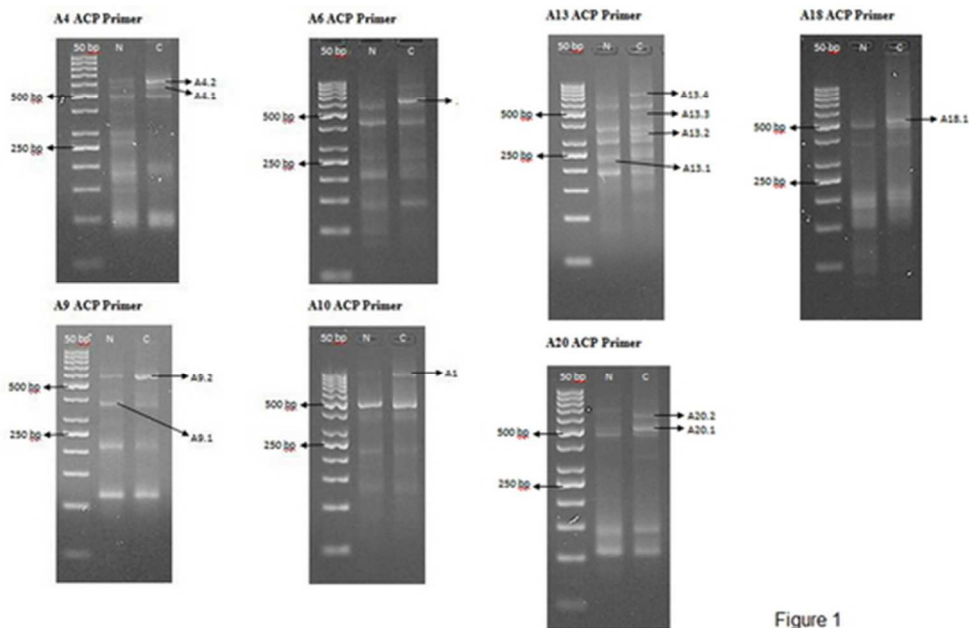


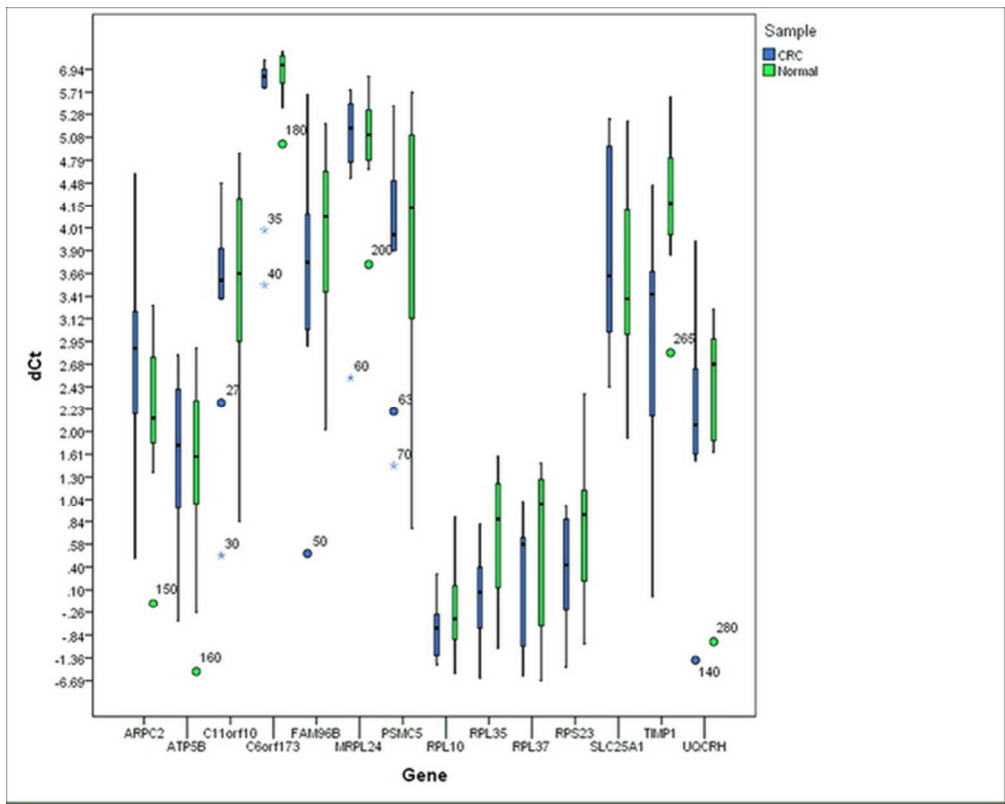
Figure 1

47x30mm (300 x 300 DPI)

review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

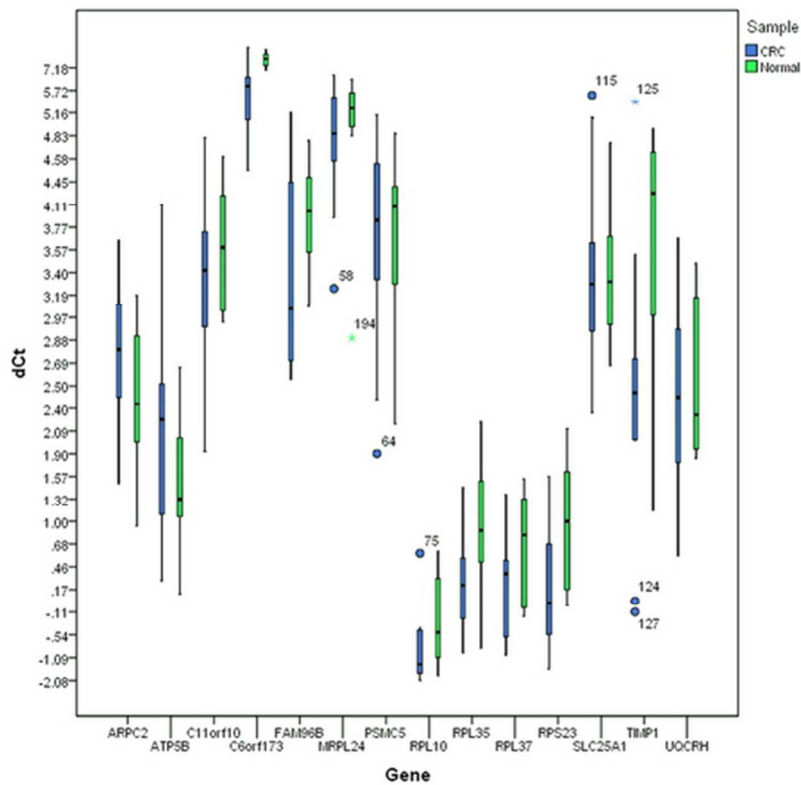
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



56x45mm (300 x 300 DPI)

ew only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

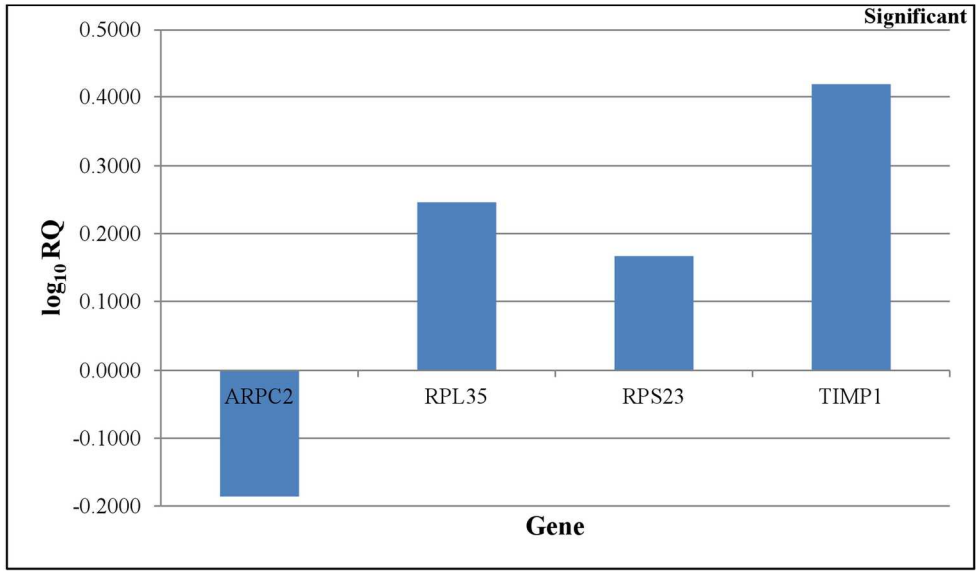
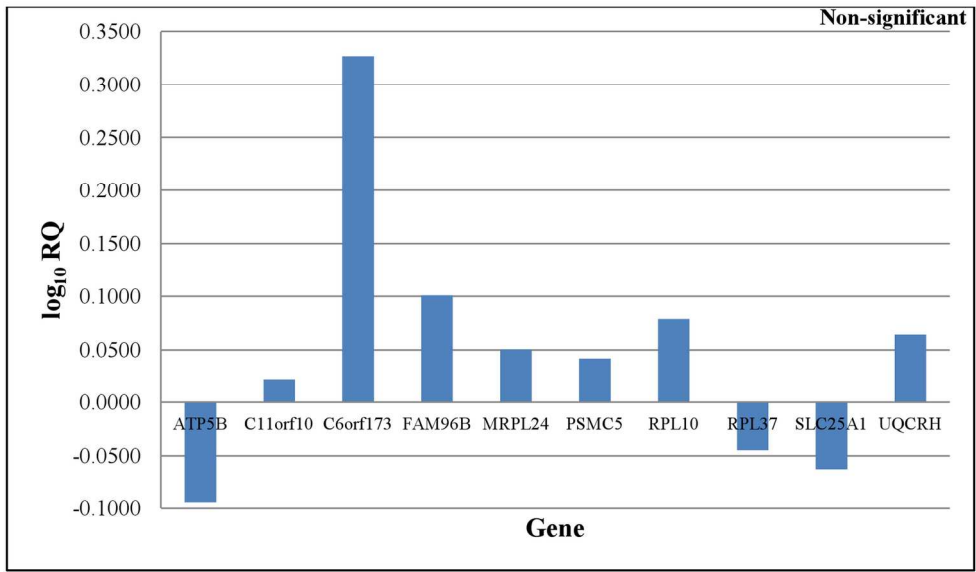


57x45mm (300 x 300 DPI)

Peer Review Only

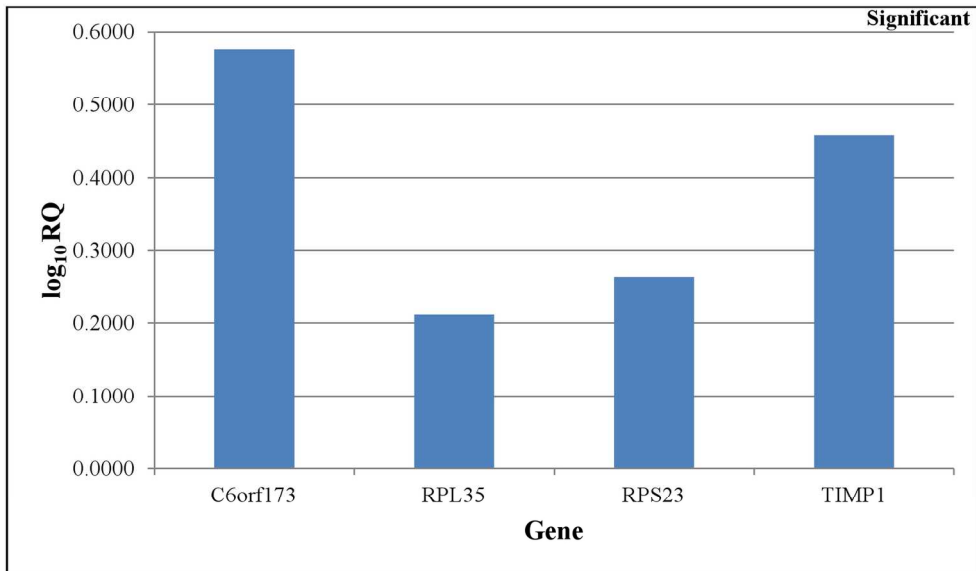
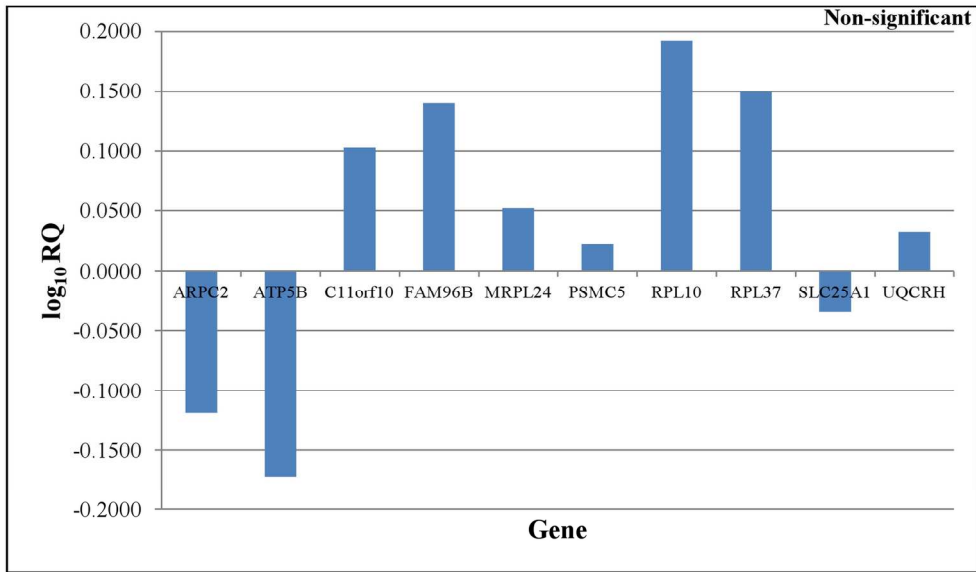


1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



186x221mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



187x224mm (300 x 300 DPI)

## STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
<b>Title and abstract</b>	1	<p>✓ (a) Indicate the study's design with a commonly used term in the title or the abstract</p> <p>✓ (b) Provide in the abstract an informative and balanced summary of what was done and what was found</p>
<b>Introduction</b>		
Background/rationale	2	✓ Explain the scientific background and rationale for the investigation being reported
Objectives	3	✓ State specific objectives, including any prespecified hypotheses
<b>Methods</b>		
Study design	4	✓ Present key elements of study design early in the paper
Setting	5	✓ Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	<p>✓ (a) <i>Cohort study</i>—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up</p> <p>✓ <i>Case-control study</i>—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls</p> <p><i>Cross-sectional study</i>—Give the eligibility criteria, and the sources and methods of selection of participants</p> <p>✓ (b) <i>Cohort study</i>—For matched studies, give matching criteria and number of exposed and unexposed</p> <p>✓ <i>Case-control study</i>—For matched studies, give matching criteria and the number of controls per case</p>
Variables	7	✓ Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	✓ For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	✓ Describe any efforts to address potential sources of bias
Study size	10	✓ Explain how the study size was arrived at
Quantitative variables	11	✓ Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	<p>✓ (a) Describe all statistical methods, including those used to control for confounding</p> <p>✓ (b) Describe any methods used to examine subgroups and interactions</p> <p>✓ (c) Explain how missing data were addressed</p> <p>✓ (d) <i>Cohort study</i>—If applicable, explain how loss to follow-up was addressed</p> <p>✓ <i>Case-control study</i>—If applicable, explain how matching of cases and controls was addressed</p> <p><i>Cross-sectional study</i>—If applicable, describe analytical methods taking account of sampling strategy</p> <p>✓ (e) Describe any sensitivity analyses</p>

Continued on next page

**Results**

Participants	13*	<p>√ (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed</p> <p>√ (b) Give reasons for non-participation at each stage</p> <p>(c) Consider use of a flow diagram</p>
Descriptive data	14*	<p>√ (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders</p> <p>(b) Indicate number of participants with missing data for each variable of interest</p> <p>(c) <i>Cohort study</i>—Summarise follow-up time (eg, average and total amount)</p>
Outcome data	15*	<p><i>Cohort study</i>—Report numbers of outcome events or summary measures over time</p> <p>√ <i>Case-control study</i>—Report numbers in each exposure category, or summary measures of exposure</p> <p><i>Cross-sectional study</i>—Report numbers of outcome events or summary measures</p>
Main results	16	<p>√ (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included</p> <p>(b) Report category boundaries when continuous variables were categorized</p> <p>(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period</p>
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses

**Discussion**

Key results	18	√ Summarise key results with reference to study objectives
Limitations	19	√ Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	√ Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results

**Other information**

Funding	22	√ Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based
---------	----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------

\*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at [www.strobe-statement.org](http://www.strobe-statement.org).

# BMJ Open

## Pair-wise comparison analysis of differential expression of mRNAs in early and advanced stage primary colorectal adenocarcinomas

Journal:	<i>BMJ Open</i>
Manuscript ID:	bmjopen-2014-004930.R2
Article Type:	Research
Date Submitted by the Author:	18-Jul-2014
Complete List of Authors:	Lau, Tze Pheng; University of Malaya, Biomedical Science Roslani, April Camilla; University of Malaya, Surgery Lian, Lay Hoong; University of Malaya, Biomedical Science Chai, Hwa Chia; University of Malaya, Biomedical Science Lee, Ping Chin; Universiti Sabah Malaysia, School of Science and Technology Hilmi, Ida; University of Malaya, Medicine Goh, Khean Lee; University of Malaya, Medicine Chua, Kek Heng; University of Malaya, Biomedical Science
<b>Primary Subject Heading</b>:	Gastroenterology and hepatology
Secondary Subject Heading:	Gastroenterology and hepatology
Keywords:	Gastrointestinal tumours < GASTROENTEROLOGY, Cancer genetics < GENETICS, PUBLIC HEALTH

SCHOLARONE™  
Manuscripts

only

1  
2  
3  
4 **1 Pair-wise comparison analysis of differential expression of mRNAs in early and**  
5  
6 **2 advanced stage primary colorectal adenocarcinomas**  
7

8  
9 3 Tze Pheng Lau<sup>1</sup>, April Camilla Roslani<sup>2</sup>, Lay Hoong Lian<sup>1</sup>, Hwa Chia Chai<sup>1</sup>, Ping Chin Lee<sup>3</sup>,  
10  
11 4 Ida Hilmi<sup>4</sup>,  
12  
13 5 Khean Lee Goh<sup>4</sup>, Kek Heng Chua<sup>1\*</sup>  
14  
15

16  
17  
18 7 <sup>1</sup>*Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala*  
19  
20 *Lumpur, Malaysia.*  
21

22 9 <sup>2</sup>*Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur,*  
23  
24 *Malaysia.*  
25

26 11 <sup>3</sup>*School of Science and Technology, Universiti Sabah Malaysia, 88400 Kota Kinabalu, Sabah,*  
27  
28 *Malaysia.*

29 13 <sup>4</sup>*Division of Gastroenterology and Hepatology, Department of Medicine, Faculty of Medicine,*  
30  
31 *University of Malaya, 50603 Kuala Lumpur, Malaysia.*  
32  
33

34  
35 16 *\*Correspondence author: Prof. Dr. Chua Kek Heng, Department of Biomedical Science,*  
36  
37 *Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur,*  
38  
39 *Malaysia.*  
40

41 19 *e-mail: khchua@um.edu.my*  
42

43  
44 20 *Phone no: 603-79676607*  
45

46  
47 21 *Fax no: 603-79676600*  
48

49 22 **Key words:** Gene expression, non-hereditary, colorectal cancer, Malaysian  
50

51  
52 23 **Word count:** 4151 words  
53  
54  
55  
56  
57  
58  
59  
60

1

2 **ABSTRACT**

3 **Objectives:** To characterise the mRNA expression patterns of early- and advanced stage  
4 colorectal adenocarcinomas of Malaysian patients.

5 **Design:** Comparative expression analysis.

6 **Setting and participants:** We performed a combination of ACP-based PCR and RT-qPCR  
7 for the identification of differentially expressed genes (DEGs) associated with early- and  
8 advanced stage primary colorectal tumours. We recruited four paired samples from CRC  
9 patients of Dukes' A and B for the preliminary differential expression study, and a total of 27  
10 paired samples, ranging from CRC Stages I – IV, for subsequent confirmatory test. The  
11 tumouric samples were obtained from CRC patients undergoing curative surgical resection  
12 without pre-operative chemoradiotherapy. The recruited CRC patients were newly-diagnosed  
13 with CRC, and were not associated with any hereditary syndromes, previously diagnosed  
14 cancer, or positive family history of CRC. The paired non-cancerous tissue specimens were  
15 excised from macroscopically normal colonic mucosa distally located from the colorectal  
16 tumours.

17 **Primary and secondary outcome measures:** The differential mRNA expression patterns of  
18 early- and advanced stage colorectal adenocarcinomas compared to macroscopically normal  
19 colonic mucosa were characterised by ACP-based PCR and RT-qPCR.

20 **Results:** The *RPL35*, *RPS23* and *TIMP1* genes were found to be over-expressed in both early-  
21 and advanced stage colorectal adenocarcinomas ( $p < 0.05$ ). On the other hand, the *ARPC2*  
22 gene was significantly under-expressed in early colorectal adenocarcinomas, while the  
23 advanced stage primary colorectal tumours exhibited an additional over-expression of the  
24 *C6orf173* gene ( $p < 0.05$ ).

25 **Conclusions:** We characterised two distinctive gene expression patterns to aid in the  
26 stratification of primary colorectal neoplasms among Malaysian CRC patients. Further work

2

1 can be done to assess and compare the mRNA expression levels of these identified DEGs  
2 between each CRC stage group, Stages I – IV.

### 3 **ARTICLE SUMMARY**

#### 4 **Article focus**

- 5     ▪ The latest staging system of colorectal tumours, which relies mainly on the  
6       clinicopathological assessment of primary tumours, is insufficient to address the  
7       complexity and heterogeneity of this disease. Hence, a non-anatomical, molecular-  
8       oriented staging system which can effectively predict the patients' outcome(s) and  
9       direct targeted treatment to different subgroups of CRC patients, is needed.
- 10    ▪ Despite continuous revisions in the CRC classification criteria and expansion in  
11      transcriptomic and proteomic studies, there is yet to be any molecular marker(s)  
12      incorporated for clinical purposes.
- 13    ▪ Article focus of this study is to characterise mRNA expression patterns of early- and  
14      advanced stage colorectal tumours of Malaysian patients.

#### 15 **Key messages**

- 16    ▪ The under-expression of *ARPC2* and over-expression of *C6orf173* gene were  
17      distinctive for the early- and advanced stage sporadic colorectal adenocarcinomas,  
18      respectively.

#### 19 **Strengths and limitations of this study**

20 This regional-based study has a relatively small sample size due to the strict sample  
21 recruitment criteria where all subjects were newly-diagnosed with CRC, and were not  
22 associated with any hereditary syndromes, previously diagnosed cancer, or positive family  
23 history of CRC. However, the findings of this study are still reliable in view of our stringent  
24 sample selection criteria, high specificity primers and probes, as well as reliable statistical  
25 analysis.

26



1

2 **Abbreviations:** CRC: colorectal cancer; TNM: tumour-node-metastasis; AJCC: American  
3 Joint Committee on Cancer; SAGE: serial analysis of gene expression; CEA:  
4 carcinoembryonic antigen; ACP: Annealing Control Primer; RT-qPCR:  
5 reverse transcription-quantitative real-time PCR; RIN: RNA integrity  
6 number; DEG: differentially expressed gene;  $\Delta\Delta C_T$ : comparative  $C_T$ ; MAD:  
7 median absolute deviation

8

## 9 INTRODUCTION

10 Cancer staging is vital for patient management, especially in prognosis prediction and  
11 planning of treatment intervention[1]. This is especially in the CRC staging system. As such,  
12 there have been many noteworthy improvements since the introduction of the classical Dukes'  
13 staging system, followed by the modified Astler-Coller staging system; to the latest 7<sup>th</sup> edition  
14 of TNM staging system published by the AJCC[2-4]. The TNM staging system allows the  
15 incorporation of various clinical information (which are obtained through histopathological  
16 examination, radiologic imaging and surgical findings), for accurate CRC stratification[5].  
17 However, these clinical assessments are greatly dependent on the expertise of pathologists,  
18 radiologists and clinicians.

19 The TNM classification is applicable for both clinical (cTNM) and pathological (pTNM)  
20 staging of primary colorectal tumours. Typically, it involves the assessment on the depth of  
21 bowel wall invasion at the time of diagnosis and the presence of regional lymph nodes  
22 metastases, as well as the presence of distant organ metastasis[4]. As a potentially worse  
23 patient outcome with more advanced disease stage is the core concept in cancer staging, AJCC  
24 revises the TNM classification system every few years with an attempt to formulate it for  
25 more accurate patient prognostication[5]. The latest 7<sup>th</sup> edition has further detailed the  
26 subclassification of the pN category and the assessment of discontinuous/satellite tumour foci.  
27 However, these revisions have increased the complexity and subjectivity during evaluation,

4

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 and thus might lead to inter-observer variability and hamper its efficiency in routine clinical  
2 practise[5,6]. In addition, current clinicopathological parameters are insufficient to address  
3 the great biologic and genetic heterogeneity of CRC in patients' outcome and treatment  
4 response prediction. From the perspective of clinical oncology, the integration of molecular  
5 biomarkers into existing clinicopathological assessment will further refine the cancer  
6 management in future.

7 Over the past decades, many researchers have attempted to establish gene expression  
8 signatures specifically for the diagnosis, prognostication and recurrence prediction of sporadic  
9 CRC. Transcriptional profiling promises a fairly dynamic view on the cellular functions,  
10 regulatory mechanisms and biochemical pathways involved in the disease pathogenesis and  
11 progression[7]. Various gene expression profiling techniques ranging from differential  
12 display, SAGE to microarrays have been utilised. Despite its wide application in gene  
13 expression profiling, microarray experiments have been subjected to various sources of  
14 variability, false-positives, as well as statistical and bioinformatic challenges. To date, none  
15 of the molecular markers described has been validated and employed in routine clinical  
16 practise owing to the poor reproducibility of the identified differentially expressed genes  
17 (DEGs) between different profiling platforms[8]. Although the KRAS mutation and  
18 mismatch repair status have showed promising prognostic and predictive values, they have yet  
19 to be incorporated into either routine pathological reporting systems or TNM staging  
20 systems[5].

21 Since most of the molecular studies on CRC were based in Western populations and different  
22 molecular changes were thought to underlie the development of sporadic CRC in populations  
23 with different genetic backgrounds, we aimed to investigate the changes in mRNA expression  
24 patterns in primary sporadic colorectal tumours with regards to our Malaysian patients. In our  
25 study, we have employed a combined approach of a two-step ACP-based PCR and real-time  
26 reverse transcription PCR to characterise the gene expression patterns for both early- and  
27 advanced stage sporadic colorectal adenocarcinomas.

28

1

## 2 MATERIALS AND METHODS

### 3 Patient selection and specimen collection

4 All patients presented with histologically confirmed colorectal adenocarcinomas and were  
5 staged accordingly to the AJCC TNM staging system (Table 1). The staging of cancer was  
6 performed by taking into consideration their histopathological reports, computed tomography  
7 images, morphological evaluations during surgery and serum CEA levels. All subjects were  
8 newly-diagnosed with CRC, and were not associated with any hereditary syndromes,  
9 previously diagnosed cancer, or positive family history of CRC. Initially, four CRC patients  
10 of Stages I - III were recruited for the preliminary ACP-based PCR analysis, while another 27  
11 patients with CRC Stages I – IV were recruited for subsequent RT-qPCR analysis. The  
12 patients' group was comprised of the three main ethnic groups in the Malaysian population,  
13 i.e., Chinese, Malays and Indians, in order to ensure a better representative of the study  
14 population.

15 The subjects were admitted to the University Malaya Medical Centre (UMMC), Kuala  
16 Lumpur, Malaysia, and underwent curative surgical resection between 2010 and 2011. None  
17 had received pre-operative chemoradiotherapy. The study protocol was approved by the  
18 Ethics Committee Board of UMMC (Ref. No.: 654.1), and written informed consent was  
19 obtained from all study subjects. The tumouric specimens were excised from the primary  
20 colorectal tumours, while the non-cancerous tissue specimens were obtained from distally  
21 located macroscopically normal colonic mucosa. Both colorectal tumour and paired non-  
22 cancerous tissue specimens were immersed in RNAlater RNA Stabilization Reagent (Qiagen)  
23 immediately after excision and stored at -80 °C.

24

25

26

27

28

29

30

31

1  
2  
3  
4  
5 1  
6  
7 2 Table 1 Cancer staging of recruited subjects.  
8

Subject	Cancer Stage
T1	Stage I / pT1N0M0
T2	Stage II / pT3N0M0
T3	Stage II / pT2N0M0
T4	Stage II / pT3N0M0
T5	Stage II / pT3N0M0
T6	Stage II / pT4N0M0
T7	Stage II / pT4N0M0
T8	Stage II / pT4N0M0
T9	Stage II / pT3N0M0
T10	Stage II / pT3N0M0
T11	Stage IV / pT3N2M1
T12	Stage IV
T13	Stage III / pT3N1M0
T14	Stage IV
T15	Stage III / pT3N1M0
T16	Stage III / pT3N2M0
T17	Stage IV / pT4N1M1
T18	Stage III / pT3N1M0
T19	Stage IV
T20	Stage III / pT4N1M0
T21	Stage III
T22	Stage II
T23	Stage III / pT3N1M0
T24	Stage II / pT3-4N0M0
T25	Stage IV / pT4N1M1
T26	Stage II / pT3N0M0
T27	Stage III / pT3N1M0

3

7

## 1 **Total RNA extraction**

2 Total RNA was extracted from homogenised colonic tissues with the RNeasy Plus Mini Kit  
3 (Qiagen) according to manufacturer's instructions. Subsequently, the RNA yield and integrity  
4 were ascertained via Agilent 2100 Bioanalyser in conjunction with Agilent RNA 6000 Nano  
5 Kits (Agilent Technologies). The values of RIN were then determined in order to assess the  
6 integrity of the isolated total RNA. In this study, only RNA samples with RIN values of 8.0 –  
7 10.0 and rRNA ratios [28S/18S] of 1.5 – 2.5 were selected for successive applications.

## 8 **ACP-based PCR analysis**

### 9 a) First-strand cDNA synthesis

10 The synthesis of first-strand cDNA was performed according to the manufacturer's protocol  
11 for the GeneFishing DEG Premix Kit (Seegene), as follows: 3 µg of total RNA was added  
12 with 2 µl of 10 µM dT-ACP1 (5'-CTGTGAATGCTGCGACTACGATXXXXX(T)<sub>18</sub>-3') and  
13 RNase-free water to a final volume of 9.5 µl. The mixture was then incubated at 80 °C for 3  
14 min, followed by chilling on ice for another 2 min. Subsequently, 4 µl of 5X RT buffer  
15 (Mbiotech), 5 µl of 2mM dNTP (Fermentas), 0.5 µl of 40 U/µl RNase inhibitor (Mbiotech)  
16 and 1 µl of 200 U/µl M-MLV reverse transcriptase (Mbiotech) were added. This mixture was  
17 then incubated at 42 °C for 90 min, heated at 94 °C for another 2 min and chilled on ice for 2  
18 min. Finally, 80 µl of DNase-free water was added to dilute the synthesised cDNA. The first-  
19 strand cDNA was stored under -20 °C until further analysis.

### 20 b) ACP-based GeneFishing PCR

21 First, all four cDNA samples within each CRC and control group samples were pooled  
22 together in equal amounts. The characterisation of DEGs was then conducted via ACP-based  
23 PCR based on 20 arbitrary ACP primers (Cat. No.: K1021) in a thermal cycler (Mastercycler  
24 Gradient, Eppendorf) according to the manufacturer's protocol (GeneFishing DEG Premix Kit,  
25 Seegene). Initially, the synthesis of second-strand cDNA was commenced in a one-cycle first-  
26 stage PCR: 94 °C for 5 min, 50 °C for 3 min and 72 °C for 1 min. Next, the constructed  
27 second-strand cDNA was subjected to second-stage PCR with 40 cycles of a denaturing step

1 at 94 °C for 40 sec, annealing step at 65 °C for 40 sec and extension step at 72 °C for 40 sec.  
2 Lastly, a final extension step at 72 °C for 5 min was carried out. The amplified products were  
3 then separated on 3 % (w/v) agarose gels stained with ethidium bromide.

#### 4 c) Cloning and sequencing

5 The identified differentially expressed bands were extracted from the agarose gel by using the  
6 PureLink Quick Gel Extraction Kit (Invitrogen). Each of these extracted DNA fragments was  
7 then individually cloned with the use of the TOPO TA Cloning Kit for Sequencing  
8 (Invitrogen). Subsequently, the plasmid containing the inserted DNA fragment was extracted  
9 from clones of interest via PureLink Quick Plasmid Miniprep Kit (Invitrogen). The isolated  
10 cloned plasmids were then sequenced with the ABI 3730xl DNA Analyser (Applied  
11 Biosystems). Finally, all the sequences obtained were analysed and matched for similarities  
12 with reference to the BLAST programme under the NCBI database.

#### 13 RT-qPCR analysis

##### 14 a) Reverse transcription

15 The total RNA isolated from 27 paired samples was reverse transcribed to first-strand cDNA,  
16 with the following protocol: 3 µg of total RNA was added with 2 µl of 0.5 µg/µl oligo(dT)<sub>12-18</sub>  
17 (Invitrogen) and RNase-free water to a final volume of 9.5 µl. The reaction mixture was then  
18 incubated at 80°C for 3 min, followed by chilling on ice for another 2 min. Next, 4 µl of 5X  
19 first strand buffer (Invitrogen), 5 µl of 2mM dNTP (Fermentas), 0.5 µl of 40 U/µl RNaseOUT  
20 recombinant RNase inhibitor (Invitrogen) and 1 µl of 200 U/µl M-MLV reverse transcriptase  
21 (Invitrogen) were added to the mixture. Finally, the reaction mixture was incubated at 42 °C  
22 for 90 min, heated at 94 °C for another 2 min and chilled on ice for 2 min. The synthesised  
23 first-strand cDNA was stored under -20 °C until further usage.

##### 24 b) $\Delta\Delta C_T$ analysis

25 The relative expression of identified DEGs in all paired colorectal tumours and control  
26 samples was determined via  $\Delta\Delta C_T$  method. The RT-qPCR was performed in a singleplex

1 reaction containing 50 ng first-strand cDNA under universal thermal cycling conditions with  
2 the ABI 7500 Fast Real-Time PCR System (Applied Biosystems). Both *ACTB* (Assay ID:  
3 Hs99999903\_m1) and *GAPDH* (Assay ID: Hs99999905\_m1) were used as reference genes  
4 and are commercially available as TaqMan Pre-designed Assays (Applied Biosystems). Prior  
5 to the analysis of gene expression, the amplification efficiency for all target and reference  
6 genes assays was measured by using the standard curve method with 2-log measurements.  
7 The amplification efficiency value of 90 – 110 % was acceptable (Applied Biosystems). In  
8 this relative quantification method, the  $2^{-\Delta\Delta C_T}$  values obtained represented the fold change in  
9 gene expression of the colorectal tumours, which was normalised with both reference genes,  
10 in relative to the calibrator (control sample)[9].

### 11 c) Statistical analysis

12 The difference in the expression level between primary colorectal tumour and paired non-  
13 cancerous tissues was analysed by using Real-Time StatMiner software (Integromics). The  
14 distribution of the  $\Delta C_T$  values obtained for each DEGs within each CRC and control group  
15 were tested for normality via the Shapiro-Wilk test. Subsequently, the paired t-test was  
16 performed to assess the statistical significance of the observed differential expression patterns.

## 18 RESULTS

### 19 DEGs between colorectal tumours and non-cancerous colonic tissues

20 This preliminary study was conducted on paired samples pooled from four patients with CRC  
21 Stages I - III. In ACP-based GeneFishing PCR, 20 sets of arbitrary ACP primers were used to  
22 randomly amplify gene products in both colorectal tumours and normal colonic samples.  
23 Upon visualisation on agarose gels, a total of 13 differentially expressed bands were observed  
24 by means of comparing bands intensity between the tumouric and non-cancerous samples, as  
25 shown in Figure 1. These bands were further sequenced for gene identification, and 16 DEGs  
26 were successfully reported. Among them, 13 were over-expressed in colorectal tumours,  
27 whilst three were under-expressed, as listed in Table 2.

Table 2 Sequence similarities and identification of DEGs.

Differentially Expressed Band	DEG	Identity	Sequence Homology (%)	Accession Number	UniGene Number	Description
<i>Over-expressed</i>						
A4.1	DEG1	<i>Homo sapiens</i> proteasome (prosome, macropain) 26S subunit, ATPase, 5 (PSMC5), mRNA	502/506 (99%)	NM_002805.4	Hs.79387	Involves in the ATP-dependent degradation of ubiquitinated proteins.
	DEG2	<i>Homo sapiens</i> ubiquinol-cytochrome c reductase hinge protein (UQCRH), mRNA	514/521 (98%)	NM_006004.2	Hs.481571	A component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex, which is part of the mitochondrial respiratory chain.
A4.2	DEG3	<i>Homo sapiens</i> ribosomal protein S23 (RPS23), mRNA	551/551 (100%)	NM_001025.4	Hs.527193	A component of the 40S subunit of human ribosomes.
A6.1	DEG4	<i>Homo sapiens</i> ribosomal protein L10 (RPL10), transcript variant 1, mRNA	554/557 (99%)	NM_006013.3	Hs.534404	A component of the 60S subunit of human ribosomes.
A9.2	DEG6	<i>Homo sapiens</i> actin related protein 2/3 complex, subunit 2, 34kDa (ARPC2), transcript variant 2, mRNA	473/473 (100%)	NM_005731.2	Hs.529303	Involves in the regulation of actin polymerization as an actin-binding component of the Arp2/3 complex, and mediates the formation of branched actin networks together with an activating nucleation-promoting factor (NPF).
	DEG7	<i>Homo sapiens</i> TIMP metalloproteinase inhibitor 1 (TIMP1), mRNA	503/511 (98%)	NM_003254.2	Hs.522632	Irreversibly inactivates the metalloproteinases by binding to their catalytic zinc cofactor.
A10.1	DEG8	<i>Homo sapiens</i> ATP synthase, H <sup>+</sup> -transporting, mitochondrial F1 complex, beta polypeptide (ATP5B), nuclear gene encoding mitochondrial protein, mRNA	917/919 (99%)	NM_001686.3	Hs.406510	A subunit of mitochondrial ATP synthase that catalyses the synthesis of ATP by utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation.



A13.2	DEG11	<i>Homo sapiens</i> chromosome 11 open reading frame 10 (C11orf10), mRNA	273/273 (100%)	NM_014206.3	Hs.437779	Unknown.
A13.3	DEG12	<i>Homo sapiens</i> mitochondrial ribosomal protein L24 (MRPL24), nuclear gene encoding mitochondrial protein, transcript variant 2, mRNA	408/411 (99%)	NM_024540.3	Hs.418233	Involves in protein synthesis within the mitochondrion.
A13.4	DEG13	<i>Homo sapiens</i> similar to OK/SW-CL.16 (LOC100288418)	635/644 (98%)	XM_002342023.1	-	Unknown.
A18.1	DEG14	<i>Homo sapiens</i> family with sequence similarity 96, member B (FAM96B), transcript variant 2, transcribed RNA	486/487 (99%)	NR_024525.1	Hs.9825	Involves in chromosome segregation as part of the mitotic spindle-associated MMXD complex.
A20.1	DEG15	<i>Homo sapiens</i> ribosomal protein L35 (RPL35), mRNA	440/446 (99%)	NM_007209.3	Hs.182825	A component of the 60S subunit of human ribosomes.
A20.2	DEG16	<i>Homo sapiens</i> chromosome 6 open reading frame173 (C6orf173), mRNA	551/554 (99%)	NM_001012507.2	Hs.486401	May be required for proper chromosome segregation during mitosis and involved with CENPT in the establishment of centromere chromatin structure.
<b><i>Under-expressed</i></b>						
A9.1	DEG5	<i>Homo sapiens</i> ribosomal protein L37 (RPL37), mRNA	284/284 (100%)	NM_000997.4	Hs.731513	A component of the 60S subunit of human ribosomes, and can bind to the 23S rRNA.
A13.1	DEG9	<i>Homo sapiens</i> solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1 (SLC25A1), nuclear gene encoding mitochondrial protein, mRNA	165/165 (100%)	NM_005984.2	Hs.111024	A mitochondrial tricarboxylate transporter which is responsible for the movement of citrate across the mitochondrial inner membrane.
	DEG10	<i>Homo sapiens</i> similar to cytochrome c oxidase subunit II (LOC100288578), miscRNA	141/146 (97%)	XR_078216.1	-	Unknown.

## 1 Differential ability of the identified DEGs on early and advanced colorectal neoplasia

2 Following the identification of DEGs, the gene sequences obtained were then used to design  
3 primers and TaqMan probes for RT-qPCR analysis by Applied Biosystems, as listed in Table  
4 3. In an attempt to assess the differential ability of identified DEGs on early and advanced  
5 colorectal adenocarcinoma, the recruited paired samples were further stratified into two  
6 groups according to the cancer stage. Among them, 13 patients with Stages I and II were  
7 grouped as early stage CRC, whilst the advanced stage CRC group comprised of 14 patients  
8 with Stages III and IV.

9  
10 Table 3 Primers and TaqMan probes for relative quantification with Comparative C<sub>T</sub>  
11 method.

DEG	Primers Sequence		TaqMan Probe Sequence
DEG1	Forward:	5'-GGGCGTGTGCACAGAAG-3'	5'-CTCGCAGGGCATAACAT-3'
	Reverse:	5'-AAGTCCTCCTGAGTGACATGGA-3'	
DEG2	Forward:	5'-GATGCTTACCGAATCCGGAGATC-3'	5'-CCTCTTCTCTCCTCCTCC-3'
	Reverse:	5'-GCATTGCTCTCTCACTGTTGTTAG-3'	
DEG3	Forward:	5'-CAACCGTCATTGGGTACAAAGG-3'	5'-ATGGCAAGAAAATCAC-3'
	Reverse:	5'-TGTAAGGGTCCAGCTGATCAAGA-3'	
DEG4	Forward:	5'-CGGCCAGGAACTTGAACCTG-3'	5'-CAGGGCCTCAATCACA-3'
	Reverse:	5'-CCGAGCTGCAGAACAAGGA-3'	
DEG5	Forward:	5'-CTGGTTCGAATGAGGCACCTAAAA-3'	5'-CATGCCTGAATCTGC-3'
	Reverse:	5'-TGGGTTTAGGTGTTTCCTCAC-3'	
DEG6	Forward:	5'-AGATTAGCGGGATGAAAACGTCTT-3'	5'-CCCCGTGATTGTTTTTC-3'
	Reverse:	5'-CGCCCAGATGCCGAGAAAA-3'	
DEG7	Forward:	5'-GGTAGTGATGTGCAAGAGTCCAT-3'	5'-CATTGCTGGAAAACCTG-3'
	Reverse:	5'-CCGCAGCGAGGAGTTTCT-3'	
DEG8	Forward:	5'-GAAGGAGACCATCAAAGGATTCCA-3'	5'-ATTCACCTGCCAAAATC-3'
	Reverse:	5'-GAAGGCCTGTTCTGGGAGATG-3'	
DEG9	Forward:	5'-GGCAGGGTGGTCTGAGA-3'	5'-CCTCTCTCCGCCCCGACA-3'
	Reverse:	5'-CCGCCATTGGCCTTAACTG-3'	
DEG11	Forward:	5'-CAGGTTTCAGTGAAGCCATCTG-3'	5'-CACCCAAGGGTAACAAC-3'
	Reverse:	5'-GGGTTGGCATCTACGTGTGA-3'	
DEG12	Forward:	5'-CCAGGTCAAACCTGTGGATCCT-3'	5'-ATGGACAGGAAACCCAC-3'
	Reverse:	5'-GCTTCAGTAAATCTCCACTCGATCT-3'	
DEG14	Forward:	5'-CCCGCTCCTTATCTGCAAGTT-3'	5'-CATGCAGTGAACAAGC-3'
	Reverse:	5'-TCAAGATGGACGTGCACATTACTC-3'	
DEG15	Forward:	5'-CGGCCTCCAAGCTCTCT-3'	5'-CCGGACGACTCGGATCT-3'
	Reverse:	5'-TGAGAACACGGGCAATGGATTT-3'	
DEG16	Forward:	5'-GGACTCTTCTGCTAATCGATGAACA-3'	5'-CAGATGGACCAATAAGTCA-3'
	Reverse:	5'-GCCTCAACTTCGTCTGGAGAAAA-3'	

1 The analysis of RT-qPCR results was performed via Real-Time StatMiner software by  
2 importing the raw Ct data. The within-group correlation of these  $\Delta C_T$  values was then  
3 determined by calculating the MAD for all the samples within the same experimental group.  
4 The biological samples which do not correlate well with other samples in the same group,  
5 were detected as group outliers and excluded from subsequent analysis. Both *ACTB* and  
6 *GAPDH* were used for normalisation in computing the  $\Delta C_T$  (Figure 2) and  $2^{-\Delta\Delta C_T}$  values by  
7 using the following formulas (Table 4).

$$C_{T(\text{Target gene})} - C_{T(\text{Reference gene})} = \Delta C_T$$

$$\Delta C_{T(\text{Sample})} - \Delta C_{T(\text{Calibrator})} = \Delta\Delta C_T$$

$$\text{Relative Fold Change in Expression (RQ)} = 2^{-\Delta\Delta C_T}$$

11 The relative fold change in the mRNA expression level between the colorectal tumours and  
12 adjacent normal colonic mucosa were shown as the  $2^{-\Delta\Delta C_T}$  values. The statistical significance  
13 of the observed fold change in expression was determined by paired t-test for all the DEGs. A  
14  $p$  value of less than 0.05 is considered as statistically significant (Table 4).

15 In both early and advanced stage CRC groups, the expression of four out of 16 DEGs was  
16 reported to be significantly differed between tumouric and non-cancerous tissues.  
17 Remarkably, the combination of this panel of four genes is different among two groups. The  
18 *RPL35*, *RPS23* and *TIMP1* genes were found to be over-expressed in both early- and  
19 advanced colorectal neoplasms ( $p < 0.05$ ) (Figures 3 and 4). It is interesting to note that, the  
20 under-expression of *ARPC2* gene ( $p < 0.05$ ) was only observed in early stage colorectal  
21 tumours (Figure 3). On the other hand, the *C6orf173* gene was found to be over-expressed ( $p$   
22  $< 0.05$ ) in advanced colorectal adenocarcinomas, but not in early stage colorectal tumours  
23 (Figure 4).

Table 4  $\Delta C_T$  mean,  $\Delta\Delta C_T$ ,  $2^{-\Delta\Delta C_T}$  and  $p$  values for all the DEGs in both early- and advanced stage CRC groups.

DEG	Early Stage CRC					Advanced Stage CRC				
	$\Delta C_T$ Mean (CRC)	$\Delta C_T$ Mean (Normal)	$\Delta\Delta C_T$	$2^{-\Delta\Delta C_T}$	$p$ value	$\Delta C_T$ Mean (CRC)	$\Delta C_T$ Mean (Normal)	$\Delta\Delta C_T$	$2^{-\Delta\Delta C_T}$	$p$ value
<i>ARPC2</i>	2.6854	2.0664	0.6190	0.6511	0.0282*	2.7240	2.3300	0.3940	0.7610	0.2424
<i>ATP5B</i>	1.5846	1.2702	0.3144	0.8042	0.3524	1.9558	1.3838	0.5720	0.6727	0.1484
<i>C11orf10</i>	3.2897	3.3639	-0.0742	1.0528	0.8333	3.3281	3.6709	-0.3428	1.2682	0.3710
<i>C6orf173</i>	6.1083	7.1943	-1.0860	2.1228	0.0905	5.9949	7.9087	-1.9138	3.7680	0.0013*
<i>FAM96B</i>	3.5602	3.8955	-0.3353	1.2616	0.2935	3.5276	3.9920	-0.4644	1.3797	0.2113
<i>MRPL24</i>	4.9171	5.0839	-0.1668	1.1226	0.3564	4.9728	5.1467	-0.1739	1.1281	0.7001
<i>PSMC5</i>	3.8232	3.9617	-0.1385	1.1008	0.6812	3.7705	3.8455	-0.0750	1.0534	0.8048
<i>RPL10</i>	-0.7462	-0.4853	-0.2609	1.1982	0.4001	-1.1576	-0.5196	-0.6380	1.5562	0.0950
<i>RPL35</i>	-0.1926	0.6222	-0.8148	1.7591	0.0024*	0.1748	0.8769	-0.7021	1.6269	0.0372*
<i>RPL37</i>	-0.0059	-0.1539	0.1480	0.9025	0.8645	0.2184	0.7143	-0.4959	1.4102	0.1537
<i>RPS23</i>	0.2176	0.7739	-0.5563	1.4705	0.0310*	0.0676	0.9431	-0.8755	1.8346	0.0250*
<i>SLC25A1</i>	3.7514	3.5430	0.2084	0.8655	0.5721	3.5565	3.4428	0.1137	0.9242	0.7991
<i>TIMP1</i>	2.9096	4.3059	-1.3963	2.6323	0.0440*	2.3330	3.8547	-1.5217	2.8713	0.0062*
<i>UQCRH</i>	2.0087	2.2216	-0.2129	1.1590	0.4108	2.3375	2.4459	-0.1084	1.0780	0.7808

\* $p < 0.05$  = statistically significant

## 1 DISCUSSION

2 Our current study has revealed two distinctive 4-gene signatures for both early- and advanced  
3 stage colorectal adenocarcinomas. The early stage sporadic CRC was characterised by the  
4 over-expression of *RPL35*, *RPS23* and *TIMP1* genes, as well as under-expression of *ARPC2*  
5 gene. On the other hand, the advanced primary colorectal tumours were reported with over-  
6 expression of *C6orf173*, *RPL35*, *RPS23* and *TIMP1* genes. Although the relative fold change  
7 for *ARPC2*, *RPL35* and *RPS23* genes is below 2, the individual result does not affect the  
8 analysis since gene expression patterns of all four genes in combination were proposed to  
9 distinguish between the early- and advanced stage colorectal neoplasms. The potential  
10 involvement of these DEGs and their altered expression levels in CRC were further supported  
11 by previous researches.

12 In fact, several proto-oncogenes and tumour suppressors are previously reported to regulate  
13 the ribosome production, i.e., the *RB*[10], *TP53*[11], *PTEN* genes[12], as well as the *MYC*  
14 gene family[13]. It is suggested that the alterations in ribosome biogenesis might affect the  
15 translation of genes that are involved in neoplastic transformation. In addition, the additional  
16 extra-ribosomal functions of the ribosomal proteins (r-proteins) in cellular apoptosis, cellular  
17 proliferation, cellular transformation, genes transcription, mRNA translation, DNA repair and  
18 inflammation, might also trigger and support the neoplastic development[14]. Hence, the  
19 over-expression of r-proteins-encoding genes observed in colorectal adenocarcinomas is not  
20 unexpected[15-17]. Our current study has revealed the significant over-expression of two r-  
21 proteins that were not previously described in colorectal tumours, i.e., the *RPL35* and *RPS23*.  
22 The observed fold changes for the *RPL35* and *RPS23* mRNA levels were comparable between  
23 the early- and advanced stage colorectal tumours in our sample cohort. This was in agreement  
24 with previous reports by Barnard et al. and Frigerio et al., where the changes in the mRNA  
25 expression levels of the r-proteins were irrespective of the cancer stage[18,19]. The  
26 hypothesis that the same ribosomal protein may contribute in different stages of cancer  
27 progression with their hitherto unknown extra-ribosomal roles might provide an explanation to  
28 these observations[20].

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 On the other hand, our present study also demonstrated an over-expression of the *TIMP1* gene  
2 in both early- and advanced stage primary colorectal tumours. This finding is supported by  
3 Zeng et al., where the over-expression of TIMP1 was reported in all stages of primary  
4 colorectal tumours[21]. Under normal physiological conditions, the proteolytic activities of  
5 MMPs are kept at bay by their natural inhibitors, the TIMPs[22]. Previous studies have  
6 reported the over-expression of MMPs in both early- and advanced stage colorectal tumours,  
7 as well as other cancer types[23-25], which is in accordance to their biological roles. Hence, a  
8 similar scenario is expected for TIMPs and indeed, their suppressive role in tumour invasion  
9 and metastasis has been demonstrated in various cancer models[26]. However, more recent  
10 studies have revealed a direct correlation between TIMP1 expression and tumour  
11 aggressiveness in cancer, including CRC[21,27]. These findings, which are contradictory to  
12 its protease-inhibiting function, have suggested a possible tumour-promoting role of TIMP1 in  
13 tumorigenesis. It is postulated that the TIMP1 exhibits the abilities to inhibit tumour cell  
14 apoptosis and promote tumour angiogenesis, as well as other growth-factor-like effects[28].  
15 In our present study, the observed comparable over-expression of TIMP1 in both early- and  
16 advanced stage sporadic colorectal neoplasms was in line with its MMP inhibitory and MMP-  
17 independent tumour-promoting activities.

18 In cancer biology, the expression of mRNAs and proteins of the ARP2/3 complex is often  
19 studied due to its role in cell migration, which contributes to cancer invasion and metastasis if  
20 aberrantly regulated[29]. We have detected a significant under-expression of ARPC2 in our  
21 cohort of early stage primary colorectal tumours. Surprisingly, this finding is contradictory  
22 with the role played by ARPC2 in cancer invasion and metastasis theoretically. Previously,  
23 Kaneda et al. has reported the decreased expression of all the seven genes encoding the  
24 subunits of ARP2/3 complex in human gastric cancers. Among them, the Arp2, ARPC2 and  
25 ARPC3 showed the most prominent reduction in their expression levels[30]. The exact  
26 mechanism underlying this observation still remains unknown, but the epigenetic alteration  
27 might potentially provide an explanation for it. For instance, promoter hypermethylation that  
28 causes gene silencing is responsible for the reduced expression of ARPC1 in human gastric

1 cancer[31]. Similarly, the epigenetic study might also offer a clue for the under-expression of  
2 ARPC2 in colorectal neoplasms.

3 *C6orf173*, which is also known as *CUG2* or *CENP-W*, is a novel oncogene that has been  
4 found to be up-regulated in many human cancer tissues. Its high expression level is  
5 profoundly reported in tumours of the ovary, liver, lung, pancreas, breast, colon, rectum and  
6 stomach. The CENP-W is a new member of the constitutive centromere-associated network,  
7 which specifically interacts with the CENP-T and plays an important role in mitosis[32]. In  
8 our current study, the CENP-W is over-expressed in advanced colorectal adenocarcinoma.  
9 This finding correlates to its function in kinetochore assembly, where its aberrant expression  
10 might lead to abnormal cell division and aneuploidy in cancer[32]. In our study, the over-  
11 expression of CENP-W was observed in both early- and advanced cohort of colorectal  
12 neoplasms but only statistically significant in the latter group. Given the fact that aneuploidy  
13 is constantly associated with a greater proportion of advanced CRC cases, the aberrant  
14 expression of CENP-W might potentially relate to a poorer prognosis of CRC[33].

15 In conclusion, we have characterised two distinctive gene expression patterns, which comprise  
16 of the *ARPC2*, *C6orf173*, *RPL35*, *RPS23* and *TIMP1* genes, for the stratification of primary  
17 colorectal adenocarcinomas among Malaysian CRC patients. It was postulated that the actin  
18 cytoskeleton might play an important role in determining the dysplastic cell morphology  
19 during the early development of CRC, while the aberrations in the assembly of functional  
20 kinetochore might be crucial for the aneuploidy of the advanced stage colorectal tumours.  
21 Nevertheless, the findings of this study were considered preliminary owing to the relatively  
22 small sample size. The main reason for this is the lack of a designated Tissue Bank in our  
23 institution. Moreover, the lack of CRC patient volunteers and our stringent criteria for patient  
24 selection have also limited the availability of suitable specimens within the short sample  
25 collection period.

26 However, our identified mRNA expression patterns specific for early- and advanced stage  
27 colorectal tumours are still convincing with our stringent sample selection criteria, high  
28 specificity primers and probes, as well as reliable statistical analysis. In future, the validation

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 of these DEGs should be performed on a larger set of clinical samples, and extensive inter-  
2 laboratory testing of their differential abilities on each CRC stage is also desired. In addition,  
3 we should also integrate other imaging and histological information to complement our  
4 identified gene expression patterns, which then hold promises for better stratification of  
5 tumours.

For peer review only



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **Funding** – This study was supported by FS176/2007C, PS172/2008C and Research  
4 Collaborative Grant, CG041-2013 from the University of Malaya.

5  
6 **Competing interests** – None.

7  
8 **Data sharing statement** – No additional data available.

9  
10 **Contributorship statement:**

11 KHC, KLG, IH, HCC and ACR had the original idea for this work and gained funding in  
12 collaboration with PCL. TPL carried out the experiment. TPL, CKH, PCL, HCC and LHL  
13 were involved in the data analysis. TPL wrote the first draft of this paper and all authors  
14 subsequently assisted in redrafting and have approved the final version.

1  
2  
3  
4  
5 1  
67 **REFERENCES**  
8

- 9  
10 3 1 Greene FL, Page D, Fleming ID, et al., eds. AJCC cancer staging manual (6th ed.).  
11  
12 4 New York: Springer 2002.
- 13  
14  
15 5 2. Dukes CE. The classification of cancer of the rectum. J Pathol Bacteriol 1932; 35:  
16  
17 6 323-32.
- 18  
19  
20  
21 7 3. Astler VB, Coller FA. The prognostic significance of direct extension of carcinoma of  
22  
23 8 the colon and rectum. Ann Surg 1954; 139: 846-52.
- 24  
25  
26 9 4. Edge SB, Byrd DR, Compton CC, et al. AJCC cancer staging manual (7th ed.). New  
27  
28 10 York: Springer 2010.
- 29  
30  
31  
32 11 5. Hu HK, Krasinskas A, Willis J. Perspectives on current tumour-node-metastasis  
33  
34 12 (TNM) staging of cancers of the colon and rectum. Semin Oncol 2011; 38: 500-10.
- 35  
36  
37 13 6. Doyle VJ, Bateman AC. Colorectal cancer staging using TNM 7: is it time to use this  
38  
39 14 new staging system? J Clin Pathol 2012; 65: 372-4.
- 40  
41  
42  
43 15 7. Russo G, Zegar C, Giordano A. Advantages and limitations of microarray technology  
44  
45 16 in human cancer. Oncogene 2003; 22: 6497-507.
- 46  
47  
48  
49 17 8. Puppa G, Sonzogni A, Colombari R, et al. TNM staging system of colorectal  
50  
51 18 carcinoma: a critical appraisal of challenging issues. Arch Pathol Lab Med 2010; 134:  
52  
53 19 837-52.
- 54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4  
5 1 9. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time  
6 quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 2001; 25: 402-8.  
7  
8  
9  
10 3 10. Voit R, Schafer K, Grummt I. Mechanism of repression of RNA polymerase I  
11 transcription by the retinoblastoma protein. *Mol Cell Biol* 1997; 17: 4230-7.  
12  
13  
14  
15 5 11. Zhai W, Cornai L. Repression of RNA polymerase I transcription by the tumour  
16 suppressor p53. *Mol Cell Biol* 2000; 20: 5930-8.  
17  
18  
19  
20  
21 7 12. Backman S, Stambolic V, Mak T. PTEN function in mammalian cell size regulation.  
22  
23  
24 8  
25  
26  
27 9 13. Greasley PJ, Bonnard C, Amati B. Myc induces the nucleolin and BN51 genes:  
28 possible implications in ribosome biogenesis. *Nucleic Acids Res* 2000; 28: 446-53.  
29  
30  
31  
32 11 14. Montanaro L, Treré D, Derenzini M. Nucleolus, ribosomes, and cancer. *Am J Pathol*  
33  
34  
35 12  
36  
37  
38 13 15. Sharp MG, Adams SM, Elvin P, et al. A sequence previously identified as metastasis-  
39 related encodes an acidic ribosomal phosphoprotein, P2. *Br J Cancer* 1990; 61: 83-8.  
40  
41  
42  
43 15 16. Chester KA, Robson L, Begent RH, et al. Identification of a human ribosomal protein  
44 mRNA with increased expression in colorectal tumours. *Biochim Biophys Acta* 1989;  
45  
46 16  
47 1009: 297-300.  
48  
49  
50  
51 18 17. Pogue-Geile K, Geiser JR, Shu M, et al. Ribosomal protein genes are overexpressed in  
52 colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein.  
53  
54 19  
55  
56 20  
57  
58  
59  
60

- 1  
2  
3  
4  
5 1 18. Barnard GF, Staniunas RJ, Mori M, et al. Gastric and hepatocellular carcinomas do not  
6 overexpress the same ribosomal protein messenger RNAs as colonic carcinoma.  
7  
8  
9 3 Cancer Res 1993; 53: 4048-52.  
10  
11  
12 4 19. Frigerio JM, Dagorn JC, Iovanna JL. Cloning, sequencing and expression of the L5,  
13 L21, L27a, L28, S5, S9, S10 and S29 human ribosomal protein mRNAs. *Biochim*  
14  
15  
16 6 *Biophys Acta* 1995; 1262: 64-8.  
17  
18  
19  
20 7 20. Lai MD, Xu J. Ribosomal proteins and colorectal cancer. *Curr Genomics* 2007; 8: 43-9.  
21  
22  
23 8 21. Zeng ZS, Cohen AM, Zhang ZF, et al. Elevated tissue inhibitor of metalloproteinase 1  
24  
25  
26 9 RNA in colorectal cancer stroma correlates with lymph node and distant metastases.  
27  
28  
29 10 *Clin Cancer Res* 1995; 1: 899-906.  
30  
31  
32 11 22. Ennis BW, Matrisian LM. Matrix degrading metalloproteinases. *J Neurooncol* 1994;  
33  
34 12 18: 105-9.  
35  
36  
37 13 23. Urbanski SJ, Edwards DR, Maitland A, et al. Expression of metalloproteinases and  
38  
39  
40 14 their inhibitors in primary pulmonary carcinomas. *Br J Cancer* 1992; 66: 1188-94.  
41  
42  
43 15 24. Boag AH, Young ID. Immunohistochemical analysis of type IV collagenase  
44  
45  
46 16 expression in prostatic hyperplasia and adenocarcinoma. *Mod Pathol* 1993; 6: 65-8.  
47  
48  
49 17 25. Newell KJ, Witty JP, Rodgers WH, et al. Expression and localisation of matrix-  
50  
51  
52 18 degrading metalloproteinases during colorectal tumourigenesis. *Mol Carcinogen* 1994;  
53  
54 19 10: 199-206.  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 1 26. Khokha R, Waterhouse P. The role of tissue inhibitor of metalloproteinase-1 in  
5  
6 specific aspects of cancer progression and reproduction. *J Neurooncol* 1994; 18: 123-7.  
7  
8  
9  
10 3 27. Lu XQ, Levy M, Weinstein IB, et al. Immunological quantitation of levels of tissue  
11  
12 inhibitor of metalloproteinase-1 in human colon cancer. *Cancer Res* 1991; 51: 6231-5.  
13  
14  
15 5 28. Egeblad M, Werb Z. New functions for the matrix metalloproteinases in cancer  
16  
17 progression. *Nat Rev Cancer* 2002; 2: 161-74.  
18  
19  
20  
21 7 29. Yamaguchi H, Wyckoff J, Condeelis J. Cell migration in tumours. *Curr Opin Cell Biol*  
22  
23 2005; 17: 559-64.  
24  
25  
26  
27 9 30. Kaneda A, Kaminishi M, Sugimura T, et al. Decreased expression of the seven  
28  
29 ARP2/3 complex genes in human gastric cancers. *Cancer Lett* 2004; 212: 203-10.  
30  
31  
32 11 31. Kaneda A, Kaminishi M, Nakanishi Y, et al. Reduced expression of the insulin-  
33  
34 induced protein 1 and p41 ARP2/3 complex genes in human gastric cancers. *Int J*  
35  
36 *Cancer* 2002; 100: 57-62.  
37  
38  
39  
40 14 32. Hori T, Amano M, Suzuki A, et al. CCAN makes multiple contacts with centromeric  
41  
42 DNA to provide distinct pathways to the outer kinetochore. *Cell* 2008; 135: 1039-52.  
43  
44  
45  
46 16 33. Chen HS, Sheen-Chen SM, Lu CC. DNA index and S-phase fraction in curative  
47  
48 resection of colorectal adenocarcinoma: analysis of prognosis and current trends.  
49  
50 *World J Surg* 2002; 26: 626-30.  
51  
52  
53  
54 19

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 1 **Figure Legends**

- 2 Figure 1 Differential banding patterns on 3 % agarose gel post ACP-based PCR  
3 amplification between normal colon and colorectal tumour samples (N: normal  
4 sample; C: CRC sample)
- 5 Figure 2 Box-plots showing  $\Delta C_T$  values of all colorectal tumours and normal colonic  
6 tissues in each early- (a) and advanced (b) stage CRC group.
- 7 Figure 3 Differential expression patterns of all the identified DEGs in early stage CRC  
8 group.
- 9 Figure 4 Differential expression patterns of all the identified DEGs in advanced stage  
10 CRC group.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 **Pair-wise comparison analysis of differential expression of mRNAs in early and**  
11 **advanced stage primary colorectal adenocarcinomas**  
12

13 Tze Pheng Lau<sup>1</sup>, April Camilla Roslani<sup>2</sup>, Lay Hoong Lian<sup>1</sup>, Hwa Chia Chai<sup>1</sup>, Ping Chin Lee<sup>3</sup>,  
14 Ida Hilmi<sup>4</sup>, Khean Lee Goh<sup>4</sup>, Kek Heng Chua<sup>1\*</sup>  
15  
16

17  
18  
19 <sup>1</sup>*Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala*  
20 *Lumpur, Malaysia.*

21 <sup>2</sup>*Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur,*  
22 *Malaysia.*

23 <sup>3</sup>*School of Science and Technology, Universiti Sabah Malaysia, 88400 Kota Kinabalu, Sabah,*  
24 *Malaysia.*

25 <sup>4</sup>*Division of Gastroenterology and Hepatology, Department of Medicine, Faculty of Medicine,*  
26 *University of Malaya, 50603 Kuala Lumpur, Malaysia.*  
27  
28

29  
30  
31  
32 *\*Correspondence author: Prof. Dr. Chua Kek Heng, Department of Biomedical Science,*  
33 *Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur,*  
34 *Malaysia.*  
35  
36

37 *e-mail: khchua@um.edu.my*

38  
39 *Phone no: 603-79676607*

40  
41 *Fax no: 603-79676600*  
42

43 **Key words:** Gene expression, non-hereditary, colorectal cancer, Malaysian  
44

45  
46 **Word count:** 4151 words  
47  
48  
49  
50  
51

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 **1 ABSTRACT**

11 **2 Objectives:** To characterise the mRNA expression patterns of early- and advanced stage  
12 colorectal adenocarcinomas of Malaysian patients.

13  
14 **4 Design:** Comparative expression analysis.

15  
16  
17 **5 Setting and participants:** We performed a combination of ACP-based PCR and RT-qPCR  
18 for the identification of differentially expressed genes (DEGs) associated with early- and  
19 advanced stage primary colorectal tumours. We recruited four paired samples from CRC  
20 patients of Dukes' A and B for the preliminary differential expression study, and a total of 27  
21 paired samples, ranging from CRC Stages I – IV, for subsequent confirmatory test. The  
22 tumouric samples were obtained from CRC patients undergoing curative surgical resection  
23 without pre-operative chemoradiotherapy. The recruited CRC patients were newly-diagnosed  
24 with CRC, and were not associated with any hereditary syndromes, previously diagnosed  
25 cancer, or positive family history of CRC. The paired non-cancerous tissue specimens were  
26 excised from macroscopically normal colonic mucosa distally located from the colorectal  
27 tumours.  
28  
29  
30  
31  
32

33  
34 **16 Primary and secondary outcome measures:** The differential mRNA expression patterns of  
35 early- and advanced stage colorectal adenocarcinomas compared to macroscopically normal  
36 colonic mucosa were characterised by ACP-based PCR and RT-qPCR.  
37

38  
39 **19 Results:** The *RPL35*, *RPS23* and *TIMP1* genes were found to be over-expressed in both early-  
40 and advanced stage colorectal adenocarcinomas ( $p < 0.05$ ). On the other hand, the *ARPC2*  
41 gene was significantly under-expressed in early colorectal adenocarcinomas, while the  
42 advanced stage primary colorectal tumours exhibited an additional over-expression of the  
43 *C6orf173* gene ( $p < 0.05$ ).  
44  
45

46  
47 **24 Conclusions:** We characterised two distinctive gene expression patterns to aid in the  
48 stratification of primary colorectal neoplasms among Malaysian CRC patients. Further work  
49 can be done to assess and compare the mRNA expression levels of these identified DEGs  
50 between each CRC stage group, Stages I – IV.  
51

52 2  
53  
54  
55  
56  
57  
58  
59  
60



## ARTICLE SUMMARY

### Article focus

- The latest staging system of colorectal tumours, which relies mainly on the clinicopathological assessment of primary tumours, is insufficient to address the complexity and heterogeneity of this disease. Hence, a non-anatomical, molecular-oriented staging system which can effectively predict the patients' outcome(s) and direct targeted treatment to different subgroups of CRC patients, is needed.
- Despite continuous revisions in the CRC classification criteria and expansion in transcriptomic and proteomic studies, there is yet to be any molecular marker(s) incorporated for clinical purposes.
- Article focus of this study is to characterise mRNA expression patterns of early- and advanced stage colorectal tumours of Malaysian patients.

### Key messages

- The under-expression of *ARPC2* and over-expression of *C6orf173* gene were distinctive for the early- and advanced stage sporadic colorectal adenocarcinomas, respectively.

### Strengths and limitations of this study

This regional-based study has a relatively small sample size due to the strict sample recruitment criteria where all subjects were newly-diagnosed with CRC, and were not associated with any hereditary syndromes, previously diagnosed cancer, or positive family history of CRC. However, the findings of this study are still reliable in view of our stringent sample selection criteria, high specificity primers and probes, as well as reliable statistical analysis.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 **Abbreviations:** CRC: colorectal cancer; TNM: tumour-node-metastasis; AJCC: American  
11 2 Joint Committee on Cancer; SAGE: serial analysis of gene expression; CEA:  
12 3 carcinoembryonic antigen; ACP: Annealing Control Primer; RT-qPCR:  
13 4 reverse transcription-quantitative real-time PCR; RIN: RNA integrity  
14 5 number; DEG: differentially expressed gene;  $\Delta\Delta C_T$ : comparative  $C_T$ ; MAD:  
15 6 median absolute deviation  
16  
17  
18  
19 7

## 20 21 8 **INTRODUCTION**

22  
23 9 Cancer staging is vital for patient management, especially in prognosis prediction and  
24 10 planning of treatment intervention[1]. This is especially in the CRC staging system. As such,  
25 11 there have been many noteworthy improvements since the introduction of the classical Dukes'  
26 12 staging system, followed by the modified Astler-Coller staging system; to the latest 7<sup>th</sup> edition  
27 13 of TNM staging system published by the AJCC[2-4]. The TNM staging system allows the  
28 14 incorporation of various clinical information (which are obtained through histopathological  
29 15 examination, radiologic imaging and surgical findings), for accurate CRC stratification[5].  
30 16 However, these clinical assessments are greatly dependent on the expertise of pathologists,  
31 17 radiologists and clinicians.  
32  
33  
34  
35

36 18 The TNM classification is applicable for both clinical (cTNM) and pathological (pTNM)  
37 19 staging of primary colorectal tumours. Typically, it involves the assessment on the depth of  
38 20 bowel wall invasion at the time of diagnosis and the presence of regional lymph nodes  
39 21 metastases, as well as the presence of distant organ metastasis[4]. As a potentially worse  
40 22 patient outcome with more advanced disease stage is the core concept in cancer staging, AJCC  
41 23 revises the TNM classification system every few years with an attempt to formulate it for  
42 24 more accurate patient prognostication[5]. The latest 7<sup>th</sup> edition has further detailed the  
43 25 subclassification of the pN category and the assessment of discontinuous/satellite tumour foci.  
44 26 However, these revisions have increased the complexity and subjectivity during evaluation,  
45 27 and thus might lead to inter-observer variability and hamper its efficiency in routine clinical  
46 28 practise[5,6]. In addition, current clinicopathological parameters are insufficient to address  
47  
48  
49  
50  
51  
52

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 | the great biologic and genetic heterogeneity of CRC in patients' outcome and treatment  
2 | response prediction. From the perspective of clinical oncology, the integration of molecular  
3 | biomarkers into existing clinicopathological assessment will further refine the cancer  
4 | management in future.

5 | Over the past decades, many researchers have attempted to establish gene expression  
6 | signatures specifically for the diagnosis, prognostication and recurrence prediction of sporadic  
7 | CRC. Transcriptional profiling promises a fairly dynamic view on the cellular functions,  
8 | regulatory mechanisms and biochemical pathways involved in the disease pathogenesis and  
9 | progression[7]. Various gene expression profiling techniques ranging from differential  
10 | display, SAGE to microarrays have been utilised. Despite its wide application in gene  
11 | expression profiling, microarray experiments have been subjected to various sources of  
12 | variability, false-positives, as well as statistical and bioinformatic challenges. To date, none  
13 | of the molecular markers described has been validated and employed in routine clinical  
14 | practise owing to the poor reproducibility of the identified differentially expressed genes  
15 | (DEGs) between different profiling platforms[8]. Although the KRAS mutation and  
16 | mismatch repair status have showed promising prognostic and predictive values, they have yet  
17 | to be incorporated into either routine pathological reporting systems or TNM staging  
18 | systems[5].

19 | Since most of the molecular studies on CRC were based in Western populations and different  
20 | molecular changes were thought to underlie the development of sporadic CRC in populations  
21 | with different genetic backgrounds, we aimed to investigate the changes in mRNA expression  
22 | patterns in primary sporadic colorectal tumours with regards to our Malaysian patients. In our  
23 | study, we have employed a combined approach of a two-step ACP-based PCR and real-time  
24 | reverse transcription PCR to characterise the gene expression patterns for both early- and  
25 | advanced stage sporadic colorectal adenocarcinomas.

## 1 MATERIALS AND METHODS

### 2 Patient selection and specimen collection

3 All patients presented with histologically confirmed colorectal adenocarcinomas and were  
4 staged accordingly to the AJCC TNM staging system (Table 1). The staging of cancer was  
5 performed by taking into consideration their histopathological reports, computed tomography  
6 images, morphological evaluations during surgery and serum CEA levels. All subjects were  
7 newly diagnosed with CRC, and were not associated with any hereditary syndromes,  
8 previously diagnosed cancer, or positive family history of CRC. Initially, four CRC patients  
9 of Stages I - III were recruited for the preliminary ACP-based PCR analysis, while another 27  
10 patients with CRC Stages I – IV were recruited for subsequent RT-qPCR analysis. The  
11 patients' group was comprised of the three main ethnic groups in the Malaysian population,  
12 i.e., Chinese, Malays and Indians, in order to ensure a better representative of the study  
13 population.

14 The subjects were admitted to the University Malaya Medical Centre (UMMC), Kuala  
15 Lumpur, Malaysia, and underwent curative surgical resection between 2010 and 2011. None  
16 had received pre-operative chemoradiotherapy. The study protocol was approved by the  
17 Ethics Committee Board of UMMC (Ref. No.: 654.1), and written informed consent was  
18 obtained from all study subjects. The tumouric specimens were excised from the primary  
19 colorectal tumours, while the non-cancerous tissue specimens were obtained from distally  
20 located macroscopically normal colonic mucosa. Both colorectal tumour and paired non-  
21 cancerous tissue specimens were immersed in RNeasy RNA Stabilization Reagent (Qiagen)  
22 immediately after excision and stored at -80 °C.

1 Table 1 Cancer staging of recruited subjects.

Subject	Cancer Stage
T1	Stage I / pT1N0M0
T2	Stage II / pT3N0M0
T3	Stage II / pT2N0M0
T4	Stage II / pT3N0M0
T5	Stage II / pT3N0M0
T6	Stage II / pT4N0M0
T7	Stage II / pT4N0M0
T8	Stage II / pT4N0M0
T9	Stage II / pT3N0M0
T10	Stage II / pT3N0M0
T11	Stage IV / pT3N2M1
T12	Stage IV
T13	Stage III / pT3N1M0
T14	Stage IV
T15	Stage III / pT3N1M0
T16	Stage III / pT3N2M0
T17	Stage IV / pT4N1M1
T18	Stage III / pT3N1M0
T19	Stage IV
T20	Stage III / pT4N1M0
T21	Stage III
T22	Stage II
T23	Stage III / pT3N1M0
T24	Stage II / pT3-4N0M0
T25	Stage IV / pT4N1M1
T26	Stage II / pT3N0M0
T27	Stage III / pT3N1M0

2

7

## 1 Total RNA extraction

2 Total RNA was extracted from homogenised colonic tissues with the RNeasy Plus Mini Kit  
3 (Qiagen) according to manufacturer's instructions. Subsequently, the RNA yield and integrity  
4 were ascertained via Agilent 2100 Bioanalyser in conjunction with Agilent RNA 6000 Nano  
5 Kits (Agilent Technologies). The values of RIN were then determined in order to assess the  
6 integrity of the isolated total RNA. In this study, only RNA samples with RIN values of 8.0 –  
7 10.0 and rRNA ratios [28S/18S] of 1.5 – 2.5 were selected for successive applications.

## 8 ACP-based PCR analysis

### 9 a) First-strand cDNA synthesis

10 The synthesis of first-strand cDNA was performed according to the manufacturer's protocol  
11 for the GeneFishing DEG Premix Kit (Seegene), as follows: 3 µg of total RNA was added  
12 with 2 µl of 10 µM dT-ACP1 (5'-CTGTGAATGCTGCGACTACGATXXXXX(T)<sub>18</sub>-3') and  
13 RNase-free water to a final volume of 9.5 µl. The mixture was then incubated at 80 °C for 3  
14 min, followed by chilling on ice for another 2 min. Subsequently, 4 µl of 5X RT buffer  
15 (Mbiotech), 5 µl of 2mM dNTP (Fermentas), 0.5 µl of 40 U/µl RNase inhibitor (Mbiotech)  
16 and 1 µl of 200 U/µl M-MLV reverse transcriptase (Mbiotech) were added. This mixture was  
17 then incubated at 42 °C for 90 min, heated at 94 °C for another 2 min and chilled on ice for 2  
18 min. Finally, 80 µl of DNase-free water was added to dilute the synthesised cDNA. The first-  
19 strand cDNA was stored under -20 °C until further analysis.

### 20 b) ACP-based GeneFishing PCR

21 First, all four cDNA samples within each CRC and control group samples were pooled  
22 together in equal amounts. The characterisation of DEGs was then conducted via ACP-based  
23 PCR based on 20 arbitrary ACP primers (Cat. No.: K1021) in a thermal cycler (Mastercycler  
24 Gradient, Eppendorf) according to the manufacturer's protocol (GeneFishing DEG Premix Kit,  
25 Seegene). Initially, the synthesis of second-strand cDNA was commenced in a one-cycle first-  
26 stage PCR: 94 °C for 5 min, 50 °C for 3 min and 72 °C for 1 min. Next, the constructed  
27 second-strand cDNA was subjected to second-stage PCR with 40 cycles of a denaturing step

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 at 94 °C for 40 sec, annealing step at 65 °C for 40 sec and extension step at 72 °C for 40 sec.  
11 2 Lastly, a final extension step at 72 °C for 5 min was carried out. The amplified products were  
12 3 then separated on 3 % (w/v) agarose gels stained with ethidium bromide.

14 c) Cloning and sequencing

15  
16  
17 5 The identified differentially expressed bands were extracted from the agarose gel by using the  
18 6 PureLink Quick Gel Extraction Kit (Invitrogen). Each of these extracted DNA fragments was  
19 7 then individually cloned with the use of the TOPO TA Cloning Kit for Sequencing  
20 8 (Invitrogen). Subsequently, the plasmid containing the inserted DNA fragment was extracted  
21 9 from clones of interest via PureLink Quick Plasmid Miniprep Kit (Invitrogen). The isolated  
22 10 cloned plasmids were then sequenced with the ABI 3730xl DNA Analyser (Applied  
23 11 Biosystems). Finally, all the sequences obtained were analysed and matched for similarities  
24 12 with reference to the BLAST programme under the NCBI database.

28  
29 **RT-qPCR analysis**

30  
31 a) Reverse transcription

32  
33 15 The total RNA isolated from 27 paired samples was reverse transcribed to first-strand cDNA,  
34 16 with the following protocol: 3 µg of total RNA was added with 2 µl of 0.5 µg/µl oligo(dT)<sub>12-18</sub>  
35 17 (Invitrogen) and RNase-free water to a final volume of 9.5 µl. The reaction mixture was then  
36 18 incubated at 80 °C for 3 min, followed by chilling on ice for another 2 min. Next, 4 µl of 5X  
37 19 first strand buffer (Invitrogen), 5 µl of 2mM dNTP (Fermentas), 0.5 µl of 40 U/µl RNaseOUT  
38 20 recombinant RNase inhibitor (Invitrogen) and 1 µl of 200 U/µl M-MLV reverse transcriptase  
39 21 (Invitrogen) were added to the mixture. Finally, the reaction mixture was incubated at 42 °C  
40 22 for 90 min, heated at 94 °C for another 2 min and chilled on ice for 2 min. The synthesised  
41 23 first-strand cDNA was stored under -20 °C until further usage.

42  
43  
44  
45  
46 b)  $\Delta\Delta C_T$  analysis

47  
48 25 The relative expression of identified DEGs in all paired colorectal tumours and control  
49 26 samples was determined via  $\Delta\Delta C_T$  method. The RT-qPCR was performed in a singleplex

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 reaction containing 50 ng first-strand cDNA under universal thermal cycling conditions with  
11 2 the ABI 7500 Fast Real-Time PCR System (Applied Biosystems). Both *ACTB* (Assay ID:  
12 3 Hs99999903\_m1) and *GAPDH* (Assay ID: Hs99999905\_m1) were used as reference genes  
13 4 and are commercially available as TaqMan Pre-designed Assays (Applied Biosystems). Prior  
14 5 to the analysis of gene expression, the amplification efficiency for all target and reference  
15 6 genes assays was measured by using the standard curve method with 2-log measurements.  
16 7 The amplification efficiency value of 90 – 110 % was acceptable (Applied Biosystems). In  
17 8 this relative quantification method, the  $2^{-\Delta\Delta C_t}$  values obtained represented the fold change in  
18 9 gene expression of the colorectal tumours, which was normalised with both reference genes,  
19 10 in relative to the calibrator (control sample)[9].

#### 11 c) Statistical analysis

12 | The difference in the expression level between **primary colorectal CRC** tumour and paired  
13 | non-cancerous tissues was analysed by using Real-Time StatMiner software (Integromics).  
14 | The distribution of the  $\Delta C_T$  values obtained for each DEGs within each CRC and control  
15 | group were tested for normality via the Shapiro-Wilk test. **Subsequently,** ~~t~~The paired t-test  
16 | was ~~then~~ performed to assess the statistical significance of the observed differential  
17 | expression patterns.

Formatted: Font color: Auto, Highlight

Formatted: Font color: Auto, Highlight

## 18 | **RESULTS**

### 19 | **DEGs between colorectal tumours and non-cancerous colonic tissues**

20 | This preliminary study was conducted on paired samples pooled from four patients with CRC  
21 | Stages I - III. In ACP-based GeneFishing PCR, 20 sets of arbitrary ACP primers were used to  
22 | randomly amplify gene products in both colorectal tumours and normal colonic samples.  
23 | Upon visualisation on agarose gels, a total of 13 **differentially expressed bands were observed**  
24 | **by means of comparing bands intensity between the tumouric and non-cancerous samples,** as  
25 | shown in Figure 1. These bands were further sequenced for gene identification, and 16 DEGs  
26 |

Formatted: Highlight



1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 were successfully reported. Among them, 13 were over-expressed in colorectal tumours,  
11 2 whilst three were under-expressed, as listed in Table 2.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

Table 2 Sequence similarities and identification of DEGs.

Differentially Expressed Band	DEG	Identity	Sequence Homology (%)	Accession Number	UniGene Number	Description
<i>Over-expressed</i>						
A4.1	DEG1	<i>Homo sapiens</i> proteasome (prosome, macropain) 26S subunit, ATPase, 5 (PSMC5), mRNA	502/506 (99%)	NM_002805.4	Hs.79387	Involves in the ATP-dependent degradation of ubiquitinated proteins.
	DEG2	<i>Homo sapiens</i> ubiquinol-cytochrome c reductase hinge protein (UQCRH), mRNA	514/521 (98%)	NM_006004.2	Hs.481571	A component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex, which is part of the mitochondrial respiratory chain.
A4.2	DEG3	<i>Homo sapiens</i> ribosomal protein S23 (RPS23), mRNA	551/551 (100%)	NM_001025.4	Hs.527193	A component of the 40S subunit of human ribosomes.
A6.1	DEG4	<i>Homo sapiens</i> ribosomal protein L10 (RPL10), transcript variant 1, mRNA	554/557 (99%)	NM_006013.3	Hs.534404	A component of the 60S subunit of human ribosomes.
A9.2	DEG6	<i>Homo sapiens</i> actin related protein 2/3 complex, subunit 2, 34kDa (ARPC2), transcript variant 2, mRNA	473/473 (100%)	NM_005731.2	Hs.529303	Involves in the regulation of actin polymerization as an actin-binding component of the Arp2/3 complex, and mediates the formation of branched actin networks together with an activating nucleation-promoting factor (NPF).
	DEG7	<i>Homo sapiens</i> TIMP metalloproteinase inhibitor 1 (TIMP1), mRNA	503/511 (98%)	NM_003254.2	Hs.522632	Irreversibly inactivates the metalloproteinases by binding to their catalytic zinc cofactor.
A10.1	DEG8	<i>Homo sapiens</i> ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, beta polypeptide (ATP5B), nuclear gene encoding mitochondrial protein, mRNA	917/919 (99%)	NM_001686.3	Hs.406510	A subunit of mitochondrial ATP synthase that catalyses the synthesis of ATP by utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation.

A13.2	DEG11	<i>Homo sapiens</i> chromosome 11 open reading frame 10 (C11orf10), mRNA	273/273 (100%)	NM_014206.3	Hs.437779	Unknown.
A13.3	DEG12	<i>Homo sapiens</i> mitochondrial ribosomal protein L24 (MRPL24), nuclear gene encoding mitochondrial protein, transcript variant 2, mRNA	408/411 (99%)	NM_024540.3	Hs.418233	Involves in protein synthesis within the mitochondrion.
A13.4	DEG13	<i>Homo sapiens</i> similar to OK/SW-CL.16 (LOC100288418)	635/644 (98%)	XM_002342023.1	-	Unknown.
A18.1	DEG14	<i>Homo sapiens</i> family with sequence similarity 96, member B (FAM96B), transcript variant 2, transcribed RNA	486/487 (99%)	NR_024525.1	Hs.9825	Involves in chromosome segregation as part of the mitotic spindle-associated MMXD complex.
A20.1	DEG15	<i>Homo sapiens</i> ribosomal protein L35 (RPL35), mRNA	440/446 (99%)	NM_007209.3	Hs.182825	A component of the 60S subunit of human ribosomes.
A20.2	DEG16	<i>Homo sapiens</i> chromosome 6 open reading frame173 (C6orf173), mRNA	551/554 (99%)	NM_001012507.2	Hs.486401	May be required for proper chromosome segregation during mitosis and involved with CENPT in the establishment of centromere chromatin structure.
<b><i>Under-expressed</i></b>						
A9.1	DEG5	<i>Homo sapiens</i> ribosomal protein L37 (RPL37), mRNA	284/284 (100%)	NM_000997.4	Hs.731513	A component of the 60S subunit of human ribosomes, and can bind to the 23S rRNA.
A13.1	DEG9	<i>Homo sapiens</i> solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1 (SLC25A1), nuclear gene encoding mitochondrial protein, mRNA	165/165 (100%)	NM_005984.2	Hs.111024	A mitochondrial tricarboxylate transporter which is responsible for the movement of citrate across the mitochondrial inner membrane.
	DEG10	<i>Homo sapiens</i> similar to cytochrome c oxidase subunit II (LOC100288578), miscRNA	141/146 (97%)	XR_078216.1	-	Unknown.

### 1 Differential ability of the identified DEGs on early and advanced colorectal neoplasia

2 Following the identification of DEGs, the gene sequences obtained were then used to design  
 3 primers and TaqMan probes for RT-qPCR analysis by Applied Biosystems, as listed in Table  
 4 3. In an attempt to assess the differential ability of identified DEGs on early and advanced  
 5 colorectal adenocarcinoma, the recruited paired samples were further stratified into two  
 6 groups according to the cancer stage. Among them, 13 patients with Stages I and II were  
 7 grouped as early stage CRC, whilst the advanced stage CRC group comprised of 14 patients  
 8 with Stages III and IV.

10 Table 3 Primers and TaqMan probes for relative quantification with Comparative C<sub>T</sub>  
 11 method.

DEG	Primers Sequence	TaqMan Probe Sequence
DEG1	Forward: 5'-GGGCGTGTGCACAGAAG-3' Reverse: 5'-AAGTCCTCCTGAGTGACATGGA-3'	5'-CTCGCAGGGCATAACAT-3'
DEG2	Forward: 5'-GATGCTTACCGAATCCGGAGATC-3' Reverse: 5'-GCATTGCTCTCTCACTGTTGTAG-3'	5'-CCTCTCCTCTCTCCTCCTCC-3'
DEG3	Forward: 5'-CAACCGTCATTGGGTACAAAGG-3' Reverse: 5'-TGTAAGGGTCCAGCTGATCAAGA-3'	5'-ATGGCAAGAAAATCAC-3'
DEG4	Forward: 5'-CGGCCAGGAACTTGAACCTG-3' Reverse: 5'-CCGAGCTGCAGAACAAGGA-3'	5'-CAGGGCCTCAATCACA-3'
DEG5	Forward: 5'-CTGGTCGAATGAGGCACCTAAAA-3' Reverse: 5'-TGGGTTTAGGTGTTGTTCCCTCAC-3'	5'-CATGCCTGAATCTGC-3'
DEG6	Forward: 5'-AGATTAGCGGGATGAAAACGTCTT-3' Reverse: 5'-CGCCAGATGCCGAGAAAA-3'	5'-CCCCGTGATTGTTTC-3'
DEG7	Forward: 5'-GGTAGTGATGTGCAAGAGTCCAT-3' Reverse: 5'-CCGCAGCGAGGAGTTTCT-3'	5'-CATTGCTGGAAAACCTG-3'
DEG8	Forward: 5'-GAAGGAGACCATCAAAGGATTCCA-3' Reverse: 5'-GAAGGCCTGTTCTGGGAGATG-3'	5'-ATTCACCTGCCAAAATC-3'
DEG9	Forward: 5'-GGCAGGGTGGTCCTGAGA-3' Reverse: 5'-CCGCCATTGGCCTTAACCTG-3'	5'-CCTCTCTCCGCCCGGACA-3'
DEG11	Forward: 5'-CAGGTTTCAGTGAAGCCATCTG-3' Reverse: 5'-GGGTTGGCATCTACGTGTA-3'	5'-CACCCAAGGGTAACAAC-3'
DEG12	Forward: 5'-CCAGGTCAAACCTGTGGATCCT-3' Reverse: 5'-GCTTCAGTAAATCTCCACTCGATCT-3'	5'-ATGGACAGGAAACCCAC-3'
DEG14	Forward: 5'-CCCGTCTTATCTGCAAGTT-3' Reverse: 5'-TCAAGATGGACGTGCACATTACTC-3'	5'-CATGCAGTGAACAAGC-3'
DEG15	Forward: 5'-CGGCCTCCAAGCTCTCT-3' Reverse: 5'-TGAGAACACGGGCAATGGATT-3'	5'-CCGGACGACTCGGATCT-3'
DEG16	Forward: 5'-GGACTCTCTGCTAATCGATGAACA-3' Reverse: 5'-GCCTCAACTTCGCTGGAGAAAA-3'	5'-CAGATGGACCAATAAGTCA-3'

1 The analysis of RT-qPCR results was performed via Real-Time StatMiner software by  
 2 importing the raw Ct data. The within-group correlation of these  $\Delta C_T$  values was then  
 3 determined by calculating the MAD for all the samples within the same experimental group.  
 4 The biological samples which do not correlate well with other samples in the same group,  
 5 were detected as group outliers and excluded from subsequent analysis. Both *ACTB* and  
 6 *GAPDH* were used for normalisation in computing the  $\Delta C_T$  (Figure 2) and  $2^{-\Delta\Delta C_T}$  values by  
 7 using the following formulas (Table 4).

$$C_{T(\text{Target gene})} - C_{T(\text{Reference gene})} = \Delta C_T$$

$$\Delta C_T(\text{Sample}) - \Delta C_T(\text{Calibrator}) = \Delta\Delta C_T$$

$$\text{Relative Fold Change in Expression (RQ)} = 2^{-\Delta\Delta C_T}$$

12 The relative fold change in the mRNA expression level between the colorectal tumours and  
 13 adjacent normal colonic mucosa were shown as the  $2^{-\Delta\Delta C_T}$  values. —The statistical  
 14 significance of the observed fold change in expression was determined by paired t-test for all  
 15 the DEGs. A *p* value of less than 0.05 is considered as statistically significant (Table 4).

16 In both early and advanced stage CRC groups, the expression of four out of 16 DEGs was  
 17 reported to be significantly differed between tumouric and non-cancerous tissues.  
 18 Remarkably, the combination of this panel of four genes is different among two groups. The  
 19 *RPL35*, *RPS23* and *TIMP1* genes were found to be over-expressed in both early- and  
 20 advanced colorectal neoplasms ( $p < 0.05$ ) (Figures 3 and 4). It is interesting to note that, the  
 21 under-expression of *ARPC2* gene ( $p < 0.05$ ) was only observed in early stage colorectal  
 22 tumours (Figure 3). On the other hand, the *C6orf173* gene was found to be over-expressed ( $p$   
 23  $< 0.05$ ) in advanced colorectal adenocarcinomas, but not in early stage colorectal tumours  
 24 (Figure 4).

Formatted: Highlight

Formatted: Highlight

Formatted: Centered

Formatted: Highlight

Formatted: Centered

Formatted: Not Superscript/ Subscript

Formatted: Highlight

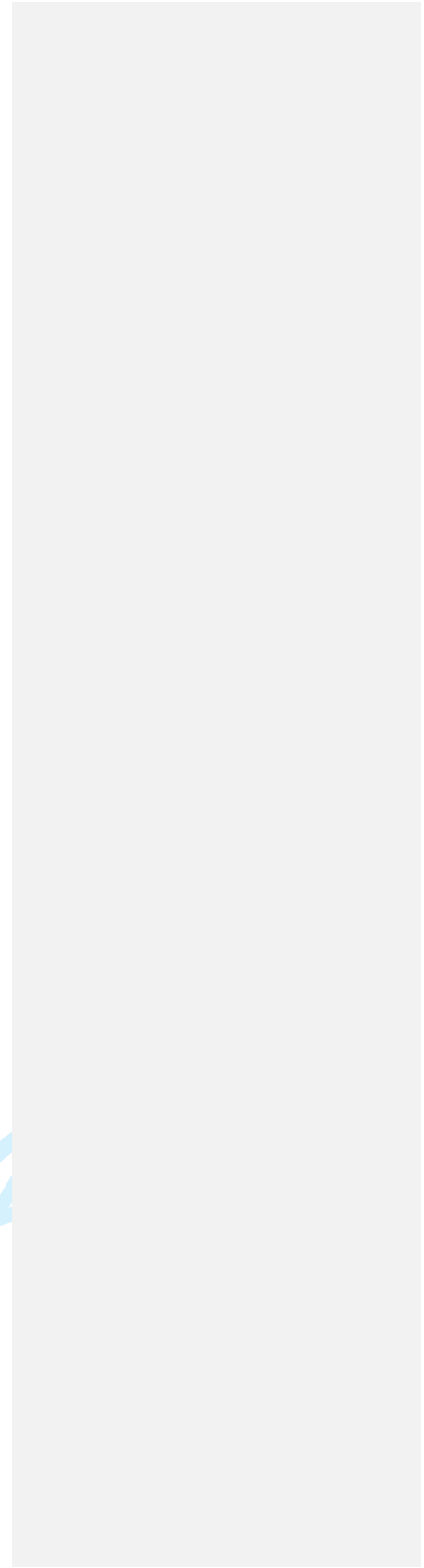
Field Code Changed

Formatted: Highlight

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1

For peer review only



16

Table 4  $\Delta C_T$  mean,  $\Delta\Delta C_T$ ,  $2^{-\Delta\Delta C_T}$  and  $p$  values for all the DEGs in both early- and advanced stage CRC groups.

DEG	Early Stage CRC					Advanced Stage CRC				
	$\Delta C_T$ Mean (CRC)	$\Delta C_T$ Mean (Normal)	$\Delta\Delta C_T$	$2^{-\Delta\Delta C_T}$	$p$ value	$\Delta C_T$ Mean (CRC)	$\Delta C_T$ Mean (Normal)	$\Delta\Delta C_T$	$2^{-\Delta\Delta C_T}$	$p$ value
<i>ARPC2</i>	2.6854	2.0664	0.6190	0.6511	0.0282*	2.7240	2.3300	0.3940	0.7610	0.2424
<i>ATP5B</i>	1.5846	1.2702	0.3144	0.8042	0.3524	1.9558	1.3838	0.5720	0.6727	0.1484
<i>C11orf10</i>	3.2897	3.3639	-0.0742	1.0528	0.8333	3.3281	3.6709	-0.3428	1.2682	0.3710
<i>C6orf173</i>	6.1083	7.1943	-1.0860	2.1228	0.0905	5.9949	7.9087	-1.9138	3.7680	0.0013*
<i>FAM96B</i>	3.5602	3.8955	-0.3353	1.2616	0.2935	3.5276	3.9920	-0.4644	1.3797	0.2113
<i>MRPL24</i>	4.9171	5.0839	-0.1668	1.1226	0.3564	4.9728	5.1467	-0.1739	1.1281	0.7001
<i>PSMC5</i>	3.8232	3.9617	-0.1385	1.1008	0.6812	3.7705	3.8455	-0.0750	1.0534	0.8048
<i>RPL10</i>	-0.7462	-0.4853	-0.2609	1.1982	0.4001	-1.1576	-0.5196	-0.6380	1.5562	0.0950
<i>RPL35</i>	-0.1926	0.6222	-0.8148	1.7591	0.0024*	0.1748	0.8769	-0.7021	1.6269	0.0372*
<i>RPL37</i>	-0.0059	-0.1539	0.1480	0.9025	0.8645	0.2184	0.7143	-0.4959	1.4102	0.1537
<i>RPS23</i>	0.2176	0.7739	-0.5563	1.4705	0.0310*	0.0676	0.9431	-0.8755	1.8346	0.0250*
<i>SLC25A1</i>	3.7514	3.5430	0.2084	0.8655	0.5721	3.5565	3.4428	0.1137	0.9242	0.7991
<i>TIMP1</i>	2.9096	4.3059	-1.3963	2.6323	0.0440*	2.3330	3.8547	-1.5217	2.8713	0.0062*
<i>UQCRH</i>	2.0087	2.2216	-0.2129	1.1590	0.4108	2.3375	2.4459	-0.1084	1.0780	0.7808

\* $p < 0.05$  = statistically significant

## DISCUSSION

Our current study has revealed two distinctive 4-gene signatures for both early- and advanced stage colorectal adenocarcinomas. The early stage sporadic CRC was characterised by the over-expression of *RPL35*, *RPS23* and *TIMP1* genes, as well as under-expression of *ARPC2* gene. On the other hand, the advanced primary colorectal tumours were reported with over-expression of *C6orf173*, *RPL35*, *RPS23* and *TIMP1* genes. Although the relative fold change for *ARPC2*, *RPL35* and *RPS23* genes is below 2, the individual result does not affect the analysis since gene expression patterns of all four genes in combination were proposed to distinguish between the early- and advanced stage colorectal neoplasms. The potential involvement of these DEGs and their altered expression levels in CRC were further supported by previous researches.

In fact, several proto-oncogenes and tumour suppressors are previously reported to regulate the ribosome production, i.e., the *RB*[10], *TP53*[11], *PTEN* genes[12], as well as the *MYC* gene family[13]. It is suggested that the alterations in ribosome biogenesis might affect the translation of genes that are involved in neoplastic transformation. In addition, the additional extra-ribosomal functions of the ribosomal proteins (r-proteins) in cellular apoptosis, cellular proliferation, cellular transformation, genes transcription, mRNA translation, DNA repair and inflammation, might also trigger and support the neoplastic development[14]. Hence, the over-expression of r-proteins-encoding genes observed in colorectal adenocarcinomas is not unexpected[15-17]. Our current study has revealed the significant over-expression of two r-proteins that were not previously described in colorectal tumours, i.e., the *RPL35* and *RPS23*. The observed fold changes for the *RPL35* and *RPS23* mRNA levels were comparable between the early- and advanced stage colorectal tumours in our sample cohort. This was in agreement with previous reports by Barnard et al. and Frigerio et al., where the changes in the mRNA expression levels of the r-proteins were irrespective of the cancer stage[18,19]. The hypothesis that the same ribosomal protein may contribute in different stages of cancer progression with their hitherto unknown extra-ribosomal roles might provide an explanation to these observations[20].



1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 On the other hand, our present study also demonstrated an over-expression of the *TIMP1* gene  
11 2 in both early- and advanced stage primary colorectal tumours. This finding is supported by  
12 3 Zeng et al., where the over-expression of TIMP1 was reported in all stages of primary  
13 4 colorectal tumours[21]. Under normal physiological conditions, the proteolytic activities of  
14 5 MMPs are kept at bay by their natural inhibitors, the TIMPs[22]. Previous studies have  
15 6 reported the over-expression of MMPs in both early- and advanced stage colorectal tumours,  
16 7 as well as other cancer types[23-25], which is in accordance to their biological roles. Hence, a  
17 8 similar scenario is expected for TIMPs and indeed, their suppressive role in tumour invasion  
18 9 and metastasis has been demonstrated in various cancer models[26]. However, more recent  
19 10 studies have revealed a direct correlation between TIMP1 expression and tumour  
20 11 aggressiveness in cancer, including CRC[21,27]. These findings, which are contradictory to  
21 12 its protease-inhibiting function, have suggested a possible tumour-promoting role of TIMP1 in  
22 13 tumorigenesis. It is postulated that the TIMP1 exhibits the abilities to inhibit tumour cell  
23 14 apoptosis and promote tumour angiogenesis, as well as other growth-factor-like effects[28].  
24 15 In our present study, the observed comparable over-expression of TIMP1 in both early- and  
25 16 advanced stage sporadic colorectal neoplasms was in line with its MMP inhibitory and MMP-  
26 17 independent tumour-promoting activities.

27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
18 In cancer biology, the expression of mRNAs and proteins of the ARP2/3 complex is often  
19 studied due to its role in cell migration, which contributes to cancer invasion and metastasis if  
20 aberrantly regulated[29]. We have detected a significant under-expression of ARPC2 in our  
21 cohort of early stage primary colorectal tumours. Surprisingly, this finding is contradictory  
22 with the role played by ARPC2 in cancer invasion and metastasis theoretically. Previously,  
23 Kaneda et al. has reported the decreased expression of all the seven genes encoding the  
24 subunits of ARP2/3 complex in human gastric cancers. Among them, the Arp2, ARPC2 and  
25 ARPC3 showed the most prominent reduction in their expression levels[30]. The exact  
26 mechanism underlying this observation still remains unknown, but the epigenetic alteration  
27 might potentially provide an explanation for it. For instance, promoter hypermethylation that  
28 causes gene silencing is responsible for the reduced expression of ARPC1 in human gastric

1 cancer[31]. Similarly, the epigenetic study might also offer a clue for the under-expression of  
2 ARPC2 in colorectal neoplasms.

3 *C6orf173*, which is also known as *CUG2* or *CENP-W*, is a novel oncogene that has been  
4 found to be up-regulated in many human cancer tissues. Its high expression level is  
5 profoundly reported in tumours of the ovary, liver, lung, pancreas, breast, colon, rectum and  
6 stomach. The CENP-W is a new member of the constitutive centromere-associated network,  
7 which specifically interacts with the CENP-T and plays an important role in mitosis[32]. In  
8 our current study, the CENP-W is over-expressed in advanced colorectal adenocarcinoma.  
9 This finding correlates to its function in kinetochore assembly, where its aberrant expression  
10 might lead to abnormal cell division and aneuploidy in cancer[32]. In our study, the over-  
11 expression of CENP-W was observed in both early- and advanced cohort of colorectal  
12 neoplasms but only statistically significant in the latter group. Given the fact that aneuploidy  
13 is constantly associated with a greater proportion of advanced CRC cases, the aberrant  
14 expression of CENP-W might potentially relate to a poorer prognosis of CRC[33].

15 In conclusion, we have characterised two distinctive gene expression patterns, which comprise  
16 of the *ARPC2*, *C6orf173*, *RPL35*, *RPS23* and *TIMPI* genes, for the stratification of primary  
17 colorectal adenocarcinomas among Malaysian CRC patients. It was postulated that the actin  
18 cytoskeleton might play an important role in determining the dysplastic cell morphology  
19 during the early development of CRC, while the aberrations in the assembly of functional  
20 kinetochore might be crucial for the aneuploidy of the advanced stage colorectal tumours.  
21 Nevertheless, the findings of this study were considered preliminary owing to the relatively  
22 small sample size. The main reason for this is Our current sample size was relatively small  
23 owing to the lack of a designated Tissue Bank in our institution. There were also not many  
24 CRC patient volunteers—Moreover, the lack of CRC patient volunteers and our stringent  
25 criteria for patient selection have also limited the availability of suitable specimens within the  
26 short sample collection period.

27 **However,** our identified mRNA expression patterns specific for early- and advanced stage  
28 colorectal tumours are still convincing with our stringent sample selection criteria, high

20

Formatted: Highlight

Formatted: Font color: Auto, Highlight

Formatted: Highlight

1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 specificity primers and probes, as well as reliable statistical analysis. In future, the validation  
11 2 of these DEGs should be performed on a larger set of clinical samples, and extensive inter-  
12 3 laboratory testing of their differential abilities on each CRC stage is also desired. In addition,  
13 4 we should also integrate other imaging and histological information to complement our  
14 5 identified gene expression patterns, which then hold promises for better stratification of  
15 6 tumours.  
16  
17  
18

19  
20 7 |  
21 8 **Funding** – This study was supported by FS176/2007C, PS172/2008C and Research  
22 9 Collaborative Grant, CG041-2013 from the University of Malaya.  
23

24 10 |  
25 11 **Competing interests** – None.  
26

27 12 |  
28 13 **Data sharing statement** – There is no additional data available.  
29

30 14  
31 15 **Contributorship statement:**

32 16 KHC, KLG, IH, HCC and ACR had the original idea for this work and gained funding in  
33 17 collaboration with PCL. TPL carried out the experiment. TPL, CKH, PCL, HCC and LHL  
34 18 were involved in the data analysis. TPL wrote the first draft of this paper and all authors  
35 19 subsequently assisted in redrafting and have approved the final version.  
36 20

37  
38 21 **REFERENCES**

- 39  
40 22 1 Greene FL, Page D, Fleming ID, et al., eds. AJCC cancer staging manual (6th ed.).  
41 42 New York: Springer 2002.  
43  
44 24 2. Dukes CE. The classification of cancer of the rectum. J Pathol Bacteriol 1932; 35:  
45 46 323-32.  
47  
48 26 3. Astler VB, Collier FA. The prognostic significance of direct extension of carcinoma of  
49 50 the colon and rectum. Ann Surg 1954; 139: 846-52.  
51 52

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 4. Edge SB, Byrd DR, Compton CC, et al. AJCC cancer staging manual (7th ed.). New  
11 York: Springer 2010.  
12 2  
13  
14 3 5. Hu HK, Krasinskas A, Willis J. Perspectives on current tumour-node-metastasis  
15 (TNM) staging of cancers of the colon and rectum. *Semin Oncol* 2011; 38: 500-10.  
16 4  
17  
18 5 6. Doyle VJ, Bateman AC. Colorectal cancer staging using TNM 7: is it time to use this  
19 new staging system? *J Clin Pathol* 2012; 65: 372-4.  
20 6  
21  
22  
23 7 7. Russo G, Zegar C, Giordano A. Advantages and limitations of microarray technology  
24 in human cancer. *Oncogene* 2003; 22: 6497-507.  
25 8  
26  
27  
28 9 8. Puppa G, Sonzogni A, Colombari R, et al. TNM staging system of colorectal  
29 carcinoma: a critical appraisal of challenging issues. *Arch Pathol Lab Med* 2010; 134:  
30 837-52.  
31 11  
32  
33  
34 12 9. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time  
35 quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 2001; 25: 402-8.  
36 13  
37  
38  
39 14 10. Voit R, Schafer K, Grummt I. Mechanism of repression of RNA polymerase I  
40 transcription by the retinoblastoma protein. *Mol Cell Biol* 1997; 17: 4230-7.  
41 15  
42  
43 16 11. Zhai W, Cornai L. Repression of RNA polymerase I transcription by the tumour  
44 suppressor p53. *Mol Cell Biol* 2000; 20: 5930-8.  
45 17  
46  
47  
48 18 12. Backman S, Stambolic V, Mak T. PTEN function in mammalian cell size regulation.  
49 *Curr Opin Neurobiol* 2002; 12: 516-22.  
50 19

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 13. Greasley PJ, Bonnard C, Amati B. Myc induces the nucleolin and BN51 genes:  
11 possible implications in ribosome biogenesis. *Nucleic Acids Res* 2000; 28: 446-53.  
12  
13  
14 3 14. Montanaro L, Treré D, Derenzini M. Nucleolus, ribosomes, and cancer. *Am J Pathol*  
15 2008; 173: 301-10.  
16  
17  
18 5 15. Sharp MG, Adams SM, Elvin P, et al. A sequence previously identified as metastasis-  
19 related encodes an acidic ribosomal phosphoprotein, P2. *Br J Cancer* 1990; 61: 83-8.  
20  
21  
22  
23 7 16. Chester KA, Robson L, Begent RH, et al. Identification of a human ribosomal protein  
24 mRNA with increased expression in colorectal tumours. *Biochim Biophys Acta* 1989;  
25 1009: 297-300.  
26  
27  
28  
29 10 17. Pogue-Geile K, Geiser JR, Shu M, et al. Ribosomal protein genes are overexpressed in  
30 colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein.  
31 *Mol Cell Biol* 1991; 11: 3842-9.  
32  
33  
34  
35  
36 13 18. Barnard GF, Staniunas RJ, Mori M, et al. Gastric and hepatocellular carcinomas do not  
37 overexpress the same ribosomal protein messenger RNAs as colonic carcinoma.  
38 *Cancer Res* 1993; 53: 4048-52.  
39  
40  
41  
42 16 19. Frigerio JM, Dagorn JC, Iovanna JL. Cloning, sequencing and expression of the L5,  
43 L21, L27a, L28, S5, S9, S10 and S29 human ribosomal protein mRNAs. *Biochim*  
44 *Biophys Acta* 1995; 1262: 64-8.  
45  
46  
47  
48 19 20. Lai MD, Xu J. Ribosomal proteins and colorectal cancer. *Curr Genomics* 2007; 8: 43-9.  
49  
50

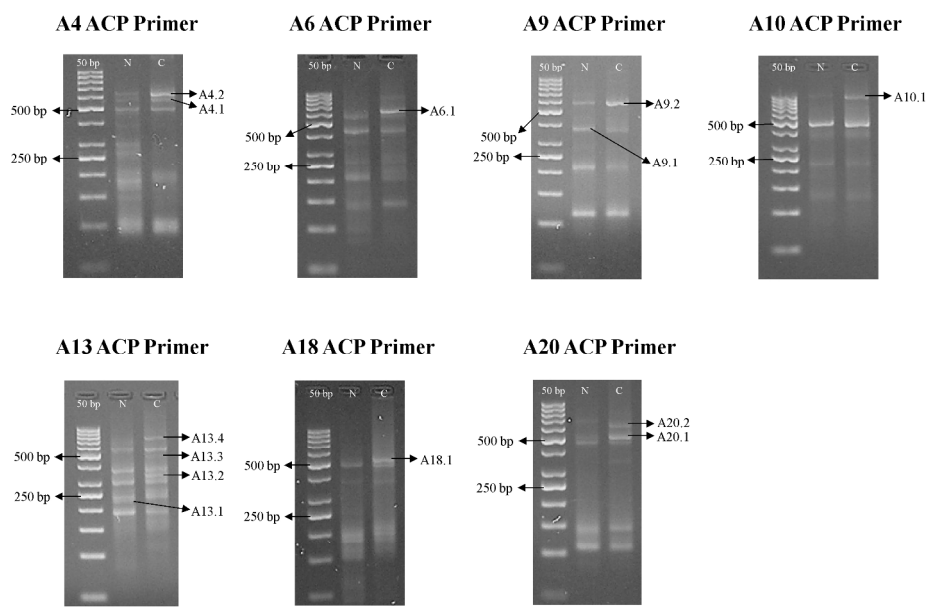
- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 21. Zeng ZS, Cohen AM, Zhang ZF, et al. Elevated tissue inhibitor of metalloproteinase 1  
11 RNA in colorectal cancer stroma correlates with lymph node and distant metastases.  
12 Clin Cancer Res 1995; 1: 899-906.  
13  
14 3  
15  
16 4 22. Ennis BW, Matrisian LM. Matrix degrading metalloproteinases. J Neurooncol 1994;  
17 18: 105-9.  
18  
19  
20 6 23. Urbanski SJ, Edwards DR, Maitland A, et al. Expression of metalloproteinases and  
21 their inhibitors in primary pulmonary carcinomas. Br J Cancer 1992; 66: 1188-94.  
22  
23 7  
24  
25 8 24. Boag AH, Young ID. Immunohistochemical analysis of type IV collagenase  
26 expression in prostatic hyperplasia and adenocarcinoma. Mod Pathol 1993; 6: 65-8.  
27  
28 9  
29 10 25. Newell KJ, Witty JP, Rodgers WH, et al. Expression and localisation of matrix-  
30 degrading metalloproteinases during colorectal tumourigenesis. Mol Carcinogen 1994;  
31 10: 199-206.  
32  
33 12  
34  
35 13 26. Khokha R, Waterhouse P. The role of tissue inhibitor of metalloproteinase-1 in  
36 specific aspects of cancer progression and reproduction. J Neurooncol 1994; 18: 123-7.  
37  
38 14  
39  
40 15 27. Lu XQ, Levy M, Weinstein IB, et al. Immunological quantitation of levels of tissue  
41 inhibitor of metalloproteinase-1 in human colon cancer. Cancer Res 1991; 51: 6231-5.  
42  
43 16  
44  
45 17 28. Egeblad M, Werb Z. New functions for the matrix metalloproteinases in cancer  
46 progression. Nat Rev Cancer 2002; 2: 161-74.  
47  
48 18  
49 19 29. Yamaguchi H, Wyckoff J, Condeelis J. Cell migration in tumours. Curr Opin Cell Biol  
50 2005; 17: 559-64.  
51 20

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10 1 30. Kaneda A, Kaminishi M, Sugimura T, et al. Decreased expression of the seven  
11 ARP2/3 complex genes in human gastric cancers. *Cancer Lett* 2004; 212: 203-10.  
12  
13  
14 3 31. Kaneda A, Kaminishi M, Nakanishi Y, et al. Reduced expression of the insulin-  
15 induced protein 1 and p41 ARP2/3 complex genes in human gastric cancers. *Int J*  
16 4 *Cancer* 2002; 100: 57-62.  
17  
18 5  
19  
20 6 32. Hori T, Amano M, Suzuki A, et al. CCAN makes multiple contacts with centromeric  
21 DNA to provide distinct pathways to the outer kinetochore. *Cell* 2008; 135: 1039-52.  
22  
23 7  
24  
25 8 33. Chen HS, Sheen-Chen SM, Lu CC. DNA index and S-phase fraction in curative  
26 resection of colorectal adenocarcinoma: analysis of prognosis and current trends.  
27 9  
28  
29 10  
30  
31  
32 11  
33

## 34 **Figure Legends**

- 35  
36 13 Figure 1 Differential banding patterns on 3 % agarose gel post ACP-based PCR  
37 14 amplification between normal colon and colorectal tumour samples (N: normal  
38 15 sample; C: CRC sample)  
39  
40 16 Figure 2 Box-plots showing  $\Delta C_T$  values of all colorectal tumours and normal colonic  
41 17 tissues in each early- (a) and advanced (b) stage CRC group.  
42  
43 18 Figure 3 Differential expression patterns of all the identified DEGs in early stage CRC  
44 19 group.  
45  
46 20 Figure 4 Differential expression patterns of all the identified DEGs in advanced stage  
47 21 CRC group.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

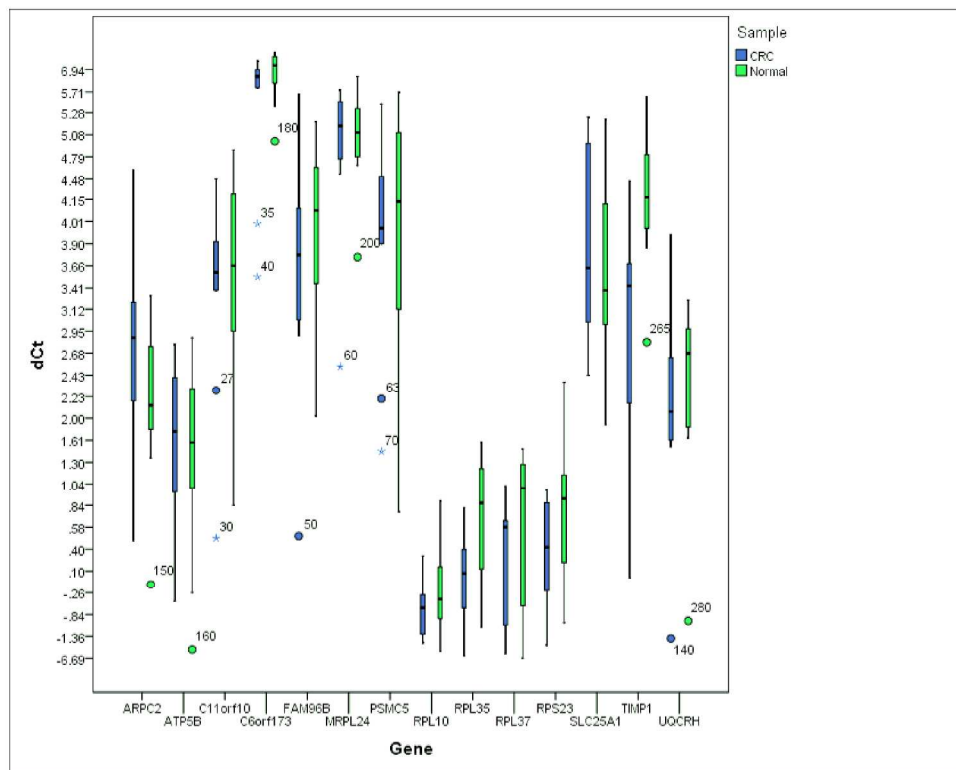


265x175mm (300 x 300 DPI)

Review only



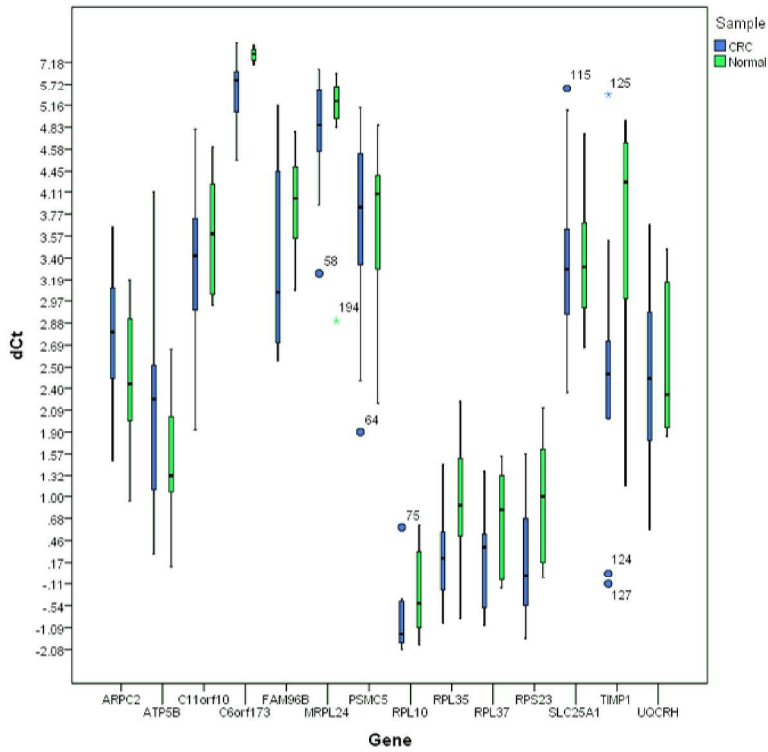
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



260x203mm (300 x 300 DPI)

ew only

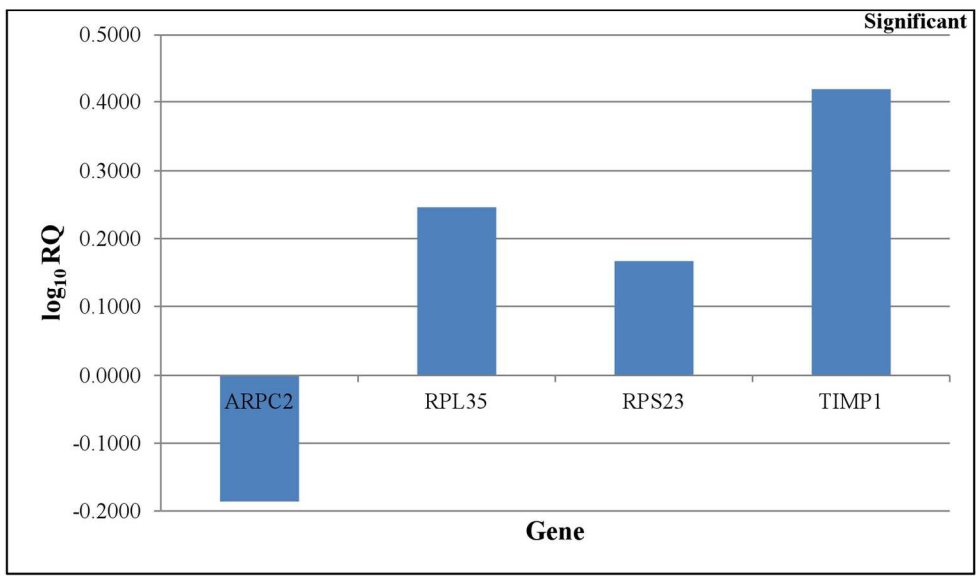
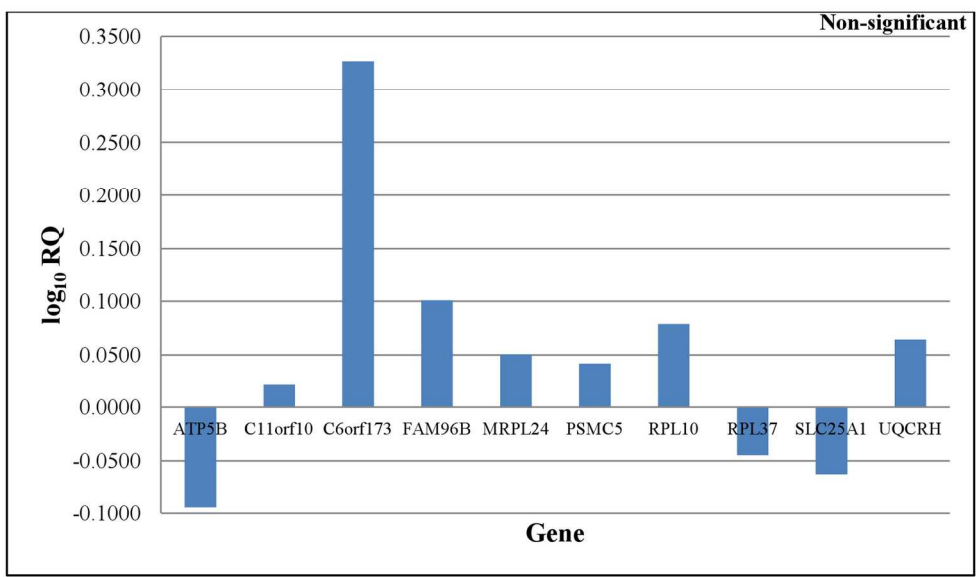
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



260x203mm (300 x 300 DPI)

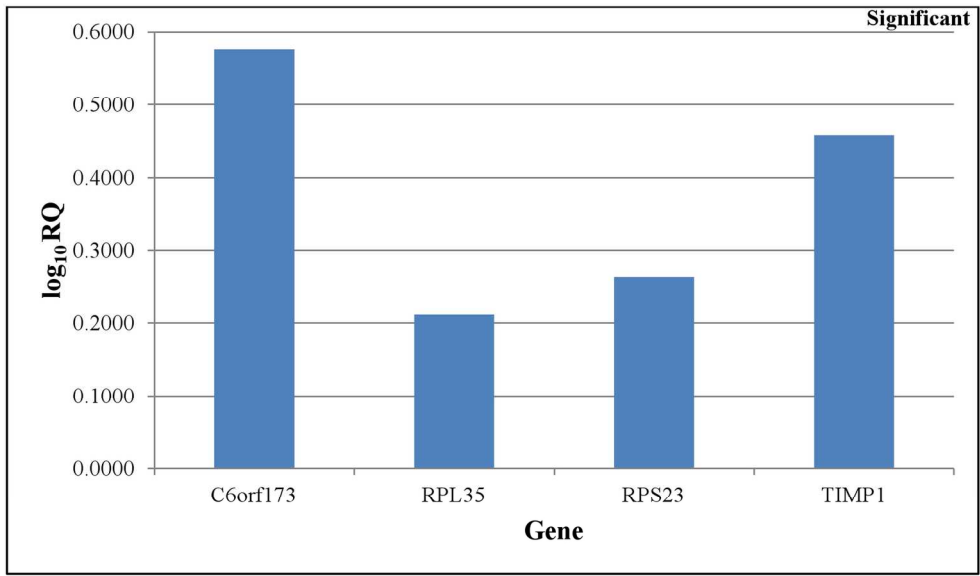
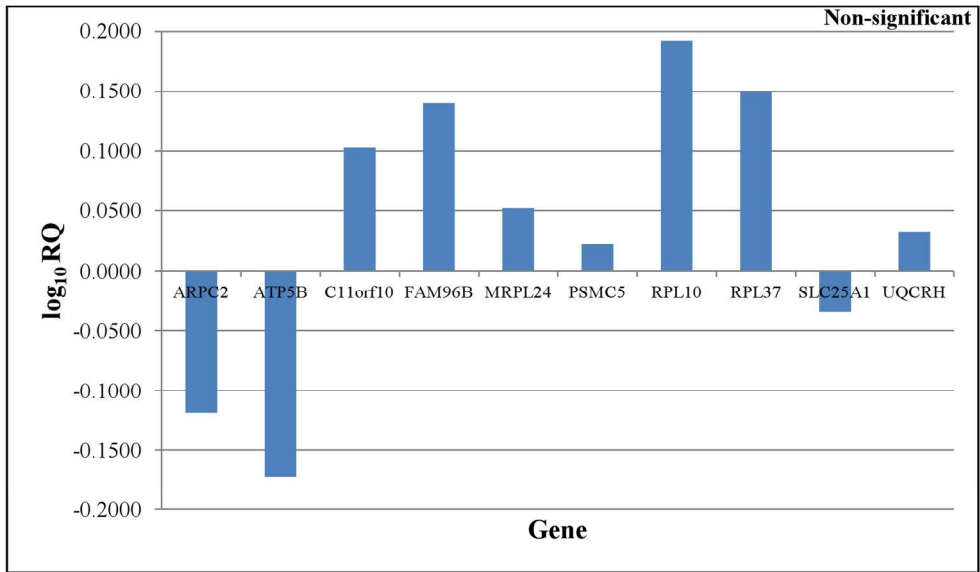
ew only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



186x221mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



187x224mm (300 x 300 DPI)

## STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
<b>Title and abstract</b>	1	<p>✓ (a) Indicate the study's design with a commonly used term in the title or the abstract</p> <p>✓ (b) Provide in the abstract an informative and balanced summary of what was done and what was found</p>
<b>Introduction</b>		
Background/rationale	2	✓ Explain the scientific background and rationale for the investigation being reported
Objectives	3	✓ State specific objectives, including any prespecified hypotheses
<b>Methods</b>		
Study design	4	✓ Present key elements of study design early in the paper
Setting	5	✓ Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	<p>✓ (a) <i>Cohort study</i>—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up</p> <p>✓ <i>Case-control study</i>—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls</p> <p><i>Cross-sectional study</i>—Give the eligibility criteria, and the sources and methods of selection of participants</p> <p>✓ (b) <i>Cohort study</i>—For matched studies, give matching criteria and number of exposed and unexposed</p> <p>✓ <i>Case-control study</i>—For matched studies, give matching criteria and the number of controls per case</p>
Variables	7	✓ Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	✓ For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	✓ Describe any efforts to address potential sources of bias
Study size	10	✓ Explain how the study size was arrived at
Quantitative variables	11	✓ Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	<p>✓ (a) Describe all statistical methods, including those used to control for confounding</p> <p>✓ (b) Describe any methods used to examine subgroups and interactions</p> <p>✓ (c) Explain how missing data were addressed</p> <p>✓ (d) <i>Cohort study</i>—If applicable, explain how loss to follow-up was addressed</p> <p>✓ <i>Case-control study</i>—If applicable, explain how matching of cases and controls was addressed</p> <p><i>Cross-sectional study</i>—If applicable, describe analytical methods taking account of sampling strategy</p> <p>✓ (e) Describe any sensitivity analyses</p>

Continued on next page

**Results**

Participants	13*	<p>√ (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed</p> <p>√ (b) Give reasons for non-participation at each stage</p> <p>(c) Consider use of a flow diagram</p>
Descriptive data	14*	<p>√ (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders</p> <p>(b) Indicate number of participants with missing data for each variable of interest</p> <p>(c) <i>Cohort study</i>—Summarise follow-up time (eg, average and total amount)</p>
Outcome data	15*	<p><i>Cohort study</i>—Report numbers of outcome events or summary measures over time</p> <p>√ <i>Case-control study</i>—Report numbers in each exposure category, or summary measures of exposure</p> <p><i>Cross-sectional study</i>—Report numbers of outcome events or summary measures</p>
Main results	16	<p>√ (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included</p> <p>(b) Report category boundaries when continuous variables were categorized</p> <p>(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period</p>
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses

**Discussion**

Key results	18	√ Summarise key results with reference to study objectives
Limitations	19	√ Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	√ Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results

**Other information**

Funding	22	√ Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based
---------	----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------

\*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at [www.strobe-statement.org](http://www.strobe-statement.org).