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

| | |
|---|---|
| Journal: | *BMJ Open* |
| 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<br>Roslani, April Camilla; University of Malaya, Surgery<br>Lian, Lay Hoong<br>Lee, Ping Chin; Universiti Sabah Malaysia, School of Science and Technology<br>Hilmi, Ida; University of Malaya, Medicine<br>Goh, Khean Lee; University of Malaya, Medicine<br>Chua, Kek Heng; University of Malaya, Biomedical Science |
| Primary Subject Heading: | Gastroenterology and hepatology |
| Secondary Subject Heading: | Gastroenterology and hepatology |
| Keywords: | Gastrointestinal tumours < GASTROENTEROLOGY, Cancer genetics < GENETICS, PUBLIC HEALTH |
| | |

SCHOLARONE™
Manuscripts

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

Tze Pheng Lau[1], April Camilla Roslani[2], Lay Hoong Lian[3], Ping Chin Lee[4], Ida Hilmi[5], Khean Lee Goh[5], Kek Heng Chua[1]*

[1]*Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*
[2]*Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*
[3]*Casa Vista, Jalan Penaga, 59100 Kuala Lumpur, Malaysia.*
[4]*School of Science and Technology, Universiti Sabah Malaysia, 88400 Kota Kinabalu, Sabah, Malaysia.*
[5]*Division of Gastroenterology and Hepatology, Department of Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

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

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

*Phone no:* 603-79676607

*Fax no:* 603-79676600

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

**Word count:** 3056 words



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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.




For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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:



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 7th 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 7th 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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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 RNA*later* RNA Stabilization Reagent (Qiagen) immediately after excision and stored at -80 $^{0}$C.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 |

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**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$(T)_{18}$-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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

c) Cloning and sequencing

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

**RT-qPCR analysis**

a) Reverse transcription

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

b) $\Delta\Delta C_T$ analysis

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 CRC 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 CRC 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 CRC 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 CRC 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 CRC tumours, whilst three were under-expressed, as listed in Table 2.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 2 Sequence similarities and identification of DEGs.

| Differentially Expressed Band | DEG | Identity | Sequence Homology (%) | Accession Number | UniGene Number | Description |
|---|---|---|---|---|---|---|
| ***Over-expressed*** | | | | | | |
| A4.1 | DEG1 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* TIMP metallopeptidase 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 | *Homo sapiens* ATP synthase, H+ 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 |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | phosphorylation. |
| A13.2 | DEG11 | *Homo sapiens* chromosome 11 open reading frame 10 (C11orf10), mRNA | 273/273 (100%) | NM_014206.3 | Hs.437779 | Unknown. |
| A13.3 | DEG12 | *Homo sapiens* 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 | *Homo sapiens* similar to OK/SW-CL.16 (LOC100288418) | 635/644 (98%) | XM_002342023.1 | - | Unknown. |
| A18.1 | DEG14 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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. |
| ***Under-expressed*** | | | | | | |
| A9.1 | DEG5 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* similar to cytochrome c oxidase subunit II (LOC100288578), miscRNA | 141/146 (97%) | XR_078216.1 | - | Unknown. |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**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' | 5'-CTCGCAGGGCATACAT-3' |
| | Reverse: | 5'-AAGTCCTCCTGAGTGACATGGA-3' | |
| DEG2 | Forward: | 5'-GATGCTTACCGAATCCGGAGATC-3' | 5'-CCTCTTCCTCTTCCTCCTCC-3' |
| | Reverse: | 5'-GCATTGCTCTCTCACTGTTGTTAG-3' | |
| DEG3 | Forward: | 5'-CAACCGTCATTGGGTACAAAGG-3' | 5'-ATGGCAAGAAAATCAC-3' |
| | Reverse: | 5'-TGTAAGGGTCCAGCTGATCAAGA-3' | |
| DEG4 | Forward: | 5'-CGGCCAGGAAACTTGAACTTG-3' | 5'-CAGGGCCTCAATCACA-3' |
| | Reverse: | 5'-CCGAGCTGCAGAACAAGGA-3' | |
| DEG5 | Forward: | 5'-CTGGTCGAATGAGGCACCTAAAA-3' | 5'-CATGCCTGAATCTGC-3' |
| | Reverse: | 5'-TGGGTTTAGGTGTTGTTCCTTCAC-3' | |
| DEG6 | Forward: | 5'-AGATTAGCGGGATGAAAACGTCTT-3' | 5'-CCCCGTGATTGTTTTC-3' |
| | Reverse: | 5'-CGCCCAGATGCCGAGAAAA-3' | |
| DEG7 | Forward: | 5'-GGTAGTGATGTGCAAGAGTCCAT-3' | 5'-CATTGCTGGAAAACTG-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'-CCAGGTCAAACTTGTGGATCCT-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' | |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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).

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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].



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

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

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

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

**Competing interests –** None.

**Data sharing statement –** There is no additional data available.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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.




For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

9. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 2001; 25: 402-8.

10. Voit R, Schafer K, Grummt I. Mechanism of repression of RNA polymerase I transcription by the retinoblastoma protein. Mol Cell Biol 1997; 17: 4230-7.

11. Zhai W, Cornai L. Repression of RNA polymerase I transcription by the tumour suppressor p53. Mol Cell Biol 2000; 20: 5930-8.

12. Backman S, Stambolic V, Mak T. PTEN function in mammalian cell size regulation. Curr Opin Neurobiol 2002; 12: 516-22.

13. Greasley PJ, Bonnard C, Amati B. Myc induces the nucleolin and BN51 genes: possible implications in ribosome biogenesis. Nucleic Acids Res 2000; 28: 446-53.

14. Montanaro L, Treré D, Derenzini M. Nucleolus, ribosomes, and cancer. Am J Pathol 2008; 173: 301-10.

15. Sharp MG, Adams SM, Elvin P, et al. A sequence previously identified as metastasis-related encodes an acidic ribosomal phosphoprotein, P2. Br J Cancer 1990; 61: 83-8.

16. Chester KA, Robson L, Begent RH, et al. Identification of a human ribosomal protein mRNA with increased expression in colorectal tumours. Biochim Biophys Acta 1989; 1009: 297-300.

17. Pogue-Geile K, Geiser JR, Shu M, et al. Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein. Mol Cell Biol 1991; 11: 3842-9.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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)



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Figure 2 Box-plots showing $\Delta C_T$ values of all CRC 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.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

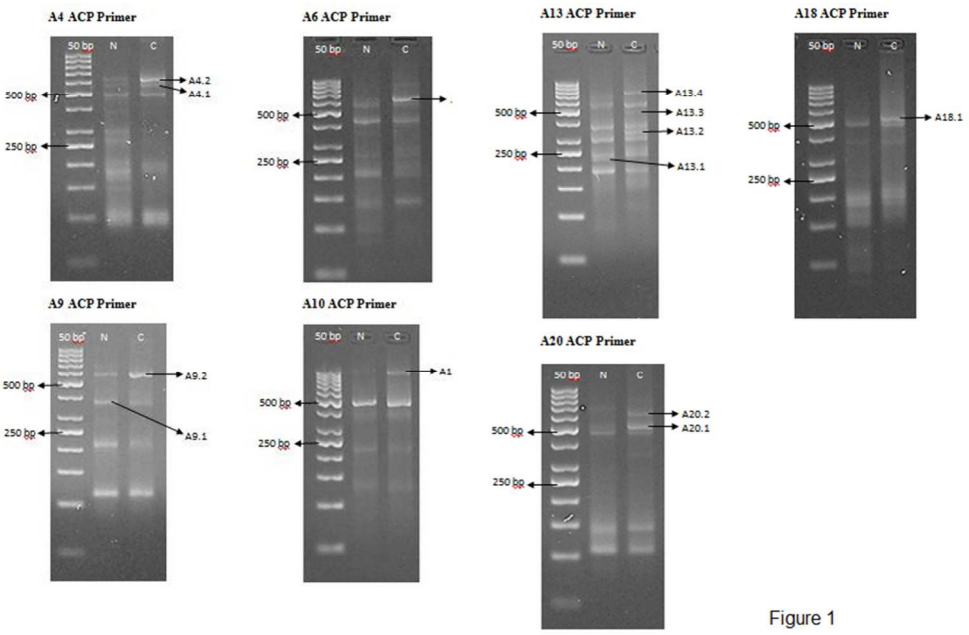



Figure 1
73x47mm (300 x 300 DPI)

eview only

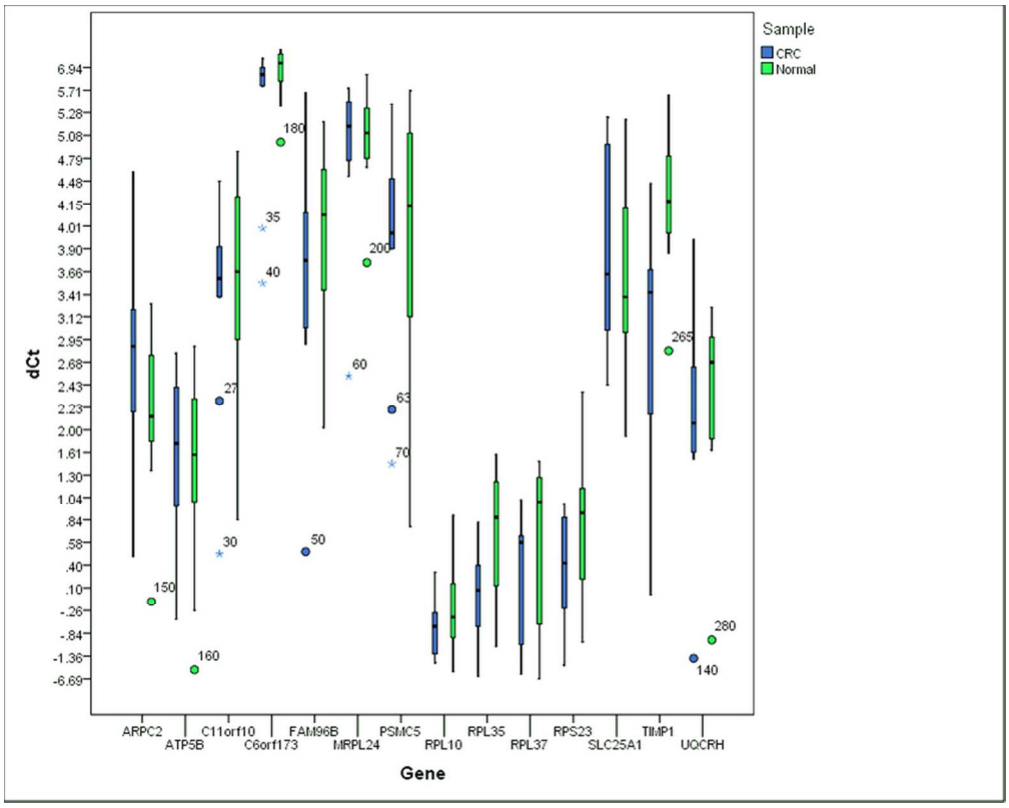



Figure 2(a)
71x56mm (300 x 300 DPI)

For peer review only

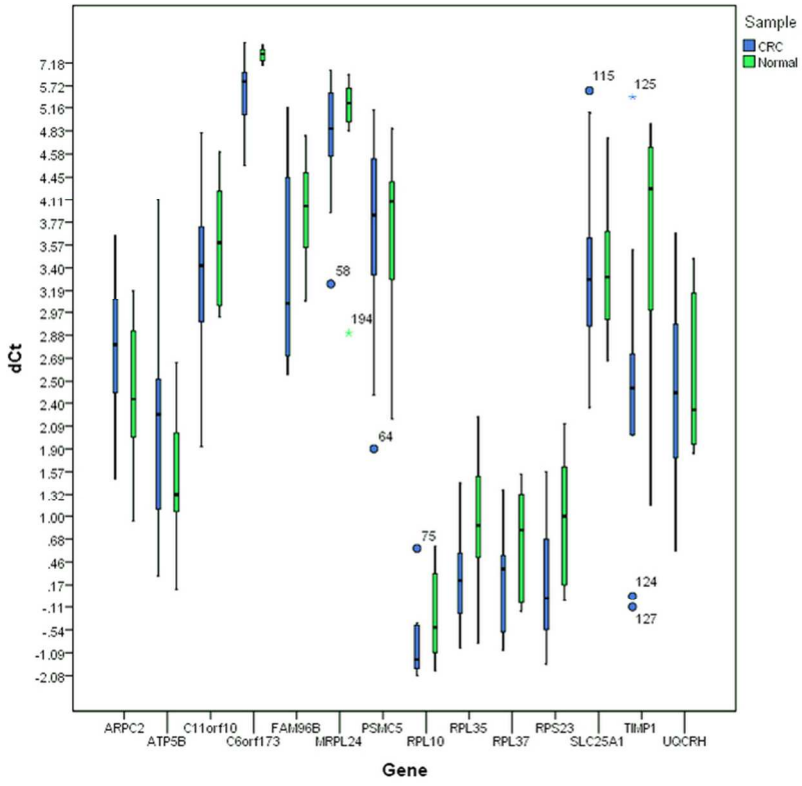



Figure 2(b)
71x57mm (300 x 300 DPI)

For peer review only

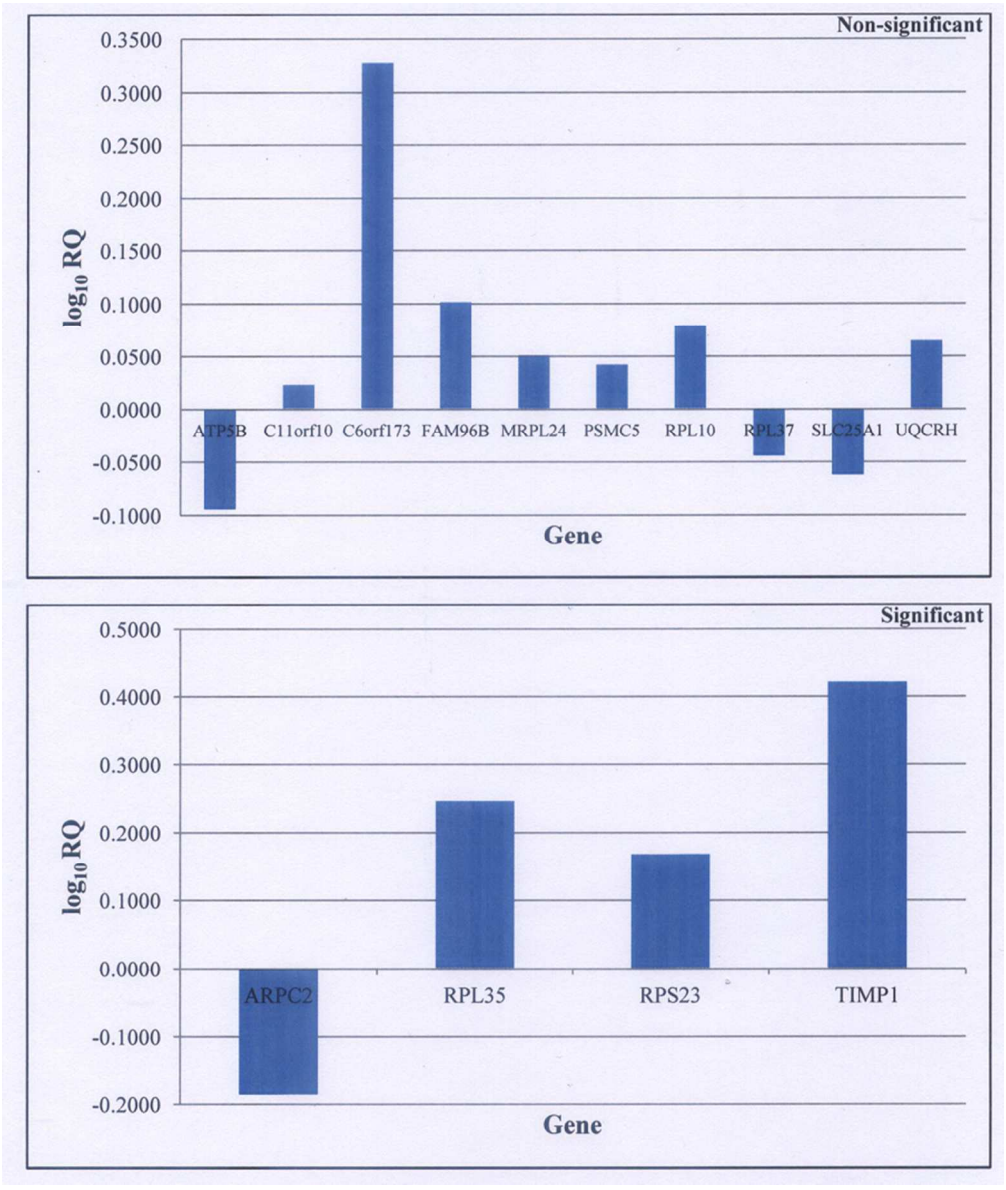



Figure 3
238x280mm (96 x 96 DPI)

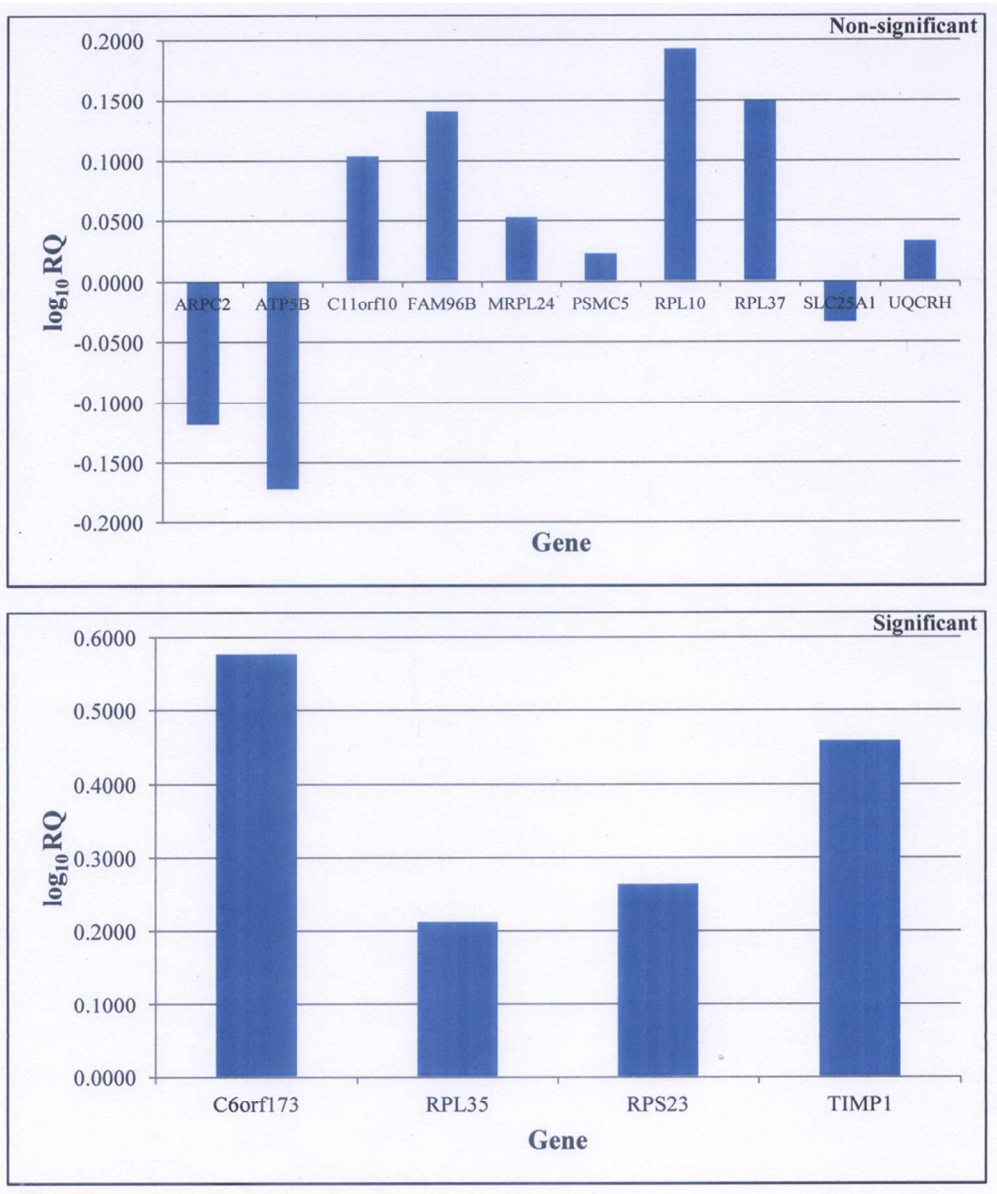



Figure 4
293x350mm (96 x 96 DPI)

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

| | Item No | Recommendation |
|---|---|---|
| **Title and abstract** | 1 | √ (*a*) Indicate the study's design with a commonly used term in the title or the abstract |
| | | √ (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found |
| **Introduction** | | |
| Background/rationale | 2 | √Explain the scientific background and rationale for the investigation being reported |
| Objectives | 3 | √State specific objectives, including any prespecified hypotheses |
| **Methods** | | |
| 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 | (*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up<br>√*Case-control study*—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<br>*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants |
| | | (*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed<br>√*Case-control study*—For matched studies, give matching criteria and the number of controls per case |
| 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 | √ (*a*) Describe all statistical methods, including those used to control for confounding |
| | | (*b*) Describe any methods used to examine subgroups and interactions |
| | | (*c*) Explain how missing data were addressed |
| | | (*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed<br>√*Case-control study*—If applicable, explain how matching of cases and controls was addressed<br>*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy |
| | | (*e*) Describe any sensitivity analyses |

Continued on next page



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

| | | |
|---|---|---|
| **Results** | | |
| Participants | 13* | √ (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 |
| | | √ (b) Give reasons for non-participation at each stage |
| | | (c) Consider use of a flow diagram |
| Descriptive data | 14* | √ (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders |
| | | (b) Indicate number of participants with missing data for each variable of interest |
| | | (c) *Cohort study*—Summarise follow-up time (eg, average and total amount) |
| Outcome data | 15* | *Cohort study*—Report numbers of outcome events or summary measures over time |
| | | √*Case-control study*—Report numbers in each exposure category, or summary measures of exposure |
| | | *Cross-sectional study*—Report numbers of outcome events or summary measures |
| Main results | 16 | √ (*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 |
| | | (*b*) Report category boundaries when continuous variables were categorized |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period |
| 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.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

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



| | |
|---|---|
| Journal: | *BMJ Open* |
| 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<br>Roslani, April Camilla; University of Malaya, Surgery<br>Lian, Lay Hoong; University of Malaya, Biomedical Science<br>Lee, Ping Chin; Universiti Sabah Malaysia, School of Science and Technology<br>Hilmi, Ida; University of Malaya, Medicine<br>Goh, Khean Lee; University of Malaya, Medicine<br>Chua, Kek Heng; University of Malaya, Biomedical Science |
| Primary Subject Heading: | Gastroenterology and hepatology |
| Secondary Subject Heading: | Gastroenterology and hepatology |
| Keywords: | Gastrointestinal tumours < GASTROENTEROLOGY, Cancer genetics < GENETICS, PUBLIC HEALTH |
| | |



SCHOLARONE™
Manuscripts

For peer review only

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

Tze Pheng Lau[1], April Camilla Roslani[2], Lay Hoong Lian[1], Ping Chin Lee[3], Ida Hilmi[4], Khean Lee Goh[4], Kek Heng Chua[1]*

[1]*Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*
[2]*Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*
[3]*School of Science and Technology, Universiti Sabah Malaysia, 88400 Kota Kinabalu, Sabah, Malaysia.*
[4]*Division of Gastroenterology and Hepatology, Department of Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

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

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

*Phone no:* *603-79676607*

*Fax no:* *603-79676600*

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

**Word count:** 3056 words



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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.




For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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:



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 7th 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 7th 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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

biomarkers into existing clinicopathological assessment will further refine the cancer management in future.

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 colorectal 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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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 RNA*later* RNA Stabilization Reagent (Qiagen) immediately after excision and stored at -80 $^{0}$C.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 |

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**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$(T)_{18}$-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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

c) Cloning and sequencing

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

**RT-qPCR analysis**

a) Reverse transcription

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

b) $\Delta\Delta C_T$ analysis

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 2 Sequence similarities and identification of DEGs.

| Differentially Expressed Band | DEG | Identity | Sequence Homology (%) | Accession Number | UniGene Number | Description |
|---|---|---|---|---|---|---|
| ***Over-expressed*** | | | | | | |
| A4.1 | DEG1 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* TIMP metallopeptidase 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 | *Homo sapiens* ATP synthase, H+ 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 |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | phosphorylation. |
| A13.2 | DEG11 | *Homo sapiens* chromosome 11 open reading frame 10 (C11orf10), mRNA | 273/273 (100%) | NM_014206.3 | Hs.437779 | Unknown. |
| A13.3 | DEG12 | *Homo sapiens* 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 | *Homo sapiens* similar to OK/SW-CL.16 (LOC100288418) | 635/644 (98%) | XM_002342023.1 | - | Unknown. |
| A18.1 | DEG14 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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. |
| ***Under-expressed*** | | | | | | |
| A9.1 | DEG5 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* similar to cytochrome c oxidase subunit II (LOC100288578), miscRNA | 141/146 (97%) | XR_078216.1 | - | Unknown. |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**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' | 5'-CTCGCAGGGCATACAT-3' |
| | Reverse: | 5'-AAGTCCTCCTGAGTGACATGGA-3' | |
| DEG2 | Forward: | 5'-GATGCTTACCGAATCCGGAGATC-3' | 5'-CCTCTTCCTCTTCCTCCTCC-3' |
| | Reverse: | 5'-GCATTGCTCTCTCACTGTTGTTAG-3' | |
| DEG3 | Forward: | 5'-CAACCGTCATTGGGTACAAAGG-3' | 5'-ATGGCAAGAAAATCAC-3' |
| | Reverse: | 5'-TGTAAGGGTCCAGCTGATCAAGA-3' | |
| DEG4 | Forward: | 5'-CGGCCAGGAAACTTGAACTTG-3' | 5'-CAGGGCCTCAATCACA-3' |
| | Reverse: | 5'-CCGAGCTGCAGAACAAGGA-3' | |
| DEG5 | Forward: | 5'-CTGGTCGAATGAGGCACCTAAAA-3' | 5'-CATGCCTGAATCTGC-3' |
| | Reverse: | 5'-TGGGTTTAGGTGTTGTTCCTTCAC-3' | |
| DEG6 | Forward: | 5'-AGATTAGCGGGATGAAAACGTCTT-3' | 5'-CCCCGTGATTGTTTTC-3' |
| | Reverse: | 5'-CGCCCAGATGCCGAGAAAA-3' | |
| DEG7 | Forward: | 5'-GGTAGTGATGTGCAAGAGTCCAT-3' | 5'-CATTGCTGGAAAACTG-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'-CCAGGTCAAACTTGTGGATCCT-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' | |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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).

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

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

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

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).



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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].



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

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

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

**Competing interests –** None.

**Data sharing statement –** There is no additional data available.

**Contributorship statement –** All authors were responsible for the design and analysis of the study. All authors were involved in drafting the manuscript, providing the intellectual input and approving the final version of the manuscript.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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.




For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

9. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 2001; 25: 402-8.

10. Voit R, Schafer K, Grummt I. Mechanism of repression of RNA polymerase I transcription by the retinoblastoma protein. Mol Cell Biol 1997; 17: 4230-7.

11. Zhai W, Cornai L. Repression of RNA polymerase I transcription by the tumour suppressor p53. Mol Cell Biol 2000; 20: 5930-8.

12. Backman S, Stambolic V, Mak T. PTEN function in mammalian cell size regulation. Curr Opin Neurobiol 2002; 12: 516-22.

13. Greasley PJ, Bonnard C, Amati B. Myc induces the nucleolin and BN51 genes: possible implications in ribosome biogenesis. Nucleic Acids Res 2000; 28: 446-53.

14. Montanaro L, Treré D, Derenzini M. Nucleolus, ribosomes, and cancer. Am J Pathol 2008; 173: 301-10.

15. Sharp MG, Adams SM, Elvin P, et al. A sequence previously identified as metastasis-related encodes an acidic ribosomal phosphoprotein, P2. Br J Cancer 1990; 61: 83-8.

16. Chester KA, Robson L, Begent RH, et al. Identification of a human ribosomal protein mRNA with increased expression in colorectal tumours. Biochim Biophys Acta 1989; 1009: 297-300.

17. Pogue-Geile K, Geiser JR, Shu M, et al. Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein. Mol Cell Biol 1991; 11: 3842-9.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**Figure Legends**

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.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

Tze Pheng Lau[1], April Camilla Roslani[2], Lay Hoong Lian[1], Ping Chin Lee[3], Ida Hilmi[4], Khean Lee Goh[4], Kek Heng Chua[1]*

[1]*Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

[2]*Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

[3]*School of Science and Technology, Universiti Sabah Malaysia, 88400 Kota Kinabalu, Sabah, Malaysia.*

[4]*Division of Gastroenterology and Hepatology, Department of Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

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

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

*Phone no:* *603-79676607*

*Fax no:* *603-79676600*

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

**Word count:** 3056 words

Formatted



For peer review only

## 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~~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 colorectal~~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 colorectal~~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 colorectal~~CRC~~ neoplasms among Malaysian CRC patients. Further



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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~~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 colorectal~~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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**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 7th 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~~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 7th 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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

the great biologic and genetic heterogeneity of CRC in patients' outcome and treatment response prediction~~, and thus are inadequate for accurate individual prognostic prediction~~. From the perspective of clinical oncology, the integration of molecular biomarkers into existing clinicopathological assessment will further refine the cancer management in future.~~a new classification scheme based on molecular biomarkers would be relatively more efficient and accurate.~~

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 colorectal~~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.



For peer review only

## 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 RNA*later* RNA Stabilization Reagent (Qiagen) immediately after excision and stored at -80 $^{0}$C.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 |



For peer review only

| | |
|---|---|
| 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$(T)_{18}$-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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

c) Cloning and sequencing

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

**RT-qPCR analysis**

a) Reverse transcription

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



For peer review only

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

b) $\Delta\Delta C_T$ analysis

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



For peer review only

randomly amplify gene products in both colorectal~~CRC~~ 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~~CRC~~ tumours, whilst three were under-expressed, as listed in Table 2.

Formatted: Highlight



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 2 Sequence similarities and identification of DEGs.

| Differentially Expressed Band | DEG | Identity | Sequence Homology (%) | Accession Number | UniGene Number | Description |
|---|---|---|---|---|---|---|
| ***Over-expressed*** | | | | | | |
| A4.1 | DEG1 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* TIMP metallopeptidase 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 | *Homo sapiens* ATP synthase, H+ 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 |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | phosphorylation. |
| A13.2 | DEG11 | *Homo sapiens* chromosome 11 open reading frame 10 (C11orf10), mRNA | 273/273 (100%) | NM_014206.3 | Hs.437779 | Unknown. |
| A13.3 | DEG12 | *Homo sapiens* 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 | *Homo sapiens* similar to OK/SW-CL.16 (LOC100288418) | 635/644 (98%) | XM_002342023.1 | - | Unknown. |
| A18.1 | DEG14 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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. |
| ***Under-expressed*** | | | | | | |
| A9.1 | DEG5 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* similar to cytochrome c oxidase subunit II (LOC100288578), miscRNA | 141/146 (97%) | XR_078216.1 | - | Unknown. |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**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' | 5'-CTCGCAGGGCATACAT-3' |
| | Reverse: | 5'-AAGTCCTCCTGAGTGACATGGA-3' | |
| DEG2 | Forward: | 5'-GATGCTTACCGAATCCGGAGATC-3' | 5'-CCTCTTCCTCTTCCTCCTCC-3' |
| | Reverse: | 5'-GCATTGCTCTCTCACTGTTGTTAG-3' | |
| DEG3 | Forward: | 5'-CAACCGTCATTGGGTACAAAGG-3' | 5'-ATGGCAAGAAAATCAC-3' |
| | Reverse: | 5'-TGTAAGGGTCCAGCTGATCAAGA-3' | |
| DEG4 | Forward: | 5'-CGGCCAGGAAACTTGAACTTG-3' | 5'-CAGGGCCTCAATCACA-3' |
| | Reverse: | 5'-CCGAGCTGCAGAACAAGGA-3' | |
| DEG5 | Forward: | 5'-CTGGTCGAATGAGGCACCTAAAA-3' | 5'-CATGCCTGAATCTGC-3' |
| | Reverse: | 5'-TGGGTTTAGGTGTTGTTCCTTCAC-3' | |
| DEG6 | Forward: | 5'-AGATTAGCGGGATGAAAACGTCTT-3' | 5'-CCCCGTGATTGTTTTC-3' |
| | Reverse: | 5'-CGCCCAGATGCCGAGAAAA-3' | |
| DEG7 | Forward: | 5'-GGTAGTGATGTGCAAGAGTCCAT-3' | 5'-CATTGCTGGAAAACTG-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'-CCAGGTCAAACTTGTGGATCCT-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' | |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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).

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

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

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~~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 colorectal~~CRC~~ tumours (Figure 4).

Formatted: Highlight
Formatted: Highlight
Formatted: Centered
Formatted: Highlight
Formatted: Highlight
Field Code Changed
Formatted: Highlight



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

For peer review only

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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].



Formatted

For peer review only

On the other hand, our present study also demonstrated an over-expression of the *TIMP1* gene in both early- and advanced stage primary colorectal~~CRC~~ tumours. This finding is supported by Zeng et al., where the over-expression of TIMP1 was reported in all stages of primary colorectal~~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 colorectal~~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 colorectal~~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 colorectal~~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



For peer review only

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

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

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



For peer review only

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

**Competing interests –** None.

**Data sharing statement –** There is no additional data available.

**Contributorship statement –** All authors were responsible for the design and analysis of the study. All authors were involved in drafting the manuscript, providing the intellectual input and approving the final version of the manuscript.

## 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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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.

9. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 2001; 25: 402-8.

10. Voit R, Schafer K, Grummt I. Mechanism of repression of RNA polymerase I transcription by the retinoblastoma protein. Mol Cell Biol 1997; 17: 4230-7.

11. Zhai W, Cornai L. Repression of RNA polymerase I transcription by the tumour suppressor p53. Mol Cell Biol 2000; 20: 5930-8.

12. Backman S, Stambolic V, Mak T. PTEN function in mammalian cell size regulation. Curr Opin Neurobiol 2002; 12: 516-22.

13. Greasley PJ, Bonnard C, Amati B. Myc induces the nucleolin and BN51 genes: possible implications in ribosome biogenesis. Nucleic Acids Res 2000; 28: 446-53.

14. Montanaro L, Treré D, Derenzini M. Nucleolus, ribosomes, and cancer. Am J Pathol 2008; 173: 301-10.



For peer review only

15. Sharp MG, Adams SM, Elvin P, et al. A sequence previously identified as metastasis-related encodes an acidic ribosomal phosphoprotein, P2. Br J Cancer 1990; 61: 83-8.

16. Chester KA, Robson L, Begent RH, et al. Identification of a human ribosomal protein mRNA with increased expression in colorectal tumours. Biochim Biophys Acta 1989; 1009: 297-300.

17. Pogue-Geile K, Geiser JR, Shu M, et al. Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein. Mol Cell Biol 1991; 11: 3842-9.

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.



For peer review only

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.

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.



For peer review only

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 colorectal~~CRC~~ tumour samples (N: normal sample; C: CRC sample)



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Figure 2 Box-plots showing $\Delta C_T$ values of all colorectal~~CRC~~ 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.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

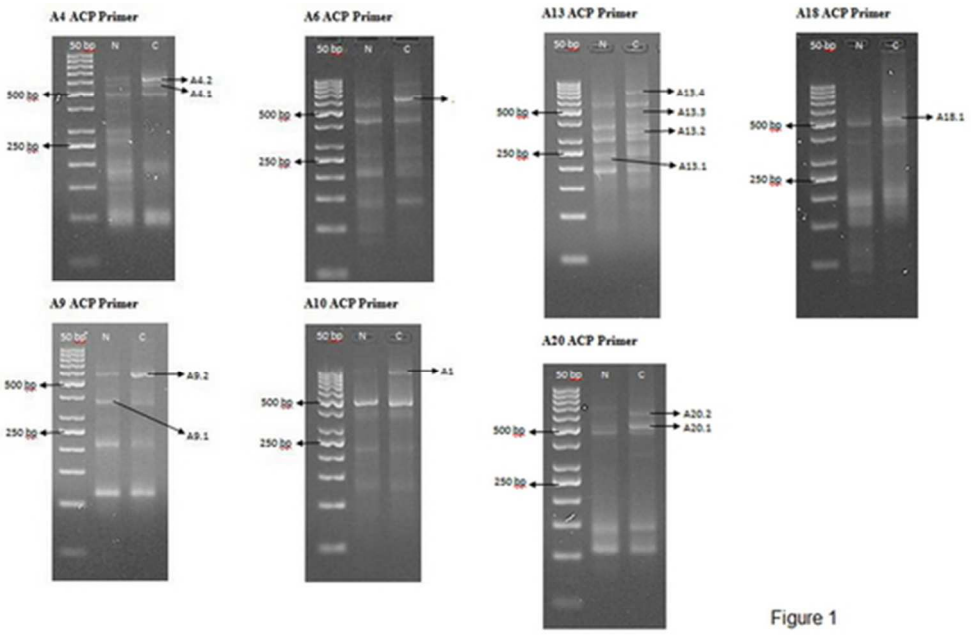
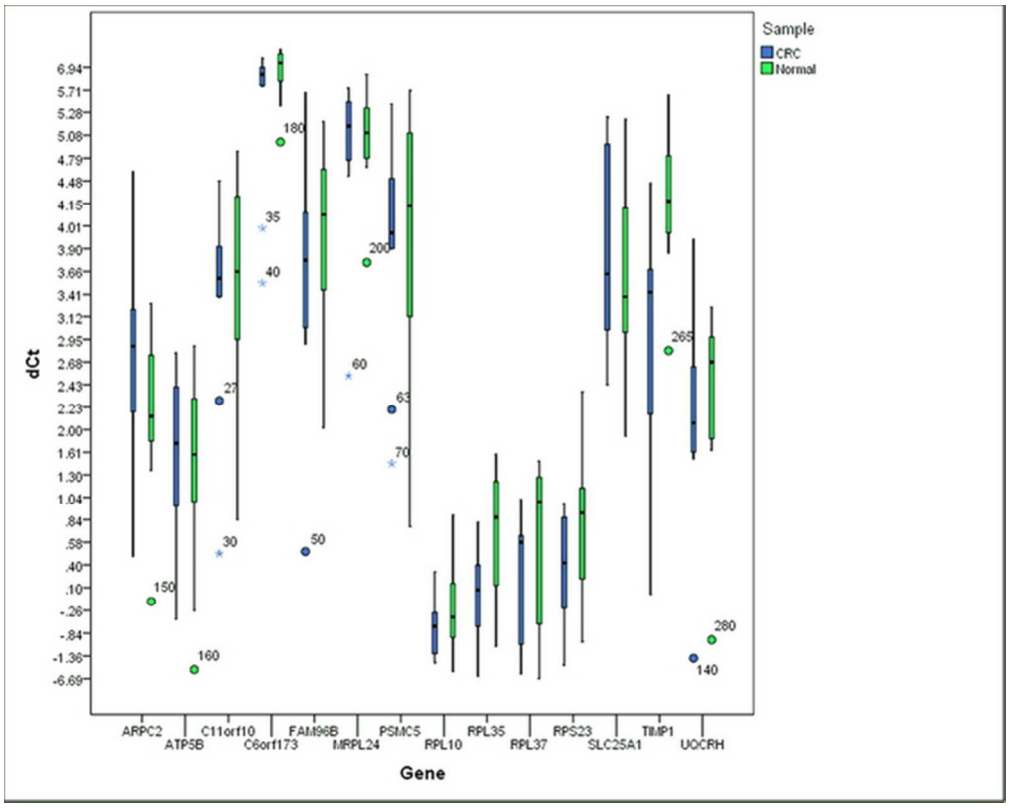



Figure 1

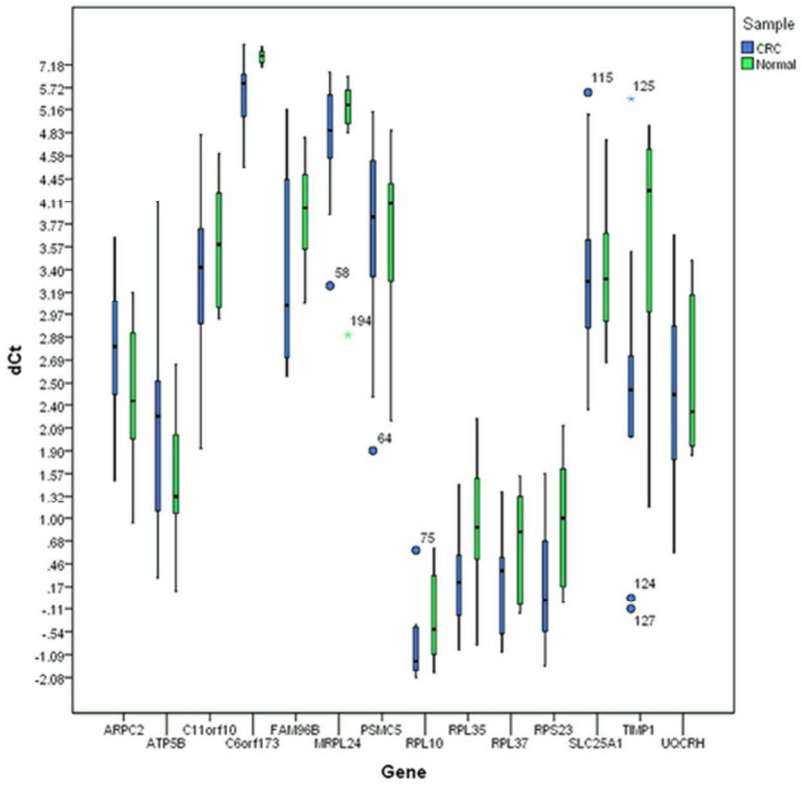
47x30mm (300 x 300 DPI)

eview only



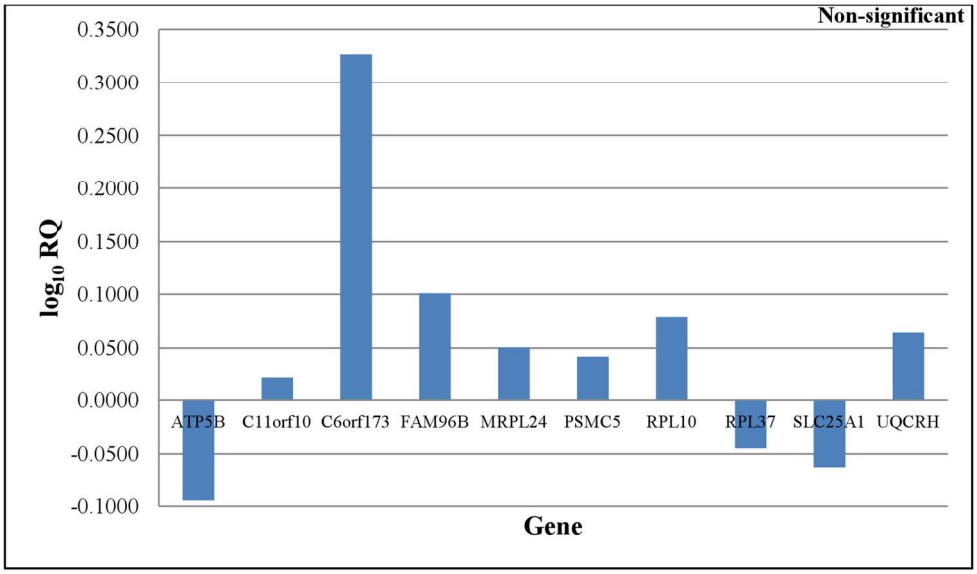

56x45mm (300 x 300 DPI)

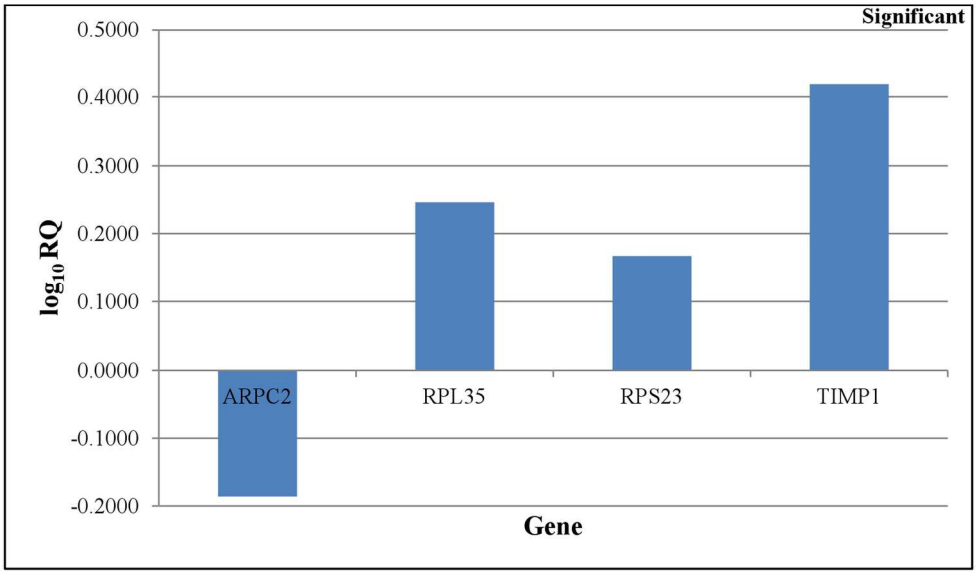
For peer review only



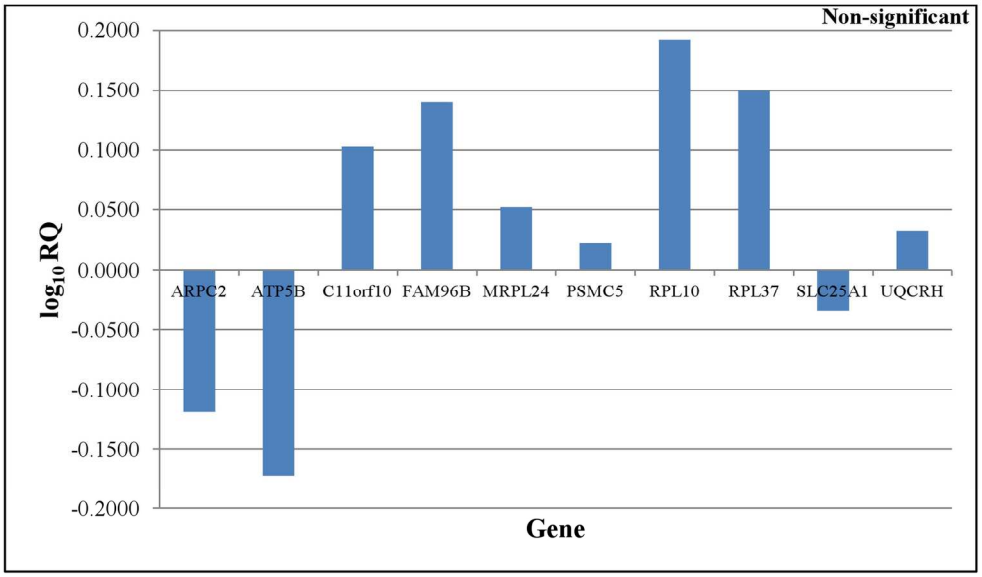

57x45mm (300 x 300 DPI)

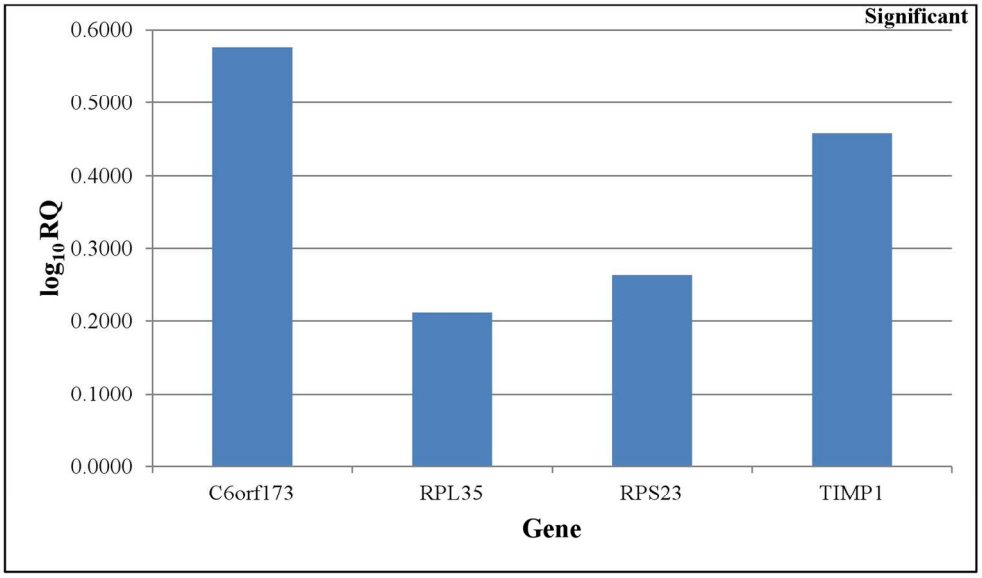
For peer review only




186x221mm (300 x 300 DPI)




187x224mm (300 x 300 DPI)

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

| | Item No | Recommendation |
|---|---|---|
| **Title and abstract** | 1 | √ (*a*) Indicate the study's design with a commonly used term in the title or the abstract |
| | | √ (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found |
| **Introduction** | | |
| Background/rationale | 2 | √Explain the scientific background and rationale for the investigation being reported |
| Objectives | 3 | √State specific objectives, including any prespecified hypotheses |
| **Methods** | | |
| 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 | (*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up<br>√*Case-control study*—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<br>*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants |
| | | (*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed<br>√*Case-control study*—For matched studies, give matching criteria and the number of controls per case |
| 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 | √ (*a*) Describe all statistical methods, including those used to control for confounding |
| | | (*b*) Describe any methods used to examine subgroups and interactions |
| | | (*c*) Explain how missing data were addressed |
| | | (*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed<br>√*Case-control study*—If applicable, explain how matching of cases and controls was addressed<br>*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy |
| | | (*e*) Describe any sensitivity analyses |

Continued on next page



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

| | | |
|---|---|---|
| **Results** | | |
| Participants | 13* | √ (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 |
| | | √ (b) Give reasons for non-participation at each stage |
| | | (c) Consider use of a flow diagram |
| Descriptive data | 14* | √ (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders |
| | | (b) Indicate number of participants with missing data for each variable of interest |
| | | (c) *Cohort study*—Summarise follow-up time (eg, average and total amount) |
| Outcome data | 15* | *Cohort study*—Report numbers of outcome events or summary measures over time |
| | | √*Case-control study*—Report numbers in each exposure category, or summary measures of exposure |
| | | *Cross-sectional study*—Report numbers of outcome events or summary measures |
| Main results | 16 | √ (*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 |
| | | (*b*) Report category boundaries when continuous variables were categorized |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period |
| 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.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

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



| | |
|---|---|
| Journal: | *BMJ Open* |
| 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<br>Roslani, April Camilla; University of Malaya, Surgery<br>Lian, Lay Hoong; University of Malaya, Biomedical Science<br>Chai, Hwa Chia; University of Malaya, Biomedical Science<br>Lee, Ping Chin; Universiti Sabah Malaysia, School of Science and Technology<br>Hilmi, Ida; University of Malaya, Medicine<br>Goh, Khean Lee; University of Malaya, Medicine<br>Chua, Kek Heng; University of Malaya, Biomedical Science |
| Primary Subject Heading: | Gastroenterology and hepatology |
| Secondary Subject Heading: | Gastroenterology and hepatology |
| Keywords: | Gastrointestinal tumours < GASTROENTEROLOGY, Cancer genetics < GENETICS, PUBLIC HEALTH |
| | |



SCHOLARONE™
Manuscripts

For peer review only

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

Tze Pheng Lau[1], April Camilla Roslani[2], Lay Hoong Lian[1], Hwa Chia Chai[1], Ping Chin Lee[3], Ida Hilmi[4],

Khean Lee Goh[4], Kek Heng Chua[1]*

[1]*Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

[2]*Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

[3]*School of Science and Technology, Universiti Sabah Malaysia, 88400 Kota Kinabalu, Sabah, Malaysia.*

[4]*Division of Gastroenterology and Hepatology, Department of Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

*****Correspondence author: Prof. Dr. Chua Kek Heng, Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.

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

*Phone no:* 603-79676607

*Fax no:* 603-79676600

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

**Word count:** 4151 words



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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




For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**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 7th 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 7th 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,



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 biomarkers into existing clinicopathological assessment will further refine the cancer management in future.

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 colorectal 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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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. All subjects were newly-diagnosed with CRC, and were not associated with any hereditary syndromes, previously diagnosed cancer, or positive family history of CRC. 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. The patients' group was comprised of the three main ethnic groups in the Malaysian population, i.e., Chinese, Malays and Indians, in order to ensure a better representative of the study population.

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. The tumouric specimens were excised from the primary colorectal tumours, while the non-cancerous tissue specimens were obtained from distally located macroscopically normal colonic mucosa. Both colorectal tumour and paired non-cancerous tissue specimens were immersed in RNA*later* RNA Stabilization Reagent (Qiagen) immediately after excision and stored at -80 $^{0}$C.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 |

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**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 Bioanalyser 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'-CTGTGAATGCTGCGACTACGATXXXXX$(T)_{18}$-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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

c) Cloning and sequencing

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

**RT-qPCR analysis**

a) Reverse transcription

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

b) $\Delta\Delta C_T$ analysis

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 primary colorectal tumour and paired non-cancerous tissues was analysed by using Real-Time StatMiner software (Integromics). The distribution of the $\Delta C_T$ values obtained for each DEGs within each CRC and control group were tested for normality via the Shapiro-Wilk test. Subsequently, the paired t-test was 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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 2 Sequence similarities and identification of DEGs.

| Differentially Expressed Band | DEG | Identity | Sequence Homology (%) | Accession Number | UniGene Number | Description |
|---|---|---|---|---|---|---|
| ***Over-expressed*** | | | | | | |
| A4.1 | DEG1 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* TIMP metallopeptidase 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 | *Homo sapiens* ATP synthase, H+ 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. |

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

| | | | | | | |
|---|---|---|---|---|---|---|
| A13.2 | DEG11 | *Homo sapiens* chromosome 11 open reading frame 10 (C11orf10), mRNA | 273/273 (100%) | NM_014206.3 | Hs.437779 | Unknown. |
| A13.3 | DEG12 | *Homo sapiens* 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 | *Homo sapiens* similar to OK/SW-CL.16 (LOC100288418) | 635/644 (98%) | XM_002342023.1 | - | Unknown. |
| A18.1 | DEG14 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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. |
| ***Under-expressed*** | | | | | | |
| A9.1 | DEG5 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* similar to cytochrome c oxidase subunit II (LOC100288578), miscRNA | 141/146 (97%) | XR_078216.1 | - | Unknown. |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**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' | 5'-CTCGCAGGGCATACAT-3' |
| | Reverse: | 5'-AAGTCCTCCTGAGTGACATGGA-3' | |
| DEG2 | Forward: | 5'-GATGCTTACCGAATCCGGAGATC-3' | 5'-CCTCTTCCTCTTCCTCCTCC-3' |
| | Reverse: | 5'-GCATTGCTCTCTCACTGTTGTTAG-3' | |
| DEG3 | Forward: | 5'-CAACCGTCATTGGGTACAAAGG-3' | 5'-ATGGCAAGAAAATCAC-3' |
| | Reverse: | 5'-TGTAAGGGTCCAGCTGATCAAGA-3' | |
| DEG4 | Forward: | 5'-CGGCCAGGAAACTTGAACTTG-3' | 5'-CAGGGCCTCAATCACA-3' |
| | Reverse: | 5'-CCGAGCTGCAGAACAAGGA-3' | |
| DEG5 | Forward: | 5'-CTGGTCGAATGAGGCACCTAAAA-3' | 5'-CATGCCTGAATCTGC-3' |
| | Reverse: | 5'-TGGGTTTAGGTGTTGTTCCTTCAC-3' | |
| DEG6 | Forward: | 5'-AGATTAGCGGGATGAAAACGTCTT-3' | 5'-CCCCGTGATTGTTTTC-3' |
| | Reverse: | 5'-CGCCCAGATGCCGAGAAAA-3' | |
| DEG7 | Forward: | 5'-GGTAGTGATGTGCAAGAGTCCAT-3' | 5'-CATTGCTGGAAAACTG-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'-CCAGGTCAAACTTGTGGATCCT-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' | |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

The analysis of RT-qPCR results was performed via Real-Time StatMiner software by importing the raw Ct data. The within-group correlation of these $\Delta C_T$ values was 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).

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

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

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

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).



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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].



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

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

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

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

**Competing interests –** None.

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

**Contributorship statement:**

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

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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.




For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

9. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 2001; 25: 402-8.

10. Voit R, Schafer K, Grummt I. Mechanism of repression of RNA polymerase I transcription by the retinoblastoma protein. Mol Cell Biol 1997; 17: 4230-7.

11. Zhai W, Cornai L. Repression of RNA polymerase I transcription by the tumour suppressor p53. Mol Cell Biol 2000; 20: 5930-8.

12. Backman S, Stambolic V, Mak T. PTEN function in mammalian cell size regulation. Curr Opin Neurobiol 2002; 12: 516-22.

13. Greasley PJ, Bonnard C, Amati B. Myc induces the nucleolin and BN51 genes: possible implications in ribosome biogenesis. Nucleic Acids Res 2000; 28: 446-53.

14. Montanaro L, Treré D, Derenzini M. Nucleolus, ribosomes, and cancer. Am J Pathol 2008; 173: 301-10.

15. Sharp MG, Adams SM, Elvin P, et al. A sequence previously identified as metastasis-related encodes an acidic ribosomal phosphoprotein, P2. Br J Cancer 1990; 61: 83-8.

16. Chester KA, Robson L, Begent RH, et al. Identification of a human ribosomal protein mRNA with increased expression in colorectal tumours. Biochim Biophys Acta 1989; 1009: 297-300.

17. Pogue-Geile K, Geiser JR, Shu M, et al. Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein. Mol Cell Biol 1991; 11: 3842-9.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## Figure Legends

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.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

Tze Pheng Lau[1], April Camilla Roslani[2], Lay Hoong Lian[1], Hwa Chia Chai[1], Ping Chin Lee[3], Ida Hilmi[4], Khean Lee Goh[4], Kek Heng Chua[1]*

[1]*Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

[2]*Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

[3]*School of Science and Technology, Universiti Sabah Malaysia, 88400 Kota Kinabalu, Sabah, Malaysia.*

[4]*Division of Gastroenterology and Hepatology, Department of Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

*Correspondence author: Prof. Dr. Chua Kek Heng, Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.

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

*Phone no:* 603-79676607

*Fax no:* 603-79676600

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

**Word count:** 4151 words



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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.




For peer review only

## 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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**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$^{th}$ 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$^{th}$ 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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 biomarkers into existing clinicopathological assessment will further refine the cancer management in future.

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 colorectal 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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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. All subjects were newly-diagnosed with CRC and were not associated with any hereditary syndromes, previously diagnosed cancer, or positive family history of CRC. 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. The patients' group was comprised of the three main ethnic groups in the Malaysian population, i.e., Chinese, Malays and Indians, in order to ensure a better representative of the study population.

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. The tumouric specimens were excised from the primary colorectal tumours, while the non-cancerous tissue specimens were obtained from distally located macroscopically normal colonic mucosa. Both colorectal tumour and paired non-cancerous tissue specimens were immersed in RNA*later* RNA Stabilization Reagent (Qiagen) immediately after excision and stored at -80 $^{0}$C.



For peer review only

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 |



For peer review only

**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 Bioanalyser 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'-CTGTGAATGCTGCGACTACGATXXXXX$(T)_{18}$-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



For peer review only

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

c) Cloning and sequencing

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

**RT-qPCR analysis**

a) Reverse transcription

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

b) $\Delta\Delta C_T$ analysis

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

Formatted: Font color: Auto, Highlight

Formatted: Font color: Auto, Highlight

## 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

Formatted: Highlight



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

were successfully reported. Among them, 13 were over-expressed in colorectal tumours, whilst three were under-expressed, as listed in Table 2.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 2 Sequence similarities and identification of DEGs.

| Differentially Expressed Band | DEG | Identity | Sequence Homology (%) | Accession Number | UniGene Number | Description |
|---|---|---|---|---|---|---|
| ***Over-expressed*** | | | | | | |
| A4.1 | DEG1 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* TIMP metallopeptidase 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 | *Homo sapiens* ATP synthase, H+ 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. |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

| | | | | | | |
|---|---|---|---|---|---|---|
| A13.2 | DEG11 | *Homo sapiens* chromosome 11 open reading frame 10 (C11orf10), mRNA | 273/273 (100%) | NM_014206.3 | Hs.437779 | Unknown. |
| A13.3 | DEG12 | *Homo sapiens* 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 | *Homo sapiens* similar to OK/SW-CL.16 (LOC100288418) | 635/644 (98%) | XM_002342023.1 | - | Unknown. |
| A18.1 | DEG14 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* 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. |
| ***Under-expressed*** | | | | | | |
| A9.1 | DEG5 | *Homo sapiens* 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 | *Homo sapiens* 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 | *Homo sapiens* similar to cytochrome c oxidase subunit II (LOC100288578), miscRNA | 141/146 (97%) | XR_078216.1 | - | Unknown. |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**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' | 5'-CTCGCAGGGCATACAT-3' |
| | Reverse: | 5'-AAGTCCTCCTGAGTGACATGGA-3' | |
| DEG2 | Forward: | 5'-GATGCTTACCGAATCCGGAGATC-3' | 5'-CCTCTTCCTCTTCCTCCTCC-3' |
| | Reverse: | 5'-GCATTGCTCTCTCACTGTTGTTAG-3' | |
| DEG3 | Forward: | 5'-CAACCGTCATTGGGTACAAAGG-3' | 5'-ATGGCAAGAAAATCAC-3' |
| | Reverse: | 5'-TGTAAGGGTCCAGCTGATCAAGA-3' | |
| DEG4 | Forward: | 5'-CGGCCAGGAAACTTGAACTTG-3' | 5'-CAGGGCCTCAATCACA-3' |
| | Reverse: | 5'-CCGAGCTGCAGAACAAGGA-3' | |
| DEG5 | Forward: | 5'-CTGGTCGAATGAGGCACCTAAAA-3' | 5'-CATGCCTGAATCTGC-3' |
| | Reverse: | 5'-TGGGTTTAGGTGTTGTTCCTTCAC-3' | |
| DEG6 | Forward: | 5'-AGATTAGCGGGATGAAAACGTCTT-3' | 5'-CCCCGTGATTGTTTTC-3' |
| | Reverse: | 5'-CGCCCAGATGCCGAGAAAA-3' | |
| DEG7 | Forward: | 5'-GGTAGTGATGTGCAAGAGTCCAT-3' | 5'-CATTGCTGGAAAACTG-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'-CCAGGTCAAACTTGTGGATCCT-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' | |



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

The analysis of RT-qPCR results was performed via Real-Time StatMiner software by importing the raw Ct data. The within-group correlation of these $\Delta C_T$ values was 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).

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

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

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

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).

Formatted: Highlight
Formatted: Highlight
Formatted: Centered
Formatted: Highlight
Formatted: Centered
Formatted: Not Superscript/ Subscript
Formatted: Highlight
Field Code Changed
Formatted: Highlight



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1

16

For peer review only

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## 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].



For peer review only

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



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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

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

In conclusion, we have characterised two distinctive gene expression patterns, which comprise of the *ARPC2*, *C6orf173*, *RPL35*, *RPS23* and *TIMP1* genes, for the stratification of primary colorectal adenocarcinomas among Malaysian CRC patients. It was postulated that the actin cytoskeleton might play an important role in determining the dysplastic cell morphology during the early development of CRC, while the aberrations in the assembly of functional kinetochore might be crucial for the aneuploidy of the advanced stage colorectal tumours. Nevertheless, the findings of this study were considered preliminary owing to the relatively small sample size. The main reason for this is ~~Our current sample size was relatively small owing to~~ the lack of a designated Tissue Bank in our institution. ~~There were also not many CRC patient volunteers.~~ Moreover, the lack of CRC patient volunteers and our stringent criteria for patient selection have also limited the availability of suitable specimens within the short sample collection period.

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

Formatted: Highlight

Formatted: Font color: Auto, Highlight

Formatted: Highlight



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

specificity primers and probes, as well as reliable statistical analysis. In future, the validation of these DEGs should be performed on a larger set of clinical samples, and extensive inter-laboratory testing of their differential abilities on each CRC stage is also desired. In addition, we should also integrate other imaging and histological information to complement our identified gene expression patterns, which then hold promises for better stratification of tumours.

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

**Competing interests –** None.

**Data sharing statement –** There is no additional data available.

**Contributorship statement:**

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

**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.



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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.

9. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 2001; 25: 402-8.

10. Voit R, Schafer K, Grummt I. Mechanism of repression of RNA polymerase I transcription by the retinoblastoma protein. Mol Cell Biol 1997; 17: 4230-7.

11. Zhai W, Cornai L. Repression of RNA polymerase I transcription by the tumour suppressor p53. Mol Cell Biol 2000; 20: 5930-8.

12. Backman S, Stambolic V, Mak T. PTEN function in mammalian cell size regulation. Curr Opin Neurobiol 2002; 12: 516-22.



For peer review only

13. Greasley PJ, Bonnard C, Amati B. Myc induces the nucleolin and BN51 genes: possible implications in ribosome biogenesis. Nucleic Acids Res 2000; 28: 446-53.

14. Montanaro L, Treré D, Derenzini M. Nucleolus, ribosomes, and cancer. Am J Pathol 2008; 173: 301-10.

15. Sharp MG, Adams SM, Elvin P, et al. A sequence previously identified as metastasis-related encodes an acidic ribosomal phosphoprotein, P2. Br J Cancer 1990; 61: 83-8.

16. Chester KA, Robson L, Begent RH, et al. Identification of a human ribosomal protein mRNA with increased expression in colorectal tumours. Biochim Biophys Acta 1989; 1009: 297-300.

17. Pogue-Geile K, Geiser JR, Shu M, et al. Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein. Mol Cell Biol 1991; 11: 3842-9.

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.



For peer review only

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.

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.



For peer review only

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**

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)

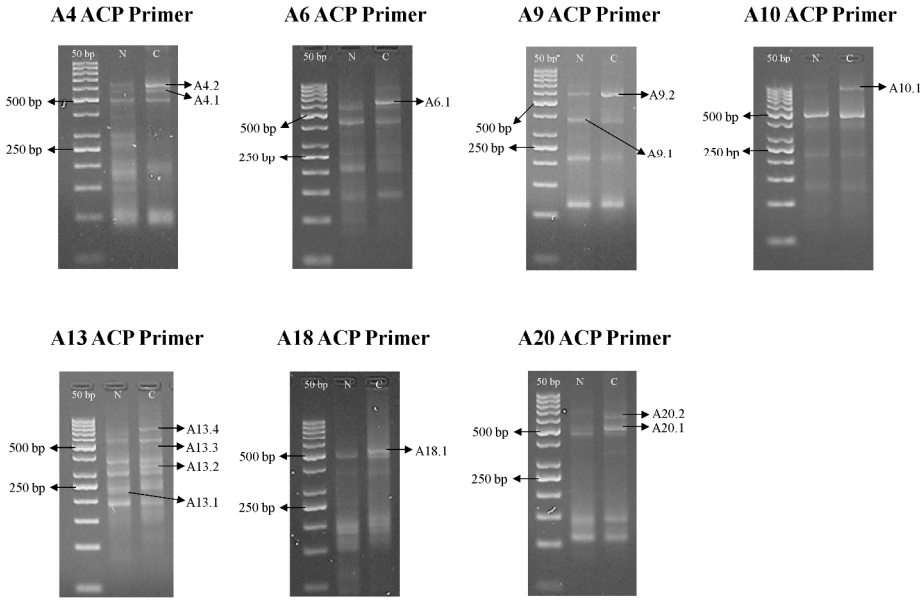
Figure 2 Box-plots showing $\Delta C_T$ values of all colorectal tumours and normal colonic tissues in each early- (a) and advanced (b) stage CRC group.

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.

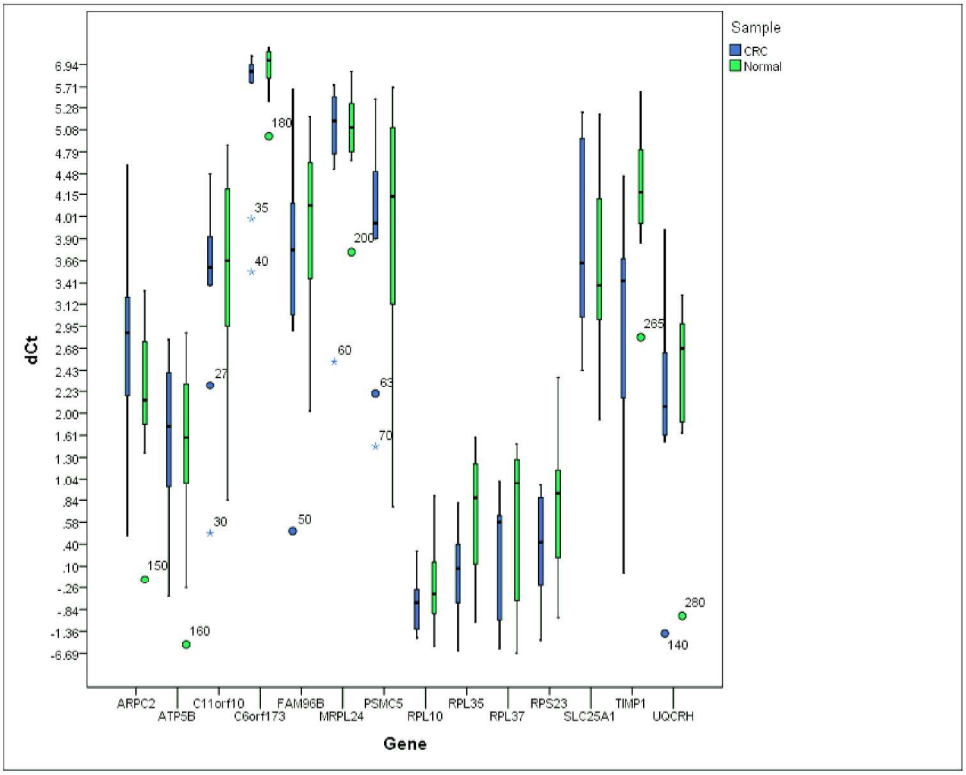


For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



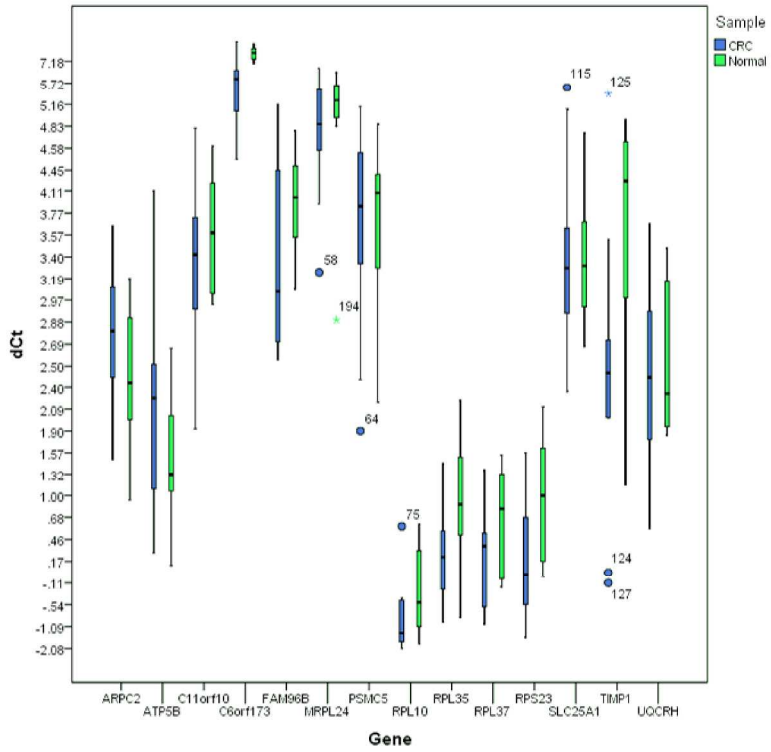

265x175mm (300 x 300 DPI)

For peer review only



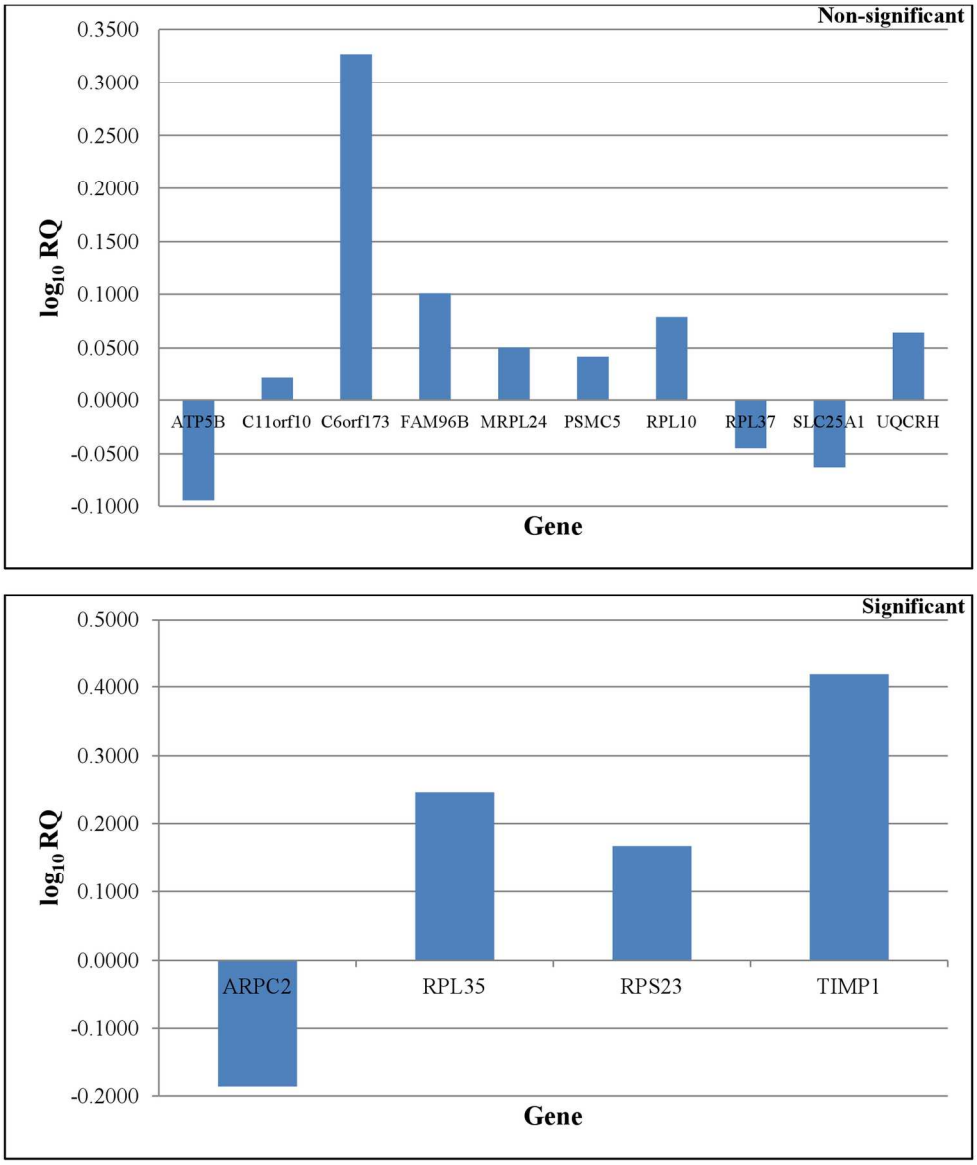

260x203mm (300 x 300 DPI)

For peer review only

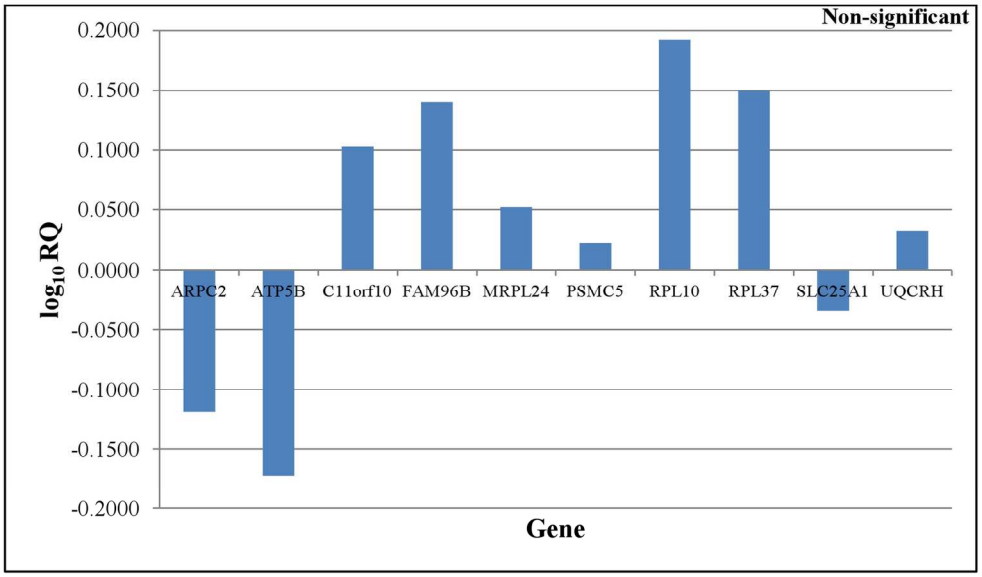


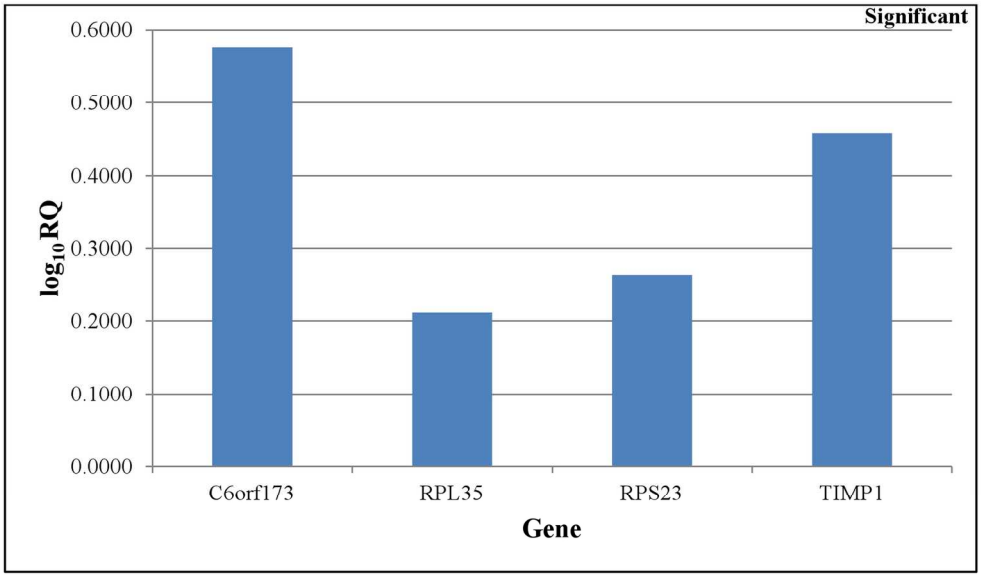

260x203mm (300 x 300 DPI)

For peer review only




186x221mm (300 x 300 DPI)




187x224mm (300 x 300 DPI)

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

| | Item No | Recommendation |
|---|---|---|
| **Title and abstract** | 1 | √ (*a*) Indicate the study's design with a commonly used term in the title or the abstract |
| | | √ (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found |
| **Introduction** | | |
| Background/rationale | 2 | √Explain the scientific background and rationale for the investigation being reported |
| Objectives | 3 | √State specific objectives, including any prespecified hypotheses |
| **Methods** | | |
| 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 | (*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up<br>√*Case-control study*—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<br>*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants |
| | | (*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed<br>√*Case-control study*—For matched studies, give matching criteria and the number of controls per case |
| 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 | √ (*a*) Describe all statistical methods, including those used to control for confounding |
| | | (*b*) Describe any methods used to examine subgroups and interactions |
| | | (*c*) Explain how missing data were addressed |
| | | (*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed<br>√*Case-control study*—If applicable, explain how matching of cases and controls was addressed<br>*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy |
| | | (*e*) Describe any sensitivity analyses |

Continued on next page



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

| | | |
|---|---|---|
| **Results** | | |
| Participants | 13* | √ (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 |
| | | √ (b) Give reasons for non-participation at each stage |
| | | (c) Consider use of a flow diagram |
| Descriptive data | 14* | √ (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders |
| | | (b) Indicate number of participants with missing data for each variable of interest |
| | | (c) *Cohort study*—Summarise follow-up time (eg, average and total amount) |
| Outcome data | 15* | *Cohort study*—Report numbers of outcome events or summary measures over time |
| | | √*Case-control study*—Report numbers in each exposure category, or summary measures of exposure |
| | | *Cross-sectional study*—Report numbers of outcome events or summary measures |
| Main results | 16 | √ (*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 |
| | | (*b*) Report category boundaries when continuous variables were categorized |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period |
| 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.

For peer review only



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml