

# RNA expression of the molecular signature genes for metastasis in colorectal cancer

LUCIANO CARVALHO<sup>1-3\*</sup>, JINSHENG YU<sup>1\*</sup>, GILBERTO SCHWARTSMANN<sup>3</sup>,  
HOWARD L. McLEOD<sup>1</sup> and JAMES W. FLESHMAN<sup>2</sup>

Departments of <sup>1</sup>Medicine and <sup>2</sup>Surgery, Washington University School of Medicine and the Siteman Cancer Center, St. Louis, MO 63110, USA; <sup>3</sup>Department of Gastroenterology, Post-Graduation in Science of Medicine, Universidade Federal do Rio, Grande do Sul, Brazil

Received October 11, 2010; Accepted December 21, 2010

DOI: 10.3892/or.2011.1208

**Abstract.** Colorectal cancer is an endemic disease in the Western world. Search for molecular signatures present in primary tumors that predict tumor metastasis potential has been proposed and in particular, a 17-gene molecular signature is associated with poor survival in breast cancer, prostate cancer, medulloblastoma and lymphoma in a recent study. Using quantitative real-time PCR assay (qPCR), our study observed tumor-normal differential RNA expression in 15 of these 17 genes in a cohort of 52 stage III colorectal cancer patients (all  $P < 0.05$ ), which signified the importance of these 17 signature genes in colorectal cancer. Although no significant correlation was found between tumor RNA levels of these 17 genes and some of clinical features (age, gender, and location, all  $P > 0.05$ ), two distinct groups among these genes were observed with Spearman correlation scores  $> 0.6$  ( $P < 0.01$ ), suggesting co-expression/interaction within these genes. Of the 37 patients for whom complete follow-up data was available, 12 patients had recurrence and 25 had no recurrence. There was no significant difference in tumor RNA levels between recurrence and non-recurrence groups for the 17 genes (all  $P > 0.05$ ), but the recurrence group had more patients with mucinous tumors (9/12 vs. 7/25,  $P < 0.05$ ) and more lymph node involvement (median 7.2 vs. 2.5,  $P < 0.05$ ) compared to the non-recurrence group. Moreover, survival analysis revealed a significant difference in patient overall survival time between low and high tumor RNA levels for 1 of the 17 genes (PTTG1,  $P = 0.024$ ). Our qPCR validation study confirms the importance of most 17-gene molecular

signature genes with differential RNA expression and suggests the relevance of PTTG1 for survival in colorectal cancers.

## Introduction

Colorectal cancer (CRC) is an endemic disease in the Western world and it is the second most common cause of death for both men and women in the United States (1). Even though there are major advances in diagnosis and treatment of the disease, the mortality remains almost unaltered in the last 20 years (2). Initial CRC treatment is surgical, where staging is the determinant factor in prognosis and therapeutic decisions. Approximately 50% of patients who have curative surgical resection will die of a disease secondary to metastasis after surgery, and 80% of those will have recurrence of the disease detectable within 2 years (1-3).

The advancement of DNA manipulation techniques have promoted a significant increase in the knowledge of molecular mechanisms involved in the development and progression of cancer. Genomic analyses have supplied new markers that can definitely help in prognostic and therapeutic prediction based on tumor biology at molecular level. However, several promising colorectal cancer progression markers, such as p53, K-Ras, and cytokeratins, have not shown uniform results (4-7). Many of these studies were based on a small, non-consecutive series of patients, and the patient selection was different in many ways such as variation in treatment time, different mutations of tumors, research methods, all making comparisons more difficult.

The studies have begun to identify groups of genes, the molecular signatures, whose alterations are associated with tissue abnormalities or altered outcome (8,9). This has allowed description of new sub-classifications of patients based on genomic analysis, which is different from traditionally clinical-pathological analysis (10). Reports of successful applications in hematological malignancies, breast cancer (11,12), bladder cancer (13), malignant melanomas (14) have recently shown the significance of molecular signatures in primary tumors as prognostic markers based upon the characteristics of gene expression that can predict malignant potential of the tumors. In the studies of CRC,

---

*Correspondence to:* Dr James W. Fleshman, Department of Surgery, Washington University School of Medicine, Campus Box 8109, 660 S. Euclid Ave, St. Louis, MO 63110, USA  
E-mail: [fleshmanj@wustl.edu](mailto:fleshmanj@wustl.edu)

\*Contributed equally

*Key words:* gene expression, metastasis, colorectal cancer, quantitative PCR



**Genes studied.** The 17 genes which are from the previous study (18) were chosen to be studied (Table I). Primers of these 17 genes were designed using the software Primer Express 2.0 (ABI, Foster City, CA) and sequence information for these primers was displayed in Table I. All primers were tested, the optimal annealing temperature was determined with gradient PCR for all 17 genes, and a single band at the right position with relevant amplicon size was confirmed with gel visualization. In addition, the dissociation curve feature in the ABI 7900 system showed the specific amplification for all 17 genes with a single peak of the melting temperature.

**Quantitative real-time reverse transcription-PCR.** Regions of high tumor cellularity were selected for RNA extraction (median 86.3%; range, 65-95%). Tissue total RNA was isolated from the colon tumor or adjacent normal colon mucosa with the TRIzol RNA isolation kit (Invitrogen, Carlsbad, CA), and the quality of RNA (i.e. A260/280>1.8 and the clear RNA bands for 28S, 18S and 5S were seen) was confirmed in the Siteman Cancer Center Tissue Procurement Core. After reverse-transcription into cDNA using Superscript II reverse transcriptase (Invitrogen), the RNAs of 17 genes were determined by real-time qPCR, which was carried out in a 10- $\mu$ l reaction mix containing 2  $\mu$ l of cDNA (10 ng/ $\mu$ l), 5  $\mu$ l of 2X SYBR-Green universal qPCR master mix (ABI), and 3  $\mu$ l of primer mix (400 nM each forward and reverse primers). All real-time qPCR assays were performed in triplicate on an ABI PRISM 7900 Sequence Detector System (ABI) with the following program: 50°C for 2 min to activate uracil N-glycosylase enzyme, and 40 cycles at 95°C for 15 sec, 54 or 6°C for 30 sec, and 72°C for 30 sec. The sequence detection program calculates a threshold cycle number (CT) at which the reporter fluorescence generated by cleavage of the probe is statistically greater than that of the background signal.

**Measurement of relative RNA expression level.** In this study, relative expression level was calculated using a modified comparative  $C_T$  method, which uses actual real-time qPCR amplification efficiency instead of assuming all sets of gene primer and probe have approximately equal efficiency (19). The gene encoding amyloid B precursor protein was used as an internal reference gene, because it had near identical expression between colon tumor and normal tissues in previous serial analysis of gene expression analysis and less than a 3-fold change between all tumor and normal sample pairs in this study. The relative expression level of an individual target gene was normalized to the reference gene and to one of all 104 colon tumor and normal RNA samples that was with the maximum CT value (i.e., the lowest expression level, called calibrator sample or 1X sample) in any target gene. A mathematical model was applied to determine the normalized relative expression level of a target gene in the individual sample, using the following formula:

$$\text{Expression} = \frac{E_{\text{target}}^{\Delta C_T(\text{calibrator-sample})}}{E_{\text{reference}}^{\Delta C_T(\text{calibrator-reference})}}$$

where E target is the real-time qPCR efficiency of target gene transcript, and E reference is the real-time qPCR efficiency of reference gene transcript (20). Therefore, relative RNA expression level of a gene in this study is a relative number based on the reference gene and the 1X sample.

**Statistical analysis.** Statistical analyses were performed with the software STATISTICA (StatSoft, Inc., Tulsa, OK). Significance of the difference of relative expression level between paired tumor and normal samples was evaluated by Wilcoxon matched-pairs test. The influence of gender, tumor location, pathological variables was evaluated with Mann-Whitney or Kruskal-Wallis test. Correlation between variables was observed with Spearman rank correlation. Survival analysis was carried out with the SAS LIFETEST procedure (Cary, NC). The significance level was set at  $P < 0.05$ . P-values from multiple tests were corrected with the Bonferroni test.

## Results

**RNA expression of the 17 genes in colorectal tumor and normal tissues.** Fifty-two samples from colorectal tumors and normal tissues each were studied simultaneously. There was a significant difference in RNA expression level between tumor and normal tissues in 15 of the 17 genes analyzed ( $P < 0.05$  all), and the RNA expression of the remaining 2 genes (DHPS and NR4A1) had no significant difference between tumor and normal tissues (Fig. 1).

**Ratio expression of the 17 genes between tumor and normal samples.** The 17 genes were classified according to the ratio of RNA expression between tumor and normal tissues; and the tumor-normal ratio of the RNA expression was considered up-regulated when the ratio was  $> 1.2$ , and down-regulated when  $< 0.8$ . Of the 17 genes, 7 genes were classified as up-regulated (SNRPF, EIF4EL3, COL1A1, COL1A2, PTTG1, RUNX1, LMNB1;  $P < 0.05$ ), another 7 were down-regulated (ACTG2, MYLK, MYH11, RBM5, CNN1, HLA-DPB1, MT3;  $P < 0.05$ ), and the rest 3 genes statistically had no difference in RNA expression between tumor and normal tissue as shown in Table II.

**Co-expression of the 17 metastasis genes.** Correlations of the RNA expression between all 17 genes were determined in 52 colorectal tumors. Statistically significant correlations were demonstrated in 2 distinct gene-groups. The first group had 4 genes with up-regulated and 2 with normal RNA expression (all Spearman coefficient  $> 0.6$ , Table III). The other group was formed with 10 genes, of which 7 were down-regulated, 2 up-regulated, and 1 had normal RNA expression (all Spearman coefficient  $> 0.6$ , Table III). The gene MT3 was not shown to have any strong correlation with the other 16 genes. No evidence of such co-expression was observed in the normal tissues.

**Tumor RNA expression and patient clinical features.** The comparisons of the RNA expression were made with patient age, gender, tumor location, histological type, pathological grade, lymph node involvement, and recurrence at the last follow-up. There was no statistically significant difference in

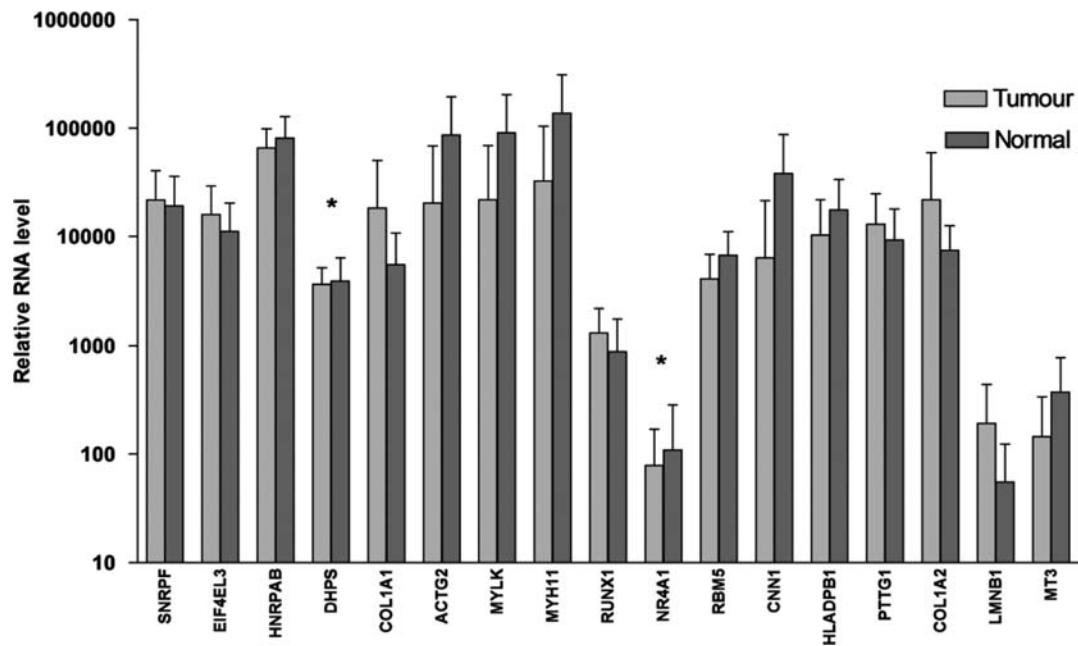


Figure 1. Differential expression in relative RNA level of the 17 genes between tumor and matched normal mucosa tissues in 52 colorectal cancer patients. ( $P > 0.05$  for DHPS and NR4A1,  $P < 0.05$  in the rest 15 genes). Error bars represent 95% CI of the mean. All assay values were normalized to the internal reference gene APP.

Table II. T/N category of the RNA expression in 52 CRC patients.

Gene symbol	No. of cases in category of T/N <0.8	No. of cases in category of T/N = 0.8-1.2	No. of cases in category of T/N >1.2	P-value
<b>Up-regulated</b>				
SNRPF	13	11	24	0.023
EIF4EL3	12	7	29	0.001
COL1A1	11	5	32	0.000
COL1A2	12	4	32	0.000
RUNX1	12	5	31	0.000
PTTG1	14	6	28	0.000
LMNB1	7	0	25	0.000
<b>Down-regulated</b>				
ACTG2	34	2	12	0.000
MYLK	36	5	7	0.000
MYH11	36	5	7	0.000
RBM5	29	14	5	0.000
CNN1	36	2	10	0.000
HLA-DPB1	34	6	8	0.000
MT3	31	6	11	0.000
<b>Normal</b>				
HNRPAB	22	11	15	0.134
DHPS	18	12	18	1.000
NR4A1	22	3	23	0.845

tumor RNA expression of these 17 genes across all different clinical-pathological features when analyzed as a single marker (all  $P > 0.05$ ). However, two groups of patients were separated using the end point of clinical recurrence, and 37 patients of these two groups had complete clinical data

available for further analysis. Of the 37 patients, 12 patients had recurrence and 25 had no recurrence. Although there was statistically no significant difference in RNA expression of the 17 genes between these 2 groups ( $P > 0.05$ ), but several pathological features such as tumor histological type and

Table III. Spearman rank correlation with RNA expression of the 17 genes in 52 CRC patients.

	SNRPF	EIF4EL3	HNRPAB	DHPS	COL1A1	ACTG2	MYLK	MYH11	RUNX1	NR4A1	RBM5	CNN1	HLADPB1	PTTG1	COL1A2	LMNB1
EIF4EL3	<b>0.84</b>															
HNRPAB	<b>0.79</b>	<b>0.62</b>														
DHPS	<b>0.62</b>	0.41	<b>0.68</b>													
COL1A1	-0.19	-0.37	0.05	0.23												
ACTG2	-0.06	-0.27	0.12	0.30	<b>0.58</b>											
MYLK	0.01	-0.28	0.19	0.30	<b>0.63</b>	<b>0.84</b>										
MYH11	-0.01	-0.29	0.19	0.28	<b>0.64</b>	<b>0.84</b>	<b>0.80</b>									
RUNX1	0.17	0.02	0.31	<b>0.50</b>	<b>0.65</b>	0.46	<b>0.56</b>	<b>0.55</b>								
NR4A1	-0.01	-0.19	0.14	0.43	<b>0.55</b>	0.49	0.42	0.41	<b>0.55</b>							
RBM5	0.18	0.03	0.39	<b>0.58</b>	<b>0.53</b>	0.55	<b>0.54</b>	0.53	<b>0.84</b>	<b>0.57</b>						
CNN1	-0.17	-0.40	0.08	0.29	<b>0.68</b>	<b>0.87</b>	<b>0.81</b>	<b>0.82</b>	0.41	<b>0.57</b>	0.43					
HLADPB1	0.02	-0.09	0.25	0.45	<b>0.59</b>	<b>0.52</b>	<b>0.61</b>	<b>0.61</b>	<b>0.51</b>	0.49	<b>0.53</b>	<b>0.63</b>				
PTTG1	<b>0.85</b>	<b>0.85</b>	<b>0.67</b>	0.43	-0.21	-0.15	-0.18	-0.19	0.01	-0.07	0.03	-0.26	-0.07			
COL1A2	-0.16	-0.37	0.05	0.20	<b>0.96</b>	<b>0.59</b>	<b>0.69</b>	<b>0.69</b>	<b>0.60</b>	<b>0.53</b>	0.45	<b>0.72</b>	<b>0.60</b>	-0.20		
LMNB1	<b>0.60</b>	<b>0.63</b>	<b>0.52</b>	0.30	-0.19	-0.28	-0.37	-0.38	-0.14	0.01	-0.10	-0.28	-0.28	<b>0.74</b>	-0.25	
MT3	0.23	0.34	0.16	0.21	-0.08	0.01	0.02	0.01	0.14	0.12	0.13	-0.06	0.26	0.22	-0.09	0.09

The score is in bold if the P<0.001. The range of the higher scores is between 0.50 and 0.96.

lymph node involvement were statistically different between the recurrence and non-recurrence groups (P=0.02 and 0.01, respectively), i.e., the recurrence group had more patients with mucinous tumor than the non-recurrence group (9/12 vs. 7/25, P<0.05) and there was a 7.2 vs. 2.5 of median lymph node involvement for the recurrence and non-recurrence groups (P<0.05), respectively. Moreover, survival analysis revealed a significant difference in patient overall survival time between low and high tumor RNA levels for 1 of the 17 genes (PTTG1, P=0.024) (Fig. 2).

*Hierarchical clustering analysis in 52 colorectal tumors.* Clustering analysis with tumor RNA expression of the 17 genes was performed in the 52 patients with colorectal cancer; and 3 distinct patient-groups were found (Fig. 3). The tumor RNA expression was statistically significantly different between these 3 groups (P<0.05) in majority of the 17 genes (14/17), but not in the other 3 genes (MT3, DPHS, and HNRPAB). No significant difference was found for each of the clinical-pathological features between these 3 patient-groups separated by clustering with tumor RNA expression (P>0.05). These correlations were tested with patient age, gender, tumor location, histological type, pathological grade, differentiation, invasion, lymph node involvement, and status of disease and recurrence. However, the hierarchical clustering analysis had not shown these 3 to be distinct or any patient-groups in 52 normal tissues.

**Discussion**

The 17-gene signature that was previously related to poor prognosis in different solid tumors (18) has been evaluated in our study for the CRC stage III patients. Our study demonstrated that the RNA expression in CRC tumors is significantly different from matched normal mucosa, independent of anatomic and clinical pathological variables in 15 out of 17 genes (P>0.05). These results might be a confirmation that this group of genes can be involved in the development of colorectal cancer, and probably in other solid tumors as well. The differential expression between tumor and normal tissues from CRC patients also suggests that these 17 genes may have important role in tumor biology which favors tumor classification and decision making for optimized therapy. However, the RNA expressions of these 17 genes in the tumors were not associated with disease recurrence or other clinical-pathological features in our study and rather the tumor type and lymph node involvement were shown to be useful for prediction of disease recurrence in the pre-therapy setting. These results did not establish a positive relationship between the 17-gene molecular signature and tumor recurrence/metastasis. Nonetheless, one of the 17 genes (PTTG1) has been indicated a significant correlation with patient overall survival, which is a potential survival biomarker although more validation studies are needed. The difficulty to compare studies of molecular analysis in literature is mainly due to the techniques utilized and different selections of cancer patients. The techniques utilized should be validated carefully in such molecular analysis and results from such studies should be interpreted with caution. Michiels and colleagues (21) have re-analyzed the metadata from seven large studies

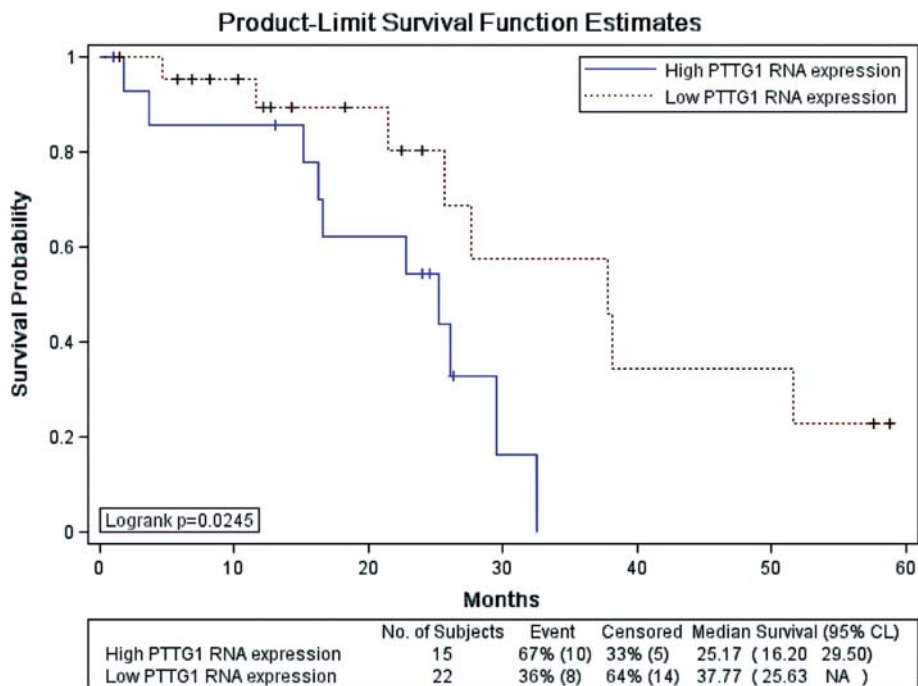


Figure 2. Survival curve for 37 colorectal cancer patients with follow-up data. The analysis was performed with the Kaplan-Meier estimate model in the SAS LIFETEST procedure.

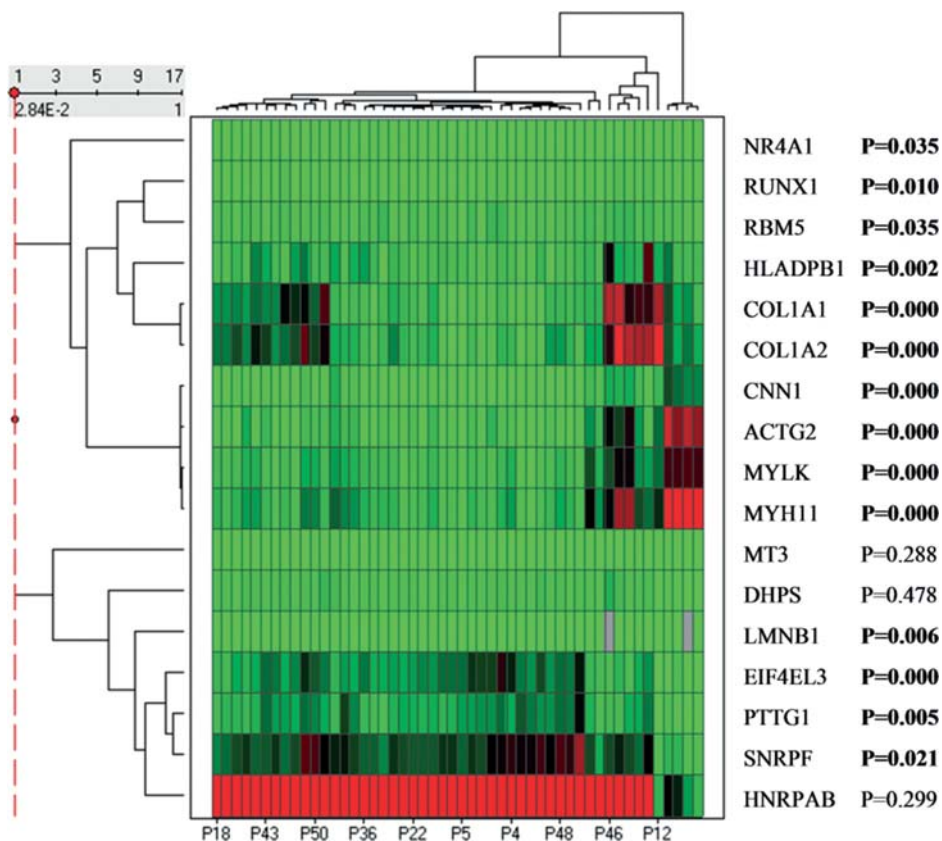


Figure 3. Hierarchical clustering heat map for tumor RNA expression in 52 CRC patients. Columns represent tumor samples and rows represent the 17 genes. Red and green represent relative high and low RNA expression levels, respectively.

that have attempted to predict prognosis of cancer patients on the basis of microarray analysis. The results from these metadata suggest that the list of genes identified as predictors

of prognosis was highly inconsistent and molecular signatures strongly depended on the selection of patients in the training sets. For all but one study, the rate of misclassification

decreased as the number of patients in the training set increased. Five of the seven studies did not classify patients better than chance. The conclusion is that the prognostic value of published microarray results in cancer studies should be considered with caution; and we advocate the use of validation by repeated random sampling (22).

Although no correlations between the tumor gene expression and clinical-pathological features were found in our study, hierarchic clustering based on the RNA expression in tumors has demonstrated the existence of 3 patient groups and all clinical-pathological parameters among those 3 groups were not significantly different, suggesting that the tumor RNA expression behaves in a different way on different groups of patients with the same stage. This might explain different biological behavior of tumors when classified by clinical-pathological parameters, even for patients in the same staging group. The characterization of molecular signatures among these groups may allow the determination of unique tumor biologic behaviors as well as different clinical evolution in the same group of patients, who are assigned into new different classifications on staging and may need different treatment options for each specific group.

Interactions between genes in a variety of cells are a major challenge for understanding of gene functions in complex cellular processes. How those genes work in a certain organ and how they interact are a complex process, which is the object of many studies. The big challenge will be to unveil how this interaction occurs among different genes biologically. In our study, there were two groups of genes that are positively correlated with each other in the RNA expression. The first group consisted of 10 genes and the other group of 6 genes. The normal tissue RNA expression of the 17 genes did not show such correlations in the groups. These findings suggest that there are interactions among those grouped genes, and require complementary studies to determine how these interactions occur.

In conclusion, our qPCR validation study confirms the importance of most 17-gene molecular signature genes with differential RNA expression, and suggests the survival relevance of PTTG1 in colorectal cancers. Additional studies should be done to define the real place and function of these 17 genes in colorectal tumors, their relationship with tumor metastasis, and the possible clinical applications.

#### Acknowledgements

This study was supported in part by the Siteman Cancer Center (P30 CA 091842), The BJH Foundation, the Pharmacogenetics Research Network (GM 63340; <http://pharmacogenetics.wustl.edu>), and the Section of Colorectal Surgery. The efforts of the members of the Siteman Cancer Center Tissue Procurement Core (Mark Watson) were greatly appreciated.

#### References

1. Jemal A, Siegel R, Ward E, Murray T, Xu J and Thun MJ: Cancer statistics, 2007. *CA Cancer J Clin* 57: 43-66, 2007.
2. Smith RA, Cokkinides V and Eyre HJ: Cancer screening in the United States, 2007: a review of current guidelines, practices, and prospects. *CA Cancer J Clin* 57: 90-104, 2007.
3. Niu R, Yan TD, Zhu JC, Black D, Chu F and Morris DL: Recurrence and survival outcomes after hepatic resection with or without cryotherapy for liver metastases from colorectal carcinoma. *Ann Surg Oncol* 14: 2078-2087, 2007.
4. Wong SC, Chan CM, Ma BB, *et al*: Clinical significance of cytokeratin 20-positive circulating tumor cells detected by a refined immunomagnetic enrichment assay in colorectal cancer patients. *Clin Cancer Res* 15: 1005-1012, 2009.
5. Watanabe T, Wu T and Catalano P: Molecular predictors of survival after adjuvant chemotherapy for colon cancer. *N Engl J Med* 344: 1196-1206, 2001.
6. Calistri D, Rengucci C, Seymour I, *et al*: KRAS, p53 and BRAF gene mutations and aneuploidy in sporadic colorectal cancer progression. *Cell Oncol* 28: 161-166, 2006.
7. Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I and Kerr D: Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer* 9: 489-499, 2009.
8. Khan A: Genomics and microarray for detection and diagnostics. *Acta Microbiol Immunol Hung* 51: 463-467, 2004.
9. Hegde P, Qi R, Gaspard R, *et al*: Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer Res* 61: 7792-7797, 2001.
10. Alizadeh A, Eisen M, Davis R, *et al*: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403: 503-511, 2000.
11. Sorlie T, Perou C, Tibshirani R, *et al*: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci* 98: 10869-10874, 2001.
12. van de Vijver M, He Y, van't Veer L, *et al*: A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347: 1999-2009, 2002.
13. Thykjaer T, Workman C, Kruhoffer M, *et al*: Identification of gene expression patterns in superficial and invasive human bladder cancer. *Cancer Res* 61: 2492-2499, 2001.
14. Jaeger J, Koczan D, Thiesen HJ, *et al*: Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues. *Clin Cancer Res* 13: 806-815, 2007.
15. Frederiksen C, Knudsen S, Laurberg S and Orntoft T: Classification of Dukes' B and C colorectal cancers using expression arrays. *J Cancer Res Clin Oncol* 129: 263-271, 2003.
16. Wang Y, Jatkoe T, Zhang Y, *et al*: Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J Clin Oncol* 22: 1564-1571, 2004.
17. Yu J, Shannon WD, Watson MA and McLeod HL: Gene expression profiling of the irinotecan pathway in colorectal cancer. *Clin Cancer Res* 11: 2053-2062, 2005.
18. Ramaswamy S: A molecular signature of metastasis in primary solid tumors. *Nat Genet* 33: 49-54, 2003.
19. Yu J, Mallon MA, Zhang W, *et al*: DNA repair pathway profiling and microsatellite instability in colorectal cancer. *Clin Cancer Res* 12: 5104-5111, 2006.
20. Pfaffl M: A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29: e45, 2001.
21. Michiels S, Koscielny S and Hill C: Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 365: 488-492, 2005.
22. Fritzmann J, Morkel M, Besser D, *et al*: A colorectal cancer expression profile that includes transforming growth factor beta inhibitor BAMBI predicts metastatic potential. *Gastroenterology* 137: 165-175, 2009.