

# A 'metastasis-prone' signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics

Yi Hong · Thomas Downey · Kong Weng Eu ·  
Poh Koon Koh · Peh Yean Cheah

Received: 26 October 2009 / Accepted: 22 January 2010  
© Springer Science+Business Media B.V. 2010

**Abstract** Metastasis is the major cause of cancer mortality. We aimed to find a metastasis-prone signature for early stage mismatch-repair proficient sporadic colorectal cancer (CRC) patients for better prognosis and informed use of adjuvant chemotherapy. The genome-wide expression profiles of 82 age-, ethnicity- and tissue-matched patients and healthy controls were analyzed using the Affymetrix U133 Plus 2 array. Metastasis-negative patients have 5 years or more of follow-up. A  $10 \times 10$  two-level nested cross-validation design was used with several families of classification models to identify the optimal predictor for metastasis. The best classification model yielded a 54 gene-set (74 probe sets) with an estimated prediction accuracy of 71%. The specificity, sensitivity, negative and positive predictive values of the signature are 0.88, 0.58, 0.84 and 0.65, respectively, indicating that the gene-set can improve prognosis for early stage sporadic CRC patients. These 54 genes, including node molecules YWHAB, MAP3K5, LMNA, APP, GNAQ, F3, NFATC2, and TGM2, integrate multiple bio-functions in various compartments into an intricate molecular network, suggesting that cell-wide perturbations are involved in metastasis transformation. Further, querying the 'Connectivity Map' with a subset (70%) of these genes shows that Gly-His-Lys

and securinine could reverse the differential expressions of these genes significantly, suggesting that they have combinatorial therapeutic effect on the metastasis-prone patients. These two perturbagens promote wound-healing, extracellular matrix remodeling and macrophage activation thus highlighting the importance of these pathways in metastasis suppression for early-stage CRC.

**Keywords** Early stage colorectal cancer · Genome-wide expression profiling · Metastasis predictor · Mismatch-repair proficient · Connectivity map query

## Introduction

Colorectal cancer (CRC) is one of the leading cancers in the developed world. In Singapore, it is the highest incidence cancer and the second leading cause of cancer death [1]. Metastasis to distant organ remains the main cause of mortality. Early stage CRC (Dukes' A/B or Stage I/II) is considered curative after surgery. However, up to 25% of these patients worldwide succumbed to metastasis and eventual early mortality [2]. Hence, other than histopathological classification, which is based solely on morphology, it is imperative to search for other means to stage tumors to improve management and reduce morbidity and mortality.

Moreover, these early stage patients are subjected to the same therapeutic regimen as advanced stage patients. Such therapeutic regimen may not yield an optimal benefit to risk ratio [3, 4]. Other chemotherapeutics with more efficiency and perhaps less toxicity may be more appropriate for early stage CRC patients.

The advent of genome-wide high-throughput technologies has brought about the possibility of using molecular

**Electronic supplementary material** The online version of this article (doi:10.1007/s10585-010-9305-4) contains supplementary material, which is available to authorized users.

Y. Hong · K. W. Eu · P. K. Koh · P. Y. Cheah (✉)  
Department of Colorectal Surgery, Singapore General Hospital,  
Singapore 169608, Singapore  
e-mail: cheah.peh.yean@sgh.com.sg

T. Downey  
Partek Incorporated, 12747 Olive Blvd, Suite 205, St. Louis,  
MO 63141, USA

biomarkers as prognostic indicators. Furthermore, such molecular-profiling studies provide a resource to link oncogenic pathways to potential chemotherapeutics [5]. Reasonable progress has been made in various cancers, notably in breast cancer [6]. Nevertheless, to date, there are only few publications pertaining to this in CRC. An early study reported a 23 gene-signature for the prediction of recurrence in Dukes' B Caucasian patients collected from several centres [2]. A 30-gene prognosis predictor for distant metastasis in Stage II Caucasian CRC patients was subsequently published [7]. A more recent study on Caucasian patients from two centres has identified a 50-gene signature for recurrence for early stage (I and II) CRC patients [8]. These gene-signatures do not share any gene in common. Further, it is not clear whether these signatures are reproducible or applicable to other populations. The 23-gene and 50-gene studies have no information on mismatch repair status of the patients.

In this retrospective study, we aimed to identify a metastasis-prone signature for early stage CRC by genome-wide expression profiling of 82 age-, ethnicity- and tissue-matched Singapore Chinese mismatch repair-proficient CRC patients and healthy controls collected from a single centre.

## Materials and methods

### Patients and healthy controls specimens

Patients of Han Chinese origin and aged 50 years or more and whose tumors were classified as early stage (Stage I/II) and microsatellite-stable were included. Patients whose tumors exhibit high microsatellite instability have different expression profiles and cancer etiology [9, 10] and were thus excluded from the study. These patients do not have clinicopathological features that fit the Bethesda's criteria for hereditary non-polyposis colorectal cancer (HNPCC). The mucosa of these patients harbored 3 or less adenomatous polyps and were therefore unlikely to be members of another familial CRC, familial adenomatous polyposis (FAP). The number of lymph nodes examined in these tumors was 10 or more to ensure accurate staging. Tumors with colonic perforation or with the resection margins not cleared were excluded. Only left-sided tumors (to the left of the splenic flexure) were included as the etiology of right and left-sided tumors are distinct [11]. These stringent inclusion criteria ensured that any differential expressions were attributable to genetic factors. Tumors were micro-dissected to enrich for tumor cells ( $\geq 90\%$ ).

Biopsies of normal-appearing colon tissues were obtained from individuals undergoing colonoscopic

examination and found to have no polyps and no known family history or previous CRC incidence. These were designated as healthy controls (HC).

Both patient and HC specimens were snap-frozen in liquid nitrogen within 30 min of removal from the colon and stored at  $-80^{\circ}\text{C}$ .

All patient and HC specimens were collected from the Singapore General Hospital (SGH). This study was approved by the Institutional Review Board of SGH.

### Genome-wide expression profiling

Total RNA was extracted from each specimen and biotinylated cRNA targets were prepared with 5  $\mu\text{g}$  of total RNA according to manufacturer's protocols (Affymetrix, Santa Clara, CA). Targets were hybridized to GeneChip Human Genome U133 Plus 2.0 Arrays. The arrays were washed and stained on the fluidics station and scanned with GeneChip scanner 3000 as previously described [12].

The microarray data set is submitted to the GEO repository (GSE9348) at <http://www.ncbi.nlm.nih.gov/geo/info/linking.html>.

### Statistical analysis

Statistical analyses were performed using Partek<sup>®</sup> Genomics Suite<sup>™</sup> (Partek GS) version 6.4, (Partek Inc., St. Louis, MO) software into which Affymetrix CEL files were directly imported using robust multi-chip average (RMA). A two-level nested cross-validation design was used to select an optimal classifier for metastasis from multiple classification models and to estimate the accuracy of the optimal classifier ( $n = 70$ ). In a  $10 \times 10$  two-level nested cross-validation approach, an 'outer' cross-validation was performed to produce an unbiased estimate of prediction error by holding out 10% of the samples as an independent test set; an additional 'inner' 10-fold cross-validation was performed on the remaining 90% samples as the training set to select the optimal model to be applied to the held out test set. This process is repeated leaving a different 10% of the samples out (to ensure that the held-out samples were not used to train the classifier) and accuracy estimates were accumulated from the results of all samples. Thus, unlike the single-level cross-validation approach which tends to overestimate the prediction accuracy, the two-level nested cross-validation reduces the bias inherent in the single-level cross-validation and gives an error estimate that is very close to that obtained from an independent test set [13]. The two-level nested cross-validation approach makes more efficient use of the limited sample size than simply splitting the samples into separate training and test sets.

Four different families of classification models were evaluated: K-nearest neighbor, nearest centroid, discriminant analysis and support vector machine. Within the inner cross-validation, the ANOVA model was re-applied with a different 10% samples held-out as test samples each time [14].

Details of the statistical analysis, microsatellite instability assay and real time PCR analysis are in the Supplementary Methods.

#### Connectivity map query

The Connectivity Map build 02 is a database of 7000 genome-wide expression profiles after the treatment of 4 human cell lines with 1,309 bioactive small molecules for a total of 6,100 instances [15]. This map provides an in silico method to connect human diseases with the genes that underlie them and drugs that treat them [16]. The Connectivity Map is freely available at: <http://www.broad.mit.edu/cmap>. Currently it is based on Affymetrix U133 A (which constitute part of the U133 Plus 2) array. The ‘metastasis-prone’ probe sets were used to query the Connectivity Map to identify any bioactive molecules or perturbagens that could possibly serve as novel therapeutics. The criterion for selection is high negative connectivity and enrichment scores which are calculated based on the Kolmogorov–Smirnov statistics. Negative scores indicate that the corresponding perturbagen reversed the expression of the query probe sets. Specificity is defined as the frequency at which enrichment of a set of instances is equalled or exceeded by the enrichment of that same set of instances produced from queries executed with 312 published, experimentally-derived signatures extracted from MSigDB.

## Results

#### Clinical features of patients and healthy control

Seventy Stage II Chinese patients were identified. There were no Stage I patient that fitted the stringent inclusion criteria. Except for one patient whose DNA was not available for MSI typing, all other tumors were typed to be microsatellite-stable. Twenty of the patients eventually succumbed to metastasis or recurrence (Table 1). Five patients from the metastasis-positive subgroup underwent chemo- and radiotherapy post-metastasis.

All patients classified as metastasis-negative had at least five (range 5.0–8.5) years of follow up. The mean age and number of lymph nodes examined was comparable between metastasis-positive and metastasis-negative patients. Although there were more male than female patients amongst the metastasis-negative patients, this was representative of the gender ratio of Singapore Chinese

CRC patients. The propensity to metastasis, however, was not significantly different between male and female patients (Table 1). Interestingly, unlike in breast carcinomas, there was no significant correlation of tumor size and metastasis status of the patients ( $P = 0.26$ , Table 1).

Twelve Han Chinese healthy controls (HC) aged 50 or more whose colonic expressions served as baselines were recruited. These HC underwent colonoscopies for reasons ranging from abdominal pain, change in bowel habits and bleeding. All those with primary bleeding were found to have hemorrhoids.

Genome-wide expression profiling and  $10 \times 10$  two-level nested cross-validation identified 54 genes (74 probesets)

The complementary RNAs generated from both HC and patient specimens were of high and comparable integrity. The mean percentage of transcripts scored as ‘present’ was very similar between HC and patient subgroups (Table 1). Furthermore, the mean 3′/5′ ratios of the internal control, GAPDH, for both the HC and patient specimens were similar and well below the threshold of 3.0 (Table 1).

The Nearest Centroid model gave the highest normalized accuracy of 71% when the  $10 \times 10$  two-level nested cross-validation was applied for several classification models (html report in supplementary Table S1). The specificity, sensitivity, negative predictive value (NPV) and positive predictive value (PPV) of the signature are 0.88, 0.58, 0.84 and 0.65, respectively (Table 2). The classifier with the best prediction identified 74 probesets. Thirty-five probe sets were up-regulated and thirty-nine probe sets were down-regulated in metastasis-positive patients compared to metastasis-negative patients, respectively, with  $P$ -values ranging from  $9.0E-8$  to  $1.8E-4$ .

Principal component analysis (PCA) indicated that the 74 probesets metastasis-prone signature separated the patients into two distinct clusters (Fig. 1a). Further, we performed PCA with the 74 probesets on a series of sporadic (aged 50 or more), Caucasian Dukes’ A/B CRC patients with no family history ( $n = 56$ ) identified from the Oncomine database (<http://www.oncomine.org>; GSE2109). The PCA indicated that nine of the specimens formed a sub-cluster away from the rest of the specimens (Fig. 1b).

There was no significant difference in the expression of these 74 probe sets between male and female individuals for all samples (with  $P$ -values ranging from 0.09 to 1.0).

Data mining reveals roles in diverse bio-functions and tumorigenesis pathways

The 74 probe sets were classified into three groups according to their expressions in the three sub-groups (HC, Met<sup>-</sup>,

**Table 1** Clinicopathological features of patients and healthy controls

Features	Patients <sup>a</sup>			Healthy controls (n = 12)
	Mets <sup>+</sup> (n = 20)	Mets <sup>-</sup> (n = 50)	P value	
Age, mean (range)	66.1 (51–93)	68.6 (52–89)	0.47	59.3 (51–67)
Sex				
Female	10	20	0.45	6
Male	10	30		6
T stage				
T2	3	7	0.34	–
T3	17	42		–
T4	0	1		–
Tumor grade				
Well differentiated (G1)	4	6	0.77	
Moderately differentiated (G2)	15	44		–
Poorly differentiated (G3)	1	0		
No. of L.N. <sup>b</sup> examined				–
10–15	13	30	0.70	
>15	7	20		
Mean	15.2	17.2		
Mean tumor size/cm <sup>3</sup> (range)	15.7 (2.3–48.8)	29.1 (1.2–270)	0.26	–
Tumor location				
Rectum	10	17	0.22	–
Left colon	10	33		
Metastasis site				
Liver	1	0	NA	
Lung	8	0		–
Multiple <sup>c</sup>	7	0		
Other <sup>d</sup>	4	0		
Mean survival/month (range)	46 (11–87)	81 (61–101) <sup>e</sup>	2.3E-12	–
Microarray report				
Mean present gene call (%)	42.0	41.5		40.1
Mean GAPDH 3'/5'	1.39	1.25		1.19

<sup>a</sup> All patients had tumors classified as Stage II (T<sub>2-4</sub>N<sub>0</sub>M<sub>0</sub>) at time of operation; metastasis occurs on average 19 months after operation for Mets<sup>+</sup> subgroup

<sup>b</sup> Lymph node

<sup>c</sup> Multiple metastasis sites that included lung or liver

<sup>d</sup> Other sites including two patients with local recurrence only

<sup>e</sup> Patients whose cause of death is 'other than cancer' are censored

**Table 2** 10 × 10 two-level nested cross-validation test summary accuracy (normalized) estimate for nearest centroid model = 70.7%

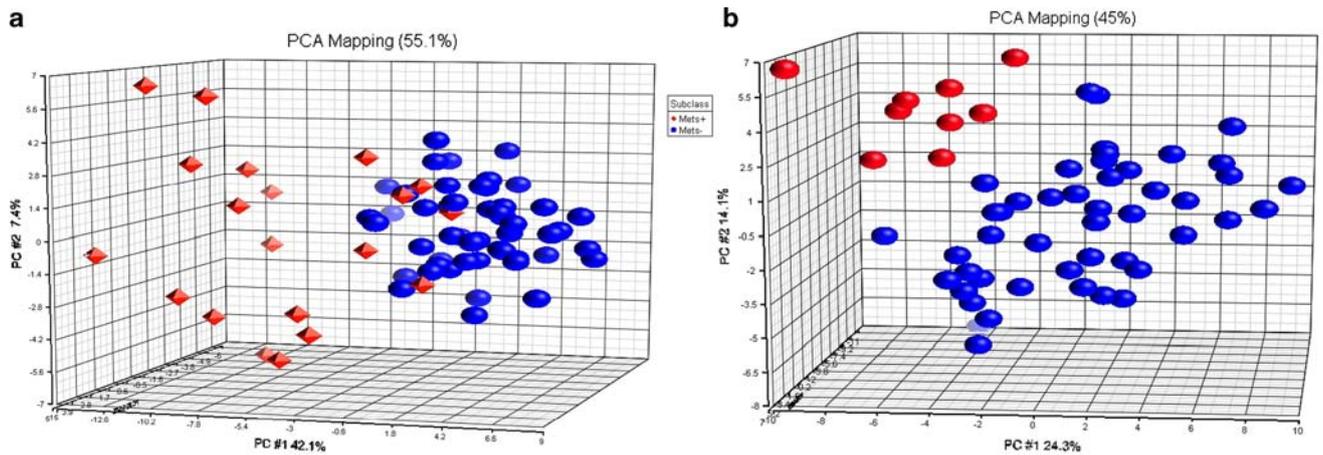
Predictive			
Real	Mets <sup>+</sup>	Mets <sup>-</sup>	
Mets <sup>+</sup>	11.5	8.5	Sensitivity
	TP	FN	0.58
Mets <sup>-</sup>	6.2	43.8	Specificity
	FP	TN	0.88
	PPV: 0.65	NPV: 0.84	

TP true positive, FN false negative, FP false positive, TN true negative

Met<sup>+</sup>) of specimens. Probe sets were classified into group 1 if there was no significant difference between expression of metastasis-negative and HC specimens, and group 2 if there

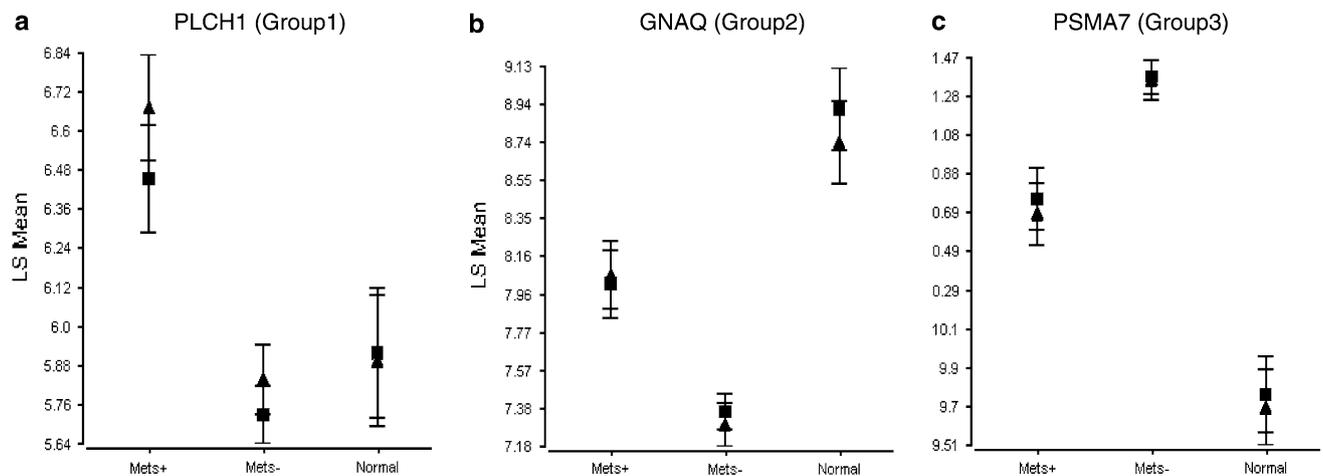
was a significant difference between expression of metastasis-negative and HC specimens (Fig. 2). They were classified into group 3 when the metastasis-negative specimens had the highest expressions. Majority (77%) of the probe sets were classified in groups 2 and 3.

These 74 probe sets representing 54 genes were fed into the NetAffy website and Ingenuity Pathway analysis (IPA) databases for further annotation. They were significantly linked to diverse biological functions in various cellular compartments such as cell morphology, cellular and embryonic development, cancer, DNA replication, recombination and repair, immune cell trafficking and nucleic acid metabolism (supplementary Table S2). These biological functions were further linked into several networks that were merged to highlight the importance of several node molecules such as YWHAB, MAP3K5, LMNA, APP,



**Fig. 1** (a) PCA plot of the 70 Singapore CRC specimens with the 54-gene (74 probesets) predictor. Diamonds and circles are clinically defined metastasis-positive and metastasis-negative specimens

respectively. (b) PCA plot of the 56 Caucasian Dukes A/B CRC specimens identified from Oncomine with the 54-gene predictor



**Fig. 2** Differential expressions of representative genes (a–c) in the 3 groups between metastasis-positive (Met<sup>+</sup>), metastasis negative (Met<sup>-</sup>) and normal (HC) specimens. Square and triangle represent male and female individuals respectively

GNAQ, F3, NFACTC2 and TGM2 (Fig. 3). These molecules play important roles in known tumorigenesis and metabolic pathways such as protein ubiquitination, ERK5, IGF-1, apoptosis, JNK, 14-3-3-mediated and PI3K/AKT signalling pathways.

Validation of gene expression with quantitative real-time PCR

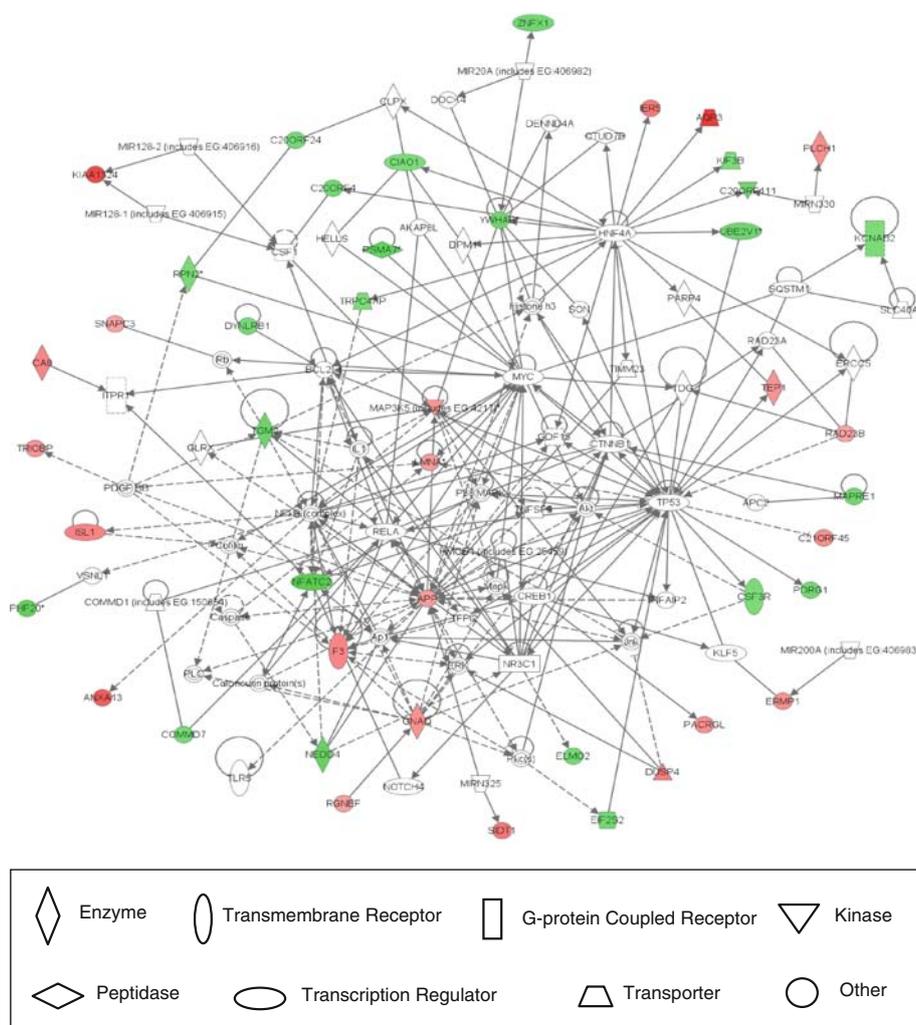
To verify that the genes were indeed differentially expressed between metastasis-positive and metastasis-negative patients (and not due to microarray artefact), two representative genes, aquaporin 3 and matrilin 2 were quantified by real-time PCR. The relative level of aquaporin 3 and matrilin 2 expression in the metastasis-positive patients were 3.5 (95% CI: 2.0–6.1) and 2.3 (95% CI: 1.5–3.5) fold that of the metastasis-negative patients respectively indicating that they

were indeed up-regulated in metastasis-positive patients. The expressions of aquaporin 3 and matrilin 2 by microarray and real-time PCR were significantly positively correlated ( $R^2 = 0.8381, P = 1.5 \text{ E-}8$ ; supplementary Fig. S1).

Connectivity map query identifies Gly-His-Lys and securinine as possible perturbagens

Thirty-eight of the 54 genes were in the Affymetrix U133 A array and thus could be used to query the ‘Connectivity Map’. The query returned Gly-His-Lys (GHK) and securinine as perturbagens with high negative connectivity scores in all seven instances (rank 5745–6057 of 6100 instances), resulting in highly significant enhancement scores (Table 3). The specificity for both molecules is 0.00000, demonstrating the uniqueness of the connectivity between the instances and the ‘metastasis-prone’ signature.

**Fig. 3** Molecular networks connecting the genes in the ‘metastasis-prone’ signature by bio-functions. *Green* and *Red* represent genes that are down- and up-regulated respectively in Met<sup>+</sup> specimens compared to Met<sup>-</sup> specimens. *Solid* and *dotted* lines represent direct and indirect interactions between the genes respectively



**Discussion**

We report a 54 gene metastasis-prone signature for sporadic early-stage mismatch-repair proficient CRC patients. The results suggest that it is possible to estimate the probability of metastasis in sporadic stage I/II CRC patients based on the expressions of genes from the primary tumors. The data thus support earlier findings that most transformation events for metastasis are already present in the primary carcinomas [17, 18] and that few further events are necessary for progression to metastasis [18].

The nearest centroid classifier with 10 × 10 two-level nested cross-validation design yielded a 54 gene-set signature with the best accuracy estimate of 71%. The NPV and PPV of the signature are 0.84 (95% CI = 0.72–0.92) and 0.65 (95% CI = 0.37–0.89), respectively. Accordingly, for a Stage II CRC patient who is predicted to be metastasis-negative by the signature, the probability that he will remain metastasis-free is 0.84 suggesting that adjuvant

therapy is probably not necessary for the individual. In contrast, for a Stage II CRC patient who is predicted to be metastasis-positive, the odds that he will eventually succumb to metastasis is 65% indicating that adjuvant therapy should be considered. The relatively lower sensitivity and PPV could be because target organs of metastasis (e.g. liver, lung or bone) were not available for profiling. Target organ microenvironment was known to influence implantation of cancer cells [19]. However, it would be highly unlikely that early stage CRC patients would consent to organ biopsies for profiling purpose and hence their applicability to prognosis would be limited. The NPV and PPV of the signature compared favorably with that of the 70-gene MammaPrint signature for breast cancer (NPV and PPV are 88 and 52%, respectively) in its original cohort [20]. Further, although no further clinical follow-up from the Caucasian series (GSE2109) was available to indicate their metastasis status, PCA on the series with the 54 genes clearly showed that a portion (16%) of the specimens was in a sub-cluster different from the rest of the specimens,

**Table 3** Connectivity map query identified Gly-His-Lys and securinine as perturbagens

Barview	Rank	cmap name	Dose ( $\mu\text{M}$ )	Cell	Score	Enrichment	<i>P</i>
	5745	Securinine	18	MCF7	-0.669	-0.942	0.00000
	5800	Securinine	18	HL60	-0.691		
	5885	Securinine	18	MCF7	-0.726		
	6043	Securinine	18	PC3	-0.846		
	5992	Gly-His-Lys	1	PC3	-0.789	-0.983	0.00002
	6004	Gly-His-Lys	1	MCF7	-0.795		
	6057	Gly-His-Lys	1	PC3	-0.867		

Rank refers to the order of the instance in the query; the instances are ranked according to the connectivity score in descending order from +1 to -1. The *barview* is a graphical representation of all these instances in descending order of connectivity score

suggesting that the 54 genes were able to separate an independent series of patients into a majority and minority clusters (Fig. 1b).

Another interesting revelation is that the expression profiles of the 74 probe sets from metastasis-positive specimens may not necessarily be significantly different from the HC (Fig. 2). There were no instances where the expression of the probe sets is the highest or the lowest in the metastasis-positive specimens compared to the metastasis-negative and HC specimens. This implies that there is a delicate homeostatic balance in the expression profiles of all genes between the three states (HC, Met<sup>-</sup>, Met<sup>+</sup>) and the expressions of these genes are very dynamic. It further suggests that the expression of any gene cannot be inferred from its expression in the other two states. For example, the expression of PSMA7 in metastasis-positive specimens was between that of the metastasis-negative and HC specimens, giving an ‘inverted-U’ profile (Fig. 2c).

There was no significant difference in the expression of these 54 genes with respect to gender indicating that hormonal pathways probably do not play important roles in metastasis progression for early stage CRC. This is also reflected in the fact that although the ratio of males to females in the metastasis-negative patients is 3:2, both genders were equally represented in the metastasis-positive group (Table 1).

Interestingly, although C-Myc, TP53 and CTNNB1 appear as node molecules on the molecular network (Fig. 3), they were not amongst the 54 genes in the signature, implying that their expressions were not

significantly perturbed between the Met<sup>-</sup> and Met<sup>+</sup> specimens and hence cannot serve as biomarkers.

Comparing the 54-gene metastasis-prone signature with that of the 23-gene signature from an earlier Caucasian study [2] revealed only one closely related gene, YWHAB and YWHAH respectively. None of the genes in the 54-gene signature from this study overlapped with that of the 30- or 50-gene signatures from two other Caucasian studies [7, 8] reaffirming the likelihood of different combination of genes with similar functions could act in concert towards common endpoints. Nevertheless, deploying these gene predictors on our series did not cluster the patient specimens as well as our 54-gene predictors (supplementary Fig. S2), indicating the possibility of real population differences.

A previous study querying the ‘Connectivity Map’ has identified inhibitors of the PI3K-AKT-MTOR pathway as potential therapeutics for mismatch repair-deficient CRC [21]. Our query of the later version of the ‘Connectivity Map’ query has returned GHK and securinine as candidate compounds or perturbagens that can significantly reverse the expressions of 70% of the 54 genes in the signature with minimum specificity indicating their unique connectivity with the signature. The negative enhancement was consistently achieved at low dosage in several different cell lines demonstrating their robustness *in vitro*. NHK is a potent wound healing agent and activator of extracellular matrix (ECM) synthesis and remodelling [22, 23]; securinine is a novel immune cell agonist [24, 25]. The identification of these two perturbagens from a host of 1,309 active bio-molecules indicates that the wound healing and

ECM remodelling pathways as well as macrophage activation are probably crucial in suppressing metastasis in early stage mismatch-repair proficient CRC patients. Since these two small molecules function at low dosage and hence low toxicity (Table 3), they are ideal candidates as adjuvant chemotherapeutics. Further experimentation is thus warranted.

**Acknowledgments** The authors thank Ms. Yu Hui Wong, Mr. Huashi Ding and Dr. Soo Chin Liew for technical assistance, the Department of Clinical Research, SGH for the use of the Affymetrix Fluidics station, and the Singapore Polyposis Registry for clinical data retrieval. This work is supported in part by a grant from the National Medical Research Council, Singapore (NMRC/0988/2005) to P.Y. Cheah.

## References

- Chia KS, Seow A, Lee HP et al (2000) Cancer incidence in Singapore 1993–1997. Singapore Cancer Registry Report: No. 5
- Wang Y, Jatkoe T, Zhang Y et al (2004) Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J Clin Oncol* 22:1564–1571
- De Gramont A, Boni C, Navarro M et al (2007) Oxaliplatin/5FU/LV in adjuvant colon cancer: updated efficacy results of the MOSAIC trial, including survival, with a median follow-up of six years. *Proc AM Soc Clin Oncol* 25:165s (suppl; abstr 4007)
- O'Dwyer PJ, Eckhardt SG, Haller DG et al (2007) Priorities in colorectal cancer research: recommendations from the Gastrointestinal Scientific Leadership Council of the Coalition of Cancer Cooperative Groups. *J Clin Oncol* 25:2313–2321
- Bild AH, Potti A, Nevins JR (2006) Linking oncogenic pathways with therapeutic opportunities. *Nat Rev Cancer* 6:735–741
- van't Veer LJ, Dai H, van de Vijver MJ et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530–536
- Barrier A, Boelle P, Roser F et al (2006) Stage II colon cancer prognosis prediction by tumor gene expression profiling. *J Clin Oncol* 24:4685–4691
- Garman KS, Acharya CR, Edelman E et al (2008) A genomic approach to colon cancer risk stratification yields biologic insights into therapeutic opportunities. *Proc Natl Acad Sci USA* 105:19432–19437
- Kruhoffer M, Jensen JL, Laiho P et al (2005) Gene expression signatures for colorectal cancer microsatellite status and HNPCC. *Br J Cancer* 92:2240–2248
- Giacomini CP, Leung SY, Chen X et al (2005) A gene expression signature of genetic instability in colon cancer. *Cancer Res* 65:9200–9205
- Watanabe T, Kobunai T, Toda E et al (2006) Distal colorectal cancers with microsatellite instability (MSI) display distinct gene expression profiles that are different from proximal MSI cancers. *Cancer Res* 66:9804–9808
- Hong Y, Ho KS, Eu KW et al (2007) A susceptibility gene set for early onset colorectal cancer that integrates diverse signalling pathways: implication for tumorigenesis. *Clin Cancer Res* 13:1107–1113
- Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. *BMC Bioinform* 7:91
- Ambroise C, McLachlan GJ (2002) Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci USA* 99:6562–6566
- Lamb J, Crawford ED, Peck D et al (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929–1935
- Lamb J (2007) The connectivity map: a new tool for biomedical research. *Nat Rev Cancer* 7:54–60
- Ramaswamy S, Ross KN, Lander ES et al (2003) A molecular signature of metastasis in primary solid tumors. *Nat Genet* 33:49–54
- Jones S, Chen W, Parmigiani G et al (2008) Comparative lesion sequencing provides insights into tumor evolution. *Proc Natl Acad Sci USA* 105:4283–4288
- Minn AJ, Gupta GP, Siegel PM et al (2005) Genes that mediate breast cancer metastasis to lung. *Nature* 436:518–524
- Wittner BS, Sgroi DC, Ryan PD et al (2008) Analysis of the MammaPrint breast cancer assay in a predominantly postmenopausal cohort. *Clin Cancer Res* 14:2988–2993
- Vilar E, Mukherjee B, Kuick R et al (2009) Gene expression patterns in mismatch repair-deficient colorectal cancers highlight the potential therapeutic role of inhibitors of the phosphatidylinositol 3-kinase-AKT-Mammalian target of Rapamycin pathway. *Clin Cancer Res* 15:2829–2839
- Simeon A, Emonard H, Hornebeck W et al (2000) The tripeptide-copper complex glycyl-L-histidyl-L-lysine-Cu<sup>2+</sup> stimulates matrix metalloproteinase-2 expression by fibroblast cultures. *Life Sci* 67:2257–2265
- Pickart L (2008) The human tri-peptide GHK and tissue remodelling. *J Biomater Sci Polym Ed* 19:969–988
- Dong NA, Gu ZL, Chou WH et al (1999) Securinine induced apoptosis in human leukemia HL-60 cells. *Zhongguo Yao Li Xue Bao* 20:267–270
- Lubick K, Radke M, Jutlia M (2007) Securinine, a GABAA receptor antagonist, enhances macrophage clearance of phase II *C.burnetii*: comparison with TLR agonists. *J Leukoc Biol* 82:1062–1069