uicc
global cancer control

IJC
International Journal of Cancer

A blood-based biomarker panel for stratifying current risk for colorectal cancer

Kenneth Wayne Marshall[1], Steve Mohr[1], Faysal El Khettabi[1], Nadejda Nossova[1], Samuel Chao[1], Weisheng Bao[1], Jun Ma[1], Xiao-jun Li[1] and Choong-Chin Liew[1,2]

[1] GeneNews Ltd., Richmond Hill, Ontario, Canada
[2] Brigham and Women's Hospital, Harvard Medical School, Boston, MA

Colorectal cancer (CRC) is often curable and preventable using current screening modalities. Unfortunately, screening compliance remains low, partly due to patient dissatisfaction with faecal/endoscopic testing. Recent guidelines advise CRC screening should begin with risk stratification. A blood-based test providing clinically actionable CRC risk information would likely improve screening compliance and enhance clinical decision making. We analyzed 196 gene expression profiles to select candidate CRC biomarkers. qRT-PCR was performed on 642 samples to develop a 7-gene biomarker panel using 112 CRC/120 controls (training set) and 202 CRC/208 controls (independent, blind test set). Panel performance characteristics and disease prevalence (0.7%) were then used to develop a scale assessing an individual's current risk of having CRC based on his/her gene signature. A 7-gene panel (*ANXA3, CLEC4D, LMNB1, PRRG4, TNFAIP6, VNN1* and *IL2RB*) discriminated CRC in the training set (area under the receiver-operating-characteristic curve (ROC AUC), 0.80; accuracy, 73%; sensitivity, 82%; specificity 64%). The independent blind test set confirmed performance (ROC AUC, 0.80; accuracy, 71%; sensitivity, 72%; specificity, 70%). Individual gene profiles were compared against the population results and used to calculate the current relative risk for CRC. We have developed a 7-gene, blood-based biomarker panel that can stratify subjects according to their current relative risk across a broad range in an average-risk population. Across the continuous spectrum of risk as defined by the current relative risk scale, it is possible to identify clinically meaningful reference points that can assist patients and physicians in CRC screening decision making.

Colorectal cancer (CRC), the third most frequently diagnosed cancer in men and women in the United States and the United Kingdom, carries an overall population lifetime risk of about 5%.[1,2] Despite being among the most preventable of neoplasms and surgically curable in early stages, cancer of the colon and rectum remains the second leading cause of cancer death in the western world. In the United States, ~150,000 people will be diagnosed with CRC in 2008 and some 50,000 will die of their disease.[1] Each year in the United Kingdom, about 36,500 people receive a diagnosis of CRC and some 16,000 die of it.[2]

Key words: colorectal cancer, biomarkers, blood test, disease risk, stratification

Abbreviations: CURR: current relative risk; CRC: colorectal cancer; NPV: negative predictive value; PPV: positive predictive value; qRT-PCR: real-time polymerase chain reaction; ROC AUC: area under the receiver-operating-characteristic curve; TNM: tumor-node-metastasis

Grant sponsor: GeneNews Ltd., Richmond Hill
DOI: 10.1002/ijc.24910
History: Received 31 Jul 2009; Accepted 8 Sep 2009; Online 30 Sep 2009

Correspondence to: Choong-Chin Liew, GeneNews Ltd., 2 East Beaver Creek Road, Building 2, Richmond Hill, Ontario L4B 2N3, Canada, Fax: +905-739-2031, E-mail: cliew@genenews.com

Most CRC arises from precursor adenomatous polyps, developing over many years.[3] Stage at detection is critically related to patient survival. Localized cancers (tumor-node-metastasis [TNM] Stages I–II) have an excellent 5-year survival prognosis (93% and 83%); regional stage (TNM Stage III) patients have a 5-year survival rate about 60%; only 8% of patients with late stage (TNM Stage IV) disease will survive 5 years.[4] These features make CRC eminently suitable for a screening program, and health authorities have long promoted screening for CRC in average-risk adults, beginning at the age of 50 years.[1,5,6]

Despite repeated recommendations and awareness campaigns, however, populations have resisted CRC screening. Paradoxically, although 90% of respondents in studies express high interest in cancer screening in general and CRC screening in particular,[7,8] screening compliance remains low. Only about one-half of age-eligible Americans are current with recommended faecal- or endoscopic-based tests.[9] In Canada, only 24% of the target groups have ever been screened and a mere 18% are up-to-date with recommendations.[10] These rates are much lower than compliance for breast and cervical cancer screening, which range from 70–79% for mammography and for the Pap test.[11]

Low compliance reflects in large part the unpleasant nature of faecal procedures with varying degrees of dietary restriction and requiring multiple stool samples, and endoscopic procedures, which require dietary restriction for colon

cleansing, sedation and the necessity for taking time off work and help getting home.[12] Endoscopic procedures can also result in serious complications. A recent Kaiser Permanente study of ~16,000 patients documented complications requiring hospital admission such as perforations and bleeding in 5 of every 1,000 patients undergoing colonoscopy.[13] It is clear that a safe, noninvasive blood test, which encourages patient compliance, would be a welcome addition to the CRC screening armamentarium.

We and others have demonstrated that RNA profiling in whole blood can be used to develop molecular signatures of disease across a broad spectrum of pathology.[14,15] Following a preliminary study,[16] for this study of 314 CRC patients and 328 controls, we characterized a 7-gene biomarker panel for discriminating CRC based on patients' blood RNA samples. These biomarkers were selected after changing the collection tubes from EDTA to Paxgene and the PCR from SYBR to duplex TaqMan® chemistries to improve the assay's performance in terms of specificity and stability. We then developed a test based on the performance characteristics of the 7-gene panel in conjunction with the prevalence of CRC (0.7%) in the average-risk population[17] that enables us to assess a patient's current relative risk of having CRC. We further demonstrated the benefit of stratifying patients based on their current relative risk of having CRC in the context of CRC screening in the general population, similar to a stratification strategy proposed for breast cancer prevention.[18]

Material and Methods

Patient samples

Blood samples were taken from screening colonoscopy subjects at 25 centers located primarily in the Greater Toronto Area and surrounding region and also in the United States. Samples were collected over a period of 3 years, from March 2005 to March 2008. The uniformity of collection procedures at the different sites was ensured through following identical study protocol, uniform training of personnel and monitoring of the sites for protocol adherence. Informed consent was obtained according to protocols approved by each institution's Research Ethics Board. Initially, all subjects were enrolled at the colonoscopy clinics; however, the low incidence rate of CRC in this population meant that most samples collected were confirmed to be noncancer. As a result, it became necessary to augment the number of cancer samples using samples from cancer clinic patients with positive colonoscopy results. Blood samples in these cases were collected before any form of treatment, including surgery. Patients enrolled at colonoscopy clinics donated blood before the colonoscopy. Samples were categorized after pathologist reports were reviewed. Controls comprised samples from subjects with no pathology at colonoscopy; disease blood samples were from colonoscopy-confirmed CRC patients, who had not undergone CRC treatment. Institutional pathologists determined cancer stage.

Blood collection and RNA isolation

For microarray study, samples of peripheral whole blood (10 ml) were collected in EDTA Vacutainer (Beckton Dickinson) tubes (to avoid the high globin transcript problem associated with the PAXgene system) and processed as described previously.[16]

For qRT-PCR, blood collected in PAXgene™ tubes (PreAnalytiX) was processed according to PAXgene™ Blood RNA Kit protocol. The PAXgene system is more suitable for RT-PCR studies and clinical applications due to its ability to immediately stabilize RNA and to keep it stable over a longer period of time, thereby providing greater flexibility in sample collection.

For all samples, RNA quality was assessed using a 2100 Bioanalyzer RNA 6000 Nano Chip (Agilent Technologies). All samples met quality criteria: RIN $\geq$ 7.0; 28S:18S rRNA ratio $\geq$ 1.0, and a validated Agilent bioanalyzer scan. RNA quantity was determined by absorbance at 260 nm in a DU640 Spectrophotometer (Beckman-Coulter).

Microarray hybridization

Microarray hybridizations were carried out on whole blood samples to generate gene expression profiles from CRC and control subjects and to identify potential CRC biomarkers for subsequent validation by qRT-PCR. Standard protocols, established in GeneNews, were followed in blood sample processing, RNA extraction and purification, probe labeling and hybridization. Five micrograms of total RNA per sample was used for hybridization, following standard Affymetrix protocol. All hybridizations were probe-level processed by GC-RMA using GeneSpring. Unreliable measurements, identified by the crossgene-error model built in GeneSpring, were removed from further analysis.

Microarray sample size calculation. The sample size calculation for microarrays was based on data published in early 2008, which estimated that 100 samples per group are required to achieve adequate power (0.80) with a Type I error less than 0.05 and a fold change over 1.2 for a large proportion (over 75%) of genes being investigated.[16] We used SAM package under R software (as described in Ref. 19).

Blood samples were collected from 97 CRC and 99 control subjects. Samples were matched for sex, age, body mass index (BMI), ethnicity, comorbidity and medication. A total of 196 blood expression profiles were generated by Affymetrix U133Plus2.0 GeneChips.

Hybridization data processing and normalization. To assess whether there was any batch effect, principal component analysis (PCA) was used, and different factors, including chip lot, hybridization date, sample collection site, were labeled on PCA plots. It was noticed that the hybridization date seemed to be the main batch effect factor, and it was decided to remove this effect by using mean-centering on GeneSpring.

Hybridization quality analysis and outlier detection. All hybridizations passed the quality thresholds for Affymetrix GeneChip suggested by the manufacturer. A number of hybridizations showed larger deviation in certain quality control parameters from the rest.

More detailed analyses using Pearson's correlation of the expression profiles and PCA plots identified 7 hybridizations, which were among the top 10 highest in GAPDH and ACTIN 3′/5′ ratios, suggesting the deviation of these samples from the rest might be caused by lower RNA integrity. The decision was made to exclude these 7 hybridizations from further analysis, resulting in a final total of 189 samples (98 controls and 91 CRC) for downstream analysis.

Quantitative reverse-transcriptase polymerase chain reaction

Calculation of sample size for RT-PCR. In the computation of the sample size, we used a significance level $\alpha = 0.05$ in each group, to detect a true difference in means of $\Delta \neq 0$ with power at least $1 - \beta$ is

$$n = \frac{(z_\alpha + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{\Delta^2},$$

where σ_1 is the standard deviation in the control group and σ_2 is the standard deviation in the CRC group, given $z_\alpha = 1.645$ and a power equal to 0.9 ($\beta = 0.1$ given $z_\beta = 1.28$). To compute the standard deviations and the difference delta, we randomly selected a number of cohorts using 15 control/15 CRC and 30 control/30 CRC in bootstrap sampling. The calculation was based on the data as described previously[16] and current data, which indicated that at least 67 samples per group are required.

Primers and probes for RT-PCR assay. Primers and probes were designed with Primer3 software.[20] Primers had to amplify the same transcript as the Affymetrix probeset that was selected from the microarray study. Preference was given to primers matching the region of the Affymetrix probeset. Primers, probe or the amplicon had to span an exon–exon junction to avoid amplification of genomic DNA. The primers and probe must also be specific to genes of interest and not able to amplify any other products. Genes were tested both in singleplex and in duplex reaction conditions, and similar Ct values were observed for the same gene in each condition, indicating that the expression levels of the genes were not affected by the presence of a duplex partner in the reaction well.

One microgram of RNA was reverse transcribed into single-stranded complementary DNA (cDNA) using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) in a 20 μL reaction volume. For PCR, 20 ng cDNA was mixed with QuantiTect® Probe PCR Master Mix (Qiagen) and TaqMan® dual-labeled probe and primers corresponding to the gene-of-interest and denominator in a 25 μL reaction volume. PCR amplification was performed using a 7500 real-time PCR system (Applied Biosystems).

Quality assurance processes included verification of negative template control for lack of amplification, review of amplification curve shape for adequate signal, difference between duplicate wells and stability of the positive reference sample. Samples that failed these quality control checks were repeated. Samples that failed a second time were excluded from the analysis.

Validation of the biomarkers. Quantitative RT-PCR experiments were performed in 2 phases. First, each gene of interest was assayed in the 232-sample training set in duplex with an endogenous reference gene (*ACTB*; beta actin) to identify genes with statistically significantly different expression levels between CRC and controls (see "Characterization of 7-gene panel" section later, for more details). Gene expression differences were estimated using the "comparative cycle threshold (ΔCt) method" of relative quantification,[21] normalizing the Ct values relative to the reference gene. This was performed by calculating $\Delta Ct_{sample} = Ct_{targetgene} - Ct_{referencegene}$. The relative fold-change (CRC *vs.* controls) was represented as $2^{-\Delta\Delta Ct}$, where $\Delta\Delta Ct = \text{mean } \Delta Ct_{CRCsamples} - \text{mean } \Delta Ct_{controlsamples}$.

In the second phase, we devised a tray format to evaluate simultaneously all 6 duplex reactions of 7 genes for each sample to confirm the results from the initial phase. We selected 6 genes that were statistically significantly overexpressed in CRC *vs.* controls ($p < 0.01$). We chose *IL2RB,* an underexpressed gene, as the common duplex partner to reassay the 232-sample training set with each of the 6 overexpressed genes. This format allows calculation of an "UP/DOWN" gene expression ratio between each overexpressed CRC biomarker gene and its duplex partner, *IL2RB,* from the difference of their Ct values. Gene expression ratios using small numbers of rationally selected genes have been established as highly accurate for distinguishing clinical groups, eliminating the need for a third reference (housekeeping) gene.[22] A nonparametric Mann–Whitney test evaluated statistical significance of differences between control and CRC mRNA levels.[23]

The qRT-PCR training set was composed of 232 disease and control samples. Cancer samples were matched for age, sex, BMI and ethnicity to an approximately equal number of control samples.

An independent blind test set was composed of 410 average-risk subjects (202 CRC/208 control): only patients aged $\geq$ 50 years with no cancer or chemotherapy history, no previous record of colorectal disease (adenomatous polyps, CRC or inflammatory bowel disease) and no first-degree relatives with CRC were enrolled. Cancer samples were matched for sex, BMI and ethnicity to an approximately equal number of control samples. The average age of disease samples was 3.6 years older than that of control samples. Furthermore, less than 10% of the subjects had advanced (Stage IV) disease (Table 1). Thus, the vast majority had local (Stage I/II) or regional (Stage III) disease, amenable to resection and appropriate adjuvant therapy. These selected samples were then

Table 1. Clinical characteristics of the patient cohorts

	Training set			Test set		
Characteristics	Control	CRC	p value[1]	Control	CRC	p value[1]
No.	120	112		208	202	
Age (mean ± SD)	66.0 ± 11.5	67.5 ± 12.5	0.29	64.7 ± 8.7	68.3 ± 10.1	<0.001
Sex, no. (%)			0.50			0.22
Male	69 (57.5)	70 (62.5)		138 (66.3)	122 (60.4)	
Female	51 (42.5)	42 (37.5)		70 (33.7)	80 (39.6)	
BMI, mean ± SD	26.7 ± 4.2	27.4 ± 4.8	0.57	26.6 ± 6.1	26.5 ± 6.8	0.75
Ethnicity, no. (%)			0.20			0.89[2]
White	101 (84.2)	91 (81.3)		162 (77.9)	138 (68.3)	
Asian	9 (7.5)	6 (5.4)		32 (15.4)	35 (17.3)	
Black	7 (5.8)	7 (6.3)		8 (3.9)	8 (4.0)	
Hispanic	3 (2.5)	3 (2.7)		3 (1.4)	3 (1.5)	
Other	–	5 (4.5)		3 (1.4)	2 (1.0)	
N/A	–	–		–	16 (7.9)	
TNM stage, no. (%)						
I	–	31 (27.7)		–	62 (30.7)	
II	–	31 (27.7)		–	55 (27.2)	
III	–	33 (29.5)		–	64 (31.7)	
IV	–	11 (9.8)		–	17 (8.4)	
Unclassified	–	6 (5.4)		–	4 (2.0)	

[1] p values for age and BMI were calculated by Mann–Whitney test; p values for sex and ethnicity were calculated by Fisher's exact test. [2] Samples of "N/A" were excluded from the calculation.

randomized and assigned blinded identification before the experiment, and data analysis was subsequently performed by scientists blinded to the disease status.

Characterization of 7-gene panel

The 7 genes were derived from a larger set of genes initially identified by microarray analysis and validated by qRT-PCR. Briefly, a candidate list of 45 genes was assembled from several previous microarray results and the current microarray results. From 45 candidate genes derived from gene profiling and cluster analysis we used multiple criteria, including level of fold-change, expression intensity and primer optimization to prioritize 20 genes for further study. These 20 genes were further validated using qRT-PCR against an initial cohort and combinations of between 2 and 14 of these 20 genes were then evaluated for predictive performance using both a standard logistic regression approach and a nested bootstrapping analysis. This revealed that a 7-gene panel was optimal in terms of unbiased prediction accuracy. One thousand iterations of randomized 5-fold and 10-fold stratified bootstrapping (subsampling with replacement) were conducted at various stages to guide the gene selection process.

Constructing a predictive model

We constructed a predictive, logistic regression model using the 6 ΔCt values from the 232 samples in the training set to determine the coefficients of these ΔCt values to the log-odd value in the predictive model[23]: $\ln[p/(1-p)] = c_0 + \sum_{g=1}^{6} c_g \times \Delta\mathrm{Ct_g}$, where p is the probability of samples being predicted as CRC and $\{C_g\}$ are coefficients of logistic regression. Optimum threshold for best accuracy on the training set was estimated using MedCalc (MedCalc software, Mariakerke, Belgium).

To check for hidden subgroups, a bootstrap technique was applied to train a total of 10,000 logistic regression predictive models.

In each of 100 iterations, 40 randomly selected control and 40 randomly selected CRC samples were set aside as testing data. From the remaining samples, 48 control and 48 CRC samples (2/3 of remaining CRC samples) were again randomly selected to train a Logistic regression candidate model. If the area under the receiver-operating-characteristic curve (ROC AUC) of the candidate model reached 0.75 or better, the candidate model was accepted as a predictive model. A total of 100 predictive models were generated in the bootstrap iteration. A total of 10,000 predictive models were generated from these 100 bootstrap iterations. The average coefficients of the 10,000 predictive models were selected as the coefficients of the 6 ΔCt values to the log-odd value in the final predictive model. The standard deviations of the coefficients yielded an estimate of the robustness of the model. The coefficients of variation (CV) were less than 1.1%

for the 6 genes. These very low CV values are indicative that the population is quite homogeneous. The models were then applied to the 6 ΔCts from the training and blind sets to predict the CRC status of the samples. As expected, both models (derived directly from the entire 232-sample set and the average from the bootstrap) generated virtually the same scores ($R^2 = 0.9998$). Log-odd values generated from the predictive model were designated logistic regression scores[24]; subjects with scores greater than 0 were classified "CRC."

Calculating current relative risk for CRC

Bayes' theorem was applied to calculate the current relative risk (CURR) for CRC using logistic regression scores.[25] The logistic regression score distributions of CRC and controls were first used to determine corresponding distributions in the average-risk population. Then, given a subject's logistic regression score, Bayes' theorem was applied to calculate the probability of the subject having CRC, using the obtained logistic regression score distributions of CRC and non-CRC in the average-risk population as conditional probability density functions and the CRC prevalence (0.7%) as the *a priori* probability.[25]

Unless otherwise specified, all statistical analyses were carried out using "R."[3]

Results

Gene profiling for identification of differentially expressed genes in CRC

Benjamini-Hochberg false discovery rate (BH-FDR) analysis with cutoff of 0.01 resulted in a total of 1,092 probesets identified as differentially expressed genes (Fig. 1).

To prioritize these biomarker candidates for real-time RT-PCR validation, the following criteria were applied to shorten the gene list: (*i*) average expression intensity above 50; (*ii*) known genes; (*iii*) fold change (mean) > 1.2 and fold change (median) > 1.15 or fold change (median) > 1.2 and fold change (mean) > 1.17 and (*iv*) high- probe design grade, low-probe crosshybridization and cluster evidence supported by mRNA. This analysis resulted in a total of 45 biomarker candidates for CRC (data not shown).

Seven-gene CRC biomarker panel: Development and validation

From the short list of candidate genes identified by the microarray studies, 20 were validated on a training set of 232 samples (112 CRC and 120 controls) using TaqMan® qRT-PCR (data not shown). Seven genes were selected for the development of our CRC biomarker panel. Six of them (*ANXA3, CLEC4D, LMNB1, PRRG4, TNFAIP6* and *VNN1*) were overexpressed (1.31- to 1.67-fold), and 1 (*IL2RB*) was underexpressed (0.84-fold) in CRC when compared with controls. We chose the underexpressed *IL2RB* as the common denominator for all overexpressed genes to compute UP/DOWN gene expression ratios and to examine the accuracy of the ratios in classifying the 232 samples with respect to group membership (Table 2). We calculated 6 UP/DOWN gene expression ratios per sample (Methods). All 6 ratios were statistically significantly different ($p < 0.001$) between the groups.

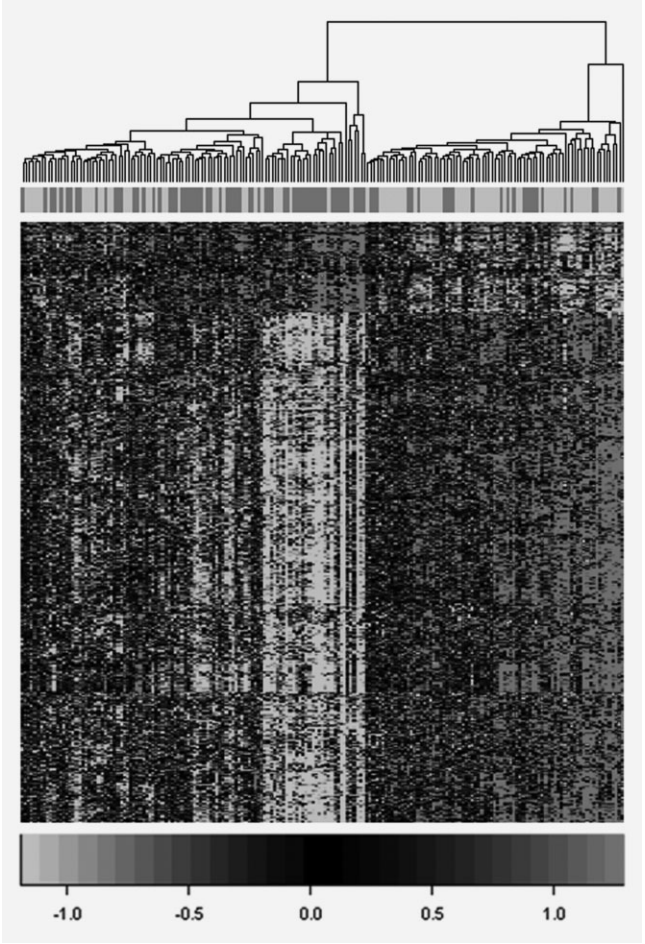

Figure 1. Heat map of gene expression and hierarchical cluster diagram showing 1,092 probe sets and the 91 CRC and 98 control samples. Dendrogram generated using "Heatmap" function in R, using default settings for the clustering algorithm.

Logistic regression multivariate analysis of the 6 ratios from the expression values of the 7 genes was used to train a predictive model for CRC (see Material and Methods section). The model correctly classified 92 of 112 CRC and 77 of 120 controls in the training set [performance characteristics: 73% accuracy; 82% sensitivity; 64% specificity; 68% positive predictive value (PPV); and 79% negative predictive value (NPV) (Fig. 2*a*). The corresponding area under the receiver-operating-characteristic curve (AUC) was 0.80 ± 0.03 (95% CI: 0.74–0.85, Fig. 2*b*).

To validate predictive performance of the 7-gene combination generated from the training set, we quantified mRNA levels for the same 7 genes by qRT-PCR using a blind independent average-risk cohort [410 samples; 202 CRC; 208 controls] (Table 1). Test set subject identifications were blinded, then scored using the logistic regression model

Table 2. Colorectal cancer (CRC) biomarker gene list and differential expression in the training set (112 CRC and 120 controls)

Gene symbol[5]	Gene name	Sequence accession ID	Fold change[1]	Fold change p value[2]	Expression ratio[3]	Expression ratio p value[2]	Expression ratio AUC[4]
ANXA3	Annexin A3	NM_005139	1.67	<0.001	1.71	<0.001	0.71
CLEC4D	C-type lectin domain family 4, member D	NM_080387	1.39	0.002	1.50	<0.001	0.66
IL2RB	Interleukin 2 receptor, beta	NM_000878	0.84	0.01	–	–	–
LMNB1	Lamin B1	NM_005573	1.31	<0.001	1.37	<0.001	0.68
PRRG4	Proline rich Gla (G-carboxyglutamic acid) 4 (transmembrane)	NM_024081	1.58	<0.001	1.72	<0.001	0.76
TNFAIP6	Tumor necrosis factor, alpha-induced protein 6	NM_007115	1.50	<0.001	1.58	<0.001	0.66
VNN1	Vanin 1	NM_004666	1.48	<0.001	1.53	<0.001	0.67

[1]Determined by qRT-PCR analysis using ACTB (reference) gene as denominator. [2]Calculated by Mann–Whitney test. [3]Determined by qRT-PCR analysis using IL2RB (underexpressed) gene as denominator. [4]Area under receiver-operating-characteristic curve. [5]Biomarker candidates were screened by microarray (5 μg of total blood RNA extracted from blood collected into EDTA tubes was hybridized to U133Plus2.0 GeneChip, Affymetrix).

generated from the training set, resulting in 146 correct CRC predictions and 145 correct control predictions. Hence, the performance characteristics showed 71% accuracy, 72% sensitivity, 70% specificity, 70% PPV and 72% NPV (Fig. 2*c*). The AUC was 0.80 ± 0.02 (95% CI: 0.76–0.84, Fig. 2*d*).

Using the biomarker panel to assess current relative risk for CRC

An assay based on the 7-gene biomarker panel was developed to assess individual CURR for having CRC. Logistic regression score distributions of CRC and non-CRC samples in the test set were used to determine the corresponding distributions of the average-risk population (Fig. 3*a*).[26] Bayes' theorem was used to calculate an individual's CURR, defined as the ratio of the probability of having CRC to the CRC prevalence, based on their blood-sample gene expression profile (see Material and Methods section). At CURR = 1, a subject has the same CRC risk as the unstratified average-risk population. At CURR = 10, the subject has a 10-fold risk increase. Similarly, at CURR = 0.1, the subject has a 10-fold risk decrease.

CURR distributions of the 202 CRC and 208 controls in the test set are plotted in Figure 3*b*. Of CRC samples, 59 (29%) had CURR < 1 and 143 (71%) had CURR > 1. By comparison, 147 (71%) controls had CURR < 1 and 61 (29%) had CURR > 1. At CURR = 1, PPV was 70%, and NPV was 72%.

Stratification of average-risk population for current CRC risk

Using CRC prevalence (0.7%) and the fitted distributions of logistic regression scores for CRC and non-CRC in Figure 3*a*, we plotted the corresponding cumulative distributions of average-risk population and of CRC patients as a function of CURR in Figure 3*c*. The distribution of average-risk population was calculated by combining the distributions of CRC and non-CRC patients with appropriate coefficients to account for CRC prevalence. The top 5% population for CRC risk is expected to have relative risk levels no less than 3.5: 32% of the CRC patients would fall within this group. The bottom 5% population for CRC relative risk is expected to have risk no greater than 0.066; only 0.2% of CRC patients would fall within this group. Hence, the most central 90% of the average-risk population spreads across a 53-fold difference in risk.

Across the continuous spectrum of risk, reference points can be identified to assist CRC screening decision making (Figure 3*d* and Table 3). For example, CURR = 2 indicates a 2-fold risk increase. Patients with CURR ≥ 2 have a current CRC risk equal-to or greater-than having a first-degree relative with CRC.[27,28] We expect 12% of the average-risk population and 51% of CRC patients have CURR ≥ 2. The corresponding PPV is 3.0%. Conversely, CURR = 0.5 correlates with a 2-fold decrease in current CRC risk. We expect 51% of average-risk patients and 12% of CRC patients to have RR ≤ 0.5. The corresponding population-based NPV is 99.8%. In comparison, the population-based PPV is only 0.7%, and the population-based NPV is 99.3% for the unstratified, average-risk population.

Discussion

Clinical practice guidelines for CRC population screening were recently updated,[29] and it was concluded that "ideally, screening should be supported in a programmatic fashion that begins with risk stratification and the results from an initial test and continues through proper follow-up based on findings." Here, we have addressed this need for risk stratification, showing that whole-blood gene expression profiling can stratify the CURR that an individual has CRC.

We recruited patients from 25 North American colonoscopic study centers. Using our extensive experience in gene profiling to identify blood-based disease biomarkers from

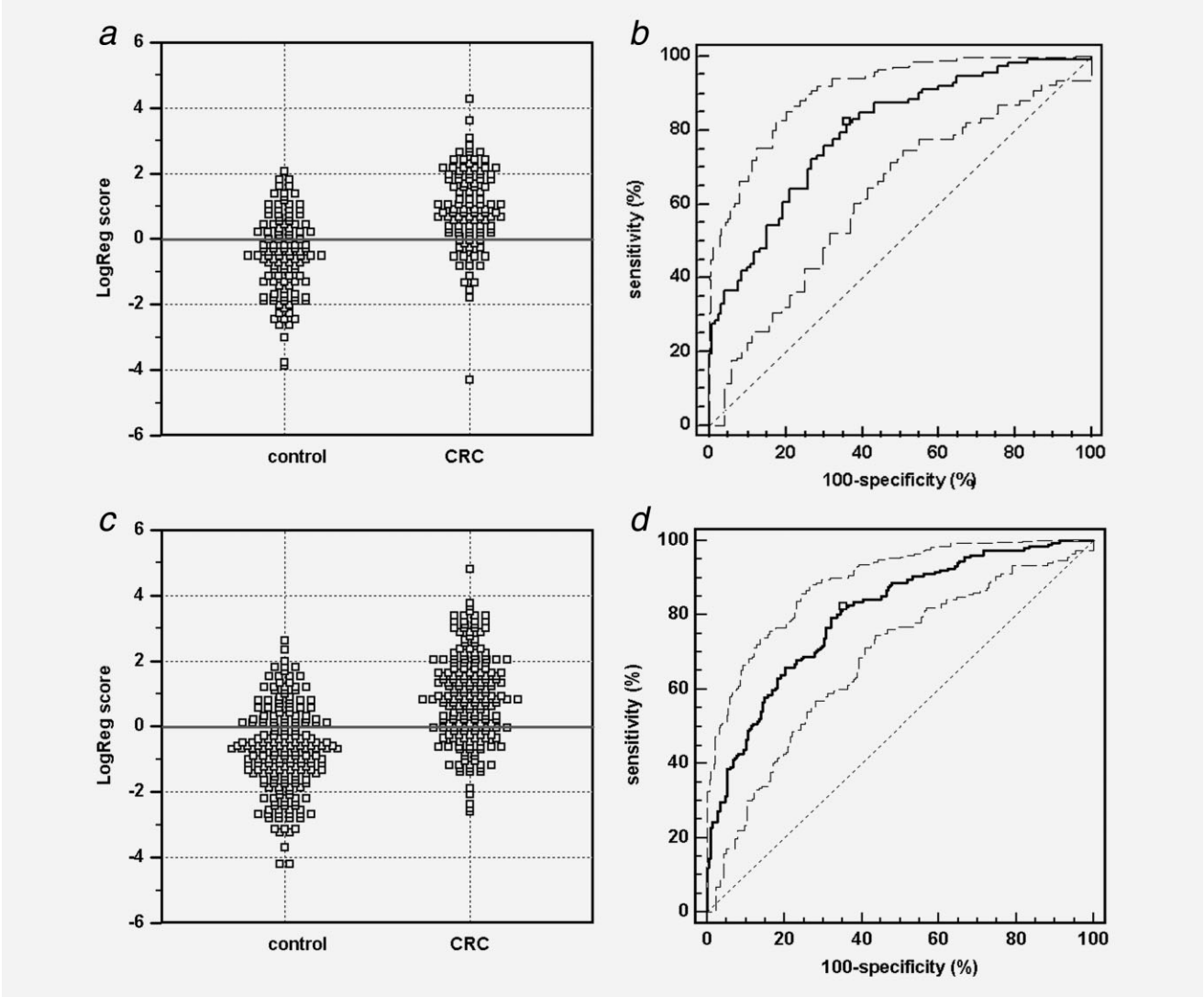

Figure 2. (*a*) Display of logistic regression (LogReg) scores of control (non-CRC) and colorectal cancer (CRC) samples in the training set (112 CRC and 120 controls). Logistic regression scores were calculated from a self-trained logistic regression model. The horizontal line at 0 indicates the CRC *vs*. control decision threshold. Samples were predicted as CRC if their logistic regression scores were equal to or greater than 0. The performance of the predictive model on the training set had the following characteristics: 73% accuracy, 82% sensitivity, 64% specificity, 68% positive predictive value (PPV) and 79% negative predictive value (NPV). (*b*) Receiver operating characteristic (ROC) curve of the training set (thick solid line). Thin lines on either side of the central thick line indicate 95% confidence interval (CI) of the ROC curve. The area under ROC curve (AUC) was 0.80, and its 95% CI was 0.74–0.85. (*c*) Displays of logistic regression scores of control and CRC samples in the test set (202 CRC and 208 controls). Logistic regression scores were calculated from the logistic regression model trained on the training set. The horizontal line at 0 indicates the CRC *vs*. control decision threshold that was fixed by the training set. Samples were predicted as CRC if their logistic regression scores were equal to or greater than 0. The performance of the predictive model on the test set had the following characteristics: 71% accuracy, 72% sensitivity, 70% specificity, 70% PPV and 73% NPV. (*d*) ROC curve of the test set (thick solid line). Thin lines on either side of central thick line indicate 95% CI of the ROC curve. The AUC was 0.80, and its 95% CI was 0.76–0.84.

microarray-derived candidate genes,[14] we identified and validated a 7-gene biomarker panel for CRC detection on 642 well-categorized, sex-, BMI- and ethnically matched CRC patients and controls. These biomarkers enabled the development of a scale to stratify average-risk patients into subgroups based on an assessment of their CURR of having CRC.

The whole-blood biomarkers identified in this study are likely not conventional tumor-derived cancer biomarkers but rather reflect subtle alterations in blood gene expression serving as a systemic response to disease, possibly acting to maintain homeostasis[30] or mediating disease pathology. Thus, for example, 1 of the biomarker genes identified in this

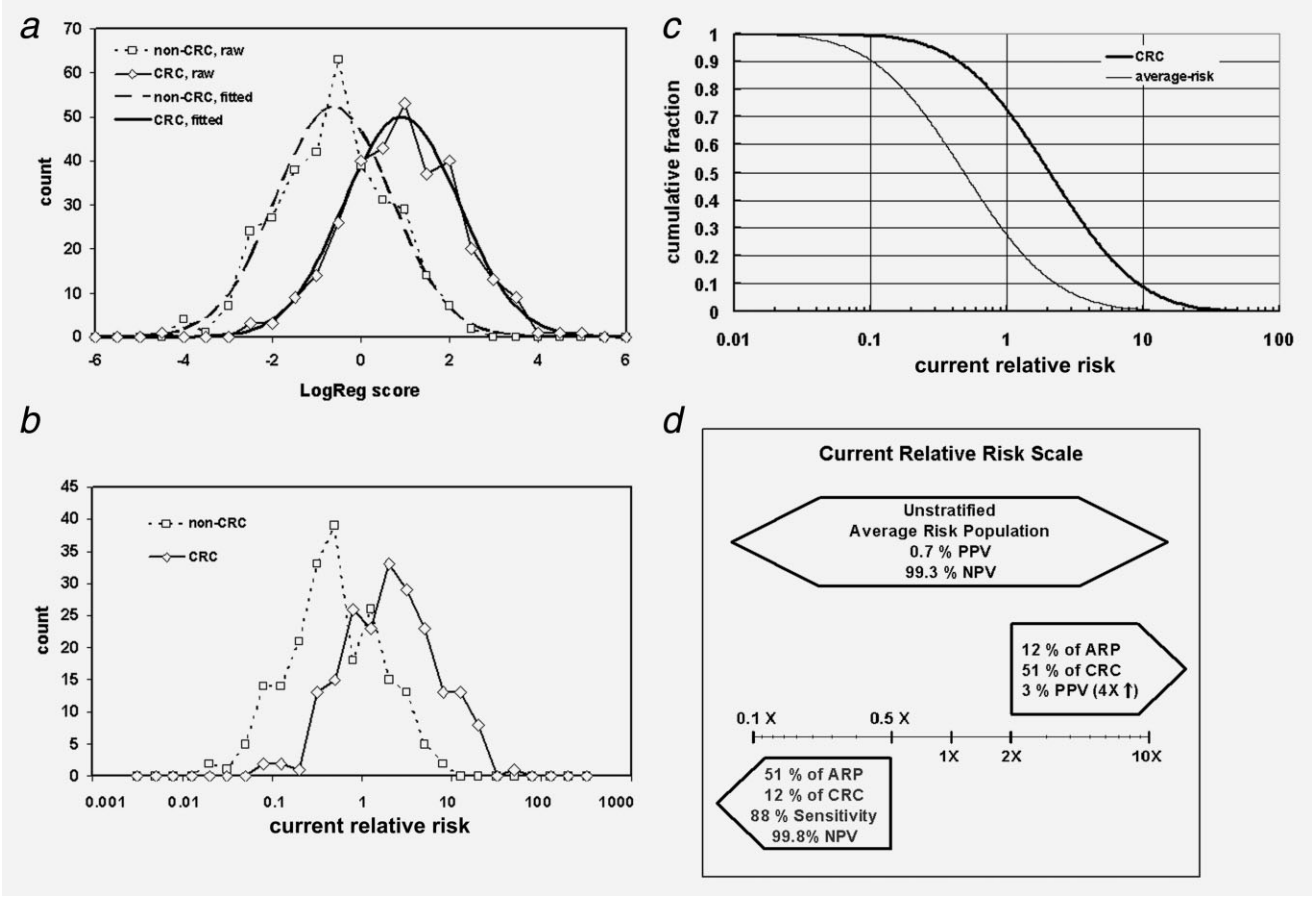

Figure 3. (*a*) Distributions of logistic regression (LogReg) scores of colorectal cancer (CRC) and control (non-CRC) samples in the test set (202 CRC and 208 controls) (Note: The test set was drawn from a population composed entirely of average-risk subjects.) The 2 distributions were tested as normal based on Shapiro-Wilk normality test (22) ($p = 0.82$ for the distribution of control samples and 0.77 for that of CRC samples). The variances of the 2 distributions were tested as equal ($p = 0.46$ by F test). The 2 distributions were fitted to normal distributions with equal variance. (*b*) Relative risk distributions of 208 control and 202 CRC samples in the average-risk population test set. (*c*) Cumulative CURR distributions of the average-risk population (thin solid line) and the CRC subpopulation (thick solid line). The distribution of average-risk population was calculated by combining the distributions of CRC and non-CRC with appropriate coefficients to account for CRC prevalence 0.7%. For any point (x, y) on a cumulative distribution curve, the y value indicates the fraction of population whose relative risks are equal to or greater than the x value. (*d*) CURR scale for stratifying patients.

Table 3. Stratification of average-risk population (ARP) for colorectal cancer (CRC)

Population	Relative risk	ARP (%)	CRC detected (%)	PPV (%)	NPV (%)
Average-risk (base state, unstratified)				0.7	99.3
Increased risk	≥2.0	12	51	3.0	
	≥1.0	28	73	1.8	
Decreased risk	≤0.5	51	12		99.8
	≤0.3	34	5		99.9

PPV: positive predictive value; NPV: negative predictive value; ARP: average risk population.

study, *ANXA3*, encodes annexin A3, a potential factor mediating angiogenesis.[31] Angiogenesis, the generation of new blood vessels from existing vasculature, is involved in the growth of tumors and may facilitate metastasis.

Another biomarker of interest in carcinogenesis, *IL2RB* encodes the beta chain of the interleukin 2 (IL2) receptor: a key factor in T-cell-mediated immune responses. Increased expression of IL2 and its receptor complex has been associated with breast tumor development and increased malignancy.[32] IL2 and IL2R expression have been reported in other types of tumors, including stomach, renal, squamous cell, melanoma and prostate (Ref. 32). In our study, *IL2RB* is

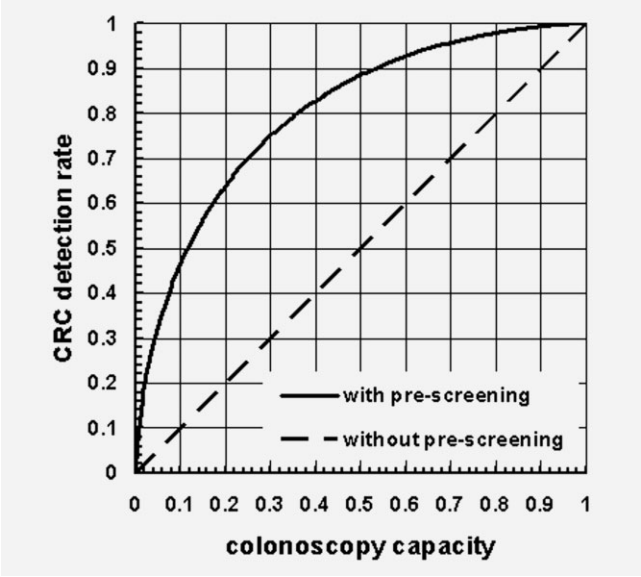

Figure 4. Colorectal cancer (CRC) detection rates as a function of colonoscopy capacity for an average-risk population with or without prescreening by gene expression profiling on their blood samples. Without prescreening, CRC detection rate equals capacity (assuming perfect sensitivity by colonoscopy). With prescreening, colonoscopy capacity can be used to determine a CURR threshold (Fig. 3*c*) to select patients whose CRC risks are equal to or greater than the threshold for colonoscopy. As shown in Figure 3*c*, the fraction of CRC patients is always higher than that of the average-risk population at any given current relative risk threshold. Hence, a greater number of CRC patients can be diagnosed by combining prescreening with colonoscopy. For example, at 10%, 20%, 30% and 40% colonoscopy capacity, the corresponding CRC detection rates for an average-risk population can be improved to 47%, 64%, 75% and 83%, respectively.

underexpressed in CRC patient blood, suggesting a homeostatic regulatory attempt to modulate this factor.

CRC screening saves lives, but patient compliance with faecal testing and endoscopy remains low.[4] Although colonoscopy is considered a CRC diagnostic "gold standard," as a screening tool the technology has limitations. Many are averse to the procedure, and most healthcare systems have limited capacity; even in the United States, colonoscopy capacity is insufficient to adequately screen the entire average-risk population.[33] Furthermore, the 0.5% incidence of significant colonoscopy-associated morbidity[13] is of concern given low CRC prevalence (0.7%) in the over 50, average-risk population. A blood-based test providing clinically actionable CRC risk information would likely enhance screening compliance and facilitate clinical decision making.

The 7-gene test can be incorporated into CRC decision making in several ways. A blood test would benefit patients who desire information about their CRC status but refuse screening due to dislike of screening options. In particular, identification of increased current CRC risk may facilitate colonoscopy decision making for these patients, who would otherwise refuse colonoscopy.

Second, in healthcare systems with limited colonoscopy capacity, this approach could help prioritize patients at greatest current risk for CRC, similar to the proposed breast cancer stratification strategy.[17] Figure 4 plots CRC detection rates, with and without prescreening, as a function of colonoscopy capacity in the average-risk population. Combining prescreening and colonoscopy can detect 2.1–4.7 times more cancers, when colonoscopy capacity is between 10% and 40%, which is the case in most countries. For example, in the CURR $\geq$ 2.0 group, PPV is 3.0%, representing a 4-fold increase in CRC detection rate per colonoscopy performed (compared with base-state PPV of 0.7% for the unstratified average-risk population; Table 3).

Furthermore, identifying patients with diminished current CRC risk can help enhance physician and patient decision making. As Table 3 shows, 34% of the average-risk population have a CRR $\leq$ 0.3. Only 0.1% of patients in this range are expected to have CRC (NPV = 99.9%). Provision of this type of novel, decreased-risk information can help facilitate subsequent screening decision making that is tailored to a patient's individual circumstances. It can also help ensure that finite colonoscopy resources are directed to those with greatest risk.

In sum, this 7-gene biomarker combination enabled development of a scale providing enriched information about an individual's CURR for having CRC. As a blood test, it addresses 1 of the greatest challenges currently limiting CRC screening effectiveness: lack of compliance. Additionally, by identifying patients with enhanced CURR (increased PPVs) and with diminished CURR (increased NPVs), this approach can help healthcare providers assess need for increased monitoring or further workup, and help tailor the use of invasive and expensive procedures to those most likely to benefit.

Acknowledgements

The authors thank Mr. Dimitri Stamatiou, Dr. Jian Guo, Ms. Pat Cotman, Dr. Jonathan Perkins, Dr. David Suria, Mr. Jay Ying and Dr. Ming Xu for their help and expertise. They also thank Ms. Isolde Prince for editorial assistance. All authors are current or former employees of GeneNews Ltd.

References

1. American Cancer Society. Detailed guide: colon and rectum cancer. Atlanta: American Cancer Society, 2008. Available at: http://www.cancer.org/docroot/CRI/CRI_2_3x.asp?dt=10. Accessed March 23, 2009.
2. Cancer Research UK. Bowel (colorectal cancer). Key facts on bowel cancer, 2008. Available at: http://

info.cancerresearchuk.org/cancerstats/types/bowel/. Accessed March 23, 2009.
3. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty RL. Natural history of untreated colonic polyps. *Gastroenterol* 1987;92:1009–13.
4. O'Connell JB, Maggard M, Ko CY. Colon cancer survival rates with the new American Joint Committee on cancer sixth edition staging. *J NCI* 2004;96:1420–5.
5. US Preventive Services Task Force. Screening for colorectal cancer: recommendations and rationale. *Ann Intern Med* 2002;137:129–31.
6. Canadian Task Force on Preventive Health Care. Colorectal cancer screening: recommendation statement from the Canadian Task Force on Preventive Health Care. *CMAJ* 2001;165:206–8.
7. Robb KA, Miles A, Campbell J, Evans P, Wardle J. Can cancer risk information raise awareness without increasing anxiety? A randomized trial. *Prev Med* 2006;43: 187–90.
8. Schwartz LM, Woloshin S, Fowler FJ, Welch HG. Enthusiasm for cancer screening in the United States. *JAMA* 2004; 291:71–8.
9. US Department of Health and Human Services Centers for Disease Control and Prevention. Colorectal cancer test use among persons aged greater than or equal to 50 years—United States, 2001. *MMWR* 2003;52:193–6.
10. Zarychanski R, Chen Y, Bernstein CN, Hebert PC. Frequency of colorectal screening and the impact of family physicians on screening behaviour. *CMAJ* 2007;177:593–97.
11. American Cancer Society. Cancer prevention and early detection facts and figures. Atlanta: American Cancer Society, 2007.
12. Klabunde CN, Lanier D, Breslau ES, Zapka JG, Fletcher RH, Ransohoff DF, Winawer SJ. Improving colorectal cancer screening in primary care practice: innovative strategies and future directions. *J Gen Int Med* 2007;22: 1195–205.
13. Levin TR, Zhao W, Conell C, Seeff LC, Manninen DL, Shapiro JA, Schulman J. Complications of colonoscopy in an integrated health care delivery system. *Ann Intern Med* 2006;145:880–6.
14. Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA. Peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J Lab Clin Med* 2006;147:126–32.
15. Burczynski ME. Transcriptional profiling of peripheral blood in oncology. In: Burczynski ME, Rockett JC, eds. Surrogate tissue analysis: genomic, proteomic and metabolomic approaches. Boca Raton, FL: Taylor and Francis, 2006.47–63.
16. Han M, Liew CT, Zhang HW, Chao S, Zheng R, Yip KT, Song ZY, Li HM, Geng XP, Zhu LX, Lin JJ, Marshall KW, et al. Novel, blood-based five-gene panel biomarker set for the detection of colorectal cancer. *Clin Cancer Res* 2008;14: 455–60.
17. Imperiale TF, Ransohoff DF, Itzkowitz SH, Turnbull BA, Ross ME;Colorectal Cancer Study Group. Faecal DNA versus faecal occult blood for colorectal-cancer screening in an average-risk population. *New Engl J Med* 2004;351:2704–14.
18. Pharoah PDP, Antoniou AC, Easton DF, Ponder BAJ. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med* 2008;358:2796–803.
19. Tibshirani R, A simple method for assessing sample sizes in microarray experiments. *BMC Bioinformatics* 2006;7: 106.
20. Rozen S, Skaletsky H.Primer 3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, eds. Bioinformatics methods and protocols: methods in molecular biology. Totowa, NJ: Humana Press, 2000:365–86.
21. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 2001;25: 402–8.
22. Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, Richards WG, Jaklitsch MT, Sugarbaker DJ, Bueno R. Using gene expression ratios to predict outcome among patients with mesothelioma. *J NCI* 2003;95:598–605.
23. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. New York: Springer-Verlag, 2001.
24. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 1947;18:50–60.
25. Duda RO, Hart PE, Stork DG. Pattern classification, 2nd edn. New York: Wiley, 2000.
26. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika* 1965;52:591–611.
27. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol* 2001;96:2992–3003.
28. Butterworth AS, Higgins JP, Pharoah P. Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis. *Eur J Cancer* 2006; 42:216–27.
29. Levin B, Lieberman DA, McFarland B, Smith RA, Brooks D, Andrews KS, Dash C, Giardiello FM, Glick S, Levin TR, Pickhardt P, Rex DK, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer and the American College of Radiology. *CA cancer. J Clin* 2008;58:130–60.
30. Mohr S, Liew CC. The peripheral blood transcriptome: new insights into disease and risk assessment. *Trends Mol Med* 2007; 13:422–32.
31. Park JE, Lee DH, Lee JA, Park SG, Kim NS, Park BC, Cho S. Annexin A3 is a potential angiogenic mediator. *Biochem Biophys Res Comm* 2005;337:1283–7.
32. García-Tuñón I, Ricote M, Ruiz A, Fraile B, Paniagua R, Royuela M Interleukin-2 and its receptor complex (alpha, beta and gamma chains) in in situ and infiltrative human breast cancer: an immunohistochemical comparative study. *Br Can Res* 2004;6:R1–6.
33. Vijan S, Inadomi J, Hayward RA, Hofer TP, Fendrick AM. Projections of demand and capacity for colonoscopy related to increasing rates of colorectal cancer screening in the United States. *Aliment Pharmacol Ther* 2004;20:507–15.