ColonSentry®

Scientific Overview of ColonSentry Test

- ColonSentry Test Overview
- Scientific Papers
- Regulatory Support
- Evaluation & Customer Base

STAGE ZERO LIFE SCIENCES

TABLE OF CONTENTS

ColonSentry Overview

ColonSentry® is a revolutionary blood test for assessing individual colorectal cancer risk (CRC). While colonoscopy, together with biopsy and confirmative pathology, is the definitive CRC diagnostic test, screening compliance remains low due to patient dissatisfaction with endoscopic testing or the alternative, fecal sampling. A blood based test, particularly one that can be drawn in-office and provide actionable CRC risk information, will improve compliance and enhance clinical decision making. Screening is advised to begin with risk stratification. Results from an initial test continue on through proper follow-up based on the findings. Patients with an elevated risk for CRC may be more amenable to undergo a colonoscopy, and thereby comply with screening guidelines. Combining pre-screening with colonoscopy finds 2.1-4.7 times more cancers.[13]

Because ColonSentry® is an easy blood test with minimal risk, yet excellent specificity,[13] it may be considered for patients who are younger than the currently recommended screening age, but who have CRC risk factors such as diabetes, obesity, or history of smoking. Screening of such individuals could detect a substantial number of CRCs and precancerous polyps; of the 588,869 CRC diagnoses made between 1998 and 2007, 19%— or over 64,000 cases—occurred in people under the age of 50. The median age for young-onset CRC was 44 years, with 75.2% of cases occurring between the ages of 40-49.[14] Studies indicate that among individuals in this age group who undergo colonoscopy screening, around 22% have abnormal findings, including hyperplastic polyps, tubular adenomas, and advanced neoplasms.[15] Although ColonSentry® itself is not a screening test, the ColonSentry® result can help physicians and patients decide on a case-by-case basis whether colonoscopy or other screening is warranted.

ColonSentry Assay Methodology

Many clinical studies have demonstrated that whole-blood RNA transcript-based profiles can be used to develop personalized gene expression signatures capable of differentiating patients with cancer from healthy patients across a broad spectrum of pathologies.[16-20] The ColonSentry® test uses quantitative real-time polymerase chain reaction (RT-PCR) to measure the RNA transcript expression of seven biomarker genes in a whole blood sample. Expression levels of six genes—ANXA3, CLEC4D, LMNB1, PRRG4, TNFAIP6, and VNN1—are each paired with the expression level of a reference gene, IL2RB, to create a genetic signature by which patients can be stratified for CRC risk.[13,21]

ColonSentry® specifically measures gene expression in whole blood. Tumors are known to affect the gene expression profiles of circulating leukocytes, including both myeloid cells (monocytes, macrophages, etc.) and lymphocytes (T cells, B cells, and natural killer cells).This occurs due to a unique interaction between tumor cells and the immune system that has been referred to as “cancer immunoediting.” Immunoediting comprises three stages: elimination (in which the immune system identifies cancerous and/or precancerous cells and attempts to eradicate them), equilibrium (in which the surviving tumor cells begin mutating rapidly under selection pressure from immune-system-mediated attacks), and escape (in which tumor cells that have acquired resistance proliferate uncontrollably, leading to tumor progression).[22] The functions of the seven genes that contribute to the ColonSentry® molecular signature are described below.

- IL2RB, the reference gene, encodes the beta subunit of the interleukin 2 (IL-2) receptor. Specific binding of the T-cell-derived cytokine IL-2 to this receptor is critical for the growth, proliferation, and differentiation of naïve T cells into effector T cells, which are able to target and destroy cancer cells.[23]
- ANXA3 encodes annexin A3, a member of the annexin protein family. Annexins play a role in the regulation of cellular growth and intracellular signaling.[24]
- CLEC4D codes for a member of the C-type lectin protein superfamily. Proteins with C-type lectin domains can bind carbohydrates and have diverse functions including cell-cell adhesion, immune response to pathogens, and apoptosis.[25,26]
- LMNB1 encodes the protein lamin-B1. The cellular lamina, a 2-dimensional protein matrix located adjacent to the inner nuclear membrane, is thought to be involved in nuclear stability, chromatin structure, and gene expression.[27]
- PRRG4 encodes one member of a family of proline-rich γ-carboxyglutamic acid (PRRG) proteins. Recent studies have identified a role for PRRG4 in cancer, allergy, and neurological disorders.[28]
- TNFAIP6 encodes the tumor necrosis factor-inducible gene 6 protein, also known as TSG-6, which is involved in cell-cell and cell-matrix interactions during inflammation and tumorigenesis.[29]
- The VNN1 (Vanin-1) gene encodes pantetheinase, an enzyme thought to be involved in regulation of the immune response to oxidative stress.30 VNN1-AB has been shown to be a potential biomarker for colorectal cancer.[46]

Development and Validation of the ColonSentry Test

The predictive ability of the ColonSentry® test was demonstrated in two key studies: a 10,000 patient North American case-control study that identified and validated the 7-gene biomarker panel, and a Malaysian case-control validation study that confirmed the findings.[13,21] Briefly, RNA profiling of whole blood samples from 642 North American cases and controls matched by sex, ethnicity, and BMI identified 7 genes whose expression pattern was associated with CRC (6 over-expressed genes and 1 under-expressed gene). A training set of 112 CRC cases and 120 controls was first used to identify the combination of genes whose expression pattern most significantly differentiated CRC cases from controls. Researchers then developed a risk scale to assess the likelihood of an individual having CRC using a logistic regression-based algorithm derived from the expression levels of these 7 genes. To validate the predictive performance of the 7-gene panel, Marshall et al. blinded 410 samples (202 CRC; 208 controls) and scored this “test set” on the basis of the logistic regression model.[13]

An independent Malaysian study then validated the 7-gene panel using blood samples from 99 CRC patients and 111 controls.[21] Blood samples were collected from an ethnically diverse Asian population of patients referred to a colonoscopy clinic between August 2007 and November 2009. All 6 genes of interest (ANXA3, CLEC4D, LMNB1, PRRG4, TNFAIP6, and VNN1) were upregulated and the reference gene (IL2RB) was downregulated in CRC as compared with control samples, confirming the pattern of differential gene expression found in the North American sample.[21] Results were highly statistically significant for both the Malaysian and North American studies.[13,21]

References

1. American Cancer Society. Colorectal Cancer Facts & Figures 2017-2019 Atlanta, GA: American Cancer Society, 2018
2. Centers for Disease Control and Prevention. Vital signs: Colorectal cancer. Atlanta, GA: 2011.
3. Levin TR. The importance of choosing colorectal cancer screening tests: comment on "Adherence to colorectal cancer screening". Arch Intern Med 2012;172(7):582-3.
4. Rex DK, Johnson DA, Anderson JC, et al. American College of Gastroenterology guidelines for colorectal cancer screening 2009 [corrected]. Am J Gastroenterol 2009;104(3):739-50.
5. Boursi B, Sella T, Liberman E, et al. The APC p.I1307K polymorphism is a significant risk factor for CRC in average risk Ashkenazi Jews. Eur J Cancer 2013;49(17):3680-5.
6. Levin B, Lieberman DA, McFarland B, 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. Gastroenterology 2008;134(5):1570-95.
7. Bernstein CN, Blanchard JF, Kliewer E, et al. Cancer risk in patients with inflammatory bowel disease: a population-based study. Cancer 2001;91:854-62.
8. Chen HF, Chen P, Su YH, et al. Age- and sex-specific risks of colorectal cancers in diabetic patients. Tohoku J Exp Med 2012;226(4):259-65.
9. Deng L, Gui Z, Zhao L, et al. Diabetes mellitus and the incidence of colorectal cancer: an updated systematic review and meta-analysis. Dig Dis Sci 2012;57(6):1576-85.
10. Larsson SC, Wolk A. Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. Am J Clin Nutr 2007;86(3):556-65.
11. Botteri E, Iodice S, Bagnardi V, et al. Smoking and colorectal cancer: a meta-analysis. JAMA 2008;300(23):2765-78.
12. Qaseem A, Denberg TD, Hopkins RH, Jr., et al. Screening for colorectal cancer: a guidance statement from the American College of Physicians. Ann Intern Med 2012;156(5):378-86.
13. Marshall KW, Mohr S, Khettabi FE, et al. A blood-based biomarker panel for stratifying current risk for colorectal cancer. Int J Cancer 2010;126(5):1177-86.
14. You YN, Xing Y, Feig BW, et al. Young-onset colorectal cancer: is it time to pay attention? Arch Intern Med 2012;172(3):287-9.
15. Imperiale TF, Wagner DR, Lin CY, et al. Results of screening colonoscopy among persons 40 to 49 years of age. N Engl J Med 2002;346(23):1781-5.
16. Burczynski ME, Twine NC, Dukart G, et al. Transcriptional profiles in peripheral blood mononuclear cells prognostic of clinical outcomes in patients with advanced renal cell carcinoma. Clin Cancer Res 2005;11(3):1181-9.
17. Liew CC, Ma J, Tang HC, et al. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. J Lab Clin Med 2006;147(3):126-32.
18. McLoughlin K, Turteltaub K, Bankaitis-Davis D, et al. Limited dynamic range of immune response gene expression observed in healthy blood donors using RT-PCR. Mol Med 2006;12(7-8):185-95.
19. Luo Y, Robinson S, Fujita J, et al. Transcriptome profiling of whole blood cells identifies PLEK2 and C1QB in human melanoma. PLoS ONE 2011;6(6):e20971.
20. Ross RW, Galsky MD, Scher HI, et al. A whole-blood RNA transcript-based prognostic model in men with castration-resistant prostate cancer: a prospective study. Lancet Oncol 2012;13(11):1105-13.
21. Yip KT, Das PK, Suria D, et al. A case-controlled validation study of a blood-based seven-gene biomarker panel for colorectal cancer in Malaysia. J Exp Clin Cancer Res 2010;29:128.
22. Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. Annu Rev Immunol 2004;22:329-60.
23. Smith KA. Interleukin-2: inception, impact, and implications. Science 1988;240(4856):1169-76.
24. Tait JF, Frankenberry DA, Miao CH, et al. Chromosomal localization of the human annexin III (ANX3) gene. Genomics 1991;10(2):441-8.
25. Drickamer K. C-type lectin-like domains. Curr Opin Struct Biol 1999;9(5):585-90.
26. Cambi A, Figdor C. Necrosis: C-type lectins sense cell death. Curr Biol 2009;19(9):R375-8.
27. Wydner KL, McNeil JA, Lin F, et al. Chromosomal assignment of human nuclear envelope protein genes LMNA, LMNB1, and LBR by fluorescence in situ hybridization. Genomics 1996;32(3):474-8.
28. Yazicioglu MN, Monaldini L, Chu K, et al. Cellular localization and characterization of cytosolic binding partners for Gla domain-containing proteins PRRG4 and PRRG2. J Biol Chem 2013;288(36):25908-14.
29. Lee TH, Wisniewski HG, Vilcek J. A novel secretory tumor necrosis factor-inducible protein (TSG-6) is a member of the family of hyaluronate binding proteins, closely related to the adhesion receptor CD44. J Cell Biol 1992;116(2):545-57.
30. Pitari G, Malergue F, Martin F, et al. Pantetheinase activity of membrane-bound Vanin-1: lack of free cysteamine in tissues of Vanin-1 deficient mice. FEBS Lett 2000;483(2-3):149-54.
31. Fast Stats: An interactive tool for access to SEER cancer statistics. Surveillance Research Program, National Cancer Institute. http://seer.cancer.gov/faststats Accessed 9.2.18.
32. American Cancer Society. Colorectal Cancer. American Cancer Society Web site. http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/index. Accessed 9.2.18.
33. Basic Information About Colorectal Cancer. CDC Web site. http://www.cdc.gov/cancer/colorectal/basic_info/ 9.2.18
34. Khorana AA, Dalal M, Lin J, et al. Incidence and predictors of venous thromboembolism (VTE) among ambulatory high-risk cancer patients undergoing chemotherapy in the United States. Cancer 2013;119(3):648-55.
35. Alcalay A, Wun T, Khatri V, et al. Venous thromboembolism in patients with colorectal cancer: incidence and effect on survival. J Clin Oncol 2006;24(7):1112-8.
36. Kawai K, Watanabe T. Colorectal cancer and hypercoagulability. Surg Today 2013.
37. Sorensen HT, Svaerke C, Farkas DK, et al. Superficial and deep venous thrombosis, pulmonary embolism and subsequent risk of cancer. Eur J Cancer 2012;48(4):586-93.
38. World Cancer Research Fund / American Institute for Cancer Research. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington, DC: 2007.
39. Pericleous M, Mandair D, Caplin ME. Diet and supplements and their impact on colorectal cancer. J Gastrointest Oncol 2013;4(4):409-23.
40. Hu J, La Vecchia C, de Groh M, et al. Dietary transfatty acids and cancer risk. Eur J Cancer Prev 2011;20(6):530-8.
41. Nolfo F, Rametta S, Marventano S, et al. Pharmacological and dietary prevention for colorectal cancer. BMC Surg 2013;13 Suppl 2:S16.
42. Komiya M, Fujii G, Takahashi M, et al. Prevention and intervention trials for colorectal cancer. Jpn J Clin Oncol 2013;43(7):685-94.
43. Shen XJ, Zhou JD, Dong JY, et al. Dietary intake of n-3 fatty acids and colorectal cancer risk: a meta-analysis of data from 489 000 individuals. Br J Nutr 2012;108(9):1550-6.
44. Abdelsattar, Z. M., Wong, S. L., Regenbogen, S. E., Jomaa, D. M., Hardiman, K. M. and Hendren, S. (2016), Colorectal cancer outcomes and treatment patterns in patients too young for average-risk screening. Cancer 2016; doi: 10.1002/cncr.2971
45. US Preventive Services Task Force. Screening for Colorectal Cancer. US Preventive Services Task Force Recommendation Statement. JAMA. 2016;315(23):2564–2575.
46. Løvf M, Nome T, Bruun J, et al. A novel transcript, VNN1-AB, as a biomarker for colorectal cancer. Int. J. Cancer. 2014; 135: 2077–2084.

ColonSentry®

Scientific Papers

STAGE ZERO LIFE SCIENCES

Scientific Papers

Stability of The ColonSentry® Colon Cancer Risk Stratification Test Samuel Chao1*, Tanya Pilcz1 , Dimitri Stamatiou1 , Jay Ying1 , Robert Burakoff2 , Leroy D. Mell3; International Journal of Disease Markers, 2019

The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. (The Sentinel Principle) J Lab Clin Med. 2006, 147: 126-132. 10.1016/j.lab.2005.10.005J L, c. Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA - (Page 9)

A blood-based biomarker panel for stratifying current risk for colorectal cancer. Int J Cancer. 2010;126: 1177–1186. doi: 10.1002/ijc.24910. pmid:19795455, b. Marshall KW, Mohr S, Khettabi FE, Nossova N, Chao S, Bao W, et alJ - (Page 16)

Blood RNA biomarker panel detects both left- and right-sided colorectal neoplasms: a case–control study Journal of Clin. Cancer Research; Samuel Chao1, Jay Ying1, Gailina Liew1, Wayne Marshall1,2, Choong-Chin Liew1* and Robert Burakoff (Page 26)

A Gene Expression Profile of Peripheral Blood in Colorectal Cancer f. Chi-Shuan Huang, et al., Microb Biochem Technol 2014, 6:2 M (Page 32)

OPEN ACCESS

GAVIN PUBLISHERS

International Journal of Disease Markers

Chao S, et al. Int J Dis Markers: IJDM-101.

Research Article

DOI: 10.29011/IJDM -101. 100001

Stability of The ColonSentry® Colon Cancer Risk Stratification Test

Samuel Chao[1*], Tanya Pilcz[1], Dimitri Stamatiou[1], Jay Ying[1], Robert Burakoff[2], Leroy D. Mell[3]

[1]GeneNews Ltd, Richmond Hill, Ontario, Canada

[2]Weill Cornell Medical College, New York, New York, USA

[3]Innovative Diagnostic Laboratory, Richmond, Virginia, USA

[*]**Corresponding author:** Samuel Chao, GeneNews Ltd, Richmond Hill, Ontario, Canada. Tel: +19052092030; Email: schao@genenews.com

Citation: Chao S, Pilcz T, Stamatiou D, Ying J, Burakoff R, et al. (2019) Stability of The ColonSentry® Colon Cancer Risk Stratification Test. Int J Dis Markers: IJDM-101. DOI: 10.29011/IJDM -101. 100001

Received Date: 03 January, 2019; **Accepted Date:** 10 January, 2019; **Published Date:** 18 January, 2019

Abstract

ColonSentry® is a molecular test for assessing the potential of colorectal cancer and pre-malignant lesions in average risk individuals. Initially developed from a clinical study involving approximately 10,000 subjects in North America, this test has been commercialized and administered to over 100,000 patients. We compare the real-life distribution of the results against the model that was initially developed and review them in the context of measurement stability, and to evaluate the validity of the assumptions made during the construction of the mathematical model. We confirm that the commercial application of the test falls well within the designed quality assurance limits and that stability was maintained over a period of multiple years. The model's assumption of two subpopulations, one with colorectal cancer at 0.7% prevalence, and the other without colorectal cancer at 99.3% prevalence, fit the data within the expected measurement tolerances. We discuss enhancement of the model to address a precancerous polyp phase subpopulation, and how the test results can be used to identify patients who should be referred directly for colonoscopy versus other modalities for colorectal cancer screening.

Keywords: Colon cancer; ColonSentry; Colorectal cancer; Diagnostic; Early detection

Introduction

In 2018, approximately 97,000 new cases of colon cancer and 43,000 new cases of rectal cancer are anticipated to be diagnosed in the United States [1]. Colorectal cancer (CRC) is the third most common cancer diagnosed in the US with a lifetime risk for men and women of approximately 4% [1]. Early diagnosis is critical to survival. The 5-year survival rate for stage I colon cancer is ~92% whereas the survival rate for stage IIIB-IV varies from 69% to 11%, depending on the extent of disease [1]. The United States Preventive Services Task Force recommends screening for colorectal cancer using stool based tests (gFOBT, FIT, FIT-DNA) or direct visualization tests (sigmoidoscopy, colonoscopy, CT colonography) in adults, beginning at age 50 years and continuing until age 75 [2]. Compliance and risks associated with these procedures vary.

GeneNews developed and validated ColonSentry, a convenient blood-based colorectal cancer risk prediction test that determines an individual's current risk of having colorectal cancer. Risk is determined by measuring the levels of 7 genes (ANXA3, CLEC4D, LMNB1, PRRG4, TNFAIP6, VNN1 and IL2RB) in the blood and inputting that information into a proprietary algorithm. Clinical validation results were published in 2009 in the International Journal of Cancer [3]. The ColonSentry model was developed on a training set consisting of 112 CRC and 120 Controls with an area under the curve (AUC) of 0.80 (95% confidence interval: 0.74 - 0.85), 64% specificity, 82% sensitivity and 73% accuracy. The predictive performance was validated on an age/gender/ethnicity balanced test set consisting of 202 CRC and 208 Controls with an AUC of 0.80 (95% confidence interval: 0.76–0.84), 70% specificity, 72% sensitivity and 71% accuracy. An analysis of the prediction distribution for location and stage of CRC shows equal sensitivity for both left and right sided lesions and a progressive increase as the cancer progresses [4]. ColonSentry has been commercially available in the US since 2012 and offered by CLIA accredited Innovative Diagnostic Laboratory (IDL), located in Richmond, VA, since 2014.

Subsequent to the launch of ColonSentry in 2008, many groups have independently validated gene expression from the

7-gene panel, autonomously or in-combination with each other, to determine the use as a CRC diagnostic marker(s). In 2010, Yip et al. validated ColonSentry in Malaysia on 99 CRC and 111 Controls reporting an AUC of 0.76 (95% confidence interval: 0.70 to 0.82), 77% specificity, 61% sensitivity and 70% accuracy [5], comparable to the data obtained from North American validation. Chang et al. developed a blood-based CRC detection assay that included the ANXA3, TNFAIP6 and IL2RB biomarkers [6]. ColonSentry has been shown to detect left and right CRC with similar sensitivity, unlike Colonoscopy which misses right-sided lesions [7-9]. To date, ColonSentry has been used to assess colon cancer risk in over 100,000 patients from the United States. We evaluate how well the model developed for ColonSentry approximated the general population and whether the assumptions that were made could be validated.

Methods

ColonSentry Logistic Regression (LogReg) Score and Relative Risk Determination

The qPCR data of each sample is specified by $\{(Ct_{g,i}, Ct'_{g,i})\}$, where $Ct_{g,i}$ are the Ct values for genes ANXA3, CLEC4D, TNFAIP6, LMNB1, PRRG4, or VNN1, $Ct'_{g,i}$ are the Ct values for the duplex partner gene IL2RB, and where $g = 1, 2, \dots, 6$ represents one of the six listed genes, and $i = 1,2$ represents the duplicate number. For convenience, we make the following definitions.

delta Ct

$$\Delta Ct_{g,i} = Ct_{g,i} - Ct'_{g,i}$$

The log-odd value of a sample being predicted as CRC was given by

$$s = \ln \frac{p}{1-p} = c_0 + \sum_{g=1}^{6} c_g * \Delta_g$$

where p is the probability of the sample being predicted as CRC

Bayes' Theorem was applied to calculate the current CRC risk using LogReg scores. The LogReg score distributions of CRC and controls in the dataset were used to determine corresponding distributions in the average-risk population. More precisely, the conditional probability of CRC patients having LogReg score s was fitted by

$$p_+(s) = \frac{1}{\sqrt{2\pi}\lambda_+} \exp\left(-\frac{(s - s_+)^2}{2\lambda_+{}^2}\right)$$

where s_+ and $\lambda_+$are parameters evaluated from the dataset

Similarly, the conditional probability of controls having LogReg score s was fitted by

$$p_-(s) = \frac{1}{\sqrt{2\pi}\lambda_-} \exp\left(-\frac{(s - s_-)^2}{2\lambda_-{}^2}\right)$$

where s_- and λ_- were fitting parameters evaluated from the test dataset

Then, given a subject's LogReg score s, Bayes' Theorem was applied to calculate the probability of the subject having CRC as

$$p(+ \mid s) = \frac{p_+(s) p_0}{p_+(s) p_0 + p_-(s)(1 - p_0)}$$

where the *a priori* probability $p_0 = 0.007$ was the CRC prevalence in average-risk population.

An individual's relative risk (RR) for CRC is reported as their "CRC Score", defined as the probability of having CRC divided by CRC prevalence, was given by

$$CRC\ Score = RR(s) = \frac{p(+ \mid s)}{p_0} = \frac{p_+(s)}{p_+(s) p_0 + p_-(s)(1 - p_0)}$$

At RR=1.0, a subject has the same CRC risk as the un-stratified average-risk population.

ColonSentry Test Procedure and Data Collection from IDL

qPCR and Plate-to-plate calibration

For qRT‑PCR, blood collected in PAXgene™ tubes (PreAnalytiX) was processed according to PAXgene™ Blood RNA Kit protocol. RNA quantity was determined by absorbance at 260nm in a NanoDrop 8000 (Thermo Scientific™).

Approximately 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, 8 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 10 µL reaction volume. PCR amplification was performed using a Viia7 Real-Time PCR Instrument (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 calibrator, positive and negative reference sample. Samples that failed these quality control checks were repeated. Samples that failed a second time were excluded from the analysis. To stabilize the qPCR measurements against variations from various sources (e.g., instrument, reagent lots), a known reference obtained from a qualified pooled RNA is placed on each plate and run alongside the patient samples.

Measured delta Ct values are then compared against the established reference values and these results are then used to calibrate the unknown samples. To evaluate the performance of this calibration procedure, two other known and qualified references are also measured on each plate: a "positive" reference known to generate a high CRC score and a "negative" reference known to generate a low CRC score. These two references are processed the same way as the unknown subjects. The "calibrated" delta Ct for these two references can then be monitored for deviations from expected values.

Data Collection

Since inception, more than 100,000 ColonSentry tests for clinical purposes have been performed in the U.S., 95,139 of which included a minimal set of clinical information to verify whether the patient would have qualified as "average risk" (i.e., no first-degree relative with CRC, no previous CRC or surgery for CRC). The age distribution by gender of these 95,139 patients is presented in Figure 2. ColonSentry scores from 95,139 patients, collected and processed as described in 2.2.1 at IDL (Richmond, Virginia) were used in this analysis.

Model Development

ColonSentry scores from 95,139 patients, collected and processed as described in 2.2.1 at IDL (Richmond, Virginia) were plotted and the distribution of these scores were compared to the model's projected score distribution for an average risk population with 0.7% CRC prevalence.

Model Comparison and Evaluation

The histogram of the distribution of accumulated patient scores was compared to the predicted distribution based on the model described above using Bayes' theorem. The bin size was set to 0.1 units on the LogReg scale. The difference between the two distributions is quantified as the RMS error which is defined as the square root of the mean of the sum of the squares of the differences at each evaluated LogReg score along the horizontal axis of the distribution chart.

Results

Comparison and Evaluation of the Predictive Model to Observed Data

By the use of Bayes' Theorem, the CRC samples from the clinical trial [3] were scaled to the known 0.7% prevalence, with the non-CRC samples scaled to represent the remaining 99.3% of the average risk population. The expected LogReg score distribution is presented in Figure 1.

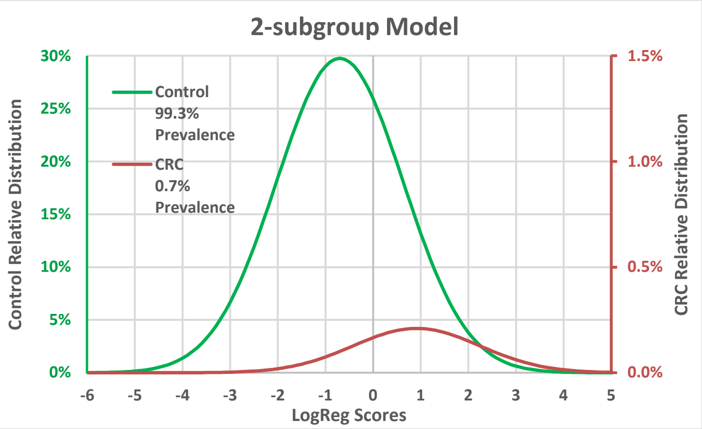

Figure 1: Expected LogReg Score Distribution for the ColonSentry Test. The distribution of LogReg scores is presented for both the non-colorectal cancer (non-CRC) and colorectal cancer (CRC) subgroups. The relative distribution of the non-CRC is indicated using the on left vertical scale, while the CRC uses the right vertical scale. The bin size to determine the distribution was set to a 0.1. Note that the secondary vertical axis for CRC group is expanded 20X compared to the Control group.

Approximately 100,000 ColonSentry tests were performed and 95,139 of them also included a minimal set of clinical information to verify whether the patient would have qualified as "average risk" (i.e., no first-degree relative with CRC, no previous CRC or surgery for CRC). The age distribution by gender of these 95,139 patients is presented in Figure 2.

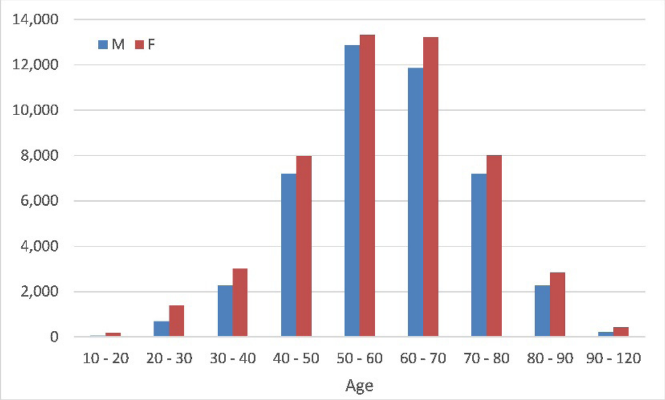

Figure 2: Patient Age Distribution for ColonSentry Tests Performed at IDL. The distribution of patients' age for the 95,139 ColonSentry tests performed at IDL separated by gender.

The model generated scores for the patients from which CRC relative risk could be predicted. The distribution of these scores were compared to the model's projected score distribution for an average risk population with 0.7% CRC prevalence (Figure 3). There was a slight displacement to the right for the actual IDL distribution (blue) relative to the model's curve (red). The asymmetrical difference curve (green) suggests that the error is mainly a relative displacement between the two curves rather than a difference in standard deviation. The total Root Mean Square (RMS) error was determined to be 0.051%.

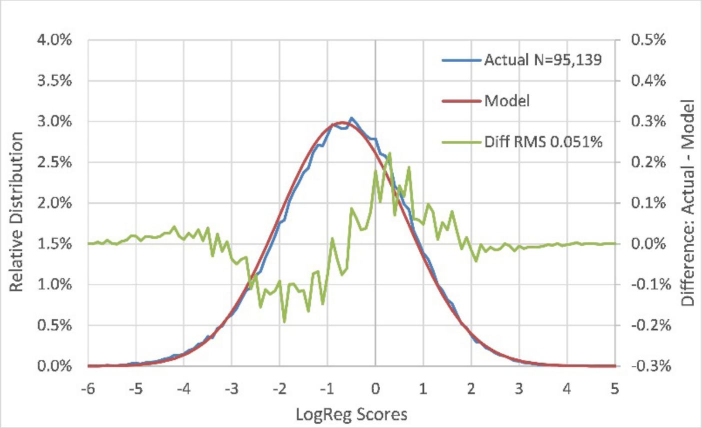

Figure 3: Comparison of LogReg Score Distribution of the ColonSentry Test between the Model and IDL Lab Results. The LogReg Score distribution was compared between the original model (red) and the IDL laboratory test results (blue) using the vertical scale on the left. The difference between the two distributions is presented as deviation (green) using the vertical scale on the right, the Root Mean Square Error value is shown in the legend as 0.051%. The bin size to determine the distribution was set to a 0.1.

One way to estimate the drift is to shift the results until the difference is minimized. The optimum shift was determined to be 0.1 units to achieve near perfect overlap throughout the range of the test results. This shift of 0.1 units magnitude is well within the allowed tolerance for the ColonSentry test which specifies a window of +/- 0.6 units for the LogReg score at 95% Confidence when all QC limits are met. Shifting the IDL lab data by 0.1 units reduced the overall RMS error to 0.033%, a factor of about 1.6X smaller than for the un-shifted results (Figure 4).

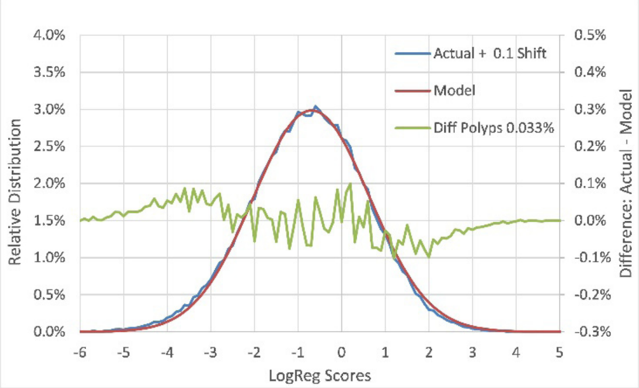

Figure 4: Optimization of IDL Lab Results. Shifting the LogReg score distribution of the IDL lab results (blue) by 0.1 units reduced the error (green) between the model (red) and the IDL lab results and the Root Mean Square value to 0.033% from 0.051%. A 0.1 unit shift, which is within allowable QA tolerance, resulted in optimal overlap of the two distributions. The bin size to determine the distribution was set to a 0.1.

Model for Early Detection

Our original model only accounted for two subgroups in the average population: subjects with CRC and subjects who are confirmed by colonoscopy and pathology to be free from CRC, polyps or advanced adenoma. However, population statistics have determined that there is a significant additional subgroup with either polyps or advanced adenomas which are non-cancerous precursor stages of CRC.

We hypothesized that this additional subgroup would have a distribution that would have prediction scores in between the CRC and Control groups, the same spread and have a prevalence in the range of 9% to 37% (Telford 2010: 9% at age 50, Frazier 2000: 21%, Imperial 2014: 37%) [10-12]. The three subgroup model is presented in Figure 5.

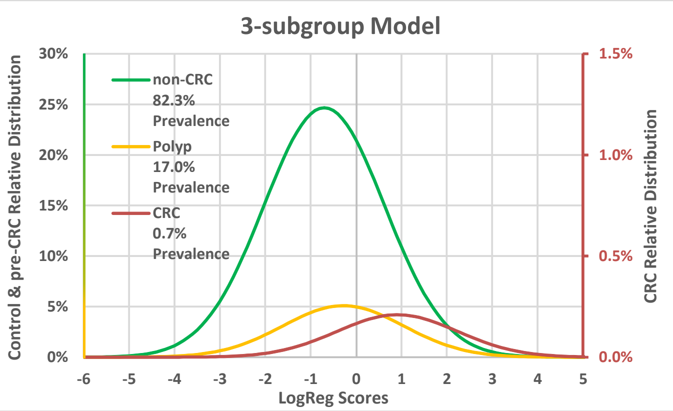

Figure 5: Expected LogReg Score Distribution for a Three Sub-Group Cancer Risk Prediction Model. The distribution of LogReg scores is presented for the non-colorectal cancer (non-CRC), non-cancerous precursor stages of CRC (pre-CRC) and colorectal cancer (CRC) subgroups. The relative distribution of the non-CRC is indicated using the left vertical scale, while the pre-CRC and CRC use the right vertical scale. Note that the secondary vertical axis for CRC group is expanded 20X compared to the Control group to preserve the relative ordinal magnitudes of the Control, pre-CRC and CRC groups.

The difference between the IDL data and the model is minimized for a shift of zero and 17% prevalence for the pre-cancer stage centered about one quarter of the distance between the CRC and the pathology-free subgroups (Figure 6). The RMS error at 0.031% is lower than the value from the unshifted 2-subgroup model and even slightly lower than the 0.1-shifted model. The zero-shift is consistent with the results of the positive and negative controls, so it is more likely that the 3-subgroup model is the more accurate representation of real-life data.

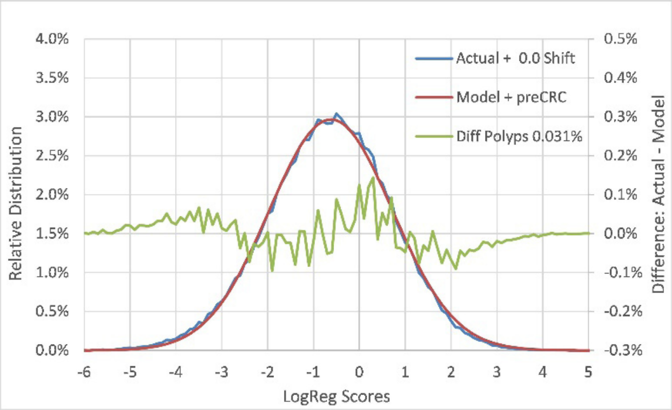

Figure 6: Comparison of LogReg Score Distribution of the Three Subgroup Model and the IDL Lab Results. The LogReg Score distribution was compared between the three subgroup model (red) and the IDL laboratory test results (blue) using the vertical scale on the left. The difference between the two distributions is presented as deviation (green) using the vertical scale on the right. The RMS was determined to be 0.031% between the two distributions. The bin size to determine the distribution was set to a 0.1.

Discussion

The long-term results that we have accumulated confirm that the ColonSentry model represented the average population with equitable accuracy. While the 2-subgroup model is a minimal representation of the real-life surveillance population, it was well within the pre-determined acceptable QC limits of the observed population distribution.

ColonSentry development began with a population of 9,199 patients recruited from multiple colorectal cancer surveillance clinics located in Canada and the US. Of these subjects, only 68 were subsequently diagnosed to have colorectal cancer by colonoscopy and pathological analysis of the biopsy [3]. This is equivalent to a prevalence rate of 0.74% which is in agreement with sources such as US SEER. This data demonstrates that the population is likely similar to the target US population.

Additional cancer samples were required to identify robust biomarkers, develop an algorithm and appropriately power statistical analysis. To achieve this, GeneNews began collecting additional samples from cancer clinics. All cancer cases were then carefully matched with subjects from the surveillance clinics for age, sex, BMI, ethnicity and cancer stage. The final cohort selected for the training set included 112 cancer cases matched to 120 pathology-free subjects. The model fitted to these data was then used to predict a test set with 202 cancer cases with a matching set of 208 control subjects.

The 7 genes included in the ColonSentry gene panel were initially selected based on microarray gene profiling of control and diseased patients [3,13]. At the time, it was unknown what, if any, role these genes had in colorectal cancer. In 2009, our analysis showed that ANXA3, CLEC4D, LMNB1, TNFAIP6, PRRG4 and VNN1 were upregulated and IL2RB was downregulated in patients with colorectal cancer but no further information on those genes was available. Since then, 6 of the 7 ColonSentry biomarkers have now been implicated in cancer, 5 of which are specifically implicated in colorectal cancer, validating their use as robust biomarkers to predict colorectal cancer risk. Multiple groups have independently studied ColonSentry, or the biomarkers within, and have validated our results [4,5,14,15].

Currently, ColonSentry can be used to identify patients at increased risk of CRC. Patients with a ColonSentry current risk scores greater than or equal to 2 are advised to pursue further evaluation with recommended screening modalities such as colonoscopy. Additional studies are underway to identify biomarkers which can predict CRC earlier, at the advanced adenoma stage. Preliminary data suggests that the ColonSentry biomarkers may play a role in the detection of advanced adenoma. Studies are underway to determine how ColonSentry can be used, or redefined, to detect CRC at the advanced adenoma stage.

The original population model only included confirmed CRC cases and control cases which were confirmed free of CRC and polyps or advanced adenoma by colonoscopy and biopsy. The decision to exclude subjects with polyps or advanced adenoma was driven by the long wait for the pathology to be confirmed and the consequently small number for which confirmation was available by the time the cancer-branch development was nearing completion. Relative displacement between the model and actual results may be attributed to the error in the estimate based on the initial training set or analytical drifts over the period of several years.

Conclusion

ColonSentry can be considered for use as an adjunct method to colon cancer screening tests in non-compliant patient populations.

Acknowledgement

- Authors S.C., T.P., D.S., J.Y., were employed by GeneNews Ltd. L.M. was employed by Innovative Diagnostic Laboratory. R.B. declares no competing interests.
- GeneNews developed and validated ColonSentry in Canada and it is commercially offered in the US by Innovative Diagnostic Laboratory. None of the authors have a financial interest in this product.
- Research presented here is covered by patent US8921074B2 where S.C. is listed as a co-inventor.
- S.C. and T.P. contributed equally to this work.
- This work was fully funded by GeneNews Ltd.
- The authors wish to thank Dr. Adam Dempsey and Dr. Joel Brill for their editorial assistance.

References

1. Colorectal Cancer Facts & Figures 2017-2019 (2017) American Cancer Society.

2. US Preventive Services Task Force K, Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, et al. (2016) Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. JAMA 315: 2564-2575.

3. Marshall KW, Mohr S, Khettabi FE, Nossova N, Chao S, et al. (2010) A blood-based biomarker panel for stratifying current risk for colorectal cancer. Int J cancer 126: 1177-1186.

4. Chao S, Ying J, Liew G, Marshall W, Liew CC, et al. (2013) Blood RNA biomarker panel detects both left- and right-sided colorectal neoplasms: A case-control study. J Exp Clin Cancer Res.

5. Yip KT, Das PK, Suria D, Lim CR, Ng GH, et al. (2010) A case-controlled validation study of a blood-based seven-gene biomarker panel for colorectal cancer in Malaysia. J Exp Clin Cancer Res.

6. Chang YT, Yao CT, Su SL, Chou YC, Chu CM, et al. (2014) Verification of gene expression profiles for colorectal cancer using 12 internet public microarray datasets. World J Gastroenterol 20: 17476-17482.

7. Baxter NN, Goldwasser MA, Paszat LF, Saskin R, Urbach DR, et al. (2009) Association of colonoscopy and death from colorectal cancer. Ann Intern Med 150: 1-8.

8. Brenner H, Hoffmeister M, Arndt V, Stegmaier C, Altenhofen L, et al. (2010) Protection from right- and left-sided colorectal neoplasms after colonoscopy: population-based study. J Natl Cancer Inst 102: 89-95.

9. Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M (2011) Protection from colorectal cancer after colonoscopy: A population-based, case-control study. Ann Intern Med 154: 22-30.

10. Telford JJ, Levy AR, Sambrook JC, Zou D, Enns RA (2010) The cost-effectiveness of screening for colorectal cancer. CMAJ 182: 1307-1313.

11. Frazier AL, Colditz GA, Fuchs CS, Kuntz KM (2000) Cost-effectiveness of screening for colorectal cancer in the general population. JAMA 284: 1954-1961.

12. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, et al. (2014) Multitarget Stool DNA Testing for Colorectal-Cancer Screening. N Engl J Med 370: 1287-1297.

13. Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA (2006) The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. J Lab Clin Med 147: 126-132.

14. Xu Y, Xu Q, Yang L, Liu F, Ye X, et al. (2015) The effect of colonoscopy on whole blood gene expression profile: an experimental investigation for colorectal cancer biomarker discovery. J Cancer Res Clin Oncol 141: 591-599.

15. Huang CS, Terng HJ, Chou YC, Su SL, Chang YT, et al. (2014) A Gene Expression Profile of Peripheral Blood in Colorectal Cancer. J Microb Biochem Technol 6: 102-109.

The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool

CHOONG-CHIN LIEW, JUN MA, HONG-CHANG TANG, RUN ZHENG, and ADAM A. DEMPSEY

TORONTO, ONTARIO, CANADA AND BOSTON, MASSACHUSETTS

In our genome-wide survey of gene expression in human peripheral blood cells using both an expressed sequence tag (EST) and a microarray hybridization approach, we identified the expression of a large proportion (approximately 80%) of the genes encoded in the human genome. Comparison of the peripheral blood transcriptome with genes expressed in nine different human tissue types revealed that expression of over 80% was shared with any given tissue. We also sought to determine whether those gene transcripts undetected by these methods were also expressed in peripheral blood cells. Using reverse-transcriptase-polymerase chain reaction, we detected additional tissue-specific gene transcripts including beta-myosin heavy chain (heart specific) and insulin (specific to pancreatic islet beta cells), in circulating blood cells. Arguably, the detection of low levels of tissue-specific transcripts could be considered products of "illegitimate" transcription; however, our study also demonstrates that environmental conditions affect the transcriptional regulation of insulin in the peripheral blood. We thus hypothesize that blood cells can act as sentinels of disease and that we could capitalize on this property of blood for the diagnosis/prognosis of disease (the "Sentinel Principle"). Peripheral blood is an ideal surrogate tissue as it is readily obtainable, provides a large biosensor pool in the form of gene transcripts, and response to changes in the macro- and micro-environments is detectable as alterations in the levels of these gene transcripts. (J Lab Clin Med 2006;147:126–132)

Abbreviations: EST = expressed sequence tag; IRB = Independent Review Board; RT-PCR = reverse-transcriptase-polymerase chain reaction

The human body is nourished by a dynamic circulatory system; the cellular components of which have a relatively rapid turnover rate.[1] Blood is classified as a fluid connective tissue, which can be defined as cells suspended in a fluid matrix functioning to connect the entire biological system at the physiological level. Blood cells also constitute the first line of the immune defense system, using an arsenal of neutrophils, eosinophils, basophils, B cells, T cells, and monocytes to defend against foreign assault and injury. Thus, the blood pervades the entire body, is in a constant state of renewal, and provides a protective barrier between the external and internal environments.

The continuous interactions between blood cells and the entire body gives rise to the possibility that subtle changes occurring in association with injury or disease, within the cells and tissues of the body, may trigger specific changes in gene expression in blood cells reflective of the initiating stimulus. We thus set out to use

From ChondroGene, Inc., Toronto, Ontario, Canada; and Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts.

Supported by ChondroGene Inc.

Submitted for publication April 15, 2005; revision submitted October 25, 2005; accepted for publication October 27, 2005.

Reprint requests: Choong-Chin Liew, 800 Petrolia Road, Unit 15, Toronto, Ontario, Canada, M3J 3K4; e-mail: cliew@chondrogene.com.

0022-2143/$ – see front matter

doi:10.1016/j.lab.2005.10.005

the "Sentinel Principle" to demonstrate that blood could serve diagnostic/prognostic purposes through profiling gene expression in blood cells. We have now demonstrated that monitoring gene expression in blood results in gene expression signatures reflective of over 35 different conditions in human subjects (see the U.S. patent applications: No. 60/115,125, No. 09/477,148, and No. 10/268,730; No. 10/601,518, No.10/802,875, and PCT application No. PCT/US04/020836).[2] Recent studies by others also demonstrate the utility of peripheral blood as a source of significant information showing that gene expression profiles of circulating blood cells were distinctive between persons,[3] and alterations in the expression profiles of blood cells were characteristic in a wide range of diseases, including juvenile arthritis,[4] hypertension,[5–7] cancer,[8,9] chronic fatigue disease[10] and neuronal injuries,[11,12] lupus,[13,14] transplantation,[15] and under various environmental pressures, such as exercise,[16] hexachlorobenzene exposure,[17] arsenic exposure,[18] and smoking.[19] A recent study by our group also demonstrated that psychiatric disorders, specifically schizophrenia and bipolar disorder, could be distinguished through specific peripheral blood gene expression profiles.[20] This rapidly growing body of evidence demonstrates the potential of using peripheral blood as a surrogate tissue for traditional tissue specimens for prognosis and diagnosis. Clearly blood provides significant advantages for this purpose, being readily available in large quantities with minimally invasive techniques.

An ideal surrogate tissue used in gene profiling analysis will be one that expresses many genes, most of which are responsive to physiological or environmental alterations. In this study, we found that genes previously believed to be restricted to non-blood tissues were in fact expressed in peripheral blood cells. Our results also suggest that the expression level of many transcripts in blood cells are responsive to, and thus indicative of, their micro-environment. These discoveries, in addition to the physical characteristics of circulating blood cells, prompted us to hypothesize that circulating blood cells can function as "sentinels" that respond to changes in the macro- or micro-environment in organs, and that blood is an ideal surrogate tissue for diagnostics.

MATERIALS AND METHODS

Isolating RNA from human circulating blood cells. Approximately 10 mL of peripheral blood was collected from each human subject. Research was carried out as required by the principles of the Declaration of Helsinki. All sample collection was approved by the collecting Institute's IRB, and written informed consents were provided in accordance with the requirements of the IRB. All samples were immediately stored on ice until RNA isolation was initiated. All RNA isolation was performed within 4 hours after blood collection. The collected blood was mixed with three volumes of hemolysis buffer (EDTA 0.6 g/L, $KHCO_3$ 1.0 g/L, NH_4Cl 8.2 g/L, pH = 7.4) to lyse red blood cells. The samples were spun at 800 rcf for 10 minutes at 4°C. The resulting pellet was washed with hemolysis buffer several times and treated with TRIZOL reagent (Invitrogen, Carlsbad, Calif) to isolate total RNA following the manufacturer's instructions. Purity and integrity of the RNA were assessed by absorbance at $UV_{260/280}$ and agarose gel electrophoresis. The quality of the RNA isolated for microarray-based expression profiling was further assessed on an Agilent Bioanalyzer 2100 using RNA 6000 Nano Chips (Agilent Technologies, Palo Alto, Calif).

Cataloguing the blood transcriptome using ESTs. The procedures of cDNA library construction and EST generation were described previously.[21] Briefly, RNAs from a pool of five adult peripheral blood and one umbilical cord blood sample were reverse transcribed into double-stranded cDNA followed by end-modification and ligation into a lambda ZAP Express vector; the assembled clones were then packaged into lambda phage, resulting in two cDNA libraries. Phage plaques were randomly picked, the cDNA insert amplified by PCR, and the product sequenced from the 5′-end. A non-redundant list of identified gene transcripts was constructed by performing sequence-based similarity clustering using the TIGR Assembler,[22] in which ESTs with an overlap of a minimum of >95% identity over 40 nt were considered to represent the same transcript and grouped together to form consensus sequences. The resultant EST cluster consensus sequences and unclustered ESTs were annotated by searching the Genbank data repositories with the BLAST algorithm (http://www.ncbi.nlm.nih.gov). Those EST clusters and unclustered ESTs that matched the same gene transcript, based on LocusLink Ids and Genbank Accession numbers, were considered redundant, and a final non-redundant list of gene transcripts was compiled.

Analysis of ESTs. The resulting 44,229 blood-derived ESTs were randomly partitioned into 44 groups, 1000 ESTs per group. A growth curve was plotted using *Sigmaplot* (Sysstat, Richmond, Calif) to estimate the total number of expressed genes in the blood transcriptome. The x-axis on the growth curve represents the total number of ESTs assessed for unique genes in increasing steps of 1000 ESTs (starting at 1000 to a total of 44,229, or 43 steps); the y-axis represents the number of unique genes identified within each group of ESTs. A regression model, $y = ax/(b + x)$, was chosen to fit the curve using *Sigmaplot*. The coefficient "a" in the regression model represents the number of genes contained in the cDNA library studied, which is based on the mathematical concept that as x approaches infinity, $y = a$. The coefficients "a" and "b" were calculated using *Sigmaplot*. The 44 groups of ESTs were randomly shuffled 20 times, resulting in 20 growth curves and 20 sets of coefficients "a" and "b." The average and standard deviation of the coefficients were calculated.

Affymetrix GeneChip profiling and data analysis. Total RNA was extracted from 248 persons. Five micrograms of each total RNA sample was used for hybridization on an

Affymetrix U133Plus2 GeneChip (Affymetrix, Santa Clara, Calif) following the manufacturer's instructions. Genes flagged as "present" or "marginal," as determined by the GeneChip Operating System (GCOS) software (Affymetrix), in at least one hybridization were considered expressed genes. LocusLink IDs were used as gene identifiers to generate a non-redundant list of expressed genes. The gene count was subsequently calculated from the non-redundant list of genes.

Comparing the tissue distribution of the circulating blood cell transcriptome. Genes expressed in nine different human tissue types, including brain, colon, heart, kidney, liver, lung, prostate, spleen, and stomach, were retrieved from the UniGene database (Build 179) at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov). UniGene IDs were used to identify corresponding LocusLink IDs. The 15,193 genes found expressed in circulating blood cells by microarray hybridization were compared with those genes identified as expressed in one of the nine tissues using LocusLink IDs.

Detecting tissue-specific transcripts in blood cells by RT-PCR. The Titan One Tube RT-PCR System (Roche Diagnostics, Indianapolis, Ind) was used for all RT-PCRs. The RT-PCR mixture contains 1 μL of total RNA from blood samples (0.1 μg/μL), 4 μL dNTP mixture (2.5mM each), 10 μL 5X RT-PCR buffer, 2.5 μL DTT-solution (100 mM), 1 μL of each primer (20 μM): (1) beta-myosin heavy chain (β-MHC) (NM_000257): Forward ′-*GCTGGAACGTAGAGACTCCCTGCT*-3′ [spans exons 21/22], Reverse 5′-*GGATCCTTCCAGATCATCCACTTG*-3′ [spans exons 24/25]; (2) insulin (INS) (NM_000207): Forward 5′-*GCCCTCTGGGGACCTGAC*-3′ [exon 2], Reverse 5′-*ACCTGCCCCACCTGCAGG*-3′ [spans exons 2/3], and 1 μL enzyme mixture.

RT-PCR was carried out at 55°C for 30 minutes for reverse transcription, followed by 30 cycles at 94°C for 30 seconds, at the appropriate annealing temperature for 20 seconds, and 72°C for 1 minute. A negative PCR control was done using the Expand High Fidelity PCR system (Roche Diagnostics, Indianapolis, Ind) to test whether any amplified product came from genomic DNA contamination in the total RNA preparation. RT-PCR products were analyzed with 1% agarose electrophoresis and purified using GeneClean (Bio101, Vista, Calif). Purified RT-PCR products were sequenced from both ends using BigDye terminator thermocycling chemistry (Applied Biosystems, Foster City, Calif) on an ABI 3770 automated sequencer (Applied Biosystems), and sequences were annotated by searching the Genbank databases at the NCBI using the BLAST program.

To assay for the gene expression of insulin, a drop of blood was collected from each subject by a finger prick. The drop of blood underwent red blood cell hemolysis as described above, and the resulting pellet was re-suspended in 10 μL water, 1 μL of which was used directly as the RNA source for the RT-PCR assays. Samples were collected at two different physiologic states (ie, fasting and non-fasting) from the same four persons for use in this assay.

The amplified insulin gene was quantified using the Quantity One v4.3.1 gel documentation system software (BioRad) and the digitally captured agarose gel image. Briefly, an equal-sized area encompassing each gel band was highlighted on the digital image and the intensity within the selected area was measured. Local background noise was determined by averaging three selected sections of the scanned image, and this value was subtracted from the experimental signal. A *t*-test was used to determine statistical significance between the intensity readings for the fasting (n = 4) and the non-fasting groups (n = 4).

RESULTS

Estimating the number of genes expressed in the blood transcriptome using expressed sequence tags. We generated a total of 44,229 blood-expressed sequence tags from two blood cDNA libraries: an adult peripheral leukocyte cDNA library (15,161 ESTs) and an umbilical cord blood cDNA library (29,068 ESTs). Overall, 10,013 unique transcripts were determined through sequence similarity-based clustering with references to the GenBank, UniGene, and LocusLink databases. The regression model, $y = ax/(b + x)$, was chosen to fit the curve because it provided the highest correlation coefficients and thus provided the best fit (Fig 1). The order of the stepwise addition of the 44 EST groups was randomized 20 times. The estimate of coefficients "a" and "b" based on each growth curve was calculated. The average of coefficient "a" was 16,409 with a standard deviation of 171. Thus, based on this model, we estimate that ~16,400 genes are expressed in human blood cells. Considering there are from 20,000 to 25,000 protein-coding genes in the human genome,[23] this indicates that approximately 66% (16,400/25,000)–82% (16,400/20,000) of the genes encoded in the human genome are expressed in human blood cells.

Estimating the number of genes expressed in the blood transcriptome using Affymetrix GeneChip profiling. For this analysis, only genes with corresponding LocusLink IDs were used to estimate the count of leukocyte-expressed genes. We investigated the approximate 19,924 unique genes with LocusLink IDs present on the *Affymetrix U133Plus2 GeneChips*. A total of 39,204 probe sets were found in at least 1 of the 248 hybridizations, representing 16,304 unique genes with LocusLink IDs. Assuming the *Affymetrix U133Plus2 GeneChip* represents an unbiased sample of the entire genome, we can estimate that approximately 81.8% (16,304/19,924) of the genes encoded in the human genome are expressed in human peripheral leukocytes. Taking into consideration the genes that have not been assigned LocusLink IDs, we can provide an estimate of approximately 16,366 (20,000 × 81.8%)–20,450 (25,000 × 81.8%) genes expressed in human blood cells.

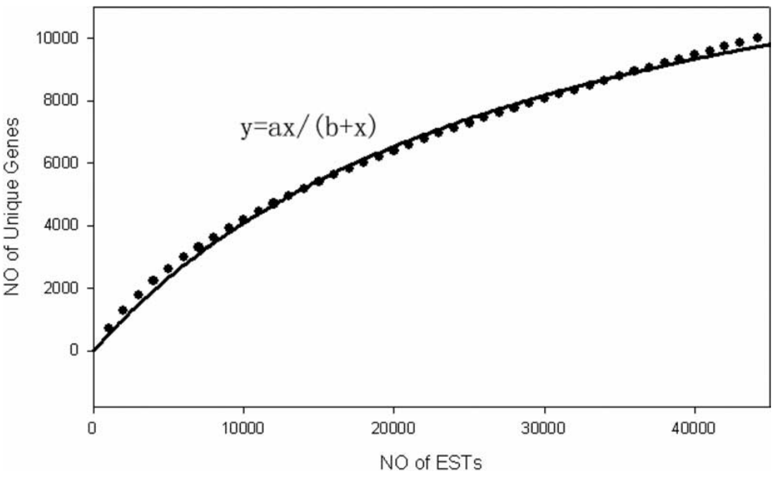

Fig 1. The EST-Gene growth curve. The x-axis represents the number of ESTs in the group used for unique gene assessment, and the y-axis represents the mean number (n = 20 reiterations) of unique genes found within the corresponding EST group. The regression model $y = ax/(b + x)$ provided the best fit to the curve. The coefficient a in the regression model represents the number of genes contained in the cDNA library studied, and b is a coefficient related to the complexity of the cDNA library. The total number of expressed genes was estimated based on the mathematical concept that as x approaches infinity, $y = a$.

Table I. Genes expressed in peripheral blood cells shared with one of nine human tissues

Tissues	Brain	Colon	Heart	Kidney	Liver	Lung	Prostate	Spleen	Stomach
Number of genes/expressed	13961	13767	12440	13428	13840	15202	11706	13224	10898
Number of co-expressed genes in blood	11428	11360	10472	11166	11490	12301	9955	10892	9408
Percentage of co-expressed genes in blood	81.9%	82.5%	84.2%	83.2%	83.0%	80.9%	83.9%	85.0%	86.3%

Comparing genes expressed in circulating blood cells and genes expressed in nine human tissue types. A non-redundant list of genes expressed in each of the nine human tissue types was retrieved by searching the UniGene database: Brain: 13,961; Colon: 13,767; Heart: 12,440; Kidney: 13,429; Liver: 13,840; Lung: 15,202; Prostate: 11,706; Spleen: 13,224; and Stomach: 10,898. The comparison between blood expressed genes (15,193 genes) and genes found in one of the nine tissue types revealed that a large proportion of genes expressed in the nine tissues are also expressed in blood cells (Table I). Approximately 80% of genes expressed in any of the nine tissues were also found expressed in blood cells.

Detecting tissue-specific transcripts in blood cells by RT-PCR. We examined the expression of two "tissue-specific" genes in peripheral blood cells: β-MHC and INS. These genes are normally associated with, and primarily expressed in, the heart and pancreas, respectively. Because their expression in the blood is likely occurring at very low levels, we could not detect their expression in our blood cell EST database or in the microarray hybridization experiment. However, in RT-PCR, we successfully detected the transcripts of these two genes in blood cells (Fig 2). The primers were designed to span exon–exon junctions, and as such, no genomic DNA should be amplified. PCR amplification of the samples without RT did not result in any amplification products, which indicates no genomic DNA contamination. The PCR products were subjected to automated DNA sequencing and confirmed that the products were generated from the targeted genes.

To determine whether environmental changes influenced expression of these low expressed genes, and to remove the possibility that detection was the result of leaky "illegitimate" transcription, we assayed the expression of the insulin gene in the blood from fasting and non-fasting subjects. A significant difference in insulin gene expression in the peripheral blood samples was observed between the fasting and the non-fasting

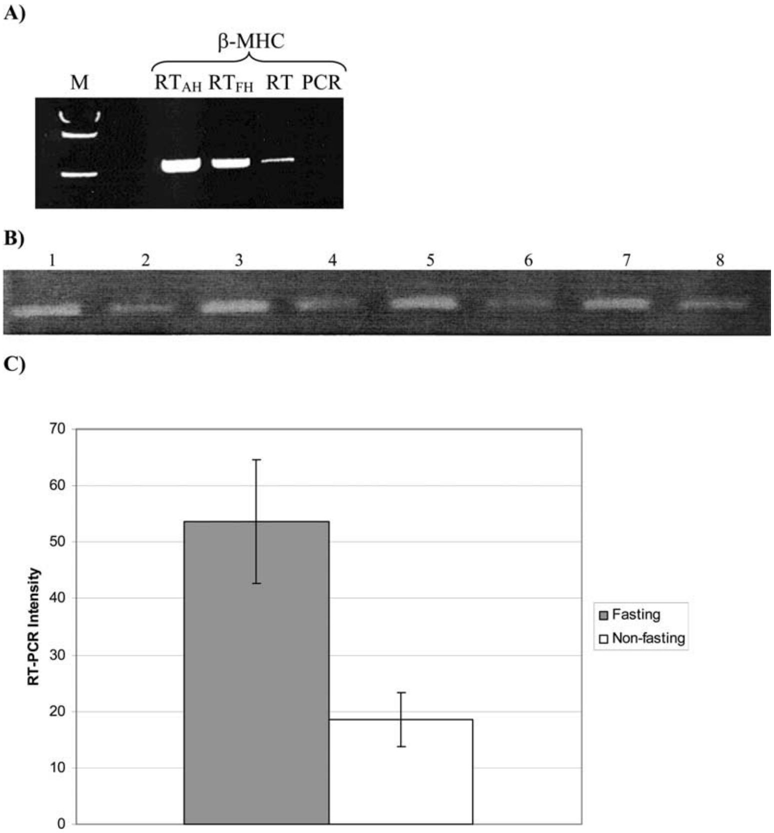

Fig 2. RT-PCR detection of "tissue-specific" gene expression in the peripheral blood. (**A**) RT-PCR was used to assay the expression of β-myosin heavy chain (β-MHC) in the peripheral blood from a single sample. The expression of β-MHC was detected in the peripheral blood (RT), whereas the negative control (PCR) did not result in amplification, which indicates no genomic DNA contamination in the samples used in the assay. RT-PCR products for beta-MHC from both human adult and fetal heart tissue ($RT_{AH}$and RT_{FH}, respectively) are also presented for reference. M indicates a molecular weight marker. (**B**) RT-PCR assay of insulin gene expression in a drop of peripheral blood. Lanes 1, 3, 5, and 7 represent samples from subjects that underwent overnight "fasting," and lanes 2, 4, 6, and 8 represent corresponding samples from subjects that did not undergo fasting or "non-fasting" samples. (**C**) Quantification of insulin RT-PCR in four subjects at fasting and non-fasting states ($P = 0.0033$, n = 4, one-sample *t*-test). The error bars indicate standard deviation from the mean. The y-axis represents signal intensity obtained from the scanned gel image in arbitrary units.

subjects ($P = 0.0033$) (Fig 2, *B* and *C*). Figures 2, *A* and *B* were modified from U.S. patent application No. 60/115,125 with permission.

DISCUSSION

Using two independent methods, we estimated the number of unique genes expressed in the blood transcriptome to be 16,000 to 20,000. The estimate derived from microarray hybridization (81.1%) is similar to the high end of the estimate derived from ESTs (82%). As microarray is more sensitive than the EST approach, we suggest peripheral blood cells express approximately 80% of the genes encoded by the human genome. Comparison of the genes expressed in the blood against a range of different tissues revealed that over 80% of the genes expressed in other tissues overlapped with the blood. Although many of these overlapping genes may be considered "housekeeping" genes, the detection of such a large number of genes shared between blood cells and other tissues cannot be explained by housekeeping functions.

Gene transcription is considered a process under strict control; only genes required by cells or tissues are expressed. However, it has been reported that tissue-specific genes may be expressed in a non-tissue-specific manner.[24–27] Ectopic expression describes "illegitimate" transcripts as products of basal transcription due to the presence of ubiquitous transcription factors and/or a net balance of negative and positive regulatory factors. It has also been suggested that in higher eukaryotes, transcription initiation occurs at a low frequency and the process is regulated in a probabilistic manner,[28] providing the opportunity for any gene to be expressed, although at varying levels. In our study, we detected the expression of two "tissue-specific" genes, insulin and β-MHC, in peripheral blood. These findings suggest that although the expression level of genes may vary among different tissue or cell types, most genes may be expressed in the blood at a detectable level using conventional methods.

An ideal surrogate tissue used in gene profiling analysis will be one that expresses many genes, many of which are responsive to physiological or environmental alterations. Proving most genes are expressed in blood cells has provided support for the first criteria of being a successful surrogate tissue. Genes, in this context, can be considered "bio-sensors." Many genes provide the potential of being able to detect various signals/stimuli. However, the overall sensitivity of these "bio-sensors" mainly relies on their capability of specific responses to various signals/stimuli. Recent blood gene expression studies have shown that the expression profiles of circulating blood cells do contain specific expression signatures in response to various physiological, pathological, and environmental changes.[4–20] In this study, using insulin as an example, we observed that insulin gene expression in blood seems to be influenced by environmental conditions, specifically fasting and nonfasting states of normal subjects.

These findings suggest that circulating blood cells have unique characteristics that make them a potential new tool for diagnostics: (1) a large proportion of the genes encoded in the human genome have detectable levels of transcripts in circulating blood cells; (2) circulating blood cells come into contact with every cell in the human body and provide an active defense against insult and injury; and (3) macro- and micro-environment changes affect gene expression in blood cells. Therefore, circulating blood cells may provide information as to the health or disease of any particular tissue by the change of gene expression pattern of their transcriptome.

In summary, we suggest that peripheral blood cells express a large proportion of the genes in the human genome, which can respond to changes occurring in the macro- and micro-environment in the body. The continuous interactions between blood cells and the entire body, combined with the fast turnover rate of blood cells, gives rise to the possibility that subtle changes occurring in association with injury or disease within the cells and tissues of the body may trigger specific changes in gene expression at a micro-level within the blood cells. These changes can then be capitalized on as biosensors for diagnostic purposes.

The authors would like to thank K. Wayne Marshall for his productive input and constructive criticism of the manuscript, Eva Cuckerman and Jim Loukides for their technical assistance, and Isolde Prince for her editorial assistance.

REFERENCES

1. Ogawa M. Differentiation and proliferation of hematopoietic stem cells. Blood 1993;81:2844–53.
2. Liew CC. Expressed genome molecular signatures of heart failure. Clin Chem Lab Med. In press.
3. Radich JP, Mao M, Stepaniants S, Biery M, Castle J, Ward T, et al. Individual-specific variation of gene expression in peripheral blood leukocytes. Genomics 2004;83:980–8.
4. Barnes MG, Aronow BJ, Luyrink LK, Moroldo MB, Pavlidis P, Passo MH, et al. Gene expression in juvenile arthritis and spondyloarthropathy: pro-angiogenic ELR+ chemokine genes relate to course of arthritis. Rheumatology (Oxford) 2004;43:973–9.
5. Bull TM, Coldren CD, Moore M, Sotto-Santiago SM, Pham DV, Nana-Sinkam SP, et al. Gene microarray analysis of peripheral blood cells in pulmonary arterial hypertension. Am J Respir Crit Care Med 2004;170:911–9.
6. Chon H, Gaillard CA, van der Meijden BB, Dijstelbloem HM, Kraaijenhagen RJ, van Leenen D, et al. Broadly altered gene expression in blood leukocytes in essential hypertension is absent during treatment. Hypertension 2004;43:947–51.
7. Okuda T, Sumiya T, Mizutani K, Tago N, Miyata T, Tanabe T, et al. Analyses of differential gene expression in genetic hypertensive rats by microarray. Hypertens Res 2002;25:249–55.
8. DePrimo SE, Wong LM, Khatry DB, Nicholas SL, Manning WC, Smolich BD, et al. Expression profiling of blood samples from an SU5416 Phase III metastatic colorectal cancer clinical trial: a novel strategy for biomarker identification. BMC Cancer 2003;3:3.
9. De Vos J, Thykjaer T, Tarte K, Ensslen M, Raynaud P, Requirand G, et al. Comparison of gene expression profiling between malignant and normal plasma cells with oligonucleotide arrays. Oncogene 2002;21:6848–57.
10. Whistler T, Unger ER, Nisenbaum R, Vernon SD. Integration of gene expression, clinical, and epidemiologic data to characterize Chronic Fatigue Syndrome. J Transl Med 2003;1:10.
11. Tang Y, Lu A, Aronow BJ, Sharp FR. Blood genomic responses differ after stroke, seizures, hypoglycemia, and hypoxia: blood genomic fingerprints of disease. Ann Neurol 2001;50:699–707.
12. Tang Y, Nee AC, Lu A, Ran R, Sharp FR. Blood genomic expression profile for neuronal injury. J Cereb Blood Flow Metab 2003;23:310–9.
13. Rus V, Atamas SP, Shustova V, Luzina IG, Selaru F, Magder LS, et al. Expression of cytokine- and chemokine-related genes in peripheral blood mononuclear cells from lupus patients by cDNA array. Clin Immunol 2002;102:283–90.
14. Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp Med 2003;197:711–23.
15. Zhang HQ, Lu H, Enosawa S, Takahara S, Sakamoto K, Nakajima T, et al. Microarray analysis of gene expression in peripheral blood mononuclear cells derived from long-surviving renal recipients. Transplantation Proc 2002;34:1757–9.
16. Connolly PH, Caiozzo VJ, Zaldivar F, Nemet D, Larson J, Hung SP, et al. Effects of exercise on gene expression in human peripheral blood mononuclear cells. J Appl Physiol 2004;97: 1461–9.
17. Ezendam J, Staedtler F, Pennings J, Vandebriel RJ, Pieters R, Boffetta P, et al. Toxicogenomics of subchronic hexachloroben-

zene exposure in Brown Norway rats. Environ Health Perspect 2004;112:782–91.

18. Wu MM, Chiou HY, Ho IC, Chen CJ, Lee TC. Gene expression of inflammatory molecules in circulating lymphocytes from arsenic-exposed human subjects. Environ Health Perspect 2003; 111:1429–38.
19. Ryder MI, Hyun W, Loomer P, Haqq C. Alteration of gene expression profiles of peripheral mononuclear blood cells by tobacco smoke: implications for periodontal diseases. Oral Microbiol Immunol 2004;19:39–49.
20. Tsuang MT, Nossova N, Yager T, Tsuang MM, Guo SC, Shyu KG, et al. Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. Am J Med Genet B Neuropsychiatr Genet 2005;133:1–5.
21. Hwang DM, Dempsey A, Wang RX, Rezvani M, Barrans JD, Dai KS, et al. A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes. Circulation 1997;96:4146–203.
22. Sutton G, White O, Adams M, Kerlavage A. TIGR Assembler: a new tool for assembling large shotgun sequencing projects. Genome Sci Technol 1995;1:9–19.
23. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004; 431:931–45.
24. Humphries S, Windass J, Williamson R. Mouse globin gene expression in erythroid and non-erythoid tissues. Cell 1976;7:267–77.
25. Chelly J, Kaplan JC, Maire P, Gantron S, Kahn A. Transcription of the dystrophin gene in human muscle and non-muscle tissues. Nature 1988;333:858–60.
26. Chelly J, Concordet JP, Kaplan JC, Kahn A. Illegitimate transcription: transcription of any gene in any cell type. Proc Natl Acad Sci USA 1989;86:2617–21.
27. Kimoto Y. A single human cell expresses all messenger ribonucleic acids: the arrow of time in a cell. Mol Gen Genet 1998; 258:233–9.
28. Hume DA. Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. Blood 2000;96:2323–8.

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^{TM} tubes (PreAnalytiX) was processed according to PAXgene^{TM} 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 $\Delta\text{Ct}_{\text{sample}} = \text{Ct}_{\text{targetgene}} - \text{Ct}_{\text{referencegene}}$. The relative fold-change (CRC *vs.* controls) was represented as $2^{-\Delta\Delta\text{Ct}}$, where $\Delta\Delta\text{Ct} = \text{mean}\ \Delta\text{Ct}_{\text{CRCsamples}} - \text{mean}\ \Delta\text{Ct}_{\text{controlsamples}}$.

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.

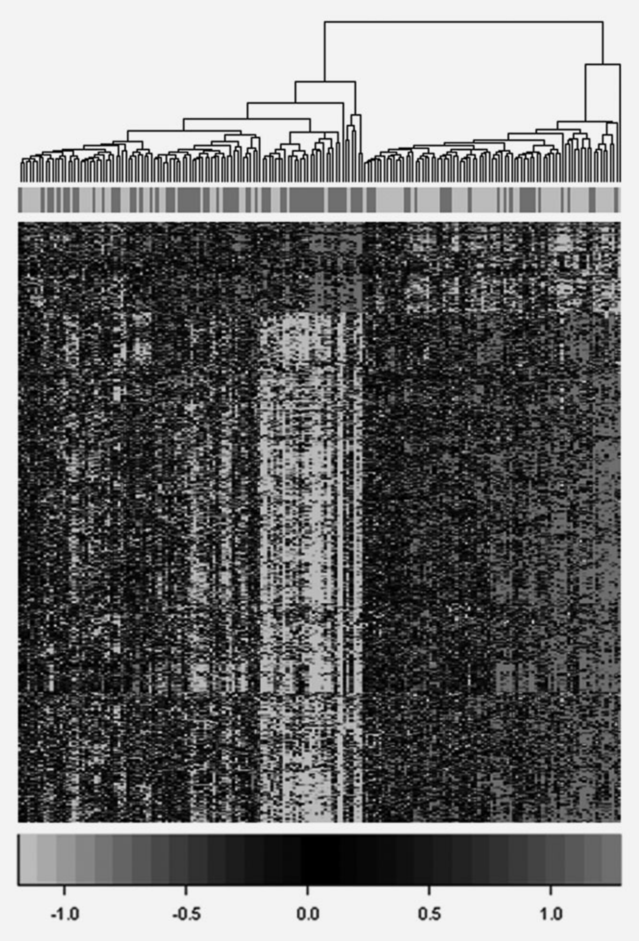

Figure 1. Heat map of gene expression and hierarchical cluster diagram showing 1,092 probe sets and the 91 CRC and 98 control samples. Dendrogram generated using "Heatmap" function in R, using default settings for the clustering algorithm.

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 $\geq$ 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 $\geq$ 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 $\leq$ 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

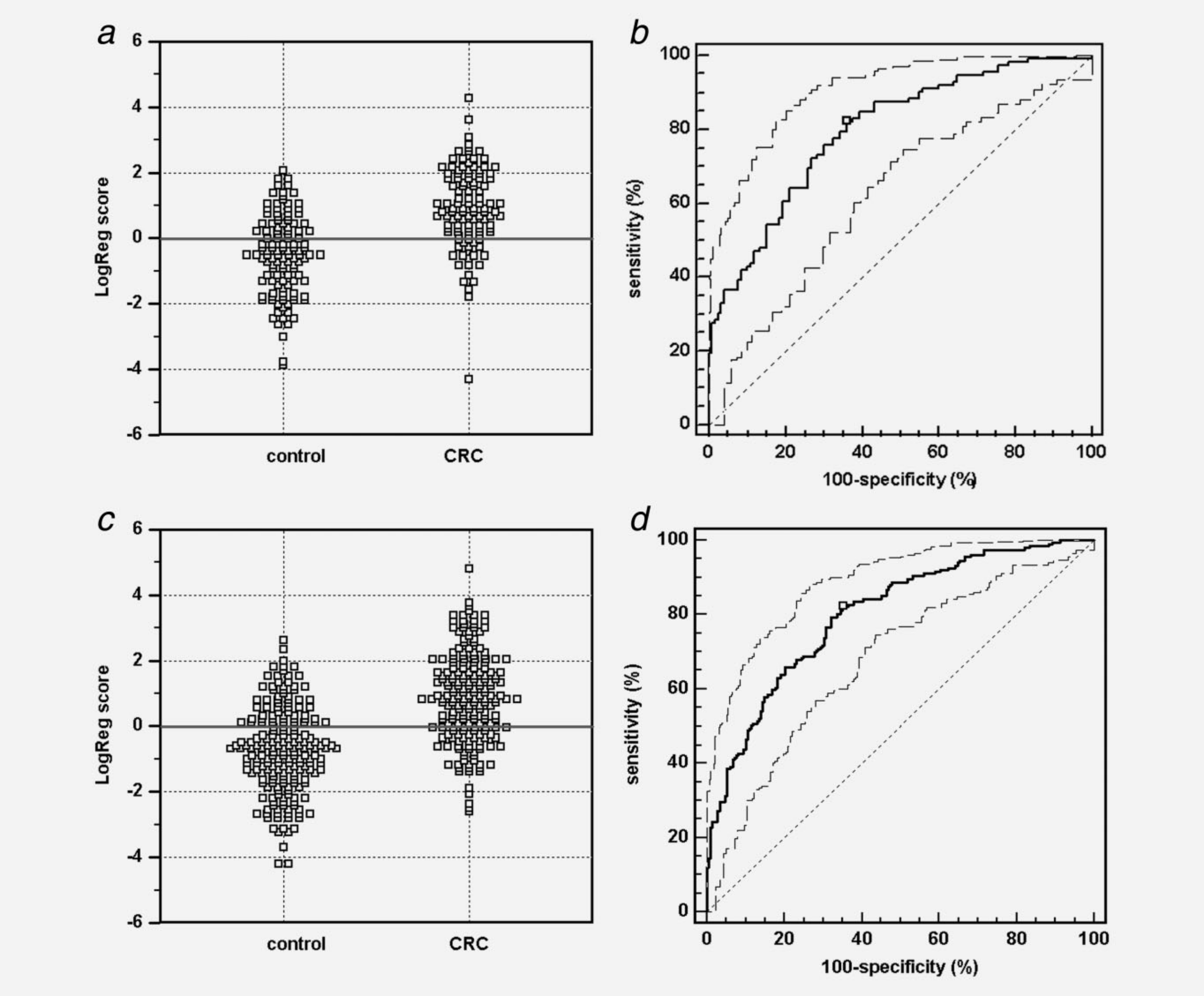

Figure 2. (*a*) Display of logistic regression (LogReg) scores of control (non-CRC) and colorectal cancer (CRC) samples in the training set (112 CRC and 120 controls). Logistic regression scores were calculated from a self-trained logistic regression model. The horizontal line at 0 indicates the CRC *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

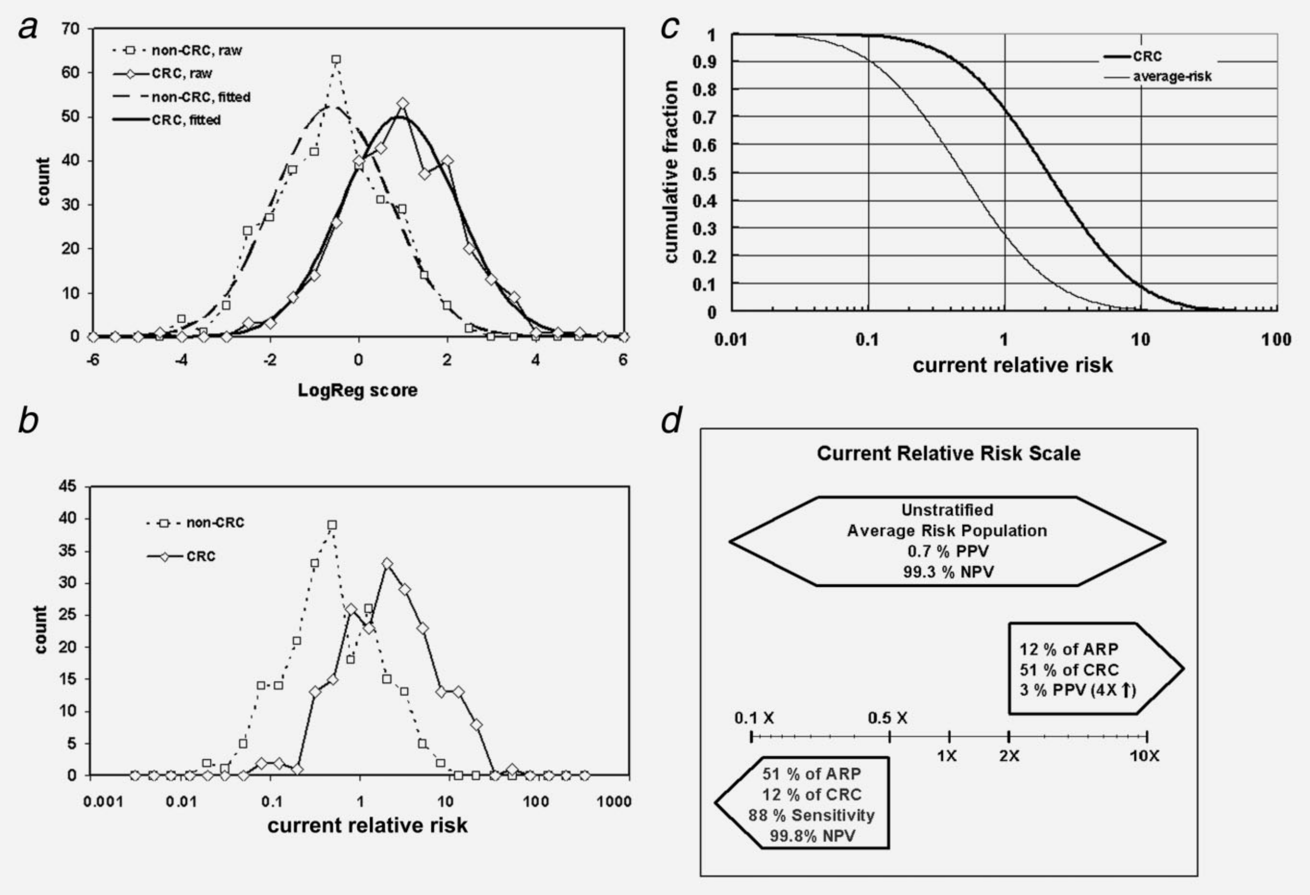

Figure 3. (*a*) Distributions of logistic regression (LogReg) scores of colorectal cancer (CRC) and control (non-CRC) samples in the test set (202 CRC and 208 controls) (Note: The test set was drawn from a population composed entirely of average-risk subjects.) The 2 distributions were tested as normal 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

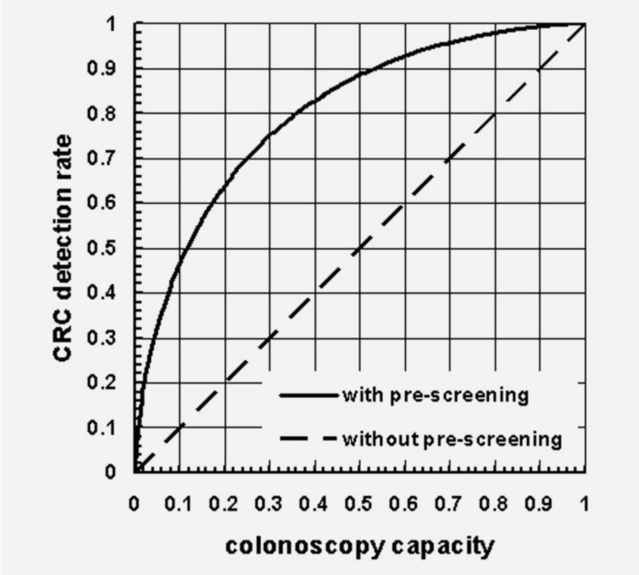

Figure 4. Colorectal cancer (CRC) detection rates as a function of colonoscopy capacity for an average-risk population with or without prescreening by gene expression profiling on their blood samples. Without prescreening, CRC detection rate equals capacity (assuming perfect sensitivity by colonoscopy). 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.

Early Detection and Diagnosis

Journal of Experimental & Clinical Cancer Research

RESEARCH Open Access

Blood RNA biomarker panel detects both left- and right-sided colorectal neoplasms: a case–control study

Samuel Chao[1], Jay Ying[1], Gailina Liew[1], Wayne Marshall[1,2], Choong-Chin Liew[1*] and Robert Burakoff[3]

Abstract

Background: Colonoscopy is widely regarded to be the gold standard for colorectal cancer (CRC) detection. Recent studies, however, suggest that the effectiveness of colonoscopy is mostly confined to tumors on the left side of the colon (descending, sigmoid, rectum), and that the technology has poor tumor detection for right-sided (cecum, ascending, transverse) lesions. A minimally invasive test that can detect both left-sided and right-sided lesions could increase the effectiveness of screening colonoscopy by revealing the potential presence of neoplasms in the right-sided "blind spot".

Methods: We previously reported on a seven-gene, blood-based biomarker panel that effectively stratifies a patient's risk of having CRC. For the current study, we assessed the effectiveness of the seven-gene panel for the detection of left- and right-sided CRC lesions. Results were evaluated for 314 patients with CRC (left-sided: TNM I, 65; TNM II, 57; TNM III, 60; TNM IV, 17; unknown, 9. right-sided: TNM I, 28; TNM II, 29; TNM III, 38; TNM IV, 12; unknown, 1 and including two samples with both left and right lesions) and 328 control samples. Blood samples were obtained prior to clinical staging and therapy. Most CRC subjects had localized disease (stages I and II, 58%); regional (stage III) and systemic (stage IV) disease represented 32% and 9%, respectively, of the study population.

Results: The panel detected left-sided (74%, 154/208) and right-sided (85%, 92/108) lesions with an overall sensitivity of 78% (215/316) at a specificity of 66% (215/328). Treatable cancer (stages I to III) was detected with left-sided lesion sensitivity of 76% (138/182) and right-sided sensitivity of 84% (80/95).

Conclusion: This seven-gene biomarker panel detected right-sided CRC lesions across all cancer stages with a sensitivity that is at least equal to that for left-sided lesions. This study supports the use of this panel as the basis for a patient-friendly, blood-based test that can be easily incorporated into a routine physical examination in advance of colonoscopy to provide a convenient companion diagnostic and a pre-screening alert, ultimately leading to enhanced CRC screening effectiveness.

Keywords: Colorectal cancer, Biomarkers, Microarray, Blood gene expression, Colonoscopy

Background

Colorectal cancer (CRC) is the third most common cancer and the second most common cause of cancer deaths in the United States and Canada. The disease is expected to be diagnosed in approximately 142,820 Americans in 2013, and an estimated 50,830 people are expected to die of CRC in that year [1]. In Canada an estimated 23,900 Canadians will be diagnosed with CRC in 2013, and 9,200 Canadians will die of the disease [2].

In the National Polyp Study, colonoscopy with adenoma removal was associated with a reduction in CRC as high as 90% [3]. Recently, however, several reports have questioned whether colonoscopy as practiced in the community reduces CRC and mortality to the same degree as that reported by highly specialized cancer centers [4-7].

Studies have found that although colonoscopy effectiveness is high for lesions that arise on the left side of the colon, the procedure fails to confer similar levels of

* Correspondence: cliew@genenews.com
[1]GeneNews Ltd, 2 East Beaver Creek Road, Building 2, Richmond Hill, Ontario, Canada
Full list of author information is available at the end of the article

protection from CRC incidence and mortality in right-sided lesions. In 2009, a case–control study of colonoscopy in Ontario, Canada, reported that although the procedure reduced mortality from left-sided lesions by about 40%, no reduction in deaths was evident when CRC originated in the right colon [4]. Similarly, in a population-based retrospective analysis from Manitoba, colonoscopy found no reduction in CRC mortality in the case of proximal lesions [5]. A large German, statewide cross-sectional study of colonoscopy found the prevalence of advanced colorectal neoplasms strongly reduced by 67% in left-sided lesions, but this protection did not extend when the lesions were right-sided [6]. A later study by the same authors, which emphasized high-quality colonoscopy, found the procedure to be associated with a reduced risk of 56% for right colonic lesions, which is an improvement over earlier reports, but is less than the 84% reduced risk for CRC the authors observed for left colonic lesions [7].

A number of suggestions have been advanced to explain why colonoscopy may be less effective in the right colon than in the left. The technology is operator-dependent and requires complete endoscopic evaluation, which is more difficult to complete in the right side of the colon. Bowel cleansing and preparation for colonoscopy may be less adequate on the right side, making lesions more difficult to visualize. Nonpolypoid flat or depressed lesions are more prevalent in the right than in the left side of the colon, and these are more challenging to identify and remove [8]. There may also be differences in biology between proximal and distal lesions; for example, distal and proximal CRCs show genetic and molecular differences [9].

We previously reported a seven-gene, blood-based biomarker panel for CRC detection [10]. For this current study, we hypothesize that this gene panel, which is a blood-based test, not dependent on localization, preparation or operator technique, can provide a non-biased method for detecting CRC arising in either the right or the left side of the colon.

The test is intended as a pre-screening tool and convenient companion diagnostic test to help those patients who are averse to colonoscopy and to the fecal occult blood test to make an informed decision based on their individual molecular profile. Because of its narrow focus, the test is not expected to alter clinical practice for patients who comply with recommended screening schedules.

Methods

Sample collection procedures and details of methodology for identification of the seven-gene blood-based biomarker panel for CRC were reported in our earlier study [10]. Briefly, 9,199 blood samples were taken from screening colonoscopy subjects at twenty-four centers located in the Greater Toronto Area and surrounding regions and in the United States, between March 2005 and March 2008. Uniformity of collection procedures at the different sites was ensured by the use of identical study protocols, uniform training of personnel, and periodic site monitoring. Informed consent was obtained according to protocols approved by the Research Ethics Board of each of the participating twenty-four clinics and hospitals.

The low incidence of CRC in the colonoscopy screening population made it necessary to recruit additional patients from cancer clinics in Toronto. In these cases, blood samples were collected prior to any treatment, including surgery. Patients enrolled in colonoscopy clinics provided blood prior to colonoscopy. Samples were categorized following review of pathology reports.

Case samples comprised blood samples taken from colonoscopy-confirmed CRC patients who had not undergone CRC treatment. Institutional pathologists determined cancer stage according to the American Joint Committee on Cancer (AJCC) Tumour, Node, and Metastases (TNM) staging system [11]. Controls comprised samples from subjects with no pathology at colonoscopy.

The qRT-PCR training set was composed of 112 well-characterized CRC and 120 control samples (total = 232) taken from the population described above. Cancer and control samples were matched for age, sex, body mass index (BMI) and ethnicity.

An independent blind test set was composed of 410 average-risk subjects following colonoscopy (202 CRC/208 control). Average risk was defined as follows: subjects aged ≥ 50 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. Cancer and control samples were matched for sex, BMI and ethnicity. The average age of patients was 3.6 years older than that of control subjects.

Most of the patients and controls who provided samples for qRT-PCR experiments had one or multiple co-morbidities, most commonly, hypertension, hypercholesterolemia, diabetes, arthritis, anemia and allergies. More than 56% of the CRC samples were diagnosed with early stage I and II CRC and 32% with stage III cancer. (Table 1) This means that approximately 90% of cases were potentially treatable CRC patients, which increases the practical value of the test.

Blood collection and RNA isolation

Samples were collected in PAXgene™ tubes (PreAnalytiX) and processed according to the manufacturer's Blood RNA Kit protocol. RNA quality for all samples was assessed using a 2100 Bioanalyzer RNA 6000 Nano Chip (Agilent Technologies). All samples met quality criteria: RIN ≥ 7.0; 28S:18S rRNA ratio ≥ 1.0 and a validated

Table 1 Available samples

Sample #	Training		Test		Combined	
Category	**Left**	**Right**	**Left**	**Right**	**Left**	**Right**
TNM I	19	12	46	16	65	28
TNM II	20	11	37	18	57	29
TNM III	21	13	39	25	60	38
TNM IV	7	5	10	7	17	12
Unknown	5	1	4	0	9	1
All Stages	72	42	136	66	208	108
Control	120		208		328	

NB Two training samples have both left and right cancer.

Agilent bioanalyzer scan. RNA quantity was determined by absorbance at 260nm in a DU-640 Spectrophotometer (Beckman Coulter).

Quantitative reverse-transcriptase polymerase chain reaction

One microgram of RNA was reverse-transcribed into single-stranded complementary DNA (cDNA) using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) in a 20μL reaction. For PCR, 20ng cDNA was mixed with QuantiTect® Probe PCR Master Mix (Qiagen), and TaqMan® dual-labeled probe and primers corresponding to the gene of interest and reference gene, in a 25 μL reaction volume. PCR amplification was performed using a 7500 Real-Time PCR System (Applied Biosystems). Each sample was tested in duplicate reactions on the same PCR plate. The run results were subjected to quality control processes, and failed samples were repeated. Samples that failed a second time were excluded from the analysis.

For the blind test set, first, we selected samples with disease status known (in order to balance the sample groups and avoid biases in clinical and demographic characteristics). Selected samples were then randomized and assigned blinded identification prior to the experiment, and data analysis was performed by scientists blinded to the disease status.

The seven-gene panel

Details of the characterization and validation of the seven-gene panel to identify CRC have been described previously [10]. In that study a seven-gene panel (*ANXA3, CLEC4D, LMNB1, PRRG4, TNFAIP6, VNN1, IL2RB*) discriminated CRC in the training set [area under the receiver-operating-characteristic curve (AUC ROC), 0.80; accuracy, 73%; sensitivity, 82%; specificity 64%]. The independent blind test set confirmed performance (AUC ROC, 0.80; accuracy, 71%; sensitivity, 72%; specificity, 70%).

For the present study we re-analyze the previously reported data in order to determine the ability of the seven gene panel not only to identify the presence of CRC but also to identify cancer stages and left- and right-sided colon cancer.

Results

The training set data was used to determine the best coefficients for a logistic regression model using 6 ratios of the 7 genes most discriminative for CRC. This model was then used to predict the CRC risk for the test set samples.

Breaking the data down by cancer stages, we were able to find the same predictive values for left- and right-sided cancers as for CRC detection as in the original paper (Table 2).

In this study, CRC detection sensitivity was generally higher for right-sided cancer except in the case of TNM stage I in the test set. However, this finding may be simply a sampling issue. To resolve this question, we combined all training and test set samples and performed 2-fold cross validation, iterated 1000 times. This process partitions the samples into 2 halves such that that the coefficients of the model are fitted to the training half and applied to the test half.

The results for all the test halves after 1000 permutations represent a less biased estimate of the performance of the gene panel. As expected, the lower sensitivity for right-sided TNM I as compared with left-sided TNM I cancers is no longer observed in the cross-validated results. Overall, right-sided lesions are detected at a higher sensitivity than left-sided lesions; however, there are fewer right-sided samples, so the observed higher sensitivity may not be statistically significant. As can be seen from the box and whisker plots of the distribution of the prediction scores, the 98% confidence intervals show considerable overlap both across all TNM stages and for left and right sided cancers (Figure 1).

The panel detected left-sided (75%, 156/208) and right-sided (85%, 92/108) lesions with an overall sensitivity of

Table 2 Correct call rate

	Training		Test		1000X 2-Fold	
					Cross validation	
Stage	**Left**	**Right**	**Left**	**Right**	**Left**	**Right**
TNM I	63%	92%	61%	44%	67%	66%
	(12/19)	(11/12)	(28/46)	(7/16)	(43.5/65)	(18.6/28)
TNM II	70%	91%	81%	89%	79%	89%
	(14/20)	(10/11)	(30/37)	(16/18)	(45.0/57)	(25.9/29)
TNM III	86%	100%	74%	84%	83%	90%
	(18/21)	(13/13)	(29/39)	(21/25)	(49.6/60)	(34.3/38)
TNM IV	86%	100%	50%	100%	66%	100%
	(6/7)	(5/5)	(5/10)	(7/7)	(11.2/17)	(12.0/12)
Unknown	80%	100%	100%	n/a	80%	100%
	(4/5)	(1/1)	(4/4)	(0/0)	(7.2/9)	(1.0/1)
All Stages	75%	95%	71%	77%	75%	85%
	(54/72)	(40/42)	(96/136)	(51/66)	(156.5/208)	(91.8/108)
Control	64% (77/120)		70% (145/208)		64% (210/328)	

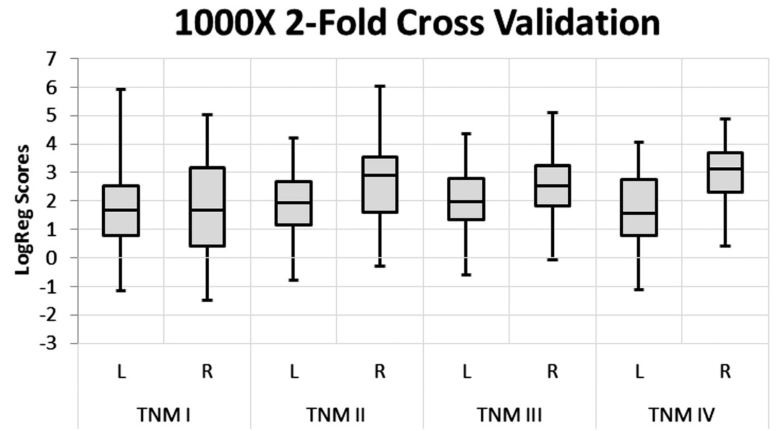

Figure 1 Distribution of prediction scores from 1000 iterations of 2-fold cross-validation analysis. Boxes indicate the central 50 percentile with whiskers showing the extent of the 98 percentile.

78% (248/316) at a specificity of 64% (210/328). Treatable cancer (stages I to III) was detected with a left-sided lesion sensitivity of 76% (138/182) and a right-sided sensitivity of 83% (79/95).

Discussion

In several studies we have shown that gene signatures obtained using blood mRNA can identify a variety of conditions occurring in various sites throughout the body, including heart failure [12], inflammatory bowel disease [13,14], psychiatric disorders [15-17] and various cancers [10,18-20]. These studies suggest that blood cells may act as "sentinels" that can mirror health or disease anywhere in the body. Blood transcriptomic signatures thus reflect molecular changes regardless of where they occur in the body.

We have also recently reported a blood test based on the performance characteristics of a seven-gene panel that enables us to assess a patient's current risk of having CRC [10]. As a blood test similar to other routine blood tests, the assay overcomes a number of reported limitations to patient acceptance of CRC screening using currently utilized tests. Such barriers include patients' fear of pain, inconvenience, unpleasantness, pre-procedure colon evacuation methods, the need to take time off work and to be sedated, risks such as bowel perforation, bleeding and other complications (for colonoscopy and other endoscopic methods) and patient embarrassment and beliefs that methods are unsanitary, unpleasant or inconvenient (fecal tests) [21-27]. By contrast, a simple, convenient blood test should encourage increased compliance with screening recommendations.

In this study we use the same seven-gene panel to address another issue limiting the effectiveness of colonoscopy: the right-sided bias observed in such technology.

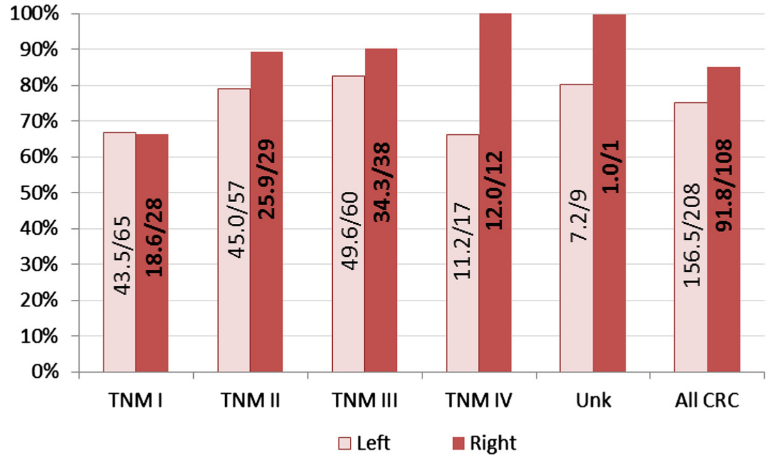

Figure 2 Prediction sensitivity for all CRC at each stage. Figures inside the bars show the ratios of average positive calls from 1000 iterations of 2-fold cross validation analysis.

CRC can arise in either the right, proximal colon or the left, distal colon. The former includes the cecum, ascending colon, hepatic flexure and transverse colon, and the latter the descending and sigmoid colon and rectum. Colonoscopy tends to bias towards detection on the left side, for reasons both technical and biological. The blood-based test for CRC reported in this study would have the effect of reducing such bias, thus potentially increasing detection rates for right sided lesions.

This pre-screening test is mainly intended for detection of TNM I to TNM III patients. For these patients, test sensitivity is 76% for left-sided cancers and 84% for right-sided cancers. TNM IV stage patients are likely to be diagnosed by conventional means and are less likely to benefit much from intervention.

Conclusion

This study finds that detection of CRCs using mRNA biomarkers from whole blood is equally sensitive to treatable TNM I – III lesions located throughout the colon (Figure 2). These findings support the use of the seven-gene panel as a non-biased method for CRC detection for both left and right-sided lesions.

Abbreviations

CRC: Colorectal cancer; BMI: Body mass index; AUC ROC: Area under the receiver-operating-characteristic curve.

Competing interests

Samuel Chao, Gailina Liew and Choong Chin Liew are all employed by GeneNews Ltd, Ontario, Canada, who funded this study. Gailina Liew is President and COO and Choong Chin Liew is Chief Scientist of GeneNews; Wayne Marshall was CEO of the company when the research was carried out. Robert Burakoff has no competing interests to declare.

Authors' contributions

CCL, WM and RB conceived and designed the study; SC and JY provided data analysis; GL and SC drafted the manuscript. All authors read and approved the final manuscript.

Author details

[1]GeneNews Ltd, 2 East Beaver Creek Road, Building 2, Richmond Hill, Ontario, Canada. [2]University Health Network, Toronto Western Hospital, Toronto, Ontario, Canada. [3]Brigham and Women's Hospital, Gastrointestinal Division, Harvard Medical School, Boston, MA, USA.

Received: 29 April 2013 Accepted: 17 July 2013
Published: 23 July 2013

References

1. American Cancer Society: *Cancer facts and figures 2013.* [http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-036845.pdf].
2. Canadian Cancer Society: *Colorectal cancer statistics.* [http://www.cancer.ca/en/cancer-information/cancer-type/colorectal/statistics/?region=on].
3. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, Waye JD, Schapiro M, Bond JH, Panish JF, Ackroyd F, Shike M, Kurtz RC, Hornsby-Lewis L, Gerdes H, Stewart ET, National Polyp Study Workgroup: **Prevention of colorectal cancer by colonoscopic polypectomy.** *N Eng J Med* 1993, **329:**1977–1981.
4. Baxter NN, Goldwasser MA, Paszat LF, Saskin R, Urbach DR, Rabeneck L: **Association of colonoscopy and death from colorectal cancer.** *Ann Intern Med* 2009, **150:**1–8.
5. Singh H, Nugent Z, Demers AA, Kliewer EV, Mahmud SM, Bernstein CN: **The reduction in colorectal cancer mortality after colonoscopy varies by site of the cancer.** *Gastroenterol* 2010, **139:**1128–1137.
6. Brenner H, Hoffmeister M, Arndt V, Stegmaier C, Altenhofen L, Haug U: **Protection from right- and left-sided colorectal neoplasms after colonoscopy: population-based study.** *J Natl Cancer Inst* 2010, **102:**89–95.
7. Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M: **Protection from colorectal cancer after colonoscopy: a population-based, case–control study.** *Ann Intern Med* 2011, **154:**22–30.
8. Soetikno RM, Kaltenbach T, Rouse RV, Park W, Maheshwari A, Sato T, Matsui S, Friedland S: **Prevalence of nonpolypoid (flat and depressed) colorectal neoplasms in asymptomatic and symptomatic adults.** *JAMA* 2008, **299:**1027–1035.
9. Azzoni C, Bottarelli L, Campanini N, Di Cola G, Bader G, Mazzeo A, Salvemini C, Morari S, Di Mauro D, Donadei E, Roncoroni L, Bordi C, Sarli L: **Distinct molecular patterns based on proximal and distal sporadic colorectal cancer: arguments for different mechanisms in the tumorigenesis.** *Int J Colorect Dis* 2007, **22:**115–126.
10. Marshall KW, Mohr S, Khettabi FE, Nossova N, Chao S, Bao W, Ma J, Li XJ, Liew CC: **Blood-based biomarker panel for stratifying current risk for colorectal cancer.** *Int J Cancer* 2010, **126:**1177–1186.
11. Greene FL, Page DL, Fleming ID, Fritz A, Balch CM, Haller DG, Morrow M (Eds): *AJCC cancer staging manual.* 6th edition. New York: Springer; 2002.
12. Vanburen P, Ma J, Chao S, Mueller E, Schneider DJ, Liew CC: **Blood gene expression signatures associate with heart failure outcomes.** *Physiol Genomics* 2011, **43:**392–397.
13. Burakoff R, Hande S, Ma J, Banks PA, Friedman S, Makrauer F, Liew CC: **Differential regulation of peripheral leukocyte genes in patients with active Crohn's disease and Crohn's disease in remission.** *J Clin Gastroenterol* 2010, **44:**120–126.
14. Burakoff R, Chao S, Perencevich M, Ying J, Friedman S, Makrauer F, Odze R, Khurana H, Liew CC: **Blood-based biomarkers can differentiate ulcerative colitis from Crohn's disease and noninflammatory diarrhea.** *Inflamm Bowel Dis* 2011, **17:**1719–1725.
15. Tsuang MT, Nossova N, Yager T, Tsuang MM, Guo SC, Shyu KG, Glatt SJ, Liew CC: **Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report.** *Am J Med Genet B Neuropsychiatr Genet* 2005, **133B:**1–5.
16. Glatt SJ, Everall IP, Kremen WS, Corbeil J, Sásik R, Khanlou N, Han M, Liew CC, Tsuang MT: **Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia.** *Proc Natl Acad Sci USA* 2005, **102:**15533–15538.
17. Glatt SJ, Stone WS, Nossova N, Liew CC, Seidman LJ, Tsuang MT: **Similarities and differences in peripheral blood gene-expression signatures of individuals with schizophrenia and their first-degree biological relatives.** *Am J Med Genet B Neuropsychiatr Genet* 2011, **156B:**869–887.
18. Osman I, Bajorin DF, Sun TT, Zhong H, Douglas D, Scattergood J, Zheng R, Han M, Marshall KW, Liew CC: **Novel blood biomarkers of human urinary bladder cancer.** *Clin Cancer Res* 2006, **12:**3374–3380.
19. Liong ML, Lim CR, Yang H, Chao S, Bong CW, Leong WS, Das PK, Loh CS, Lau BE, Yu CG, Ooi EJJ, Nam RK, Allen PD, Steele GS, Wassmann K, Richie JP, Liew CC: **Blood-based biomarkers of aggressive prostate cancer.** *PLoS One* 2012, **7:**e45802.
20. Zaatar AM, Lim CR, Bong CW, Lee MML, Ooi JJ, Suria D, Raman R, Chao S, Yang H, Neoh SB, Liew CC: **Whole blood transcriptome correlates with treatment response in nasopharyngeal carcinoma.** *J Exp Clin Cancer Research* 2012, **31:**76.
21. von Wagner C, Good A, Smith SG, Wardle J: **Responses to procedural information about colorectal cancer screening using faecal occult blood testing: the role of consideration of future consequences.** *Health Expect* 2011, **15:**176–186.
22. de Wijkerslooth TR, de Haan MC, Stoop EM, Bossuyt PM, Thomeer M, van Leerdam ME, Essink-Bot ML, Fockens P, Kuipers EJ, Stoker J, Dekker E: **Reasons for participation and nonparticipation in colorectal cancer screening: a randomized trial of colonoscopy and CT colonography.** *Am J Gastroenterol* 2012, **107:**1777–1783.
23. Klabunde CN, Vernon SW, Nadel MR, Breen N, Seeff LC, Brown ML: **Barriers to colorectal cancer screening: a comparison of reports from primary care physicians and average-risk adults.** *Medical Care* 2005, **43:**939–944.
24. Vernon SW: **Participation in colorectal screening: a review.** *JNCI* 1997, **89:**1406–1422.

25. Worthley DL, Cole SR, Esterman A, Mehaffy S, Roosa NM, Smith A, Turnbull D, Young GP: **Screening for colorectal cancer by faecal occult blood test: why people choose to refuse.** *Intern Med J* 2006, **36:**607–610.
26. Lewis SF, Jensen NM: **Screening sigmoidoscopy. Factors associated with utilization.** *J Gen Intern Med* 1996, **11:**542–544.
27. Colorectal Association of Canada: *Screening and diagnostics. A guide to screening tests.* [http://www.colorectal-cancer.ca/en/screening/screening-tests].

doi:10.1186/1756-9966-32-44
Cite this article as: Chao *et al.*: Blood RNA biomarker panel detects both left- and right-sided colorectal neoplasms: a case–control study. *Journal of Experimental & Clinical Cancer Research* 2013 **32**:44.

Microbial & Biochemical Technology

Huang et al., J Microb Biochem Technol 2014, 6:2
http://dx.doi.org/10.4172/1948-5948.1000129

Research Article Open Access

A Gene Expression Profile of Peripheral Blood in Colorectal Cancer

Chi-Shuan Huang[1], Harn-Jing Terng[2], Yu-Chin Chou[3], Sui-Lung Su[3], Yu-Tien Chang[3,4], Chin-Yu Chen[2], Woan-Jen Lee[2], Chung-Tay Yao[5], Hsiu-Ling Chou[6], Chia-Yi Lee[3,4], Chien-An Sun[7], Ching-Huang Lai[3], Lu Pai[3,8], Chi-Wen Chang[9]*, Kang-Hwa Chen[9], Thomas Wetter[10],Yun-Wen Shih[3,4] and Chi-Ming Chu[3,4]*

[1]Division of Colorectal Surgery, Cheng Hsin Rehabilitation Medical Center, Taiwan
[2]Advpharma, Inc., Taipei, Taiwan
[3]School of Public Health, National Defense Medical Center, Taiwan
[4]Division of Biomedical Statistics and informatics, School of Public Health, National Defense Medical Center, Taiwan
[5]Department of Surgery, Cathay General Hospital, Taipei, Taiwan.
[6]Department of Nursing, Far Eastern Memorial Hospital & Oriental Institute of Technology, New Taipei, Taiwan.
[7]Department of Public Health, Fu Jen Catholic University, Taiwan
[8]Taipei Medical University, Taiwan
[9]Department of Nursing, College of Medicine, Chang Gung University, Taiwan
[10]Institute of Medical Biometry and Informatics, Heidelberg University, Germany and Department of Biomedical Informatics and Medical Education, University of Washington Seattle, USA

Abstract

Background: Optimal molecular markers for detecting colorectal cancer (CRC) in a blood-based assay were evaluated. Microarray technology has shown a great potential in the colorectal cancer research. Genes significantly associated with cancer in microarray studies, were selected as candidate genes in the study. Pooling Internet public microarray data sets can overcome the limitation by the small number of samples in previous studies.

Objective: Using public microarray data sets verifies gene expression profiles for colorectal cancer

Methods: Logistic regression analysis was performed, and odds ratios for each gene were determined between CRC and controls. Public microarray datasets of GSE 4107, 4183, 8671, 9348, 10961, 13067, 13294, 13471, 14333, 15960, 17538, and 18105 included 519 cases of adenocarcinoma and 88 controls of normal mucosa, which were used to verify the candidate genes from logistic models and estimated its external generality.

Results: A 7-gene model of CPEB4, EIF2S3, MGC20553, MAS4A1, ANXA3, TNFAIP6 and IL2RB was pairwise selected that showed the best results in logistic regression analysis (H-L p=1.000, R2=0.951, AUC=0.999, accuracy=0.968, specificity=0.966 and sensitivity=0.994).

Conclusions: A novel gene expression profile was associated with CRC and can potentially be applied to blood-based detection assays.

Keywords: Colorectal cancer; Gene expression; Microarray; Internet

Background

Colorectal cancer (CRC) is a common cancer worldwide [1]. An estimated 146,970 new cases of colon and rectal cancer and 49,920 deaths are expected to occur in 2009 in the United States [2]. CRC screening can possibly reduce the incidence of advanced disease and provide better overall, progression-free survival. Conventional CRC screening tests include fecal occult blood testing, flexible sigmoidoscopy, double-contrast barium enema X-ray, and colonoscopy [3]. Although they are commonly used, these tests have limitations, including highly variable sensitivity (i.e., 37% to 80%) and diet-test interactions [4].

The dissemination of malignant cells from a primary neoplasm is the pivotal event in cancer progression. In many clinical cases, tumor cells metastasize before the primary tumor is diagnosed. Individual circulating tumor cells may be the earliest detectable form of metastasis [5]. PCR-based analyses of mRNA from cytokeratins, the carcinoembryonic antigen (CEA), and epidermal growth factor receptor (EGFR) genes in peripheral blood samples from CRC patients have been reported [6]. However, the low sensitivities and specificities for these well-known genes are not considered acceptable for the detection of colorectal cancer. Recently, multiple biomarkers were reported for the detection of colorectal cancer that delivered a better sensitivity or specificity [7,8].

In the present study, expression levels of 28 cancer-associated candidate genes in the peripheral blood samples from 111 colorectal cancer patients and 227 non-cancer controls were analyzed using quantitative real time-PCR. Genes correlated with CRC were selected, and a discrimination model was constructed using multivariate logistic regression. Sensitivity, specificity, accuracy, positive and negative predictive values, and the AUC of the discrimination model are reported. Meanwhile, from this study (Model 1: 5 genes), Marshall et al. [7] (Model 2: 7 genes) and Han et al. [8] (Model 3: 5 genes), the 17 candidate genes were validated by pooling12 public microarray data sets as well as the external validation.

***Corresponding authors:** Prof. Dr. Chi-Ming Chu, Division of Bioinformatics and Statistics, Department of Epidemiology, School of Public Health, National Defense Medical Center, Taiwan, Tel: 886-2-87923100 X18438; Fax: 886-2-87923147; E-mail: chuchiming@web.de

Dr. Chi-Wen Chang, Department of Nursing, College of Medicine, Chang Gung University, Taiwan, E-mail: chuchiming@web.de

Received January 15, 2014; **Accepted** February 12, 2014; **Published** February 19, 2014

Citation: Huang CS, Terng HJ, Chang CW, Chou YC, Chang YT, et al. (2014) A Gene Expression Profile of Peripheral Blood in Colorectal Cancer. J Microb Biochem Technol 6: 102-109. doi:10.4172/1948-5948.1000129

Methods

Patients, controls, and blood samples

One hundred eleven patients with histologically confirmed colorectal cancer were enrolled (2006-2009) in a prospective investigational protocol, which was approved by the Institutional Review Board at Cheng Hsin Rehabilitation Medical Center (Taipei, Taiwan). CRC patients at different stages were classified according to the TNM system (Table 1). Peripheral blood samples (6-8 ml) were drawn from patients before any therapeutic treatment, including surgery, but after written informed consent was obtained. All blood samples were collected using BD vacutainer CPT™ tubes containing sodium citrate as an anti-coagulant (Becton Dickinson, NJ, USA) and were stored at 4°C.

The healthy controls were 227 volunteers who had come in for a routine health examination and had no evidence of any clinically detectable cancer disease. Each participant gave informed consent for the analysis. The same volume of peripheral blood was collected from controls as from patients. Samples were randomly divided into a training set (n=162) and a testing set (n=176). There were no significant differences in age, sex, cancer stage or tumor site between the two sets (Table 1).

RNA isolation and reverse transcription

The mononuclear cell (MNC) fraction was isolated within three hours after blood collection using BD vacutainer CPT™ tubes (Becton Dickinson, NJ, USA), according to the manufacturer's instructions. Total RNA was then extracted from the MNC fraction using the Super RNApure™ kit (Genesis, Taiwan) according to the manufacturer's instructions. The average yield of total RNA per milliliter of peripheral blood was 1.6 µg. The mRNA quality was assessed by the electrophoresis of total RNA, followed by staining with ethidium bromide, which showed two clear rRNA bands of 28S and 18S. Using a spectrophotometer, the ratio of the absorbance of each RNA at 260 and 280 nm (A260:A280) was confirmed to be greater than 1.7, which is an indicator of RNA purity [9]. One microgram of total RNA was used for cDNA synthesis with random hexamer primers (Amersham Bioscience, UK) and Superscript™ II reverse transcriptase (Invitrogen, USA).

Quantitative real-time PCR

Real-time PCR was performed using pre-designed, gene-specific amplification primer sets purchased from Advpharma, Inc. (Taiwan), nucleotide probes from Universal ProbeLibrary™ (Roche, Germany) and TaqMan® Master Mix (Roche) on a Roche LightCycler® 1.5 instrument. The hypoxanthine phosphoribosyltransferase 1 (HPRT1) gene was used as the internal control because its expression accurately reflects the mean expression of multiple commonly used normalization genes [10,11]. The cycle number for each candidate gene, Ct(test), was normalized against the cycle number of HPRT1, that is, C_t(HK). The calculation is performed as follows: ΔC_t(test)=Ct(HK) - Ct(test). The derived (normalized) value, ΔC_t(test), for each candidate gene is presented as the relative difference as compared to the mRNA expression level of the reference gene [12].

Preliminary screening of investigating genes

In the preliminary screening for CRC-associated genes, we selected candidate genes from the published microarray study [14], and tested for the relative range of expression levels using real-time PCR. There were totally 28 gene candidates for first run of screening, including 12 genes, which were reported as risk for cancer prognosis [14], 14 genes identified as correlated with the incidence of tumor tissues (unpublished data), and two genes with elevated expression in colon cancer patients, A3 adenosine receptor and CCSP-2 [15,16]. Since the measurement of a higher cycle number (i.e., Ct greater than 30) generally implies lower amplification efficiency [17,18], 15 genes were used for further analysis (Table 2) after eliminating genes with low amplification efficiency.

Statistical analysis

The chi-square test and *t*-test were performed to characterize sex and age distributions between cases and controls. The transcript levels of candidate genes were tested statistically for differences between the case and control samples, using the t-test. A logistic regression was performed, and odds ratios were determined in order to study the association of candidate genes with CRC. The power of the study was 100% for each candidate gene [13]. The statistical alpha level was 0.05.

Multivariate logistic regression was used to analyze the relationship of the cases and controls to the ΔC_t(test) values of candidate genes. The logistic probabilities were calculated using the modeling equations from logistic regression analysis. Diagnostic performances were further used to evaluate multivariate logistic models, including sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). We used the Hosmer-Lemeshow test to check goodness-of-fit. A receiver operating characteristic (ROC) curve analysis was

	Training set (n=162)			**Testing set (n=176)**			**P-value**	
	CRC (n=55)	Non-CRC (n=107)	**P-value**	**CRC (n=56)**	**Non-CRC (n=120)**	**P-value**	**Cases**	**Controls**
Age, yr (S.E.)	66.47 (1.50)	68.31 (1.12)	0.335	67.38 (1.83)	69.99 (1.03)	0.216	0.704	0.270
Gender, no. (%)								
Male	32 (58.2)	58(54.2)	0.630	28 (50.0)	73 (60.8)	0.176	0.387	0.313
Female	23 (41.8)	49(45.8)		28 (50.0)	47 (39.2)			
Stage, no. (%)								
I	21 (38.2)	-		15 (26.8)	-			
II	10 (18.2)	-	-	9 (16.1)	-	-	0.447	-
III	14 (25.5)	-		21 (37.5)	-			
IV	10 (18.2)	-		11 (19.6)	-			
Tumor site, no. (%)								
Colon	28 (50.9)			30 (53.6)				
Rectum	22 (40.0)	-	-	16 (28.6)	-	-	0.286	-
Cecum	4 (7.3)			5 (8.9)				
Colon+Rectum	1 (1.8)			5 (8.9)				

CRC: ColonRectal Cancer; *Data are given as means (SE) or as the number of cases (%); §P values were estimated using the t-test

Table 1: Characteristics of the training and testing sets*§.

			95% CI of OR		
	B	OR	Upper	Lower	P-value
Sex	0.577	1.780	7.582	0.418	0.435
Age	0.028	1.028	1.083	0.976	0.293
MCM4	0.142	1.152	4.504	0.295	0.838
ZNF264	1.450	4.265	18.208	0.999	0.050
RNF4	-0.550	0.577	5.146	0.065	0.622
GRB2	2.009	7.456	37.131	1.497	0.014
MDM2	1.359	3.892	15.166	0.999	0.050
STAT2	-1.178	0.308	1.466	0.065	0.139
WEE1	1.264	3.540	14.784	0.848	0.083
DUSP6	2.465	11.769	40.330	3.435	<0.001
CPEB4	2.045	7.725	27.695	2.155	0.002
MMD	-1.067	0.344	0.865	0.137	0.023
NF1	-1.417	0.243	1.517	0.039	0.130
IRF4	0.057	1.059	3.350	0.335	0.923
EIF2S3	-2.105	0.122	0.718	0.021	0.020
EXT2	-1.933	0.145	1.235	0.017	0.077
POLDIP2	-1.294	0.274	1.515	0.050	0.138

B: coefficient of logistic regression; OR: odds ratio; CI: confidence interv

Table 2: Multivariate analysis of colorectal cancer-related molecular markers and the discrimination model based on age, sex, and 15 genes using the logistic regression model on the training set.

performed to determine the cut-off logistic probabilities and the areas under the ROC curves (AUC), to identify the performance of each candidate gene and combinations of multiple genes. A sensitivity analysis demonstrated the influence on performance of different cut-off logistic probabilities [Logit(P)] in the logistic model.

Internet public microarray data sets

The microarray gene expression data are from searches using “colon cancer” AND “human [organism]” AND “expression profiling by array [dataset type]” as the key words in the GEO database of the National Center for Biotechnology Information (NCBI). The eligible criteria were 1) the examined samples were frozen tissue sections of normal human colorectal mucosa, primary colorectal cancer or hepatic metastases from colorectal cancer; 2) the microarray platform used was limited to single-color, whole genome gene chips from Affymetrix; and 3) the data were presented as gene expression level. The exclusion criteria were 1) data from cultured cell lines or other in vitro assays; 2) datasets without the original gene expression level data files; and 3) those with redundant sub-datasets. A total of 175 GEO series (GSE) datasets were finally excluded, leaving 12 public microarray dataset of GSE 4107, 4183, 8671, 9348, 10961, 13067, 13294, 13471, 14333, 15960, 17538, and 18105, which included 519 cases of adenocarcinoma and 88 controls of normal mucosa.

Furthermore, we validated the 17 CRC-associated genes from the studies (Model 1: 5 genes), Marshall et al. [7] (Model 2: 7 genes) and Han et al. [8] (Model 3: 5 genes) and performed the multivariate logistic regression analysis using the pooled 12 public microarray data sets as well as the external validation.

Results

Genes correlated with colorectal cancer

A multivariate analysis based on age, sex and 15 genes was used in a logistic regression model in the training set because the peripheral blood samples were drawn from patients before any therapeutic treatment. Although this full model seemed capable of discriminating between the CRC cases and controls, it may have resulted in over fitting(Table 2). The logistic regression analysis further resulted in the selection of five genes of significance (i.e., P-value<0.05), MDM2, DUSP6, CPEB4, MMD, and EIF2S3, with odds ratios of 2.978, 6.029, 3.776, 0.538, and 0.138, respectively. This model was reduced to a panel of five genes in a forward stepwise regression, which statistical powers of the five genes were 1.00 between case and control groups in training and testing sets.

Discrimination of colorectal cancer and non-cancer controls using five genes

Five genes, i.e., MDM2, DUSP6, CPEB4, MMD, and EIF2S3, were significantly associated with CRC. A five-gene logistic regression model provided good discriminative performance with 87.0% accuracy, 78.0% sensitivity, 92.0% specificity, 90.7% positive predictive value (PPV), and 80.7% negative predictive value (NPV) in the training set. The five-gene model performed with 94.9% accuracy, 97.1% sensitivity, 81.8% specificity, 96.9% PPV, 82.8% NPV, and an area under the ROC (receiver operating characteristic) curve of 0.978 (0.912-1) in the external validation. Discrimination models can be constructed with one of the five genes selected, based on forward multivariate logistic regression analysis using the training set. AUCs were used to compare the performance of discrimination models for single genes and combinations of two, three, four, or five marker genes. The DUSP6 model (Table 3) displayed the best discrimination ability, with an AUC of 0.804 (95% C.I.: 0.730-0.879), as compared to other one-gene models (AUC: 0.49-0.69). Distinct increases in the AUC of up to 0.905 (95% C.I.: 0.849-0.960) resulted from the combination of five genes (Table 3). The five-gene model fulfilled the criteria of good performance for diagnostic tests as well as accuracy (87%), sensitivity (78%), and specificity (92%); in addition, the Hosmer-Lemeshow test was non-significant (P-value=0.108).

The cut-off value of Logit(P) for the five-gene model could also be adjusted to achieve high sensitivity or specificity, i.e., 99%, 95% or 90% (Table 4). The five-gene model performed stably for the discrimination between CRC cases and controls in the training set, with accuracies

				95% CI	
Genes used for models	**AUC**	**S.E.**	**P-value**	**Lower**	**Upper**
DUSP6, CPEB4, EIF2S3, MDM2, MMD	0.905	0.028	<0.001	0.849	0.960
DUSP6, CPEB4, EIF2S3, MDM2	0.895	0.030	<0.001	0.838	0.953
DUSP6, CPEB4, EIF2S3	0.882	0.032	<0.001	0.820	0.945
DUSP6, CPEB4	0.855	0.032	<0.001	0.791	0.919
DUSP6	0.804	0.038	<0.001	0.730	0.879

ROC: receiver operating characteristic; AUC: area under the ROC curve; S.E.: Standard Error; CI: confidence interval. P-values for AUC were estimated using the Z test

Table 3: Discrimination power and ROC analysis of different combinations of CRC-associated genes in training set.

a: Logistic probabilities for the training set

Logit(P)	Sensitivity	Specificit	PPV	NPV	Accuracy
0.0198	99.0%	16.0%	2.3%	99.9%	44.2%
0.0511	95.0%	63.0%	12.1%	99.6%	73.9%
0.1783	90.0%	72.0%	41.1%	97.1%	78.1%
0.5	78.0%	92.0%	90.7%	80.7%	87.0%
0.4747	80.0%	90.0%	87.8%	83.3%	86.6%
0.6845	61.0%	95.0%	96.4%	52.9%	83.5%
0.9012	25.0%	99.0%	99.6%	12.6%	73.9%

Logit(P): Logistic Probabilities; PPV: Positive Predictive Value; NPV: Negative Predictive Value

b: Performance of the statistical model on the training and testing sets with Logit(P)=0.5

	Training set	Testing set	External validation
Non-Cancers	107	120	88
True negative	98	110	72
False positive	9	10	16
Colorectal Cancers	55	56	519
False negative	12	19	15
True positive	43	37	504
Total	162	176	519
Sensitivity	78.0%	66.0%	97.1%
Specificit	92.0%	92.0%	81.8%
PPV	90.7%	89.2%	96.9%
NPV	80.7%	73.0%	82.8%
Accuracy	87.0%	83.5%	94.9%

Logit(P): Logistic Probabilities; PPV: Positive Predictive Value; NPV: Negative Predictive Value

Table 4: Performance of the statistical model based on the five-gene profil

ranging from 73.9% to 87.0%, with sensitivity of 95%, or with specificity of 95%. In addition, a well performance in the testing set was obtained using the discrimination model, with 84% accuracy, 66% sensitivity, 92% specificity, 89% PPV and 73% NPV (Table 4b).

Pooling 12 microarray studies to verify the 17 candidate genes and estimate its external generality

Furthermore, we performed the multivariate logistic regression analysis for pooled 12 public microarray data sets as well as the external validation to verify the CRC-associated genes from 3 studies (the present one of Chu et al., Marshall et al. and Han et al.) [7,8]. As the Table 5, we validated the 17 CRC-associated genes from this study (Model 1: 5 genes), Marshall et al. [7] (Model 2: 7 genes) and Han et al. [8] (Model 3: 5 genes) by pooling 12 public microarray dataset of GSE 4107, 4183, 8671, 9348, 10961, 13067, 13294, 13471, 14333, 15960, 17538, and 18105, which included 519 cases of adenocarcinoma and 88 controls of normal mucosa. The goodness-of-fit test of Hosmer-Lemeshow (H-L) showed statistical significance (p=0.044) for Model 2 of Marshall et al. [7], which observed event rates did not match expected event rates in subgroups of the model population. Models for which expected and observed event rates in subgroups are similar are called well calibrated (Model 1, 3 and 4). A 7-gene model (Model 4 with genes CPEB4, EIF2S3, MGC20553, MAS4A1, ANXA3, TNFAIP6 and IL2RB) was pairwise selected from genes of Model 1, 2 and 3 that showed the best results in logistic regression analysis (H-L p=1.000, R2=0.951, AUC=0.999, accuracy=0.968, specificity=0.966 and sensitivity=0.994).

Discussion

Common serum tumor markers used in primary care practice have not demonstrated a survival benefit in randomized controlled trials for screening in the general population. Most of them showed elevated levels only in some early-stage or late-stage cancer patients [19]. A recent review of real-time PCR-based assays with single molecular markers, such as CEA, CK19, and CK20, demonstrated low sensitivity, was ranging from 4% to 35.9%, 25.9% to 41.9%, and 5.1% to 28.3%, respectively [6]. One study, performed with a newly identified molecular marker known as ProtM [20], also attained unsatisfactory sensitivity.

Circulating tumor cells from any cancer type are capable of disseminating from solid tumor tissues, penetrating and invading blood vessels and circulating in the peripheral blood [21,22]. The number of circulating tumor cells has been used to predict the clinical outcome of

	Model 1			**Model 2**			**Model 3**			**Model4**		
	B	**S.E.**	**p**	**B**	**S.E.**	**p**	**B**	**S.E.**	**p**	**B**	**S.E.**	**p**
5 Candidate genes of this study;												
MDM2	6.069	1.461	<0.001									
DUSP6	1.360	0.235	<0.001									
CPEB4	-3.177	0.383	<0.001							-4.423	1.160	<0.001
MMD	0.335	0.442	0.448									
EIF2S3	1.462	0.244	<0.001							2.604	0.856	0.002
7 Candidate genes of Marshall et al. [7]												
ANXA3				0.559	0.212	0.008				1.566	0.485	0.001
CLEC4D				46.259	9.918	<0.001						
LMNB1				1.883	0.330	<0.001						
PRRG4				-1.284	0.371	0.001						
TNFAIP6				1.787	0.377	<0.001				2.031	0.572	<0.001
VNN1				0.207	0.159	0.194						
IL2RB				0.269	0.216	0.213				1.824	0.637	0.004
5 Candidate genes of Han et al. [8]												
CDA							-0.496	0.090	<0.001			
MGC20553							-1.386	0.197	<0.001	-1.751	0.619	0.005
BANK1							0.565	0.373	0.129			
BCNP1							-0.944	1.148	0.411			
MAS4A1							-1.483	0.457	0.001	-1.907	0.590	0.001
Constant	-32.758	6.001	<0.001	-124.678	25.437	<0.001	16.601	2.995	<0.001	-14.268	6.968	0.041
H-L		0.460			0.044			0.194			1.000	
R2		0.853			0.841			0.693			0.951	
AUC		0.978			0.985			0.957			0.999	
Accuracy		0.949			0.974			0.939			0.990	
Specificit		0.818			0.886			0.716			0.966	
Sensitivity		0.971			0.988			0.977			0.994	

Model 1: 5 candidate genes of this study; Model 2: 7 candidate genes of Marshall et al. ; Model 3: 5 candidate genes of Han et al. ; Model 4: stepwise 7 candidate genes from model 1, 2 and 3; B: logistic regression coifficient beta; S.E.: standard error of B; p: p value with statistical significance; H-L: Hosmer and Lemeshow test p value R2: Nagelkerke R Square; AUC: area under ROC

Table 5: The logistic regression models for pooled 12 microarray data sets as the external validation of CRC-associated genes from 3 studies.

cancer patients [23]. On the basis of the presence of circulating tumor cells, we identified five molecular markers, MDM2, DUSP6, CPEB4, MMD, and EIF2S3, which were differentially expressed between peripheral blood samples of CRC patients and healthy controls. The application of multivariate logistic regression analysis resulted in a five-gene discrimination model, which achieved good diagnostic performance and provided stable conditions with accuracies ranging from 73.9% to 87.0%, with sensitivity of 95%, or with specificity of 95%.

Both mRNAs and proteins in the peripheral blood have been tested for diagnostic use to detect circulating tumor cells of different solid tumors or to determine prognoses of various cancers. We confirmed, in our study, that the AUCs of the discrimination models greatly improved from 0.80 for the model based on a single gene (DUSP6) to 0.91 for the combined model with all five genes. More and more clinical studies show improvements in the sensitivity of cancer detection by assaying transcript levels of multiple genes in patient peripheral blood [7,8,24].

A higher sensitivity or specificity of the discriminatory performance of our five-gene model was achieved by adjusting the cut-off value of Logit(P) (Table 4a). This five-gene discrimination model with Logit(P)=0.0511 had a sensitivity of 95%, a specificity of 63%, and an accuracy of 74%, which is ideal for screening colorectal cancer. However, setting Logit(P) to 0.4747 resulted in specificity of 90%, sensitivity of 80% and an accuracy of 86%, which indicates that our five-gene model is robust and highly accurate for discriminating CRC from healthy or benign conditions. Similar accuracy rates (i.e., 80% to 86%) were achieved with Logit(P), ranging from 0.0511 to 0.4747. In the testing set, the five-gene model performed with satisfactory accuracy, sensitivity, and specificity.

Two reports [7,8] with similar screening approaches used different gene sets to detect CRC (Table 5). The two gene sets were obtained by direct selection from differentially expressed genes in peripheral blood samples using microarray techniques followed by real-time PCR. The biomarkers they selected may more or less reflect the static and dynamic changes of the immune system in response to cancer. The strategy of our study was to choose genes clinically confirmed to be cancer-associated in tumor tissues and to validate in peripheral blood samples. Five genes (MDM2, DUSP6, CPEB4, MMD and EIF2S3) identified here for discrimination between CRC patients and healthy controls showed strong association with CRC. MDM2 (Mouse double minute 2 homolog) gene, also known as HDM2 gene in human, is a negative regulator of the tumor suppressor protein p53 [25]. Overexpression of MDM2 gene was reported in several human tumor types, including osteosarcomas, melanoma, non-small cell lung cancer (NSCLC), esophageal cancer, leukemia, and non-Hodgkin's lymphoma [26-31]. Inhibition of MDM2 can restore p53 activity in cancers containing wild-type p53 and has recently become a strategy to develop anti-tumor drug [32-35].

DUSP6, the dual-specificity MAP kinase phosphatase 3 (also known as MKP3), inactivates ERK1/ERK2 [36,37]. Clinical studies based on tumor tissues demonstrated that elevated DUSP6 transcript (mRNA) level was a risk factor for clinical outcome in non-small cell lung cancer (NSCLC) patients (hazard ratio=2.2) [14] and stronger protein level was identified in 31% of primary human NSCLC tumor using Immunohistochemistry [38]. Furthermore, overexpression of DUSP6 was associated with papillary and poorly differentiated thyroid carcinoma both at the mRNA and protein level [39,40] and with KRAS mutant colon cancer [41]. In addition, higher expression level of DUSP6 was found in the tamoxifen-resistant breast tumors group compared with the tamoxifen-sensitive tumor group [42] and tumor growth promotion in glioblastoma [43]. The DUSP6 function might vary in different cancer types. On the contrary, some other reports demonstrated that DUSP6 gene was a candidate tumor suppressor gene, for instance, in pancreatic cancer [44], esophageal squamous cell carcinoma [45], and lung cancer with 17.7% cases of study sample [46].

CPEB4, cytoplasmic polyadenylation element binding protein, targets mRNAs and promote translation by inducing cytoplasmic polyadenylation [47,48]. Overexpressed CPEB4 was identified in pancreatic ductal adenocarcinomas and glioblastomas compared with its corresponding normal tissue [49]. Increased CPEB4 mRNA was considered as a prognostic marker for poor clinical outcome in non-small cell lung cancer (NSCLC) patients (hazard ratio=1.8) [14]. In contrast, reduced or weaker CPEB4 expression was observed in most of hepatocellular carcinoma samples compared with normal tissues using IHC staining [50]. In addition, Xu and Liu [51] proposed that CPEB4 gene might be selectively overexpressed in metastasic cancers, such as metastatic prostate cancer, and potentially as a biomarker for chemotherapy resistance.

MMD is an integral membrane protein with seven putative transmembrane segments [52]. Its biological function is still unclear. EIF2S3 is the largest subunit (gamma) of eukaryotic translation initiation factor 2 (EIF2) [53] and might be indirectly involved in inhibition of prostate cancer metastasis through N-myc downstream regulated gene 1 [54].

Our study has firstly presented that four expressing genes in PBMC-derived fractions, including MDM2, CPEB4, EIF2S3 and MMD, have the direct association with CRC with significance. As many clinical studies have been reported, MDM2, DUSP6 and CPEB4 have been showed their association with other pathologies, especially different cancer types. These observations might provide the evidence that these biomarkers play central roles during carcinogenesis or malignance of tumor, but with different strength depending on cancer type. Indeed, it is important to have multiple biomarkers integrated in developed diagnostic or prognostic methods, while each candidate gene has its independent power and the efficacy to discriminate cancer and normal subject (Table 3).

There are several limitations of current study. Since the small number of different stages in the study CRC sample, we were not able to know whether individual gene expression or the five gene signature is stage-dependent. Secondly, the change of gene expression level in the BPMC fraction of CRC patient before and after treatment was not studied due to the restriction of single blood drawing of IRB-approved clinical protocol. Thirdly, the collection of survival status information is not completed and prognostic value of biomarkers could not be evaluated in this study due to many censored cases (over 50%; 5-year survival rate of CRC patients is around 50%).

Further investigation is warranted on the potential of gene signature for evaluation of clinical staging, metastatic probability and survival in CRC patients, when information for the disease progression and survival is completely collected. In addition, the application of currently identified gene signature for CRC detection is very important and it is also the goal for assay development. Discussion with physicians will be planning for integration the test of our CRC-specific gene signature into the national screening program for CRC. Especially, the diagnostic performance between this CRC-specific gene signature and current screening method, such fecal occult blood test (FOBT) or colonoscopy should be approached. As well as the potential of the individual gene expression or gene signature for evaluation of therapeutic response should be planned as future work.

Furthermore, we verified the CRC-associated genes by pooling 12 public microarray data sets that the four logistic models performed similar AUCs without statistically significant difference. In the future, the 7-gene logistic regression model (Model 4: CPEB4, EIF2S3, MGC20553, MAS4A1, ANXA3, TNFAIP6 and IL2RB) showed the best results that can be further verified for more samples. Meanwhile, the causal relations are needed to confirm among the selected genes and CRC. The expression signature of these CRC-associated genes should be evaluated for early detection of CRC, with more samples randomly screened from the population; in addition, subjects who eventually receive a diagnosis of CRC should be evaluated as well. Early CRC detection could provide inherent benefits to the patient and could also enable screening for post-operative residual tumor cells and occult metastases, an early indicator of tumor recurrence. Early detection could thus improve survival in patients before symptoms are detectable, during treatment, or during remission.

Conclusion

In conclusion, we found that the expression profile of 7 genes, CPEB4, EIF2S3, ANXA3, TNFAIP6, IL2RB, MGC20553 and MAS4A1, is highly associated with colorectal cancer. Detection of cancer cell-specific biomarkers in the peripheral blood can be an effective screening strategy for CRC.

Competing Interests

No other potential conflict of interest relevant to this article was reported. The authors have declared that no competing interests exist. Dr. Terng, Woan-Jen Lee, and Chin-Yu Chen report being employees of Advpharma. Authors of Advpharma involved in partial sample collection, laboratoryexperiment and manuscript.

Financial Disclosure

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Finance was supported by Taiwan's SBIR promoting program from the Department of Industrial Technology of Ministry of Economic Affairs and by research grants from National Defense Medical Center of Ministry of Defense.

Ethics Statement

Patients with histologically confirmed colorectal cancer were enrolled (2006-2009) in a prospective investigational protocol, which was approved by the Institutional Review Board at Cheng Hsin Rehabilitation Medical Center (Taipei, Taiwan). CRC patients at different stages were classified according to the TNM system (Table 1). Peripheral blood samples (6-8 ml) were drawn from patients before any therapeutic treatment, including surgery, but after written informed consent were obtained.

Authors' contributions

Conception and design: Chi-Ming Chu and Harn-Jing Terng. Administrative support: Chi-Ming Chu, Harn-Jing Terng, Woan-Jen Lee and Chin-Yu Chen. Provision of study materials or patients: Woan-Jen Lee and Chin-Yu Chen. Collection and assembly of data: Chi-Suan Huang, Woan-Jen Lee, Chin-Yu Chen, and Yun-Wen Shih. Data analysis and interpretation: Chi-Ming Chu and Yun-Wen

Shih. Manuscript writing: Chi-Ming Chu, Harn-Jing Terng, and Yun-Wen Shih. Final approval of manuscript: Chi-Ming Chu, Harn-Jing Terng, Chi-Suan Huang, Woan-Jen Lee, Chin-Yu Chen, and Mark L. Wahlqvist.

Acknowledgments

Research Funding: This work was supported by Taiwan's SBIR promoting program from the Department of Industrial Technology of the Ministry of Economic Affairs, Advpharma, Inc., and the National Defense Medical Center (NDMC), Bureau of Military Medicine, Ministry of Defense, Taiwan. Employment or Leadership Position: Chi-Ming Chu, NDMC; Harn-Jing Terng, Advpharma, Inc. Taiwan. Honoraria: Wu-ChienChien; Ching-Huang Lai, NDMC; Mark Sarno, Stason Pharmaceuticals, Inc., CA, USA.Expert Testimony: None. Other Remuneration: None.

References

1. Parkin DM, Bray F, Ferlay J, Pisani P (2005) Global cancer statistics, 2002. CA Cancer J Clin 55: 74-108.
2. Jemal A, Siegel R, Ward E, Hao Y, Xu J, et al. (2008) Cancer statistics, 2008. CA Cancer J Clin 58: 71-96.
3. Smith RA, Cokkinides V, Eyre HJ (2006) American Cancer Society guidelines for the early detection of cancer, 2006. CA Cancer J Clin 56: 11-25.
4. Levin B, Lieberman DA, McFarland B, Smith RA, Brooks D, et al. (2008) 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 58: 130-160.
5. Fidler IJ (1990) Critical factors in the biology of human cancer metastasis: twenty-eighth G.H.A. Clowes memorial award lecture. Cancer Res 50: 6130-6138.
6. Sergeant G, Penninckx F, Topal B (2008) Quantitative RT-PCR detection of colorectal tumor cells in peripheral blood--a systematic review. J Surg Res 150: 144-152.
7. Marshall KW, Mohr S, Khettabi FE, Nossova N, Chao S, et al. (2010) A blood-based biomarker panel for stratifying current risk for colorectal cancer. Int J Cancer 126: 1177-1186.
8. Han M, Liew CT, Zhang HW, Chao S, Zheng R, et al. (2008) Novel blood-based, five-gene biomarker set for the detection of colorectal cancer. Clin Cancer Res 14: 455-460.
9. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd edn 1989, NY Cold Spring Harbor Laboratory Press.
10. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: RESEARCH0034.
11. de Kok JB, Roelofs RW, Giesendorf BA, Pennings JL, Waas ET, et al. (2005) Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. Lab Invest 85: 154-159.
12. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408.
13. The Survey System: Sample Size Calculator.
14. Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, et al. (2007) A five-gene signature and clinical outcome in non-small-cell lung cancer. N Engl J Med 356: 11-20.
15. Xin B, Platzer P, Fink SP, Reese L, Nosrati A, et al. (2005) Colon cancer secreted protein-2 (CCSP-2), a novel candidate serological marker of colon neoplasia. Oncogene 24: 724-731.
16. Gessi S, Cattabriga E, Avitabile A, Gafa' R, Lanza G, et al. (2004) Elevated expression of A3 adenosine receptors in human colorectal cancer is reflected in peripheral blood cells. Clin Cancer Res 10: 5895-5901.
17. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45.
18. Hunt M: Real Time Pcr. In Book Real Time Pcr (Editor Ed.^Eds.). City.
19. Perkins GL, Slater ED, Sanders GK, Prichard JG (2003) Serum tumor markers. Am Fam Physician 68: 1075-1082.
20. Schuster R, Max N, Mann B, Heufelder K, Thilo F, et al. (2004) Quantitative real-time RT-PCR for detection of disseminated tumor cells in peripheral blood of patients with colorectal cancer using different mRNA markers. Int J Cancer 108: 219-227.
21. Bogenrieder T, Herlyn M (2003) Axis of evil: molecular mechanisms of cancer metastasis. Oncogene 22: 6524-6536.
22. Carmeliet P, Jain RK (2000) Angiogenesis in cancer and other diseases. Nature 407: 249-257.
23. Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, et al. (2004) Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med 351: 781-791.
24. Shen C, Hu L, Xia L, Li Y (2008) Quantitative real-time RT-PCR detection for survivin, CK20 and CEA in peripheral blood of colorectal cancer patients. Jpn J Clin Oncol 38: 770-776.
25. Reifenberger G, Liu L, Ichimura K, Schmidt EE, Collins VP (1993) Amplification and overexpression of the MDM2 gene in a subset of human malignant gliomas without p53 mutations. Cancer Res 53: 2736-2739.
26. Marchetti A, Buttitta F, Girlando S, Dalla Palma P, Pellegrini S, et al. (1995) mdm2 gene alterations and mdm2 protein expression in breast carcinomas. J Pathol 175: 31-38.
27. Bueso-Ramos CE, Yang Y, deLeon E, McCown P, Stass SA, et al. (1993) The human MDM-2 oncogene is overexpressed in leukemias. Blood 82: 2617-2623.
28. Ladanyi M, Cha C, Lewis R, Jhanwar SC, Huvos AG, et al. (1993) MDM2 gene amplification in metastatic osteosarcoma. Cancer Res 53: 16-18
29. Momand J, Jung D, Wilczynski S, Niland J (1998) The MDM2 gene amplification database. Nucleic Acids Res 26: 3453-3459.
30. Onel K, Cordon-Cardo C (2004) MDM2 and prognosis. Mol Cancer Res 2: 1-8.
31. Heist RS, Zhou W, Chirieac LR, Cogan-Drew T, Liu G, et al. (2007) MDM2 polymorphism, survival, and histology in early-stage non-small-cell lung cancer. J Clin Oncol 25: 2243-2247.
32. Wasylyk C, Salvi R, Argentini M, Dureuil C, Delumeau I, et al. (1999) p53 mediated death of cells overexpressing MDM2 by an inhibitor of MDM2 interaction with p53. Oncogene 18: 1921-1934.
33. Chène P, Fuchs J, Bohn J, García-Echeverría C, Furet P, et al. (2000) A small synthetic peptide, which inhibits the p53-hdm2 interaction, stimulates the p53 pathway in tumour cell lines. J Mol Biol 299: 245-253.
34. Wang H, Nan L, Yu D, Lindsey JR, Agrawal S, et al. (2002) Anti-tumor efficacy of a novel antisense anti-MDM2 mixed-backbone oligonucleotide in human colon cancer models: p53-dependent and p53-independent mechanisms. Mol Med 8: 185-199.
35. Tortora G, Caputo R, Damiano V, Bianco R, Chen J, et al. (2000) A novel MDM2 anti-sense oligonucleotide has anti-tumor activity and potentiates cytotoxic drugs acting by different mechanisms in human colon cancer. Int J Cancer 88: 804-809.
36. Keyse SM (2008) Dual-specificity MAP kinase phosphatases (MKPs) and cancer. Cancer Metastasis Rev 27: 253-261.
37. Zhou B, Wu L, Shen K, Zhang J, Lawrence DS, et al. (2001) Multiple regions of MAP kinase phosphatase 3 are involved in its recognition and activation by ERK2. J Biol Chem 276: 6506-6515.
38. Zhang Z, Kobayashi S, Borczuk AC, Leidner RS, Laframboise T, et al. (2010) Dual specificity phosphatase 6 (DUSP6) is an ETS-regulated negative feedback mediator of oncogenic ERK signaling in lung cancer cells. Carcinogenesis 31: 577-586.
39. Degl'Innocenti D, Romeo P, Tarantino E, Sensi M, Cassinelli G, et al. (2013) DUSP6/MKP3 is overexpressed in papillary and poorly differentiated thyroid carcinoma and contributes to neoplastic properties of thyroid cancer cells. Endocr Relat Cancer 20: 23-27.
40. Lee JU, Huang S, Lee MH, Lee SE, Ryu MJ, et al. (2012) Dual specificity phosphatase 6 as a predictor of invasiveness in papillary thyroid cancer. Eur J Endocrinol 167: 93-101.
41. De Roock W, Janssens M, Biesmans B, Jacobs B, De Schutter J, et al. (2009) DUSPs as markers of MEK/Erk activation in primary colorectal cancer. J Clin Oncol 27: 4064.
42. Cui Y, Parra I, Zhang M, Hilsenbeck SG, Tsimelzon A, et al. (2006) Elevated

expression of mitogen-activated protein kinase phosphatase 3 in breast tumors: a mechanism of tamoxifen resistance. Cancer Res 66: 5950-5959.

43. Messina S, Frati L, Leonetti C, Zuchegna C, Di Zazzo E, et al. (2011) Dual-specificity phosphatase DUSP6 has tumor-promoting properties in human glioblastomas. Oncogene 30: 3813-3820.

44. Furukawa T, Yatsuoka T, Youssef EM, Abe T, Yokoyama T, et al. (1998) Genomic analysis of DUSP6, a dual specificity MAP kinase phosphatase, in pancreatic cancer. Cytogenet Cell Genet 82: 156-159.

45. Ma J, Yu X, Guo L, Lu SH (2013) DUSP6, a tumor suppressor, is involved in differentiation and apoptosis in esophageal squamous cell carcinoma. Oncol Lett 6: 1624-1630.

46. Okudela K, Yazawa T, Woo T, Sakaeda M, Ishii J, et al. (2009) Down-Regulation of DUSP6 Expression in Lung Cancer. Am J Pathol 175:867-881.

47. Huang YS, Kan MC, Lin CL, Richter JD (2006) CPEB3 and CPEB4 in neurons: analysis of RNA-binding specificity and translational control of AMPA receptor GluR2 mRNA. EMBO J 25: 4865-4876.

48. Hake LE, Richter JD (1994) CPEB is a specificity factor that mediates cytoplasmic polyadenylation during Xenopus oocyte maturation. Cell 79: 617-627.

49. Ortiz-Zapater E, Pineda D, Martínez-Bosch N, Fernández-Miranda G, Iglesias M, et al. (2011) Key contribution of CPEB4-mediated translational control to cancer progression. Nat Med 18: 83-90.

50. Tian Q, Liang L, Ding J, Zha R, Shi H, et al. (2012) MicroRNA-550a acts as a pro-metastatic gene and directly targets cytoplasmic polyadenylation element-binding protein 4 in hepatocellular carcinoma. PLoS One 7: e48958.

51. Xu H, Liu B (2013) CPEB4 is a candidate biomarker for defining metastatic cancers and directing personalized therapies. Med Hypotheses 81: 875-877.

52. Rehli M, Krause SW, Schwarzfischer L, Kreutz M, Andreesen R (1995) Molecular cloning of a novel macrophage maturation-associated transcript encoding a protein with several potential transmembrane domains. Biochem Biophys Res Commun 217: 661-667.

53. Gaspar NJ, Kinzy TG, Scherer BJ, Hümbelin M, Hershey JW, et al. (1994) Translation initiation factor eIF-2. Cloning and expression of the human cDNA encoding the gamma-subunit. J Biol Chem 269: 3415-3422.

54. Tu LC, Yan X, Hood L, Lin B (2007) Proteomics analysis of the interactome of N-myc downstream regulated gene 1 and its interactions with the androgen response program in prostate cancer cells. Mol Cell Proteomics 6: 575-588.

Citation: Huang CS, Terng HJ, Chang CW, Chou YC, Chang YT, et al. (2014) A Gene Expression Profile of Peripheral Blood in Colorectal Cancer. J Microb Biochem Technol 6: 102-109. doi:10.4172/1948-5948.1000129

ColonSentry®

Regulatory Support

STAGE ZERO
LIFE SCIENCES

COLONSENTRY NEW YORK STATE'S DEPARTMENT OF HEALTH APPROVAL

On February 21, 2012 GeneNews announced that Enzo Clinical Labs, a division of Enzo Biochem, Inc. (NYSE:ENZ), had received approval from New York State's Department of Health to market ColonSentry, a blood-based test to assess an individual's current risk for colorectal cancer.

ColonSentry®

Evaluation & Customer Base

STAGE ZERO LIFE SCIENCES

VALIDATION OF THE SCIENTIFIC PRINCIPLE

The Sentinel Principle is an award winning technology developed by GeneNews based on the scientific observation that circulating white blood cells reflect, in a detectable way, what is occurring throughout the body. Many clinical studies have demonstrated that whole-blood RNA transcript-based profiles can be used to develop personalized gene expression signatures capable of differentiating patients with cancer from healthy patients across a broad spectrum of pathologies. ColonSentry® specifically measures gene expression in white blood cells, or leukocytes. Tumors are known to affect the gene expression profiles of circulating leukocytes, including both myeloid cells (monocytes, macrophages, etc.) and lymphocytes (T cells, B cells, and natural killer cells). This occurs due to a unique interaction between tumor cells and the immune system that has been referred to as “cancer immunoediting.”

Immunoediting comprises three stages: elimination (in which the immune system identifies cancerous and/or precancerous cells and attempts to eradicate them), equilibrium (in which the surviving tumor cells begin mutating rapidly under selection pressure from immune-system-mediated attacks), and escape (in which tumor cells that have acquired resistance proliferate uncontrollably, leading to tumor progression).Each of these stages induces leukocyte gene expression changes that constitute a unique, detectable molecular signature.

CLINICAL DATA TO BACKUP TEST

The ColonSentry assay was originally developed and clinically and analytically validated using the blood samples taken from screening colonoscopy subjects at twenty-four centers who recruited 10,000 patients located primarily in the Greater Toronto Area and surrounding region, and also in the United States. Classification model was trained on 232-sample reference set (112 CRC and 120 controls). Clinical validation was performed on an independent set of 410 average-risk samples (202 CRC and 208 controls).

A panel of 7 genes was selected from a genome-wide analysis using Affymetrix U133 Plus 2 microarrays. The expression level for these genes was determined using qPCR on a training set with 112 CRC and 120 Control samples. A logistic regression equation was fitted to the data. This equation was then applied to the gene expression levels for a test set with 202 CRC and 208 Control samples. The discriminatory performance of the 7-gene panel was confirmed upon validation with the independent test set (ROC AUC = 0.80 and Accuracy = 73% in the training set; ROC AUC = 0.80 and Accuracy = 71% in the test set) [1]. The training and test sets used in this validation are hereafter collectively referred to as the “IJC Set”. The distribution of the scores from the logistic regression equation for the CRC and the Control groups was then used as the basis of determining the Current Risk of CRC for an individual sample using Baye’s theorem. The CRC and Control populations were scaled according to the known prevalence of CRC (0.7% in the average risk population). The proportion of CRC to Control samples at any given logistic regression score is compared to the 0.7% disease risk and is the estimate of the individual sample’s Current Risk for CRC.

Customer Base & Support

- **New York State Health Department:** Validation completed February 2012
- **Geisinger Health System:** Validation completed May of 2015
- **Mt. Sinai:** Validation completed 2015
- **More than 1,000 practices** have used the ColonSentry test throughout the U.S.

30-70 East Beaver Creek Rd
Richmond Hill, ON L4B 3B2

STAGE
ZERO
LIFE SCIENCES

8751 Park Central Drive, Suite 200
Richmond, VA 23227