



Published in final edited form as:

*Cancer Epidemiol Biomarkers Prev.* 2008 August ; 17(8): 2188–2193. doi:  
10.1158/1055-9965.EPI-07-2767.

## Assessment of serum proteomics to detect large colon adenomas

David F. Ransohoff<sup>1</sup>, Christopher Martin<sup>1</sup>, Wesley S. Wiggins<sup>2</sup>, Ben A. Hitt<sup>2</sup>, Temitope O. Keku<sup>1</sup>, Joseph A. Galanko<sup>1</sup>, and Robert S. Sandler<sup>1</sup>

<sup>1</sup>*Departments of Medicine, Epidemiology, and Biostatistics; University of North Carolina at Chapel Hill; Chapel Hill, NC 27514*

<sup>2</sup>*Correlogic Systems, Inc.; 1405 Research Blvd. Suite 220; Rockville, MD 20850*

### Abstract

A non-invasive blood test that could reliably detect early CRC or large adenomas would provide an important advance in colon cancer screening. The purpose of this study was to determine whether a serum proteomics assay could discriminate among persons with and without a large ( $\geq 1$ cm) colon adenoma. To avoid problems of ‘bias’ that have affected many studies about molecular markers for diagnosis, specimens were obtained from a previously-conducted study of CRC etiology in which bloods had been collected before the presence or absence of neoplasm had been determined by colonoscopy, helping to assure that biases related to differences in sample collection and handling would be avoided. Mass spectra of 65 unblinded serum samples were acquired using a nano-electrospray ionization source on a QSTAR-XL mass spectrometer. Classification patterns were developed using the ProteomeQuest® algorithm, performing measurements twice on each specimen, and then applied to a blinded validation set of 70 specimens. After removing 33 specimens that had discordant results, the “test group” comprised 37 specimens that had never been used in training. Although in the primary analysis no discrimination was found, a single post-hoc analysis, done after hemolyzed specimens had been removed, showed sensitivity of 78%, specificity of 53%, and an accuracy of 63% (95% CI: 53% to 72%). The results of this study, although preliminary, suggest that further study of serum proteomics, in a larger number of appropriate specimens, could be useful. They also highlight the importance of understanding sources of ‘noise’ and ‘bias’ in studies of proteomics assays.

### Keywords

serum profiling; screening; diagnosis; mass spectrometry; colon neoplasm

## BACKGROUND AND PURPOSE

In the USA colorectal cancer (CRC) is responsible for about 150,000 cancers and 75,000 deaths per year. (1) The large majority of CRCs are thought to arise from adenomatous colon polyps; large adenomas ( $\geq 1$ cm) are thought to become clinical cancer at a rate of roughly 1% per year (2) and, along with ‘early’ and curable CRC, constitute a major target of screening.

A non-invasive test for CRC would be very useful clinically. Among screening tests currently recommended, colonoscopy and sigmoidoscopy are invasive, require laxative prep, and may incur risks of bleeding, perforation, or complications of conscious sedation. Fecal occult blood testing is non-invasive but has very limited sensitivity and must be done yearly or every-other-

year, involving a collection process that may be bothersome and lead to low compliance among some people. There is an urgent need for a non-invasive procedure to identify patients with adenoma or carcinoma. One promising emerging technology is the use of serum profiling for the detection of cancers.

No markers for colon adenoma or carcinoma have been well-demonstrated, although a number of preliminary reports have suggested serum-based signals may be associated with these growths. (3-11) Profiling of serum using SELDI, followed by applying artificial neural network and support vector machine analysis, was used to identify patterns of markers that differentiated carcinoma, adenoma, and normal healthy people; (8,12) however it is not clear that results were assessed in subjects totally independent of those used in 'training,' to rule out the possibility of overfitting. (13)

Although potential serum markers for CRC have been studied by a number of groups, as noted above, the interpretation of such studies may be substantially limited by threats to validity. 'Overfitting,' a problem caused by chance, can occur when patterns or a list of analytes is derived from a large number of candidates that is 'fit' to a small number of subjects. Demonstrating that overfitting did not occur can be done by assessing the model derived in the 'training set' on subjects in a 'validation set' that is totally independent of those used in training. (13) Bias can occur when systematic differences among the compared groups account for the 'discrimination' found. (14)

This study was designed to determine if serum could be used to discriminate between people with and without large adenomatous colon polyps. To achieve the goal of reducing or eliminating the possibility of bias, (14) the population used was one undergoing screening colonoscopy in which bloods were drawn before the procedure and before the 'true state' was known. This feature of study design helps prevent bias that could occur from differential handling of specimens. To help achieve the goal of avoiding chance as the explanation for results, total independence of the 'validation' set was maintained throughout the entire study.

Bias, the most serious problem in non-experimental clinical research, (14) can come from many sources including the study population (e.g. if there are age, ethnic, or gender differences between the adenoma and adenoma-free subjects), the metabolic status of the subjects prior to serum collection (e.g. fed, fasted, GI evacuated), the way the bloods are collected (e.g. tube type, site of collection, timing relative to an invasive procedure), the way the serum is processed (e.g. length of time for clotting, temperature, clot removal) and stored (e.g. time to freezer, freezer temperature, freeze-thaw history), and the way the samples are analyzed (e.g. order in which samples are analyzed, days on which different sample types are analyzed). In an ideal study all these kinds of factors are controlled and variations are accounted for in a balanced experimental design; such control rarely happens in retrospectively collected sample collections, and it may be logistically difficult to achieve in a prospective study if the disease under study has a low incidence.

This study utilizes a set of already-collected sera that, because they were collected before the diagnosis was known, were believed to have considerably less bias in their collection and handling than samples in many other studies. The clinical question in this study - whether a serum proteomics approach can discriminate persons with and without large adenomatous polyps of the colon - was chosen in part because of the availability of rigorously collected set of specimens; however, the question is also clinically very relevant because large adenomas (over 1cm) are important precursors of CRC and constitute important targets that clinicians would like to discover and remove.

## METHODS

### Overall strategy

The overall strategy was to use specimens from a group of subjects who had been colonoscopically screened for colon neoplasia and in whom blood samples had been drawn in standardized manner before the colonoscopy procedure, so that bloods would be collected and handled in a uniform and 'blinded' manner.

### Study population

All aspects of the study population, their blood sample quality, selection, collection, and processing to serum were under the direction of UNC investigators (CM, JG, TK). This included the association of pathology with the samples, the creation of sample IDs, and the selection of those samples to use in model development as opposed to blinded validation. The key to the blinded validation set were known only to UNC investigators (CM, JG) and have not, even post-analysis, been unblinded to other participants, ensuring the complete independence of the data analysis and the scoring of the blinded samples. The source of specimens was two large cross-sectional studies conducted by the same research team for the Diet and Health Studies III and IV conducted between 1998 and 2002 at the University of North Carolina Hospital (UNCH), a large referral center in Chapel Hill, NC. (15,16). Both studies were designed to assess neoplasia etiology in relation to lifestyle and biologic risk factors, as measured by questionnaires, blood, and biopsies. Both studies were approved by the Committee for the Protection of the Rights of Human Subjects at the University of North Carolina School of Medicine, and all participants gave written informed consent.

Participants were recruited from consecutive patients who underwent screening colonoscopy at UNCH during the recruitment periods. Eligible patients were 30 years of age or older; sufficiently proficient cognitively and in the English language to complete a telephone risk-factor interview; and had no known history of familial polyposis, colitis, previous colonic resection, previous colon cancer or adenoma. Age, ethnicity, and gender were recorded along with a reference number anonymized to the researchers. Bloods were drawn prior to initiation of the procedure and were transported to a laboratory for processing of the serum without knowledge of the colonoscopy result. Polyps were removed at the time of colonoscopy by board-certified gastroenterologists or supervised gastroenterology trainees and were sent to a central laboratory for histologic interpretation. A single pathologist reviewed all slides and classified polyps as adenomatous, hyperplastic or other (e.g. lymphoid nodules, inflammatory, no pathological diagnosis). The anonymized records were then annotated, retrospectively, with the outcomes of all patients - normal, healthy; normal with hyperplastic; small adenoma +/- hyperplastic; medium adenoma +/- hyperplastic; and large adenoma +/- hyperplastic. For the purposes of this study, large adenomas were those equal to or greater than 1 cm in diameter as estimated by the colonoscopist. Participants with multiple adenomas were categorized according to largest adenoma. Participants whose colonoscopy procedure did not achieve complete visualization of the colon to the cecum, or who had unsatisfactory preparation, were excluded.

In both studies, prior to the day of the examination, all patients followed the same regimen which involved a 24 hour fast and bowel cleansing using the laxative Go-Lytely, a proprietary mixture of polyethylene glycol and electrolytes (sodium sulfate, sodium bicarbonate, sodium chloride and potassium chloride), or Phospho-Soda (Fleet), a sodium phosphate saline solution, according to a standard protocol. On the day of the procedure, prior to administration of medication, an IV catheter was inserted into the patient's arm and 10ml of blood was immediately withdrawn using a royal blue top mineral free vacutainer and stored temporarily in a refrigerator at 4C in the clinical gastroenterology unit to allow clotting. Tubes were then

transported to the lab in an adjacent building for serum separation within 2-6 hours. Diagnoses (e.g. “normal” or “adenoma”) were not noted on the specimen label, and all personnel handling the specimens were unaware of the diagnosis. Specimens from patients seen late in the day were processed the next morning after specimens had been stored at 4C overnight. Because of this broad range of times from collection to freezing, it is possible that ‘noise’ (that would obscure or degrade signal from adenoma) could be introduced into specimens, thus preventing detection of a signal or difference. (17,18) Tubes were centrifuged for 5 minutes at 2000 RPM using a fixed angle Adams compact clinical centrifuge (Becton Dickinson) with a standard rotor, to separate serum that was then collected in 3.5 ml cryogenic vials that were labeled, placed in freezer boxes, and stored at -80C. The time from spinning to freezing was not strictly controlled but was generally done within 20 minutes. All samples were aliquotted at a single time so that there was only a single freeze/thaw cycle prior to thawing for the current analysis. To prepare aliquots for analysis, vials were thawed on ice and then vortexed to mix contents thoroughly, before 250 microliter aliquots were withdrawn and placed in sterile 1.5 ml Eppendorf tubes and stored at -80C until shipped.

### Sample Allocation

Prior to the start of any analysis, investigators at UNC-CH randomly assigned the relevant sera of the two kinds of subjects (large adenoma +/- hyperplastic; normal, no hyperplastic) into two groups. In the model development group, the sample identity of large adenoma vs normal was provided; in the blinded validation group, only an anonymous identifier was provided. The intent of the random selection to the training and validation groups was to minimize possible biases arising from unequal distribution of factors, such as age, sex, smoking status, and sera separation/processing times, which might contribute to analytical variability. Once assigned, the samples were shipped on dry ice to Correlogic Systems Inc. for mass spectroscopy and data analysis.

Following model development and selection of the best model from the development phase, spectra from the blinded validation set were classified using that model, and the classification results were sent to the holder of the blinding key at UNC for scoring. The subsequent statistical analysis of the results and interpretation of the statistical significance of the classification were performed solely by JG and CM. A point estimate and 95% confidence interval for overall accuracy (true positives plus true negatives divided by the total number of predictions) was calculated to assess discrimination.

### Analytical mass spectrometry

Analysis of the serum by mass spectrometry was performed as a service by an independent research organization following detailed protocols established and provided by Correlogic Systems Inc. Briefly, serum samples, stored in a -80C freezer prior to use, were thawed at room temperature for 30 minutes and then mixed gently by vortex to ensure a complete suspension. For dilution, 2  $\mu$ l of serum was pipetted into 1.5 ml tubes containing 498  $\mu$ l of mobile phase consisting of 50% (v/v) Acetonitrile (Burdick & Jackson), 0.2% (v/v) Formic acid (Suprapur, EM Science), to give a final 250-fold dilution. Samples were then mixed well and held at 4C, overnight. Prior to analysis, samples were then centrifuged at 13,000 g for 15 min at 4C, and 150  $\mu$ l of supernatant was transferred into individual wells of a prewashed 96-well microtitre plate. Duplicates of each serum preparation were placed in adjacent wells. The plates were then covered to prevent evaporation with a heat sensitive film (ABgene) and sealed for 4 seconds using a thermo-sealer (ABgene). Prior to use the 96-well Sample Plates (NUNC 267245, 0.5ml polypropylene) were washed twice with deionized water, twice with mobile phase, and finally air dried in an inverted orientation. For spectral analysis, sera were distributed evenly across each 96-well autosampler plate so that spectra representing the adenoma, non-

adenoma and blinded samples were acquired in a positionally and temporally independent manner.

Mass spectra were acquired using an ABI QSTAR-XL mass spectrometer, with an Advion Nanomate 100 automated nanoelectrospray system. Tuning and calibration of the spectrometer were performed using  $7.5 \times 10^{-6}$  M CsI and  $1 \times 10^{-6}$  M Sex Pheromone Inhibitor iPD1 Octapeptide (ALILTLVS), according to the manufacturer's recommendation. Samples were held at room temperature for analysis. The spray pressure on the source was set to 0.6 psi and the voltage to 1.6 KV. Five  $\mu$ l of sample was picked up with a 1  $\mu$ l of air gap and sprayed for 1 min 10 seconds. Contact closure started 5 seconds after spray initiation, when a stable spray had been established. The spectra were acquired in positive TOF MS mode from 500 m/z to 14000 m/z, using 30 two-second cycles with MCA on. Duplicate spectra were acquired for all samples.

### Deriving discriminating patterns in the training data set

Analysis of the raw mass spectral data and the generation of potential classification models were performed by Correlogic Systems Inc. using methods previously established. Prior to modeling, all spectra were aligned by linear binning at 100ppm over the range 500 - 1100 m/z.

Three methods of data analysis were conducted and compared. Two were traditional classification schemes - k-nearest neighbor (kNN) and oneR; the third was a non-linear pattern recognition algorithm, ProteomeQuest® (PQ; Correlogic Systems, Inc.). All methods were performed using a 10-fold cross-validation strategy which holds out 10% of the model development set as a validation set. The results reported are the mean validation performances and standard deviations.

OneR (One Rule) is a simple classification algorithm that generates a one-level decision tree able to identify simple, yet accurate, classification rules that have been shown to be only slightly less accurate than state-of-the-art learning schemes. The strength of oneR is that it attempts to classify samples by identifying multiple cut-offs for a single feature. The level of classification obtained by this approach represents the extent to which a single feature can classify and provides a useful benchmark that any classification using multiple features must exceed. In contrast, both kNN and ProteomeQuest® use multiple features to classify.

To derive the best model using the training set provided, a set of classification rules or models was developed using 65 unblinded samples: 37 large adenomas (with or without hyperplastic polyp) and 28 normals without a hyperplastic polyp. The 'best discriminating' model was chosen and then was used to classify the blinded validation set consisting of 20 large adenoma samples and 50 normals. Before any analyses were done, it was recognized that the overall sample size was suboptimal and did not meet the usual recommendations of Correlogic Systems, Inc. who recommended a minimum of 100 samples each for the normal and diseased subjects for model building in a proof of principle study. The reason for proceeding, though, was the potential advantage of the rigorous control of collection and handling of the specimens that were available.

### Modeling, using ProteomeQuest®

Modeling was performed using the ProteomeQuest® (Correlogic Systems Inc) algorithm, which has been described. (19,20) Briefly, the algorithm uses an iterative procedure combining lead cluster mapping and a genetic algorithm to identify combinations of m/z features whose relative intensity ratios define a particular state. The resulting model is a centroid map. Each centroid is associated with a given state (diseased or normal) and is surrounded by a defined

decision boundary. The centroid and decision boundary define a node. To score unknown samples, the intensities of the features used to form the map are extracted from the spectrum of interest and are plotted relative to each other on the map. The classification of the unknown is then defined by the identity of the node that the data points fall within.

Modeling using a strategy such as ProteomeQuest® is especially difficult when small numbers of samples are available to build models, because it becomes increasingly difficult to differentiate between truly meaningful patterns from ‘accidental’ ones. One mitigating methodology to guard against ‘accidental’ patterns is a 10-fold cross-validation strategy that was implemented, in which the model building set was divided into 10 unique, non-overlapping subsets. Duplicate spectra were always kept together so that the spectra from any individual appeared only in a single subset. These were then assembled into ten unique groups in which all but one subset were grouped for model building and the tenth subset was held out for model validation. Then, using a single combination of modeling parameters, a model was built for each group. The resulting 10 models were designated a ‘cross-validation group’ of models, whose combined performance should more accurately reflect the performance of those parameter settings than any single model built with those parameters. Each sample was scored as 1 (large adenoma) or 0 (normal) using each of the 10 models in the set. The 10 scores were then summed to generate a final classification of each sample as positive (sum >5), negative (sum <5), or indeterminate (sum=5, a tie). Scoring was repeated twice for all samples, once for each of the two duplicate spectra from that sample. If discordant results were obtained from duplicate spectra, that specimen was categorized as indeterminate.

After making a set of models, the final model that would be used to score the blinded validation set was selected by assessing the performance characteristics of each cross-validation group of models; this assessment was done by calculating the means and standard errors of three parameters of discrimination (accuracy, sensitivity and specificity) across the ten models in each cross-validation group. The cross-validation group model having the best discrimination in the model development set (model 10CVF9M75G100) consisted of 9 m/z features; this was the model used to classify the blinded samples.

## RESULTS

### Results of model-building: the training set

**OneR results**—The best model generated by oneR identified 3 cutoff values for a single feature that yielded a mean validation accuracy of  $61.2 \pm 7.4\%$ ; sensitivity of  $66.1 \pm 19.1\%$ ; and specificity of  $54.2 \pm 20.5\%$ ; a non-significant classification over random assignment.

**kNN results**—The best model by kNN generated a 9-NN model with a mean validation accuracy of  $57.3 \pm 17.0\%$ ; sensitivity of  $66.1 \pm 21.4\%$ ; and specificity of  $46.3 \pm 29.9\%$ ; a non-significant classification over random assignment.

**PQ Results**—The PQ model selected in the development phase had an accuracy of  $63.0 \pm 11.0\%$ , sensitivity of  $72.0 \pm 16.7\%$ , and specificity  $52.0 \pm 22.6\%$ . While these results appear better than those of either the oneR or kNN methods superficially, the large standard deviations demonstrate there is no significant classification power in this model.

### Results of model-testing: the validation set

The PQ model was selected to predict the adenoma status for the 70 blinded validation samples. Since duplicate spectra were acquired for each sample, there were four possible outcomes: (1) concordant positive; (2) concordant negative; (3) discordant; or (4) indeterminate as a result

of one or both acquisitions for a given sample failing due to spray blockage or technical problems.

Only 37 of 70 samples had concordant positive or concordant negative results. These concordant results are shown in Table 1. Predicted status was compared to the known adenoma status from colonoscopy to assess discriminatory ability of the model. Model accuracy was 51% and was not sufficiently different from 50% to reject the null hypothesis of no discrimination.

### **Results of model-testing: post-hoc analysis of the validation set**

One concern raised at Correlogic during the preparation of samples was the discoloration of a number of sera that indicated significant hemolysis had occurred during collection in the clinical laboratory. To address this concern, which might be reflected as ‘noise’ in the spectra, one post-hoc analysis was done after the primary pre-specified analysis (shown in Table 1) had been completed. To conduct the post-hoc analysis, an independent technician identified hemolyzed specimens in the validation set (while still blinded regarding status as adenoma or normal); these were then removed from the validation set in the post-hoc analysis. The samples remaining in the now-smaller “validation set” remained totally blinded to the Correlogic investigators during this post-hoc analysis. The remaining samples were then classified by UNC-CH investigators. In this post-hoc analysis of unhemolyzed specimens only, a modest but statistically significant degree of discrimination is shown (Table 2). Model accuracy was 63%. The 95% confidence interval excludes the null value (50%), permitting rejection of the null hypothesis of no discrimination.

## **DISCUSSION**

This study utilized a rigorously collected set of specimens to assess whether molecular signals in serum can be used to discriminate between persons with adenomas and those with normal colons. This study did not find, in the primary analysis, that serum proteomics could discriminate between persons with large adenomas and normal persons; however, the single post-hoc analysis (done to reduce possible noise from hemolyzed specimens) did suggest discrimination. This finding of discrimination, if demonstrated in other high-quality specimens, could be clinically important.

Several reasons may explain failure to find ‘discrimination.’ First, of course, may be that, biologically, there is no ‘signal’ in serum that distinguishes people with or without an adenoma.

Second, the very small number of subjects limited the ability of any approach to find signal, even if it was there. In other words, the study was so small that, in the cross-validation strategy used to minimize overfitting, a relatively low level of classification was found in the training set. In this setting, it simply was not expected that much discrimination would be ‘confirmed’ in the validation set. This expectation was borne out in the blinded validation in the primary analysis, although discrimination was suggested in the one post-hoc analysis (See Table 2).

Third, it is possible that ‘noise’ in the samples overwhelmed any signal that may have been present. While the sample set displays no apparent systematic biases - that could lead to discrimination due to non-adenoma causes - a weakness of the study was that, at multiple points, specimens may have been handled in ways that were suboptimal for preserving ‘signal’ in serum and so may have caused ‘noise.’ The presence of hemolyzed samples indicates that this is a clear possibility. This situation might be understandable in the sense that, in the original studies, preservation of uncharacterized proteomic signals in serum was not a priority of the study. Possible important sources of noise in this study include how samples were handled (the time window from collection to spinning and was sometimes long, although the time from

spinning to freezing was consistent - about 20 minutes with little variation). Another concern was instrument performance, in that in the blinded sample set only 37 of 70 samples produced a reproducible spray. Of the 33 other samples, 10 had discordant sprays, but 23 failed to spray appropriately. This problem probably occurred because of the use of a different spray nozzle for each sample when using the Advion nano-ESI chip; more recent experience has shown much better consistency when using a more traditional ESI source using the same nozzle for each spray.

The negative results of this study should not be considered ‘conclusively’ negative, for the reasons discussed above. Further, the single post-hoc analysis done suggests that there might be signal in serum that distinguishes subjects with large adenomas from those without. The results of this study also highlight the importance of understanding sources of ‘noise’ in a proteomics study, in addition to understanding and addressing sources of ‘bias.’

## Acknowledgements

### *Financial support:*

We wish to acknowledge support from the following grants:

- National Cancer Institute: 2-R01-CA044684, The Epidemiology of Rectal Mucosal Proliferation;
- National Institute of Diabetes and Digestive and Kidney Diseases: 5P30DK034987, The Center for Gastrointestinal Biology and Disease;
- a Population Sciences Research Award from the UNC Lineberger Comprehensive Cancer Center.

## REFERENCES

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2006. *CA Cancer J Clin* 2006;56:106–30. [PubMed: 16514137]
2. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty RL. Natural history of untreated colonic polyps. *Gastroenterology* 1987;93:1009–13. [PubMed: 3653628]
3. Campo E, Palacin A, Benasco C, Condom E, Cardesa A. Ferritin immunohistochemical localization in normal and neoplastic colonic mucosa. *Int J Biol Markers* 1987;2:177–83. [PubMed: 2453593]
4. Chung YC, Chang YF. Serum interleukin-6 levels reflect the disease status of colorectal cancer. *J Surg Oncol* 2003;83:222–6. [PubMed: 12884234]
5. Flamini E, Mercatali L, Nanni O, et al. Free DNA and carcinoembryonic antigen serum levels: an important combination for diagnosis of colorectal cancer. *Clin Cancer Res* 2006;12:6985–8. [PubMed: 17145818]
6. Lu H, Goodell V, Disis ML. Targeting serum antibody for cancer diagnosis: a focus on colorectal cancer. *Expert Opin Ther Targets* 2007;11:235–44. [PubMed: 17227237]
7. Wilson S, Wakelam MJ, Hobbs RF, et al. Evaluation of the accuracy of serum MMP-9 as a test for colorectal cancer in a primary care population. *BMC Cancer* 2006;6:258. [PubMed: 17076885]
8. Chen YD, Zheng S, Yu JK, Hu X. Artificial neural networks analysis of surface-enhanced laser desorption/ionization mass spectra of serum protein pattern distinguishes colorectal cancer from healthy population. *Clin Cancer Res* 2004;10:8380–5. [PubMed: 15623616]
9. Engwegen JY, Helgason HH, Cats A, et al. Identification of serum proteins discriminating colorectal cancer patients and healthy controls using surface-enhanced laser desorption ionisation-time of flight mass spectrometry. *World J Gastroenterol* 2006;12:1536–44. [PubMed: 16570345]
10. Leung WK, To KF, Man EP, et al. Quantitative detection of promoter hypermethylation in multiple genes in the serum of patients with colorectal cancer. *Am J Gastroenterol* 2005;100:2274–9. [PubMed: 16181380]
11. Wallner M, Herbst A, Behrens A, et al. Methylation of serum DNA is an independent prognostic marker in colorectal cancer. *Clin Cancer Res* 2006;12:7347–52. [PubMed: 17189406]



12. Yu JK, Chen YD, Zheng S. An integrated approach to the detection of colorectal cancer utilizing proteomics and bioinformatics. *World J Gastroenterol* 2004;10:3127–31. [PubMed: 15457557]
13. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* 2004;4:309–314. [PubMed: 15057290]
14. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. *Nat Rev Cancer* 2005;5:142–149. [PubMed: 15685197]
15. Martin C, Connelly A, Keku TO, et al. Nonsteroidal anti-inflammatory drugs, apoptosis, and colorectal adenomas. *Gastroenterology* 2002;123:1770–7. [PubMed: 12454832]
16. Lackey L, Keku TO, Simmons J, et al. Local IGFBP-3 expression is associated with increased apoptosis and reduced colorectal adenoma risk. *Gastroenterology* 2005;128(Supplement 2):W1585.
17. Aristoteli LP, Molloy MP, Baker MS. Evaluation of endogenous plasma peptide extraction methods for mass spectrometric biomarker discovery. *J Proteome Res* 2007;6:571–81. [PubMed: 17269714]
18. Villanueva J, Philip J, Chaparro CA, et al. Correcting common errors in identifying cancer-specific serum peptide signatures. *J Proteome Res* 2005;4:1060–72. [PubMed: 16083255]
19. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002;359:572–7. [PubMed: 11867112]
20. Krebs MD, Mansfield B, Yip P, et al. Novel technology for rapid species-specific detection of *Bacillus* spores. *Biomolecular Engineering* 2006;23:119–27. [PubMed: 16542873]

**Table 1**

**Primary analysis: Classification of blinded validation set of sera from asymptomatic colonoscopy-screened persons, based on analysis of spectra generated by a nano-electrospray ionization source using a QSTAR-XL mass spectrometer and the ProteomeQuest® algorithm**

		Colonoscopy Status		Totals
		Adenoma	No Adenoma	
Model Prediction	Positive	10	16	26
	Negative	2	9	11
	Totals	12	25	37

Sensitivity:  $10/12 = 83\%$

Specificity:  $9/25 = 36\%$

Accuracy: 51% (95% CI: 38% to 65%)

**Table 2**  
**Post-hoc analysis: classification of blinded validation set after hemolyzed specimens were excluded**

		Colonoscopy Status		Totals
		Adenoma	No Adenoma	
Model Prediction	Positive	7	7	14
	Negative	2	8	10
	Totals	9	15	24

Sensitivity:  $7/9 = 78\%$  (95% CI: 40% to 96%)

Specificity:  $8/15 = 53\%$  (95% CI: 27% to 78%)

Accuracy: 63% (95% CI: 53% to 72%)