

# Quantitative Multiplex Protein Assay for Serum Samples to Rule Out Colon Cancer

Randall Brock<sup>1</sup>, Ruth VanBogelen<sup>2</sup>, Bob Xoing<sup>3</sup>, Lily Li<sup>3</sup>, Dennis King<sup>4</sup>, Lori Hoffman<sup>4</sup>

<sup>1</sup>Onconome, Inc., Redmond WA, <sup>2</sup>Biomarker Science Group, Manchester, MI, <sup>3</sup>Tandem Labs, Woburn, MA, <sup>4</sup>STATKING Consulting, Inc., Fairfield, OH

## Background

Colon cancer is highly curable when diagnosed at an early stage. However, compliance with screening guidelines is low for a variety of reasons. A serum based test that demonstrates high sensitivity and specificity and is convenient for both patients and physicians could increase screening compliance. We are focusing on such a test now that addresses these issues.

We identified numerous serum proteins that are differentially expressed in serum from individuals with colon cancer versus individuals that are known to be colon cancer free as determined by colonoscopy. By measuring the levels of these proteins present in a serum sample and comparing those levels found in colon cancer patients and normal control samples, a determination can be made as to whether colon cancer may be present or is absent in the test sample.

## Materials and Methods

### Sample selection

Serum samples were obtained from a commercial biobank using consistent methods for collection, preparation and storage of samples. The donors of all 200 samples had had a colonoscopy the day the serum sample was collected. The breakdown of samples was:

- 91 males; 109 females
- Age range from 36 – 80 with 78% over age 50
- 100 negative for colon cancer, 100 with colon cancer (Stage I= 13, Stage II= 39, Stage III= 40, Stage IV= 8)

### Multiplexed Peptide MRM Assay

**Sample preparation** – Serum samples were diluted 1:25 into Millipore’s Direct-Q water, denatured, reduced, alkylated and digested with trypsin, and after centrifugation the supernatant was dried down. Samples were reconstituted in 50 µL of 95/5/0.1 water/ACN/formic acid before analysis. An IS (internal standard peptide, unrelated to the study proteins) was added to each sample.

**Sample testing** – The multiplexed Peptide MRM assay measured the area under the curve for each transition. The analysis was performed using a Shimadzu LC-10ADvp HPLC system and an API5000™ LC/MS/MS System in MRM mode. All samples were analyzed in analytical duplicate. A total of six analytical batches were run. Each batch took 15 hours of instrument time.

**Calibration curves** – Eleven standard solutions were made using diluted, digested human serum as the proxy matrix. Each standard solution included the heavy isotope labeled derivative of each of the peptides and a single IS. The concentration range was 1 – 2500 ng/mL. For each batch of samples, the standard solutions were analyzed twice (front and back).

**Quality control** – Three QC samples were prepared by adding the heavy isotope derivatives of the peptides and the IS to diluted, digested human serum. The concentrations of the heavy peptides were 30, 900 and 1800 ng/mL in the different QC samples and each QC sample was run in duplicate in each batch.

**Calculation of peptide concentration in samples** – The equation generated with the calibration curves was used to convert the area ratio (endogenous light peptide/IS) of each peptide in each sample to ng/mL. This number was multiplied by 25 to correct for the dilution factor. The average of the analytical replicates was calculated and used as the concentration value for the peptide in each sample.

### Statistical Analysis

A wide variety of statistical models were fit to this data using discriminant analysis and logistic regression techniques. Discriminant analysis finds the best separation between the different sample classifications based on the peptide concentration values, under the assumption that the peptide concentrations are from a multivariate normal distribution. The result of this method is a discriminant function, which is a rule that uses the peptide concentrations to determine how to classify each sample. The logistic regression function also gives a rule for classifying observations based on the peptide data, without the added assumption that the peptide data are multivariate normally distributed. Table 1 illustrates the wide variety of statistical models that were applied to this data.

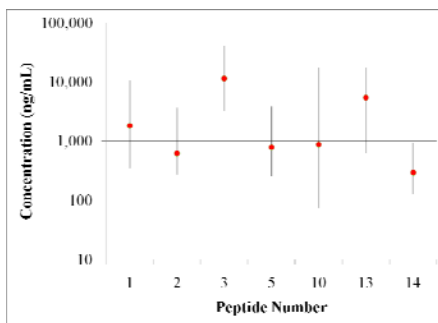
Analysis Method	Data Comparisons	Peptides
Discriminant Analysis	Cancer vs. Normal	01, 03, 05, 13, 14
Discriminant Analysis	Cancer vs. Normal	01, 03, 10
Discriminant Analysis	Cancer vs. Normal	05, 13
Logistic Regression	Cancer vs. Normal	03, 13, 14
Logistic Regression	Cancer vs. Normal	02, 10, 13

## Assay Results

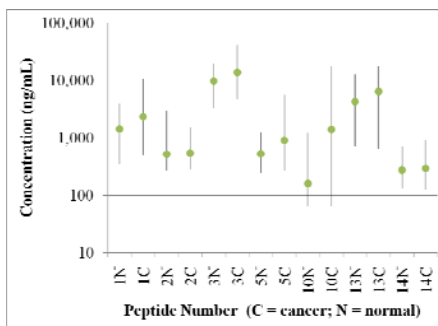
The seven proteins included in this multiplex assay are listed in the table below. One peptide for each of these proteins is the actual analyte being measured in the assay.

Peptide Number	Protein Name/Gene Name	Accession no. Swiss-prot Gene ID
1	a-1-acid glycoprotein 1/ ORM1	P02673 5004
2	Gelsolin/ GSN	P06396 2934
3	C9 Complement Component/ C9	P02748 735
5	Plasma serine protease inhibitor/ SERPINA3	P05154 5104
10	Hyaluronan-binding protein 2/ HABP2	Q14520 3026
13	Serum amyloid A protein/ SAA2	P02735 6288,6289
14	LOC653879 similar to complement C3 precursor/ LOC653879	P01024 653879

The endogenous concentration of seven peptides was determined in 200 samples. The figure below provides a summary of the results. The concentration range for each peptide in the 200 samples is shown by the line and the average concentration is shown by the solid red circle.



Individually, these peptides have indistinct concentration ranges for normal versus cancer groups. The figure below shows the concentration range and the average concentration for the normal and cancer group for each of the seven peptides.



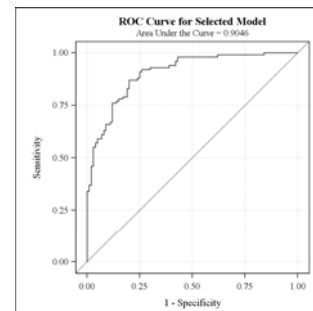
The P-values for 5 of the 7 peptides indicate statistically significant differences in the data for the 100 normal versus 100 cancer samples. These results suggested that a model could be built with this data to predict whether a sample falls into the normal or cancer group.

Peptide Number	P-Value
1	4.87 E-07
2	6.81 E-01
3	7.08 E-09
5	4.60 E-05
10	3.18 E-04
13	2.78 E-09
14	3.31 E-01

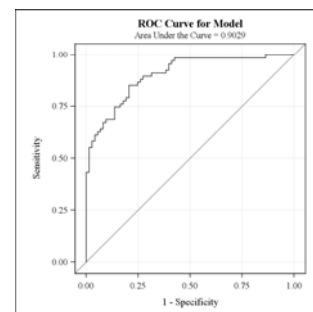
## Statistical Analysis Results

Results of the models indicate that there is information in the peptide concentration data that is useful for classifying samples as either cancer or normal. To assess the model fit, 70% of the data were randomly selected as the training data on which the model was built, and the remaining 30% of the data were set aside as test data.

Forward Selection Logistic Regression analysis with all 7 peptides yielded the ROC curve below which has an Area Under the Curve of 0.9046.



In terms of classification accuracy, the best model is that of a logistic regression fit to the cancer and normal data using peptides 2, 10, and 13. The figure below displays the ROC curve of the model constructed from the training data.



The above figure indicates that the model performs very well, as the AUC is 0.9029 (where an AUC of 1.0 is the best possible model). Each point on the ROC curve represents a possible combination of sensitivity and specificity that the model can achieve. The table below highlights the key points along this curve. The optimal cut-point is that which results in the highest combination of sensitivity and specificity.

Model Cut-Point	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
0.4407 (optimal)	85.07%	79.45%	79.17%	85.29%
0.1961	95.52%	60.27%	68.82%	93.62%
0.1853	98.51%	57.53%	68.04%	97.67%

## Conclusions

These results indicate a benefit could be derived from using a sensitive, non-invasive proteomics approach to rule out colon cancer in individuals who should be screened for colon cancer under current guidelines but choose to forgo a colonoscopy.

## Next Steps in Development

1. Perform key validation tests on the assay.
2. Analyze 315 samples (300 normal, 15 cancer) to obtain a better assessment of the NPV of this Rule-Out test using a sample set that better matches the prevalence of this disease.
3. Analyze 500 samples (300 normal, 200 cancer) to obtain a better assessment of the sensitivity of this Rule-Out test.