Scientific Rationale

Hepatocellular carcinoma (HCC) is a major worldwide health problem and a cancer with increasing incidence in the United States [1;2]. The increasing incidence of HCC in the U.S. has been associated with hepatitis C viral infection (HCV), and a further increase in HCC is predicted to occur over the next few decades [3]. In the Egyptian population, up to 90% of HCC cases were attributed to HCV infection [4]. Approximately 14% of the population in Egypt is infected with HCV and 7 million people are believed to suffer from a chronic liver disease (CLD) [5]. HCC is third in incidence among the cancer diseases in men with more than 8,000 new cases predicted by 2012 in this population [6]. Studies of HCV progression to HCC in Egypt are expected to provide new insights into the management of this increasingly significant health problem [7].

The current diagnosis of HCC relies on clinical information, liver imaging, and measurement of serum alpha-fetoprotein (AFP). The reported sensitivity (41-65%) and specificity (80-94%) of AFP are not effective for early diagnosis due to the high proportion of false negatives [8]. The identification of effective markers for the early detection of HCC is an active area of research with several new marker candidates reported within the last few years [9;10]. It has been pointed out that many currently used cancer biomarkers, including alpha-fetoprotein (AFP), are glycoproteins [11]. Fucosylated AFP was introduced as a marker of hepatocellular carcinoma with improved specificity [12;13], while other glycoproteins including GP73 are currently under evaluation as markers of HCC [14;15]. The analysis of protein glycosylation seems particularly relevant to liver pathology because of the major influence of this organ on the homeostasis of blood glycoproteins [16;17].

An alternative strategy to the analysis of glycoproteins is the analysis of protein-associated glycans [18;19]. Results from our laboratory [24-27] and other groups show that the characterization of glycans in serum of patients with liver disease is a promising strategy for biomarker discovery [20]. We used matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometric (MS) quantification of permethylated N-glycan structures, enzymatically released from serum glycoproteins, to identify novel biomarker candidtates [21;26]. We were able to identify three N-glycans that detect HCC in the background of chronic liver disease with 90% prediction accuracy. In this study, we propose to validate the biomarker candidates in serum samples from the HALT C trial.

Methods and Preliminary Data

Study Population and Sample Collection: We compared N-glycans in serum of HCC cases (n=73) with two groups of controls, controls without manifest liver disease (n=77) and controls with CLD (n=52). HCC cases and matched age- and gender- controls free of liver disease were enrolled in collaboration with the Cairo University School of Medicine, Egypt as described previously [4]. Briefly, adults who were seen at the cancer institute with newly diagnosed HCC, aged 17 and older, without a previous history of cancer, were eligible for the study. Diagnosis of HCC was confirmed by pathology, cytology, imaging (CT, ultrasound), and serum AFP levels. Controls were recruited from the orthopedic department of the school of medicine [4]. The characteristics of this population are summarized in **Table 1** which shows, as expected, increased prevalence of markers of viral infections (HCV RNA, anti HCV, and anti-HBV) in cancer cases compared to controls (see also Supplemental Table 2) [4]. This comparison group allowed us to study the changes in glycans associated with HCV and HBV viral infections. The controls with CLD (n=52), fibrosis (n=22) and cirrhosis

Table 1	Cases (n=73)		Controls (n=77)	
Mean Age (SD)	54.2	9.3	52.3	11.9
Male Gender	53	72%	52	67%
Smokers	40	55%	41	53%
HCV RNA+	49	67%	13	17%
anti HCV+	64	87%	25	32%
anti HBV+	56	77%	42	54%
HBsAg+	4	5%	2	3%

(n=25), were recruited from Ain Shams University Specialized Hospital and Tropical Medicine Research Institute, Cairo, Egypt, during the same period. The diagnosis of liver disease in this group of controls was confirmed by ultrasound-guided liver biopsy; 5 remaining CLD controls did not have sufficient clinical information. The CLD controls were all positive for HCV RNA. This comparison group served to evaluate the ability of selected glycans to identify HCC in the background of chronic liver disease. All participants signed informed consent, provided a blood sample, and answered a questionnaire with demographic information, personal habits, medical history, and occupational history. The study protocol was approved by the institutional review committees of all participating institutions and conformed to the ethical guidelines of the 1975 Helsinki Declaration. We obtained cancer stage information on 51 cases, with 18 cases classified as early (Stage I and II) and 33 cases as advanced (Stage III and IV) according to the AJCC staging system [22]; for the remaining cases, the available information was not sufficient to assign the stage.

Mass Spectrometric Quantification of N-glycans. Serum samples were processed as described previously [18;21]. Briefly, samples were reduced, alkylated, and the reaction mixture was diluted with 100 µl of ammonium bicarbonate pH 7.7 for an enzymatic release of N-glycans using 5 mU PNGase F overnight at 37 °C. N-glycans were cleaned up by solid phase extraction on C18 Sep-Pak® cartridges (Waters, Milford, MA) and the combined eluents containing N-glycans were then passed over activated charcoal micro columns (Harvard Apparatus, Holliston, MA). The purified N-glycans were eluted with 1 ml of 50% aqueous ACN with 0.1% TFA and dried prior to the solid-phase permethytion. Sodium hydroxide powder was suspended in ACN and packed into Peek tubes (1 mm i.d.; Polymicro Technologies, Phoenix, AZ) Purified N-glycans were resuspended in a 50-µl aliquot of DMSO and 22 µl methyl iodide and passed through a 100 µl syringe from Hamilton (Reno, NV) using a syringe pump from KD Scientific, Inc. (Holliston, MA) at 2 µl/min. This permethylation procedure has been shown to minimize oxidative degradation and peeling reactions and to eliminate excessive clean-up. All eluents were combined and permethylated N-glycans were extracted using 200 µl chloroform and dried. Permethylated glycans were resuspended in 2 µl of (50:50) methanol:water solution. A 0.5-µl aliquot of the sample was spotted on a MALDI plate and mixed with an equal volume of DHB-matrix. The MALDI plate was dried under vacuum to ensure uniform crystallization. Mass spectra were acquired using an Applied Biosystems 4800 MALDI TOF/TOF Analyzer (Applied Biosystems Inc., Framingham, MA) equipped with a Nd:YAG 355-nm laser. MALDI-spectra were recorded in the positive-ion mode, since permethylation eliminates the negative charge normally associated with sialylated glycans [23].

Data Processing and Analysis: Analyses were carried out in MATLAB (MathWorks, Natick, MA) and SAS (SAS Inc., Cary, NC) software packages. Raw MALDI-TOF mass spectra were processed as described previously [24;26]; the data are available at http://microarray.georgetown.edu/web/files/glycans.zip . Briefly, baseline-corrected spectra were normalized and N-glycan peak identification was carried out on a randomly selected training set of 75 samples (25 HCC, 25 disease free controls, 25 controls with CLD) of the total of 202 spectra analyzed. After scaling the peak intensities to an overall maximum intensity of 100, local maximum peaks above a specified threshold were identified and nearby peaks within 300 ppm mass were coalesced into a single window to account for drift in m/z location. The threshold intensity for peak identification was set so that isotopic clusters were represented by a single peak. This procedure identified 85 peak-containing windows; the maximum intensity in each window was used as the variable of interest. For final analysis, we used 64 of 83 peaks (glycan intensities) associated with known N-glycan structure. Logistic regression models were used to determine association of the glycans and covariates including age, gender, and viral infections (independent variables) with HCC status (dependent variable).

For determination of prediction accuracy and construction of ROC curves, the training dataset (75 spectra) was randomly selected. The remaining 127 spectra (48 HCC, 52 disease free controls, 27 controls with CLD) served as a blinded validation set. Spectra in the blinded validation set were processed using the same criteria described above for the training samples. An algorithm that integrates ant colony optimization with support vector machine (ACO-SVM) was used to select three N-glycan peaks for the classification of HCC as described previously . Individual N-glycans and a combined SVM classifier were used to classify the blinded spectra. Sensitivity and specificity of the marker-candidates were evaluated on the blinded validation dataset including the set of controls with CLD.

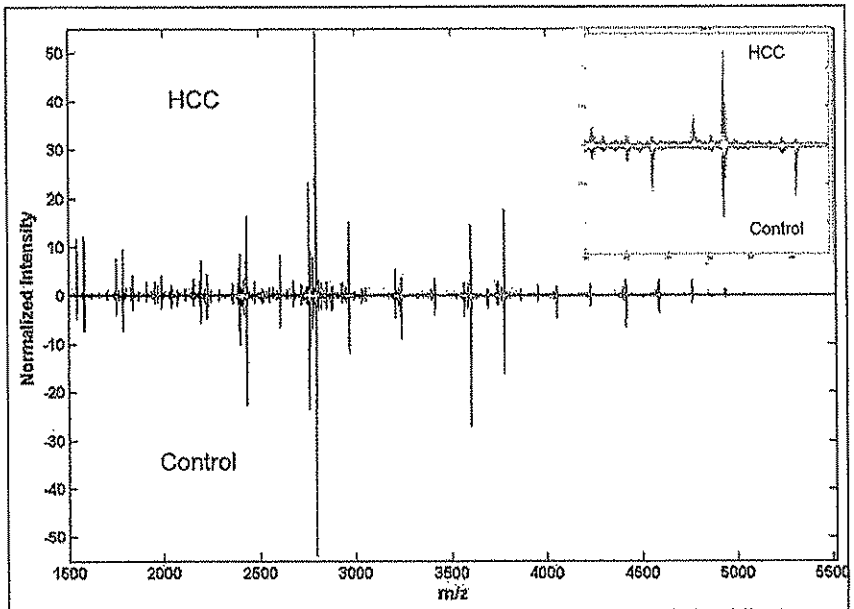

Figure 1. Average MALDI-TOF spectra in the mass range of 1.5-5.5 kDa in HCC cases (n=73) and population controls (n=77, inverted).
Inset: example of individual reflector spectra of a case (red) and a control (blue, inverted) zoomed into the mass region 2,800-3,100 Da

Comparison of average spectra in HCC cases (n=73) and controls (n=77) showed marked differences in glycan abundance **(Figure 1)**. About half of the glycans were significantly different by t-test at $p<0.01$. Three of the N-glycans were selected with greater than 50% frequency in 100 repeats of the ACO-SVM algorithm; they are evaluated below as candidate markers. Association of the glycans and covariates (age, gender, HCV and HBV viral infections, and smoking) with HCC (dependent variable) was analyzed by logistic regression. Glycan intensities were dichotomized by the median value in population controls; the analysis of glycans as continuous variables did not substantially affect the outcome (data not shown). The three selected glycans remain associated with HCC after adjustment for all covariates including viral infections; this is true for the comparison with disease free controls (n=77; data not shown) and for the comparison of HCC cases (n=73) and controls with CLD (n=52) **(Table 2)**.

Table 2.	Mass (Da)	Structure	OR	95% CI		p value
Glycan 1	3241.9		0.03	0.01	0.24	**0.0007**
Glycan 2	2472.9		6.07	1.51	24.34	**0.0110**
Glycan 3	4052.2		0.09	0.01	0.6	**0.0135**

Association of glycans with HCC; multivariate logistic regression controlled for gender, age and HCV infection comparing 73 HCC cases and 52 CLD controls.

Figure 2 shows ROC curves of the three individual glycans that are different compared to CLD controls, and their combination, in a blinded, independent validation set of HCC cases (n=48) ad CLD controls (n=27). AuROC for individual glycans ranged from 89-93%, while the combined classifier has a sensitivity of 90% and specificity of 89% in the blinded independent validation set. Glycan 1 and 6 are tri- and tetra-antennary complex glycan that decrese in HCC. Glycan 5 is a bisecting glycan that increase

in HCC patients. This is consistent with the general trends of changes observed in our study. To further evaluate the potential of the three glycans for early detection of HCC, we analyzed the progressive glycan changes from fibrosis to cirrhosis and early cancer, stage 1 and 2 disease **(Figure 3)**. It is interesting to note that 17 of the 18 early cancers and 23 out of 25 CLD controls were classified correctly by the three glycans. Both prediction accuracy results and **Figure 3** suggest that the three selected glycans separate efficiently cirrhosis controls from early stage HCC; we verified by regression analysis that all three glycans are significantly different at $p < 0.05$ between cirrhosis controls (n=25) and early stage HCC (n=18). Determination of AFP in our study showed that 30% of the cases had AFP <200 ng/ml; of these 22 HCC cases 18 were correctly classified by the three glycans. Our results strongly suggests that these N-glycans should be further evaluated as markers for early detection of HCC.

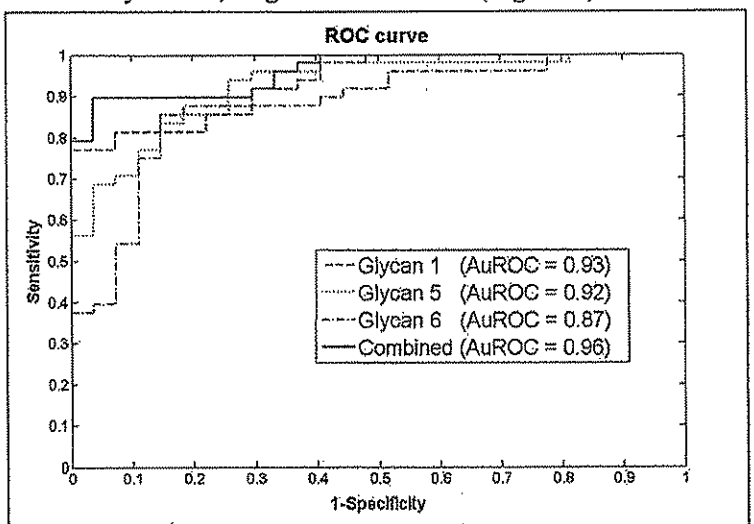

Figure 2. Area under the ROC curve for three glycans comparing a blinded validation set of HCC cases (n=47) and CLD controls (n=27).

Patient sample size: We propose to analyze 50 HCC cases and 100 controls matched on fibrosis and treatment from the HALT C study. In addition to the 150 samples from the HALT C trial, we will analyze in parallel 25 HCC cases and 25 CLD controls from the Egyptian population. This will allow us to demonstrate long term stability of our analytical measurements and define appropriate cutoff values for the determination of prediction accuracy in the HALTC samples. The order of the analysis of the total 200 samples will be randomized. Matching of cases and controls on age, gender, and viral infection status could be considered but can be adjusted at analysis. Previously, we have used frequency matching to avoid an imbalanced study design. The effect of age, gender, and viral infections on N-glycosylation of serum proteins is not well studied; in our analyses, we were able to adjust for effects other than HCC in the regression models.

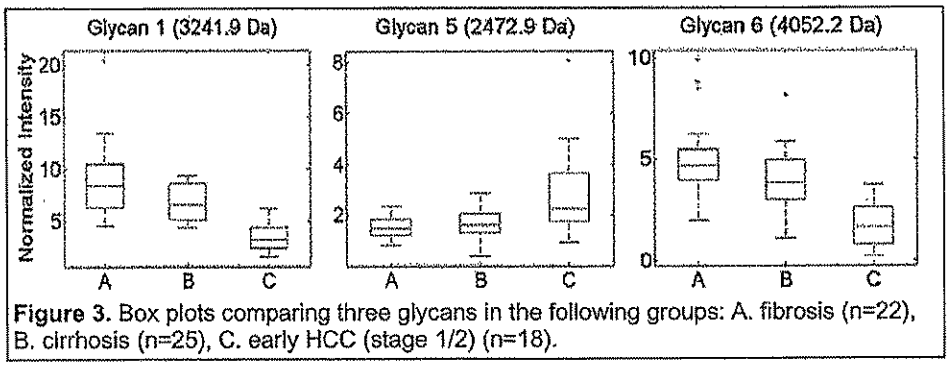

Figure 3. Box plots comparing three glycans in the following groups: A. fibrosis (n=22), B. cirrhosis (n=25), C. early HCC (stage 1/2) (n=18).

Suggested data analysis: D5.5 Sample size considerations: The sample size should be as large as possible to assess better the variability of these new markers in the population. The required sample size can be approximated on the basis of comparisons of a single glycan for two different classes and the ability to control the control false-discovery rate. From the preliminary data set, we have found that the standard deviation of the $\sqrt[3]{}$(intensity)-transformation for MALDI-TOF is around 0.465. Therefore, we deduce that if there are 20 patients in each group, we have 99% power at $\alpha = 0.0001$ level to detect a 2-fold difference in glycan concentration. Additional estimation can be based on the power to detect associations of a glycan classifier with disease-status and progression. We used the PS software

developed by Dupont {Dupont, 1990 9884 /id}, based on Schlesselman's method {Schlesselman, 1982 6076 /id} for determining statistical power in case control studies. We have estimated the power we will have to detect odds ratios of 5.0 or 10.0 with $\alpha = 0.05$ for a given glycan classifier that is expressed in at least 5% of the controls **(Table 3)**. The main comparison in sufficiently powered; power for the analysis of subgroups decreases accordingly. At this point, it is not clear whether HCC with different etiology has identical glycan changes; this could be further evaluated in the proposed study using regression analysis. The performance of the markers in etiological subgroups can be evaluated as well.

Table 3.

	Prevalence (%) of controls with the glycan							
	5		10		15		20	
Odds ratio	5	10	5	10	5	10	5	10
100 HCC vs. 100 cirrhosis	94	100	100	100	100	100	100	100
50 HCC vs. 50 cirrhosis	76	99	92	100	96	100	99	100

The major focus of the study should be the determination of sensitivity and specificity for the detection of early stage HCC in the background of chronic liver disease. We propose to evaluate the N-glycans individually and in combination. To determine sensitivity and specificity of the markers, an appropriate cutoff needs to be defined. The cutoff values will be defined at GU based on our existing data in the Egyptian population. In addition, we will combine the three N-glycans in a joint SVM classifier or a classifier based on a liner combination of the variables. In our experience, the SVM classifier works equally well or better than other methods in terms of prediction accuracy. However, the SVM is less transparent than other possible combinations of the three variables that can be used and achieve a comparable prediction accuracy. The determination of sensitivity, specificity, ROC curve analysis, and regression analysis should be sufficient to evaluate the performance of the N-glycan markers in the HALT C dataset.

Reference List

1. Montalto, G., Cervello, M., Giannitrapani, L., Dantona, F., Terranova, A., and Castagnetta, L. A. M. Epidemiology, Risk Factors, and Natural History of Hepatocellular Carcinoma. Annals of the New York Academy of Sciences 2003;963:13-20.

2. El-Serag, H. B. and Mason, A. C. Rising Incidence of Hepatocellular Carcinoma in the United States [See Comments]. N.Engl.J Med 1999;340(10):745-50.

3. Wong, J. B., McQuillan, G. M., McHutchison, J. G., and Poynard, T. Estimating Future Hepatitis C Morbidity, Mortality, and Costs in the United States. Am J Public Health 2000;90(10):1562-9.

4. Ezzat, S., Abdel-Hamid, M., Eissa, S. A., Mokhtar, N., Labib, N. A., El Ghorory, L., Mikhail, N. N., Abdel-Hamid, A., Hifnawy, T., Strickland, G. T., and Loffredo, C. A. Associations of Pesticides, HCV, HBV, and Hepatocellular Carcinoma in Egypt. Int J Hyg.Environ.Health 2005;208(5):329-39.

5. Perz, J. F. and Alter, M. J. The Coming Wave of HCV-Related Liver Disease: Dilemmas and Challenges. J.Hepatol. 2006;44(3):441-3.

6. Deuffic-Burban, S., Mohamed, M. K., Larouze, B., Carrat, F., and Valleron, A. J. Expected Increase in Hepatitis C-Related Mortality in Egypt Due to Pre-2000 Infections. J.Hepatol. 2006;44(3):455-61.

7. Di Bisceglie, A. M., Lyra, A. C., Schwartz, M., Reddy, R. K., Martin, P., Gores, G., Lok, A. S., Hussain, K. B., Gish, R., Van Thiel, D. H., Younossi, Z., Tong, M., Hassanein, T., Balart, L., Fleckenstein, J., Flamm, S., Blei, A., and Befeler, A. S. Hepatitis C-Related Hepatocellular Carcinoma in the United States: Influence of Ethnic Status. Am.J.Gastroenterol. 2003;98(9):2060-3.

8. Gupta, S., Bent, S., and Kohlwes, J. Test Characteristics of Alpha-Fetoprotein for Detecting Hepatocellular Carcinoma in Patients With Hepatitis C. A Systematic Review and Critical Analysis. Ann.Intern.Med. 7-1-2003;139(1):46-50.

9. Filmus, J. and Capurro, M. Glypican-3 and Alphafetoprotein As Diagnostic Tests for Hepatocellular Carcinoma. Mol.Diagn. 2004;8(4):207-12.

10. Chignard, N., Shang, S., Wang, H., Marrero, J., Brechot, C., Hanash, S., and Beretta, L. Cleavage of Endoplasmic Reticulum Proteins in Hepatocellular Carcinoma: Detection of Generated Fragments in Patient Sera. Gastroenterology 2006;130(7):2010-22.

11. Ludwig, J. A. and Weinstein, J. N. Biomarkers in Cancer Staging, Prognosis and Treatment Selection. Nat.Rev.Cancer 2005;5(11):845-56.

12. Taketa, K., Endo, Y., Sekiya, C., Tanikawa, K., Koji, T., Taga, H., Satomura, S., Matsuura, S., Kawai, T., and Hirai, H. A Collaborative Study for the Evaluation of Lectin-Reactive Alpha-Fetoproteins in Early Detection of Hepatocellular Carcinoma. Cancer Res. 11-15-1993;53(22):5419-23.

13. Shiraki, K., Takase, K., Tameda, Y., Hamada, M., Kosaka, Y., and Nakano, T. A Clinical Study of Lectin-Reactive Alpha-Fetoprotein As an Early Indicator of Hepatocellular Carcinoma in the Follow-Up of Cirrhotic Patients. Hepatology 1995;22(3):802-7.

14. Marrero, J. A., Romano, P. R., Nikolaeva, O., Steel, L., Mehta, A., Fimmel, C. J., Comunale, M. A., D'Amelio, A., Lok, A. S., and Block, T. M. GP73, a Resident Golgi Glycoprotein, Is a Novel Serum Marker for Hepatocellular Carcinoma. J Hepatol. 2005;43(6):1007-12.

15. Comunale, M. A., Lowman, M., Long, R. E., Krakover, J., Philip, R., Seeholzer, S., Evans, A. A., Hann, H. W., Block, T. M., and Mehta, A. S. Proteomic Analysis of Serum Associated Fucosylated Glycoproteins in the Development of Primary Hepatocellular Carcinoma. J Proteome Res 2-3-2006;5(2):308-15.

16. Turner, G. A. N-Glycosylation of Serum Proteins in Disease and Its Investigation Using Lectins. Clin.Chim.Acta 6-30-1992;208(3):149-71.

17. Lee, S. J., Evers, S., Roeder, D., Parlow, A. F., Risteli, J., Risteli, L., Lee, Y. C., Feizi, T., Langen, H., and Nussenzweig, M. C. Mannose Receptor-Mediated Regulation of Serum Glycoprotein Homeostasis. Science 3-8-2002;295(5561):1898-901.

18. Kyselova, Z., Mechref, Y., Bataineh, M. M., Dobrolecki, L. E., Hickey, R. J., Vinson, J., Sweeney, C. J., and Novotny, M. V. Alterations in the Serum Glycome Due to Metastatic Prostate Cancer. J.Proteome.Res. 4-14-2007.

19. Zhao, J., Qiu, W., Simeone, D. M., and Lubman, D. M. N-Linked Glycosylation Profiling of Pancreatic Cancer Serum Using Capillary Liquid Phase Separation Coupled With Mass Spectrometric Analysis. J.Proteome.Res. 2007;6(3):1126-38.

20. Callewaert, N., Van Vlierberghe, H., Van Hecke, A., Laroy, W., Delanghe, J., and Contreras, R. Noninvasive Diagnosis of Liver Cirrhosis Using DNA Sequencer-Based Total Serum Protein Glycomics. Nat.Med 2004;10(4):429-34.

21. Kang, P., Mechref, Y., Klouckova, I., and Novotny, M. V. Solid-Phase Permethylation of Glycans for Mass Spectrometric Analysis. Rapid Commun.Mass Spectrom. 2005;19(23):3421-8.

22. AJCC Cancer Staging Manual, 6th Edition. American College of Surgeons . 2002. Philadelphia, Lippincott-Raven.

23. Mechref, Y. and Novotny, M. V. Mass Spectrometric Mapping and Sequencing of N-Linked Oligosaccharides Derived From Submicrogram Amounts of Glycoproteins. Anal.Chem. 1998;70(3):455-63.

24. Ressom, H. W., Varghese, R. S., Drake, S. K., Hortin, G. L., Abdel-Hamid, M., Loffredo, C. A., and Goldman, R. Peak Selection From MALDI-TOF Mass Spectra Using Ant Colony Optimization. Bioinformatics. 2007;23(5):619-26.

25. Ressom, H. W., Varghese, R.S., Goldman, L., An, Y., Loffredo, C.A., Abdel-Hamid, M, Kyselova, Z., Mechref, Y., Novotny, M., Drake, S., Goldman, R. Analysis of MALDI-TOF mass spectrometry data for detection of glycan biomarkers. Pac Symp Bioc. 2008; 216-227. Altman, R.S., Dunker, A.K., Hunter, L., Murray, T., and Klein, T.E. eds., World Scientific Publishing Company, Singapore

26. Ressom, H. W., Varghese, R.S., Goldman, L., Loffredo, C.A., Abdel-Hamid, M, Kyselova, Z., Mechref, Y., Novotny, M., Goldman, R. Analysis of MALDI-TOF mass spectrometry data for discovery of peptide and glycan biomarkers of hepatocellular carcinoma. Journal of Proteome Research. 2008;7:603-610

27. Isailovic, D., Kurulugama, R.T., Plasencia, M.D., Stokes, S.T., Kyselova, Z., Goldman, R., Mechref, Y., Novotny, M.V., and Clemmer, D.E. Profiling of Human Serum Glycans Associated with Liver Cancer and Cirrhosis by IMS-MS. Journal of Proteome Research. 2008;7:1109-1117