1

NOVEL PLASMA GLYCOPROTEIN BIOMARKERS PREDICT PROGRESSION-FREE SURVIVAL IN SURGICALLY RESECTED CLEAR CELL RENAL CELL CARCINOMA

Daniel Serie¹, Amanda A. Myers, MD², Daniela A. Haehn, MD², Alexander S. Parker, PhD³, Essa Bajalia², Giovanni Gonzalez, MD², Qiongyu Li¹, Maurice Yu Wong¹, Kaitlynn Moser¹, Bo Zhou¹, David D. Thiel, MD²

¹ InterVenn Biosciences

² Mayo Clinic Florida Department of Urology

³ University of Florida, College of Medicine - Jacksonville

Keywords: Renal cell carcinoma, glycoproteins, biomarker, progression free survival

Corresponding author: David D. Thiel, M.D. Mayo Clinic Florida 3 East Urology 4500 San Pablo Road Jacksonville, Fl 32224 <u>Thiel.David@mayo.edu</u> 904-953-7330 Fax: 904-953-2218

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosures: Alex Parker received compensation from InterVenn as a consultant in 2016 and 2017

Abstract

Introduction: Limited data exists on utilization of protein post-translational modifications as biomarkers for clear cell renal cell carcinoma (ccRCC). We employed high-throughput glycoproteomics to evaluate differential expression of glycoprotein-isoforms as novel markers for ccRCC progression-free survival (PFS).

Methods: Plasma samples were obtained from 77 patients treated surgically for ccRCC. Glycoproteomic analyses were carried out after liquid chromatography tandem mass spectrometry. Age-adjusted Cox proportional hazard models were constructed to evaluate PFS. Optimized Harrells c-index was employed to dichotomize the collective for the construction of Kaplan-Meier curves.

Results: The average length of follow-up was 3.4 (range: 0.04-9.83) years. Glycoproteomic analysis identified 39 glycopeptides and 14 non-glycosylated peptides that showed statistically significant (false discovery rate $p \le 0.05$) differential expression associated with PFS. Five of the glycosylated peptides conferred continuous hazard ratio of > 6 (range 6.3-11.6). These included prothrombin A2G2S glycan motif (HR=6.47, P=9.53E-05), immunoglobulin J chain FA2G2S2 motif (HR=10.69, P=0.001), clusterin A2G2 motif (HR=7.38, P=0.002), complement component

2

C8A A2G2S2 motif (HR=11.59, P=0.002), and apolipoprotein M glycopeptide with non-fucosylated and non-sialylated hybrid-type glycan (HR=6.30, P=0.003). Kaplan-Meier curves based on dichotomous expression of these five glycopeptides resulted in hazard ratios of 3.9-10.7, all with p-value < 0.03. Kaplan-Meyer plot using the multivariable model comprising 3 of the markers yielded HR of 11.96 (p <0.0001).

Conclusion: Differential glyco-isoform abundance of plasma proteins may be a useful source of biomarkers for the clinical course and prognosis of ccRCC.

Introduction:

The standard of care for clinically localized clear cell renal cell carcinoma (ccRCC) is partial or radical nephrectomy. Distant metastatic disease develops in 20-30% of patients with initially localized disease, despite surgical resection of the primary tumor[1]. Current surveillance strategies for ccRCC recurrence rely primarily on stage-based risk assessment to determine intervals of monitoring between 3 and 12 months. Methods to precisely risk stratify patients represents a clear unmet need in patients with ccRCC. Non-invasive risk stratification tests have potential to guide clinical decision making on follow up. Moreover, in the era of targeted molecular therapeutics, there is an urgent need for improved prognostication in ccRCC patients who would potentially benefit from adjuvant therapy. While several potential biomarkers have been explored in this context, none have demonstrated superiority to pathologic stage [2].

Protein glycosylation is the most common form of post-translational protein modification (PTM) and has profound effects on protein structure, conformation, and function. Changes in PTM are a well-recognized hallmark of malignant transformation, observed in a variety of cancers, and have been associated with severity and progression of disease [3]. While the majority of clinically utilized protein biomarkers (PSA, CEA, or CA125) are glycosylated [4], currently used immuno-tests for these markers only recognize their peptide backbone but do not consider their glycoprotein-isoforms. Early studies evaluating glycoproteins showed promise in the development of biomarkers for various cancers [5, 6], the technical challenges of accurately measuring individual glyco-isoforms have so far precluded clinical use. Recent development of a highly accurate and scalable platform has now made it possible to carry out rapid and accurate analysis of site-specific changes in glycan motifs of multiple serum glycopeptides[7]. We deployed this platform to evaluate whether differential expression of glycosylated peptides may provide insights into the variable natural history and disease course of ccRCC.

Methods:

2.1 Sample Procurement. Plasma samples obtained from patients with newly diagnosed ccRCC prior to their surgical treatment, and stored at -80°C, were provided by the Mayo Clinic Renal Registry, an IRB-approved biorepository and clinical database on ccRCC patients. Histopathologic diagnosis of ccRCC was confirmed for all patients by a genitourinary pathologist who served as an independent reviewer, as previously described [8].

2.2 Sample Preparation. Serum samples were reduced with dithiothreitol (DTT), alkylated with iodoacetamide (IAA)then digested with trypsin at 37°C for 18 hours. To quench

3

the digestion, formic acid was added to each sample after incubation to a final concentration of 1% (v/v).

2.3 LC-MS Analysis. Digested serum samples were separated over an Agilent ZORBAX Eclipse Plus C18 column (2.1 mm x 150 mm i.d., 1.8 µm particle size) using an Agilent 1290 Infinity UHPLC system. The mobile phase A consisted of 3% acetonitrile, 0.1% formic acid in water (v/v), and the mobile phase B of 90% acetonitrile 0.1% formic acid in water (v/v), with the flow rate set at 0.5 mL/minute. The binary solvent composition was set at 100% mobile phase A at the beginning of the run, linearly shifting to 20% B at 20 minutes, 30% B at 40 minutes, and 44% B at 47 minutes. The column was flushed with 100% B and equilibrated with 100% A for a total run time of 70 minutes. After electrospray ionization (ESI), samples were injected into Agilent 6495B triple quadrupole mass spectrometer operated in dynamic multiple reaction monitoring (dMRM) mode. The MRM transitions comprised of 408 glycopeptides. Then normalized by comparing them with the abundance of 68 non-glycosylated peptide fragments from which the glycopeptides monitored were derived, as previously published [7].

2.4 Data Analyses. PeakBoundaryNet, an in-house-developed spectrogram feature recognition and integration software based on recurrent neural networks [9] was used to integrate chromatogram peaks and obtain molecular abundance quantification for each peptide and glycopeptide. To account for differential protein abundance, the ratio of glycosylated peptides to non-glycosylated peptides ("relative abundance") was utilized. To correct for assay-related artifacts, pooled human serum sample was interspersed every ten samples. The glycosylated/non-glycosylated peptide ratios in these calibration samples were used to normalize the patient sample ratios.

The association of relative abundance ratio values with progression-free survival was assessed using age-adjusted Cox Proportional Hazards models. P-values for multiple comparisons were determined using the Benjamini-Hochberg false discovery rate (FDR) [10] procedure. P-value of 0.05 or less was defined as statistically significant. Biomarker expression values were dichotomized at cutoff values that maximized Harrell's C-index. Kaplan-Meier plots were generated to assess the differential progression-free survival between groups with high and low marker expression. Multivariable predictive models were generated from glycopeptides with HR > 6 and FDR < 0.05, by trimming features via backwards selection, then generating models applied to each sample as a holdout via leave-one-out cross-validation (LOOCV). All statistical analyses were run in R, version 3.4.4.

2.5 Ingenuity Pathway Analysis. Bioinformatic analysis was performed to identify canonical pathways and associated protein networks by using Ingenuity® Pathway Analysis software (QIAGEN Inc.). Right-tailed Fisher's exact test was used to calculate a P value of overlap to determine the significance of each canonical pathway, and P <0.01 was considered significant. Ingenuity® Knowledge Base was used to predict the upstream regulators of candidate biomarkers. The molecular filter was applied to analysis transcription factors, cytokines, transmembrane and G-coupled protein receptors. Associated networks with candidate biomarkers were algorithmically generated.

Results:

4

We analyzed plasma samples of 77 ccRCC patients, 48 (62%) were male and 29 were female (38%). Mean age was 61 years (range: 33-79 years). Fifty-four patients (70%) had stage I disease, 9 (11.7%) stage II, 12 (15.6%) stage III, 1 (1.3%) had stage IV disease, and 1 was missing stage. The average length of follow-up was 3.4 (range: 0.04 - 9.83) years. No patients with T1 or T2 disease had positive margin. Four patients with pT3 disease had positive renal vein margins, of these two had disease progression. In total, thirteen patients suffered recurrent disease (Table 1). The sites of progression were 7 (54%) lung, 2 (15%) bone and 4 (31%) local.

Variable	Summary (N=77)
Age at diagnosis, years	mean 61, range 33-79
Sex, n (%)	
Female	29 (38%)
Male	48 (62%)
Surgery Performed	
Partial Nephrectomy	30 (39%)
Laparoscopic	19
Open	11
Radical Nephrectomy	47 (61%)
Laparoscopic	34
Open	13
Pathologic Stage	
T1	54
T2	9
Т3	12
T4	1
Preoperative eGFR, ml/min/1.73m ²	79 (65, 92)
Furhman Grade	
1	7
2	43
3	20
4	7
Disease Progression	13 (16%)

Table 1. Patient Characteristic	Table	atient Characte	ristics
---------------------------------	-------	-----------------	---------

Glycoproteomic analysis identified 39 glycopeptides and 14 non-glycosylated peptides (Appendix A), that showed statistically significant (false discovery rate $p \le 0.05$) differential expression among patients with shorter and longer PFS, respectively. Five of the glycosylated peptides conferred a continuous hazard ratio (HR) of > 6 (range 6.3-11.6). These included the prothrombin A2G2S glycan motif (HR=6.47, P=<0.001), immunoglobulin J chain FA2G2S2 motif (HR=10.69, P=0.001), clusterin A2G2 motif (HR=7.38, P=0.002), complement component C8A A2G2S2 motif (HR=11.59, P=0.002), and apolipoprotein M glycopeptide with non-

5

fucosylated and non-sialylated hybrid-type glycan (HR=6.30, P=0.003). Kaplan-Meier (KM) curves based on dichotomous expression of these five glycopeptides into high score and low score resulted in HR of 3.9-10.7, all with p-value < 0.03 (Figure 1). The median difference between high score and low score was 1.07 for THRB_A2G2S glycan motif, 0.46 for IGJ_FA2G2S2, 0.82 for CLUS_A2G2 motif, 0.86 for CO8A_A2G2S2, and 0.58 for APOM_Hybrid (Figure 1). Three glycoproteins had statistically significant correlation with increasing stage. Prothrombin A2G2S glycan motif (Linear estimate = 0.37, P=<0.001), clusterin A2G2 motif (Linear estimate = 0.17, P=0.04) associated with increasing tumor stage (Appendix B). Figure 2 represents other glycoprotein markers with continuous HR under 6, but dichotomization resulted in KM curves with HR > 10.



Figure 1. Kaplan Meier curve of Glycoprotein markers utilizing Harrell's C index to optimize cutoff values. This figure demonstrates the five glycoprotein markers with continuous hazard ratio >6 (resulting in dichotomized hazard ratios from 3.9-10.7).



Figure 2. Kaplan Meier curve of glycoprotein markers with dichotomized HR > 10

Glycopeptides with continuous HR > 6 and FDR < 0.05 were utilized for multivariable model building. Three markers (prothrombin A2G2S, immunoglobulin J chain G2FS, and complement component C8A A2G2S2) were retained via backwards selection. The three markers were used to generate multivariable LOOCV prediction models to determine progression. Predicted scores were dichotomized at cutoff that maximized Harrell's c-index and resulting Kaplan-Meier curve is plotted in Figure 3a (HR = 11.96, p = <0.001). Stratification by tumor stage demonstrates applicability in low stage (Figure 3b, HR = 14.56, p = 0.001) and high stage tumors (Figure 3c, HR =9.72, p = 0.057). One patient was not included in stratification by tumor stage due to missing stage.



Figure 3a. Kaplan-Meier (KM) curve of predicted survival score generated via LOOCV, utilizing prothrombin A2G2S, immunoglobulin J chain FA2G2S2, and complement component C8A A2G2S2. Figure 3b. KM curve stratified by stages 1-2. Figure 3c. KM curve stratified by stages 3-4.

To further assess biological significance of the biomarkers identified and to investigate their putative association with cancer biology, these biomarkers were subjected to pathway analysis to identify canonical pathways, protein networks, and predicted upstream regulators of relevance. The 10 statically most significantly enriched pathways with an overlapping *p*-value of < 0.01 are shown in Figure 4A. Three candidate biomarkers, APOM, CLU, and F2, are associated with the clathrin-mediated endocytosis signaling pathway. APOM and CLU are involved in the liver X receptor and retinoid acid X (LXR/RXR) pathway as well. The five statically most significantly enriched pathways also include the FXR/RXR pathway, atherosclerosis signaling, and IL-12 signaling and production in macrophages. Upstream regulators analysis identified three

9

transcription factors, HNF1A, JUNB and TAF12, at *p* values of <0.001 (Figure 4B). Among them, JunB was reported to modulate tumor invasion and angiogenesis of ccRCC through matrix metalloproteinase-2 (MMP-2), MMP-9 and chemokine (C-C motif) ligand-2 (CCL2)[11]. A GWAS study of the human N-glycome has shown that HNF1 α acts as a master regulator of plasma protein fucosylation through regulating the expression of major fucosyltranferase and fucose biosynthesis genes [12]. Proteins interacting directly or indirectly with the statistically significant glycopeptides in our study are consolidated into the network. The network is overlaid with functions and presented as a graph indicating the molecular relationships between proteins (Figure 4C). The results suggest that 24 (including C8A, JCHAIN, F2 and CLU) of 31 proteins in the network are associated with cancer development and progression.



Figure 4. IPA analysis of candidate biomarkers. (A) Top 10 canonical pathways enriched with candidate biomarkers (P < 0.01). (B) Top 3 upstream transcription factor regulators predicted. (C) Molecular protein network generated in IPA. The network is overlaid with disease and functions. The nodes highlighted with cyan are the candidate biomarkers. The molecules outlined with magenta are involved in cancer development based on knowledge findings. Solid lines represent direct interactions between two proteins and dotted lines show indirect interactions.

Discussion:

Protein glycosylation characteristics are affected by structural and regulatory modulations at the DNA and RNA level of enzymes, as well as by metabolic, inflammatory, and immunologic changes within the intra- and extracellular microenvironment. All of these have been proposed as factors in carcinogenesis. Changes of protein glycosylation may be envisioned as uniquely informative biomarkers as they represent the ultimate downstream product of cellular anabolism. In the case of ccRCC, downregulation of oxidative phosphorylation was recently demonstrated without observed changes at the RNA level, indicating that identifying drivers of oncogenesis must include downstream regulators, including evaluation of PTM[13].

Despite the evidence of PTM essential role in oncogenesis, biomarkers have been limited by the challenges associated with analysis. These are a complexly heterogeneous class of molecules that are difficult to characterize. Determining their specific localization within a given protein backbone represents an additional layer of challenges. Thus, the commonly used approach towards glycoproteomic characterization, namely the quantification in aggregate of all individual glycans released by enzymatic treatment of one or more proteins, ignores sitespecificity aspects of glycosylation and materially abandons the most critical information about the modification [6]. Likewise, studies that interrogate protein glycosylation sites in the absence of glycan information -- due to glycan degradation-- are not capturing the depth of information present [14]. Other approaches, such as depletion of abundant proteins to improve analysis of lower abundance proteins are similarly limiting for many reasons [1]. The site-specific glycoprotein analysis developed by Li et al. allows rapid, analysis of integrated proteomic and glycan data [7], without the bias introduced by depletion or enrichment. This has dramatically improved quantification efficiency of N-glycans in serum [5] and with use of advanced softwareenhanced platform has fundamentally changed the scalability.

The results of our study provide, for the first time to our knowledge, a set of biomarkers that highly correlates with prognosis and disease course of individuals diagnosed with ccRCC and treated with surgical tumor resection. We identified 36 glycosylated proteins in the serum of ccRCC patients that exhibited statistically significant expression differences in accordance with subsequent disease course. Five of these markers demonstrated hazard ratios above 3.5 (from 3.9-10.7) for PFS.

Most early studies of glycosylated protein markers have focused on early diagnosis and screening with limited evaluation of prognostication. A recent study in ovarian cancer demonstrated differences in the global glycoproteome as well as site specific differences between benign and malignant ovarian tumors[15], highlighting the importance of quantifying site-specific changes that would be otherwise missed. Likewise, site-specific glycosylation changes demonstrated in liver cirrhosis may be helpful as screening tools for hepatocellular carcinoma [16, 17].

Glycosylated protein biomarkers are still in their infancy. Renal cell carcinoma is a heterogeneous disease with multiple molecular pathways leading to differential tumorigenesis. External validation on a larger cohort is essential given the multiple subsets of proteomic alterations identified in ccRCC[13]. The major limitation of this study is the relatively large number of individuals lost to follow-up, impacting the applicability to larger cohorts. Also, the underrepresentation of more advanced-stage disease limits this subset. Another challenge is the lack of functional insights into the nature of the glycoprotein-isoform differences associated with phenotype. To our knowledge, none of the glycoproteins identified have been previously associated with ccRCC biology. It is unclear whether these glycopeptides are byproducts of the tumor released from the cell surface by matrix metalloproteases or whether they represent a more systemic inflammatory or immunologic reaction to the tumor. Clearly, additional experimentation will be needed to interrogate these questions. However robust association of glycoprotein profiles with tumor phenotype, if confirmed, can be a valuable tool for

prognostication and management of patients. Thus, glycoprotein profiling may help identify patients who may benefit from closer surveillance or adjuvant therapy after surgical treatment for ccRCC. Additionally, this approach may be explored in characterizing small renal masses which currently represents another diagnostic challenge.

Conclusions

Glycoprotein isoform differences measured in serum samples of patients with ccRCC prior to undergoing surgical treatment were found to be strongly associated with subsequent disease course and tumor recurrence. Our results warrant further studies to examine the findings more closely, and in larger cohorts. If confirmed, this may open the door for a more targeted, and potentially more effective, approach to the management of ccRCC.

12

References:

- 1. Lane, B.R. and M.W. Kattan, *Prognostic models and algorithms in renal cell carcinoma*. Urol Clin North Am, 2008. **35**(4): p. 613-25; vii.
- 2. Iafolla, M.A.J., et al., *Systematic review and REMARK scoring of renal cell carcinoma prognostic circulating biomarker manuscripts.* PLOS ONE, 2019. **14**(10): p. e0222359.
- 3. Pinho, S.S. and C.A. Reis, *Glycosylation in cancer: mechanisms and clinical implications*. Nat Rev Cancer, 2015. **15**(9): p. 540-55.
- 4. Drake, P.M., et al., *Sweetening the Pot: Adding Glycosylation to the Biomarker Discovery Equation.* Clinical Chemistry, 2010. **56**(2): p. 223-236.
- 5. Alley, W.R., et al., *N-linked Glycan Structures and Their Expressions Change in the Blood Sera of Ovarian Cancer Patients.* Journal of Proteome Research, 2012. **11**(4): p. 2282-2300.
- de Leoz, M.L., et al., *Glycomic approach for potential biomarkers on prostate cancer: profiling of N-linked glycans in human sera and pRNS cell lines*. Dis Markers, 2008.
 25(4-5): p. 243-58.
- Li, Q., et al., Site-Specific Glycosylation Quantitation of 50 Serum Glycoproteins Enhanced by Predictive Glycopeptidomics for Improved Disease Biomarker Discovery. Anal Chem, 2019. 91(8): p. 5433-5445.
- 8. Jorns, J.J., et al., *Kidney size and cancer-specific survival for patients undergoing nephrectomy for pT1 clear cell renal cell carcinoma*. Urology, 2012. **80**(1): p. 147-50.
- 9. Wu, Z., et al., *PB-Net: Automatic peak integration by sequential deep learning for multiple reaction monitoring*. J Proteomics, 2020. **223**: p. 103820.
- Benjamini, Y. and Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), 1995. 57(1): p. 289-300.
- 11. Kanno, T., et al., *JunB promotes cell invasion and angiogenesis in VHL-defective renal cell carcinoma*. Oncogene, 2012. **31**(25): p. 3098-110.
- Lauc, G., et al., Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1α as a master regulator of plasma protein fucosylation. PLoS Genet, 2010. 6(12): p. e1001256.
- 13. Clark, D.J., et al., *Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma*. Cell, 2019. **179**(4): p. 964-983.e31.
- 14. Zeng, X., et al., *Lung cancer serum biomarker discovery using glycoprotein capture and liquid chromatography mass spectrometry*. J Proteome Res, 2010. **9**(12): p. 6440-9.
- 15. Li, Q.K., et al., An integrated proteomic and glycoproteomic approach uncovers differences in glycosylation occupancy from benign and malignant epithelial ovarian tumors. Clinical Proteomics, 2017. **14**(1).
- 16. Lee, H.J., et al., *Abundance-ratio-based semiquantitative analysis of site-specific N-linked glycopeptides present in the plasma of hepatocellular carcinoma patients.* J Proteome Res, 2014. **13**(5): p. 2328-38.

13

					CoxPH Models - I	Dichotomized by	C-index				
Marker	Protein	Marker Type	Glycan Structure	Hazard	Confidence	P Value	FDR	Cutoff	Hazard	Confidence	P Value
APOD FA2G2	APOD	alvcopeptide	5410	2 29	144-3.62	4 23E-04	1 19E-02	2 85	14.54	4 28-49 31	1 75E-05
AGP12 FA4G4S4	AGP12	glycopeptide	7614	2.23	1.35-3.69	1.73E-03	2.36E-02	2.78	11.09	3.00-40.95	3.08E-04
APOB A2G2S1	APOB	alvcopentide	5401	1.52	1 24-1 86	5.59E-05	5 12E-03	4 66	10.89	3 34-35 46	7.41E-05
TRFE A3G3S2	TRFE	glycopeptide	6502	1.34	1.17-1.52	1.44E-05	5.12E-03	5.23	10.79	3.28-35.47	8.98E-05
IGJ FA2G2S2	IGJ	alvcopeptide	5411	10.69	2 51-45 49	1.35E-03	2 03E-02	0.89	10.65	2 31-49 04	2 39E-03
A2GL GQTLLAVAK	A2GL	peptide		1.85	1.24-2.77	2.81E-03	3.29E-02	2.66	10.63	2.89-39.12	3.75E-04
APOD 42G2S2	APOD	alvcopentide	5402	1.98	1 38-2 84	2 13E-04	6 74E-03	3.06	9.96	3 06-32 46	1 37E-04
APOD hybrid A6G6	APOD	glycopeptide	9800	2.17	1.46-3.23	1.38E-04	5.34E-03	2.35	9.46	3.01-29.8	1.23E-04
CLUS A2G2	CLUS	glycopeptide	5400	7.38	2.12-2.57	1.72E-03	2.36E-02	1.35	9.19	2.84-29.73	2.14E-04
AGP12 WFYIASAFR	AGP12	peptide		3.13	1.61-6.08	7.51E-04	1.72E-02	1.80	9.07	2.87-28.61	1.69E-04
AGP01 EQLGEFYEALDCLR	AGP01	peptide		2.43	1.37-4.29	2.30E-03	2.90E-02	1.47	9.07	2.87-28.61	1.69E-04
THRB A2G2S2	THRB	glycopeptide	5402	4.78	2.30-9.92	2.65E-05	5.12E-03	1.75	8.12	2.55-25.89	3.99E-04
HPT ILGGHLDAK	HPT	peptide		3.39	1.63-7.04	1.08E-03	1.87E-02	1.64	8.02	2.56-25.15	3.55E-04
KLKB1 A3G3S3	KLKB1	glycopeptide	6503	1.19	1.05-1.35	5.16E-03	4.91E-02	6.27	7.79	1.94-31.34	3.84E-03
UN13A hybrid	UN13A	glycopeptide	7420	1.31	1.15-1.50	3.59E-05	5.12E-03	4.57	7.45	2.34-23.75	6.90E-04
AGP2 EHVAHLLFLR	AGP2	peptide		9.73	2.90-32.6	2.27E-04	6.74E-03	1.30	7.45	1.98-28.07	3.02E-03
THRB A2G2S1	THRB	glycopeptide	5401	6.47	2.53-16.5	9.53E-05	5.12E-03	1.43	7.36	2.31-23.41	7.31E-04
CAN3 A3G3S3	CAN3	glycopeptide	6503	1.47	1.16-1.86	1.19E-03	1.94E-02	2.43	7.30	2.30-23.15	7.36E-04
AGP01 YVGGQEHFAHLLILR	AGP01	peptide		2.16	1.26-3.71	5.06E-03	4.91E-02	1.56	7.28	2.30-22.97	7.16E-04
AACTADLSGITGAR	AACT	peptide		4.92	1.79-13.5	2.06E-03	2.73E-02	1.41	7.23	2.25-23.24	8.99E-04
CO8A A2G2S2	CO8A	glycopeptide	5402	11.59	2.41-55.7	2.24E-03	2.89E-02	1.53	7.12	2.15-23.52	1.29E-03
PON1 A3G3S2	PON1	glycopeptide	6502	1.26	1.10-1.46	1.37E-03	2.03E-02	3.78	7.02	2.21-22.32	9.55E-04
PON1 FA2G2F1	PON1	glycopeptide	5420	1.51	1.19-1.91	6.99E-04	1.68E-02	3.08	6.99	2.13-22.94	1.35E-03
APOD FA3G3F2	APOD	glycopeptide	6530	2.22	1.48-3.33	1.18E-04	5.25E-03	3.52	6.91	2.21-21.65	9.03E-04
CLUS A3G3S3	CLUS	glycopeptide	6503	1.68	1.29-2.19	1.25E-04	5.25E-03	2.01	6.83	2.17-21.47	1.01E-03
APOC3 O-glycan F1S1	APOC3	glycopeptide	1111	1.84	1.34-2.53	1.76E-04	6.35E-03	2.33	6.81	2.16-21.52	1.08E-03
APOC3 O-glycan S2	APOC3	glycopeptide	1102	4.47	2.02-9.91	2.27E-04	6.74E-03	1.05	6.81	2.16-21.52	1.08E-03
HPT FA3G3S3	HPT	glycopeptide	6513	2.21	1.30-3.74	3.35E-03	3.67E-02	1.94	6.79	2.14-21.52	1.14E-03
HPT LPECEAVCGKPK	HPT	peptide		7.18	2.23-23.0	9.51E-04	1.82E-02	0.79	6.47	2.05-20.37	1.43E-03
PON1 A3G3S1	PON1	glycopeptide	6501	1.30	1.14-1.49	7.51E-05	5.12E-03	5.88	6.15	1.81-2.83	3.55E-03
APOM FA2G2F1S1	APOM	glycopeptide	5421	3.16	1.54-6.46	1.64E-03	2.36E-02	1.60	5.74	1.26-26.27	2.43E-02
TRFE A2G2	TRFE	glycopeptide	5400	1.37	1.17-1.60	8.10E-05	5.12E-03	3.38	5.20	1.70-15.89	3.80E-03
CAN3 FA3G3S3	CAN3	glycopeptide	6513	1.78	1.33-2.38	1.02E-04	5.12E-03	4.66	4.93	1.50-16.17	8.45E-03
AGP2 EQLGEFYEALDCLCIPR	AGP2	peptide	5410	11.69	2.61-52.4	1.32E-03	2.03E-02	1.05	4.90	1.49-16.07	8.72E-03
C15 A2G252	C1S	glycopeptide	5402	3.33	1.51-7.31	2.75E-03	3 29E-02	1.22	4.50	1.38-14.87	1.00E-02
A2GL DLLLPQPDLR	A2GL	peptide		1.76	1.20-2.59	3.94E-03	4.14E-02	1.33	4.23	1.31-13.62	1.56E-02
APOC3 O-glycan E1	APOC3	alvcopeptide	2110	121	1 08-1 35	7.01F-04	1.68E-02	6.30	4 19	1 20-14 70	2 51E-02
KLKB1 A2G2S2	KLKB1	glycopeptide	5402	3.08	1.57-8.01	1.01E-03	1.82E-02	0.66	4.01	1.27-12/69	1.83E-02
APOM hybrid	APOM	glycopeptide	8500	6.30	1.87-21.2	2.97E-03	3.33E-02	1.20	3.86	1.16-12.85	2.80E-02
APOC3 O-glycan	APOC3	glycopeptide	1300	1.41	1.15-1.72	1.01E-03	1.82E-02	1.71	3.71	1.07-12.87	3.92E-02
AFAM A2G2S2	AFAM	glycopeptide	5402	3.21	1.44-7.18	4.47E-03	4.51E-02	0.89	3.36	1.01-11.11	4.74E-02
UN13A FA2G2F2S1	UN13A	glycopeptide	5431	1.55	1.25-1.93	9.02E-05	5.12E-03	3.40	3.16	1.03-9.72	4.50E-02
APOD FA3G3	APOD	glycopeptide	6510	1.99	1.35-2.93	4.90E-04	1.30E-02	1.10	3.15	0.85-11.68	8.55E-02
FETUA A3G3S3	FETUA	glycopeptide	6503	3.91	1.50-10.1	5.15E-03	4.91E-02	1.36	3.15	0.94-10.50	6.22E-02
FETUA FA3G3S3	FETUA	glycopeptide	6513	2.01	1.44-2.81	4.26E-05	5.12E-03	0.99	3.14	0.86-11.47	8.38E-02
FETUA A2G2S2	FETUA	glycopeptide	5402	6.63E-03	2.43E-04- 1.81E-01	2.95E-03	3.33E-02	0.93	0.22	0.07-0.68	9.10E-03
TRFE DSAHGFLK	TRFE	peptide		2.50E-04	1.84E-06- 3.41E-02	9.42E-04	1.82E-02	0.76	0.17	0.05-0.63	7.74E-03
ITIH1 LDAQASFLPK	ITIH1	peptide		7.26E-03	2.47E-04- 2.13E-01	4.29E-03	4.42E-02	0.71	0.13	0.04-0.39	3.50E-04
APOM AFLLTPR	APOM	peptide		3.58E-04	3.41E-06-	8.28E-04	1.74E-02	0.57	0.09	0.03-0.30	5.44E-05

17. Zhu, J., et al., *Glycoproteomic markers of hepatocellular carcinoma-mass spectrometry based approaches.* Mass Spectrom Rev, 2019. **38**(3): p. 265-290.

1	Λ
ᆂ	-

					3.75E-02						
APOM SLTSCLDSK	APOM	peptide		3.10E-03	9.62E-05- 9.99E-02	1.12E-03	1.87E-02	0.66	0.09	0.03030	7.12E-05
AACT A4G4S3	AACT	glycopeptide	7603	6.27E-04	8.33E-06- 4.72E-02	8.22E-04	1.74E-02	0.27	0.08	0.02-0.25	1.65E-05
IGG1 FA2G2S2	IGG1	glycopeptide	5411	0.16	0.05-0.55	3.48E-03	3.73E-02	1.51	0.05	0.01-0.43	5.59E-03

Appendix A. Glycoproteomic analysis identified 39 glycopeptides and 14 non-glycosylated peptides. FDR = False discovery rate.





Appendix B. Linear Regression Analysis of Association of the five glycoprotein markers with continuous hazard ratio >6 with increasing tumor stage.