Analysis of N-linked Glycopeptides Derived from Human Liver Tissues by LC-MS/MS

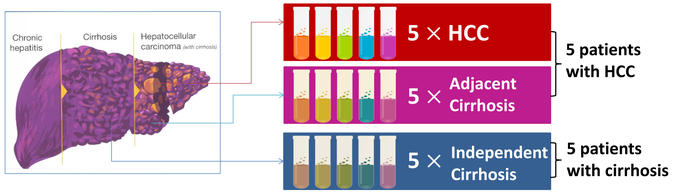
Minkun Wang[1,2], Cristina Di Poto[1], Ehwang Song[3], Rui Zhu[3], Yehia Mechref[3], Habtom W. Ressom[1]

[1]Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC;

[2]Department of Electrical and Computer Engineering, Virginia Tech, Arlington, VA; [3]Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, TX

Overview

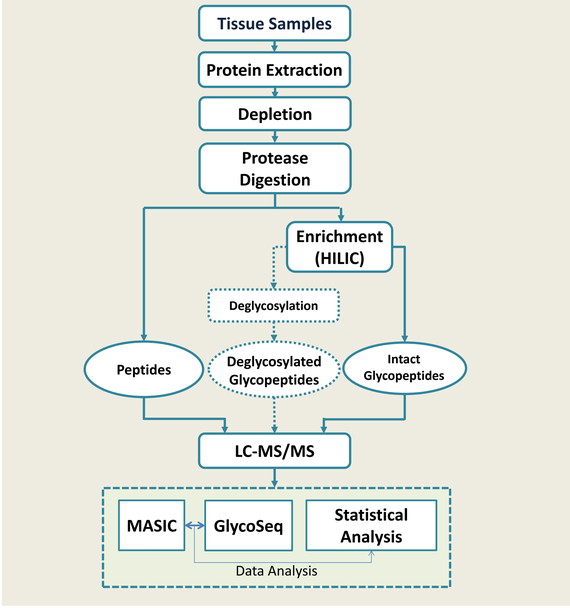
- Hepatocellular carcinoma (HCC) is a significant worldwide health problem with as many as 600,000 new cases diagnosed each year.
- Primary liver cancer is the fifth most common cancer worldwide and the third most common cause of cancer mortality.
- Patients with cirrhosis have an annual risk of 1-2% for developing HCC.
- Malignant conversion of cirrhosis to HCC is often diagnosed at a late stage.
- Glycoproteins as new HCC markers include AFP-L3, GP73, etc.
- Quantitative alterations in glycosylation of proteins are associated with many human diseases including liver cancer.
- LC-MS/MS allows us to evaluate the levels of N-linked glycopeptides digested from proteins in biological samples to characterize the alterations.
- The microheterogeneity of glycosylation sites, however, represents significant challenges in the analysis of glycoproteins.
- In this study, we aim to investigate the changes in levels of intact N-linked glycopeptides among different disease groups.



- This study aims to investigate the changes in glycosylation by comparing the levels of N-linked glycopeptides digested from proteins in human liver tissues of HCC patients with those of cirrhotic patients using LC-MS/MS.

Method

- We analyzed 15 liver tissues from 10 participants recruited at MedStar Georgetown University Hospital. The tissues represent 5 HCC cases (5 tumor and 5 adjacent cirrhotic tissues) and 5 patients with liver cirrhosis.
- Following protein extraction, depletion, digestion, and enrichment by cotton hydrophilic interaction liquid chromatography (HILIC) microcolumns, the samples were analyzed using LTQ Orbitrap Velos mass spectrometer coupled with an ultimate 3000 nano-LC system.



- MS/MS data were acquired using CID and ETD.
- With CID mode of fragmentation, the diagnostic sugar oxonium ions, sequential neutral losses of glycosyl residues, and Y1 ions of peptide backbones were detected and utilized to infer the complete glycopeptide information.
- Accurate peptide backbone information can be obtained by electron transfer dissociation (ETD), which were simultaneously generated from the same samples.
- Parallelly, we carried out proteomic studies before and after enrichment on the same samples, with expectation of providing specific background information for glycosylated proteins.

Data Analysis

- **Identification:**
  - The acquired LC-MS/MS data were analyzed by **GlycoSeq** (Indiana University) to identify putative glycopeptide assigned with attached N-glycans.
  - A pre-specified list of glycoproteins/glycopeptides was provided as background knowledge in order to facilitate the identification.
  - GlycoSeq created a glycopeptide database by tryptically digesting the provided glycoproteins in silico with maximum missed cleavages at 2. Together with N-linked glycan database, the tool identified putative glycopeptides using combined information from the two databases.
- **Quantification:**
  - The acquired LC-MS/MS data were also analyzed by **MASIC** (Pacific Northwest National Lab) for fast quantitation of any detected features, i.e., extracted ion chromatograms (EICs) for all of the parent ions selected for fragmentation.
- **Statistical Analysis:**
  - We applied paired t-test (HCC vs. Adjacent Cirrhosis) and unpaired t-test (HCC vs. Cirrhosis) on log-transformed intensities of glycopeptides detected in the majority of the samples.
  - We also consider glycopeptides that are present in one group but absent in the other group as candidates.

Experimental Results

- **Identification:**
  - **3,355** putative N-linked glycopeptides were identified in all 15 tissues combined (~ 1,000 per tissue) consisting of **934** unique peptide backbones, originating from 82 glycoproteins. At this stage, we provided a list of 85 glycoproteins (related to the studied diseases) as background knowledge to GlycoSeq.
- **Quantification:**
  - ~12,000 precursor ions were quantified by MASIC based on their EICs in all 15 tissues combined (about 2,800 per tissue)
- **Statistical Analysis:**
  - **588** analytes (**254** with putative glycopeptide IDs) had quantitative information with < 4 missing values
  - **30** significant analytes (16 with putative glycopeptide ID) with $p < 0.05$
  - **15** glycopeptides that are based on present and absent calls.
  - **17** glycoproteins are also reported with significant alteration in serum proteomic study.

HCC vs. adjacent CIRR & HCC vs. CIRR

- P13598 (GNETLHYETFGK) *
- P04275 (IGEADFNR)
- P02774 (LCDNLSTK) *
- P19320 (LDNGNLQHLSGNATLTLIAMR) *
- P03952(IYPGVDFGGEELNVTFVK)*

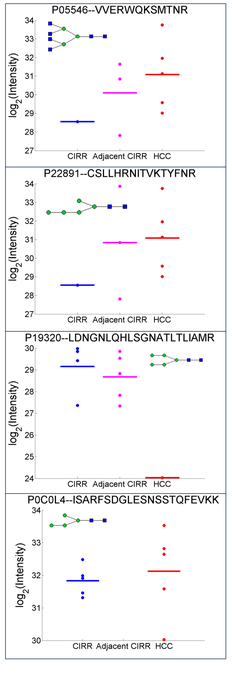
HCC vs. CIRR only

- P02790(GHGHRNGTGHGNSTHHGPEYMR, ALPQPQNVTSLLGCTH)
- P04278 (LDVDQALNR)
- P08571(CMWSSALNSLNLSFAGLEQVPK)
- Q13201 (LTLLQKKIDNISLTVNDVR)*
- Q9Y6R7 (KVTVRPGGESVMVNISAK) *
- O75636 (VELEDFNGNR)*
- P01877(MAGKPTHVNVSVVMAEVDGTCY)*
- P05546(VVERWQKSMTNR)*
- P22891(CSLLHRNITVKTYFNR)*
- P41222(FLGRWFSAGLASNSSWLR)*
- P51884(LHINHNNLTESVGPLPK)*
- Q10588(NKNCTAIWEAFKVALDK)*
- Q9UNN8(AHVFFEVAVNGSSFVSFR)
- Q9Y5Y7(KANQQLNFTEAKEACR)*

HCC vs. adjacent CIRR only

- P01011 (NGTRGK)
- P01033(FVGTPEVNQTTLYQRYEIKMTK)
- P10909 (KKEDALNETR, HNSTGCLR)*
- Q9Y6R7 (VVTVAALGTNISIHKDEIGK) *
- P0C0L4(ISARFSDGLESNSSTQFEVKK)
- P14151(FCRDNYTDLVAIQNK)*
- Q16610(QGNNHTCTWK)

** significant in serum proteomic study*



Summary

- In this study, we investigated N-linked glycopeptide-level alterations in human liver tissues of HCC and cirrhosis patients using LC-MS/MS.
- We found ~30 glycopeptides with statistical significance between HCC and cirrhosis/adjacent cirrhosis.

Future Work

- Proteins/peptides detected in the same tissue samples will be used as background database for identification of glycopeptides.
- ETD data will be analyzed for more accurate identification of peptide backbones.
- Various glycoforms on the same glycosylation site will be investigated.
- Levels of protein glycosylation and corresponding proteins will be compared.

Acknowledgements

This work was supported by NIH Grants R01CA143420 and R01GM086746.