An N-glycoproteomics Analysis Reveals Glycoprotein Alterations in Esophageal Squamous Cell Carcinoma

Yingzhen Gao

Shanxi Medical University

Liuyi Shen

Shanxi Medical University

Tianyue Dong

Shanxi Medical University

Xin Yang

Shanxi Medical University

Heyang Cui

Peking University Shenzhen Hospital

Yanlin Guo

Shanxi Medical University

Yanchun Ma

Shanxi Medical University

Pengzhou Kong

Shanxi Medical University

Xiaolong Cheng

Shanxi Medical University

Ling Zhang (✉ zhanglingty@sxmu.edu.cn)

Shanxi Medical University <https://orcid.org/0000-0002-4100-0475>

Yongping Cui

Shanxi Medical University

Research Article

Keywords: ESCC, glycoprotein, N-glycosylation, N-glycoproteomics, lymph node metastasis

Posted Date: January 27th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1262095/v1>

Abstract

Background: Aberrant glycosylation has been recognized as a hallmark of cancer. It usually includes the changes of expression level and function of glycoproteins caused by abnormal glycosylation modification and changes of glycan structure of glycoproteins. N-glycosylation is one of the main types of glycosylation in all eukaryotes.

Methods: We integrated global proteomics and N-glycoproteomics analysis of ten cases to reveal the N-glycoproteomics profiles in esophageal squamous cell carcinoma (ESCC) compared with normal tissues.

Results: A series of differentially expressed glycoproteins (e.g. LAMP2, PLOD2) and some differentially glycoproteins enriched signaling pathway (e.g. metabolism-related pathway, ECM-receptor interaction, focal adhesion) were identified in ESCC. Seven significantly enriched motifs were found from all of the identified N-glycosylation sites. The dynamic profiling changes of glycoprotein during lymph node metastasis progression in ESCC was also investigated and three clusters showed the obvious increasing or decreasing trend with lymph node metastasis.

Conclusion: In summary, we presented the characters of N-glycosylation and N-glycoprotein alterations associated with ESCC. Specific changes in the expression of glycoproteins or glycosylation occupancy have the potential to be used as the biomarker to discriminate ESCC from normal tissues and to improve our understanding of ESCC biology.

Introduction

Esophageal cancer is one of the most common malignant tumors worldwide and about half of global esophageal cancer patients are in China, where the main histologic type is esophageal squamous cell carcinoma (ESCC). ESCC patients tend to have a very poor prognosis with the 5-year survival rate less than 30% [1]. Although the combined detection of some markers such as squamous cell carcinoma antigen (SCC), carcinoembryonic antigen (CEA) and cytokeratin-19-fragment (CYFRA21-1) is helpful for the diagnosis and prognosis prediction of ESCC patients [2], the tumor markers used in the early diagnosis of ESCC are not specific and mature. Radical surgery is considered to be the best choice to treat ESCC patients. However, most ESCC patients are in the middle and late stage at initial diagnosis, and recurrence remains still the main reason for treatment failure. Although the application of some drugs such as trastuzumab and ramucirumab have entered the clinic, there is still a lack of specific and effective therapeutic target in ESCC compared with other types of cancer[3]. Thus, there is a long way to go to further study the molecular mechanism of ESCC.

Aberrant glycosylation has been recognized as one of the hallmarks of cancer, which conferred a new perspective for cancer research [4]. As a prominent post-translational modification of proteins, glycosylation plays key roles in a variety of biological process such as cell transformation, differentiation, growth, invasion, and immune surveillance of tumors [5]. There are two main ways of glycosylation modification, N-glycosylation and mucin-type O-glycosylation occurring in cell endoplasmic reticulum and

Golgi complex respectively [5]. Most tumor biomarkers currently used in clinical practice are glycoproteins (e.g. CA125 for ovarian cancer, AFP for liver cancer and CEA for colon cancer) or glycan-related such as CA19-9 for gastrointestinal and pancreatic cancer [6]. Fucosylated haptoglobin (Fut-Hpt) could be a biomarker for the detection of liver metastasis of colorectal or pancreatic cancer [7]. Moreover, antibodies of the fucosylated Le Y have also been used as potential drugs for immunotherapy in epithelial-derived tumors [8]. Thus, the study of the specific glycosylation changes in cancer especially ESCC may provide a new way for clinical translation and application.

Abnormal glycosylation usually includes the changes of glycoprotein expression level caused by abnormal glycosylation modification and changes of glycan structure of glycoproteins determined by glycogene alteration. However, the complex characteristics of glycans and the limitation of research methodology made glycoproteomics and glycomics study severely lag behind. To date, there is no information about abnormal glycosylation signatures in ESCC. Here, we performed N-glycoproteomics analysis to investigate differentially expressed glycoproteins in ESCC compared with matched normal tissues. Because changes of glycoproteins may be a result of changes in protein concentration [9], we integrated proteomics data and N-glycoproteomics data to evaluate the effect of protein normalization on identified glycosylation changes. Moreover, we investigated dynamic profiling changes of glycoprotein expression during lymph node metastasis progression of ESCC. This study might throw new light on the early diagnosis and molecular targeted treatment of ESCC.

Materials And Methods

Patient tissue samples

The ten pairs of tumor tissues and adjacent normal tissues at least 5 cm from the center of the tumor used for N-glycoproteomics and proteomics analysis in this study were obtained from patients diagnosed with ESCC who underwent surgical resection at the Shanxi Cancer Hospital (Taiyuan, China). There was no preoperative chemotherapy, radiotherapy or other therapies done on these patients. All the recruited patients gave their informed consent before they participated in our study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the Shanxi Medical University (Approval No. 2017LL037). Details of clinicopathological features of cases are summarized in **Table S1**.

Protein extraction and trypsin digestion

After the samples were ground into cell powder in liquid nitrogen, four times the volume of lysis buffer (1% protease inhibitor, 3 μ M TSA, 8 M urea, 50 mM NAM and 2 mM EDTA) was added and the samples were sonicated on ice using an ultrasonic processor. Once removing the remaining debris, the supernatant was collected and BCA kit was used to detect the protein concentration. Then the samples were reduced with 5 mM dithiothreitol for 30 min at 56°C, iodoacetamide was added to make the final concentration 11 mM and samples were incubated at room temperature for 15 min in darkness. Finally, the urea concentration

was diluted to less than 2 M. Trypsin was added at 1:50 trypsin-to-protein mass ratio overnight for the first digestion and 1:100 trypsin-to-protein mass ratio for 4 hours for the second digestion.

TMT Labeling and PTM enrichment for N-glycosylation

The digested peptides were desalted using Strata X C18 SPE column (Phenomenex) and vacuum-dried. Then, the peptides were reconstituted in 0.5 M TEAB buffer and labeled with TMT kit according to the manufacturer's protocol. Briefly, the peptides were incubated with TMT reconstituted in acetonitrile for 2 hours at room temperature, then pooled, desalted and dried. The peptides were dissolved in 40 µL enrichment buffer containing 80% acetonitrile/1% trifluoroacetic acid. Then the supernatant was transferred to a hydrophilic (HILIC) micro-column. After centrifugation at 4000 g for 15 minutes, the HILIC micro-column was washed three times. Subsequently, glycopeptides were eluted using 10% acetonitrile and the eluate was collected. After being vacuum freeze-dried, samples were dissolved in 50 µl of ammonium bicarbonate buffer dissolved in 50 µl heavy oxygen water. Finally, 2 µl of PNGase was added and samples were digested overnight at 37 ° C.

HPLC Fractionation and LC-MS/MS analysis

The tryptic peptides were fractionated using high pH reverse-phase HPLC in Thermo Betasil C18 column (5 µm particles, 250 mm length, 10 mm ID). Firstly, the peptides were separated with a gradient of 8–32% acetonitrile (pH = 9.0) for more than 60 min into 60 fractions, followed by being dried by vacuum centrifuging. All of the above processes, LC-MS/MS and subsequent bioinformatics analyses were performed in Jingjie PTM Biolabs (Hangzhou in China). The peptides were dissolved in 0.1% formic acid (solvent A) and directly loaded into a reversed-phase analytical column (75 µm i.d, 15-cm length.). The gradient consisted of a gradual increase from 6–23% of solvent B (0.1% formic acid in 98% acetonitrile) for more than 26 min, 23–35% in 8 min and climbing to 80% in 3 min, followed by being held at 80% for the last 3 min. Above all were operated at a constant flow rate of 400 nL / min on EASY-nLC 1000 UPLC system. The peptides were submitted to NSI source and were analyzed by tandem mass spectrometry (MS/MS) in Q Exactive™ Plus (Thermo). The m/z scan range was 300 - 1600 for full scan, and Orbitrap was used to detect intact peptides at a resolution of 70,000. Then, peptides were selected and analyzed by MS/MS using NCE setting as 28 and Orbitrap was used to detect the fragments at a resolution of 17,500. Automatic gain control (AGC) was set at 5E4. Fixed first mass was set as 100 m/z.

Database Search.

Maxquant search engine (v1.5.2.8) was used to retrieve the mass spectrometer data. Parameter settings are as follow. The database is SwissProt Human (20130 sequences), and the reversed database is added to calculate the false positive rate (FDR) caused by random matching. The cleavage enzyme method was set as Trypsin/P allowing up to 4 missing cleavages. The minimum required peptide length was set to seven amino acids and the maximum modification number of the peptide was set to 5. In First search, the mass tolerance for precursor ions was set as 20 ppm. In Main search, the mass tolerance for precursor ions was set as 5 ppm. The mass tolerance for fragment ions was set as 0.02 Da. FDR was adjusted to < 1% and minimum score for modified peptides was set > 40.

Gene Ontology (GO) and KEGG pathway annotation.

The annotation of Gene Ontology (GO) of proteins comes from the UniProt-GOA database ([www.http://www.ebi.ac.uk/GOA/](http://www.ebi.ac.uk/GOA/)). Firstly, the system converts protein ID to UniProt ID, then uses UniProt ID to match GO ID, and extracts corresponding information from UniProt-GOA database based on GO ID. If there is no according protein information in UniProt-GOA database, the InterProScan soft based on protein sequence, will be used to annotate protein's GO function. The proteins were then classified according to cellular component, biological process and molecular function. Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to annotate protein signaling pathway. Firstly, we used KEGG online service tools KAAS to annotate KEGG database description of proteins. Then KEGG mapper, KEGG online service tools was used to map the annotation result on the KEGG pathway database. Wolfpsort, an updated version of PSORT/PSORT II, was applied to predict subcellular localization.

Motif Analysis.

Software MoMo (motif-x algorithm) was applied to analyze the motif characteristics of N-glycosylation sites. Among them, the peptide sequences of 10 amino acids upstream and downstream of all the N-glycosylation site were analyzed. And all the database protein sequences were used as background database parameter. When the number of peptides in a specific sequence is more than 20 and the p value is less than 0.000001, it is considered that the characteristic sequence is a motif of modified peptide.

Soft-clustering analysis.

The Mfuzz package (version 2.36.0) based on the open-source statistical language R (version 3.6.3) was used to perform soft-clustering analysis. The fuzzy c-means algorithm, one of the most widely used soft-clustering method, provided by Mfuzz package, is more noise robust and a priori pre-filtering of genes/proteins can be avoided, so it avoided the exclusion of biologically relevant genes/proteins from the data analysis[10, 11]. FCM clustering uses Euclidean distance as the distance metric, and demands two main parameters (c = number of clusters, m = fuzzification parameter). Different from k-means clustering, each element has a set of membership coefficients corresponding to the degree of being in a given cluster by FCM clustering.

For the quantitative information of N-glycosites, we removed the N-glycosites containing missing values in the raw data at first, to make sure that each gene could be quantified in each sample. Calculated the quantitative ratio of N-glycosites in each sample pair, and the quantitative ratios were log2 transformed and then z-score normalized for each pair of sample, to make the mean was zero and standard deviation was one. The normalized-data makes sure that N-glycosites with similar temporal patterns are close in Euclidean space [12]. The transformed profiles were then clustered using the Mfuzz package. For this analysis, the optimal values of c and m were derived. Final clustering was done with the parameters $c = 15$ and $m = 1.48$. Then we screened out the groups related to the degree of LNM, set the appropriated membership coefficients to make the remaining genes in these groups had a similar trend of change (membership ≥ 0.30).

Results

Study design and global proteomics analysis of ESCC

Figure 1 shows the workflow for the analysis of global N-glycoproteomics and proteomics of ESCC and matched normal tissues. The separately matched samples for potential biomarkers discovery were used and expected to reflect the level of different individuals. Twenty protein samples from 10 patients were assigned to two experiments with TMT10plex markers and digested with trypsin for N-glycoproteomics and proteomics analysis. A detailed description of the clinical characteristics of the analyzed samples is presented in **Table S1**.

In order to exclude the possibility that the changes of glycoproteins may be a result of changes in protein concentration, we firstly investigated global proteomic profile in 10 cases with paired tumor tissue and adjacent normal tissues using TMT labeling-based quantitative proteomics. On one hand, quality control of global proteomic data indicated that the proteomics analysis system was robust (Figure 2). On the other hand, we exclude the data of two samples (#8 and #9) because principal component analysis and Pearson's correlation coefficient analysis indicated that the extreme values of the two samples were probably caused by sample collection problems. Altogether, 6856 proteins were identified from these pairs of ESCC tissues, among which 6113 proteins were quantified. The mean value of 6113 protein quantitative ratios was 1.11, indicating expression levels of most proteins were unchanged. With the threshold of >1.5 fold change and p-value <0.05, it was found that 596 proteins were up-regulated while 495 proteins were down-regulated when comparing ESCC samples with their matched normal tissues (Figure 3A). All these data were listed in **Table S2**. Proteins showing significant up/down-regulation in all paired comparisons were exhibited in detail in the heat-map of Figure 3B. In terms of cellular component, the up-regulated proteins were mainly localized in chromosome, nuclear lumen, spliceosomal complex and intracellular organelle lumen, while most of the down-regulated proteins were found in extracellular region, extracellular matrix and vesicle (Figure 3C). Accordingly, KEGG-based enrichment analysis showed that the prominent pathways enriched in the increased proteins included spliceosome, DNA replication, RNA transport, transcriptional mis-regulation in cancer, cell cycle, base excision repair and mismatch repair. By contrast, protein expression in the cellular pathways including ECM-receptor, focal adhesion, metabolism (e.g. protein digestion and absorption; tyrosine, histidine, phenylalanine, and tryptophan metabolism; drug metabolism-cytochrome P450), interaction and proteoglycans was decreased (Figure 3D). Because proteomic analysis of more ESCC samples will be reported in our another study, here, we focused on the N-glycoproteomics analysis and all the following N-glycoproteomics data were normalized by corresponding proteomics data.

Motif analysis of N-glycoproteome

As one type of classical post-transcriptional modification (PTM), glycosylation has been proven to be essential for regulating cellular functions. Although N-glycoproteomics study has been applied in many cancers, it remains unreported in ESCC. In this study, the quality control results of N-glycoproteomics showed that our optimized methods were highly confident. Most identified peptides are distributed in 8-20

amino acids, which is consistent with the common pattern of trypsin digestion (Figure 4A). The mass error of most spectrograms is within 5 ppm, which indicates that the accuracy of the mass spectrometer is reliable and the qualitative analysis of protein will not be affected by the large mass deviation (Figure 4B). Principal component analysis and Pearson's correlation coefficient analysis demonstrated a good reproducibility and accuracy in LC-MS/MS analysis (Figure 4C-D). In total, N-glycoproteomics data quality control indicated that the analysis system was robust.

To characterize the possible specific sequence, a motif analysis was performed to indicate the likelihood of amino acid types being over- or under-represented at the positions surrounding the N-glycosylation sites. Seven significantly enriched motifs were found from all of the identified N-glycosylation sites. As shown in Figure 4E, motif logos according to the score ranking included N-X-T, N-C-S, R-XXXXXX-N-X-S, N-A-S, R-N-X-S, N-X-S, N-G-X (X represents a random amino acid residue; N, T, A, G, C, S, R represent asparagine, threonine, alanine, glycine, cystine, serine, arginine respectively). The amino acid frequencies in the sequences flanking N-glycosylation sites were assessed by motif model to confirm whether there were position-specific amino acids surrounding N-glycosylation sites (Figure 4F). In addition to canonical N-linked glycopeptide sequon of N-X-T/S, we found that cysteine residue at +1 and +3, glycine residue at -1 and +1, arginine residue at -1, -2 and -7, alanine residue at +1 position were overrepresented surrounding N-glycosylation sites. Interestingly, despite N-X-C and N-X-V have been discovered to be atypical N-glycosylation sequons in recent years [13], cystine and valine were obviously depleted at +3 position of N-glycosylation site in our study.

Characteristics of the N-glycoproteins: identification and quantitative analysis

On average in this study, a total of 1,839 N-glycosylation sites in 1021 glycoproteins were identified, of which 1588 sites in 901 glycoproteins were quantitative. Figure 5A demonstrated the distribution of the number of identified N-glycosylation sites in each glycoprotein. Among all identified glycoproteins, approximately 62.8% (641/1021) contained only one N-glycosylation site, 20.3% (207/1021) possessed two N-glycosylation sites, and 5.5% (56/1021) contained five or more N-glycosylation sites. Notably, six glycoproteins were found to have ten or more N-glycosylation sites: LDL receptor-related protein 1 (23 sites), laminin subunit alpha-2 (14 sites), fibrillin-1 (13 sites), laminin subunit gamma 1 (12 sites), tenascin C (10 sites) and laminin subunit alpha 5 (10 sites).

In order to explore the abundance difference between the global N-glycoproteins and proteins, we compared datasets of the quantified N-glycoproteomics and the proteomics. A case-by-case review demonstrated that the numbers of quantified proteins are more than glycoproteins for individual ESCC. Moreover, the numbers of quantified proteins have no trend while the numbers of quantified glycoproteins increased with disease development and lymph node metastasis (Figure 5B). Figure 5C shows the overlap in the protein groups identified in both N-glycoproteomics and proteomics. There are 542 overlapped proteins identified in both N-glycoproteomics and proteomics. In other words, only 60.2% (542/901) of all identified glycosylated proteins in N-glycoproteomics were also identified in the

proteomics. In fact, relatively low abundance of some glycosylated proteins, such as surface receptors and secreted proteins, make it difficult to detect them in the proteomics. This is supported by the results that the intensity of deglycoside proteins only identified in the N-glycoproteomics data is obviously and generally lower than that of corresponding proteins also identified in the global proteomics data (Figure 5D).

N-glycoproteomics in this study revealed that a total of 716 differentially expressed N-glycosylation sites (with a cut-off change of 1.5-fold and p-value < 0.05) in 441 proteins, in which 512 N-glycosylation sites in 326 glycoproteins were observed significantly upregulated and 204 N-glycosylation sites in 115 glycoproteins were observed significantly down-regulated (Figure 5E). Differential abundances of PTM's peptides represent changes in PTM's status and corresponding protein expression affected by PTM. Thus, comprehensive PTM changes require normalization by total protein expression changes to exclude that differential PTMs are attributed to changes in total protein expression. So, we mainly focused on the glycoproteins quantified in both N-glycoproteomics and proteomics to investigate the effect of protein normalization on glycosylation change. We evaluated the quantitative N-glycoproteomic ratios of 1107 N-glycosites in 542 proteins which were normalized by corresponding protein quantity. There is a significant correlation between the N-glycosite ratios and protein ratios (Pearson $r=0.84$), demonstrating that changes of N-glycosylation modification of a protein are probably a reflection of protein abundance change. A total of 264 differentially expressed N-glycosylation sites in 180 proteins were identified, in which 243 N-glycosylation sites in 162 glycoproteins were observed significantly upregulated and 21 N-glycosylation sites in 18 glycoproteins were observed significantly down-regulated. Differentially expressed glycoproteins and glycosylation sites were listed in **Table S3**. Twenty-four glycosylation sites of 20 proteins were consistently upregulated in all samples and nineteen glycosylation sites of 17 proteins were consistently down-regulated in more than four samples were shown in Figure 5F. Among them, procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 (PLOD2), collagen alpha-3(VI) chain (COL6A3), peptidyl-prolyl cis-trans isomerase (FKBP10) and Tenascin (TNC) have two or more differential glycosylation sites in all samples. The number of differentially expressed N-glycosylation sites that were identified in different individuals of both experiments was exhibited in Figure 5G.

Characteristics of the N-glycoproteins: functional analysis

Subcellular localization analysis revealed that the majority differentially expressed glycoproteins (fold change > 1.5 and p value < 0.05), whether up-regulated or down-regulated, distributed in the extracellular region (53.6%), plasma membrane (22.4%) and endoplasmic reticulum (10.6%), while few glycoproteins were localized in cytoplasm (2.2%) and nucleus (3.4%). It is highly consistent with the property of glycoproteins, completely different from the protein localization in proteomics (Figure 6A). Then we performed enrichment analysis of differentially expressed glycoproteins to identify significantly Gene Ontology terms and KEGG pathways and to investigate their functions. Most up-regulated glycoproteins were enriched in cell-substrate-related metabolic processes including collagen metabolic process, receptor metabolic process, protein activation cascade. For the KEGG pathway enrichment, some known glycosylation-affected cancer-associated pathways including ECM-receptor interaction, focal adhesion,

phagosome and lysosome were enriched in up-regulated glycoproteins. Interestingly, some amino acid metabolism-related pathways including lysine degradation, tyrosine metabolism, phenylalanine metabolism, valine, leucine and isoleucine degradation were obviously glycosylated and the levels of according glycoproteins were significantly upregulated in ESCC (Figure 6B). Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 (PLOD2), which catalyzes the hydroxylation of lysyl residues in collagen-like peptides, was identified with three N-glycosites (N63, N209 and N696). Other up-regulated glycoproteins related to amino acid metabolism included procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1 and 3 (PLOD1 and PLOD3), amine oxidase copper containing 3 (AOC3), interleukin 4 induced 1 (IL4I1), 3-hydroxy-3-methylglutaryl- CoA synthase 1 (HMGCS1) and so on. In contrast, these up-regulated pathways in N-glycoproteomics were found to be down-regulated in proteomics, indicating these pathways can be actually regulated by glycosylation modification in ESCC. Many down-regulated terms were related to peptidase regulator activity. The KEGG pathway enrichment result showed ECM-receptor interaction, focal adhesion and PI3K-Akt signaling pathway as negatively enriched (Figure 6C). Differential glycosylated proteins and differential proteins in some amino acid metabolism-related pathways are marked in Figure 6D.

Stage progression related dynamic clusters analysis in ESCC

As the lymph node status is one of the main factors affecting the clinical stage and survival of ESCC patients, we were interested in the dynamic process of glycoprotein expression during ESCC development, especially in lymph node metastasis (LNM). To investigate the types of dynamic change, we performed dynamic clusters analysis on the basis of glycoprotein quantity. We captured a series of glycoprotein clusters whose fold change increased or decreased as cancer progresses with lymph node metastasis.

For glycoprotein clusters whose fold change increased or decreased as cancer progression, the fifteen clusters were detected. Clusters #3, #7 and #9 represented a series of glycoproteins whose expression quantity showed an increasing or decreasing trend (Figure 7A). The enrichment pathways and interaction networks of these glycoproteins were exhibited in Figure 7B and 7C. And the glycoproteins of Clusters #9 were enriched in HIF-1 signaling pathway (INSR, IGF1R, TFRC), RAP1 signaling pathway (INSR, IGF1R, ITGB3), and lysine degradation pathway (PLOD3 and COLGALT1), indicating that these pathways might play an important role in the LNM process in ESCC. Furthermore, the glycoproteins of Clusters #3 and #7 were enriched in Cell adhesion molecules (CAMs), complement and coagulation cascades, and leukocyte trans-endothelial migration. Decreased glycosylation levels of CAMs may promote tumor cells to detach from the primary tumor, then facilitate tumor invasion and metastasis; and for the leukocyte migration and complement system, low glycosylation levels may affect their immune surveillance and clearance capabilities for tumors.

Discussion

Aberrant protein glycosylation is well known to be associated with the occurrence and development of cancer including cell transformation, invasion and metastasis[14]. Glycans are usually attached to proteins on the cell membrane or in the extracellular matrix. N-linked glycans are the addition of an oligosaccharide chain to an asparagine (Asn) residue within an amino acid sequence Asn-X-Ser/Thr (X should not be proline), and O-linked glycans are the transfer to the oxygen atom of the serine(Ser), hydroxyl (Tyr) or threonine(Thr) residues [15]. Although great progresses have recently been made to identify the glycoproteins and glycosylation patterns as potential biomarkers for various types of cancers, this study is the first unbiased analysis of N-glycoproteomics study in ESCC.

As expected, most of the changes in glycoproteins were accompanied by changes in according protein expression. After normalization by protein expression changes, a total of 901 glycoproteins were identified and quantified from ESCC tissues. According to the localization, the identified glycoproteins were mostly located in the subcellular compartments, consistent with the sites of the N-glycoprotein formation process. Regarding enriched related signaling pathways, the glycoproteins involved in the ECM-receptor interaction, focal adhesion (such as laminin, integrin and fibronectin family) and lysosome (e.g. LAMP1, GNS, LGMN, MAN2B1) were largely enriched which was consistent with previous reports[9, 16]. Interestingly, the proteins involved in amino acid metabolism were obviously upregulated and largely enriched. PLOD2, PLOD3 and PLOD1, which belong to lysyl hydroxylase and catalyze collagen lysine hydroxylation, were found to be up-regulated glycoproteins in all cases of ESCC. Although increased glycosylation of these proteins was only also observed in colorectal cancer [17], the detailed mechanism and impact of glycosylation in cancer remained elucidated. Glycosylation of HMGCS1 and AOC3 proteins has not been reported in cancer. HMGCS1 mediates the mevalonate pathway, ketogenesis [18] and is involved in cholesterol biosynthesis [19]. AOC3 catalyzes the conversion of primary amines into aldehydes [20]. In this study, AOC3 protein was up-regulated in the N-glycosylation level at Asn137 and down-regulated in its protein expression, indicating that an overall increase of N-glycosylation occupancy of AOC3 has an important impact on the biological function. All these results supported the well-known experience that glycosylation participated in cellular metabolism in the onset and progression of ESCC [21].

The majority of differential glycoproteins with multiple glycosylation sites were laminin, integrin, collagen, fibronectin associated proteins. Among differential glycoproteins in all samples, lysosome-associated membrane glycoprotein 1 (LAMP1) has been reported to be a pro-invasive factor in cancer progression through abnormal localization on the plasma membrane of cancer cells [22]. On one hand, this could be a membrane repair mechanism through lysosome fusion and exocytosis [23]. On the other hand, LAMP1 localization on the plasma membrane provided the binding ability to E-selectin through sialyl-LeX residues, thereby promoting the cancer cells adhering to extracellular matrix [24]. LAMP1 over-expression could also influence cancer development inside the lysosomal membrane through increasing lysosomal exocytosis and/or lysosomal size [25, 26]. In this study, N-glycosylation level of LAMP1 was extremely up-regulated but protein expression level was unchanged, indicating it is glycosylation modification but not expression change affects the biological function. Whether glycosylation of LAMP1 affects localization and how the six differential glycosylation sites of LAMP1 (Asn84, 103, 261, 76, 62, 249) in this study work

in ESCC needs further study. Transferrin receptors encoded by TFRC is a membrane glycoprotein, which can import iron by binding transferrin. TFRC displayed moderate to strong cytoplasmic expression in various cancer tissues and expression increased with increasing cancer stage [27]. TFRC knockdown decreased intracellular total iron, suppressed tumor growth and metastasis in human and mouse mammary adenocarcinomas [27]. Matrix remodeling associated 5 (MXRA5) was a novel extracellular protein that was also upregulated in several types of cancers [28]. MXRA5 was identified to be frequently mutated in non-small cell lung carcinoma and the high MXRA5 expression was correlated with tumor progression [29, 30]. However, there are no previous studies that focus on the role of the N-glycosylation modification of this protein in cancer including ESCC.

Lymphatic metastasis (LNM) related dynamic analysis captured 15 glycoprotein clusters whose fold change increased or decreased as cancer progresses with LNM. And the glycoproteins that increased with LNM status were enriched in HIF-1 signaling pathway, RAP1 signaling pathway and lysine degradation pathway. As is well known, HIF-1 signaling is a classical oncogenic pathway associated with tumor growth, angiogenesis, metastasis, and mortality. Recent studies showed HIF-1 signaling was involved in lymphatic invasion through induction of platelet-derived growth factor B (PDGF-B) in breast cancer[31] and vascular endothelial growth factor C (VEGF-C)[32] and SP1 in ESCC[33]. INSR and IGF1R belong to Insulin/IGF System, which was known to affect the malignant behavior of cancer cells and was regulated by Glycans[34]. Increased INSR/IGF1R were related with LNM in cancers[35]. Inhibition of N-linked glycosylation impaired the receptors (INSR and IGF1R) glycosylation and reduced their abundance at the cell surface[36]. As an important cell adhesion molecular, ITGB3 expression has been reported to be associated with LNM in several types of cancers[37] and its N-glycosylation was required for the attachment of cells to ECM[38]. RAP1 pathway is also an important regulator of cell adhesions and junctions, cellular migration, and polarization[39]. RAP1 expression was reported to be correlated with LNM in ESCC and its knockdown decreased the migratory and invasive capacities via AKT signaling in ESCC cells[40]. The role of lysine degradation pathway in cancer was rarely reported. In colorectal cancer, thrombopoietin (TPO) was reported to promote self-renewal and metastasis of CD110+ tumor-initiating cells (TICs) to the liver by activating lysine degradation[41]. In our study, the two genes involved in lysine degradation, PLOD3 and COLGALT1, were involved in collagen glycosylation. PLOD3 was upregulated in some types of cancers and promoted tumor malignant progression[42–46]. COLGALT2 was also reported to be overexpressed in metastatic ovarian cancer and interacts with PLOD3[47].

Conclusion

Our study revealed a series of differentially expressed glycoproteins and signaling pathways in ESCC tissues compared with paired normal tissues. Especially we found several metastasis-related glycoproteins and signaling pathways, which might be potential therapeutic targets for the prevention and treatment of ESCC patients with LN metastasis.

Abbreviations

ESCC: Esophageal squamous cell carcinoma; SCC: Squamous cell carcinoma antigen; CEA: Carcinoembryonic antigen; CYFRA21-1: Cytokeratin-19-fragment; Fut-Hpt Fucosylated haptoglobin; MS/MS: Tandem mass spectrometry; AGC: Automatic gain control; FDR: False positive rate; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; FCM: Fuzzy c-means algorithm; PTM: Post-transcriptional modification; PLOD2: Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2; COL6A3: Collagen alpha-3(VI) chain; FKBP10: peptidyl-prolyl cis-trans isomerase; TNC: Tenascin; PLOD1: Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1; PLOD3: Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3; ACO3: Amine oxidase copper containing 3; IL4I1: Interleukin 4 induced 1; HMGCS1: 3-hydroxy-3-methylglutaryl- CoA synthase 1; LMN: Lymph node metastasis; CAMs: Cell adhesion molecules; MXRA5: Matrix remodeling associated 5; LAMP1: Lysosome-associated membrane glycoprotein 1; TFRC: Transferrin receptor protein 1; PDGF-B: Platelet-derived growth factor B; VEGF-C: Vascular endothelial growth factor C; TICs: Tumor-initiating cells (TICs).

Declarations

Acknowledgements

We would like to thank Jingjie PTM Biolabs (Hangzhou) for performing proteomics and N-glycoproteomics analysis. We are grateful to Professor Xiuqian Mu (Department of Ophthalmology/Ross Eye Institute, University at Buffalo, Buffalo, NY, USA.) for helpful discussions.

Authors' Contributions

CYP and CXL designed the study. ZL, GYZ and CHY integrated and analyzed the data. CYP, CXL and ZL obtained funding. SLY, DTY and YX prepared the figures. ZL and KPZ wrote the manuscript. MYC and GYL helped in the process of sample sorting. All authors read and approved the final manuscript

Funding

This research was funded by the National Natural Science Foundation of China (81773150, 81972613, 82072746), the Fund for Shanxi "1331 Project", the funding for "Sanjin Scholars", and the Program for the Outstanding Innovative Teams of Higher Learning Institutions of Shanxi (OIT 2017). Research Project Supported by Shanxi Scholarship Council of China (2021-082).

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the Shanxi Medical University (Approval No. 2017LL037).

Consent for publication

Not applicable.

Competing interests

References

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J: **Cancer statistics in China, 2015.** *CA Cancer J Clin* 2016, **66**:115-132.
2. Zheng X, Xing S, Liu XM, Liu W, Liu D, Chi PD, Chen H, Dai SQ, Zhong Q, Zeng MS, Liu WL: **Establishment of using serum YKL-40 and SCCA in combination for the diagnosis of patients with esophageal squamous cell carcinoma.** *BMC Cancer* 2014, **14**:490.
3. Fatehi Hassanabad A, Chehade R, Breadner D, Raphael J: **Esophageal carcinoma: Towards targeted therapies.** *Cell Oncol (Dordr)* 2020, **43**:195-209.
4. Peixoto A, Relvas-Santos M, Azevedo R, Santos LL, Ferreira JA: **Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks.** *Front Oncol* 2019, **9**:380.
5. Ohtsubo K, Marth JD: **Glycosylation in cellular mechanisms of health and disease.** *Cell* 2006, **126**:855-867.
6. Silsirivanit A: **Glycosylation markers in cancer.** *Adv Clin Chem* 2019, **89**:189-213.
7. Miyoshi E, Kamada Y: **Application of glycoscience to the early detection of pancreatic cancer.** *Cancer Sci* 2016, **107**:1357-1362.
8. Burvenich IJ, Lee FT, O'Keefe GJ, Makris D, Cao D, Gong S, Rigopoulos A, Allan LC, Brechbiel MW, Liu Z, et al: **Engineering anti-Lewis-Y hu3S193 antibodies with improved therapeutic ratio for radioimmunotherapy of epithelial cancers.** *EJNMMI Res* 2016, **6**:26.
9. Raghunathan R, Sethi MK, Klein JA, Zaia J: **Proteomics, Glycomics, and Glycoproteomics of Matrisome Molecules.** *Mol Cell Proteomics* 2019, **18**:2138-2148.
10. Futschik ME, Carlisle B: **Noise-robust soft clustering of gene expression time-course data.** *J Bioinform Comput Biol* 2005, **3**:965-988.
11. Kumar L, M EF: **Mfuzz: a software package for soft clustering of microarray data.** *Bioinformatics* 2007, **2**:5-7.
12. Song Y, Wang J, Cheng Z, Gao P, Sun J, Chen X, Chen C, Wang Y, Wang Z: **Quantitative global proteome and lysine succinylome analyses provide insights into metabolic regulation and lymph node metastasis in gastric cancer.** *Sci Rep* 2017, **7**:42053.
13. Sun S, Zhang H: **Identification and Validation of Atypical N-Glycosylation Sites.** *Anal Chem* 2015, **87**:11948-11951.
14. Hakomori S: **Glycosylation defining cancer malignancy: new wine in an old bottle.** *Proc Natl Acad Sci U S A* 2002, **99**:10231-10233.
15. Yang S, Hu Y, Sokoll L, Zhang H: **Simultaneous quantification of N- and O-glycans using a solid-phase method.** *Nat Protoc* 2017, **12**:1229-1244.
16. Munkley J, Mills IG, Elliott DJ: **The role of glycans in the development and progression of prostate cancer.** *Nat Rev Urol* 2016, **13**:324-333.

17. Nicastrì A, Gaspari M, Sacco R, Elia L, Gabriele C, Romano R, Rizzuto A, Cuda G: **N-glycoprotein analysis discovers new up-regulated glycoproteins in colorectal cancer tissue.** *J Proteome Res* 2014, **13**:4932-4941.
18. Talman V, Teppo J, Poho P, Movahedi P, Vaikkinen A, Karhu ST, Trost K, Suvitaival T, Heikkonen J, Pahikkala T, et al: **Molecular Atlas of Postnatal Mouse Heart Development.** *J Am Heart Assoc* 2018, **7**:e010378.
19. Liu S, Jing F, Yu C, Gao L, Qin Y, Zhao J: **AICAR-Induced Activation of AMPK Inhibits TSH/SREBP-2/HMGCR Pathway in Liver.** *PLoS One* 2015, **10**:e0124951.
20. Yang H, Ralle M, Wolfgang MJ, Dhawan N, Burkhead JL, Rodriguez S, Kaplan JH, Wong GW, Haughey N, Lutsenko S: **Copper-dependent amino oxidase 3 governs selection of metabolic fuels in adipocytes.** *PLoS Biol* 2018, **16**:e2006519.
21. Pinho SS, Reis CA: **Glycosylation in cancer: mechanisms and clinical implications.** *Nat Rev Cancer* 2015, **15**:540-555.
22. Alessandrini F, Pezze L, Ciribilli Y: **LAMPs: Shedding light on cancer biology.** *Semin Oncol* 2017, **44**:239-253.
23. Corrotte M, Castro-Gomes T, Koushik AB, Andrews NW: **Approaches for plasma membrane wounding and assessment of lysosome-mediated repair responses.** *Methods Cell Biol* 2015, **126**:139-158.
24. Agarwal AK, Srinivasan N, Godbole R, More SK, Budnar S, Gude RP, Kalraiya RD: **Role of tumor cell surface lysosome-associated membrane protein-1 (LAMP1) and its associated carbohydrates in lung metastasis.** *J Cancer Res Clin Oncol* 2015, **141**:1563-1574.
25. Gotink KJ, Rovithi M, de Haas RR, Honeywell RJ, Dekker H, Poel D, Azijli K, Peters GJ, Broxterman HJ, Verheul HM: **Cross-resistance to clinically used tyrosine kinase inhibitors sunitinib, sorafenib and pazopanib.** *Cell Oncol (Dordr)* 2015, **38**:119-129.
26. Machado E, White-Gilbertson S, van de Vlekkert D, Janke L, Moshiah S, Campos Y, Finkelstein D, Gomero E, Mosca R, Qiu X, et al: **Regulated lysosomal exocytosis mediates cancer progression.** *Sci Adv* 2015, **1**:e1500603.
27. Takahashi M, Shibutani M, Woo GH, Inoue K, Fujimoto H, Igarashi K, Kanno J, Hirose M, Nishikawa A: **Cellular distributions of molecules with altered expression specific to the tumor promotion process from the early stage in a rat two-stage hepatocarcinogenesis model.** *Carcinogenesis* 2008, **29**:2218-2226.
28. Robins JE, Capehart AA: **Matrix remodeling associated 5 expression in trunk and limb during avian development.** *Int J Dev Biol* 2018, **62**:335-340.
29. Xiong D, Li G, Li K, Xu Q, Pan Z, Ding F, Vedell P, Liu P, Cui P, Hua X, et al: **Exome sequencing identifies MXRA5 as a novel cancer gene frequently mutated in non-small cell lung carcinoma from Chinese patients.** *Carcinogenesis* 2012, **33**:1797-1805.
30. He Y, Chen X, Liu H, Xiao H, Kwapong WR, Mei J: **Matrix-remodeling associated 5 as a novel tissue biomarker predicts poor prognosis in non-small cell lung cancers.** *Cancer Biomark* 2015, **15**:645-651.

31. Schito L, Rey S, Tafani M, Zhang H, Wong CC, Russo A, Russo MA, Semenza GL: **Hypoxia-inducible factor 1-dependent expression of platelet-derived growth factor B promotes lymphatic metastasis of hypoxic breast cancer cells.** *Proc Natl Acad Sci U S A* 2012, **109**:E2707-2716.
32. Katsuta M, Miyashita M, Makino H, Nomura T, Shinji S, Yamashita K, Tajiri T, Kudo M, Ishiwata T, Naito Z: **Correlation of hypoxia inducible factor-1alpha with lymphatic metastasis via vascular endothelial growth factor-C in human esophageal cancer.** *Exp Mol Pathol* 2005, **78**:123-130.
33. Hu X, Lin J, Jiang M, He X, Wang K, Wang W, Hu C, Shen Z, He Z, Lin H, et al: **HIF-1alpha Promotes the Metastasis of Esophageal Squamous Cell Carcinoma by Targeting SP1.** *J Cancer* 2020, **11**:229-240.
34. de-Freitas-Junior JCM, Andrade-da-Costa J, Silva MC, Pinho SS: **Glycans as Regulatory Elements of the Insulin/IGF System: Impact in Cancer Progression.** *Int J Mol Sci* 2017, **18**.
35. Sun J, Lu Z, Deng Y, Wang W, He Q, Yan W, Wang A: **Up-regulation of INSR/IGF1R by C-myc promotes TSCC tumorigenesis and metastasis through the NF-kappaB pathway.** *Biochim Biophys Acta Mol Basis Dis* 2018, **1864**:1873-1882.
36. Klaver E, Zhao P, May M, Flanagan-Steet H, Freeze HH, Gilmore R, Wells L, Contessa J, Steet R: **Selective inhibition of N-linked glycosylation impairs receptor tyrosine kinase processing.** *Dis Model Mech* 2019, **12**.
37. Meves A, Nikolova E, Heim JB, Squirewell EJ, Cappel MA, Pittelkow MR, Otley CC, Behrendt N, Saunte DM, Lock-Andersen J, et al: **Tumor Cell Adhesion As a Risk Factor for Sentinel Lymph Node Metastasis in Primary Cutaneous Melanoma.** *J Clin Oncol* 2015, **33**:2509-2515.
38. Singh C, Shyanti RK, Singh V, Kale RK, Mishra JPN, Singh RP: **Integrin expression and glycosylation patterns regulate cell-matrix adhesion and alter with breast cancer progression.** *Biochem Biophys Res Commun* 2018, **499**:374-380.
39. Zhang YL, Wang RC, Cheng K, Ring BZ, Su L: **Roles of Rap1 signaling in tumor cell migration and invasion.** *Cancer Biol Med* 2017, **14**:90-99.
40. Li Q, Xu A, Chu Y, Chen T, Li H, Yao L, Zhou P, Xu M: **Rap1A promotes esophageal squamous cell carcinoma metastasis through the AKT signaling pathway.** *Oncol Rep* 2019, **42**:1815-1824.
41. Wu Z, Wei D, Gao W, Xu Y, Hu Z, Ma Z, Gao C, Zhu X, Li Q: **TPO-Induced Metabolic Reprogramming Drives Liver Metastasis of Colorectal Cancer CD110+ Tumor-Initiating Cells.** *Cell Stem Cell* 2015, **17**:47-59.
42. Xie D, Li J, Wei S, Qi P, Ji H, Su J, Du N, Zhang X: **Knockdown of PLOD3 suppresses the malignant progression of renal cell carcinoma via reducing TWIST1 expression.** *Mol Cell Probes* 2020, **53**:101608.
43. Wang B, Xu L, Ge Y, Cai X, Li Q, Yu Z, Wang J, Wang Y, Lu C, Wang D, et al: **PLOD3 is Upregulated in Gastric Cancer and Correlated with Clinicopathologic Characteristics.** *Clin Lab* 2019, **65**.
44. Baek JH, Yun HS, Kwon GT, Lee J, Kim JY, Jo Y, Cho JM, Lee CW, Song JY, Ahn J, et al: **PLOD3 suppression exerts an anti-tumor effect on human lung cancer cells by modulating the PKC-delta signaling pathway.** *Cell Death Dis* 2019, **10**:156.

45. Baek JH, Yun HS, Kwon GT, Kim JY, Lee CW, Song JY, Um HD, Kang CM, Park JK, Kim JS, et al: **PLOD3 promotes lung metastasis via regulation of STAT3.** *Cell Death Dis* 2018, **9**:1138.
46. Tsai CK, Huang LC, Tsai WC, Huang SM, Lee JT, Hueng DY: **Overexpression of PLOD3 promotes tumor progression and poor prognosis in gliomas.** *Oncotarget* 2018, **9**:15705-15720.
47. Guo T, Li B, Kang Y, Gu C, Fang F, Chen X, Liu X, Lu G, Feng C, Xu C: **COLGALT2 is overexpressed in ovarian cancer and interacts with PLOD3.** *Clin Transl Med* 2021, **11**:e370.

Figures

Figure 1

Schematic representation of the N-glycoproteomics and proteomics approach for the discovery of differential expressed glycoproteins and proteins in ESCC tissues compared with matched normal tissue.

Figure 2

Identification of global proteins in proteomics analysis. (A) Principal component analysis of protein quantity in proteomics data. (B) Pearson correlation coefficient heat map of protein quantification. (C) The boxplot of RSD distribution. (D) Basic statistical graph of mass spectrometry data.

Figure 3

Functional analysis of global differential expressed proteins in proteomics analysis. (A) The volcano plot of differentially-expressed proteins. (B) The heat map of differentially expressed protein in all paired comparisons. (C) Cellular component enrichment analysis of up-regulated (left) and down-regulated proteins. (D) KEGG enrichment analysis of up-regulated (left) and down-regulated proteins.

Figure 4

Motif analysis of the identified glycosylation peptides. (A) Statistic analysis of the whole identified peptide length in N-glycoproteomics analysis in two experiments. (B) The mass error of the whole identified peptides in N-glycoproteomics analysis. (C) Principal component analysis of glycoproteins quantity. (D) Pearson correlation coefficient heat map of glycoproteins quantification. (E) Sequence logo

of glycosylation motifs. (F) Heatmap of amino acid frequencies of the sequences flanking N-glycosylation sites.

Figure 5

Comparative analyses of N-glycoproteomics and proteomics between ESCC and paired normal tissue. (A) Number of N-glycosylation sites identified per protein in N-glycoproteomics analysis. (B) Overview of quantified proteins and N-glycoproteins in different stage of ESCC. (C) The overlap of the proteins identified in both proteomics and N-glycoproteomics. (D) Intensity distribution of all glycosites only identified in the N-glycoproteomics data (red bars) overlaid with that of N-glycosites whose corresponding proteins were also identified in the proteomics data (blue bars). (E) Log2 ratio distributions for N-glycosites analyzed, the value greater than 0.584 or less than -0.584 were considered to be significant. (F) Heat map of 37 common proteins that exhibited significant up-regulation in all eight comparisons and down-regulation in more than four comparisons. Color represents the expression levels of proteins, among which red color denotes the high expression while blue color denotes the low expression level. (G) Number of N-glycosylation sites identified in different samples.

Figure 6

Identification and functional analysis of global normalized N-glycoproteomics in ESCC. (A) The comparison of the global proteomics and N-glycoproteomics in subcellular locations. (B) GO and KEGG enrichment analysis of up-regulated glycosylated proteins. (C) GO and KEGG enrichment analysis of down-regulated glycosylated proteins. (D) Differential expressed glycoproteins and proteins associated with amino acid metabolism-related pathways. Red box represents upregulated glycosylated proteins in N-glycoproteomics data. Blue font represents downregulated proteins in proteomics data.

Figure 7

Enrichment-based dynamic cluster analysis of glycoprotein expression profile in ESCC development. (A) A total of 15 dynamic clusters analyses in the process of LNM. The abscissa represents the samples and the ordinate represents the quantitative values of each sample in log2 conversion. (B) KEGG enrichment analysis of increased and decreased fold-change glycoprotein dynamic clusters. (C) IPA analysis of the genes associated with cluster #3, #7 and #9.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)