74816_Auto_Edited.docx

Basic Study

N-linked glycoproteomic profiling in esophageal squamous cell carcinoma

Qi-Wei Liu, Hao-Jie Ruan, Wei-Xia Chao, Meng-Xiang Li, Ye-Lin Jiao, Douglas G. Ward, She-Gan Gao, Yi-Jun Qi

Abstract

BACKGROUND

Mass spectrometry-based proteomics and glycomics reveal post-translational modifications providing significant biological insights beyond the scope of genomic sequencing.

AIM

To characterize the N-linked glycoproteomic profile in esophageal squamous cell carcinoma *via* two complementary approaches.

METHODS

Using tandem multilectin affinity chromatography for enrichment of N-linked glycoproteomes, we performed N-linked glycoproteomic profiling in esophageal squamous cell carcinoma (ESCC) tissues by 2-dimensional gel electrophoresis (2-DE)-based and iTRAQ-based mass spectrometry quantitation in parallel, followed by validation of candidate glycoprotein biomarkers by Western blotting.

RESULTS

Two-DE-based and iTRAQ-labeling-based quantitation identified 24 and 402 differentially expressed N-linked glycoproteins, respectively, with 15 in common, demonstrating the outperformance of iTRAQ-labeling-based quantitation over 2-DE and complementarity of these 2 approaches. Proteomaps showed the distinct compositions of functional categories between proteins and glycoproteins with differential expression associated with ESCC. Western blotting validated the up-regulation of total procathepsin D and high-mannose procathepsin D, and the down-regulation of total haptoglobin, high-mannose clusterin, and GlcNAc/sialic acid-containing fraction of 14-3-3ζ in ESCC tissues. The serum levels of glycosylated fractions of clusterin, PRELP and haptoglobin in patients with ESCC were remarkably higher than those in healthy controls.

CONCLUSION
Our study provides insights into the aberrant N-linked glycoproteome associated with ESCC, which will be a valuable resource for future investigations.

Key Words: Esophageal squamous cell carcinoma; N-linked glycoprotein; Post-translational modification; Lectin; Cathepsin D; Haptoglobin; 14-3-3ζ

Core Tip: Comprehensive N-linked glycoproteome was profiled by 2-dimensional gel electrophoresis-based and iTRAQ labeling-based MS quantitation in parallel after N-linked glycoproteins enrichment by a tandem of multilectin affinity chromatography. The iTRAQ-based quantitative proteomic profiling outperformed protein spot intensity quantification used by 2-dimensional gel electrophoresis. A total of 411 N-linked glycoproteins were identified, including 128 up-regulated and 283 down-regulated

glycoproteins with differential expression, which provides the scientific community with a dataset of glycoprotein associated with ESCC for in-depth investigation.

INTRODUCTION

Esophageal cancer is a severe global health issue with 572,000 new cases and 509,000 deaths in 2018, half of which occur in China [1]. The two most common histological subtypes of esophageal cancer are esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma, with ESCC being the predominant histological subtype both in China and worldwide [2, 3]. The long-term outcome of ESCC is correlated with clinical stage at diagnosis with the best outcome in the early stages [4-6]. Unfortunately, the current endoscopic screening for ESCC suffers from low patient compliance and high-cost despite the fact that an early detection rate of 70% was achieved in a high-risk region in northern central China [7]. Furthermore, serum biomarkers for ESCC, including squamous cell carcinoma antigen, carcinoembryonic antigen, CYFRA21-1, and carbohydrate antigen 19-9, lack sufficient sensitivity and specificity for early detection [8-11]. The overall five-year survival rate ranges from 15% to 25% because most cases of ESCC present at an advanced stage [6, 12]. It is well accepted that novel biomarkers would hold great promise to improve the clinical outcome of ESCC.

Recent large-scale "omics" studies in ESCC have identified a myriad of aberrations at the levels of genome, epigenome, transcriptome, proteome, *etc.*, revealing the high molecular heterogeneity of ESCC [13-20]. Beyond the scope of other techniques, mass spectrometry (MS)-based proteomics has the ability to measure post-translational modifications that provide additional significant biological insights, for example, therapeutically targetable signaling proteins and pathways [21, 22]. To date, comprehensive glycoproteomic studies of ESCC are lacking. Glycosylation is present on more than 50% of human proteins and aberrant glycosylation has been implicated in the development and progression of various cancers [23-32]. The sugar chains of glycoproteins are involved in numerous physiological and pathological conditions including cell growth, adhesion, differentiation, migration, signal regulation, immune responses, and tumor invasion [23, 33, 34]. Notably, the clinical prognosis of cancer has been correlated with certain structures of glycans attached to proteins [35, 36]. The bisecting GlcNAc structure catalyzed by N-Acetylglucosaminyltransferase IIIplays

crucial roles in suppression of cancer metastasis, which contrasts with the tumor-promoting role of β1,6-branched N-glycan catalyzed by N-Acetylglucosaminyltransferase V [37]. It has been reported that E-cadherin modified by complex N-glycans exhibits weakened adherens junctions, whereas high mannose or less N-glycans attached to E-cadherin produces stable adherens junctions [38]. Furthermore, N-Acetylglucosaminyltransferase IIImodified α5β1 integrin suppresses cell spreading, migration and focal adhesion kinase activity [39]. The fraction of ConA-binding procathepsin D in hepatocellular carcinoma serum is significantly increased and shows a sensitivity of 85% and a specificity of 80% for hepatocellular carcinoma diagnosis [30]. Not surprisingly, over 50% of current cancer biomarkers are glycoproteins, such as AFP, AFP-L3, CA19-9, CA125 and CEA [40-43].

In this study, we used MS-based proteomic analysis to identify N-linked glycoproteins associated with ESCC after isolation of N-linked glycoproteins using tandem multilectin affinity chromatography. Our study unraveled a comprehensive landscape of N-linked glycosylation dysregulation associated with ESCC and bioinformatics analysis provided insights into the functional significance of these aberrant glycosylation. The glycosylated fractions of procathepsin D, clusterin, 14-3-3ζ, proline-arginine-rich end leucine-rich repeat protein (PRELP) and haptoglobin may play important roles in malignant progression of ESCC and are potential diagnostic biomarkers for ESCC, suggesting that N-linked glycosylation aberrations may promote the malignant progression of ESCC.

MATERIALS AND METHODS

Clinical samples

All patients provided written informed consent, and the study was approved by the Ethics Committee of the Medical School, Henan University, China (ethics ref: 108) and conducted in accordance with the ethical guidelines of the 1975 Declaration of Helsinki. Forty-seven pairs of resected ESCC and adjacent normal esophageal mucosa tissue samples, and 10 pre-operative serum samples were collected from Linzhou Cancer

Hospital, Henan, China, between 2010 and 2011, and stored in liquid nitrogen or at -70 °Cprior to protein extraction. Among the tissue samples, 15 pairs served as the discovery set and the other 32 pairs as the validation set. All tissues were histopathologically confirmed as normal esophageal mucosa or ESCC. The adjacent non-cancerous samples were at least 3 cm distant from the edge of neoplastic mass. No patient in this study had received preoperative radiotherapy or chemotherapy.

Isolation of N-linked glycoproteins by lectin affinity chromatography

Extraction of total tissue proteins and N-linked glycoproteins were performed as previously reported with modifications [30]. Briefly, the esophageal tissue was homogenized in lysis buffer (50 mmol/L Tris, 150 mmol/L NaCl, 1% NP-40, pH 7.2) supplemented with complete proteinase inhibitor cocktail (Roche Diagnostics GmbH, Mannheim, Germany) followed by centrifugation and supernatant collection. Enrichment of N-linked glycoproteins for 3 pairs of pooled protein samples from 15 ESCC and adjacent non-tumor tissues (5 samples for each pair) was performed by tandem lectin affinity chromatography. The multilectin affinity chromatography used a combination of three agarose-bound lectins, Con A (Vector Laboratories, AL-1003), lentil lectin (LCH, Vector Laboratories, AL-1003) and snowdrop lectin (GNA, Vector Laboratories, AL-1243), and a combination of wheat germ agglutitin lectin (WGA, Vector Laboratories, AL-X) and elderberry lectin (SNA, Vector Laboratories, AL-X) for the first and second affinity chromatography, respectively. Sixty milligrams of pooled proteins were diluted with binding buffer (30 mmol/L Bis-Tris, 150 mmol/L NaCl, 1 mmol/L $CaCl_2$, 1 mmol/L $MnCl_2$, pH 7.2) and incubated with the lectins in the first affinity chromatography columns at RT for 2h. After incubation, 200 mmol/L methyl-α-D-mannopyranoside was used for elution of high-mannose N-linked glycoproteins. The flowthrough of the first three-lectin column was subjected to the second chromatography for isolation of GlcNAc/sialic acid containing N-linked glycoproteins using 200 mmol/L N-Acetyl-D-glucosamine as elution buffer. The eluted samples were

desalted and concentrated by acetone precipitation followed by solubilization in rehydration buffer.

Two-dimensional gel electrophoresis and image analysis

Two-dimensional gel electrophoresis (2-DE) was performed as previously reported [30, 44]. Fifty micrograms of glycoproteins were used for the first-dimension isoelectrofocusing electrophoresis using IPG strips with a linear pH range of 3-10, followed by the second dimension separation using 15% SDS-PAGE gels. Silver stained gels were scanned with an ImageScanner. Image analysis was carried out with ImageMaster 2-D Elite software 4.01.

In-gel digestion and MS identification

The excised protein spots with differential expression were reduced with DTT, alkylated with iodoacetamide, dried in a speedvac, and digested with trypsin (12.5 μg/mL in 50 mmol/L ammonium bicarbonate). LC-MS/MS analysis of tryptic peptides was performed using a Maxis Impact Q-TOF mass spectrometer (BrukerDaltonics, Bremen, Germany). The MS/MS data were searched against the SwissProt human sequence database using MASCOT *via* Proteinscape (BrukerDaltonics, Bremen, Germany) and reverse database searching was used for estimating the false discovery rate. All proteins were identified by ≥ 2 unique peptides.

iTRAQ labeling and MS identification

Isobaric Tags for Relative and Absolute Quantification (iTRAQ) labeling of tryptic peptides was performed according to the protocol provided by the supplier. The iTRAQ labeled peptides derived from pooled protein samples of 10 ESCC and adjacent non-tumor tissues as previously described in duplicate[45] were separated into 16 fractions by mixed-mode anion exchange/reverse-phase chromatography. Each fraction was dried, dissolved in 0.1% formic acid and analyzed by LC-MS/MS using a Maxis Impact Q-TOF mass spectrometer (BrukerDaltonics, Bremen, Germany). Data were searched

against the SwissProt human sequence database using MASCOT and expression ratios calculated by WARP-LC *via* Proteinscape (BrukerDaltonics, Bremen, Germany). The false discovery rate was estimated using a reverse database search strategy. All proteins were identified by ≥ 2 unique peptides. The data have been deposited at https://www.iprox.cn/ with Project ID IPX0004371000.

Functional enrichment analysis

Gene-set enrichment analyses of differentially expressed glycoproteins (DEGs) and previously identified differentially expressed proteins (DEPs) [45] were performed using Enrichment Map, which organizes the enriched gene-sets into a network with links representing the overlap of member genes [46]. The gene-sets with P value < 0.001 and false discovery rate < 5% were selected for display in the enrichment map. To gain an insight into the functional makeup of DEGs and DEPs, Proteomaps was used to visualize the functional distribution of DEGs or DEPs in cellular processes, including genetic information processing, metabolism, cellular processes, signaling and others [47].

Western blotting

Total tissue proteins were extracted using radioimmunoprecipitation assay lysis buffer, and glycoproteins were isolated by lectin affinity chromatography. Total proteins or glycoproteins were separated by SDS polyacrylamide gel electrophoresis and then transferred to PVDF membrane. The blots were blocked in 5% non-fat milk, incubated with primary antibodies and then corresponding secondary antibodies and visualized using Supersignal West chemiluminescent substrate (Pierce Biotechnology, Rockford, IL, USA). Semi-quantitation of the band intensity was performed using image analysis software and corrected by reference to GAPDH. The antibodies used in this study were: haptoglobin (1:2000, 16665-1-AP, Proteintech), cathepsin D (1:2500, ab75852, Abcam), clusterin (1:5000, 12289-1-AP, Proteintech), SOD3 (1:1500, T1799, Epitomics), PRELP (1:2000, AP6665b, Abgent) and 14-3-3-ζ (1:2500, ab85268, Abcam).

Statistical analysis

All statistical analyses were performed with SPSS 24.0 software (SPSS, Chicago, IL, USA). Comparisons of quantitated protein band density between ESCC and adjacent non-cancerous tissue were done with paired Student *t* tests or Wilcoxon signed-rank sum tests. All statistical tests were 2-sided, and *P* < 0.05 was considered statistically significant.

RESULTS

N-linked glycoprotein profiling by 2-dimensional gel electrophoresis

This study utilized two rounds of multilectin affinity chromatography to isolate N-linked glycoproteins. The first multilectin affinity column comprised Con A, LCH and GNA lectins to isolate high-mannose N-linked glycoproteins, and the second affinity column included WGA and SNA lectins for GlcNAc/sialic acid-containing glycoprotein enrichment. 2-DE was used to separate the isolated N-linked glycoproteins. Figure 1A & 1B show the representative 2-DE images of high-mannose glycoproteins isolated from ESCC and non-cancerous tissues. An average of 742 ± 45 protein spots could be detected on each 2-DE gel analyzed by Image Master 2-D Elite. A total of 35 differential protein spots with ≥ 1.5 fold-changes in protein spot intensities (*P* < 0.05) were identified, including 20 up-regulated and 15 down-regulated protein spots in ESCC compared with corresponding non-cancerous tissues, respectively. For GlcNAc/sialic acid-containing glycoprotein profiling, the patterns of protein spots on 2-DE gels were markedly different compared with those of high-mannose glycoproteins (Figure 1C & 1D). There were 23 protein spots with differential expression, including 5 up-regulated and 18 down-regulated protein spots in ESCC compared to non-cancerous tissues.

These protein spots were identified by LC-SI-MS/MS of the trypsin-digested spots excised from preparative gels. The identities of proteins are shown in Table 1 & 2, in which 22 proteins and 23 proteins are derived from the first and second lectin affinity chromatography, respectively. There are 6 proteins in common between these 2 isolated fractions. In agreement with our previous reports [30, 44], many gel spots were identified

as the same protein, indicating the presence of different proteoforms due to post-translational modifications. Using NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc) to predict the glycosylation sites on proteins, 72.7% (16/22) of high-mannose N-linked glycoproteins and 56.5% (13/23) of GlcNAc/sialic acid-containing glycoproteins were found to have potential glycosylation sites, indicating some non-specific binding to lectins by non-glycosylated proteins.

N-linked glycoprotein profiling by iTRAQ labeling

To comprehensively characterize the N-linked glycoproteome in ESCC, we further performed iTRAQ-based analysis of the N-linked glycoproteome in ESCC and non-cancerous tissues. One pair of protein pools from 10 ESCC and matched non-cancerous tissues, respectively, were subjected to tandem lectin affinity chromatography enrichment, in-solution digestion with trypsin, iTRAQ labeling and LC-MS/MS identification in duplicate. In total, 1464 and 1298 proteins from the first affinity column were identified in the two technical replicates (Figure 2A, Supplementary Table S1 & S2). Using the cut-off criteria of ≥ 2 unique peptides and a fold-change of 1.5, 189 and 138 proteins with differential expression were identified from the first and second independent pools, respectively, with 65 proteins in common, resulting in identification of 262 differentially expressed proteins in total from the high-mannose binding lectin affinity chromatography, of which 203 (77.5%) of these proteins have predicted N-linked glycosylation sites (Figure 2B). For N-linked glycoproteins enriched by the downstream GlcNAc/sialic acid binding lectin chromatography, 2 independent iTRAQ-labeled pools produced 1174 and 936 proteins following isolation of high-mannose glycoproteins (Figure 2C, Supplementary Table S3 & S4). There were 237 of 1174 and 154 of 936 proteins with differential expression, respectively, with 74 proteins in common (Figure 2D). A total of 317 unique proteins, of which 246 (77.6 %) of these proteins are predicted to be N-linked glycoproteins, were identified from the GlcNAc/sialic acid binding lectin chromatography enrichment. There were 58

overlapping glycoproteins found between 262 high-mannose glycoproteins and 317 GlcNAc/sialic acid-containing glycoproteins.

Comparison of 2-dimensional gel electrophoresis-based and iTRAQ-based profiling

In this study, 2 protocols for identifying N-linked glycoproteins based on different quantitative methods yielded quite different results. There were only 11 high-mannose glycoproteins and 4 GlcNAc/sialic acid-containing glycoproteins in common between 2-DE-based and iTRAQ-based glycoproteomic profiling, respectively. Our results demonstrate that iTRAQ-based quantitative proteomic profiling outperformed protein spot intensity quantification used by 2-DE. There were, however, common glycoproteins found by both approaches. Therefore, the combination of these 4 fractions produced 411 unique N-linked glycoproteins. These findings indicate that a combination of these 2 complementary protocols could generate a more comprehensive landscape of N-linked glycoproteome implicated in ESCC because iTRAQ-based and 2-DE-based methods were able to isolate unique glycoproteins, thus providing a good representation of the N-linked glycoproteome.

Functional analysis of differentially expressed glycoproteins

To explore the biological function of the 411 DEGs associated with ESCC, KEGG pathway enrichment analysis was initially performed using Enrichment Map. Figures 3A to 3C show the 64 enriched pathways, which include complement and coagulation cascades, focal adhesion, ECM-receptor interaction, cellular interactions, immune response and infection, metabolism of fructose and mannose, glycolysis and gluconeogenesis, splicesome, and PI3K-Akt and HIF-1 signaling pathways. In addition, Proteomaps were constructed to visualize the composition of these DEGs in the 5 functional categories, which comprise genetic information processing, environmental information processing, cellular processes, metabolism, and organismal systems. Figures 4A to 4C show that the genetic information processing accounts for the largest component of the DEGs, with protein processing, translation, and transcription being

the biggest contributors to this functional category. In the category of signaling events, both the up-regulated and down-regulated DEGs were involved in PI3K-Akt, MAPK, Ras, Rap1, and HIF-1 signaling pathways (Figures 4A & 4B). The layouts of cellular process between up-regulated and down-regulated DEGs were different, with common components of exosome, lysosome, phagosome, cytoskeleton, and cell cycle, whereas the down-regulated DEGs implicated focal adhesion, tight junction, endocytosis, and p53 signaling pathway (Figure 4A & 4B). In glycolysis, there were more up-regulated DEGs than down-regulated DEGs. Notably, the complement and coagulation cascades in immune system include exclusively down-regulated DEGs (Figure 4A & 4B).

Functional comparison between glycoproteins and proteins with differential expression
Because there were only 82 overlapping proteins between DEGs and DEPs, the distinct subproteomes in each functional category may affect the structure of proteomaps. As shown in Figure 4, the compositions of proteomaps differ. In the category of genetic information processing, the chaperones, spliceosome and peptidases are the predominant components in DEGs whereas the spliceosome, proteasome, and ribosome are the main contributors in DEPs. Although there are some signaling pathways (PI3K/Akt, MAPK, HIF1, Ras, Rap1) in common in DEGs and DEPs, distinct components (neuroactive ligand-receptor interaction, calcium signaling pathway, GPT-binding signaling proteins, steroid hormone biosynthesis and cytokines in DEPs vs. NF kappa B pathway, FoxO pathway, ECM-receptor interaction, CD molecules, lectins and enzyme-linked receptors in DEGs) are present in environment interaction (Figure 4C & 4F). Along with these differences, distinct compositions of the other three categories were found as well, in agreement with uniquely characterized proteomes. Likewise, Enrichment Map analysis showed that marked differences were found between DEGs and DEPs (Figure 3C & 3F). In the same line, there are remarkable distinctions between DEGs and DEPs identified using iTRAQ-based quantitation of Indian ESCC samples (Data not shown) [20].

Validation of potential glycoprotein biomarkers

Western blotting was used to validate the differential expression of potential glycoprotein biomarkers for ESCC, including haptoglobin, cathepsin D, clusterin, superoxide dismutase 3 (SOD3), PRELP, and 14-3-3ζ. Consistent with the corresponding protein spots of 748 and 1001 on 2-DE images, haptoglobin expression in ESCC was significantly down-regulated in tumor compared with adjacent non-cancerous tissues (Figure 5A). In contrast, the expression of procathepsin D instead of mature cathepsin D was remarkably enhanced in ESCC compared with the corresponding non-cancerous tissues (Figure 5B). There were no significant differences in the expression levels of the other 4 selected proteins (Supplementary Figure 1A to 1D). These data suggests that the differential expression of glycoproteins identified in this study could be caused by up-regulation of protein expression, or up-regulation of N-linked glycosylation due to aberrant expression or activities of certain glycosyltransfereases.

To assess the contribution of high-mannose or GlcNAc/sialic acid components attached to the selected glycoproteins, enrichment of corresponding glycoproteins *via* glycan chain binding to specific lectins was performed followed by Western blot detection. Consistent with the up-regulation of procathepsin D protein, the high-mannose subtype of procathepsin D was pronouncedly increased in ESCC as well (Figure 5C). In contrast to the expression pattern of total clusterin protein, the high-mannose fraction of clusterin was markedly down-regulated in ESCC as compared with non-cancerous tissues (Figure 5C). However, the glycosylated fractions of SOD3, PRELP and haptoglobin did not show significant differences between ESCC and non-cancerous tissues after enrichment of high-mannose glycan structure of glycoproteins using lectin affinity chromatography. In agreement with the quantitative result of glycosylated 14-3-3ζ, the GlcNAc/sialic acid-containing fraction of 14-3-3ζ was strongly reduced in ESCC (Figure 5C).

Serum levels of potential glycoprotein biomarkers

As isolation of N-linked glycoproteins from serum reduces the high-complexity of the serum proteome, we characterized the serum levels of potential glycoprotein biomarkers in patients with ESCC. As expected, the total protein levels of the 6 selected proteins were not significantly different in the serum of ESCC patients and healthy controls (Supplementary Figure 2). In contrast, the serum levels of glycosylated forms of clusterin, PRELP and haptoglobin in ESCC patients were much higher than those of healthy controls (Figure 5D).

DISCUSSION

Aberrant glycosylation of proteins, which is not correlated with mRNA levels, has been linked to a variety of cancers and reflects disease status and progression [23-32]. MS-based proteomics and glycomics holds considerable potential to identify novel glycosylation-based biomarkers. To the best of our knowledge, this study is the first to use tandem lectin affinity chromatography to enrich N-linked glycoproteins, followed by 2-DE-based and iTRAQ-based proteomic quantitative profiling, and protein identification by mass spectrometry. In total, 411 differentially expressed N-linked glycoproteins were identified, including 128 up-regulated and 283 down-regulated DEGs. These DEGs may provide the scientific community a dataset of glycoproteins associated with ESCC for in-depth investigation.

Depending on the linkage of carbohydrate to the protein backbone, glycoproteins fall into two main types, N-glycosylated and O-glycosylated, in which glycans are covalently attached to the amide nitrogen of asparagine residues and to the hydroxyl group of serine or threonine residues, respectively [48]. Lectins are defined as proteins that bind to the particular oligosaccharide epitope structures attached to proteins or lipids *via* selective affinities [49]. Due to these natural features, lectins are frequently used to isolate glycoproteins from a variety of origins, in particular from blood, to reduce sample complexity and to enrich low-abundance proteins. Using 3 Lectins in parallel, i.e. ConA, WGA and Jacalin, a total of 423 proteins were fractionated from membrane proteins, among which 202 (73%) have predicted N-linked glycosylation sites. As

expected, there were unique and shared proteins enriched by these three lectins. To comprehensively isolate N-linked glycoproteins involved in ESCC, 5 Lectins allocated to a two sequential affinity chromatographies were used in this study to increase the coverage of captured N-linked glycoproteins. In addition, 2 complementary methods for quantitative profiling of enriched glycoproteins were used. In total, 262 and 317 nonredundant differentially expressed proteins were identified by iTRAQ-based profiling from the first and second lectin affinity chromatography, respectively. Two-DE-based MS/MS profiling led to fewer differentially expressed proteins being identified compared with the iTRAQ-based approach. Under a high stringency of selection criteria (fold-change ≥1.5, unique peptides ≥ 2), 519 proteins were identified in total, of which 402 (77.5%) carry predicted N-linked glycosylation sites. Identification of a large number of proteins without predicated glycosylation sites indicates the existence of non-specific binding to certain lectins needing further in-depth study to elucidate.

Functional analysis based on KEGG pathway annotations by Proteomap revealed that there were striking differences in the compositions of DEGs distributed in 5 functional entities. In line with previous findings, the DEGs in genetic information processing accounts for the largest fraction of N-linked glycoproteome, supporting the survival and rapid growth of cancer cells [47]. In further support of this, a number of signaling pathways, including PI3K-Akt, MAPK, Rap1, and HIF-1 pathways, which play important roles in oncogenesis and progression, were significantly enriched. In metabolism, glycolysis was significantly enriched in DEGs identified in this study. The increased rate of glucose metabolism and aerobic glycolysis, also known as the Warburg effect, are recognized hallmarks of cancer [50]. In agreement with this, HIF signaling proteins were over-represented in ESCC and may contribute to increased glycolysis. Besides energy generation, aerobic glycolysis enhances rapid growth and unlimited proliferation of cancer cells through providing precursors for certain biosynthetic processes as well as alteration of the tumor microenvironment [51]. Moreover, extracellular lactate has inhibitory effects on human cytotoxic T lymphocytes [52, 53]. Furthermore, metabolic phenotype of cancer stem cell seems distinct from bulk

cancer cells and is characterized by glycolysis and/or XoPhos [51]. The increased expression of fructose-bisphosphate aldolase A, triosephosphate isomerase, alpah-enolase, all-trans-retinol dehydrogenase [NAD(+)] ADH1B and glyceraldehyde-3-phosphate dehydrogenase in ESCC may contribute to the progression of ESCC *via* increased glycolysis and could serve as therapeutic targets for the management of ESCC. All together, these distinct functional categories and related effector molecules may represent potential therapeutic targets in ESCC.

Cathepsin D, a lysosomal protease, is expressed in nearly all cells and tissues of mammals. However, pro-cathepsin D but not mature cathepsin D is reportedly associated with development and progression of a variety of cancers [54]. In line with up-regulation of ConA-binding cathepsin D and procathepsin D, glycosylated procathepsin D was increased in ESCC as well, suggesting it as a potential biomarker for ESCC. In addition, we also demonstrated that both the total haptoglobin and N-linked haptoglobin in serum were differentially expressed in ESCC, as seen in other malignancies [26, 55, 56]. Although there were no alterations in the levels of total protein expression, the high-mannose clusterin and sialic acid glycoforms of 14-3-3ζ were down-regulated in ESCC but need further study to characterize their glycan moieties. After probing the serum levels of these candidate biomarkers, we provide several potential biomarkers for ESCC diagnosis. Additionally, our findings reveal that differential protein expression is a confounding factor for identification of aberrant glycosylation.

The current study has several limitations. Firstly, the main limitation is that sample pooling used in this study may cause a loss of inter-individual information. However, this is partly compensated by the validation of candidate biomarkers using individual tissue samples. Secondly, affinity chromatography comprising multiple lectins for glycoprotein isolation yielded different subglycoproteomes. Further studies should characterize the composition and structure of glycans associated with proteins. Nevertheless, our study represents, to our knowledge, the first study to characterize the

glycoproteome associated with ESCC, and thus provides the scientific community a resource of glycoproteins for in-depth investigation.

CONCLUSION

In conclusion, this study catalogued a total of 411 N-linked glycoproteins implicated in ESCC after glycoprotein enrichment using two stage multilectin affinity chromatography and MS/MS identification. Proteomap analysis revealed the distribution of differential glycoproteins in five functional categories. A preliminary validation demonstrated the differential expression of glycoprotein biomarker candidates, including total procathepsin D and haptoglobin in ESCC, high-mannose procathepsin D and clusterin and sialiated 14-3-3ζ in ESCC, and glycosylated fractions of clusterin, PRELP and haptoglobin in serum. Consequently, this study yielded a comprehensive landscape of N-linked glycoproteome associated with ESCC for further characterization.

74816_Auto_Edited.docx

0%
SIMILARITY INDEX

PRIMARY SOURCES