World Journal of

Gastroenterology

World Journal of Gastroenterology

Contents

Contents

ABOUT COVER

Editorial Board Member of *World Journal of Gastroenterology*, Nikolaos Papadopoulos, MD, PhD, Consultant, The First Department of Internal Medicine, 417 Army Share Fund Hospital, Monis Petraki 10-12, Athens 11521, Greece. nipapmed@gmail.com

AIMS AND SCOPE

The primary aim of *World Journal of Gastroenterology* (*WJG*, *World J Gastroenterol*) is to provide scholars and readers from various fields of gastroenterology and hepatology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online. *WJG* mainly publishes articles reporting research results and findings obtained in the field of gastroenterology and hepatology and covering a wide range of topics including gastroenterology, hepatology, gastrointestinal endoscopy, gastrointestinal surgery, gastrointestinal oncology, and pediatric gastroenterology.

INDEXING/ABSTRACTING

The *WJG* is now abstracted and indexed in Science Citation Index Expanded (SCIE, also known as SciSearch®), Current Contents/Clinical Medicine, Journal Citation Reports, Index Medicus, MEDLINE, PubMed, PubMed Central, Scopus, Reference Citation Analysis, China National Knowledge Infrastructure, China Science and Technology Journal Database, and Superstar Journals Database. The 2022 edition of Journal Citation Reports® cites the 2021 impact factor (IF) for *WJG* as 5.374; IF without journal self cites: 5.187; 5-year IF: 5.715; Journal Citation Indicator: 0.84; Ranking: 31 among 93 journals in gastroenterology and hepatology; and Quartile category: Q2. The *WJG*'s CiteScore for 2021 is 8.1 and Scopus CiteScore rank 2021: Gastroenterology is 18/149.

RESPONSIBLE EDITORS FOR THIS ISSUE

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

Qi-Wei Liu, Hao-Jie Ruan, Wei-Xia Chao, Meng-Xiang Li, She-Gan Gao, Yi-Jun Qi, Henan Key Laboratory of Microbiome and Esophageal Cancer Prevention and Treatment; Henan Key Laboratory of Cancer Epigenetics; Cancer Hospital, The First Affiliated Hospital, College of Clinical Medicine, Henan University of Science and Technology, Luoyang 471003, Henan Province, China

Ye-Lin Jiao, Department of Pathology, The First People's Hospital of Luo Yang, Luoyang 471000, Henan Province, China

Douglas G Ward, Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, United Kingdom

Corresponding author: Yi-Jun Qi, MD, PhD, Professor, Henan Key Laboratory of Microbiome and Esophageal Cancer Prevention and Treatment; Henan Key Laboratory of Cancer Epigenetics; Cancer Hospital, The First Affiliated Hospital, College of Clinical Medicine, Henan University of Science and Technology, No. 29 Jinghua Road, Luoyang 471003, Henan Province, China. qiyijun@haust.edu.cn

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 (ESCC) *via* two complementary approaches.

METHODS
Using tandem multilectin affinity chromatography for enrichment of N-linked glycoproteins, we performed N-linked glycoproteomic profiling in ESCC tissues by two-dimensional gel electrophoresis (2-DE)-based and isobaric tags for relative and absolute quantification (iTRAQ) labeling-based mass spectrometry quantitation in parallel, followed by validation of candidate glycoprotein biomarkers by Western blot.

RESULTS

2-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 two approaches. Proteomaps showed the distinct compositions of functional categories between proteins and glycoproteins with differential expression associated with ESCC. Western blot analysis 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, proline-arginine-rich end leucine-rich repeat protein, 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: The N-linked glycoproteome was comprehensively profiled by two-dimensional gel electrophoresis-based and isobaric tags for relative and absolute quantification (iTRAQ) labeling-based mass spectrometry quantitation in parallel after N-linked glycoprotein enrichment by a tandem of multilectin affinity chromatography. The iTRAQ labeling-based quantitative proteomic profiling outperformed protein spot intensity quantification used by two-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 provide the scientific community with a dataset of glycoproteins associated with esophageal squamous cell carcinoma for in-depth investigation.

INTRODUCTION

Esophageal cancer is a severe global health issue with 572000 new cases and 509000 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 5-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 [13-17]. 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[18,19]. 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[20-24]. The sugar chains of glycoproteins are involved in numerous physiological and pathological functions including cell growth, adhesion, differentiation, migration, signal regulation, immune responses, and tumor invasion[20,25,26]. Notably, the clinical prognosis of cancer has been correlated with certain structures of glycans attached to proteins[27,28]. The bisecting GlcNAc structure

catalyzed by N-acetylglucosaminyltransferase III plays 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[29]. 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[30]. Furthermore, N-acetylglucosaminyltransferase III modified α3β1 integrin suppresses cell spreading and migration and focal adhesion kinase activity[31]. The fraction of ConA-binding procathepsin D in the serum of hepatocellular carcinoma patients is significantly increased and shows a sensitivity of 85% and specificity of 80% for hepatocellular carcinoma diagnosis [24]. Not surprisingly, over 50% of current cancer biomarkers are glycoproteins, such as AFP, AFP-L3, CA19-9, CA125, and CEA[32-35].

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 such 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 ten pre-operative serum samples were collected at Linzhou Cancer Hospital, Henan, China between 2010 and 2011 and stored in liquid nitrogen or at -70 °C prior to protein extraction. Among the tissue samples, 15 pairs served as a discovery set and the other 32 pairs as a 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 was performed as previously reported with modifications[24]. 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 three pairs of pooled protein samples from 15 ESCC and adjacent non-tumor tissues (five 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-1043), and snowdrop lectin (GNA, Vector Laboratories, AL-1243), and a combination of wheat germ agglutitin lectin (WGA, Vector Laboratories, AL-1023) and elderberry lectin (SNA, Vector Laboratories, AL-1303) 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 room temperature for 2 h. 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.

Two-dimensional gel electrophoresis and image analysis
Two-dimensional gel electrophoresis (2-DE) was performed as previously reported[24,36]. 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.

Isobaric tags for relative and absolute quantification 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 ten ESCC and adjacent non-tumor tissues as previously described in duplicate[37] 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)[37] were performed using Enrichment Map, which organizes the enriched gene-sets into a network with links representing the overlap of member genes [38]. The gene-sets with a *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, Proteomap was used to visualize the functional distribution of DEGs or DEPs in cellular processes, including genetic information processing, metabolism, cellular processes, signaling, and others[39].

Western blot analysis

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-PAGE and then transferred to PVDF membranes. 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, United States). 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, United States). Comparisons of quantitated protein band density between ESCC and adjacent non-cancerous tissues were done by paired Student's *t*-tests or Wilcoxon signed-rank sum tests. All statistical tests were two-sided, and $P < 0.05$ was considered statistically significant.

RESULTS

N-linked glycoprotein profiling by two-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 and B shows 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 and D). 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-MS/MS of the trypsin-digested spots excised from preparative gels. The identities of proteins are shown in Tables 1 and 2, in which 22 proteins and 23 proteins were derived from the first and second lectin affinity chromatography, respectively. There

Table 1 Summary of high-mannose N-linked glycoproteins in esophageal squamous cell carcinoma by 2-dimensional gel electrophoresis

Spot No.	Accession	Protein identity (Theoretical MW/pI)	Ratio	*P* value	Scores	Unique Peptides	N-glycosylation sites
608	ENOA_HUMAN	Alpha-enolase (47.1/7.7)	2.2 (↑)	0.008	2050.2 (M:2050.2)	33	1 (102)
638	CATD_HUMAN	Cathepsin D (44.5/6.1)	2.6 (↑)	0.009	1351.8 (M:1351.8)	20	2 (134, 263)
645	G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase (36/9.3)	3.2 (↑)	0.009	319.9 (M:319.9)	4	0
646	G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase (36/9.3)	5.2 (↑)	0.002	611.5 (M:611.5)	10	0
648	G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase (36/9.3)	8.9 (↑)	0.009	730.1 (M:730.1)	14	0
660	CO1A2_HUMAN	Collagen alpha-2(I) chain (129.2/9.7)	1.7 (↑)	0.008	1207.8 (M:1207.8)	20	1 (207)
664	CATD_HUMAN	Cathepsin D (44.5/6.1)	12.5 (↑)	0.005	1337.1 (M:1337.1)	18	2 (134, 263)
668	ALDOA_HUMAN	Fructose-bisphosphate aldolase A (39.4/9.2)	1.6 (↑)	0.004	182.0 (M:182.0)	4	0
692	HSPB1_HUMAN	Heat shock protein beta-1 (22.8/6)	1.6 (↑)	0.009	1015.9 (M:1015.9)	15	0
694	TPIS_HUMAN	Triosephosphate isomerase (26.7/6.5)	1.8 (↑)	0.004	1277.6 (M:1277.6)	19	1 (233)
724	FABP5_HUMAN	Fatty acid-binding protein, epidermal (15.2/7.5)	2 (↑)	0.002	818.4 (M:818.4)	14	0
731	HPT_HUMAN	Haptoglobin (45.2/6.1)	2.1 (↑)	0.006	418.7 (M:418.7)	11	2 (207, 241)
734	HSPB1_HUMAN	Heat shock protein beta-1 (22.8/6)	1.5 (↑)	0.009	464.0 (M:464.0)	7	0
737	APOA1_HUMAN	Apolipoprotein A-I (30.8/5.5)	1.6 (↑)	0.002	200.4 (M:200.4)	4	0
738	COF1_HUMAN	Cofilin-1 (18.5/9.1)	1.5 (↑)	0.009	381.8 (M:381.8)	7	0
739	HBB_HUMAN	Hemoglobin subunit beta (16/6.9)	2.7 (↑)	0.008	720.5 (M:720.5)	10	0
753	CATD_HUMAN	Cathepsin D (44.5/6.1)	4.6 (↑)	0.009	702.0 (M:702.0)	13	2 (134, 263)
755	HBA_HUMAN	Hemoglobin subunit alpha (15.2/9.4)	3.9 (↑)	0.005	377.6 (M:377.6)	7	0
756	FIBB_HUMAN	Fibrinogen beta chain (55.9/9.3)	1.8 (↑)	0.009	3911.9 (M:3911.9)	55	0
791	FIBG_HUMAN	Fibrinogen gamma chain precursor (51.5/5.3)	1.5 (↑)	0.009	1356.7 (M:1356.7)	26	1 (78)
676	FIBB_HUMAN	Fibrinogen beta chain (55.9/9.3)	4.5 (↓)	0.006	1071.8 (M:1071.8)	18	0
678	CO3_HUMAN	Complement C3 (187/6)	6.2 (↓)	0.009	383.9 (M:383.9)	12	2 (85, 939)
688	HEMO_HUMAN	Hemopexin (51.6/6.6)	5.3 (↓)	0.002	698.2 (M:698.2)	13	2 (64, 240)
744	A1AT_HUMAN	Alpha-1-antitrypsin (46.7/5.3)	12.7 (↓)	0.009	784.8 (M:784.8)	14	3 (70, 107, 271)
748	HPT_HUMAN	Haptoglobin (45.2/6.1)	7.5 (↓)	0.008	1067.0 (M:1067.0)	19	2 (207, 241)
750	FIBB_HUMAN	Fibrinogen beta chain (55.9/9.3)	4.9 (↓)	0.009	3276.7 (M:3276.7)	48	0
772	A1AT_HUMAN	Alpha-1-antitrypsin (46.7/5.3)	1.8 (↓)	0.005	678.2 (M:678.2)	12	3 (70, 107, 271)
779	CLUS_HUMAN	Clusterin (52.5/5.9)	2.9 (↓)	0.009	619.3 (M:619.3)	10	3 (86, 103, 291)
801	FIBB_HUMAN	Fibrinogen beta chain (55.9/9.3)	2.1 (↓)	0.009	714.1 (M:714.1)	13	0
864	SODE_HUMAN	Extracellular superoxide dismutase [Cu-Zn] (25.8/6.2)	6.4 (↓)	0.004	798.6 (M:798.6)	12	1 (107)
902	APOA1_HUMAN	Apolipoprotein A-I (30.8/5.5)	4.3 (↓)	0.002	1206.2	24	0

| Spot No.	Accession	Protein identity (Theoretical MW/pI)	Ratio	P value	Scores	Unique Peptides	N-glycosylation sites
					(M:1206.2)		
932	TRFE_HUMAN	Serotransferrin precursor (77/7)	3.8 (↓)	0.004	187.0 (M:187.0)	4	1 (630)
936	HPTR_HUMAN	Haptoglobin-related protein (39/6.7)	1.5 (↓)	0.002	226.4 (M:226.4)	4	1 (149)
1000	A1AT_HUMAN	Alpha-1-antitrypsin (46.7/5.3)	15.2 (↓)	0.006	1091.2 (M:1091.2)	20	3 (70, 107, 271)
1001	HPT_HUMAN	Haptoglobin (45.2/6.1)	9.8 (↓)	0.009	1054.6 (M:1054.6)	18	2 (207, 241)

Table 2 Summary of sialic N-linked glycoproteins identified in esophageal squamous cell carcinoma by 2-dimensional gel electrophoresis

Spot No.	Accession	Protein identity (Theoretical MW/pI)	Ratio	P value	Scores	Unique Peptides	N-glycosylation sites
3569	ENOA_HUMAN	Alpha-enolase (47.1/7.7)	2 (↑)	0.002	802.0 (M:802.0)	18	2 (17, 102)
3570	IGKC_HUMAN	Ig kappa chain C region (11.6/5.5)	2.4 (↑)	0.006	54.4 (M:54.4)	2	0
3612	TRFE_HUMAN	Serotransferrin (77/7)	1.5 (↑)	0.004	927.5 (M:927.5)	25	1 (630)
3687	KV305_HUMAN	Ig kappa chain V-III region WOL (11.7/9.5)	1.7 (↑)	0.009	125.4 (M:125.4)	2	0
3720	A1AT_HUMAN	Alpha-1-antitrypsin (46.7/5.3)	1.9 (↑)	0.002	49.3 (M:49.3)	2	3 (70, 107, 271)
4746	FIBB_HUMAN	Fibrinogen beta chain (55.9/9.3)	2.8 (↓)	0.004	817.3 (M:817.3)	17	0
4784	ACTB_HUMAN	Actin, cytoplasmic 1 (41.7/5.2)	3.4 (↓)	0.002	43.6 (M:43.6)	2	1 (12)
4948	HBB_HUMAN	Hemoglobin subunit beta (16/6.9)	1.8 (↓)	0.006	129.3 (M:129.3)	4	0
4967	ACTA_HUMAN	Actin, aortic smooth muscle (42/5.1)	1.7 (↓)	0.009	402.1 (M:402.1)	9	1 (14)
4985	FHR2_HUMAN	Complement factor H-related protein 2 (30.6/6)	2 (↓)	0.002	271.1 (M:271.1)	5	0
4998	1433Z_HUMAN	14-3-3 protein zeta/delta (27.7/4.6)	6.6 (↓)	0.009	62.1 (M:62.1)	2	2 (173, 224)
5010	FIBA_HUMAN	Fibrinogen alpha chain (94.9/5.6)	2.1 (↓)	0.002	83.3 (M:83.3)	3	3 (288, 419, 831)
5019	ACTH_HUMAN	Actin, gamma-enteric smooth muscle (41.8/5.2)	1.6 (↓)	0.009	92.7 (M:92.7)	3	1 (13)
5031	KV204_HUMAN	Ig kappa chain V-II region TEW (12.3/5.6)	1.5 (↓)	0.008	84.9 (M:84.9)	2	0
5057	CATB_HUMAN	Cathepsin B (37.8/5.9)	1.6 (↓)	0.009	84.8 (M:84.8)	2	2 (192, 289)
5069	FABP5_HUMAN	Fatty acid-binding protein, epidermal (15.2/7.5)	1.5 (↓)	0.005	88.7 (M:88.7)	2	0
5076	CRYAB_HUMAN	Alpha-crystallin B chain (20.1/6.9)	3.8 (↓)	0.009	240.1 (M:240.1)	6	0
5081	TAGL_HUMAN	Transgelin (22.6/9.4)	4.1 (↓)	0.009	134.7 (M:134.7)	3	0
5083	HSP71_HUMAN	Heat shock 70 kDa protein 1A/1B (70/5.4)	1.9 (↓)	0.004	105.8 (M:105.8)	3	1 (432)
5098	K1C17_HUMAN	Keratin, type I cytoskeletal 17 (48.1/4.8)	2.2 (↓)	0.002	198.2 (M:198.2)	5	0
5128	CYC_HUMAN	Cytochrome c (11.7/10.1)	2.6 (↓)	0.006	97.4 (M:97.4)	3	0
5164	S10A9_HUMAN	Protein S100-A9 (13.2/5.7)	1.7 (↓)	0.009	85.7 (M:85.7)	3	0
5173	S10A8_HUMAN	Protein S100-A8 (10.8/6.6)	1.5 (↓)	0.002	421.3 (M:421.3)	8	0
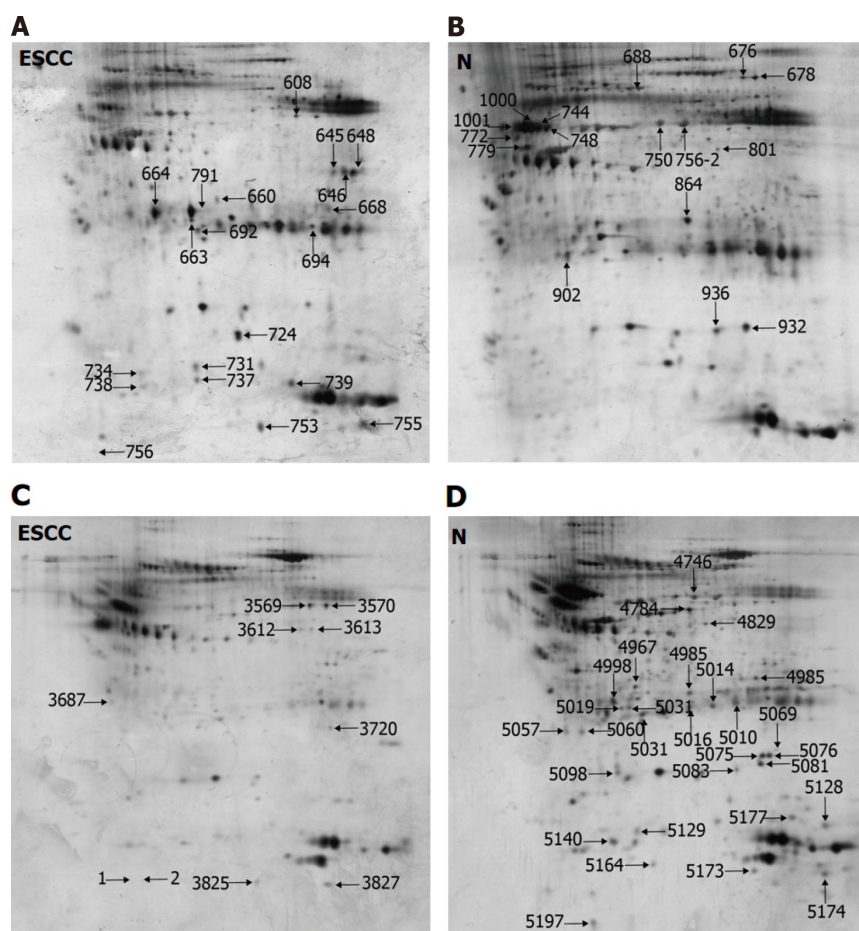
DOI: 10.3748/wjg.v28.i29.3869 **Copyright** ©The Author(s) 2022.

Figure 1 Representative two-dimensional gel electrophoresis images of N-linked glycoproteins from esophageal squamous cell carcinoma (ESCC) and adjacent non-cancerous tissues (N), in which the denoted numbers represent protein spots with differential expression. A and B: Representative two-dimensional gel electrophoresis (2-DE) images of high-mannose glycoproteins from ESCC (A) and N (B); C and D: Representative 2-DE images of GlcNAc/sialic acid-containing glycoproteins from ESCC (C) and N (D). ESCC: Esophageal squamous cell carcinoma; N: Non-cancerous tissues.

were six proteins in common between these two isolated fractions. In agreement with our previous reports[24,36], 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 ten 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 Tables 1 and 2). 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%) have predicted N-linked glycosylation sites (Figure 2B). For N-linked glycoproteins enriched by the downstream GlcNAc/sialic acid binding lectin chromatography, two independent iTRAQ-labeled pools produced 1174 and 936 proteins following isolation of high-mannose glycoproteins (Figure 2C, Supplementary Tables 3 and 4). 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%) were predicted to be N-linked glycoproteins, were identified from the GlcNAc/sialic acid binding lectin chromatography enrichment.
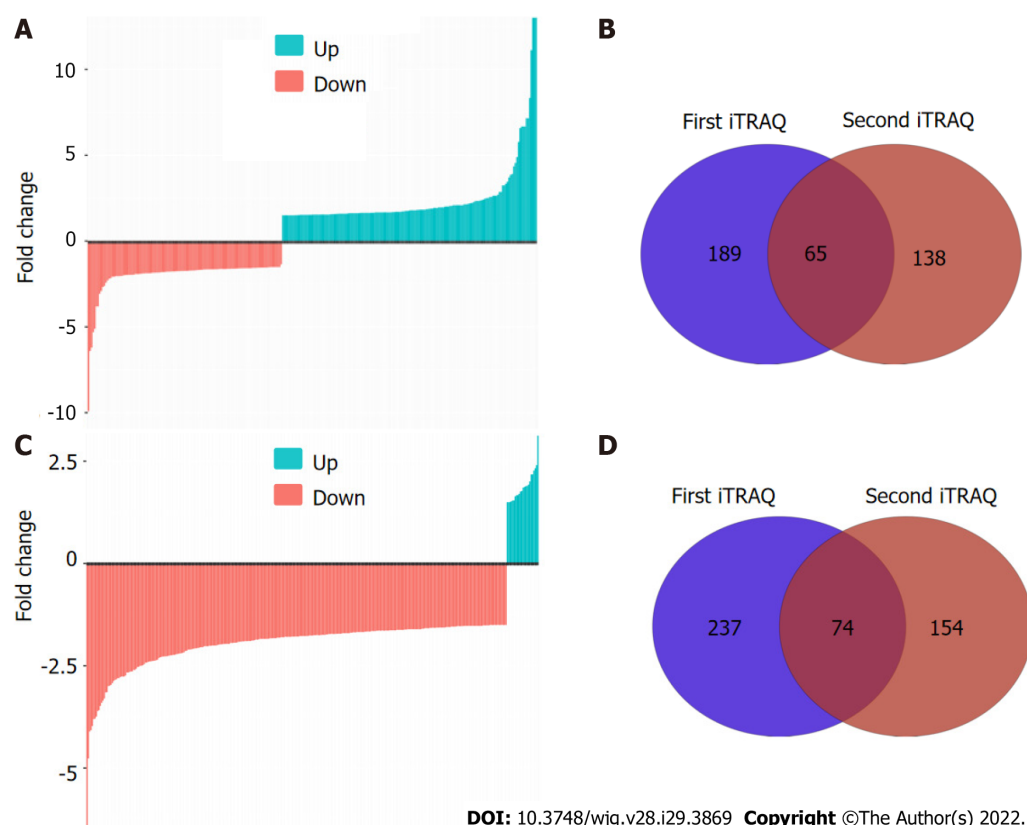
Figure 2 N-linked glycoprotein profiling by isobaric tags for relative and absolute quantification labeling and liquid chromatography electrospray ionisation tandem mass spectrometry/mass spectrometry identification. A: Waterfall plot shows the differentially expressed high-mannose glycoproteins; B: Venn diagram depicts the unique and overlapped high-mannose glycoproteins identified from two independent replicates; C: Waterfall plot shows the differentially expressed GlcNAc/sialic acid-containing glycoproteins; D: Venn diagram depicts the unique and overlapped GlcNAc/sialic acid-containing glycoproteins identified from two independent replicates. iTRAQ: Isobaric tags for relative and absolute quantification.

There were 58 overlapping glycoproteins found between 262 high-mannose glycoproteins and 317 GlcNAc/sialic acid-containing glycoproteins.

Comparison of two-dimensional gel electrophoresis-based and iTRAQ labeling-based profiling

In this study, two 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. Our results demonstrate that iTRAQ-based quantitative proteomic profiling outperformed protein spot intensity quantification by 2-DE. There were, however, common glycoproteins found by both approaches. Therefore, the combination of these four fractions produced 411 unique N-linked glycoproteins. These findings indicate that a combination of these two 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. Figure 3A-C shows 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 five functional categories, which comprise genetic information processing, environmental information processing, cellular processes, metabolism, and organismal systems. Figure 4A-C shows 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 (Figure 4A and B). 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 are implicated in focal adhesion, tight
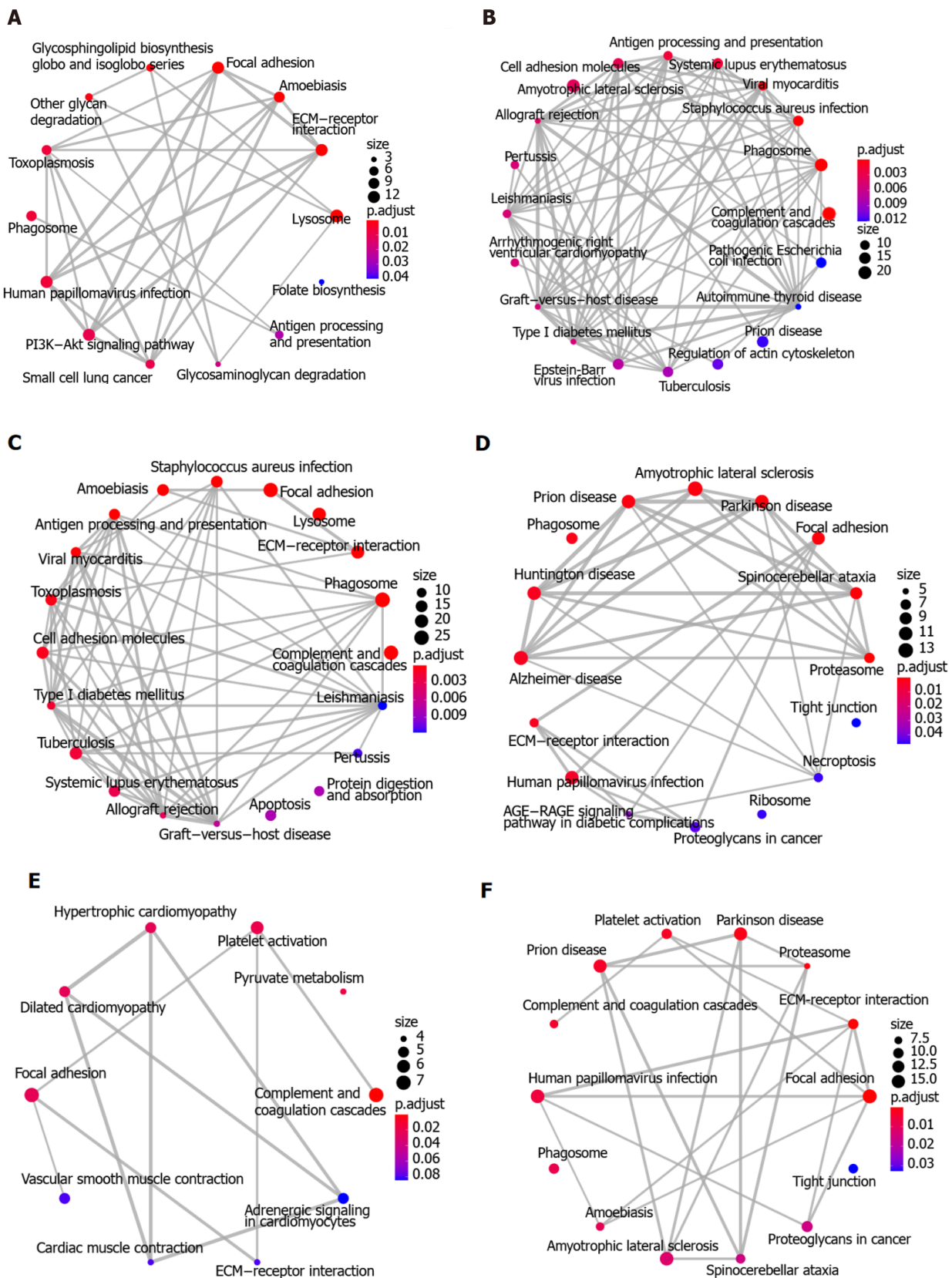
DOI: 10.3748/wjg.v28.i29.3869 **Copyright** ©The Author(s) 2022.

Figure 3 Functional analyses of proteins and glycoproteins with differential expression using enrichment map. A-C: Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis using enrichment map shows the significantly enriched biological pathways of up-regulated differentially expressed glycoproteins (DEGs), down-regulated DEGs, and total DEGs, respectively; D-F: KEGG pathway enrichment analysis using enrichment map shows the significantly enriched biological pathways of up-regulated DEPs, down-regulated DEPs, and total DEPs, respectively. KEGG: Kyoto encyclopedia of genes and genomes; DEGs: Differentially expressed glycoproteins; DEPs: Differentially expressed proteins.
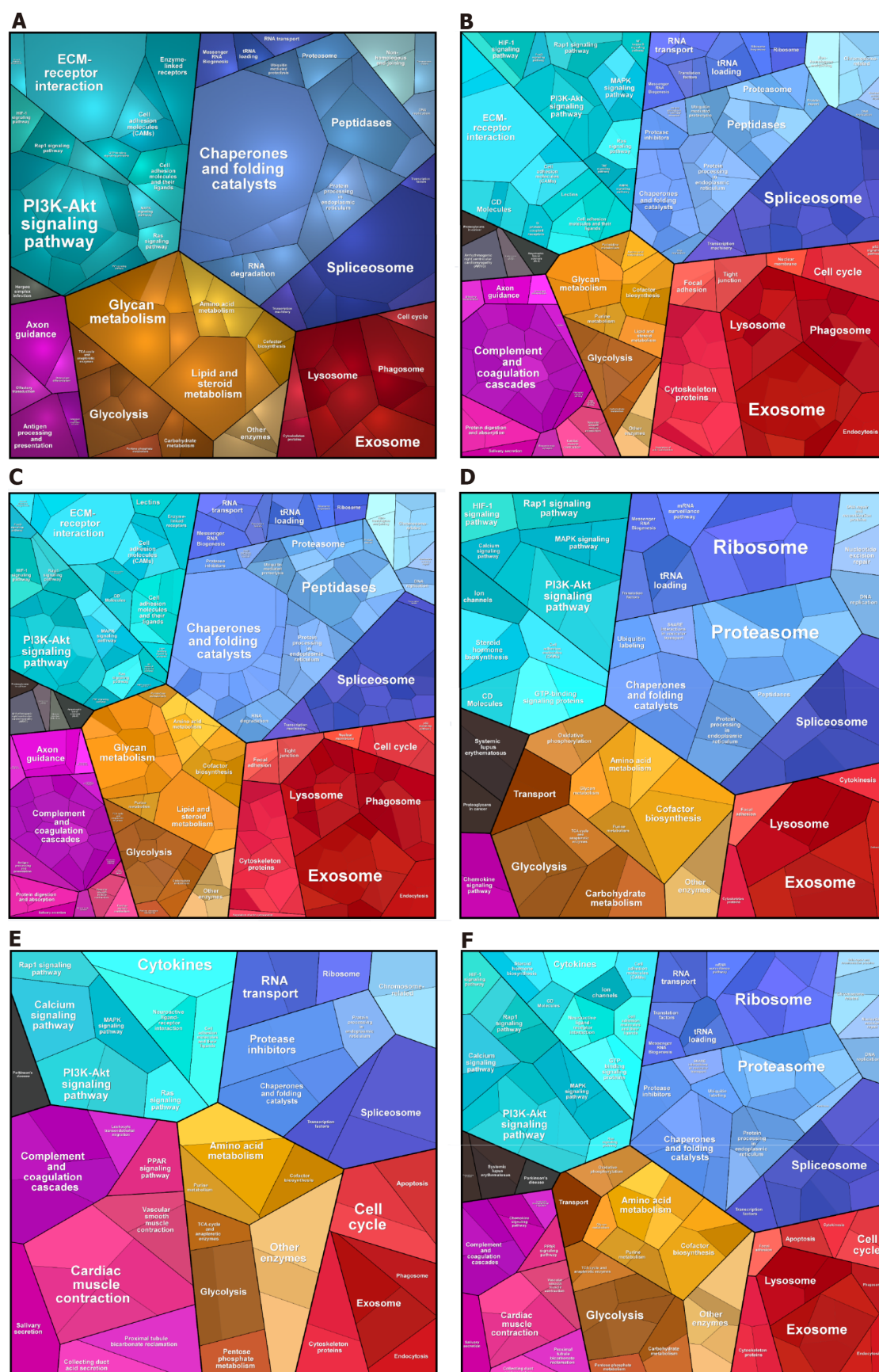
Figure 4 Functional analyses of proteins and glycoproteins with differential expression using Proteomap. A-C: Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis using Proteomap shows the significantly enriched biological pathways of up-regulated differentially expressed glycoproteins (DEGs), down-regulated DEGs, and total DEGs, respectively; D-F: KEGG pathway enrichment analysis using Proteomap shows the significantly enriched biological pathways of up-regulated DEPs, down-regulated DEPs, and total DEPs, respectively. KEGG: Kyoto encyclopedia of genes and genomes; DEGs:

Differentially expressed glycoproteins; DEPs: Differentially expressed proteins.

junction, endocytosis, and p53 signaling pathway (Figure 4A and B). In glycolysis, there were more up-regulated than down-regulated DEGs. Notably, the complement and coagulation cascades in the immune system include exclusively down-regulated DEGs (Figure 4A and B).

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, and 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 and F). 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 and F). In the same line, there were remarkable distinctions between DEGs and DEPs identified using iTRAQ-based quantitation of Indian ESCC samples (data not shown)[40].

Validation of potential glycoprotein biomarkers

Western blot analysis was performed 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 four selected proteins (Supplementary Figure 1). These data suggest 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 glycosyltrans-fereases.

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 six 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[20,41-44]. MS-based proteomics and glycomics hold 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 labeling-based proteomic quantitative profiling, and protein identification by MS. In total, 411 differentially expressed N-linked glycoproteins were
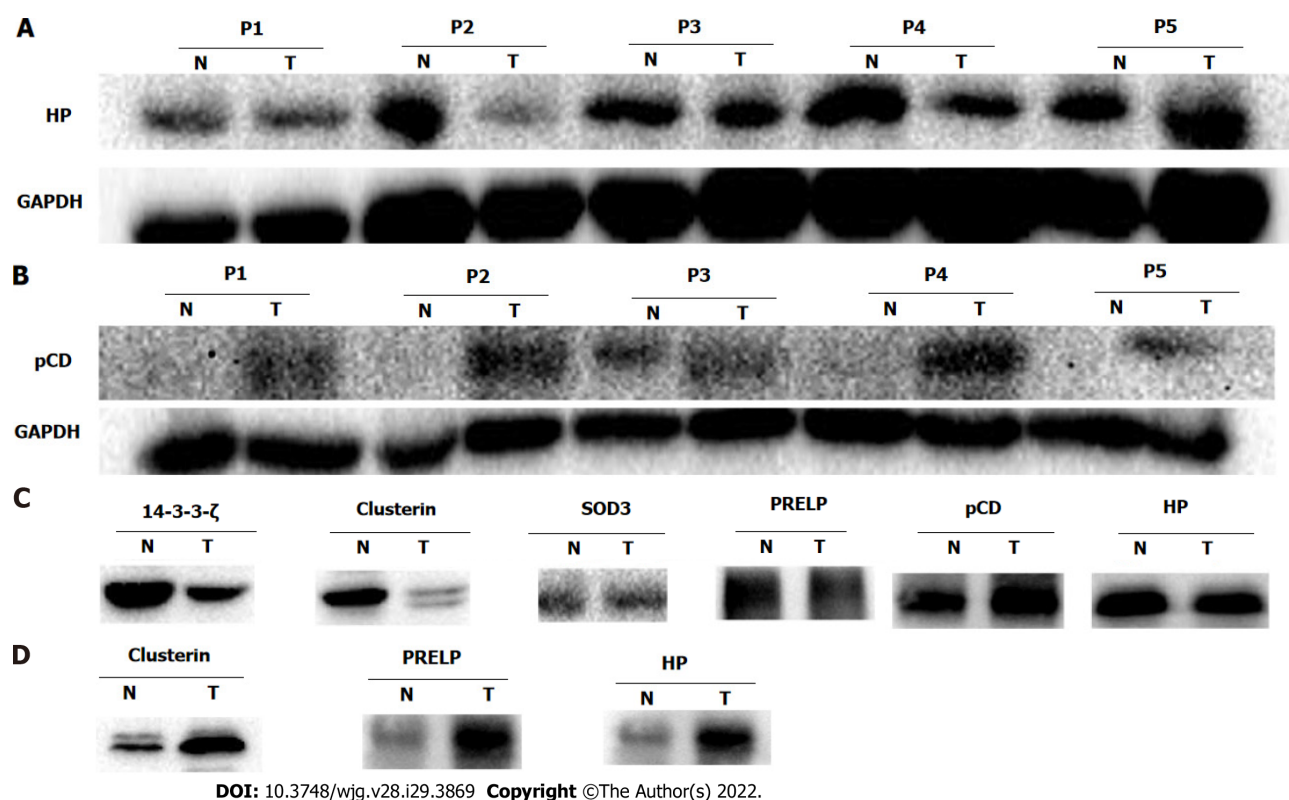
DOI: 10.3748/wjg.v28.i29.3869 **Copyright** ©The Author(s) 2022.

Figure 5 Western blot validations of potential glycoprotein biomarkers. A and B: Representative Western blot results show haptoglobin (HP) (A) and procathepsin D (pCD) (B) with differential expression between esophageal squamous cell carcinoma (ESCC/T) and adjacent non-cancerous tissues (N); C: Representative Western blot results show the N-linked glycosylated fraction of HP, pCD, clusterin, superoxide dismutase 3 (SOD3), proline-arginine-rich end leucine-rich repeat protein (PRELP), and 14-3-3ζ in ESCC/T and N enriched by corresponding lectins; D: Representative Western blot results show the N-linked glycosylated fractions of clusterin, PRELP, and HP in serum of patients with ESCC/T and healthy controls. The results are representative of three independent experiments. ESCC/T: Esophageal squamous cell carcinoma; HP: Haptoglobin; pCD: Procathepsin D; SOD3: Superoxide dismutase 3; PRELP: Proline-arginine-rich end leucine-rich repeat protein.

identified, including 128 up-regulated and 283 down-regulated DEGs. These DEGs 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 [45]. Lectins are defined as proteins that bind to the particular oligosaccharide epitope structures attached to proteins or lipids *via* selective affinities[46]. 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 three lectins in parallel, *i.e.*, ConA, WGA, and Jacalin, a total of 423 proteins were fractionated from membrane proteins, among which 202 (71%) have predicted N-linked glycosylation sites[47]. As expected, there were unique and shared proteins enriched by these three lectins. To comprehensively isolate N-linked glycoproteins involved in ESCC, five lectins allocated to two sequential affinity chromatographies were used in this study to increase the coverage of captured N-linked glycoproteins. In addition, two complementary methods for quantitative profiling of enriched glycoproteins were used. In total, 262 and 317 nonredundant DEPs were identified by iTRAQ labeling-based profiling from the first and second lectin affinity chromato-graphy, respectively. 2-DE-based MS/MS profiling led to fewer DEPs being identified compared with the iTRAQ labeling-based approach. Under a high stringency of selection criteria (foldchange ≥ 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 five functional entities. In line with previous findings, the DEGs in genetic information processing account for the largest fraction of N-linked glycoproteome, supporting the survival and rapid growth of cancer cells[39]. 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[48]. 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[49]. Moreover, extracellular lactate has inhibitory effects on human cytotoxic T lymphocytes[50,51]. Furthermore, metabolic phenotype of cancer stem cell seems distinct from bulk cancer cells and is characterized by glycolysis and/or XoPhos [49]. The increased expression of fructose-bisphosphate aldolase A, triosephosphate isomerase, alpha-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 the development and progression of a variety of cancers[52]. 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 total haptoglobin and N-linked haptoglobin in serum were differentially expressed in ESCC, as seen in other malignancies[41,53,54]. 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. First, 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. Second, 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 fractions of procathepsin D, 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.

ARTICLE HIGHLIGHTS

Research background

Recent large-scale "omics" studies in esophageal squamous cell carcinoma (ESCC) have identified a myriad of aberrations at the levels of genome, epigenome, transcriptome, proteome, *etc.*, revealing the high molecular heterogeneity of ESCC. However, protein post-translational modifications, such as glycosylation and phosphorylation, which provide additional significant biological insights, are missing.

Research motivation

The sugar chains of glycoproteins are involved in numerous physiological and pathological conditions. More than 50% of current cancer biomarkers are glycoproteins.

Research objectives

To identify N-linked glycoproteins associated with ESCC after isolation of N-linked glycoproteins using tandem multilectin affinity chromatography.

Research methods

N-linked glycoproteins were isolated from ESCC and adjacent non-tumor tissue samples using multilectin affinity chromatography. Two-dimensional gel electrophoresis (2-DE)-based and isobaric tags for relative and absolute quantification (iTRAQ) labeling-based mass spectrometry quantitation were performed in parallel to profile the N-linked glycoproteome in ESCC, followed by validation of candidate glycoprotein biomarkers using Western blot.

Research results

A total of 411 differentially expressed N-linked glycoproteins (DEGs) with potential glycosylation sites on proteins were identified by 2-DE-based and iTRAQ labeling-based quantitation, demonstrating the outperformance of iTRAQ labeling-based quantitation over 2-DE. These DEGs exhibited distinctive compositions in functional categories from differentially expressed proteins in ESCC. Western blot analysis 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, proline-arginine-rich end leucine-rich repeat protein, and haptoglobin in patients with ESCC were remarkably higher than those in healthy controls.

Research conclusions

This study identified the aberrant N-linked glycoproteome associated with ESCC, which will be a valuable resource for future investigations.

Research perspectives

In-depth characterization of the composition and structure of glycans associated with proteins can shed more lights on biological insights and clinical relevance of the identified DEGs in ESCC.

FOOTNOTES

ORCID number: Qi-Wei Liu 0000-0002-7874-4056; Hao-Jie Ruan 0000-0002-3201-3864; Wei-Xia Chao 0000-0003-0700-658X; Meng-Xiang Li 0000-0002-6987-545X; Ye-Lin Jiao 0000-0002-4920-0738; Douglas G Ward 0000-0002-2328-1445; She-Gan Gao 0000-0002-7720-8729; Yi-Jun Qi 0000-0003-0239-5862.

REFERENCES

1   **Bray F**, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; **68**: 394-424 [PMID: 30207593 DOI: 10.3322/caac.21492]

2   **Abnet CC**, Arnold M, Wei WQ. Epidemiology of Esophageal Squamous Cell Carcinoma. *Gastroenterology* 2018; **154**: 360-373 [PMID: 28823862 DOI: 10.1053/j.gastro.2017.08.023]

3   **Song Y**, Li L, Ou Y, Gao Z, Li E, Li X, Zhang W, Wang J, Xu L, Zhou Y, Ma X, Liu L, Zhao Z, Huang X, Fan J, Dong L, Chen G, Ma L, Yang J, Chen L, He M, Li M, Zhuang X, Huang K, Qiu K, Yin G, Guo G, Feng Q, Chen P, Wu Z, Wu J, Zhao J, Luo L, Fu M, Xu B, Chen B, Li Y, Tong T, Wang M, Liu Z, Lin D, Zhang X, Yang H, Zhan Q. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* 2014; **509**: 91-95 [PMID: 24670651 DOI: 10.1038/nature13176]

4   **Zeng H**, Zheng R, Guo Y, Zhang S, Zou X, Wang N, Zhang L, Tang J, Chen J, Wei K, Huang S, Wang J, Yu L, Zhao D, Song G, Shen Y, Yang X, Gu X, Jin F, Li Q, Li Y, Ge H, Zhu F, Dong J, Guo G, Wu M, Du L, Sun X, He Y, Coleman MP, Baade P, Chen W, Yu XQ. Cancer survival in China, 2003-2005: a population-based study. *Int J Cancer* 2015; **136**: 1921-1930 [PMID: 25242378 DOI: 10.1002/ijc.29227]

5   **Wang GQ**, Jiao GG, Chang FB, Fang WH, Song JX, Lu N, Lin DM, Xie YQ, Yang L. Long-term results of operation for 420 patients with early squamous cell esophageal carcinoma discovered by screening. *Ann Thorac Surg* 2004; **77**: 1740-1744 [PMID: 15111177 DOI: 10.1016/j.athoracsur.2003.10.098]

6   **Pennathur A**, Gibson MK, Jobe BA, Luketich JD. Oesophageal carcinoma. *Lancet* 2013; **381**: 400-412 [PMID: 23374478 DOI: 10.1016/S0140-6736(12)60643-6]

7   **He Z**, Liu Z, Liu M, Guo C, Xu R, Li F, Liu A, Yang H, Shen L, Wu Q, Duan L, Li X, Zhang C, Pan Y, Cai H, Ke Y. Efficacy of endoscopic screening for esophageal cancer in China (ESECC): design and preliminary results of a population-based randomised controlled trial. *Gut* 2019; **68**: 198-206 [PMID: 29306867 DOI: 10.1136/gutjnl-2017-315520]

8   **Kosugi S**, Nishimaki T, Kanda T, Nakagawa S, Ohashi M, Hatakeyama K. Clinical significance of serum carcinoembryonic antigen, carbohydrate antigen 19-9, and squamous cell carcinoma antigen levels in esophageal cancer patients. *World J Surg* 2004; **28**: 680-685 [PMID: 15383868 DOI: 10.1007/s00268-004-6865-y]

9   **Mealy K**, Feely J, Reid I, McSweeney J, Walsh T, Hennessy TP. Tumour marker detection in oesophageal carcinoma. *Eur J Surg Oncol* 1996; **22**: 505-507 [PMID: 8903494 DOI: 10.1016/s0748-7983(96)92998-4]

10   **Shimada H**, Nabeya Y, Okazumi S, Matsubara H, Miyazawa Y, Shiratori T, Hayashi H, Gunji Y, Ochiai T. Prognostic significance of CYFRA 21-1 in patients with esophageal squamous cell carcinoma. *J Am Coll Surg* 2003; **196**: 573-578 [PMID: 12691934 DOI: 10.1016/s1072-7515(02)01905-1]

11   **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 [PMID: 25001061 DOI: 10.1186/1471-2407-14-490]

12   **Enzinger PC**, Mayer RJ. Esophageal cancer. *N Engl J Med* 2003; **349**: 2241-2252 [PMID: 14657432 DOI: 10.1056/NEJMra035010]

13   **Cui Y**, Chen H, Xi R, Cui H, Zhao Y, Xu E, Yan T, Lu X, Huang F, Kong P, Li Y, Zhu X, Wang J, Zhu W, Ma Y, Zhou Y, Guo S, Zhang L, Liu Y, Wang B, Xi Y, Sun R, Yu X, Zhai Y, Wang F, Yang J, Yang B, Cheng C, Liu J, Song B, Li H, Wang Y, Zhang Y, Cheng X, Zhan Q, Liu Z. Whole-genome sequencing of 508 patients identifies key molecular features associated with poor prognosis in esophageal squamous cell carcinoma. *Cell Res* 2020; **30**: 902-913 [PMID: 32398863 DOI: 10.1038/s41422-020-0333-6]

14   **Gao YB**, Chen ZL, Li JG, Hu XD, Shi XJ, Sun ZM, Zhang F, Zhao ZR, Li ZT, Liu ZY, Zhao YD, Sun J, Zhou CC, Yao R, Wang SY, Wang P, Sun N, Zhang BH, Dong JS, Yu Y, Luo M, Feng XL, Shi SS, Zhou F, Tan FW, Qiu B, Li N, Shao K, Zhang LJ, Xue Q, Gao SG, He J. Genetic landscape of esophageal squamous cell carcinoma. *Nat Genet* 2014; **46**: 1097-1102 [PMID: 25151357 DOI: 10.1038/ng.3076]

15   **Li J**, Chen Z, Tian L, Zhou C, He MY, Gao Y, Wang S, Zhou F, Shi S, Feng X, Sun N, Liu Z, Skogerboe G, Dong J, Yao R, Zhao Y, Sun J, Zhang B, Yu Y, Shi X, Luo M, Shao K, Li N, Qiu B, Tan F, Chen R, He J. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut* 2014; **63**: 1700-1710 [PMID: 24522499 DOI: 10.1136/gutjnl-2013-305806]

16   **Cancer Genome Atlas Research Network**; Analysis Working Group: Asan University; BC Cancer Agency; Brigham and Women's Hospital; Broad Institute; Brown University; Case Western Reserve University; Dana-Farber Cancer Institute; Duke University; Greater Poland Cancer Centre; Harvard Medical School; Institute for Systems Biology; KU Leuven; Mayo Clinic; Memorial Sloan Kettering Cancer Center; National Cancer Institute; Nationwide Children's Hospital; Stanford University; University of Alabama; University of Michigan; University of North Carolina; University of Pittsburgh; University of Rochester; University of Southern California; University of Texas MD Anderson Cancer Center; University of Washington; Van Andel Research Institute; Vanderbilt University; Washington University; Genome Sequencing Center: Broad Institute; Washington University in St. Louis; Genome Characterization Centers: BC Cancer Agency; Broad Institute; Harvard Medical School; Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University; University of North Carolina; University of Southern California Epigenome Center; University of Texas MD Anderson Cancer Center; Van Andel Research Institute; Genome Data Analysis Centers: Broad Institute; Brown University:; Harvard Medical School; Institute for Systems Biology; Memorial Sloan Kettering Cancer Center; University of California Santa Cruz; University of Texas MD Anderson Cancer Center; Biospecimen Core Resource: International Genomics Consortium; Research Institute at Nationwide Children's Hospital; Tissue Source Sites: Analytic Biologic Services; Asan Medical Center; Asterand Bioscience; Barretos Cancer Hospital; BioreclamationIVT; Botkin Municipal Clinic; Chonnam National University Medical School; Christiana Care Health System; Cureline; Duke University; Emory University; Erasmus University; Indiana University School of Medicine; Institute of Oncology of Moldova; International Genomics Consortium; Invidumed; Israelitisches Krankenhaus Hamburg; Keimyung University School of Medicine; Memorial Sloan Kettering Cancer Center; National Cancer Center Goyang; Ontario Tumour Bank; Peter MacCallum Cancer Centre; Pusan National University Medical School; Ribeirão Preto Medical School; St. Joseph's Hospital &Medical Center; St. Petersburg Academic University; Tayside Tissue Bank; University of Dundee; University of Kansas Medical Center; University of Michigan; University of North Carolina at Chapel Hill; University of Pittsburgh School of Medicine; University of Texas MD Anderson Cancer Center; Disease Working Group: Duke University; Memorial Sloan Kettering Cancer Center; National Cancer Institute; University of Texas MD Anderson Cancer Center; Yonsei University College of

Medicine; Data Coordination Center: CSRA Inc; Project Team: National Institutes of Health. Integrated genomic characterization of oesophageal carcinoma. *Nature* 2017; **541**: 169-175 [PMID: 28052061 DOI: 10.1038/nature20805]

17   **Cao W**, Lee H, Wu W, Zaman A, McCorkle S, Yan M, Chen J, Xing Q, Sinnott-Armstrong N, Xu H, Sailani MR, Tang W, Cui Y, Liu J, Guan H, Lv P, Sun X, Sun L, Han P, Lou Y, Chang J, Wang J, Gao Y, Guo J, Schenk G, Shain AH, Biddle FG, Collisson E, Snyder M, Bivona TG. Multi-faceted epigenetic dysregulation of gene expression promotes esophageal squamous cell carcinoma. *Nat Commun* 2020; **11**: 3675 [PMID: 32699215 DOI: 10.1038/s41467-020-17227-z]

18   **Zhang B**, Whiteaker JR, Hoofnagle AN, Baird GS, Rodland KD, Paulovich AG. Clinical potential of mass spectrometry-based proteogenomics. *Nat Rev Clin Oncol* 2019; **16**: 256-268 [PMID: 30487530 DOI: 10.1038/s41571-018-0135-7]

19   **Gao Q**, Zhu H, Dong L, Shi W, Chen R, Song Z, Huang C, Li J, Dong X, Zhou Y, Liu Q, Ma L, Wang X, Zhou J, Liu Y, Boja E, Robles AI, Ma W, Wang P, Li Y, Ding L, Wen B, Zhang B, Rodriguez H, Gao D, Zhou H, Fan J. Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell* 2019; **179**: 561-577.e22 [PMID: 31585088 DOI: 10.1016/j.cell.2019.08.052]

20   **Ohtsubo K**, Marth JD. Glycosylation in cellular mechanisms of health and disease. *Cell* 2006; **126**: 855-867 [PMID: 16959566 DOI: 10.1016/j.cell.2006.08.019]

21   **Mechref Y**, Hussein A, Bekesova S, Pungpapong V, Zhang M, Dobrolecki LE, Hickey RJ, Hammoud ZT, Novotny MV. Quantitative serum glycomics of esophageal adenocarcinoma and other esophageal disease onsets. *J Proteome Res* 2009; **8**: 2656-2666 [PMID: 19441788 DOI: 10.1021/pr8008385]

22   **Sethi MK**, Hancock WS, Fanayan S. Identifying N-Glycan Biomarkers in Colorectal Cancer by Mass Spectrometry. *Acc Chem Res* 2016; **49**: 2099-2106 [PMID: 27653471 DOI: 10.1021/acs.accounts.6b00193]

23   **Li CW**, Lim SO, Xia W, Lee HH, Chan LC, Kuo CW, Khoo KH, Chang SS, Cha JH, Kim T, Hsu JL, Wu Y, Hsu JM, Yamaguchi H, Ding Q, Wang Y, Yao J, Lee CC, Wu HJ, Sahin AA, Allison JP, Yu D, Hortobagyi GN, Hung MC. Glycosylation and stabilization of programmed death ligand-1 suppresses T-cell activity. *Nat Commun* 2016; **7**: 12632 [PMID: 27572267 DOI: 10.1038/ncomms12632]

24   **Qi YJ**, Ward DG, Pang C, Wang QM, Wei W, Ma J, Zhang J, Lou Q, Shimwell NJ, Martin A, Wong N, Chao WX, Wang M, Ma YF, Johnson PJ. Proteomic profiling of N-linked glycoproteins identifies ConA-binding procathepsin D as a novel serum biomarker for hepatocellular carcinoma. *Proteomics* 2014; **14**: 186-195 [PMID: 24259486 DOI: 10.1002/pmic.201300226]

25   **Alper J**. Glycobiology. Turning sweet on cancer. *Science* 2003; **301**: 159-160 [PMID: 12855785 DOI: 10.1126/science.301.5630.159]

26   **Varki A**. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* 1993; **3**: 97-130 [PMID: 8490246 DOI: 10.1093/glycob/3.2.97]

27   **Hakomori S**. Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. *Cancer Res* 1996; **56**: 5309-5318 [PMID: 8968075 DOI: 10.1097/00002820-199612000-00006]

28   **Kobata A**. A retrospective and prospective view of glycopathology. *Glycoconj J* 1998; **15**: 323-331 [PMID: 9613818 DOI: 10.1023/a:1006961532182]

29   **Xu Q**, Isaji T, Lu Y, Gu W, Kondo M, Fukuda T, Du Y, Gu J. Roles of N-acetylglucosaminyltransferase III in epithelial-to-mesenchymal transition induced by transforming growth factor β1 (TGF-β1) in epithelial cell lines. *J Biol Chem* 2012; **287**: 16563-16574 [PMID: 22451656 DOI: 10.1074/jbc.M111.262154]

30   **Liwosz A**, Lei T, Kukuruzinska MA. N-glycosylation affects the molecular organization and stability of E-cadherin junctions. *J Biol Chem* 2006; **281**: 23138-23149 [PMID: 16682414 DOI: 10.1074/jbc.M512621200]

31   **Zhao Y**, Nakagawa T, Itoh S, Inamori K, Isaji T, Kariya Y, Kondo A, Miyoshi E, Miyazaki K, Kawasaki N, Taniguchi N, Gu J. N-acetylglucosaminyltransferase III antagonizes the effect of N-acetylglucosaminyltransferase V on alpha3beta1 integrin-mediated cell migration. *J Biol Chem* 2006; 281: 32122-32130 [PMID: 16940045 DOI: 10.1074/jbc.M607274200]

32   **Kim JH**, Jun KH, Jung H, Park IS, Chin HM. Prognostic Value of Preoperative Serum Levels of Five Tumor Markers (Carcinoembryonic Antigen, CA19-9, Alpha-fetoprotein, CA72-4, and CA125) in Gastric Cancer. *Hepatogastroenterology* 2014; **61**: 863-869 [PMID: 26176088 DOI: 10.5754/hge131018]

33   **Terzi H**, Kale E, Kale A, Turkay U, Chong GO, Lee YS. New method: Are tumor markers in vaginal-washing fluid significant in the diagnosis of primary ovarian carcinoma? *Eur J Gynaecol Oncol* 2015; **36**: 560-563 [PMID: 26513883 DOI: 10.12892/ejgo2743.2015]

34   **Zhang D**, Yu M, Xu T, Xiong B. Predictive value of serum CEA, CA19-9 and CA125 in diagnosis of colorectal liver metastasis in Chinese population. *Hepatogastroenterology* 2013; **60**: 1297-1301 [PMID: 23933921 DOI: 10.5754/hge121125]

35   **Kumada T**, Nakano S, Takeda I, Kiriyama S, Sone Y, Hayashi K, Katoh H, Endoh T, Sassa T, Satomura S. Clinical utility of Lens culinaris agglutinin-reactive alpha-fetoprotein in small hepatocellular carcinoma: special reference to imaging diagnosis. *J Hepatol* 1999; **30**: 125-130 [PMID: 9927159 DOI: 10.1016/s0168-8278(99)80016-6]

36   **Qi Y**, Chiu JF, Wang L, Kwong DL, He QY. Comparative proteomic analysis of esophageal squamous cell carcinoma. *Proteomics* 2005; **5**: 2960-2971 [PMID: 15986332 DOI: 10.1002/pmic.200401175]

37   **Gao SG**, Liu RM, Zhao YG, Wang P, Ward DG, Wang GC, Guo XQ, Gu J, Niu WB, Zhang T, Martin A, Guo ZP, Feng XS, Qi YJ, Ma YF. Integrative topological analysis of mass spectrometry data reveals molecular features with clinical relevance in esophageal squamous cell carcinoma. *Sci Rep* 2016; **6**: 21586 [PMID: 26898710 DOI: 10.1038/srep21586]

38   **Merico D**, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 2010; **5**: e13984 [PMID: 21085593 DOI: 10.1371/journal.pone.0013984]

39   **Liebermeister W**, Noor E, Flamholz A, Davidi D, Bernhardt J, Milo R. Visual account of protein investment in cellular functions. *Proc Natl Acad Sci U S A* 2014; **111**: 8488-8493 [PMID: 24889604 DOI: 10.1073/pnas.1314810111]

40   **Pawar H**, Kashyap MK, Sahasrabuddhe NA, Renuse S, Harsha HC, Kumar P, Sharma J, Kandasamy K, Marimuthu A, Nair B, Rajagopalan S, Maharudraiah J, Premalatha CS, Kumar KV, Vijayakumar M, Chaerkady R, Prasad TS, Kumar RV, Pandey A. Quantitative tissue proteomics of esophageal squamous cell carcinoma for novel biomarker discovery. *Cancer Biol Ther* 2011; **12**: 510-522 [PMID: 21743296 DOI: 10.4161/cbt.12.6.16833]

41   **Zhang Y**, Zhu J, Yin H, Marrero J, Zhang XX, Lubman DM. ESI-LC-MS Method for Haptoglobin Fucosylation Analysis

in Hepatocellular Carcinoma and Liver Cirrhosis. *J Proteome Res* 2015; **14**: 5388-5395 [PMID: 26503433 DOI: 10.1021/acs.jproteome.5b00792]

42   **Darebna P**, Novak P, Kucera R, Topolcan O, Sanda M, Goldman R, Pompach P. Changes in the expression of N- and O-glycopeptides in patients with colorectal cancer and hepatocellular carcinoma quantified by full-MS scan FT-ICR and multiple reaction monitoring. *J Proteomics* 2017; **153**: 44-52 [PMID: 27646713 DOI: 10.1016/j.jprot.2016.09.004]

43   **Ahn JM**, Sung HJ, Yoon YH, Kim BG, Yang WS, Lee C, Park HM, Kim BJ, Lee SY, An HJ, Cho JY. Integrated glycoproteomics demonstrates fucosylated serum paraoxonase 1 alterations in small cell lung cancer. *Mol Cell Proteomics* 2014; **13**: 30-48 [PMID: 24085812 DOI: 10.1074/mcp.M113.028621]

44   **Lin Z**, Yin H, Lo A, Ruffin MT, Anderson MA, Simeone DM, Lubman DM. Label-free relative quantification of alpha-2-macroglobulin site-specific core-fucosylation in pancreatic cancer by LC-MS/MS. *Electrophoresis* 2014; **35**: 2108-2115 [PMID: 24285556 DOI: 10.1002/elps.201300376]

45   **Christiansen MN**, Chik J, Lee L, Anugraham M, Abrahams JL, Packer NH. Cell surface protein glycosylation in cancer. *Proteomics* 2014; **14**: 525-546 [PMID: 24339177 DOI: 10.1002/pmic.201300387]

46   **Ghosh D**, Krokhin O, Antonovici M, Ens W, Standing KG, Beavis RC, Wilkins JA. Lectin affinity as an approach to the proteomic analysis of membrane glycoproteins. *J Proteome Res* 2004; **3**: 841-850 [PMID: 15359739 DOI: 10.1021/pr049937f]

47   **Lee A**, Kolarich D, Haynes PA, Jensen PH, Baker MS, Packer NH. Rat liver membrane glycoproteome: enrichment by phase partitioning and glycoprotein capture. *J Proteome Res* 2009; **8**: 770-781 [PMID: 19125615 DOI: 10.1021/pr800910w]

48   **Hanahan D**, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; **144**: 646-674 [PMID: 21376230 DOI: 10.1016/j.cell.2011.02.013]

49   **Ganapathy-Kanniappan S**. Molecular intricacies of aerobic glycolysis in cancer: current insights into the classic metabolic phenotype. *Crit Rev Biochem Mol Biol* 2018; **53**: 667-682 [PMID: 30668176 DOI: 10.1080/10409238.2018.1556578]

50   **Fischer K**, Hoffmann P, Voelkl S, Meidenbauer N, Ammer J, Edinger M, Gottfried E, Schwarz S, Rothe G, Hoves S, Renner K, Timischl B, Mackensen A, Kunz-Schughart L, Andreesen R, Krause SW, Kreutz M. Inhibitory effect of tumor cell-derived lactic acid on human T cells. *Blood* 2007; **109**: 3812-3819 [PMID: 17255361 DOI: 10.1182/blood-2006-07-035972]

51   **Kumagai S**, Koyama S, Itahashi K, Tanegashima T, Lin YT, Togashi Y, Kamada T, Irie T, Okumura G, Kono H, Ito D, Fujii R, Watanabe S, Sai A, Fukuoka S, Sugiyama E, Watanabe G, Owari T, Nishinakamura H, Sugiyama D, Maeda Y, Kawazoe A, Yukami H, Chida K, Ohara Y, Yoshida T, Shinno Y, Takeyasu Y, Shirasawa M, Nakama K, Aokage K, Suzuki J, Ishii G, Kuwata T, Sakamoto N, Kawazu M, Ueno T, Mori T, Yamazaki N, Tsuboi M, Yatabe Y, Kinoshita T, Doi T, Shitara K, Mano H, Nishikawa H. Lactic acid promotes PD-1 expression in regulatory T cells in highly glycolytic tumor microenvironments. *Cancer Cell* 2022; **40**: 201-218.e9 [PMID: 35090594 DOI: 10.1016/j.ccell.2022.01.001]

52   **Choi H**, Ko Y, Lee CY. Pro-cathepsin D as a diagnostic marker in differentiating malignant from benign pleural effusion: a retrospective cohort study. *BMC Cancer* 2020; **20**: 825 [PMID: 32867726 DOI: 10.1186/s12885-020-07327-w]

53   **Okuyama N**, Ide Y, Nakano M, Nakagawa T, Yamanaka K, Moriwaki K, Murata K, Ohigashi H, Yokoyama S, Eguchi H, Ishikawa O, Ito T, Kato M, Kasahara A, Kawano S, Gu J, Taniguchi N, Miyoshi E. Fucosylated haptoglobin is a novel marker for pancreatic cancer: a detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation. *Int J Cancer* 2006; **118**: 2803-2808 [PMID: 16385567 DOI: 10.1002/ijc.21728]

54   **Park SY**, Yoon SJ, Jeong YT, Kim JM, Kim JY, Bernert B, Ullman T, Itzkowitz SH, Kim JH, Hakomori SI. N-glycosylation status of beta-haptoglobin in sera of patients with colon cancer, chronic inflammatory diseases and normal subjects. *Int J Cancer* 2010; **126**: 142-155 [PMID: 19551866 DOI: 10.1002/ijc.24685]