Identification of plasma proteins associated with oesophageal cancer chemotherapeutic treatment outcomes using SWATH-MS

Naici Guo [a], Giorgos Minas [a], Silvia A. Synowsky [b], Margaret R. Dunne [c,d,e], Hasnain Ahmed [b,g], Rhiannon McShane [b,g], Anshul Bhardwaj [c], Noel E. Donlon [c,d], Cliona Lorton [c,d,f], Jacintha O'Sullivan [c,d], John V. Reynolds [c,d], Peter D. Caie [g], Sally L. Shirran [b], Andy G. Lynch [a,g], Alan J. Stewart [b,g,*], Swati Arya [b,g,*]

[a] School of Mathematics and Statistics, University of St Andrews, St Andrews KY16 9SS, United Kingdom
[b] Biomedical Sciences Research Complex, University of St Andrews, St Andrews KY16 9ST, United Kingdom
[c] Department of Surgery, Trinity Translational Medicine Institute, Trinity College Dublin, St James's Hospital, Dublin D08 W9RT, Ireland
[d] Trinity St James's Cancer Institute, St James's Hospital, Dublin D08 W9RT, Ireland
[e] Department of Applied Science, Technological University Dublin, Tallaght, Dublin 24 D24 FKT9, Ireland
[f] Our Lady's Hospice & Care Services, Harold's Cross, Dublin 6w, Ireland
[g] School of Medicine, University of St Andrews, St Andrews KY16 9TF, United Kingdom

ARTICLE INFO

ABSTRACT

Oesophageal adenocarcinoma (OAC) is an aggressive cancer with a five-year survival of <15%. Current chemotherapeutic strategies only benefit a minority (20-30%) of patients and there are no methods available to differentiate between responders and non-responders. We performed quantitative proteomics using Sequential Window Acquisition of all THeoretical fragment-ion spectra-Mass Spectrometry (SWATH-MS) on albumin/IgG-depleted and non-depleted plasma samples from 23 patients with locally advanced OAC prior to treatment. Individuals were grouped based on tumour regression (TRG) score (TRG1/2/3 vs TRG4/5) after chemotherapy, and differentially abundant proteins were compared. Protein depletion of highly abundant proteins led to the identification of around twice as many proteins. SWATH-MS revealed significant quantitative differences in the abundance of several proteins between the two groups. These included complement c1q subunit proteins, C1QA, C1QB and C1QC, which were of higher abundance in the low TRG group. Of those that were found to be of higher abundance in the high TRG group, glutathione S-transferase pi (GSTP1) exhibited the lowest *p*-value and highest classification accuracy and Cohen's kappa value. Concentrations of these proteins were further examined using ELISA-based assays. This study provides quantitative information relating to differences in the plasma proteome that underpin response to chemotherapeutic treatment in oesophageal cancers.

Significance: Oesophageal cancers, including oesophageal adenocarcinoma (OAC) and oesophageal gastric junction cancer (OGJ), are one of the leading causes of cancer mortality worldwide. Curative therapy consists of surgery, either alone or in combination with adjuvant or neoadjuvant chemotherapy or radiation, or combination chemoradiotherapy regimens. There are currently no clinico-pathological means of predicting which patients will benefit from chemotherapeutic treatments. There is therefore an urgent need to improve oesophageal cancer disease management and treatment strategies. This work compared proteomic differences in OAC patients who responded well to chemotherapy as compared to those who did not, using quantitative proteomics prior to treatment commencement. SWATH-MS analysis of plasma (with and without albumin/IgG-depletion) from OAC patients prior to chemotherapy was performed. This approach was adopted to determine whether depletion offered a significant improvement in peptide coverage. Resultant datasets demonstrated that depletion increased peptide coverage significantly. Additionally, there was good quantitative agreement between commonly observed peptides. Data analysis was performed by adopting both univariate as well as multivariate analysis strategies. Differentially abundant proteins were identified between treatment response groups based on tumour regression grade. Such proteins included complement C1q sub-components and GSTP1. This study provides a platform for further work, utilising larger sample sets across different treatment regimens for oesophageal

---

* Corresponding authors at: Both School of Medicine, University of St Andrews, St Andrews KY16 9TF, UK.
E-mail addresses: ajs21@st-andrews.ac.uk (A.J. Stewart), sa212@st-andrews.ac.uk (S. Arya).

cancer, that will aid the development of 'treatment response prediction assays' for stratification of OAC patients prior to chemotherapy.

1. Introduction

Oesophageal cancers, including oesophageal adenocarcinoma (OAC) and oesophageal gastric junction cancer (OGJ), are one of the leading causes of cancer mortality worldwide, with a poor 5-year survival rate (globally <20%; developed countries <10%) [1]. It is asymptomatic in early stages and has no biomarkers for its early detection which leads to late diagnosis and poor prognosis resulting in high mortality rate. In 2018, there were an estimated 85,000 new cases of OAC globally (age-standardised incidence rates 0.9 per 100,000, both sexes combined) [2]. The burden of OAC is expected to rise dramatically and surpass oesophageal squamous cell carcinoma, in high-income countries by 2020 [3], with an accompanying rise in the number of patients seeking treatment. Curative therapy consists of surgery, either alone or in combination with adjuvant or neoadjuvant chemotherapy or radiation, or combination chemoradiotherapy regimens (neoCRT) [4,5]. However, only a minority of patients respond to such treatments, meaning the majority (70–80%) of patients will experience treatment-related toxicity and delay to surgery with no clinical benefit [6,7]. There are currently no clinico-pathological means of predicting which patients will benefit from chemotherapeutic treatments. There is therefore an urgent need to improve oesophageal cancer disease management and treatment strategies.

Cancer progression and treatment response involve multi-step processes originating from interactions between the tumour and various host cell types and extracellular factors [8]. Moreover, persistent systemic inflammation has been implicated as a possible mechanism of resistance to neoCRT [9]. Circulating plasma proteins reflect systemic events such as inflammation and angiogenesis and are more amenable to routine interrogation than tumour tissues. Plasma proteins such as C-reactive protein (CRP; [10]), soluble interleukin-6 receptor (sIL6R; [11]) and vascular endothelial growth factor (VEGF; [12]) have all demonstrated prognostic potential in oesophageal cancer. In addition to this, there is evidence that plasma proteins can serve as potential prognostic markers for other forms of cancers such as transforming growth factor beta 1 (TGFβ1) in triple negative breast cancer [13].

Although useful, the previous studies have not comprehensively examined changes in the concentrations of plasma proteins prior to oesophageal cancer treatment. In this study we quantitatively profiled the plasma proteomes of individuals undergoing chemotherapy treatment. This includes (mostly MAGIC trial regimen using etoposide, cisplatin or oxaliplatin, and fluorouracil or capecitabine [14], or in a few cases the FLOT regimen consisting of docetaxel, oxaliplatin, leucovirin, and 5-fluorouracil [15] or Folfox regimen consisting of folinic acid 5-fluoruracil and oxaliplatin [16]. Despite some differences in the drug combinations these chemotherapeutic regimens are known to target rapidly dividing cells and have been shown to be useful for the treatment of OAC, leading to tumour regression. Quantitative proteomic analysis was performed using SWATH-MS (Sequential Window Acquisition of All Theoretical Mass Spectra), a data-independent acquisition (DIA)-based method. DIA-based quantitative proteomics represents a powerful tool for identifying the cellular and molecular changes that underlie a disease; this is particularly useful in understanding the pathophysiology of cancers and options for their treatment. Here we have used both albumin/IgG-depleted and non-depleted pre-treatment plasma samples to identify differentially abundant proteins in samples grouped according to Mandard tumour regression grade (TRG, with TRG 1 indicating no residual cancer cells, TRG 2 indicating rare residual cells, TRG 3 representing an increase in the number of residual cancer cells, but fibrosis still predominates, TRG 4 where cancer outgrows fibrosis, and TRG 5 represents a complete absence of regressive changes; [17]). The patients

were surgically treated (after chemo) and the TRG was performed on resected specimens. This work details proteomic differences that may underpin treatment responses. These in the future could be used to predict response to treatment and allow the development of diagnostic tools to guide personalised treatment options for oesophageal cancer.

2. Methods

2.1. Ethical approvals and clinical cohort information

This study was approved by the School of Medicine Ethics Committee, University of St Andrews (Ref: MD9202). Samples were obtained from the Upper Gastrointestinal Trinity Biobank, Trinity Translational Medicine Institute, Trinity College Dublin, St James' Hospital, Dublin, with approval from Tallaght/St James's Hospital committee (reference: 2011/27/01). All methods outlined were performed in accordance with the relevant ethical guidelines and regulations. A cohort of 23 patients (mostly males) with locally advanced OAC or OGJ cancers were included for clinical analysis, and these underwent proteomic analysis upon giving informed consent (see Table 1). This cohort received pre- and post-operative chemotherapy as per the MAGIC, FLOT or Folfox regimen. The cohort was grouped based on TRG status (TRG1/2/3 vs TRG 4/5). The clinical TNM (tumour/node/metastasis) stage for each patient where known is provided in Table S1.

2.2. Experimental design and statistical rationale for SWATH-MS

This experiment aimed to identify proteomic differences in plasma that may underlie (or potentially predict) chemotherapy treatment outcomes in OAC patients. Plasma was prepared from whole blood collected in EDTA Vacutainer tubes (BD, Wokingham, UK) taken after diagnosis but prior to chemotherapy. Both non-depleted and albumin/IgG-depleted plasma samples were used. Albumin/IgG depletion was performed according to manufacturer's instructions using albumin/IgG depletion spin trap columns (Cytiva, Marlborough, MA, USA). These were analysed independently and grouped based upon the TRG score (grades 1/2/3 vs grades 4/5) after treatment. It was not possible to group samples based on post-treatment survival time as detailed data for this parameter was not available for all samples. A sample-specific library was made by pooling all samples for a comprehensive representation of the cohort.

Table 1
Demographic information on the patient cohort by study group.

	TRG 1/2/3 (n = 8)	TRG 4 and 5 (n = 15)
Mean age (years)	62.9 ± 14.0	63.0 ± 8.9
Sex (no. of males)	7	15
Cancer type (no. of individuals)		
Oesophageal adenocarcinoma	6	13
Gastro-oesophageal junction cancer	2	2
Neo-adjuvant treatment (no. of individuals)		
MAGIC	6	13
FLOT	2	1
Folfox	0	1
TRG Score (no. of individuals)		
=1	3	0
=2	4	0
=3	1	0
=4	0	9
=5	0	6
Lymphovascular invasion (no. of individuals)	1	11
Disease recurrence (no. of individuals)	1	12

2.3. Sample preparation for mass spectrometry

Plasma containing 30 μg of total protein was prepared for mass spectrometric analysis. For this, plasma proteins were denatured in a final concentration of 5 M urea, 2% sodium deoxycholate and 50 mM ammonium bicarbonate. Proteins were reduced and alkylated with 5 mM tris (2-carboxyethyl) phosphine followed by 5 mM iodoacetamide. The reaction was quenched with 10 mM dithiothreitol. Samples were diluted with 50 mM ammonium bicarbonate to a final urea concentration of 1.5 M. The resulting samples were then digested with trypsin (0.2 μg/μl trypsin, 1:50 ratio (*w/w*); Promega, Southampton, UK) overnight at 30 °C. Finally, 0.5% (*v/v*) trifluoroacetic acid (TFA) was added. Peptides were desalted using a C18 SepPak cartridge (Waters, Elstree, UK). An aliquot of the peptides in each sample was kept separate for the SWATH analysis. The remaining peptides were pooled for a spectral library generation. The solvent in each sample was removed using a SpeedVac (Thermo Fisher Scientific, Loughborough, UK). SWATH-MS was performed on equal amounts of peptides (3 μg) from both albumin/IgG-depleted and non-depleted plasma.

2.4. LC-ESI-MSMS analysis for spectral library generation

Peptides used for the library generation were separated using high pH reversed phase fractionation. The pooled, dried peptides were then resuspended in 100 μl buffer A, consisting of 10 mM ammonium formate, 2% acetonitrile, pH 10.0. Peptides were then fractionated on a XBridge C18 column (4.6 × 100 mm, 5 μm, Waters) at 1 ml/min with the following gradient: linear gradient of 4–28% buffer B (10 mM ammonium formate, 90% acetonitrile, pH 10.0) for 36 min, then 28–50% B for 8 min, followed by 100% B for a further 5 min to wash the column, before re-equilibration in 100% buffer A for 10 min. Fractions of 0.5 ml were collected every 30 s. The UV chromatogram was inspected and fractions pooled to give 25 fractions across the elution profile. The pooled fractions were dried and resuspended in 0.1% FA for mass spectrometric analysis.

The fractions were analysed individually on a Sciex TripleTOF 5600+ system mass spectrometer (Sciex, Framingham, MA, USA) coupled to an Eksigent nanoLC AS-2/2Dplus system, in data dependent mode, to achieve in depth identification of proteins. Additionally, 1 μg of peptides from each individually digested sample for the SWATH analysis were combined and analysed in data dependent mode. Prior to mass spectrometric analysis, reference iRT peptides (Biognosys, Schlieren, Switzerland) were added to each sample according to the manufacturer's specifications to allow correction of retention times. The samples were loaded in loading buffer (2% acetonitrile, 0.05% trifluoroacetic acid) and bound to an Acclaim Pepmap 100 μm × 2 cm trap (Thermo Fisher Scientific), and washed for 10 min to waste, after which the trap was turned in-line with the analytical column (Acclaim Pepmap RSLC 75 μm × 15 cm). The analytical solvent system consisted of buffer A (2% acetonitrile, 0.1% FA in water) and buffer B (2% water, 0.1% FA in acetonitrile) at a flow rate of 300 nl/min, with the following gradient: linear 1–20% of buffer B over 90 min, linear 20–40% of buffer B over 30 min, linear 40–99% of buffer B over 10 min, isocratic 99% of buffer B for 5 min, linear 99–1% of buffer B over 2.5 min and isocratic 1% solvent buffer B for 12.5 min. The mass spectrometer was operated in data-dependent analysis (DDA) top 20 positive ion mode, with 250 and 150 ms acquisition time for the MS1 (*m/z* 400–1200) and MS2 (m/z 230–1800) scans respectively, and 15 s dynamic exclusion. Rolling collision energy with a collision energy spread of 5 eV was used for fragmentation. The data was searched against SwissProt (accessed January 2019), restricted to human proteins (42,407 proteins) using the Mascot Algorithm with the following search parameters: trypsin as the cleavage enzyme (/K-\P and /R-\P) and carbamidomethylation as a fixed modification of cysteines and oxidation as variable modification. Peptide mass tolerance was set to 20 ppm with fragment mass tolerance set to 0.1 Da. Note that the iRT peptides were included in this database.

2.5. SWATH-MS data acquisition

For SWATH-MS data acquisition, the same mass spectrometer and LC-MS/MS setup was used essentially as described above but operated in SWATH mode. The method uses 100 windows of variable Da effective isolation width with a 1 Da overlap as specified by Sciex. Each window has a dwell time of 150 ms to cover the mass range of 400–1250 *m/z* in TOF-MS mode and MS/MS data is acquired over a range of 230–1800 m/z with high sensitivity setting and a dwell time of 35 ms, resulting in a cycle time of 3.7 s. The collision energy for each window was set using the collision energy of a 2+ ion centred in the middle of the window with a spread of 5 eV.

2.6. Data processing and availability

Identified proteins within Mascot were imported into Skyline 21.10.146 for spectral library generation. SWATH-MS results were mapped against this library. Skyline parameters were chosen as follows: Peptide settings (trypsin digestion with 1 missed cleavage, human as background proteome, carbamidomethylation (C) and oxidation (M) as modifications); Transition settings (2^+ and 3^+ precursor charges, 1^+ and 2^+ ion charges, b and y ions and precursor. Six transitions were selected per peptide. The peak area was normalized based on the median value. Moderated *t*-tests with an empirical Bayes estimate of the variance were performed in R (version 4.1.1) with the limma (version 3.48.3) package [18]. The implementation of the Linear discriminant analysis (LDA) classifier with repeated stratified K-fold cross-validation was implemented with a scikit-learn python package (version 0.24.1) [19]. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE Partner Repository [20], with the dataset identifier PXD032137.

2.7. Enzyme-linked immunosorbent assays (ELISAs)

ELISAs were performed on the patient plasma samples using commercially available kits for human complement component C1q and human glutathione S-transferase pi-1 (GSTP1; Cloud-Clone Corp., Buckingham, UK). Samples were centrifuged at 13,000 ×g for 10 min and the supernatant used for analysis. The assay was performed in accordance with the manufacturer's instructions.

3. Results

3.1. Comparison of proteomic data from depleted vs non-depleted plasma samples taken from patients prior to undergoing chemotherapy for OAC

For SWATH-MS quantitation, cohort-specific reference spectral libraries were generated for both depleted and non-depleted sample sets (for use as a reference in generating peptide query parameters for the peptide centric analysis of the data) by data-dependent acquisition analysis of the plasma proteins. To determine the differences in levels of peptides/proteins present across the different TRG groups, we measured quantitative proteome profiles using albumin/IgG-depleted and non-depleted plasma for each individual sample by SWATH-MS. The resultant data were log$_2$ transformed and then normalized using median normalization at the protein and peptide level for non-depleted data and depleted data, respectively. As shown in Figs. S1, S2, normalization effectively scaled the patient samples to the same median and stabilised variance across the mean abundance of proteins and peptides. Comparison of the obtained quantitative data identified 1640 and 2870 peptides (1% FDR) in non-depleted and depleted samples, respectively. These represented 330 proteins in non-depleted samples and 537 proteins in depleted samples. There was a significant overlap between the two datasets where 1529 peptides (from 295 proteins) were consistently found in both non-depleted and depleted data (Fig. 1A). The Bland-Altman analysis was performed for estimating the degree of
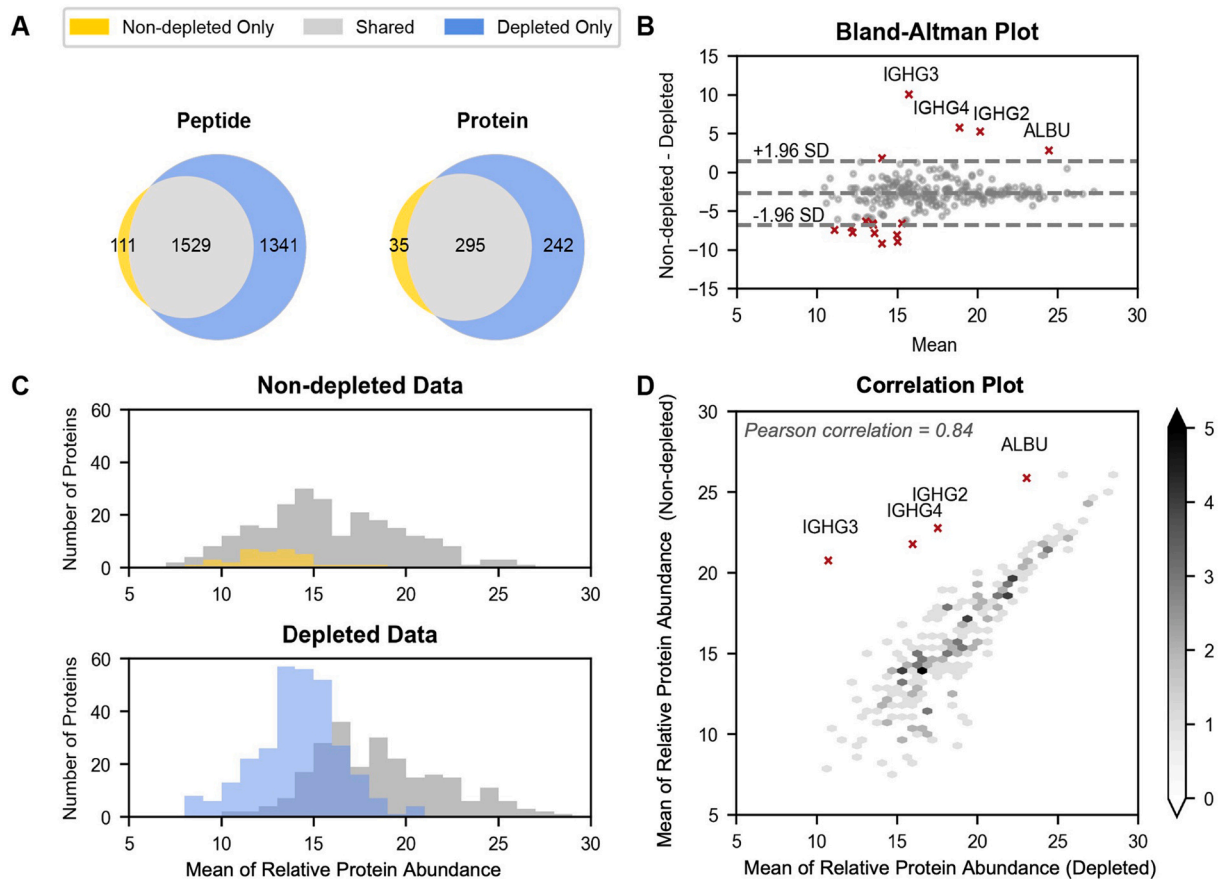
Fig. 1. Comparison of proteomic data from depleted vs. non-depleted plasma samples. Grey represents peptides or proteins that are detected in both non-depleted data and depleted data. Yellow and blue indicate peptides or proteins that are uniquely existing in non-depleted data and depleted data, respectively. A. Venn diagrams of 1640 and 2870 peptides, 330 and 537 proteins in non-depleted data and depleted data, respectively. B. Bland-Altman Plot showing agreement between depleted and non-depleted data ($\log_2$ relative abundance) for commonly found proteins. C. Overlay histograms of means of commonly found proteins from non-depleted data and depleted data. D. Scatterplot coloured by density shows the correlation between depleted and non-depleted data for all common proteins. Note that albumin and immunoglobins have a higher abundance in non-depleted data than in depleted data as expected. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
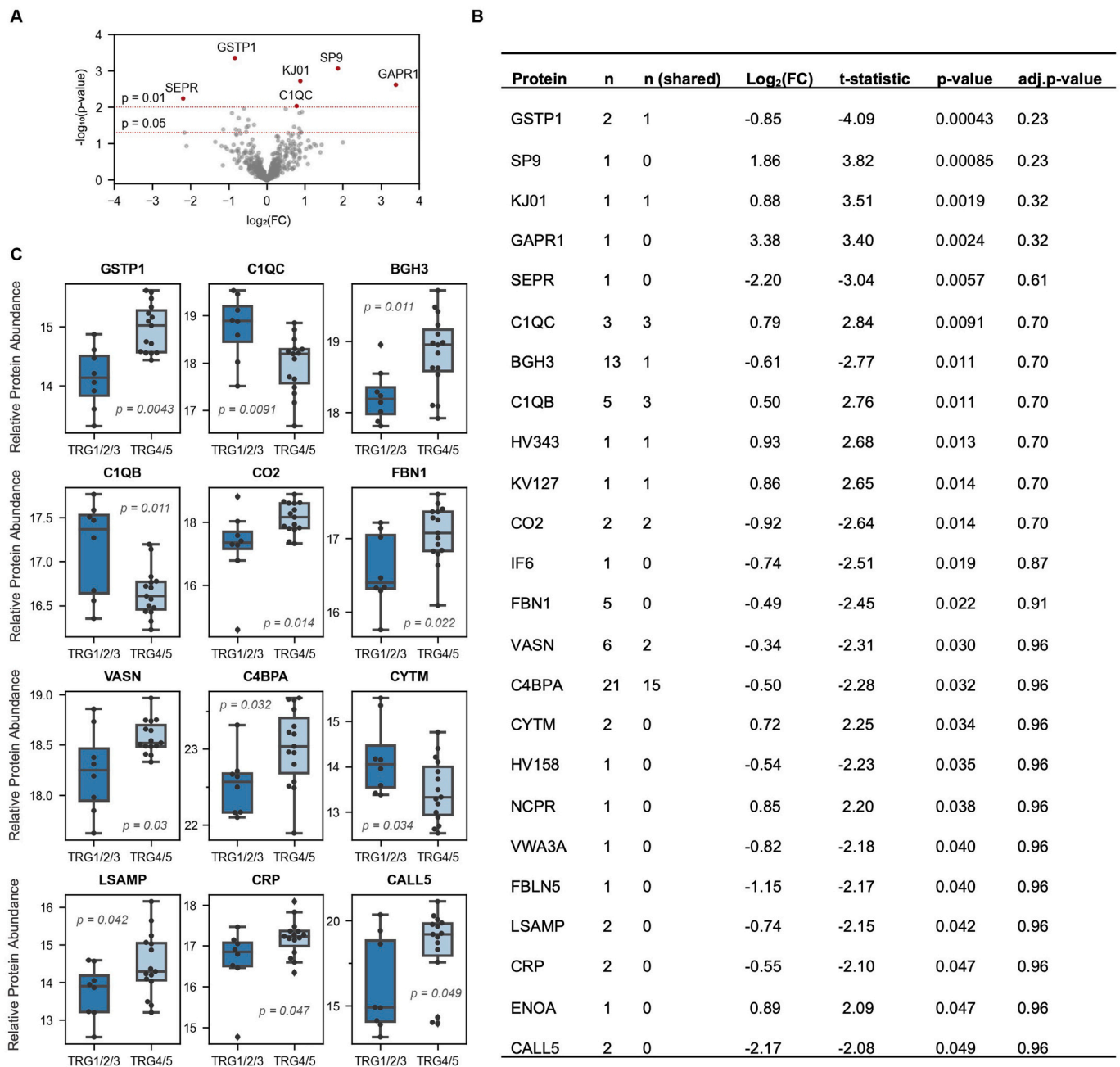
concordance between the measurement of relative protein abundance level for commonly found 295 proteins. As shown in Fig. 1B, the mean difference is -2.7, which implies, on average depleted samples have 2.7 units higher in relative protein abundance than non-depleted samples. Most of the proteins lie in 95% limits of agreement, which are [-6.85, 1.45], except for outlier proteins that are likely to have been affected by depletion, as shown in Table S2. Comparing the distributions of the means of proteins between non-depleted and depleted data (Fig. 1C), consistently found proteins showed higher mean in depleted data than non-depleted data. The correlation of relative protein abundance for consistently found proteins between non-depleted and depleted samples is shown in Fig. 1D (Pearson correlation coefficient $= 0.84$). Following both protein and peptide normalization steps, serum albumin and immunoglobin proteins/peptides were relatively more abundant in the non-depleted data than in the depleted data, as expected.

3.2. Quantitative analysis of proteins from plasma taken from OAC patients prior to undergoing chemotherapy

3.2.1. Univariate analysis

Relative protein abundance in non-depleted plasma samples was compared between low and high TRG groups (Fig. S4). Note that in our study most patients exhibited tumour staging at T3 (specifically, 18 out of the 23 patients). As such it is extremely unlikely that this had any implications on treatment success prediction analysis in our study. Of the proteins consistently quantified, the relative abundance of a total of

14 unique proteins were statistically significantly altered (*p*-value <0.05) between TRG1/2/3 and TRG4/5 groups. The quantitative data for all differentially abundant proteins is presented. The proteins hemoglobin subunit beta (HBB), complement C1r like subcomponent (C1RL) and Fc gamma binding protein (FCGBP), where quantitation was based on more than one peptide, exhibited differences in relative abundance between the two groups. Relative protein abundance in depleted plasma samples was also compared between low and high TRG groups. A volcano plot showing a comparison of the relative abundance of plasma proteins between TRG1/2/3 and TRG4/5 is presented in Fig. 2A. Of the proteins consistently quantified, the relative abundance of a total of 24 unique proteins (12 based on data from more than one peptide) were statistically significantly altered (*p*-value <0.05) between TRG1/2/3 and TRG4/5 groups. Multiple hypothesis testing was also performed with the respective q-values calculated. However, these were too stringent for this study given the sample size and as such were not used in determining significance. Adjusted *p*-values (q-values) were computed to control the False Discovery Rate (FDR) using the Benjamini-Hochberg method. The quantitative data for all differentially abundant proteins is presented in Fig. 2B. Box-and-whisker plots showing the relative abundance of differentially abundant proteins in the two groups where data is based on more than one peptide [glutathione S-transferase pi (GSTP1), complement c1q subunit C (C1QC), transforming growth factor-beta induced protein (BGH3), complement c1q subunit B (C1QB), cytochrome C oxidase II (CO2), fibrillin-1 (FBN1), vasorin (VASN), complement 4 binding protein alpha (C4BPA), cystatin

A

B

Protein	n	n (shared)	$\text{Log}_2(\text{FC})$	t-statistic	p-value	adj.p-value
GSTP1	2	1	-0.85	-4.09	0.00043	0.23
SP9	1	0	1.86	3.82	0.00085	0.23
KJ01	1	1	0.88	3.51	0.0019	0.32
GAPR1	1	0	3.38	3.40	0.0024	0.32
SEPR	1	0	-2.20	-3.04	0.0057	0.61
C1QC	3	3	0.79	2.84	0.0091	0.70
BGH3	13	1	-0.61	-2.77	0.011	0.70
C1QB	5	3	0.50	2.76	0.011	0.70
HV343	1	1	0.93	2.68	0.013	0.70
KV127	1	1	0.86	2.65	0.014	0.70
CO2	2	2	-0.92	-2.64	0.014	0.70
IF6	1	0	-0.74	-2.51	0.019	0.87
FBN1	5	0	-0.49	-2.45	0.022	0.91
VASN	6	2	-0.34	-2.31	0.030	0.96
C4BPA	21	15	-0.50	-2.28	0.032	0.96
CYTM	2	0	0.72	2.25	0.034	0.96
HV158	1	0	-0.54	-2.23	0.035	0.96
NCPR	1	0	0.85	2.20	0.038	0.96
VWA3A	1	0	-0.82	-2.18	0.040	0.96
FBLN5	1	0	-1.15	-2.17	0.040	0.96
LSAMP	2	0	-0.74	-2.15	0.042	0.96
CRP	2	0	-0.55	-2.10	0.047	0.96
ENOA	1	0	0.89	2.09	0.047	0.96
CALL5	2	0	-2.17	-2.08	0.049	0.96

C

Fig. 2. Relative abundance of proteins in depleted plasma based on the univariate analysis comparing samples from individuals with low (1/2/3) and high (4/5) TRG scores. A. Volcano plot showing $\log_2$-fold change (x-axis) versus negative $\log_{10}$ of *p*-value (y-axis). Red highlighted dots represent proteins with p-value <0.01 identified by the empirical Bayes test. B. Quantitative data for proteins with p-value <0.05. The columns represent the following: n = number of peptides identified from depleted plasma sample; n (shared) = number of peptides that are commonly found in non-depleted plasma samples. C. Boxplots showing the relative expressions of differentially expressed proteins between TRG1/2/3 and TRG4/5 groups. Only those proteins where quantification is based on more than one peptide are shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

M (CYTM), limbic system associated membrane protein (LSAMP), C-reactive protein (CRP) and calmodulin-like protein 5 (CALL5)], is shown in Fig. 2C.
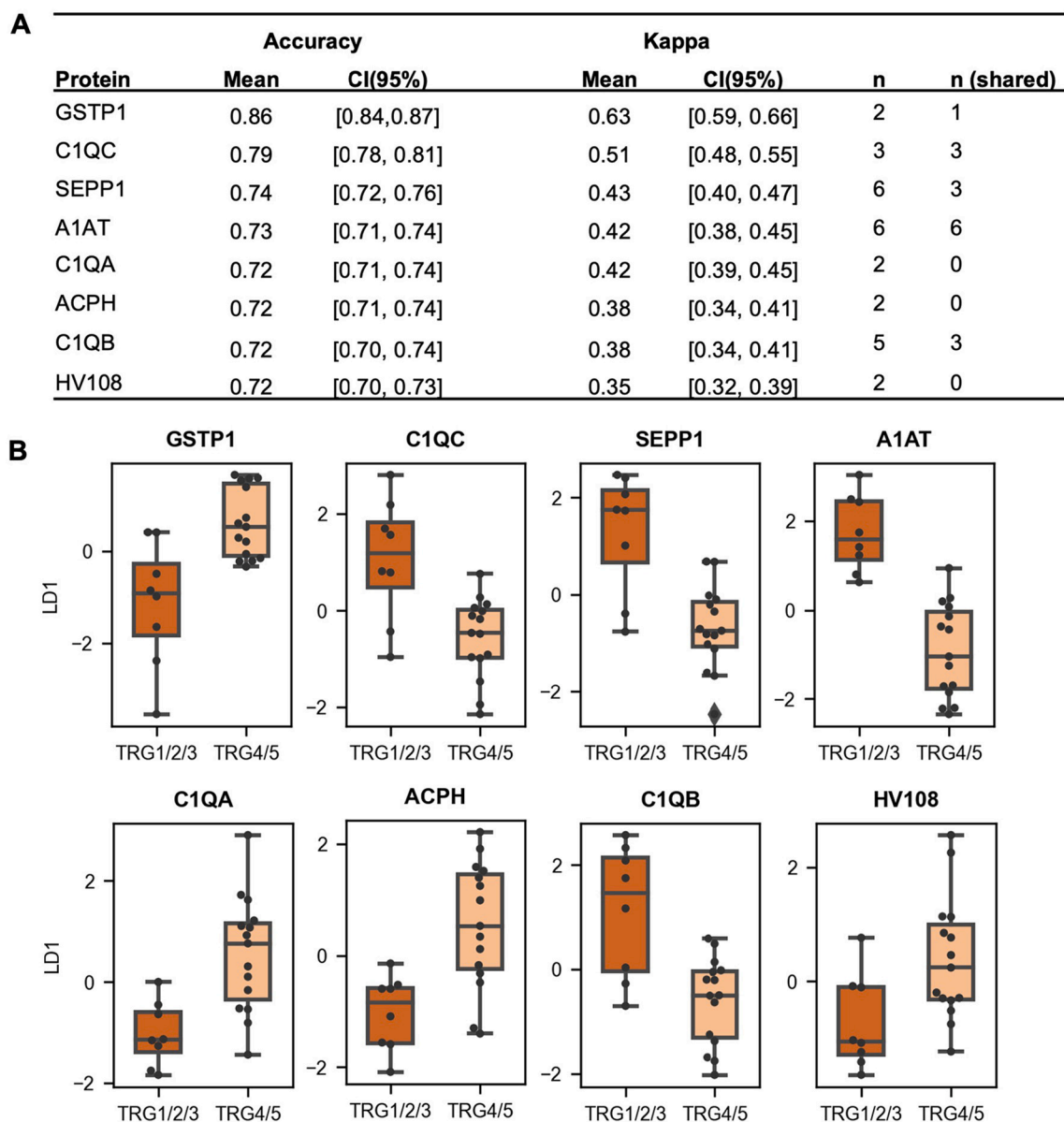
3.2.2. Multivariate analysis

In addition to the univariate analysis, the LDA classifier that predicts sample class based on multiple predictor variables is applied. To fit this classifier, relative peptide abundance that are from the same protein are considered as predictor variables and TRG1/2/3 and TRG4/5 as two sample classes. Cross-validation was used to test the ability of the LDA classifier to make predictions on data that are not used for fitting the model. Thus, for each protein, an LDA model is fitted with a repeated stratified K-fold cross-validator to examine the model performance. Patient samples are randomly partitioned into K (K = 4) segments. K-1 segments (75% of patients) are used for training the LDA model and the remaining one segment (25% of patients) is used as a held-out segment for testing the predictions. The testing here is a comparison of predictions with actual TRG groups. The fraction of correctly classified samples is recorded as a classification accuracy score, e.g., a classification accuracy score of 0.8 implies that 80% of patients in the held-out segment are allocated correctly. The partitioning into K groups of samples is done using a stratified sampler to ensure the proportional

representation of the low and high TRG patients in the K segments. After K iterations, each K segment has been held out for testing, thus K classification accuracy scores are generated for each cross-validator. This K-fold cross-validation by stratified-random partitioning is repeated 100 times. The accuracy score presented in Fig. S4A and Fig. 3A represents the average of 400 classification accuracy scores that are from 100 repeats of four held-out testing sets. Cohen's kappa statistics, which compares the accuracy of LDA with an accuracy of purely random guessing is also computed [21]. The average kappa statistics and confidence interval are calculated to measure the consistency of predictions in 400 held out sets, ranging from −1(total disagreement) through 0 (random classification to 1(total agreement). Three proteins exhibited an average Cohen's kappa value >0.35 in non-depleted plasma samples; these were IGI, APOD and TETN. Box-and-whisker plots showing the

relative abundance of these three proteins is shown in Fig. S5B. Multi-variate analysis of the depleted plasma peptide data was also performed using the LDA model. Fourteen proteins exhibited an average kappa value >0.35. These were GSTP1, Golgi-associated plant pathogenesis related protein 1 (GAPR1), immunoglobulin kappa joining 1 (KJ01), Sp9 transcription factor (SP9), C1QC, alpha-1 antitrypsin (A1AT), prolyl endopeptidase FAP (SEPR), selenoprotein P (SEPP1), complement c1q subunit A (C1QA), immunoglobulin kappa variable 1–27 (KV127), alpha-enolase (ENOA), immunoglobulin heavy-chain variable 1–8 protein (HV108), acyl carrier protein phosphodiesterase (ACPH) and C1QB (Fig. 3A). Box-and-whisker plots showing the relative abundance of proteins, where data is based on more than one peptide is shown in Fig. 3B. Interestingly, nine proteins displayed both significant differential abundance at the protein level and high accuracy at the peptide level

A

Protein	Accuracy		Kappa		n	n (shared)
	Mean	CI(95%)	Mean	CI(95%)		
GSTP1	0.86	[0.84, 0.87]	0.63	[0.59, 0.66]	2	1
C1QC	0.79	[0.78, 0.81]	0.51	[0.48, 0.55]	3	3
SEPP1	0.74	[0.72, 0.76]	0.43	[0.40, 0.47]	6	3
A1AT	0.73	[0.71, 0.74]	0.42	[0.38, 0.45]	6	6
C1QA	0.72	[0.71, 0.74]	0.42	[0.39, 0.45]	2	0
ACPH	0.72	[0.71, 0.74]	0.38	[0.34, 0.41]	2	0
C1QB	0.72	[0.70, 0.74]	0.38	[0.34, 0.41]	5	3
HV108	0.72	[0.70, 0.73]	0.35	[0.32, 0.39]	2	0

B

Fig. 3. Prediction of low or high TRG score groups from proteins with multiple peptides in the depleted data. A linear discriminant analysis (LDA) classifier is applied to all peptides from the protein. A. Table presents the results of the LDA classifier for the proteins that have the highest average Cohen's kappa. The proteins are ranked by the average Cohen's kappa and the top hits are listed. The columns represent the following: CI = confidence interval; n = the number of peptides identified from depleted plasma sample; n (shared) = the number of peptides that are commonly found in non-depleted plasma samples. B. Linear discriminant scores regarding low (1/2/3) and high (4/5) TRG scores groups. Significant differences along the first linear discriminant LD1 were detected for all the selected proteins, suggesting that low and high TRG score groups can be very well separated by the weighted sum of relative peptide expression level from each protein, as shown in Fig. S6.

when comparing the two groups. These include GSTP1, GAPR1, KJ01, SP9, C1QC, SEPR, KV127, ENOA and C1QB. A list of all differentially abundant proteins (abbreviations and protein descriptions) is provided in Table S3.

3.3. Functional annotation of differentially expressed proteins

Pathway enrichment analysis of plasma proteins (depleted and non-depleted) displaying differential abundance between TRG 1/2/3 vs TRG 4/5 groups was performed using Metascape (https://metascape.org; [22]). This is an over-representation analysis that statistically determines whether proteins from certain pathways are enriched in the data. Top clusters along with their representative enriched terms were selected. Resultant regulated pathways in TRG 1/2/3 group are shown in Fig. 4A and those that are regulated in TRG 4/5 are shown in Fig. 4B. The results show that the proteins that are expressed at higher levels in patients under TRG 1/2/3 group are enriched in biological processes related to "C1Q complex", "Platelet degranulation", "Selenium micronutrient network", "Negative regulation of endopeptidase activity" and "Monocarboxylic acid metabolic process". Proteins that are of higher abundance in patients within the TRG4/5 group are involved in processes such as "Regulation of superoxide metabolic process", "Naba extracellular matrix glycoproteins", "Complement cascade", "Response to reactive oxygen species", "Angiogenesis", "Negative regulation of Immune system", "Hemostasis", "Negative regulation of cell differentiation" and "Response to hormone".

3.4. Determination of plasma concentrations of complement component C1q complex in patient samples

Of the proteins that exhibited differing mean relative abundances between groups, C1q (complex) and GSTP1 were chosen for further analysis. To validate the findings from the SWATH-MS analysis, ELISA assays were carried out on the non-depleted plasma samples. The resultant data are shown in Fig. 5. Statistically significant differences in the concentrations of C1q complex and GSTP1 were observed in agreement with the SWATH-MS data. Moreover, the predictive

performance of c1q complex was examined and compared with C1QA, C1QB and C1QC subunits. As illustrated in Fig. 5B, greater separation along the first linear discriminant was observed when using relative abundance of all the peptides from c1q protein complex as features.

4. Discussion

To our knowledge, this is the first SWATH-MS-based clinical study to directly compare results obtained from both depleted and non-depleted plasma samples. Depletion led to the identification and relative quantitation of many more peptides/proteins (~2-times as many) than could be observed when analysing the non-depleted plasma. We assume that removal of highly abundant proteins such as albumin and IgG facilitated the enhanced detection of lower abundant peptides, thereby increasing proteomic coverage. A comparison of peptides identified in both depleted and non-depleted samples showed a high level of agreement in the measurement obtained using linear regression, suggesting that although depletion would be expected to alter the variability of some proteins (i.e., those that bound to the column which included albumin, IgG proteins and ATP binding cassette subfamily member 1) it did not significantly alter the relative abundance of the vast majority of proteins in plasma. Thus, our study supports the utilisation of protein depletion as a viable strategy to gain higher peptide/protein coverage in plasma using SWATH-MS or similar approaches. However, one of the limitations of the depletion approach is that the relative quantitation of depleted proteins (i.e. albumin or IgGs) is unlikely to be reflective of the pre-depletion levels of these specific proteins. It is noteworthy that we did not observe any differential abundance of these proteins between high and low TRG groups in our non-depleted samples.

It was evident from the clinical data that patients who would go on to have a higher TRG score (which equates to a lower degree of tumour regression after chemotherapy) were more likely to have tumours that developed lymphovascular invasion and have recurrence of disease. To examine whether there are any plasma proteomic differences between groups with high and low TRG, the quantitative proteomic data for individuals in each group were compared and differences in several proteins were observed between these groups. Univariate analysis of the
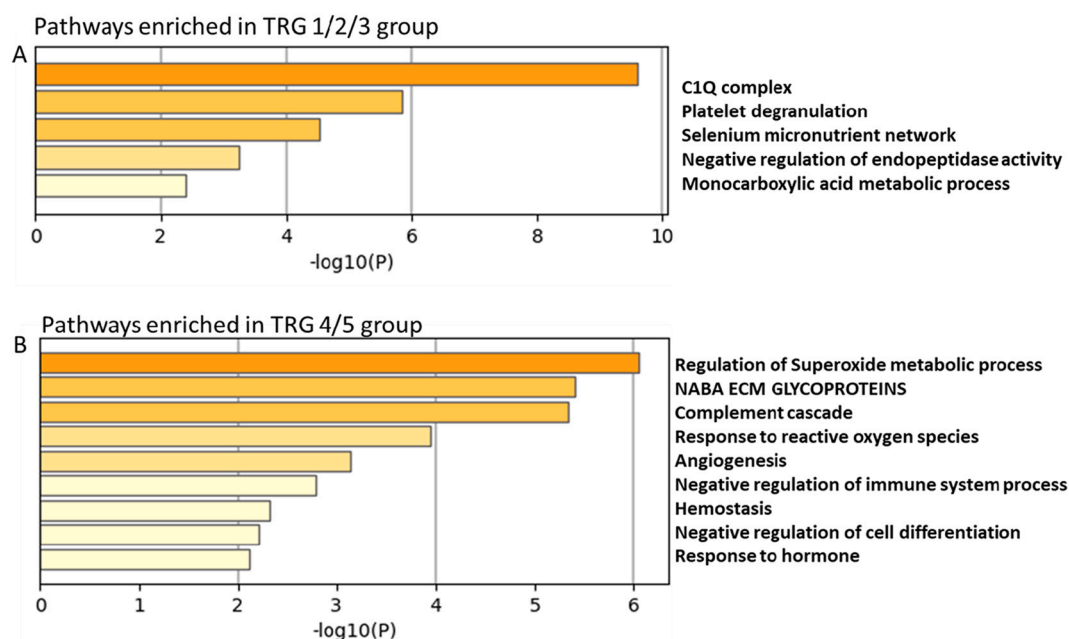
Fig. 4. Enriched Ontology Clusters of significantly altered proteins.
Enriched pathway clusters developed from all plasma proteins that display significantly altered expression levels between individuals with low (1/2/3) and high (4/5) TRG scores. A. Pathways related to all upregulated proteins in individuals with TRG1/2/3 and B. Pathways related to all upregulated proteins in individuals with TRG 4/5.
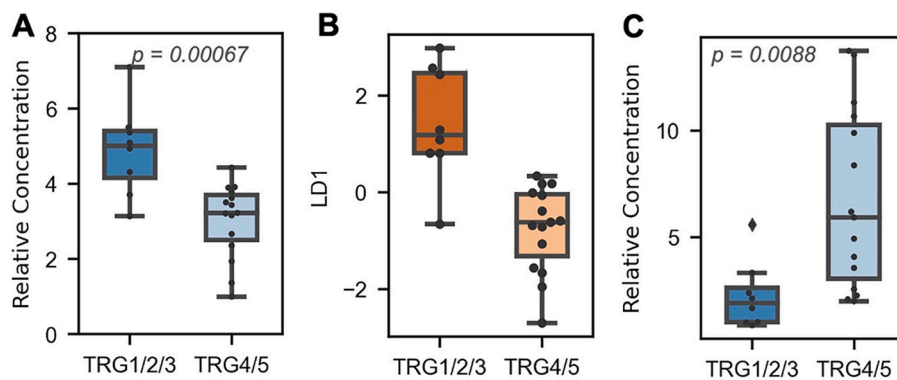
Fig. 5. Quantitative analysis of C1q and GSTP1 in plasma samples by enzyme-linked immunosorbent assay (ELISA) and comparison of C1q protein complex and subunit proteins in respect of predicting TRG score groups. Statistical comparisons were carried out by an independent two-sample *t*-test. A. Relative concentrations of C1q in samples associated with a low (1/2/3) vs high TRG score (4/5). B. Linear discriminant scores regarding low (1/2/3) and high (4/5) TRG scores for C1q protein complex. Compared with individual C1q subunit proteins, greater difference was observed along LD1 axis, indicating improved classification performance based on weighted sum of relative peptide abundance level from c1q complex. C. Relative concentrations of GSTP1 in samples associated with a low (1/2/3) vs high TRG score (4/5).

non-depleted data revealed 14 plasma proteins that had significantly different abundance between those with a low TRG compared to a high TRG (*p*-value <0.05). However, it is important to add that the results of 11 of these were based on only one detected peptide. Multivariate analysis of proteins (where more than one peptide was detected) was also performed but did not identify any proteins with high classification accuracy in the non-depleted dataset.

With the albumin/IgG-depleted sample set, we obtained much richer data characterised by increased peptide/protein coverage and a higher signal-to-noise ratio, which allowed the identification of differentially abundant proteins between low and high TRG groups. Differentially abundant proteins based on the measurement of multiple peptides were prioritised for further discussion here. Proteins that exhibited differential abundance between groups using multi-variate analysis included GSTP1, SEPR, BGH3, CO2 and eukaryotic translation initiation factor 6 (IF6), which were present at a lower concentration, and C1QA, C1QB, C1QC, SP9, GAPR1, immunoglobulin heavy variable 3–43 (HV343) and KV127, which were present at a higher concentration in patients with a lower TRG score after chemotherapy. Among these, the three C1q proteins (C1QA, C1QB and C1QC) were perhaps the most notable as they together form the functional C1q complex. Complement factor C1q, together with C1r and C1s form the C1 complex, the first component in the classical complement pathway. C1q is composed of six identical subunits, each of which is composed of three different polypeptide chains, C1qA, C1qB and C1qC [23]. All three C1q component proteins were more abundant in the plasma of individuals with low TRG scores compared to high TRG scores, suggesting it may play a role in reducing oesophageal cancer severity and could serve as a positive prognostic indicator for this cancer type. Using an ELISA-based assay, the levels of C1q complex were also found to be higher in samples from those who would go on to develop a lower TRG score after chemotherapy. Activation of the classical pathway occurs when C1q binds to an antigen-antibody complex [23]. However, C1q can also exert functions unrelated to complement activation. In cancer, C1q is expressed in the stroma and vascular endothelium of several human malignant tumours [24]. Other proteomic studies have identified C1q as a potential prognostic indicator. In lung and kidney cancer and in glioma, higher C1q levels are associated with a negative prognosis [25,26]. Conversely, in breast cancer and some types of lung cancer, as with our study, higher C1q levels are associated with a more positive prognosis [27,28]. C1q is an important modulator of inflammation and cytokine secretion. What is not clear is whether (and when) C1q is protective against or supportive of cancer progression [27]. C1q can be detrimental to cancer cell viability via cell lytic mechanisms [29,30]. However, C1q can also exert tumour-promoting functions independent of classical pathway activation, as observed for some types of cancers [23,31,32].

Glutathione S-transferase pi-1 (GSTP1) was identified from our multivariate analysis as the protein exhibiting most significantly different abundance in plasma between groups (highest accuracy and Cohen's kappa value). This finding was further corroborated by ELISA using non-depleted plasma samples. Higher GSTP1 levels were found in individuals with a high TRG score and was thus associated with a poorer prognosis from those in our study. GSTP1 is a member of a family of enzymes that play an important role in detoxification by catalysing the conjugation of hydrophobic and electrophilic compounds with reduced glutathione [33]. *GSTP1* is a polymorphic gene encoding active, functionally different GSTP1 variant proteins that function in xenobiotic metabolism and play a role in susceptibility to cancer, and other diseases [33]. Various members of the glutathione S-transferase (GST) family have been reported as being overexpressed in several cancers and in most cases have been linked to poor prognosis and chemoresistance [34–37]. GSTP1 is involved in cell maintenance, cell survival and the cellular stress response via the NF-κB and MAP kinase pathways during tumour progression. Furthermore, *GSTP1* knockdown in cervical cancer cells showed that evasion of apoptosis was affected, and tumour proliferation was significantly reduced [38]. Further work is required to fully understand the role of GSTP1 and its prognostic relevance in OAC and other cancer types.

Among the other proteins that displayed altered abundance between the two groups based on the univariate analysis, BGH3, CO2, C4BPA, LSAMP, CALL5, CRP, FBN1, and VASN were found to be of higher abundance on average in plasma from patients with a high TRG score after treatment. BGH3 is thought to play a role in angiogenesis which is important for tumour growth and expansion [39]. A microarray study conducted by He et al. identified significant up-regulation of BGH3 in oesophageal squamous cell carcinoma (OSCC) [40]. CO2 is predicted to be involved in regulation of mitochondrial electron transport and has been implicated in cancer progression [41]. C4BP (complement component 4 binding protein) is a glycoprotein responsible for regulating the classical pathway of the complement system, which exists as distinct isoforms made up of alpha and beta subunits; C4BPA, which exhibited differential plasma abundance between groups in our study, encodes the alpha chain [42]. A recent study by Sogawa et al. demonstrated that the sera levels of C4BPA was significantly higher in the preoperative Pancreatic ductal adenocarcinoma (PDAC) patients than in the postoperative ones, C4BPA was also significantly higher in PDAC patients in comparison to healthy controls [43]. LSAMP has been reported to be a candidate tumour suppressor gene in a range of cancer types [44–48]. Interestingly, the downregulation of LSAMP expression in cancer cell lines is associated with poor survival prognosis by promoting lung cancer progression [49]. CALL5 is also thought to have tumour suppressor activities and is silenced at an early stage of carcinogenesis in squamous cell carcinoma of uterine cervix [50]. However, in our study a higher abundance of LSAMP and CALL5 in plasma were associated with a relatively poorer prognosis for oesophageal cancer. How plasma LSAMP and CALL5 levels relate to the cellular expression/activities of these proteins is unknown but it is possible that they may not correlate positively. CRP is a classical inflammatory marker

identified to be related to progression of oesophageal cancer and, consistent with our study, elevated pre-treatment serum CRP levels are associated with poorer prognosis in oesophageal cancer patients [51]. FBN1 is an extracellular matrix glycoprotein that provides structural and regulatory support to both elastic and non-elastic connective tissues via calcium-binding microfibrils [52]. This protein has previously been associated with poor survival in ovarian cancer [53] and been proposed as a germ cell tumour marker [54]. VASN is a classic type 1 trans-membrane protein which is linked with vascular injury repair and is overexpressed in a range of tumours [55]. Serum vasorin levels are upregulated in patients with colon cancer, with levels further increasing as cancer progresses [56]. In addition to the C1q subunit proteins, uni-variate analysis revealed one other protein found to be present at higher abundance in plasma from patients with a low TRG score after treatment, CYTM. Several studies have found this protein to have anti-tumour activity and to be a positive prognostic indicator for treatment of various cancers [57–59].

From the multivariate analysis of proteomic data, proteins that displayed altered abundance between the two groups (other than GSTP1 and C1q subunit proteins) included ACPH and HV108, which were found to be of lower abundance in those with a low TRG score after treatment, and A1AT and SEPP1 which were more abundant in plasma from patients in this group. ACPH and HV108 have not, to our knowledge, previously been linked to cancer progression or prognosis. A1AT is a circulating glycoprotein that inhibits neutrophil elastase and other serine proteases in blood and tissues. The role of this protein in cancer progression is controversial, while deficiency of A1AT is a risk factor for several cancer types, increased plasma/serum expression has been reported for multiple cancers compared to healthy controls [60]. It has also been reported that A1AT levels in the blood are significantly decreased in non-small cell lung cancer and prostate cancer after chemotherapeutic treatment compared with those before the treatment started [61]. SEPP1 is a protein that is positively associated with cancer risk and can regulate tumorigenesis and progression through its effects on cancer-related signalling pathways [62]. However, a protective effect of selenium supplementation on OAC risk, mediated through the anti-oxidant activity of selenoenzymes, has been reported [63].

Functional annotation of proteins displaying differential abundance between the groups highlighted several pathways that were significantly represented by these proteins. Unsurprisingly, "C1q complex" was the most significantly enriched pathway, being associated with proteins upregulated in the TRG1/2/3 group. Interestingly, pathways classically associated with cancer progression including angiogenesis and negative regulation of both the immune system and cell differentiation were highlighted as involving upregulated proteins in the TRG4/5 group and thus, were associated with poorer treatment outcomes. Pathways associated with oxidative stress ("Regulation of superoxide metabolic process" and "response to reactive oxygen species") were enriched in the TRG4/5 group. This is interesting as generation of reactive oxygen species and an altered redox status are common biochemical aspects in cancer cells and are known to be associated with drug-resistant mechanisms [64–66]. Several studies have suggested that targeting oxidative stress in OAC treatment is vital to maximising positive outcomes [67,68]. Although complement C1q was associated with better treatment outcomes, "Complement cascade" was identified as a functional group for proteins associated with poor treatment outcomes. This was interesting as a previous (serum) proteomic study by our group showed that pre-treatment serum C4a and C3a levels were significantly higher in poor responders versus good responders [69].

5. Conclusions

In conclusion, this study has utilised a combination of quantitative proteomics and statistical methods to identify proteins that have predictive power to display differential abundance between OAC patients that respond favourably and less favourably to chemotherapy. In our study, we analysed both albumin/IgG-depleted and non-depleted plasma in parallel. It was clear that depletion of very highly abundant proteins led to an improvement in signal-to-noise and the consequent detection and quantitation of many more proteins. A direct comparison between proteins common to both depleted and non-depleted sample sets showed a high degree of agreement. From the proteins that exhibited differential abundance, the c1q family proteins and GSTP1 had the highest potential to predict likelihood of tumour regression in response to neoadjuvant chemotherapy in OAC. These observations point towards the use of such markers for stratifying patients prior to treatment and the further use of quantitative proteomics in personalised medicine to predict treatment outcomes. This study provides a platform for further work, utilising larger sample sets across different treatment regimens for oesophageal cancer, that will aid the development of prognostic assays for use in the clinic.

Data availability

The proteomic data are available and have been deposited to the ProteomeXchange Consortium via the PRIDE Partner Repository with the dataset identifier PXD032137.

Declaration of Competing Interest

The authors have no competing interests to declare.

Data availability

No

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2022.104684.

References

[1] GBD 2017 Oesophageal Cancer Collaborators, The global, regional, and national burden of oesophageal cancer and its attributable risk factors in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017, Lancet Gastroenterol. Hepatol. 5 (2020) 582–597.

[2] M. Arnold, J. Ferlay, M.I. van Berge Henegouwen, I. Soerjomataram, Global burden of oesophageal and gastric cancer by histology and subsite in 2018, Gut 69 (2020) 1564–1571.

[3] M. Arnold, M. Laversanne, L.M. Brown, S.S. Devesa, F. Bray, Predicting the future burden of esophageal cancer by histological subtype: international trends in incidence up to 2030, Am. J. Gastroenterol. 112 (2017) 1247–1255.

[4] N.E. Donlon, M. Davern, A. Sheppard, R. Power, F. O'Connell, A.B. Heeran, R. King, C. Hayes, A. Bhardwaj, J.J. Phelan, M.R. Dunne, N. Ravi, C.L. Donohoe, J. O'Sullivan, J.V. Reynolds, J. Lysaght, The prognostic value of the lymph node in oesophageal adenocarcinoma; incorporating clinicopathological and immunological profiling, Cancers 13 (2021) 4005.

[5] J.I. Geh, S.J. Bond, S.M. Bentzen, R. Glynne-Jones, Systematic overview of preoperative (neoadjuvant) chemoradiotherapy trials in oesophageal cancer: evidence of a radiation and chemotherapy dose response, Radiother. Oncol. 78 (2006) 236–244.

[6] D.P. Kelsen, R. Ginsberg, T.F. Pajak, D.G. Sheahan, L. Gunderson, J. Mortimer, N. Estes, D.G. Haller, J. Ajani, W. Kocha, B.D. Minsky, J.A. Roth, Chemotherapy followed by surgery compared with surgery alone for localized esophageal cancer, N. Engl. J. Med. 339 (1998) 1979–1984.

[7] R.J. Korst, Kansler Al, J.L. Port, P.C. Lee, Y. Kerem, N.K. Altorki, Downstaging of T or N predicts long-term survival after preoperative chemotherapy and radical resection for esophageal carcinoma, Ann. Thorac. Surg. 82 (2006) 480–484.

[8] F. Castro-Giner, N. Aceto, Tracking cancer progression: from circulating tumor cells to metastasis, Genome Med. 12 (2020) 13.

[9] S. Marcone, A. Buckley, C.J. Ryan, M. McCabe, N. Lynam-Lennon, D. Matallanas, J. O'Sullivan, S. Kennedy, Proteomic signatures of radioresistance: Alteration of inflammation, angiogenesis and metabolism-related factors in radioresistant oesophageal adenocarcinoma, Cancer Treat. Res. Commun. 27 (2021) 100376.

[10] C.M. Lorton, L. Higgins, N. O'Donoghue, C. Donohoe, J. O'Connell, D. Mockler, J. V. Reynolds, D. Walsh, J. Lysaght, C-reactive protein and C-reactive protein-based scores to predict survival in esophageal and junctional adenocarcinoma: systematic review and meta-analysis, Ann. Surg. Oncol. 29 (2022) 1853–1865.

[11] Y. Makuuchi, K. Honda, Y. Osaka, K. Kato, T. Kojima, H. Daiko, H. Igaki, Y. Ito, S. Hoshino, S. Tachibana, T. Watanabe, K. Furuta, S. Sekine, T. Umaki, Y. Watabe, N. Miura, M. Ono, A. Tsuchida, T. Yamada, Soluble interleukin-6 receptor is a serum biomarker for the response of esophageal carcinoma to neoadjuvant chemoradiotherapy, Cancer Sci. 104 (2013) 1045–1051.

[12] A. Kleespies, M. Guba, K.W. Jauch, C.J. Bruns, Vascular endothelial growth factor in esophageal cancer, J. Surg. Oncol. 87 (2004) 95–104.

[13] H. Katayama, P. Tsou, M. Kobayashi, M. Capello, H. Wang, F. Esteva, M.L. Disis, S. Hanash, A plasma protein derived TGFβ signature is a prognostic indicator in triple negative breast cancer, NPJ Precision Oncol. 3 (2019) 10.

[14] D. Cunningham, W.H. Allum, S.P. Stenning, J.N. Thompson, C.J. Van de Velde, M. Nicolson, J.H. Scarffe, F.J. Lofts, S.J. Falk, T.J. Iveson, D.B. Smith, R.E. Langley, M. Verma, S. Weeden, Y.J. Chua, MAGIC trial participants, perioperative chemotherapy versus surgery alone for resectable gastro-esophageal cancer, N. Engl. J. Med. 355 (2006) 11–20.

[15] S.E. Al-Batran, R.D. Hofheinz, C. Pauligk, H.G. Kopp, G.M. Haag, K.B. Luley, J. Meiler, N. Homann, S. Lorenzen, H. Schmalenberg, S. Probst, M. Koenigsmann, M. Egger, N. Prasnikar, K. Caca, J. Trojan, U.M. Martens, A. Block, W. Fischbach, R. Mahlberg, M. Clemens, G. Illerhaus, K. Zirlik, D.M. Behringer, W. Schmiegel, M. Pohl, M. Heike, U. Ronellenfitsch, M. Schuler, W.O. Bechstein, A. Königsrainer, T. Gaiser, P. Schirmacher, W. Hozaeel, A. Reichart, T.O. Goetze, M. Sievert, E. Jäger, S. Mönig, A. Tannapfel, Histopathological regression after neoadjuvant docetaxel, oxaliplatin, fluorouracil, and leucovorin versus epirubicin, cisplatin, and fluorouracil or capecitabine in patients with resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4-AIO): results from the phase 2 part of a multicentre, open-label, randomised phase 2/3 trial, Lancet Oncol. 17 (2016) 1697–1708.

[16] T. Conroy, M.P. Galais, J.L. Raoul, O. Bouché, S. Gourgou-Bourgade, J.Y. Douillard, P.L. Etienne, V. Boige, I. Martel-Lafay, P. Michel, C. Llacer-Moscardo, E. François, G. Créhange, M.B. Abdelghani, B. Juzyna, L. Bedenne, A. Adenis, Fédération Francophone de Cancérologie Digestive and UNICANCER-GI Group, Definitive chemoradiotherapy with FOLFOX versus fluorouracil and cisplatin in patients with oesophageal cancer (PRODIGE5/ACCORD17): final results of a randomised, phase 2/3 trial, Lancet Oncol. 15 (2014) 305–314.

[17] A.-M. Mandard, F. Dalibard, J.-C. Mandard, J. Marnay, M. Henry-Amar, J.-F. Petiot, A. Roussel, J.-H. Jacob, P. Segol, G. Samama, J.-M. Ollivier, S. Bonvalot, M. Gignoux, Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations, Cancer 73 (1994) 2680–2686.

[18] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, *Limma* powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (2015), e47.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[20] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Perez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, Ş. Yilmaz, S. Tiwary, J. Cox, E. Audin, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaíno, The PRIDE database and related tools and resources in 2019: improving support for quantification data, Nucleic Acids Res. 47 (2019) D442–D450.

[21] J.A. Cohen, Coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46.

[22] Y. Zhou, B. Zhou, L. Pache, M. Chang, A.H. Khodabakhshi, O. Tanaseichuk, C. Benner, S.K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, Nat. Commun. 3 (2019) 10–1523.

[23] N.M. Thielens, F. Tedesco, S.S. Bohlson, C. Gaboriaud, A.J. Tenner, C1q: a fresh look upon an old molecule, Mol. Immunol. 89 (2017) 73–83.

[24] R. Bulla, C. Tripodo, D. Rami, G.-S. Ling, C. Agostinis, C. Guarnotta, S. Zorzet, P. Durigutto, M. Botto, F. Tedesco, C1q acts in the tumour microenvironment as a cancer-promoting factor independently of complement activation, Nat. Commun. 7 (2016) 10346.

[25] A. Mangogna, B. Belmonte, C. Agostinis, P. Zacchi, D.G. Iacopino, A. Martorana, V. Rodolico, D. Bonazza, F. Zanconati, U. Kishore, R. Bulla, Prognostic implications of the complement protein C1q in gliomas, Front. Immunol. 10 (2019) 2366.

[26] D. Zhang, Y. Li, H. Li, T. Tang, Y. Zheng, X. Guo, X. Xu, A preliminary study of the complement component 1q levels in predicting the efficacy of combined immunotherapy in patients with lung cancer, Cancer Manag. Res. 13 (2021) 7131–7137.

[27] A. Mangogna, C. Agostinis, D. Bonazza, B. Belmonte, P. Zacchi, G. Zito, A. Romano, F. Zanconati, G. Ricci, U. Kishore, R. Bulla, Is the complement protein C1q a pro- or anti-tumorigenic factor? Bioinformatics analysis involving human carcinomas, Front. Immunol. 10 (2019) 865.

[28] W. Kou, B. Li, Y. Shi, Y. Zhao, Q. Yu, J. Zhuang, Y. Xu, W. Peng, High complement protein C1q levels in pulmonary fibrosis and non-small cell lung cancer associated with poor prognosis, BMC Cancer 22 (2022) 110.

[29] A. Kaur, S.H.A. Sultan, V. Murugaiah, A.A. Pathan, F.S. Alhamlan, E. Karteris, U. Kishore, Human C1q induces apoptosis in an ovarian cancer cell line via tumour necrosis factor pathway, Front. Immunol. 7 (2016) 599.

[30] Q. Hong, C.-I. Sze, S.-R. Lin, M.-H. Lee, R.-Y. He, L. Schultz, J.-Y. Chang, S.-J. Chen, R.J. Boackle, L.-J. Hsu, N.-S. Chang, Complement C1q activates tumour suppressor WWOX to induce apoptosis in prostate cancer cells, PLoS One 4 (2009), e5755.

[31] C. Agostinis, R. Vidergar, B. Belmonte, A. Mangogna, L. Amadio, P. Geri, V. Borelli, F. Zanconati, F. Tedesco, M. Confalonieri, C. Tripodo, U. Kishore, R. Bulla, Complement protein C1q binds to hyaluronic acid in the malignant pleural mesothelioma microenvironment and promotes tumor growth, Front. Immunol. 8 (2017) 1559.

[32] R. Zhang, Q. Liu, T. Li, Q. Liao, Y. Zhao, Role of the complement system in the tumour microenvironment, Cancer Cell Int. 19 (2019) 300.

[33] J. Cui, G. Li, J. Yin, L. Li, Y. Tan, H. Wei, B. Liu, L. Deng, J. Tang, Y. Chen, L. Yi, GSTP1 and cancer: expression, methylation, polymorphisms and signalling, Int. J. Oncol. 56 (2020) 867–878.

[34] J. Huang, P.-H. Tan, J. Thiyagarajan, B.-H. Bay, Prognostic significance of glutathione S-transferase-pi in invasive breast cancer, Mod. Pathol. 16 (2003) 558–565.

[35] S. Meding, B. Balluff, M. Elsner, C. Schöne, S. Rauser, U. Nitsche, M. Maak, A. Schäfer, S.M. Hauck, M. Ueffing, R. Langer, H. Höfler, H. Friess, R. Rosenberg, A. Walch, Tissue-based proteomics reveals FXYD3, S100A11 and GSTM3 as novel markers for regional lymph node metastasis in colon cancer, J. Pathol. 228 (2012) 459–470.

[36] A. Cabelguenne, M.A. Loriot, I. Stucker, H. Blons, E. Koum-Besson, D. Brasnu, P. Beaune, O. Laccourreye, P. Laurent-Puig, I. De Waziers, Glutathione-associated enzymes in head and neck squamous cell carcinoma and response to cisplatin-based neoadjuvant chemotherapy, Int. J. Cancer 93 (2001) 725–730.

[37] D. Pectasides, K. Kamposioras, G. Papaxoinis, E. Pectasides, Chemotherapy for recurrent cervical cancer, Cancer Treat. Rev. 34 (2008) 603–613.

[38] A. Checa-Rojas, L.F. Delgadillo-Silva, MdC Velasco-Herrera, A. Andrade-Domínguez, J. Gil, O. Santillán, L. Lozano, A. Toledo-Leyva, A. Ramirez-Torres, P. Talamas-Rohana, S. Encarnación-Guevara, GSTM3 and GSTP1: novel players driving tumor progression in cervical cancer, Oncotarget 9 (2018) 21696–21714.

[39] M. Aitkenhead, S.-J. Wang, M.N. Nakatsu, J. Mestas, C. Heard, C.C.W. Hughes, Identification of endothelial cell genes expressed in an in vitro model of angiogenesis: induction of ESM-1, βig-h3, and NrCAM, Microvasc. Res. 63 (2002) 159–171.

[40] Y. He, F. Shao, W. Pi, C. Shi, Y. Chen, D. Gong, B. Wang, Z. Cao, K. Tang, Largescale transcriptomics analysis suggests over-expression of BGH3, MMP9 and PDIA3 in oral squamous cell carcinoma, PLoS One 11 (2016), e0146530.

[41] S. Srinivasan, M. Guha, D.W. Dong, K.A. Whelan, G. Ruthel, Y. Uchikado, S. Natsugoe, H. Nakagawa, N.G. Avadhani, Disruption of cytochrome c oxidase function induces Warburg effect and metabolic reprogramming, Oncogene 35 (2016) 1585–1595.

[42] P. Sánchez-Corral, O. Criado García, S. Rodríguez de Córdoba, Isoforms of human C4b-binding protein. I. Molecular basis for the C4BP isoform pattern and its variations in human plasma, J. Immunol. 155 (1995) 4030–4036.

[43] K. Sogawa, S. Takano, F. Iida, M. Satoh, S. Tsuchida, Y. Kawashima, H. Yoshitomi, A. Sanda, Y. Kodera, H. Takizawa, R. Mikata, M. Ohtsuka, H. Shimizu, M. Miyazaki, O. Yokosuka, F. Nomura, Identification of a novel serum biomarker for pancreatic cancer, C4b-binding protein α-chain (C4BPA) by quantitative proteomic analysis using tandem mass tags, Br. J. Cancer 115 (2016) 949–956.

[44] I. Pasic, A. Shlien, A.D. Durbin, D.J. Stavropoulos, B. Baskin, P.N. Ray, A. Novokmet, D. Malkin, Recurrent focal copy-number changes and loss of heterozygosity implicate two noncoding RNAs and one tumor suppressor gene at chromosome 3q13.31 in osteosarcoma, Cancer Res. 70 (2010) 160–171.

[45] C.C. Yen, W.M. Chen, T.H. Chen, W.Y.K. Chen, P.C.H. Chen, H.J. Chiou, G.Y. Hung, H.T.H. Wu, C.J. Wei, C.Y. Shiau, Y.C. Wu, T.C. Chao, C.H. Tzeng, P.M. Chen, C. H. Lin, Y.J. Chen, J.A. Fletcher, Identification of chromosomal aberrations associated with disease progression and a novel 3q13.31 deletion involving LSAMP gene in osteosarcoma, Int. J. Oncol. 35 (2009) 775–788.

[46] E. Ntougkos, R. Rush, D. Scott, T. Frankenberg, H. Gabra, J.F. Smyth, G.C. Sellar, The IgLON family in epithelial ovarian cancer: expression profiles and clinicopathological correlates, Clin. Cancer Res. 11 (2005) 5764–5768.

[47] H.O. Kanayama, J.D. Chen, W.O. Lui, M. Takahashi, S. Kagawa, B.T. Teh, LSAMP and norei: novel genes in hereditary and sporadic clear cell renal cell carcinomas (CCRCC), J. Urol. 171 (2004) 200.

[48] J. Chen, W.O. Lui, M.D. Vos, G.J. Clark, M. Takahashi, J. Schoumans, S.K. Khoo, D. Petillo, T. Lavery, J. Sugimura, D. Astuti, C. Zhang, S. Kagawa, E.R. Maher, C. Larsson, A.S. Alberts, H.O. Kanayama, B.T. Teh, The t(1;3) breakpoint- spanning involved in clear cell renal cell genes LSAMP and NORE1 are carcinomas, Cancer Cell 4 (2003) 405–413.

[49] C.Y. Chang, K.L. Wu, Y.Y. Chang, Y.W. Liu, Y.C. Huang, S.F. Jian, Y.S. Lin, P. H. Tsai, J.Y. Hung, Y.M. Tsai, Y.L. Hsu, The downregulation of LSAMP expression promotes lung cancer progression and is associated with poor survival prognosis, J. Personal. Med. 11 (2021) 578–592.

[50] S. Kitazawa, Y. Takaoka, Y. Ueda, R. Kitazawa, Identification of calmodulin-like protein 5 as tumor-suppressor gene silenced during early stage of carcinogenesis in squamous cell carcinoma of uterine cervix, Int. J. Cancer 149 (2021) 1358–1368.

[51] T.-L. Zheng, K. Cao, C. Lang, K. Zhang, H.-Z. Guo, D.-P. Li, S. Zhao, Prognostic value of C-reactive protein in esophageal cancer: a meta-analysis, Asian Pac. J. Cancer Prev. 15 (2014) 8075–8081.

[52] D.A. Yadin, I.B. Robertson, J. McNaught-Davis, P. Evans, D. Stoddart, P. A. Handford, S.A. Jensen, C. Redfield, Structure of the fibrillin-1 N-terminal domains suggests that heparan sulfate regulates the early stages of microfibril assembly, Structure. 21 (2013) 1743–1756.

[53] R. Kerslake, M. Hall, P. Vagnarelli, J. Jeyaneethi, H.S. Randeva, G. Pados, I. Kyrou, E. Karteris, A pancancer overview of FBN1, asprosin and its cognate receptor OR4M1 with detailed expression profiling in ovarian cancer, Oncol. Lett. 22 (2021) 650.

[54] Z. Cierna, M. Mego, I. Jurisica, K. Machalekova, M. Chovanec, V. Miskovska, D. Svetlovska, K. Kalavska, K. Rejlekova, K. Kajo, J. Mardiak, P. Babal, Fibrillin-1 (FBN-1) a new marker of germ cell neoplasia in situ, BMC Cancer 16 (2016) 597.

[55] F.-L. Cui, A.-N. Mahmud, Z.-P. Xu, Z.-Y. Wang, J.-P. Hu, VASN promotes proliferation of prostate cancer through the YAP/TAZ axis, Eur. Rev. Med. Pharmacol. Sci. 24 (2020) 6589–6596.

[56] M. Aydin, R. Kızıltan, S. Algul, O. Kemik, The utility of serum vasorin levels as a novel potential biomarker for early detection of colon cancer, Cureus 14 (2022), e21653.

[57] J. Zhang, R. Shridhar, Q. Dai, J. Song, S.C. Barlow, L. Yin, B.F. Sloane, F.R. Miller, C. Meschonat, B.D. Li, F. Abreo, D. Keppler, Cystatin m: a novel candidate tumor suppressor gene for breast cancer, Cancer Res. 64 (2004) 6957–6964.

[58] R. Shridhar, J. Zhang, J. Song, et al., Cystatin M suppresses the malignant phenotype of human MDA-MB-435S cells, Oncogene 23 (2004) 2206–2215.

[59] K. Daniel, Towards novel anti-cancer strategies based on cystatin function, Cancer Lett. 235 (2006) 159–176.

[60] S. Pérez-Holanda, I. Blanco, M. Menéndez, L. Rodrigo, Serum concentration of alpha-1 antitrypsin is significantly higher in colorectal cancer patients than in healthy controls, BMC Cancer 14 (2014) 355.

[61] Z.J. El-Akawi, A.M. Abu-awad, N.A. Khoun, Alpha-1 antitrypsin blood levels as indicator for the efficacy of cancer treatment, World J. Oncol. 4 (2013) 83–86.

[62] S.P. Short, C.S. Williams, Selenoproteins in tumorigenesis and cancer progression, Adv. Cancer Res. 136 (2017) 49–83.

[63] Y. Takata, A.R. Kristal, R.M. Santella, I.B. King, D.J. Duggan, J.W. Lampe, M. P. Rayman, P.L. Blount, B.J. Reid, T.L. Vaughan, U. Peters, Selenium, selenoenzymes, oxidative stress and risk of neoplastic progression from Barrett's esophagus: results from biomarkers and genetic variants, PLoS One 7 (2012), e38612.

[64] J.H. Song, Y.M. Han, W.H. Kim, J.M. Park, M. Jeong, E.J. Go, S.P. Hong, K. B. Hahm, Oxidative stress from reflux esophagitis to esophageal cancer: the alleviation with antioxidants, Free Radic. Res. 50 (2016) 1071–1079.

[65] G. Barrera, Oxidative stress and lipid peroxidation products in cancer progression and therapy, ISRN Oncol. 2012 (2012), 137289.

[66] J.D. Hayes, A.T. Dinkova-Kostova, K.D. Tew, Oxidative stress in cancer, Cancer Cell 38 (2020) 167–197.

[67] J. Liu, Z. Wang, Increased oxidative stress as a selective anticancer therapy, Oxidative Med. Cell. Longev. 2015 (2015), 294303.

[68] Y. Hu, X. Ye, R. Wang, K. Poon, Current research progress in the role of reactive oxygen species in esophageal adenocarcinoma, Transl. Cancer Res. 10 (2021) 1568–1577.

[69] S.G. Maher, D.T. McDowell, B.C. Collins, C. Muldoon, W.M. Gallagher, J. V. Reynolds, Serum proteomic profiling reveals that pretreatment complement protein levels are predictive of esophageal cancer patient response to neoadjuvant chemoradiation, Ann. Surg. 254 (2011) 809–816.