Identification of proteins associated with development of metastasis from cutaneous squamous cell carcinomas (cSCCs) via proteomic analysis of primary cSCCs

Running head: Proteomic analysis and ANXA5 and DDOST in squamous cell carcinoma

[1]A Shapanis* (ORCID 0000-0003-4147-6956), [1,2]C Lai* (ORCID 0000-0003-2282-5655), [3]S Smith (ORCID 0000-0001-7744-3238), [1,2]G Coltart, [4]M Sommerlad, [5]J Schofield, [5]E Parkinson, [5]P Skipp[#], [1]E Healy (ORCID 0000-0001-5591-6970)[#]

[1]Dermatopharmacology, Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, UK

[2]Dermatology, University Hospital Southampton NHS Foundation Trust, UK

[3]Department of Pathology, University of Cambridge, Cambridge, UK

[4]Histopathology, University Hospital Southampton NHS Foundation Trust, UK

[5]Centre for Proteomic Research, Institute for Life Sciences, University of Southampton, UK

*A.S. and C.L are joint first authors.

[#]E.H. and P.S are joint senior authors.

Corresponding authors: (1) Eugene Healy; email: ehealy@soton.ac.uk (2) Paul Skipp; email: pjss@soton.ac.uk

Funding sources: Wessex Medical Research (PhD studentship for A Shapanis), Wellcome Trust (Research Training Fellowship for C Lai), National Institute for Health Research

(Clinical Lectureship for C Lai, Academic Clinical Fellowship for G Coltart), Medical Research Council (Post-Doctoral Clinical Research Fellowship for S Smith).

Conflict of interest: The authors declare no potential conflicts of interest.

Word count: 3,498 words.

Figure count: 5 figures, 2 tables, 4 supplementary figures and 3 supplementary tables.

What's already known about this topic?

- Keratinocyte cancer is the most common cancer in the UK, and the capacity for cSCCs to metastasise presents a clinical problem.
- Although there are known clinical risk factors for cSCC metastasis, current staging systems are inaccurate at predicting the development of metastasis in patients with cSCC.
- It has been shown that mass spectrometry-based proteomic analysis can quantify and uncover potential key proteins in cancer development and metastasis.

What does this study add?

- This study has identified a number of proteins that are differentially expressed between primary cSCCs which metastasise and primary cSCCs which do not metastasise.
- Expression of the genes encoding for several of these proteins influence outcome in SCCs of other organs (lung, oropharynx, cervix and oesophagus).
- Higher abundance of two key proteins, ANXA5 and DDOST, are associated with the development of, and reduced time to, cSCC metastasis.

What is the translational message?

- This is the first study to undertake proteomic profiling using mass spectrometry to investigate proteins that are differentially expressed between human primary cSCCs that metastasise and those that don't metastasise.
- The results of this proteomic analysis of cSCCs will be useful for identifying potential therapeutic targets in this cancer.
- A prediction model incorporating ANXA5 and DDOST showed higher sensitivity and specificity than cSCC clinical staging systems for estimating likelihood of cSCC metastases.

Summary

Background Cutaneous squamous cell carcinoma (cSCC) is one of the most common cancers capable of metastasising. Proteomic analysis of cSCCs can provide insight into biological processes responsible for metastasis as well as future therapeutic targets and prognostic biomarkers.

Objectives This study aimed to identify proteins associated with development of metastasis in cSCC.

Methods A proteomic-based approach was employed on 105 completely-excised, primary cSCCs, comprising 52 which metastasised (P-M) and 53 which had not metastasised at 5 years post-surgery (P-NM). Formalin-fixed, paraffin-embedded cSCCs were microdissected and subjected to proteomic profiling after one dimensional (1D), and separately two dimensional (2D), liquid chromatography fractionation.

Results A discovery set of 24 P-Ms and 24 P-NMs identified 144 significantly differentially expressed proteins, including 33 proteins identified via both 1D and 2D separation, between P-Ms and P-NMs. Several differentially expressed proteins were also associated with survival in SCCs of other organs. Findings were verified by multiple reaction monitoring on 6 peptides from 2 proteins, Annexin A5 (ANXA5) and Dolichyl-diphosphooligosaccahride-protein glycosyltransferase non-catalytic subunit (DDOST), in the discovery group and validated on a separate cohort (n=57). Increased expression of ANXA5 and DDOST was associated with reduced time to metastasis in cSCC and decreased survival in cervical and oropharyngeal cancer. A prediction model using ANXA5 and DDOST had an area under the curve (AUC) of 0.929 (CI=0.8277-1), an accuracy of 91.18% and higher sensitivity and specificity than cSCC staging systems currently in clinical use.

Conclusions This study highlights that increased expression of two proteins, ANXA5 and DDOST, is significantly associated with poorer clinical outcomes in cSCC.

Introduction

The number of keratinocyte cancers in the United Kingdom is >211,120 annually, with cutaneous squamous cell carcinoma (cSCC) accounting for >44,672, constituting one of the most common types of cancer capable of metastasising.[1,2] The risk of metastasis for cSCC depends on clinical and histological parameters, including site, depth of invasion, diameter, differentiation of the tumour, the presence of lymphovascular or perineural invasion, and host immunosuppression.[3] Following surgical excision, cSCC metastasises in 16% of cases with tumour depth >6mm,[4] and in 30% of tumours >2cm diameter.[5] Whereas the 3-year disease specific survival rate for patients with cSCC is 85%,[6] for patients with distant metastasis the median survival is <2 years.[7]

Staging systems assist identification of patients at greater risk of metastases after excision of primary cSCC.[8,9] However, current staging systems distinguish "poorly to moderately" between patients who do and those who don't develop cSCC metastases[8] and one-third of patients are classified incorrectly using these staging systems.[10] There is a need to undertake research into factors which contribute to more aggressive tumours[11], to understand the mechanisms responsible for development of metastases in cSCC and to identify more accurately those patients at risk of metastases.

Proteomic analysis can aid in understanding the aetiology of cancer progression and provide information of prognostic relevance.[12] In this study we used a mass spectrometry-based proteomic approach on cSCCs to identify proteins involved in development of metastases. The results highlight a number of differentially expressed proteins that associate with occurrence of metastases from cSCC, and reduced survival in lung, cervical, oropharyngeal and oesophageal SCC.

Materials and methods

Tissue samples

Formalin-fixed paraffin-embedded (FFPE) human primary cSCCs were acquired from Histopathology, University Hospital Southampton NHS Foundation Trust (UHS-NHSFT) under ethics committee approval (South Central Hampshire B National Research Ethics Service Committee; LREC number 07/H0504/187). Samples were categorised as primary cSCCs that metastasised (P-M) or primary cSCCs that had not metastasised at 5 years post-surgery (P-NM), with the latter based on no evidence of metastasis during 5 years follow-up and/or patient review for another reason after 5 years in Dermatology UHS-NHSFT.

Sample preparation for mass spectrometry

FFPE tissue sections were mounted onto glass slides, and tumour and surrounding immune infiltrate microdissected and transferred into protein extraction buffer (see supplementary material and methods). Samples were heated to 105°C for 30 minutes, cooled, then heated to 80°C for 2 hours before reduction using dithiothreitol and alkylation with iodoacetamide. Samples were digested with sequencing grade trypsin overnight and resulting peptides desalted using C18 reverse phase clean-up plates.

Immunostaining

Standard immunostaining protocols were used. Briefly, slides were deparaffinised and rehydrated, and endogenous peroxidase blocked, before incubation overnight at 4ºC with primary antibody (CD1a, 1:50, Dako M3571; Lplastin, 1:200, Abcam ab109129; ANXA5, Abcam EPR3979; DDOST, LSBio C340633; CD8, 1:50, Invitrogen 998254C). Subsequent incubation with biotinylated secondary antibody (anti-mouse, 1:400, JIR 315-066-045; anti-rabbit, 1:400, Dako E0731) was followed by addition of avidin-biotin-horseradish peroxidase

complex (Vector) and DAB as chromogen. Slides were imaged using an Olympus VS110 virtual microscopy system.

Discovery liquid chromatography mass spectrometry (LC-MS^E)

Samples were fractionated using a nanoACQUITY UPLC system (Waters) and electrosprayed into a Waters Synapt-G2-Si mass spectrometer operating in MS^E mode with ion mobility activated (supplementary materials and methods). Estimates of absolute quantification using the Top3 approach[13] were obtained using one-dimensional (1D) and two-dimensional (2D) LC separation strategies. Data from 1D and 2D LC procedures were analysed separately. Three blank runs were conducted between samples to ensure avoidance of carry-over into subsequent samples.

Multiple reaction monitoring

A spectral library from the discovery proteomic data was generated using Skyline software[14] to identify unique peptides for proteins of interest. Heavy stable isotope labelled (SIL) peptides were synthesised by Cambridge Research Biochemicals. Calibration curves were created using 1µg cSCC "proteomic-ready" sample as background. High Definition MRM acquisition mode was used for targeted acquisition. Transitions for each peptide were identified using Skyline and imported into MassLynx (Waters) for targeted acquisition. Samples were analysed containing 100 fmol of each heavy SIL peptide. Raw data was imported into Skyline for interpretation and calculation of native peptide quantity.

Gene expression in other SCCs

Expression levels of relevant genes were analysed in publicly available RNA sequencing data from the TCGA Research Network: http://cancergenome.nih.gov/.[15] Computational analysis

and statistical testing of NGS data was conducted using R statistical programming language.[16] Filtered and log2 normalised RNA expression data, alongside available clinical data, were downloaded from the GDAC firehose database (run: stddata__2015_06_01). Plotting of TCGA data was performed using ggplot2 R package.[17] Survival analysis was performed using survminer and survival R packages.[18] Kaplan-Meier survival curves were constructed using TCGA clinical data. Statistical testing of differences between survival curves used G-rho family of tests, as implemented in the survdiff function of the survival package.

Data analysis

1% FDR was applied for searching for peptide identification. Each protein was inferred from identification of at least one unique peptide. Only proteins detected in ≥50% of samples were subsequently analysed. Data was normalised to median protein concentration for each sample and *P* values obtained by Mann-Whitney U test. Topological data analysis, using Ayasdi, was performed on complete, normalised proteomic data with a hamming metric and 2 neighbourhood lenses. For Kaplan Meier survival analysis, *P* values were obtained by Log Rank test. Machine learning was performed using the statistical programming language, R, with packages caret and caretensemble.

Results

Discovery proteomics

This study investigated proteomic differences between P-M and P-NM cSCCs to identify proteins associated with metastasis in cSCC. As expected, more patients in the P-M than P-NM group had poorly differentiated tumours, perineural invasion or were immunosuppressed (Table 1). A discovery group of 24 P-M and 24 P-NM samples was subjected to proteomic profiling using 1D, and independently 2D, separation to identify and quantify differences in

protein abundance between P-Ms and P-NMs. Microdissected cSCC samples included tumour keratinocytes and stromal regions containing the immune cell infiltrate (Fig. 1a). Volcano plots demonstrated higher numbers of upregulated than downregulated proteins in P-M compared with P-NM cSCCs (Fig. 1b-c). Overall, 4,018 unique proteins were identified in the cSCCs (Fig. 1d), of which 144 were significantly differentially expressed between P-Ms and P-NMs ($P < 0.05$, Supplementary Tables 1 and 2), including 33 proteins identified both via 1D and 2D proteomics (Fig. 1e and Table 2). Topological data analysis of the 48 proteomes from the discovery set of 24 P-M and 24 P-NM cSCCs, performed without including input information on metastases or any other clinical data, demonstrated separation of samples in both 1D and 2D analyses according to development of metastases (Fig. 1f-g), providing support for distinct proteomic profiles of P-M and P-NM cSCCs.

Pathway analysis

Weighted gene co-expression network analysis (WGCNA) of the proteomics data was conducted and, following construction of a signed topological overlap matrix (TOM) of corresponding dissimilarity, hierarchical clustering was used on the dissimilarity TOM to produce modules of genes (Supplementary Fig. 1a-b). Modules were examined for correlation with clinical and immunohistochemical characteristics (Supplementary Fig. 1c) in addition to analysing for pathway enrichment (Supplementary Fig. 1d). Immunohistochemical characterisation showed significantly fewer CD8+ T cells and CD1a+ Langerhans cells in P-M than in P-NM samples (Fig. 2a-b, Supplementary Table 3), with lower numbers of CD8+, and separately CD1a+, cells significantly associated with reducing time to metastasis ($P = 0.0041$ and $P = 0.0057$ respectively, Fig. 2c-d). In WGCNA, one module (denoted by the colour "blue") correlated inversely with intratumoral CD1a+ Langerhans cell numbers ($P = 0.04$) but positively with FOXP3+ regulatory T cell (Tregs) numbers ($P = 0.005$) and with

development of metastasis (P = 0.04). Conversely, another module of proteins (represented by the colour "brown") correlated positively with number of intratumoral CD1a+ Langerhans cells (P = 0.03) but inversely with CD3+ T cell numbers (P = 0.03). A different module ("turquoise") demonstrated strong correlation with greater Clark's level invasion, an inverse correlation with peritumoral CD1a+ cell numbers, and also showed increased pathway enrichment, including neutrophil degranulation (P = 3.3e-23). Another module ("yellow"), which was heavily enriched in the keratinisation pathway (P = 2.07e-17), correlated with CD3+ and CD8+ cell numbers, but inversely with tumour differentiation and CD20+ B cell numbers.

To identify cell signalling pathways associated with development of cSCC metastasis, Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) analysis with Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping of significantly differentially expressed proteins in the 1D and 2D data was conducted. STRING analysis demonstrated highly connected structures with clusters (Supplementary Fig. 2a-b), with KEGG pathway enrichment highlighting ribosomal proteins, protein processing in the endoplasmic reticulum, focal adhesion, extracellular matrix/receptor interactions, PI3K-Akt signalling, and antigen processing and presentation as key differences between P-Ms and P-NMs (Supplementary Fig. 2c).

Comparison to the cancer genome atlas

To determine whether proteins involved in development of cSCC metastases influence development of metastases in other SCC types, the 33 significantly differentially expressed proteins in the 1D and 2D proteomic data were compared against gene expression in cervical, oropharyngeal, oesophageal and lung SCC using The Cancer Genome Atlas (TCGA). Expression of genes encoding for several proteins differentially expressed between P-M and P-NM cSCCs were identified as having significant effects on survival in SCCs arising at these

other sites, with reduced survival associated, separately, with high expression of POSTN, DDOST, HNRNPK, COL6A3, ANXA5, and LCP1 and with low expression of CALML5 (Fig. 3a-n, Supplementary Table 3). Furthermore, as immune dysfunction is important for cSCC development, and as LCP1 (L-plastin) can stimulate the T cell receptor and activate T-cells,[19] immunohistochemistry for LCP1 was conducted on the discovery group of cSCCs and demonstrated more LCP1+ cells in P-Ms than in P-NMs (Fig. 4o-p, Supplementary Table 3).

Multiple reaction monitoring

Multiple reaction monitoring (MRM) was used to validate the discovery proteomics. MRM is a highly sensitive and specific mass spectrometry method that involves filtering the mass spectrometer on specific peptides of interest and quantifying these against known concentrations of isotopically labelled peptides spiked into the samples, enabling greater sensitivity and more accurate quantification of protein concentrations. Firstly, machine learning (using a generalised linear model, GLM) was conducted on significantly differentially expressed proteins between P-M and P-NM cSCCs, in which a model predicting cSCC metastases was produced for every combination of two proteins on a training set and tested on a holdout cohort (2/3rds and 1/3 split, respectively). From >300 models, the combination of ANXA5 and DDOST gave one of the best area under curve (AUC) results, and because expression of both these genes had been identified via TCGA as important in reducing survival in SCCs of other organs, ANXA5 and DDOST were selected for targeted verification and validation using MRM. Three unique peptides per protein were identified using Skyline software and synthesised as stable isotope-labelled (SIL) peptides (Supplementary Fig. 3). MRM of the discovery cSCC group (22 P-M and 22 P-NM) verified that there was more DDOST and ANXA5 in P-M than P-NM cSCCs (DDOST $P = 0.0036$, ANXA5 $P = 0.0046$, Fig. 4a-d, Supplementary Table 3). MRM for DDOST and ANXA5 was then conducted in a

different (i.e. validation) group of cSCCs, comprising 28 P-Ms and 29 P-NMs. Again, DDOST and ANXA5 levels were significantly higher in the P-M than P-NM cSCCs (DDOST P = 0.0004, ANXA5 P = 0.0004, Fig. 4e-h, Supplementary Table 3).

Survival analyses were conducted to investigate the relationship between ANXA5 and DDOST expression and clinical outcome. High expression of ANXA5 and DDOST was associated with reduced time to cSCC metastasis (P = 0.00058, Fig. 5a). P-M cSCCs were associated with a reduced time to death compared to P-NM cSCCs (P < 0.0001, Fig 5b) and high expression of ANXA5 and DDOST was also associated with reduced 5-year overall survival (P = 0.0236, Fig. 5c). Moreover, TCGA analysis demonstrated that high co-expression of ANXA5 and DDOST significantly reduces survival in cervical and oropharyngeal SCC (P = 0.046 and P = 0.0072 respectively, Fig. 5d-e).

A stacked ensemble prediction model with the ANXA5 and DDOST MRM data was created using R software and base level algorithms comprising k-Nearest Neighbors, naïve Bayes, glmnet, AdaBoost, xgbDART and the stochastic gradient boosting GBM. The predictions of these individual algorithms were then subjected to a top layer algorithm, xgbTree, to form final predictions for each sample. Data was split into 2/3 (n=67) for training and 1/3 (n=34) for testing and models were trained using 10-fold cross validation repeated 3 times. The resulting prediction model ROC curve gave an AUC = 0.929 (Fig. 5f). This ANXA5-DDOST prediction model was compared on the same cSCC samples with cSCC clinical staging systems, including American Joint Committee on Cancer 7th and 8th editions,[20,21] Brigham and Women's Hospital,[9] British Association of Dermatologists,[22] Breuninger et al.,[23] European Dermatology Forum,[7] Union for International Cancer Control,[24] and with results of the validation study of some of these staging systems by Roscher et al. on their patient cohort.[8] This comparison showed that the ANXA5-DDOST prediction model has higher sensitivity and specificity than each of these staging systems.

Discussion

This proteomics-based study identified multiple proteins associated with development of cSCC metastases and ascertained that high expression of several respective genes encoding for these proteins associate with reduced survival in SCCs of the cervix, oropharynx, oesophagus and lung. Although mass spectrometry for proteomic analysis of cSCCs has been employed previously,[25] to our knowledge, the current study is the first to investigate differential expression of proteins in primary cSCC with respect to metastasis/clinical outcome. Our topological data analysis was largely able to separate cSCCs according to development of metastases, providing strong support for involvement of the detected proteins in the metastatic process, although it is not possible to conclude from this study what proportion of these are drivers or passengers in this process. Some differences in protein expression between P-M and P-NM cSCCs may be due to variation in tumour parameters (e.g. cell proliferation, differentiation status) or composition of the immune infiltrate between the two tumour groups. However, bioinformatic analysis highlighted several pathways/processes likely to be causally involved in permitting cSCC metastases. STRING/KEGG identified differences between P-Ms and P-NMs in PI3K-Akt signalling, which influences development of cancer metastasis[26] and can affect cSCC growth.[27] Indeed, PI3K-Akt signalling pathways differ between well-differentiated and moderately/poorly-differentiated cSCCs,[28] and oncogenic mutations affecting PI3K signalling are frequent in metastatic cSCCs.[29] STRING/KEGG also identified extracellular matrix-receptor interaction and enrichment of focal adhesion, important for cancer invasion and metastases,[30,31] in P-M compared to P-NM samples. Additionally, STRING/KEGG identified "antigen processing and presentation" differences between P-M and P-NM, consistent with our observations that lower numbers of CD1a+ Langerhans cells and CD8+ T cells in cSCCs associate with metastasis, and our previous work demonstrating

that cSCC Tregs suppress effector T cells in this tumour.[32] Furthermore, the current study shows that P-Ms have higher levels of TGFβ1, which exerts immunosuppressive effects via Tregs[33] and inducing PD-1 on CD8+ T cells.[34]

More proteins were upregulated than downregulated in the comparison of P-M with P-NM cSCCs, which may relate to limitations with mass spectrometry in detecting reduced protein expression below the sensitivity threshold. There were also substantial variations between samples, confirming our previous observations that cSCCs and their immune infiltrates are highly heterogeneous.[32] In addition, although many proteins that were differentially expressed between P-Ms and P-NMs were identified using both 1D and 2D separation, the 1D and 2D separation methodologies yielded differences in the overall numbers of unique proteins. Moreover, correction for multiple parameters was not feasible given the large number of variables, including varying levels of infiltration of different immunocyte populations. However, we processed cSCC samples which included tumour and surrounding stroma/immune infiltrate instead of microdissecting the tumour without the stroma because there is evidence that immune, as well as tumour, parameters are determinants of clinical outcome in cSCC.[3,4,32,35] We acknowledge there is likely to have been a loss of resolution with this approach, and that future studies undertaking proteomic profiling of cSCCs following purification of separate tumour regions, and deconvolution of data based on heterogeneous cell populations, would allow identification of additional pathways relevant to development of metastases and clinical outcome.

MRM verified differential expression of ANXA5 and DDOST in the discovery group of P-M and P-NM cSCCs and validated this in a separate cohort of tumours, highlighting the relevance of ANXA5 and DDOST in development of cSCC metastasis. However, as both proteins were expressed in tumour and immune cells (Supplementary Fig. 4), it is unclear whether the mechanism underlying this association is due to expression of the proteins in the

tumour, or immune infiltrate, or both these sites. High ANXA5 expression is associated with metastases from colorectal cancer,[40] and reduced survival in renal cell carcinoma.[41] Additionally, the Human Protein Atlas indicates that, using TCGA data, ANXA5 is an unfavourable prognostic marker in renal, liver, urothelial, and head and neck cancers, but favourable marker in endometrial and stomach cancers.[42] ANXA5 has also been identified as a potential biomarker in a DNA microarray study of cSCC cell lines and tissue [43] and a proteomic analysis of head and neck SCC.[44] The mode of action of ANXA5 in relation to development of metastases is not fully understood, but it has been shown to promote migration and invasion of keratinocyte,[45] oral SCC,[45] renal cell carcinoma [41] and hepatocarcinoma [46] cell lines in ANXA5 knockdown experiments. Potential mechanisms for this include effects of ANXA5 on regulation of genes implicated in cell motility (including *S100A4*, *TIMP-3*, *RHOC*),[45] activation of PI3K/Akt/mTOR signalling leading to tumour cell proliferation,[41] promotion of migration and invasion via upregulation of MMP2 and MMP9,[41] and effects on integrin signalling and MEK-ERK pathways.[46] Conversely, ANXA5 may have a protective role in some cancers because ANXA5 overexpression can inhibit proliferation and metastasis, including in uterine and cervical carcinoma cell lines.[47] In addition, administration of ANXA5 in a murine model of HPV16-associated cancer augmented anti-tumour immunity by binding to phosphatidylserine externalised by apoptotic tumour cells, which enhanced immunogenicity of tumour antigens.[48]

While there is limited published research on DDOST in cancer, the Human Protein Atlas documents DDOST as an unfavourable prognostic marker in renal, liver, and head and neck cancers but favourable marker in endometrial cancer.[49] Gene expression profiling interactive analysis of TCGA and genome-scale CRISPR-Cas9 knockout screening data have demonstrated *DDOST* as an essential gene across many cancer cell lines, with *DDOST* upregulated in colon adenocarcinoma and overlapping with expression of genes required for

cell growth and viability (although in that study, higher DDOST expression was associated with increased survival in colon adenocarcinoma).[50] Furthermore, another study investigating susceptibility variants for oesophageal SCC reported missense variants in DDOST in two cases.[51] The mechanism whereby DDOST permits metastasis is unclear, but may involve protein glycosylation and the impact of this via various biological processes relevant to cancer.[52] For example, DDOST functions as a subunit for an accessory protein required for stabilisation of the STT3 protein subunits of oligosaccharyltransferase (OST),[53,54] which promotes tumour immune evasion via PD-L1.[55,56] Moreover, STT3, which is induced by epithelial mesenchymal transition, is required for PD-L1 N-glycosylation, which stabilises and upregulates PD-L1 in breast cancer stem cells.[57] OST is also required for EGFR cell surface localisation and signalling in non-small lung cancer cells and, in EGFR-driven tumour cells, OST inhibition induces senescence.[58] Likewise, OST inhibition reduces tumour growth in EGFR-mutant non-small lung cancer[59] and glioma[60] xenografts.

The absolute quantification of ANXA5 and DDOST via MRM in primary cSCCs in this study, and confirmation of higher levels of these proteins in P-M tumours in the discovery and validation groups, suggest that they may have potential for use as biomarkers for development of metastasis in cSCC following surgical excision of the tumour. This is supported by our findings that high expression of ANXA5 and DDOST are associated with shorter time to metastasis and reduced 5-year overall survival in patients with cSCCs, and similarly, reduced survival in cervical and oropharyngeal SCC. Indeed, the incorporation of our ANXA5 and DDOST MRM data in a prediction model demonstrated higher sensitivity and specificity than commonly used clinical staging systems for cSCC, indicating that ANXA5 and DDOST offer potential to provide additional useful information on the likelihood of metastatic spread in this cancer. As MRM was conducted on FFPE cSCC samples in the current study, future evaluation of ANXA5 and DDOST in larger cohorts of FFPE samples, and their

subsequent study/use in clinical practice as an adjunct to current staging systems which use FFPE samples, would be possible.[10] Although conjectural, based on evaluation of ANXA5 and DDOST in larger cohorts of patients, the future incorporation of these markers with other relevant clinicopathological risk factors into a prediction model may offer clinical benefits through improved staging and consequently more personalised treatment and/or follow up of patients with cSCC.

In conclusion, this proteomics study has identified multiple proteins associated with cSCC metastasis, with several of our findings relevant to other types of SCC. Importantly, high expression of ANXA5 and DDOST in primary cSCCs is associated with subsequent metastatic spread. The results highlight that proteomic analysis has potential to offer useful insight into biological factors which influence development of metastases from primary cSCCs, and can be a useful adjunct to other 'omics' approaches aimed at identifying potential biomarkers in this cancer.

Acknowledgements

Instrumentation in the Centre for Proteomic Research is supported by the BBSRC (BM/M012387/1) and Wessex Medical Trust.

REFERENCES

1 Venables ZC, Nijsten T, Wong KF *et al.* Epidemiology of basal and cutaneous squamous cell carcinoma in the U.K. 2013-15: a cohort study. *Br J Dermatol* 2019; **181**: 474-82.

2 Nehal KS, Bichakjian CK. Update on Keratinocyte Carcinomas. *N Engl J Med* 2018; **379**: 363-74.

3 Thompson AK, Kelley BF, Prokop LJ *et al.* Risk Factors for Cutaneous Squamous Cell Carcinoma Recurrence, Metastasis, and Disease-Specific Death: A Systematic Review and Meta-analysis. *JAMA Dermatol* 2016; **152**: 419-28.

4 Brantsch KD, Meisner C, Schonfisch B *et al.* Analysis of risk factors determining prognosis of cutaneous squamous-cell carcinoma: a prospective study. *Lancet Oncol* 2008; **9**: 713-20.

5 Alam M, Ratner D. Cutaneous squamous-cell carcinoma. *N Engl J Med* 2001; **344**: 975-83.

6 Clayman GL, Lee JJ, Holsinger FC *et al.* Mortality risk from squamous cell skin cancer. *J Clin Oncol* 2005; **23**: 759-65.

7 Stratigos A, Garbe C, Lebbe C *et al.* Diagnosis and treatment of invasive squamous cell carcinoma of the skin: European consensus-based interdisciplinary guideline. *Eur J Cancer* 2015; **51**: 1989-2007.

8 Roscher I, Falk RS, Vos L *et al.* Validating 4 Staging Systems for Cutaneous Squamous Cell Carcinoma Using Population-Based Data: A Nested Case-Control Study. *JAMA Dermatol* 2018; **154**: 428-34.

9 Karia PS, Jambusaria-Pahlajani A, Harrington DP *et al.* Evaluation of American Joint Committee on Cancer, International Union Against Cancer, and Brigham and Women's Hospital tumor staging for cutaneous squamous cell carcinoma. *J Clin Oncol* 2014; **32**: 327-34.

10 Abraham I, Curiel-Lewandrowski C. Staging Systems to Predict Metastatic Cutaneous Squamous Cell Carcinoma: Unsatisfactory for Clinical Use, but Some Less So? *JAMA Dermatol* 2018; **154**: 1391-2.

11 Nagarajan P, Asgari MM, Green AC *et al.* Keratinocyte Carcinomas: Current Concepts and Future Research Priorities. *Clin Cancer Res* 2019; **25**: 2379-91.

12 Iglesias-Gato D, Thysell E, Tyanova S *et al.* The Proteome of Prostate Cancer Bone Metastasis Reveals Heterogeneity with Prognostic Implications. *Clin Cancer Res* 2018; **24**: 5433-44.

13 Silva JC, Gorenstein MV, Li GZ *et al.* Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* 2006; **5**: 144-56.

14 MacLean B, Tomazela DM, Shulman N *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010; **26**: 966-8.

15 Cancer Genome Atlas Research N, Weinstein JN, Collisson EA *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013; **45**: 1113-20.

16 R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing,. In. Vienna, Austria. 2013.

17 Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. 2016.

18 Therneau TM GP. *Modeling survival data: extending the Cox model*. New York: Springer. 2000.

19 Wabnitz GH, Kocher T, Lohneis P *et al.* Costimulation induced phosphorylation of L-plastin facilitates surface transport of the T cell activation molecules CD69 and CD25. *Eur J Immunol* 2007; **37**: 649-62.

20 Farasat S, Yu SS, Neel VA *et al.* A new American Joint Committee on Cancer staging system for cutaneous squamous cell carcinoma: creation and rationale for inclusion of tumor (T) characteristics. *J Am Acad Dermatol* 2011; **64**: 1051-9.

21 Lydiatt WM, Patel SG, O'Sullivan B *et al.* Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. *CA Cancer J Clin* 2017; **67**: 122-37.

22 Motley RJ, Preston PW, Lawrence CM. Multi-professional Guidelines for the Management of the Patient with Primary Cutaneous Squamous Cell Carcinoma. In: British Association of Dermatologists. 2009.

23 Breuninger H, Brantsch K, Eigentler T *et al.* Comparison and evaluation of the current staging of cutaneous carcinomas. *J Dtsch Dermatol Ges* 2012; **10**: 579-86.

24 Sobin LH, Gospodarowicz MK, Wittekind C. *TNM Classification of Malignant Tumours*, 7th edn. New York: Wiley-Blackwell. 2009.

25 Azimi A, Yang P, Ali M *et al.* Data Independent Acquisition Proteomic Analysis Can Discriminate between Actinic Keratosis, Bowen's Disease, and Cutaneous Squamous Cell Carcinoma. *J Invest Dermatol* 2020; **140**: 212-22 e11.

26 Pierobon M, Ramos C, Wong S *et al.* Enrichment of PI3K-AKT-mTOR Pathway Activation in Hepatic Metastases from Breast Cancer. *Clin Cancer Res* 2017; **23**: 4919-28.

27 Zou Y, Ge M, Wang X. Targeting PI3K-AKT-mTOR by LY3023414 inhibits human skin squamous cell carcinoma cell growth in vitro and in vivo. *Biochem Biophys Res Commun* 2017; **490**: 385-92.

28 Inman GJ, Wang J, Nagano A *et al.* The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature. *Nat Commun* 2018; **9**: 3667.

29 Li YY, Hanna GJ, Laga AC *et al.* Genomic analysis of metastatic cutaneous squamous cell carcinoma. *Clin Cancer Res* 2015; **21**: 1447-56.

30 Hamidi H, Pietila M, Ivaska J. The complexity of integrins in cancer and new scopes for therapeutic targeting. *Br J Cancer* 2016; **115**: 1017-23.

31 Hamidi H, Ivaska J. Every step of the way: integrins in cancer progression and metastasis. *Nat Rev Cancer* 2018; **18**: 533-48.

32 Lai C, August S, Albibas A *et al.* OX40+ Regulatory T Cells in Cutaneous Squamous Cell Carcinoma Suppress Effector T-Cell Responses and Associate with Metastatic Potential. *Clin Cancer Res* 2016; **22**: 4236-48.

33 Cuende J, Lienart S, Dedobbeleer O *et al.* Monoclonal antibodies against GARP/TGF-beta1 complexes inhibit the immunosuppressive activity of human regulatory T cells in vivo. *Sci Transl Med* 2015; **7**: 284ra56.

34 Park BV, Freeman ZT, Ghasemzadeh A *et al.* TGFbeta1-Mediated SMAD3 Enhances PD-1 Expression on Antigen-Specific T Cells in Cancer. *Cancer Discov* 2016; **6**: 1366-81.

35 Lai C, August S, Behar R *et al.* Characteristics of immunosuppressive regulatory T cells in cutaneous squamous cell carcinomas and role in metastasis. *Lancet* 2015; **385 Suppl 1**: S59.

36 Gruosso T, Gigoux M, Manem VSK *et al.* Spatially distinct tumor immune microenvironments stratify triple-negative breast cancers. *J Clin Invest* 2019; **129**: 1785-800.

37 Tian C, Clauser KR, Ohlund D *et al.* Proteomic analyses of ECM during pancreatic ductal adenocarcinoma progression reveal different contributions by tumor and stromal cells. *Proc Natl Acad Sci U S A* 2019; **116**: 19609-18.

38 Zemek RM, De Jong E, Chin WL *et al.* Sensitization to immune checkpoint blockade through activation of a STAT1/NK axis in the tumor microenvironment. *Sci Transl Med* 2019; **11**.

39 Moeini A, Torrecilla S, Tovar V *et al.* An Immune Gene Expression Signature Associated With Development of Human Hepatocellular Carcinoma Identifies Mice That Respond to Chemopreventive Agents. *Gastroenterology* 2019; **157**: 1383-97 e11.

40 Xue G, Hao LQ, Ding FX *et al.* Expression of annexin a5 is associated with higher tumor stage and poor prognosis in colorectal adenocarcinomas. *J Clin Gastroenterol* 2009; **43**: 831-7.

41 Tang J, Qin Z, Han P *et al.* High Annexin A5 expression promotes tumor progression and poor prognosis in renal cell carcinoma. *Int J Oncol* 2017; **50**: 1839-47.

42 TPHA ANXA5: The Human Protein Atlas. Available at: https://www.proteinatlas.org/ENSG00000164111-ANXA5/pathology (last accessed 24/01/20).

43 Dooley TP, Reddy SP, Wilborn TW *et al.* Biomarkers of human cutaneous squamous cell carcinoma from tissues and cell lines identified by DNA microarrays and qRT-PCR. *Biochem Biophys Res Commun* 2003; **306**: 1026-36.

44 Melle C, Ernst G, Schimmel B *et al.* Biomarker discovery and identification in laser microdissected head and neck squamous cell carcinoma with ProteinChip technology, two-dimensional gel electrophoresis, tandem mass spectrometry, and immunohistochemistry. *Mol Cell Proteomics* 2003; **2**: 443-52.
45 Wehder L, Arndt S, Murzik U *et al.* Annexin A5 is involved in migration and invasion of oral carcinoma. *Cell Cycle* 2009; **8**: 1552-8.
46 Sun X, Liu S, Wang J *et al.* Annexin A5 regulates hepatocarcinoma malignancy via CRKI/II-DOCK180-RAC1 integrin and MEK-ERK pathways. *Cell Death Dis* 2018; **9**: 637.
47 Li X, Ma W, Wang X *et al.* Annexin A5 overexpression might suppress proliferation and metastasis of human uterine cervical carcinoma cells. *Cancer Biomark* 2018; **23**: 23-32.
48 Kang TH, Park JH, Yang A *et al.* Annexin A5 as an immune checkpoint inhibitor and tumor-homing molecule for cancer treatment. *Nat Commun* 2020; **11**: 1137.
49 THPA DDOST: The Human Protein Atlas. Available at: https://www.proteinatlas.org/ENSG00000244038-DDOST/pathology (last accessed 24/01/20).
50 Hu M, Fu X, Si Z *et al.* Identification of Differently Expressed Genes Associated With Prognosis and Growth in Colon Adenocarcinoma Based on Integrated Bioinformatics Analysis. *Front Genet* 2019; **10**: 1245.
51 Donner I, Katainen R, Tanskanen T *et al.* Candidate susceptibility variants for esophageal squamous cell carcinoma. *Genes Chromosomes Cancer* 2017; **56**: 453-9.
52 Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* 2015; **15**: 540-55.
53 Harada Y, Ohkawa Y, Kizuka Y *et al.* Oligosaccharyltransferase: A Gatekeeper of Health and Tumor Progression. *Int J Mol Sci* 2019; **20**.
54 Roboti P, High S. The oligosaccharyltransferase subunits OST48, DAD1 and KCP2 function as ubiquitous and selective modulators of mammalian N-glycosylation. *J Cell Sci* 2012; **125**: 3474-84.
55 Li CW, Lim SO, Xia W *et al.* Glycosylation and stabilization of programmed death ligand-1 suppresses T-cell activity. *Nat Commun* 2016; **7**: 12632.
56 Hsu JM, Li CW, Lai YJ *et al.* Posttranslational Modifications of PD-L1 and Their Applications in Cancer Therapy. *Cancer Res* 2018; **78**: 6349-53.
57 Hsu JM, Xia W, Hsu YH *et al.* STT3-dependent PD-L1 accumulation on cancer stem cells promotes immune evasion. *Nat Commun* 2018; **9**: 1908.
58 Lopez-Sambrooks C, Shrimal S, Khodier C *et al.* Oligosaccharyltransferase inhibition induces senescence in RTK-driven tumor cells. *Nat Chem Biol* 2016; **12**: 1023-30.
59 Lopez Sambrooks C, Baro M, Quijano A *et al.* Oligosaccharyltransferase Inhibition Overcomes Therapeutic Resistance to EGFR Tyrosine Kinase Inhibitors. *Cancer Res* 2018; **78**: 5094-106.
60 Baro M, Lopez Sambrooks C, Quijano A *et al.* Oligosaccharyltransferase Inhibition Reduces Receptor Tyrosine Kinase Activation and Enhances Glioma Radiosensitivity. *Clin Cancer Res* 2019; **25**: 784-95.

LEGENDS FOR FIGURES

Fig 1. Proteomic analysis of cSCCs reveals multiple significantly differentially-abundant proteins between P-M and P-NM tumours. (a) Representative photo of microdissected tumour, scale bar = 5 mm. Volcano plot of proteins identified by mass spectrometry in discovery group following (b) 1D and (c) 2D separation; values are shown as P-M relative to P-NM. *P* values were calculated using Mann Whitney U test. Fold changes for individual proteins were calculated by dividing the mean of P-M by the mean of P-NM; blue $P > 0.05$, green $P < 0.05$, red $P < 0.01$. Venn diagrams of (d) total number of unique proteins identified in 1D and 2D proteomes and (e) number of significantly differentially expressed proteins between P-M and P-NM cSCCs. Topological data analysis (which analyses datasets using systems derived from topology) of (f) whole 1D proteome and (g) whole 2D proteome demonstrates separation of samples according to metastasis status; nodes represent a cluster of samples (2 or more) with highly similar proteomes; edges (lines between nodes) indicate similarity between the clusters.

Fig 2. Lower CD8+ and CD1a+ immune cell frequencies in the primary tumour associate with development of cSCC metastases. (a, b) Immunohistochemical staining for CD8 and CD1a revealed significantly more CD8+ cells and CD1a+ cells in P-NM than P-M. (c, d). Decreased frequencies of CD8+ cells and CD1a+ cells in the tumour/tumoral immune infiltrate are significantly associated with reduced time to metastasis.

Fig 3. Expression of genes encoding for proteins which were significantly differentially expressed between P-M and P-NM were identified as markers of survival in other types of SCC. (a, c, e, g, i, k, m) Examples of proteins that were significantly differentially expressed between P-M and P-NM cSCCs; median with interquartile range; Mann Whitney U test for significance. (b, d, f, h, j, l, n) TCGA data demonstrates that expression of genes encoding for

relevant protein have significant effects on survival in cervical, oropharyngeal, oesophageal and lung SCC. High/low expression was defined as above and below median, respectively. (o) Representative immunohistochemistry stains of L-Plastin in P-M and P-NM cSCCs. (p) Immunohistochemical quantification of L-Plastin in cSCCs corroborated proteomic results, identifying significantly more L-Plastin+ cells in P-M than in P-NM tumour groups.

Fig 4. Multiple reaction monitoring (MRM) mass spectroscopy confirms higher ANXA5 and DDOST expression in P-M than P-NM cSCCs. MRM of (a – d) discovery group cSCCs and (e – h) validation group cSCCs demonstrated that ANXA5 and DDOST protein levels are increased in P-M as compared with P-NM. Data for individual peptides in (a, e) DDOST, (c, g) ANXA5 and mean+/-SD for (b, f) DDOST and (d, h) ANXA5.

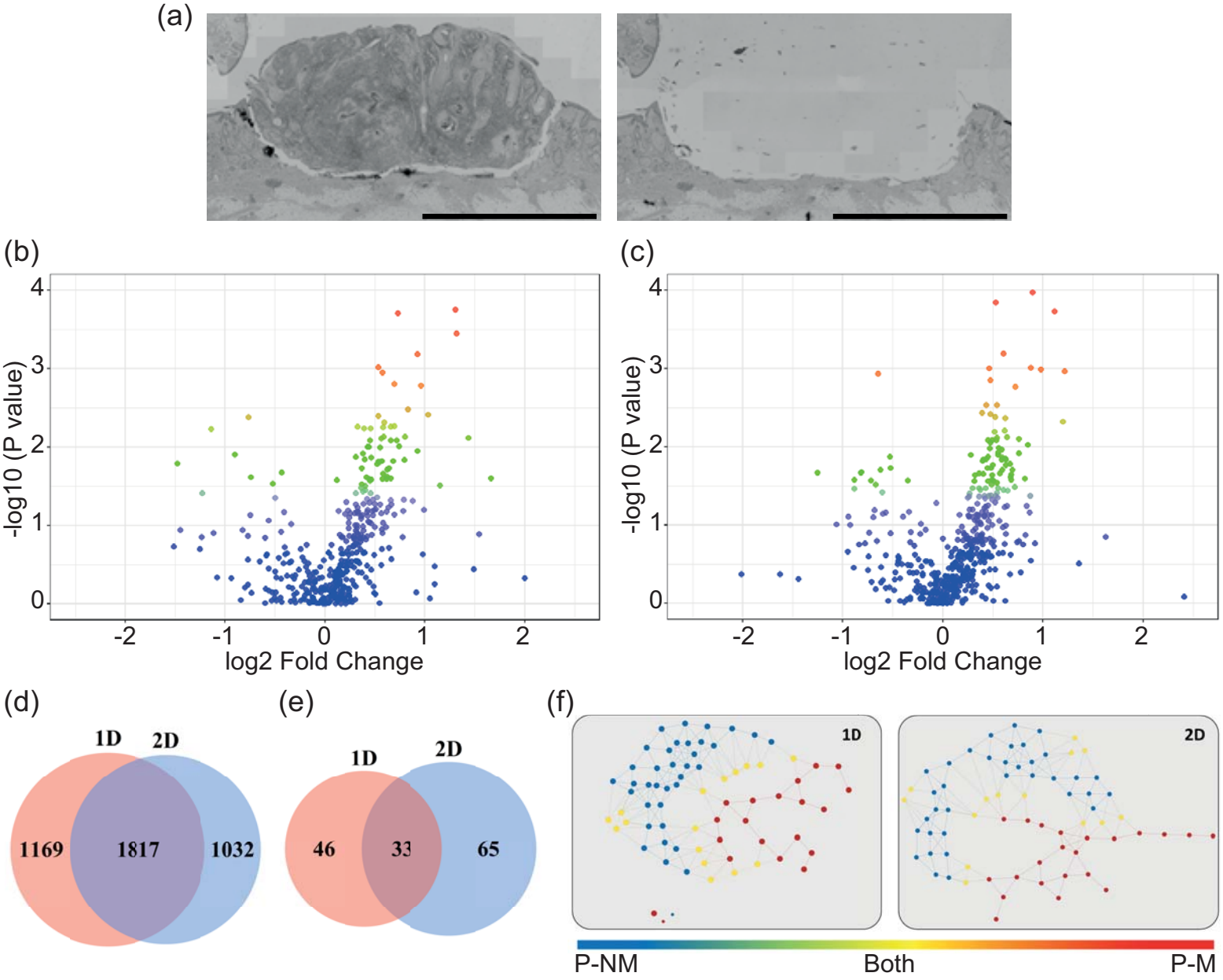
Fig 5. High ANXA5 and DDOST expression is associated with reduced time to metastasis in cutaneous, cervical and oropharyngeal SCC. (a) ANXA5 and DDOST levels have a significant effect on time to cSCC metastasis. (b, c) Kaplan-Meier plots showing 5-year overall survival for cSCCs based on (b) P-M and P-NM status and (c) expression level of ANXA5 and DDOST. (d, e) TCGA data signifying that expression of genes encoding for ANXA5 and DDOST has a significant effect on survival in (d) cervical SCC and (e) oropharyngeal SCC. High denotes both ANXA5 and DDOST protein abundance or gene expression above median. (f) ROC curve of model produced from MRM data performs better than current guidelines in clinical use; stacked ensemble model was created using all peptide MRM data as predictors.

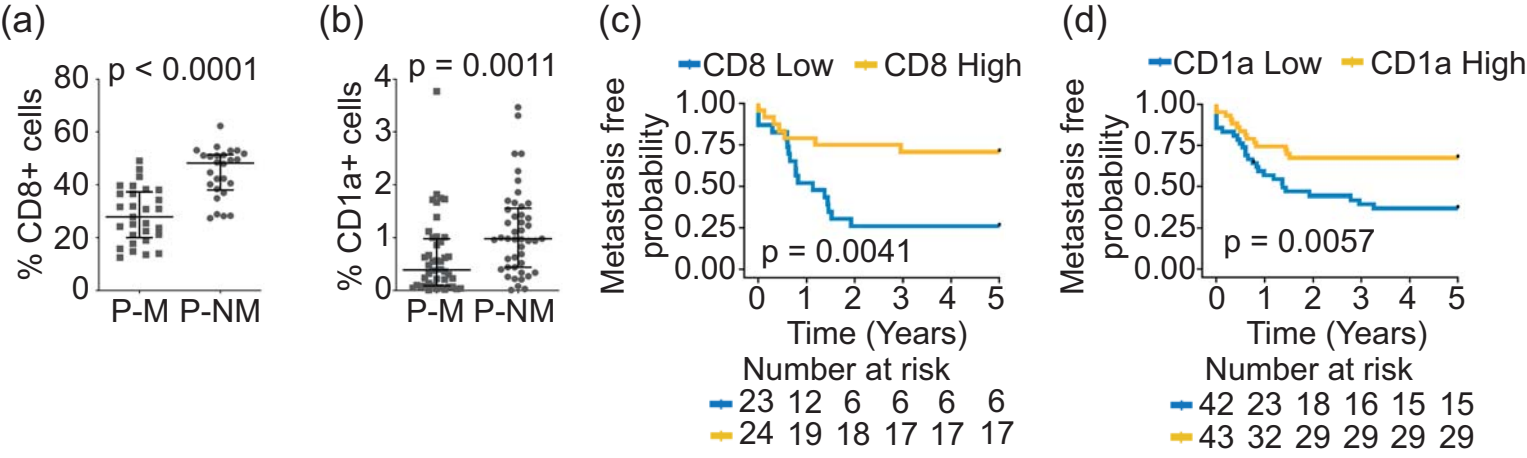
Table 1. Clinicopathological characteristics of study subjects.

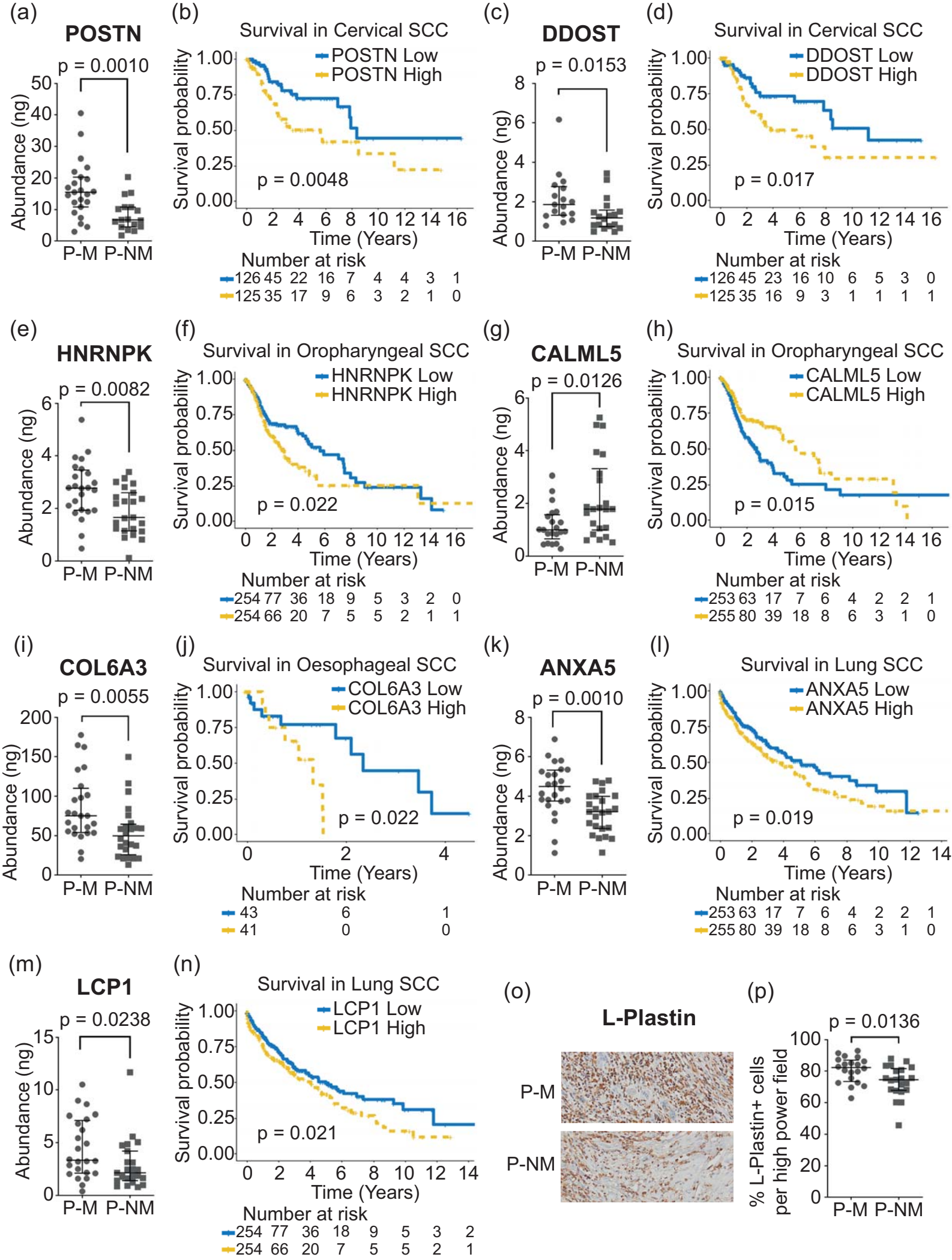
*Some samples were used for proteomic and immunohistochemistry analysis, whereas other samples were used for proteomic or immunohistochemistry analysis according to amount of

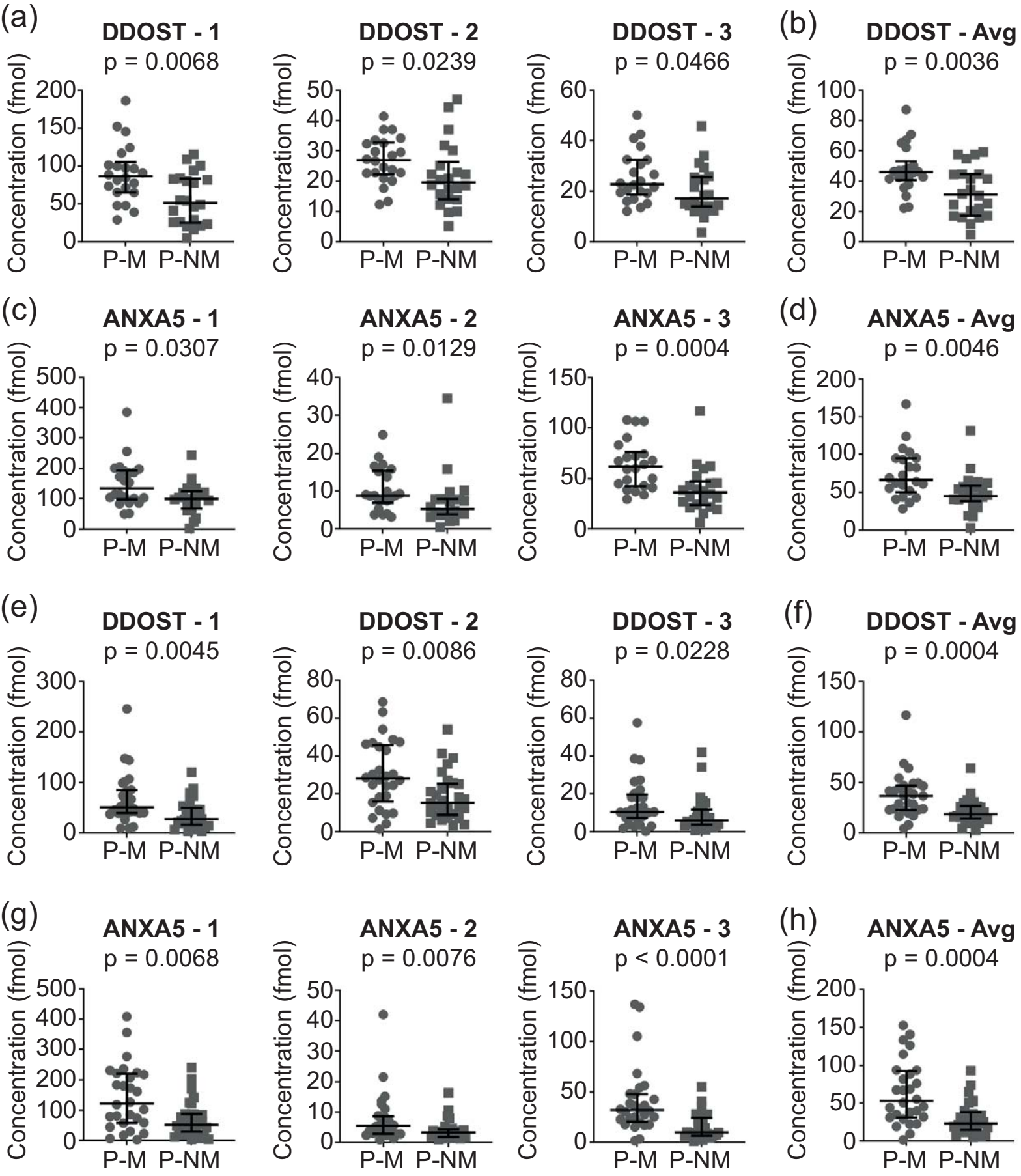
tissue available; **2 samples from each group were removed during MRM analysis due to limited amount of tissue available.

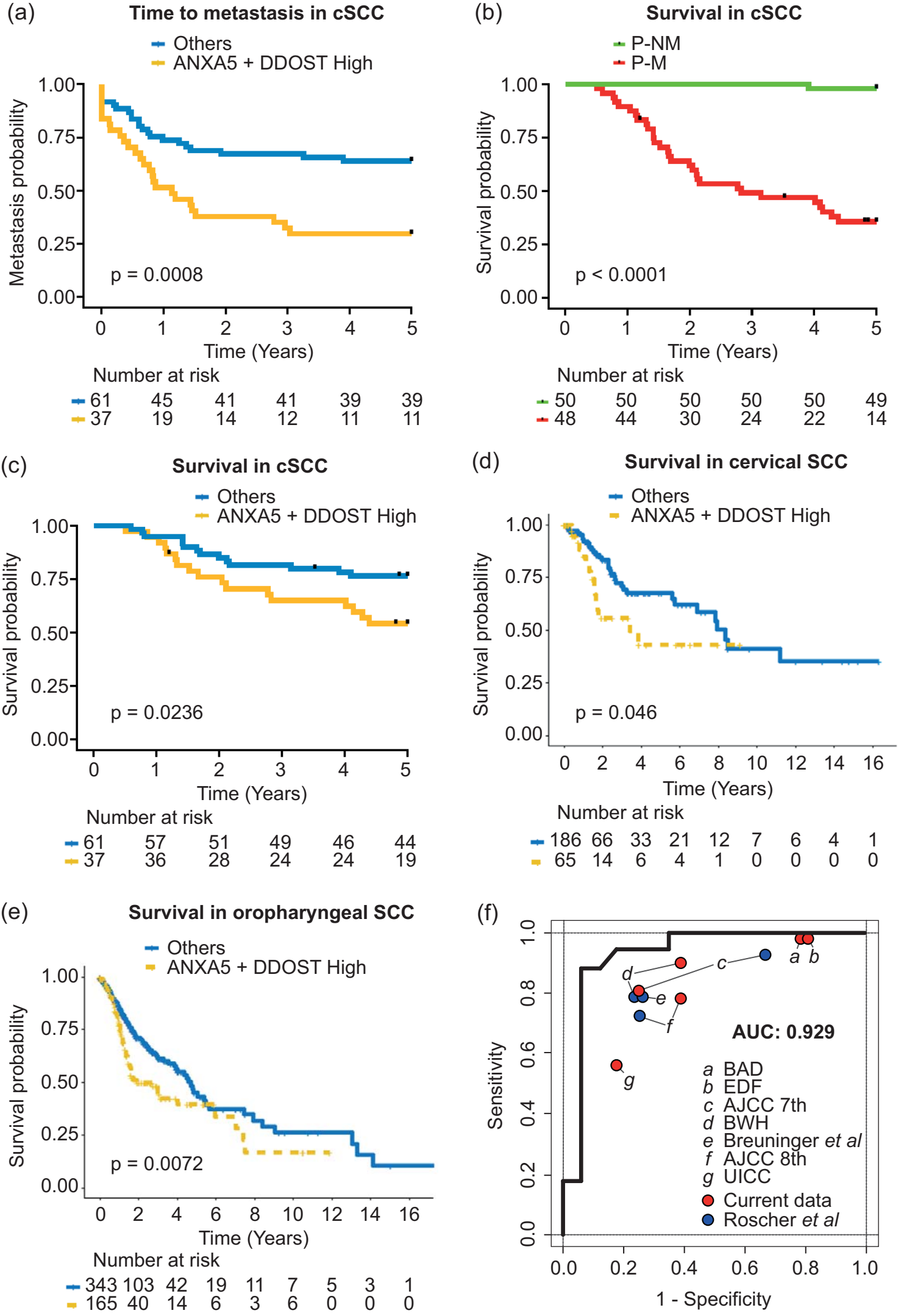
Table 2. Significantly differentially expressed proteins identified in both 1D and 2D proteomics and a summary of their respective Uniprot descriptions.


(a)
(b)
-log10 (P value)
log2 Fold Change
(c)
-log10 (P value)
log2 Fold Change
(d)
1D
2D
1169
1817
1032
(e)
1D
2D
46
33
65
(f)
1D
2D
P-NM
Both
P-M


(a)
p < 0.0001
% CD8+ cells
80
60
40
20
0
P-M P-NM
(b)
p = 0.0011
% CD1a+ cells
4
3
2
1
0
P-M P-NM
(c)
CD8 Low
CD8 High
Metastasis free probability
1.00
0.75
0.50
0.25
0.00
p = 0.0041
0 1 2 3 4 5
Time (Years)
Number at risk
23 12 6 6 6 6
24 19 18 17 17 17
(d)
CD1a Low
CD1a High
Metastasis free probability
1.00
0.75
0.50
0.25
0.00
p = 0.0057
0 1 2 3 4 5
Time (Years)
Number at risk
42 23 18 16 15 15
43 32 29 29 29 29


(a)
POSTN
p = 0.0010
Abundance (ng)
P-M P-NM
(b)
Survival in Cervical SCC
POSTN Low
POSTN High
Survival probability
p = 0.0048
Time (Years)
Number at risk
126 45 22 16 7 4 4 3 1
125 35 17 9 6 3 2 1 0
(c)
DDOST
p = 0.0153
Abundance (ng)
P-M P-NM
(d)
Survival in Cervical SCC
DDOST Low
DDOST High
Survival probability
p = 0.017
Time (Years)
Number at risk
126 45 23 16 10 6 5 3 0
125 35 16 9 3 1 1 1 1
(e)
HNRNPK
p = 0.0082
Abundance (ng)
P-M P-NM
(f)
Survival in Oropharyngeal SCC
HNRNPK Low
HNRNPK High
Survival probability
p = 0.022
Time (Years)
Number at risk
254 77 36 18 9 5 3 2 0
254 66 20 7 5 5 2 1 1
(g)
CALML5
p = 0.0126
Abundance (ng)
P-M P-NM
(h)
Survival in Oropharyngeal SCC
CALML5 Low
CALML5 High
Survival probability
p = 0.015
Time (Years)
Number at risk
253 63 17 7 6 4 2 2 1
255 80 39 18 8 6 3 1 0
(i)
COL6A3
p = 0.0055
Abundance (ng)
P-M P-NM
(j)
Survival in Oesophageal SCC
COL6A3 Low
COL6A3 High
Survival probability
p = 0.022
Time (Years)
Number at risk
43 6 1
41 0 0
(k)
ANXA5
p = 0.0010
Abundance (ng)
P-M P-NM
(l)
Survival in Lung SCC
ANXA5 Low
ANXA5 High
Survival probability
p = 0.019
Time (Years)
Number at risk
253 63 17 7 6 4 2 2 1
255 80 39 18 8 6 3 1 0
(m)
LCP1
p = 0.0238
Abundance (ng)
P-M P-NM
(n)
Survival in Lung SCC
LCP1 Low
LCP1 High
Survival probability
p = 0.021
Time (Years)
Number at risk
254 77 36 18 9 5 3 2
254 66 20 7 5 5 2 1
(o)
L-Plastin
P-M
P-NM
(p)
p = 0.0136
% L-Plastin+ cells per high power field
P-M P-NM


(a)
DDOST - 1
p = 0.0068
DDOST - 2
p = 0.0239
DDOST - 3
p = 0.0466
(b)
DDOST - Avg
p = 0.0036
(c)
ANXA5 - 1
p = 0.0307
ANXA5 - 2
p = 0.0129
ANXA5 - 3
p = 0.0004
(d)
ANXA5 - Avg
p = 0.0046
(e)
DDOST - 1
p = 0.0045
DDOST - 2
p = 0.0086
DDOST - 3
p = 0.0228
(f)
DDOST - Avg
p = 0.0004
(g)
ANXA5 - 1
p = 0.0068
ANXA5 - 2
p = 0.0076
ANXA5 - 3
p < 0.0001
(h)
ANXA5 - Avg
p = 0.0004
Concentration (fmol)
P-M
P-NM


(a)
Time to metastasis in cSCC
Others
ANXA5 + DDOST High
Metastasis probability
1.00
0.75
0.50
0.25
0.00
p = 0.0008
0 1 2 3 4 5
Time (Years)
Number at risk
61 45 41 41 39 39
37 19 14 12 11 11
(b)
Survival in cSCC
P-NM
P-M
Survival probability
1.00
0.75
0.50
0.25
0.00
p < 0.0001
0 1 2 3 4 5
Time (Years)
Number at risk
50 50 50 50 50 49
48 44 30 24 22 14
(c)
Survival in cSCC
Others
ANXA5 + DDOST High
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.0236
0 1 2 3 4 5
Time (Years)
Number at risk
61 57 51 49 46 44
37 36 28 24 24 19
(d)
Survival in cervical SCC
Others
ANXA5 + DDOST High
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.046
0 2 4 6 8 10 12 14 16
Time (Years)
Number at risk
186 66 33 21 12 7 6 4 1
65 14 6 4 1 0 0 0 0
(e)
Survival in oropharyngeal SCC
Others
ANXA5 + DDOST High
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.0072
0 2 4 6 8 10 12 14 16
Time (Years)
Number at risk
343 103 42 19 11 7 5 3 1
165 40 14 6 3 6 0 0 0
(f)
Sensitivity
0.0 0.2 0.4 0.6 0.8 1.0
1 - Specificity
0.0 0.2 0.4 0.6 0.8 1.0
a b c d e f g
AUC: 0.929
a BAD
b EDF
c AJCC 7th
d BWH
e Breuninger et al
f AJCC 8th
g UICC
Current data
Roscher et al

	P-M	P-NM
n	58	65
Male/Female	44 / 14 (75.9% / 24.1%)	47 / 18 (72.3% / 27.7%)
Age (median)	82 (51 - 98)	75 (47 - 94)
Site - Head and neck	43 (74.1%)	44 (67.7%)
Site - Trunk	3 (5.2%)	5 (7.7%)
Site - Upper limb	8 (13.8%)	9 (13.9%)
Site - Lower limb	4 (6.9%)	7 (10.8%)
Well differentiated	1 (1.7%)	22 (33.9%)
Moderately differentiated	20 (34.5%)	33 (50.8%)
Poorly differentiated	37 (63.8%)	10 (15.4%)
Perineural invasion	12 (20.7%)	3 (4.6%)
Immunosuppressed	8 (13.8%)	5 (7.7%)
Clarks I	0	0
Clarks II	0	2 (3.1%)
Clarks III	2 (3.4%)	10 (15.4%)
Clarks IV	21 (36.2%)	39 (60%)
Clarks V	34 (58.6%)	9 (13.8%)
Mean tumour diamter (mm)	28.56 (±28.95)	13.23 (±8.26)
Mean tumour depth (mm)	7.46 (±5.78)	3.91 (±2.49)

Gene ID	Uniprot ID	Description †	1D Fold Change*	1D P value	2D Fold Change*	2D P value
ANXA5	P08758	Anticoagulant protein.	0.4416	0.010	0.4631	0.001
CALML5	Q9NZT1	Binds to calcium; may be involved in terminal differentiation of keratinocytes.	-0.9011	0.013	-0.8085	0.021
CCT8	P50990	Assists folding of proteins after ATP hydrolysis.	0.6436	0.019	0.5699	0.022
COL6A3	P12111	Cell-binding protein.	0.6616	0.006	0.5020	0.017
DDOST	P39656	Essential subunit of OST. Catalyses transfer of oligosaccharide to asparagine residues.	0.6863	0.015	0.5416	0.029
FGB	P02675	Major function in haemostasis. Guides cell migration during re-epithelialisation.	0.6621	0.025	0.6947	0.020
GANAB	Q14697	Subunit of glucosidase 2.	0.7464	0.010	0.4810	0.019
GDI2	P50395	Regulates GDP/ GTP.	0.4178	0.024	0.4447	0.019
HIST1H	P62805	Inhibits GDP dissociation from Rab proteins to allow binding of GTP.	0.3125	0.014	0.2653	0.044
HNRNPA2B1	P22626	Helps package other nascent hnRNPs.	0.6627	0.008	0.7237	0.002
HNRNPK	P61978	Major pre-mRNA binding protein. Important for P53s response to DNA damage.	0.4549	0.008	0.5364	0.008
HSP90AA1	P07900	Molecular chaperone.	0.4242	0.047	0.4200	0.042
HSP90AB1	P08238	Molecular chaperone.	0.6965	0.005	0.4977	0.008
KRT2	P35908	Keratinocyte activation, proliferation and keratinisation. Role in epidermal barrier / terminal cornification.	-1.4734	0.016	-1.2476	0.022
KRT6B	P04259	Expressed in filiform papillae of tongue, epithelial lining of oral mucosa and oesophagus and outer sheath of hair follicles.	-0.4318	0.021	-0.5240	0.014
LCP1	P13796	Co-stimulates activation of T-cells with CD3, CD2 and CD28.	0.6266	0.024	0.6851	0.026
LUM	P51884	Extracellular protein involved in collagen fibril organisation and epithelial cell migration and tissue repair.	0.8040	0.007	0.5327	0.042
MSN	P26038	Involved in cytoskeletal structuring. Helps regulate the proliferation, migration and adhesion of lymphoid cells.	0.3722	0.019	0.4947	0.023
MYL6	P60660	Involved in muscle contraction and ATP-dependant actin based motility.	0.5658	0.008	0.4369	0.014
NCL	P19338	Major constituent of nucleolus in growing cells.	0.5705	0.015	0.3665	0.044
P4HB	P07237	Catalyses formation, breakage and rearrangement of disulphide bonds. Can promote Th2 T helper cell migration.	0.7337	0.000	0.6312	0.017
PHB2	Q99623	Recruits histone deacetylases to mediate transcriptional repression by hormone receptors.	0.6266	0.020	0.6745	0.017
POSTN	Q15063	Secreted EMP associated with epithelial-mesenchymal transition. Binds to integrins, activating Akt-PKB and FAK signalling pathways.	1.0328	0.004	0.9814	0.001
PPIA	P62937	Catalyses folding of proteins. Helps induce inflammatory response in the presence of ROS.	0.5223	0.017	0.5397	0.008
PRDX5	P30044	Reduces hydrogen peroxide to water. Helps protect against oxidative stress.	0.6618	0.048	0.6838	0.023
RPS13	P62277	Ribosomal protein. Catalyses protein synthesis.	0.5897	0.007	0.4798	0.004
RPS20	P60866	Ribosomal protein. Catalyses protein synthesis.	0.3125	0.039	0.3192	0.027
RPS7	P62081	Ribosomal protein. Catalyses protein synthesis. Required for rRNA maturation.	0.5779	0.001	0.3631	0.034
SFPQ	P23246	Required for pre-mRNA splicing.	0.5868	0.006	0.6435	0.008

TGFBI	Q15582	Involved in cell adhesion and possibly cell-collagen adhesion. Binds several integrins.	0.9283	0.001	1.1167	0.000
TKT	P29401	Connects glycolysis to pentose phosphate pathway. Important for NADPH production in tissues undergoing biosynthesis.	0.4651	0.026	0.5216	0.004
TNC	P24821	EMP that guides migrating neurones and axons during development. Thought to stimulate angiogenesis in cancers.	0.9295	0.011	0.8501	0.009
TUBB	P07437	Major constituent of microtubules.	0.4267	0.010	0.4720	0.011

† Sourced from Uniprot. *Log2 Fold change. OST= N-oligosaccharyl transferase. EMP= extracellular matrix protein

SUPPLEMENTARY MATERIALS AND METHODS

Sample preparation for mass spectrometry

Three 10µm sections of each FFPE sample were cut and mounted onto glass slides. Sections were deparaffinised, rehydrated, then stained with Mayer's hematoxylin. Tumor and tumoral immune infiltrate were microdissected from surrounding skin and transferred to 100µl protein extraction buffer (containing 0.2% RapiGest SF (Waters), 50mM ammonium bicarbonate, 5mM dithiothreitol) and kept on ice for 45 minutes. Samples were heated at 105°C for 30 minutes, cooled on ice for 5 minutes, then heated to 80°C for 2 hours, cooled on ice for 5 minutes before being reduced in 5mM dithioerythritol at 60°C for 30 minutes. Samples were alkylated with 15mM iodacetamide for 30 minutes in the dark at room temperature and then digested overnight in 1µg trypsin at 37°C. Following addition of 0.5% trifluoroacetic acid (TFA), samples were incubated at 37°C for 30 minutes, then centrifuged at 15,000g for 15 minutes and supernatant collected and lyophilised in an Eppendorf Concentrator-5301 before reconstitution in 150µl 0.5% TFA in water. Samples were cleaned using an Empore^{TM} C18 plate (Sigma, 66875-U) and washed twice with 0.5% TFA/water before eluting with 80% acetonitrile/water. Samples were then lyophilised and reconstituted in 0.5% TFA/water and peptide concentration of resulting cSCC "proteomic-ready" sample determined using a Direct Detect Spectrometer (Merck).

Discovery liquid chromatography mass spectrometry (LC-MS^E)

100fm digested enolase standard (Waters) was added to 3.75µg cSCC "proteomic-ready" sample for absolute quantification.[10] Peptides were introduced to a nanoACQUITY UPLC system (Waters) and injected into a 5µl loop before trapping onto a Symmetry-C18 180µm x 20mm trap column (Waters). For one-dimensional (1D) LC, the sample was eluted off the trap column and separated on a 75µm I.D x 250mm, 1.7µm particle size C18 analytical column (Waters) using buffer A1:buffer B mixture (buffer A1 0.1% formic acid in water, buffer B 0.1% formic acid in acetonitrile) with linear gradient of 1 to 50% organic buffer B over a 150 minute run, with final 60% buffer B wash. A constant flow rate of 300nl/min was used and 20µl/min for trapping. Two-dimensional (2D) LC was employed by adsorbing the sample to a high pH column (XBridge-BEH130 C18 5µm 300x50 nano) at constant flow rate of 1µl/min with buffer A2 (20mM ammonium formate in water) before eluting aliquots at buffer B compositions; 11.1%, 14.5%, 17.4%, 20.8%, 45% and 65%. These aliquots were then trapped and separated as per one-dimensional LC. After LC separation, samples were ionised using electrospray ionisation into a Waters Synapt-G2-Si mass spectrometer operating in MS^E mode. Ion mobility mode utilising low (5v) and high (20-40v) collision energy was enabled and data between 50 to 2000 m/z was acquired. Three blank runs were conducted between each sample to ensure no carry over between samples. Samples were randomly batched into groups of 12. Standards were run at the beginning and end of every batch to assess instrument performance. MS data was searched against the human SwissProt database (November 2016) allowing for deamidation of asparagine and glutamine, oxidation of methionine, and hydroxymethylation of cysteine with fixed modifications of carbamidomethylation of cysteine.

SUPPLEMENTARY FIGURES

Supplementary Table 1. List of significantly differentially expressed proteins in 1D discovery proteomics between P-M and P-NM cSCC. Grey shading indicates proteins identified in both 1D and 2D proteomic analysis.

Uniprot ID	Gene ID	log2 Fold Change	p-value
P61158	ACTR3	0.961506898	5.21E-06
P50991	CCT4	1.30927292	0.000178
P07237	P4HB	0.733735226	0.000197
P50454	SERPINH1	1.317776969	0.000362
Q15582	TGFBI	0.928341791	0.000667
P15880	RPS2	0.535540549	0.000977
P62081	RPS7	0.577872021	0.001138
Q15019	SEPT2	0.699315588	0.00159
P09382	LGALS1	0.966564061	0.001659
P13010	XRCC5	0.831577845	0.003321
Q15063	POSTN	1.032788051	0.003898
P29692	EEF1D	0.535534134	0.004054
Q9H299	SH3BGRL3	-0.761111358	0.004194
P62857	RPS28	0.594810883	0.004916
P08238	HSP90AB1	0.696523855	0.005466
P12111	COL6A3	0.6616087	0.00553
P60709	ACTB	0.331314871	0.00553
P23246	SFPQ	0.58683941	0.005641
P08133	ANXA6	0.457416134	0.00585
O43707	ACTN4	0.39691863	0.005863
Q9NV66	TYW1	-1.134978807	0.005908
P51884	LUM	0.80396623	0.007435
P62277	RPS13	0.589690303	0.007468
Q9NSB2	KRT84	1.438285492	0.007721
P60660	MYL6	0.565832754	0.007811
P22626	HNRNPA2B1	0.662690619	0.008254
P61978	HNRNPK	0.454874485	0.008254
P35222	CTNNB1	0.721477453	0.008257
Q14697	GANAB	0.746390079	0.0099
P07437	TUBB	0.426709653	0.009969
P08758	ANXA5	0.441600309	0.010052
P04844	RPN2	0.554418246	0.010152
P24821	TNC	0.929523896	0.011412
P11142	HSPA8	0.383725658	0.012193
Q9NZT1	CALML5	-0.901133441	0.012585
P62805	HIST1H	0.312532197	0.013583
P59998	ARPC4	0.466567609	0.013831
P36578	RPL4	0.402361344	0.014564
P16403	HIST1H1C	0.79468494	0.014741
P19338	NCL	0.570498	0.015221
P39656	DDOST	0.686281	0.015341
O00571	DDX3X	0.55198	0.015599
P46783	RPS10	0.533801	0.015599
P35908	KRT2	-1.47336	0.016373
P62937	PPIA	0.522324	0.01663
P26038	MSN	0.372223	0.018779
P50990	CCT8	0.643611	0.018874
Q99623	PHB2	0.626632	0.019505
P04259	KRT6B	-0.43177	0.021146
P62140	PPP1CB	0.582665	0.021844
P13796	LCP1	0.626618	0.023763
P50395	GDI2	0.417755	0.024282
P21810	BGN	-0.73761	0.024587
P02675	FGB	0.662066	0.025039
P35900	KRT20	1.661553	0.025217
P04792	HSPB1	0.43912	0.025692
P10599	TXN	0.578065	0.025747
Q07065	CKAP4	0.620822	0.025817
P29401	TKT	0.465121	0.026105
P09651	HNRNPA1	0.119659	0.026825
P52597	HNRNPF	0.450961	0.027294
Q9HCY8	S100A14	-0.51908	0.029751
P13639	EEF2	0.36846	0.030267
P35580	MYH10	1.153696	0.030978
P07741	APRT	0.406844	0.031729
Q02878	RPL6	0.365138	0.033587
O00148	DDX39A	0.408352	0.03601
P46940	IQGAP1	0.379869	0.036832
Q6KB66	KRT80	-1.22693	0.039027
P12109	COL6A1	0.461111	0.039192

P60866	RPS20	0.312545	0.039198
O75369	FLNB	0.524255	0.044172
P15088	CPA3	-0.49888	0.044723
P07900	HSP90AA1	0.424161	0.046572
P16144	ITGB4	0.791541	0.047448
P62318	SNRPD3	0.484703	0.047584
P30044	PRDX5	0.661837	0.047683
P42224	STAT1	0.882976	0.048652

Supplementary Table 2. List of significantly differentially expressed proteins in 2D discovery proteomic data between P-M and P-NM cSCCs. Grey shading indicates proteins identified in both 1D and 2D proteomic analysis.

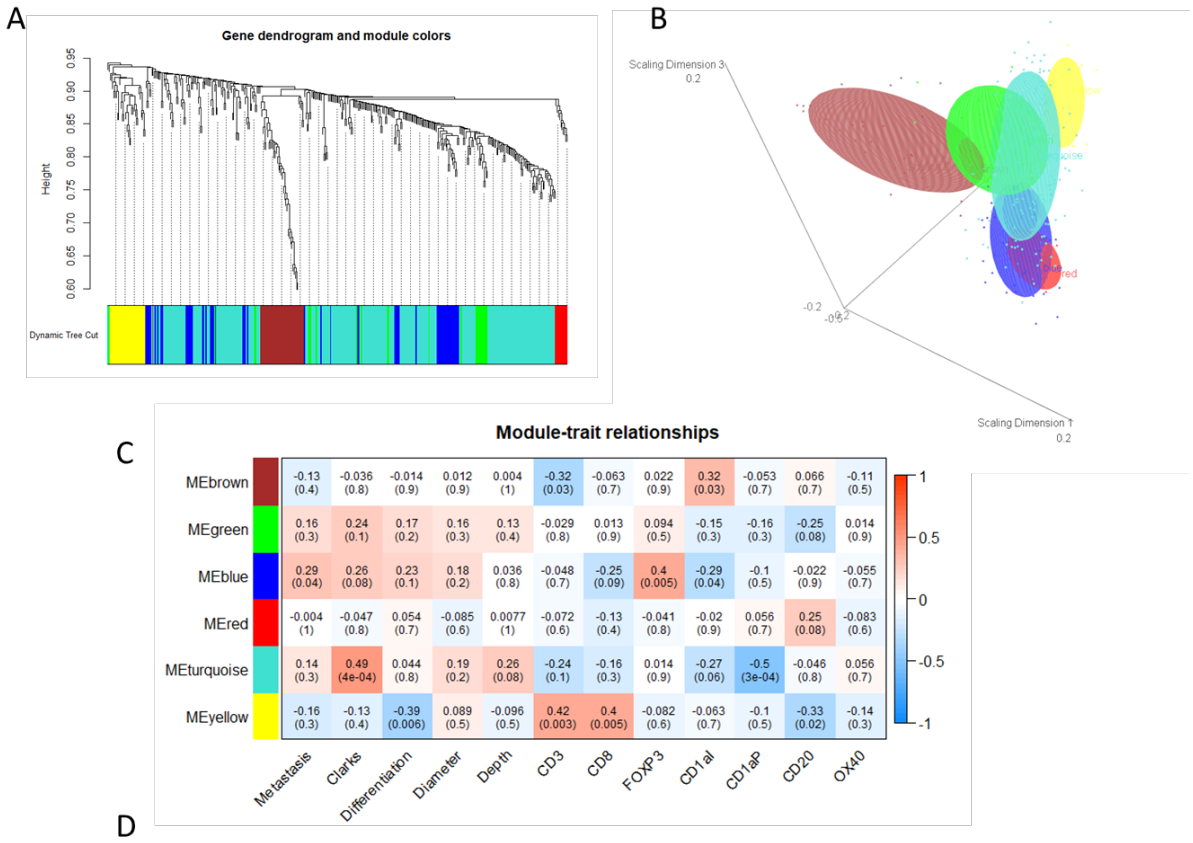
Uniprot ID	Gene ID	log2 Fold Change	p-value
P25398	RPS12	0.899331	1.08E-04
P61981	YWHAG	0.529794	0.000145
Q15582	TGFBI	1.116691	0.000187
P06396	GSN	0.609908	0.000655
P02751	FN1	0.883513	0.00098
P08758	ANXA5	0.463055	0.001
Q15063	POSTN	0.981432	0.001047
P31949	S100A11	1.220557	0.001104
P08779	KRT16	-0.64341	0.001175
P63000	RAC1	0.475662	0.001428
P22626	HNRNPA2B1	0.723681	0.001739
P02545	LMNA	0.436365	0.002937
P36957	DLST	0.541366	0.002971
P18206	VCL	0.391585	0.003743
P62277	RPS13	0.479793	0.003863
P29401	TKT	0.521632	0.004185
P46782	RPS5	0.62297	0.004337
P12110	COL6A2	1.202369	0.004823
P40121	CAPG	0.6193	0.006275
Q99497	PARK7	0.521044	0.006502
P23246	SFPQ	0.643473	0.007564
P62937	PPIA	0.53968	0.007713
P04179	SOD2	0.762658	0.007975
P08123	COL1A2	0.450851	0.00823
P08238	HSP90AB1	0.497668	0.00823
P61978	HNRNPK	0.536419	0.008247
P62158	CALM	0.477913	0.008814
P24821	TNC	0.850068	0.009468
Q07960	ARHGAP1	0.559179	0.010409
P07437	TUBB	0.471989	0.010616
P62314	SNRPD1	0.592314	0.010933
P60174	TPI1	0.556856	0.012021
P31146	CORO1A	0.759081	0.012238
P68104	EEF1A1	0.283546	0.012781
P09525	ANXA4	0.553413	0.01287
P04259	KRT6B	-0.52402	0.013583
P08670	VIM	0.454294	0.013583
P14625	HSP90B1	0.59337	0.013831
P02671	FGA	0.62589	0.014133
P60660	MYL6	0.436868	0.014429
Q03252	LMNB2	0.590182	0.014685
Q99878	HIST1H2AJ	0.407634	0.014719
P29590	PML	0.628868	0.014724
P23396	RPS3	0.399818	0.015319
Q99623	PHB2	0.67447	0.016553
P07237	P4HB	0.631181	0.017242
P12111	COL6A3	0.501978	0.017242
P27482	CALML3	-0.51899	0.018889
P50395	GDI2	0.444733	0.018889
Q14697	GANAB	0.481031	0.018922
P27824	CANX	0.5971	0.019724
P16615	ATP2A2	-0.62138	0.019818
P02675	FGB	0.694736	0.019933
Q9NZT1	CALML5	-0.80848	0.021169
P04264	KRT1	-0.81594	0.021718
P35908	KRT2	-1.24764	0.021718
P62249	RPS16	0.365977	0.021847
P50990	CCT8	0.569868	0.022002
Q96FW1	OTUB1	0.406082	0.022399
P30044	PRDX5	0.683793	0.022719
P26038	MSN	0.494652	0.02298
P63104	YWHAZ	0.323326	0.02298
P20700	LMNB1	0.676225	0.024282
P05141	SLC25A5	0.823864	0.025666
Q562R1	ACTBL2	0.495019	0.026073
P13796	LCP1	0.685076	0.026452
P01871	IGHM	-0.88295	0.026604
P35555	FBN1	0.508086	0.027037
P48668	KRT6C	-0.71584	0.027113
P02538	KRT6A	-0.34777	0.027147
P37802	TAGLN2	0.455563	0.027434
P60866	RPS20	0.319156	0.027496
P39656	DDOST	0.541621	0.029103
P01011	SERPINA3	0.490469	0.030512
P29508	SERPINB3	-0.66772	0.030512
Q99715	COL12A1	0.722631	0.032952

P00338	LDHA	0.298966	0.033683
O43390	HNRNPR	0.430315	0.034032
P01009	SERPINA1	0.554795	0.034206
P62081	RPS7	0.363127	0.034206
Q02388	COL7A1	-0.88471	0.034513
P11021	HSPA5	0.377159	0.035508
P07195	LDHB	0.6105	0.036002
Q05707	COL14A1	0.650654	0.03669
P55795	HNRNPH2	-0.60532	0.038253
O00299	CLIC1	0.539654	0.039171
P21333	FLNA	0.267202	0.039408
P00558	PGK1	0.485908	0.041488
P62899	RPL31	0.458861	0.042321
P30041	PRDX6	0.877492	0.042339
P07900	HSP90AA1	0.419973	0.042481
P51884	LUM	0.532743	0.042486
P19338	NCL	0.366465	0.043658
P62805	HIST4H	0.265299	0.043658
Q71UI9	H2AFV	0.546786	0.044845
P62269	RPS18	0.250738	0.045023
P30101	PDIA3	0.421946	0.045921
P27816	MAP4	0.496461	0.046927

Supplementary Table 3. Medians, interquartile ranges and P values for comparing P-M and P-NM cSCC groups.

Marker	P-M median (interquartile range)	P-NM median (interquartile range)	Fold change	P value
Cells				
CD8+ cells	27.9% (20.0 – 37.4%)	48.3% (38.1 – 51.4 ng)	0.578	< 0.0001
CD1a+ cells	0.39% (0.09 – 0.98%)	0.98% (0.44 – 1.56 ng)	0.398	0.0011
L-plastin + cells	82.3% (73.4 – 87.0%)	74.5% (67.7 – 87.9 ng)	1.105	0.0136
Proteomics discovery group				
POSTN	15.5 ng (10.8 - 20.3 ng)	6.7 ng (4.5 – 10.8 ng)	2.313	0.0010
DDOST	1.86 ng (1.32 – 2.77 ng)	1.18 ng (0.74 – 1.58 ng)	1.576	0.0153
HNRNPK	2.78 ng (1.94 – 3.46 ng)	1.66 ng (1.15 – 2.61 ng)	1.675	0.0082
CALML5	1.00 ng (0.66 – 1.58 ng)	1.79 ng (0.53 – 3.31 ng)	0.559	0.0126
COL6A3	75.2 ng (54.0 – 109.7 ng)	49.7 ng (25.6 – 54.6 ng)	1.513	0.0055
ANXA5	4.48 ng (3.76 – 5.31 ng)	3.22 ng (2.37 – 4.00 ng)	1.391	0.0010
LCP1	3.35 ng (2.12 – 7.13 ng)	2.14 ng (1.41 – 4.22 ng)	1.565	0.0238
MRM discovery group				
DDOST peptide 1	86.5 fmol (65.1 – 105.3 fmol)	51.6 fmol (25.1 – 83.2 fmol)	1.676	0.0068
DDOST peptide 2	26.9 fmol (22.2 – 32.8 fmol)	19.6 fmol (14.1 – 26.4 fmol)	1.372	0.0239
DDOST peptide 3	22.8 fmol (18.6 – 32.4 fmol)	17.2 fmol (13.9 – 25.6 fmol)	1.326	0.0466
DDOST average	46.0 fmol (40.6 – 53.3 fmol)	31.1 fmol (17.2 – 44.7 fmol)	1.479	0.0036
ANXA5 peptide 1	134.5 fmol (97.8 – 193.0 fmol)	99.1 fmol (68.6 – 125.0 fmol)	1.357	0.0307
ANXA5 peptide 2	8.83 fmol (6.92 – 15.32 fmol)	5.29 fmol (3.88 – 7.92 fmol)	1.669	0.0129
ANXA5 peptide 3	61.9 fmol (42.3 – 76.4 fmol)	36.3 fmol (23.6 – 47.4 fmol)	1.705	0.0004
ANXA5 average	66.7 fmol (50.1 – 95.1 fmol)	45.2 fmol (38.3 – 58.8 fmol)	1.476	0.0046
MRM validation group				
DDOST peptide 1	50.6 fmol (39.7 – 85.2 fmol)	27.8 fmol (16.4 – 47.6 fmol)	1.820	0.0045
DDOST peptide 2	28.1 fmol (16.2 – 45.9 fmol)	15.3 fmol (9.0 – 25.4 fmol)	1.837	0.0086
DDOST peptide 3	10.4 fmol (7.31 – 19.7 fmol)	5.94 fmol (3.76 – 11.7 fmol)	1.751	0.0228
DDOST average	36.9 fmol (22.9 – 47.0 fmol)	21.4 fmol (14.6 – 26.8 fmol)	1.724	0.0004
ANXA5 peptide 1	122.6 fmol (57.8 – 220.0 fmol)	51.7 fmol (28.6 – 87.5 fmol)	2.371	0.0068
ANXA5 peptide 2	5.48 fmol (2.87 – 8.62 fmol)	3.25 fmol (1.78 – 4.28 fmol)	1.686	0.0076
ANXA5 peptide 3	32.2 fmol (20.5 – 48.0 fmol)	9.80 fmol (6.67 – 24.4 fmol)	3.286	< 0.0001
ANXA5 average	53.1 fmol (31.3 – 92.3 fmol)	23.1 fmol (14.6 – 38.1 fmol)	2.299	0.0004

Supplementary Figure 1



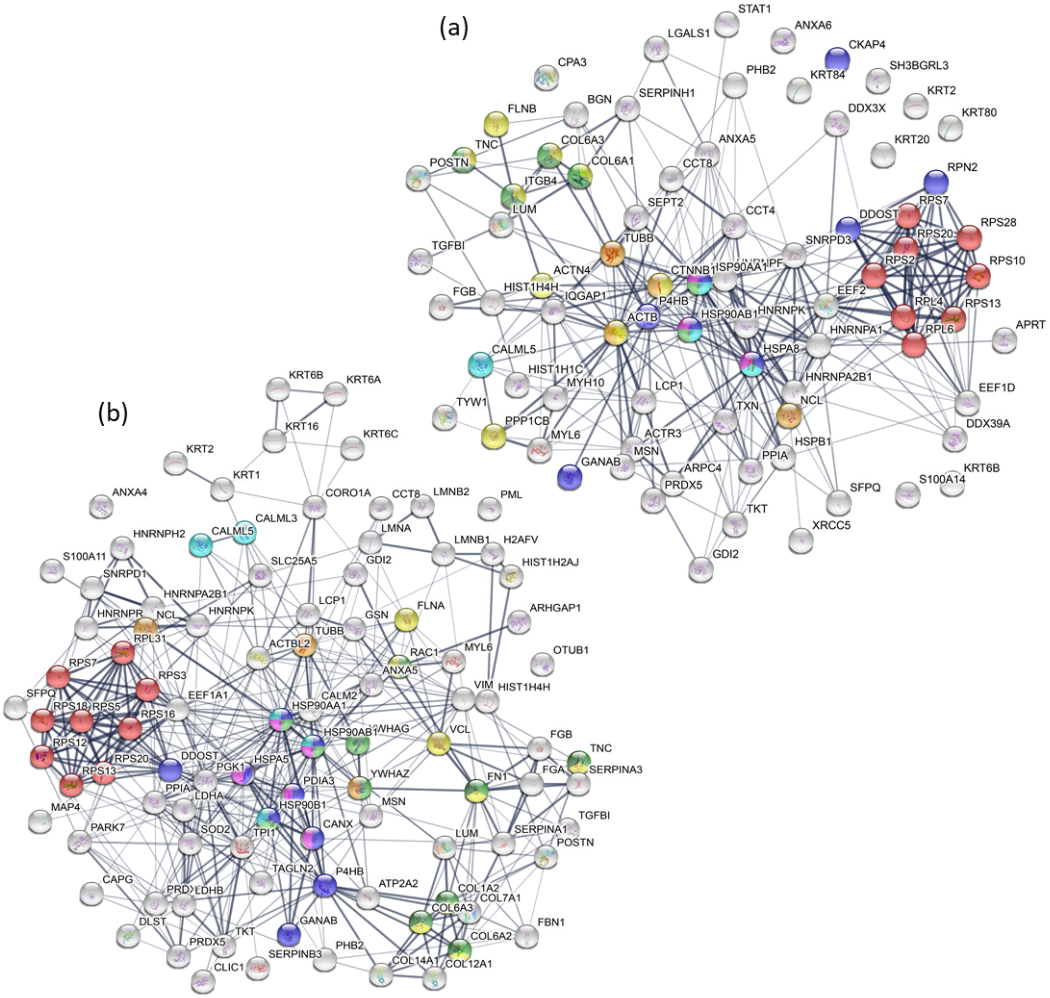
D

source	term name Protein databases (CORUM protein complexes)	term ID	n. of term genes	corrected p-value	BLUE	TURQUOISE	YELLOW	RED	BROWN	GREEN
cor	40S ribosomal subunit, cytoplasmic	CORUM:305	33	6.13e-09	9	8				
cor	iNOS-S100A8/A9 complex	CORUM:6827	3	1.86e-04			2			
cor	CCT complex (chaperonin containing TCP1 complex)	CORUM:126	8	1.96e-07		7				
cor	BBS-chaperonin complex	CORUM:6247	10	1.26e-04		6				
cor	Ribosome, cytoplasmic	CORUM:306	80	7.81e-14	15	17				
cor	Nop56p-associated pre-rRNA complex	CORUM:3055	104	3.22e-09	13	16				
cor	(S100A8/S100A9)2 heterotetramer	CORUM:6826	2	6.19e-05			2			
cor	40S ribosomal subunit, cytoplasmic	CORUM:338	31	3.25e-09	9	8				
cor	Emerin complex 52	CORUM:5615	23	2.79e-04		8				
source	term name Biological pathways (KEGG)	term ID	n. of term genes	corrected p-value	BLUE	TURQUOISE	YELLOW	RED	BROWN	GREEN
keg	Protein processing in endoplasmic reticulum	KEGG:04141	165	2.22e-05		17				
keg	Regulation of actin cytoskeleton	KEGG:04810	213	7.84e-04		17				
keg	Ribosome	KEGG:03010	136	1.03e-15	15	17				
keg	Metabolic pathways	KEGG:01100	1297	3.62e-05						17
keg	Glycolysis / Gluconeogenesis	KEGG:00010	69	4.16e-05		11				
keg	Legionellosis	KEGG:05134	55	3.95e-05		10				
keg	Biosynthesis of amino acids	KEGG:01230	76	8.44e-04		10				4
keg	Pathogenic Escherichia coli infection	KEGG:05130	55	2.18e-08		13				
keg	Salmonella infection	KEGG:05132	84	3.13e-04		11				
keg	Tight junction	KEGG:04530	172	2.02e-04		16				
source	term name Biological pathways (Reactome)	term ID	n. of term genes	corrected p-value	BLUE	TURQUOISE	YELLOW	RED	BROWN	GREEN
rea	Platelet degranulation	R-HSA-114608	127	1.51e-13		20				
rea	Apoptosis	R-HSA-109581	169	8.85e-05		16				
rea	Glycolysis	R-HSA-70171	72	5.41e-04		10				
rea	Interleukin-12 family signaling	R-HSA-447115	59	5.19e-07		12				3
rea	Peptide chain elongation	R-HSA-156902	90	1.47e-18	15	19				
rea	Neutrophil degranulation	R-HSA-6798695	476	3.30e-23	53	18				
rea	Viral mRNA Translation	R-HSA-192823	90	1.47e-18	15	17				
rea	Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	R-HSA-975956	96	4.13e-18	15	16				
rea	Host Interactions of HIV factors	R-HSA-162909	130	9.38e-04						6
rea	Selenocysteine synthesis	R-HSA-2408557	94	2.95e-18	15	17				
rea	Keratinization	R-HSA-6805567	216	2.07e-17			15			
rea	HSF1 activation	R-HSA-3371511	12	8.00e-04		5				
rea	Post-translational protein phosphorylation	R-HSA-8957275	108	5.51e-04		12				
rea	Beta oxidation of palmitoyl-CoA to myristoyl-CoA	R-HSA-77305	3	9.66e-06						3
rea	Apoptotic cleavage of cell adhesion proteins	R-HSA-351906	11	5.25e-04		3				
rea	RHO GTPases activate PAKs	R-HSA-5627123	21	3.88e-05		7				

Supplementary Figure 1. Weighted gene co-expression network analysis (WGCNA) reveals clusters of proteins which can be related to clinical and histological characteristics. The WGCNA package in R was

used with whole proteomic data to identify modules. A soft power threshold of 5 and a minimum module size of 10 was used. (a) Dendrogram and hierarchical clustering of proteins. (b) Multidimensional scaling plot of identified modules confirms clusters are separate. (c) Correlation matrix of modules to clinical/histopathological traits. (d) Pathway analysis of modules. CD1aI, CD1a intratumoural; CD1aP, CD1a peritumoural.

Supplementary Figure 2



(c)

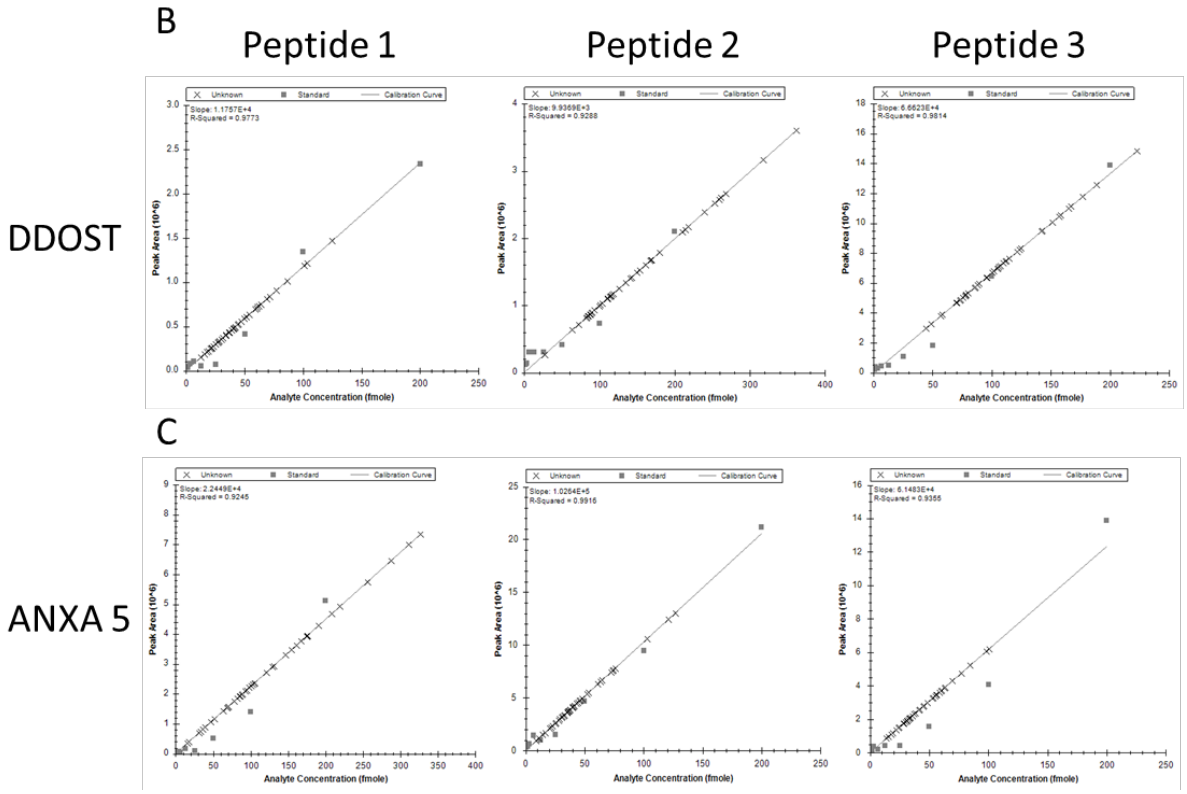
	KEGG Pathway	1D FDR	2D FDR
	Ribosome	8.78e-06	2.85e-06
	Protein processing in ER	1.82e-05	1.11e-05
	PI3K-Akt signalling	0.0342	5.39e-05
	Focal adhesion	1.08e-05	0.000308
	Antigen processing and presentation	0.0368	0.000735
	ECM-receptor interaction	0.00899	0.00215
	Oestrogen signalling pathway	0.0115	0.0029
	Pathogenic *E.coli* infection	0.0022	0.0426

Supplementary Figure 2. STRING analysis with KEGG pathway mapping identified several pathways significantly enriched in both 1D and 2D data. (a) STRING analysis of 1D significantly differentially expressed proteins. (b) STRING analysis of 2D significantly differentially expressed proteins. (c) KEGG pathway enrichment of significantly differentially expressed proteins comparing P-M against P-NM as base. FDR, false discovery rate.

Supplementary Figure 3

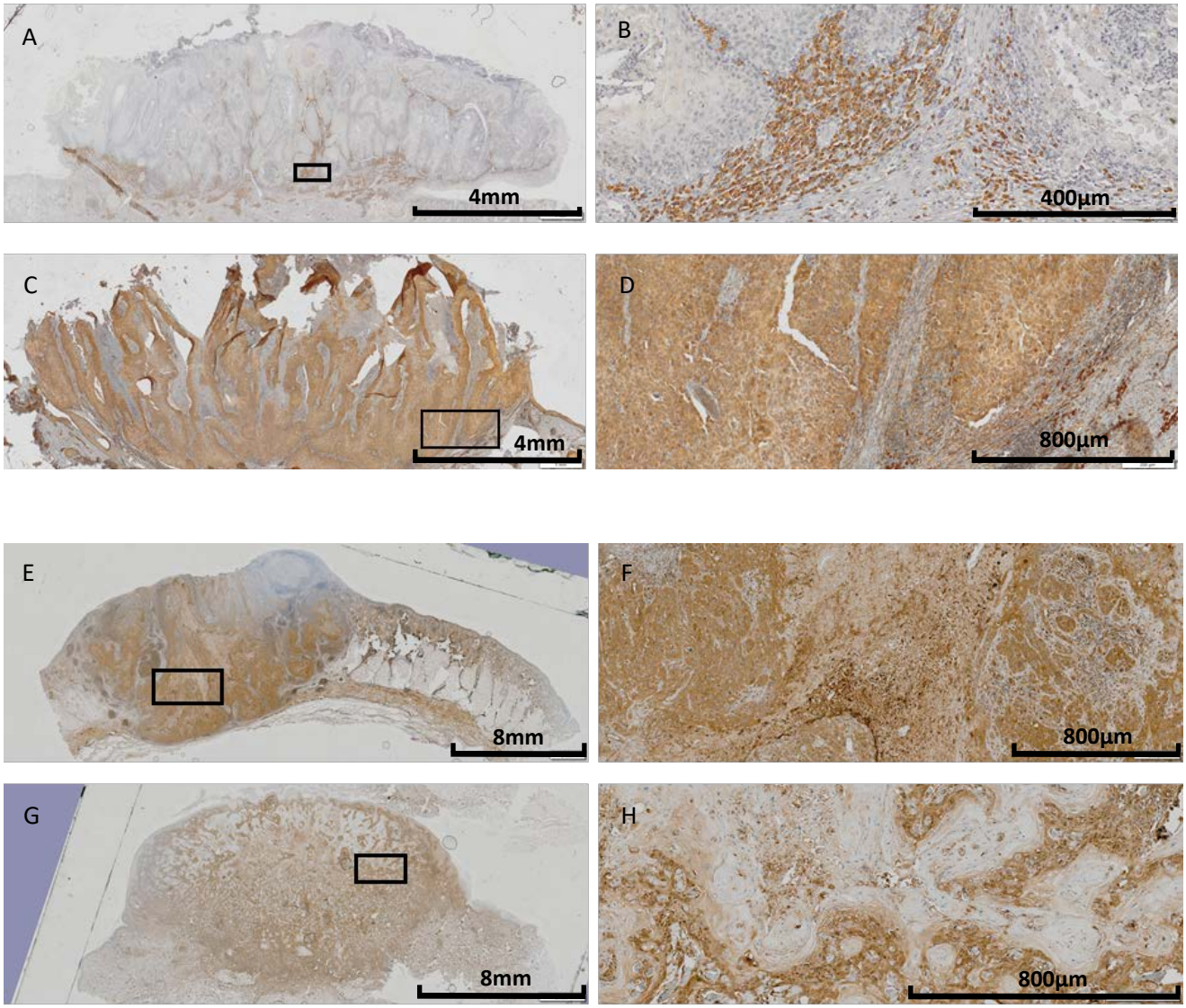
A

protein	peptide sequence	heavy labelled peptide sequence
DDOST	TLVLLDNLNVR	TLVLLDNLNV-[U-13C6,15N4-Arg]-acid
	GFELTFK	GFELTF-[U-13C6,15N2-Lys]-acid
	SSLNPILFR	SSLNPILF-[U-13C6,15N4-Arg]-acid
ANXA5	GLGTDEESILTLLTSR	GLGTDEESILTLLTS-[U-13C6,15N4-Arg]-acid
	FITIFGTR	FITIFGT-[U-13C6,15N4-Arg]-acid
	SEIDLFNIR	SEIDLFNI-[U-13C6,15N4-Arg]-acid



Supplementary Figure 3. Multiple reaction monitoring of cSCC samples. (a) Unique peptides and the heavy labelled versions used for MRM analysis. Unique peptides were identified using Skyline software. Calibration curves of (b) DDOST and (c) ANXA5. Each peptide was run at various concentrations to get a linear regression which was later used to predict concentration of native peptide.

Supplementary Figure 4



Supplementary Figure 4. Immunostaining of DDOST and ANXA5 shows presence of these proteins in tumour cells and in cells within the surrounding immune infiltrate. (a-d) Representative images of DDOST staining. (e-h) Representative images of ANXA5 staining.