



Proteomic profiling of soft tissue sarcomas with SWATH mass spectrometry

Martina Milighetti^{a,1}, Lukas Krasny^{a,1}, Alex T.J. Lee^{a,b,c,1}, Gabriele Morani^{c,d},
Cornelia Szecsei^a, Yingtong Chen^{c,d}, Nafia Guljar^a, Frank McCarthy^a, Christopher P. Wilding^a,
Amani Arthur^a, Cyril Fisher^b, Ian Judson^b, Khin Thway^{a,b}, Maggie C.U. Cheang^{c,d},
Robin L. Jones^{b,c}, Paul H. Huang^{a,*}

^a Division of Molecular Pathology, The Institute of Cancer Research, 237 Fulham Road, SW3 6JB London, UK

^b Sarcoma Unit, The Royal Marsden NHS Foundation Trust, 203 Fulham Road, SW3 6JJ London, UK

^c Division of Clinical Studies, The Institute of Cancer Research, 15 Cotswold Road, SM2 5NG Sutton, London, UK

^d Clinical Trials and Statistics Unit, The Institute of Cancer Research, 15 Cotswold Road, SM2 5NG Sutton, London, UK

ARTICLE INFO

Keywords:

Soft tissue sarcoma
Proteomics
Mass spectrometry
SWATH MS
FFPE
Biomarkers

ABSTRACT

Soft tissue sarcomas (STS) are a group of rare and heterogeneous cancers. While large-scale genomic and epigenomic profiling of STS have been undertaken, proteomic analysis has thus far been limited. Here we utilise sequential window acquisition of all theoretical fragment ion spectra mass spectrometry (SWATH-MS) for proteomic profiling of formalin fixed paraffin embedded (FFPE) specimens from a cohort of STS patients ($n = 36$) across four histological subtypes (leiomyosarcoma, synovial sarcoma, undifferentiated pleomorphic sarcoma and dedifferentiated liposarcoma). We quantified 2951 proteins across all cases and show that there is a significant enrichment of gene sets associated with smooth muscle contraction in leiomyosarcoma, RNA splicing regulation in synovial sarcoma and leukocyte activation in undifferentiated pleomorphic sarcoma. We further identified a subgroup of STS cases that have a distinct expression profile in a panel of proteins, with worse survival outcomes when compared to the rest of the cohort. Our study highlights the value of comprehensive proteomic characterisation as a means to identify histotype-specific STS profiles that describe key biological pathways of clinical and therapeutic relevance; as well as for discovering new prognostic biomarkers in this group of rare and difficult-to-treat diseases.

Significance

Soft tissue sarcomas (STS) are rare tumours that account for 1% of adult cancers and comprise of >100 histological subtypes. While genomic and transcriptomic analyses of multiple subtypes have been reported, few comprehensive proteomic analyses by mass spectrometry have been undertaken. Herein we used sequential window acquisition of all theoretical fragment ion spectra mass spectrometry (SWATH-MS) to characterise a cohort of 36 formalin fixed paraffin embedded (FFPE) specimens from STS patients across four major histological subtypes. Proteomic profiling data defines underlying biological pathways enriched in each histological subtype including both well-established and new signalling networks of functional importance. A subset of proteins were subjected to orthogonal validation using immunohistochemistry. Further analysis of the proteomic dataset identifies a panel of

133 proteins associated with patient outcome in the STS cohort under study. We anticipate the future application of SWATH-MS to additional STS subtypes and larger cohorts has the potential to facilitate diagnosis, deliver new therapeutic targets and define prognostic signatures in these rare and difficult-to-treat diseases.

1. Introduction

STS are a group of rare and heterogeneous malignancies of mesenchymal origin, comprising more than 100 distinct diagnostic subtypes that are primarily defined by histological characteristics [1]. Multiple gene expression-based studies across different histological subtypes have led to a deeper understanding of the oncogenic processes driving STS development and progression as well as the identification of new prognostic signatures for these cancers [2–5]. Whilst current diagnostic

* Corresponding author at: Division of Molecular Pathology, The Institute of Cancer Research, London SW3 6JB, United Kingdom.

E-mail address: paul.huang@icr.ac.uk (P.H. Huang).

¹ These authors contributed equally to this work.

evaluation of STS is reliant on histological assessment by specialist sarcoma histopathologists, supplemented by specific molecular tests in selected subtypes, recent large-scale genomic and epigenetic analyses have demonstrated the feasibility of refining the classification system of STS based on intrinsic underlying biology [3,4]. Collectively, these studies highlight the promise of molecular profiling strategies in improving our knowledge of drivers of sarcoma pathogenesis, providing complementary information to aid in the molecular classification of these heterogeneous tumours, identify new therapeutic targets and develop clinically relevant prognostic biomarkers.

While the advent of next generation sequencing (NGS) has accelerated the use of genomic and epigenetic profiling in STS, proteomic analysis in this disease has been limited [6]. Comprehensive analysis of the tumour proteome is highly informative as proteins represent the largest class of druggable targets and directly reflect the functional state of biological pathways [7]. Proteomic analysis of STS has been performed by The Cancer Genome Atlas (TCGA) consortium in a cohort of 173 flash-frozen sarcoma specimens across six histological subtypes [3]. However, this study utilised the reverse phase protein array (RPPA) platform which is limited to the analysis of 192 proteins/phosphoproteins. Other studies that utilise mass spectrometry (MS) have largely been limited to 1–2 subtypes including recent proteomic analyses of gastrointestinal stromal tumours (GIST) and undifferentiated pleomorphic sarcomas (UPS) which identified distinct clinical and molecular subgroups [6,8,9]. To date few MS-based analysis has been conducted across multiple histological subtypes in STS.

In this study, we utilise sequential window acquisition of all theoretical fragment ion spectra (SWATH)-MS to undertake proteomic profiling of FFPE tissue specimens in a cohort of STS patients across four histological subtypes ($n = 36$). SWATH-MS is a next-generation proteomics method that offers high reproducibility in protein identification across multiple samples [10]. Due to the stochastic nature of precursor ion selection for fragmentation in conventional proteomic approaches such as data-dependent acquisition (DDA) mass spectrometry, there are routinely many missing values in large-scale proteomic studies, resulting in a reduction in the reproducibility of protein identification and quantification. In contrast, SWATH-MS selects and fragments all detected precursor ions in a sample generating a highly reproducible digital proteome map for each specimen in the cohort. Herein we characterise the intrinsic biological pathways associated with specific histotypes and explore the association of proteomic data with patient outcome. This study represents, to our knowledge, the most comprehensive analysis of the proteome in multiple STS subtypes to date and highlights the potential of proteomic profiles as predictors for patient outcome in these rare cancers.

2. Material and methods

2.1. Patients and tumour specimens

Use of archival FFPE tumour samples and linked anonymised patient data was approved by Institutional Review Board as part of the PRO-SPECTUS study, a Royal Marsden-sponsored non-interventional translational protocol (CCR 4371, REC 16/EE/0213). FFPE tissue from surgically resected primary tumours from four STS subtypes (LMS, SS, UPS and DDLPS) and accompanying annotation of baseline clinicopathological variables were identified and retrieved through retrospective review of departmental database and medical notes at a single specialist cancer centre. In line with standard management approaches, primary LMS, UPS and DDLPS tumours were naïve to any pre-operative therapy, while pre-operative exposure to chemotherapy and/or radiotherapy was varied in SS. The histological diagnosis was confirmed in all cases by experienced soft tissue pathologists (KT, CF). For each tumour, a single FFPE tissue block containing representative viable tumour was selected through review of haematoxylin and eosin (H&E)-stained sections. 20 μm sections were cut from each selected tumour block and,

where indicated, macrodissected to enrich to >75% viable tumour content. In liposarcoma tumours that contained both well-differentiated and de-differentiated components, slide review and macrodissection ensured the dedifferentiated areas were sampled.

2.2. Protein extraction and sample preparation

20 μm tissue sections from each sample were deparaffinised by three washing steps in xylene, rehydrated by washes with decreasing ethanol gradient (100%, 96%, 70%) and then dried in a SpeedVac concentrator (Thermo Scientific). The lysis buffer (0.1 M Tris-HCl pH 8.8, 0.50% (w/v) sodium deoxycholate, 0.35% (w/v) sodium lauryl sulphate) was added at a ratio of 200 μl /mg of dry tissue. The sample was homogenised using a LabGen700 blender (ColeParmer) with $3 \times 30\text{s}$ pulses and sonicated on ice for 10 min, then heated at 95 $^{\circ}\text{C}$ for 1 h to reverse formalin crosslinks. Lysis was carried out by shaking at 750 rpm at 80 $^{\circ}\text{C}$ for 2 h. The sample was then centrifuged for 15 min at 4 $^{\circ}\text{C}$ at 14,000 rpm and the supernatant collected. Protein concentration in the homogenate was measured by bicinchoninic acid (BCA) assay (Pierce) The extracted protein sample was digested using the Filter-Aided Sample Preparation (FASP) protocol as previously described in [11]. Briefly, each sample was placed into an Amicon-Ultra 4 (Merck) centrifugal filter unit and detergents were removed by several washes with 8 M urea. The concentrated sample was then transferred to Amicon-Ultra 0.5 (Merck) filters to be reduced with 10 mM dithiothreitol (DTT) and alkylated with 55 mM iodoacetamide (IAA). The sample was washed with 100 mM ammonium bicarbonate (ABC) and digested with trypsin overnight (Promega, trypsin to starting protein ratio 1:100 μg). Peptides were collected by two successive centrifugations with 100 mM ABC and desalted on C18 SepPak columns (Waters). The desalted peptide samples were then dried in a SpeedVac concentrator and stored at -80°C .

2.3. SWATH-MS data acquisition and processing

Samples were resuspended in a buffer of 2% ACN/ 0.1% formic acid, spiked with iRT calibration mix (Biognosys AG) and analysed on an Agilent 1260 HPLC system (Agilent Technologies) coupled to a TripleTOF 5600+ mass spectrometer with NanoSource III (AB SCIEX). 1 μg of peptides for each sample was loaded onto a ZORBAX C18 (Agilent Technologies) trap column and separated on a 75 $\mu\text{m} \times 15\text{ cm}$ long analytical column with an integrated manually pulled tip packed with Reprosil Pur C18AQ beads (3 μm , 120 \AA particles, Dr. Maisch). A linear gradient of 2–40% of Buffer B (98% ACN, 0.1% FA) in 120 min and a flow rate of 250 nl/min was used. Each sample was analysed in 2 technical replicates. Full profile MS scans were acquired in the mass range of m/z 340–1400 in positive ion mode. 8 data points per elution peak were set up for calculation of 60 precursor isolation windows with a fixed size of 13 Da across the mass range of m/z 380–1100 with 1 Da overlap. MS/MS scans were acquired in the mass range of m/z 100–1500. Maximum filling time for MS scans was 250 ms and for MS/MS scans 100 ms, resulting in a cycle time of 3.1 s. SWATH spectra were analysed using Spectronaut 11 (Biognosys AG) against a published human library [12]. FDR was restricted to 1% and only the top 6 peptides were used for quantification of a protein. Peak area of 2 to 6 fragment ions was used for peptide quantification, and the mean value of the peptides was used to quantify proteins. 2 peptides were set as minimum requirement for inclusion of a protein in the analysis.

2.4. Data processing and statistical methods

Data were log₂ transformed, quantile normalised at sample level, followed by feature level (protein) centering across the samples to remove technical bias such as batch effect. Hierarchical clustering and Principle Component Analyses were used as visualisation tools to assess presence of batch effects. In brief, quantile normalisation was applied to transform the statistical distributions across samples to be the same,

based on the assumption that the statistical distribution of each sample should be the same (or have the same distributional shape) within biological groups or conditions, but allowing that they may differ between groups. This procedure was performed using *proBatch* package in R [13]. Feature level (median centering) was performed in order to remove further technical bias. The proteomics data was visualised using 3D-t-Distributed Stochastic Neighbour Embedding. To assess similarity of protein expression profiles, pairwise Pearson correlations between all samples were visualised using the 'ComplexHeatmap' R package [14]. Dendrograms were created using complete linkage clustering and the Euclidean distances between the resulting correlations. Differential expression analysis was performed using Significant Analysis of Microarray upon a false discovery rate (FDR) less than 0.1% [15]. Gene Set Enrichment Analysis (GSEA) version 4.0.1 was used to identify gene sets from the Hallmark database (v7.0, 'c5.bp.v7.0.symbols.gmt') that were significantly enriched between each subtype and rest of the groups respectively. Protein-Protein Interaction (PPI) network analysis and visualisation was performed using NetworkAnalyst 3.0 pipeline against the STRING interactome database, with settings to confidence score > 900 and experimental evidence required [16–18]. The level 4 (log2 transformed with loading and batch corrected) RPPA dataset from the TCGA-SARC study was downloaded from The Cancer Proteome Atlas portal and clinical data downloaded from the TCGA Pan-cancer Clinical Data Resource (TCGA-CDR) within the NCI Genomic Data Commons [19,20]. The RPPA dataset was feature level (protein) median centred across samples and plotted along with the SWATH data using box-and-whisker plot. Univariable survival analysis of overall survival was estimated by Kaplan-Meier curve and Log-rank test for significance. Event was defined as death. Time was defined as surgery to death or last follow-up. Association between molecular groups and prognostic factors (grade, tumour size and sex) were tested by Fisher's exact test while association between molecular groups with age was tested by Kruskal-Wallis test. Univariable and multivariable Cox proportional hazards analyses were used to assess the prognostic significance of the different variables. Assessments of the proportional hazard were performed to check that the proportional hazard assumption is valid. Overrepresentation analysis was performed on the 133 identified proteins using PANTHER v14 [21].

2.5. Immunohistochemical staining and scoring

For validation of SWATH-MS results, the same cohort of specimens was stained for vinculin (ab219649, Abcam, 1:1000 dilution) and decorin (ab151988, Abcam, 1:500) in tissue microarray (TMA) format. To generate the TMAs, two to four 1 mm diameter cores were sampled from areas of viable tumour within FFPE donor blocks, re-embedded in an arrayed recipient paraffin block and sectioned at 4.0 µm thickness. Sections were deparaffinised in xylene and rehydrated using decreasing concentrations of ethanol in water (once in 100%, 96% and 80%). Antigen retrieval was performed in Tris-EDTA buffer (pH 6.0) for 8 min in a microwave oven and cooled for 45 mins at room temperature. Sections were then washed once in Tris-buffered saline buffer (TBS), twice in Tris-buffered saline-Tween buffer (TBST) and incubated with blocking buffer for 90 min in a humidity chamber at room temperature. The blocking buffer is 3% (m/v) Bovine Serum Albumin (Sigma-Aldrich) in TBST. After blocking, sections were incubated with the primary antibody in a humidity chamber at 4 °C overnight.

The next day, slides were washed once in TBS and twice in TBST. Endogenous peroxidase activity was blocked with DAKO Peroxidase blocking solution (DAKO, Agilent Technologies) for 60 min at room temperature and sections were then washed once in TBS and twice in TBST. The sections were incubated with the goat anti-rabbit secondary antibody (7074S, Cell Signalling, 1:100) in the humidity chamber for 60 min at room temperature. The slides were washed again once in TBS and twice in TBST and then incubated with diaminobenzidine (DAKO, Agilent Technologies) for 25 s and 40s for vinculin and decorin staining,

respectively. The sections were rinsed in water, counterstained with Modified Mayer's haematoxylin (Abcam) and dehydrated by taking them through washes in graded ethanol (once in 80%, 96% and 100%) and xylene. Finally, the slides were mounted in Pertex mounting medium (Pioneer) and scanned on Hamamatsu Nanozoomer slide scanner.

Scanned TMA cores were independently evaluated on a semi-quantitative basis by three investigators. Staining intensity was scored on the scale from 0 to 3 where 0: no staining; 1: weak staining; 2: moderate staining; 3: strong staining. Staining representation in the tumour area was scored on the scale from 0 to 3 where 0: <1% of tumour cells stained; 1: 1–10% of tumour cells stained; 2: 10–50% tumour cells stained; 3: >50% of tumour cells stained. Total score for each core was calculated as a sum of intensity and representation scores. For each case, median value of total scores was calculated across all cores scored by all three investigators and cases were further classified by a three tier system where "no or low staining": total score median of 0–1; "weak staining": total score median of 2–4; "strong staining": total score median of 5–6. Borderline median scores (e.g. 1.5) were resolved by reviewing the IHC stained TMA cores and reaching a consensus classification between the three investigators.

2.6. Availability of data and materials

The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [22] partner repository with the dataset identifier PXD019719.

3. Results

3.1. Patient characteristics

The cohort is comprised of FFPE tumour material from 36 patients treated at The Royal Marsden Hospital. These specimens were obtained from surgical resections of primary tumours from four of the more common STS subtypes: leiomyosarcoma (LMS) ($n = 12$), synovial sarcoma (SS) ($n = 7$), UPS ($n = 10$) and dedifferentiated liposarcoma (DDLPS) ($n = 7$). Baseline clinico-pathological characteristics of these patients are summarised in Table 1.

3.2. Quantitative analysis of the STS proteome

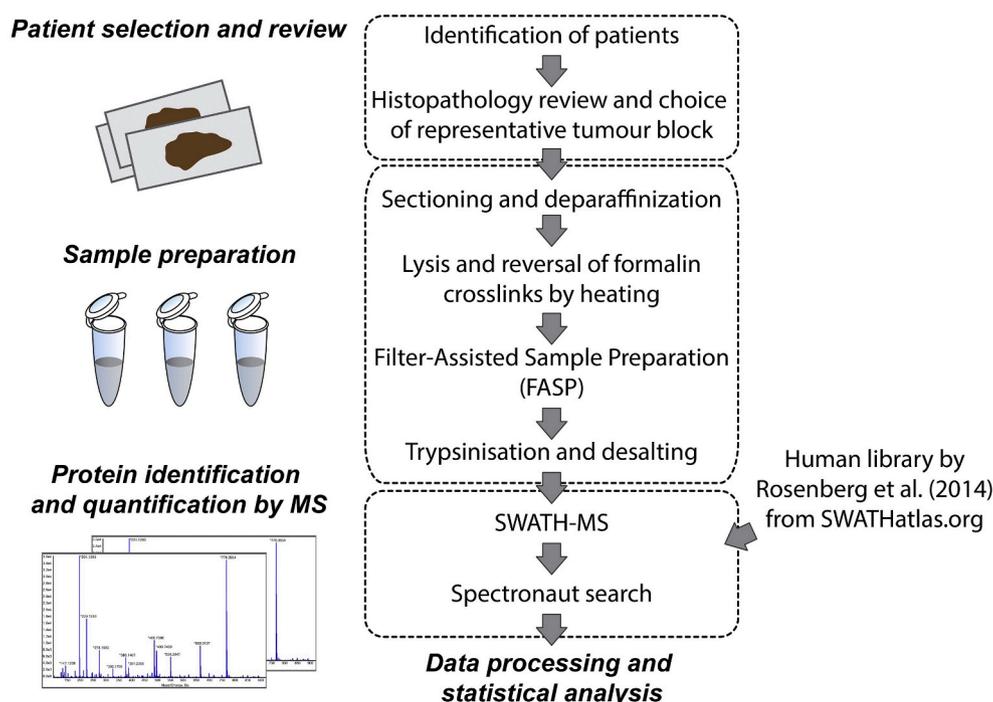
STS specimens were subjected to protein extraction and SWATH-MS analysis in technical duplicates as outlined in Fig. 1. This analysis led to the identification and quantification of 2951 proteins across all samples (Fig. 2A and Table S1). Comparisons of quantitative data between technical duplicates in each sample resulted in a median Pearson's correlation coefficient of 0.994 (range 0.968–0.998), demonstrating exceptional reproducibility of the SWATH-MS methodology. Applying t-Distributed Stochastic Neighbour Embedding (t-SNE) on the full dataset, four major groups can be visualised on the 3D-tSNE plot (Fig. 2B), which correspond to the four distinct histological subtypes. These results demonstrate that SWATH-MS profiling can reveal intrinsic biology associated with distinct histological subtypes which are characterised by unique proteomic signatures.

3.3. Defining biological processes that are enriched in STS histological subtypes

To gain an understanding of the underlying gene set annotations in each histological subtype, we undertook gene set enrichment analysis (GSEA) of the full proteomic dataset [23]. The top 20 ranked enriched gene sets are shown in Fig. 3. Gene sets significantly enriched in LMS versus the rest of the cohort comprise of those involved in muscle development and contraction. This finding is consistent with the smooth muscle lineage of this histological subtype which has also been reported in published gene expression studies of LMS [2,24]. The utility of

Table 1
Clinico-pathological characteristics of STS cohort.

		Overall	LMS	SS	UPS	DDLPS
Number of cases		36	12	7	10	7
Age		62 (28–84)	67 (35–75)	49 (43–77)	76 (28–84)	61 (37–78)
Gender	Female	25	12	4	5	4
	Male	11	0	3	5	3
Anatomical site	Intracavity	14	5	2	2	5
	Limbs	12	3	4	4	1
	Trunk/head	6	0	1	4	1
	Uterus	4	4	0	0	0
Disease stage	Localised	36	12	7	10	7
	Local recurrent	0	0	0	0	0
Histological grade	1	0	0	0	0	0
	2	16	6	4	4	2
	3	20	6	3	6	5
Pre-op treatment	No treatment	32	12	3	10	7
	Chemotherapy	1	0	1	0	0
	Chemo & Radiotherapy	3	0	3	0	0

**Fig. 1.** Schematic of the experimental workflow highlighting the key steps that were undertaken for sample selection and preparation as well as SWATH-MS data acquisition and analysis.

SWATH-MS data to rediscover known molecular processes highlights the validity of this approach for pathway discovery. In the case of SS, GSEA identifies RNA splicing and processing to be key gene sets significantly enriched in this histological subtype. This result is in line with a previous study which showed that the SS18-SSX fusion in SS binds to the ribonucleoprotein SYT-interacting protein/co-activator activator (SIP/CoAA) which is a key modulator of RNA splicing [25]. Meanwhile, gene sets significantly enriched in UPS finds a number of biological processes linked to leukocyte activation and metal ion transport. Notably, Phase II STS clinical trials of immune checkpoint inhibitors have reported evidence of clinical activity in UPS [26,27], underscoring the value of SWATH-MS in identifying biologically meaningful gene sets with potential clinical utility. No ontologies were found to be significantly enriched in DDLPS.

3.4. Identification of key protein complexes and signalling networks operating in STS histotypes

Using the multiclass Significance Analysis of Microarray (SAM) method, 277 proteins (FDR <0.1%) were found to be significantly differentially expressed across these four histotypes (Fig. S1, Table S2 and S3). Of these proteins, 95 were found to be significantly upregulated only in LMS, and 23 seed proteins were significantly mapped to directly interact with each other to form a zero-order network according to the STRING interactome database (Fig. 4A). These 23 seed proteins formed a subnetwork including the core machinery required for the regulation of smooth muscle contractile activity such as the myosin light chains (MYL1, MYL6, MYL9, MYL12B), myosin heavy chains (MYH10, MYH11), tropomyosin alpha chains (TPM1, TPM4), Myosin Light Chain Kinase (MYLK) and Protein Phosphatase 1 Regulatory Subunit 12A or Myosin Phosphatase Target Subunit 1 (PPP1R112A) (Fig. 4A). Evaluation of the protein expression levels of the myosin heavy chain MYH11

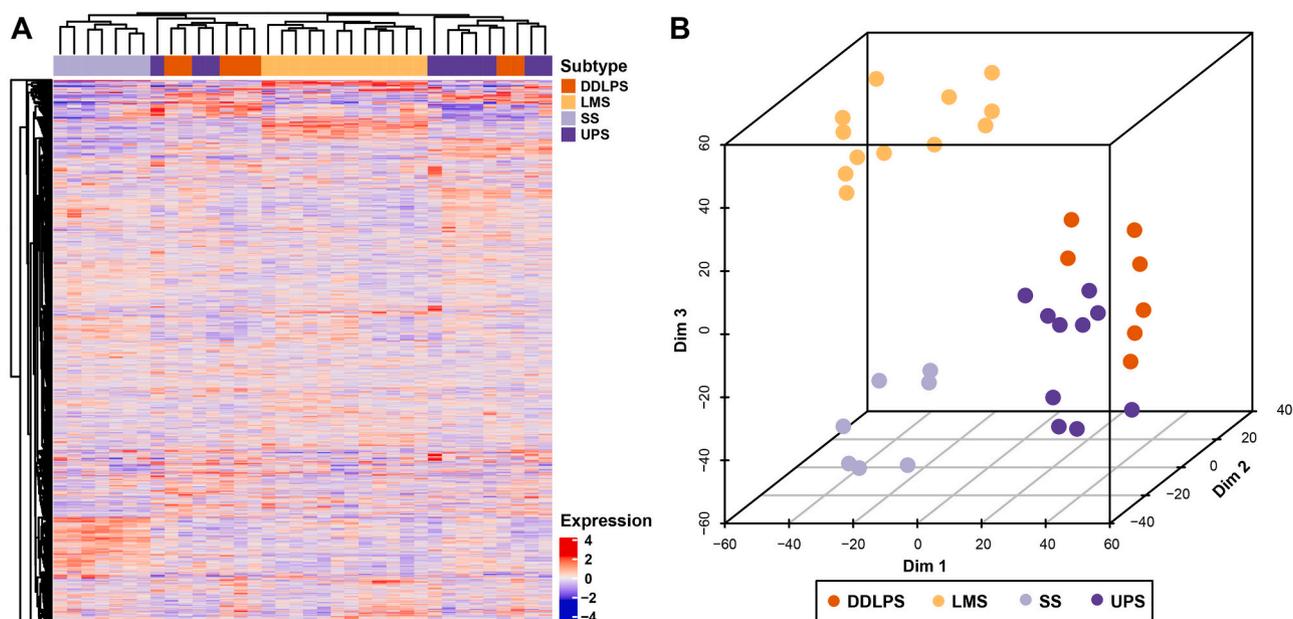


Fig. 2. (A) Hierarchical clustering of 2951 proteins across 36 STS cases. The full list of proteins is listed in Table S1. (B) 3D-tSNE plot depicts four distinct groups of STS cases corresponding to the distinct histological subtypes. LMS is leiomyosarcoma, SS is synovial sarcoma, UPS is undifferentiated pleomorphic sarcoma and DDLPS is dedifferentiated liposarcoma.

in the RPPA dataset from the independent TCGA sarcoma cohort (DDLPS $n = 54$, LMS $n = 81$, SS $n = 6$, UPS $n = 49$) confirms that this protein is significantly upregulated in LMS, providing independent validation of our SWATH-MS results (Fig. S2, Table S4) [3]. A second protein sub-network comprising extracellular matrix components (matrisome) and their cognate adhesion receptors and downstream regulatory proteins (adhesome) was also found to be upregulated in LMS. This network includes the integrin receptors (ITGA5, ITGB1, ITGB5), intracellular adhesion signalling proteins (VCL, FLNB, ILK, TNS1) and extracellular matrisome components (HSPB1, HSPG1, NID1, LAMB2, TNC, TGFBI1) [28,29] (Fig. 4A).

In SS, 103 proteins were found to be uniquely significantly upregulated; 28 of which were identified as seed proteins that significantly mapped to directly interact with each other to form a zero-order network. Twenty-four of these proteins are key components of mRNA splicing regulation. Cross-referencing these proteins to the SpliceosomeDB of spliceosome components [30] showed that the proteins enriched in SS comprise of the SR protein class (SRSF1, SRSF3, SRSF7, SRSF9, SRSF11) involved in constitutive and alternative pre-mRNA splicing, the heterogeneous ribonucleoprotein particle (hnRNPs) class and proteins involved in the formation of the spliceosome A Complex (HNRNPA1, U2AF2, SNRNP70) (Fig. 4B). Twenty-nine proteins were found to be upregulated only in UPS, and 3 seed proteins directly interact with each other, all components of the MHC class 1 complex (HLA-A, HLA-B and B2M). In DDLPS, 13 proteins were uniquely upregulated with no zero-order network identified. Notably, CDK4 is one of these 13 upregulated proteins which is in agreement with the molecular pathology of this disease where CDK4 is amplified in ~90% of DDLPS; and CDK4/6 inhibitors (palbociclib, ribociclib and abemaciclib) are currently being evaluated in the treatment of this histological subtype [31–33].

3.5. Identification of proteomic profiles that are associated with STS patient outcome

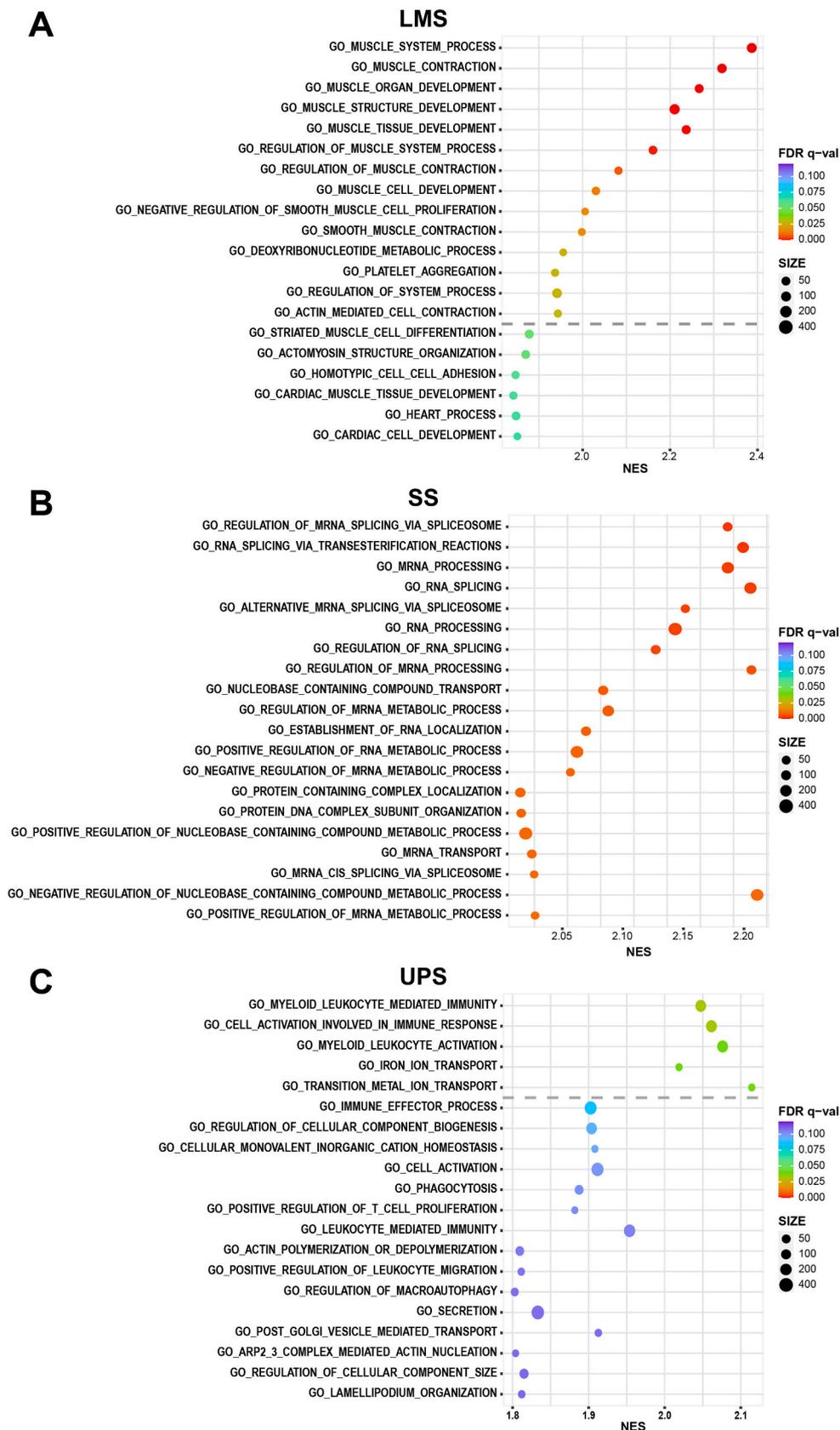
In order to evaluate if there are proteins within our dataset that are associated with overall survival (OS) in our cohort of patients, univariable Cox regression analysis was performed for each of the 2951

proteins as exploratory analyses, and we selected a total of 133 proteins with $p < 0.05$ (Fig. 5A and Table S5). Subsequently, by hierarchical clustering of the 36 cases based on these 133 protein expression values, three subgroups of mixed histological subtypes were identified (Fig. 5B). These three groups were associated with significantly differential OS, with Group 2 comprising cases demonstrating the worst survival estimate (Fig. 5C, Log-rank $p < 0.00001$). There were no statistically significant associations between the molecular subgroups with prognostic factors grade, tumour size, sex and age (Table S6). However, there was a significant association between the molecular subgroup and sarcoma histological subtype (Table S6). Adjusting for other clinicopathological factors including age, tumour size, grade, sex and histological subtypes, the molecular subgroups remained an independent prognostic factor (Table S7) in the multivariable Cox regression analysis. After adjustment for other prognostic factors, patients in Group 2 were 72.6 times more likely to die (Hazard ratio 72.6; 95% Confidence Interval 3.71–1432; $p = 0.005$) compared to patients in Group 1 and 20.6 times more likely to die (Hazard ratio 20.6; 95% Confidence Interval 3.08–137; $p = 0.002$) compared to patients in Group 3. It should be noted that Group 2 is not a single cluster but is comprised of multiple smaller clusters. No statistically significant biological processes were identified to be enriched in the 133 proteins based on over-representation analysis [21].

3.6. Validation of vinculin and decorin protein expression by immunohistochemistry

To validate the SWATH-MS data using an orthogonal approach, we performed immunohistochemical (IHC) analysis of tumour cell expression levels of two proteins, vinculin (VCL) and decorin (DCN), in our cohort (Fig. 6). VCL is an intracellular adhesome signalling protein which was shown by SWATH-MS to be significantly upregulated in LMS compared to the other subtypes (Fig. 6A). Consistent with the SWATH-MS data, IHC analysis showed that half of all LMS cases (6/12) displayed strong tumour cell staining of this protein with the remaining cases either having weak (3/12) or no/low staining (3/12). This is in contrast to the other three subtypes where only weak or no/low staining was observed with no cases showing strong tumour cell expression of this protein. Decorin (DCN) is a proteoglycan that was found to be

Fig. 3. Plot of Gene Set Enrichment Analysis results showing the top ranked 20 positively enriched gene sets for (A) leiomyosarcoma (LMS), (B) synovial sarcoma (SS) and (C) undifferentiated pleomorphic sarcoma (UPS). The dashed line indicates a False Discovery Rate (FDR) = 0.05 threshold. The colour of the circles represents the FDR q-value while the size of the circle indicates the number of proteins within the dataset that is in each gene set. NES is the normalised enrichment score from the GSEA analysis.



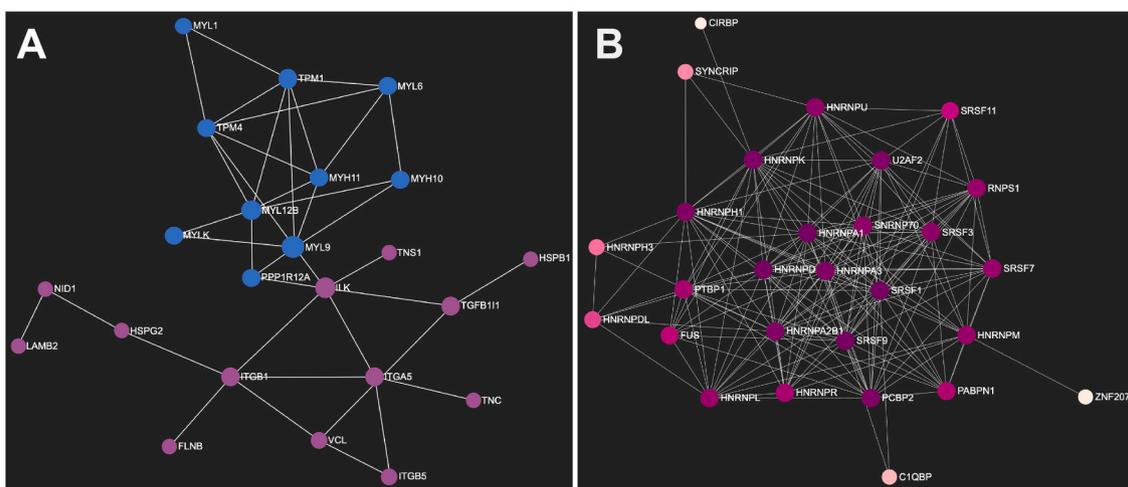


Fig. 4. Network diagrams in Force Atlas layouts depicting protein-protein interaction maps for proteins which are significantly upregulated in (A) leiomyosarcoma and (B) synovial sarcoma. In (A), proteins in blue are components involved in the regulation of smooth muscle contraction while those in purple are components of the matrisome and adhesion. In (B), the majority of significantly upregulated proteins in synovial sarcoma are involved in RNA splicing regulation and the different colours indicate the number of interactions between identified proteins. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

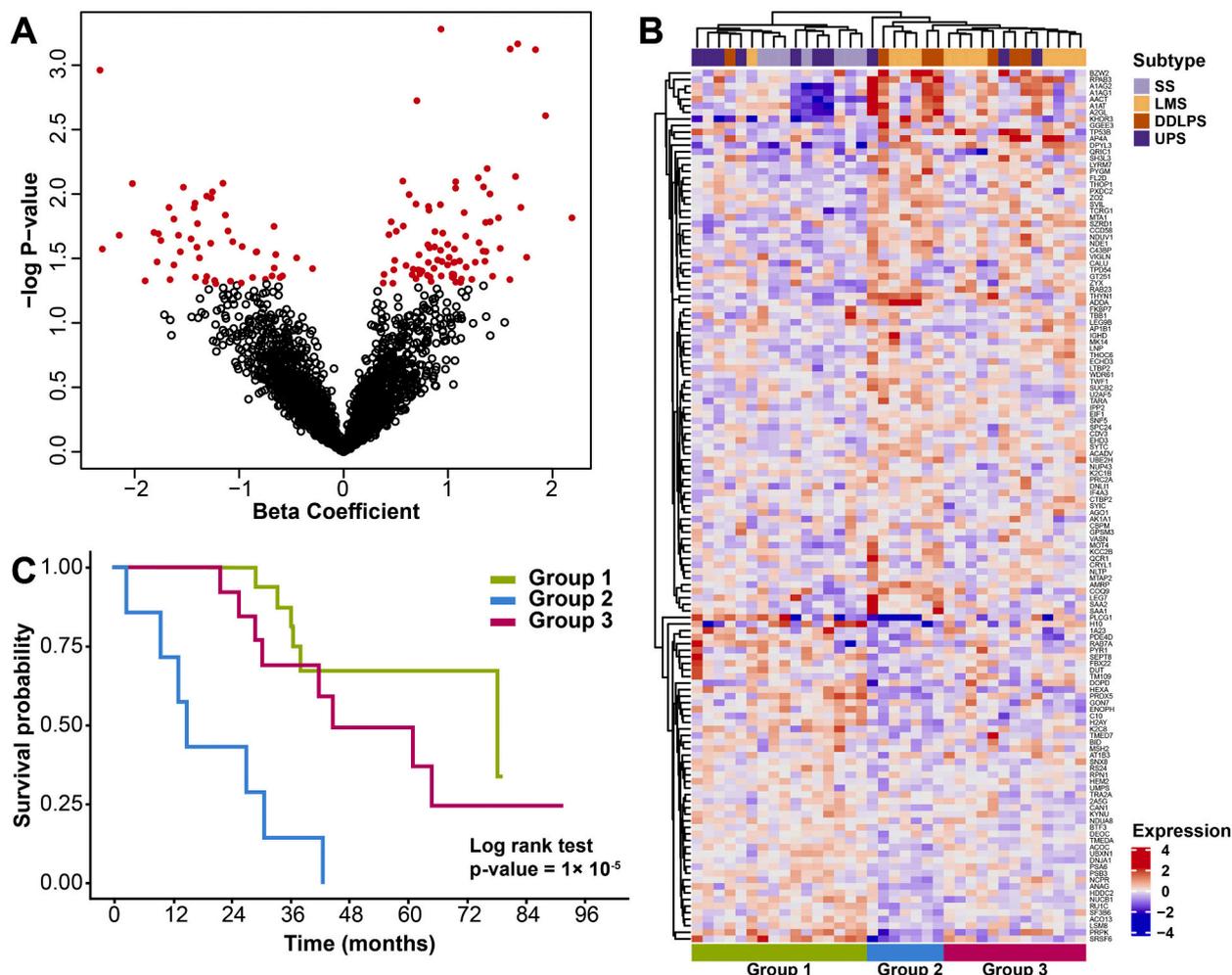


Fig. 5. (A) Volcano plot of the beta coefficients from univariable Cox regression analysis for each of the 2951 proteins in the proteomic dataset and their associated $-\log p$ -value. Red circles indicate 133 proteins with $p < 0.05$. The full list of proteins is listed in Table S3. (B) Hierarchical clustering of 133 proteins across 36 STS cases identifies 3 subgroups of mixed histological subtypes. (C) The Kaplan Meier curves for overall survival (OS) of the 3 subgroups identified in (B). LMS is leiomyosarcoma, SS is synovial sarcoma, UPS is undifferentiated pleomorphic sarcoma andDDLPS is dedifferentiated liposarcoma. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

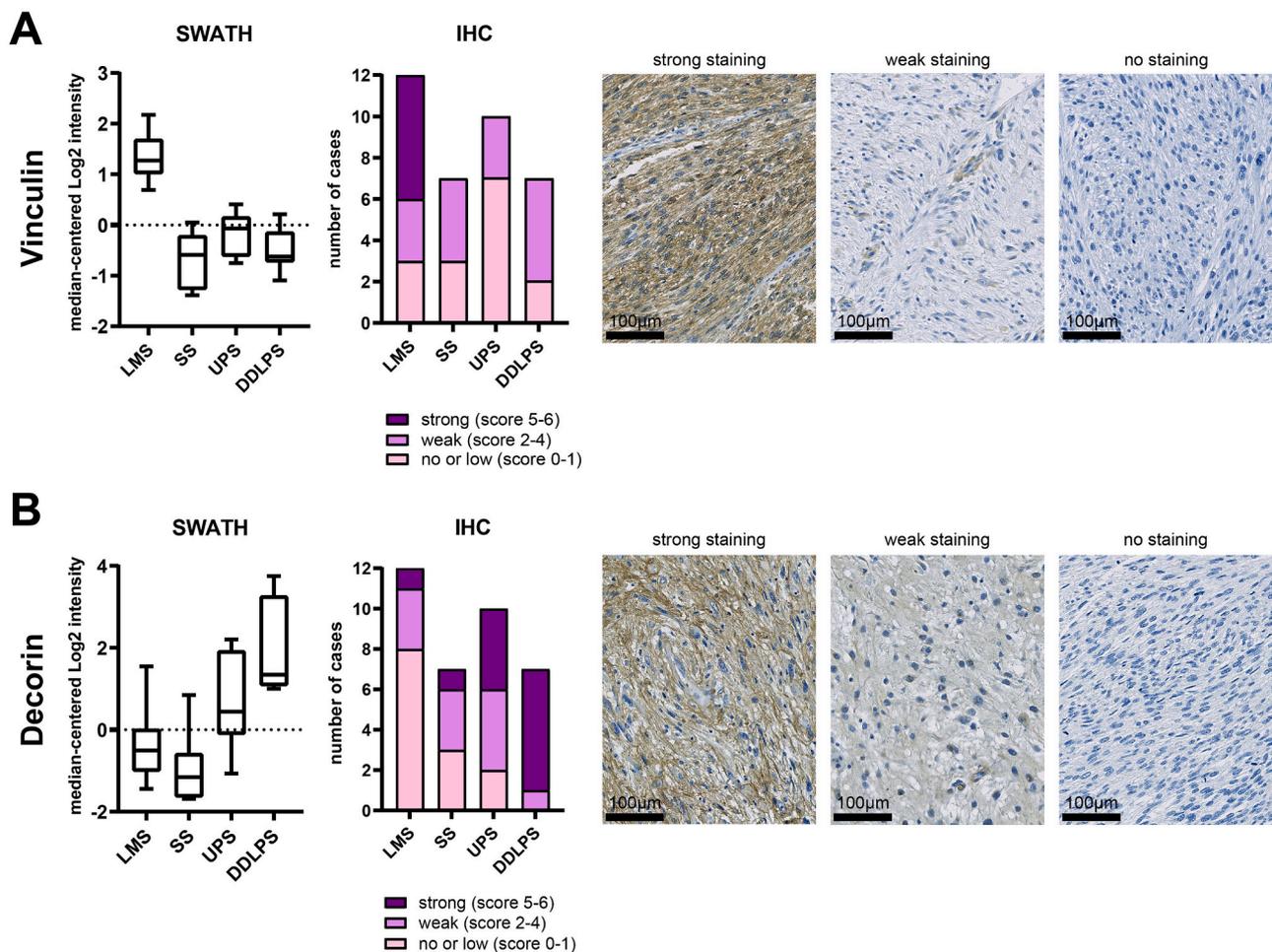


Fig. 6. Comparison of expression profile of (A) vinculin and (B) decorin analysed by SWATH-MS (SWATH) with immunohistochemical (IHC) staining of TMA cores generated from the same tissue specimens in the cohort. Boxplots for SWATH-MS data shows 1st quartile, 3rd quartile and median value for each subtype, whiskers indicate interquartile range. Stacked bar charts indicate immunoreactivity of cases in each of the sarcoma subtypes. Photomicrographs of representative TMA cores with strong, weak and no staining for both proteins are shown.

significantly upregulated in DDLPS compared to the other three histological subtypes (Fig. 6B). IHC analysis showed that 85% (6/7) of the DDLPS cases had strong protein expression levels of DCN with the remaining case having weak staining. In contrast, only a small proportion of cases in LMS (1/12, 8%), UPS (4/10, 40%) and SS (1/7, 14%) had strong tumour cell staining of this protein. Taken together, the IHC results of these two proteins are consistent with the SWATH-MS data which provides independent validation of the proteomic profiling dataset.

4. Discussion

In this study we have undertaken a comprehensive proteomic analysis of multiple STS subtypes and have demonstrated the utility of SWATH-MS in the following applications: 1. Defining unique proteomics signatures associated with distinct histological subtypes, 2. Identifying, within histological subtypes, biological processes and key protein networks and 3. Defining candidate proteins which are associated with predicting patient outcomes. Previous studies have shown that transcriptomic profiling can aid in the molecular classification of distinct sarcoma subtypes [34,35]. More recently, DNA methylation profiling has similarly revealed subtype-specific signatures that have utility in molecular classification, particularly in diagnostically challenging histological subtypes [36–38]. Our study shows for the first time that different histological subtypes harbour distinct proteomic signatures

indicative of inherent subtype-specific biology driven by specific protein networks. We anticipate that expanded proteomic analysis incorporating larger STS cohorts with an increased number of histological subtypes will enable further refinement of this approach and the development of robust proteomic classifiers to aid in accurate molecular diagnosis of sarcomas.

In our study, we have identified a number of histotype-specific gene sets that are consistent with prior transcriptomic analyses. In LMS, ontology analysis of our proteomic data identifies biological processes associated with muscle development and contraction. Furthermore, analysis of protein-protein interactions identified a subnetwork comprising the core machinery involved in the regulation of smooth muscle contractile activity. This finding is in line with previous gene expression-based studies which have consistently demonstrated that smooth muscle functional gene ontologies are enriched in LMS cohorts from multiple independent transcriptomic datasets [4,24,39]. Interestingly, it has previously been shown that LMS can be classified into three distinct molecular subtypes based on transcriptomic data [24,39]. We anticipate that future studies integrating proteomics with transcriptomic analysis may shed light on the specific oncogenic pathways and candidate drug targets that are enriched in these LMS molecular subtypes which could be exploited for therapy. In UPS, our data finds that this subgroup is enriched in gene sets associated with leukocyte activation. This result is consistent with a very recent transcriptomic study which showed that UPS can be classified into two distinct molecular

subgroups, one of which is characterised by genes involved in inflammatory response and immune cell signatures [9]. The enrichment of an immune cell proteomic signature in UPS, where immune checkpoint inhibitors have showed some clinical activity [26,27], suggests that future inclusion of proteomic profiling in cancer immunotherapy trials may lead to new predictive biomarkers that complement currently available approaches [40].

Several transcriptomic studies have led to the development of prognostic signatures in sarcoma. These include gene expression panels based on chromosome instability, hypoxia and stromal signatures [5,41–43]. Here we have performed an exploratory analysis which has for the first time identified a panel of proteins that is capable of stratifying a subgroup of patients (Group 2) with worse outcomes when compared with the rest of the cohort. Notably, Group 2 is composed of patients with mixed histological subtypes (3 DDLPS, 3 LMS and 1 UPS), and has a distinct expression profile of 133 proteins. Interestingly, Group 2 shared a subgroup of proteins that were also overexpressed within Group 3 (comprising mainly of DDLPS and LMS), while having low expression in a subset of proteins compared to both Group 1 and 3. These data suggest that these patients may share molecular characteristics that transcend histotypes. Over-representation analysis did not identify any enriched biological processes in the 133 proteins. It is important to note that this analysis was performed on a small patient cohort treated within a single institution and our findings should be considered hypothesis generating.

Together with a recently published SWATH-MS analysis in prostate cancer and diffuse large B-cell lymphoma [44], our study demonstrates the versatility of this label-free MS strategy in the analysis of FFPE specimens, a tissue preservation technique which has historically posed technical challenges for conventional proteomic workflows and analyses [45]. By utilising archival material from a tissue bank which is typically an abundant tissue resource, our study describing the proteomic profiling of STS from FFPE specimens opens new opportunities for gaining new insights into the biology and molecular drivers in rare cancers without the need for logistically challenging prospective collection of fresh tissue.

5. Conclusions

We have employed SWATH MS to profile FFPE tissue specimens across four sarcoma subtypes and have identified histotype-specific proteomic profiles that describe key biological pathways as well as discovered a panel of candidate proteins associated with patient outcome in STS. We anticipate the future application of this strategy to additional STS subtypes and clinical trial cohorts has the potential to deliver new therapeutic targets and define predictive and prognostic signatures in these rare and difficult-to-treat diseases.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jprot.2021.104236>.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgements

This study is supported by grants from the National Institute for Health Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, The Royal Marsden Cancer Charity Sarcoma Research Fund, Breast Cancer Now, and The Pathological Society as well as a charitable donation from Geoff Crocker and Bristol Care Homes.

References

- [1] WHO, Classification of Tumours Editorial Board, *Soft Tissue and Bone Tumours*, 5th ed, World Health Organization, 2020.
- [2] A.H. Beck, R.B. West, M. van de Rijn, Gene expression profiling for the investigation of soft tissue sarcoma pathogenesis and the identification of diagnostic, prognostic, and predictive biomarkers, *Virchows Arch.* 456 (2) (2010) 141–151.
- [3] e.d.s.c. Cancer Genome Atlas Research Network. Electronic address, N, Cancer genome atlas research, comprehensive and integrated genomic characterization of adult soft tissue sarcomas, *Cell* 171 (4) (2017) 950–965, e28.
- [4] P. Chudasama, S.S. Mughal, M.A. Sanders, D. Hubschmann, I. Chung, K.I. Deeg, S. H. Wong, S. Rabe, M. Hlevnjak, M. Zapatka, A. Ernst, K. Kleinheinz, M. Schlesner, L. Sieverling, B. Klink, E. Schrock, R.M. Hoogenboezem, B. Kasper, C.E. Heilig, G. Egerer, S. Wolf, C. von Kalle, R. Eils, A. Stenzinger, W. Weichert, H. Glimm, S. Groschel, H.G. Kopp, G. Omlor, B. Lehner, S. Bauer, S. Schimmack, A. Ulrich, G. Mechtersheimer, K. Rippe, B. Brors, B. Hutter, M. Renner, P. Hohenberger, C. Scholl, S. Frohling, Integrative genomic and transcriptomic analysis of leiomyosarcoma, *Nat. Commun.* 9 (1) (2018) 144.
- [5] S. Le Guellec, T. Lesluyes, E. Sarot, C. Valle, T. Filleron, P. Rochaix, T. Valentin, G. Perot, J.M. Coindre, F. Chibon, Validation of the Complexity Index in SARComas prognostic signature on formalin-fixed, paraffin-embedded, soft-tissue sarcomas, *Ann. Oncol.* 29 (8) (2018) 1828–1835.
- [6] J. Burns, C.P. Wilding, R.L. Jones, P.H. Huang, Proteomic research in sarcomas - current status and future opportunities, *Semin. Cancer Biol.* 61 (2020) 56–70.
- [7] J. Noujaim, L.S. Payne, I. Judson, R.L. Jones, P.H. Huang, Phosphoproteomics in translational research: a sarcoma perspective, *Ann. Oncol.* 27 (5) (2016) 787–794.
- [8] Y. Liu, Z. Li, Z. Xu, X. Jin, Y. Gong, X. Xia, Y. Yao, Z. Xu, Y. Zhou, H. Xu, S. Li, Y. Peng, X. Wu, L. Dai, Proteomic maps of human gastrointestinal stromal tumor subgroups, *Mol. Cell. Proteomics* 18 (5) (2019) 923–935.
- [9] M. Toulmonde, C. Lucchesi, S. Verbeke, A. Crombe, J. Adam, D. Geneste, V. Chaire, A. Laroche-Clary, R. Perret, F. Bertucci, F. Bertolo, L. Bianchini, B. Dadone-Montaudie, T. Hembrough, S. Sweet, Y.J. Kim, F. Cecchi, F. Le Loarer, A. Italiano, High throughput profiling of undifferentiated pleomorphic sarcomas identifies two main subgroups with distinct immune profile, clinical outcome and sensitivity to targeted therapies, *EBioMedicine* 62 (2020) 103131.
- [10] L. Krasny, P.H. Huang, Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology, *Mol. Omics* 17 (1) (2021) 29–42.
- [11] J.R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nat. Methods* 6 (5) (2009) 359–362.
- [12] G. Rosenberger, C.C. Koh, T. Guo, H.L. Rost, P. Kouvonon, B.C. Collins, M. Heusel, Y. Liu, E. Caron, A. Vichalkovski, M. Faini, O.T. Schubert, P. Faridi, H.A. Eberhardt, M. Matondo, H. Lam, S.L. Bader, D.S. Campbell, E.W. Deutsch, R.L. Moritz, S. Tate, R. Aebersold, A repository of assays to quantify 10,000 human proteins by SWATH-MS, *Sci. Data* 1 (2014) 140031.
- [13] J. Cuklina, C.H. Lee, E.G. Williams, B.C. Collins, T. Sajic, P. Pedrioli, M. Rodriguez-Martinez, R. Aebersold, Computational Challenges in Biomarker Discovery from High-Throughput Proteomic Data, ETH Zurich, 2018.
- [14] Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data, *Bioinformatics* 32 (18) (2016) 2847–2849.
- [15] V.G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, *Proc. Natl. Acad. Sci. U. S. A.* 98 (9) (2001) 5116–5121.
- [16] G. Zhou, O. Soufan, J. Ewald, R.E.W. Hancock, N. Basu, J. Xia, NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis, *Nucleic Acids Res.* 47 (W1) (2019) W234–W241.
- [17] J. Xia, E.E. Gill, R.E. Hancock, NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data, *Nat. Protoc.* 10 (6) (2015) 823–844.
- [18] D. Szklarczyk, J.H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N.T. Doncheva, A. Roth, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, *Nucleic Acids Res.* 45 (D1) (2017) D362–D368.
- [19] J. Liu, T. Lichtenberg, K.A. Hoadley, L.M. Poisson, A.J. Lazar, A.D. Cherniack, A. J. Kovatich, C.C. Benz, D.A. Levine, A.V. Lee, L. Omberg, D.M. Wolf, C.D. Shriver, V. Thorsson, N. Cancer Genome Atlas Research, H. Hu, An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics, *Cell* 173 (2) (2018) 400–416, e11.
- [20] J. Li, Y. Lu, R. Akbani, Z. Ju, P.L. Roebuck, W. Liu, J.-Y. Yang, B.M. Broom, R.G. W. Verhaak, D.W. Kane, C. Wakefield, J.N. Weinstein, G.B. Mills, H. Liang, TCPA: a resource for cancer functional proteomics data, *Nat. Methods* 10 (11) (2013) 1046–1047.
- [21] H. Mi, A. Muruganujan, D. Ebert, X. Huang, P.D. Thomas, PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools, *Nucleic Acids Res.* 47 (D1) (2019) D419–D426.
- [22] 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, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaino, The PRIDE database and related tools and resources in 2019: improving support for quantification data, *Nucleic Acids Res.* 47 (D1) (2019) D442–D450.
- [23] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.* 102 (43) (2005) 15545–15550.
- [24] X. Guo, V.Y. Jo, A.M. Mills, S.X. Zhu, C.H. Lee, I. Espinosa, M.R. Nucci, S. Varma, E. Forgo, T. Hastie, S. Anderson, K. Ganjoo, A.H. Beck, R.B. West, C.D. Fletcher,

- M. van de Rijn, Clinically relevant molecular subtypes in leiomyosarcoma, *Clin. Cancer Res.* 21 (15) (2015) 3501–3511.
- [25] M. Perani, P. Antonson, R. Hamoudi, C.J. Ingram, C.S. Cooper, M.D. Garrett, G. H. Goodwin, The proto-oncoprotein SYT interacts with SYT-interacting protein/co-activator activator (SIP/CoAA), a human nuclear receptor co-activator with similarity to EWS and TLS/FUS family of proteins, *J. Biol. Chem.* 280 (52) (2005) 42863–42876.
- [26] S.P. D'Angelo, M.R. Mahoney, B.A. Van Tine, J. Atkins, M.M. Milhem, B. N. Jahagirdar, C.R. Antonescu, E. Horvath, W.D. Tap, G.K. Schwartz, H. Streicher, Nivolumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2 trials, *Lancet Oncol.* 19 (3) (2018) 416–426.
- [27] H.A. Tawbi, M. Burgess, V. Bolejack, B.A. Van Tine, S.M. Schuetz, J. Hu, S. D'Angelo, S. Attia, R.F. Riedel, D.A. Priebat, S. Movva, L.E. Davis, S.H. Okuno, D. R. Reed, J. Crowley, L.H. Butterfield, R. Salazar, J. Rodriguez-Canales, A.J. Lazar, I. I. Wistuba, L.H. Baker, R.G. Maki, D. Reinke, S. Patel, Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial, *Lancet Oncol.* 18 (11) (2017) 1493–1501.
- [28] S.E. Winograd-Katz, R. Fassler, B. Geiger, K.R. Legate, The integrin adhesome: from genes and proteins to human disease, *Nat. Rev. Mol. Cell Biol.* 15 (4) (2014) 273–288.
- [29] A. Naba, K.R. Clauser, S. Hoersch, H. Liu, S.A. Carr, R.O. Hynes, The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices, *Mol. Cell. Proteomics* 11 (4) (2012) M111–014647.
- [30] I. Cvitkovic, M.S. Jurica, Spliceosome database: a tool for tracking components of the spliceosome, *Nucleic Acids Res.* 41 (Database issue) (2013) D132–D141.
- [31] K. Thway, R. Flora, C. Shah, D. Olmos, C. Fisher, Diagnostic utility of p16, CDK4, and MDM2 as immunohistochemical panel in distinguishing well-differentiated and dedifferentiated liposarcomas from other adipocytic tumors, *Am. J. Surg. Pathol.* 36 (3) (2012) 462–469.
- [32] A.T.J. Lee, K. Thway, P.H. Huang, R.L. Jones, Clinical and molecular spectrum of liposarcoma, *J. Clin. Oncol.* 36 (2) (2018) 151–159.
- [33] M.A. Dickson, G.K. Schwartz, M.L. Keohan, S.P. D'Angelo, M.M. Gounder, P. Chi, C.R. Antonescu, J. Landa, L.X. Qin, A.M. Crago, S. Singer, A. Koff, W.D. Tap, Progression-free survival among patients with well-differentiated or dedifferentiated liposarcoma treated with CDK4 inhibitor palbociclib: a phase 2 clinical trial, *JAMA Oncol.* 2 (7) (2016) 937–940.
- [34] N.H. Segal, P. Pavlidis, C.R. Antonescu, R.G. Maki, W.S. Noble, D. DeSantis, J. M. Woodruff, J.J. Lewis, M.F. Brennan, A.N. Houghton, C. Cordon-Cardo, Classification and subtype prediction of adult soft tissue sarcoma by functional genomics, *Am. J. Pathol.* 163 (2) (2003) 691–700.
- [35] S. Watson, V. Perrin, D. Guillemot, S. Reynaud, J.M. Coindre, M. Karanian, J. M. Guinebretiere, P. Freneaux, F. Le Loarer, M. Bouvet, L. Galmiche-Rolland, F. Larousserie, E. Longchamp, D. Ranchere-Vince, G. Pierron, O. Delattre, F. Tirede, Transcriptomic definition of molecular subgroups of small round cell sarcomas, *J. Pathol.* 245 (1) (2018) 29–40.
- [36] C. Koelsche, D. Schrimpf, D. Stichel, M. Sill, F. Sahn, D.E. Reuss, M. Blattner, B. Worst, C.E. Heilig, K. Beck, P. Horak, S. Kreutzfeldt, E. Paff, S. Stark, P. Johann, F. Selt, J. Ecker, D. Sturm, K.W. Pajtlar, A. Reinhardt, A.K. Wefers, P. Sievers, A. Ebrahimi, A. Suwala, F. Fernandez-Klett, B. Casalini, A. Korshunov, V. Hovestadt, F.K.F. Kommos, M. Kriegsmann, M. Schick, M. Bewerunge-Hudler, T. Milde, O. Witt, A.E. Kulozik, M. Kool, L. Romero-Perez, T.G.P. Grunewald, T. Kirchner, W. Wick, M. Platten, A. Unterberg, M. Uhl, A. Abdollahi, J. Debus, B. Lehner, C. Thomas, M. Hasselblatt, W. Paulus, C. Hartmann, O. Staszewski, M. Prinz, J. Hench, S. Frank, Y.M.H. Versleijen-Jonkers, M.E. Weidema, T. Mentzel, K. Griewank, E. de Alava, J.D. Martin, M.A.I. Gastearena, K.T. Chang, S.Y.Y. Low, A. Cuevas-Bourdier, M. Mittelbronn, M. Mynarek, S. Rutkowski, U. Schuller, V. F. Mautner, J. Schittenhelm, J. Serrano, M. Snuderl, R. Buttner, T. Klingebiel, R. Buslei, M. Gessler, P. Wesseling, W.N.M. Dinjens, S. Brandner, Z. Jaunmuktane, I. Lyskjaer, P. Schirmacher, A. Stenzinger, B. Brors, H. Glimm, C. Heining, O. M. Tirado, M. Sainz-Jaspeado, J. Mora, J. Alonso, X.G. Del Muro, S. Moran, M. Esteller, J.K. Benhamida, M. Ladanyi, E. Wardelmann, C. Antonescu, A. Flanagan, U. Dirksen, P. Hohenberger, D. Baumhoer, W. Hartmann, C. Vokuhl, U. Flucke, I. Petersen, G. Mechttersheimer, D. Capper, D.T.W. Jones, S. Frohling, S. M. Pfister, A. von Deimling, Sarcoma classification by DNA methylation profiling, *Nat. Commun.* 12 (1) (2021) 498.
- [37] C. Koelsche, W. Hartmann, D. Schrimpf, D. Stichel, S. Jabar, A. Ranft, D.E. Reuss, F. Sahn, D.T.W. Jones, M. Bewerunge-Hudler, M. Trautmann, T. Klingebiel, C. Vokuhl, M. Gessler, E. Wardelmann, I. Petersen, D. Baumhoer, U. Flucke, C. Antonescu, M. Esteller, S. Frohling, M. Kool, S.M. Pfister, G. Mechttersheimer, U. Dirksen, A. von Deimling, Array-based DNA-methylation profiling in sarcomas with small blue round cell histology provides valuable diagnostic information, *Mod. Pathol.* 31 (8) (2018) 1246–1256.
- [38] S.P. Wu, B.T. Cooper, F. Bu, C.J. Bowman, J.K. Killian, J. Serrano, S. Wang, T. M. Jackson, D. Gorovets, N. Shukla, P.A. Meyers, D.J. Pisapia, R. Gorlick, M. Ladanyi, K. Thomas, M. Snuderl, M.A. Karajannis, DNA methylation-based classifier for accurate molecular diagnosis of bone sarcomas, *JCO Precis. Oncol.* 2017 (2017).
- [39] M.L. Hemming, C. Fan, C.P. Raut, G.D. Demetri, S.A. Armstrong, E. Sicinska, S. George, Oncogenic gene-expression programs in leiomyosarcoma and characterization of conventional, inflammatory, and uterogenic subtypes, *Mol. Cancer Res.* 18 (9) (2020) 1302–1314.
- [40] G.T. Gibney, L.M. Weiner, M.B. Atkins, Predictive biomarkers for checkpoint inhibitor-based immunotherapy, *Lancet Oncol.* 17 (12) (2016) e542–e551.
- [41] S.L. Volchenboum, J. Andrade, L. Huang, D.A. Barkauskas, M. Krailo, R.B. Womer, A. Ranft, J. Potratz, U. Dirksen, T.J. Triche, E.R. Lawlor, Gene expression profiling of Ewing sarcoma tumors reveals the prognostic importance of tumor-stromal interactions: a report from the children's oncology group, *J. Pathol. Clin. Res.* 1 (2) (2015) 83–94.
- [42] N. Aggerholm-Pedersen, B.S. Sorensen, J. Overgaard, K. Toustrup, S. Baerentzen, O.S. Nielsen, K. Maretty-Kongstad, M. Nordmark, J. Alsner, A. Safwat, A prognostic profile of hypoxia-induced genes for localised high-grade soft tissue sarcoma, *Br. J. Cancer* 115 (9) (2016) 1096–1104.
- [43] L. Yang, L. Forker, J.J. Irlam, N. Pillay, A. Choudhury, C.M.L. West, Validation of a hypoxia related gene signature in multiple soft tissue sarcoma cohorts, *Oncotarget* 9 (3) (2018) 3946–3955.
- [44] Y. Zhu, T. Weiss, Q. Zhang, R. Sun, B. Wang, X. Yi, Z. Wu, H. Gao, X. Cai, G. Ruan, T. Zhu, C. Xu, S. Lou, X. Yu, L. Gillet, P. Blattmann, K. Saba, C.D. Fankhauser, M. B. Schmid, D. Rutishauser, J. Ljubcic, A. Christiansen, C. Fritz, N.J. Rupp, C. Poyet, E. Rushing, M. Weller, P. Roth, E. Haralambieva, S. Hofer, C. Chen, W. Jochum, X. Gao, X. Teng, L. Chen, Q. Zhong, P.J. Wild, R. Aebersold, T. Guo, High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification, *Mol. Oncol.* 13 (11) (2019) 2305–2328.
- [45] M.B. O'Rourke, M.P. Padula, Analysis of formalin-fixed, paraffin-embedded (FFPE) tissue via proteomic techniques and misconceptions of antigen retrieval, *Biotechniques* 60 (5) (2016) 229–238.