Proteomic profiling of soft tissue sarcomas with SWATH mass spectrometry

Martina Milighetti1, Lukas Krasny1, Alex T.J. Lee1,2,3*, Frank McCarthy1, Gabriele Morani3,4, Cornelia Szecsei1, Yingtong Chen3,4, Cyril Fisher2, Ian Judson2, Khin Thway1,2, Maggie, C.U. Cheang3,4, Robin L. Jones2,3 and Paul H. Huang1.

1 Division of Molecular Pathology, The Institute of Cancer Research, London, UK.
2 Sarcoma Unit, The Royal Marsden NHS Foundation Trust, London, UK.
3 Division of Clinical Studies, The Institute of Cancer Research, London, UK.
4 Clinical Trials and Statistics Unit, The Institute of Cancer Research, London, UK.

*These authors contributed equally to this work.

Key words: Soft tissue sarcoma, Proteomics, Mass spectrometry, SWATH MS, FFPE, biomarkers.

Correspondence:

Dr Paul Huang
Division of Molecular Pathology
The Institute of Cancer Research
London SW3 6JB
United Kingdom
Tel: +44-207-153-5554
E-mail: paul.huang@icr.ac.uk
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 (independent of histological subtype) that have a distinct expression profile in a panel of 133 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.
Introduction

Soft tissue sarcomas (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 prognostic factors for these cancers (2-5). Whilst current diagnostic 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 a recent proteomic analysis of 13 gastrointestinal stromal tumours (GIST) which utilised mass spectrometry (MS) to identify three clinical subgroups (6, 8). To date no 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). 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.

Results

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), undifferentiated pleomorphic sarcoma (UPS) (n=10) and dedifferentiated liposarcoma (DDLPS) (n=7). Baseline clinico-pathological characteristics of these patients are summarised in Table 1.

Quantitative analysis of the STS proteome

STS specimens were subjected to protein extraction and SWATH-MS analysis in technical duplicates as outlined in Figure 1. This analysis led to the identification and quantification of 2951 proteins across all samples (Figure 2A and Table S1). Applying t-Distributed Stochastic Neighbour Embedding (t-SNE) on the full dataset, four major groups can be visualised on the 3D-tsNE plot (Figure 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.
Defining biological processes that are enriched in STS histological subtypes

To gain an understanding of the underlying biological pathways in each histological subtype, we undertook gene set enrichment analysis (GSEA) of the full proteomic dataset (9). The top 20 ranked enriched biological processes are shown in Figure 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, 10). The utility of 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 synovial sarcomas binds to the ribonucleoprotein SYT-interacting protein/co-activator activator (SIP/CoAA) which is a key modulator of RNA splicing (11). 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 (12, 13), underscoring the value of SWATH-MS in identifying biologically meaningful processes with potential clinical utility. No ontologies were found to be significantly enriched in DDLPS.

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 (Figure S1 and Table S2). Of these proteins, 103 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 (Figure 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) (Figure 4A). A second protein subnetwork 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) (14, 15) (Figure 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 (16) 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) (Figure 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 (17-19).

Identification of proteomic profiles that are predictive of 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 (Figure 5A and Table S3). Subsequently, by hierarchical clustering of the 36 cases based on these 133 protein expression values, three subgroups of mixed histological subtypes were identified (Figure 5B). These three groups were associated with significantly differential OS, with Group 2 comprising cases demonstrating the worst survival estimate (Figure 5C, Log-rank p<0.00001). Adjusting for other clinicopathological factors including age, tumour size, grade and histological subtypes, the molecular subgroups remained an independent prognostic factor. After adjustment for other prognostic factors, patients in Group 2 were 81 times more likely to die (Hazard ratio 81; 95% Confidence Interval 6-1085; p=0.0009) compared to patients in Group 1. No statistically significant biological pathways were identified to be enriched in the 133 proteins based on over-representation analysis (20).

Discussion

In this study we have undertaken the most comprehensive proteomic analysis of multiple STS subtypes to date 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 histotype-independent candidate biomarkers for predicting patient outcomes. Expanded proteomic analysis incorporating larger STS cohorts with an increased number of histological subtypes will enable further refinement of this approach as a means of gaining new insights into the biology and molecular drivers of this rare group of cancers. In particular, the enrichment of an immune response signature in UPS, where immune checkpoint inhibitors have showed some clinical activity (12, 13), suggests that future inclusion of proteomic profiling in cancer immunotherapy trials may lead to new predictive biomarkers that complement currently available approaches (21).
Several transcriptomic studies have led to the development of prognostic biomarkers in sarcoma. These include gene expression panels based on chromosome instability, hypoxia and stromal signatures (5, 22-24). 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 pathways 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. Further assessment of these 133 proteins in additional validation cohorts is required to confirm any prognostic association.

Together with a recently published SWATH-MS analysis in prostate cancer and diffuse large B-cell lymphoma (25), 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 (26). 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 such proteomic studies in rare cancers without the need for logistically challenging prospective collection of fresh tissue.

In conclusion, 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 new candidate prognostic biomarker protein panel 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 biomarkers in these rare and difficult-to-treat diseases.

Materials and Methods

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 PROSPECTUS 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. 20µm sections were cut from each selected tumour block and, where indicated, macrodissected to enrich to >75% viable tumour content. In LPS tumours that contained both well-differentiated and de-differentiated components, slide review and macrodissection ensured the dedifferentiated areas were sampled.

Protein extraction and sample preparation

20µm 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 (0.1M Tris-HCl pH
8.8, 0.50% (w/v) sodium deoxycholate, 0.35% (w/v) sodium lauryl sulphate) was added at a ratio of 200μl/mg of dry tissue. The sample was homogenised using a LabGen700 blender (ColeParmer) with 3x 30s pulses and sonicated on ice for 10 min, then heated at 95°C for 1 h to reverse formalin crosslinks. Lysis was carried out by shaking at 750rpm at 80°C for 2 h. The sample was then centrifuged for 15min at 4°C at 14,000rpm 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 (27). Briefly, each sample was placed into an Amicon-Ultra 4 (Merck) centrifugal filter unit and detergents were removed by several washes with 8M urea. The concentrated sample was then transferred to Amicon-Ultra 0.5 (Merck) filters to be reduced with 10mM dithiothreitol (DTT) and alkylated with 55mM iodoacetamide (IAA). The sample was washed with 100mM 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 100mM ABC and desalted on C18 SepPak columns (Waters). The desalted peptide samples were then dried in a SpeedVac concentrator and stored at -80°C.

**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 μm×15 cm long analytical column with an integrated manually pulled tip packed with Reprosil Pur C18AQ beads (3 μm, 120 Å 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 analyzed 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 (28). 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.

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. 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 visualized using the ‘ComplexHeatmap’ R package (29). 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% (30). 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 (31-33). 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. Univariable and multivariable Cox proportional hazards analyses were used to assess the prognostic significance of the different variables.
Over-representation analysis were performed on the 133 identified proteins using PANTHER v14 (20).

Availability of data and materials
The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (34) partner repository with the dataset identifier PXD019719

Competing interests
The authors declare that they have no competing interests.

Funding
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.

Acknowledgements
We thank Chris Wilding, Nafia Guljar and Jess Burns for assistance in the curation of clinical data and specimens.
References

1. WHO Classification of Tumours Editorial Board. Soft Tissue and Bone Tumours. 5th ed: World Health Organization; 2020. 368 p.

2. Beck AH, West RB, van de Rijn M. Gene expression profiling for the investigation of soft tissue sarcoma pathogenesis and the identification of diagnostic, prognostic, and predictive biomarkers. Virchows Arch. 2010;456(2):141-51.

3. Cancer Genome Atlas Research Network. Electronic address edsc, Cancer Genome Atlas Research N. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. Cell. 2017;171(4):950-65 e28.

4. Chudasama P, Mughal SS, Sanders MA, Hubschmann D, Chung I, Deeg KI, et al. Integrative genomic and transcriptomic analysis of leiomyosarcoma. Nat Commun. 2018;9(1):144.

5. Le Guellec S, Lesluyes T, Sarot E, Valle C, Filleron T, Rochaix P, et al. Validation of the Complexity INdex in SARComas prognostic signature on formalin-fixed, paraffin-embedded, soft-tissue sarcomas. Ann Oncol. 2018;29(8):1828-35.

6. Burns J, Wilding CP, Jones RL, Huang PH. Proteomic research in sarcomas - current status and future opportunities. Semin Cancer Biol. 2019.

7. Noujaim J, Payne LS, Judson I, Jones RL, Huang PH. Phosphoproteomics in translational research: a sarcoma perspective. Ann Oncol. 2016;27(5):787-94.

8. Liu Y, Li Z, Xu Z, Jin X, Gong Y, Xia X, et al. Proteomic Maps of Human Gastrointestinal Stromal Tumor Subgroups. Mol Cell Proteomics. 2019;18(5):923-35.

9. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.

10. Guo X, Jo VY, Mills AM, Zhu SX, Lee CH, Espinosa I, et al. Clinically Relevant Molecular Subtypes in Leiomyosarcoma. Clin Cancer Res. 2015;21(15):3501-11.

11. Perani M, Antonson P, Hamoudi R, Ingram CJ, Cooper CS, Garrett MD, et al. 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. 2005;280(52):42863-76.

12. D'Angelo SP, Mahoney MR, Van Tine BA, Atkins J, Milhem MM, Jahagirdar BN, et al. Nivolumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2 trials. Lancet Oncol. 2018;19(3):416-26.

13. Tawbi HA, Burgess M, Bolejack V, Van Tine BA, Schuetze SM, Hu J, et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a
multicentre, two-cohort, single-arm, open-label, phase 2 trial. Lancet Oncol. 2017;18(11):1493-501.

14. Winograd-Katz SE, Fassler R, Geiger B, Legate KR. The integrin adhesome: from genes and proteins to human disease. Nat Rev Mol Cell Biol. 2014;15(4):273-88.

15. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics. 2012;11(4):M111 014647.

16. Cvitkovic I, Jurica MS. Spliceosome database: a tool for tracking components of the spliceosome. Nucleic Acids Res. 2013;41(Database issue):D132-41.

17. Thway K, Flora R, Shah C, Olmos D, Fisher C. Diagnostic utility of p16, CDK4, and MDM2 as an immunohistochemical panel in distinguishing well-differentiated and dedifferentiated liposarcomas from other adipocytic tumors. Am J Surg Pathol. 2012;36(3):462-9.

18. Lee ATJ, Thway K, Huang PH, Jones RL. Clinical and Molecular Spectrum of Liposarcoma. J Clin Oncol. 2018;36(2):151-9.

19. Dickson MA, Schwartz GK, Keohan ML, D'Angelo SP, Gounder MM, Chi P, et al. Progression-Free Survival Among Patients With Well-Differentiated or Dedifferentiated Liposarcoma Treated With CDK4 Inhibitor Palbociclib: A Phase 2 Clinical Trial. JAMA Oncol. 2016;2(7):937-40.

20. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 2019;47(D1):D419-D26.

21. Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. Lancet Oncol. 2016;17(12):e542-e51.

22. Volchenboum SL, Andrade J, Huang L, Barkauskas DA, Kralio M, Womer RB, et al. 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. 2015;1(2):83-94.

23. Aggerholm-Pedersen N, Sorensen BS, Overgaard J, Toustrup K, Baerentzen S, Nielsen OS, et al. A prognostic profile of hypoxia-induced genes for localised high-grade soft tissue sarcoma. Br J Cancer. 2016;115(9):1096-104.

24. Yang L, Forker L, Irlam JJ, Pillay N, Choudhury A, West CML. Validation of a hypoxia related gene signature in multiple soft tissue sarcoma cohorts. Oncotarget. 2018;9(3):3946-55.

25. Zhu Y, Weiss T, Zhang Q, Sun R, Wang B, Yi X, et al. High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification. Mol Oncol. 2019;13(11):2305-28.

26. O'Rourke MB, Padula MP. Analysis of formalin-fixed, paraffin-embedded (FFPE) tissue via proteomic techniques and misconceptions of antigen retrieval. Biotechniques. 2016;60(5):229-38.
27. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods. 2009;6(5):359-62.

28. Rosenberger G, Koh CC, Guo T, Rost HL, Kouvonen P, Collins BC, et al. A repository of assays to quantify 10,000 human proteins by SWATH-MS. Sci Data. 2014;1:140031.

29. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847-9.

30. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001;98(9):5116-21.

31. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 2019;47(W1):W234-W41.

32. Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nat Protoc. 2015;10(6):823-44.

33. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017;45(D1):D362-D8.

34. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 2019;47(D1):D442-D50.
Figure Legends

Figure 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.

Figure 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.

Figure 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 represent 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.

Figure 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 adhesome. 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.
Figure 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 indicates 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 and DDLPS is dedifferentiated liposarcoma.

Figure S1. (A) Hierarchical clustering of 277 proteins that were found to be significantly differentially expressed across the four histotypes by multiclass Significance Analysis of Microarray (SAM) method (FDR <0.1%) across 36 STS cases. The full list of proteins is listed in Table S2. LMS is leiomyosarcoma, SS is synovial sarcoma, UPS is undifferentiated pleomorphic sarcoma and DDLPS is dedifferentiated liposarcoma.
Table 1: Clinico-pathological characteristics of STS cohort

|                          | Overall | LMS    | SS     | UPS    | DDLPS  |
|--------------------------|---------|--------|--------|--------|--------|
| **Number of cases**      | 36      | 12     | 7      | 10     | 7      |
| **Age**                  | 62      | 67     | 49     | 76     | 61     |
|                          | (28 - 84) | (35 - 75) | (37 - 77) | (28 - 84) | (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      |
A

LMS

GO_MUSCLE_SYSTEM_PROCESS
GO_MUSCLE_CONTRACTION
GO_MUSCLE_ORGAN_DEVELOPMENT
GO_MUSCLE_STRUCTURE_DEVELOPMENT
GO_MUSCLE_TISSUE_DEVELOPMENT
GO_REGULATION_OF_MUSCLE_SYSTEM_PROCESS
GO_REGULATION_OF_MUSCLE_CONTRACTION
GO_MUSCLE_CELL_DEVELOPMENT
GO_NEGATIVE_REGULATION_OF_SMOOTH_MUSCLE_CELL_PROLIFERATION
GO_SMOOTH_MUSCLE_CONTRACTION
GOpaque
GO_REGULATION_OF_SYSTEM_PROCESS
GOACTIN_MEDIATED_CONTRACTION
GO_STRIATED_MUSCLE_CELL_DIFFERENTIATION
GO_ACTOMYOSIN_STRUCTURE_ORGANIZATION
GO_HOMOTYPIC_CELL_CELL_ADSHION
GO_CARDIAC_MUSCLE_TISSUE_DEVELOPMENT
GO_HEART_PROCESS
GO_CARDIAC_CELL_DEVELOPMENT

FDR q-val

B

SS

GO_REGULATION_OF_MRNA_SPLICING_VIA_SPLICEOSOME
GO_RNA_SPLICING_VIA_TRANSESTERIFICATION_REACTIONS
GO_MRNA_PROCESSING
GO_RNA_SPLICING
GO_ALTERNATIVE_MRNA_SPLICING_VIA_SPLICEOSOME
GO_RNA_PROCESSING
GO_REGULATION_OF_MRNA_SPLICING
GO_REGULATION_OF_MRNA_PROCESSING
GO_NUCLEOBASE_CONTAINING_COMPOUND_TRANSPORT
GO_REGULATION_OF_MRNA_METABOLIC_PROCESS
GO_ESTABLISHMENT_OF_RNA.Localization
GO_POSITIVE_REGULATION_OF_RNA_METABOLIC_PROCESS
GO_NEGATIVE_REGULATION_OF_MRNA_METABOLIC_PROCESS
GO_PROTEIN_CONTAINING_COMPLEX_LOCALIZATION
GO_PROTEIN_DNA_COMPLEX_SUBUNIT_ORGANIZATION
GO_POSITIVE_REGULATION_OF_NUCLEOBASE_CONTAINING_COMPOUND_METABOLIC_PROCESS
GO_MRNA_TRANSPORT
GO_MRNA_CIS_SPLICING_VIA_SPLICEOSOME
GO_NEGATIVE_REGULATION_OF_NUCLEOBASE_CONTAINING_COMPOUND_METABOLIC_PROCESS
GO_POSITIVE_REGULATION_OF_MRNA_METABOLIC_PROCESS

FDR q-val

C

UPS

GO_MYOELID_LEUCOCYTE_MEDIATED_IMMUNITY
GO_CELL_ACTIVATION_INVOLVED_IN_IMMUNE_RESPONSE
GO_MYOELID_LEUCOCYTE_ACTIVATION
GO_IRON_ION_TRANSPORT
GO_TRANSITION_METAL_ION_TRANSPORT
GO_IMMUNE_EFFECTOR_PROCESS
GO_REGULATION_OF_CELLULAR_COMPONENT_BIOMENERGY
GO_CELLULAR_MONOVALENT_INORGANIC_CATION_HOMEOASTASIS
GO_CELL_ACTIVATION
GO_PHAGOCYTOSIS
GO_POSITIVE_REGULATION_OF_T_CELL_PROLIFERATION
GO_LEUCOCYTE_MEDIATED_IMMUNITY
GO_ACTIN_POLYMORPHIZATION_OR_DEPOLYMORPHIZATION
GO_POSITIVE_REGULATION_OF_LEUCOCYTE_MIGRATION
GO_REGULATION_OF_MACROAUTOPHAGY
GO_SECRETION
GO_POST_GOLGI_VESICLE_MEDIATED_TRANSPORT
GO_ARP2_ARP3_COMPLEX_MEDIATED_ACTIN_NUCLEATION
GO_REGULATION_OF_CELLULAR_COMPONENT_SIZE
GO_LAMELLIPODIUM_ORGANIZATION

FDR q-val
