

Proteomic profiling identifies co-regulated expression of splicing factors as a characteristic feature of intravenous leiomyomatosis

Lukas Krasny¹, Chris P Wilding¹, Emma Perkins^{1,2}, Amani Arthur¹, Nafia Guljar¹, Andrew D Jenks¹, Cyril Fisher³, Ian Judson², Khin Thway^{1,2}, Robin L Jones^{2,4} and Paul H Huang¹

Affiliations:

¹ Division of Molecular Pathology, The Institute of Cancer Research, Sutton, United Kingdom.

² The Royal Marsden NHS Foundation Trust, London, United Kingdom.

³ University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK.

⁴ Division of Clinical Studies, The Institute of Cancer Research, London, United Kingdom.

Correspondence to:

Paul H Huang
Institute of Cancer Research
15 Cotswold Road
Sutton, SM2 5NG
United Kingdom
Email: paul.huang@icr.ac.uk
Tel: +44 207 153 5554

Keywords: Proteomics, Spliceosome, Splicing Factors, Leiomyoma, Intravenous Leiomyomatosis

Abstract

Intravenous leiomyomatosis (IVLM) is a rare benign smooth muscle tumour that is characterised by intravenous growth in the uterine and pelvic veins. Previous DNA copy number and transcriptomic studies have shown that IVLM harbours unique genomic and transcriptomic alterations when compared to uterine leiomyoma (uLM), which may account for their distinct clinical behaviour. Here we undertake the first comparative proteomic analysis of IVLM and other smooth muscle tumours (comprising uLM, soft tissue leiomyoma and benign metastasising leiomyoma) utilising data-independent acquisition mass spectrometry. We show that, at the protein level, IVLM is defined by the unique co-regulated expression of splicing factors. In particular, IVLM is enriched in two clusters composed of co-regulated proteins from the hnRNP, LSM, SR and Sm classes of the spliceosome complex. One of these clusters (Cluster 3) is associated with key biological processes including nascent protein translocation and cell signalling by small GTPases. Taken together, our study provides evidence of co-regulated expression of splicing factors in IVLM compared to other smooth muscle tumours which suggests a possible role for alternative splicing in the pathogenesis of IVLM.

Introduction

Intravenous leiomyomatosis (IVLM) is a rare histologically benign smooth muscle tumour which is characterised by intravenous growth in the uterine and pelvic veins [1,2]. In some instances, it can extend into the inferior vena cava and the right heart which in rare cases may cause death [3,4]. IVLM is usually present with concomitant uterine leiomyoma (uLM) and one theory is that it originates from a pre-existing uLM where it extends and invades into the vessel wall [4,5]. Given that there are some instances where IVLM arises in the absence of a uLM [2,6], an alternate theory is that this tumour originates from the smooth muscle cells of the vessel wall. In addition to IVLM, there are other rare smooth muscle tumours with unusual quasi-malignant clinical behaviour such as benign metastasising leiomyoma (BML) and disseminated peritoneal leiomyomatosis [7,8].

Previous studies have undertaken comparative analysis of the molecular features of IVLM versus uLM to gain a better understanding of its underlying biology as well as the relationship between the two entities [9-15]. Some of the system-wide comprehensive profiling studies that have been reported include array comparative genomic profiling (aCGH) and transcriptomic analysis [9,11,13,14]. Collectively, these focused and system-wide studies indicate that IVLM share some cytogenetic and protein expression features with uLM (e.g. translocations in (12;14) and HMGA2 protein expression) [11,12,14,15], while at the same time harbour genetic and transcriptomic alterations that are unique. These unique alterations include distinct *MED12* mutations and elevated *HOXA13* gene expression in IVLM [10,12,13]. Given its rarity, all of the published Omics-based IVLM molecular profiling studies, with the exception of a recent study by Ordulu et al.[11], have been limited to a small number of cases (typically <5).

To date no proteomic profiling analyses have been undertaken in IVLM. Proteins are the critical drivers of cellular communication in normal cells and dysregulation of protein function is causative of many diseases including cancer [16,17]. We hypothesized that, unlike genomic and transcriptomic analysis, proteomic profiling will provide a more direct readout of the biological pathways and protein complexes that may play a role in the pathogenesis of IVLM [18,19]. Here we undertake a comparative mass spectrometry-based proteomic analysis of IVLM and other smooth muscle tumours (uLM, soft tissue leiomyoma (stLM) and BML), and demonstrate that at the protein level, IVLM is characterised by the unique co-regulated expression of splicing factors that comprise the spliceosome.

Materials and methods

Patients and tumour specimens

Use of archival formalin fixed paraffin embedded (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). One of the IVLM cases in this series has previously been described in a case report [20]. FFPE tissue from surgically resected primary tumours and accompanying annotation of baseline clinico-pathological variables were identified and retrieved through retrospective review of departmental database and medical notes at the Royal Marsden NHS Foundation Trust. 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. Five 20 μ m sections were cut from each selected tumour block and, where indicated, macroadissected to enrich to >75% viable tumour content.

Protein extraction and sample preparation

The samples were processed as previously described [18]. Briefly, 20 μ m tissue sections from each sample were deparaffinised in xylene, rehydrated by washes with decreasing ethanol gradient and then dried. Samples were homogenized in lysis buffer (0.1M Tris-HCl pH 8.8,

0.50% (w/v) sodium deoxycholate, 0.35% (w/v) sodium lauryl sulphate) at a ratio of 200ul/mg of dry tissue using a LabGen700 blender (ColeParmer) with 3x 30s pulses. Homogenates were sonicated on ice for 10 min and then incubated at 95°C for 1 h to reverse formalin crosslinks. Lysis was continued by shaking at 750rpm at 80°C for 2 h. The resulting homogenate was then centrifuged for 15min at 4°C at 15,000rpm, the supernatant was collected and protein concentration of the supernatant was measured by bicinchoninic acid (BCA) assay (Pierce). The extracted proteins were digested using the Filter-Aided Sample Preparation (FASP) protocol as previously described [21]. 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, reduced with 10mM dithiothreitol (DTT) and alkylated with 55mM iodoacetamide (IAA). The sample was washed with 100mM ammonium bicarbonate (ABC) and digested by trypsin (Promega, trypsin to starting protein ratio 1:100 µg) overnight at 37°C. Peptides were desalted on C18 SepPak columns (Waters), dried in a SpeedVac concentrator and stored at -80°C.

SWATH-MS data acquisition and processing

Quantitative proteomic profiling was performed by sequential window acquisition of all theoretical fragments mass spectrometry (SWATH-MS) which is also known as data-independent acquisition mass spectrometry. Dried, desalted peptides were resuspended in a buffer A (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 self-made trap column packed with a 10 µm ReprosilPur C18AQ beads (Dr. Maisch) and washed for 5 minutes by buffer A. Peptides were then 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% formic acid) in 120 min and a flow rate of 250 nl/min was used. Each sample was acquired in 2 technical replicates. Acquisition parameters were as follows: 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. Cycle time of 3.1 s was used resulting in average 8 datapoints per elution peak. SWATH-MS spectra were analysed using Spectronaut 15.2 (Biognosys AG) against a published human library [22]. FDR was restricted to 1% on both protein and PSM level. Peak area of 3 to 6 fragment ions was used for peptide quantification. The mean value of max 6 peptides was used to quantify proteins while 2 unique peptides was set as a minimum requirement for inclusion of a protein in the subsequent analysis.

Data processing and statistical methods

The proteomics dataset was further processed using R, Perseus 1.5.6 [23,24] and GraphPad 8.2.1. Protein quantities were log2 transformed and quantile normalised at sample level using proBatch package [25] in R followed by protein median centering across the samples. The normalized dataset was then visualized by hierarchical clustering using ComplexHeatmap package in R [26]. Gene Set Enrichment Analysis (GSEA) was applied using GenePattern online tool [27] to identify gene sets obtained from the MSigDB (c5.gobp.v7.5) [28] that were significantly enriched in IVLM samples. Similarly, single sample GSEA (ssGSEA) was applied using GenePattern to score sample-specific enrichment of the Spliceosome gene set from the KEGG pathways database [29]. To identify spliceosome components, the list of all identified proteins in this study was cross-referenced with the annotated spliceosome protein interaction dataset published by Hegele et al. [30]. Mutual co-expression of the splicing factors was assessed by Pearson's correlation coefficient that was calculated in Perseus for all possible combinations of the identified splicing factors. The resulting similarity matrix was analysed and visualised by ConsensusClusterPlus [31] and ComplexHeatmap packages in R respectively.

To study association of the splicing factors identified in clusters 1-3 with known biological pathways, Pearson's correlation coefficients between splicing factors and all other proteins in the proteomic dataset (after removal of all proteins annotated in the Spliceosome Database [32]) were calculated in Perseus. The resulting similarity matrices were hierarchically clustered and visualized by ComplexHeatmap package in R, where rows of each matrix were split into 4 clusters using k-means partitioning, Euclidean distance and 1000 repetitions. Subsets of proteins from the clusters with the highest and lowest average correlation were then used for over-representation analysis using DAVID 6.8 Functional analysis online tool [33].

Results

Quantitative proteomic profiling of smooth muscle tumours

The cohort is comprised of FFPE tumour material from 14 patients treated at The Royal Marsden Hospital. These specimens were obtained from surgical resections of IVLM (n = 3), uLM (n = 3), stLM (n = 7) and BML (n=1). Tumour specimens were subjected to sample preparation and protein extraction as depicted in Figure 1. Digested peptides then underwent proteomic profiling with SWATH-MS in technical duplicates. This analysis resulted in the identification and quantification of 2,473 proteins (Table S1). Unsupervised clustering of the full dataset shows that the IVLM cases largely cluster together separate from the stLM and uLM cases (Figure 2A). Interestingly the only BML case in the cohort clusters most closely to the IVLM cases.

Assessment of proteins that are significantly different in IVLM cases compared to uLM, stLM and BML cases identified 162 proteins of which 109 and 53 proteins are upregulated (>2 fold) or downregulated (<2 fold) in IVLM respectively (Fig 2B). Consistent with published immunohistochemical analysis studies [12], expression of the chromatin factor HMGA2, a protein which is highly expressed in IVLM due to the breakpoint on 12q14-15 [11,12,14], was not significantly different between IVLM and the other smooth muscle tumours in the cohort (Fig S1). Interestingly we find that 29/162 (18%) of the differentially expressed proteins are components of the spliceosome complex (Figure 2B).

Enrichment of splicing processes in IVLM

To further investigate the biological processes that are enriched in IVLM compared to the other smooth muscle tumours, we undertook gene set enrichment analysis (GSEA) of the full proteomic dataset (Figure 3A). We show that the majority of the top 20 ranked enriched gene sets are processes associated with RNA splicing, processing, transport or metabolism. Beyond RNA-related biological processes, other enriched gene sets include protein targeting and localisation to membrane, regulation of gene transcription and translation. In line with the observation that a significant proportion of proteins enriched in IVLM are components of the spliceosome complex (Figure 2B), single sample GSEA (ssGSEA) of the proteomic data for each specimen in the cohort using the KEGG spliceosome gene set showed that the IVLM cases had significantly higher ssGSEA spliceosome scores compared to the other smooth muscle tumours in the cohort (Figure 3B). Taken together, our data indicate that both the spliceosome complex and biological processes involving RNA biology are enriched in IVLM specimens.

Identification of co-regulated expression of splicing factors in the proteomic profiling dataset

It is well-established that the spliceosome is a highly dynamic macromolecular complex where more than 200 splicing factors are assembled into distinct complexes that vary in their composition in space and time [30,34]. We therefore hypothesized that despite the overall enrichment of spliceosome components in IVLM (Figure 3B), it is possible that subsets of co-regulated splicing factors may be responsible for the distinct clinical behaviour of IVLM versus leiomyomas. Indeed, unsupervised hierarchical clustering of 116 spliceosome components in the proteomic dataset showed that the spliceosome complex as a whole was not upregulated

in IVLM (Figure 4A). Rather, there appeared to be subsets of splicing factors that were differentially expressed in IVLM, uLM and stLM.

Inspired by a previous study which showed that co-regulation of splicing factors is important in regulating breast cancer progression [35], we performed a Pearson's correlation coefficient analysis of the protein expression levels of all possible combinations of 116 splicing factors in our dataset. Consensus clustering identified 3 clusters of splicing factors which is shown in the similarity matrix in Figure 4B (composition of each cluster provided in Table S2). In particular, Clusters 2 (n=43) and 3 (n=40) contain splicing factors which are negatively correlated between clusters but are positively correlated within clusters. Cluster 1 (n=33) is mixed with both positively and negatively correlated splicing factors.

Distinct co-regulated clusters are comprised of splicing factors which are differentially expressed in IVLM versus the other smooth muscle tumours.

An evaluation of the composition of splicing factors showed that each cluster is comprised of different proportions of core and non-core spliceosome proteins with Cluster 2 having the highest proportion of core proteins (65%) and Cluster 3 having the least core proteins (25%) (Figure 4C). Furthermore, assessment of the splicing factor classes based on nomenclature defined by Hegele et al., [30] finds that the splicing factor class composition of Clusters 1 and 3 is similar with the majority of proteins coming from the hnRNP, LSm, SR and Sm protein classes (Figure 4C). In contrast, the composition of cluster 2 is very different with U2, U2 rel and U5 protein classes dominating.

Quantitative assessment of the proteomic data showed that when broken down by cluster assignment, the IVLM specimens were significantly enriched in co-regulated splicing factors from Clusters 1 and 3 versus the other smooth muscle tumours in the cohort (Figure 4C). No significant difference between IVLM and the other smooth muscle tumours was seen in co-regulated splicing factors in Cluster 2. Collectively, this analysis indicates that at the protein level, IVLM is characterised by the co-regulated expression of specific classes of splicing factors that comprise the spliceosome.

Co-regulated splicing factors are associated with multiple biological pathways, including protein translocation and signal transduction by small GTPases.

We sought to determine if the expression of splicing factors in each of these clusters was linked to specific biological process. To do this, the Pearson's correlation coefficient was calculated between all the proteins in the dataset (excluding spliceosomal proteins) and splicing factors in each of the three clusters. Unsupervised hierarchical clustering finds that 537 and 585 proteins were positively or negatively correlated with the splicing factors in Cluster 2, respectively (Figure 5A, clusters C and A). The same analysis in Cluster 3 identified positive and negative correlation in 545 and 738 proteins, respectively (Figure 5B, clusters C and B). Unsurprisingly, since Cluster 1 comprised of both positively and negatively correlated splicing factors, no significantly correlated proteins were found in our dataset (data not shown). Given that Clusters 2 and 3 have opposing profiles in co-regulated splicing factors (Figure 4B), it is expected that proteins correlating with these clusters would follow the same trend. Indeed, we demonstrate that there was substantial overlap of proteins which show opposite co-expression patterns (i.e. positively correlated proteins in Cluster 2 and negatively correlated proteins in Cluster 3), and vice versa (Figure 5C).

Focusing on Cluster 3 which is significantly upregulated in IVLM (Figure 4C), over-representation analysis finds 4 ontologies that are enriched in the proteins that are positively correlated with the splicing factors in this cluster (Figure 5D). These ontologies include nascent protein targeting to the endoplasmic reticulum (SRP-dependent cotranslational protein targeting to membrane), signal transduction mediated by small GTPases, hydrolysis of proteins by peptidases (negative regulation of endopeptidase activity) and proteins involved

in viral transcription. The positively coregulated proteins in these ontologies is shown in the chord diagram in Figure 5E.

Discussion

IVLM is a rare benign smooth muscle tumour with quasi-malignant clinical behaviour. Previous profiling studies characterising its molecular features have focused on DNA copy number and transcriptomic alterations [9,11,13,14]. Here we performed the first proteome level analysis of IVLM and compare it to other smooth muscle tumours including uLM, stLM and BML. We show that IVLM is characterised by a differential expression of spliceosome complex components. In particular, by utilising a bioinformatics approach to delineate co-regulation of splicing factors, we find that there are two specific clusters of co-regulated splicing factors in the hnRNP, LSm, SR and Sm protein classes that are enriched in IVLM compared to the other smooth muscle tumours in this cohort. Finally, we demonstrate that one of these clusters (Cluster 3) is associated with high expression of proteins involved in key biological processes such as nascent protein translocation and signalling by small GTPases. To our knowledge this is the first demonstration that IVLM is characterised by a distinct group of co-regulated splicing factors, which may contribute to its unique clinical behaviour. It highlights the utility of proteomics to provide novel insights into IVLM tumour biology beyond the current state-of-the-art gained from published aCGH and gene expression studies.

Splicing occurs through a complex series of well-regulated steps mediated by the spliceosome machinery [36]. It has been shown that aberrations in specific splicing factors disrupt the composition of the spliceosome complex and drive carcinogenesis [37,38]. For instance, mutations in the splicing factor SF3B1 in both solid and liquid cancers initiate oncogenic alternative splicing reprogramming that is key to cancer development and progression [39-43]. Furthermore, it has been recently shown that some of these splicing factor mutations may induce new vulnerabilities that can be therapeutically exploited in a synthetic lethal fashion [44-46]. In the same vein, it is possible that the distinct co-regulation of splicing factors observed in IVLM may result in dysregulated alternative splicing that could account for its intravenous growth patterns. Unfortunately, due to the highly fragmented nature of total RNA extracted from FFPE specimens, we were unsuccessful in our efforts to measure alternative splicing profiles by RT-PCR from the cases in this series despite multiple repeated attempts. Future RNASeq or RT-PCR analysis on prospectively collected flash frozen specimens would be key to establishing if differential alternative splicing occurs in IVLM versus uLM. Identifying such alternatively spliced genes could offer a mechanistic explanation into the quasi-malignant behaviour of IVLM.

This study is limited by the small number of IVLM cases that were studied. IVLM is a rare condition and the vast majority of profiling studies to date comprise a small number of cases (typically <5). Despite the limited numbers, we were able to demonstrate that there was a statistically significant enrichment of co-regulated spliceosome components in IVLM. Interestingly, we show that the sole BML case in our cohort clustered most closely to the IVLM cases (Figure 2A). BML is another rare unusual variant of leiomyoma that often manifests as multiple nodules in the lungs and other sites [47]. A recent aCGH analysis finds that IVLM and BML share recurrent copy number alterations that are rarely seen in uLM [11]. Consistent with this finding, our data shows that at the proteomic level, BML is more similar to IVLM compared to uLM. It is however important to note that our proteomic analysis was performed on a small case series treated within a single institution and any findings will need to be independently validated.

Conclusions

In summary, we have undertaken a comparative proteomic profiling study of IVLM and other smooth muscle tumours (uLM, stLM and BML) and describe the selective enrichment of co-

regulated splicing factors which are associated with distinct biological pathways. We anticipate that future work integrating proteomics with complementary Omics-based profiling approaches such as RNAseq will shed further insights into the possible role of alternative splicing in the pathogenesis of IVLM.

Funding: This research was funded by The Institute of Cancer Research.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of The Royal Marsden NHS Foundation Trust (PROSPECTUS study, CCR 4371, REC 16/EE/0213)

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [48] partner repository with the dataset identifier PXD031637.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results

References

1. Valdes Devesa, V.; Conley, C.R.; Stone, W.M.; Collins, J.M.; Magrina, J.F. Update on intravenous leiomyomatosis: report of five patients and literature review. *Eur J Obstet Gynecol Reprod Biol* **2013**, *171*, 209-213, doi:10.1016/j.ejogrb.2013.09.031.
2. Clement, P.B.; Young, R.H.; Scully, R.E. Intravenous leiomyomatosis of the uterus. A clinicopathological analysis of 16 cases with unusual histologic features. *Am J Surg Pathol* **1988**, *12*, 932-945.
3. Ling, F.T.; David, T.E.; Merchant, N.; Yu, E.; Butany, J.W. Intracardiac extension of intravenous leiomyomatosis in a pregnant woman: A case report and review of the literature. *Can J Cardiol* **2000**, *16*, 73-79.
4. Lo, K.W.; Lau, T.K. Intracardiac leiomyomatosis. Case report and literature review. *Arch Gynecol Obstet* **2001**, *264*, 209-210, doi:10.1007/s004040000115.
5. Konrad, P.; Mellblom, L. Intravenous leiomyomatosis. *Acta Obstet Gynecol Scand* **1989**, *68*, 371-376, doi:10.3109/00016348909028675.
6. Carr, R.J.; Hui, P.; Buza, N. Intravenous leiomyomatosis revisited: an experience of 14 cases at a single medical center. *Int J Gynecol Pathol* **2015**, *34*, 169-176, doi:10.1097/PGP.0000000000000127.
7. Pacheco-Rodriguez, G.; Taveira-DaSilva, A.M.; Moss, J. Benign Metastasizing Leiomyoma. *Clin Chest Med* **2016**, *37*, 589-595, doi:10.1016/j.ccm.2016.04.019.
8. Altinok, G.; Usubutun, A.; Kucukali, T.; Gunalp, S.; Ayhan, A. Disseminated peritoneal leiomyomatosis. A benign entity mimicking carcinomatosis. *Arch Gynecol Obstet* **2000**, *264*, 54-55, doi:10.1007/s004049900055.
9. Wang, W.; Wang, Y.; Chen, F.; Zhang, M.; Jia, R.; Liu, X.; Zhang, C.; Shao, J.; Cheng, N.; Ma, G.; et al. Intravenous leiomyomatosis is inclined to a solid entity different from uterine leiomyoma based on RNA-seq analysis with RT-qPCR validation. *Cancer Med* **2020**, *9*, 4581-4592, doi:10.1002/cam4.3098.
10. Wang, L.; Hu, S.; Xin, F.; Zhao, H.; Li, G.; Ran, W.; Xing, X.; Wang, J. MED12 exon 2 mutation is uncommon in intravenous leiomyomatosis: clinicopathologic features and molecular study. *Hum Pathol* **2020**, *99*, 36-42, doi:10.1016/j.humpath.2020.03.011.
11. Ordulu, Z.; Chai, H.; Peng, G.; McDonald, A.G.; De Nictolis, M.; Garcia-Fernandez, E.; Hardisson, D.; Prat, J.; Li, P.; Hui, P.; et al. Molecular and clinicopathologic

characterization of intravenous leiomyomatosis. *Mod Pathol* **2020**, 33, 1844-1860, doi:10.1038/s41379-020-0546-8.

12. Lu, B.; Liu, Q.; Tang, L.; Ma, Y.; Shi, H. Intravenous leiomyomatosis: molecular analysis of 17 cases. *Pathology* **2020**, 52, 213-217, doi:10.1016/j.pathol.2019.10.009.

13. Zhang, X.; Wu, L.; Xu, R.; Zhu, C.; Ma, G.; Zhang, C.; Liu, X.; Zhao, H.; Miao, Q. Identification of the molecular relationship between intravenous leiomyomatosis and uterine myoma using RNA sequencing. *Sci Rep* **2019**, 9, 1442, doi:10.1038/s41598-018-37452-3.

14. Ordulu, Z.; Nucci, M.R.; Dal Cin, P.; Hollowell, M.L.; Otis, C.N.; Hornick, J.L.; Park, P.J.; Kim, T.M.; Quade, B.J.; Morton, C.C. Intravenous leiomyomatosis: an unusual intermediate between benign and malignant uterine smooth muscle tumors. *Mod Pathol* **2016**, 29, 500-510, doi:10.1038/modpathol.2016.36.

15. Buza, N.; Xu, F.; Wu, W.; Carr, R.J.; Li, P.; Hui, P. Recurrent chromosomal aberrations in intravenous leiomyomatosis of the uterus: high-resolution array comparative genomic hybridization study. *Hum Pathol* **2014**, 45, 1885-1892, doi:10.1016/j.humpath.2014.05.010.

16. Burns, J.; Wilding, C.P.; R, L.J.; P, H.H. Proteomic research in sarcomas - current status and future opportunities. *Semin Cancer Biol* **2020**, 61, 56-70, doi:10.1016/j.semcan.2019.11.003.

17. Noujaim, J.; Payne, L.S.; Judson, I.; Jones, R.L.; Huang, P.H. Phosphoproteomics in translational research: a sarcoma perspective. *Ann Oncol* **2016**, 27, 787-794, doi:10.1093/annonc/mdw030.

18. Milighetti, M.; Krasny, L.; Lee, A.T.J.; Morani, G.; Szecsei, C.; Chen, Y.; Guljar, N.; McCarthy, F.; Wilding, C.P.; Arthur, A.; et al. Proteomic profiling of soft tissue sarcomas with SWATH mass spectrometry. *J Proteomics* **2021**, 241, 104236, doi:10.1016/j.jprot.2021.104236.

19. Krasny, L.; Huang, P.H. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. *Mol Omics* **2021**, 17, 29-42, doi:10.1039/d0mo00072h.

20. Judson, I.; Messiou, C. Vitamin D deficiency in the pathogenesis of leiomyoma and intravascular leiomyomatosis: A case report and review of the literature. *Gynecol Oncol Rep* **2021**, 35, 100681, doi:10.1016/j.gore.2020.100681.

21. Wisniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **2009**, 6, 359-362, doi:10.1038/nmeth.1322.

22. Rosenberger, G.; Koh, C.C.; Guo, T.; Rost, H.L.; Kouvonen, P.; Collins, B.C.; Heusel, M.; Liu, Y.; Caron, E.; Vichalkovski, A.; et al. A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci Data* **2014**, 1, 140031, doi:10.1038/sdata.2014.31.

23. Tyanova, S.; Cox, J. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. *Methods Mol Biol* **2018**, 1711, 133-148, doi:10.1007/978-1-4939-7493-1_7.

24. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M.Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **2016**, 13, 731-740, doi:10.1038/nmeth.3901.

25. Cuklina, J.; Lee, C.H.; Williams, E.G.; Sajic, T.; Collins, B.C.; Rodriguez Martinez, M.; Sharma, V.S.; Wendt, F.; Goetze, S.; Keele, G.R.; et al. Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial. *Mol Syst Biol* **2021**, 17, e10240, doi:10.15252/msb.202110240.

26. Gu, Z.; Eils, R.; Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **2016**, 32, 2847-2849, doi:10.1093/bioinformatics/btw313.

27. Reich, M.; Liefeld, T.; Gould, J.; Lerner, J.; Tamayo, P.; Mesirov, J.P. GenePattern 2.0. *Nat Genet* **2006**, 38, 500-501, doi:10.1038/ng0506-500.

28. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; 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*, 15545-15550, doi:10.1073/pnas.0506580102.
29. Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **2000**, *28*, 27-30, doi:10.1093/nar/28.1.27.
30. Hegele, A.; Kamburov, A.; Grossmann, A.; Sourlis, C.; Wowro, S.; Weimann, M.; Will, C.L.; Pena, V.; Luhrmann, R.; Stelzl, U. Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell* **2012**, *45*, 567-580, doi:10.1016/j.molcel.2011.12.034.
31. Wilkerson, M.D.; Hayes, D.N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **2010**, *26*, 1572-1573, doi:10.1093/bioinformatics/btq170.
32. Cvitkovic, I.; Jurica, M.S. Spliceosome database: a tool for tracking components of the spliceosome. *Nucleic Acids Res* **2013**, *41*, D132-141, doi:10.1093/nar/gks999.
33. Huang da, W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **2009**, *4*, 44-57, doi:10.1038/nprot.2008.211.
34. Wahl, M.C.; Will, C.L.; Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **2009**, *136*, 701-718, doi:10.1016/j.cell.2009.02.009.
35. Koedoot, E.; Smid, M.; Foekens, J.A.; Martens, J.W.M.; Le Devedec, S.E.; van de Water, B. Co-regulated gene expression of splicing factors as drivers of cancer progression. *Sci Rep* **2019**, *9*, 5484, doi:10.1038/s41598-019-40759-4.
36. Wilkinson, M.E.; Charenton, C.; Nagai, K. RNA Splicing by the Spliceosome. *Annu Rev Biochem* **2020**, *89*, 359-388, doi:10.1146/annurev-biochem-091719-064225.
37. Dvinge, H.; Kim, E.; Abdel-Wahab, O.; Bradley, R.K. RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer* **2016**, *16*, 413-430, doi:10.1038/nrc.2016.51.
38. Kahles, A.; Lehmann, K.V.; Toussaint, N.C.; Huser, M.; Stark, S.G.; Sachsenberg, T.; Stegle, O.; Kohlbacher, O.; Sander, C.; Cancer Genome Atlas Research, N.; et al. Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell* **2018**, *34*, 211-224 e216, doi:10.1016/j.ccr.2018.07.001.
39. El Marabti, E.; Younis, I. The Cancer Spliceome: Reprograming of Alternative Splicing in Cancer. *Front Mol Biosci* **2018**, *5*, 80, doi:10.3389/fmemb.2018.00080.
40. Alsafadi, S.; Houy, A.; Battistella, A.; Popova, T.; Wassef, M.; Henry, E.; Tirode, F.; Constantinou, A.; Piperno-Neumann, S.; Roman-Roman, S.; et al. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat Commun* **2016**, *7*, 10615, doi:10.1038/ncomms10615.
41. Furney, S.J.; Pedersen, M.; Gentien, D.; Dumont, A.G.; Rapinat, A.; Desjardins, L.; Turajlic, S.; Piperno-Neumann, S.; de la Grange, P.; Roman-Roman, S.; et al. SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov* **2013**, *3*, 1122-1129, doi:10.1158/2159-8290.CD-13-0330.
42. Maguire, S.L.; Leonidou, A.; Wai, P.; Marchio, C.; Ng, C.K.; Sapino, A.; Salomon, A.V.; Reis-Filho, J.S.; Weigelt, B.; Natrajan, R.C. SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J Pathol* **2015**, *235*, 571-580, doi:10.1002/path.4483.
43. Rossi, D.; Bruscaggin, A.; Spina, V.; Rasi, S.; Khiabanian, H.; Messina, M.; Fangazio, M.; Vaisitti, T.; Monti, S.; Chiaretti, S.; et al. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood* **2011**, *118*, 6904-6908, doi:10.1182/blood-2011-08-373159.
44. Dalton, W.B.; Helmenstine, E.; Walsh, N.; Gondek, L.P.; Kelkar, D.S.; Read, A.; Natrajan, R.; Christenson, E.S.; Roman, B.; Das, S.; et al. Hotspot SF3B1 mutations

induce metabolic reprogramming and vulnerability to serine deprivation. *J Clin Invest* **2019**, *129*, 4708-4723, doi:10.1172/JCI125022.

45. Lee, S.C.; North, K.; Kim, E.; Jang, E.; Obeng, E.; Lu, S.X.; Liu, B.; Inoue, D.; Yoshimi, A.; Ki, M.; et al. Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. *Cancer Cell* **2018**, *34*, 225-241 e228, doi:10.1016/j.ccr.2018.07.003.

46. Singh, S.; Ahmed, D.; Dolatshad, H.; Tatwavedi, D.; Schulze, U.; Sanchi, A.; Ryley, S.; Dhir, A.; Carpenter, L.; Watt, S.M.; et al. SF3B1 mutations induce R-loop accumulation and DNA damage in MDS and leukemia cells with therapeutic implications. *Leukemia* **2020**, *34*, 2525-2530, doi:10.1038/s41375-020-0753-9.

47. Awonuga, A.O.; Shavell, V.I.; Imudia, A.N.; Rotas, M.; Diamond, M.P.; Puscheck, E.E. Pathogenesis of benign metastasizing leiomyoma: a review. *Obstet Gynecol Surv* **2010**, *65*, 189-195, doi:10.1097/OGX.0b013e3181d60f93.

48. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **2019**, *47*, D442-D450, doi:10.1093/nar/gky1106.

Figures

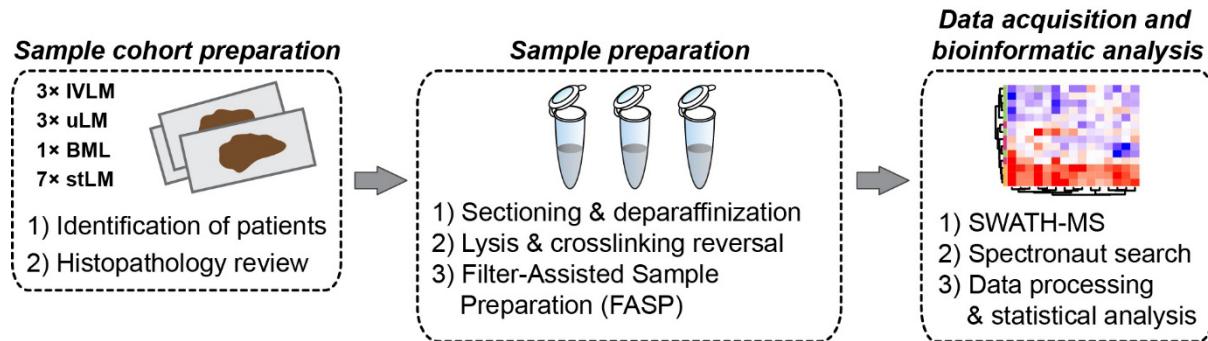


Figure 1: Experimental workflow depicting key procedures of sample selection and preparation, proteomic data acquisition and subsequent data processing and analysis.

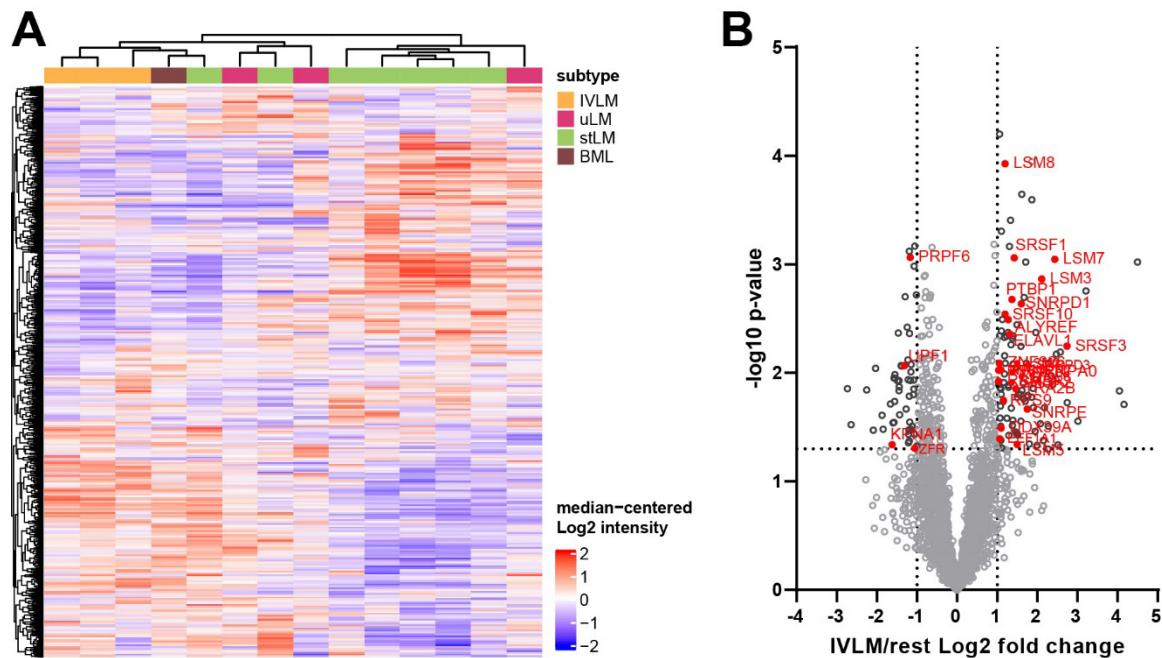


Figure 2: (A) Heatmap depicting unsupervised hierarchical clustering of 2,478 proteins that were quantified across all samples. The distance measure used for clustering is Pearson's correlation. The full protein list is provided in Table S1. (B) Volcano plot depicting difference in protein expression between IVLM cases and all the other smooth muscle tumours (rest). Splicing factors with significantly different expression levels (>2 fold or <2 fold) are highlighted in red.

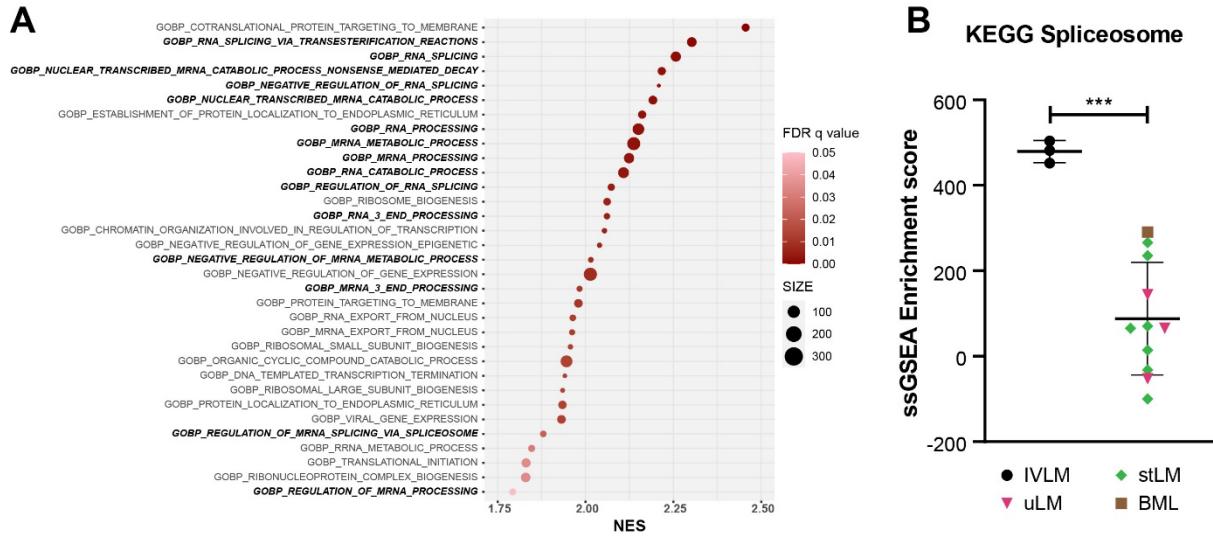


Figure 3: (A) Plot of Gene Set Enrichment Analysis (GSEA) results showing all the gene sets that are significantly enriched in IVLM samples. FDR q-value is represented by the colour of the circles while the size of the circles represents number of identified genes within each gene set. NES – normalized enrichment score. (B) Plot of single sample GSEA scores for the spliceosome gene set as defined by KEGG. The line and whiskers in plots represent mean and standard deviation. Statistical significance was calculated by two-sample t-test. *** p<0.001.

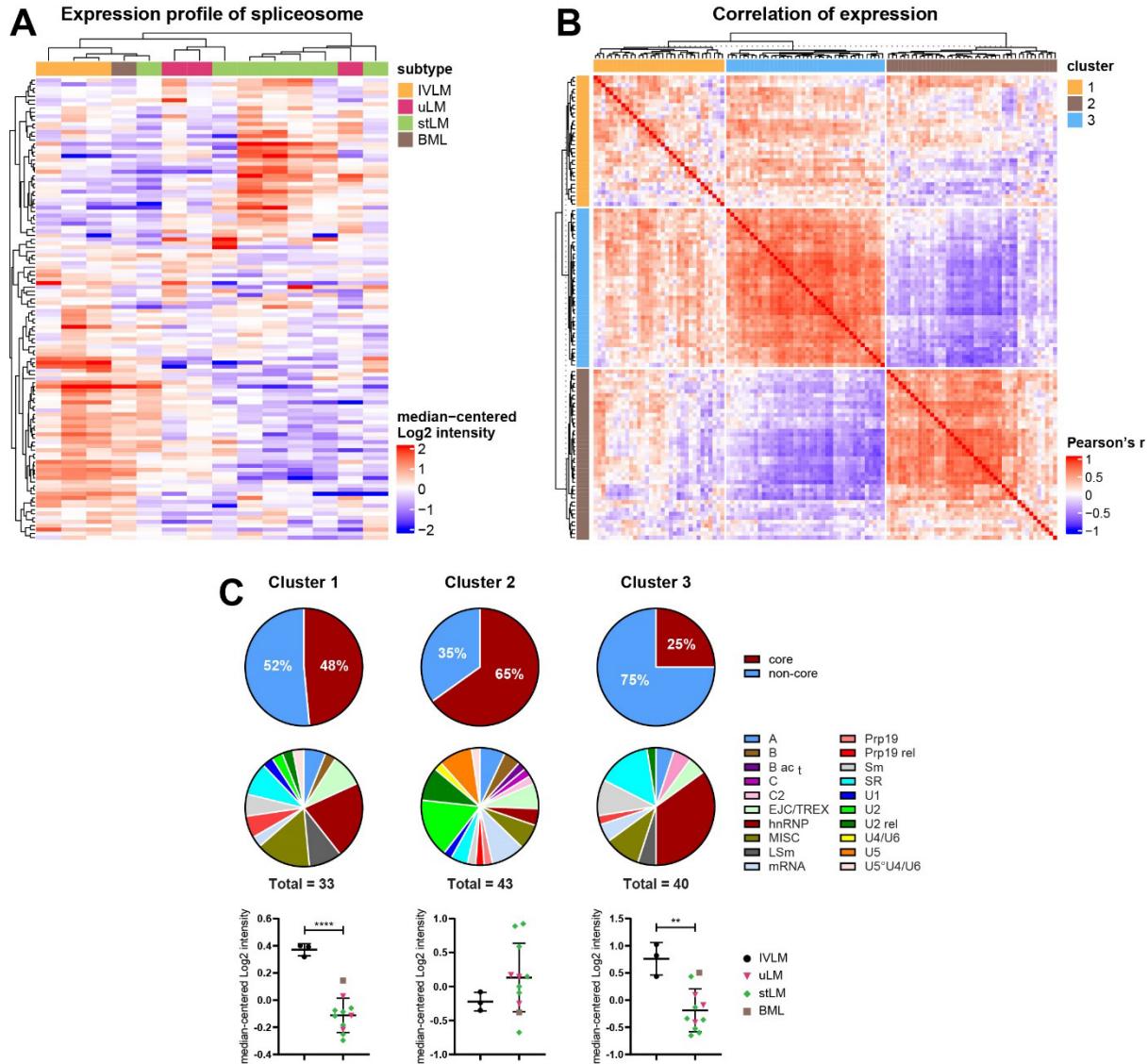


Figure 4: (A) Heatmap depicting unsupervised hierarchical clustering of 116 proteins of the spliceosome complex as defined by Hegele et al. [30]. The distance measure used for clustering is Pearson's correlation. (B) Heatmap depicting similarity matrix of Pearson's correlation coefficients of all possible pairwise combinations of the 116 splicing factors. Three clusters were identified by consensus clustering analysis. (C) Annotation and expression profile of the spliceosomal proteins belonging to clusters shown in Figure 4B. Venn diagrams depict spliceosome composition (core versus non-core, and distinct splicing factor classes) in each cluster while plots below show average expression levels of spliceosome components in each sample for a given cluster. Detailed composition of clusters and identity of individual proteins are listed in Table S2. The line and whiskers in plots represent mean and standard deviation. Statistical significance was calculated by two-sample t-test. ** p<0.01, **** p<0.0001.

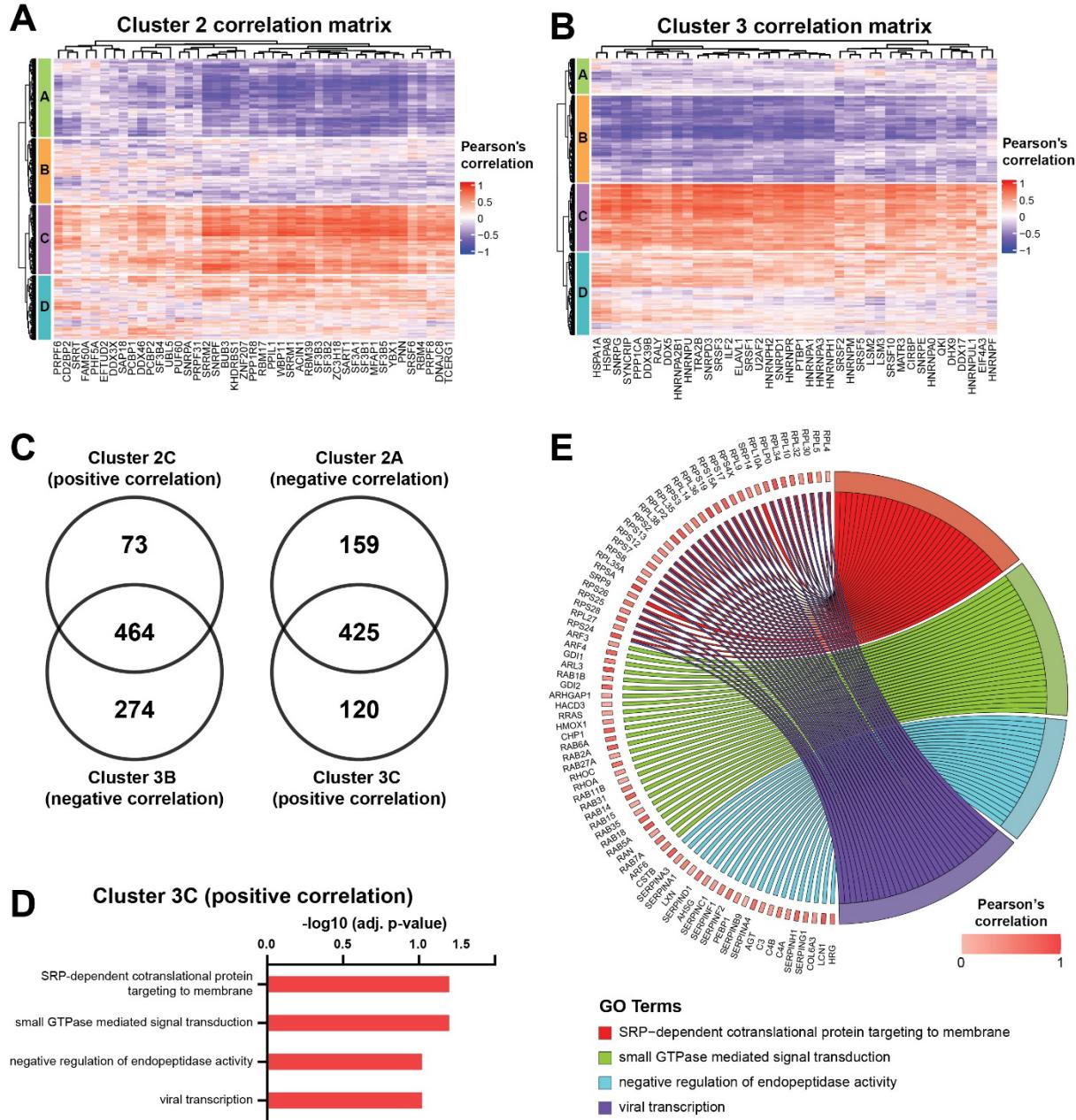


Figure 5: Heatmaps depicting correlation matrix of Pearson's correlation coefficient calculated between the splicing factors in (A) Cluster 2 or (B) Cluster 3 and all the other proteins in the dataset that are not part of the spliceosome complex. Heatmaps are split into four clusters based on k-means partitioning. (C) Venn diagrams depicting the overlap between the positively and negatively correlated proteins in Cluster 2 and 3 respectively, and vice versa. (D) Plot of overrepresentation analysis results showing ontologies which are positively correlated with the splicing factors in Cluster 3 (FDR < 0.1). (E) Chord plot depicting all positively correlated proteins identified by overrepresentation analysis in Fig 5D.

Supplemental figure

HMG2 expression

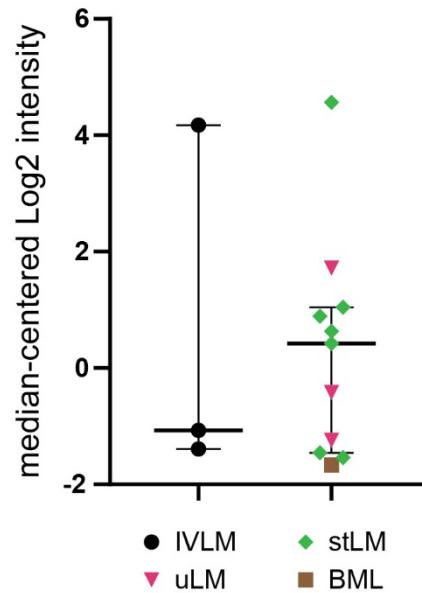


Figure S1: Expression levels of HMG2 protein of each case in the cohort. The line and whiskers in plots represent mean and standard deviation. Expression levels of this protein was not significantly different between IVLM and the other smooth muscle tumours in the cohort

Supplemental table legends

Table S1: Full proteomic dataset for the cohort. The dataset was log2 transformed and quantile normalized. Reported values represent median centred protein expression levels.

Table S2: List of splicing factors found in the three individual consensus clusters which are annotated based on spliceosome complex protein classes as defined by Hegele *et al.*, Mol. Cell. 2012