Accepted Manuscript

Quantitative mass spectrometry to identify protein markers for diagnosis of malignant pleural mesothelioma

Reuben White, Emily Pulford, David J. Elliot, Lauren A. Thurgood, Sonja Klebe

PII: S1874-3919(18)30359-2
DOI: doi:10.1016/j.jprot.2018.09.018
Reference: JPROT 3224
To appear in: Journal of Proteomics

Received date: 9 April 2018
Revised date: 27 September 2018
Accepted date: 30 September 2018

Please cite this article as: Reuben White, Emily Pulford, David J. Elliot, Lauren A. Thurgood, Sonja Klebe , Quantitative mass spectrometry to identify protein markers for diagnosis of malignant pleural mesothelioma. Jprot (2018), doi:10.1016/j.jprot.2018.09.018

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Quantitative mass spectrometry to identify protein markers for diagnosis of malignant pleural mesothelioma

Reuben White, Emily Pulford, David J Elliot, Lauren A Thurgood*, lauren.thurgood@flinders.edu.au, and Sonja Klebe

aDepartment of Anatomical Pathology, College of Medicine and Public Health, Flinders University, Bedford Park, SA 5042, Australia
bDepartment of Clinical Pharmacology, Flinders Medical Centre, Bedford Park, SA 5042, Australia
cMolecular Medicine and Pathology, College of Medicine and Public Health, Flinders University, Bedford Park, SA 5042, Australia
dDepartment of Surgical Pathology, SA Pathology, Flinders Medical Centre, Bedford Park, SA 5042, Australia

*Corresponding author.

Significance
Malignant pleural mesothelioma (MPM) is a devastating malignancy with a prognosis of less than 12 months. Even with bans on the use of asbestos in most Western countries, the incidence is still increasing due to the long latency periods between exposure and development of the disease. Diagnosis is often delayed due to invasive biopsies and lack of distinguishable markers. Patients frequently present with pleural effusions months to years before a radiologically detectable mass appears. This study aimed to investigate the proteome of pleural effusions taken from patients with MPM, adenocarcinoma and benign conditions in an attempt to identify a biomarker for early diagnosis. We identified several proteins that may be possible targets and warrant further investigation. Due to the predominance of up regulated proteins involved in VEGF signalling in MPM, we analysed VEGFA levels in effusions and found a strong correlation between VEGFA levels and survival in MPM.

Keywords:
Mesothelioma, VEGFA, lung adenocarcinoma, pleural effusions, quantitative proteomics, SWATH MS

Introduction:
Malignant mesothelioma is a highly aggressive tumour of the mesothelial cells located on serosal membranes of the thoracic cavity, abdominal cavity and testis. Mesothelioma development is intimately linked with asbestos exposure [1, 2], however, whilst occupational exposure is still the most common source of contact, there are cases of non-occupational environmental exposure, such as individuals developing mesothelioma from living in close proximity to factories and previous mining sites such as Libby, Montana [3]. Despite bans on asbestos in most developed countries the incidence of mesothelioma is still increasing worldwide. This is attributable to the latency period, around 10 - 40 years, between exposure to asbestos and the development of the disease [4]. Asbestos is still present in the built environment, and further research is essential because exposures continues [5, 6].

When asbestos fibres are inhaled a proportion is deposited in the lung tissue, and some translocate to the pleura. The fibres, especially amphiboles, are characterised by biopersistence in the tissues, with a rate of clearance of only about 10-15% per year[7, 8]. It is now generally acknowledged that highly reactive free radicals such as reactive oxygen species (ROS) generated from the surface of asbestos fibres or from their interaction with macrophages have the capacity to damage DNA and induce gene mutations that contribute to mesothelioma development [9]. This is also supported by an animal model study conducted by Kodavanti et al. which demonstrated that asbestos inhalation causes systemic inflammation and a predisposition to malignancy [10].

Malignant pleural mesothelioma (MPM) comprises the majority of mesothelioma cases, approximately 90%, with an estimated 9% of tumours located in the peritoneal cavity and less than 1% located in the pericardium and tunica vaginalis testis [11, 12]. There are three main histological subtypes of mesothelioma which have implications for diagnosis and prognosis. The epithelioid subtype accounts for 60%, the sarcomatoid subtype between 10-20%, and biphasic (mixed epithelioid and sarcomatoid) accounts for 30% [13, 14]. Kadota et al. reported the median survival length of epithelioid, biphasic and sarcomatoid subtypes are 16.2, 7.0 and 3.8 months respectively [15]. Overall, the reported median survival time after diagnosis is about 12 months or less [16]. Treatment options are limited and patients with mesothelioma present with non-specific symptoms such as dyspnoea and chest wall pain. The lack of specific symptoms often contributes to a delay in diagnosis [17]. Mesothelioma
patients commonly present with pleural effusions months to years before a radiologically detectable mass develops. Effusions are drained for comfort and diagnosis, but cytology-only diagnosis is rarely possible, and usually requires an invasive biopsy for confirmation [18]. If mesothelioma in situ could be identified at an early stage, targeted therapy may be possible and outcomes for patients might be improved. Whilst survival time at diagnosis is short, the long latency period also offers an opportunity to prevent the progression of the disease from initial exposure highlighting the need for improvements in diagnostic testing methods.

Current mesothelioma biomarkers, such as BAP1, only have consistent utility for diagnosing MPM with an epithelioid component from pleural effusions in isolation, since cells are needed for diagnosis, and sarcomatoid mesotheliomas commonly do not shed cells [18]. Commonly proposed soluble biomarkers include soluble mesothelin-related protein (SMRP) [19-22], osteopontin (OPN) [22-24], CA125 [25], megakaryocyte potentiating factor [22, 26], hyaluronan [27], galectin-1 [28] and fibulin [29]. Many of these biomarkers have reasonable specificity and can help to differentiate mesothelioma from other pleural diseases. However, these proteins currently lack the sensitivity and specificity to be diagnostic alone [18] and negative results for these markers cannot exclude the possibility of mesothelioma. One study by Rai et al. using an ELISA assay analysed serum samples from MPM patients, healthy individuals and patients with other malignancies found consistently elevated concentrations of SMRP and OPN in mesothelioma, and the concentration of SMRP returned to normal following tumour resection [24].

As pleural effusions are routinely taken from patients, we aimed to investigate the use of quantitative mass spectrometry as a tool to identify novel protein biomarkers in pleural effusion fluid associated with mesothelioma which could ultimately permit accurate and early diagnosis. Here we aim to use a new highly sensitive label free quantitative mass spectrometry (MS) technique called ‘sequential windowed acquisition of all theoretical fragment ion mass spectra (SWATH MS/MS) to compare pleural effusions from patients with benign reactive (BR) conditions, lung adenocarcinoma (AC) and MPM. We hope to validate this technology by identifying previous diagnostic and prognostic markers for mesothelioma and to identify potentially new diagnostic and treatment targets for MPM. To the best of our knowledge, there have only been 2 proteomic studies on the pleural effusions of mesothelioma patients [30, 31]. In the most recent study, Mundt et al. analysed pleural effusions from patients with mesothelioma (n=6), lung adenocarcinoma (n=6) and benign mesotheliosis (n=7). From the 1300 proteins identified, aldo-keto reductase 1B10, apolipoprotein C-I, galectin 1, myosin-VIIb, superoxide dismutase 2, tenascin C, and
thrombospondin 1 were found to be mesothelioma-specific markers and were validated using ELISA in a larger patient cohort [31]. The other study Hegmans et al. examined pleural effusions from malignant mesothelioma (n=41) and compared them to effusions taken from patients with a range of conditions including other malignancies and non-infective inflammatory exudates (n=48). They identified SMRP as a reliable marker for mesothelioma and apolipoprotein CI as a potential discriminating marker using SELDI-TOF MS [30].

Method:

Pleural effusion samples
Pleural effusions were collected from patients presenting at Flinders Medical Centre, South Australia. They were drained using thoracic thoracentesis and stored at -80°C until use. A total of 15 pleural effusion samples were selected from a series of 90 matched, clinically characterised effusion specimens based on shortest survival time. There were 5 MPM, 5 AC and 5 BR samples. All MPM samples were of the epithelioid subtype. Patient demographics are presented in Table 1. This study was approved by the Southern Adelaide Health clinical human research Ethics Committee (Approval number 381/09) and the RAH Research Ethics Committee (Approval number 111215).

Proteome purify removal of abundant proteins
A summary of all experimental methods can be seen in Figure 1. In order to remove the top 12 most abundant proteins which include albumin and immunoglobulins, 10 µL of sample were subject to sample clean up using a Protein Purify™ depletion kit, (RnD systems, Minneapolis, USA). Briefly, this involved the addition of 1.0 mL immunodepletion resin to the sample, suspension using rotary shaker for 30 minutes and centrifuged for 2 min at 1000 x g. An EZQ protein assay (Life Technologies, California, USA) was carried out pre- and post-depletion, according to the manufacturer’s recommendations. Each sample was analysed in triplicate, imaged using a Gel Doc™ EZ gel documentation system (Biorad, USA) and standard curves generation with the assistance of ImageLab (Biorad).

Sample preparation for mass spectrometry
In order to confirm equal loading of samples onto the mass spectrometer for quantitation, a secondary conformation of protein concentration was carried out. Ten µg of each sample was analysed on mini-PROTEIN® TGX Stain-free™ precast gels (Biorad, California, USA). Following activation of the gel with UV light for 5 minutes, the fluorescent intensity of each lane was measured using a Gel Doc™ EZ gel documentation system (Biorad, USA) coupled with Image Lab Software™ to ensure equal concentration of proteins per sample. Following determination of protein concentration of the pleural effusion, 5µg of lysates were diluted in
100mM ammonium bicarbonate and reduced using 50µM dithiothreitol (DTT) for 30 minutes at 65°C and alkylated using 100µM iodoacetamide (IAA) in the dark for 60 minutes. MS grade Trypsin Gold was added a ratio of 1:25 and digested overnight at 37°C.

**SWATH MS/MS analysis of samples**

Samples were analysed in two phases, a data-dependent acquisition (DDA) followed by SWATH acquisition (data-independent analysis) on the same sample using the same gradient conditions. All samples were analysed using a Triple TOF 5600+ mass spectrometry (AB SCIEX, MA, USA) fitted with a nano-LC source. For the DDA run, peptides were separated using a Polar 3 µm precoloumn (0.3 X 10 mm, SGE Analytical Science) and eluted onto a 5 µm C18column (75 mm x 150 mm with a bead pore size of 100 Å, NikkyoTechnos), using an Eksigent Ekspert 416 nanoLC. The reverse phase LC solvents include solvent A (99.9% water + 0.1% formic acid) and solvent B (99.9% acetonitrile + 0.1% formic acid). A 140 minute gradient was used at a flow rate of 0.3µL/minute with the following elution conditions: 0-85 min, 97% A; 85-92 min, 75% A; 92-97 min, 5% A; 107-110, 5% B; 110-140, 95% B. Between runs, the column was equilibrated by running 97% of solvent for 20 minutes before the next run.

Peptides were injected into the MS using PicoTip™ SilicaTip™ electrospray emitters. For DDA acquisition, experiments were set to obtain a high-resolution TOF-MS scan over a mass range of 350 – 1250 m/z, followed by 100 to 1250 m/z for MS/MS scans operating in high sensitivity mode. The selection criteria for the parent ions included the intensity of the ions had to be >200 cps, with a charge state of between 2 and 5. The dynamic exclusion was set for 30s after 2 occurrences. A 50mDa mass tolerance was set and 100 maximum candidate ions were monitored per cycle. The ion accumulation time was set to 0.057s (MS) and 0.05s (MS/MS).

Using the same conditions as described above, a SWATH acquisition was carried out using data-independent analysis (DIA). Using a 140 minute gradient as detailed above and a cycle time of 5.861 seconds, a variable isolation width including a 1Da overlap, a set of 60 windows were constructed, which covered a precursor mass range from 350 to 1250 Da. The high resolution mode was used to allow accurate extraction of the fragment ion masses. The total running time for each DDA and DIA injection was 120 minutes.

**Spectral library generation**

Data generated in the DDA run was searched against the Swiss-Prot protein database released on April 2017 using ProteinPilot (version 4.5 beta, AB SCIEX, USA) with the paragon algorithm and the following parameters: sample type – identification, alkylation - Iodoacetic acid, digestion -
trypsin, Instrument – TripleTOF5600, Special factors – nil, species – homo sapiens, ID focus with biological modifications, results quality of 0.05% confidence and false discovery rate analysis. The resulting Protein Pilot group file was used as the ion library file for all SWATH analysis and quantification.

A table of precursor masses and fragment ions was created, this was used as a reference spectral library for targeted extraction to generate ion chromatograms (XIC) of fragment ions of targeted proteins and peptides in the SWATH processing microapp contained within the PeakView® software program (AbSciex, USA). Each peptide was manually examined and peak areas were integrated from the SWATH data files. The following search parameters were used: ion library mass tolerance – 50ppm; number of peptides –4; number of transitions –6; peptide confidence –95% and shared peptides were excluded. False discovery rates were performed. The output of this search (.group file) was used as the reference spectral library containing the following information: protein name and UniProt accession, peptide sequences, Q1 and Q3 ion detection, retention time, relative intensity, precursor charge, fragment type, score, confidence and decoy result.

Swath acquisition, processing and targeted data extraction
A 5µg aliquot of the peptides were used for each injection. Three samples were selected at random and technical replicates performed (Supplementary Figure 1). Targeted data extraction and spectral alignment of the SWATH samples (DIA run) was carried out using PeakView (AB SCIEX, USA) using the spectral library generated above. All SWATH files were loaded and exported in .txt format using the following parameters: extraction window of 20 minutes, 4 peptides per protein, 6 transitions per peptide, XIC extraction window of 10 minutes and a width of 75ppm.

Data export and analysis in OneOmics and BaseSpace
In order to perform relative quantitation, the SWATH files were processed in PeakView and the output uploaded to BaseSpace (Illumina, California, USA) through the CloudConnect MicroApp in PeakView®. A three step process was carried out using the OneOmics microapps hosted in Basespace. 1) Samples were first processed using the Protein Expression Extractor Application (AB SCIEX, USA) and the following process parameters: 4 peptides per protein, 6 transitions per peptide, XIC extraction window of 10 minutes and a width of 75ppm. During this process, each .wiff file was extracted and interrogated using the uploaded spectral library. The metadata is defined in this application. 2) Protein Expression Assembler (AB SCIEX, USA) was used to select files and assign sample grouping. The data was normalized using the most likely ratio (MLR) by creating a ratio for
pairs of biological replicate within an experimental group. Using the MLR, the software makes a ratio histogram between each single experimental group. The sample ratio histogram with the centroid closest to zero is used to align the other histograms for each of the samples within that experimental group (see supplementary data 1). This is used to create a normalized peak for each of the sample, from which a MLR weight is computed for each sample (measurement of reproducibility). The normalisation protocol was then repeated between all of the experimental groups. 3) In the final step, the Protein Expression Workflow Application (AB SCIEX, USA) was used to calculate fold changes, analyse sample clustering (PCA plots), FDR and protein ontology functions and processes. Proteins were considered as ‘significant changes’ if they fulfilled the predefined criteria of differential expression which was set at p value <0.05 (Benjamini-Hochberg correction) and a fold change of greater than >1.5 in ≥2 biological replicates. All fold changes are reported as log2-fold change. For a detailed description of the statistics carried out within OneOmics see Lambert et al, 2013 [32].

Pathway analysis
Pathway analysis was carried out using Enrichr (http://amp.pharm.mssm.edu/Enrichr/enrich). Differentially expressed proteins were used as input data and three different pathway databases were interrogated, KEGG 2016, Reactome 2016 and HumanCyc 2016. Data was exported as a table, and only pathways where 2 or more target proteins had been identified, and pathways with p values <0.05 were examined further.

VEGF marker quantification using ELISA
A larger cohort of patients was examined to investigate VEGF expression in MPM compared to AC and BR (Table 2). The pleural effusion samples were centrifuged at 500 x g for 5 minutes at 25°C. The supernatant was then stored at -80°C until further use. The VEGFA ELISA was performed on pleural effusion supernatants using a human VEGFA DuoSet ELISA Development Kit (R&D Systems, Minneapolis, MN) according to manufacturer’s instructions. The plate was read at 450nm on the Versamax ELISA Microplate Reader (Molecular Devices, CA, US). Sample concentrations were calculated by comparison to the standard curve that was generated. A Kruskal-Wallis statistical analysis was used to determine whether any significant difference existed amongst the groups, followed by post-hoc Mann-Whitney-U analysis. Significance was defined as p<0.05.

Results and Discussion
Pleural effusions, similar to the serum and plasma have a wide dynamic range of protein concentrations over at least 10 orders of magnitude. This makes the identification of low
abundance proteins, such as those as the pg/mL level difficult to detect using proteomic approaches. Depletion of high abundant proteins such as albumin, transferrin, haptoglobin and immunoglobulins is an essential step to improve the resolution of less abundant proteins. Our effusion spectral library clearly shows a predominance of proteins involved in the immune response, such as immunoglobulins, suggesting that depletion was not highly effective which may also be reflected by the low number of proteins identified in the pleural effusions, which totalled 284. To date, no other spectral libraries have been published on pleural effusions making comparison of our library with other studies impossible. Our library did include well known proteins present in malignant effusions including OPN and mesothelin. In order to increase the depth of the spectral library, we combined it with other libraries generated within the laboratory which were collected from patients with haematological malignancies. This brought the total number of proteins in the spectral library to 1425 (full library details in Supplementary Table 1).

All samples were further acquired using a SWATH (DIA) approach. The first step was data inter- and intragroup normalization of all samples (Supplementary Figure 1A). One sample showed low overall counts and was subsequently reanalysed. The overall SWATH data quality including retention time, percentage of the coefficient of variation (CV) of transitions, peptides, and proteins across all biological and technical groups that matched the ion library were compared. The global expression profiles among the three groups were similar in terms of average signal intensity, which spanned more than four orders of magnitude in all three proteomes (Supplementary Figure 1B). This information indicated that a high quality data set suitable for SWATH quantification has been obtained. We performed technical replicates on three randomly selected samples from each group and the resulting PCA plot showed that the variation between technical replicates was low (Supplementary Figure 1C).

A PCA plot of the individual samples were plotted and showed that the samples clustered into their respective groups. The benign group showed more of a spread, which is expected because the patients had different underlying causes of their effusions and therefore the protein expression between patients in this group is likely to be more variable (Figure 2A). Protein changes were also plotted as a volcano plot (Figure 2B) showing the spread of up- and down-regulated proteins at different confidence levels. In order to identify differentially expressed proteins, two stringency filters can be set based on either confidence values or corrected P-values. The heat map (Figure 2C) shows the key differentially expressed proteins between each group, these proteins met the highest filter stringency of >65% confidence, reproducibility of >0.1 and proteins must have had >2 peptides identified. We
identified a large number of immunoglobulin and complement related proteins, which have previously been reported to be found at high concentrations in malignant effusions [33, 34].

When we relaxed the confidence settings, but analysed the data based on Benjamini-Hochberg corrected p-value we identified more differentially expressed proteins between the groups. We first analysed the protein differences between MPM and AC and identified 87 differentially expressed proteins (p<0.05, Benjamini-Hochberg correction applied). Comparatively, we analysed the differences between MPM and BR effusions, where we identified 62 differentially expressed proteins. A full list of all proteins can be found in Supplementary Table 2 and the key protein differences for each can be seen in Tables 3 and 4.

Table 3 shows selected upregulated and downregulated proteins for MPM samples relative to lung AC samples respectively. Protein S100 was elevated 2.5 fold in MPM samples relative to AC samples. Similarly Table 4 showed Protein S100 was 2.1 fold greater in MPM relative to BR effusions. Protein S100 is normally present in cells derived from the neural crest and fat and performs a range of regulatory functions including phosphorylation, regulation of transcription factors, calcium homeostasis, cytoskeleton changes, enzyme activities, cell growth and differentiation and inflammation. The protein S100 family has been associated with a broad range of malignancies and there is evidence to suggest it has a role in tumorigenesis and metastasis [35]. Previous studies have not indicated protein S100 has a role in malignant mesothelioma: 2012 Guidelines for MPM diagnosis using 52 lung AC and 51 MPM histological samples found S100 staining was positive in 19% of AC but completely absent in MPM samples [36]. Therefore, further research is required to assess the potential value of novel markers such as S100.

Tables 3 and 4 showed cytokeratin 5 and 6 were significantly elevated in MPM relative to AC and cytokeratin 6 to BR samples. Both of these proteins have a key role in the diagnosis of MPM [18]. Cytokeratins 5/6 have been well established as a means of differentiating between MPM and lung AC [36, 37]. A review by Husain et al. states that they are expressed in 75% to 100% of mesotheliomas, compared to 2% to 20% of lung adenocarcinomas as determined by immunohistochemical staining. The use of cytokeratin 5/6 for distinguishing MPM from lung AC is supported by Chu and Weiss 2002.[37] However, it should be noted that work by Chu and Weiss demonstrated that other malignancies including transitional cell carcinoma, endometrial adenocarcinoma and breast adenocarcinoma have high rates of positive staining and can present challenges for diagnosis when metastatic disease is present in the pleura.
Fibronectin was elevated 2.2 fold in MPM relative to lung AC. This protein has a broad range of functions including cell adhesion, growth, migration and differentiation. Previous work by Klominek et al. indicates that fibronectin is locally synthesised by malignant mesothelial cells [38], indicating this molecule could be potential marker. At 2.1 and 2.4 fold respectively MPM effusions showed elevated 14-3-3 protein epsilon and 14-3-3 protein theta compared to AC effusions. 14-3-3 protein epsilon was also found elevated compared to BR effusions at 3.0 fold. This family of proteins is ubiquitous in all eukaryotic organisms and is known to be involved in regulation of signalling pathways and 14-3-3 sigma has been strongly implicated in malignancies including breast, invasive ductal carcinoma, hepatocellular carcinoma and basal-cell carcinoma [39]. Typically the 14-3-3 proteins have an inhibitory role in the cell cycle [40]. Although the epsilon subtype has not been implicated in mesothelioma, 14-3-3 protein theta was found upregulated in malignant mesothelial lines in vitro relative to non-malignant lines suggesting a link in tumorigenesis [41].

Serum amyloid A-1 protein and serum amyloid A-2 protein were significantly upregulated in MPM compared to BR samples at 3.3 and 2.5 fold respectively. Serum amyloid proteins are a family of apolipoproteins associated with high density lipoproteins. They can be triggered by interleukin-6 which is known to be elevated with malignant mesothelioma, suggesting a link [42, 43]. The association between MPM-related inflammatory molecules and serum amyloid A-2 suggests this may have a role in prognosis. This result was partially consistent with MPM relative to AC where serum amyloid A-2 protein was significantly elevated in MPM (3.9 fold) but serum amyloid A-1 protein was higher in AC samples (1.8 fold). Galectins are well documented to play a role in MPM, with several studies identifying the importance of galectin-1 in MPM [31, 44] and the role of galectin-3 [45, 46]. Our results showed that galectin-3-binding protein was lower in MPM compared to BR. Galectin-3-binding protein is an endogenous β-galactoside-binding protein and has a role in cell growth, differentiation and malignant transformation. Studies have demonstrated a link between raised levels and MPM [47]. However, galectin-3-binding protein was elevated in MPM samples at 1.7 fold relative to AC samples.

Figure 3A shows protein pathway differences between MPM and AC samples. Interestingly, signalling pathways predominate the analysis, including EGF receptor, FGF signalling, G-protein signalling and VEGF signalling. Vascular endothelial growth factor (VEGF) promotes angiogenesis and has a well-established role in promoting MPM [48-51]. VEGF staining correlates with short survival (p = 0.0002), tumour stage (p = 0.046) and prognosis (p = 0.001) in MPM [51]. Another study by Hirayama et al. found pleural effusion VEGF levels
were greater in patients with MPM compared to lung cancer and inflammatory conditions [52]. The study also found VEGF concentrations correlated with the stage of MPM. Due to the importance of VEGF signalling in MPM, we further characterised the role of this protein in our cohort (see below).

We also noted a difference in several glycolysis proteins, like glyceraldehyde-3-phosphate dehydrogenase which catalyses the conversion of glyceraldehyde-3-phosphate to D-glyceraldehyde-1,3-bisphosphate. This enzyme has a well-established role in tumorigenesis as it protects telomeres and increases metabolic rate. Additionally, this protein has found to be elevated in MPM exosomes [53]. Another enzyme involved in glycolysis, fructose-bisphosphate aldolase was similarly found to be elevated in MPM samples relative AC samples: this is inconsistent with previous literature as this protein has been found to be moderately reduced in MPM relative to AC [54].

Figure 3B shows the bioinformatic analysis of MPM compared to BR. The predominant pathways in this analysis were integrin cell surface interactions, ephrin B reverse signalling, urokinase-type plasminogen activation (uPA) and uPAR mediated signalling. Integrins are a family of adhesion molecules and play fundamental roles in cell growth, migration, survival and differentiation and modulate signalling events initiated by growth factors [55]. Integrins are expressed in MM cell lines [56] and they mediate migration towards ECM proteins like collagen and fibronectin. The ephrin B2 receptor is overexpressed in MPM but not in benign mesothelial cells and has been suggested as a potential therapeutic target in MPM [57]. uPAR (CD87) signalling plays an important role in numerous malignancies, including solid and haematological malignancies (16, 17). uPAR is increased in asbestos-induced rat MPM models and was associated with a significantly shorter survival [58].

**VEGF expression**

We identified VEGFA at elevated levels in MPM compared BR effusions (p < 0.001) (Figure 5A) but did not identify any statistically significant difference between the MPM and AC groups (p = 0.789). This is somewhat contradictory to our proteomic results. Whilst we did not identify VEGFA in our proteomic analysis, we did identify several proteins involved in VEGF signalling (e.g. PRKCB and PRKCD) but these were only differentially expressed in AC vs MPM. VEGFA was also significantly elevated in AC compared to the BR control group (p = 0.001), suggesting VEGFA and angiogenesis may have a role in lung cancers [52]. VEGF is an endothelial cell specific mitogen and has a well-documented role in MPM. Almost all MPM cases are known to stain positive for VEGF [59] and it
appears to be a useful prognostic serum based marker for screening individuals with known asbestos-exposure [60].

**MPM Survival and VEGFA levels**

Using the median level of VEGFA in MPM samples as a cut-off (1482 pg/ml), a statistically significant difference between the survival of low and high groups was found (p = 0.003). However, Hirayama et al. determined 2000 pg/ml as the optimal cut-off value between low and high VEGFA level in MPM pleural effusion samples [52]. This value also detected a significant difference in survival between low and high VEGFA in MPM pleural effusion samples (p = 0.02). In both cases high VEGFA was associated with a markedly reduced life expectancy, confirming VEGFA as a marker for poor MPM prognosis. Figure 5B shows a Kaplan-Meier curve for low and high VEGFA MPM groups using 2000 pg/ml as the cut-off value. The median survival length for the low VEGFA group was 12 months (CI 3.93-20.07, \( n = 15 \)) whereas the median survival length for high VEGFA group was significantly reduced at 4 months (CI 1.65-6.35, \( n = 15 \)). High VEGFA levels were significantly associated with poorer survival times (p = 0.002).

**Comparison to previous studies**

Whilst SWATH MS is a powerful label-free quantitation approach, combining the advantages of high-throughput shotgun proteomics seen in traditional data-dependent acquisition (DDA) approaches with the reproducibility of selected reaction monitoring (SRM). Its data independent acquisition (DIA) approach provides a more reproducible coverage, for example, SWATH MS conferred a 15-20% increase in reproducible peptide identification and a 54% increase in the number of proteins identified in liver samples compared to traditional DDA approaches [61]. However, the method does come with limitations as retrospective interrogation of the samples is needed using spectral libraries, therefore, the identification of proteins is dependent on their initial identification in these libraries.

To date, there have been two other proteomic studies of pleural effusions in MPM [30, 31]. Using SELDI-TOF, cytokeratin 19 fragment (CYAFR 21-1) and SMRP were found to be significantly higher in MPM samples compared to non-MPM effusions, whereas ApoC1 was found at lower levels [30]. They did not identify any difference in OPN levels between MPM and non-MPM samples. In our study, we failed to identify SMRP, but we did find several differences for the cytokeratin family and in most cases found higher levels of keratin in MPM compared to AC and benign effusions. We also identified several Apo protein family
members, although all failed to meet our high confidence criteria from the Protein Expression workflow.

The most detailed study of pleural effusions in MPM was carried out by Mundt et al (2014), who used an iTRAQ based quantitation approach. They identified galectin-1 as an important diagnostic marker to distinguish between AC and MPM, with lower levels identified in MPM. They also found high levels of aldo-keto reductase and lower levels of APOC-1 confer a shorter survival. Whilst several galectin 1 peptides were identified in our initial spectral library, galectin-1 failed to be detected in the quantitation runs. This is an interesting, but not surprising observation. Jylha et al (2018)[62] found there was only a 60% overlap in proteins identified from the same samples that had been analysed using both iTRAQ and SWATH, which may explain the discrepancy noted between our study and that of Mundt et al [62]. However, we did note a higher level of aldo-keto reductase in AC vs MPM (+2.6 fold).

Conclusion:
In the era of liquid biopsies and personalised medicine approaches it is becoming clear that the ability to diagnose and monitor disease is fundamental to the success of treatments. In MPM, this is further reiterated by complexities in current diagnostic methods leading to delays in essential treatment. The identification of a predictive biomarker could be useful to monitor individuals who have had prior asbestos exposure in their profession or environmental exposure, such as Libby, Montana (USA). This study did not identify some of the previously proposed diagnostic or biomarkers, such as calretinin and hyaluronic acid in MPM samples. However, cytokeratin 5/6 was significantly elevated in MPM compared to AC and BR effusions. Serum amyloid protein-2 was found consistently elevated in MPM compared to both AC and BR indicating this peptide may have a role in prognosis and diagnosis. Additionally novel molecules protein S100, 14-3-3 epsilon, 14-3-3 theta and fibronectin were found elevated in MPM samples and are worthy of further exploration. Pathway analysis has shown there are differences in glycolysis and VEGF signalling between MPM and lung AC, and this is an important finding. The relevance of VEGF to tumour progression and ultimately prognosis in metastatic tumours is well established, including for this cohort of patients. However, clinical response to anti-VEGF therapies has been limited, and a deeper understanding of the processes involved may lead to improved therapy. However, it is important to acknowledge the limitations of this study as the small sample size could impact on the reliability and further work with more samples and high longevity patients is required. Furthermore, this study highlights that the depletion of dominant effusion proteins such as immunoglobulins and complement factors is essential to increase the sensitivity and identification of lower abundant protein species.
**Supplementary Figure 1:** a) The first step in the SWATH analysis is data normalization, which was carried out using the most likely ratio (MLR) within the Protein Extractor and Assembler applications using BaseSpace. The sample depicted in the pink line showed low counts and was re-acquired. b) The distribution of peptide signal intensities for MPM (blue line), AC (orange line) and reactive effusions (green line) plotted in a semi-logarithmic scale using OneOmics. The dynamic range plot shows the distribution of peptide intensities across the samples that have at least one confident detection (<1% FDR). c) 3 technical replicates of a randomly chosen sample from each experimental group were analysed. The tight clustering suggests that the acquisition of the replicates was reproducible.

**Supplementary Table 1: Spectral library.** Complete list of proteins isolated from pleural effusions (listed on first tab) as well as libraries previously constructed in the lab from patients with haematological malignancies (tab 2 – other spectral library). There were identified using data dependent acquisition on the TripleTOF 5600+ (AbSciex) and searched using ProteinPilot with the Pargon algorithm (for further details see methods section).

**Supplementary Table 2: Full list of all differentially expressed proteins between MPM, AC and BR effusions including Uniprot Assession ID.** Differential expression was measured as reported in the methods section and presented as signed fold change (sfc), log2 sign fold change, MLR ratios and P-values. MLR ratios and P-values were determined by the Protein Expression Assembler application hosted on the BaseSpace cloud based analysis software (Illumina). Data is presented across three tabs for the various comparisons (AC vs MPM; BR vs MPM and BR vs AC).

**Acknowledgements:**
The Authors wish to thank the Royal College of Pathologists of Australasia for their generous support of this research by a Medical Student Scholarship. This work was supported by FCIC foundation. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD010797 [63].

**References:**


**Figure 1: Summary of the methods used in this study.** Pleural effusions were drained by thoracentesis. Five samples were selected from each of the patient groups and abundant proteins were depleted from each effusion. Each of the individual samples was acquired using DDA methods and the results were analysed using ProteinPilot to generate a .group file containing the spectral library information. Further to this, an in-house generated spectral library was combined with the effusion library to increase the number of total proteins of the library to 1425 with 2203 peptides and 13,109 transitions. All 15 samples were also acquired using a DIA (SWATH). This generated 15 .wiff files containing all of the quantitative information. Using the CloudConnect uploader the .wiff files were loaded into BaseSpace (Illumina). Protein Extractor application was used to extract and integrate the peptide peaks.
using the ion spectral library to generate .qresult file. Within Protein Extractor, the experimental metadata is defined, the data is normalized using the method described in Lambert et al. [32] and the protein fold changes are computed. This is further refined in Protein Assembler to generate an .fca file. The .fca file first undergoes analytics assessment to ensure the MS data is of a high quality (Supplementary Figure 1). The data is then analysed in the OneOmics Browser (Figure 2). OneOmics uses a more sophisticated method where fold changes and confidence values are computed for each protein across the groups, P-values are then corrected (Bejamini-Hochberg correction).

**Figure 2: OneOmics data analysis.** (A) A PCA plot was carried out with each individual sample. This statistical method compares the data across multiple samples and reveal groupings among the data sets, as shown in the score plot. The biological replicates show more variance, although the MPM and AC samples do cluster into different regions on the score plot. The reactive group is more variable, likely due to the different underlying diseases represented in this group. The settings used for the PCA plot are a Weighting of None and Pareto Scaling. (B) A pseudo-volcano plot showing the degree of fold change (log2) and the confidence of the protein expressions. Each dot represents a pairwise comparison of the experimental group (AC and reactive) to the defined control group (MPM). (C) A heat map generated in OneOmics showing proteins that met the highest filter stringency of >65% confidence, reproducibility of >0.1 and proteins must have had >2 peptides identified. Log2 fold change values are depicted.

**Figure 3: Pathway analysis of protein differences analysed using Enrichr.** (A) Analysis of MPM vs AC effusions. SWATH analysis revealed 87 proteins as differentially expressed between MPM and AC effusions. Bioinformatic analysis using the Panther database shows that these proteins clustered predominately into signalling pathways, such as FGF, EGR and VEGF. (B) Pathway analysis of MPM vs BR effusions. SWATH analysis revealed 62 differentially expressed proteins between MPM and BR effusions. Bioinformatic analysis was carried out using the Reactome database. The differentially expressed proteins clustered into integrin cell surface interactions, ephrin B signalling and urokinase-type plasminogen activator (uPAR) signalling.

**Figure 4:** (A) Using an ELISA based approach, the levels of VEGFA were measured in pleural effusions isolated from BR (n = 30), MPM (n = 30) and AC (n = 30) patients. VEGFA was found at higher levels in patients with MPM and AC compared to those with BR effusions. ○ represent outliers (> 1.5x the interquartile range); ▲ represents extreme outlier (> 3x the interquartile range); * indicates p < 0.05. (B) Kaplan-Meier
survival curves for low and high VEGFA in MPM effusions. The median survival time for the low VEGFA group was 12 months (CI 3.93-20.07, n = 15), compared to 4 months for the high VEGFA group (CI 1.65-6.35, n = 15). This difference was found to be statistically significant (p = 0.002).

Table 1: Demographics of patients used in this study

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Malignant Pleural Mesothelioma</th>
<th>Lung Adenocarcinoma</th>
<th>Benign Reactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number (genders)</td>
<td>5 (3M, 2F)</td>
<td>N = 5 (3M, 2F)</td>
<td>N = 5 (3M, 2F)</td>
</tr>
<tr>
<td>Age (Range) in years</td>
<td>76 (67-77)</td>
<td>74 (64-78)</td>
<td>70 (65-77)</td>
</tr>
<tr>
<td>Mean Survival time (months)</td>
<td>3.2</td>
<td>4.2</td>
<td>16.6</td>
</tr>
</tbody>
</table>

Table 2: Patient demographics used for quantification of VEGF expression

<table>
<thead>
<tr>
<th>Total number (Males)</th>
<th>30 (26)</th>
<th>30 (25)</th>
<th>30 (26)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Range) in years</td>
<td>75 (55-94)</td>
<td>75 (58-89)</td>
<td>74 (55-94)</td>
</tr>
<tr>
<td>Subtype</td>
<td>Epithelioid (27)</td>
<td>Lung (25)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Biphasic (2)</td>
<td>Breast (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sarcomatoid (1)</td>
<td>Endometrial (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Colon (1)</td>
<td>Colon (1)</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Selected relative protein change for MPM and AC effusions. Benjamini-Hochberg corrections were applied and only proteins that were significant are reported.

<table>
<thead>
<tr>
<th>Protein Symbol</th>
<th>Protein Name</th>
<th>Gene</th>
<th>Tissue</th>
<th>Fold (log2)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>R4GN98</td>
<td>Protein S100</td>
<td>S100A6</td>
<td>MPM</td>
<td>2.5</td>
<td>5.40E-05</td>
</tr>
<tr>
<td>K2C6C</td>
<td>Keratin, type II cytoskeletal 6C</td>
<td>KRT6C</td>
<td>MPM</td>
<td>3.9</td>
<td>0.0192</td>
</tr>
<tr>
<td>SAA2</td>
<td>Serum amyloid A-2 protein</td>
<td>SAA2</td>
<td>MPM</td>
<td>3.9</td>
<td>0.0370</td>
</tr>
<tr>
<td>V9HVZ4</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>GAPDH</td>
<td>MPM</td>
<td>3.1</td>
<td>0.0350</td>
</tr>
<tr>
<td>K2C5</td>
<td>Keratin, type II cytoskeletal 5</td>
<td>KRT5</td>
<td>MPM</td>
<td>3.1</td>
<td>0.0157</td>
</tr>
<tr>
<td>1433T</td>
<td>14-3-3 protein theta</td>
<td>YWHAQ</td>
<td>MPM</td>
<td>2.4</td>
<td>0.0286</td>
</tr>
<tr>
<td>K2C6A</td>
<td>Keratin, type II cytoskeletal 6A</td>
<td>KRT6A</td>
<td>MPM</td>
<td>2.3</td>
<td>0.0250</td>
</tr>
<tr>
<td>FINC</td>
<td>Fibronectin</td>
<td>FN1</td>
<td>MPM</td>
<td>2.2</td>
<td>0.0479</td>
</tr>
<tr>
<td>1433E</td>
<td>14-3-3 protein epsilon</td>
<td>YWHAE</td>
<td>MPM</td>
<td>2.1</td>
<td>0.0184</td>
</tr>
<tr>
<td>V9HWN7</td>
<td>Fructose-bisphosphate aldolase</td>
<td>ALDOA</td>
<td>MPM</td>
<td>2.6</td>
<td>0.0131</td>
</tr>
<tr>
<td>LG3BP</td>
<td>Galectin-3-binding protein</td>
<td>LGALS3BP</td>
<td>MPM</td>
<td>1.7</td>
<td>0.0489</td>
</tr>
<tr>
<td>SAA1</td>
<td>Serum amyloid A-1 protein</td>
<td>SAA1</td>
<td>AC</td>
<td>1.8</td>
<td>0.0281</td>
</tr>
<tr>
<td>KPCB</td>
<td>Protein kinase C beta type</td>
<td>PRKCB</td>
<td>AC</td>
<td>2.4</td>
<td>5.4E-05</td>
</tr>
</tbody>
</table>
Table 4: Selected relative protein change for MPM and BR effusions. Benjamini-Hochberg corrections were applied and only proteins that were significant are reported.

<table>
<thead>
<tr>
<th>Protein Symbol</th>
<th>Protein Name</th>
<th>Gene</th>
<th>Effusion</th>
<th>Fold (log2)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>K2C6C</td>
<td>Keratin, type II cytoskeletal 6C</td>
<td>KRT6C</td>
<td>MPM</td>
<td>6.6</td>
<td>0.0167</td>
</tr>
<tr>
<td>SAA1</td>
<td>Serum amyloid A-1 protein</td>
<td>SAA1</td>
<td>MPM</td>
<td>3.3</td>
<td>0.0326</td>
</tr>
<tr>
<td>1433E</td>
<td>14-3-3 protein epsilon</td>
<td>YWHAE</td>
<td>MPM</td>
<td>3.0</td>
<td>0.00438</td>
</tr>
<tr>
<td>PPIB</td>
<td>Peptidyl-prolyl cis-trans isomerase B</td>
<td>PPIB</td>
<td>MPM</td>
<td>2.7</td>
<td>0.0257</td>
</tr>
<tr>
<td>SAA2</td>
<td>Serum amyloid A-2 protein</td>
<td>SAA2</td>
<td>MPM</td>
<td>2.5</td>
<td>0.0349</td>
</tr>
<tr>
<td>R4GN98</td>
<td>Protein S100</td>
<td>S100A6</td>
<td>MPM</td>
<td>2.1</td>
<td>0.0350</td>
</tr>
<tr>
<td>K2C6A</td>
<td>Keratin, type II cytoskeletal 6A</td>
<td>KRT6A</td>
<td>MPM</td>
<td>1.9</td>
<td>0.0228</td>
</tr>
<tr>
<td>LG3BP</td>
<td>Galectin-3-binding protein</td>
<td>LGALS3BP</td>
<td>BR</td>
<td>3.3</td>
<td>0.0426</td>
</tr>
</tbody>
</table>
Highlights:

- Key proteins upregulated in mesothelioma include S100 proteins and cytokeratin 5/6
- Serum amyloid protein-2 may have a role in prognosis and diagnosis of mesothelioma
- Proteins involved in VEGF signalling are increased in mesothelioma
- VEGFA is a marker for poor prognosis in mesothelioma
Figure 1

- Generation of Peptide Ion Spectral Library
- In house spectral library, generated from over 30 patient samples of various tissue sources (plasma, serum, peripheral blood mononuclear cells)

- Analytics
  - Supplementary Figure 1

- Browser
  - Peptide confidence > 65%
  - Reproducibility > 0.1
  - P<0.05 (Benjamini-Hochberg correction)
  - Figure 2
Pathway analysis MPM vs AC

- EGF receptor signaling pathway_Homo sapiens_P00018
- FGF signaling pathway_Homo sapiens_P00021
- Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway_Homo sapiens_P00027
- Glycolysis_Homo sapiens_P00024
- Oxytocin receptor mediated signaling pathway_Homo sapiens_P04391
- Thyrotropin-releasing hormone receptor signaling pathway_Homo sapiens_P04394
- Histamine H1 receptor mediated signaling pathway_Homo sapiens_P04385
- 5HT2 type receptor mediated signaling pathway_Homo sapiens_P04374
- Muscarinic acetylcholine receptor 1 and 3 signaling pathway_Homo sapiens_P00042
- VEGF signaling pathway_Homo sapiens_P00056

Pathway analysis MPM vs Reactive

- Beta2 integrin cell surface interaction_Homo sapiens_95b6b434-618d-11e5-8ac5-06603eb7f303
- Ephrin B reverse signaling_Homo sapiens_149a63dc-6192-11e5-8ac5-06603eb7f303
- Urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling_Homo sapiens_503076a2-6196-11e5-8ac5-
- Beta3 integrin cell surface interactions_Homo sapiens_c2800165-618d-11e5-8ac5-06603eb7f303
- Beta1 integrin cell surface interactions_Homo sapiens_2fd0bc63-618d-11e5-8ac5-06603eb7f303
- S1P5 pathway_Homo sapiens_845321c3-6195-11e5-8ac5-06603eb7f303
- Glucocorticoid receptor regulatory network_Homo sapiens_dfba0dfb-6192-11e5-8ac5-06603eb7f303
- S1P4 pathway_Homo sapiens_821b0c12-6195-11e5-8ac5-06603eb7f303
- Validated nuclear estrogen receptor beta network_Homo sapiens_5fb983a4-6196-11e5-8ac5-06603eb7f303
Figure 4

A

VEGFA protein level (pg/ml)

BR  MPM  AC

B

Cumulative Survival

Months survival after diagnosis

Low VEGFA  High VEGFA  Low-censored  High-censored
Figure 5

(A) Pre-Normalization Ratio Histograms
(B) Dynamic Range
(C) Principal Components Analysis (PCA)