Identification of bio-markers in sarcoidosis BAL using multidimensional Differential Display Proteome Analysis coupled with nano-HPLC tandem mass spectrometry

Otto F Villa
Oscar E Guevara
Avelino Teixeira
Kusuman Joseph
Paul DelaRipa
Marc Judson
Robert P Baughman
Michael Iannuzzi

Sarcoidosis is a granulomatous disease with worldwide prevalence estimated as high as 300 per million\textsuperscript{1,2}. Sarcoidosis affects all organs of the body, but the lungs and intra-thoracic lymph nodes are affected in approximately 90\% of patients accounting for the major morbidity and mortality\textsuperscript{3,4}.

The clinical progression of sarcoidosis is well documented\textsuperscript{5}. Indeed, it is known that patients who present with hilar adenopathy on chest roentgenogram as the sole abnormality and those with erythema nodosum do well two years later\textsuperscript{6,7}. In contrast sarcoidosis patients with lupus pernio only had an 20\% chance of resolution at two years\textsuperscript{7}. Although there is evidence of genetic predisposition\textsuperscript{8}, no test helps the clinician predict the outcome of sarcoidosis patients on presentation or at times of relapse. Therefore, identification of sarcoidosis biomarkers remains an unmet need.

The molecular pathogenesis of sarcoidosis remains unsettled, however it is clear that activated T cells and macrophages (mf) are key players\textsuperscript{9,10,11,12}. These cells accumulate in the tissues forming granulomas and can also be identified in the bronchoalveolar lavage (BAL).

Because proteins govern cell physiology, a comprehensive analysis of sarcoidosis BAL cell-associated proteins may identify much needed biomarkers. A comprehensive analysis of the protein equivalent of a gnome is known as “the proteome”. Therefore, we studied the sarcoidosis CD14+ derived proteome.

Proteomics workflow includes: protein separation and assessment of fold differences followed by identification of selected proteins via mass spectrometry (MS). Protein separation can be accomplished by two dimensional gel electrophoresis (2DE\textsuperscript{13}) or multidimensional, high performance liquid chromatography (HPLC) protein identification technology (MUDPIT\textsuperscript{14,15}). Under MUDPIT, peptides are produced from the entire extract followed by several in-line, modes of HPLC [cation exchanges (SCX), and reverse phase (RP)]. The last step is always an RP, so that peptides elute in organic solvent.
which is easily vaporized at the front end of mass spectrometer. Mudpit can identify several thousand proteins but does not allow up-front selection of differentially expressed proteins; hence called “shotgun proteomics”.

On the contrary 2DE produces a mathematical matrix for each experimental condition where a spot can be identified by its x-y coordinates corresponding to specific Pi and Mr. values. Spots are defined as gray pixel areas fitting a Gaussian model while non-Gaussian pixels surrounding spots are considered “background”. After background subtraction, spots are quantified using a transform of diameter and pixel intensity. For analysis two gels are overlaid at a time (termed differential display). To correct for variations in spot position due to (shift during scanning, swelling and shrinking during staining, inhomogeneities in the gel, temperature or current variations), matching vectors are created and a warping algorithm executed.

To address the sarcoidosis proteome, patients who required bronchoscopy and BAL as part of the workup for an interstitial lung disease (ILD) were recruited. The protocol was approved by IRBs, and all patients signed informed consent. BAL was performed as described. A total of 1516 spots comprised the sarcoidosis proteome, of these 383 were up-regulated over two-fold and 294 over 10 fold when compared with the proteome of the patients without ILD (Fig. 1).

Comparisons between two samples can also be performed using differential gel electrophoresis (DIGE). Under DIGE experimental and control samples are each covalently linked to a fluorochrome (Cy3 and Cy5, Amersham), mixed and resolved in a single gel, so that spots will be located at the same x-y coordinates regardless of gel distortion. However, since only three dyes exist, the number of comparisons is limited. On the other hand, using our approach, a proteome simply becomes a mathematical matrix, allowing unlimited comparisons.

This feature is very practical for our studies because the optimal control in sarcoidosis is not clear. Indeed, in an attempt to control for “normal” cells, we utilized cells from patients who presented with hemoptis but had normal bronchoscopy since this procedure is not standard of care for healthy individuals. Be-
identification of bio-markers in sarcoidosis BAL using multidimensional differential display proteome analysis coupled with nano-HPLC tandem mass spectrometry
Otto F Villa, Oscar E Guevara, Avelino Teixeira, Kusuman Joseph, Paul DelaRipa, Marc Judson, Robert P Baughman, Michael Iannuzzi

Because of this, the mφ may have been resting, thus accounting for the dissimilar proteomes. Therefore, to narrow the array of upregulated sarcoidosis spots, the proteome of mφ naturally activated in the human lung were purified from biopsy proven desquamative interstitial pneumonia (DIP, Fig. 2).

Surprisingly, the proteomes were diametrically different. However the DIP mφ becomes stimulated by cigarette smoking and never form granulomas suggesting that the activation pathways between the two diseases are different. To control for activated normal human CD14+ cells, the procedure described in Fig. 1 was used to purify CD14+ from the peripheral blood (PBL) of healthy volunteers (n=10). However, neither resting nor PMA/ionomycin stimulated PBL-derived mφ-proteomes changed the sarcoidosis CD14+-BAL array (data not shown).

However, alveolar mφ and PBL-mφ are obviously different cells and it is not difficult to differentiate sarcoidosis from a non-ILD patient. A more relevant clinical challenge is to exclude other ILDs, and cancer. Therefore proteomes from patients with biopsy proven ILD n=9 were analyzed. After 2DE, scanning, spot detection, background subtraction and spot quantification, a mathematical matrix was prepared for each condition and all spots integrated into a large matrix n=23960 spots. This matrix was next used to identify spots shared between sarcoidosis and hypersensitivity pneumonitis (HP; both drug induced and caused by animal droppings) since these diseases could be clinically and sometimes pathologically similar. Indeed a large number of sarcoidosis spots had comparable x,y coordinates and spot intensities in the HP proteomes (Fig. 3).
These shared spots may relate to molecules involved in granuloma formation and will be the focus of further analysis. To narrow this array of “granuloma” spots three patients with tuberculosis were studied but their cells did not resolve well on 2DE (data not shown). Interestingly, DIP and sarcoidosis did not have many shared molecules (Fig 3) confirming the analysis presented in Fig. 2 and reinforcing the conclusion that the DIP and sarcoidosis mφ have different activation pathways. This large mathematical matrix was next used to identify sarcoidosis spots upregulated over 3 fold over all (Fig. 4). The importance of this analysis is that, it identified 1516 spots in the sarcoidosis proteome, of which 383 were not present in non-ILD proteome but only 97 persisted after exclusion of spots present with ratios as low as 0.3 in 7 ILDs. Obviously the analysis is not complete because more proteomes from other granulomatous and disease additional ILDs should be added to fine-tune the sarcoidosis array. In conclusion, multi-dimensional differential display allows comparison of sarcoidosis spots with several ILD and granulomatous diseases thereby providing robustness to the identification of spots with potential importance as biomakers of sarcoidosis over diseases with similar clinical presentation. Protein identification of this array of spots and biological validation is the focus of a separate manuscript.

Future directions: Our current efforts are directed towards a more comprehensive proteome analysis using MudPit since we recognize that the current approach is limited to the most abundant BAL species that were resolved on 2DE-gels. Our long term goal is to identify biomarkers able to assist the clinician in the care of sarcoidosis patients.

Acknowledgements
Work supported by NHLBI-K01 grant (OV), Massachusetts Thoracic Society RT award (OV) and the Fennessey Foundation (OV). We thank Matthias Bert (Decodon.com) for computational advice.

References
Identification of bio-markers in sarcoidosis BAL using multidimensional differential display proteome analysis coupled with nano-HPLC tandem mass spectrometry

Otto F Villa, Oscar E Guevara, Avelino Teixeira, Kusuman Joseph, Paul DelaRipa, Marc Judson, Robert P Baughman, Michael Iannuzzi