SUPPLEMENTARY MATERIAL

DOI: 10.4081/monaldi.2020.1231

Proteomic characterization of idiopathic pulmonary fibrosis patients: stable *versus* acute exacerbation

Alfonso Carleo[1,2], Claudia Landi[4], Antje Prasse[1,2], Laura Bergantini[3], Miriana d'Alessandro[3], Paolo Cameli[3], Sabina Janciauskiene[2], Paola Rottoli[3], Luca Bini[4], Elena Bargagli[3]

[1]Department of Pneumology, Hannover Medical School (MHH), Hannover, Germany
[2]DZL BREATH, Hannover, Germany
[3]Unit of Respiratory Diseases and Lung Transplantation, Department of Internal and Specialist Medicine, University of Siena, Italy
[4]Functional Proteomic Section, Department of Life Sciences, University of Siena, Italy

Corresponding author: Claudia Landi, Department of Life Sciences, University of Siena, Via Aldo Moro 2, 53100 Siena, Italy. Tel. +39.057.234937. Email: landi35@unisi.it

Keywords: Idiopathic pulmonary fibrosis; acute exacerbation; bronchoalveolar lavage; proteomics; protease/antiprotease imbalance.

Sample preparation

BAL samples were filtered through sterile gauze and BAL cells count and lymphocyte phenotyping was carried out using BD Facs-Caliburflow cytometry (BD Biosciences Becton, Dickinson and Company, San Jose, CA, USA). BALF samples were separated from cellular component through centrifugation at 800x g for 5 minutes and were filtered again through filter with cut-off of 0.2 µm (Filtropur S, Sarstedt AC&Co, Numbrecht, DE). BALFs were dialyzed against purified water (18.2 MΩ) at 4°C for 24 hours, then lyophilized and dissolved in lysis buffer solution containing 8M urea, 4% w/v 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate hydrate (CHAPS), 40mM Tris base and 65mM dithioerythritol (DTE).

2D-electrophoresis

After Bradford assay, lysis buffer (and trace of bromophenol blue) was added until the protein concentrations were adjusted to 60 µg in 100 µl and 700 µg of protein in 200 µl of solution for analytical and MS-preparatory 2D-Electophoresis (2DE), respectively. The Isoelectric focusing was performed through immobiline polyacrylamide strips with 18 cm in length and immobilized non-linear pH gradient 3-10 (GE Healthcare, Uppsala, Sweden). In particular, the strips were rehydrated with a solution containing urea 8M, 4% w/v CHAPS, 1% w/v DTE for 12 hours and the samples were loaded by cup-loading method. For isoelettric focusing, the samples were added of 0,2% and 2% v/v of carrier ampholyte for analytical and MS-preparatory run, respectively. First electrophoretic dimension was carried out using Ettan™ IPGphor™ system (Amersham Biosciences) at 16°C, with the following electrical conditions: 0V for 1 hour, 30V for 8 hours, 200V for 1 hour, from 300 to 3500V in 30 minutes, 3500V for 3 hours, from 3500 to 8000V in 30 minutes, 8000V, for a total of 80,000Vh. at constant temperature of 16°C. Before the SDS-PAGE separation, strips were incubated for 12 minutes with a buffer containing 6M urea, 2% w/v Sodium Dodecyl Sulphate (SDS), 2% w/v DTE, 30% v/v glycerol and 0.05M Tris-HCl pH 6.8, and, subsequently, with the same solution where DTE was replaced by 2.5% w/v iodoacetamide and a trace of bromophenol blue for further 5 minutes. The SDS-PAGE run was performed using 9-16% SDS polyacrylamide linear gradient gels with size of 18x20x1.5cm and constant current of 40mA/gel at 9°C until the dye front reached the bottom of the gel.

The gels for MS-preparatory were stained with SYPRO Ruby (Bio-rad headquarters, Hercules, California) according to the manufacturer's instructions. Bind-silane (γ-methacryloxypropyltrimethoxysilane) (LKBProdukter AB, Brommo, Sweden) was used to attach polyacrylamide gels covalently to a glass surface for those undergoing SYPRO Ruby staining and digitized with Typhoon 9400 laser densitometer (GE Healthcare). Ammoniacal silver nitrate staining is used to stain the analytical gels that were digitalized thanks to Molecular Dynamicas 300S laser densitometer (4000 x 5000 pixels, 12 bits/pixel; Sunnyvale, CA, USA).

2D-image and statistical analysis

Spots were detected and quantified (in term of relative percentage of volume, %V) in each gel using Image Master Platinum 7.0 software (GE Healthcare). Gels were then grouped in respective class and a reference gel for each class (called Master gel) was selected. Each gel was compared with appropriate Master (Intra-class matching) and consequently the Master gels were compared each other (Inter-class matching). Spots were considered differentially abundant between two conditions when the ratio of the %V means was greater than ±2 folds. Student's T, Wilcoxon-Mann-Whitney, and False Discovery Rate (FDR) tests were performed (using $p \leq 0.05$ as threshold) by RStudio

Desktop 1.1.463 (Integrated Development for RStudio, Inc., Boston, USA, https://www.rstudio.com) in order to validate the statistical significance of comparisons between two classes.

Protein identification

After excision from MS-preparatory gels by Ettan Spot Picker (GE Healthcare), the spots of interest were destained in 5mM ammonium bicarbonate and 50% acetonitrile solution and then completely dehydrated in acetonitrile solution. The gel spots were incubated at 37°C overnight in 50mM ammonium bicarbonate and trypsin solution for protein digestion. Peptide masses were acquired using UltrafleXtreme™ MALDI-ToF/ToF (Brucher Corporation, Billerica, MA, United States), arranging each digested protein solution onto a MALDI support and embedding it with α-cyano-4-hydroxycinnamic acid (CHCA), 50% v/v ACN and 0.5% v/v TFA matrix. Protein identification was carried out by Peptide Mass Fingerprinting (PMF) approach by Mascot online tool (Matrix Science Ltd., London, UK, www.matrixscience.com), setting up Homo Sapiens (Taxonomy), Swiss-Prot/TrEMBL and NCBInr (databases), 100 ppm (mass tolerance), one acceptable missed cleavage site, carbamidomethylation - due to iodacetamide alkylation - of cysteine (fixed modification), and oxidation of methionine (possible modification) as research parameters.

In case of uncertain identification, an additional MS/MS ion search is also performed using nanoscale LC-ESI-IT-MS2 system (Phoenix40 chromatography, TheroQuestLtd., Hemel Hepstead, UK, is combined with LCQ DECA IonTrap mass mass spectrometer, Finnigan, San Jose, CA, USA). Digested peptides were separated by linear gradient reverse-phase chromatography through a C18 column (Nanoseparations, Nieuwkoop, NL). Chromatography operations and the mass spectrometer parameters for spectra acquisition were supervised by Xcalibur 1.2 software (Thermo Fische Scientific Inc, Waltham, MA, USA).

Tree clustered heatmap

Through RStudio, the unsupervised tree clustered heatmap was performed using the centered and scaled %V of differentially abundant proteins using Ward's minimum variance method and Euclidean distance measure.

3D-principal component analysis

Three-dimensional principal component analysis (3D-PCA) was carried out by RStudio using the unscaled relative %V values of differentially abundant spots. In particular variances were linearly transformed and the samples plotted into a 3-dimensional cartesian space. The pairwise confidence interval of ellipsoids around each group was defined as 0.75.

Clustering, pathway and enrichment analysis

Functional clustering analysis was carried out by DAVID 6.7 (Database for Annotation, Visualization and Integrated Discovery), (Frederick, MD, USA, david.abcc.ncifcrf.gov) using Biological Process, Cellular Component, and Molecular Function Gene Onthology (GO) terms together with REACTOME_PATHWAY, and CGAP_SAGE_QUARTILE as database. We considered only terms satisfying p-value and Benjamini tests (<0.05) and consisting of at least three members. QIAGEN's Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, CA, USA, www.qiagen.com/ingenuity) was used to perform the pathway and enrichment analysis. The set parameters were: human (species), experimentally observed or high predicted (confidence), immune

cells, vascular entothelial cells, lung epithelial cells, lung cells, fibroblast, or lung cancer cell lines (tissues/cell lines), Ingenuity Knowledge Base (reference set), direct and indirect relationship including endogenous chemicals. The network was built growing each experimentally observed protein as maximum 5 interactors, building and connecting the overall network.

Dot-plot analysis

The diagnostic criteria of IPF and AE, as well as the initial sample preparation, were as above mentioned. The Dot-Blot was carried out in duplicate, in native condition, at room temperature and 0.83µg/0.5µL of BALF protein were blotted on nitrocellulose membrane and dried in the hood for 2 hours. Primary immunodetection was carried out by 2 hours hybridization at room temperature with goat polyclonal anti-A1AT (Bethyl Laboratories, Montgomery, USA) or rabbit monoclonal anti-C36-A1AT (Cell Signaling Technologies, Danvers, USA) antibodies. Rabbit anti-goat and goat anti-rabbit IgG-HRP Conjugate (Dako Agilent Pathology Solutions, Glostrup, Denmark), as secondary antibody, was incubated for 1 hour. Immunostained bands were detected using ChemiDoc (BioRad) through chemiluminescence evaluated by ImageJ (National Institute of Mental Health, NIMH).

Table S1. Protein spots identified by MS analysis. The table reports the MS protein identification results in term of number of matched peptides, percentage of coverage and Mascot score of PMF identification as well as the sequences of peptides confirmed by MS2. The theoretical and experimentally observed molecular weights (Mw) in kDa and isoelectric point (pI) are also reported.

#	Protein name	Symbol	AC	Theoretical		Experimental		Peptide matches	% cover	Score
				pI	Mw	pI	Mw			
1	Leucine-rich alpha-2-glycoprotein	**LRG1**	P02750	6.5	38.38	4.6	45.13	10/17	29	135
2	Ig γ-1 chain C region	**IGHG1**	P01857	8.5	36.60	6.0	42.00	9/22	36	107
3	Serum albumin	**ALBU**	P02768	5.9	71.32	8.4	96.43	20/34	35	205
4	Haptoglobin	**HPT**	P00738	6.1	45.86	4.8	47.34	8/19	16	62
5	Ig γ-1 chain C region	**IGHG1**	P01857	8.5	36.60	7.8	94.44	8/19	35	99
6	Serum albumin (C-term)	**ALBU**	P02768	5.9	71.32	5.8	54.50	13/21	22	147
7	Annexin A1	**ANXA1**	P04083	6.6	38.92	6.3	37.32	15/27	52	184
8	Serum albumin (C-term)	**ALBU**	P02768	5.9	71.32	6.3	39.24	7/14	9	69
9	α-1-antitrypsin (C-term)	**A1AT**	P01009	5.4	46.88	7.1	18.64	8/12	20	91
10	Serum albumin	**ALBU**	P02768	5.9	71.32	5.4	43.41	20/32	33	201
11	Serum albumin (C-term)	**ALBU**	P02768	5.9	71.32	5.4	34.20	9/28	18	68
12	Serum albumin (C-term)	**ALBU**	P02768	5.9	71.32	5.4	34.65	8/9	16	110
13	Serum albumin (C-term)	**ALBU**	P02768	5.9	71.32	6.8	37.50	9/20	19	91
14	Protein S100-A9	**S100A9**	P06702	5.7	13.29	5.6	11.81	5/14	46	72
15	α-1-antitrypsin	**A1AT**	P01009	5.4	46.88	4.8	49.82	9/18	31	109
16	Serum albumin	**ALBU**	P02768	5.9	71.32	5.0	51.02	21/36	27	210
17	Ig γ-1 chain C region	**IGHG1**	P01857	8.5	36.60	7.6	98.45	9/22	37	108
18	Ig κ chain C region	**IGKC**	P01834	5.6	11.77	7.6	28.76	.DIQMTQSPSSLSASVGDR.V + Ox(M) K.SGTASVVCLLNNFYPR.E + Dea(NQ) .TVAAPSVFIFPPSDEQLK.S K.VYACEVTHQGLSSPVTK.S K.VYACEVTHQGLSSPVTK.S + Dea(NQ) K.VDNALQSGNSQESVTEQDSK.D		
19	Protein S100-A9	**S100A9**	P06702	5.7	13.29	5.5	10.92	11/26	84	129
20	Ig γ-1 chain C region	**IGHG1**	P0185	8.5	36.60	7.3	98.45	8/21	32	94

			7							
21	Fructose-bisphosphate aldolase A	**ALDOA**	P04075	8.3	39.85	8.3	38.61	12/23	56	161
22	Ig γ-1 chain C region	**IGHG1**	P01857	8.5	36.60	7.6	93.46	8/17	32	103
23	Serum albumin (N-term)	**ALBU**	P02768	5.9	71.32	5.0	6.54	7/17	11	59
24	Ig γ-1 chain C region	**IGHG1**	P01857	8.5	36.60	7.6	39.43	8/15	35	109
25	Ig γ-1 chain C region	**IGHG1**	P01857	8.5	36.60	7.4	100.45	6/20	23	64
26	Serum albumin	**ALBU**	P02768	5.9	71.32	5.6	40.93	19/39	35	178
27	Serum albumin	**ALBU**	P02768	5.9	71.32	6.0	52.10	13/22	24	135
28	Ig γ-1 chain C region	**IGHG1**	P01857	8.5	36.60	7.4	94.93	7/23	31	76
29	Serum albumin	**ALBU**	P02768	5.9	71.32	6.0	51.95	8/20	17	70
30	Ig γ-1 chain C region	**IGHG1**	P01857	8.5	36.60	7.3	93.95	8/32	32	78
31	α-1-antitrypsin	**A1AT**	P01009	5.4	46.88	4.9	49.96	20/39	50	211
32	Haptoglobin	**HPT**	P00738	6.1	45.86	5.0	41.73	17/44	33	134
33	Serotransferrin	**TRFE**	P02787	6.8	79.29	6.3	60.34	17/42	27	132
34	Serum albumin	**ALBU**	P02768	5.9	71.32	6.0	51.95	14/34	27	125
35	Ig α-1 chain C region	**IGHA1**	P01876	6.1	38.49	6.4	59.80	7/20	28	78
36	α-1-antitrypsin	**A1AT**	P01009	5.4	46.88	4.9	48.64	8/9	24	129
37	Polymeric immunoglobulin receptor	**PIGR**	P01833	5.6	84.43	5.1	90.12	18/48	26	142
38	Protein S100-A8	**S100A8**	P05109	6.5	10.89	5.9	52.10	5/7	48	97
39	Selenium-binding protein 1	**SBP1**	Q13228	5.9	52.93	5.9	50.72	16/36	39	168
40	Serum albumin	**ALBU**	P02768	5.9	71.32	5.2	40.34	39/62	56	393
41	α-1-antitrypsin	**A1AT**	P01009	5.4	46.88	4.9	52.89	18/39	48	179
42	Ceruloplasmin	**CERU**	P00450	5.4	122.98	5.6	131.17	11/24	13	87
43	Serotransferrin	**TRFE**	P02787	6.8	79.29	6.1	80.80	36/60	43	347
44	Polymeric immunoglobulin receptor	**PIGR**	P01833	5.6	84.43	5.1	86.45	19/31	31	198
45	α-1B-glycoprotein	**A1BG**	P04217	5.6	54.79	5.1	73.58	13/36	37	123
46	Pigment epithelium-derived factor	**PEDF**	P36955	6.0	46.46	5.6	44.86	14/27	38	167
47	Complement C3 (C-term)	**C3**	P01024	6.0	188.57	4.8	40.47	16/25	9	99

48	Serum albumin	**ALBU**	P02768	5.9	71.32	5.6	89.65	24/47	39	219
49	Ig γ-1 chain C region	**IGHG1**	P01857	8.5	36.60	7.3	37.20	9/16	36	123
50	Serum albumin (C-term)	**ALBU**	P02768	5.9	71.32	5.6	48.20	20/31	37	211
51	Serum albumin	**ALBU**	P02768	5.9	71.32	5.8	88.27	28/48	46	267

Table S2. DAVID Functional clustering analysis according 5 different Gene Ontology (GO) Categories. The table reports the significant GO terms with Fold Enrichment (FE), p-value of hypergeometric distribution test, the Benjamini-Hochberg procedure values (FDR), and the matched genes.

GO Term	FE	p-value	FDR	Genes
BIOLOGICAL PROCESS				
Innate immune response	13.0	5.16E-05	2.41E-03	*IGKC, S10A9, ANXA1, IGHG1, IGHA1, S10A8*
Retina homeostasis	116.6	6.42E-08	1.20E-05	*IGKC, TRFE, IGHA1, ALBU, PIGR*
Platelet degranulation	45.3	2.99E-06	2.79E-04	*TRFE, ALBU, ALDOA, A1AT, A1BG*
Defense response to bacterium	32.2	1.16E-05	7.25E-04	*IGKC, S10A9, IGHG1, HPT, S10A8*
Complement activation, classical pathway	37.7	1.27E-04	4.75E-03	*IGKC, IGHG1, IGHA1, CO3*
Receptor-mediated endocytosis	20.1	8.11E-04	1.67E-02	*IGKC, IGHA1, ALBU, HPT*
Positive regulation of B cell activation	107.6	3.09E-04	9.59E-03	*IGKC, IGHG1, IGHA1*
Phagocytosis, recognition	100.0	3.59E-04	9.55E-03	*IGKC, IGHG1, IGHA1*
Phagocytosis, engulfment	80.0	5.63E-04	1.31E-02	*IGKC, IGHG1, IGHA1*
B cell receptor signaling pathway	51.8	1.34E-03	2.47E-02	*IGKC, IGHG1, IGHA1*
Cellular oxidant detoxification	40.0	2.24E-03	3.43E-02	*S10A9, ALBU, HPT*
Complement activation	32.2	3.43E-03	4.49E-02	*IGKC, IGHG1, CO3*
CELLULAR COMPONENT				
Extracellular space	13.5	5.34E-20	4.00E-18	*A2GL, S10A9, TRFE, ANXA1, IGHA1, ALBU, ALDOA, S10A8, A1BG, IGKC, PEDF, IGHG1, A1AT, HPT, CERU, CO3, PIGR, SBP1*
Extracellular exosome	6.5	1.52E-14	5.70E-13	*A2GL, S10A9, TRFE, ANXA1, IGHA1, ALBU, ALDOA, S10A8, A1BG, IGKC, PEDF, IGHG1, A1AT, HPT, CERU, CO3, PIGR, SBP1*
Extracellular region	10.1	1.68E-14	4.22E-13	*A2GL, S10A9, TRFE, ANXA1, IGHA1, ALBU, ALDOA, S10A8, A1BG, IGKC, PEDF, IGHG1, A1AT, HPT, CERU, CO3*
Blood microparticle	59.9	4.44E-13	8.32E-12	*IGKC, TRFE, IGHG1, IGHA1, ALBU, HPT, CERU, A1BG, CO3*
Platelet alpha granule lumen	73.6	1.72E-05	2.57E-04	*ALBU, ALDOA, A1AT, A1BG*
MOLECULAR FUNCTION				
Antioxidant activity	149.0	1.58E-04	8.83E-03	*S10A9, ALBU, HPT*
Immunoglobulin receptor binding	114.6	2.70E-04	7.54E-03	*IGKC, IGHG1, IGHA1*
Serine-type endopeptidase activity	15.6	1.65E-03	3.03E-02	*IGKC, IGHG1, HPT, CO3*
Antigen binding	28.9	4.18E-03	4.59E-02	*IGKC, IGHG1, IGHA1*
PATHWAY				
Platelet degranulation	29.1	1.23E-05	4.42E-04	*TRFE, ALBU, ALDOA, A1AT, A1BG*
Scavenging of heme from plasma	43.8	6.65E-05	1.20E-03	*IGKC, IGHA1, ALBU, HPT*
TISSUE EXPRESSION				
Lung poor differentiated adenocarcinoma with lymphoplasmatic infiltration	6.6	8.64E-06	2.35E-03	*IGKC, S10A9, PEDF, IGHG1, ALDOA, A1AT, HPT, CO3, PIGR*
Liver normal bulk liver	6.7	2.03E-04	2.72E-02	*TRFE, ALBU, A1AT, HPT, S10A8, A1BG, CO3*