

Deep proteome profiling reveals common prevalence of MZB1-positive plasma B cells in human lung and skin fibrosis

Herbert B. Schiller^{1, 2*}, Christoph H. Mayr¹, Gabriela Leuschner^{1,5}, Maximilian Strunz¹, Claudia Staab-Weijnitz^{1,2}, Stefan Preisendörfer¹, Beate Eckes³, Pia Moinzadeh³, Thomas Krieg³, David A. Schwartz⁷, Rudolf A. Hatz⁴, Jürgen Behr^{5,2}, Matthias Mann⁶, Oliver Eickelberg^{1,2,7 *}

* corresponding authors

- 1 Comprehensive Pneumology Center, Helmholtz Zentrum München, Munich, Germany
- 2 German Center for Lung Research (DZL)
- 3 Department of Dermatology, University of Cologne, Cologne, Germany
- 4 Center for Thoracic Surgery, Munich Lung Transplant Group, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany
- 5 Department of Internal Medicine V, University Hospital Grosshadern, Ludwig-Maximilians-University, and Asklepios Fachkliniken München-Gauting, Comprehensive Pneumology Center, Munich, Germany.
- 6 Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany
- 7 Division of Respiratory Sciences and Critical Care Medicine, Department of Medicine, University of Colorado, Denver, CO, USA

Author contributions

HBS and OE initiated, conceptualized and designed the study; HBS acquired and analyzed the proteomics data and wrote the paper; CHM, GL, MS and CSW performed immunofluorescence and immunoblot analysis of tissue samples; SP performed MZB1 analysis on plasma cell differentiation; GL evaluated clinical patient data; BE, PM, TK, RAH, JB, DAS, and MM provided patient material and/or important analytical tools. All authors read and approved the final version of the manuscript.

Support

This work was supported by the German Center for Lung Research (DZL), the Max Planck Society for the Advancement of Science, by the German Research Foundation (DFG) through a grant to TK (KR558), and the Helmholtz Association.

Running title

Proteomic profiling of human lung and skin fibrosis

Word count

3912

Online Data Supplement

This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org

At a Glance

Scientific Knowledge on the Subject

Organ fibrosis is a major clinical problem with limited to no therapeutic options, depending on organ manifestation. Fibrosis can occur as a result of persistent tissue injury and inflammation, impaired regeneration or repair pathways, distorted proteostasis during e.g. aging, or auto-immunity. It is unclear, whether organ-specific fibrotic diseases, such as idiopathic pulmonary fibrosis (IPF), have a common underlying pathophysiology compared with other fibrotic syndromes, or whether tissue-specific mechanisms of fibrosis exist that allow targeted therapeutic intervention.

What This Study Adds to the Field

While the analysis of tissue fibrosis has mainly relied on gene expression data to date, full-scale quantitative proteome approaches to fibrosis are limited. It is well documented that changes in protein abundance are not necessarily reflected at the messenger RNA level, and novel therapeutic compounds largely act on proteins. We provide the most comprehensive proteomic resource of human tissue fibrosis, containing information about the abundance, stoichiometry, and detergent solubility of proteins. We identified common and distinct features of lung fibrosis, in comparison with skin fibrosis of patients with localized scleroderma. The most significant commonality of different interstitial lung diseases and skin fibrosis was the prevalent occurrence of MZB1+ plasma B cells, which points to a common involvement of antibody-mediated autoimmunity in at least two forms of tissue fibrosis.

Abstract

Rationale: Analyzing the molecular heterogeneity of different forms of organ fibrosis may reveal common and specific factors and thus identify potential future therapeutic targets.

Objectives: We sought to use proteome-wide profiling of human tissue fibrosis to (1) identify common and specific signatures across endstage interstitial lung disease (ILD) cases, (2) characterize ILD subgroups in an unbiased fashion, and (3) identify common and specific features of lung and skin fibrosis.

Methods: We collected samples of ILD tissue (n=45) and healthy donor controls (n=10), as well as fibrotic skin lesions from localized scleroderma and uninvolved skin (n=6). Samples were profiled by quantitative label-free mass spectrometry, Western blotting, or confocal imaging.

Measurements and Main Results: We determined the abundance of >7900 proteins and stratified these proteins according to their detergent solubility profiles. Common protein regulations across all ILD cases, as well as distinct ILD subsets, were observed. Proteome comparison of lung and skin fibrosis identified a common upregulation of MZB1, the expression of which identified MZB1+/CD38+/CD138+/CD27+/CD45-/CD20- plasma B cells in fibrotic lung and skin tissue. MZB1 levels correlated positively with tissue IgG, and negatively with diffusing capacity of the lung for carbon monoxide (DLCO).

Conclusions: Despite the presumably high molecular and cellular heterogeneity of ILD, common protein regulations are observed, even across organ boundaries. The surprisingly high prevalence of MZB1+ plasma B cells in tissue fibrosis warrants future investigations regarding the causative role of antibody-mediated autoimmunity in idiopathic cases of organ fibrosis, such as idiopathic pulmonary fibrosis (IPF).

Keywords: Fibrosis, Proteomics, ILD, localized scleroderma (Morphea), MZB1, Autoimmunity

Introduction

Tissue fibrosis is a major health burden, accounting for about 45% of deaths in the developed world, both directly and indirectly [1]. Replacement of normal tissue architecture by extracellular matrix (ECM)-rich scar tissue in fibrosis impedes organ functionality and regeneration after injury. More than 200 different chronic lung disorders are characterized by lung fibrosis. Many of these ILD exhibit poor prognosis, such as Idiopathic Pulmonary Fibrosis (IPF), with a median survival time of 3-5 years after diagnosis [2]. In localized scleroderma (morphea), an autoimmune-mediated chronic inflammation leads to severe fibrotic plaques restricted to the skin [3]. This disease represents a particularly good model system to study fibrotic reactions, as involved and uninvolved areas can be directly identified and compared in the same patient.

In many cases, the true origin and cause of fibrosis remains unknown. Possible causes for idiopathic fibrosis discussed today include persistent tissue injury or inflammation, impaired tissue regeneration or repair, distorted proteostasis during e.g. aging, or autoimmunity [4]. While ILD caused by autoimmunity is well known in connective tissue disease (CTD) [5], the involvement of autoimmunity in IPF has also been discussed, due to the presence of circulating immune complexes [6]. However, definitive evidence still remains limited due to a lack of specific diagnostic tests. Recent experimental evidence shows that autoimmunity to a lung-specific autoantigen can drive pulmonary fibrosis [7], suggesting that the presence of other unidentified autoantigens may drive IPF. Furthermore, impaired regeneration and subsequent fibrosis upon injury is associated with dysregulated developmental pathways, such as the Wnt-, Bmp/TGFbeta- or sonic hedgehog (Shh) signaling pathways [8-11]. The interactions of secreted morphogens of these pathways with the ECM affect its function as an “instructing niche” [12], which motivates the recently growing interest in ECM structure and function [13].

The tissue- and disease-specific composition of the ECM-proteome (matrisome) *in vivo*, as well as its specific architecture and dynamic association with secreted proteins is still largely unexplored, due to challenging technical limitations. We recently developed a quantitative detergent solubility profiling (QDSP) method, which largely improved the *in depth* analysis of tissue proteomes and matrisomes [14]. In this study, we used the QDSP method to characterize human tissue proteomes from lung and skin fibrosis to identify common and distinct molecular alterations in cases of ILD and morphea. We provide a

comprehensive resource of protein regulation in human tissue fibrosis and describe a surprisingly high prevalence of MZB1-positive plasma B cells in IPF.

Methods

Human patient material

Resected human lung tissue and explant material was obtained from the bioarchive at the Comprehensive Pneumology Center (CPC) in Munich. Biopsies were obtained from 10 healthy donors and 45 endstage ILD patients (see supplementary Table S3 and S7 in the online data supplement for clinical baseline characteristics). Segments of the resected fresh frozen lung tissue that were histologically characterized with fibrosis were used for the proteome analysis. All participants gave written informed consent and the study was approved by the local ethics committee of Ludwig-Maximilians University of Munich, Germany (333-10). Skin biopsies were taken from 6 patients (3 females, 3 male; mean age: 66) with localized scleroderma (morphea). From each patient, one biopsy was obtained from an involved area that was clinically characterized by sclerosis and inflammation, and another one from a distant, clinically uninvolved site. Samples were immediately snap-frozen in liquid nitrogen. All patients gave written informed consent. The study was approved by the local ethics committee at University Hospital of Cologne, Germany (08144). Human lung tissue derived proteins for the UC Denver cohort were obtained from the National Jewish Health-Interstitial Lung Disease Program, including IPF (n=4) and non-fibrotic control (n=5) samples. Control tissue was obtained from transplant specimens that failed regional lung selection (NJH). The diagnosis of IPF was determined by a pathology core consisting of 2 pulmonary pathologists, a radiology core consisting of 3 pulmonary radiologists, and a clinical core consisting of 5 pulmonary physicians. All diagnoses were made in accordance with established criteria. The Institutional Review Board (IRB) at National Jewish Health approved the collection and the use of tissue.

Online data supplement

This article has an online data supplement, which contains a detailed description of all experimental methods and xlsx tables to all proteomics experiments, accessible from this issue's table of content online at www.atsjournals.org

Results

Quantitative detergent solubility profiling of human fibrotic lung and skin

We used mass spectrometry to analyze human tissue fibrosis biopsies. Segments of the resected lung and skin tissue were histologically analyzed to confirm fibrosis in this region and then used for the proteome analysis. From each sample the proteins were extracted with increasing stringency into four fractions by changing the detergent and buffer conditions as described in the QDSP protocol [14]. We then subjected each protein fraction individually to our shotgun proteomics analysis pipeline, using a four hour gradient measurement on a Quadrupole/Orbitrap mass spectrometer (Q-Exactive) and subsequent label free protein quantification and data analysis with the MaxQuant [15] and Perseus [16] software packages, as well as custom built analysis scripts (Figure 1A). We quantified 7907 proteins in the ILD analysis (Table S1 in the online data supplement) and 5826 proteins in the analysis of the morphea biopsies (Table S2 in the online data supplement). The QDSP method adds an additional dimension to the tissue proteome by separating proteins by their detergent solubility. As expected we observed a significant separation of cytoplasmic, membrane, nuclear and ECM proteins, with ECM proteins being most insoluble (Figure 1B). This analysis is particularly interesting for secreted proteins, which might stay soluble upon secretion, or become incorporated into the ECM, which renders them insoluble. We used Uniprot keywords and the Matrisome annotations [17] to identify 550 proteins in our dataset that were previously annotated to be secreted by cells, and performed a principal component analysis (PCA) with this subset of the data. This analysis efficiently separated the four protein fractions in component one, which accounted for 38.8% of the data variability, and separated healthy donor controls from endstage ILD in component four, which accounted for 4.4% of the data variability (Figure 1C). Principal component four was significantly enriched for the gene categories 'antimicrobial' and 'innate immunity', which were higher in the healthy controls and 'proteoglycans' and 'extracellular matrix', which were higher in the ILD proteomes. A scatter plot of the loadings of the PCA revealed the position of individual proteins in the data space (Figure 1D).

We next determined the total abundance of proteins in the tissue biopsies by summing up the MS-intensities of the four individual protein fractions. We performed a t-test to compare ILD and donor lung tissue proteomes (Figure 2A), as well as the skin lesions

from patients with localized scleroderma with the respective healthy skin from the same patient (Figure 2B). In order to identify common factors in different forms of ILD, we began our proteomic investigation with a heterogeneous group of patients (see Table S3 in the online data supplement for clinical baseline characteristics). Irrespective of the expected heterogeneity of the patient biopsies we observed significant alterations in both ILD and localized scleroderma to the respective controls. At a false discovery rate of 10%, 44 proteins were regulated in the ILD cohort (Figure 2A). Hierarchical clustering analysis (Pearson correlation) of these 44 proteins sorted patients by diagnostic classes (Figure 2C). The most significant common factor in all forms of ILD analyzed was MMP19, which was previously shown to be upregulated in pulmonary fibrosis in both mice and humans [18]. MMP19 was mostly enriched in the detergent insoluble fraction, indicating its association with the ECM upon secretion (Figure 2D). We also found common upregulation of the collagen chaperone FKBP10 that we previously identified to be upregulated in the bleomycin model of lung fibrosis and IPF [19]. Furthermore, we also observed the increased expression of the prolyl 3-hydroxylase 1 protein (Lepre1), which is involved in collagen hydroxylation [20, 21] and thus may serve the increased production of collagen in fibrotic tissue. We confirmed the upregulation of KRT17 and SDF4 using Western blot analysis of IPF samples from an independent US cohort (Figure S1A, B).

Finally, we compared our ILD proteome dataset with the currently (to our knowledge) largest available transcriptomic dataset of human ILD (n=194) and control tissue (n=91) (Gene Expression Omnibus dataset GSE47460), published by the Lung Tissue Research Consortium, and identified many proteins that are both regulated on the RNA and proteome level, including the proteins KRT17 and MZB1 (Figure S1C). In this comparison, the Pearson correlation of protein and mRNA copy numbers were weak (Figure S1D, E), confirming the known fact that protein and mRNA abundances do not always correlate well, even in matched samples. Of note, some of the proteins we identified as upregulated in ILD by mass spectrometry, such as LEPRE1 and MMP19, were not found increased in total ILD tissue mRNA abundance by microarray analysis (Figure S1C).

In localized scleroderma, 1 protein (LTBP2) was detected at <1% false discovery rate (FDR), 10 proteins were at <5% FDR and a total of 27 proteins were at <20% FDR (Figure 2B). One of the most upregulated proteins in the fibrotic skin lesions was the cartilage oligomeric matrix protein (COMP) (Figure 2B), which we previously showed to be increased in skin

fibrosis to regulate dermal collagen ultrastructure [22, 23] and collagen secretion [24]. The analysis identified several interesting proteins that are not well studied in the context of fibrosis, including LTBP2 and CPXM2. Interestingly, LTBP2 was recently shown to bind FGF-2 in hypertrophic scars, thereby blocking cell proliferation [25, 26].

In summary, we provide a comprehensive biochemical characterization of the ECM proteome in human lung and skin fibrosis and identified previously known, as well as novel alterations in protein abundance.

Molecular heterogeneity of ILD tissue proteomes

ILD pathophysiology can be highly heterogeneous. Thus, it was conceivable that we would encounter a large variability between patients, even though all 11 lung biopsies were taken from diseased areas that underwent fibrotic remodeling and showed uniform upregulation of fibrosis markers such as MMP19 (Figure 2). In order to identify the proteins with highest differences between ILD samples we calculated the coefficient of variation (CV) for each protein and plotted it against the protein abundance rank (Figure 3A). We identified 133 proteins with a high coefficient of variation between patient samples, which were quantified in at least 5 out of 11 ILD biopsies (Figure 3A). To reveal gene categories that show high variation between patients, we performed two dimensional annotation enrichment analyses [27] for protein abundance ranks versus coefficient of variation ranks. We also calculated the enrichment score of the CV quantiles, showing that there is a mild increase in data dispersion with decreasing abundance (Figure 3B; Table S4 in the online data supplement). Interestingly, the upper 20% quantile (Q1) with highest coefficient of variation breaks the trend and shows a slightly higher abundance rank than Q2. This indicates that many highly abundant proteins also showed high data dispersion. In the upper right quadrant of the 2D annotation enrichment plot, the gene categories enriched within highly abundant proteins with high coefficient of variation (including ECM proteins, antibodies and antimicrobial peptides) are depicted (Figure 3B).

Interestingly, many of the high CV proteins are cell-type specific genes, such as BPIFB1, MUC5B (goblet cells), AGER (type 1 pneumocytes), KRT5, or KRT14 (basal cells), indicating that we do observe differences in cellular composition between samples. The

average levels of AGER were only mildly reduced in ILD samples compared with donor lungs, however, a few patient samples showed drastic changes with at least 60-fold reduced protein levels, explaining its high CV. Interestingly, a subset of three ILD patient biopsies showed at least 60-fold increased levels of the mucin-5B (MUC5B), which is normally expressed by goblet cells in the bronchi. The same three patient samples were strongly enriched for matrilysin (MMP7), which was shown to be a key regulator of pulmonary fibrosis in mice and humans [28], and one of the most upregulated genes in microarray studies of IPF [29]. Interestingly, these samples did not display significantly different levels of neutrophil defensin 3 (DEFA3), which was shown to be a marker of acute exacerbations of IPF [29], compared with donor lungs (Figure 3C).

We next determined the Pearson correlation coefficients between the MS-intensity profiles of the 133 proteins (Table S5 in the online data supplement) to group proteins by similarity. Unsupervised hierarchical clustering (Pearson correlation) of these correlation coefficients revealed three main groups of proteins, which were anti-correlated (Figure 3C). In group 1, we detected markers for type one (AGER) and type two (SFTPC) pneumocytes, while in group 2, we found markers of lung fibrosis, such as MMP7 [28], as well as the basal stem cell markers KRT5 and KRT14 [30]. The third distinct group showed higher correlation with group 2 compared to group 1 and contained proteins with functions in innate immune defense (DEFA3, ELANE), as well as immunoregulatory proteins (CXCL13) (Figure 3D).

To visualize patient heterogeneity, we used principal component analysis (PCA) of 1037 proteins representing the upper 20% quantile (Q1) of the CV (Figure 4A), and selected 2 ILD subgroups that were characterized by a distinct protein profile compared with healthy donor controls (Figure 4B and C). Clinically, the ILD group1 (1 IPF and 2 HP) had a lower DLCO than group2 (three unclassifiable ILDs) (see Table S3). To determine which proteins in these patient subsets were significantly different compared with healthy donor controls, we used a two sided t-test, which produced 272 significantly regulated proteins (FDR < 10%) in group 1 and 262 significantly regulated proteins (FDR < 10%) in group 2 (Figure 4D and E, and Table S1 in the online data supplement).

In summary, the application of unsupervised exploratory statistics on the proteome data uncovered correlated groups of proteins and enabled the stratification of ILD patients, revealing patient groups with distinct protein composition.

Enrichment of MZB1+ tissue resident plasma B cells is a highly prevalent feature of lung and skin fibrosis

We matched the two proteomic datasets and compared fold changes in ILD and localized scleroderma biopsies. To reveal common and distinct gene categories in lung and skin fibrosis, we first performed 2D annotation enrichment analysis [27] (Table S6 in the online data supplement). We observed common upregulation of extracellular matrix genes, complement activation, N-glycan biosynthesis, plasma-lipoprotein particles and most significantly, we found a common increase in the abundance of antibodies (Figure 5A, B). Comparing the protein outliers with the highest fold changes, we identified a number of interesting differences as well as similarities between both datasets. For instance, the ECM protein Tenascin-C (TNC), which is known to be increased in IPF [31], was upregulated in both datasets. Surprisingly, the most significant similarity with highest fold changes in both lung and skin fibrosis was an upregulation of the Marginal zone B- and B1-cell-specific protein (MZB1) (Figure 5B), which is known to be expressed in certain B-cell subsets to diversify peripheral B cell functions by regulating Ca(2+) stores, antibody secretion, and integrin activation [32]. We validated this finding by staining MZB1 in tissue sections from ILD and scleroderma (Figure 5C and S3 – S7). MZB1 localized to cells with a considerable volume of cytoplasm that were found in higher numbers in fibrotic tissue compared with controls and were localized typically in perivascular regions.

MZB1 is an ER resident protein, which is important for antibody secretion and is thus upregulated in cells with high antibody secretory activity [33]. We confirmed this finding by treating human peripheral blood mononuclear cells (PBMC) with interleukin-2 (IL2) and the TLR7/8 ligand R848, which induces differentiation of memory B cells to Ig-secreting cells [34], and analyzing MZB1 expression. IL2/R848 treatment induced IgG production and expression of BLIMP1, a transcription factor essential for plasma cell function [35]. Indeed, expression of MZB1 on both transcript and protein levels was drastically increased under these conditions (Figure S2), confirming its specific expression in antibody secreting cells.

To establish MZB1 as marker for antibody-secreting plasma B cells in human ILD tissue, we performed co-immunostainings with several lineage markers. The MZB1+ cells were negative for the T-lymphocyte lineage marker CD3 (Figure S3A, S6A), and also completely negative for the B-cell lineage marker CD20 (Figure S3B, S6B) or the leukocyte lineage marker CD45 (Figure S4A, S7A). Co-expression of MZB1 with CD38 (Figure 5C), CD138

(Figure S4B), and CD27 (Figure S5A), clearly identified the MZB1+ cells as terminally differentiated plasma B cells in ILD tissues [36]. Similarly, we also found MZB1+/CD38+ double positive cells that were CD20/CD45 negative (Figure S6 and S7) in the skin. Furthermore, consistent with the notion that we identified tissue resident plasma B cells, we also found positive staining of MZB1+ cells with an antibody against human IgG (Figure S5B).

Finally, to increase the overall number of samples and validate our findings in an independent cohort, we performed Western blot analysis of 34 additional ILD tissues (IPF n=14, HP n=7, CTD-ILD n=2, NSIP n=3, unclassifiable ILD n=12, other ILD n=3) and 7 healthy donor controls (see Table S7 in the online data supplement for clinical baseline characteristics). We found MZB1 significantly increased in both IPF and non-IPF ILD compared with healthy donor tissue (Figure 6A and 6B). We also re-confirmed our finding by Western blotting of MZB1 protein in 3 additional localized scleroderma patients (Figure 6C). Importantly, the quantification of total tissue IgG and MZB1 levels showed a significant positive correlation, again indicating that MZB1 amounts are predictive for local antibody secretion (Figure 6D). Increased abundance of MZB1 transcripts in ILD compared with healthy donor controls and COPD cases were also found in an independent large US cohort microarray study (Gene Expression Omnibus dataset GSE47460) (Figure 6E). Of note, the same dataset shows increased abundance of CD38 in ILD tissues (Figure S1), thus providing an independent confirmation for the prevalence of plasma B cells. MZB1 levels were independent of age, vital capacity, gender, treatment with steroids or antifibrotics (Figure S8), but showed a significant negative correlation with DLCO (%) in both cohorts analyzed (Figure 6F and 6G).

In summary, the unbiased proteomic analysis of human lung and skin fibrosis uncovered a surprising prevalence of MZB1+/CD38+/CD138+/CD27+/CD20-/CD45- plasma B cells, which is an indication for a common involvement of antibody-mediated autoimmunity in idiopathic organ fibrosis.

Discussion

Mass spectrometry driven proteomics evolved into a highly sensitive and accurate technology that enables the precise quantification of thousands of proteins at once [37]. In this study, we used the recently developed QDSP method to analyze tissue biopsies of human lung and skin fibrosis at a depth of >7900 proteins quantified. We provide the most comprehensive proteomic resource of human tissue fibrosis, containing information about the abundance, stoichiometry, and detergent solubility of proteins, and the first cross-organ comparison of tissue fibrosis. Profiling lung biopsies from a heterogeneous cohort of human ILD together with skin biopsies from patients with localized scleroderma (morphea) enabled the identification of common and distinct protein regulation in various forms and stages of fibrotic remodeling.

Proteomic analysis of tissue composition is particularly powerful for secreted proteins, whose protein abundances very often do not correlate with total tissue mRNA quantification in RNA sequencing assays [14]. Thus, our data represents an essential addition to existing transcriptomic studies of human lung and skin fibrosis. Furthermore, the QDSP method captures the interactions of morphogens and other secreted proteins with the ECM in an unbiased way, revealing those that are bound to the matrix by their decreased detergent solubility. Thereby, we added an additional dimension to the human lung and skin proteome, which for the first time revealed the association of secreted proteins with the ECM.

The use of unsupervised statistical tests, such as principal component analysis clearly showed the high degree of molecular heterogeneity between samples. This was no surprise since we intentionally selected a diverse patient cohort to screen for molecular events that are commonly present in all forms of fibrosis. We made use of patient heterogeneity by grouping proteins with high abundance variation across samples by their Pearson correlation. The correlation of proteins in this analysis can be for instance explained by their cell-type specific expression and thus their capacity to report differences in the relative amount of cell types in the respective biopsies. Along these lines, the recent development of high throughput technologies in the field of single cell mRNA sequencing [38] will likely enable future attempts to study cellular heterogeneity in chronic lung disease in great detail. Given the high amount of biological variation and heterogeneity of cellular composition

observed in our ILD cohort, it is remarkable that we identified a substantial number of common factors that were increased in all samples.

The most significant common factor across ILD and scleroderma samples was the protein MZB1, which we localized to CD38+/CD138+/CD27+/CD20-/CD45- B cells. In respiratory immunity, B cells can be recruited to tertiary lymphoid organs around the bronchi where they are organized in B cell follicles [39, 40]. In our analysis, MZB1 positive cells were found to be quite dispersed in the tissue and not necessarily associated with tertiary lymphoid structures, however with predominant perivascular abundance. MZB1 has important functions in the endoplasmic reticulum (ER) of B cells that undergo ER-stress upon high antibody secretory activity [32, 33]. In the immunostainings we observed MZB1 high and low cells, indicating that the expression level is tightly regulated in B cells. MZB1 high cells had a large cytoplasm and were positive for a comprehensive panel of known mature plasma B cell markers, which clearly identifies them as terminally differentiated antibody producing tissue resident plasma cells. Since most of the samples in our study were biopsies from idiopathic forms of ILD, we believe that this observation warrants future investigation regarding the causative role of antibody mediated autoimmunity in idiopathic cases of organ fibrosis. Of note, it has been recognized that many idiopathic interstitial pneumonia (IIP) patients have clinical features that suggest an underlying autoimmune process, but do not meet established criteria for a connective tissue disease (CTD). To meet this problem, an ERS/ATS task force recently proposed the term “interstitial pneumonia with autoimmune features” (IPAF) and offered several classification criteria [41].

Circulating autoantibodies in IPF have been described long time ago [6], and a causative role for B-cell mediated autoimmunity for idiopathic ILD has been discussed [42, 43]. In localized scleroderma the role of autoantibodies is unclear but the histology of the fibrotic reaction, involving a strong inflammatory infiltrate around the blood vessels, is identical to the lesions found in systemic scleroderma patients, who all have circulating autoantibodies [44]. A recent study demonstrated that autoantibodies against the lung specific protein Bpifb1 occur in 12% of patients with idiopathic ILD [7]. Importantly, the authors of this landmark study also demonstrated that T cells specific for a single autoantigen (Bpifb1) are sufficient to induce full blown and irreversible lung fibrosis in mice [7]. It is thus conceivable that (1) the presence of autoantibodies and autoreactive T cells against unknown antigens may cause or at least perpetuate many if not most IIPs and (2)

that the identification of these unknown autoantigens in patient plasma may serve as a powerful tool for both patient stratification and future immunotherapy based approaches to treatment of ILD. The recent use of chimeric antigen receptor T cells specific for autoantigen producing B-cells for targeted therapy of autoimmune disease [45] introduces an exciting new avenue for eliminating autoreactive B-cell clones, while maintaining protective adaptive immunity. Such future therapeutic approaches will depend on the identification of disease specific autoantigens, and appropriate pre-clinical models, to test if indeed certain antigens have causative roles in idiopathic forms of organ fibrosis.

Figure legends

Figure 1. Quantitative detergent solubility profiling (QDSP) of human lung and skin fibrosis.

(A) Experimental design (Graphs adapted from [14]). (B) Proteins in the four indicated QDSP detergent solubility fractions were quantified individually and the z-score of relative protein MS-intensities across 14 ILD proteomes and the four protein fractions was used for unsupervised hierarchical clustering (using Pearson correlation of rows). Clusters A-D were significantly enriched for the indicated gene categories. Core matrisome and matrisome associated proteins are assigned to the cluster by the indicated color code. (C) A principal component analysis (PCA) of the relative MS-intensities of 550 secreted proteins was used to separate the four QDSP protein fractions (indicated by the color code) in Component 1 and endstage ILD tissues (closed circles) from healthy donor lungs (open circles) in Component 4. (D) The scatter plot depicts the protein loadings used for the PCA in panel (C).

Figure 2. Total tissue protein abundance changes. The depicted volcano plots show significantly altered proteins relative to controls at the indicated false discovery rates (FDR) in (A) lung biopsy samples from ILD patients, and (B) fibrotic skin lesions from localized scleroderma patients. (C) Hierarchical clustering (Pearson correlation of z-score) of the 44 significant proteins in panel (A). Gene names and clinical classification of patients are shown. (D) In the left panel the summed up MS-intensities of MMP19 (total tissue protein abundance) are shown on log2 scale for donor and ILD samples. In the right panel, the relative MS-intensities across detergent solubility fractions of MMP19 are shown.

Figure 3. Identification of proteins and gene categories with high variance in ILD proteomes. (A) The scatter plot depicts proteins sorted from highest to lowest abundance (MS-intensity normalized by number of theoretical tryptic peptides; IBAQ) and their coefficient of variation (CV) across the ILD tissue proteomes. The indicated color code shows the 5 quantiles of CV. The dashed line shows the chosen cutoff resulting in 133 proteins with high CV used for analysis in panel D. (B) The scatter plot depicts significantly enriched gene categories along two dimensions, with annotation enrichment scores (-1 to +1) for both CV and protein abundance. The indicated gene categories in the upper right quadrant are highly

abundant while nevertheless having a high CV. GOCC – Gene Ontology Cellular Compartment; GOMF – Gene Ontology Molecular Function. (C) Normalized MS-intensities of the indicated proteins are shown on log2 scale. (D) The Correlogram visualizes the color coded Pearson correlation of 133 high CV proteins across the ILD proteomes. Gene names within the three main clusters of proteins are indicated in the boxes.

Figure 4. Distinct molecular signatures of ILD patient subsets. (A) Principal component analysis of 5707 quantified proteins was used to separate the 14 human lung tissue proteomes. A three dimensional visualization of components 1-3 enabled separation of patient subsets and healthy donor controls as indicated. Clinical diagnosis is color coded as indicated. Two distinct ILD subsets and the donor controls are labeled with a circle. (B) Two dimensional PCA plot of component 1 and 3. Two distinct ILD subsets and the donor controls are labeled with a circle. (C) The loadings of the principal component analysis in panel (B) are shown. Proteins with the highest loadings are labeled with their gene names. (D) The volcano plot depicts 272 significantly regulated proteins (FDR < 10%) in the ILD subset 1 (indicated in panels A & B) compared to healthy donor controls. (E) The volcano plot depicts 262 significantly regulated proteins (FDR < 10%) in the ILD subset 2 (indicated in panels A & B) compared to healthy donor controls.

Figure 5. High prevalence of MZB1+ tissue resident plasma B cells is a common feature in human lung and skin fibrosis. (A) The scatter plot depicts significantly enriched gene categories along two dimensions, with annotation enrichment scores (-1 to +1) for both lung fibrosis [ILD / donor] and skin fibrosis [scleroderma / control]. The indicated gene categories in the upper right quadrant are common factors in ILD and localized scleroderma. (B) The scatter plot shows the t-test difference in human skin and lung fibrosis respectively. T-test significant proteins in the ILD comparison (orange) and the scleroderma comparison (blue) are color coded. MZB1 protein was significant in both comparisons. (C) Representative confocal image of four-color immunostainings with antibodies to the indicated proteins in a FFPE tissue section of an ILD patient. Arrows indicate MZB1/CD38 double positive cells.

Nuclei were stained with DAPI and Desmin stains vascular smooth muscle cells and some mesenchymal cells in fibrotic tissues.

Figure 6. The number of MZB1+ cells correlates with tissue IgG and predicts lower diffusing capacity of the lung for carbon monoxide (DLCO). (A) Tissue homogenates of the indicated groups were subjected to Western blot analysis with antibodies against MZB1 and human IgG. Blots were stained with AmidoBlack for quantification of total protein loading. (B) The box and whiskers plot shows densitometric quantification of the MZB1 bands in the Western blot in panel (A), which were normalized to AmidoBlack staining. (C) Tissue homogenates of the indicated groups were subjected to Western blot analysis with antibodies against MZB1 and human IgG. Blots were stained with AmidoBlack for quantification of total protein loading. (D,F,G) Linear regression analyses - clinical classifications are color coded as indicated and the p-values of the linear regressions and the Pearson correlation coefficients (r) are shown. (D) Positive correlation of IgG and MZB1 levels (normalized to total protein analyzed using amidoBlack quantification). (E) Quantification of MZB1 tissue abundance by microarray in a large US cohort (Gene Expression Omnibus dataset GSE47460 published by the Lung Tissue Research Consortium). (F) Negative correlation of DLCO (%) and MZB1 levels (normalized to total protein analyzed using amidoBlack quantification). (G) Negative correlation of DLCO (%) and MZB1 levels (quantified by mass spectrometry).

Acknowledgments

We thank Silvia Weidner for technical assistance and Christoph Schaab for providing data analysis scripts. We also thank Gabi Sowa, Igor Paron and Korbinian Mayr for expert support of the proteomics pipeline. We also thank Britta Peschel and Marion Frankenberger for support with BioArchive material. This work was supported by the German Center for Lung Research (DZL), the Max Planck Society for the Advancement of Science, by the German Research Foundation (DFG) through a grant to TK (KR558), and the Helmholtz Association.

References

1. Cox, T.R. and J.T. Erler, *Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer*. Dis Model Mech, 2011. **4**(2): p. 165-78.
2. Fernandez, I.E. and O. Eickelberg, *New cellular and molecular mechanisms of lung injury and fibrosis in idiopathic pulmonary fibrosis*. Lancet, 2012. **380**(9842): p. 680-8.
3. Kreuter, A., et al., *German guidelines for the diagnosis and therapy of localized scleroderma*. J Dtsch Dermatol Ges, 2016. **14**(2): p. 199-216.
4. Thannickal, V.J., et al., *Fibrosis: ultimate and proximate causes*. J Clin Invest, 2014. **124**(11): p. 4673-7.
5. Vij, R. and M.E. Strek, *Diagnosis and treatment of connective tissue disease-associated interstitial lung disease*. Chest, 2013. **143**(3): p. 814-24.
6. Dreisin, R.B., et al., *Circulating immune complexes in the idiopathic interstitial pneumonias*. N Engl J Med, 1978. **298**(7): p. 353-7.
7. Shum, A.K., et al., *BPIFB1 is a lung-specific autoantigen associated with interstitial lung disease*. Sci Transl Med, 2013. **5**(206): p. 206ra139.
8. Peng, T., et al., *Hedgehog actively maintains adult lung quiescence and regulates repair and regeneration*. Nature, 2015.
9. Konigshoff, M., et al., *WNT1-inducible signaling protein-1 mediates pulmonary fibrosis in mice and is upregulated in humans with idiopathic pulmonary fibrosis*. J Clin Invest, 2009. **119**(4): p. 772-87.
10. Lee, J.H., et al., *Lung Stem Cell Differentiation in Mice Directed by Endothelial Cells via a BMP4-NFATc1-Thrombospondin-1 Axis*. Cell, 2014. **156**(3): p. 440-55.
11. Hogan, B.L., et al., *Repair and Regeneration of the Respiratory System: Complexity, Plasticity, and Mechanisms of Lung Stem Cell Function*. Cell Stem Cell, 2014. **15**(2): p. 123-138.
12. Martino, M.M., et al., *Growth factors engineered for super-affinity to the extracellular matrix enhance tissue healing*. Science, 2014. **343**(6173): p. 885-8.
13. Hynes, R.O., *Stretching the boundaries of extracellular matrix research*. Nat Rev Mol Cell Biol, 2014. **15**(12): p. 761-3.
14. Schiller, H.B., et al., *Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair*. Mol Syst Biol, 2015. **11**(7): p. 819.
15. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification*. Nature Biotechnology, 2008. **26**(12): p. 1367-72.

16. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data*. Nat Methods, 2016.
17. Naba, A., et al., *The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices*. Mol Cell Proteomics, 2012. **11**(4): p. M111 014647.
18. Yu, G., et al., *Matrix metalloproteinase-19 is a key regulator of lung fibrosis in mice and humans*. Am J Respir Crit Care Med, 2012. **186**(8): p. 752-62.
19. Staab-Weijnitz, C.A., et al., *FK506-Binding Protein 10, a Potential Novel Drug Target for Idiopathic Pulmonary Fibrosis*. Am J Respir Crit Care Med, 2015. **192**(4): p. 455-67.
20. Willaert, A., et al., *Recessive osteogenesis imperfecta caused by LEPRE1 mutations: clinical documentation and identification of the splice form responsible for prolyl 3-hydroxylation*. J Med Genet, 2009. **46**(4): p. 233-41.
21. Vranka, J.A., L.Y. Sakai, and H.P. Bachinger, *Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes*. J Biol Chem, 2004. **279**(22): p. 23615-21.
22. Agarwal, P., et al., *Collagen XII and XIV, new partners of cartilage oligomeric matrix protein in the skin extracellular matrix suprastructure*. J Biol Chem, 2012. **287**(27): p. 22549-59.
23. Agarwal, P., et al., *Enhanced deposition of cartilage oligomeric matrix protein is a common feature in fibrotic skin pathologies*. Matrix Biol, 2013. **32**(6): p. 325-31.
24. Schulz, J.N., et al., *COMP-assisted collagen secretion--a novel intracellular function required for fibrosis*. J Cell Sci, 2016. **129**(4): p. 706-16.
25. Menz, C., et al., *LTBP-2 Has a Single High-Affinity Binding Site for FGF-2 and Blocks FGF-2-Induced Cell Proliferation*. PLoS One, 2015. **10**(8): p. e0135577.
26. Sideek, M.A., et al., *Co-localization of LTBP-2 with FGF-2 in fibrotic human keloid and hypertrophic scar*. J Mol Histol, 2016. **47**(1): p. 35-45.
27. Cox, J. and M. Mann, *1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data*. BMC Bioinformatics, 2012. **13 Suppl 16**: p. S12.
28. Zuo, F., et al., *Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans*. Proc Natl Acad Sci U S A, 2002. **99**(9): p. 6292-7.
29. Konishi, K., et al., *Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis*. Am J Respir Crit Care Med, 2009. **180**(2): p. 167-75.
30. Smirnova, N.F., et al., *Detection and quantification of epithelial progenitor cell populations in human healthy and IPF lungs*. Respir Res, 2016. **17**(1): p. 83.

31. Estany, S., et al., *Lung fibrotic tenascin-C upregulation is associated with other extracellular matrix proteins and induced by TGFbeta1*. BMC Pulm Med, 2014. **14**: p. 120.
32. Flach, H., et al., *Mzb1 protein regulates calcium homeostasis, antibody secretion, and integrin activation in innate-like B cells*. Immunity, 2010. **33**(5): p. 723-35.
33. Rosenbaum, M., et al., *MZB1 is a GRP94 cochaperone that enables proper immunoglobulin heavy chain biosynthesis upon ER stress*. Genes Dev, 2014. **28**(11): p. 1165-78.
34. Pinna, D., et al., *Clonal dissection of the human memory B-cell repertoire following infection and vaccination*. European Journal of Immunology, 2009. **39**(5): p. 1260-70.
35. Teller, J., et al., *Blimp-1 controls plasma cell function through the regulation of immunoglobulin secretion and the unfolded protein response*. Nature Immunology, 2016. **17**(3): p. 323-30.
36. Nutt, S.L., et al., *The generation of antibody-secreting plasma cells*. Nat Rev Immunol, 2015. **15**(3): p. 160-71.
37. Aebersold, R. and M. Mann, *Mass-spectrometric exploration of proteome structure and function*. Nature, 2016. **537**(7620): p. 347-55.
38. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets*. Cell, 2015. **161**(5): p. 1202-14.
39. John-Schuster, G., et al., *Cigarette smoke-induced iBALT mediates macrophage activation in a B cell-dependent manner in COPD*. Am J Physiol Lung Cell Mol Physiol, 2014. **307**(9): p. L692-706.
40. Moyron-Quiroz, J.E., et al., *Role of inducible bronchus associated lymphoid tissue (iBALT) in respiratory immunity*. Nat Med, 2004. **10**(9): p. 927-34.
41. Fischer, A., et al., *An official European Respiratory Society/American Thoracic Society research statement: interstitial pneumonia with autoimmune features*. Eur Respir J, 2015. **46**(4): p. 976-87.
42. Xue, J., et al., *Plasma B lymphocyte stimulator and B cell differentiation in idiopathic pulmonary fibrosis patients*. J Immunol, 2013. **191**(5): p. 2089-95.
43. Donahoe, M., et al., *Autoantibody-Targeted Treatments for Acute Exacerbations of Idiopathic Pulmonary Fibrosis*. PLoS One, 2015. **10**(6): p. e0127771.
44. Fleischmajer, R. and A. Nedwich, *Generalized morphea. I. Histology of the dermis and subcutaneous tissue*. Arch Dermatol, 1972. **106**(4): p. 509-14.
45. Ellebrecht, C.T., et al., *Reengineering chimeric antigen receptor T cells for targeted therapy of autoimmune disease*. Science, 2016.

A



Figure 2

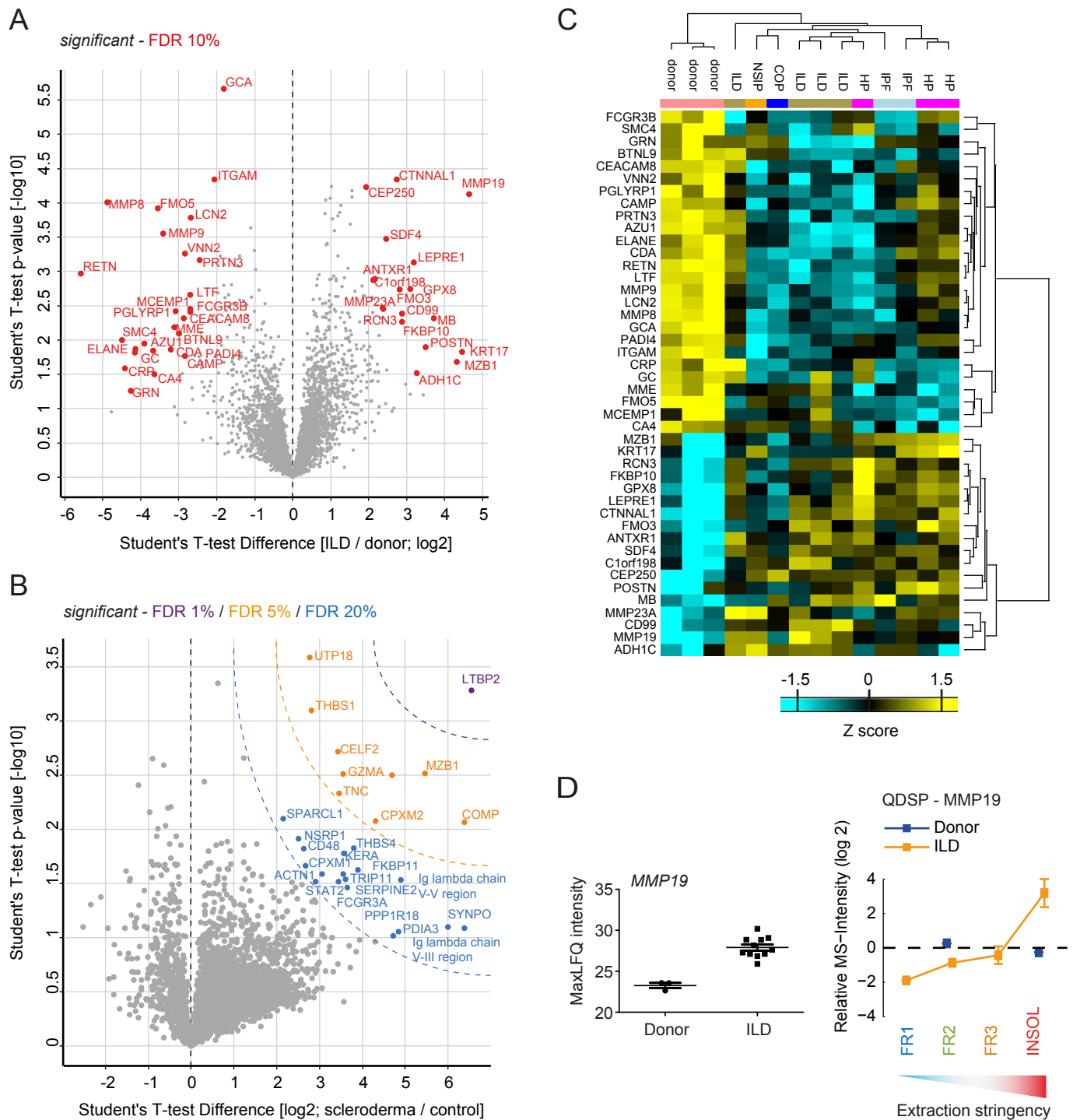


Figure 3

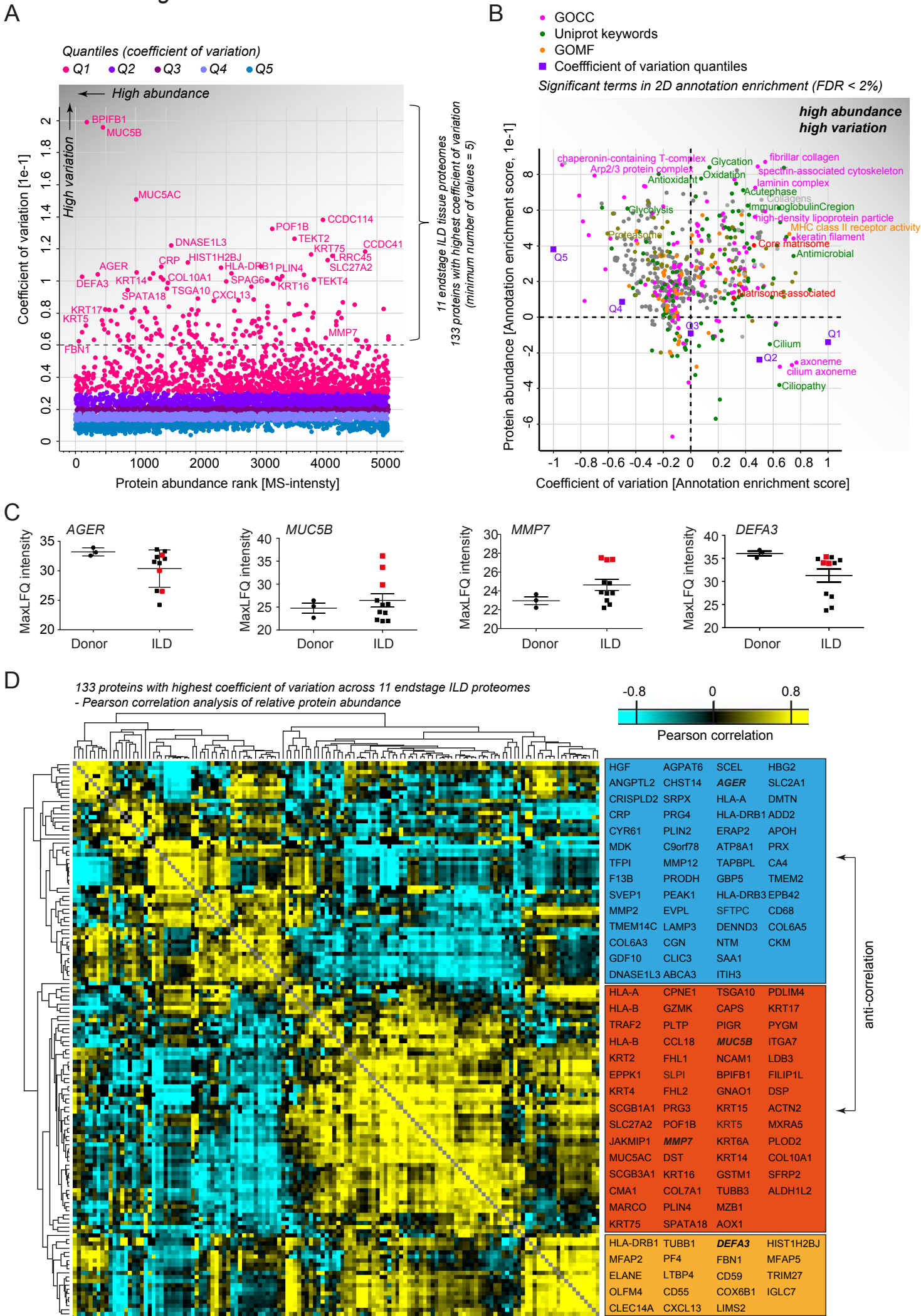


Figure 4

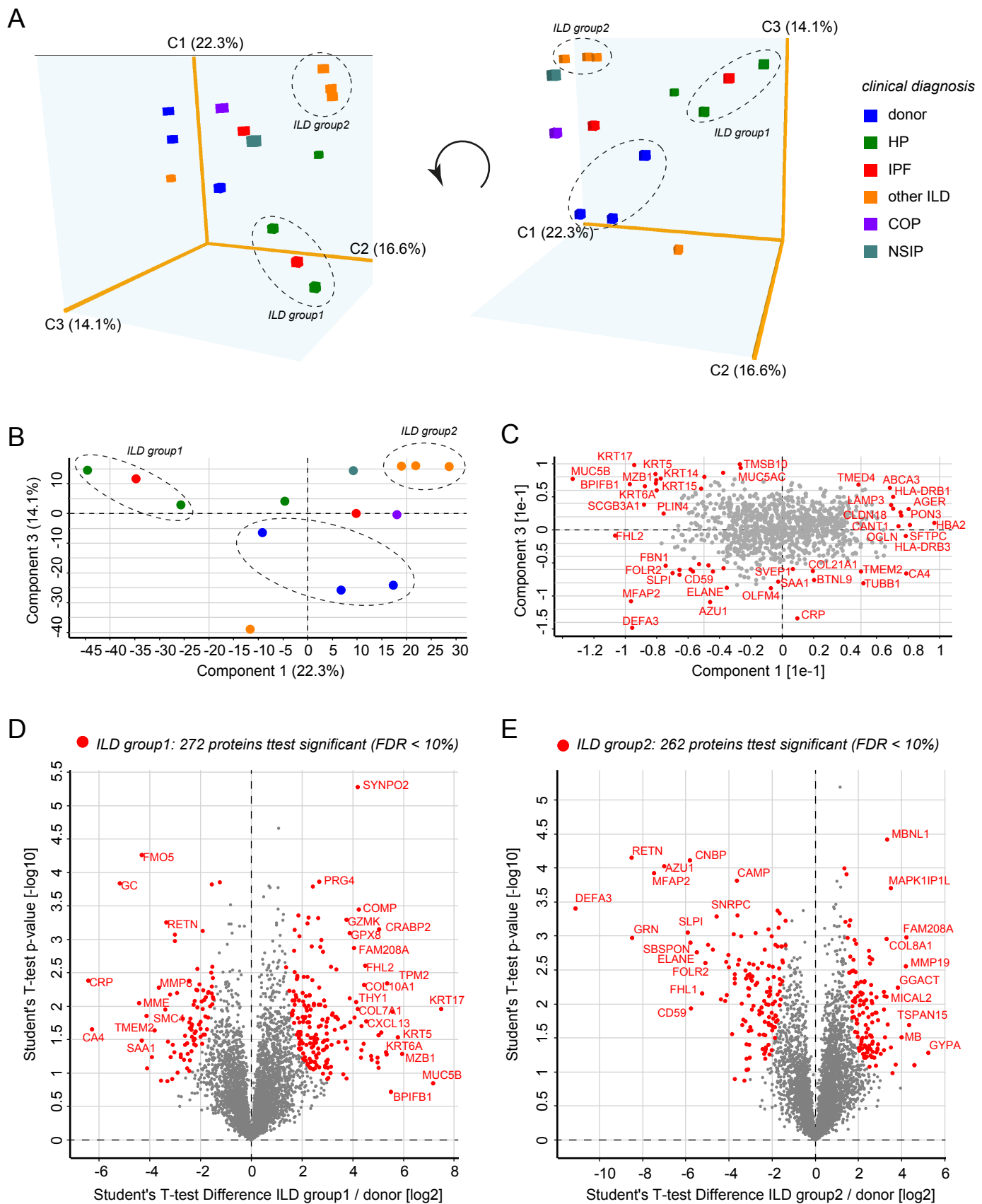


Figure 5

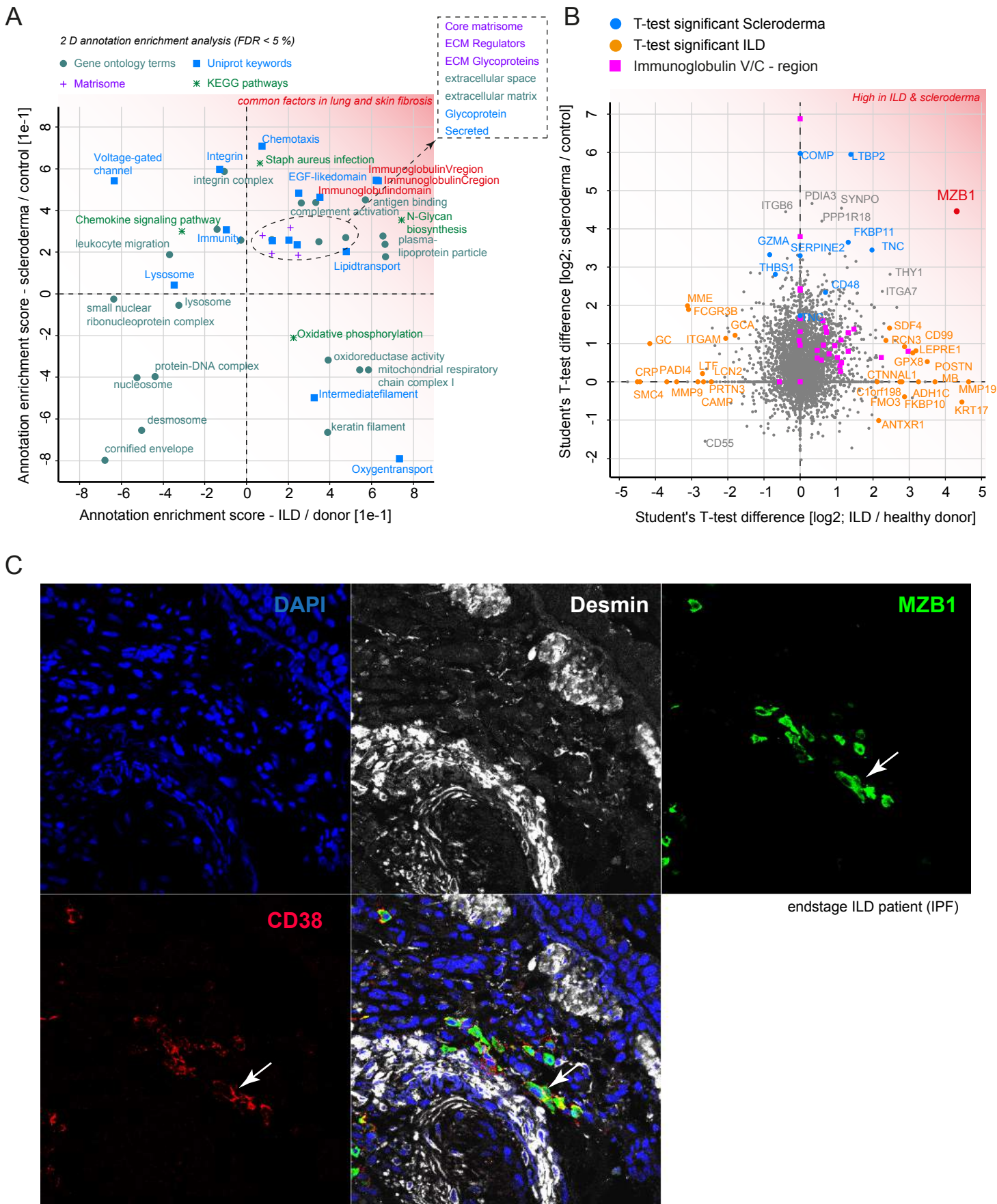
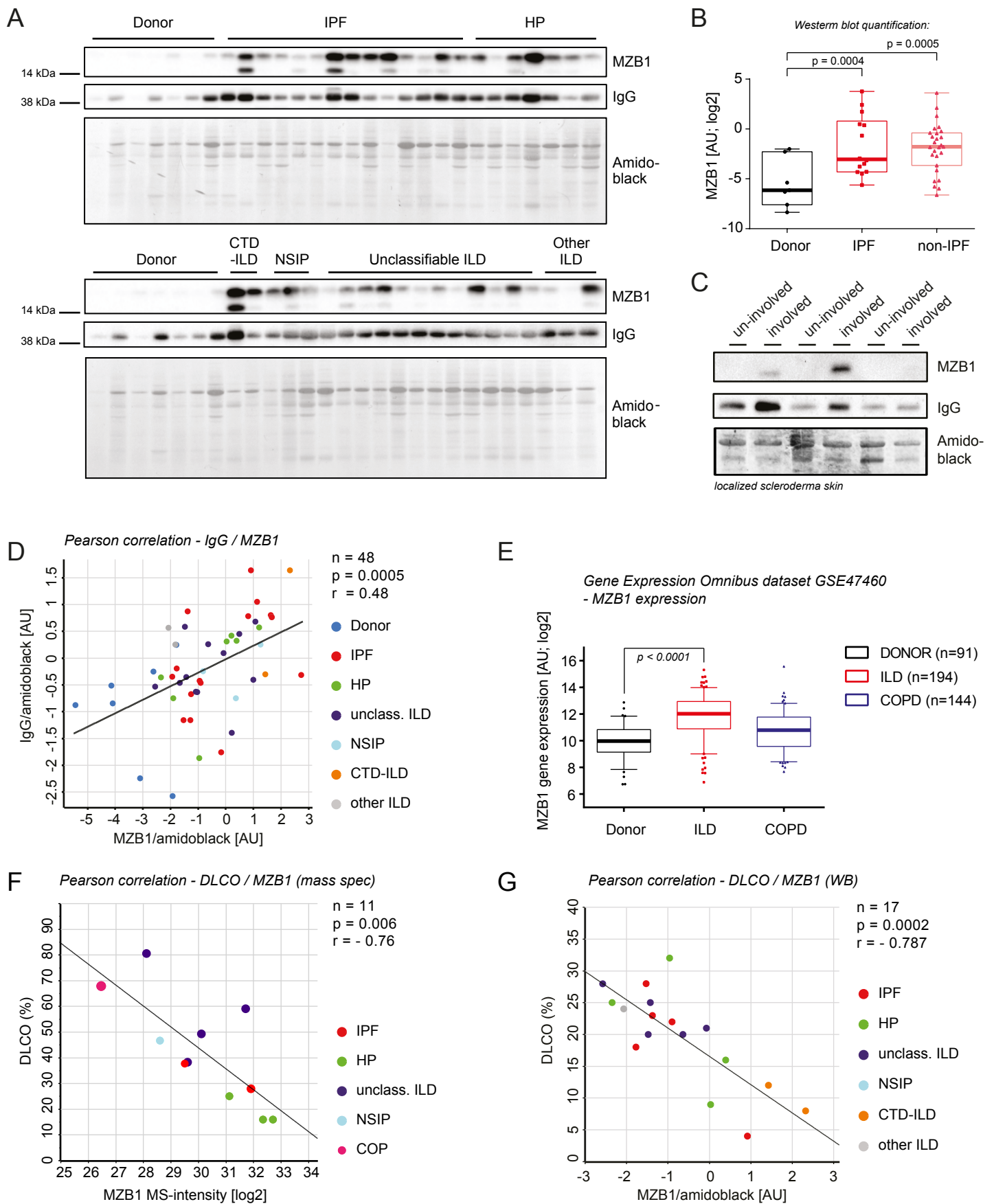


Figure 6



ONLINE DATA SUPPLEMENT**Deep proteome profiling reveals common prevalence of MZB1-positive plasma B cells in human lung and skin fibrosis**

Herbert B. Schiller^{1, 2*}, Christoph H. Mayr¹, Gabriela Leuschner^{1,5}, Maximilian Strunz¹, Claudia Staab-Weijnitz^{1,2}, Stefan Preisendörfer¹, Beate Eckes³, Pia Moinzadeh³, Thomas Krieg³, David A. Schwartz⁷, Rudolf A. Hatz⁴, Jürgen Behr^{5,2}, Matthias Mann⁶, Oliver Eickelberg^{1,2,7 *}

MATERIAL AND METHODS

Antibodies, immunohistochemistry and microscopy

FFPE tissue samples were sectioned and stained as previously described [1]. The following primary (1) and secondary (2) antibodies were used: (1) MZB1 rabbit (Sigma-Aldrich, HPA043745), CD3 rabbit (Abcam, ab16669), CD20 mouse (Dako, MO755), CD38 mouse (Santa Cruz, sc-374650), CD45 mouse (Sigma-Aldrich, AMAb90518), Desmin goat (Santa Cruz, sc-7559), Human IgG [EPR4421] (Abcam, ab109489), CD27 (Abcam, ab49518), CD138 (Sigma, SAB4700486); (2) donkey anti-mouse Alexa Fluor (AF) 647 (Invitrogen, A-31571), donkey anti-rabbit AF 568 (Invitrogen, A10042), donkey anti-goat AF 488 (Invitrogen, A11055).

Plasma cell differentiation and MZB1 qPCR

QPCR and Western Blot analysis was performed as described previously [2]. Primers used are listed in the table below. Antibody against human IgG was goat anti-Human IgG (Fc specific, Sigma-Aldrich).

Target	Species	Forward (5'->3')	Reverse (5'->3')
MZB1	human	GGA ACT GGC AGG ACT AC	CAA ACA TGT CCT GGA GAG
Mzb1	mouse	AAC TGG CAG TCC TAT GG	GAA ACA CGT CTT GGA GAG
GAPDH	human	TGA CCT CAA CTA CAT GGT TTA CAT G	TTG ATT TTG GAG GGA TCT CG
Gapdh	mouse	TGT GTC CGT CGT GGA TCT GA	CCT GCT TCA CCA CCT TCT TGA
BLIMP1	human	GAT GAA TCT CAC ACA AAC AC	GAT TTC TTT CAC GCT GTA CT

Differentiation of memory B cells to Ig-secreting cells was performed according to Pinna *et al* [3]. Briefly, human peripheral blood mononuclear cells (PBMCs) were stimulated with interleukin-2 (IL2) and the TLR7/8 ligand R848 for 7 days and compared to unstimulated cells cultured for the same time period.

Sample preparation procedures for proteome analysis

For proteome analysis of human tissue biopsies ~100mg of fresh frozen total tissue (wet weight) was homogenized in 500 μ l PBS (with protease inhibitor cocktail) using an Ultra-turrax homogenizer. After centrifugation the soluble proteins were collected and proteins were extracted from the insoluble pellet in 3 steps using buffers with increasing stringency as described in the QDSP protocol [1]. Peptides from LysC and trypsin proteolysis of the four protein fractions in guanidium hydrochloride (enzyme/protein ratio 1:50), were purified as previously described on SDB-RPS material stage-tips [1].

LC-MS/MS analysis

Data was acquired on a Quadrupole/Orbitrap type Mass Spectrometer (Q-Exactive, Thermo Scientific) as previously described [1]. Approximately 2 μ g of peptides were separated in a four hour gradient on a 50-cm long (75- μ m inner diameter) column packed in-house with ReproSil-Pur C18-AQ 1.9 μ m resin (Dr. Maisch GmbH). Reverse-phase chromatography was performed with an EASY-nLC 1000 ultra-high pressure system (Thermo Fisher Scientific), which was coupled to a Q-Exactive Mass Spectrometer (Thermo Scientific). Peptides were loaded with buffer A (0.1% (v/v) formic acid) and eluted with a nonlinear 240-min gradient of 5–60% buffer B (0.1% (v/v) formic acid, 80% (v/v) acetonitrile) at a flow rate of 250 nl/min. After each gradient, the column was washed with 95% buffer B and reequilibrated with buffer A. Column temperature was kept at 50 °C by an in-house designed oven with a Peltier element [4] and operational parameters were monitored in real time by the SprayQc software [5]. MS data were acquired with a shotgun proteomics method, where in each cycle a full scan, providing an overview of the full complement of isotope patterns visible at that particular time point, is followed by up-to ten data-dependent MS/MS scans on the most abundant not yet sequenced isotopes (top10 method) [6]. Target value for the full scan MS spectra was 3×10^6 charges in the 300–1,650 m/z range with a maximum injection time of 20 ms and a resolution of 70,000 at m/z 400. Isolation of precursors was performed with the quadrupole at window of 3 Th. Precursors were fragmented by higher-energy collisional dissociation (HCD) with normalized collision energy of 25 % (the appropriate energy is calculated using this percentage, and m/z and charge state of the precursor). MS/MS scans were acquired at a resolution of 17,500 at m/z 400 with an ion target value of 1×10^5 a

maximum injection time of 120 ms, and fixed first mass of 100 Th. Repeat sequencing of peptides was minimized by excluding the selected peptide candidates for 40 seconds.

Bioinformatic analysis and statistics

MS raw files were analyzed by the MaxQuant software [7] (version 1.4.3.20) and peak lists were searched against the human Uniprot FASTA database (version May 2013), and a common contaminants database (247 entries) by the Andromeda search engine [8] as previously described [1]. As fixed modification cysteine carbamidomethylation and as variable modifications, hydroxylation of proline and methionine oxidation was used. False discovery rate was set to 0.01 for proteins and peptides (minimum length of seven amino acids) and was determined by searching a reverse database. Enzyme specificity was set as C-terminal to arginine and lysine, and a maximum of two missed cleavages were allowed in the database search. Peptide identification was performed with an allowed precursor mass deviation up to 4.5 ppm after time-dependent mass calibration and an allowed fragment mass deviation of 20 ppm. For label-free quantification in MaxQuant the minimum ratio count was set to two. For matching between runs, the retention time alignment window was set to 30 min and the match time window was 1 min. QDSP data analysis was performed with a custom made matlab script as previously described [1]. Box plots, t-test statistics and correlation analysis was performed using the software GraphPad Prism. All other statistical and bioinformatics operations, such as normalization, pattern recognition, cross-omics comparisons and multiple-hypothesis testing corrections, were performed with the Perseus software package [9].

1. Schiller, H.B., et al., *Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair*. Mol Syst Biol, 2015. **11**(7): p. 819.
2. Staab-Weijnitz, C.A., et al., *FK506-Binding Protein 10, a Potential Novel Drug Target for Idiopathic Pulmonary Fibrosis*. American Journal of Respiratory and Critical Care Medicine, 2015. **192**(4): p. 455-67.
3. Pinna, D., et al., *Clonal dissection of the human memory B-cell repertoire following infection and vaccination*. European Journal of Immunology, 2009. **39**(5): p. 1260-70.
4. Thakur, S.S., et al., *Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation*. Mol Cell Proteomics, 2011. **10**(8): p. M110 003699.
5. Scheltema, R.A. and M. Mann, *SprayQc: a real-time LC-MS/MS quality monitoring system to maximize uptime using off the shelf components*. J Proteome Res, 2012. **11**(6): p. 3458-66.
6. Michalski, A., et al., *Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer*. Mol Cell Proteomics, 2011. **10**(9): p. M111 011015.
7. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification*. Nature Biotechnology, 2008. **26**(12): p. 1367-72.
8. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment*. Journal of proteome research, 2011. **10**(4): p. 1794-805.
9. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data*. Nat Methods, 2016.

Supplement Legends

Figure S1. Cross-omics analysis of molecular alterations in ILD. (A, B) Western blot analysis of KRT17 (B) and SDF4 (C) expression in IPF samples from an independent US cohort compared to Donor lungs. (C) The mean log₂ mRNA abundance ratios [ILD n=194 / CTRL n=91] from a published gene expression dataset (Gene Expression Omnibus dataset GSE47460 published by the Lung Tissue Research Consortium) was plotted against the mean log₂ protein abundance ratios [ILD n=11 / CTR n=3] of the ILD patient proteomes described in Figure 2. Outliers are labeled with gene names. (D, E) Correlation analysis of mRNA abundance (Gene Expression Omnibus dataset GSE47460) and protein abundance (mass spectrometry) for (B) healthy controls and (C) ILD tissues. Core matrisome proteins are labeled and the Pearson correlation coefficient (r) is shown.

Figure S2: MZB1 expression correlates with plasma B cell differentiation. (A) Treatment of human peripheral blood mononuclear cells (PBMCs) with interleukin-2 (IL2) and the TLR7/8 ligand R848 induces differentiation of memory B cells to Ig-secreting plasma cells [34], as monitored by immunofluorescent stainings for human IgG and (B) increased transcript levels of BLIMP1, a transcription factor essential for plasma cell function [35]. (C) MZB1 transcript (upper panel) and protein (lower panel, representative Western Blot) levels are drastically increased upon treatment of PBMCs with IL2/R848 in comparison to unstimulated cells. Results are based on three independent experiments using PBMCs derived from the same donor and are given as mean $\pm$ SEM. Statistical analysis was performed using paired t-test (*, $p < 0.05$). Scale bar: 20 μ m.

Figure S3. MZB1+ cells in human lungs are CD3 and CD20 negative. (A, B) Representative confocal image of four-color immunostainings with antibodies to the indicated proteins in a FFPE tissue section of an ILD patient. Nuclei were stained with DAPI and Desmin stains vascular smooth muscle cells and some mesenchymal cells in fibrotic tissues. MZB1+ cells are (A) negative for the T lymphocyte specific cell surface marker protein CD3, and (B) the B cell specific cell surface marker protein CD20. Arrows indicate MZB1+ cells.

Figure S4. MZB1+ cells in human lungs are CD45 negative and CD138 positive. (A, B) Representative confocal image of four color immunostainings with antibodies to the indicated proteins in a FFPE tissue section from an ILD patient. Nuclei were stained with DAPI and Desmin stains vascular smooth muscle cells and some mesenchymal cells in fibrotic tissues. (A) MZB1+ cells are negative for the leukocyte specific cell surface marker protein CD45. Arrows indicate MZB1+ cells. (B) MZB1+ cells are positive for the plasma B cell marker CD138. Arrows indicate MZB1/CD138 double-positive cells.

Figure S5. MZB1+ cells in human lungs stain positive for CD27 and IgG. (A, B) Representative confocal image of four color immunostainings with antibodies to the indicated proteins in a FFPE tissue section from an ILD patient. Nuclei were stained with DAPI and Desmin stains vascular smooth muscle cells and some mesenchymal cells in fibrotic tissues. (A) MZB1+ cells are positive for the plasma B cell specific cell surface marker protein CD27. Arrows indicate MZB1/CD27 double positive cells. (B) MZB1+ cells are positive for IgG. Arrows indicate MZB1/IgG double-positive cells.

Figure S6. MZB1+ cells in human skin are CD3 and CD20 negative. (A, B) Representative confocal image of four-color immunostainings with antibodies to the indicated proteins in a FFPE tissue section of localized scleroderma. Nuclei were stained with DAPI and Desmin stains vascular smooth muscle cells and some mesenchymal cells in fibrotic tissues. MZB1+ cells are (A) negative for the T lymphocyte specific cell surface marker protein CD3, and (B) the B cell specific cell surface marker protein CD20. Arrows indicate MZB1+ cells.

Figure S7. MZB1+ cells in human skin are CD45 negative and CD38 positive. (A, B) Representative confocal image of four color immunostainings with antibodies to the indicated proteins in a FFPE tissue section of localized scleroderma. Nuclei were stained with DAPI and Desmin stains vascular smooth muscle cells and some mesenchymal cells in fibrotic tissues. MZB1+ cells are (A) negative for the leukocyte specific cell surface marker protein

CD45, and (B) positive for the plasma B cell marker CD38. Arrows indicate MZB1/CD38 double-positive cells.

Figure S8. No effect of age, vital capacity, treatment or gender on MZB1 tissue levels. (A, B) Linear regression analyses - clinical classifications are color coded as indicated and the p-values of the linear regressions and the Pearson correlation coefficients (r) are shown. (A) Correlation of age and MZB1 levels (normalized to total protein analyzed using amidoblack quantification). (B) Correlation of vital capacity (VC) in % and MZB1 levels (normalized to total protein analyzed using amidoblack quantification). (C-E) The box and whisker plots show MZB1 levels in the indicated groups as quantified by Western blotting.

Table S1. ILD proteomes. The xlsx table contains three tabs with different types of information. **Tab1** ('total proteome') shows the quantification of total protein abundance. Expression columns as well as numerical and categorical annotations are shown. **Tab2** ('QDSP full dataset') shows the quantification of proteins separately for each fraction in the QDSP protocol [14]. Individual columns representing the MS-intensities for each sample are labeled with the fraction names ('FR1-FR4') and the group names ('ILD' or 'donor'). Additional columns show gene annotations. **Tab3** ('QDSP normalized') shows normalized protein intensities across the QDSP fractions. The solubility profiles across the four fractions are compared by first normalizing the intensities such that the mean log 2 intensities of the groups ('ILD', 'ILD subset' and 'donor') are zero.

Table S2. Localized scleroderma proteomes. The xlsx table contains three tabs with different types of information. **Tab1** ('total proteome') shows the quantification of total protein abundance. Expression columns as well as numerical and categorical annotations are shown. **Tab2** ('QDSP full dataset') shows the quantification of proteins separately for each fraction in the QDSP protocol (Schiller et al., 2015). Individual columns representing the MS-intensities for each sample are labeled with the fraction names ('FR1-FR4') and the group names ('control' or 'sclerotic'). Additional columns show gene annotations. **Tab3** ('QDSP normalized') shows normalized protein intensities across the QDSP fractions. The solubility

profiles across the four fractions are compared by first normalizing the intensities such that the mean log 2 intensities of the groups (‘control’, ‘sclerotic’) are zero.

Table S3. Baseline characteristics of patients with ILD included in the mass spectrometry analysis. Abbreviations: PFT: pulmonary function test, VC: vital capacity, TLC: total lung capacity, RV: residual volume, FEV1: forced expiratory volume in 1 second, DLCO: diffusing capacity of the lung for carbon monoxide, LTOT: long term oxygen therapy.

Table S4. Two dimensional annotation enrichment analyses for protein abundance ranks versus coefficient of variation ranks. The table shows significantly enriched gene categories along two dimensions, with annotation enrichment scores (-1 to +1) for both CV and protein abundance.

Table S5. Pearson correlation matrix of high CV proteins. The xlsx table shows the Pearson correlation coefficients of 133 proteins that were selected by their high coefficient of variation across the 11 ILD proteomes. Alongside the correlation matrix several numerical and categorical annotations to every protein are shown.

Table S6. Two dimensional annotation enrichment analyses for lung versus skin fibrosis. The table shows significantly enriched gene categories along two dimensions, with annotation enrichment scores (-1 to +1) for both lung fibrosis [ILD / donor] and skin fibrosis [scleroderma / control].

Table S7. Baseline characteristics of patients with ILD analyzed by Western blotting. Abbreviations: PFT: pulmonary function test, VC: vital capacity, TLC: total lung capacity, RV: residual volume, FEV1: forced expiratory volume in 1 second, DLCO: diffusing capacity of the lung for carbon monoxide, LTOT: long term oxygen therapy.

Patient characteristics

Characteristics	IPF n=2	HP n=3	COP n=1	NSIP n=1	Unclassifiable ILD n=4	All n=11
Male gender, no. (%)	1 (50,0)	1 (33,3)	0 (0)	1 (100,0)	3 (75,0)	8 (72,7)
Age, years, $\pm$ SD	66,0 $\pm$ 4,2	47,3 $\pm$ 13,3	48,0	84,0	54,0 $\pm$ 4,2	58 $\pm$ 13,1
PFT						
VC, l (% pred)	3,41 $\pm$ 0,33 (97,8 $\pm$ 26,6)	1,06 $\pm$ 0,07 (31,7 $\pm$ 4,04)	3,2 (109)	2,32 (69,4)	3,27 $\pm$ 0,62 (81,1 $\pm$ 10,7)	2,72 $\pm$ 1,09 (76,1 $\pm$ 29,9)
TLC, l (% pred)	5,61 $\pm$ 0,70 (95,1 $\pm$ 36,8)	2,23 $\pm$ 0,09 (43,0 $\pm$ 5,2)	4,91 (105,9)	4,25 (66,2)	5,46 $\pm$ 0,28 (86,4 $\pm$ 13,4)	4,62 $\pm$ 1,5 (79,0 $\pm$ 27,1)
RV, l (% pred)	2,20 $\pm$ 1,01 (99,9 $\pm$ 59,2)	1,17 $\pm$ 0,16 (31,7 $\pm$ 27,1)	1,71 (104,9)	1,94 (68,6)	2,09 $\pm$ 0,37 (100,5 $\pm$ 27,0)	1,87 $\pm$ 0,58 (82,3 $\pm$ 41,7)
FEV1, l (% pred)	2,72 $\pm$ 0,73 (97,1 $\pm$ 5,7)	0,93 $\pm$ 0,01 (33,7 $\pm$ 5,77)	2,86 (117,2)	1,79 (75,5)	2,73 $\pm$ 0,59 (86,3 $\pm$ 11,8)	2,25 $\pm$ 0,93 (79,0 $\pm$ 29,8)
DLCO, % pred.	32,8 $\pm$ 6,9	19,0 $\pm$ 5,2	67,9	46,7	56,8 $\pm$ 18,0	40,8 $\pm$ 21,0
pO ₂ at rest, mmHg	61,25 $\pm$ 22,98	46,67 $\pm$ 9,81	68,7	73,4	75,25 $\pm$ 12,39	63,70 $\pm$ 16,48
Therapy *						
LTOT, no. (%)	1 (50,0)	3 (100,0)	x	x	x	4/5 (80,0)
Steroids, no. (%)	1 (50,0)	3 (100,0)	x	x	x	4/5 (80,0)
Immunosuppressant, no. (%)	1 (50,0)	1 (33,3)	x	x	x	2/5 (40,0)
Antifibrotic drugs, no. (%)	0 (0)	0 (0)	x	x	x	0/5 (0)

Table S3: Baseline characteristics of patients with ILD included in the mass spectrometry analysis

Abbreviations: PFT: pulmonary function test, VC: vital capacity, TLC: total lung capacity, RV: residual volume, FEV1: forced expiratory volume in 1 second, DLCO: diffusing capacity of the lung for carbon monoxide, LTOT: long term oxygen therapy.

Characteristics	IPF n=14	HP n=7	CTD-ILD n=2	NSIP n=3	Unclassifiable ILD n=12	Other ILD n=3	All n=41
Male gender, no. (%)	12 (85,7)	3 (42,9)	2 (100)	2 (66,7)	7 (58,3)	2 (66,7)	28 (58,3)
Age, years, ±SD	56,1±10,0	53,7±10,5	42,5±4,9	56,3±8,1	51,0±7,0	55,1±6,9	54,4±8,8
PFT							
VC, l (% pred)	1,8±0,2 (38,0±2,8)	1,5±0,7 (42,3±22,5)	2,1±0,01 (39,5±9,2)	1,8±0,2 (38,0±2,8)	1,5±0,4 (39,8±9,3)	1,9±1,3 (43,7±28,6)	1,7±0,6 (40,6±15,0)
TLC, l (% pred)	3,11±0,5 (41,5±0,7)	3,2±1,2 (54,8±20,7)	3,1±0,5 (41,5±0,7)	2,67 (36) *	2,7±0,5 (45,7±6,0)	4,5±1,7 (63,0±28,3)	3,2±1,0 (49,2±14,2)
RV, l (% pred)	1,06±0,5 (50,0±21,2)	1,6±0,5 (70,2±32,29)	1,06±0,5 (50±21,2)	0,95 (38) *	1,2±0,3 (59,3±14,3)	2,3±0,2 (100,0±4,2)	1,5±0,6 (68,5±35,2)
FEV1, l (% pred)	1,69±0,3 (40,5±2,1)	1,2±0,5 (42,7±16,5)	1,69±0,3 (40,5±2,1)	1,73±0,1 (46,5±0,7)	1,3±0,4 (43,1±11,6)	1,1±0,3 (33,3±9,1)	1,4±0,5 (42,7±14,5)
DLCO, % pred.	10±2,8 n=5	20,5±10,1 n=4°	10 ±2,8	x	22,8±3,6 n=5°	24 n=1°	19,7±7,7
pO2 at rest, mmHg	36,0±7,1	49,2±8,2	36±7,1	x	50,6±6,8	56±1,4	48,3±8,5
Therapy							
LTOT, no. (%)	14 (100)	7 (100)	2 (100)	3 (100)	10 (83,3)	3 (100)	39 (95,1)
Steroids, no. (%)	7 (50,0)	7 (100)	2 (100)	3 (100)	10 (83,3)	3 (100)	32 (78,0)
Immunosuppressant, no. (%)	2 (14,3)	3 (42,9)	1 (50)	0 (0)	2 (16,7)	0 (0)	13 (31,7)
Antifibrotic drugs, no. (%)	7 (50,0)	2 (28,6)	0	1 (33,3)	7 (58,3)	1 (33,3)	13 (31,7)

Table S7: Baseline characteristics of patients with ILD analyzed by Western blotting

Abbreviations: PFT: pulmonary function test, VC: vital capacity, TLC: total lung capacity, RV: residual volume, FEV1: forced expiratory volume in 1 second, DLCO: diffusing capacity of the lung for carbon monoxide, LTOT: long term oxygen therapy.

* n=1

° not all patients were able to perform a DLCO

Figure S1

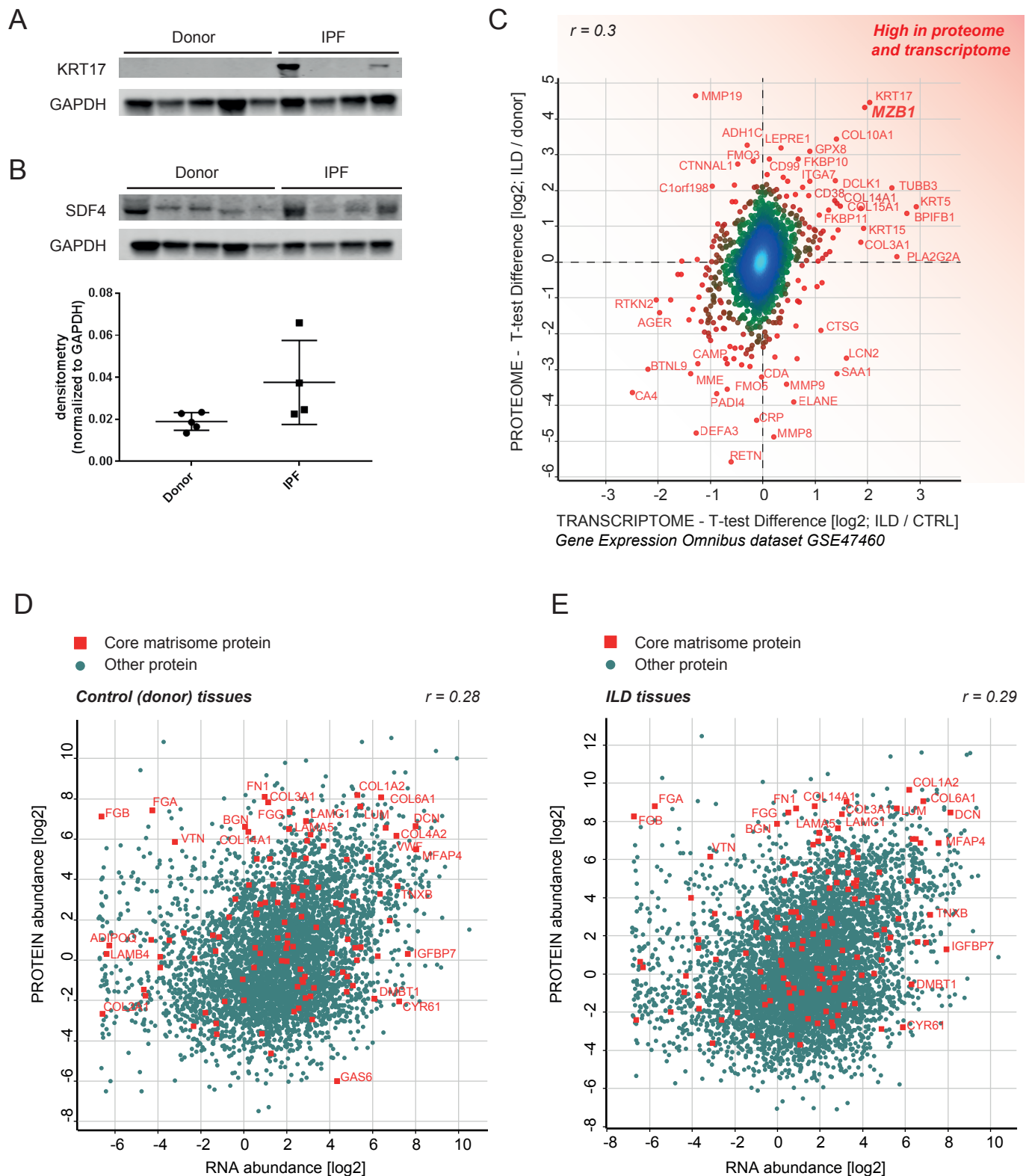
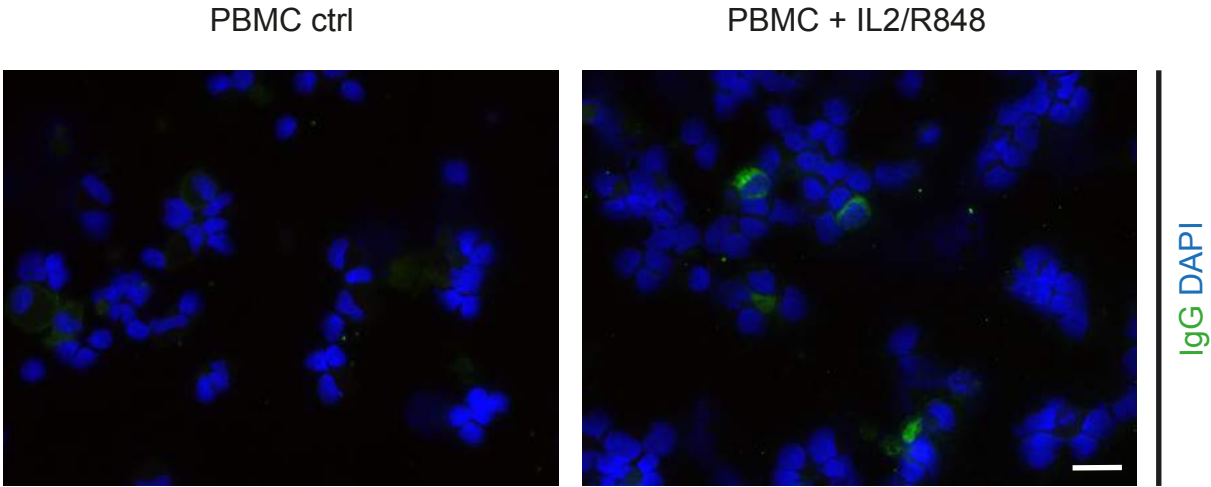
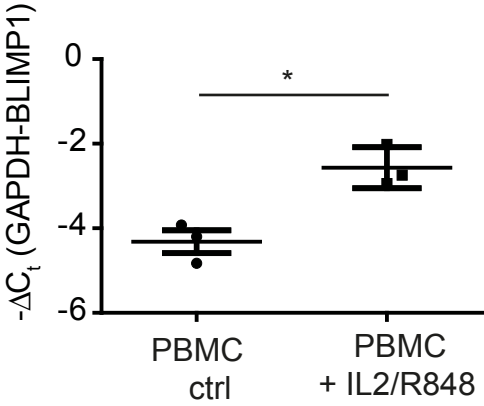


Figure S2

A



B



C

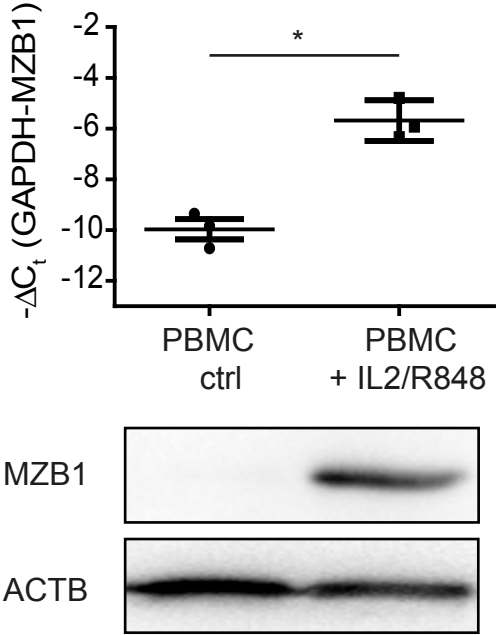
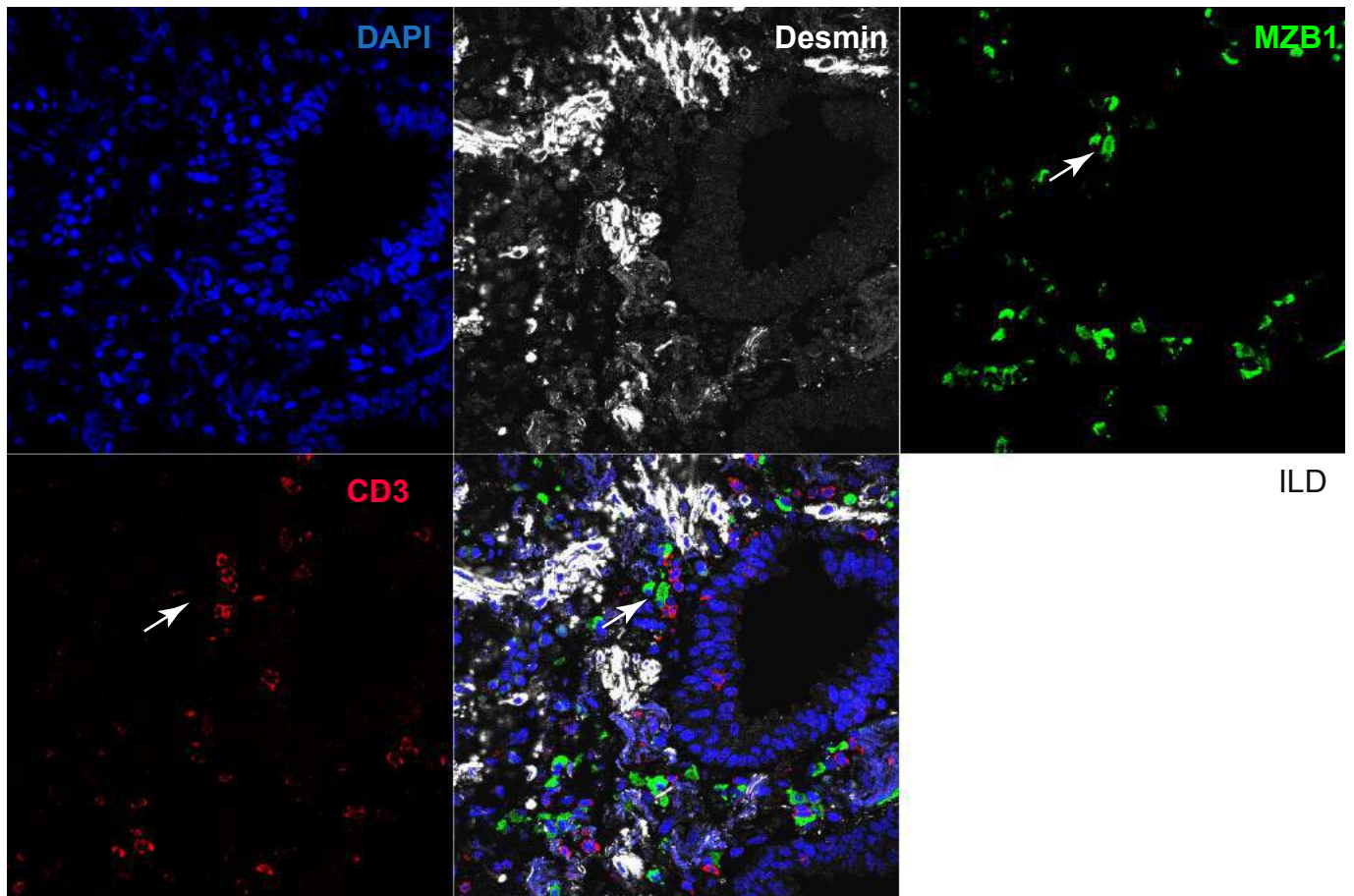


Figure S3

A



B

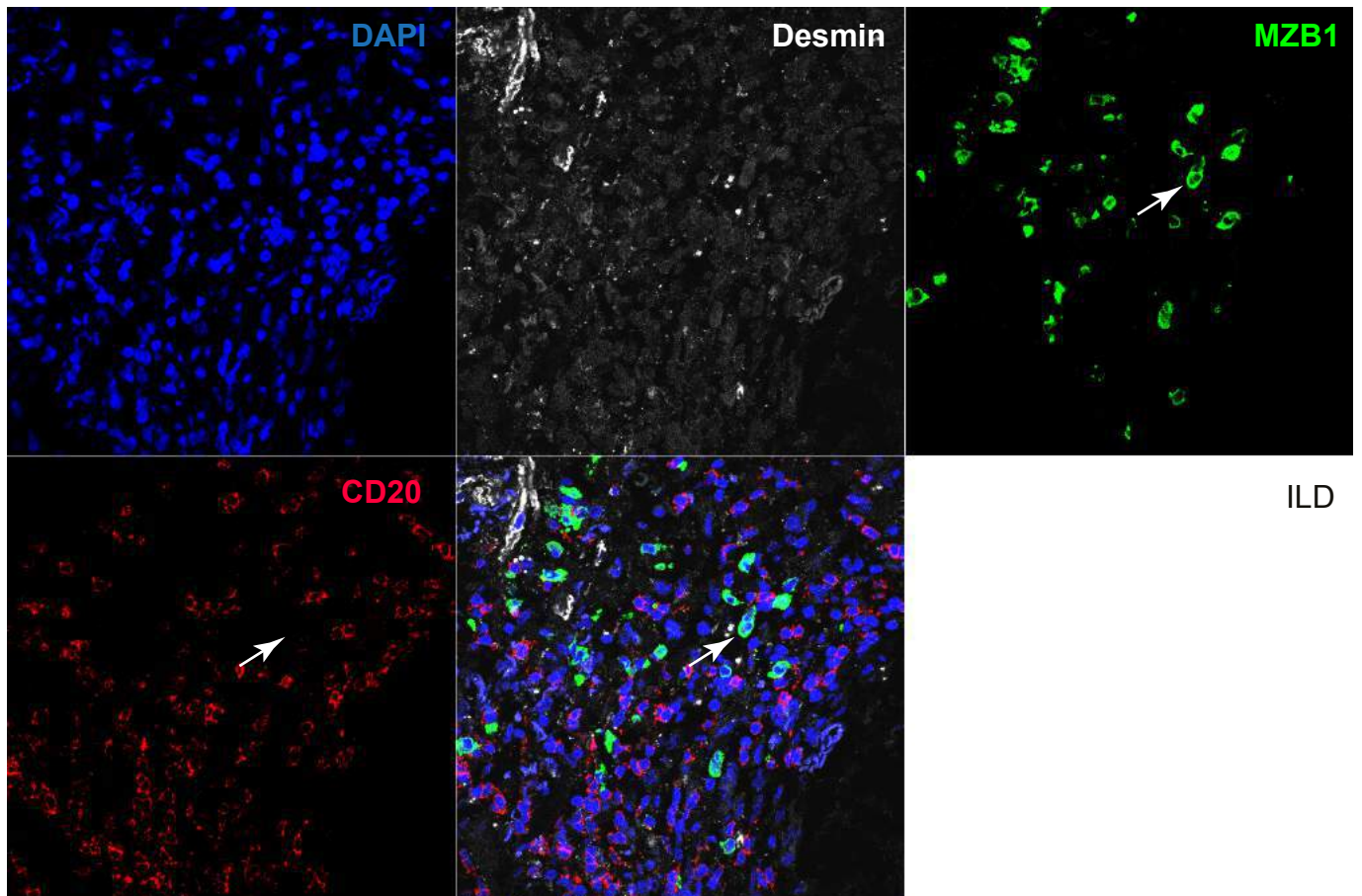
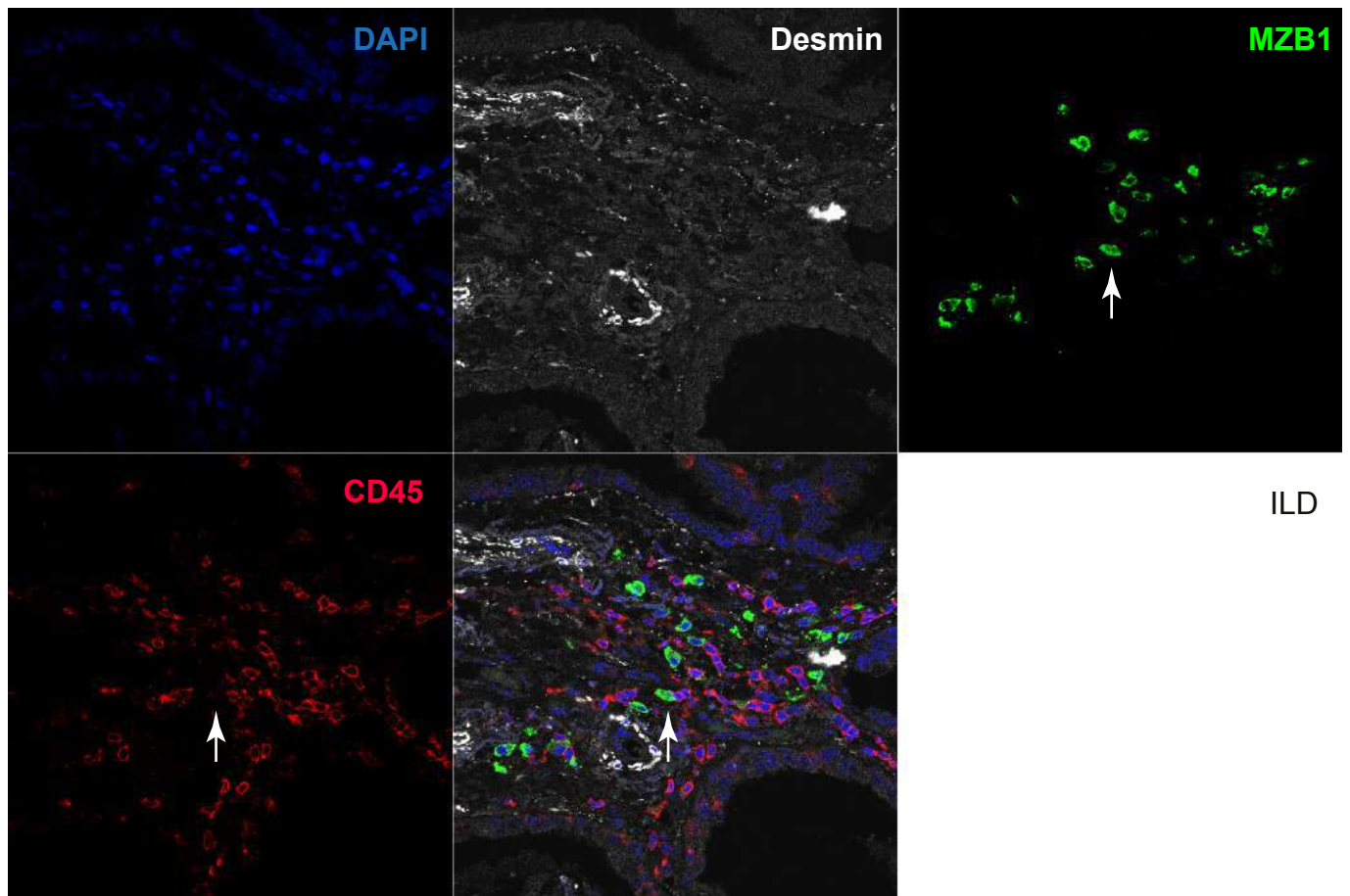


Figure S4

A



B

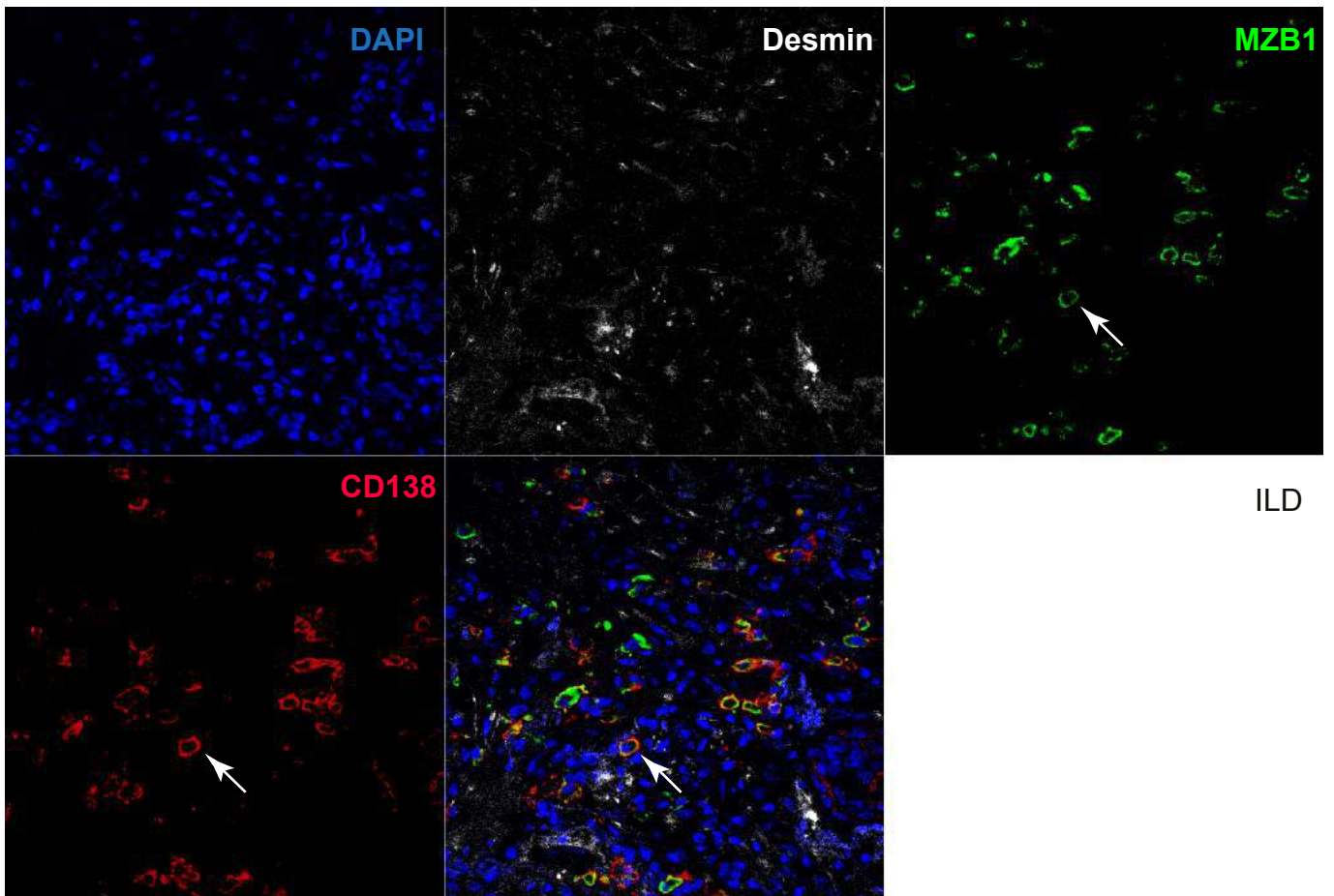
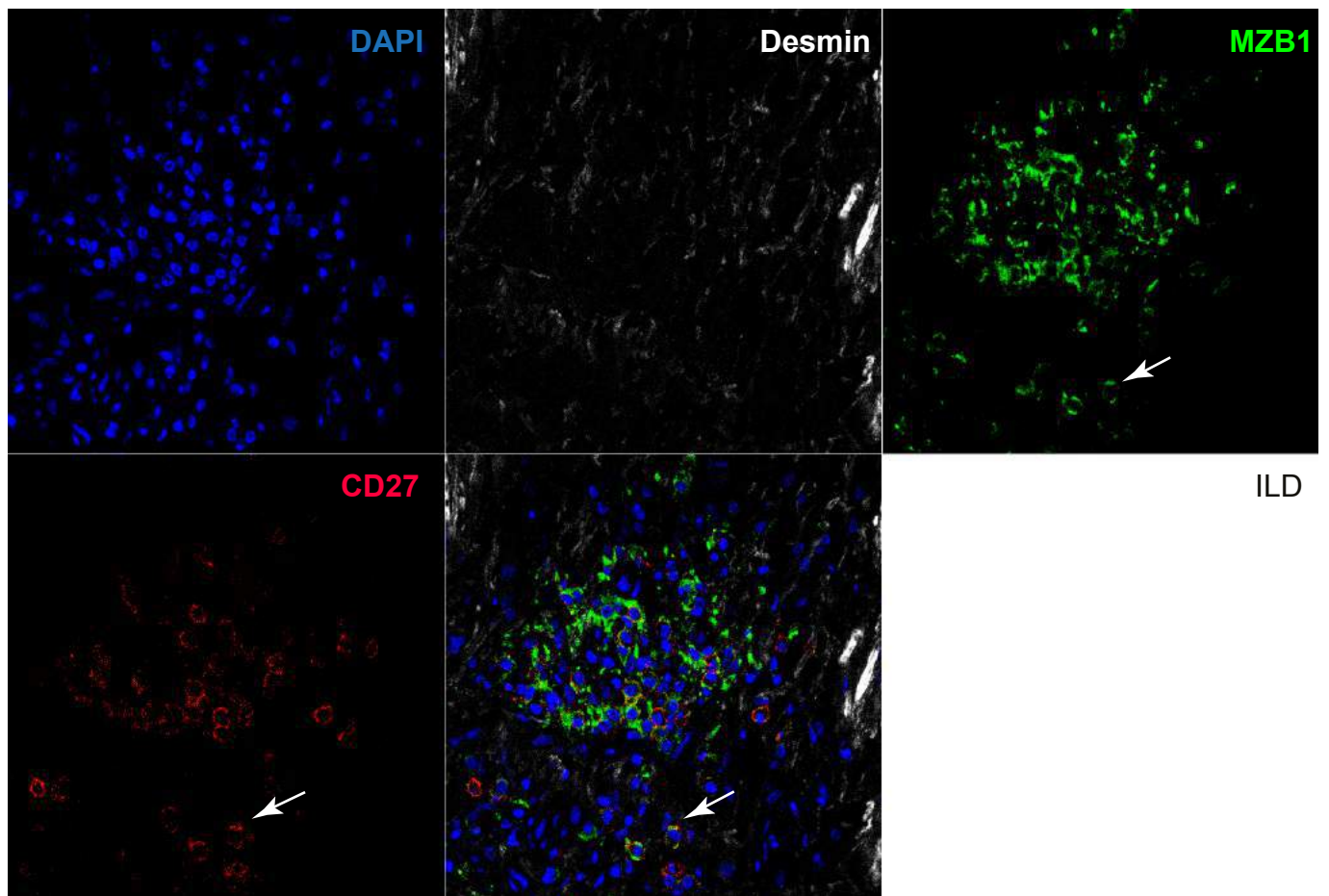


Figure S5

A



B

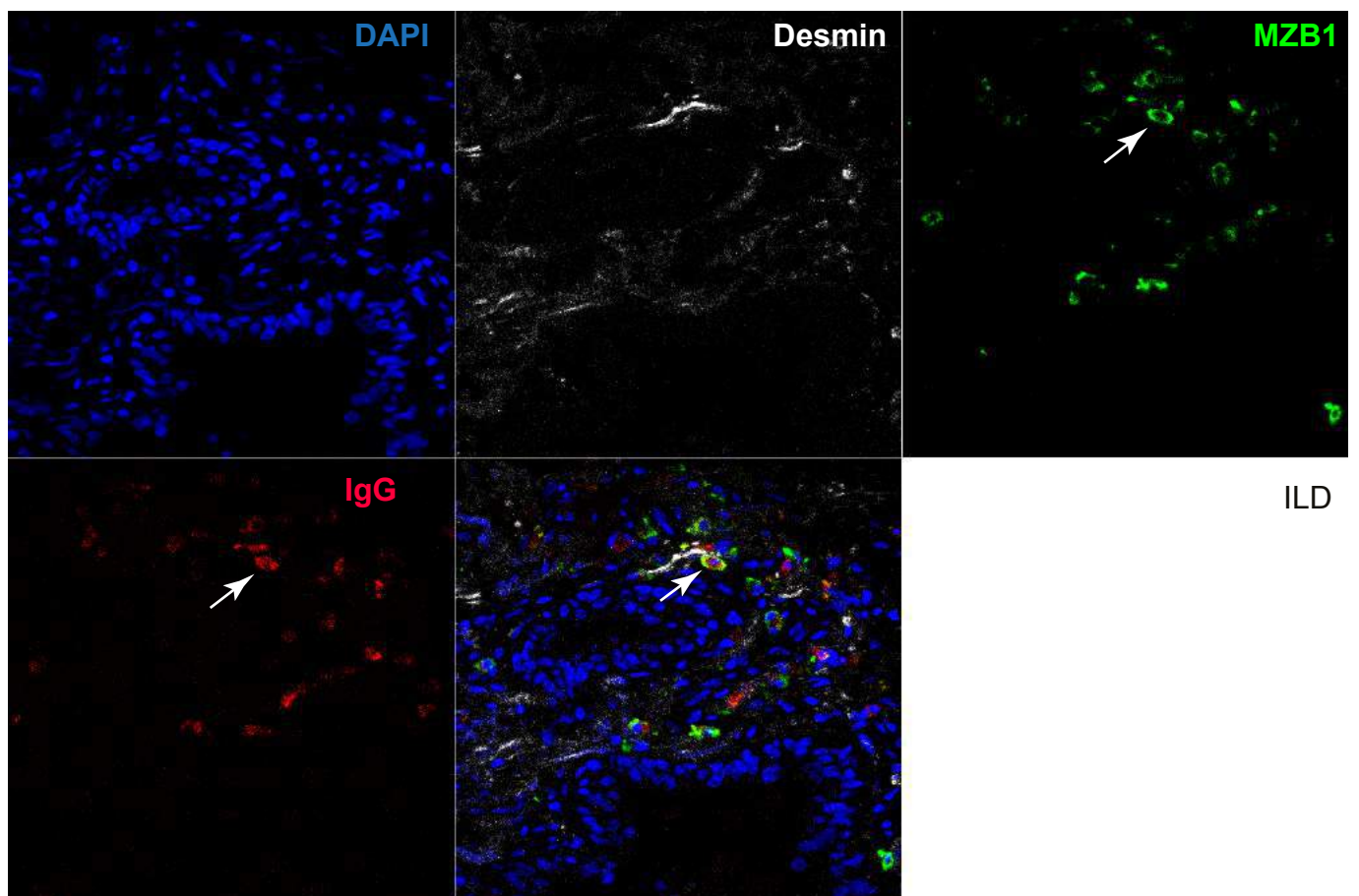


Figure S6

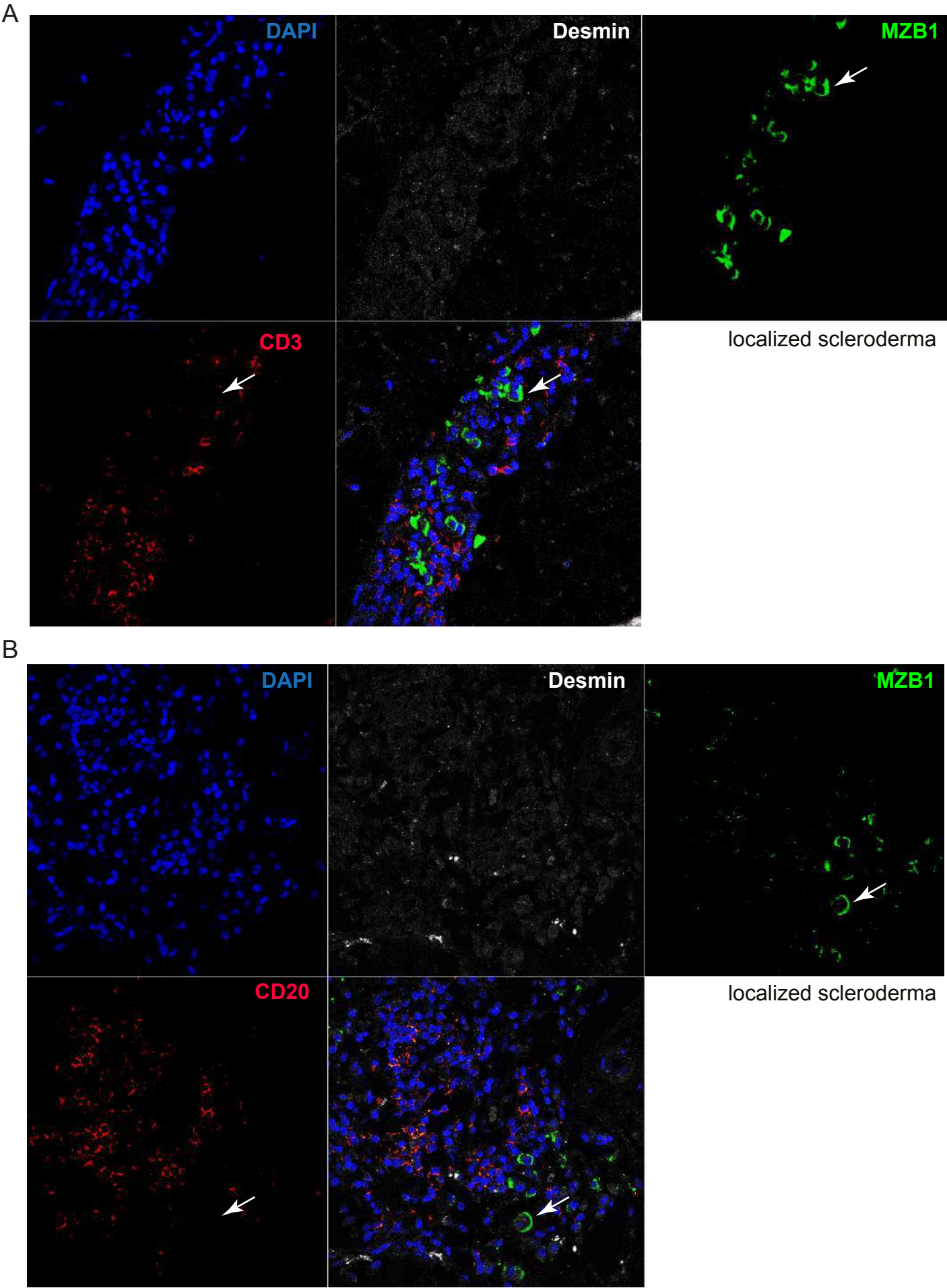
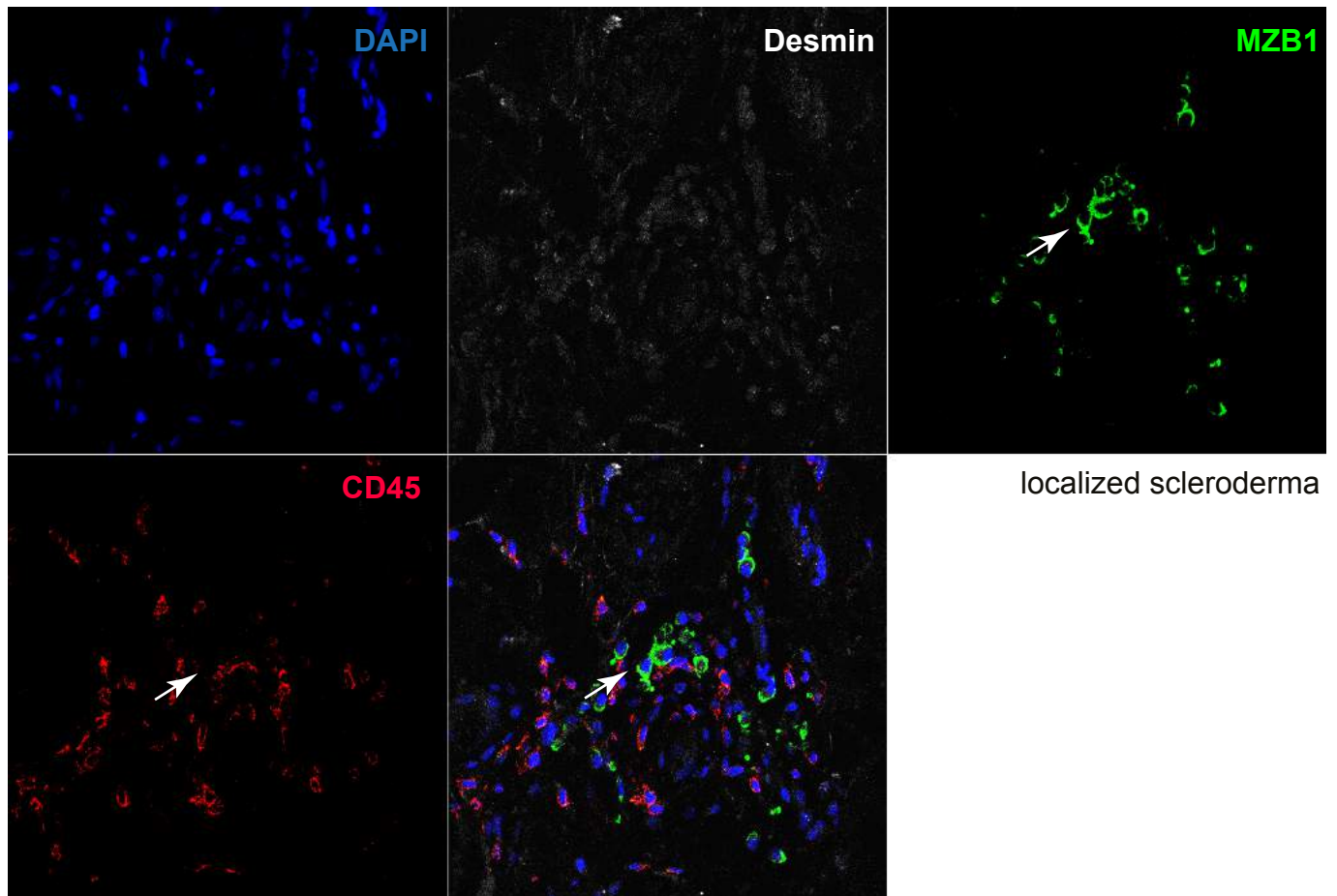


Figure S7

A



B

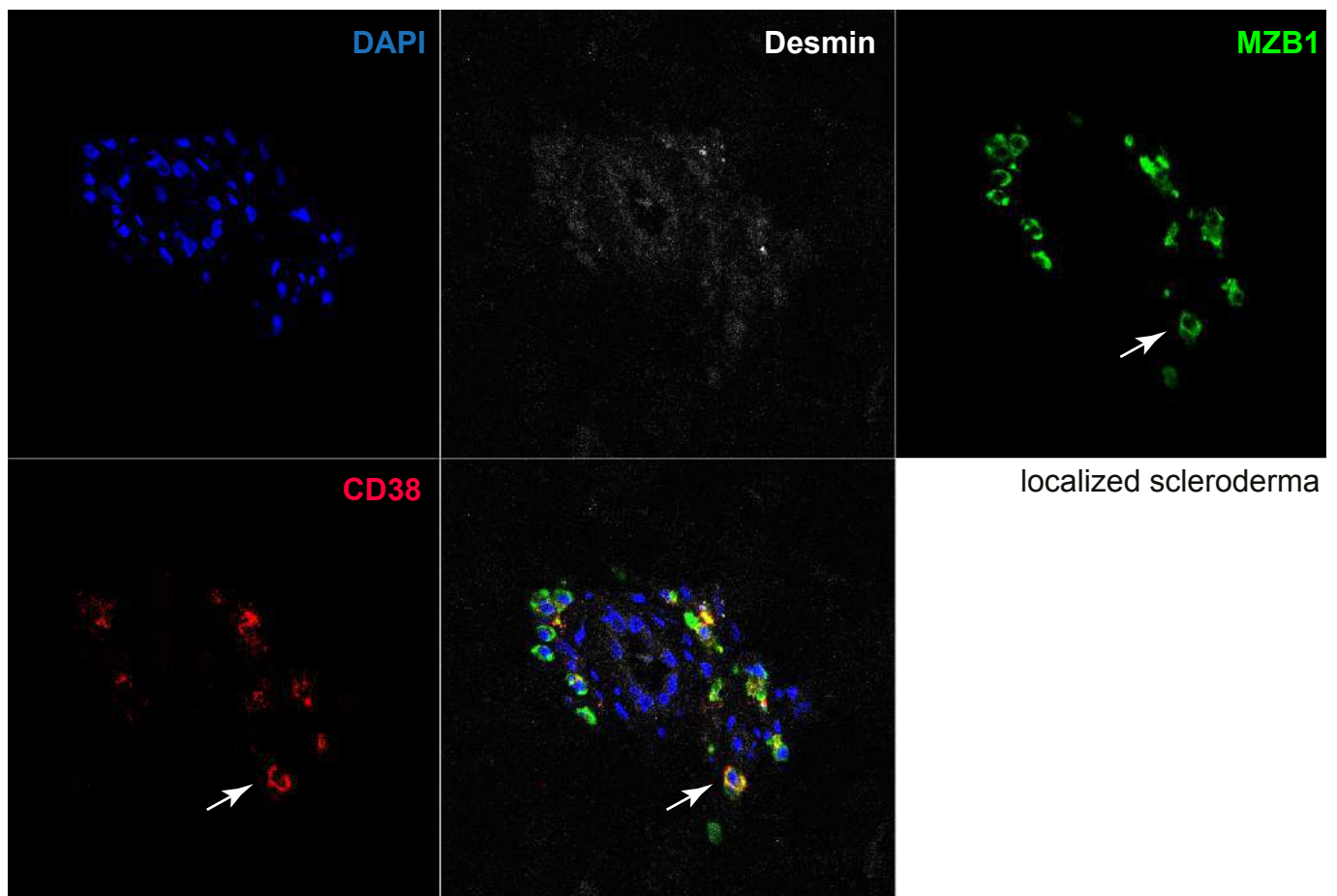


Figure S8

