

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
pO2 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, \pm SD	56,1 \pm 10,0	53,7 \pm 10,5	42,5 \pm 4,9	56,3 \pm 8,1	51,0 \pm 7,0	55,1 \pm 6,9	54,4 \pm 8,8
PFT							
VC, l (% pred)	1,8 \pm 0,2 (38,0 \pm 2,8)	1,5 \pm 0,7 (42,3 \pm 22,5)	2,1 \pm 0,01 (39,5 \pm 9,2)	1,8 \pm 0,2 (38,0 \pm 2,8)	1,5 \pm 0,4 (39,8 \pm 9,3)	1,9 \pm 1,3 (43,7 \pm 28,6)	1,7 \pm 0,6 (40,6 \pm 15,0)
TLC, l (% pred)	3,11 \pm 0,5 (41,5 \pm 0,7)	3,2 \pm 1,2 (54,8 \pm 20,7)	3,1 \pm 0,5 (41,5 \pm 0,7)	2,67 (36) *	2,7 \pm 0,5 (45,7 \pm 6,0)	4,5 \pm 1,7 (63,0 \pm 28,3)	3,2 \pm 1,0 (49,2 \pm 14,2)
RV, l (% pred)	1,06 \pm 0,5 (50,0 \pm 21,2)	1,6 \pm 0,5 (70,2 \pm 32,29)	1,06 \pm 0,5 (50 \pm 21,2)	0,95 (38) *	1,2 \pm 0,3 (59,3 \pm 14,3)	2,3 \pm 0,2 (100,0 \pm 4,2)	1,5 \pm 0,6 (68,5 \pm 35,2)
FEV1, l (% pred)	1,69 \pm 0,3 (40,5 \pm 2,1)	1,2 \pm 0,5 (42,7 \pm 16,5)	1,69 \pm 0,3 (40,5 \pm 2,1)	1,73 \pm 0,1 (46,5 \pm 0,7)	1,3 \pm 0,4 (43,1 \pm 11,6)	1,1 \pm 0,3 (33,3 \pm 9,1)	1,4 \pm 0,5 (42,7 \pm 14,5)
DLCO, % pred.	10 \pm 2,8 n=5	20,5 \pm 10,1 n=4°	10 \pm 2,8	x	22,8 \pm 3,6 n=5°	24 n=1°	19,7 \pm 7,7
pO2 at rest, mmHg	36,0 \pm 7,1	49,2 \pm 8,2	36 \pm 7,1	x	50,6 \pm 6,8	56 \pm 1,4	48,3 \pm 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

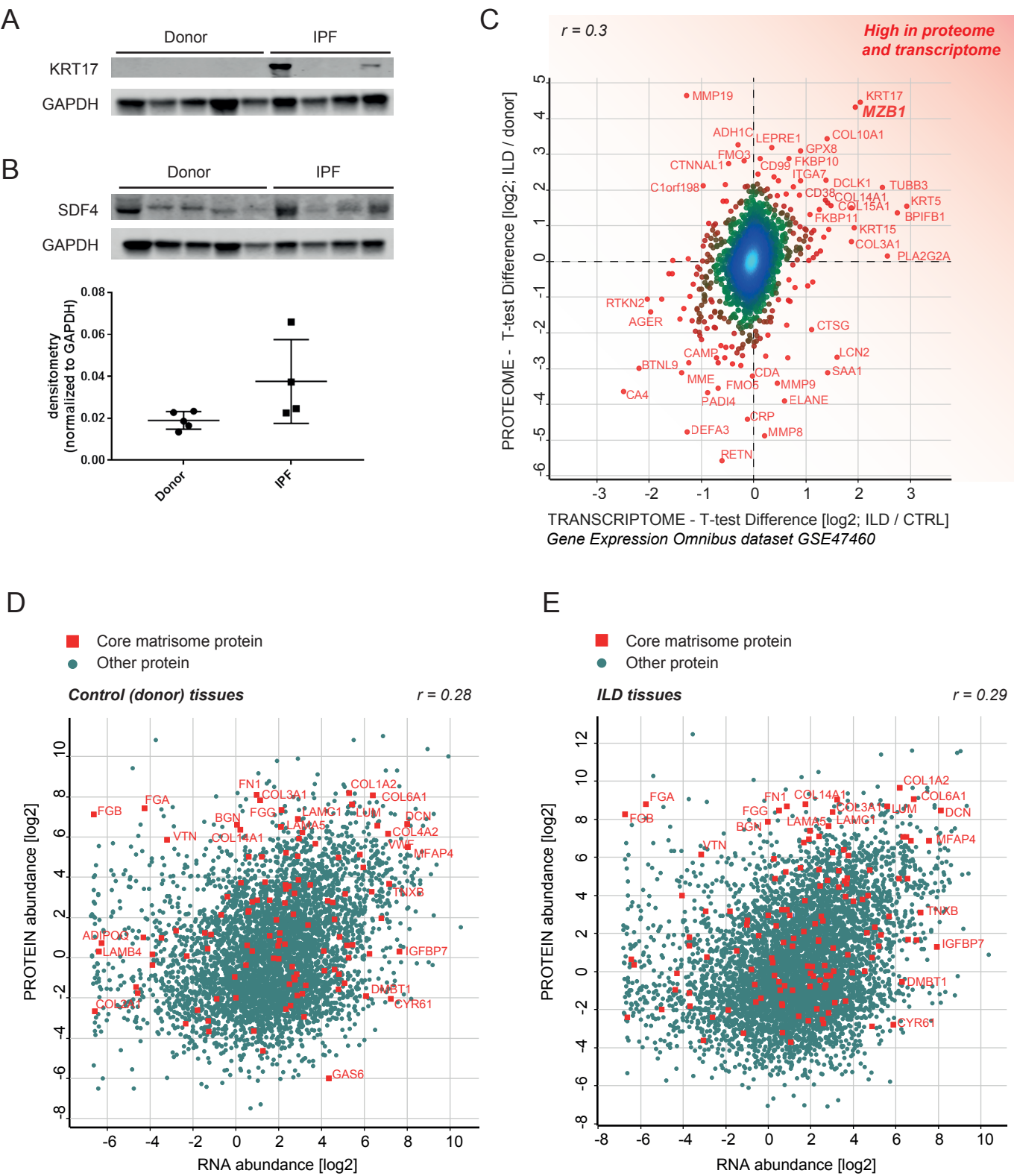
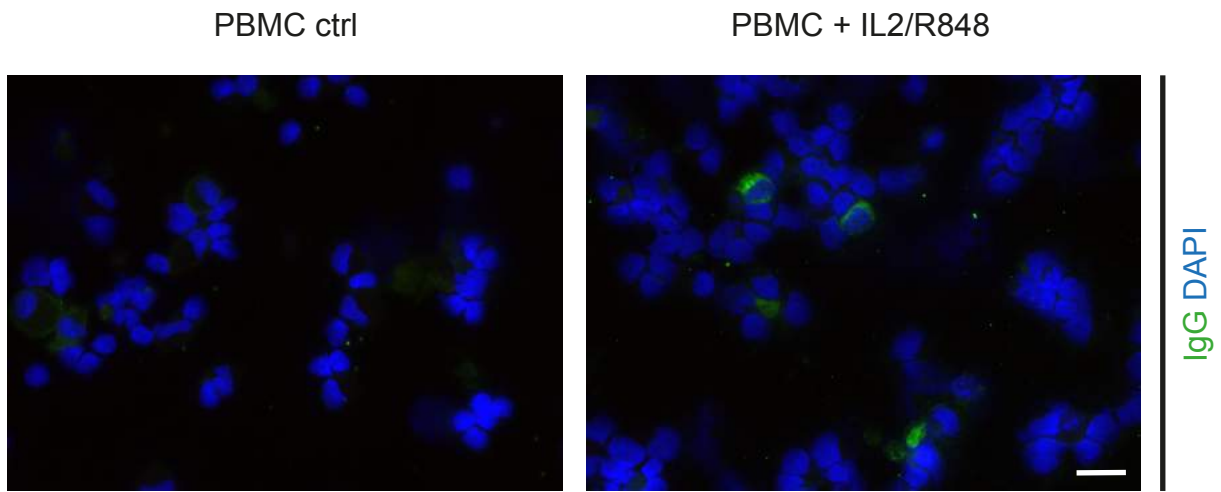
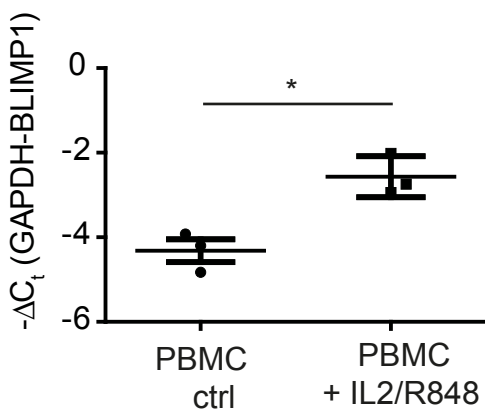


Figure S2

A



B



C

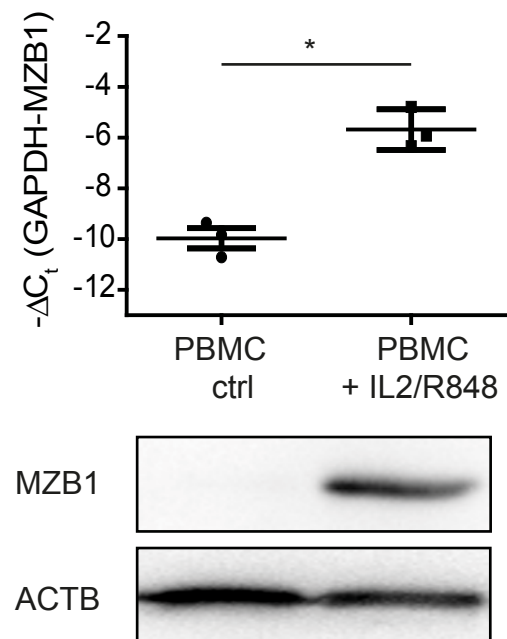
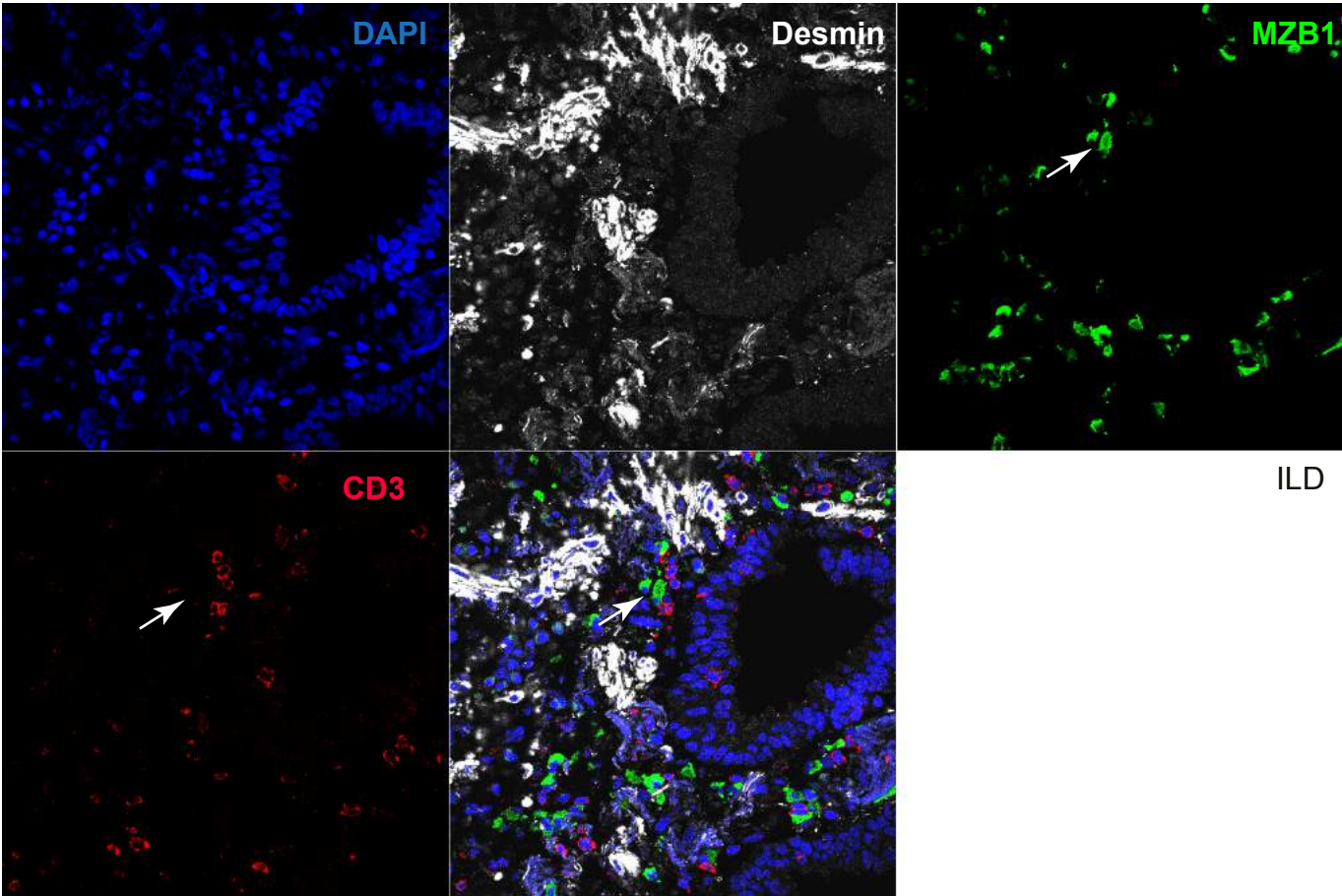


Figure S3

A



B

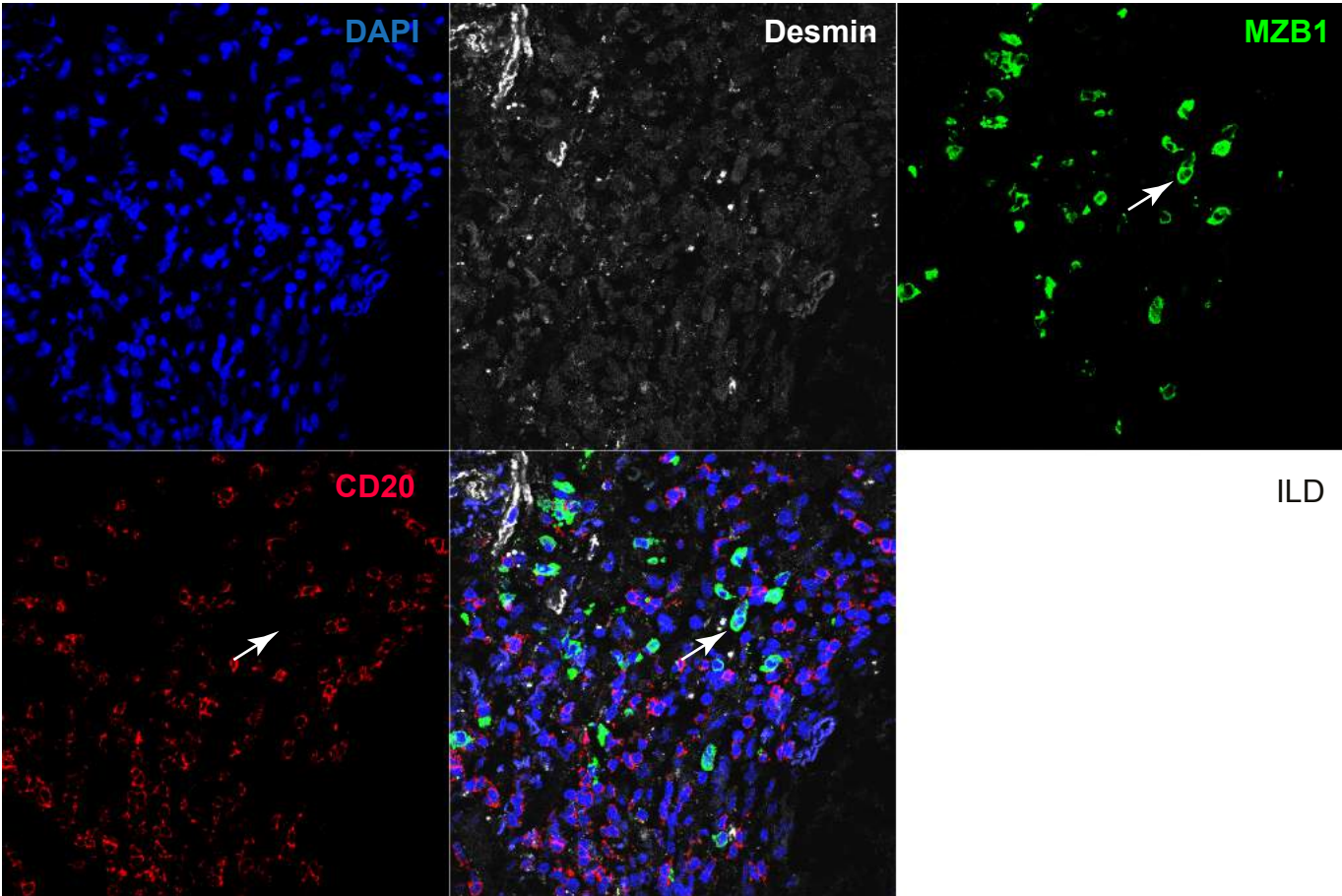
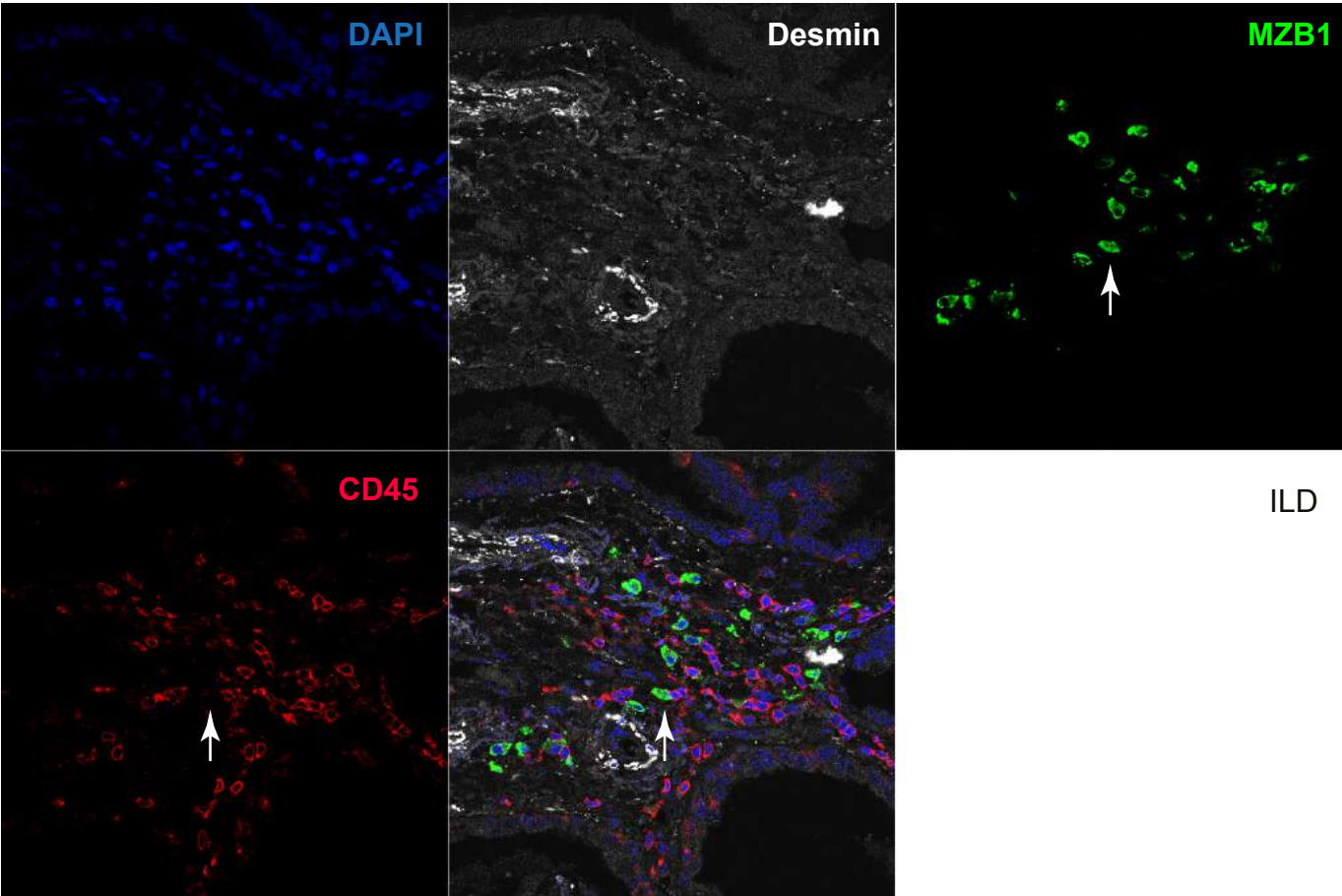


Figure S4

A



B

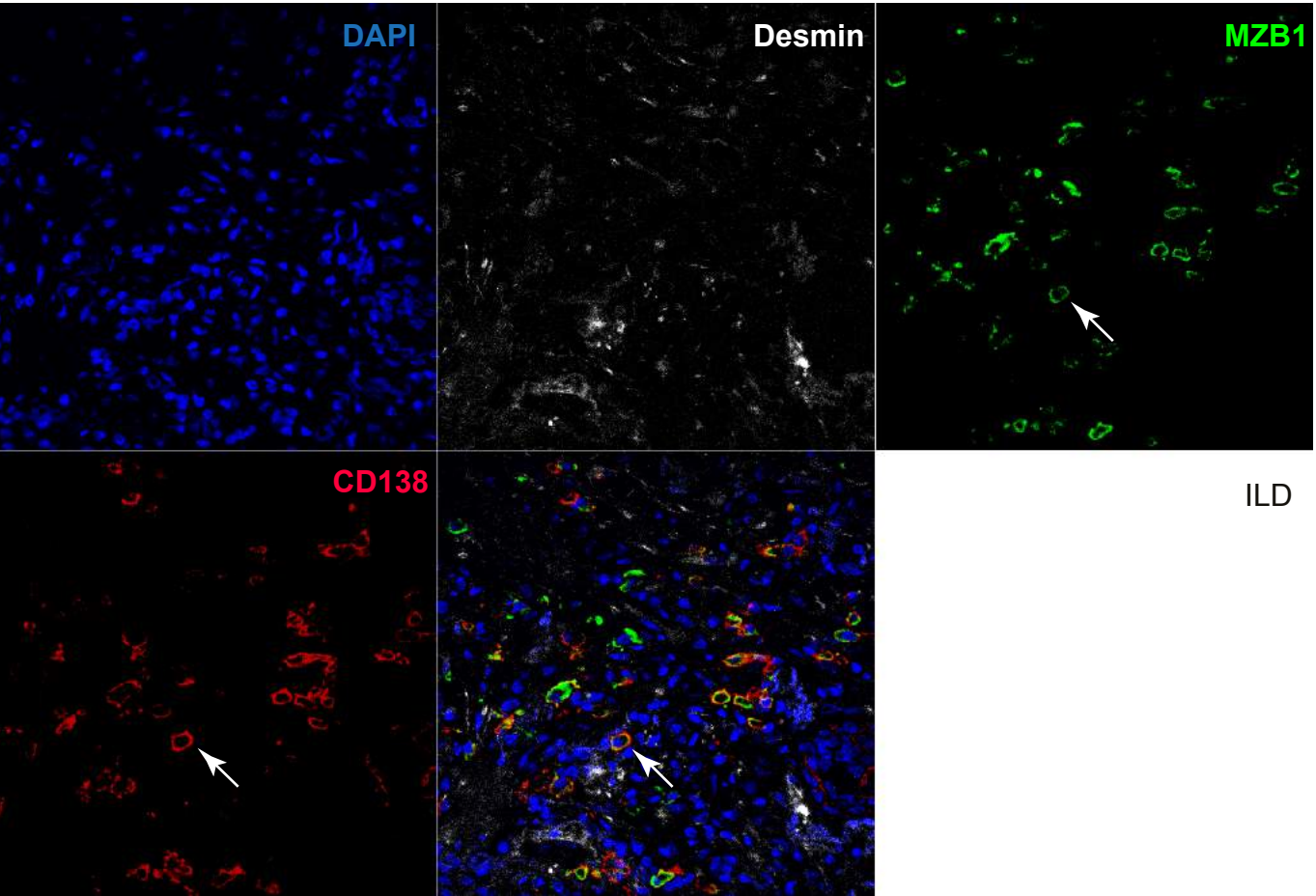
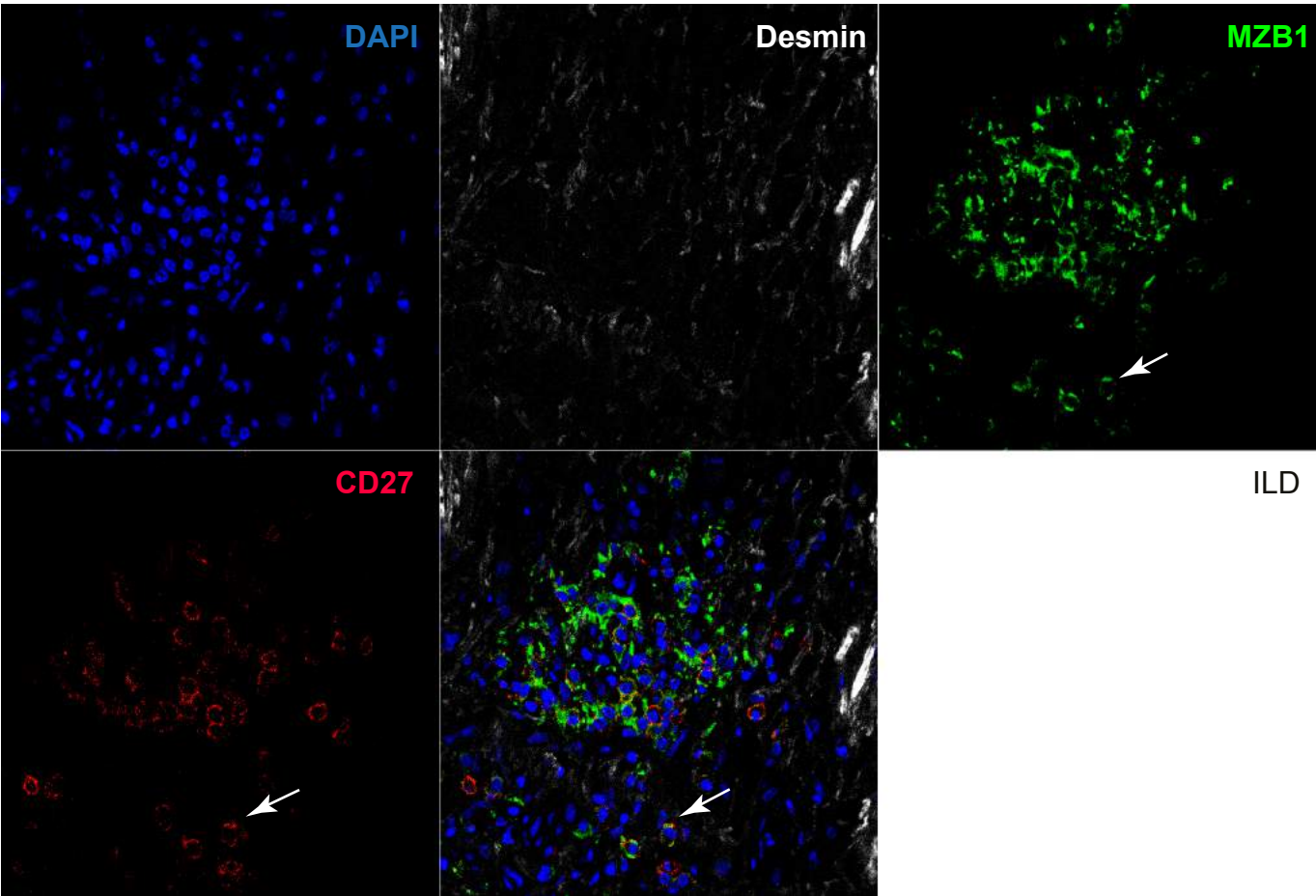


Figure S5

A



B

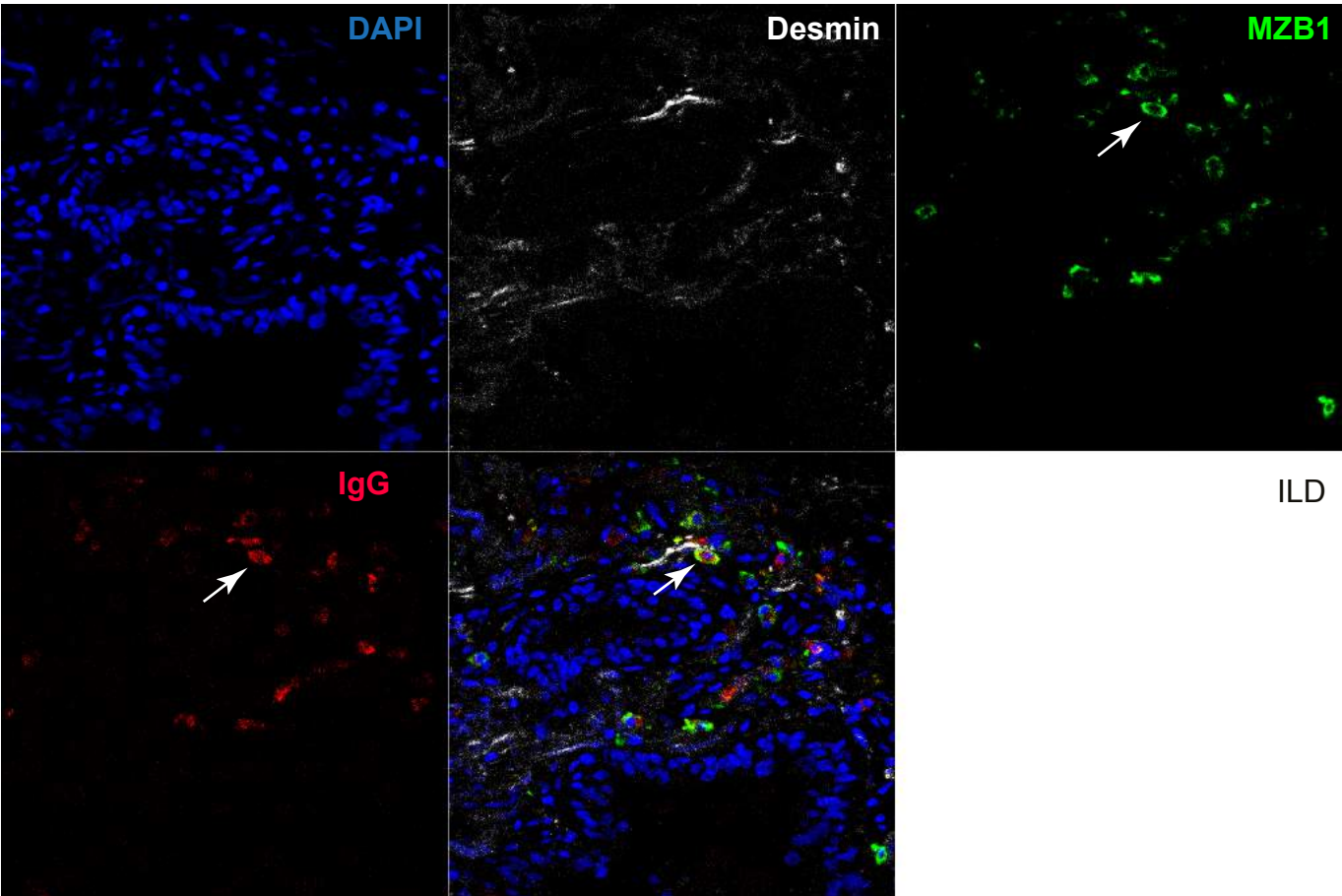


Figure S6

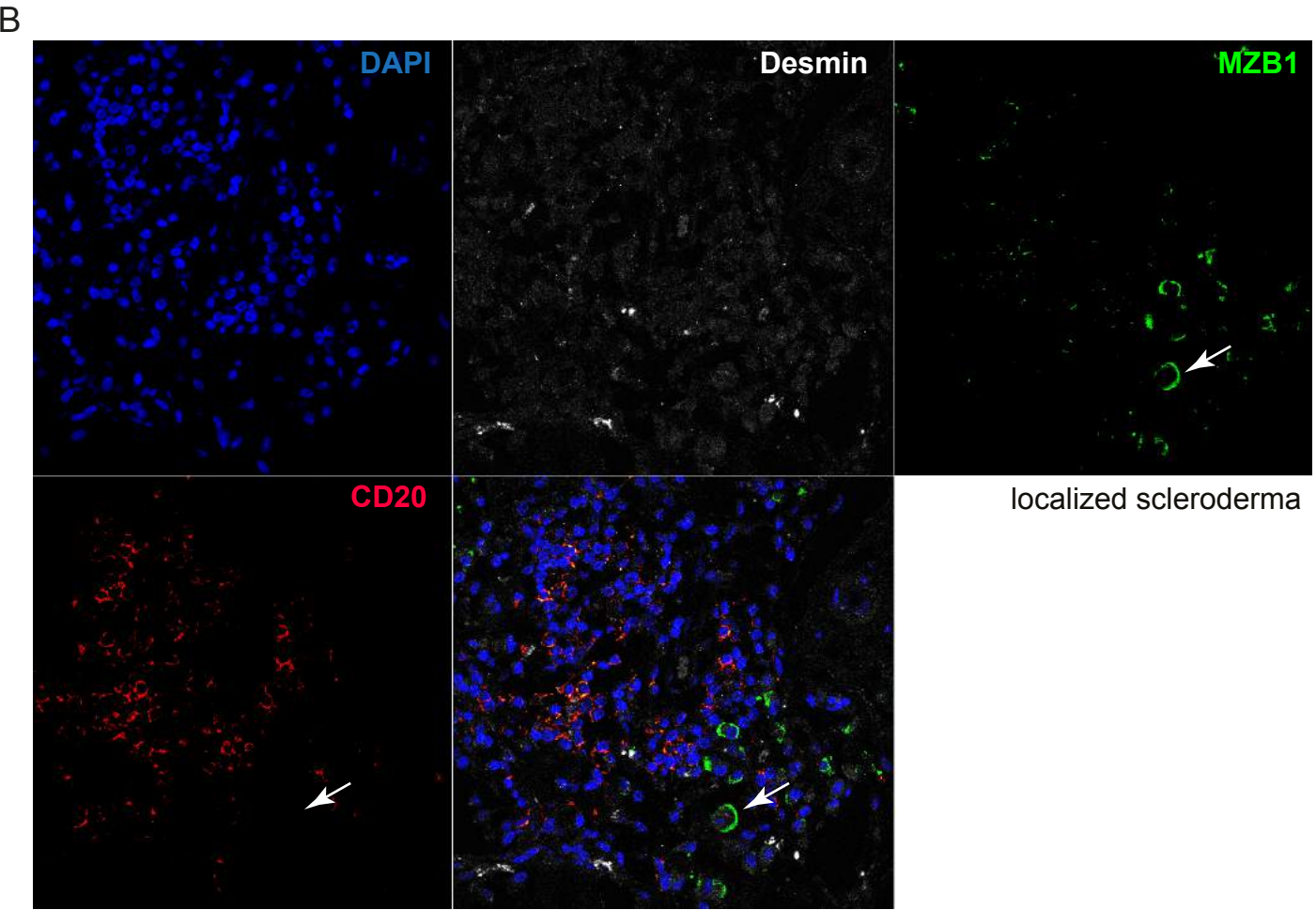
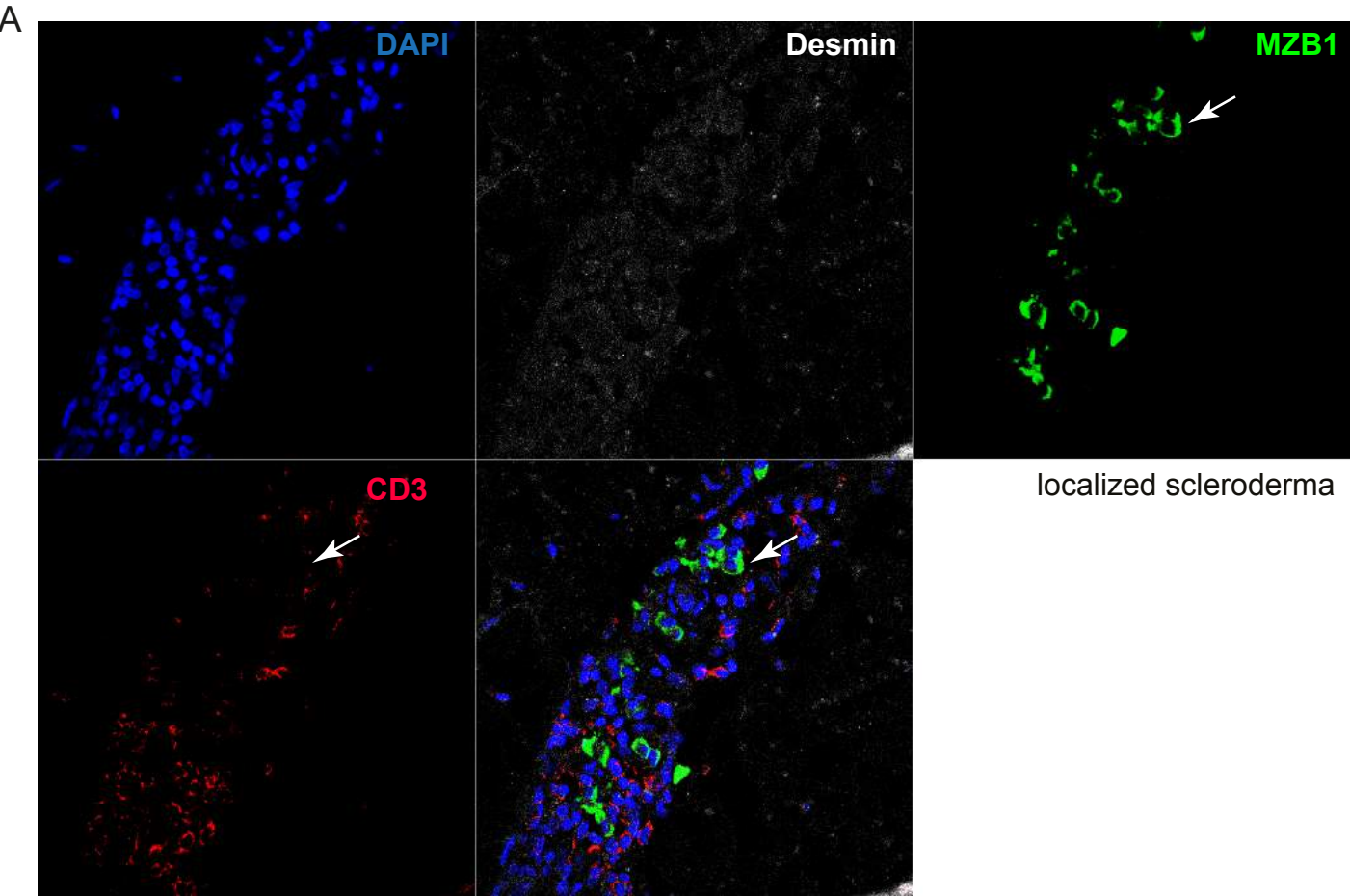
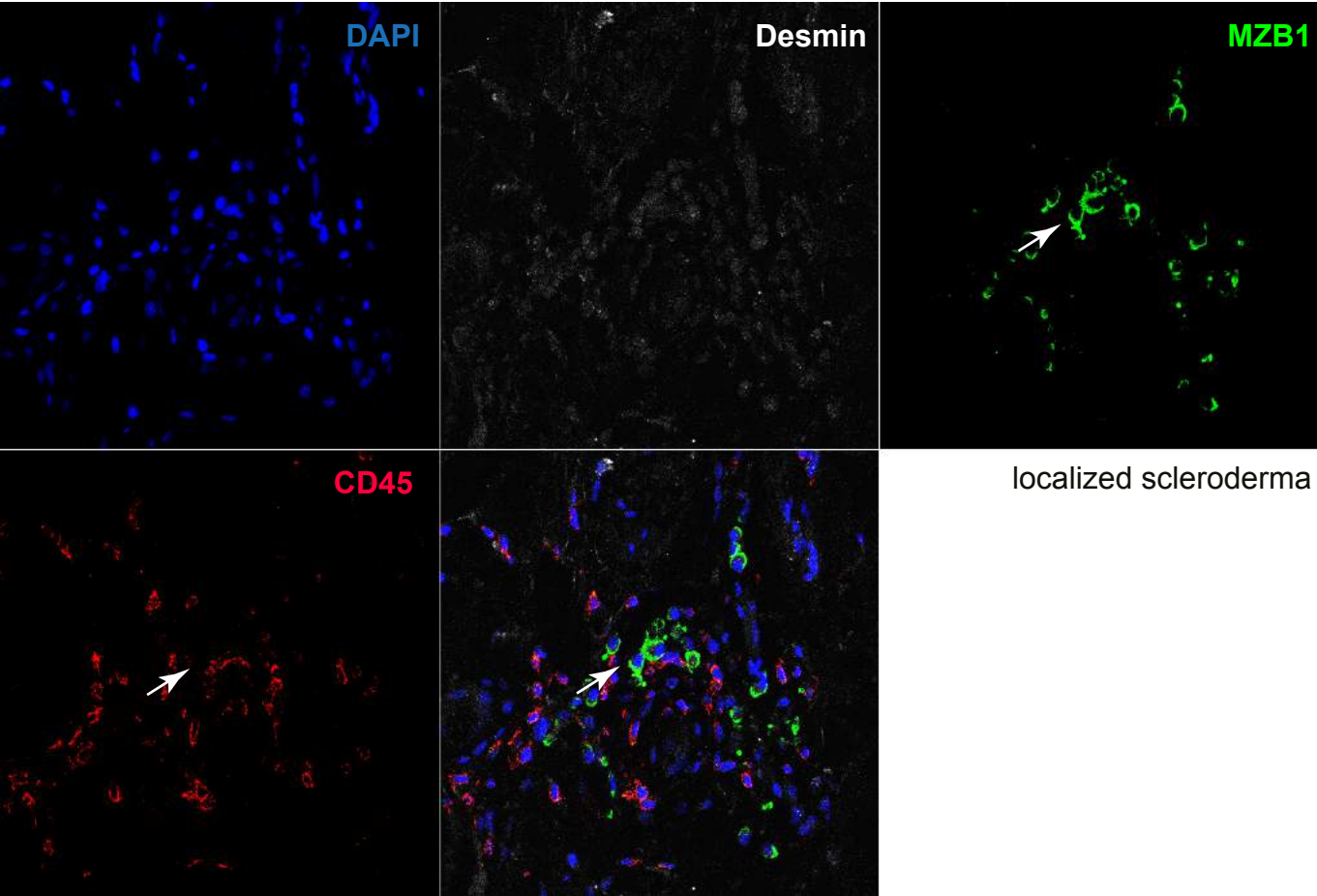


Figure S7

A



B

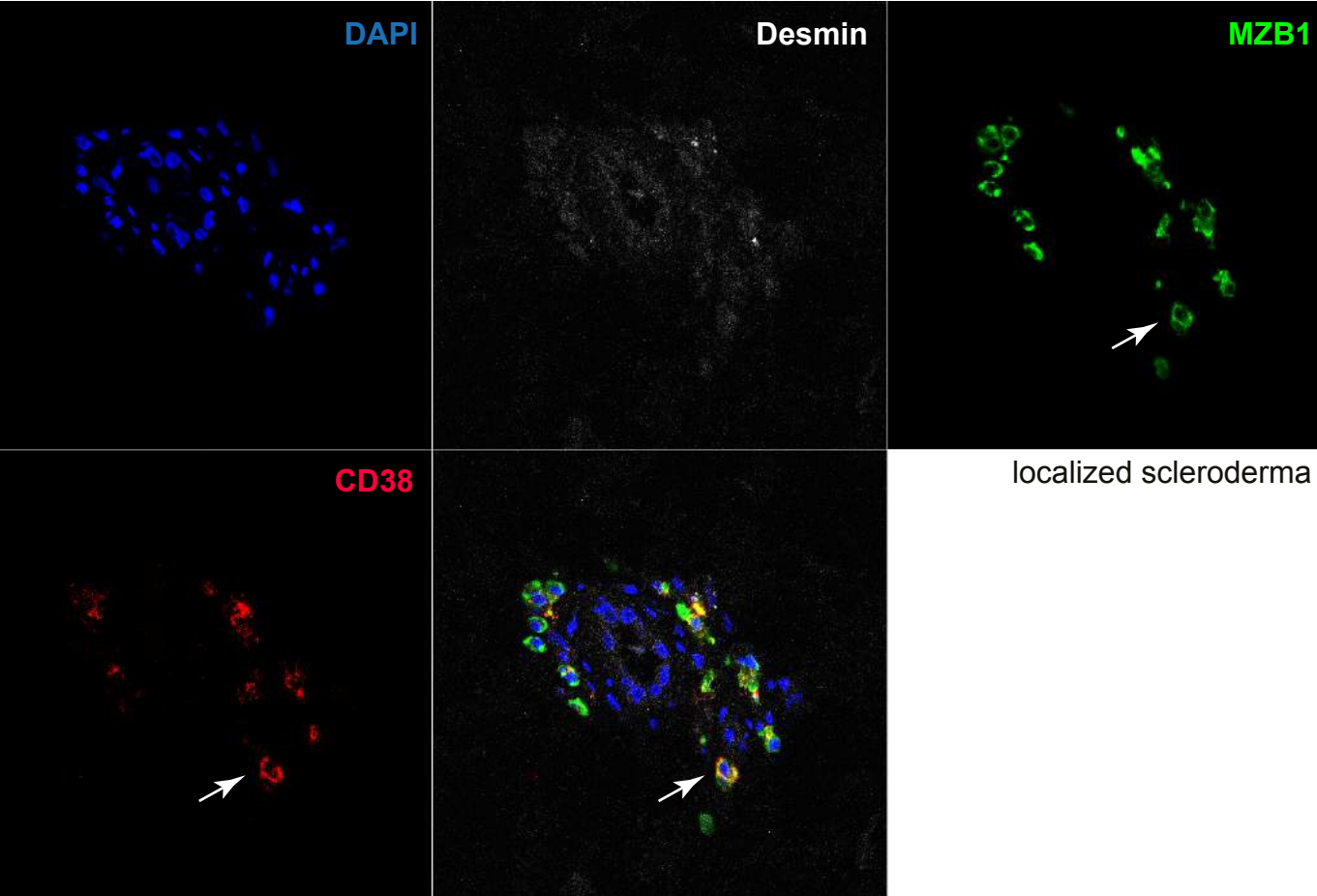


Figure S8

