Data Supplement

Supplementary Methods

Patient Cohorts and Clinical Data

In this study, 90 patients (76.7% female, median age 57.5 years) from the University Hospital Zurich's (derivation cohort) and 66 patients (75.8% female, median age 61.0 years) from the Oslo University Hospital's prospective SSc patient cohorts (external validation cohort) were included. Both centres are part of the EUSTAR (European Scleroderma Trial and Research) network [1]. Patients were retrospectively selected based on the following criteria:

(1) Fulfilment of diagnosis of early/mild SSc according to the Very Early Diagnosis of Systemic Sclerosis (VEDOSS) criteria [2] or established disease according to the 2013 American College of Rheumatology//European league against rheumatism (ACR/EULAR) classification criteria [3],
(2) Presence of ILD on HRCT as determined by a senior radiologist, and
(3) Availability of an HRCT scan with the following settings:
    (a) Slice thickness between 0.6 and 3 mm,
    (b) One of the following lung kernels available (B60f, B70f, Bl64d, LUNG),
    (c) Filtered-back projection as reconstruction algorithm, and
    (d) CT image acquired in full inspiration.

For each patient, demographic and clinical parameters, including age, sex, SSc disease duration and subset, the extent of skin involvement, autoantibody status, CRP levels, presence of pulmonary hypertension according to right heart catheterisation or echocardiography as judged by the local investigators, and pulmonary function test (PFT) parameters were retrieved from the local patients' records. The recorded PFT parameters (expressed as % predicted values) included forced vital capacity (FVC), forced expiratory volume in 1 second (FEV1), and diffusing capacity for carbon monoxide (DLCO). Data from the 6-minute walk test (6-MWT), including walk distance, oxygen saturation (% SpO_2) before and after the test, and Borg scale of perceived exertion (Borg CR-10) [4], were only available for the derivation cohort. Disease duration of SSc was calculated as the difference between the date of first available CT and the date of manifestation of the first non-Raynaud's symptom. The follow-up period was defined as the time interval between the baseline visit and the last available follow-up visit for every patient. The mean follow-up time for the derivation cohort was 66.1 (± 30.1) months and 43.9 (± 30.9) months for the external validation cohort. All outcome events occurring in this period were considered in this study. As outcomes for SSc-ILD, we selected progression-free survival, which was defined as the time from the date of the HRCT to the date of the first occurrence of ILD progression. The primary endpoint for progression-free survival was the progression of ILD defined as a relative decline in FVC% predicted from baseline to follow-up of ≥ 15% based on the criteria recommended for idiopathic pulmonary fibrosis trials by the American Thoracic Society/European Respiratory Society and previous clinical trials in SSc-ILD [5–8]. As a secondary and exploratory endpoint, we used a recently proposed FVC-DLCO composite index, in which progression is defined as either

a relative decrease in FVC% predicted of ≥15% or a relative decline in FVC% predicted of ≥10% combined with DLCO% predicted of ≥15% [9]. As further exploratory and non-lung function-based outcome measures for SSc-ILD, we selected 1) visual ILD progression on HRCT and 2) overall survival, which were defined as the time from the date of the HRCT to the date of the first occurrence of visual ILD progression on HRCT or all-cause death, respectively.

The vital status was determined based on the electronic patients' records.

The local ethics committees approved the study (approval numbers: pre-BASEC-EK-839 (KEK-no.-2016-01515), KEK-ZH-no. 2010-158/5, BASEC-no. 2018-02165, BASEC-no. 2018-01873) and written informed consent was obtained from every patient.

Pulmonary Function Tests

In brief, spirometry, body plethysmography, and DLCO measurements were performed by trained technicians in the Department of Pneumology of the University Hospital Zurich and Oslo. Measures included, among others FVC, FEV1, TLC, VC, and DLCO. The PFTs were performed following established protocols [10–13]. Since the PFTs were performed as part of the routine diagnostics, the respective pulmonologist on call interpreted the results and provided a written report, including the measured values and their interpretation.

HRCT Image Acquisition and Visual CT Analysis

The settings used for the acquisition of HRCT images are summarized in Supplementary Table S9. All HRCT images were assessed for the presence of characteristic visual

features of ILD, including ground glass opacification (GGO), reticular changes, traction bronchiectasis, emphysema, and honeycombing. In addition, the radiological subtype (usual interstitial pneumonia (UIP), nonspecific interstitial pneumonia (NSIP), or diffuse interstitial pneumonia (DIP)) was determined. UIP includes the radiological diagnosis of both, “definite” and “probable” UIP [14, 15]. The extent of lung fibrosis was determined visually by the clinical radiologists in charge of routine diagnostics. All sections from the lung apex to the hemidiaphragm were assessed. All CT scans from both cohorts were re-evaluated by a long-standing expert on chest radiology (T.F.). The extent of lung fibrosis on HRCT, defined as the presence of reticular changes and/or honeycombing, was categorized as either <20% or ≥ 20% in relation to the total lung volume. For visual analysis of ILD progression on HRCT, all available follow-up HRCT scans from every patient were extracted from the electronic patient’s records. Due to the differences in the types of scanners and kernels used, ILD progression on HRCT was visually assessed by a senior radiologist and expert in chest radiology (T.F.). ILD progression was defined as an increase in ground-glass, reticulation or honeycombing including more than a second lobule or the transition of ground-glass into reticulation or honeycombing. All visual analyses were performed using a standard picture archiving and communication system workstation (Impax, Version 6.5.5.1033; Agfa-Gevaert, Mortsel, Belgium) and a high definition liquid crystal display monitor (BARCO; Medical Imaging Systems, Kortrijk, Belgium).

CT Segmentation and Extraction of Radiomic Features

The left and right lung lobes of each patient were semi-automatically segmented by two readers (J.S., M.B.) using the “region grow” function (lower threshold -950 HU, upper threshold: -300 HU) of MIM software (version 6.9.2, MIM Software Inc., Cleveland, Ohio, United States). Manual corrections were applied when computationally defined tissue borders did not coincide with the actual lung borders. In addition, pulmonary hilar vessels and atelectatic lung areas were carefully excluded from the regions of interest.

Radiomic analysis was performed on merged structures of both lung lobes using the in-house developed software Z-Rad based on Python programming language 2.7. For radiomics analysis, CT images were resized to isotropic voxels of 2.75 mm and discretized to a fixed bin size of 50 HU. In total, 1,386 radiomic features were calculated per lung (HU limits: -1000 HU to 200 HU), corresponding to the following radiomic feature classes:

(1) Intensity or histogram features (*n* = 17),

(2) Texture features (*n* = 137) of the *Gray Level Co-occurrence Matrix* (*n* = 52; GLCM), the *Neighborhood Gray Tone Difference Matrix* (*n* = 5; NGTDM), the *Gray Level Run Length Matrix* (*n* = 32); GLRLM), the *Gray Level Size Zone Matrix* (*n* = 16; GLSZM), the *Gray Level Distance Matrix* (*n* = 16; GLDZM) and the *Neighboring Gray Level Dependence Matrix* (*n* = 16; NGLDM), and

(3) Wavelet features (*n* = 1,232).

The first class of radiomic features relates to the histogram or distribution of voxel intensities using first-order statistics (e.g. mean, standard deviation, skewness and kurtosis) and quantifies tissue intensity characteristics. The second category, including the texture features, describes the intra-tissue heterogeneity by calculating the statistical, spatial inter-relationship between neighbouring voxel intensities [16]. The third group of features, the wavelet features, calculates the intensity and texture features after wavelet decompositions of the original image using eight different coiflet filters (high-pass to low-pass filters), thereby focusing the features on different frequency ranges [17].

A list of all radiomic features is provided in Supplementary File 1. Radiomic feature definitions were based on the Imaging Biomarker Standardization Initiative report by Zwanenburg *et al. [18]*.

Assessment of Radiomic Feature Stability

Intraclass correlation (ICC) analysis was performed to assess the stability of radiomic features against intra- and inter-operator variability in the semi-automated segmentation process (Supplementary Figure S1). For inter-operator ICC analysis, three examiners (J.S., M.B., C.B.), and for intra-operator ICC analysis, one examiner (J.S.) twice, independently contoured 15 randomly selected SSc patients from the derivation (Zurich) cohort, and radiomic features were extracted from the multiple delineation structures. The ICC coefficient for every radiomic feature was quantified using two-way mixed effect models and applying the "consistency" method (ICC(3,1)) according to [19] using "irr" package of R. Only features with good reproducibility defined as ICC ≥ 0.75 [20] were considered in further analyses.

Unsupervised Clustering

Unsupervised clustering was performed to identify groups of patients with similar radiomic feature patterns in the derivation cohort (Zurich; n=90). After confirmation of data clusterability by visual assessment of cluster tendency (VAT) and calculation of the Hopkin's statistic *H* (with $H > 0.5$ indicating clusterability) [21], the *k*-Means clustering algorithm [22] was applied to the z-scored radiomic data. Only robust radiomic features (ICC ≥ 0.75) entered the cluster analyses. The optimal number of clusters was determined by varying the number of k-clusters between 2 and 10 and selecting the optimal k concerning best visual separation and stability as determined by Jaccard bootstrapping (n = 1,000 iterations).

Building a Quantitative Radiomic Risk Score for SSc-ILD

The Zurich cohort was used as a derivation cohort to build and train the radiomic risk score for ILD progression (qRISSc). Patients with no follow-up and survival data available on the electronic patients' records were excluded from the analysis, resulting in a final dataset of 75 patients. For score building, we adapted a recently described approach by Lu et al. [23] for *z*-scored, radiomic features. Following Lu and colleagues, we selected radiomic features in two steps: 1) Cox regression and 2) penalized LASSO regression using "cox" family with 10-fold cross-validation. In the first step, we applied univariate Cox regression per radiomic feature only considering features with FDR of $p<0.005$. Features selected in step 1) underwent further reduction by LASSO. Only features with non-zero coefficients were retained, thereby removing strongly inter-correlated, redundant

features. Since limited by the modest sample size of the derivation cohort, we did not perform weighting of score features according to the coefficients from LASSO regression and assigned the same importance to each feature by dividing each standardized feature by the total number of features *j.* The final radiomic score was constructed as follows: $qRISSc = \sum_{i=1}^{j} \alpha_i f_i$ with α = $\frac{1}{j}$ being the feature weight and *f* being the values of *z*-transformed radiomic features.

After having selected features in steps 1) and 2) we searched for the significant cut-off value for Cox regression by applying the “cox” function from the “cutoff” package of R. Due to the modest sample size, we searched for two groups, i.e. “low” and “high” risk patients composed of at least 25% of subjects for the minority group. We selected the one that was significant after correction for multiple testing from the proposed pairs of cut-offs. Once a score was built, we fitted a univariate Cox regression model on the external validation cohort (Oslo). Kaplan-Meier plots were used to visualize the Cox regression results. As a reference model to qRISSc, we analogously build a radiomic score composed only of less complex, first-order densitometric (intensity) features, which have been previously explored for the quantification of disease extent and progression of SSc-ILD [24–27].

Multivariable Cox regression analyses were applied to analyse the predictive ability of conventional risk factors and qRISSc for progressive ILD in the pooled cohorts (n=156). Ten events per variable were required in the multivariable analyses, and the variables were selected based on literature evidence and expert opinion [28–30]. We reported the concordance index (C-index) as the general assessment of the quality of the model, the p-value of the whole model, and the hazard ratio (HR) with 95% confidence intervals for

the quantitative radiomic risk score. The C-index is equivalent to the area under the curve in ROC analysis and can also be used in Cox regression analysis [31].

Association Analyses with Clinical Characteristics

Association analyses were performed to explore associations of identified patient groups (*k*-Means clusters and risk groups) with clinical parameters.

Fisher's exact test was used to compare categorical, and Mann-Whitney U for comparison of continuous clinical variables, respectively.

Association Analyses with Biological Data

To reveal possible associations of the radiomic risk score with the underlying pathophysiology of ILD, correlation analysis with histological, proteomics and quantitative PCR data was performed. Since lung biopsies are only very rarely performed in SSc-ILD and thus matched patient tissue samples have not been available for molecular analyses, we conducted a cross-species correlation approach, using the mouse model of bleomycin-induced lung fibrosis as a model system for SSc-ILD. For this animal model, we have recently confirmed the transferability of radiomics signatures between mice and humans [32].

Animal Model of Experimental ILD

We applied the well-established preclinical model of bleomycin-induced lung fibrosis to model human SSc-ILD as described previously [33, 34]. In brief, 30 female, 8-week-old C57BL/6J-rj (Janvier Labs, Le Genest-Saint-Isle, France) were randomized and

intratracheally instilled with 2 U/kg bleomycin sulfate (BLM, Baxter 15,000 I.U., pharmacy of the canton Zurich, Switzerland) to induce ILD. For molecular and histological analyses, mice were sacrificed with carbon dioxide and subsequently transcardially perfused with ice-cold phosphate-buffered saline (PBS) solution to remove residual blood. All animal experiments were approved by the cantonal veterinary office (approval number ZH235-2018) and performed in strict compliance with the Swiss law for animal protection.

Proteomic Data

For proteomic analyses, frozen left lung lobes (blood-free) collected from PBS-perfused BLM-treated mice were homogenized in 8M urea/100mM Tris (pH 8.0) buffer supplemented with protease inhibitors using the FastPrep system (MP Biomedicals). After reduction and alkylation, and overnight protein precipitation with ice-cold acetone, 10 ug of the cleaned protein mixture were digested into peptides using a two-step digestion protocol (LysC for 2 h at 37 °C followed by Trypsin at room temperature overnight) and then subjected to liquid-chromatography-based tandem mass spectrometric analysis (LC-MS/MS). For LC-MS/MS, mouse samples were randomly allocated to the analysis by loading 800 ng onto a pre-column (C18 PepMap 100, 5 µm, 100 A, 300 µm i.d. x 5 mm length) at a flow rate of 50µL/min with solvent C (0.05% TFA in water/acetonitrile 98:2). After loading, peptides were eluted in backflush mode onto a home packed analytical Nano-column (Reprosil Pur C18-AQ, 1.9 µm, 120 A, 0.075 mm i.d. x 500 mm length) using an acetonitrile gradient of 5% to 40% solvent B (0.1% Formic Acid in water/acetonitrile 4,9:95) in 180 min at a flow rate of 250 nL/min. The column effluent was directly coupled to a Fusion LUMOS mass spectrometer (Thermo Fisher, Bremen;

Germany) via a nano-spray ESI source. Data acquisition was done in data-dependent mode with precursor ion scans recorded in the orbitrap with a resolution of 120’000 (at m/z=250) parallel to top speed HCD fragment spectra of the most intense precursor ions in the Linear trap for a cycle time of 3 seconds. Mass spectrometry data were processed by MaxQuant software, and set parameters are available in Supplementary Table S10. MaxQuant experimental design was such that the two repeated injections were combined, and match between runs allowed between all samples.

Histological and Immunohistochemical Data

Formalin-fixed paraffin-embedded lung sections (4 µm thick) from all BLM-treated mice were stained with Hematoxylin and Eosin (HE) for the examination of the overall tissue architecture, and the presence of cellular infiltrates and stained with Picrosirius Red (PSR) to visualize collagen deposition using standard protocols. Furthermore, specific immunohistochemical stainings for the pan-leukocyte marker CD45 and the myofibroblast marker alpha-smooth muscle actin (αSMA) were performed as described in [33, 34]. Whole slide images of histological and immunohistological stainings were obtained with the AxioScan.Z1 slide scanner (Zeiss, Feldbach, Switzerland) in bright-field mode using a Plan-Apochromat 20x/0.8 M27 objective. Stainings were automatically quantified on whole slide images using the open-source Orbit Image Analysis software (License: GPLv3; Actelion Pharmaceuticals Ltd) as described in [35, 36]. Furthermore, for histopathological scoring of pulmonary fibrosis, the Ashcroft score [37] was applied on PSR stained lung sections by two experienced blinded examiners (J.S., M.B.) as previously described [38].

Gene Expression Data

Total RNA was isolated from perfused cranial lobes of the right mouse lung with the RNeasy Tissue Mini Kit from Qiagen (Hombrechtikon, Switzerland), reverse-transcribed into complementary DNA, and messenger RNA (mRNA) expressions of inflammatory (*Il6*, *Mcp1*) and fibrotic (*Col1a1*, *Col3a1*, *Fn1*) genes were analyzed by SYBR Green quantitative real-time PCR as described in [33]. mRNA expression was expressed as ΔCt values (Ct (gene-of-interest) - Ct (reference gene)) with 60S acidic ribosomal protein P0 (*Rplp0*) as a reference gene, with a lower ΔCt indicating higher target gene expression. A list of primers used in this study is provided in Supplementary Table S11.

Micro-CT imaging, Radiomics Analysis and Score Calculation in Mice

CT images were acquired in free-breathing mice with prospective respiratory gating on a state-of-the-art micro-CT scanner (Skyscan 1176; Bruker-microCT, Kontich, Belgium) under isoflurane anaesthesia at the following time points: day 0, 7, 14, 21, 28, and 35. The following scan parameters were used: tube voltage 50 kV, tube current 500 µA, filter Al 0.5 mm, averaging (frames) 3, rotation step 0.7 degrees, sync with event 50 ms, X-ray tube rotation 360 degrees, resolution 35 µm, and slice thickness 35 µm. Images were reconstructed with NRecon reconstruction software (v.1.7.4.6; Bruker) using the built-in filtered-back projection Feldkamp algorithm and applying misalignment compensation, ring artefact reduction, and a beam hardening correction of 10% to the images.

Analogous to the radiomics analysis in patients, mouse lungs were segmented, resized to isotropic voxels (150 µm) and discretized to a fixed bin size of 50 HU, and all 1,386 radiomic features were extracted (HU limits: -1000 HU to 200 HU).

The Hounsfield units depend on the tube voltage, and the Hounsfield scale is normalized for 120 keV for patient diagnostics. Our microCT scanner allows a maximum tube voltage of 80keV. Thus, the Hounsfield units can be transferred to a limited extent. We addressed this by post-processing the microCT scans to adjust the pixel values to match the human patient data. This has been done by plotting the intensity histograms of several mice and patients from the Zurich cohort with the subsequent estimation of optimal parameters for linear transformation based on visual assessment. Specifically, the intercept value has been changed from -1000 to -1024, whereas the slope was changed from 1.0 to 0.6. These parameters were applied to all microCT scans. The choice of 2.75 mm voxel size in patients was dictated by the voxel size in mice and the difference in lung size between mice and humans. Since the voxel size in mice was 0.15 mm and the total lung capacity in humans was estimated to be 6000x greater than in mice [39], a comparable voxel size in patients was set to 2.75 mm.

For calculating the quantitative radiomic ILD risk score, the respective radiomic features were z-transformed and summed up as for patients.

Correlation Analysis and Pathway Enrichment Analysis

Spearman's rank correlation coefficient *rho* was calculated between the quantitative radiomic ILD risk score and the different biological features for correlation analysis with established inflammatory and fibrotic markers on the tissue level.

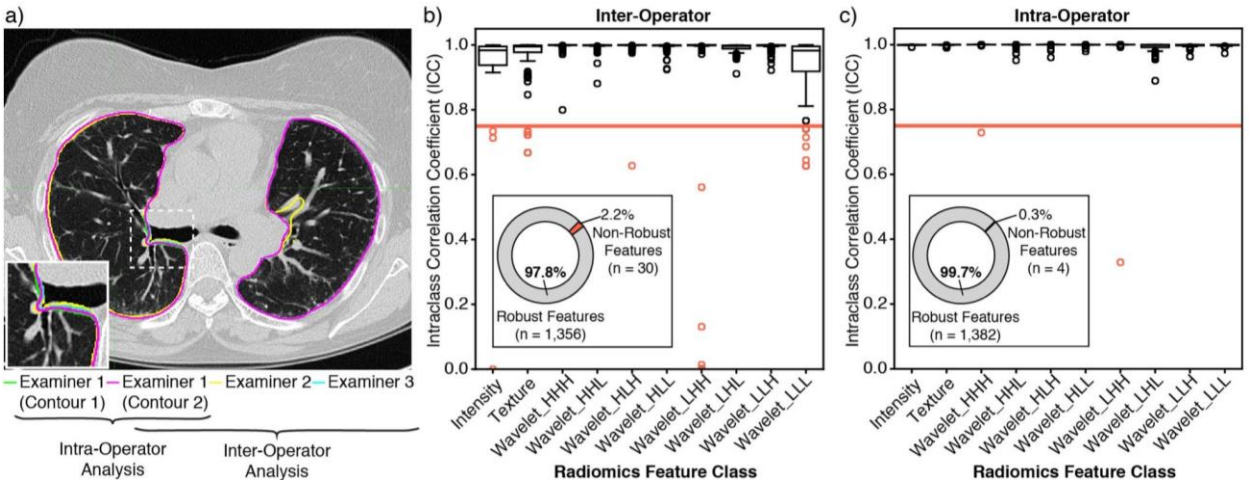
For pathway enrichment analyses, *rho* was calculated between qRISSc and the LFQ intensity value of every protein identified in at least 50% of mice in the proteomics analyses, and only proteins with $p < 0.05$ and $rho \geq |0.3|$ entered further analyses. The

resulting list of proteins, and their coding genes, were used as input for the pathway analysis using the ‘ClusterProfiler’ package of Bioconductor. Protein names were converted to gene IDs using the UniProt mapping tool (https://www.uniprot.org/uploadlists/). We investigated pathway enrichment searching against “Reactome” and “GO Biological Process” databases and retained results after adjustment ($p < 0.05$).

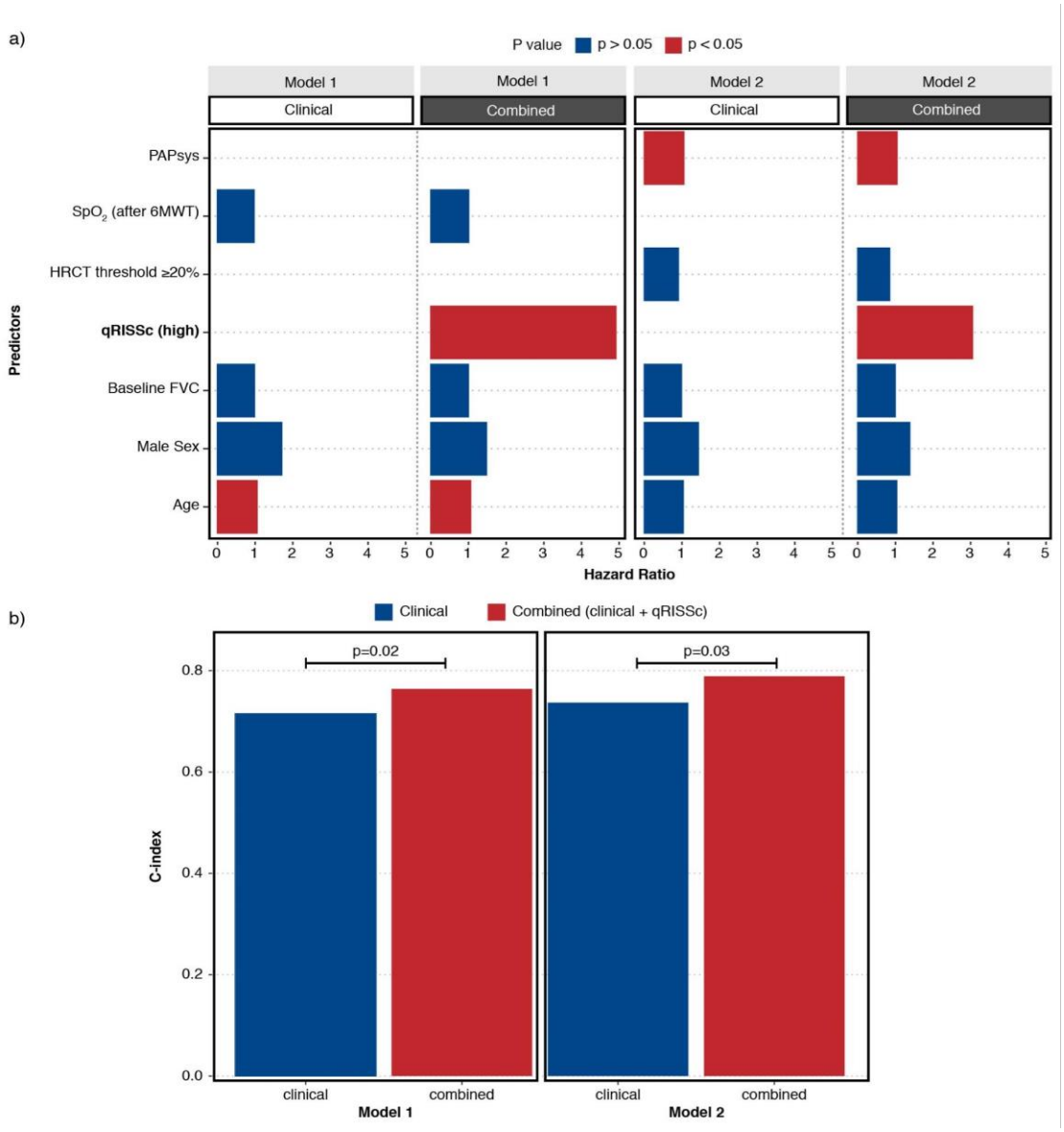
Statistical Analyses

All statistical analyses were conducted in R using the following packages: "ggplot2", "tidyverse", "ggsci", "corrplot", "readxl", "clusterSim", "dplyr", "readxl", "survival", "glmnet", "cutoff", "survminer", "cluster", "fpc", "factoextra", "clustvarsel", "clustertend". For all analyses, a p-value of < 0.05 was considered statistically significant.

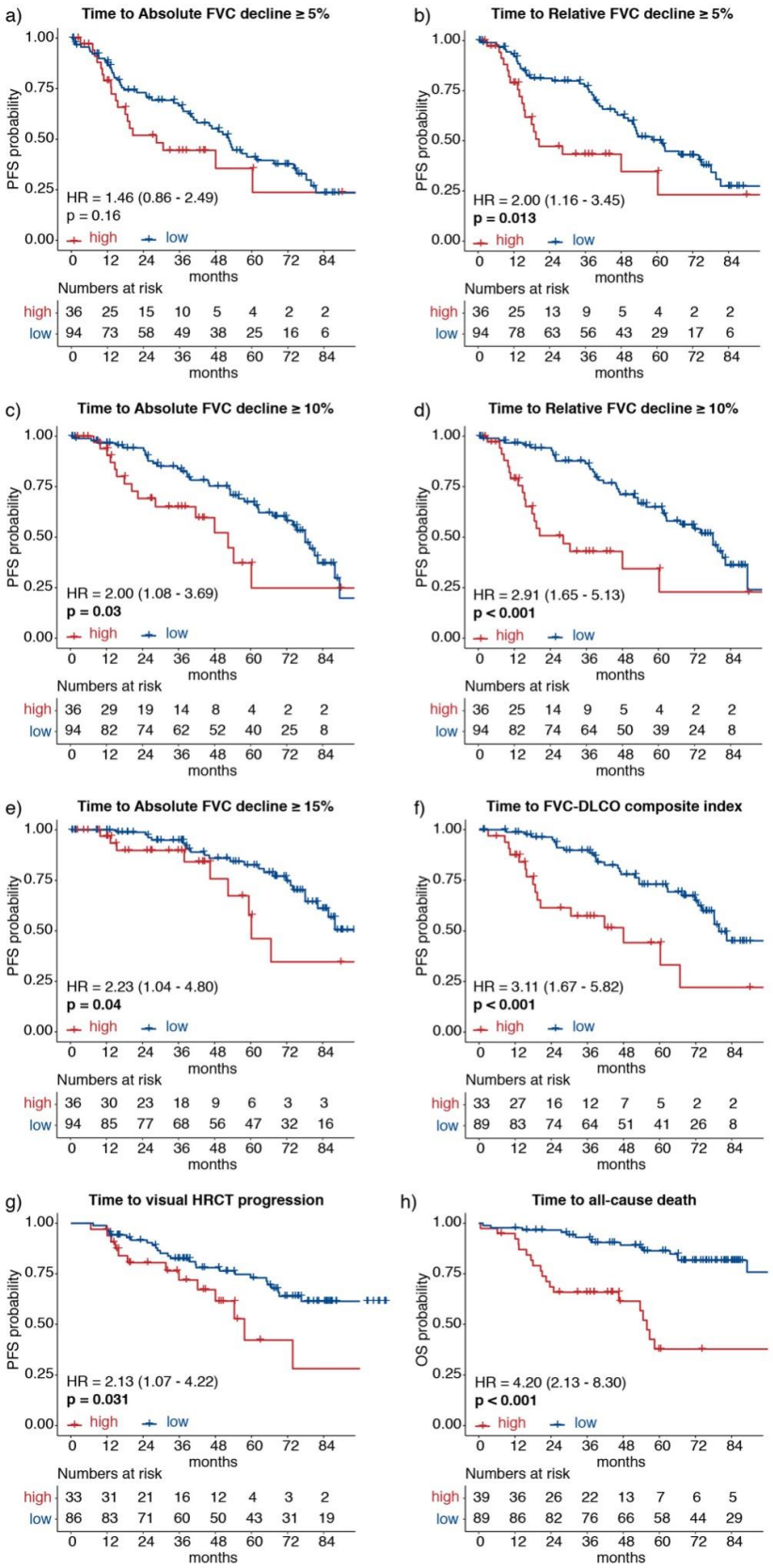
Supplementary Figures



Supplementary Figure S1: Assessment of radiomic feature robustness against inter-and intra-operator variability in the semi-automated lung segmentation process. (a) Representative transversal HRCT image showing excellent agreement and overlap in the semi-automatically delineated lung structures of the three different examiners (examiner 1: green and magenta, examiner 2: yellow, examiner 3: cyan) for the intra- and inter-operator ICC analyses. This confirmed the reproducibility and validity of our lung segmentation protocol. (**b**) Boxplots showing the distribution of the ICC coefficient per radiomic feature category for inter-operator ICC analysis and (**c**) intra-operator ICC analysis. In (b, c), the bright red line indicates the threshold defined for the ICC analyses (ICC = 0.75; corresponding to good reproducibility [20]). The pie charts summarize the respective percentage and total numbers of robust (ICC ≥ 0.75) and non-robust (ICC < 0.75) radiomic features.

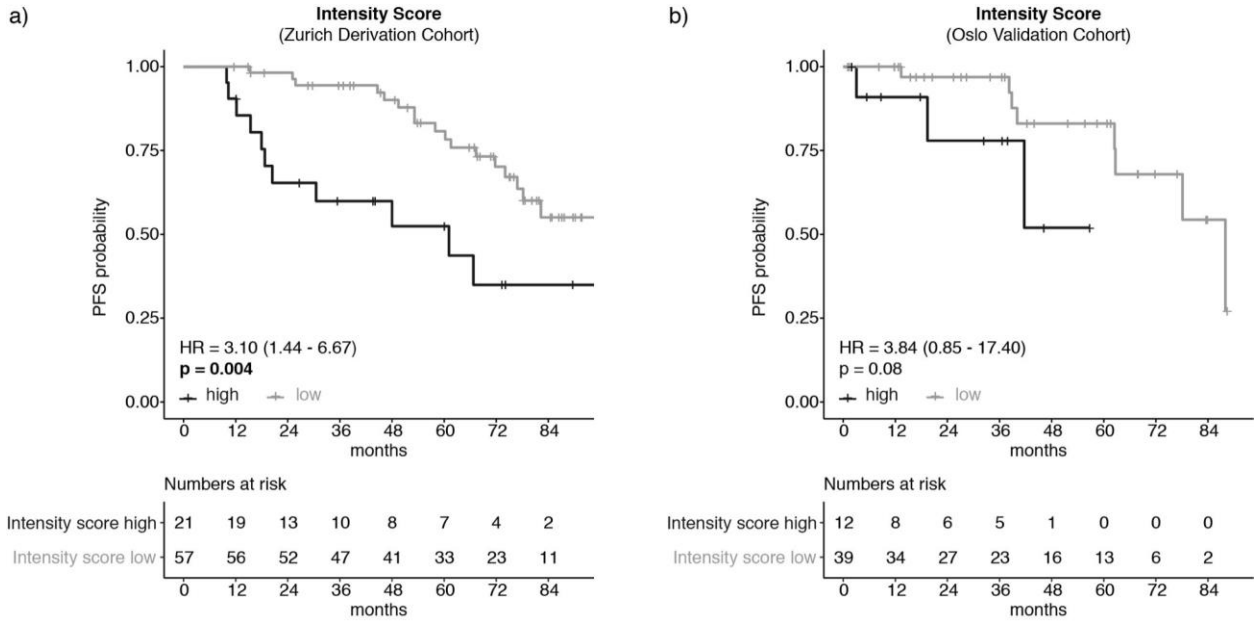


Supplementary Figure S2. Prognostic performance of qRISSc compared to other SSc-ILD risk factors. (**a**) Bar plot indicating the results of the multivariable Cox regression analysis incorporating qRISSc (combined models) versus multivariable models composed of clinical risk factors alone (clinical model). Bars represent hazard radios for each predictor in each model, whereas colours indicate the nominal p-value of the predictors. Covariates for Cox regression were selected based on literature evidence [29] and expert opinion. Due to missing data for the systolic pulmonary artery pressure *(*PAPsys, in mmHg) and the oxygen saturation at the end of the 6-min walk test (SpO_2 after 6MWT, in percent) in the validation cohort from Oslo, we only fitted the multivariable models on the derivation cohort from Zurich. (**b**) Bar plot comparing the predictive power (C-index) of multivariable models composed of clinical risk factors of SSc-ILD progression alone (clinical models) versus models also incorporating qRISSc (combined models). Two-way ANOVA was used to compare model performances.

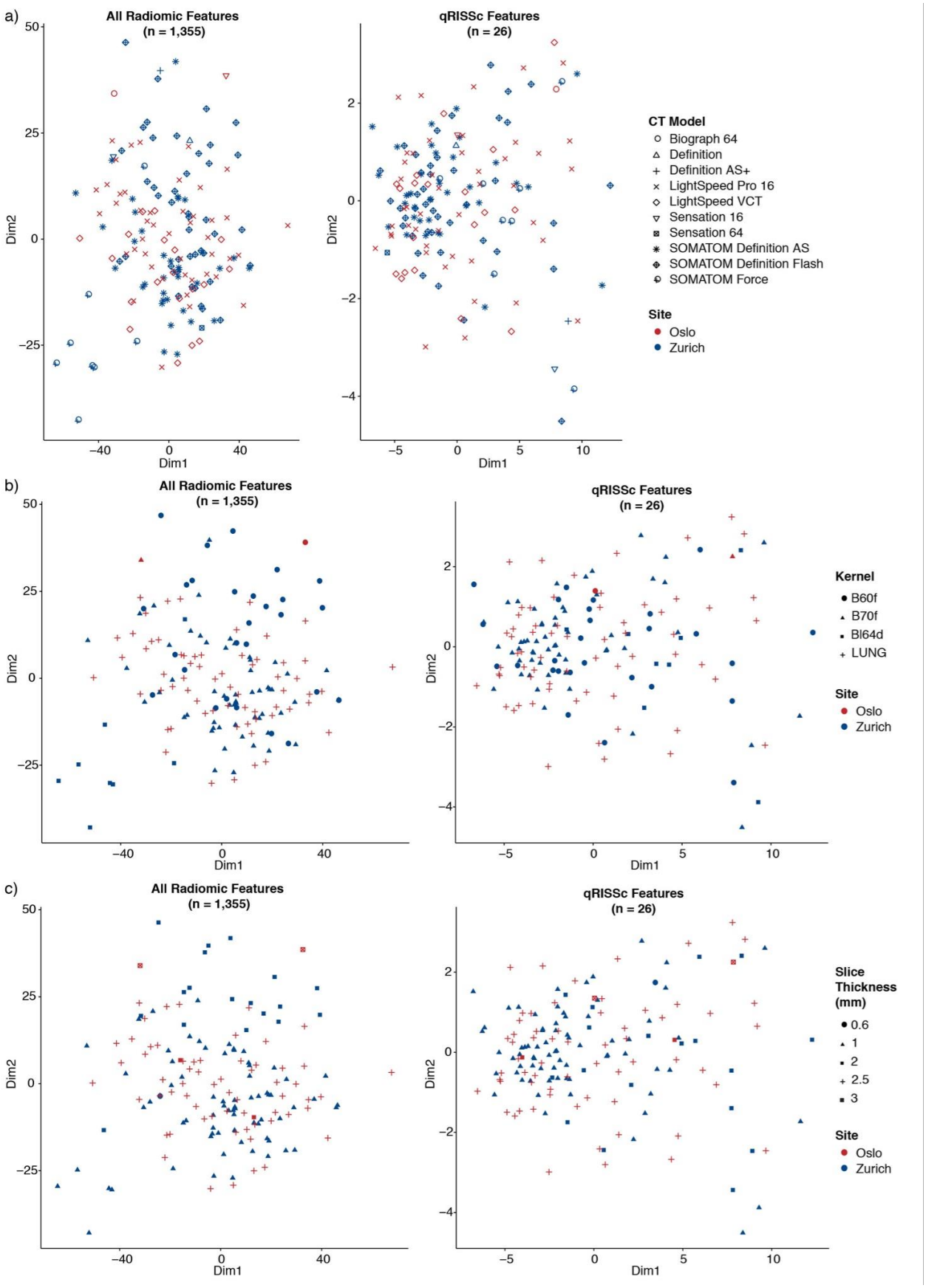


Supplementary Figure S3: Associations of qRISSc-stratified patient groups with different clinical

outcomes. Kaplan Meier curves for progression-free survival (PFS) defined as the time to (**a**) an absolute decline of FVC predicted ≥ 5%, (**b**) a relative decline of FVC predicted ≥ 5%, (**c**) an absolute decline of FVC predicted ≥ 10%, (**d**) a relative decline of FVC predicted ≥ 10%, (**e**) an absolute decline of FVC predicted ≥ 15%, (**f**) the time to the FVC-DLCO composite index (= relative decrease in FVC% predicted of ≥15% or a relative decline in FVC% predicted of ≥10% combined with DLCO% predicted of ≥15% according to [9]), or (**g**) the time to the visual ILD progression on HRCT. (**h**) Kaplan Meier curves for overall survival (OS) defined as the time to all-cause death. The Hazard ratios (HR) with 95% confidence intervals and p-value of the univariate Cox regression for the combined study cohorts are shown.



Supplementary Figure S4: Assessment of the prognostic potential of a quantitative radiomics score that is only composed of less complex, first-order intensity features. Kaplan Meier curves of the constructed intensity score for progression-free survival (PFS) defined as the time to relative FVC decline ≥ 15% in (**a**) the derivation cohort from Zurich and (**b**) in the external validation cohort from Oslo. The Hazard ratios (HR) with 95% confidence intervals and p-values of the univariate Cox regression are shown. The intensity score was statistically constructed analogously to qRISSc yet only taking first-order intensity features instead of all radiomic features into consideration.


a)
All Radiomic Features
(n = 1,355)
qRISSc Features
(n = 26)
Dim1
Dim2
CT Model
Biograph 64
Definition
Definition AS+
LightSpeed Pro 16
LightSpeed VCT
Sensation 16
Sensation 64
SOMATOM Definition AS
SOMATOM Definition Flash
SOMATOM Force
Site
Oslo
Zurich
b)
All Radiomic Features
(n = 1,355)
qRISSc Features
(n = 26)
Kernel
B60f
B70f
Bl64d
LUNG
Site
Oslo
Zurich
c)
All Radiomic Features
(n = 1,355)
qRISSc Features
(n = 26)
Slice Thickness (mm)
0.6
1
2
2.5
3
Site
Oslo
Zurich

Supplementary Figure S5: Impact of different CT image acquisition and reconstruction settings on radiomic feature values and qRISSc.
Multidimensional scaling of *z*-transformed radiomic profiles of all robust radiomic features (left panel) or only qRISSc features (right panel) combined for all SSc-ILD patients from the Zurich (*n* = 90) and Oslo cohort (*n* = 66) for (**a**) the different CT scanner types, (**b**) different lung reconstruction kernels, and (**c**) different slice thicknesses.

Supplementary Tables

Supplementary Table S1: Associations of the identified patients' clusters based on their radiomic profile with clinical parameters for the Zurich cohort. Continuous variables are described as median ± interquartile range, and categorical variables are presented as absolute values with relative frequencies (percent). P-values of univariate comparisons of baseline characteristics between the two clusters are shown. Fisher's exact test was used to compare categorical, and Mann-Whitney U to compare continuous variables, respectively. *Abbreviations: UIP = usual interstitial pneumonia, NSIP = nonspecific interstitial pneumonia, DIP = diffuse interstitial pneumonia, PAPsys = systolic pulmonary artery pressure, FVC = forced vital capacity, FEV1 = forced expiratory volume in 1 second, DLCO = diffusing capacity for carbon monoxide, 6-MWT = 6-min walk test, CRP = C-reactive protein*

Characteristics	**Cluster 1** (n=59)	**Cluster 2** (n=31)	**P-value**
Age (years)	58.0 ± 17.0	57.0 ± 16.9	0.693
Sex			
Male	14 (23.7%)	7 (22.6%)	1.000
Female	45 (76.3%)	24 (77.4%)	
SSc disease duration (years)*	5.0 ± 8.0	4.3 ± 8.3	0.507
SSc subset (LeRoy 1988)			
Limited cutaneous SSc	30 (50.8%)	11 (35.5%)	0.234
Diffuse cutaneous SSc	26 (44.1%)	16 (51.6%)	
No skin involvement	3 (5.1%)	4 (12.9%)	
Skin involvement			
Limited cutaneous	20 (33.9%)	11 (35.5%)	0.224
Diffuse cutaneous	27 (45.8%)	16 (51.6%)	
No skin involvement	5 (8.5%)	4 (12.9%)	
Only sclerodactyly	7 (11.9%)	0 (0.0%)	
Autoantibodies			
Anti-centromere positive	10 (16.9%)	3 (9.7%)	0.530
Anti-topoisomerase I positive	28 (47.5%)	13 (41.9%)	0.661
Anti-RNA polymerase III positive	4 (6.8%)	3 (9.7%)	0.602
Anti-PMScl positive	14 (23.7%)	4 (12.9%)	0.496
FVC (% predicted)	97.0 ± 26.0	65.5 ± 22.2	
FVC ≥70% predicted	54 (91.5%)	10 (32.3%)	**<0.001**
FVC <70% predicted	4 (6.8%)	20 (64.5%)	
DLCO (% predicted)	75.0 ± 24.0	48.0 ± 25.5	**<0.001**
FEV1 (% predicted)	95.8 ± 19.0	65.5 ± 25.5	**<0.001**
Pulmonary hypertension†	7 (11.9%)	13 (41.9%)	**0.001**
PAPsys (mmHg)‡	25.0 ± 7.0	32.0 ± 18.0	**<0.001**
CRP (mg/l)	2.4 ± 5.6	4.2 ± 6.1	0.071
6 min walk distance (m)	543.5 ± 109.2	407.0 ± 173.0	**<0.001**
SpO_2 before 6-MWT (%)	97.0 ± 1.2	96.0 ± 3.0	**0.011**
SpO_2 after 6-MWT (%)	96.0 ± 3.0	88.5 ± 9.8	**<0.001**
Borg scale (unit; range 0-10)	2.0 ± 2.0	4.0 ± 3.0	**<0.001**

Extent of lung fibrosis on CT			
<20%	40 (67.8%)	10 (32.3%)	**0.002**
≥20%	19 (32.2%)	21 (67.7%)	
Ground glass opacification	30 (50.8%)	15 (48.4%)	1.000
Reticular changes	58 (98.3%)	29 (93.5%)	0.272
Traction bronchiectasis	30 (50.8%)	20 (64.5%)	0.267
Honeycombing	9 (15.3%)	13 (41.9%)	**0.009**
Bullae	1 (1.7%)	2 (6.5%)	0.272
Radiological subtype			
NSIP	33 (55.9%)	16 (51.6%)	0.662
UIP[#]	24 (40.7%)	13 (41.9%)	
DIP	0 (0.0%)	1 (3.2%)	
Unclassifiable	2 (3.4%)	1 (3.2%)	
Immunomodulatory therapy[§]	29 (49.2%)	22 (71.0%)	0.073
Smoking status			
Never	35 (59.3%)	20 (64.5%)	0.868
Former	14 (23.7%)	7 (22.6%)	
Current	9 (15.3%)	3 (9.7%)	

*Disease duration of SSc was calculated as the difference between the date of baseline CT and the date of manifestation of the first non-Raynaud's symptom.

[†]Pulmonary hypertension was assessed by echocardiography or right heart catheterisation.

‡PAPsys was determined by right heart catheterisation.

[#]UIP includes the radiological diagnosis of both, "definite" and "probable" UIP.

[§]Immunomodulatory therapy included prednisone, methotrexate, rituximab, cyclophosphamide, mycophenolate mofetil, hydroxychloroquine, tocilizumab, imatinib, azathioprine, adalimumab, leflunomid, cyclosporine.

Supplementary Table S2: Radiomic features used to construct the quantitative radiomic risk score for SSc-ILD (qRISSc). A complete list of all radiomics features, including standardized feature names is provided in Supplementary File 1. *Abbreviations: GLCM = Gray Level Co-occurrence Matrix, NGTDM = Neighborhood Gray Tone Difference Matrix, GLRLM = Gray Level Run Length Matrix, GLDZM = Gray Level Distance Matrix and NGLDM = Neighboring Gray Level Dependence Matrix*

Feature ID	Feature Name	Feature Class	Feature Subclass	Wavelet Filter	LASSO Coeff.
V3	COV	Intensity	Intensity	Unfiltered	15.14
V10	iqr	Intensity	Intensity	Unfiltered	14.41
V12	mad	Intensity	Intensity	Unfiltered	17.86
V13	rmad	Intensity	Intensity	Unfiltered	-42.79
V26	variance	Texture	GLCM	Unfiltered	-11.36
V29	sum_variance	Texture	GLCM	Unfiltered	0.01
V40	autocorrelation	Texture	GLCM	Unfiltered	-5.41
V41	clust_tendency	Texture	GLCM	Unfiltered	11.66
V66	M_autocorrelation	Texture	mGLCM	Unfiltered	-0.29
V84	len_sshge	Texture	GLRLM	Unfiltered	-9.68
V86	len_lshge	Texture	GLRLM	Unfiltered	-19.25
V102	M_len_lshge	Texture	mGLRLM	Unfiltered	-0.0024
V146	NGLDM_hgse	Texture	NGLDM	Unfiltered	59.55
V588	GLDZM_sizeVar_n.3	Wavelet	GLDZM	HLH	-0.3
V641	idiff_n.4	Wavelet	GLCM	HLL	1.84
V665	M_homogenity_n.4	Wavelet	mGLCM	HLL	-2.79
V686	coarseness.4	Wavelet	NGTDM	HLL	0.96
V687	neighContrast.4	Wavelet	NGTDM	HLL	0.03
V840	coarseness.5	Wavelet	NGTDM	LHH	-0.83
V994	coarseness.6	Wavelet	NGTDM	LHL	1.03
V998	strength6	Wavelet	NGTDM	LHL	1.26
V1082	skewness.7	Wavelet	Intensity	LLH	1.55
V1235	COV.8	Wavelet	Intensity	LLL	6.6
V1236	skewness.8	Wavelet	Intensity	LLL	8.6
V1242	iqr.8	Wavelet	Intensity	LLL	16.99
V1302	coarseness.8	Wavelet	NGTDM	LLL	0.38

Supplementary Table S3: Associations of the patients' risk groups based on qRISSc with clinical parameters for the derivation (Zurich) dataset. Continuous variables are described as median ± interquartile range, and categorical variables are presented as absolute values with relative frequencies (percent). P-values of univariate comparisons of baseline characteristics between the two risk groups are shown. Fisher's exact test was used to compare categorical, and Mann-Whitney U to compare continuous variables, respectively. *Abbreviations: UIP = usual interstitial pneumonia, NSIP = nonspecific interstitial pneumonia, DIP = diffuse interstitial pneumonia, PAPsys = systolic pulmonary artery pressure, FVC = forced vital capacity, FEV1 = forced expiratory volume in 1 second, DLCO = diffusing capacity for carbon monoxide, 6-MWT = 6-min walk test, CRP = C-reactive protein*

Characteristics	**Low risk** (n=54)	**High risk** (n=21)	**P-value**
Age (years)	56.5 ± 16.8	56.0 ± 18.0	0.939
Sex			
Male	14 (25.9%)	5 (23.8%)	1.000
Female	40 (74.1%)	16 (76.2%)	
SSc disease duration (years)*	4.3 ± 6.6	5.0 ± 9.3	0.915
SSc subset (LeRoy 1988)			
Limited cutaneous SSc	27 (50.0%)	9 (42.9%)	
Diffuse cutaneous SSc	23 (42.6%)	10 (47.6%)	0.797
No skin involvement	4 (7.4%)	2 (9.5%)	
Skin involvement			
Limited cutaneous	18 (33.3%)	9 (42.9%)	
Diffuse cutaneous	24 (44.4%)	10 (47.6%)	0.481
No skin involvement	6 (11.1%)	2 (9.5%)	
Only sclerodactyly	6 (11.1%)	0 (0.0%)	
Autoantibodies			
Anti-Centromere positive	12 (22.2%)	0 (0.0%)	**0.016**
Anti-Topoisomerase I positive	28 (51.9%)	9 (42.9%)	0.609
Anti-RNA polymerase III positive	3 (5.6%)	3 (14.3%)	0.343
Anti-PMScl positive	12 (22.2%)	4 (19.0%)	1.000
FVC (% predicted)	97.4 ± 28.5	65.0 ± 18.0	
FVC ≥70% predicted	48 (88.9%)	6 (28.6%)	**<0.001**
FVC <70% predicted	6 (11.1%)	15 (71.4%)	
DLCO (% predicted)	74.4 ± 24.2	51.0 ± 25.0	**<0.001**
FEV1 (% predicted)	95.8 ± 21.0	65.0 ± 27.0	**<0.001**
Pulmonary hypertension†	3 (5.6%)	10 (47.6%)	**<0.001**
PAPsys (mmHg)‡	24.0 ± 7.0	31.0 ± 12.5	**<0.001**
CRP (mg/l)	2.4 ± 5.4	6.1 ± 6.8	**0.006**
6 min walk distance (m)	543.0 ± 118.0	421.0 ± 126.5	**<0.001**
SpO_2 before 6MWT (%)	97.0 ± 1.0	96.0 ± 3.2	0.087
SpO_2 after 6MWT (%)	96.0 ± 3.0	85.5 ± 5.2	**<0.001**
Borg (unit; range 0-10)	2.0 ± 2.0	5.5 ± 3.2	**<0.001**
Extent of lung fibrosis on CT			
<20%	40 (74.1%)	5 (23.8%)	**<0.001**

≥20%	14 (25.9%)	16 (76.2%)	
Ground glass opacification	25 (46.3%)	11 (52.4%)	0.797
Reticular changes	52 (96.3%)	20 (95.2%)	1.000
Traction bronchiectasis	21 (38.9%)	17 (81.0%)	**0.002**
Honeycombing	5 (9.3%)	11 (52.4%)	**<0.001**
Bullae	2 (3.7%)	1 (4.8%)	1.000
Radiological subtype			
NSIP	29 (53.7%)	12 (57.1%)	**0.461**
UIP[#]	23 (42.6%)	8 (38.1%)	
DIP	0 (0.0%)	1 (4.8%)	
Unclassifiable	2 (3.7%)	0 (0.0%)	
Immunomodulatory therapy[§]	30 (55.6%)	14 (66.7%)	0.441
Smoking status			
Never	35 (64.8%)	13 (61.9%)	
Former	12 (22.2%)	5 (23.8%)	1.000
Current	7 (13.0%)	2 (9.5%)	

*Disease duration of SSc was calculated as the difference between the date of baseline CT and the date of manifestation of the first non-Raynaud's symptom.

[†]Pulmonary hypertension was assessed by echocardiography or right heart catheterisation.

[‡]PAP sys was determined by right heart catheterisation.

[#]UIP includes the radiological diagnosis of both, "definite" and "probable" UIP.

[§]Immunomodulatory therapy included prednisone, methotrexate, rituximab, cyclophosphamide, mycophenolate mofetil, hydroxychloroquine, tocilizumab, imatinib, azathioprine, adalimumab, leflunomid, cyclosporine.

Supplementary Table S4: Associations of the patients' risk groups based on qRISSc with clinical parameters for the external and independent validation (Oslo) cohort. Continuous variables are described as median ± interquartile range, and categorical variables are presented as absolute values with relative frequencies (percent). P-values of univariate comparisons of baseline characteristics between the two risk groups are shown. Fisher's exact test was used to compare categorical, and Mann-Whitney U to compare continuous variables, respectively. *Abbreviations: UIP = usual interstitial pneumonia, NSIP = nonspecific interstitial pneumonia, DIP = diffuse interstitial pneumonia, PAPsys = systolic pulmonary artery pressure, FVC = forced vital capacity, FEV1 = forced expiratory volume in 1 second, DLCO = diffusing capacity for carbon monoxide, 6-MWT = 6-min walk test, CRP = C-reactive protein*

Characteristics	**Low risk** (n=47)	**High risk** (n=19)	***p* value**
Age (years)	58.0 ± 22.0	64.0 ± 19.0	0.311
Sex			
Male	12 (25.5%)	4 (21.1%)	1.000
Female	35 (74.5%)	15 (78.9%)	
SSc disease duration (years)*	4.3 ± 9.1	6.1 ± 9.1	0.325
SSc subset (LeRoy 1988)			
Limited cutaneous SSc	26 (55.3%)	11 (57.9%)	
Diffuse cutaneous SSc	21 (44.7%)	8 (42.1%)	1.000
No skin involvement	0 (0.0%)	0 (0.0%)	
Skin involvement			
Limited cutaneous	26 (55.3%)	11 (57.9%)	
Diffuse cutaneous	21 (44.7%)	8 (42.1%)	1.000
No skin involvement	0 (0.0%)	0 (0.0%)	
Only sclerodactyly	0 (0.0%)	0 (0.0%)	
Autoantibodies			
Anti-Centromere positive	5 (10.6%)	2 (10.5%)	1.000
Anti-Topoisomerase I positive	17 (36.2%)	7 (36.8%)	1.000
Anti-RNA polymerase III positive	8 (17.0%)	0 (0.0%)	**0.046**
Anti-PMScl positive	3 (6.4%)	1 (5.3%)	1.000
FVC (% predicted)	92.0 ± 25.5	60.0 ± 20.0	
FVC ≥70% predicted	36 (76.6%)	8 (42.1%)	**<0.001**
FVC <70% predicted	6 (12.8%)	9 (47.4%)	
DLCO (% predicted)	66.0 ± 17.5	35.0 ± 20.0	**<0.001**
FEV1 (% predicted)	82.0 ± 22.0	64.0 ± 18.0	**<0.001**
Pulmonary hypertension†	1 (2.1%)	5 (26.3%)	**0.008**
PAPsys (mmHg)‡	15.0 ± 10.0	35.0 ± 18.8	0.054
CRP (mg/l)	2.9 ± 5.7	5.4 ± 9.0	0.121
6 min walk distance (m)	n/a	n/a	n/a
SpO_2 before 6MWT (%)	n/a	n/a	n/a
SpO_2 after 6MWT (%)	n/a	n/a	n/a
Borg (unit; range 0-10))	n/a	n/a	n/a
Extent of lung fibrosis on CT			
<20%	30 (63.8%)	0 (0.0%)	**<0.001**

≥20%	17 (36.2%)	19 (100.0%)	
Ground glass opacification	33 (70.2%)	9 (47.4%)	0.097
Reticular changes	34 (72.3%)	17 (89.5%)	0.198
Traction bronchiectasis	12 (25.5%)	15 (78.9%)	**<0.001**
Honeycombing	6 (12.8%)	10 (52.6%)	**0.001**
Bullae	2 (4.3%)	2 (10.5%)	0.573
Radiological subtype			
NSIP	27 (57.4%)	7 (36.8%)	0.175
UIP[#]	16 (34.0%)	11 (57.9%)	
DIP	0 (0.0%)	0 (0.0%)	
Unclassifiable	4 (8.5%)	1 (5.3%)	
Immunomodulatory therapy[§]	16 (34.0%)	12 (63.2%)	0.053
Smoking status			
Never	16 (34.0%)	8 (42.1%)	0.578
Former	17 (36.2%)	8 (42.1%)	
Current	5 (10.6%)	0 (0.0%)	

*Disease duration of SSc was calculated as the difference between the date of baseline CT and the date of manifestation of the first non-Raynaud's symptom.

[†]Pulmonary hypertension was assessed by echocardiography or right heart catheterisation.

[‡]PAP sys was determined by right heart catheterisation.

[#]UIP includes the radiological diagnosis of both, "definite" and "probable" UIP.

[§]Immunomodulatory therapy included prednisone, methotrexate, rituximab, cyclophosphamide, mycophenolate mofetil, hydroxychloroquine, tocilizumab, imatinib, azathioprine, adalimumab, leflunomid, cyclosporine.

Supplementary Table S5: Summary of the univariable Cox regression analysis for qRISSc and the previously proposed clinical risk factors for SSc-ILD progression. Covariates for univariable Cox regression were selected based on literature evidence [29] and expert opinion.

Predictor	HR (95% CI)	P-value	C-Index (SE), p-value
Zurich			
Age	1.03 (0.99, 1.06)	0.17	0.59 (0.06), p=0.17
Male Sex	1.38 (0.60, 3.16)	0.45	0.53 (0.05), p=0.46
Anti-Topoisomerase 1 positive	1.04 (0.49, 2.22)	0.92	0.46 (0.05), p=0.92
Baseline FVC (% predicted)	0.98 (0.96, 1.00)	**0.04**	0.62 (0.07), **p=0.04**
Baseline DLCO (% predicted)	0.98 (0.96, 1.00)	**0.02**	0.69 (0.05), **p=0.02**
HRCT Threshold (≥20%)	1.94 (0.90, 4.19)	0.09	0.61 (0.05), p=0.10
UIP Subtype*	0.88 (0.40, 1.92)	0.74	0.49 (0.05), p=0.74
Diffuse cutaneous skin involvement	1.81 (0.84, 3.89)	0.13	0.56 (0.05), **p=0.13**
CRP	1.01 (0.99, 1.04)	0.31	0.63 (0.06), p=0.35
qRISSc (high)	4.10 (1.87, 9.03)	**<0.001**	0.67 (0.05), **p=0.001**
Oslo			
Age	1.02 (0.97, 1.07)	0.45	0.61 (0.11), p=0.44
Male Sex	0.22 (0.03, 1.71)	0.15	0.61 (0.04), p=0.08
Anti-Topoisomerase 1 positive	1.07 (0.25, 4.55)	0.93	0.48 (0.10), p=0.93
Baseline FVC (% predicted)	1.00 (0.97, 1.03)	0.87	0.60 (0.12), p=0.87
Baseline DLCO (% predicted)	0.95 (0.92, 0.99)	**0.01**	0.85 (0.06), **p=0.008**
HRCT Threshold (≥20%)	2.04 (0.52, 8.00)	0.31	0.65 (0.06), p=0.29
UIP Subtype*	1.38 (0.42, 4.55)	0.60	0.55 (0.09), p=0.60
Diffuse cutaneous skin involvement	0.69 (0.19, 2.59)	0.58	0.48 (0.09), p=0.59
CRP	1.00 (0.94, 1.07)	0.98	0.52 (0.10), p=0.98
qRISSc (high)	5.14 (1.14, 23.2)	**0.03**	0.71 (0.07), **p=0.04**
Combined Cohorts			
Age	1.02 (1.00, 1.05)	0.11	0.59 (0.05), p=0.10
Male Sex	0.96 (0.46, 2.04)	0.92	0.50 (0.04), p=0.92
Anti-Topoisomerase 1 positive	1.05 (0.54, 2.02)	0.89	0.47 (0.05), p=0.89
Baseline FVC (% predicted)	0.98 (0.97, 1.00)	**0.04**	0.62 (0.06), **p=0.04**
Baseline DLCO (% predicted)	0.97 (0.96, 0.99)	**0.001**	0.72 (0.04), **p=0.001**
HRCT Threshold (≥20%)	1.98 (1.04, 3.76)	**0.04**	0.62 (0.04), **p=0.04**
UIP Subtype*	1.01 (0.53, 1.94)	0.98	0.53 (0.05), p=0.98
Diffuse cutaneous skin involvement	1.67 (0.88, 3.17)	0.12	0.54 (0.05), p=0.12
CRP	1.01 (0.99, 1.04)	0.40	0.60 (0.05), p=0.43
qRISSc (high)	4.07 (2.07, 8.00)	**<0.001**	0.68 (0.04), **p<0.001**

*UIP includes the radiological diagnosis of both, “definite” and “probable” UIP.

Supplementary Table S6: Summary of the multivariable Cox regression analysis of the clinical models composed of previously proposed risk factors for SSc-ILD progression. Covariates for multivariable Cox regression were selected based on literature evidence [29] and expert opinion.

Predictor	HR (95% CI)	P-value	FDR	C-Index (SE), p-value
Model 1				
Age	1.03 (1.00, 1.06)	0.09	0.18	0.64 (0.05) p=0.16
Male Sex	0.92 (0.42, 2.03)	0.84	0.88	
Baseline FVC (% predicted)	0.98 (0.97, 1.00)	**0.04**	0.15	
Anti-Topoisomerase 1 positive	1.07 (0.55, 2.11)	0.84	0.88	
Model 2				
Age	1.03 (1.00, 1.06)	0.06	0.15	0.66 (0.05) **p=0.04**
Male Sex	0.85 (0.39, 1.86)	0.69	0.83	
Baseline FVC (% predicted)	0.98 (0.96, 1.00)	**0.02**	0.12	
Model 3				
Age	1.03 (1.00, 1.06)	0.09	0.18	0.71 (0.04) **p=0.006**
Male Sex	0.98 (0.45, 2.11)	0.96	0.96	
Baseline DLCO (% predicted)	0.97 (0.95, 0.99)	**0.003**	0.06	
HRCT Threshold (≥20%)	1.08 (0.52, 2.25)	0.84	0.88	
Model 4				
Age	1.02 (0.99, 1.05)	0.15	0.28	0.66 (0.05) p=0.12
Male Sex	0.83 (0.38, 1.82)	0.64	0.83	
Baseline FVC (% predicted)	0.98 (0.97, 1.00)	0.06	0.15	
HRCT Threshold (≥20%)	1.40 (0.68, 2.90)	0.36	0.62	
UIP Subtype*	0.79 (0.39, 1.59)	0.50	0.77	
Model 5				
Age	1.04 (1.01, 1.07)	**0.02**	0.12	0.69 (0.04) **p=0.02**
Male Sex	0.78 (0.36, 1.73)	0.54	0.77	
Baseline FVC (% predicted)	0.98 (0.97, 1.00)	**0.03**	0.15	
Diffuse cutaneous skin involvement	2.01 (1.00, 4.04)	**0.05**	0.15	
Model 6				
Age	1.03 (1.00, 1.06)	0.06	0.15	0.68 (0.06) **p=0.03**
Male Sex	0.77 (0.35, 1.72)	0.53	0.77	
Baseline FVC (% predicted)	0.98 (0.96, 0.99)	**0.005**	0.06	
CRP (mg/l)	1.01 (0.98, 1.04)	0.67	0.83	

*UIP includes the radiological diagnosis of both, “definite” and “probable” UIP.

Supplementary Table S7: Summary of the multivariable Cox regression analysis of the combined models, i.e. incorporating qRISSc and the previously proposed clinical risk factors for SSc-ILD progression. Covariates for multivariable Cox regression were selected based on literature evidence [29] and expert opinion.

Predictor	HR (95% CI)	p-value	FDR	C-Index (SE), p-value
Model 1				
Age	1.03 (1.00, 1.06)	0.06	0.15	0.71 (0.05), **p=0.009**
Male Sex	1.03 (0.46, 2.30)	0.95	0.95	
Baseline FVC (% predicted)	0.99 (0.97, 1.01)	0.37	0.58	
Anti-Topoisomerase 1 positive	1.29 (0.65, 2.56)	0.46	0.70	
qRISSc (high)	3.48 (1.60, 7.55)	**0.002**	**0.01**	
Model 2				
Age	1.03 (1.00, 1.06)	0.06	0.15	0.74 (0.04), **p=0.001**
Male Sex	0.96 (0.43, 2.13)	0.92	0.95	
Baseline FVC (% predicted)	0.99 (0.97, 1.01)	0.30	0.56	
qRISSc (high)	3.59 (1.72, 7.50)	**0.001**	**0.01**	
Model 3				
Age	1.03 (1.00, 1.07)	**0.04**	0.15	0.77 (0.04), **p=2.72E-04**
Male Sex	0.95 (0.43, 2.10)	0.90	0.95	
Baseline DLCO (% predicted)	0.98 (0.96, 1.00)	**0.05**	0.15	
HRCT Threshold (≥20%)	0.92 (0.44, 1.93)	0.83	0.95	
qRISSc (high)	3.42 (1.58, 7.41)	**0.002**	**0.01**	
Model 4				
Age	1.03 (0.99, 1.06)	0.12	0.25	0.72 (0.05), **p=0.006**
Male Sex	0.96 (0.42, 2.19)	0.92	0.95	
Baseline FVC (% predicted)	0.99 (0.97, 1.01)	0.33	0.56	
HRCT Threshold (≥20%)	1.04 (0.48, 2.25)	0.92	0.95	
UIP Subtype*	0.96 (0.47, 1.98)	0.91	0.95	
qRISSc (high)	3.49 (1.60, 7.61)	**0.002**	**0.01**	
Model 5				
Age	1.04 (1.01, 1.07)	**0.01**	**0.05**	0.75 (0.04) **p<0.001**
Male Sex	0.91 (0.41, 2.03)	0.81	0.95	
Baseline FVC (% predicted)	0.99 (0.98, 1.01)	0.33	0.56	
Diffuse cutaneous skin involvement	2.48 (1.23, 5.01)	**0.01**	**0.05**	
qRISSc (high)	4.23 (2.03, 8.83)	**<0.001**	**0.001**	
Model 6				
Age	1.03 (1.00, 1.06)	0.07	0.16	0.72 (0.05) **p=0.003**
Male Sex	0.89 (0.39, 1.99)	0.77	0.95	
Baseline FVC (% predicted)	0.99 (0.97, 1.01)	0.18	0.35	
CRP (mg/l)	1.01 (0.98, 1.04)	0.69	0.95	
qRISSc (high)	3.07 (1.38, 6.85)	**0.006**	**0.03**	

*UIP includes the radiological diagnosis of both, “definite” and “probable” UIP.

Supplementary Table S8. Summary of the multivariable Cox regression analysis for the clinical and combined models, incorporating systolic pulmonary artery pressure or oxygen saturation at the end of the 6-min walk test as previously proposed risk factors for SSc-ILD progression, respectively. Covariates for multivariable Cox regression were selected based on literature evidence [29] and expert opinion. Due to missing data for the systolic pulmonary artery pressure (PAPsys, in mmHg) and the oxygen saturation at the end of the 6-min walk test (SpO_2 after 6MWT, in percent) in the validation cohort from Oslo, we only fitted the multivariable models on the derivation cohort from Zurich.

Predictor	HR (95% CI)	p-value	C-Index (SE), p-value
Clinical 1			
Age	1.05 (1.00, 1.10)	**0.05**	0.72 (0.05), p=0.06
Male Sex	1.70 (0.58, 5.01)	0.33	
Baseline FVC (% predicted.)	0.98 (0.96, 1.01)	0.15	
SpO_2 after 6-MWT (%)	0.97 (0.91, 1.03)	0.34	
Clinical 2			
Age	1.03 (0.99, 1.08)	0.13	0.74 (0.05), **p=0.006**
Male Sex	1.43 (0.49, 4.16)	0.51	
Baseline FVC (% predicted)	0.98 (0.96, 1.00)	0.10	
HRCT Threshold (≥20%)	0.90 (0.37, 2.22)	0.82	
PAPsys (mmHg)*	1.05 (1.02, 1.08)	**0.003**	
Combined 1			
Age	1.05 (1.00, 1.10)	**0.04**	0.76 (0.06), **p=0.01**
Male Sex	1.47 (0.52, 4.18)	0.47	
Baseline FVC (% predicted)	0.99 (0.96, 1.02)	0.67	
SpO_2 after 6-MWT (%)	1.00 (0.92, 1.10)	0.96	
qRISSc (high)	4.91 (1.19, 20.26)	**0.03**	
Combined 2			
Age	1.03 (0.99, 1.08)	0.13	0.79 (0.05), **p=0.002**
Male Sex	1.38 (0.46, 4.15)	0.56	
Baseline FVC (% predicted)	0.99 (0.97, 1.02)	0.47	
HRCT Threshold (≥20%)	0.85 (0.35, 2.06)	0.71	
PAPsys (mmHg)*	1.04 (1.01, 1.08)	**0.008**	
qRISSc (high)	3.05 (1.13, 8.20)	**0.03**	

*PAPsys was determined by right heart catheterization

Supplementary Table S9: Summary of HRCT image acquisition parameters for the two study cohorts. For slice thickness and tube voltage, data are presented as median and range of minimal and maximal values.

CT parameter	**Discovery (Zurich) cohort** (n=90)	**Validation (Oslo) cohort** (n=66)
Manufacturer*	Siemens	Siemens, GE Medical Systems
Acquisition Model	Inspiration (breath hold)	Inspiration (breath hold)
Slice thickness (mm)	1 (range 0.6 - 2)	2.5 (range 2 - 3)
Reconstruction kernels	B60f, B70f, Bl64	B60f, B70f, LUNG
Tube voltage (kVP)	120 (range 80 - 150)	120

*HRCT scanners included SOMATOM Definition AS, SOMATOM Definition Flash, SOMATOM Force, SOMATOM Sensation 64, SOMATOM Sensation 16, Biograph 64, LightSpeed Pro 16, LightSpeed VCT.

Supplementary Table S10: Parameter settings for MaxQuant analysis.

Parameter	Value
Version	1.6.6.0
Machine name	PROXMOX-W10
PSM FDR	0.01
PSM FDR Crosslink	0.01
Protein FDR	0.01
Site FDR	0.01
Use Normalized Ratios For Occupancy	TRUE
Min. peptide Length	7
Min. score for unmodified peptides	0
Min. score for modified peptides	40
Min. delta score for unmodified peptides	0
Min. delta score for modified peptides	6
Min. unique peptides	0
Min. razor peptides	2
Min. peptides	2
Use only unmodified peptides and	FALSE
Peptides used for protein quantification	Razor
Discard unmodified counterpart peptides	TRUE
Label min. ratio count	2
Use delta score	FALSE
iBAQ	TRUE
iBAQ log fit	TRUE
Match between runs	TRUE
Matching time window [min]	0.7
Match ion mobility window [indices]	0.05
Alignment time window [min]	20
Alignment ion mobility window [indices]	1
Find dependent peptides	FALSE
Fasta file	MusMusculus_SP_2019_10.fasta
Decoy mode	revert
Include contaminants	TRUE
Fixed modification	Carbamidomethylation of Cys
Variable modifications	Oxidation on Met; Acetyl on protein N-term
Advanced ratios	FALSE
Second peptides	TRUE
Stabilize large LFQ ratios	TRUE
Separate LFQ in parameter groups	FALSE
Require MS/MS for LFQ comparisons	TRUE
Calculate peak properties	FALSE
Main search max. combinations	200
Advanced site intensities	FALSE
Write msScans table	FALSE

Write msmsScans table	FALSE
Write ms3Scans table	FALSE
Write allPeptides table	FALSE
Write mzRange table	FALSE
Write pasefMsmsScans table	FALSE
Write accumulatedPasefMsmsScans table	FALSE
Max. peptide mass [Da]	5500
Min. peptide length for unspecific search	8
Max. peptide length for unspecific search	25
Razor protein FDR	TRUE
Max mods in site table	3
Match unidentified features	FALSE
Evaluate variant peptides separately	TRUE
Variation mode	None
MS/MS tol. (FTMS)	20 ppm
Top MS/MS peaks per Da interval. (FTMS)	6
Da interval. (FTMS)	20
MS/MS deisotoping (FTMS)	TRUE
MS/MS deisotoping tolerance (FTMS)	7
MS/MS deisotoping tolerance unit (FTMS)	ppm
MS/MS higher charges (FTMS)	TRUE
MS/MS water loss (FTMS)	TRUE
MS/MS ammonia loss (FTMS)	TRUE
MS/MS dependent losses (FTMS)	TRUE
MS/MS recalibration (FTMS)	FALSE
MS/MS tol. (ITMS)	0.4 Da
Top MS/MS peaks per Da interval. (ITMS)	12
Da interval. (ITMS)	100
MS/MS deisotoping (ITMS)	FALSE
MS/MS deisotoping tolerance (ITMS)	0.15
MS/MS deisotoping tolerance unit (ITMS)	Da
MS/MS higher charges (ITMS)	TRUE
MS/MS water loss (ITMS)	TRUE
MS/MS ammonia loss (ITMS)	TRUE
MS/MS dependent losses (ITMS)	TRUE
MS/MS recalibration (ITMS)	FALSE
MS/MS deisotoping (Unknown)	FALSE
MS/MS deisotoping tolerance (Unknown)	0.15
MS/MS deisotoping tolerance unit (Unknown)	Da
MS/MS higher charges (Unknown)	TRUE
MS/MS water loss (Unknown)	TRUE
MS/MS ammonia loss (Unknown)	TRUE
MS/MS dependent losses (Unknown)	TRUE
MS/MS recalibration (Unknown)	FALSE

Supplementary Table S11: Murine primer sequences used for qRT-PCR.

Gene	Forward primer (5' - 3')	Reverse primer (5' - 3')
Collagen 1 alpha 1 (*Col1a1*)	GAT GAC GTG CAA TGC AAT GAA	CCC TCG ACT CCT ACA TCT TCT GA
Collagen 3 alpha 1 (*Col3a1*)	AGC TTT GTG CAA AGT GGA ACC	ATA GGA CTG ACC AAG GTG GC
Fibronectin 1 (*Fn1*)	ATG TGG ACC CCT CCT GAT AGT	GCC CAG TGA TTT CAG CAA AGG
Interleukin 6 (*Il6*)	TGA TGG ATG CTA CCA AAC TGG	GGT ACT CCA GAA GAC CAG AG
Monocyte chemoattractant protein 1 (*Mcp-1*)	CCA CTC ACC TGC TGC TAC TCA T	TGG TGA TCC TCT TGT AGC TCT CC
60S acidic ribosomal protein P0 (*Rplp0*)	GCA GGT GTT TGA CAA CGG CAG	GAT GAT GGA GTG TGG CAC CGA

Supplementary References

1. Tyndall A, Ladner UM, Matucci-Cerinic M. The EULAR Scleroderma Trials And Research group (EUSTAR): an international framework for accelerating scleroderma research [Internet]. Current Opinion in Rheumatology 2008. p. 703–706Available from: http://dx.doi.org/10.1097/bor.0b013e328311f841.

2. Minier T, Guiducci S, Bellando-Randone S, Bruni C, Lepri G, Czirják L, Distler O, Walker UA, Fransen J, Allanore Y, Denton C, Cutolo M, Tyndall A, Müller-Ladner U, Matucci-Cerinic M, EUSTAR co-workers, EUSTAR co-workers. Preliminary analysis of the very early diagnosis of systemic sclerosis (VEDOSS) EUSTAR multicentre study: evidence for puffy fingers as a pivotal sign for suspicion of systemic sclerosis. *Ann. Rheum. Dis.* 2014; 73: 2087–2093.

3. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, Matucci-Cerinic M, Naden RP, Medsger TA Jr, Carreira PE, Riemekasten G, Clements PJ, Denton CP, Distler O, Allanore Y, Furst DE, Gabrielli A, Mayes MD, van Laar JM, Seibold JR, Czirjak L, Steen VD, Inanc M, Kowal-Bielecka O, Müller-Ladner U, Valentini G, Veale DJ, Vonk MC, Walker UA, Chung L, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Arthritis Rheum.* 2013; 65: 2737–2747.

4. Borg G. Borg's Perceived Exertion and Pain Scales. Human Kinetics 1; 1998.

5. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, Arriola E, Silver R, Strange C, Bolster M, Seibold JR, Riley DJ, Hsu VM, Varga J, Schraufnagel DE, Theodore A, Simms R, Wise R, Wigley F, White B, Steen V, Read C, Mayes M, Parsley E, Mubarak K, Connolly MK, Golden J, Olman M, Fessler B, Rothfield N, et al. Cyclophosphamide versus placebo in scleroderma lung disease. *N. Engl. J. Med.* 2006; 354: 2655–2666.

6. Tashkin DP, Roth MD, Clements PJ, Furst DE, Khanna D, Kleerup EC, Goldin J, Arriola E, Volkmann ER, Kafaja S, Silver R, Steen V, Strange C, Wise R, Wigley F, Mayes M, Riley DJ, Hussain S, Assassi S, Hsu VM, Patel B, Phillips K, Martinez F, Golden J, Connolly MK, Varga J, Dematte J, Hinchcliff ME, Fischer A, Swigris J, et al. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomised controlled, double-blind, parallel group trial. *Lancet Respir. Med.* Elsevier BV; 2016; 4: 708–719.

7. American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). *Am. J. Respir. Crit. Care Med.* 2000; 161: 646–664.

8. Wu W, Jordan S, Becker MO, Dobrota R, Maurer B, Fretheim H, Ye S, Siegert E, Allanore Y, Hoffmann-Vold A-M, Distler O. Prediction of progression of interstitial

lung disease in patients with systemic sclerosis: the SPAR model. *Ann. Rheum. Dis.* 2018; 77: 1326–1332.

9. Wu W, Jordan S, Becker MO, Dobrota R, Maurer B, Fretheim H, Ye S, Siegert E, Allanore Y, Hoffmann-Vold A-M, Distler O. Prediction of progression of interstitial lung disease in patients with systemic sclerosis: the SPAR model. *Ann. Rheum. Dis.* 2018; 77: 1326–1332.

10. Graham BL, Steenbruggen I, Miller MR, Barjaktarevic IZ, Cooper BG, Hall GL, Hallstrand TS, Kaminsky DA, McCarthy K, McCormack MC, Oropez CE, Rosenfeld M, Stanojevic S, Swanney MP, Thompson BR. Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. *Am. J. Respir. Crit. Care Med.* 2019; 200: e70–e88.

11. Graham BL, Brusasco V, Burgos F, Cooper BG, Jensen R, Kendrick A, MacIntyre NR, Thompson BR, Wanger J. 2017 ERS/ATS standards for single-breath carbon monoxide uptake in the lung. *Eur. Respir. J.* [Internet] 2017; 49Available from: http://dx.doi.org/10.1183/13993003.00016-2016.

12. Holland AE, Spruit MA, Troosters T, Puhan MA, Pepin V, Saey D, McCormack MC, Carlin BW, Sciurba FC, Pitta F, Wanger J, MacIntyre N, Kaminsky DA, Culver BH, Revill SM, Hernandes NA, Andrianopoulos V, Camillo CA, Mitchell KE, Lee AL, Hill CJ, Singh SJ. An official European Respiratory Society/American Thoracic Society technical standard: field walking tests in chronic respiratory disease. *Eur. Respir. J.* 2014; 44: 1428–1446.

13. Wanger J, Clausen JL, Coates A, Pedersen OF, Brusasco V, Burgos F, Casaburi R, Crapo R, Enright P, van der Grinten CPM, Gustafsson P, Hankinson J, Jensen R, Johnson D, Macintyre N, McKay R, Miller MR, Navajas D, Pellegrino R, Viegi G. Standardisation of the measurement of lung volumes. *Eur. Respir. J.* 2005; 26: 511–522.

14. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, Behr J, Cottin V, Danoff SK, Morell F, Flaherty KR, Wells A, Martinez FJ, Azuma A, Bice TJ, Bouros D, Brown KK, Collard HR, Duggal A, Galvin L, Inoue Y, Jenkins RG, Johkoh T, Kazerooni EA, Kitaichi M, Knight SL, Mansour G, Nicholson AG, Pipavath SNJ, Buendía-Roldán I, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am. J. Respir. Crit. Care Med.* 2018; 198: e44–e68.

15. Lynch DA, Sverzellati N, Travis WD, Brown KK, Colby TV, Galvin JR, Goldin JG, Hansell DM, Inoue Y, Johkoh T, Nicholson AG, Knight SL, Raoof S, Richeldi L, Ryerson CJ, Ryu JH, Wells AU. Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. *Lancet Respir Med* 2018; 6: 138–153.

16. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, Bellomi M. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp* 2018; 2: 36.

17. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* 2014; 5: 4006.

18. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga RJ, Boellaard R, Bogowicz M, Boldrini L, Buvat I, Cook GJR, Davatzikos C, Depeursinge A, Desseroit M-C, Dinapoli N, Dinh CV, Echegaray S, El Naqa I, Fedorov AY, Gatta R, Gillies RJ, Goh V, Götz M, Guckenberger M, Ha SM, Hatt M, Isensee F, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* 2020; : 191145.

19. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability [Internet]. Psychological Bulletin 1979. p. 420–428Available from: http://dx.doi.org/10.1037/0033-2909.86.2.420.

20. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* 2016; 15: 155–163.

21. Lawson RG, Jurs PC. New index for clustering tendency and its application to chemical problems [Internet]. Journal of Chemical Information and Modeling 1990. p. 36–41Available from: http://dx.doi.org/10.1021/ci00065a010.

22. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm [Internet]. Applied Statistics 1979. p. 100Available from: http://dx.doi.org/10.2307/2346830.

23. Lu H, Arshad M, Thornton A, Avesani G, Cunnea P, Curry E, Kanavati F, Liang J, Nixon K, Williams ST, Hassan MA, Bowtell DDL, Gabra H, Fotopoulou C, Rockall A, Aboagye EO. A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic- and molecular-phenotypes of epithelial ovarian cancer. *Nat. Commun.* 2019; 10: 764.

24. Ufuk F, Demirci M, Altinisik G. Quantitative computed tomography assessment for systemic sclerosis-related interstitial lung disease: comparison of different methods. *Eur. Radiol.* 2020; 30: 4369–4380.

25. Bocchino M, Bruzzese D, D'Alto M, Argiento P, Borgia A, Capaccio A, Romeo E, Russo B, Sanduzzi A, Valente T, Sverzellati N, Rea G, Vettori S. Performance of a new quantitative computed tomography index for interstitial lung disease assessment in systemic sclerosis. *Sci. Rep.* 2019; 9: 9468.

26. Saldana DC, Hague CJ, Murphy D, Coxson HO, Tschirren J, Peterson S, Sieren JP, Kirby M, Ryerson CJ. Association of Computed Tomography Densitometry with Disease Severity, Functional Decline, and Survival in Systemic Sclerosis-associated Interstitial Lung Disease. *Ann. Am. Thorac. Soc.* 2020; 17: 813–820.

27. Ariani A, Silva M, Seletti V, Bravi E, Saracco M, Parisi S, De Gennaro F, Idolazzi L, Caramaschi P, Benini C, Bodini FC, Scirè CA, Carrara G, Lumetti F, Alfieri V, Bonati E, Lucchini G, Aiello M, Santilli D, Mozzani F, Imberti D, Michieletti E, Arrigoni E, Delsante G, Pellerito R, Fusaro E, Chetta A, Sverzellati N. Quantitative chest computed tomography is associated with two prediction models of mortality in interstitial lung disease related to systemic sclerosis. *Rheumatology* 2017; 56: 922–927.

28. Lydersen S. Statistical review: frequently given comments. *Ann. Rheum. Dis.* BMJ; 2015; 74: 323–325.

29. Distler O, Assassi S, Cottin V, Cutolo M, Danoff SK, Denton CP, Distler JHW, Hoffmann-Vold A-M, Johnson SR, Müller Ladner U, Smith V, Volkmann ER, Maher TM. Predictors of progression in systemic sclerosis patients with interstitial lung disease. *Eur. Respir. J.* [Internet] 2020; 55Available from: http://dx.doi.org/10.1183/13993003.02026-2019.

30. Heinze G, Dunkler D. Five myths about variable selection. *Transpl. Int.* 2017; 30: 6–10.

31. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Cham: Springer International Publishing; 2015.

32. Schniering J, Gabrys H, Brunner M, Distler O, Guckenberger M, Bogowicz M, Vuong D, Karava K, Müller C, Frauenfelder T, Tanadini-Lang S, Maurer B. Computed-tomography-based radiomics features for staging of interstitial lung disease – transferability from experimental to human lung fibrosis - a proof-of-concept study [Internet]. Imaging 2019.Available from: http://dx.doi.org/10.1183/13993003.congress-2019.pa4806.

33. Schniering J, Benešová M, Brunner M, Haller S, Cohrs S, Frauenfelder T, Vrugt B, Feghali-Bostwick C, Schibli R, Distler O, Müller C, Maurer B. F-AzaFol for Detection of Folate Receptor-β Positive Macrophages in Experimental Interstitial Lung Disease-A Proof-of-Concept Study. *Front. Immunol.* 2019; 10: 2724.

34. Schniering J, Benešová M, Brunner M, Haller S, Cohrs S, Frauenfelder T, Vrugt B, Feghali-Bostwick CA, Schibli R, Distler O, Mueller C, Maurer B. Visualisation of interstitial lung disease by molecular imaging of integrin αvβ3 and somatostatin receptor 2. *Ann. Rheum. Dis.* 2019; 78: 218–227.

35. Stritt M, Stalder AK, Vezzali E. Orbit Image Analysis: An open-source whole slide image analysis tool. *PLoS Comput. Biol.* 2020; 16: e1007313.

36. Seger S, Stritt M, Vezzali E, Nayler O, Hess P, Groenen PMA, Stalder AK. A fully automated image analysis method to quantify lung fibrosis in the bleomycin-induced rat model. *PLoS One* 2018; 13: e0193057.

37. Ashcroft T, Simpson JM, Timbrell V. Simple method of estimating severity of pulmonary fibrosis on a numerical scale. *J. Clin. Pathol.* 1988; 41: 467–470.

38. Schniering J, Guo L, Brunner M, Schibli R, Ye S, Distler O, Béhé M, Maurer B. Evaluation of Tc-rhAnnexin V-128 SPECT/CT as a diagnostic tool for early stages of interstitial lung disease associated with systemic sclerosis. *Arthritis Res. Ther.* 2018; 20: 183.

39. Irvin CG, Bates JHT. Measuring the lung function in the mouse: the challenge of size. *Respir. Res.* 2003; 4: 4.