

DEVELOPING A HISTOLOGY-BASED COMPOSITE INDEX TO ASSESS THE
DEGREE OF LUNG FIBROSIS USING QUANTITATIVE HALO® MODULES

USING QUANTITATIVE HISTOLOGICAL READOUTS OF PULMONARY
FIBROSIS TO PRODUCE A SCORE OF FIBROTIC LUNG DISEASE SEVERITY

By SPENCER D. REVILL, B. Tech. (Co-op)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the
Requirements for the Degree Master of Science

McMaster University © Copyright by Spencer D. Revill, June 2022

MASTER OF SCIENCE (2022)
(Medical Sciences)

McMaster University
Hamilton, Ontario

TITLE: Using Quantitative Histological Readouts
of Pulmonary Fibrosis to Produce a
Score of Fibrotic Lung Disease Severity

AUTHOR: Spencer D. Revill, B.Tech. (Co-op)

SUPERVISOR: Dr. Kjetil Ask, Ph.D.

NUMBER OF PAGES: xiii, 81

LAY ABSTRACT

Lung diseases involving excess scar tissue formation resulting in lung failure are difficult to study, especially when the most used method for assessing the progression of the disease, the Ashcroft score, is imprecise and subject to bias. The goal with this project is to update this current gold standard for rating disease severity in a portion of lung from the purely visual Ashcroft score to one that uses robust computational methods.

This novel method uses digital pathology software, HALO®, to quantify components of the lung tissue, such as population densities of specific cell types, the proportion of alveolar airspace to alveolar interstitium, and the extent of scarring in the lung according to collagen content. Then, machine learning is used to generate a severity score that resembles the standard method, and should feel familiar to lung researchers, but that we believe is more reliable and would be consistent between research groups.

ABSTRACT

Histological and visual assessment of pulmonary fibrosis are used as a gold standard to determine severity of fibrotic lung disease. The Ashcroft scoring technique and histological stains like Masson's trichrome and picrosirius red (PSR) have been used for decades as primary readouts, often alongside a biochemical assay estimating the amount of hydroxyproline in the tissue. As pulmonary fibrosis is a heterogeneous disease and collagen content rarely shows more than a two-fold increase between a non-diseased and a diseased lung, therapeutic assessment is challenging due to the variability animal models.

To improve the assessment of the severity of fibrotic lung disease, we have developed models using multivariate linear regression and XGBoost machine learning that consider digital quantifications of histological features from 73 clinically diagnosed cases of IPF and 11 control tissues, obtained using the HALO® platform, that correlate with the Ashcroft score and combine them into a composite index that we believe represents a better, more objective way of scoring fibrotic lung tissues.

The tissues were stained with H&E for structural assessment, Trichrome (TRI), and Picrosirius Red (PSR) for collagen content, alpha-smooth muscle actin (α SMA) to assess the amount of ECM-producing myofibroblasts (MFBs), and CD68, CD163, and CD206 for macrophage content and activation as they have been correlated with lung fibrosis and are considered a therapeutic target.

ACKNOWLEDGEMENTS

I am truly thankful to my supervisor, Dr. Kjetil Ask, who took me under his wing as a co-op student during my undergraduate career and brought me where I am today. I could not imagine a more generous and thoughtful individual as my guide throughout my university career. Only you could have, and did, change my mind regarding the pursuit of science, and I am immensely thankful for that, as well as the many opportunities you've afforded me, including my new position in industry that was only possible thanks to you. I am glad that we both could move onto bigger things, and I hope we can continue working together in the future.

Next, I would like to thank current members of the Ask Lab who've been of immense help and great friends to me, including Megan Vierhout, Anmar Ayaub, and Safaa Naiel, without whom I could not have accomplished my experiments nor navigate the tumultuous path that is a graduate career. Thank you also to Olivia Mekhael, Pari Yazdanshenas, and Vaishna Kumaran for inviting me to partake in your own projects and help them move forward.

To Dr. Anna Dvorkin-Gheva for the incredible work she's put into this project, enabling us to pursue paths of inquiry and analysis only possible with someone of her caliber. Thank you for your care, your kindness, and expertise.

To Mary Jo Smith and Mary Bruni at the McMaster Immunology Research Centre John Mayberry histology facility for their technical help with the various histological stains required for this project to proceed.

To the pathologists at the Cooperman Barnabas Medical Center and at the Mayo Clinic, Dr. Uzma Zafar and Dr. Julian Villalba, respectively, who, were excellent teachers in pathology, always patient and willing to elaborate on complex topics in pathology, who always treated me with empathy, and provided Ashcroft scores. Also, to Dr. Asghar Naqvi for the H-Score methodology and for selecting IPF lung biopsy regions for TMA construction.

To the Firestone Institute for Respiratory Health and the Research Institute of St. Joe's Hamilton for the technologies that were made available in the Molecular Phenotyping and Imaging Core facility and were critical for my project.

To those who trained me: Ehab Ayaub and Jewel Imani, scientists at Novartis and Vertex, respectively, and my former friends and colleagues, Nafis Wazed , James Murphy, Hemisha Patel, Manreet Padwal, and Jane Ann Smith.

To my friends and family, both Canadian and Mexican, who were always there for me when I needed them, who assisted with our move to Brantford and again to Quebec, and who helped us through a tough two-year pandemic.

Finally, to my wife Azucena Gonzalez Gomez, who supported me emotionally, financially, and even performed statistical analyses using machine learning to assist with the creation of the composite index. I am forever grateful.

TABLE OF CONTENTS

PRELIMINARY PAGES	ii-xiii
CHAPTER 1 : INTRODUCTION	1
1.1 Interstitial Lung Disease	1
1.1.1 Collagen and the Extracellular Matrix	1
1.1.2 Fibrotic Disease	1
1.1.3 Idiopathic Pulmonary Fibrosis	2
1.2 Pulmonary Fibrosis Research.....	3
1.2.1 Rodent models of Pulmonary Fibrosis	3
1.2.2 Composite Indices from Clinical Data	3
1.2.3 Composite Indices from Histology Data	4
1.2.4 Commonly Measured Fibrosis-Related Endpoints	5
1.2.5 ECM and Collagen Stains Trichrome and Picrosirius Red	5
1.2.6 Myofibroblast and Macrophage Receptors IHC	6
1.3 HALO® Software	7
1.3.1 Digital Pathology Platform	7
1.3.2 Tissue Segmentation	7
1.3.3 Tissue, Cellular, and Protein Analysis.....	8

1.4 Composite Index Derived from Machine Learning	8
1.4.1 eXtreme Gradient Boosting (XGBoost)	8
1.4.2 XGBoost Basic Principles	9
1.5 Hypothesis	10
1.5.1 Perform Ashcroft Scoring on Human Tissues	11
1.5.2 Develop Quantitative H&E Assessments	11
1.5.3 Compare TRI and PSR assessments	12
1.5.4 Develop Myofibroblast and Macrophage IHC Assessments	12
1.5.5 Create Statistical Models using the Ashcroft Score	13
CHAPTER 2 : METHODS.....	14
2.1 Human Lung Tissues.....	14
2.2 Tissue Microarrays	14
2.3 Histological Staining and Slide Digitization	16
2.4 Visual Assessment of Fibrosis Severity	17
2.5 Tissue Segmentation	17
2.6 Statistical Analysis and Index Development	26
CHAPTER 3 : RESULTS	30
3.1 Perform Ashcroft Scoring on Human Tissues.....	30

3.2 Develop quantitative H&E assessments using HALO®	34
3.3 Compare TRI and PSR assessments using HALO®	37
3.4 Develop quantitative α SMA assessments using HALO®.....	39
3.5 Repeat IHC assessments on CD68, CD163, and CD206.....	42
3.6 Create Statistical Models using the Ashcroft Score	48
CHAPTER 4 : DISCUSSION	56
4.1 Results.....	57
4.2 Limitations	64
4.3 Future Directions	66
4.3.1 Characterization of Vasculature in Alveolar Interstitium	67
4.3.2 Characterization of Mucin Secretions in Alveolar Airways	67
4.3.3 Characterization of Adipocytes in the Alveolar Interstitium.....	68
4.3.4 Quantification of Honeycombing of the Alveolar Interstitium	68
4.3.5 Larger Training and Testing Datasets	69
4.3.6 Translation to Pre-Clinical Models.....	69
CHAPTER 5 : CONCLUSION.....	70
CHAPTER 6 : SUPPLEMENTARY FIGURES	72
CHAPTER 7 : REFERENCES	75

LIST OF FIGURES AND TABLES

Figure 1 : TMA 1	15
Figure 2 : TMA 2.....	15
Figure 3 : TRI and PSR	17
Figure 4 : Airspaces.....	19
Figure 5 : TRI.....	20
Figure 6 : PSR.....	21
Figure 7 : H-Score	22
Figure 8 : Hemorrhage.....	23
Figure 9 : Secretions	24
Figure 10 : Adipocytes.....	25
Figure 11 : TMA 1 Ashcroft.....	30
Figure 12 : TMA 1 Ashcroft Comparison.....	31
Figure 13 : TMA 2 Ashcroft.....	32
Figure 14 : TMA 2 Ashcroft Comparison.....	33
Figure 15 : H&E Cells	34
Figure 16 : H&E Airspaces	35
Figure 17 : H&E Airspace Properties.....	36
Figure 18 : TRI.....	37
Figure 19 : PSR.....	38
Figure 20 : α SMA Cells.....	39

Figure 21 : α SMA Area	40
Figure 22 : α SMA H-Score	41
Figure 23 : Macrophage IHC Cells in Parenchyma.....	43
Figure 24 : Macrophage IHC Cells in Interstitium	44
Figure 25 : Macrophage IHC Area Stains in Parenchyma	45
Figure 26 : Macrophage IHC Area Stains in Interstitium.....	46
Figure 27 : Macrophage IHC H-Score	47
Table 1 : Model 1 Validation	49
Table 2 : Model 2 Validation	50
Figure 28 : XGB Model 1 F-Scores.....	51
Figure 29 : XGB Model 2 F-Scores.....	52
Table 3 : XGB Model 1 Validation.....	53
Table 4 : XGB Model 2 Validation.....	54
Supplementary Figure 1 : Ashcroft Score Criteria	72
Supplementary Figure 1 : XGBoost Model 1	73
Supplementary Figure 2 : XGBoost Model 2	74

LIST OF ALL ABBREVIATIONS AND SYMBOLS

Alpha-Smooth Muscle Actin	αSMA
Alveolar Epithelial Cell	AEC
Alveolar Macrophage	AM
Brightfield	BF
Clinical, Radiographic, and Physiologic	CRP
Extracellular Matrix.....	ECM
Forced Expiratory Volume.....	FEV₁
Forced Vital Capacity	FVC
Formalin-Fixed Paraffin-Embedded	FFPE
Hematoxylin & Eosin	H&E
Idiopathic Interstitial Pneumonia	IIP
Idiopathic Pulmonary Fibrosis.....	IPF
Immunohistochemistry	IHC
Multivariate Linear Regression.....	MLR
Pearson’s Correlation Coefficient	R²
Pentachrome.....	PTC
Picrosirius Red	PSR
Polarized Monochrome	POL-MONO
Red Blood Cell	RBC
Registered	®
Tissue Microarray.....	TMA
Trichrome	TRI
Univariate Linear Regression.....	ULR
Usual Interstitial Pneumonia	UIP
eXtreme Gradient Boost.....	XGBoost

DECLARATION OF ACADEMIC ACHIEVEMENT

This is a declaration that the work included in this thesis was completed by Spencer D. Reville, under the supervision of Dr. Kjetil Ask. The linear regression models were completed by Dr. Anna Dvorkin-Gheva and the XGBoost models were completed by Azucena Gonzalez Gomez.

CHAPTER 1 : INTRODUCTION

1.1 Interstitial Lung Disease

1.1.1 Collagen and The Extracellular Matrix

Connective tissue serves the critical function of providing structural support to tissues and can be classified into three categories: proper, specialized, and embryonic. Proper connective tissues are predominantly comprised of extracellular matrix (ECM), a protein scaffold-fluid complex that surrounds the primary cell type of the ECM, fibroblasts. These generate additional ECM components and maintain its homeostasis. The fluid component is a gel-like mixture of water, proteoglycans, glycoproteins and glycosaminoglycans, enabling nutrient dispersal and adequate hydration of the tissue. The protein-scaffold component is made of collagen fibers and elastic fibers, providing a balanced combination of strength and flexibility, especially important for the lung as it constantly expands and compresses (Vasković & McLaren, 2022).

1.1.2 Fibrotic Disease

Fibroblasts will convert into their activated myofibroblast form to generate and distribute additional collagen into the ECM in response to tissue damage until the damage has been repaired, a scar has been formed, and the cell is no longer required. This process can become unregulated leading to the overproduction of scar tissue and progressive fibrotic disease. In the case of fibrotic lung disease, as this process descends into further disrepair, myofibroblasts produce and

deposit excessive ECM into the interstitial space, resulting in interstitial thickening, tissue architecture remodelling, impaired gas exchange and eventually, total fibrous obliteration and organ failure (S. Wang et al., 2022).

1.1.3 Idiopathic Pulmonary Fibrosis

Interstitial lung disease (ILD) is a term that comprises many disorders from mild reduction of inflammatory processes to severe crippling fibrosis of the lungs (Weerakkody & Jones, 2011). Idiopathic interstitial pneumonias (IIP) are a subset of ILDs whereby the etiology is unknown, characterized by permeation of immune cells into the alveolar interstitium and accompanied by a fibrotic phenotype (Rock & Amini, 2008). Idiopathic pulmonary fibrosis is an IIP that is currently recognized as the deadliest manifestation of pulmonary fibrosis to date, characterized by both the presence of the histological pattern known as usual interstitial pneumonia (UIP) and the absence of causes for other lung diseases, such as drug toxicity for drug induced ILD, environmental exposures for silicosis, or collagen vascular diseases like lupus (Bickle & Weerakkody, 2012; Rasuli & Weerakkody, 2012). IPF predominantly affects those middle-aged and older, and the irreparable loss of lung function lends to a poor survivability, with a median survival of 3 to 5 years from the time of diagnosis (Glass et al., 2020). There are two accepted treatments that have been shown to slow the progression rate of the disease, Nintedanib and Pirfenidone, both FDA-approved in 2014, though neither reverse tissue damage and both have noted side effects (Liu et al., 2017).

1.2 Pulmonary Fibrosis Research

1.2.1 Rodent models of Pulmonary Fibrosis

Pre-clinical models can attempt to recapitulate fibrotic lung diseases. Models include asbestos, silica, age-related, cytokine overexpression, targeted cell injury, acid/hyperoxia, bleomycin, fluorescent isothiocyanate, radiation-induction, familial, cell-transfer, humanized, and infectious exacerbations. They are used to test drug therapies and deepen our understanding of these diseases, however IPF's yet undiscovered etiology renders any singular model incomplete (B. Moore et al., 2013; Moore & Hogaboam, 2008; Tashiro et al., 2017).

1.2.2 Composite Indices from Clinical Data

The idea of using a composite index, that is a multi-pronged approach to assessing a given topic, isn't new. The composite index scoring methodology most used by clinicians is the gender, age, and physiology (GAP) score for predicting patient mortality. The physiology variables are forced vital capacity (FVC) and carbon monoxide diffusing capacity (D_{LCO}) (Ley et al., 2012).

Another less commonly used composite index methodology is called the clinical, radiographic, and physiologic (CRP) score, which combines information from physician-led questionnaires, imaging information from chest radiographs, and physiology variables FVC and forced expiratory volume (FEV_1), to generate an overall score of fibrotic disease and even predict survivability (KING et al., 2001; Watters et al., 1986; Zisman et al., 2000).

1.2.3 Composite Indices from Histology Data

A multi-faceted approach that is made possible once tissue biopsy samples have been taken and used by pathologists is a fully quantitative method known as the Pathology score which requires several histological stains be performed, including a hematoxylin and eosin (H&E), a pentachrome stain (PTC), a Prussian blue, and a toluidine blue. Each stain is used to help assess specific parameters within the tissue and are combined in a composite index with a grading system on a 0 to 5 scale. These include specific observations in the airways, in the alveolar walls such as cellularity, honeycombing, percentage smooth muscle, cell metaplasia, and overall cellularity of inflammatory cells, among other endpoints. This system was specially designed for usage by pathologists as only they would have the knowledge and the training necessary to assess each parameter accurately and consistently (Cherniack et al., 1991).

Another method was designed by histopathologists that uses a semi-quantitative approach instead and is currently used for pre-clinical research. The Ashcroft score was developed by histopathologists, whereby histological staining is performed, such as H&E or a trichrome stain (TRI), the tissues are visualized at high magnification under a microscope and graded on a scale of 0 to 8, from healthy to total obliteration of the lung. Fibrosis in the lung is the only factor considered when determining the score (Ashcroft et al., 1988). This has since become a gold standard in lung research.

1.2.4 Commonly Measured Fibrosis-Related Endpoints

The Ashcroft score ignores other potential markers of disease, such as inflammatory cell aggregates, cellular secretions, and the presence of specific cell populations, such as myofibroblasts and immune cells. It does not make use of any quantitative measurements. Therefore, most fibrosis researchers perform further analysis using additional histology and/or the physical measurements of the animals. These can include lung resistance and compliance by *in vivo* measurements, collecting bronchoalveolar lavage fluid (BALF) upon sacrifice of the animal to perform cell counts and cell differentials (Leitz et al., 2021; Ley et al., 2012; Tsitoura et al., 2021).

Post-lung extraction protocols include performing biochemical assays using lung homogenate, BALF or collected blood, acquiring various histological stains such as H&E for structural assessment, TRI, PTC or PSR stain for collagen and extracellular matrix assessment, immunohistochemistry (IHC) or immunofluorescence (IF) for protein examination, in situ hybridization (ISH) or fluorescent in situ hybridization (FISH) for mRNA transcript examination, RNA assessments via qPCR and differential gene expression (Humphries et al., 2021; Sul et al., 2022; Westermann-Clark et al., 2022).

1.2.5 ECM and Collagen Stains Trichrome and Picrosirius Red

Masson's or Lillie's trichrome stain, literally meaning 3 colours, is a connective tissue stain that utilizes two or more acid dyes of contrasting colours

that differentiate the tissue by colour: nuclei appear black-brown-purple, in red is keratin, muscle fibers and red blood cells, in pink is fibrin and cytoplasm, and in either blue, from the aniline blue dye, or green, from the fast green FCF dye, are collagen fibers and the extracellular matrix (Llewellyn, 2019).

Picrosirius red (PSR) is a collagen stain made by combining yellow picric acid with the molecule Sirius red. When viewed in brightfield, the colour palette is primarily reds and yellows, however when visualized using a polarizer in colour light, only collagens type I and III, found in bronchi, alveolar interstitium and blood vessels, are visible, with a black background. Collagen I will shine orange-yellow while collagen III shines green. This is due to the collagen's birefringent properties at the molecular level being enhanced by the PSR stain (Junqueira et al., 1979; López De Padilla et al., 2021).

1.2.6 Myofibroblast and Macrophage Receptors IHC

Alpha-Smooth Muscle Actin (α SMA) is a cell marker for myofibroblasts (Yang et al., 2022) and myofibroblast-related activity (Tugcu et al., 2022), which is responsible for ECM restoration, and fibrosis progression (Shinde et al., 2017).

Macrophages have been implicated as active contributors to the fibrotic progression in IPF such as producing the profibrotic cytokine transforming growth factor beta (TGF- β) and promoting fibroblast-to-myofibroblast differentiation (McErlean et al., 2021; Mou et al., 2022). Macrophages have demonstrated an innate plasticity, regarded as being polarized towards a proinflammatory

phenotype, labelled M1, or a profibrotic, or pro-wound-healing, phenotype, labelled M2 (McErlean et al., 2021).

Data concerning M2-like macrophages suggest that they are involved and upregulated in IPF and possess the receptors CD163 (Li et al., 2021) for bacterial and inflammatory identification (GeneCards, n.d.-a), and CD206 (Cao et al., 2022; Fabrick et al., 2005; Komohara et al., 2008) for viral, bacterial, and fungal phagocytosis (GeneCards, n.d.-b). Meanwhile, all macrophages are thought to express the CD68 scavenger receptor, which serves as a pan-macrophage marker (Wu et al., 2022).

1.3 HALO® Software

1.3.1 Digital Pathology Platform

HALO® is a software licensed by Indica Labs used to perform quantitative analysis on digitized histological slides. It is a modular platform, currently available as version 3.4.2986.185, whereby each user can have access to the specific tools they require for their research. HALO® can provide morphological data for cell populations within the tissue and overall tissue classification data.

1.3.2 Tissue Segmentation

Digitized tissues can be annotated using HALO®'s built-in tools, such as the brush, scissors, and flood tools, or using HALO®'s modules, such as the TMA (Indica Labs, 2019) and Classifier (Indica Labs, 2016) modules.

1.3.3 Tissue, Cellular, and Protein Analysis

Segmented tissues can be analyzed using any of HALO®'s quantitative modules. Those relevant to this project include the Vacuole module to measure alveolar area in the parenchyma (Indica Labs, 2018), the Classifier module to measure the area of the pulmonary compartments (Indica Labs, 2016), the Multiplex IHC module to count the number of cells in the tissue and identify the proportion of those cells that are found positive for a protein of interest as stained with IHC (Indica Labs, 2017), the Area Quantification module to measure the area occupied by a protein of interest as stained with IHC (Indica labs, 2014), and finally the Area Quantification FL module to measure the area occupied by birefringent collagen in the lung digitized under polarized monochromatic light (Indica Labs, 2014).

1.4 Composite Index Derived from Machine Learning

1.4.1 XGBoost Overview

eXtreme Gradient Boosting (XGBoost) is an algorithm that creates decision trees that utilize independent variables, such as experimentation data, to generate a model where the variables are weighted according to their predictive power and can produce a value that is as similar as possible to the dependent variable against which it is being compared. The decision trees are added together to improve the performances of other trees, and models are fitted in a manner akin to a neural network (Trunfio et al., 2022).

1.4.2 XGBoost Basic Principles

Bagging is a principle whereby many of these decision trees are being created at once, each one using different random criteria, combining those decision trees, and voting on which is most correct. Random forest is a form of bagging whereby some decision trees are limited to testing and analyzing no more than a subset of randomly selected parameters at a time (Morde, 2019).

Boosting is a principle whereby information garnered from some decision trees are passed on to other decision trees to increase the efficiency of the analytical process, enhancing previously less accurate decision trees. Gradient boosting is a form of boosting where errors are minimized using a gradient descent algorithm, whereby target outputs are specified and every decision tree reduces prediction errors by gradually improving their accuracy (Morde, 2019).

This process, in combination with software optimizations, is what lends the eXtreme portion of the name XGBoost. This form of machine learning also benefits from parallelization, where many processes are done simultaneously, and regularization, where XGBoost penalizes decision trees that might be considered too accurate based solely on the training dataset, as it would likely fail when used on any test sample (Morde, 2019).

The result is a composite index in the form of an algorithm that takes experimentation data as inputs and outputs a single score,

1.5 Hypothesis

Overall Hypothesis

I hypothesize that by developing a composite index, whereby several primary histological endpoints are collected, measured, and combined, we can assess the severity of fibrosis in human lung tissues with increased accuracy and reduced bias as compared to relying on any single primary readout or a semiquantitative method like the Ashcroft score.

Specific Aims

I believe a novel and objective method of assessing lung fibrosis severity is warranted due to the subjective nature of the current gold standard; the Ashcroft scoring method involves visually assessing histological slides and assigning a score from 0 (normal lung) to 8 (total fibrous obliteration of the field). This method was developed by pathologists and is currently being used in pulmonary research laboratories.

I aim to develop a tool, a composite index, that would be provided objective quantifications, as opposed to potentially biased visual assessments, from histological samples as inputs, compute them using statistical models or machine learning algorithms that were trained using data from pathologists and data from researchers who were trained by pathologists, and output a result for fibrosis severity that would be accurate and reliable.

1.5.1 Perform Ashcroft Scoring on Human Tissues

Two tissue microarrays (TMAs) were made comprised of a total of 73 confirmed IPF tissues and 12 control tissues. These tissues would need to be scored using the currently accepted gold standard, the Ashcroft score. Each scorer was provided with Table 1 from the (Ashcroft et al., 1988) publication (**Supplementary Figure 1**) outlining the tiers of severity upon which the Ashcroft score is based. The IDs of each tissue from both TMAs were hidden from the 6 pulmonary scientists and 2 pulmonary pathologists, and they were shown in a randomized order. Each scorer kept their scores private. Everyone's results were averaged per tissue ID such that each possessed an average Ashcroft score (n=8).

1.5.2 Develop quantitative H&E assessments

A visual assessment of the tissues stained with H&E provides a clear demonstration of which tissues appear healthy and which appear diseased (**Figures 1, 2, and 4**). We hypothesized that this phenomenon must be measurable. Using the HALO® (v3.4.2986.185) digital pathology platform by Indica Labs, specifically the TMA, Classifier, Multiplex IHC, and Vacuole modules (Indica Labs, 2016, 2017, 2018, 2019), various quantifications were performed on H&E slides. Both TMAs were annotated quantified for the total number of cells, the proportions of the parenchyma that comprise airway versus interstitium, as well as the measurements of the airways in the parenchyma.

1.5.3 Compare TRI and PSR assessments

TRI and PSR measure similar variables: TRI stains all the collagen in the ECM blue, while PSR, under polarized light, only shows collagens I and III. It is possible that there are other collagens, such as type IV, that are not being measured in PSR, while it is also possible that TRI is measuring more than just collagen; perhaps some other portions of the ECM are being picked up as well.

We hypothesized that measuring collagen using TRI and PSR would yield the same measurements and the same conclusions regarding lung fibrosis severity. HALO®'s Classifier (Indica Labs, 2016) module enable us to quantify the different colours in the TRI and PSR brightfield (PSR-BF) slides, while the Area Quantification FL (Indica Labs, 2014) enables us to quantify the white signal, collagens I and III, from the polarized monochrome PSR (PSR-POL-MONO) image. Then, we can determine the proportion of the parenchyma that is positive for TRI's blue stain and the proportion that is positive for PSR-POL-MONO signal, respectively.

1.5.4 Develop Myofibroblast and Macrophage IHC assessments

α SMA is present naturally on vasculature, specifically the smooth muscle cells that surround them for support. After excluding all vessels larger than capillaries, cross-sectional area $\geq 250 \mu\text{m}^2$, from analysis, HALO®'s Multiplex IHC (Indica Labs, 2017) module enables us to localize cell nuclei and identify whether an IHC stain is present within cellular vicinity, as defined by the

cytoplasm, in sufficient intensity as to be considered positive for the protein of interest. The Area Quantification (Indica labs, 2014) module can quantify the total area that is positive for the protein of interest. These were used to quantify the number of cells, which ones are protein-positive, to what degree they are positive, weak, moderate, or strong, as defined by the user. I have had training by two pathologists on the topic of binning cells according to their staining intensity, with examples provided in **Figure 7**.

The process is the same for CD68, CD163, and CD206, except that vasculature does not express any of these markers, so removing the vasculature as was done with α SMA was not needed.

1.5.5 Create Statistical Models using the Ashcroft Score

Experts in the fields of bioinformatics and data science were recruited for the final portion of this experiment, to create multivariate linear regression (MLR) models and XGBoost machine learning models, respectively, using the data collected throughout the project and using the Ashcroft scores. The first MLR model used every variable we had collected, and we found that 16 of 95 variables were contributing the most to the overall model. With only 16 variables remaining, we kept running both the MLR and XGB models with additional caveats, such as only using TRI or PSR or neither, but not both, for example. Two models each were made using the MLR approach and using the XGB approach.

CHAPTER 2 : METHODS

2.1 Human Lung Tissues

All work conducted using human tissues was approved by the Hamilton Integrated Research Ethics Board (11-3559 and 13-523-C). Formalin-fixed paraffin-embedded (FFPE) human lung tissues were obtained from a biobank for interstitial lung diseases at St. Joseph's Healthcare in Hamilton, Ontario, Canada. IPF lung biopsies were selected based on clinical, radiological, and a pattern of usual interstitial pneumonia determined by a trained molecular pathologist. Non-cancerous regions of tissue from biopsies from lung cancer cases were used as non-disease controls.

2.2 Tissue Microarrays

Fibrotic regions from FFPE IPF tissues and non-tumorous regions from FFPE control tissues were selected to be placed into two tissue microarrays (TMAs) (**Figures 1 and 2**). The TMAs were created using a 3D Histech TMA Master II semi-automated tissue microarrayer. Specific regions within the IPF lung biopsies were identified by a trained molecular pulmonary pathologist, extracted using the tissue microarrayer's equipped 2-millimetre diameter tissue core extractor, and placed into pre-drilled receiving paraffin blocks. The finished TMAs were placed in a 60°C oven for 1 hour to meld the wax to the inserted tissue cores, then placed in a 4°C overnight to harden. Finally, excess paraffin wax was trimmed using a microtome to ensure the TMAs were flat for histological staining.

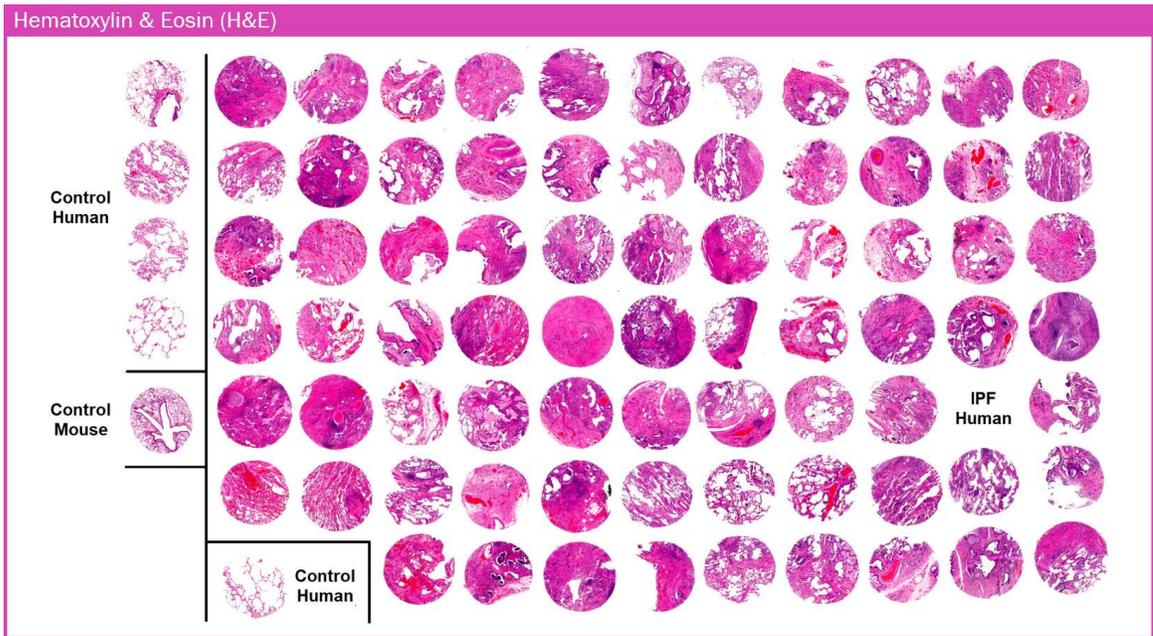


Figure 1 : TMA 1

A tissue microarray stained with H&E containing 75 cases of IPF and 5 cases of control tissues.

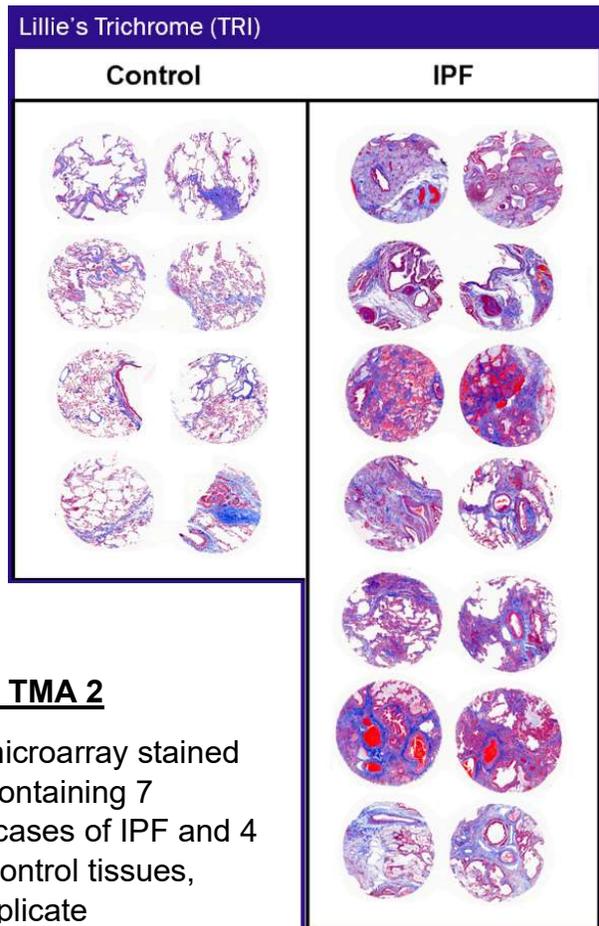


Figure 2 : TMA 2

A tissue microarray stained with TRI containing 7 duplicate cases of IPF and 4 cases of control tissues, both in duplicate

2.3 Histological Staining and Slide Digitization

The TMAs were processed for histological staining at the McMaster Immunology Research Centre John Mayberry histology facility using a Leica Bond Rx auto-immunostainer. The reagent kits are instrument- and application-specific (Richmond Hill, Ontario). Brightfield microscopy was performed using an Olympus BX-61 motorized system microscope which is a component of the Olympus VS120 slide loader system. H&E, immunohistochemistry (IHC), TRI, and PSR slides were digitized at 20× objective magnification (**Figure 3**). High definition images of all stains were acquired in this view using an Allied Vision Pike F-505C CCD camera with a pixel resolution of 2452×2054 and a pixel size of 3.45×3.45 μm. Additional visualization and imaging of PSR was performed using polarized microscopy (**Figure 3**) with a Hamamatsu ORCA-Flash4.0 V3 sCMOS monochrome camera possessing a pixel resolution of 2048×2048 and a pixel size of 2.55×2.55 μm, and an Olympus U-Pot drop-in polarizer was used in conjunction with the transmitted light. The VS120 uses the VS-ASW v2.9.2 software that can perform auto focusing, shading correction, auto white balance, and image-stitching, which enables digitization of whole-slides.

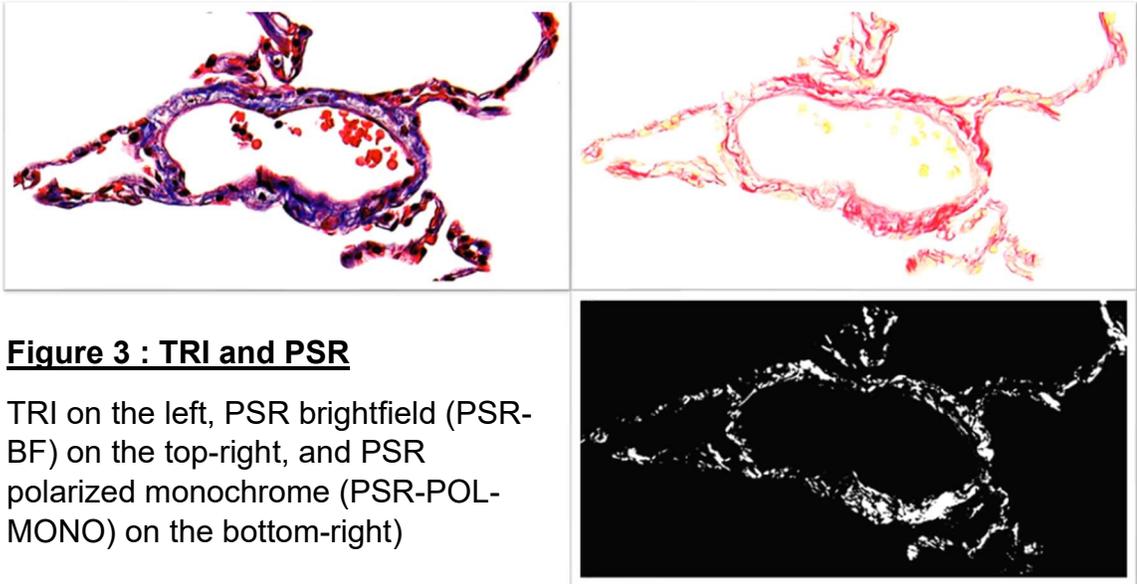


Figure 3 : TRI and PSR

TRI on the left, PSR brightfield (PSR-BF) on the top-right, and PSR polarized monochrome (PSR-POL-MONO) on the bottom-right)

2.4 Visual Assessment of Fibrosis Severity

Ashcroft scores were performed for each case by 8 individuals: six trained scientists with experience in lung fibrosis and two molecular pulmonary pathologists, experts in interstitial lung disease (ILD). Guidelines on scoring were provided according to the instructions published (Ashcroft et al., 1988) (**Supplementary Figure 1**). Cases were randomized, scoring was kept confidential between scorers, and the 8 scores per case were pooled into a single average Ashcroft score.

2.5 Tissue Segmentation and Quantification

Each TMA was subjected to HALO® analysis (v3.4.2986.185). The images were annotated to include only lung parenchyma, excluding airways and vessels with a cross-sectional area $\geq 0.05\text{mm}^2$, as anything more stringent yielded little-

to-no difference in measurements while immensely increasing the amount of work, and their support structures, including, but not limited to cartilage, smooth muscle, and collagen in the adventitia. These structures were identified and subsequently excluded in accordance with three molecular pulmonary pathologists. Analyses took place in the parenchyma, the combination of the alveolar interstitium and the alveolar airspaces, the region of the lung where gas exchange formally takes place. The outer perimeter of the parenchyma was drawn using HALO®'s flood tool which automatically outlines areas of tissue with similar RGB values, creating an annotation along contiguous stretches of tissue. Any artifacts present on the image because of particulate or other foreign matter trapped beneath the slide's coverslip was also excluded from analysis. For alpha-smooth muscle actin (α SMA) specifically, all vasculature besides capillaries were excluded due to vessel smooth muscle cells being high expressors of α SMA.

The digitized H&E slides were analyzed using HALO®'s Classifier (Indica Labs, 2016) module to bin the regions within the parenchyma into two categories: alveolar airspace, defined by RGB values that produce an approximate white colour, and the other is alveolar interstitium, defined by RGB values that produce non-white colours. HALO®'s Multiplex IHC (Indica Labs, 2017) was used to detect and count the number of cells present within the parenchyma by detecting the nuclei. HALO®'s Vacuole (Indica Labs, 2018) module is designed to quantify alveoli in the lung and characterize them according to their average area, diameter, and perimeter (**Figure 4**).

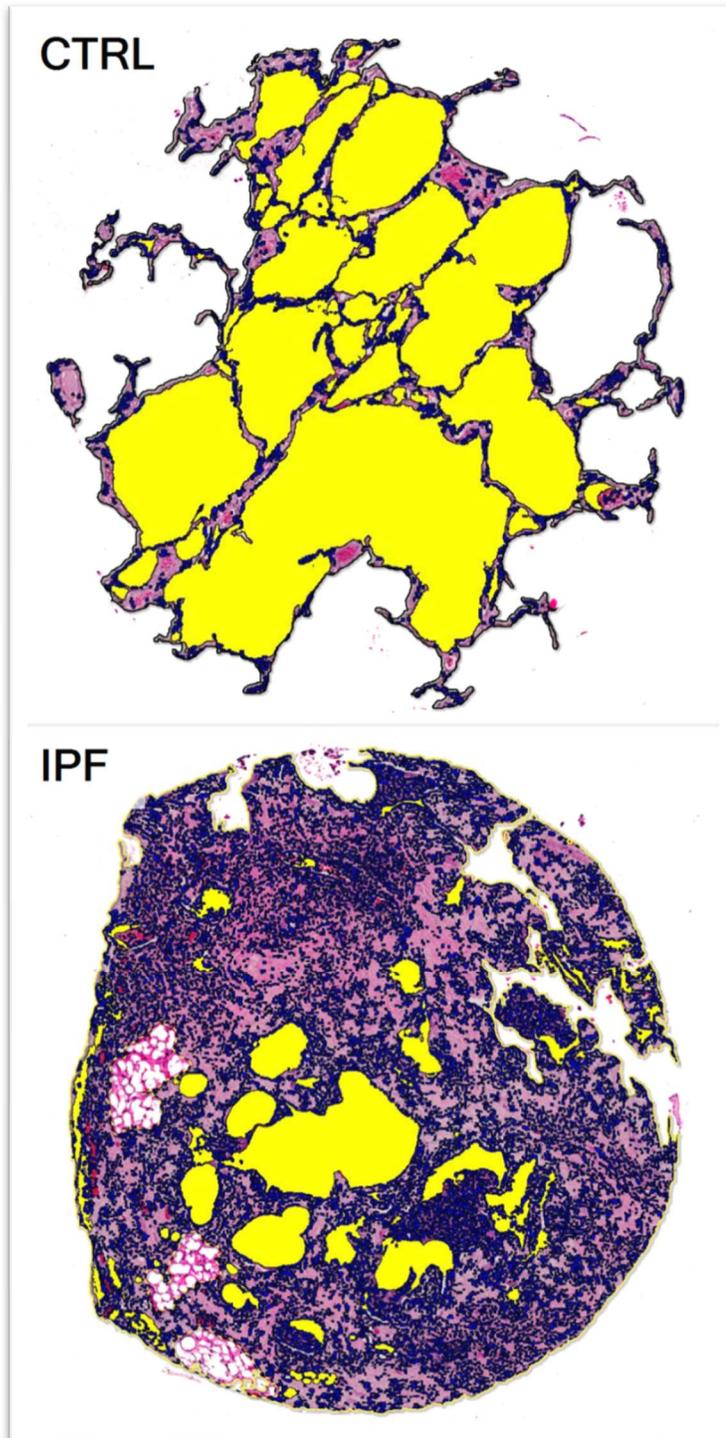


Figure 4 : Airspaces

A control case (CTRL) and an IPF case, both stained with H&E and analyzed for the characteristics of their airspaces (highlighted in yellow), including the number of airspaces, their average area (mm^2), average diameter (mm) and average perimeter (mm).

Note 1: As mentioned, the flood tool creates annotations along contiguous stretches of similar RGB values, or tissue, hence why some outer airspaces are left incomplete; there was insufficient tissue to completely close them. Attempting to close them manually by altering the annotation using the brush or pen tool would introduce bias, so this method was avoided.

Note 2: 3 regions of adipose tissues seen in bottom-left corner of IPF tissue. These were excluded from analysis.

The digitized TRI slides were also analyzed using the Classifier (Indica Labs, 2016) module, but the parenchyma was divided into three bins: alveolar airspace, collagenous alveolar interstitium, (coloured blue), and non-collagenous alveolar interstitium (coloured shades of pink, red, purple, and grey) (**Figure 5**).

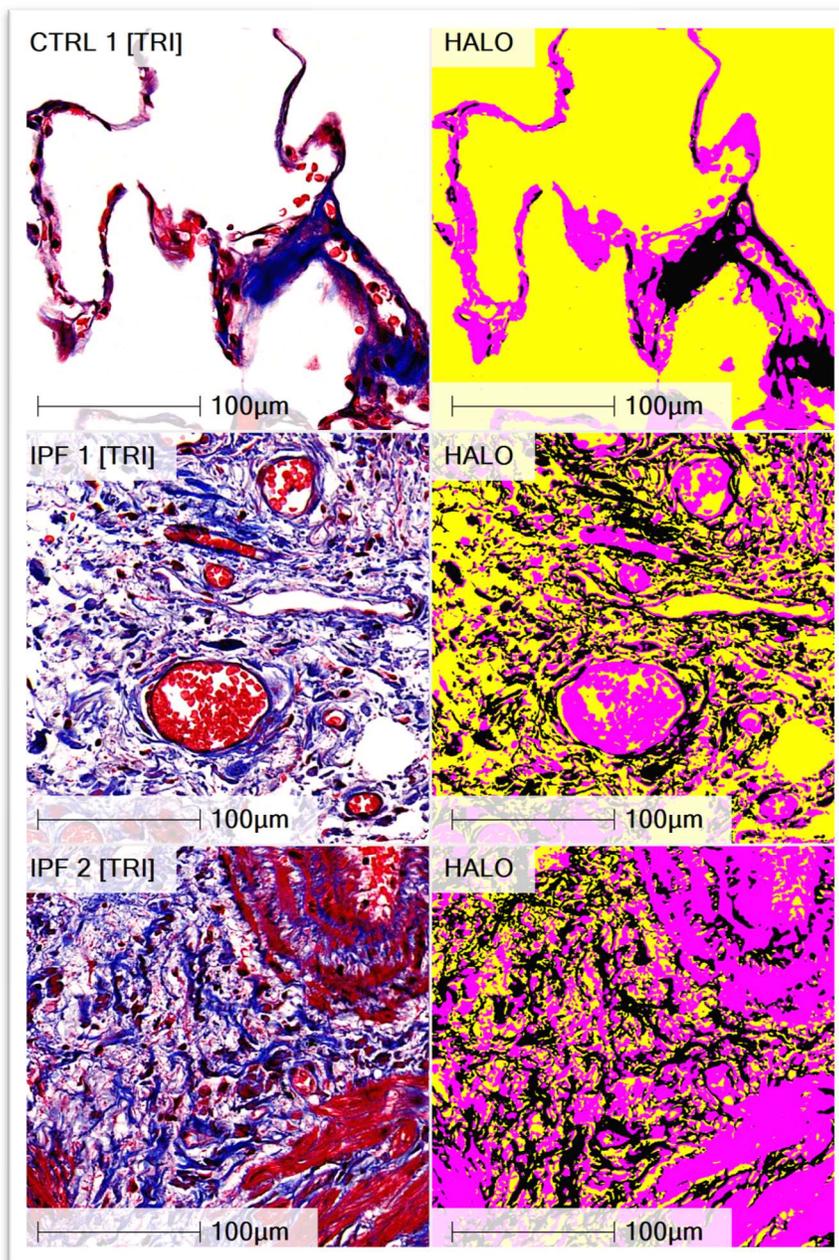


Figure 5 : TRI

TRI stain from 3 different tissues at high magnification, 1 per row. Left column shows brightfield, right column shows how the Classifier module binned the images into 3 regions: airspace (yellow), collagenous interstitium (black), and non-collagenous interstitium (pink).

The digitized PSR slides were acquired in brightfield and monochrome polarized settings (Figure 3). The latter shows collagen I and III fibers in brilliant white, due to their birefringent properties, on a pure black background. The brightfield image was dually binned into alveolar airspace and alveolar interstitium with the Classifier (Indica Labs, 2016) module, while the polarized image was analyzed using the Area Quantification FL (Indica Labs, 2014) module to quantify the collagen fibers present within the parenchyma (**Figure 6**).

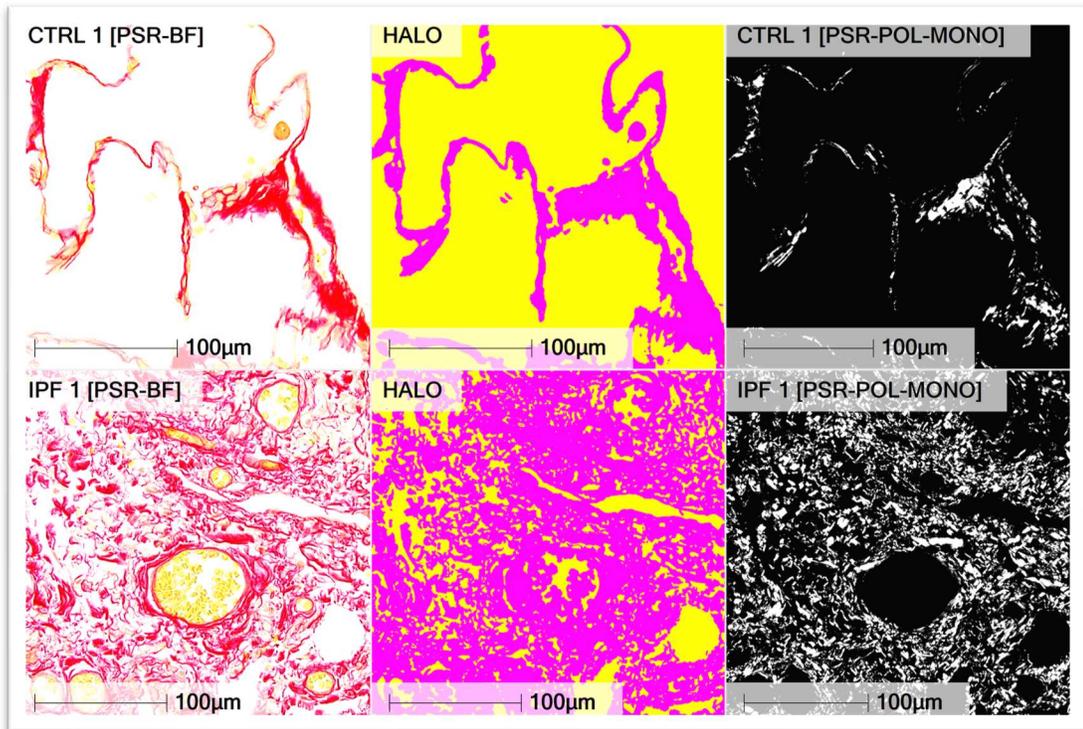


Figure 6 : PSR

PSR stain from 2 different tissues at high magnification, 1 per row. Left column shows brightfield, center column shows how the Classifier module binned the images into 2 regions: airspace (yellow) and interstitium (pink), right column shows PSR under polarized monochrome whereby only the collagens I and III are visible as white signal.

The digitized immunohistochemistry (IHC) slides, including α SMA, CD68, CD163, and CD206, were also analyzed using HALO®'s Classifier (Indica Labs, 2016) and Multiplex IHC (Indica Labs, 2017) modules to bin the parenchyma, count the total number of cells, and determine which cells were positive for the IHC stain and to what intensity. This staining intensity binning was done in accordance with a molecular pulmonary pathologist. Cells can be negative (0), weak expressors (+1), moderate expressors (+2) or strong expressors (+3). The overall cell percentages are used to calculate the H-Score (**Figure 7**).

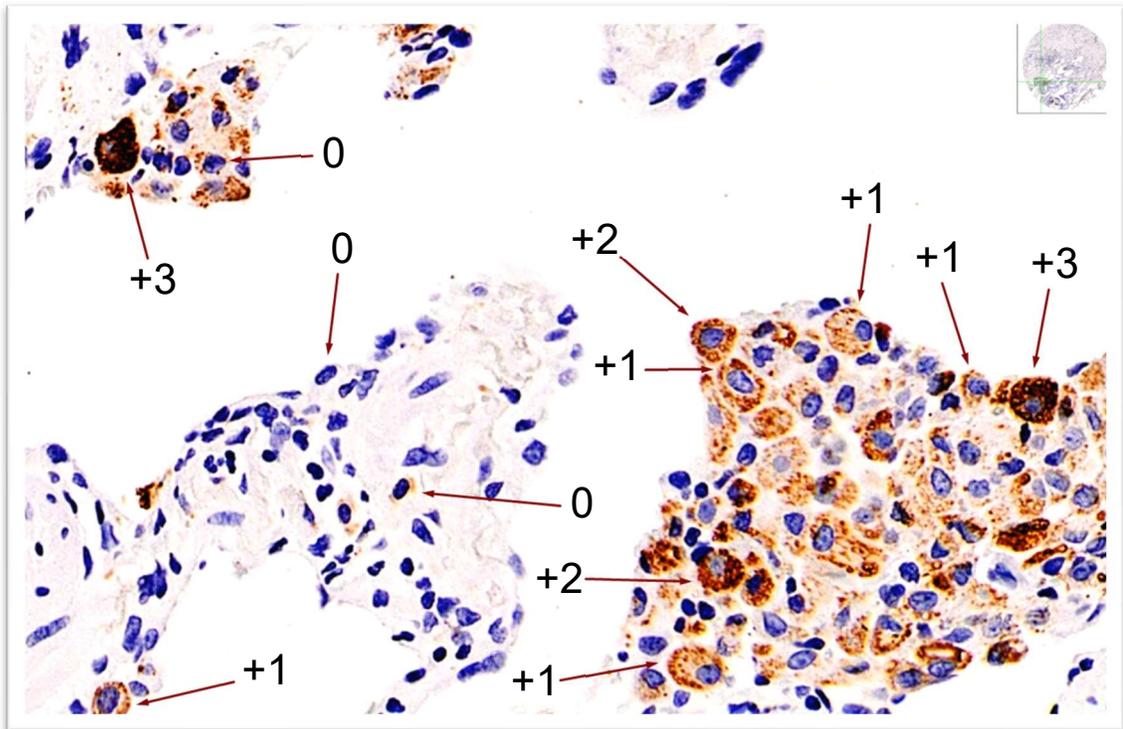


Figure 7 : H-Score

Generic IHC stain where cells are rated based on their staining intensity.

Cases predominantly characterized by hemorrhaging (**Figure 8**), where there were abundant red blood cells throughout the parenchyma, were excluded from analysis as it was unknown whether the blood was present naturally from the IPF patient or if its presence originated due to the surgical biopsy procedure.

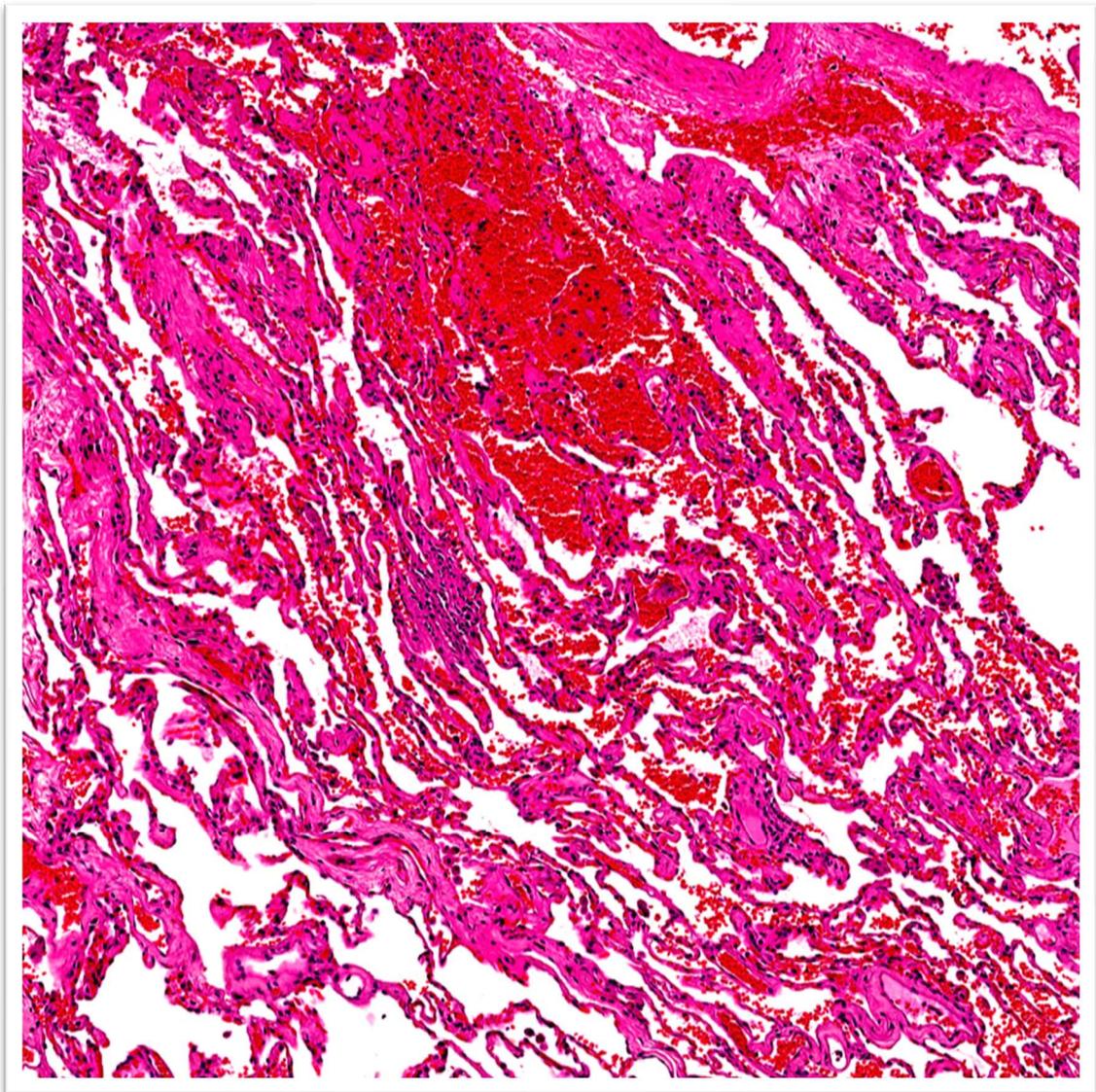


Figure 8 : Hemorrhage

H&E tissue with significant hemorrhaging throughout the parenchyma.

Similarly, cases predominantly characterized by secretions in the alveolar airspaces (**Figure 9**) were also excluded, as it was unknown whether the mucin secretions were related to the IPF diagnosis, though the role of mucins in profibrotic intracellular signalling has yet to be elucidated (Calabrese et al., 2022), or if their presence was instigated by an infection or allergies.

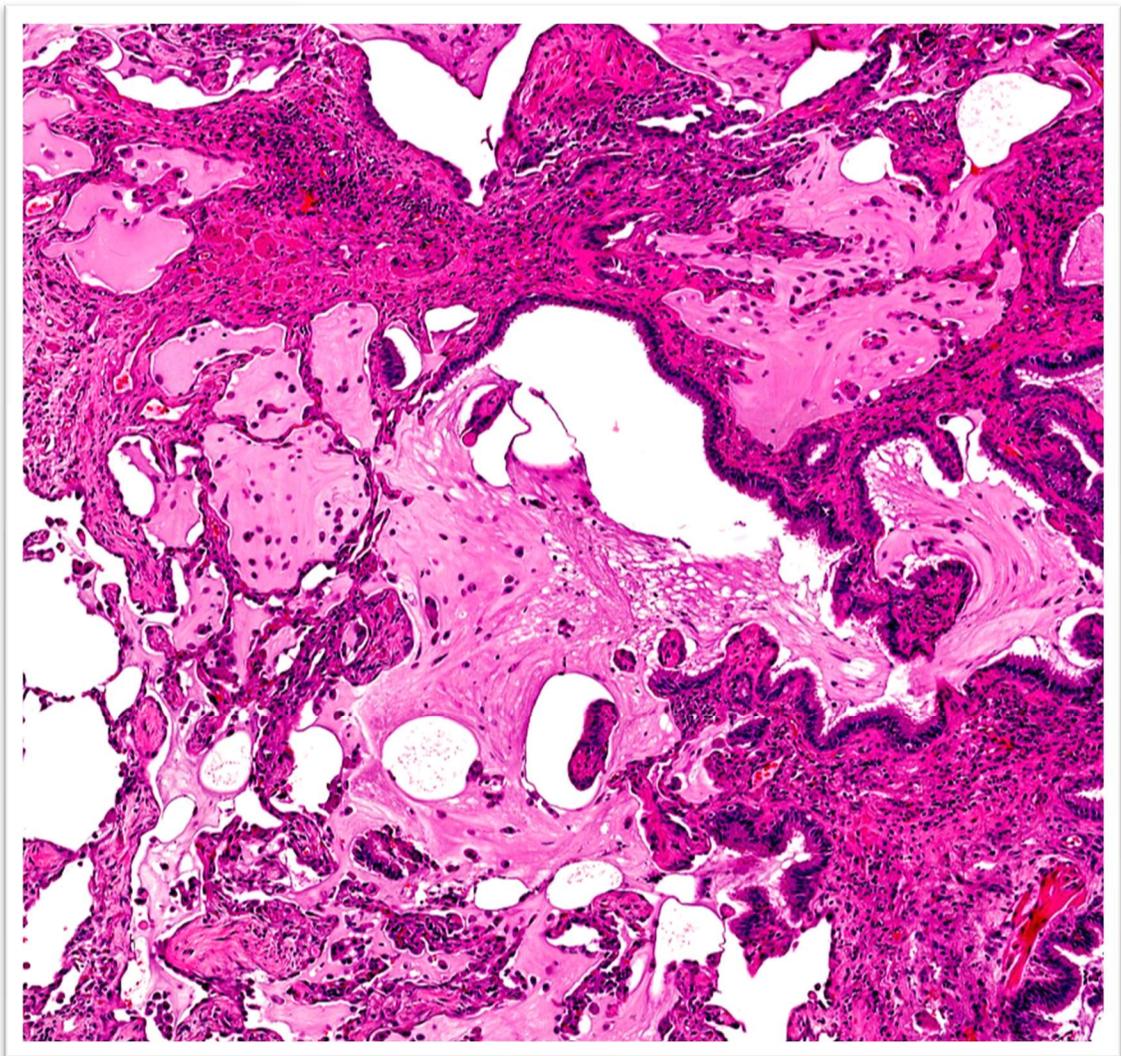


Figure 9 : Secretions

H&E tissue with significant protein secretion within the alveolar airspaces.

Finally, some cases were characterized by the presence of adipocytes (Figure 10) within the alveolar interstitium. These fat cells were excluded from analysis since their translucent and circular appearance would have been confounded by HALO®, mistaking them for alveolar airspace.

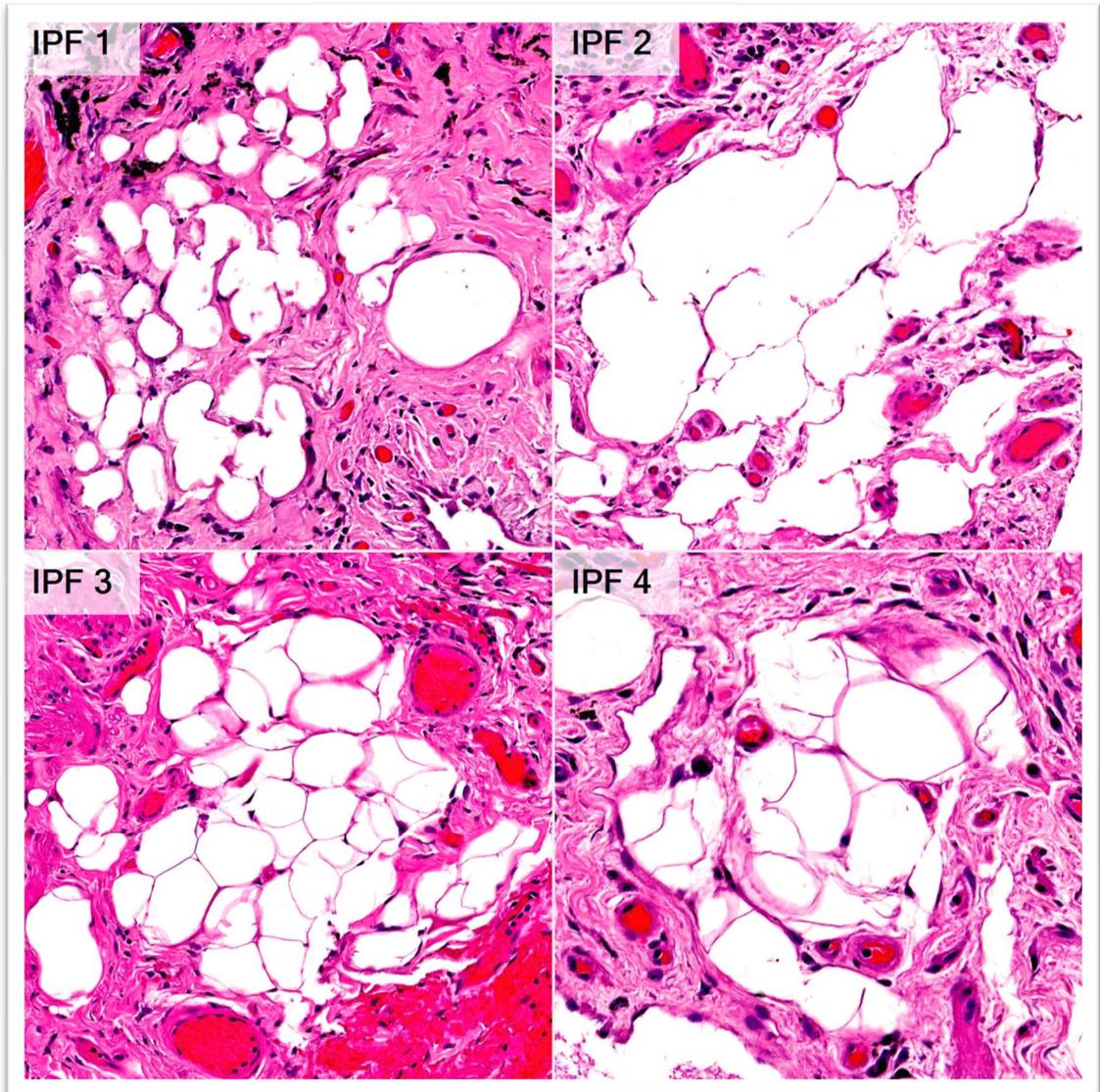


Figure 10 : Adipocytes

H&E tissue with adipose tissues (adipocytes) within the alveolar interstitium.

2.6 Statistical Analysis and Index Development

Among the results collected, total cells (count), α SMA-positive area (mm^2), and number of airspaces (count) were measured. These type of data was deemed ill-suited for analysis since they are absolute values that would increase proportionally to the overall size of the tissue that was assessed. To proceed with the data, it needed to be normalized, such as by changing total cells (count) to cellular density (count/mm^2) by using parenchyma area or interstitium area (mm^2) as a denominator. All variables used in the statistical analyses were normalized as ratios or percentages, in this manner.

All results prior to the creation of the statistical models (or composite indexes) were expressed as median \pm (1.5 * IQR). All graphs were statistically analyzed using Wilcoxon's Rank-Sum test, also known as Mann-Whitney's test. This test was chosen as its utility is derived from comparing the differences between two independent samples (i.e., control tissues and IPF tissues) when the sample distributions are not normally distributed, and sample sizes are small. This is important, since there is no expectation for the results to be distributed normally (i.e., shaped like a bell-curve) as these cases were all chosen by a pathologist specifically because they were either diseased or non-diseased. As such, there is, in fact, an expectation that one group will specifically possess values that are statistically greater or lower than the other group in most cases. The exceptions are Figures 12 and 14, which were analyzed with the similar Kruskal-Wallis test since there were 3 groups instead of 2.

Statistical models were created using two separate approaches, the first used the R programming language to perform multivariate linear regression (MLR) and second used the Python programming language to perform machine learning using eXtreme Gradient Boosting (XGBoost).

Prior to both approaches, all the Ashcroft scores and the data acquired from the HALO® quantifications were pooled into a single dataset, creating a pool of 73 IPF cases and 12 control cases in our possession. Then, we used function in python to semi-randomly split the single pooled dataset into 2 smaller datasets, one comprised of 80% of the tissues, which would serve as the training dataset for both approaches, and the other comprised of the remaining 20% of the tissues. The caveat for this random segregation was that the overall proportion of the distribution of Ashcroft scores between both smaller datasets should be near equal. This would result in two datasets that have similar proportions of tissues with low Ashcroft scores to tissues with high Ashcroft scores. This is essential for the training dataset so as not to inadvertently create a model that is exceedingly capable at distinguishing between tissues with Ashcroft scores from 4 to 8, as is typical with IPF cases, but poor at distinguishing between tissues with Ashcroft scores from 0 to 3, typical of non-disease control cases.

The former approach utilized the entirety of the data provided by the first TMA alone as the training dataset to perform univariate linear regression to identify variables that were seemingly predictive of the pooled Ashcroft scores for each case, a process known as feature selection. These features were then used

in a multivariate regression based on backward elimination yielding the final set of variables and their coefficients to be used in a $y = mx + b$ style of formula. This was then repeated with two specific restrictions set into place: first, preventing the inclusion of more than one of TRI and PSR simultaneously in the final model, and second, preventing the inclusion of more than one of CD68, CD163, and CD206 simultaneously in the final model. The rationale behind these model restrictions was to make the task of assessing fibrosis severity using this novel method easier for the researchers by reducing the time required and the costs of staining, slide digitization, tissue quantification, and data analysis required to run the model while also reducing possible redundancies; TRI and PSR are both said to measure collagen, while CD68, CD163, and CD206 are all macrophage markers. The variables that were chosen for each model because of these restrictions was determined by whichever was found to provide greater contribution to each specific model's overall performance. This makes it possible for one variable to be included in one model while a different one is included in another model.

The latter approach was done by running an eXtreme Gradient Boosting (XGBoost) machine learning algorithm in python with scikit-learn (Pedregosa et al., 2011). 16 variables were chosen as independent variables to conduct the regression analysis, the selection criteria were the following:

(1) The dependent variable was the average of the observations by all individuals, each given the same weight. (2) The sci-kit learn `train_test_split()` method was used to split the samples randomly into training and testing sets

using an 80%/20% split due to having a reduced sample. The distribution of “control samples” (scored below 3) and “disease samples” was conserved.

Hyperparameter tuning was conducted for XGBoost, the objective was set to squarederror

```
param_tuning = {  
    'learning_rate': [0.01, 0.1, 0.3],  
    'max_depth': [10, 3, 2],  
    'min_child_weight': [0.4, 0.3, 3, 5],  
    'subsample': [ 0.4, 0.6, 0.2],  
    'colsample_bytree': [0.4, 0.8, 0.6, 0.8],  
    'n_estimators' : [800, 500, 1000, 300, 1500],  
    'objective': ['reg:squarederror']  
}
```

The ideal parameters for our training set were determined to be

```
objective = 'reg:squarederror',  
colsample_bytree = 0.8,  
learning_rate = 0.1,  
max_depth = 10,  
min_child_weight = 3,  
n_estimators = 500,  
subsample = 0.8)
```

We tested the model in the following ways: (1) Could the model use only one of TRI or PSR while remaining accurate and (2) how few variables does it need?

CHAPTER 3 : RESULTS

3.1 Perform Ashcroft Scoring on Human Tissues

3.1.1 Scientists and pathologists, both independently and pooled, generated statistically identical sets of Ashcroft scores in a test cohort.

Ashcroft scores were acquired as a basis for our future models against which our quantitative data would be compared. One tissue microarray (TMA) was used that possessed 5 control tissues and 75 IPF tissues, 13 were omitted due to hemorrhaging and secretions, leaving only 62 IPF tissues. The scientists (n=6), the pulmonary pathologists (n=2) and all participants pooled (n=8) independently yielded highly statistically significant results.

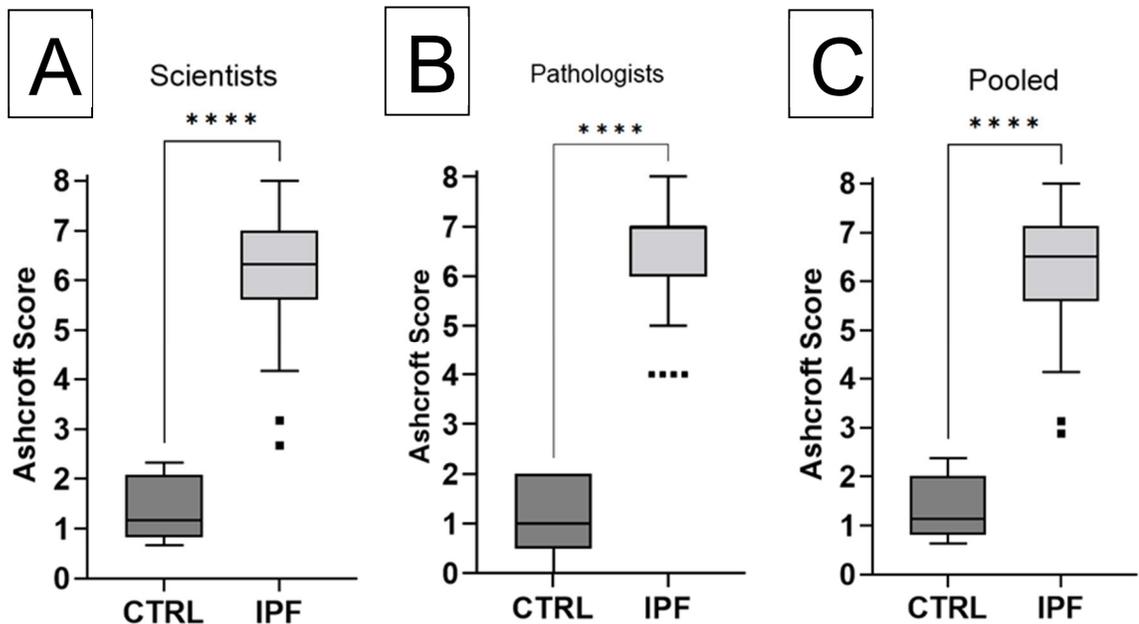


Figure 11 : TMA 1 Ashcroft

TMA 1 Ashcroft Scores (One-Tailed Wilcoxon Rank-Sum Test, $p < 0.0001$)

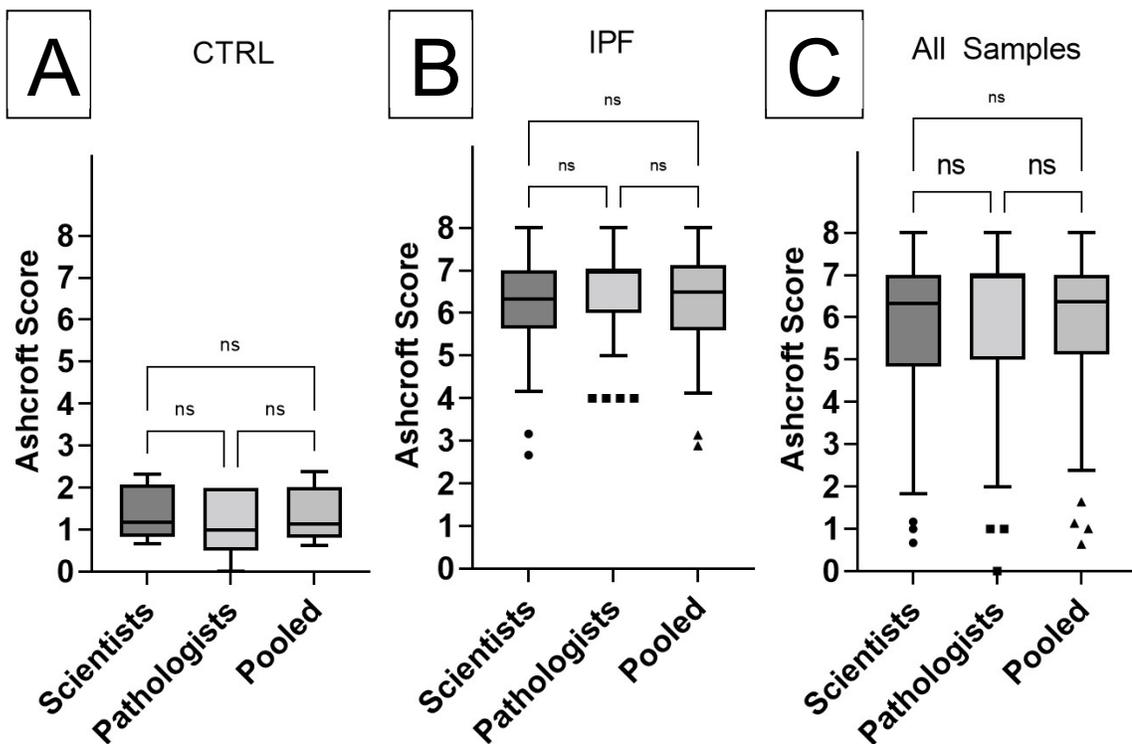


Figure 12 : TMA 1 Ashcroft Comparison

TMA 1 Ashcroft Score (Kruskal-Wallis Test, $p < 0.05$). No significant differences were found between the scientists and pathologists, nor their pooled averages, whether split into disease groups (A and B) or compiled together (C).

3.1.2 Scientists and pathologists generated statistically identical sets of Ashcroft scores in a validation cohort.

To serve as a validation cohort once the models have run, a second TMA possessing 8 control tissues and 14 IPF tissues were also scored using the Ashcroft method, with scientists and pathologists independent, then pooled.

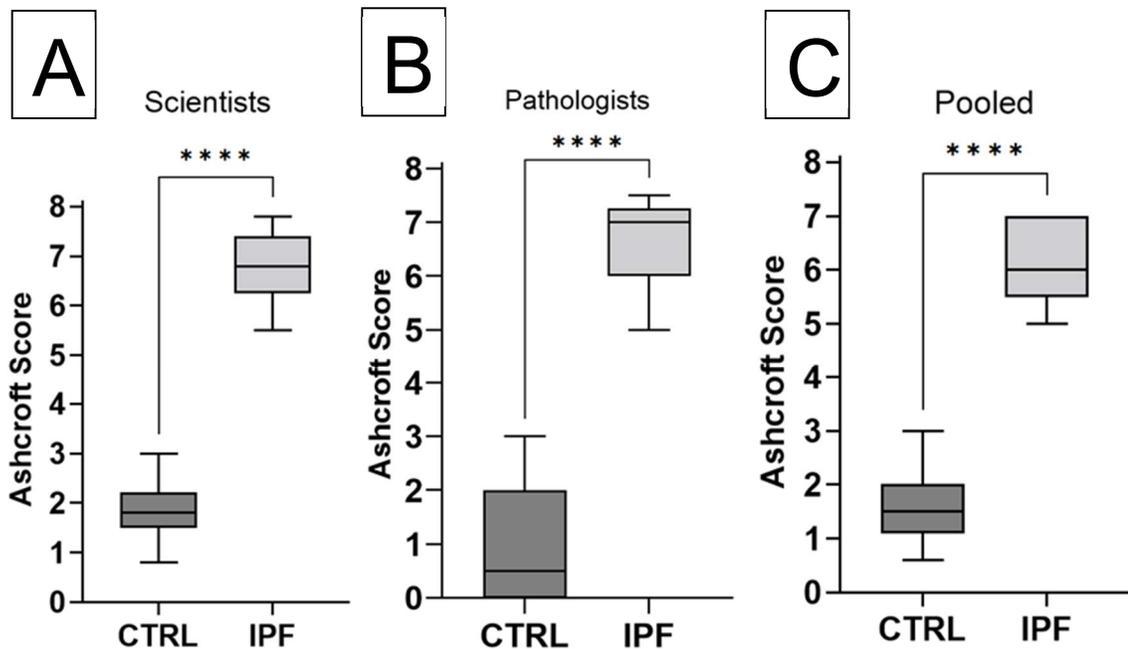


Figure 13 : TMA 2 Ashcroft

TMA 2 Ashcroft Scores (One-Tailed Wilcoxon Rank-Sum Test, $p < 0.0001$)

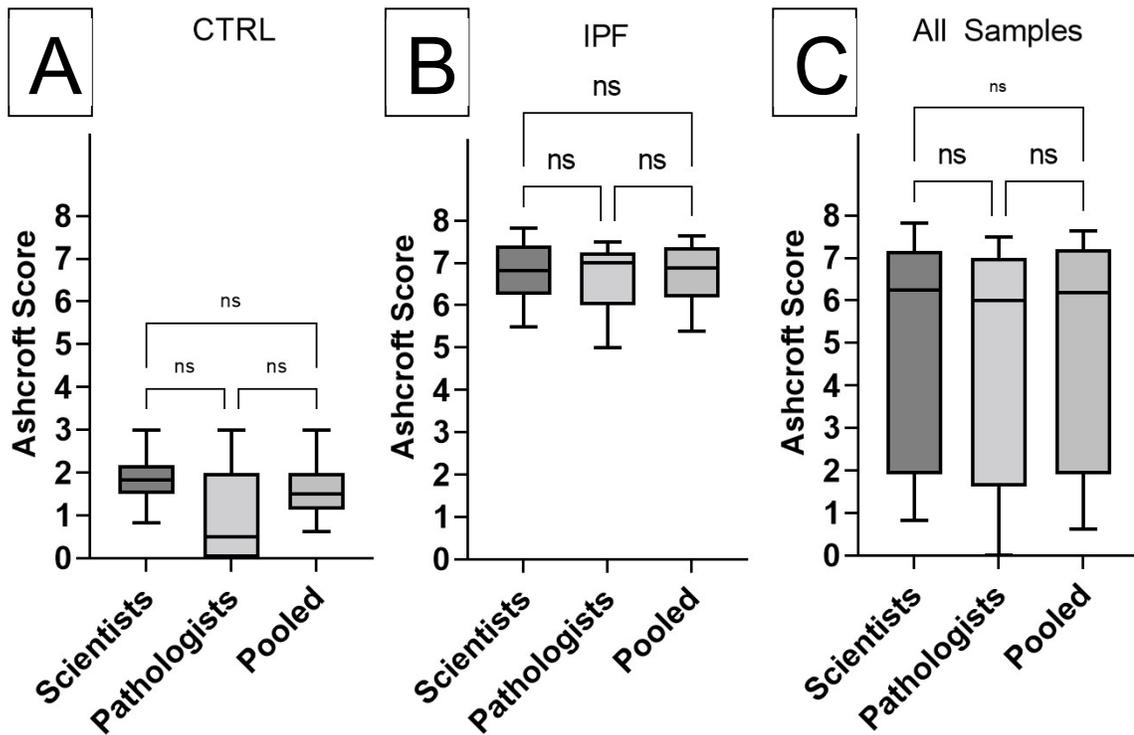


Figure 14 : TMA 2 Ashcroft Comparison

TMA 2 Ashcroft Score (Kruskal-Wallis Test, $p < 0.05$). No significant differences were found between the scientists and pathologists, nor their pooled averages, whether split into disease groups (A and B) or compiled together (C).

3.2 Develop quantitative H&E assessments

3.2.1 Cellularity, as defined by cellular density normalized to parenchyma area (mm^2), is elevated in IPF compared to CTRL cases.

To determine if cellularity might be predictive of fibrotic disease severity, the total number of cells was determined and normalized by the parenchyma area and again by the interstitium area. Only the former showed significant differences between the control and IPF tissues.

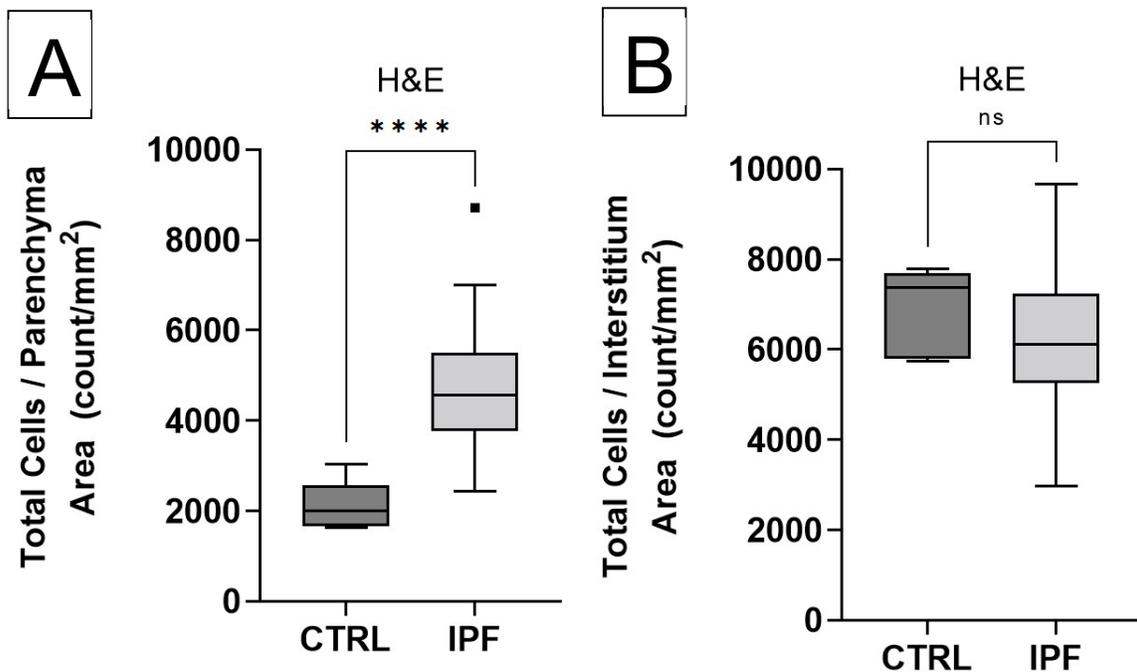


Figure 15 : H&E Cells

H&E Cellular Density (One-Tailed Wilcoxon Rank-Sum Test, $p < 0.0001$)

3.2.2 Airspace measured in the parenchyma is decreased in IPF compared to CTRL cases.

To determine if measurable airspace might be predictive of fibrotic disease severity, the area of the image comprised of white space was measured and normalized by the parenchyma area as a percentage and again by the interstitium area as a ratio. Both demonstrate highly significant differences between the control tissues and IPF tissues.

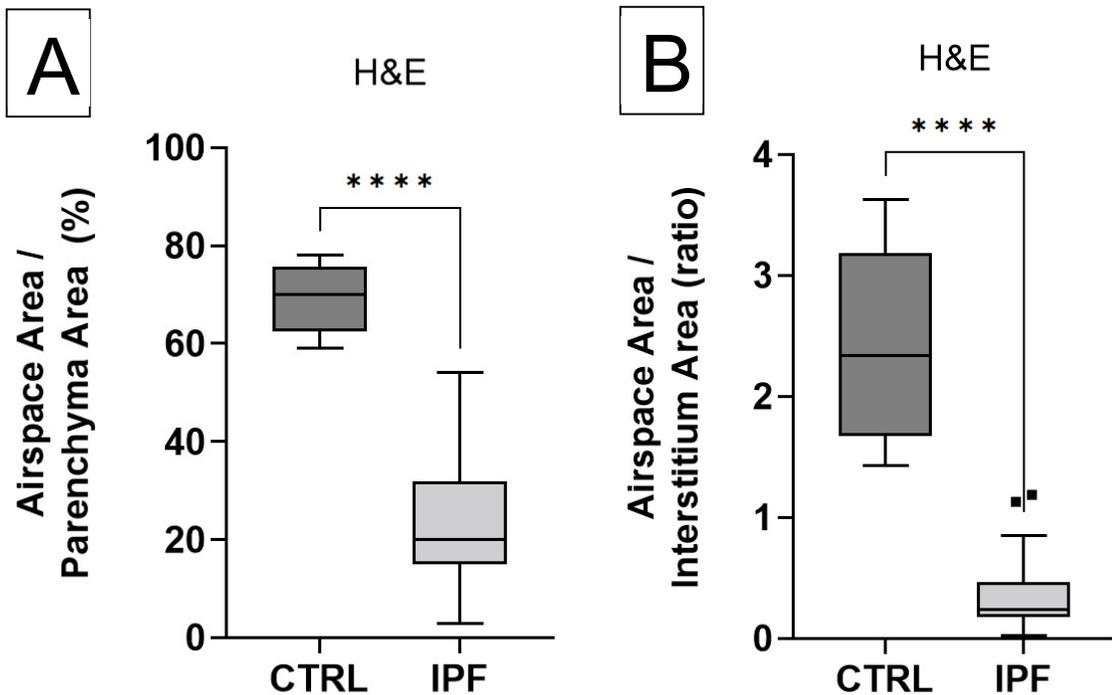


Figure 16 : H&E Airspaces

H&E Total Airspace mm² (One-Tailed Wilcoxon Rank-Sum Test, p<0.0001)

3.2.3 The properties of the airspaces in the parenchyma are decreased in IPF compared to CTRL cases.

The airspaces in the lung, in terms of their average area (A), diameter (C) and perimeter (D), though not count (B), after being normalized to the parenchyma area, are significantly different between the control tissues and IPF tissues.

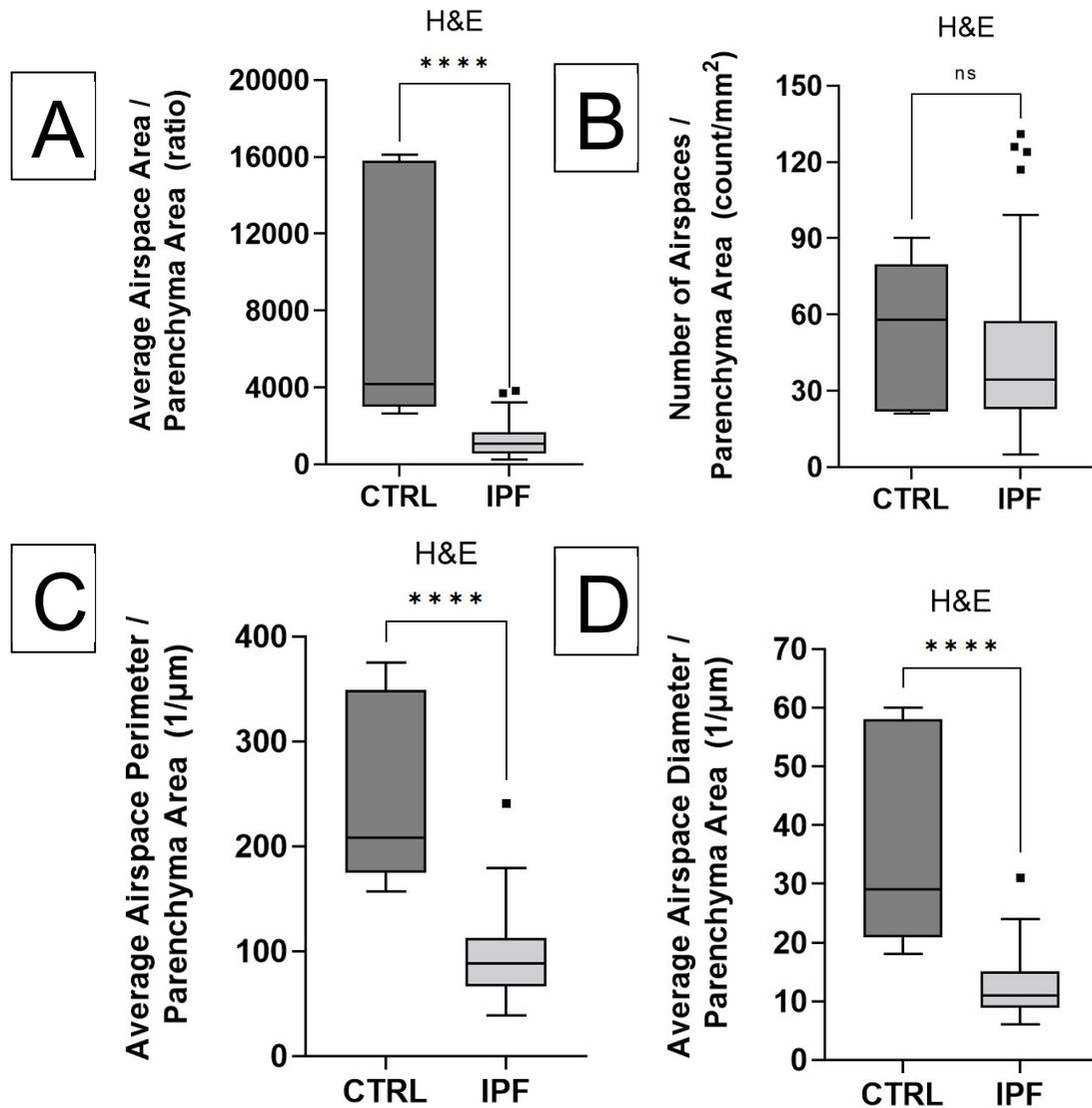


Figure 17 : H&E Airspace Properties

H&E Airspaces attributes (One-Tailed Wilcoxon Rank-Sum Test, $p < 0.0001$)

3.3 Compare TRI and PSR assessments

3.3.1 TRI positive area normalized to parenchyma area is elevated in IPF compared to CTRL cases.

The blue-stained collagenous space, referred to as TRI positive area, is elevated in IPF compared to control parenchyma ($p < 0.0001$), while the degree to which it was significant dropped when normalized with interstitium area ($p < 0.005$).

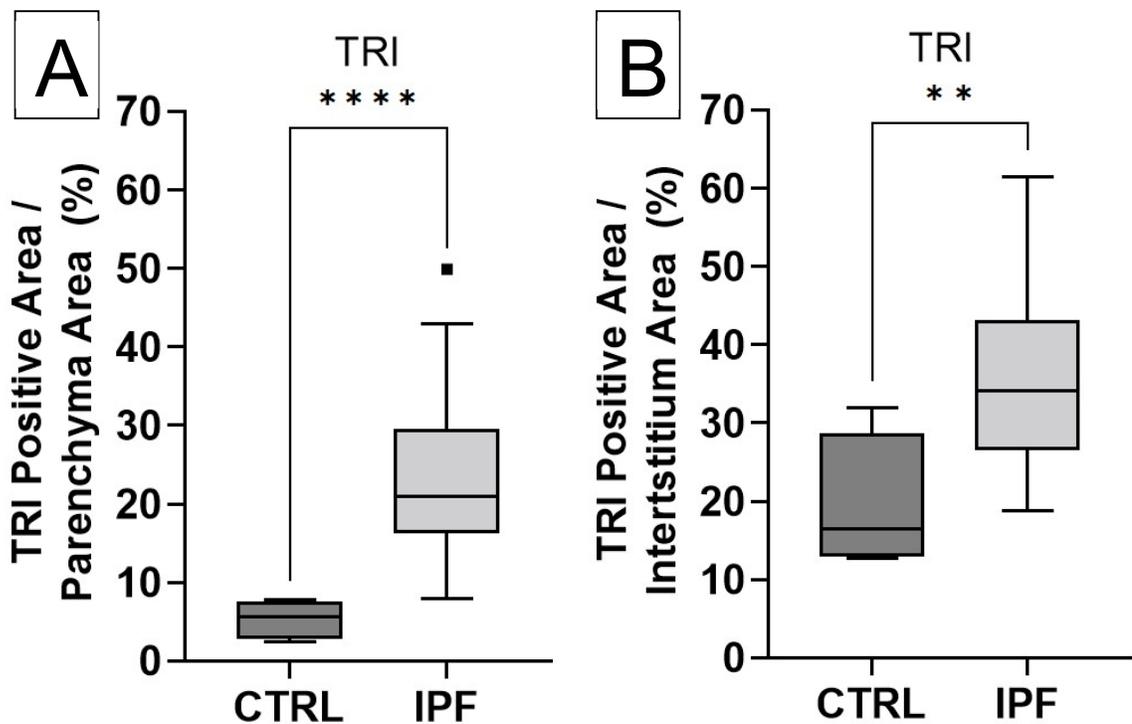


Figure 18 : TRI

Percent TRI positive area (One-Tailed Wilcoxon Rank-Sum Test, $p < 0.0001$)

3.3.2 PSR positive area normalized to parenchyma area is elevated in IPF compared to CTRL cases.

The brilliant white birefringent collagenous space measured in the polarized monochrome PSR slide, referred to as PSR positive area, is elevated in IPF parenchyma compared to controls ($p < 0.0001$). Only the former showed statistically significant differences.

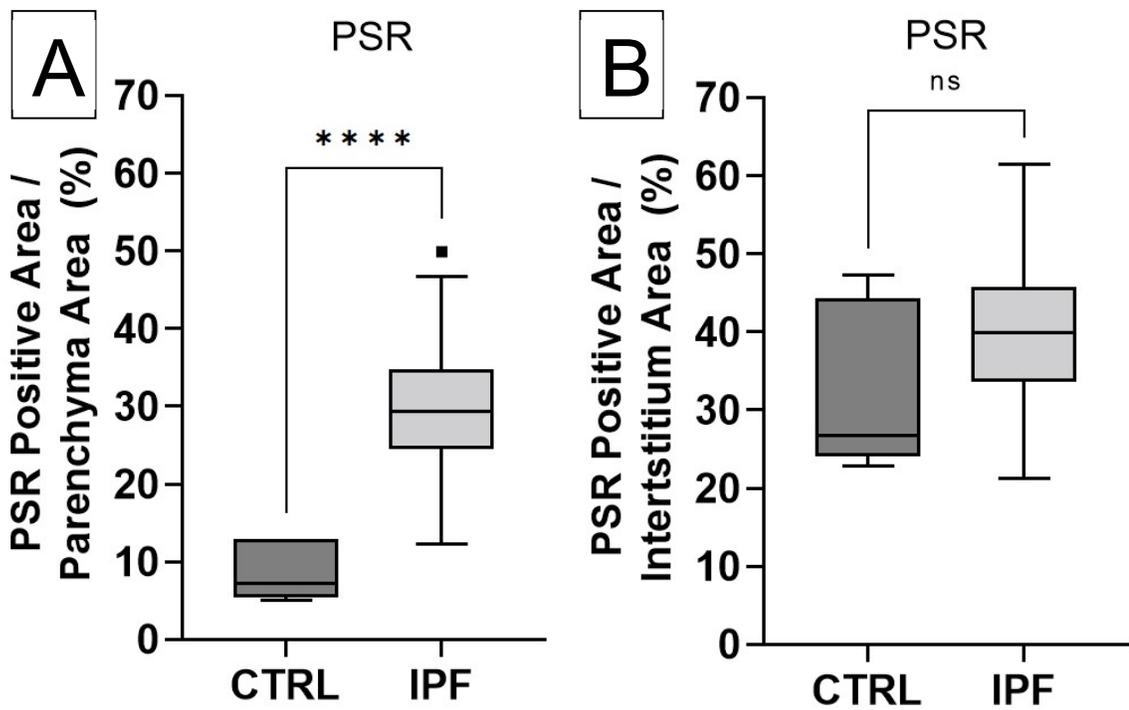


Figure 19 : PSR

Percent PSR positive area (One-Tailed Wilcoxon Rank-Sum Test, $p < 0.0001$)

3.4 Develop quantitative α SMA assessments

3.4.1 α SMA positive cell count normalized to parenchyma area is elevated in IPF compared to CTRL cases.

The alpha-smooth muscle actin (α SMA) positive cells in the parenchyma, minus the smooth muscle cells surrounding blood vessels as all vasculature was excluded from α SMA quantification, are elevated in IPF parenchyma compared to controls ($p < 0.0001$). This effect was reduced when normalized with interstitium area ($p < 0.005$).

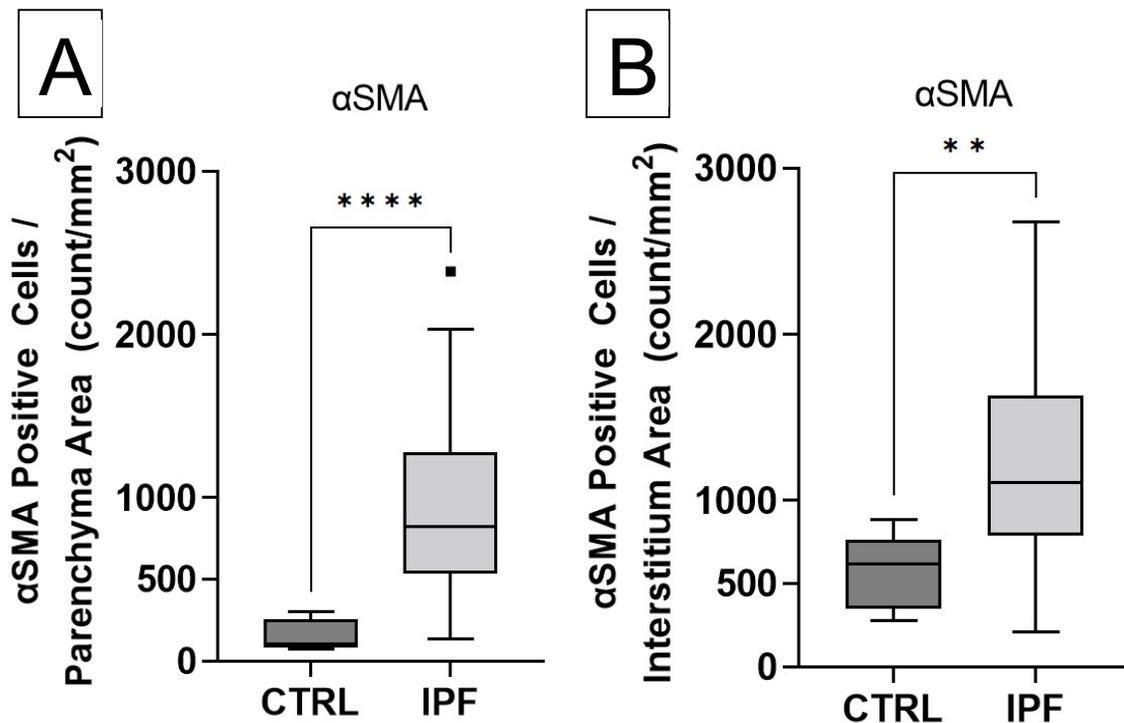


Figure 20 : α SMA Cells

α SMA positive cell density (One-Tailed Wilcoxon Rank-Sum Test, $p < 0.0001$)

3.4.2 α SMA positive area normalized to parenchyma area is elevated in IPF compared to CTRL cases.

Measuring α SMA positive area, the portion of the parenchyma binned by the HALO® Classifier (Indica Labs, 2016) module according to the various shades of brown, minus the area occupied by vasculature and their smooth muscle cell support structure and expressing it as a percentage of parenchyma area is elevated in IPF compared to controls ($p < 0.0001$) and less so when expressed as a percentage of interstitium area ($p < 0.0025$).

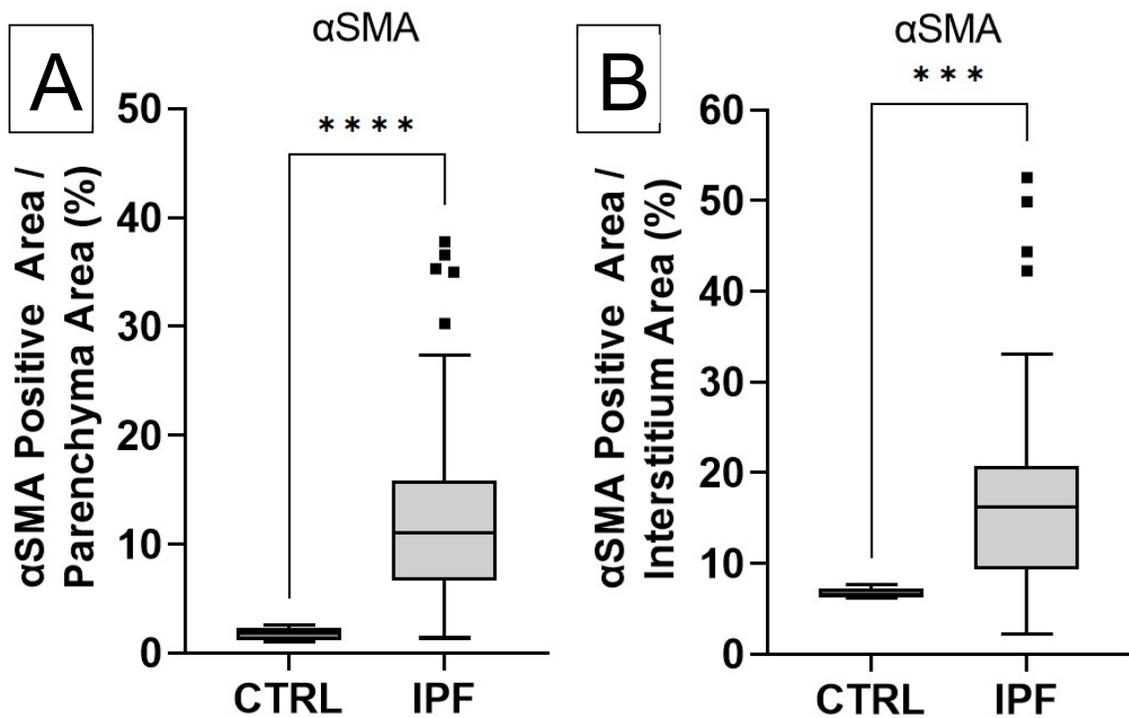


Figure 21 : α SMA Area

α SMA positive area (One-Tailed Wilcoxon Rank-Sum Test, $p < 0.0001$)

3.4.3 α SMA H-Score is elevated in IPF compared to CTRL cases.

Binning the α SMA positive cells into 3 categories of weak expressor, moderate expressor, and strong expressor, and calculating the H-Score using the formula demonstrates higher scores in the IPF cases compared to controls ($p < 0.0001$).

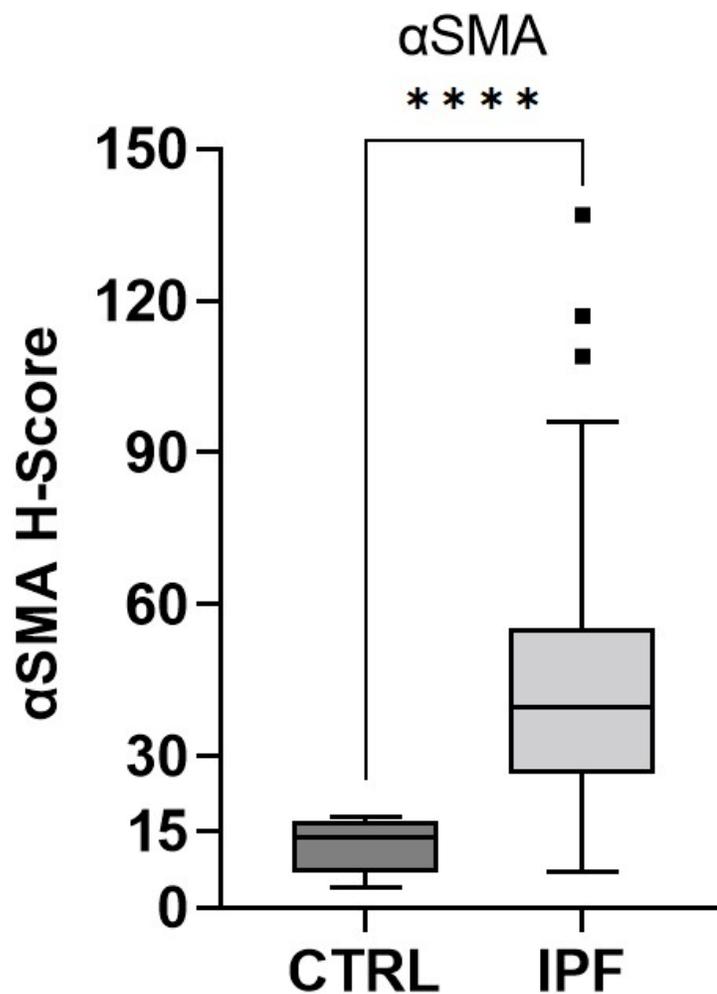


Figure 22 : α SMA H-Score

α SMA H-Score (One-Tailed Wilcoxon Rank-Sum Test, $p < 0.0001$)

3.5 Repeat IHC assessments on CD68, CD163, and CD206

3.5.1 CD163 and CD206 positive cell counts normalized to parenchyma area are highly elevated in IPF compared to CTRL cases.

CD163 and CD206 positive cell densities are both highly elevated in IPF compared to control parenchyma ($p < 0.0001$), while only CD163 remains highly correlated in the interstitium ($p < 0.00075$), CD206 does remain significantly different in the interstitium as well ($p < 0.05$). Finally, CD68 was found to be significantly different in parenchyma ($p < 0.02$), this slight difference was no longer seen in the interstitium (**Figures 21 and 22**).

3.5.2 CD163 positive area is highly correlated with Ashcroft score when normalized with parenchyma area.

Percent CD163 positive area is highly elevated in IPF parenchyma ($p < 0.0001$) compared to controls, and this effect is lessened in the interstitium ($p < 0.02$). CD206 is also found elevated in the parenchyma ($p < 0.005$), but not it nor CD206 are significantly elevated in the interstitium (**Figures 23 and 24**).

3.5.3 CD163 H-Score is highly correlated with Ashcroft score.

CD163 staining intensity as measured by H-Score is highly elevated in IPF cases compared to controls ($p < 0.0005$) as well as with CD206 ($p < 0.02$) (**Figure 25**).

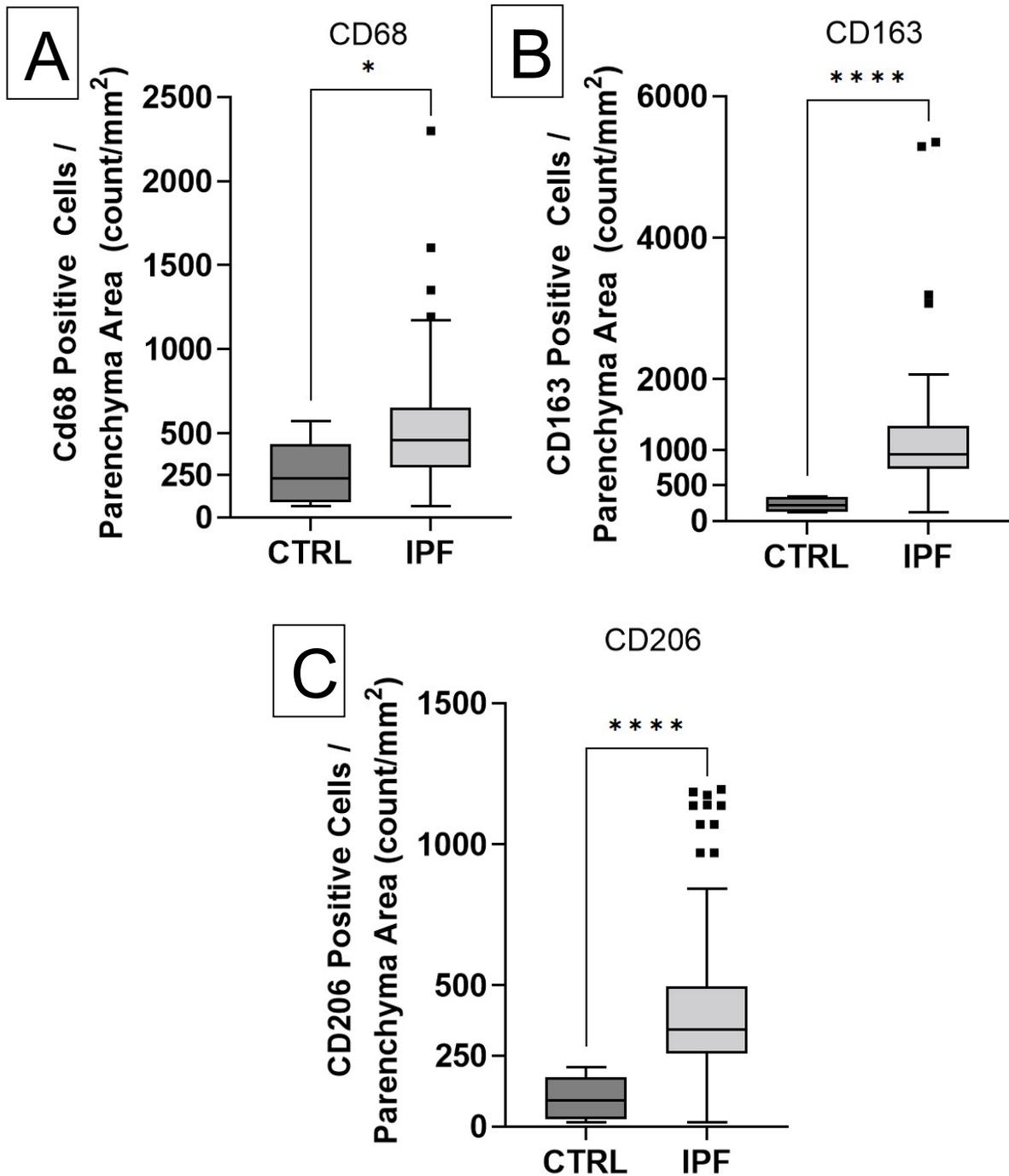


Figure 23 : Macrophage IHC Cells in Parenchyma

CD68, CD163, and CD206 positive cells (parenchyma)

(One-Tailed Wilcoxon Rank-Sum Test, * p<0.02, **** p<0.0001)

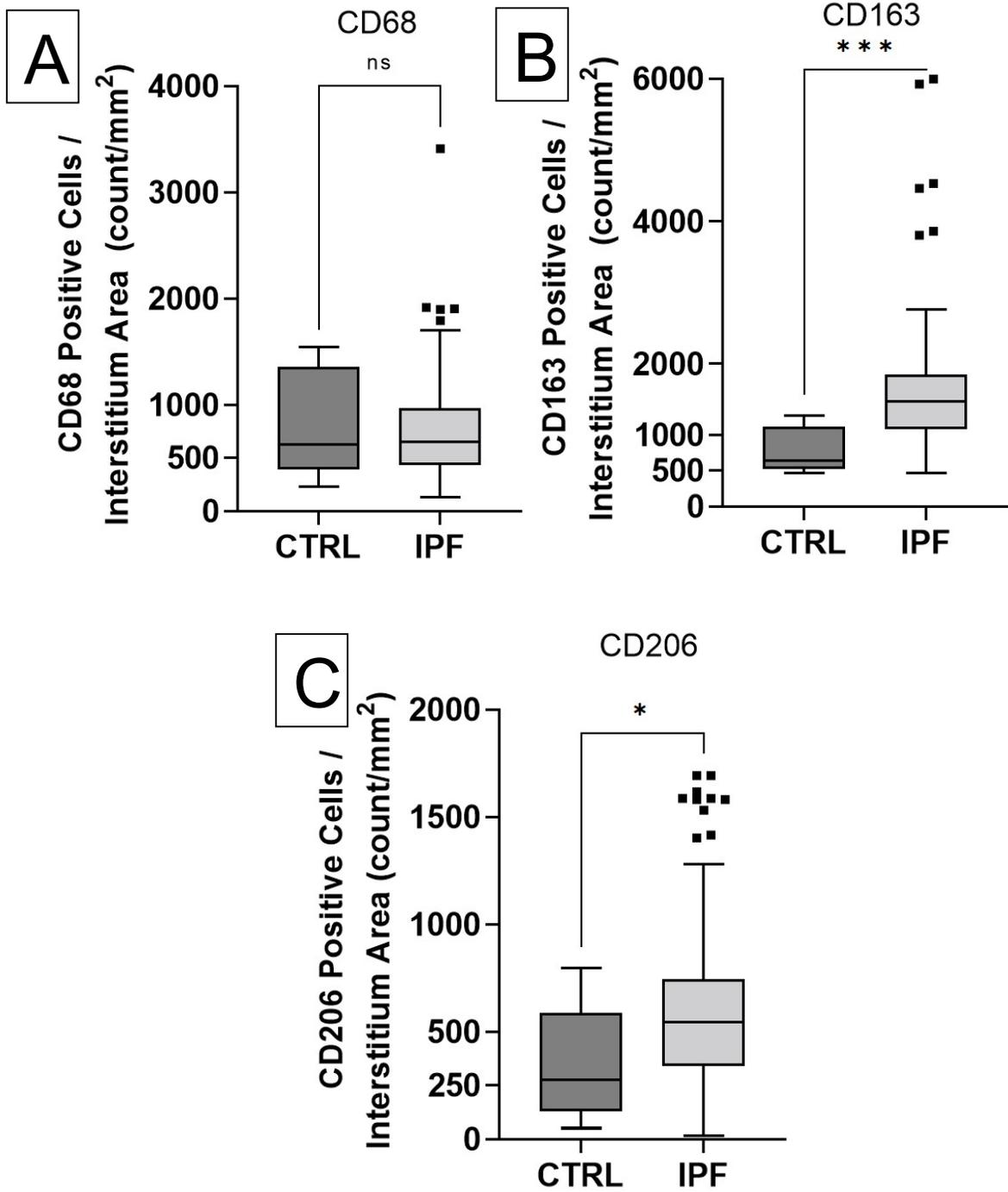


Figure 24 : Macrophage IHC Cells in Interstitium

CD68, CD163, and CD206 positive cells (interstitium)

(One-Tailed Wilcoxon Rank-Sum Test, * $p < 0.05$, *** $p < 0.00075$)

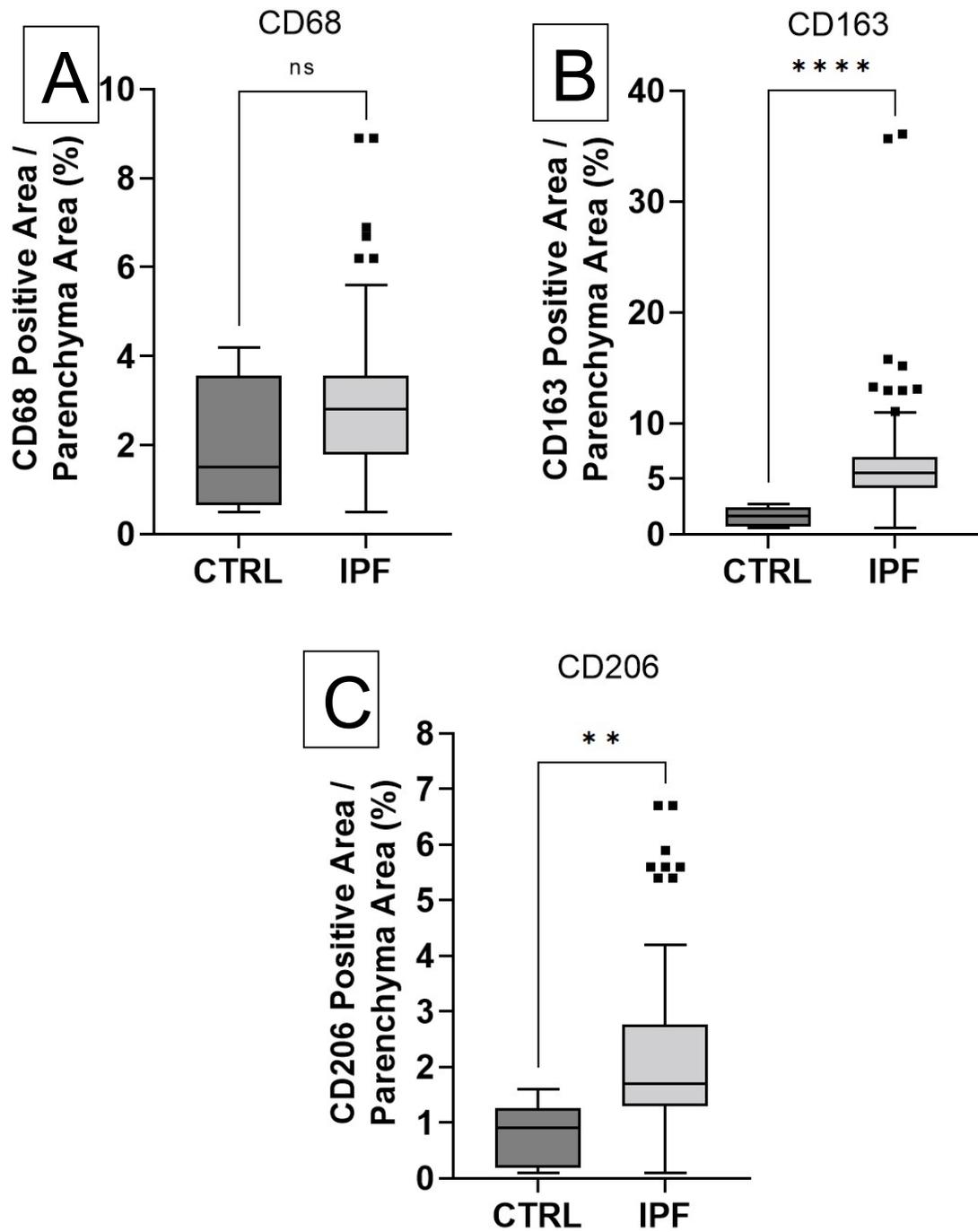


Figure 25 : Macrophage IHC Area in Parenchyma

CD68, CD163, and CD206 positive area (parenchyma)

(One-Tailed Wilcoxon Rank-Sum Test, ** $p < 0.005$ **** $p < 0.0001$)

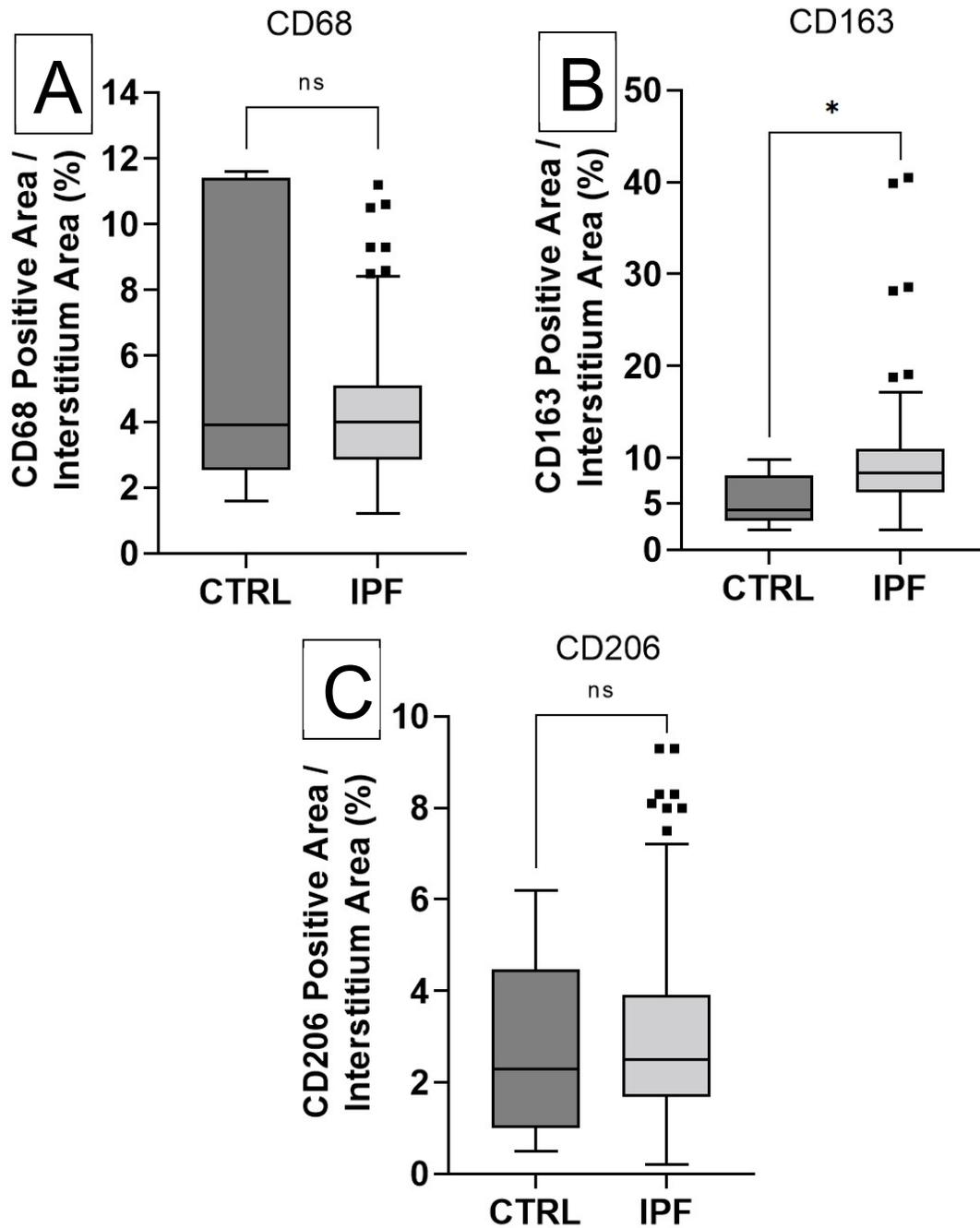


Figure 26 : Macrophage IHC Area in Interstitium

CD68, CD163, and CD206 positive area (interstitium)

(One-Tailed Wilcoxon Rank-Sum Test, * $p < 0.02$)

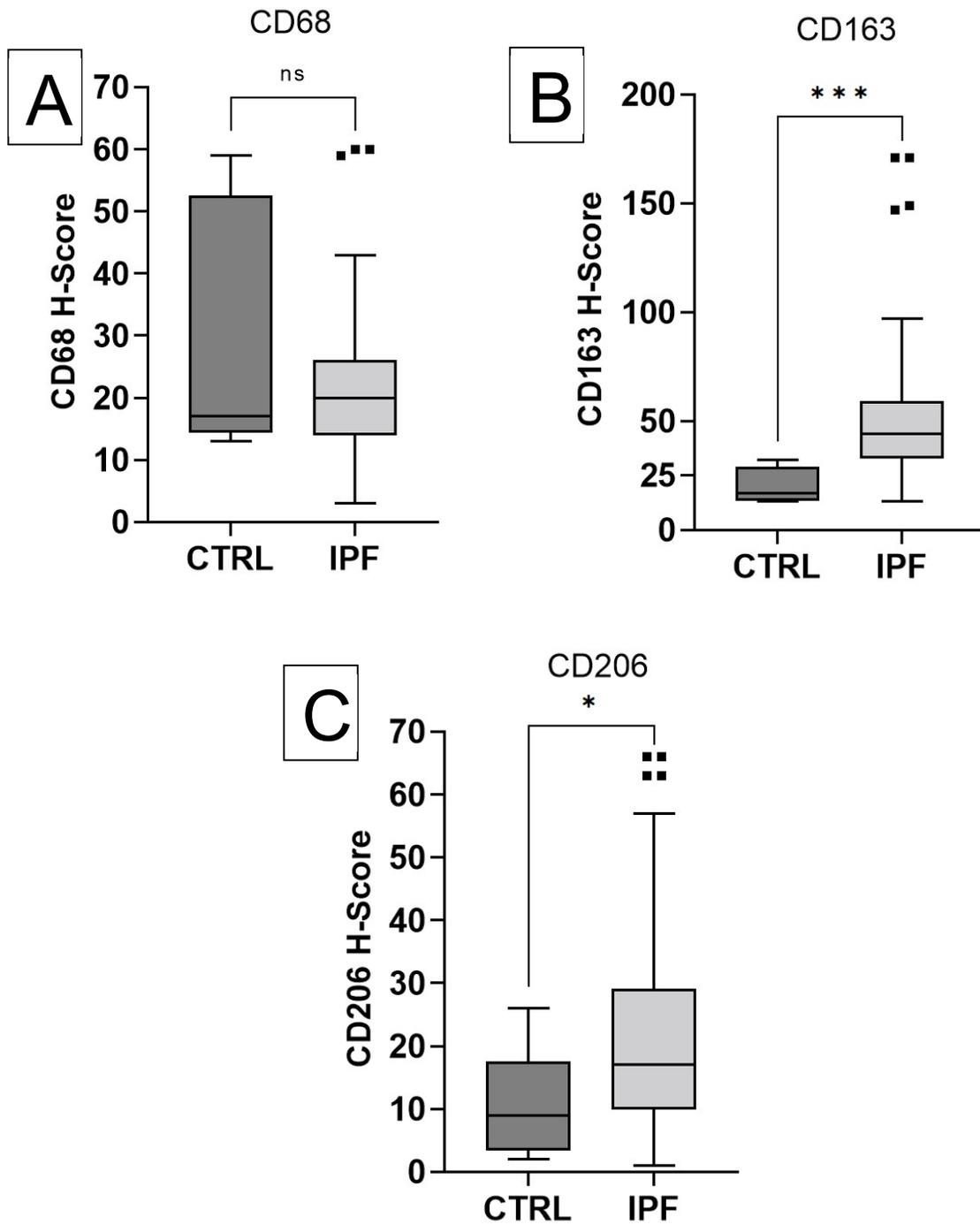


Figure 27 : Macrophage IHC H-Scores

CD68, CD163, and CD206 H-Score

(One-Tailed Wilcoxon Rank-Sum Test, * $p < 0.02$, *** $p < 0.0005$)

3.6 Create Statistical Models using the Ashcroft Score

3.6.1 Linear Regression Feature Selection Backward Elimination, No Restrictions

The first scenario posed no restrictions on which combination of markers to use to create the best model possible. This was generated using the data from the first TMA (training dataset) and has an adjusted $R^2 = 0.7333$ (**Table 1**).

$$\begin{aligned} \text{Askvill Score} = & \left(-1.725 * \frac{\text{total airspace area}}{\text{interstitium area}}\right) + \left(5.904 * \frac{\text{collagen area (PSR)}}{\text{parenchyma area}}\right) \\ & + \left(6.925 * \frac{\text{aSMA cells (count)}}{\text{parenchyma area}}\right) + \left(0.00036 * \frac{\text{H\&E cells (count)}}{\text{parenchyma area}}\right) \\ & + \left(-9.807 * \frac{\text{aSMA area}}{\text{parenchyma area}}\right) + \left(0.1403 * \frac{\text{avg airspace diameter}}{\text{parenchyma area}}\right) \\ & + \left(-0.0189 * \frac{\text{avg airspace perimeter}}{\text{parenchyma area}}\right) + 2.87266 \end{aligned}$$

3.6.2 Linear Regression Feature Selection Backward Elimination, a maximum of 4 Stains, one of TRI or PSR, but not both

For the second scenario, we wanted to know if we could generate a model using, at most, only a single collagenous stain and no more than four total stains. The same methodology was applied with these in mind, and the following model was derived with an adjusted $R^2 = 0.7018$ (**Table 2**).

$$\text{Askvill Score} = \left(-2.3822 * \frac{\text{total airspace area}}{\text{interstitium area}}\right) + \left(-0.0178 * \frac{\text{avg airspace perimeter}}{\text{parenchyma area}}\right) + 7.5604$$

Table 1 : MLR Model 1 Validation

A data table of the validation experiment which used the first multivariate linear regression model (without variable restrictions), comparing its quantitated scores (Model Prediction) against the pooled Ashcroft scores generated by the scientists and pathologists.

Multivariate Linear Regression MLR Model 1		Training : R ² = 0.7215 Validation : R ² = 0.9251
Validation Tissue	Ashcroft (Average)	Model Prediction
1	2.00	3.08
2	1.00	2.33
3	1.13	1.50
4	6.88	6.30
5	8.00	6.75
6	2.88	3.35
7	5.88	6.28
8	6.50	6.53
9	6.88	7.57
10	5.00	4.78
11	7.38	6.83
12	5.38	6.42
13	5.13	5.97
14	7.63	7.77
15	6.88	6.44
16	7.88	7.24
17	6.63	6.90
18	6.88	7.26
19	2.00	3.08
20	1.00	2.33

Table 2 : MLR Model 2 Validation

A data table of the validation experiment which used the second multivariate linear regression model (with variable restrictions), comparing its quantitated scores (Model Prediction) against the pooled Ashcroft scores generated by the scientists and pathologists.

Multivariate Linear Regression MLR Model 2		Training : $R^2 = 0.7018$ Validation : $R^2 = 0.9601$
Validation Tissue	Ashcroft (Average)	Model Prediction
1	2.00	3.12
2	1.00	1.90
3	1.13	1.84
4	6.88	6.15
5	8.00	6.74
6	2.88	4.07
7	5.88	6.13
8	6.50	5.61
9	6.88	6.33
10	5.00	5.20
11	7.38	7.02
12	5.38	5.32
13	5.13	5.58
14	7.63	7.28
15	6.88	6.52
16	7.88	7.43
17	6.63	6.68
18	6.88	6.37
19	2.00	3.12
20	1.00	1.90

3.6.3 eXtreme Gradient Boosting Feature Selection with Machine Learning, No Restrictions

The first model tested combinations of as many variables as it was provided, but we gave it the rule to use either TRI, or PSR, or neither. The training dataset was generated using a random 80% of the total cases from both TMAs ($R^2 = 0.9999$) and was validated using the remaining random 20% of the cases ($R^2 = 0.9190$) as shown in **Table 3**.

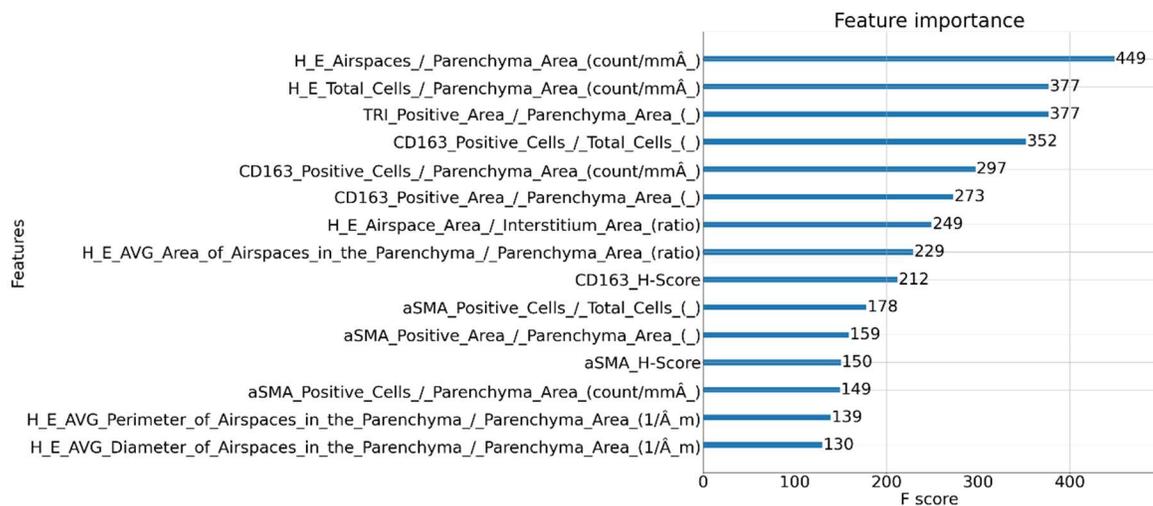


Figure 28 : XGB Model 1 F-Scores

The 15 variables across 4 stains utilized by the XGB model 1, sorted according to the number of times each variable was used (F-Score) in the decision-making process by any of XGB model 1's 500 decision trees (a variable may be considered more than once in a single tree, though this was not observed).

3.6.4 eXtreme Gradient Boosting Feature Selection with Machine Learning, a maximum of 4 Stains, one of TRI or PSR, but not both

The second model tested combinations of variables while prioritizing decision trees that accurately predicted the Ashcroft score while using as few variables as possible in conjunction with the rule to use either TRI, or PSR, or neither. The training dataset was the same 80% of the total cases from both TMAs ($R^2 = 0.9779$) as Section 3.6.3 and was also validated with the same 20% of cases ($R^2 = 0.9525$) as shown in **Table 4**.

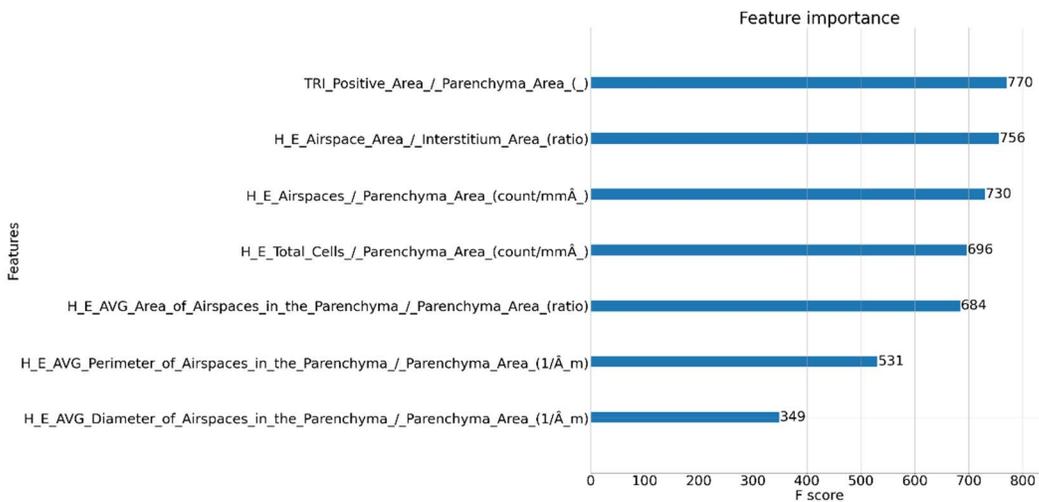


Figure 29 : XGB Model 2 F-Scores

The 7 variables across 2 stains utilized by the XGB model 2, sorted according to the number of times each variable was used (F-Score) in the decision-making process by any of XGB model 2's 500 decision trees (a variable may be considered more than once in a single tree, as was the case with every variable shown above except for the last one).

Table 3 : XGB Model 1 Validation

A data table of the validation experiment which used the first XGBoost model (without variable restrictions), comparing its quantitated scores (Model Prediction) against the pooled Ashcroft scores generated by the scientists and pathologists.

eXtreme Gradient Boosting XGBoost Model 1		Training : R ² = 0.9999 Validation : R ² = 0.9190
Random Test Case	Ashcroft (Average)	Model Prediction
0	7.88	6.85
1	8.00	7.18
2	6.88	6.94
3	7.38	6.92
4	6.88	6.87
5	6.50	6.81
6	6.88	6.88
7	6.88	6.83
8	6.63	6.35
9	2.00	2.04
10	7.63	6.77
11	5.13	5.08
12	2.88	2.17
13	5.38	7.07
14	5.88	5.98
15	5.00	3.63
16	1.13	1.08
17	1.00	0.64
18	7.88	6.85
19	8.00	7.18

Table 4 : XGB Model 2 Validation

A data table of the validation experiment which used the second XGBoost model (with variable restrictions), comparing its quantitated scores (Model Prediction) against the pooled Ashcroft scores generated by the scientists and pathologists.

eXtreme Gradient Boosting XGBoost Model 2		Training : R ² = 0.9779 Validation : R ² = 0.9525
Random Test Case	Ashcroft (Average)	Model Prediction
0	7.88	6.90
1	8.00	6.97
2	6.88	6.59
3	7.38	7.00
4	6.88	6.73
5	6.50	6.65
6	6.88	6.76
7	6.88	6.92
8	6.63	6.44
9	2.00	2.40
10	7.63	6.93
11	5.13	5.16
12	2.88	2.78
13	5.38	6.15
14	5.88	6.19
15	5.00	4.33
16	1.13	1.40
17	1.00	1.07
18	7.88	6.90
19	8.00	6.97

3.6.5 Askvill Score Implementation using the following Python script

```
# Importing Libraries

import pandas as pd
import joblib

# Getting your sample data
# Here you would use your own CSV file with your sample data, for this example we used the file
named: "askvill_sample_HE_TRI_aSMA_CD163.csv"

ValData=pd.read_csv('askvill_sample_HE_TRI_aSMA_CD163.csv',index_col=None, encoding =
'unicode_escape' )
ValData.columns = ValData.columns.str.replace('[#,@,&, ,%, ^, µ]', '_')
X_val, y_val = ValData.iloc[:,3:],ValData["Ashcroft_(Average)"]

# Loading askvill model from file and making predictions
# Here you would use your preferred model, for this example we used the model named
"Askvill_Score_HE_TRI_aSMA_CD163.sav" but you could also choose the other model

loaded_model = joblib.load("Askvill_Score_HE_TRI_aSMA_CD163.sav")
result = loaded_model.score(X_val, y_val)
preds = loaded_model.predict(X_val)

# Printing results

print("r2 score = ", result)
target = pd.DataFrame(y_val)
predictions = pd.DataFrame(preds, columns=['Model Prediction'])
table = pd.concat([target, predictions], axis=1)

table
```

Both XGBoost models from this project were uploaded to GitHub, labelled “Askvill_Score_HE_TRI_aSMA_CD163.sav” and “Askvill_Score_HE_TRI.sav” for models 1 and 2, respectively, along with example sample data files for each. The example python code shown on the previous page was uploaded and labelled “Implementation_Example.ipynb”. <https://github.com/azuhds/AskvillScoreModels>

To perform the Askvill Scoring method using one of the models and one’s own dataset, download the model you would like to use (.sav), as well as the associated sample dataset (.csv). Within the sample data, each row beyond the first, as those are the exact required titles for each column, can be completely replaced with your own data, and each row represents a single tissue ID. Your data can have more or fewer tissues IDs than what is provided in the sample file.

Open an integrated development environment (IDE) of your choice, such as a standalone program like Microsoft Visual Studio or an online resource like Google Collaboratory, provide the IDE with the model you have downloaded and the data file that you’ve just edited to contain your specific data. Provide the IDE with the python code (.ipynb) and edit the .csv file being read from “askvill_sample...csv” to the filename with your data, and the model loaded from “Askvill_Score...sav” to the name of the model you downloaded. Note, the model you chose may be the same as the one shown in the python code above and would thus not need to be changed.

Finally, run the program and read the Askvill scores for your tissues.

CHAPTER 4 : DISCUSSION

4.1 Results

There is a need in pulmonary research for a method of assessing the severity of fibrotic disease using a quantitative approach – a composite index – as opposed to the currently accepted semiquantitative scoring approach that is standard in the field (Ashcroft et al., 1988). While quantitative approaches do exist for physicians working directly with patients (KING et al., 2001; Ley et al., 2012; Watters et al., 1986) and pathologists assessing patient tissues (Cherniack et al., 1991), no such method is found in pulmonary research, though they are not absent from research altogether. Composite indices have been used in other areas of research where many endpoints can be measured and the combination of at least some of those endpoints yields a more robust conclusion than utilizing only one. Such composite indices can be found in fields of research such as genetics (Tanner et al., 2021), age and obesity (Schafer et al., 2019), environmental exposure (Si et al., 2016), biochemical pathways (Lin et al., 2008), inflammation (Sumariwalla et al., 2002), and lung cancer (Jones et al., 2011).

This project aims to create a quantitative scoring system using human tissues and employs the recruitment of molecular pathologists to provide their expertise in the foundation of the model.

A composite index will produce an overall score based on numerous independent variables but developing a system of weights for each of those

requires a dependent variable against which the system can be validated. We used the existing framework of the Ashcroft score as the basis for the models. Combining the expertise of pulmonary pathologists and the experience of pulmonary researchers, we scored 62 IPF tissue samples and 5 non-disease control tissue samples from the first TMA and found no significant differences between the average scores provided by researchers and the average scores provided by the pathologists (**Figures 11 and 12**), choosing to pool them together for use by the model. This was also done with the second TMA comprised of 12 IPF tissue and 8 non-disease controls as the validation cohort once the MLR models were completed, where no statistically significant differences were found between scoring groups (**Figures 13 and 14**).

Cellular density in the parenchyma of IPF tissues was found to be 1.6x to 3.3x greater than that of control tissues (**Figure 1**), while no significant difference was observed in the interstitium. This would suggest that although the number of cells in the parenchyma is increased in IPF cases, as the fraction of the parenchyma comprised by interstitium increases through ECM deposition and tissue architecture remodelling, the cellular density within the interstitium remains proportionate.

Alveolar airspaces, whole units of non-interstitial space (**Figure 4**), were measured in several manners. The proportion of the parenchyma occupied by airspace in control cases was decreased in IPF cases by 49% to 80%, based on the interquartile ranges of both groups (**Figure 16a**), while the ratio of alveolar

airspace to alveolar interstitium was decreased in IPF cases by 73% to 94% (**Figure 16b**). When assessing the airspaces as individual units, the average airspace cross-sectional area was decreased in IPF compared to control cases by 45% to 96% (**Figure 17a**), the average diameter of airspaces was decreased in IPF cases by 29% to 84% (**Figure 17d**), and the average airspace perimeter was decreased in IPF cases by 36% to 81% (**Figure 17c**). These data, in conjunction with the discovery that there was no statistically significant difference in the number of airspaces normalized to the parenchyma between IPF and control tissues (**Figure 17b**), suggest that tissue enlargement and loss of functional airspace seen in IPF can be measured quantitatively. These data also demonstrate the ranges that can be expected from these measurements.

Collagen and ECM deposition measured in the parenchyma of control cases using TRI was increased in IPF cases by 2.2 to 10.5 times (**Figure 18a**), while in the interstitium there was an increase in IPF cases by 0 to 3.3 times (**Figure 18b**). In contrast, the collagen measurements in the parenchyma of control cases using PSR was increased in IPF by 1.9 to 6.3 times (**Figure 19a**), while in the interstitium there was no statistically significant difference measured (**Figure 19b**). This suggests that TRI could be measuring more within the tissue than PSR. This difference may be due to PSR-POL-MONO only highlighting collagens I and III whereas TRI does not have this distinction. Consequently, with respect to the collagen in the parenchyma, both stains demonstrated similar increases of 3.7 and 4.1 times, respectively, between the median IPF and the

median control tissues, suggesting that either one could be used to measure fibrosis in the parenchyma. There is no apparent reason to use both stains simultaneously within a single experiment nor project since they both produce results that fulfil the same purpose.

For the α SMA analysis, all blood vessels were excluded thus removing any endothelial smooth muscle cells from the quantification. α SMA-positive cell density in the parenchyma of control cases was found elevated in IPF cases by 2.1 to 15.4 times (**Figure 20a**), while in the interstitium it was increased in IPF cases by 1.04 to 4.7 times (**Figure 20b**), and the overall staining intensity of those α SMA positive cells, as measured using the H-Score, was increased in IPF cases by 1.6 to 7.9 times. When measuring α SMA-positive area in the parenchyma of control cases, there was an increase in IPF cases by 2.9 to 12.7 times (**Figure 21a**), while in the interstitium there was an increase in IPF cases of 1.3 to 3.3 times. These data suggest that myofibroblasts and their footprint are elevated in IPF tissues compared to controls, which supports the evidence that IPF is characterized by the proliferation of myofibroblasts which are the active contributors of the unregulated collagen deposition (Korfei et al., 2022).

For the various macrophage markers, CD68, CD163, and CD206, the blood vessels were not excluded. This was decided after the fact because the results with and without those exclusions were not significantly different from one another, and not having to perform exclusions is both quicker and less work for the end user.

CD68-positive cell density in the parenchyma of control cases was increased in IPF cases by 0 to 6.9 times (**Figure 23a**). With respect to the cell density in the interstitium, the H-Score, as well as the CD68-positive area in both the parenchyma and interstitium, showed no statistically significant differences between the control cases and IPF cases (**Figures 24a-27a**). These suggest that the most measurable aspects of CD68 are poor predictors of fibrosis severity. This may be because CD68 is a marker for all macrophages without being specific for those that are polarized towards pro-inflammatory phenotype (M1-like) or those polarized towards anti-inflammatory, or pro-fibrotic, phenotype (M2-like) which are thought to play an important role in the progression of fibrotic lung diseases (Lee et al., 2021).

In contrast, CD163 and CD206-positive cell densities in the parenchyma of control cases were increased in IPF by 2.2 to 9.9 times (**Figure 23b**) and 1.5 to 17.4 times (Figure 23c), respectively, while in the interstitium they were increased in IPF by 0 to 3.5 times (Figure 24b) and 0 to 5.7 times (**Figure 24c**), respectively, and the H-Scores were increased in IPF by 1.1 to 4.4 times (**Figure 27b**) and 0 to 8.3 times (**Figure 27c**), respectively. For the CD163 and CD206 area quantifications in the parenchyma of control cases, they were increased in IPF cases by 1.8 to 9.3 times (**Figure 25b**) and 1.04 to 13.8 times (**Figure 25c**) respectively, while in the interstitium, CD163 was increased in IPF by 0 to 3.5 times (**Figure 26b**) but CD206 was not found to be significantly different (**Figure 26c**). These data show that some cases of IPF do not have much increase in

expression of CD206 at all compared to control tissues while other cases demonstrate an order of magnitude increase in expression. In comparison, CD163 is somewhat more consistently elevated in IPF cases with a lower ceiling with respect to the degree of expression. While both CD163 and CD206 are shown to be elevated in IPF cases compared to control cases, I suggest using CD163 as the default M2-like macrophage marker due to its improved reliability in terms of elevated expression in IPF cases.

The first MLR model developed was made without any intervention from the researchers' concern of practicality. The first model uses both TRI and PSR as they were both highly correlated with disease severity as defined by the Ashcroft score, which makes sense, as well as H&E, α SMA, and CD163. The second model was developed with specific requests in mind, including: (1) only requiring, at most, one of the collagen stains TRI or PSR, but not both, and (2) only requiring, at most, one of the macrophage markers CD68, CD163, or CD206, but no combination. This would reduce the overall cost and processing time for data acquisition on the researchers' part, with the understanding that there would likely be some loss in precision in exchange. For the XGBoost models, the first was given the TRI-or-PSR rule that the second MLR was given, while the second was given that and the 1-macrophage-marker rule as well.

The MLR model 1 and XGBoost model 1 recognized that the stains possessing the greatest predictive power were H&E, one (or both) collagen stain(s), α SMA and CD163, which mirrors what was observed from the HALO®

quantifications. With the addition of further limitations, the MLR model 2 and XGBoost model 2 narrowed those stains even further to only H&E and one collagen stain, MLR used PSR and XGBoost used TRI. This suggests that the models are successfully quantitatively identifying which aspects of the histology that people who perform Ashcroft scoring are recognizing visually.

To calculate the prediction accuracy, we contrasted the measured Ashcroft scores and the model prediction using the Pearson correlation coefficient (R^2), first with the training dataset used to build the model, which is expected to be high, then verify the model's accuracy using a validation dataset. The first model's predictions of the training and validation data were 0.73 and 0.93, respectively, and the second model's predictions of the training and validation data were 0.70 and 0.96, respectively. Both models performed similarly with both the training datasets and validation datasets, though the predictive variables used in the latter model comprised less than half that of the first. Simultaneously, both validation runs performed significantly better than training runs, suggesting that the training dataset may have several outliers, or the validation dataset was simply coincidentally very well predicted. Other datasets would likely perform closer to 0.7. This was also done for the XGBoost models: the first model's predictions of the training and validation data were 0.99 and 0.92, respectively, and the second model's predictions of the training and validation data were 0.98 and 0.95, respectively. The implication of the R^2 of 0.99 is that the model became too specific for the dataset that it was given as training, and this is reflected by

the lower R^2 of 0.92 seen from the validation set, though it is still very highly correlated. As for the second model, it possessed a lower training R^2 of 0.98, yet the validation R^2 was only 0.95, 0.03 points lower, suggesting that this model is the one that should be chosen as the composite index for scoring severity of pulmonary fibrosis.

4.2 Limitations

This project is not without its flaws. As with all histological stains, proper assessment entirely depends on the quality of the stain. Even a poor H&E stain can lead to variability and misinterpretations. As such, validated and reproducible stains are required. An advantage of developing a composite index based on fewer stains, as is the case with both the second MLR and XGBoost models, is that it reduces the chance of variability.

One example is regarding using TRI as a marker for collagen deposition in the lung, one author noted that in some TRI protocols, the stain used to differentiate the collagen fibers is accentuated and may inadvertently give some muscle and cytoplasm a blueish hue (Llewellyn, 2019). If this is the case, then it would likely be detected by the HALO® Classifier (Indica Labs, 2016) module that categorizes blue pixels as collagen, as this was the definition provided by the TRI stain protocol. This, however, could be remedied by increasing the minimum threshold for blue pixel detection to ensure that the weaker staining intensities of any muscle or cytoplasm would not be detected.

Another similar limitation is with the determination of IHC-positive cells using the Multiplex IHC (Indica Labs, 2017) module. The classification required for dividing the cells identified by the software into the 4 categories of IHC-negative (0), weak expressor (+1), moderate expressor (+2), and strong expressor (+3) requires careful attention. The two thresholds that separate weak from moderate and moderate from strong expressors impact the H-Score for that IHC stain, which is a clinically used parameter in some cases. As such, learning the difference is critical, and best done with the accompaniment of a trained pathologist. However, what is arguably more important is properly setting the threshold that separates IHC-negative cells from weak expressors, as this will directly impact the overall number of IHC-positive cells. Some minor brown stain surrounding or simply near a cell is not sufficient for it to be considered positive, according to teachings of the pathologists. Some examples of this are shown in Figure 7, specifically the cells labelled “0”. Setting the threshold too low would cause an increase in cells adjacent to even slight IHC stain to be considered IHC-positive cells, thus artificially inflating the overall cellular positivity. The opposite is also true, where setting the threshold too high would artificially deflate the overall cellular positivity.

Another aspect of creating a composite index using these tissues is that such analyses can be limiting when sample sizes are small. These types of models benefit greatly by having hundreds of samples providing equal proportions of all possible severities.

4.3 Future Directions

This model could be expanded upon to include other likely relevant components of IPF but would require further testing to first determine what measurements could be extracted from any additional stain or already included stains, followed by assessing whether these variables are correlated, or inversely correlated, with disease severity. Another direction would be to attempt to generate a fibrotic score that would be superior to the Ashcroft score. We believe that the identification of histological features that is associated with fibrosis, could be added to generate a fibrotic index. The works in this thesis have identified the following fibrotic features in the parenchyma described above, including:

1. increased cellularity (Figure 15)
2. reduced airspace size (Figure 16 to 17)
3. increased collagen (Figures 18 to 19)
4. increased myofibroblasts (Figures 20 to 22)
5. increased CD163 positive macrophages (Figures 23b to 27b)

In addition to these five main fibrotic features, we also observe the following phenotypes, often associated within the UIP/IPF lung:

4.3.1 Quantification of Vasculature in the Alveolar Interstitium

We and others (C. Wang & Yang, 2022) have observed that in the fibrotic parenchyma there is also an increase in the number of vessels. We believe that these vessels should be quantified in the same or a similar manner as the alveolar airspaces, such as by repurposing the Vacuole module (Indica Labs, 2018), and considered a fibrotic feature within a new composite index.

One publication on sheep fibrosis measured vascular density as a percentage of the total area occupied by CD34 and collagen type IV positive endothelial cells per total area of parenchyma. This approach could be replicated using the Multiplex IHC module (Indica Labs, 2017) to quantify cells and their positivity for markers CD34 and collagen type IV (Derseh et al., 2021)

4.3.2 Quantification of Mucin Secretions in Alveolar Airspace

Another component of lung histology that was evident was the presence of mucin secretions in the alveolar airspace. Confirming with two pulmonary pathologists on the nature of these secretions, they could be present in response to an infection or due to allergies, both of which would fall outside the scope of fibrotic lung disease, or perhaps it's directly related to IPF. This should be further elucidated with the stains we currently possess, and perhaps with a marker for mucin to specifically target it for quantification to see what, if any, relationship its presence in the airways has with IPF severity. It would be logical to conclude that alveoli filled with proteinaceous secretions would be unable to perform gas

exchange, though they would not be considered part of the interstitium. Those cores that were predominantly mucin-filled were omitted from our analyses, but with more information, perhaps they should be retained.

4.3.3 Quantification of Adipocytes in the Alveolar Interstitium

Adipocytes in the lung are an abnormal phenomenon. Fat cells are easily distinguished from the rest of the lung since they collect in bundles and possess separations much thinner and string-like compared to even the smallest alveoli. Their apparent non-existence in our control tissues but infrequent presence in IPF tissues indicate that adipose tissues would be a great target for study. They may be due to obesity in the patient (European Lung Foundation, 2019), but they could also be related to disease severity and add a unique insight to the disease.

4.3.4 Quantification of Honeycombing of the Alveolar Interstitium

Honeycombing is characterized by airspace enlargement combined with interstitial thickening comprised of bronchiolar epithelium and often filled with mucin and inflammatory cells. This phenotype is a well-known fibrotic feature of lung fibrosis, specifically UIP (Bell, D., 2021). Determining a robust methodology for the quantification of this characteristic would likely be beneficial.

Based on this work, we propose that a future index could be developed containing the 9 main fibrotic features discussed above. These should be quantified and included in a summative fibrotic index. Therapeutic treatments that

would affect these features in a positive way could be considered for future therapeutic developments.

4.3.5 Larger Training and Testing Datasets

As mentioned, machine learning models improve with increased sizes of training datasets. If more were collected, the analyses performed during this project could always be re-run in the future, by combining the previously used datasets in with any new datasets amassed over time, thus increasing the sample sizes, and assisting the machine learning by providing a greater variety with which to create a potentially better model.

4.3.6 Translation to Pre-Clinical Models

Each model generated from this project was made using measurements on human tissues. In their current states, they would more than likely fail to accurately score murine lung tissues since parameters like average airway diameter would be much smaller in mice than humans. This project could be repeated in its entirety using the same methodology with a strict focus on murine cell populations and tissues. Ideally, the variables that were deemed predictive in the human models would also be found predictive in mice.

CHAPTER 5 : CONCLUSION

This project has demonstrated a novel approach for assessing the severity of fibrotic lung disease using lung histology. Our overall hypothesis was to quantify various components of several varieties of digitized lung histology to generate a method of scoring lung fibrosis in a manner that was more objective than the current visual semiquantitative standard.

This was accomplished by assessing H&E for lung cellularity and airspace properties, TRI and PSR for collagen content of the parenchyma, α SMA for myofibroblast populations, and CD68, CD163, or CD206 for macrophage populations. Prior to developing a composite index, we found several quantifiable fibrotic phenotypes in the parenchyma including (1) increased cellularity, (2) reduced airspace size, (3) increased collagen content, (4) increased myofibroblast and (5) CD163-positive macrophage populations.

Several composite indices were developed using multivariate linear regression models and separately with XGBoost machine learning based on the semiquantitative Ashcroft scoring methodology. These indices generate a fibrotic severity score, an Askvill score, from 0 to 8, akin to the Ashcroft score, but the methods used to acquire these scores utilize objective quantifications derived from histological samples using HALO® software instead of biased visual assessments.

Through testing and validating the histological stains across 73 IPF and 12 control lung tissues, we confirmed that fibrosis severity is highly inversely correlated with the proportion of the parenchyma occupied by airspace, as well as the average area, diameter, and perimeter of the airspaces in the parenchyma. We also confirmed that fibrosis severity is highly correlated with overall cellular density within the parenchyma and the proportion of the parenchyma occupied by collagen. We further confirmed that it is highly correlated with the cellular density, H-Score, and proportion of the parenchyma occupied by both α SMA-positive myofibroblasts and CD163-positive macrophages, and to a lesser extent, the cellular density of CD206-positive macrophages in the parenchyma.

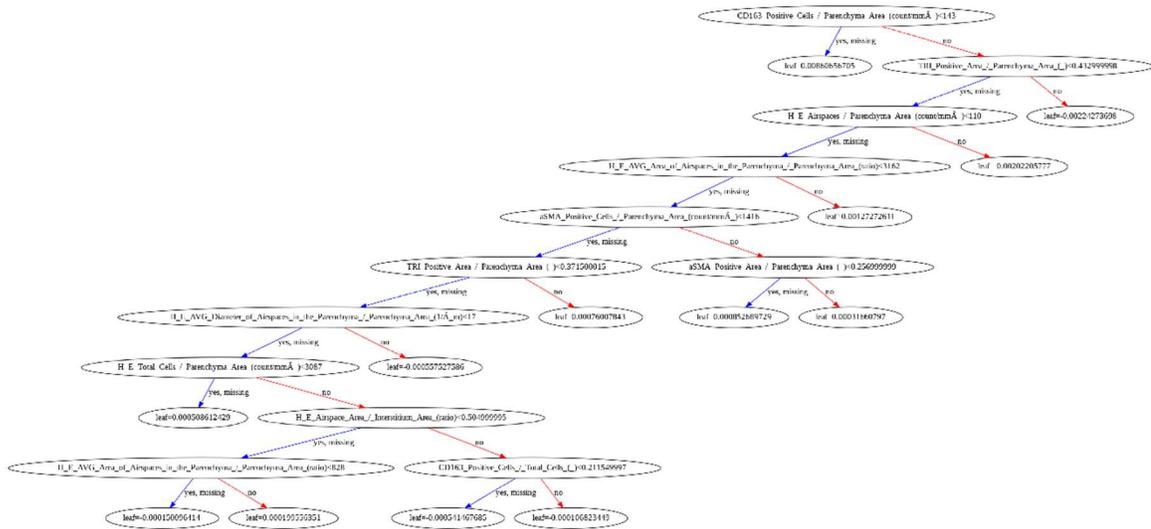
This thesis work has demonstrated that a multivariate linear regression model, MLR model 2, is capable of reproducing Ashcroft scores produced by several scientists and pathologists using quantitative measurements of alveolar airspaces from digitized lung H&E, possessing a significant training accuracy of 70.18% and a highly significant validation accuracy of 96.01%. Furthermore, it has been demonstrated that a machine learning model, XGBoost model 2, is also capable using quantitative measurements of alveolar airspaces and cellularity from digitized lung H&E and collagen content from digitized lung Trichrome, possessing a highly significant training accuracy of 97.79% and a highly significant validation accuracy of 95.25%. The latter has been uploaded to GitHub (<https://github.com/azuhds/AskvillScoreModels>).

CHAPTER 6 : SUPPLEMENTARY FIGURES

<i>Grade of fibrosis</i>	<i>Histological features</i>
0	Normal lung
1	Minimal fibrous thickening of alveolar or bronchiolar walls
2*	
3	Moderate thickening of walls without obvious damage to lung architecture
4*	
5	Increased fibrosis with definite damage to lung structure and formation of fibrous bands or small fibrous masses
6*	
7	Severe distortion of structure and large fibrous areas; "honeycomb lung" is placed in this category
8	Total fibrous obliteration of the field

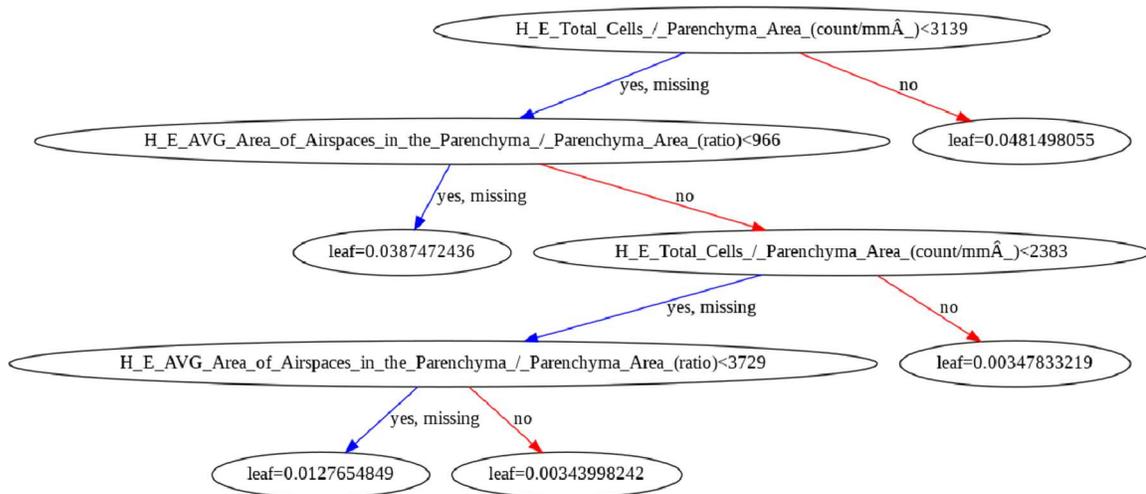
Supplementary Figure 1 : Ashcroft Score Criteria

This is the Ashcroft scoring criteria outlined in Table 1 from the original publication (Ashcroft et al., 1988) that was used by all those who scored tissues for this thesis.



Supplementary Figure 2 : XGBoost Model 1

One of the 500 decision trees performed by the model. The result from this tree is pooled with the other 499 trees to generate the final score predicted by the model. All decision trees start at the top, and this specific tree starts with analyzing the result for `CD133 Positive Cells / Parenchyma Area (countnorm)` and compares it to the decision node: is the value greater than or smaller than 143? If it's lesser, than this tree is done and the answer that will be pooled with the other trees is 0.0086. If it's greater, then it moves on to the next node where it checks the result for `H E Total Cells / Parenchyma Area (countnorm)` and moves either left or right down the rest of the tree depending on whether it's greater or smaller than 43.29%. It stops once it reaches a terminal leaf.



Supplementary Figure 3 : XGBoost Model 2

One of the 500 decision trees performed by the model. The result from this tree is pooled with the other 499 trees to generate the final score predicted by the model. The logic is the same as Supplementary Figure 1, however this branching pathway is much smaller due to the prioritization given this model to using as few variables as possible while retaining an accurate prediction power.

CHAPTER 7 : REFERENCES

- Ashcroft, T., Simpson, J. M., & Timbrell, V. (1988). Simple method of estimating severity of pulmonary fibrosis on a numerical scale. *Journal of Clinical Pathology*, 41(4), 467–470. <https://doi.org/10.1136/jcp.41.4.467>
- B. Moore, B., Lawson, W. E., Oury, T. D., Sisson, T. H., Raghavendran, K., & Hogaboam, C. M. (2013). Animal Models of Fibrotic Lung Disease. *American Journal of Respiratory Cell and Molecular Biology*, 49(2), 167–179. <https://doi.org/10.1165/rcmb.2013-0094TR>
- Bell, D. (2021) Honeycombing (lungs). *Radiopaedia.org*. Radiopaedia.org. <https://doi.org/10.53347/rID-1462>
- Bickle, I., & Weerakkody, Y. (2012). Idiopathic pulmonary fibrosis. In *Radiopaedia.org*. Radiopaedia.org. <https://doi.org/10.53347/rID-16929>
- Calabrese, F., Lunardi, F., Tauro, V., Pezzuto, F., Fortarezza, F., Vedovelli, L., Faccioli, E., Balestro, E., Schiavon, M., Esposito, G., Vuljan, S. E., Giraud, C., Gregori, D., Rea, F., & Spagnolo, P. (2022). RNA Sequencing of Epithelial Cell/Fibroblastic Foci Sandwich in Idiopathic Pulmonary Fibrosis: New Insights on the Signaling Pathway. *International Journal of Molecular Sciences*, 23(6), 3323. <https://doi.org/10.3390/ijms23063323>
- Cao, Y., Rudrakshala, J., Williams, R., Rodriguez, S., Sorkhdini, P., Yang, A. X., Mundy, M., Yang, D., Palmisciano, A., Walsh, T., Delcompare, C., Caine, T., Tomasi, L., Shea, B. S., & Zhou, Y. (2022). CRTH2 Mediates Pro-fibrotic Macrophage Differentiation and Promotes Lung Fibrosis. *American Journal of Respiratory Cell and Molecular Biology*. <https://doi.org/10.1165/rcmb.2021-0504OC>
- Cherniack, R. M., Colby, T. v., Flint, A., Thurlbeck, W. M., Waldron, J., Ackerson, L., & King, T. E. (1991). Quantitative Assessment of Lung Pathology in Idiopathic Pulmonary Fibrosis. *American Review of Respiratory Disease*, 144(4), 892–900. <https://doi.org/10.1164/ajrccm/144.4.892>
- Derseh, H. B., Perera, K. U. E., Dewage, S. N. V., Stent, A., Koumoundouros, E., Organ, L., Pagel, C. N., & Snibson, K. J. (2021). Tetrathiomolybdate Treatment Attenuates Bleomycin-Induced Angiogenesis and Lung Pathology in a Sheep Model of Pulmonary

Fibrosis. *Frontiers in Pharmacology*, 12.
<https://doi.org/10.3389/fphar.2021.700902>

European Lung Foundation. (2019, October 21). Fat Accumulates Inside Lungs of Overweight People – May Cause Asthma Symptoms. *Independent*, 2. <https://www.independent.co.uk/news/science/fat-lungs-overweight-asthma-wheezing-obesity-a9159846.html>

Fabrick, B. O., Dijkstra, C. D., & van den Berg, T. K. (2005). The macrophage scavenger receptor CD163. *Immunobiology*, 210(2–4), 153–160. <https://doi.org/10.1016/j.imbio.2005.05.010>

Glass, D. S., Grossfeld, D., Renna, H. A., Agarwala, P., Spiegler, P., Kasselmann, L. J., Glass, A. D., DeLeon, J., & Reiss, A. B. (2020). Idiopathic pulmonary fibrosis: Molecular mechanisms and potential treatment approaches. *Respiratory Investigation*, 58(5), 320–335. <https://doi.org/10.1016/j.resinv.2020.04.002>

Humphries, D. C., Mills, R., Dobie, R., Henderson, N. C., Sethi, T., & Mackinnon, A. C. (2021). Selective Myeloid Depletion of Galectin-3 Offers Protection Against Acute and Chronic Lung Injury. *Frontiers in Pharmacology*, 12. <https://doi.org/10.3389/fphar.2021.715986>

Indica labs. (2014). *Area Quantification Brightfield*.
<https://indicalab.com/products/area-quantification-brightfield/>

Indica Labs. (2014). *Area Quantification FL*.
<https://indicalab.com/products/area-quantification-fl/>

Indica Labs. (2016). *Tissue Classifier Add-On*.
<https://indicalab.com/products/tissue-classification/>

Indica Labs. (2017). *Multiplex IHC*. <https://indicalab.com/products/multiplex-ihc/>

Indica Labs. (2018). *Vacuole Module*.

Indica Labs. (2019). *TMA (Tissue Microarray) Add-On*.
<https://indicalab.com/products/tissue-microarray-tma/>

Jones, L. W., Eves, N. D., Spasojevic, I., Wang, F., & Il'yasova, D. (2011). Effects of Aerobic Training on Oxidative Status in Postsurgical Non-Small Cell Lung Cancer Patients: A Pilot Study. *Lung Cancer*, 72(1), 45–51. <https://doi.org/10.1016/j.lungcan.2010.08.002>

- Junqueira, L. C. U., Bignolas, G., & Brentani, R. R. (1979). Picrosirius staining plus polarization microscopy, a specific method for collagen detection in tissue sections. *The Histochemical Journal*, *11*(4), 447–455. <https://doi.org/10.1007/BF01002772>
- KING, T. E., TOOZE, J. A., SCHWARZ, M. I., BROWN, K. R., & CHERNIACK, R. M. (2001). Predicting Survival in Idiopathic Pulmonary Fibrosis. *American Journal of Respiratory and Critical Care Medicine*, *164*(7), 1171–1181. <https://doi.org/10.1164/ajrccm.164.7.2003140>
- Komohara, Y., Ohnishi, K., Kuratsu, J., & Takeya, M. (2008). Possible involvement of the M2 anti-inflammatory macrophage phenotype in growth of human gliomas. *The Journal of Pathology*, *216*(1), 15–24. <https://doi.org/10.1002/path.2370>
- Korfei, M., Mahavadi, P., & Guenther, A. (2022). Targeting Histone Deacetylases in Idiopathic Pulmonary Fibrosis: A Future Therapeutic Option. *Cells*, *11*(10), 1626. <https://doi.org/10.3390/cells11101626>
- Lee, J.-W., Chun, W., Lee, H. J., Min, J.-H., Kim, S.-M., Seo, J.-Y., Ahn, K.-S., & Oh, S.-R. (2021). The Role of Macrophages in the Development of Acute and Chronic Inflammatory Lung Diseases. *Cells*, *10*(4), 897. <https://doi.org/10.3390/cells10040897>
- Leitz, D. H. W., Duerr, J., Mulugeta, S., Seyhan Agircan, A., Zimmermann, S., Kawabe, H., Dalpke, A. H., Beers, M. F., & Mall, M. A. (2021). Congenital Deletion of Nedd4-2 in Lung Epithelial Cells Causes Progressive Alveolitis and Pulmonary Fibrosis in Neonatal Mice. *International Journal of Molecular Sciences*, *22*(11), 6146. <https://doi.org/10.3390/ijms22116146>
- Ley, B., Ryerson, C. J., Vittinghoff, E., Ryu, J. H., Tomassetti, S., Lee, J. S., Poletti, V., Buccioli, M., Elicker, B. M., Jones, K. D., King, T. E., & Collard, H. R. (2012). A Multidimensional Index and Staging System for Idiopathic Pulmonary Fibrosis. *Annals of Internal Medicine*, *156*(10), 684. <https://doi.org/10.7326/0003-4819-156-10-201205150-00004>
- Li, S., Gao, S., Jiang, Q., Liang, Q., Luan, J., Zhang, R., Zhang, F., Ruan, H., Li, X., Li, X., Zhou, H., & Yang, C. (2021). Clevudine attenuates bleomycin-induced early pulmonary fibrosis via regulating M2 macrophage polarization. *International Immunopharmacology*, *101*, 108271. <https://doi.org/10.1016/j.intimp.2021.108271>

- Lin, C.-Y., Chin, C.-H., Wu, H.-H., Chen, S.-H., Ho, C.-W., & Ko, M.-T. (2008). Hubba: hub objects analyzer—a framework of interactome hubs identification for network biology. *Nucleic Acids Research*, *36*(suppl_2), W438–W443. <https://doi.org/10.1093/nar/gkn257>
- Liu, Y.-M., Nepali, K., & Liou, J.-P. (2017). Idiopathic Pulmonary Fibrosis: Current Status, Recent Progress, and Emerging Targets. *Journal of Medicinal Chemistry*, *60*(2), 527–553. <https://doi.org/10.1021/acs.jmedchem.6b00935>
- Llewellyn, B. (2019, January). *Trichrome Stains*. Stains File. http://stainsfile.info/theory/tri_gen.htm
- López De Padilla, C. M., Coenen, M. J., Tovar, A., de la Vega, R. E., Evans, C. H., & Müller, S. A. (2021). Picrosirius Red Staining: Revisiting Its Application to the Qualitative and Quantitative Assessment of Collagen Type I and Type III in Tendon. *Journal of Histochemistry & Cytochemistry*, *69*(10), 633–643. <https://doi.org/10.1369/00221554211046777>
- McErlean, P., Bell, C. G., Hewitt, R. J., Busharat, Z., Ogger, P. P., Ghai, P., Albers, G. J., Calamita, E., Kingston, S., Molyneaux, P. L., Beck, S., Lloyd, C. M., Maher, T. M., & Byrne, A. J. (2021). DNA Methylome Alterations Are Associated with Airway Macrophage Differentiation and Phenotype during Lung Fibrosis. *American Journal of Respiratory and Critical Care Medicine*, *204*(8), 954–966. <https://doi.org/10.1164/rccm.202101-0004OC>
- Moore, B. B., & Hogaboam, C. M. (2008). Murine models of pulmonary fibrosis. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, *294*(2), L152–L160. <https://doi.org/10.1152/ajplung.00313.2007>
- Morde, V. (2019, April 7). *XGBoost Algorithm: Long May She Reign!* Towards Data Science. <https://towardsdatascience.com/https-medium-com-vishalorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d>
- Mou, Y., Wu, G., Wang, Q., Pan, T., Zhang, L., Xu, Y., Xiong, W., Zhou, Q., & Wang, Y. (2022). Macrophage-targeted delivery of siRNA to silence *Mecp2* gene expression attenuates pulmonary fibrosis. *Bioengineering & Translational Medicine*, *7*(2). <https://doi.org/10.1002/btm2.10280>

- Rasuli, B., & Weerakkody, Y. (2012). Usual interstitial pneumonia. In *Radiopaedia.org*. Radiopaedia.org. <https://doi.org/10.53347/rID-16895>
- Schafer, M. J., Mazula, D. L., Brown, A. K., White, T. A., Atkinson, E., Pearsall, V. M., Aversa, Z., Verzosa, G. C., Smith, L. A., Matveyenko, A., Miller, J. D., & LeBrasseur, N. K. (2019). Late-life time-restricted feeding and exercise differentially alter healthspan in obesity. *Aging Cell*, 18(4). <https://doi.org/10.1111/accel.12966>
- Shinde, A. v., Humeres, C., & Frangogiannis, N. G. (2017). The role of α -smooth muscle actin in fibroblast-mediated matrix contraction and remodeling. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1863(1), 298–309. <https://doi.org/10.1016/j.bbadis.2016.11.006>
- Si, W., He, X., Li, A., Liu, L., Li, J., Gong, D., Liu, J., Liu, J., Shen, W., & Zhang, X. (2016). Application of an integrated biomarker response index to assess ground water contamination in the vicinity of a rare earth mine tailings site. *Environmental Science and Pollution Research*, 23(17), 17345–17356. <https://doi.org/10.1007/s11356-016-6728-8>
- Sul, O. J., Kim, J. H., Lee, T., Seo, K. W., Cha, H. J., Kwon, B., Ahn, J.-J., Cho, Y. S., Oh, Y.-M., Jegal, Y., & Ra, S. W. (2022). GSPE Protects against Bleomycin-Induced Pulmonary Fibrosis in Mice via Ameliorating Epithelial Apoptosis through Inhibition of Oxidative Stress. *Oxidative Medicine and Cellular Longevity*, 2022, 1–16. <https://doi.org/10.1155/2022/8200189>
- Sumariwalla, P. F., Cao, Y., Wu, H.-L., Feldmann, M., & Paleolog, E. M. (2002). The angiogenesis inhibitor protease-activated kringle 1–5 reduces the severity of murine collagen-induced arthritis. *Arthritis Res Ther*, 5(1), R32. <https://doi.org/10.1186/ar608>
- Tanner, M. K., Tang, Z., & Thornton, C. A. (2021). Targeted splice sequencing reveals RNA toxicity and therapeutic response in myotonic dystrophy. *Nucleic Acids Research*, 49(4), 2240–2254. <https://doi.org/10.1093/nar/gkab022>
- Tashiro, J., Rubio, G. A., Limper, A. H., Williams, K., Elliot, S. J., Ninou, I., Aidinis, V., Tzouvelekis, A., & Glassberg, M. K. (2017). Exploring Animal Models That Resemble Idiopathic Pulmonary Fibrosis. *Frontiers in Medicine*, 4. <https://doi.org/10.3389/fmed.2017.00118>

- Trunfio, T. A., Borrelli, A., & Improta, G. (2022). Is It Possible to Predict the Length of Stay of Patients Undergoing Hip-Replacement Surgery? *International Journal of Environmental Research and Public Health*, 19(10), 6219. <https://doi.org/10.3390/ijerph19106219>
- Tsitoura, E., Trachalaki, A., Vasarmidi, E., Mastrodemou, S., Margaritopoulos, G. A., Kokosi, M., Fanidis, D., Galaris, A., Aidinis, V., Renzoni, E., Tzanakis, N., Wells, A. U., & Antoniou, K. M. (2021). Collagen 1a1 Expression by Airway Macrophages Increases In Fibrotic ILDs and Is Associated With FVC Decline and Increased Mortality. *Frontiers in Immunology*, 12. <https://doi.org/10.3389/fimmu.2021.645548>
- Tugcu, B., Bayraktar, H., Ekinci, C., Kucukodaci, Z., Tunali, M., & Nuhoglu, F. (2022). The effect of platelet-rich fibrin on wound healing following strabismus surgery. *Cutaneous and Ocular Toxicology*, 1–6. <https://doi.org/10.1080/15569527.2022.2081173>
- Vasković, J., & McLaren, N. (2022, March 28). *Overview and types of connective tissue*. <https://www.kenhub.com/en/library/anatomy/overview-and-types-of-connective-tissue>.
- Wang, C., & Yang, J. (2022). Mechanical forces: The missing link between idiopathic pulmonary fibrosis and lung cancer. *European Journal of Cell Biology*, 101(3), 151234. <https://doi.org/10.1016/j.ejcb.2022.151234>
- Wang, S., Liang, Y., & Dai, C. (2022). Metabolic Regulation of Fibroblast Activation and Proliferation during Organ Fibrosis. *Kidney Diseases*, 8(2), 115–125. <https://doi.org/10.1159/000522417>
- Watters, L. C., King, T. E., Schwarz, M. I., Waldron, J. A., Stanford, R. E., & Cherniack, R. M. (1986). A Clinical, Radiographic, and Physiologic Scoring System for the Longitudinal Assessment of Patients with Idiopathic Pulmonary Fibrosis¹⁻³. *American Review of Respiratory Disease*, 133(1), 97–103. <https://doi.org/10.1164/arrd.1986.133.1.97>
- Weerakkody, Y., & Jones, J. (2011). Interstitial lung disease. In *Radiopaedia.org*. Radiopaedia.org. <https://doi.org/10.53347/rID-14479>
- Westermann-Clark, E., Soundararajan, R., Fukumoto, J., Patil, S. S., Stearns, T. M., Saji, S., Czachor, A., Hernandez-Cuervo, H., Breitzig, M., Krishnamurthy, S., Lockey, R. F., & Kolliputi, N. (2022). Matrix Metalloproteinase 7 Expression and Apical Epithelial Defects in Atp8b1

Mutant Mouse Model of Pulmonary Fibrosis. *Biomolecules*, 12(2), 283.
<https://doi.org/10.3390/biom12020283>

Wu, X.-R., Peng, H.-X., He, M., Zhong, R., Liu, J., Wen, Y.-K., Li, C.-C., Li, J.-F., Xiong, S., Yu, T., Zheng, H.-B., Chen, Y.-H., He, J.-X., Liang, W.-H., & Cai, X.-Y. (2022). Macrophages-based immune-related risk score model for relapse prediction in stage I–III non-small cell lung cancer assessed by multiplex immunofluorescence. *Translational Lung Cancer Research*, 11(4), 523–542. <https://doi.org/10.21037/tlcr-21-916>

Zisman, D. A., Lynch, J. P., Toews, G. B., Kazerooni, E. A., Flint, A., & Martinez, F. J. (2000). Cyclophosphamide in the Treatment of Idiopathic Pulmonary Fibrosis. *Chest*, 117(6), 1619–1626.
<https://doi.org/10.1378/chest.117.6.1619>