BLINDED TITLE PAGE

Automated quantitative CT versus visual CT scoring in idiopathic pulmonary fibrosis: validation against pulmonary function

PURPOSE: To determine whether a novel CT post-processing software technique (CALIPER) is superior to visual CT scoring as judged by functional correlations in idiopathic pulmonary fibrosis.

MATERIALS and METHODS: 283 consecutive patients with idiopathic pulmonary fibrosis had CT parenchymal patterns evaluated quantitatively with CALIPER and by visual scoring. These two techniques were evaluated against: FEV1, FVC, DLco, Kco and a composite physiological index (CPI), with regard to extent of interstitial lung disease, extent of emphysema and pulmonary vascular abnormalities.

RESULTS: CALIPER-derived estimates of interstitial lung disease extent demonstrated stronger univariate correlations than visual scores for most pulmonary function tests: (FEV1: CALIPER R^2=0·29, visual R^2=0·18, FVC: CALIPER R^2=0·41, visual R^2=0·27; DLco: CALIPER R^2=0·31, visual R^2=0·35; CPI: CALIPER R^2=0·48, visual R^2=0·44). Correlations between CT measures of emphysema extent and pulmonary function tests were weak and did not differ significantly between CALIPER and visual scoring. Intriguingly, the pulmonary vessel volume provided similar correlations to total interstitial lung disease extent scored by CALIPER for FVC, DLco and CPI (FVC: R^2=0·45; DLco: R^2=0·34; CPI: R^2=0·53).

CONCLUSIONS: CALIPER was superior to visual scoring as validated by functional correlations with pulmonary function tests. The pulmonary vessel volume, a novel CALIPER CT parameter with no visual scoring equivalent, has the potential to be a CT feature in the assessment of patients with idiopathic pulmonary fibrosis and requires further exploration.

KEY WORDS

QUANTITATIVE COMPUTER ANALYSIS

IDIOPATHIC PULMONARY FIBROSIS

VISUAL CT ANALYSIS

PULMONARY VESSEL VOLUME

INTRODUCTION

Computer-based quantitative CT evaluation has the potential for greater precision than visual scoring in the estimation of the extent of diffuse parenchymal diseases. A new generation of computer-based CT software tools have demonstrated similar results between computer quantitation and visual quantitative scoring in small-scale studies in patients with idiopathic pulmonary fibrosis (IPF), [1, 2] with distinct improvement in performance on older, less sophisticated software programs.[3, 4]

Pulmonary damage in IPF is the consequence of pathological involvement of three components of the lung: the parenchyma, the vasculature (largely due to pulmonary hypertension)[5] and, in a large proportion of IPF patients who are cigarette smokers, co-existent emphysema.[6, 7] Our study assessed baseline involvement of these three compartments using traditional visual CT evaluation and a sophisticated quantitative CT software tool, CALIPER (Computer-Aided Lung Informatics for Pathology Evaluation and Rating), developed at the (_____). In our study, we set out to validate computer-based CT scoring by examining correlations between CT patterns against pulmonary function tests.

MATERIALS and METHODS

Study population and clinical information:

A retrospective analysis of an interstitial lung disease database identified all new consecutive patients, (over a four and a half year period (January 2007 to July 2011)), with a multidisciplinary team diagnosis of IPF according to published guidelines.[8] Patients with a non-contrast, supine, volumetric thin section CT were collected, and subsequent exclusions are shown (as per CONSORT in Figure 1). Pulmonary function and echocardiography protocols are explained in the online appendix. The DICOM images for the CT scans were transferred to the (_____) for blinded CALIPER processing. Approval for this study of clinically indicated CT and pulmonary function data was obtained from the Institutional Ethics Committee of the (_____) and the Institutional Review Board of (_____).

CT protocol:

The CT scans were obtained using a 64-slice multiple detector CT scanner (Somatom Sensation 64, Siemens, Erlangen, Germany) or a 4-slice multiple detector CT scanner (Siemens Volume Zoom, Siemens, Erlangen, Germany. To satisfy requirements for processing by the CALIPER algorithm, all scans were reconstructed using a high spatial frequency, B70 kernel (Siemens, Munich, Germany). All patients were scanned from lung apices to bases, at full inspiration, using a peak voltage of 120kVp with tube current modulation (range 30-140 mA). Images of 1mm thickness were viewed at window settings optimized for the assessment of the lung parenchyma (width 1500 H.U.; level -500 H.U.).

CT visual evaluation:

Each CT scan was evaluated independently by two radiologists (___,___) with 5 and 7 years thoracic imaging experience respectively, blinded to all clinical information. An initial training dataset of 15 non-study cases was used to help to identify pre-existing biases. The scores of the test cases were reviewed and the most widely discrepant results discussed with a third radiologist (__).

CTs were scored on a lobar basis using a continuous scale. The total interstitial lung disease (ILD) extent was estimated to the nearest 5%, and sub-classified into four patterns: reticular pattern, ground glass opacification, honeycombing and consolidation, using definitions from the Fleischner Society glossary of terms for thoracic imaging.[9] To derive a lobar percentage for each parenchymal pattern, the total lobar ILD extent was multiplied by individual lobar parenchymal pattern extents and divided by 100. The percentage (to the nearest 5%) of each lobe that contained mosaicism (decreased attenuation component) or emphysema was recorded. The individual lobar percentages of each parenchymal pattern were summed for each radiologist and divided by six to create an averaged lobar score per pattern, per scorer per case.

Traction bronchiectasis, as defined in the Fleischner society glossary of terms, [9] was assigned with a categorical "severity" score that took into account the average degree of airway dilatation within areas of fibrosis as well as the extent of dilatation throughout the lobe and was given an overall score of: none=0, mild=1, moderate=2, severe=3. An index of pulmonary hypertension (main pulmonary artery:ascending aorta ratio) was assessed by a single scorer using electronic caliper diameter measurements of the ascending aorta and pulmonary artery diameters at the level of the pulmonary artery bifurcation.[10]

The identification of systematic biases in visual scores was achieved by plotting the spread of differences in parenchymal pattern scores between observers. The most disparate 5% (two standard deviations) of values were arbitrated by a third scorer for all parameters except traction bronchiectasis, thereby minimizing bias within the original scorers. The original scorers derived a consensus for the traction bronchiectasis score. If a single parenchymal subtype extent was changed at consensus, the other parameters were modified, following CT review, to retain an overall sum of 100% for the four parenchymal subtypes. Similarly, if the lobar percentages of total interstitial disease, emphysema or mosaicism varied, the other two parameter extents were rescored.

CALIPER CT evaluation:

Data processing: Initial data processing steps involved extraction of the lung from the surrounding thoracic structures and segmentation into upper, middle and lower zones. Lung segmentation was performed with an adaptive density-based morphological approach,[11] whilst airway segmentation involved iterative three-dimensional region growing, density thresholding (thresholds including -950HU and -960HU) and connected components analysis.

Parenchymal tissue type classification was applied to 15x15x15 voxel volume units using texture analysis, computer vision-based image understanding of volumetric histogram signature mapping features and 3D morphology.[12] The CALIPER tool was trained by sub-specialty thoracic radiologist consensus assessment of pathologically confirmed datasets.[2, 12]

Pattern evaluation: CALIPER and visual CT were used to quantify pulmonary variables in three domains:

i) total extent of interstitial disease

ii) total extent of emphysema

iii) pulmonary vessels

CALIPER evaluation of CT data involved algorithmic identification and volumetric quantification of every voxel volume unit into one of eight radiological parenchymal features: normal lung, three grades of decreased lung attenuation (grade 1=mild, 2=moderate 3=marked), ground glass opacification, reticular pattern, honeycombing and the pulmonary vessels (Figure 2). Volumes for all eight parenchymal features were converted into a percentage using the total lung volume also measured by CALIPER.

The final CALIPER emphysema index was defined by pre-processing the CALIPER variables against PFTs and visual comparison of CT scans. Using univariate linear correlations, Grade 1 decreased attenuation (DA) demonstrated no fit against the visual emphysema extent score or against Kco and was therefore not included within the CALIPER emphysema score. On inspection of colour maps, Grade 1 DA was shown to encompass areas of patchy centrilobular emphysema and considerable

amounts of normal lung (Figure 2a-c), which accounted for its large extent and lack of functional impact (Table 1). The sum of Grade 2 and 3 DA was taken to represent CALIPER emphysema and the suitability of the new variable was confirmed on analysis of individual colour maps (Figure 2a-c) that demonstrated that Grade 2 and 3-DA corresponded to discrete and conglomerate foci of emphysema.

Segmentation of pulmonary vessels, prior to their extraction, was achieved using an optimized multi-scale tubular structure enhancement filter based on the eigenvalues of the Hessian matrix. The filters calculated the 2nd-order derivatives that occurred in the regions that surrounded each pulmonary voxel. The eigenvalues of the Hessian matrix that were constructed from the derivatives were then analyzed, and from these values, it was possible to determine the likelihood that an underlying voxel was connected to a dense tubular structure and therefore represented a vessel.[12, 13] The pulmonary vessel volume (PVV) score quantified the volumes of pulmonary arteries and veins excluding vessels at the lung hilum as a percentage of lung volume (Figure 3). Total ILD extent represented the sum of ground glass, reticular and honeycomb percentages.

Statistical analysis:

Data are given as means with standard deviations, or numbers of patients with percentages where appropriate. Interobserver variation for visual scores was assessed using the single determination standard deviation.

Correlations between the extents of parenchymal patterns and individual PFTs were examined using Pearson's product moment correlation. Univariate and multivariate analyses were undertaken to investigate relationships between CALIPER or visual CT evaluation and pulmonary function tests. In multivariate analyses, robustness of relationships were tested by bootstrapping the dataset with 1000 samples. In all study analyses, a p-value of <0.01 was considered significant. Models were formally tested for heteroscedasticity to confirm that the assumptions of parametric analysis had been satisfied. Statistical analyses were performed with STATA (version 12, StatCorp, College Station, TX, USA).

RESULTS:

Baseline data

The final study group comprised 283 consecutive patients with a multidisciplinary diagnosis of IPF. Age, gender, mean visual scores, CALIPER scores and pulmonary function tests are shown in Table 1. Discordances between CALIPER and visual estimations of total ILD extent, emphysema and the various individual parenchymal patterns are shown in Table 1 and an illustrative example is shown in Figure 2. Visual scores identified on average 1.6 times more ILD than CALIPER and 10 times more honeycombing than CALIPER. Interobserver variation values for the visual scores are provided in Table 2.

Univariate relationships

The relationships between pulmonary function parameters (FEV1, FVC, DLco, Kco, CPI) and CT variables are shown in Table 3. Taken across the PFTs, CALIPER ILD extent was either superior to (FEV1, FVC) or comparable with (DLco, CPI) visual scoring (Table 3). CALIPER ILD extent and PVV had very similar correlations with pulmonary function indices. The PVV correlations were either superior to (FEV1, FVC, CPI) or comparable with (DLco) visual ILD extent scores. PVV increased with ILD extent (Quadratic R^2=0.76), but less so with more advanced disease. The visual pulmonary artery:aorta diameter ratio demonstrated no relationship to any functional index.

Multiple regression

Multivariate regression analyses of visual and CALIPER scores of pulmonary vasculature, total interstitial disease extent and emphysema extent were analysed against pulmonary function tests (Table 4). PVV and ILD extent could not be included in the same model due to major co-linearity and so were examined in separate models (Model 1 containing CALIPER ILD and Model 2 containing PVV).

In Model 1, CALIPER ILD extent was clearly superior for two PFTs (FEV1, FVC), (confirmed on bootstrapping the dataset with 1000 samples), with visual ILD extent discarded. CALIPER and visual ILD extents were complementary for two PFTs (DLco, CPI) with both variables retained (again

confirmed on bootstrapping the dataset with 1000 samples). In the second model, CALIPER PVV was the strongest determinant of all examined PFTs. Although visual ILD extent was retained for DLco and CPI, there was minimal effect on model fit when it was discarded (DLco change: R^2 0.04; CPI change R^2 0.03). Similarly, the inclusion of individual CT patterns (e.g. ground glass opacity, reticular pattern and honeycombing), whether quantified by CALIPER or visually, resulted in only a minimal improvement in correlations with PFTs (R^2 values increasing by <0.02).

Given the overall strong correlations between PVV and the various pulmonary function tests, relationships between PVV and RVSP were explored. Univariate analysis of a subgroup of 150 patients with concurrent right ventricular systolic pressure (RVSP) measured on echocardiography was performed. RVSP was found to explain 20% of the variability of the PVV (R^2=0·20, P<0·0001). In the same 150 patients, the extent of ILD measured by both CALIPER (R^2=0·73, P<0·0001) and visually (R^2=0·47, P<0·0001), better explained PVV variability. Furthermore, with these two CT variables included in separate models, RVSP had no independent linkage with PVV.

Post-hoc evaluation: CALIPER-based validation of CPI

On the basis of the strong linkages between CALIPER ILD extent and PFTs, the robustness of the CPI was validated using CALIPER-scored ILD extent in the current study group. This showed that the same three best-fit PFTs, as used in the original CPI model, predicted CALIPER ILD extent:

CALIPER-derived CPI = 66.0 – (0.47 x DLco) – (0.67 x FVC) + (0.32 x FEV1)

Correlation of the new CPI against the old CPI score demonstrated a strong linkage as shown in the graph in Figure 4 (R^2=0.95).

DISCUSSION

Our study has shown that in an IPF population, CALIPER derived interstitial and vascular parameters correlated more strongly with pulmonary function indices (FEV1, FVC) or were at least comparable to visual scores (DLco, CPI). Consequently, it was felt logical to explore the robustness of the CPI using an objective scoring methodology (CALIPER) and, in so doing, the CPI was vindicated. Importantly, we have also shown that the pulmonary vessel volume, a novel CALIPER-derived percentage of the lung composed of pulmonary arteries and veins, is surprisingly strongly linked to the extent of interstitial disease.

The increase in the size and number of drug trials in IPF has necessitated the development of new automated computer-based algorithms capable of analysing hundreds of CTs per study. The accuracy of the new computer-based techniques requires validation to ensure that they are at least comparable to visual CT scoring. One of the steps in validating the accuracy of CT in assessing disease extent is by examining the relationship between CT estimates, however obtained, and pulmonary function measures of disease severity. There are numerous structure-function studies in interstitial lung disease, but these have almost exclusively relied on visual scoring of total and individual pattern extents, with all the inherent problems of interobserver agreement.[14-16]

In our study, CALIPER-derived interstitial and vascular markers clearly demonstrated stronger or comparable correlations with all cardinal pulmonary function tests than visual scores. The strong univariate correlations between both visual and CALIPER-scored total ILD extent and pulmonary function parameters are in line with previous IPF studies. [16, 17] Whilst the primary purpose of this study was a validation exercise of CALIPER against visual scores, the strong correlations between CALIPER scores and PFTs made it possible to examine historical derivation of the CPI. The CPI, originally derived using a subjective visual scoring system, was validated as a robust variable following replication of the CPI by using scores from the objective CALIPER system.

CALIPER was less sensitive to the extent of emphysema as compared to visual scoring and less accurate as judged by correlations with pulmonary function indices. Nevertheless, the positive correlations between visually and CALIPER scored emphysema and FVC, identified in the current study, are consistent with previous studies in IPF patients with co-existent emphysema. [18, 19]

On multivariate analysis of CALIPER and visual variables, it was striking that the strongest independent parameter predictive of pulmonary function tests was the PVV. We had assumed a priori that PVV was a measure of pulmonary vessel involvement, but intriguingly our investigations revealed that PVV was not only a ILD marker per se, but that it was co-linear with both CALIPER and visual ILD extents. PVV was at least as strong as CALIPER ILD extent in its linkage with key PFTs (FVC, CPI). Furthermore, after correcting for the extent of ILD, no linkage between PVV and RVSP remained, thereby establishing that PVV was not an index of pulmonary hypertension.

Discordances between CALIPER and visual scores of total ILD extent and honeycombing were assessed by analyzing outlying cases and were found to relate largely to differences in scoring methodologies. On visual evaluation, each lobe represented a sixth of the total lung volume, regardless of the extent of lobar disease. CALIPER however, assessed parenchymal patterns, as a proportion of the total lung volume. Lower lobes, contracted by the retractile fibrosis of UIP, contribute a smaller percentage of disease than, for example, non-fibrotic upper lobes and were thus under-represented by CALIPER when compared to visual scores. The disparity in visual and CALIPER estimates of reticular pattern and ground glass opacity reflect differences in categorization of a pattern of intermixed fine reticulation and ground glass opacity (Figure 2d-f). Furthermore, review of individual cases showed that quite frequently, CALIPER characterized visually scored honeycombing as reticular pattern and ground glass opacity (Figure 2g-i).

Our findings suggest that a computer-based quantitative CT tool such as CALIPER, has several valuable roles for the evaluation of patients with IPF. The improved sensitivity of CALIPER in evaluating ILD extent, when compared to visual scoring, has the potential to enhance understanding of the natural history of IPF by improving the accuracy of identifying serial change. In the sphere of drug trials,

computer-based CT evaluation has several possible applications. CALIPER could be used to correct for baseline CT disease extent in patients at the start of a trial and it could also be applied as a monitoring tool in the context of end-points. In the current study, recalibration of CALIPER software was required to analyse the Siemens B70f algorithms performed in our department, with the result that post-recalibration, similar results for parenchymal pattern extents were achieved when CALIPER analysed both edge-enhancing algorithms such as a Siemens B70f and "less edge-enhancing" algorithms such as a Siemens B46f, (the algorithm constituting most of CALIPER's original training dataset). A consequence of CALIPERs recalibration was an improvement in its versatility, which has relevance for multicentre drug trials, where CTs in different centres can be reconstructed with a range of different algorithms.

CALIPER could also make a useful contribution to the investigation of combined fibrosis and emphysema (CPFE). The strong correlations with functional indices we have demonstrated suggest that CALIPER would be a more suitable tool than visual scores to quantify ILD extent in CPFE, although conversely visual scoring may be best placed to quantify emphysema extent. Both methodologies used together may better delineate the contribution of each component of CPFE.

With regard to PVV, individual colour overlay maps demonstrated some contamination of the variable by areas of reticulation (Figure 3), particularly in cases with extensive pulmonary fibrosis. Nonetheless, the variable primarily reflects the quantitation, by CALIPER, of large and small vessels in the lung, in a way that has not, to date, been possible by human scorers. Evaluating this new parameter has resulted in a credible additional CT measure to visual ILD extent scores when quantifying interstitial involvement in IPF. A possible explanation for the relationship between PVV and ILD extent relates to the increased negative intra-thoracic pressure that non-compliant fibrotic lungs generate during inspiration. The transmission of high negative pressures into the lung parenchyma could in turn affect compliant vessels, resulting in vascular dilatation throughout the lung and an increase in capacitance. However deciphering the exact pathophysiological mechanisms that link interstitial damage to vascular volume requires further investigation. Furthermore, the potential prognostic role of PVV in patients with fibrosing lung disease is worthy of exploration.

Some limitations to the CALIPER technique are evident. One lies in the poor correlations we have identified between CALIPER emphysema extent and functional indices, (in particular Kco). Improving the detection of non-conglomerate emphysema by CALIPER would be a preliminary yet feasible objective. The complexities in scoring CT parenchymal pattern extents, be that visual or computer-based, in a disease that is inevitably associated with volume loss are considerable. Whilst discrepancies between CALIPER and visual scores for parenchymal patterns such as honeycombing have been partly explained, further studies directed towards clarifying the reasons behind the differences in disease extent scores are needed. Lastly, the minor contamination of the PVV signal by reticulation in cases with severe fibrosis, might be considered a limitation, as it could be thought to dilute the relationships between PVV and functional indices. However improvement and greater sophistication of the algorithm to detect vessels (versus reticulation) may result in simply strengthening the correlations we have already shown.

In conclusion, we have shown that CALIPER measures of lung disease are more strongly related to pulmonary function tests than visual scores. Strong links between CALIPER estimation of pulmonary vessel volume and pulmonary function tests suggests that evaluation of pulmonary vessel volume may be an important new index when assessing disease severity in patients with IPF.

REFERENCES

1. Kim HJ, Brown MS, Chong D, et al. Comparison of the Quantitative CT Imaging Biomarkers of Idiopathic Pulmonary Fibrosis at Baseline and Early Change with an Interval of 7 Months. *Acad Radiol*. 2015;**22**(1):70-80.
2. Maldonado F, Moua T, Rajagopalan S, et al. Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. *Eur Respir J*. 2014;**43**(1):204-12.
3. Best AC, Meng J, Lynch AM, et al. Idiopathic Pulmonary Fibrosis: Physiologic Tests, Quantitative CT Indexes, and CT Visual Scores as Predictors of Mortality. *Radiology*. 2008;**246**(3):935-40.
4. Iwasawa T, Asakura A, Sakai F, et al. Assessment of prognosis of patients with idiopathic pulmonary fibrosis by computer-aided analysis of CT images. *J Thorac Imaging*. 2009;**24**(3):216-22.
5. Nadrous HF, Pellikka PA, Krowka MJ, et al. Pulmonary hypertension in patients with idiopathic pulmonary fibrosis. *Chest*. 2005;**128**(4):2393-9.
6. Cottin V, Nunes H, Brillet PY, et al. Combined pulmonary fibrosis and emphysema: a distinct underrecognised entity. *European Respiratory Journal*. 2005;**26**(4):586-93.
7. Sugino K, Ishida F, Kikuchi N, et al. Comparison of clinical characteristics and prognostic factors of combined pulmonary fibrosis and emphysema versus idiopathic pulmonary fibrosis alone. *Respirology*. 2014;**19**(2):239-45.
8. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med*. 2011;**183**(6):788-824.
9. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology*. 2008;**246**(3):697-722.
10. Ng CS, Wells AU, Padley SP. A CT sign of chronic pulmonary arterial hypertension: the ratio of main pulmonary artery to aortic diameter. *J Thorac Imaging*. 1999;**14**(4):270-8.
11. Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans Med Imaging*. 2001;**20**(6):490-8.
12. Bartholmai BJ, Raghunath S, Karwoski RA, et al. Quantitative CT imaging of interstitial lung diseases. *J Thorac Imaging*. 2013;**28**(5):298-307.
13. Shikata H, McLennan G, Hoffman EA, Sonka M. Segmentation of pulmonary vascular trees from thoracic 3D CT images. *Int J Biomed Imaging*. 2009:11.
14. Mogulkoc N, Brutsche MH, Bishop PW, Greaves SM, Horrocks AW, Egan JJ. Pulmonary function in idiopathic pulmonary fibrosis and referral for lung transplantation. *Am J Respir Crit Care Med*. 2001;**164**:103-8.
15. Lynch DA, David GJ, Safrin S, et al. High-resolution computed tomography in idiopathic pulmonary fibrosis: diagnosis and prognosis. *Am J Respir Crit Care Med*. 2005;**172**:488-93.
16. Xaubet A, AgustÍ C, Luburich P, et al. Pulmonary Function Tests and CT Scan in the Management of Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 1998;**158**(2):431-6.

17. Wells AU, Desai SR, Rubens MB, et al. Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. *Am J Respir Crit Care Med*. 2003;**167**:962-9.
18. Akagi T, Matsumoto T, Harada T, et al. Coexistent emphysema delays the decrease of vital capacity in idiopathic pulmonary fibrosis. *Respir Med*. 2009;**103**:1209-15.
19. Kim YJ, Shin SH, Park J-W, et al. Annual Change in Pulmonary Function and Clinical Characteristics of Combined Pulmonary Fibrosis and Emphysema and Idiopathic Pulmonary Fibrosis: Over a 3-Year Follow-up. *Tuberc Respir Dis*. 2014;**77**:18-23.

FIGURE CAPTIONS

Figure 1. CONSORT diagram illustrating the selection of patients for the final study population. ILD = interstitial lung disease, CTD = connective tissue disease, IPF = idiopathic pulmonary fibrosis, LCH = Langerhans cell histiocytosis, LAM = Lymphangioleiomyomatosis, CT = computed tomography.

Figure 2A-C. Axial CT slices, axial CALIPER-derived colour image overlays and three-dimensional CALIPER coronal rendering of the lungs displaying parenchymal patterns as various colours (Dark green=normal lung, light green=grade 2 decreased attenuation area, light blue=grade 2 decreased attenuation area, dark blue=grade 3 decreased attenuation area, yellow=ground glass opacity, orange=reticular pattern, brown=honeycombing).

2A(i-iii) 75-year-old male ex-smoker with a 40-pack year smoking history. Mean visual scores of the CT: 7% reticular pattern, 3% honeycombing, no ground glass opacity and 24% emphysema (i). CALIPER characterized 5% reticular pattern, 1% honeycombing, 2% ground glass opacity, 43% Grade 1, 10% Grade 2 and 4% Grade 3 decreased attenuation and 3% pulmonary vessel volume (PVV). A large proportion of the areas with interspaced patches of centrilobular emphysema were characterized as Grade I decreased attenuation by CALIPER as demonstrated on overlaid axial ii) and 3D rendered images (iii) but characterized as predominantly normal lung with only minor emphysema on visual scores.

2B(i-iii) 48-year-old female never-smoker. Mean visual scores of the CT: 34% reticular pattern, 0.5% honeycombing, 45% ground glass opacity with no emphysema (i). CALIPER characterized 6% reticular pattern, 1% honeycombing, 62% ground glass opacity, no Grade 2 and 3 decreased attenuation and 8% PVV. A large proportion of the areas visually labeled reticular pattern were characterized as ground glass opacity by CALIPER as demonstrated on overlaid axial (ii) and 3D rendered images (iii) reflecting a pattern of textured ground glass opacity that is often difficult to classify.

2C(i-iii) 76-year-old male ex-smoker with a 20-pack year smoking history. Mean visual scores of the CT: 14% reticular pattern, 49% honeycombing, 5% ground glass opacity with no emphysema (i). CALIPER characterized 16% reticular pattern, 2% honeycombing, 24% ground glass opacity, 0·5% Grade 2 and 3 decreased attenuation and 9% PVV. A substantial proportion of the areas visually

labeled honeycombing were characterized as reticular pattern and/or ground glass opacity by CALIPER as shown on overlaid axial (ii) and 3D rendered images (iii).

Figure 3. A 75-year-old ex-smoker diagnosed with idiopathic pulmonary fibrosis. Axial CT slice with CALIPER-derived colour overlay demonstrating intraparenchymal pulmonary arteries and veins (red). During the initial extraction process separating the lungs from the mediastinum and chest wall, the hilar structures including the central pulmonary arteries and veins were removed. Pulmonary vessels are classified by CALIPER using structure and textural analysis and computer vision-based image understanding of volumetric histogram signature mapping features for 9x9x9 voxel volume units. The pulmonary vessel volume is calculated by dividing the total lung vessel volume by the total lung volume and multiplying by 100. The caliber of vessels in the spared right lung are increased when compared to vessels within areas of fibrosis.

Figure 4. Graph demonstrating the relationship between CPI derived from correlations between pulmonary function tests and ILD extent scored visually in a separate, historic cohort of IPF patients, and the CPI calculated using a new formula derived from correlations between pulmonary function tests and ILD extent scored by CALIPER in the current cohort (R^2=0·95).

Table 1. Patient age, gender and mean and standard deviations of pulmonary function tests, CALIPER and visually scored CT parameters and echocardiography data. Data represent mean values with standard deviations. CT = computed tomography, FEV1 = forced expiratory volume in one second, FVC = forced vital capacity, DLco = diffusing capacity for carbon monoxide, Kco = carbon monoxide transfer coefficient, TLC = total lung capacity, CPI = composite physiological index, ILD = interstitial lung disease, RVSP = right ventricular systolic pressure, TxBx = traction bronchiectasis.

Table 2. Single determination standard deviation values of visual CT scores for idiopathic pulmonary fibrosis cases. CT = computed tomography.

Table 3. Univariate linear regression demonstrating relationships between disease in three compartments characterized by visual and CALIPER-derived scores and pulmonary function tests. CT=Computed tomography, ILD=interstitial lung disease, FEV1=forced expiratory volume in one second, FVC=forced vital capacity, DLco=diffusing capacity for carbon monoxide, Kco = carbon monoxide transfer coefficient, CPI=composite physiological index, NS=not significant.

Table 4. Multivariate linear regression demonstrating relationships significant to a level of 0·01 between parenchymal patterns characterized by both visual and CALIPER scores with pulmonary function tests. Model 1 contained CALIPER ILD extent but excluded PVV. Model 2 contained PVV but excluded CALIPER ILD extent. CT= Computed tomography, FEV1=forced expiratory volume in one second, FVC=forced vital capacity, DLco=diffusing capacity for carbon monoxide, CPI=composite physiological index, ILD=interstitial lung disease, PVV= pulmonary vessel volume.

Figure 1

Click here to download Figure Figure 1.tiff

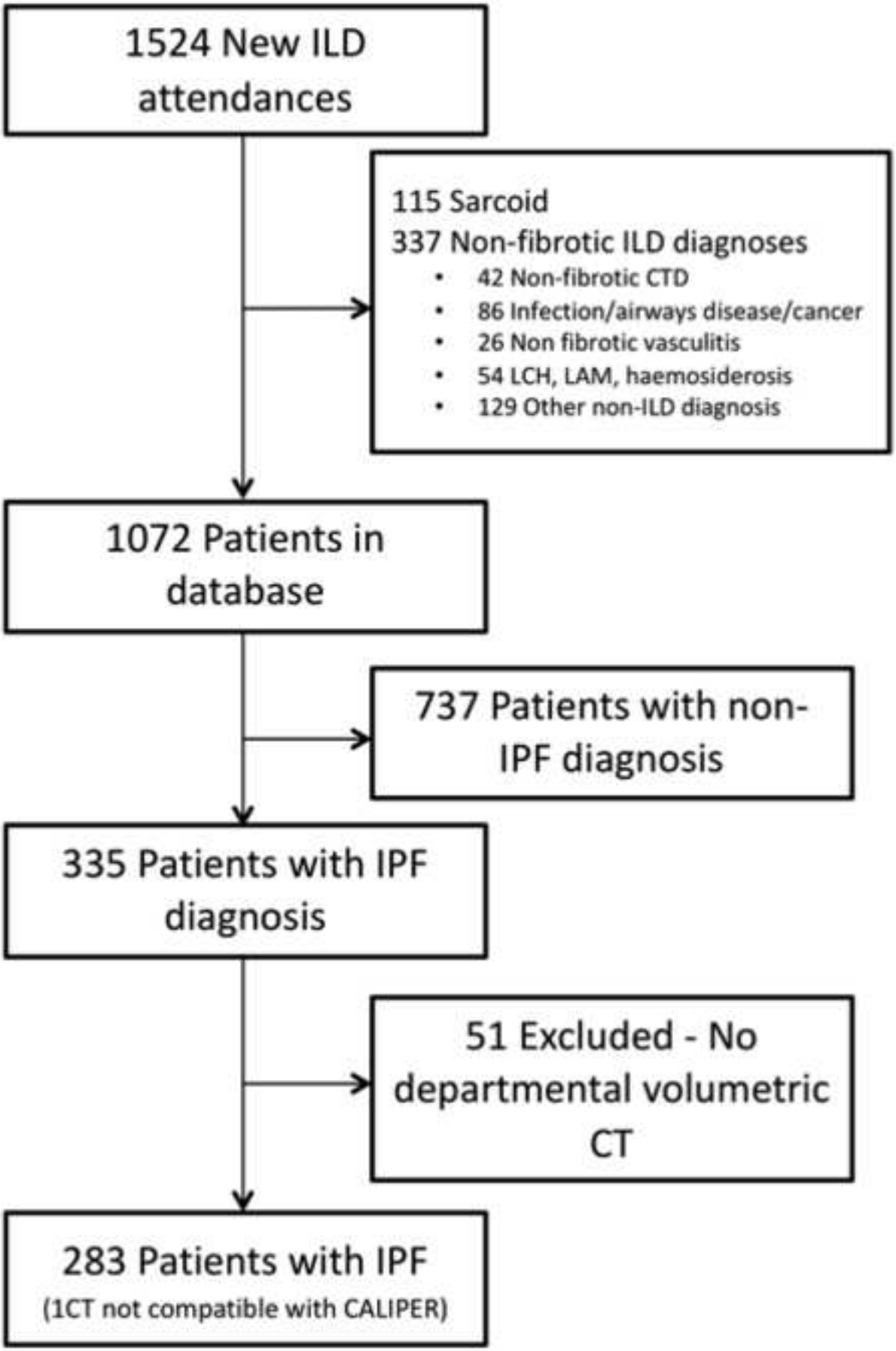

Figure 2ai

Click here to download Figure IMAGE 4H CASE005 crop.tif

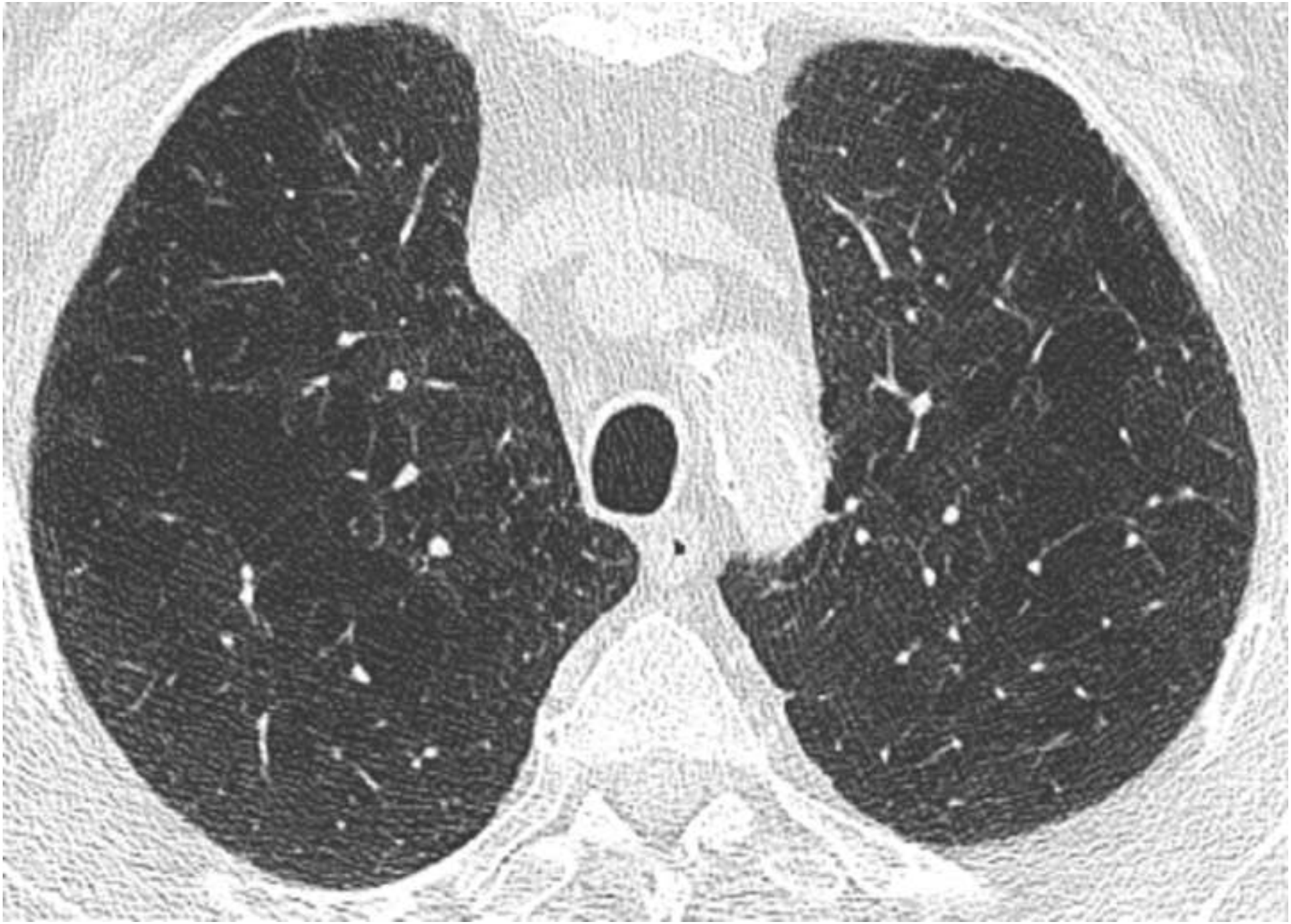

Figure 2aii

Click here to download Figure IMAGE4I CASE005 crop.tif

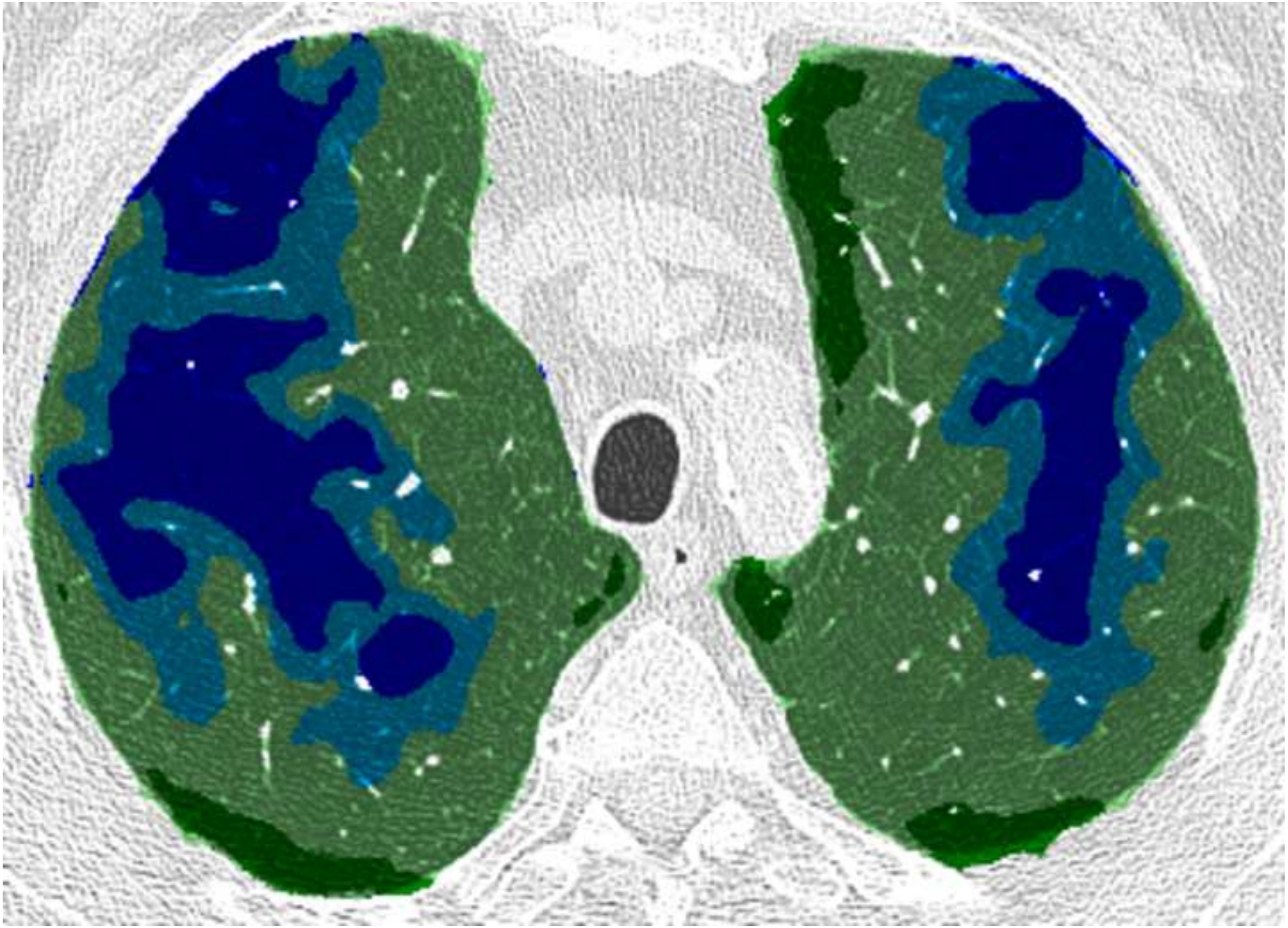

Figure 2aiii

Click here to download Figure IMAGE 4GCASE 005crop.tiff

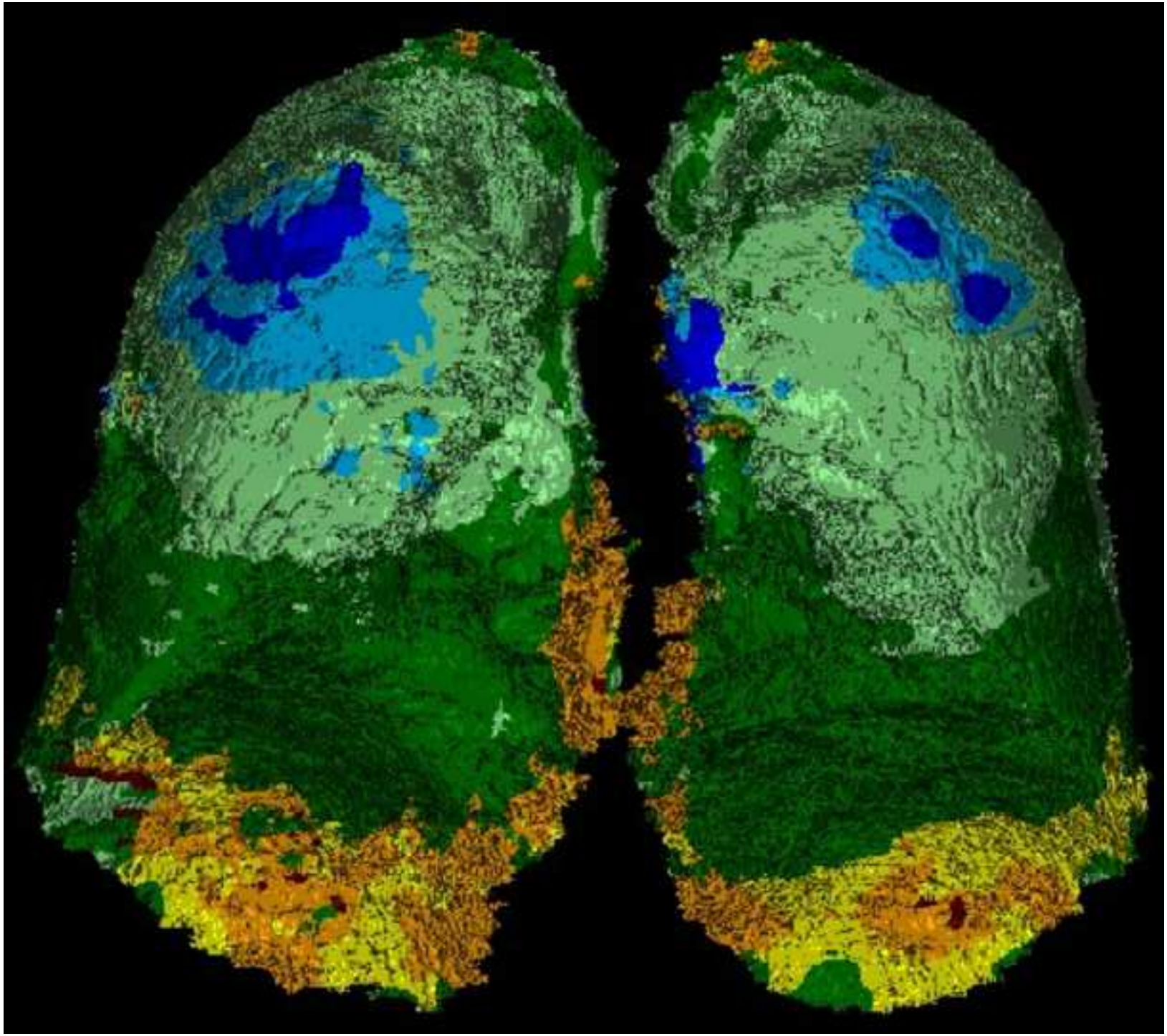

Figure 2bi

Click here to download Figure IMAGE 4E CASE080_crop.tif

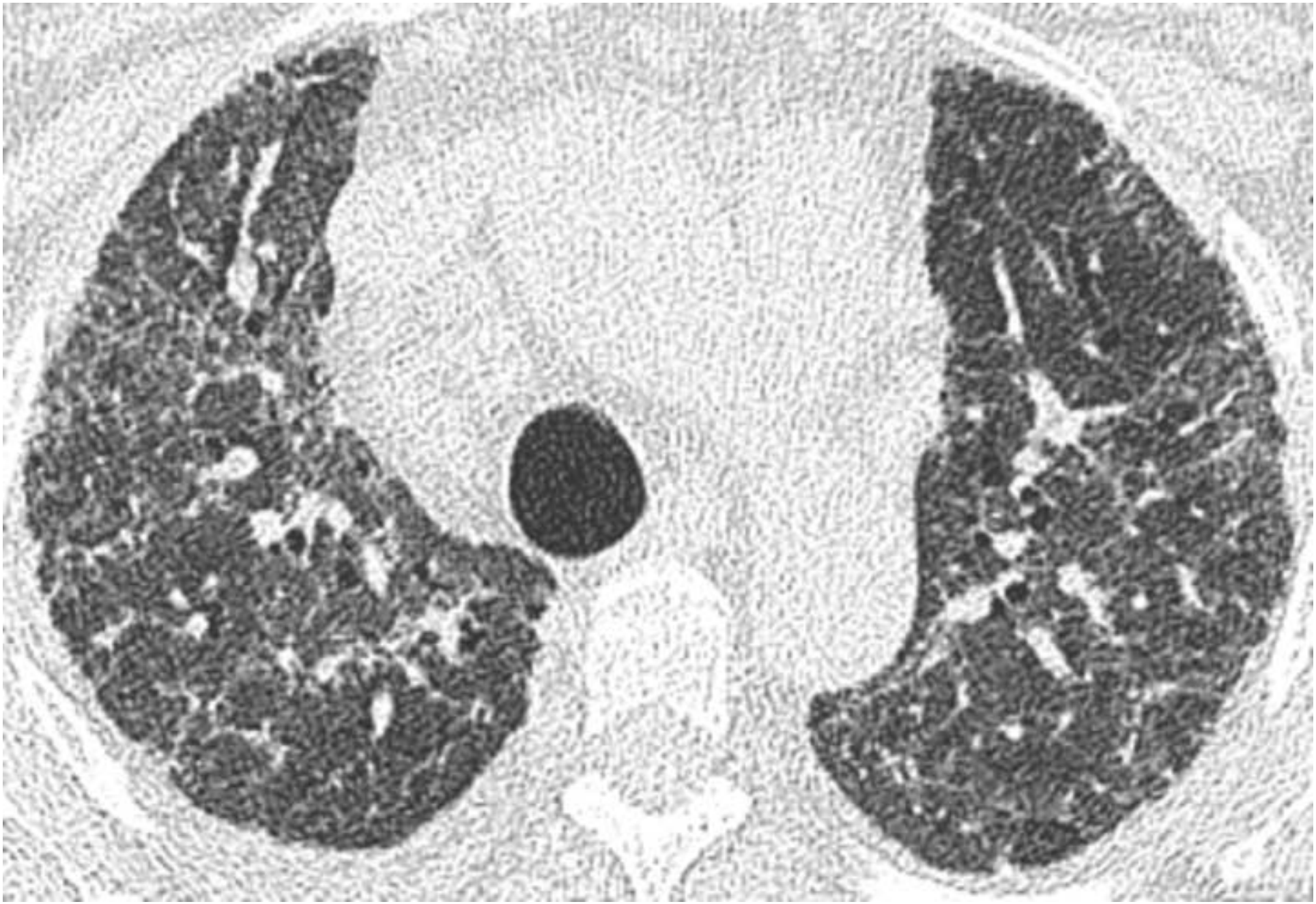

Figure 2bii

Click here to download Figure IMAGE 4F CASE080_crop.tif

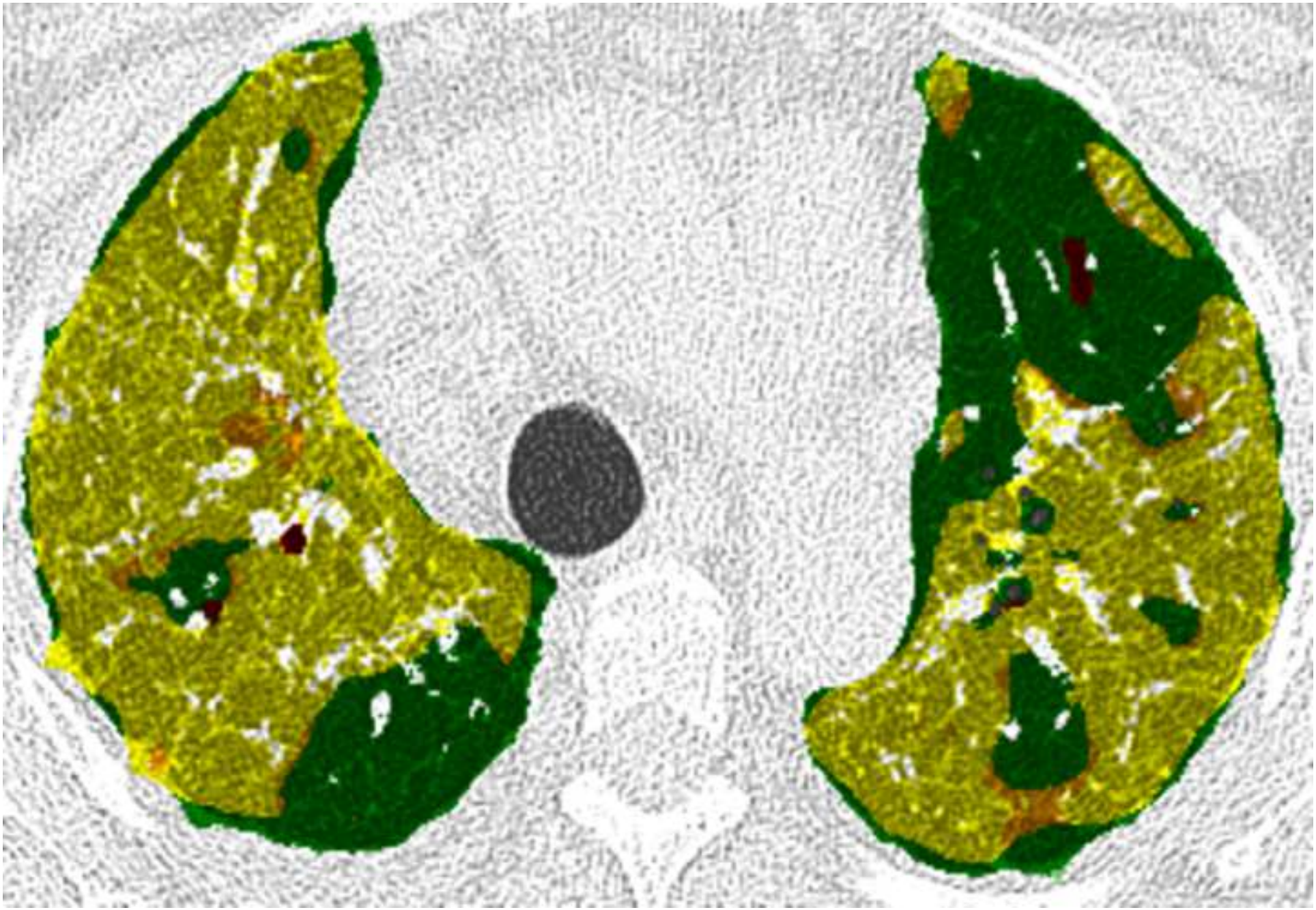

Figure 2biii

Click here to download Figure IMAGE 4D CASE 080_crop.tif

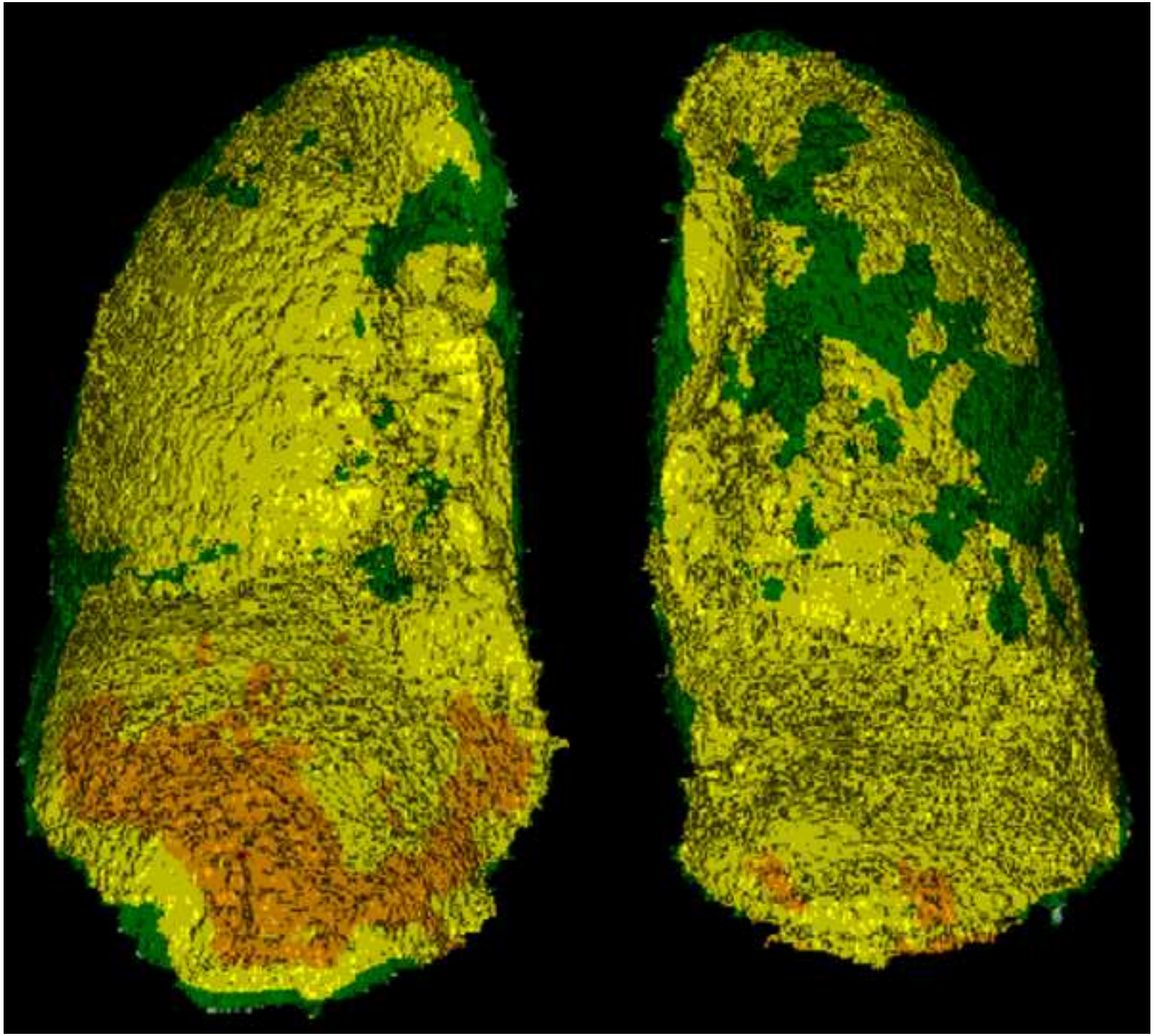

Figure 2ci

Click here to download Figure IMAGE 4B CASE078_crop.tif

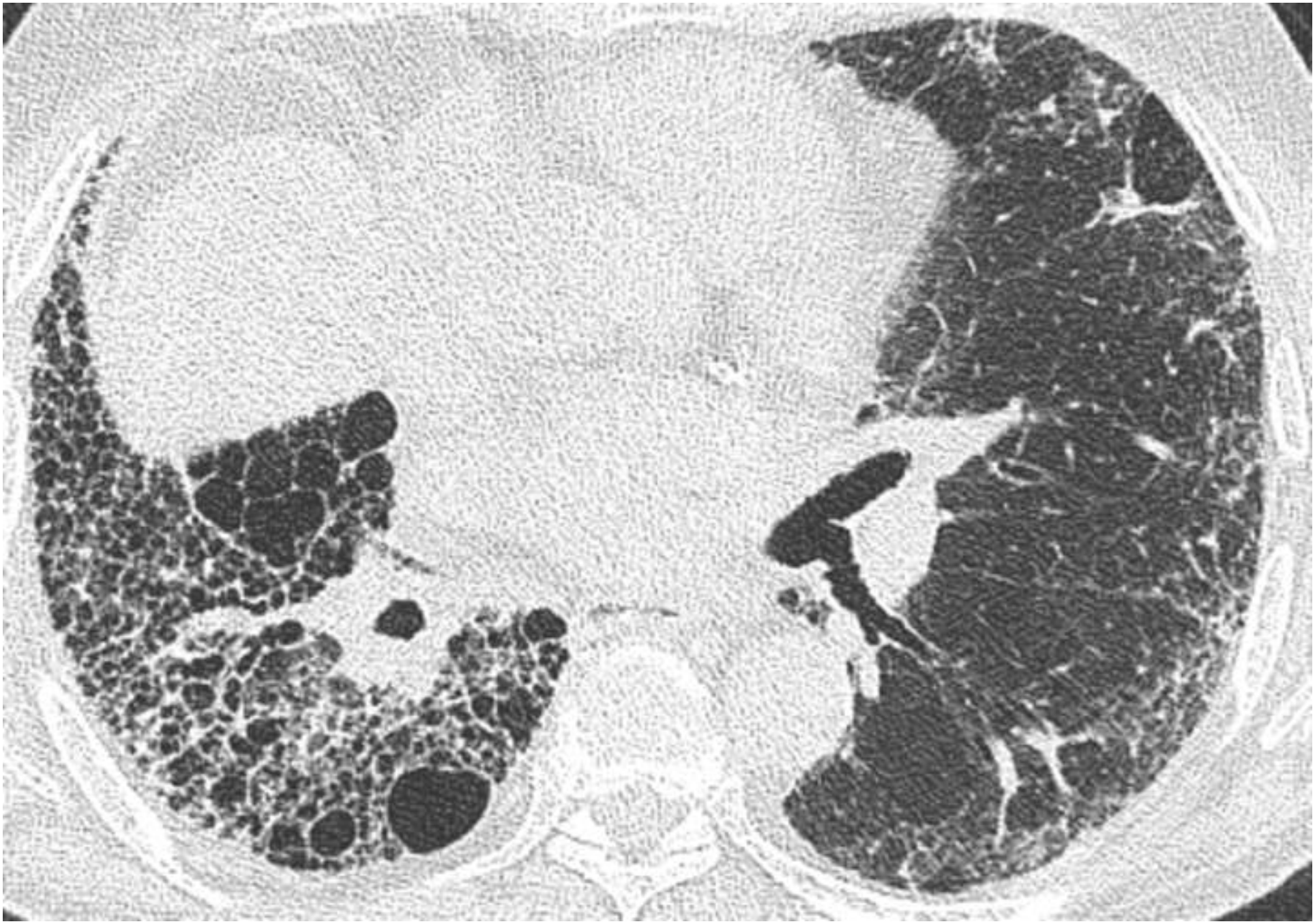

Figure 2cii

Click here to download Figure IMAGE 4C CASE078_crop.tif

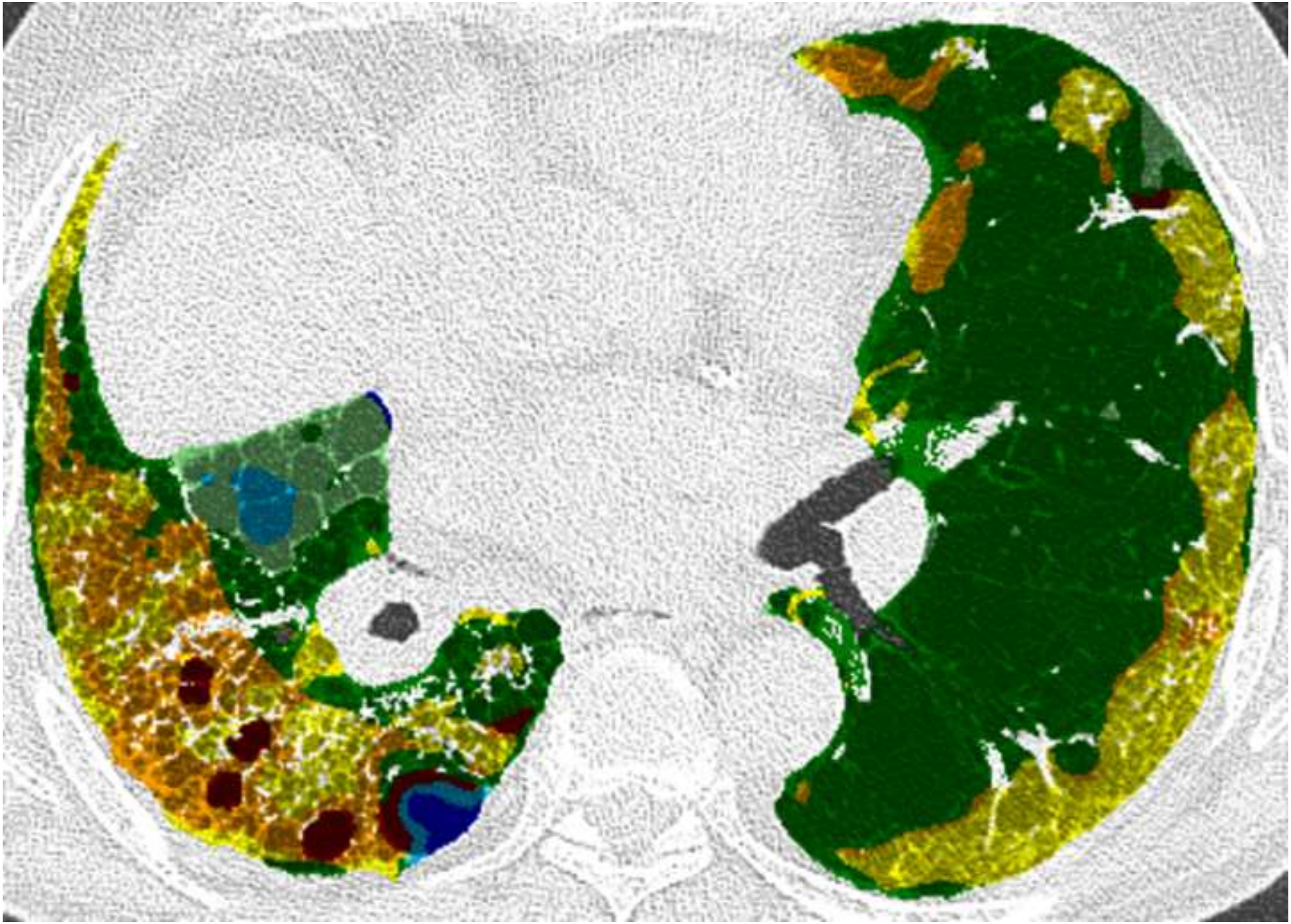

Figure 2ciii

Click here to download Figure IMAGE 4A CASE 078_Crop.tif

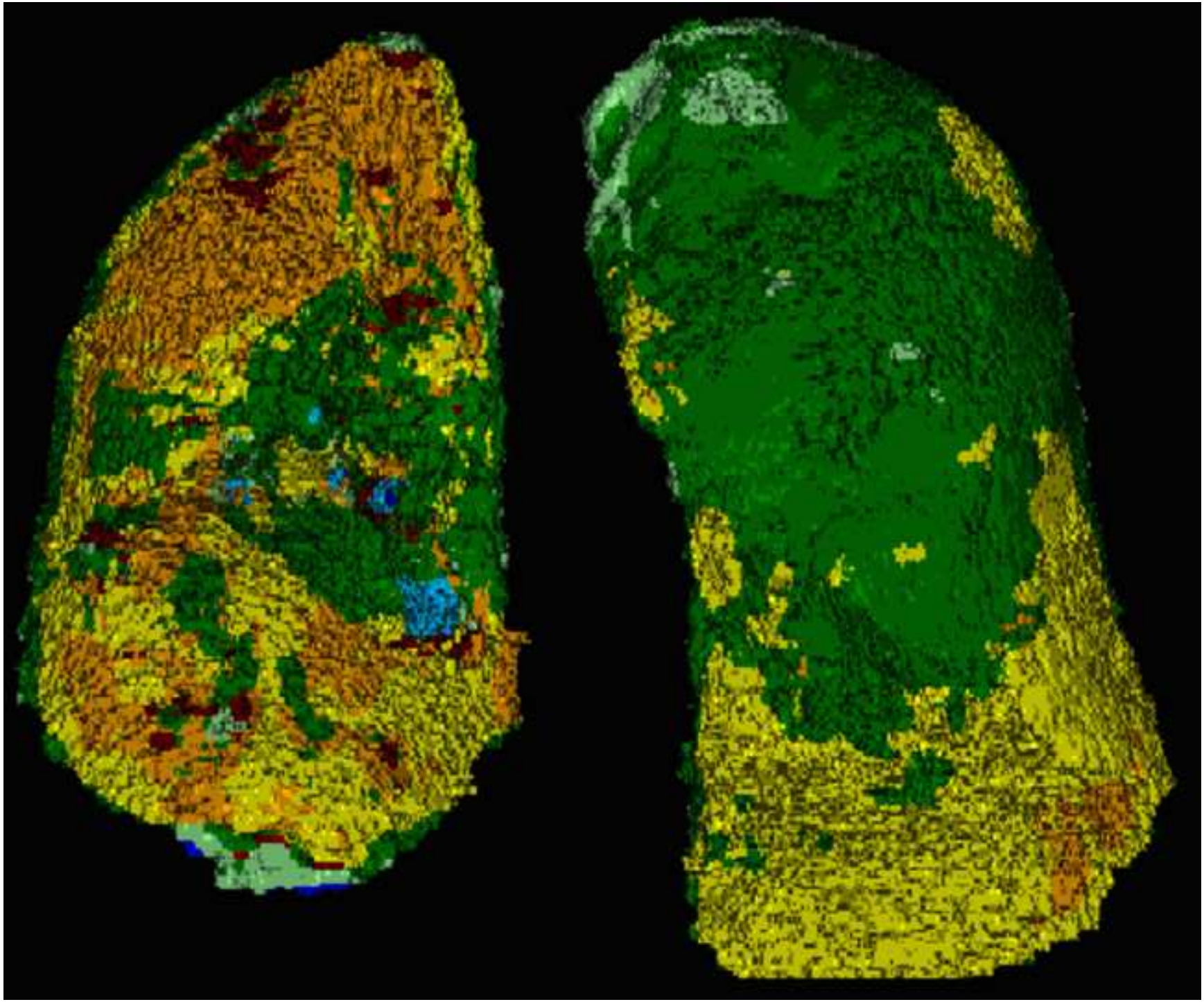

Figure 3

Click here to download Figure Vessels_150 crop .tif

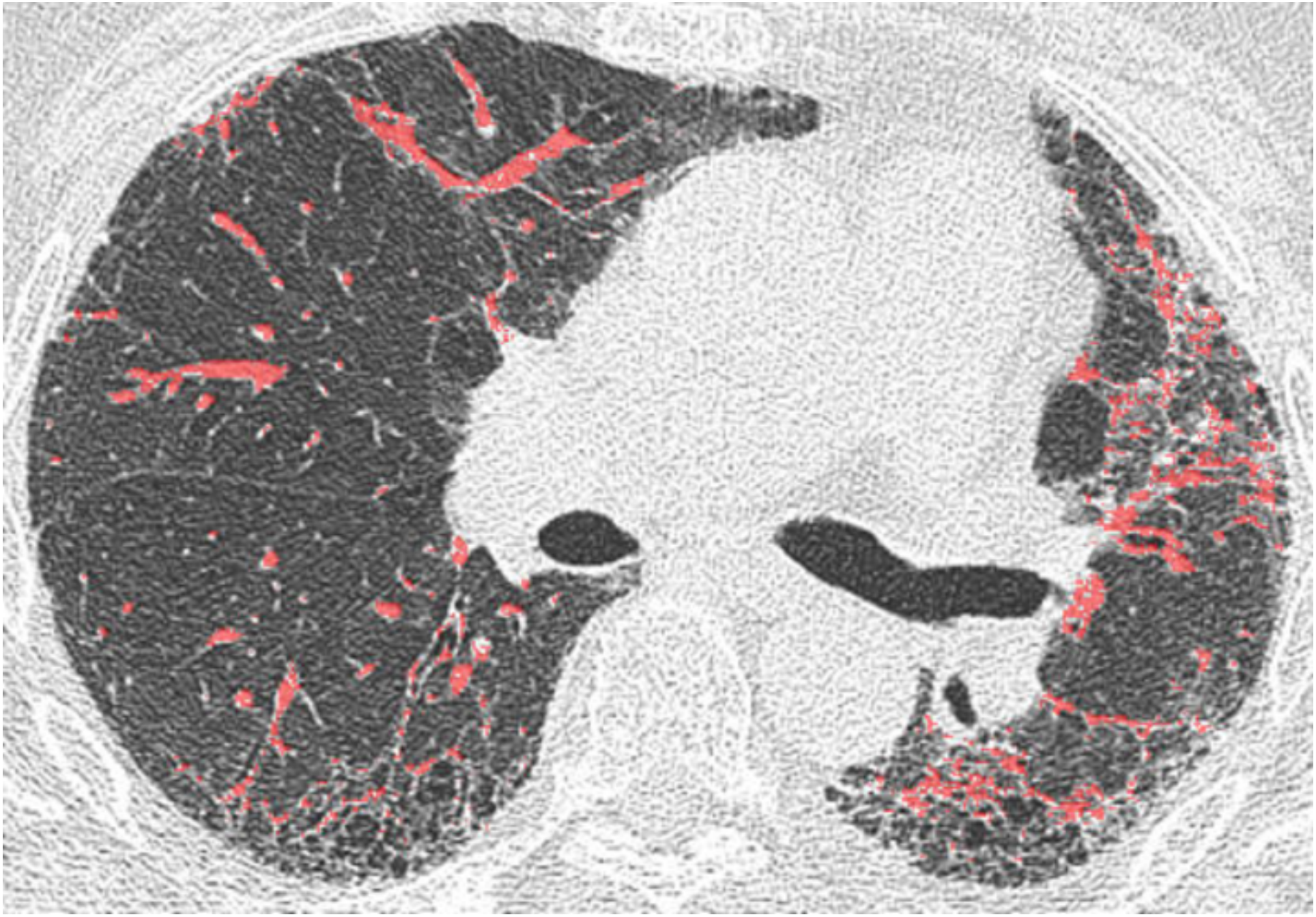

Figure 4

Click here to download Figure Figure 4.tiff

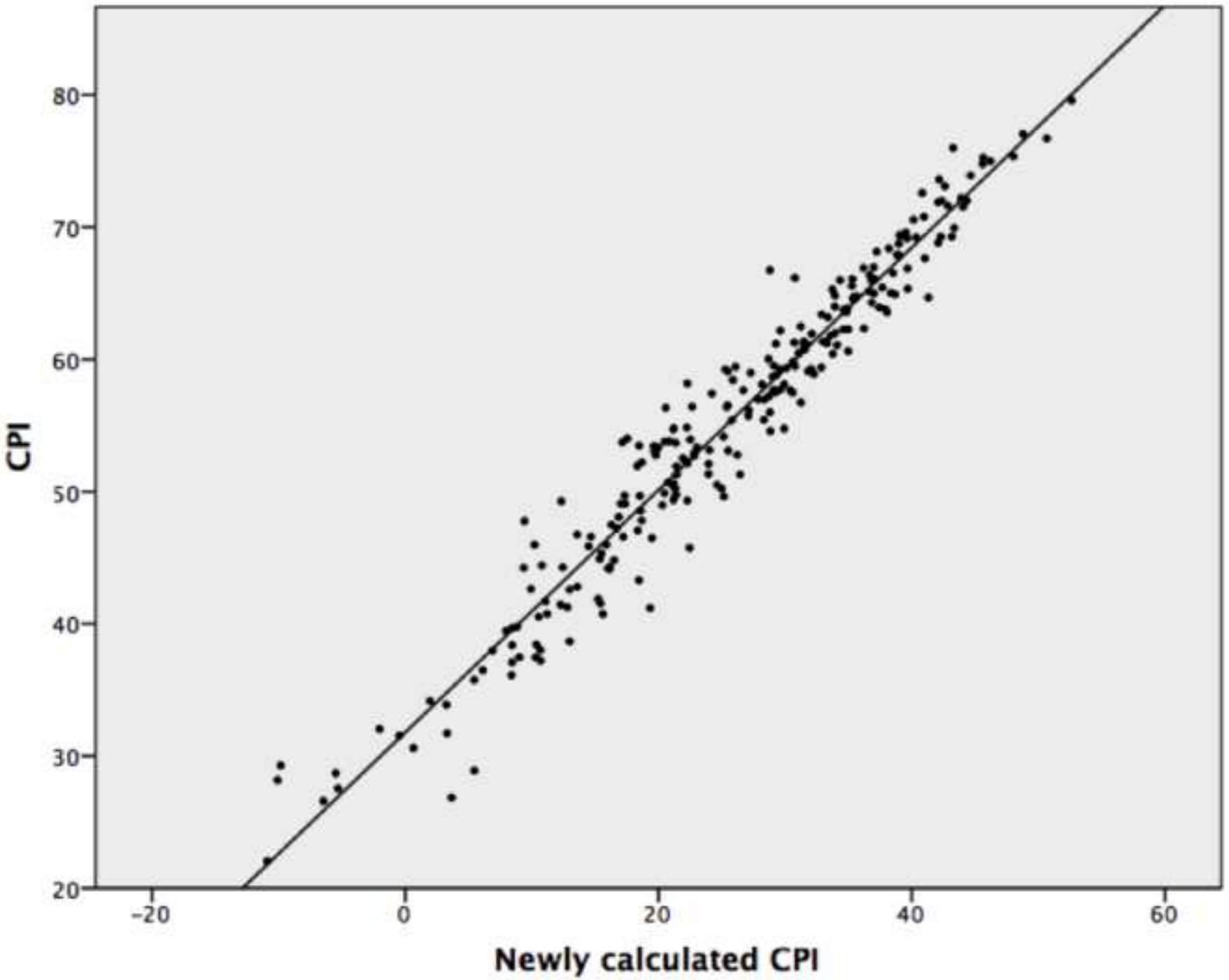

Table 1

Variable (n = 283 unless stated) **Units are percentage unless stated**	**Value**
Median Age (years)	67
Male/female (ratio)	219/64
FEV1 % predicted (n=257)	70·8 ± 19·1
FVC % predicted (n=257)	68·8 ± 20·5
DLco % predicted (n=254)	36·1 ± 12·9
Kco % predicted (n=254)	69·0 ± 19·2
CPI (n=249)	55·1 ± 11·7
Echocardiography RVSP (mmHg) (n=150)	45·1 ± 16·8
CALIPER ILD extent	26·5 ± 18·1
CALIPER Ground glass opacity	17·0 ± 14·7
CALIPER Reticular pattern	8·5 ± 6·0
CALIPER Honeycombing	1·0 ± 1·7
CALIPER Grade 1 decreased attenuation	20·8 ± 20·7
CALIPER Grade 2 decreased attenuation	0·8 ± 2·6
CALIPER Grade 3 decreased attenuation	0·5 ± 2·8
CALIPER pulmonary vessel volume	5·1 ± 1·7
Visual ILD extent	43·1 ± 17·8
Visual Ground glass opacity	10·4 ± 11·4
Visual Reticular pattern	21·7 ± 10·9
Visual Honeycombing	9·8 ± 12·6
Visual Consolidation	1·1 ± 3·3
Visual Emphysema	4·7 ± 10·9
Visual TxBx severity (max score 18)	7·0 ± 3·3
Main pulmonary artery diameter (mm)	30·3 ± 4·8
Ascending aorta diameter (mm)	34·8 ± 4·2

Table 2

Visual CT Variable (n = 283)	Single determination standard deviation
CT Interstitial lung disease extent	7·24
CT Ground glass opacity	6·15
CT Reticular pattern	5·24
CT Honeycombing	7·88
CT Consolidation	2·69
CT Total emphysema	4·99
CT Mosaic attenuation	3·83
CT Traction bronchiectasis severity	1·43

Table 3

	CT Variable	FEV1	FVC	DLco	Kco	CPI
ILD	Visual	0.19, <0.0001	0.27, <0.0001	0.35, <0.0001	0.07, <0.0001	0.44, <0.0001
	CALIPER	0.29, <0.0001	0.41, <0.0001	0.31, <0.0001	NS	0.48, <0.0001
Emphysema	Visual	NS	0.13, <0.0001	0.05, <0.0001	0.26, <0.0001	NS
	CALIPER	NS	0.06, <0.0001	NS	0.08, <0.0001	0.03, 0.004
Vessels	Visual	NS	NS	0.03, 0.008	0.04, 0.001	NS
	CALIPER	0.31, <0.0001	0.45, <0.0001	0.34, <0.0001	NS	0.53, <0.0001

Table 4

	Pulmonary function test	CT Pattern	Beta Coefficient	95% Confidence Interval	P value	Model R^2
Model 1	FEV1	CALIPER Emphysema	-0·98	-1·49, -0·46	0·0003	0·33
		CALIPER ILD extent	-0·59	-0·70, -0·48	<0·0001	
		Visual Emphysema	0.33	0.09, 0.57	0.008	
	FVC	CALIPER ILD extent	-0·67	-0·78, -0·56	<0·0001	0·46
		Visual Emphysema	0·40	0·23, 0·57	<0·0001	
	DLco	CALIPER ILD extent	-0·29	-0·38, -0·20	<0·0001	0·51
		Visual Emphysema	-0·40	-0·50, -0·29	<0·0001	
		Visual ILD extent	-0·27	-0·36, -0·18	<0·0001	
	CPI	CALIPER ILD extent	0·32	0·24, 0·40	<0·0001	0·53
		Visual ILD extent	0·21	0·14, 0·29	<0·0001	
Model 2	FEV1	CALIPER Emphysema	-0·63	-1·01, -0·25	0·001	0·34
		CALIPER PVV	-6·83	-8·02, -5·65	<0·0001	
	FVC	CALIPER PVV	-7·56	-8·68, -6·43	<0·0001	0.48
		Visual Emphysema	0·33	0·17, 0·50	0·0001	
	DLco	CALIPER PVV	-3·81	-4·81, -2·80	<0·0001	0·53
		Visual Emphysema	-0·43	-0·54, -0·33	<0·0001	
		Visual ILD extent	-0·22	-0·31, -0·13	<0·0001	
	CPI	CALIPER PVV	3·88	3.03, 4·74	<0·0001	0·56
		Visual ILD extent	0·18	0·10, 0·25	<0·0001	