Manuscript word count – 3090~~2858~~

Abstract word count - 19~~98~~

Automated computer-based CT stratification as a predictor of outcome in hypersensitivity pneumonitis

Joseph Jacob[1], B.J. Bartholmai[2], S. Rajagopalan[3], R. Karwoski[3], S.M. Mak[1], W. Mok[1], G. Della Casa[1], K. Sugino[1], S.L.F. Walsh[1], A.U. Wells[4], D.M. Hansell[1].

[1]Department of Radiology, Royal Brompton Hospital, Royal Brompton and Harefield NHS Foundation Trust, London, UK.

[2]Division of Radiology, Mayo Clinic Rochester, Rochester, Minnesota, USA.

[3] Biomedical Imaging Resource, Mayo Clinic Rochester, Rochester, Minnesota, USA.

[4]Interstitial Lung Disease Unit, Royal Brompton Hospital, Royal Brompton and Harefield NHS Foundation Trust, London, UK.

Corresponding Author: Joseph Jacob; akelajacob@gmail.com; Tel - 0044 7511033666

Automated computer-based CT stratification as a predictor of outcome in hypersensitivity pneumonitis

ABSTRACT

BACKGROUND: Hypersensitivity pneumonitis (HP) has a variable clinical course. Modelling of quantitative CALIPER-derived CT data can identify distinct disease phenotypes. Mortality prediction using CALIPER analysis was compared to the interstitial lung disease gender, age, physiology (ILD-GAP) outcome model.

METHODS: CALIPER CT analysis of parenchymal patterns in 98 consecutive HP patients was compared to visual CT scoring by two radiologists and functional indices including forced vital capacity (FVC) and diffusion capacity for carbon monoxide (DLco) in univariate and multivariate Cox mortality models. Automated stratification of CALIPER scores was evaluated against outcome models.

RESULTS: Univariate predictors of mortality included visual and CALIPER CT fibrotic patterns, and all functional indices. Multivariate analyses identified only two independent predictors of mortality: CALIPER reticular pattern ($p=0.001$) and DLco ($p<0.0001$).

Automated stratification distinguished three distinct HP groups (log-rank test $p<0.0001$). Substitution of automated stratified groups for FVC and DLco in the ILD-GAP model demonstrated no loss of model strength (C-Index=0.73 for both models). Model strength

improved when automated stratified groups were combined with the ILD-GAP model (C-Index=0.77).

CONCLUSIONS: CALIPER-derived variables are the strongest CT predictors of mortality in HP. Automated CT stratification is equivalent to functional indices in the ILD-GAP model for predicting outcome in HP.

KEY WORDS:

Hypersensitivity pneumonitis

Interstitial lung disease

Computed Tomography

Computer assisted image analysis

Staging

KEY POINTS

Computer CT analysis better predicts mortality than visual CT analysis in HP

Quantitative CT analysis is equivalent to functional indices for prognostication in HP

Prognostication using the ILD-GAP model improves when combined with quantitative CT analysis

ABBREVIATIONS:

ILD	interstitial lung disease
CT	computed tomography
PVV	pulmonary vessel volume
FEV1	forced expiratory volume in one second
FVC	forced vital capacity
DLco	diffusing capacity for carbon monoxide
Kco	Carbon monoxide transfer coefficient
CPI	composite physiologic index
TLC	total lung capacity
RV	residual volume
ILD-GAP	interstitial lung disease gender, age, physiology model
HP	hypersensitivity pneumonitis
CALIPER	Computer-Aided Lung Informatics for Pathology Evaluation and Rating
HRCT	high-resolution computed tomography
DA	decreased attenuation
ANOVA	analysis of variance
PFI	pulmonary function index

HR	hazard ratio
CI	confidence interval
TxBx	traction bronchiectasis
GGO	ground glass opacity
HU	Hounsfield unit

INTRODUCTION

Recently developed sophisticated computer analytic tools such as CALIPER are able to quantify a range of parenchymal pattern extents in a whole lung CT dataset ~~.~~[1]. Considerable ~~morphologic~~ information is generated by such tools, including quantitation of parenchymal patterns that ~~have no~~ cannot be similarly characterised using visual CT analysis, ~~visual~~ such as the volume of intraparenchymal vessels ~~equivalent.~~[2]. To date however, there has been a paucity of computer-based analytic studies in non-idiopathic fibrosing lung diseases such as hypersensitivity pneumonitis (HP). Yet quantitative analysis has the potential to estimate disease burden in a condition such as HP that may present with highly variable morphological [3] and histopathological features [4; 5] ~~features~~.

Traditionally, the evaluation of disease severity in patients with interstitial lung disease (ILD) has focused on the identification of individual CT parenchymal patterns that predict mortality. The result has been the identification of a few key features such as honeycombing [6; 7] and traction bronchiectasis [8; 9] that are associated with a worse outcome.

In tandem with the development of computer quantitation, mathematical modelling has also evolved. New advanced automated computational techniques are able to analyse information contained within an entire CT dataset and identify patient groups that share common or idiosyncratic disease phenotypes ~~.~~[10; 11]. By examining the spectrum of morphological appearances on an individual patients CT, the resulting analysis facilitates individualised medicine. Automated stratification of patients into phenotypically similar

groups has been shown to correlate well with functional indices [12]. However, the association between stratified patient subgroups and differing patient outcomes has not been validated.

In the current study, CT imaging scored visually and by CALIPER in patients with HP was analysed to identify variables predictive of mortality using proportional hazards regression analysis. The population was also stratified using advanced automated computational techniques. Patient subgroups derived using automated stratification were evaluated against mortality and an alternative risk prediction model - the interstitial lung disease gender, age physiology (ILD-GAP) model [13]. The ability of automated stratification to substitute for functional indices in the ILD-GAP model was also explored.

MATERIALS AND METHODS

Study population

A previously defined cohort of consecutive patients diagnosed with subacute and chronic HP was identified using the clinical coding database of the ~~Royal Brompton Hospital~~ for the period January 2000 to December 2006~~.~~ [14]. All patients with a non-contrast, interspaced supine HRCT (1mm sections at 10mm intervals) reconstructed with a Siemens B70 edge-enhancing algorithm were chosen for analysis (Figure 1)[n=98]. Patients were diagnosed by multidisciplinary team clinical, radiological and when available pathological consensus. 68/98 patients (69%) had both a positive exposure history and serum precipitants to relevant antigens. The remainder were diagnosed using broncho-alveolar lavage lymphocytosis results, appropriate CT findings [15; 16] and histopathological confirmation~~.~~ [17; 18]. Eleven cases with concurrent volumetric CT imaging were evaluated with CALIPER to analyse differences in interpretation of interspaced and volumetric scans (Figure 1, Supplementary appendix). Approval for this study of clinically indicated CT and pulmonary function index (PFI) data was obtained from the Institutional Ethics Committee of the ~~Royal Brompton Hospital~~ and the Institutional Review Board of the ~~Mayo Clinic~~.

CT, CALIPER and PFI protocols

Protocols are described in detail in the supplementary appendix. PFI's analysed included forced expiratory volume in one second (FEV1), forced vital capacity (FVC), total lung capacity (TLC), residual volume (RV), transfer coefficient of the lung for carbon monoxide

(Kco), single breath carbon monoxide diffusing capacity corrected for hemoglobin concentration (DLco) and the composite physiologic index (CPI)~~.~~ [19].

CT evaluation

Visual CT evaluation on a lobar basis was performed independently by two radiologists (__~~MM~~, __~~WM~~) each with 3 years thoracic imaging experience, blinded to all clinical information. CT parameters evaluated included: ground glass opacity, reticular pattern, honeycombing, and consolidation which were summed to calculate total ILD extent~~.~~ [2]. Total fibrosis extent represented the sum of reticular pattern and honeycombing. Emphysema, mosaicism (decreased attenuation component) and traction bronchiectasis were scored as previously described ~~.~~ [2].

CALIPER CT evaluation

CALIPER segmented the lung into 6 zones: right and left upper, middle and lower zones, demarcated with reference to the lung hilum, with each of the 6 zones further divided into an inner half (central region) and outer half (peripheral region). 8 initial CT patterns were classified by CALIPER,~~(described below) were evaluated for all the 12 areas (see supplementary appendix).~~ however preliminary analyses demonstrated that honeycombing was not identified as a distinct pattern on CALIPER evaluation of interspaced datasets as honeycombing requires three-dimensional information for characterisation. Consequently, CALIPER honeycombing was not further analysed in the study. In the final analysis therefore,

seven CT patterns (described below) were evaluated for all 12 anatomic areas of the lung (see supplementary appendix).

CALIPER evaluation of the lungs was pictorially expressed as volume rendered three-dimensional images or as a glyph (Figure 2). Each glyph comprised six wedges, or zones, the size of which reflected the volume of the zone relative to the total lung volume. Within each lung zone, every voxel was classified into one of 8 separately colour coded CALIPER parenchymal patterns: ground glass opacity=yellow, reticular pattern=orange, honeycombing=brown, Grade 1 decreased attenuation (DA)=light green, Grade 2 DA=light blue, Grade 3 DA=dark blue, Normal lung=dark green, pulmonary vessel volume (PVV)[pulmonary arteries and veins, excluding vessels at the lung hilum]=white. The relative volumes of the patterns within a zone determined the proportions of each colour in a zone.

~~Honeycombing was not identified as a distinct pattern on CALIPER evaluation of interspaced datasets as honeycombing requires three-dimensional information for characterisation. Consequently, CALIPER honeycombing was not further analysed.~~ All CT variables were expressed as a percentage of the total lung volume. CALIPER grade~~type~~ 2 and 3 DA lung were corresponded to areas of emphysema~~,~~ [2], whilst ILD extent represented the sum of ground glass opacity and reticular pattern.

Automated stratification of CALIPER-variables:

Global and regional dissimilarities in distribution of the 8 patterns quantified by CALIPER within 12 zones were evaluated by a dissimilarity metric as previously described [12]. Regional dissimilarities were discerned using 3 components. Within a single lung, differences in regional lung volume as a proportion of the total lung volume were calculated. Between any two lungs, dissimilarities in the proportions of absolute lung volumes in corresponding regions and dissimilarities in the proportions of specific parenchymal patterns in the corresponding regions were calculated.

The dissimilarity metric was used to compare all 98 HP cases in a pairwise manner and the resultant 98x98 matrix was stratified using unsupervised affinity propagation [20] to identify unique clusters representing patient groups with shared parenchymal characteristics. An a priori specification of the number of expected clusters was not imposed, as affinity propagation derives naturally occurring clusters using real-valued message exchange [20].

Statistical analysis:

Data are given as means with standard deviations, or numbers of patients with percentages where appropriate. Interobserver variation for the visual scores was calculated using the single determination standard deviation [21]. CALIPER analysis of 11 interspaced and volumetric CTs was compared using the independent samples T test, (significance =$p<0.05$). Univariate and multivariate Cox regression analyses were used to investigate relationships within and between: CALIPER and visual CT evaluation and PFIs. Linear regression analyses

were used to characterise relationships between cardinal pulmonary function indices and CT scores of ground glass opacity and reticular pattern.

Comparisons of functional and morphological indices between automated stratified groups were examined using one-way analysis of variance (ANOVA) and post-ANOVA pairwise T test analyses using the Bonferroni correction for multiple analyses. Cox regression analysis and Kaplan Meier survival curves were used to identify survival differences between automated stratification results and the ILD-GAP staging system. Survival distributions were compared using the Log rank test, and bootstrapped with 1000 randomly generated samples. The automated stratified groups were then substituted for PFIs in the ILD-GAP staging system resulting in the creation of a Stratified-CT model containing the following weighted variables: automated stratified group score, patient age and gender. Finally the automated stratified groups were combined with the ILD-GAP model to form a third final model termed the Stratified-GAP model. Model strength for the ILD-GAP, Stratified-CT and Stratified-GAP models was compared using receiver operating characteristic (ROC) curves, and measuring the area under the ROC curve (AUROCC) and Harrells C-Index [22]. Statistical analyses were performed with STATA (version 12, StatCorp, College Station, TX, USA).

RESULTS

Baseline results

The median age of the cohort was 59 years, with 52% having died during the average follow up period of 69 months. Demographic data and average visual and CALIPER CT scores and PFI data are provided in Table 1. Interobserver variation for the visual scores is demonstrated in Supplementary Table 1. On average, visual scores identified more ILD than CALIPER. ILD mainly compromised reticular opacities on visual scoring as opposed to ground glass opacities as scored by CALIPER. CALIPER scores for ground glass opacity and reticular pattern correlated more strongly with cardinal pulmonary function indices than equivalent visual CT scores (Supplementary Table 3).

~~Honeycombing was not identified as a distinct pattern on CALIPER evaluation of interspaced datasets as honeycombing requires three-dimensional information for characterisation. Consequently, CALIPER honeycombing was not further analysed.~~ Evaluation of the 11 cases with concurrent interspaced and volumetric CT imaging demonstrated a significant difference in PVV extent between groups (p=0.03), with CALIPER classifying more vessels on interspaced images than volumetric scans (Figure 2)[Figure 1, Supplementary appendix].

Mortality analyses

On univariate visual CT analysis reticular pattern, honeycombing, mosaicism and traction bronchiectasis were strongly predictive of mortality. All CALIPER-scored patterns except

emphysema and all PFIs were predictive of mortality on univariate analysis (Table 2). The results were maintained after adjusting for patient age and gender (Supplementary Table 2). When visual parameters alone were analysed in a multivariate model, reticular pattern (HR=1.05, CI 1.02-1.08, p=0.001) and ILD extent (HR=1.03, CI 1.01-1.04, p=0.01) were independent predictors of mortality. Reticular pattern (HR=1.10, CI 1.02-1.18, p=0.02) and PVV (HR=1.08, CI 1.01-1.15, p=0.02) were independently predictive of mortality on multivariate analysis of CALIPER variables. Of the PFIs, DLco alone best described mortality on multivariate analysis. When CALIPER, visual and PFI variables were analysed together in a multivariate model, DLco and CALIPER reticular pattern were the only two independent predictors of mortality (Table 2). The results were maintained after adjusting for patient age and gender (Supplementary Table 2).

Automated stratification of CALIPER-derived groups:

Automated stratification of HP patients identified 3 distinct groups with similar distributions of CALIPER parenchymal patterns within each group (represented as glyphs in Figure 3). With progression from group 1 to 3, the proportion of the lung comprised of reticular pattern, ground glass density and PVV increased, while the extent of normal lung and grade 1 DA decreased (Table 1). Mean PFIs also worsened with progression from group 1 to 3 (Table 1).

Significant functional differences across all groups were identified with FVC, TLC and CPI, with differences in at least two groups identified with FEV1 and DLco (~~Figure 4 and~~ Table 3).

Visual CT parameters demonstrated good separation between group 1 and groups 2 and 3 for parenchymal patterns indicative of fibrosis, and for mosaicism. However only ILD and consolidation extents identified group separations between groups 2 and 3. CALIPER scores for patterns indicating ILD and PVV demonstrated clear differences across all three automated stratified groups. When the eleven cases with concurrent interspaced and volumetric imaging were analysed by CALIPER and stratified, the volumetric cases mapped to the same outcome groups as the interspaced cases (Figure 1, Supplementary appendix).

Survival distributions between automated stratified groups are demonstrated in Figure 45a (p<0.0001 Log rank test). Cox regression analysis demonstrated that separation of patients into automated stratified groups was strongly predictive of mortality (HR=2.74, CI 1.86-4.05, p<0.0001). A mortality effect from automated stratification was maintained following correction for age, gender and baseline disease severity using the CPI (group stratification = HR 1.95, CI 1.15-3.29, p=0.01) and DLco (group stratification = HR 2.05, CI 1.23-2.41, p=0.006). When automated stratified groups were evaluated against DLco tertiles in a Cox proportional Hazards analysis, both DLco tertiles (p=0.001) and automated stratified groups (p=0.002) were equivalent in their ability to predict outcome following bootstrapping of 1000 samples.

The ILD-GAP model separated patients according to age, gender, FVC and DLco values (scored on a nine point scale) into four outcome groups (Figure 45b). A bivariate Cox regression analysis, bootstrapped with 1000 samples demonstrated no difference between

the ILD-GAP model (p=0.002, CI 0.34-1.13) and automated stratification groups (p=0.001, CI 0.36-1.22).

To compare the ability of the automated stratified groups to substitute for FVC and DLco in the ILD-GAP model, the three-point automated stratified group scale was converted into a five-point scale analogous to the five point scale for FVC (0-2) and DLco (0-3) in the ILD-GAP model. Automated stratified group 1 was converted to a score of 0, group 2 remained unchanged, whilst group 3 was converted to a score of 4. When the five-point automated stratified group scale was combined with patient age and gender (weighted on a three (0-2) and two-point (0-1) scale respectively, in accordance with the ILD-GAP model) an eight point scale was derived and converted into a four point automated stratified model (Stratified-CT model) using the same group divisions as the ILD-GAP model (scores of 0-1=1; scores of 2-3=2; scores of 4-5=3; scores >5=4) with good separation of outcome groups (Figure 4~~5~~c; Log rank test p<0.0001).

The ordinal predictive power of the four-point Stratified-CT model was 0.73 as judged by Harrells C-Index, which was identical to the C-Index value of the ILD-GAP model (0.73). When the three-point scale of the automated stratified groups were combined with the four-point scale of the ILD-GAP model, the new Stratified-GAP model had a C-Index value of 0.77. Sensitivity and specificity for mortality prediction using the three models are demonstrated using ROC curves analysis in Figure 5~~6~~, and was greatest with the Stratified-GAP model.

DISCUSSION

In the current study, a computer derived variable, CALIPER reticular pattern was stronger than all visual CT scores at predicting mortality in patients with HP. The current study has demonstrated for the first time that in conjunction with computer quantitation, automated stratification techniques can separate HP patients into prognostic groups that are functionally distinct and comparable to the ILD-GAP model for risk prediction. When automated stratification is combined with functional indices, patient age and gender, the resulting model is stronger than the ILD-GAP model alone at predicting mortality. Furthermore, survival across groups defined using automated stratification remains independent of baseline disease severity.

Automated stratification identified 3 unique HP groups that had distinct functional characteristics. The recognition of disease sub-groups is clinically desirable if it allows identification of high-risk patients that may benefit from aggressive interventions, yet limits unnecessary treatment in patients with quiescent disease~~.~~ [23]. Computer analysis of CT imaging is attractive as a tool to generate repeatable and reproducible information across disease cohorts given its lack of interobserver variation and reproducibility. Furthermore, quantitative tools such as CALIPER can evaluate the entirety of a CT dataset providing a comprehensive analysis of an individual patient's CT. Computer analysis has been used in emphysema cohorts with the aim of identifying distinct phenotypic groups [12; 24] with limited success ~~,~~[25], but similar studies in individual ILD populations are sparse~~.~~ [26].

The benefits of combining CT quantitation with automated stratification techniques lie in the emphasis on data to drive the identification of distinct phenotypic groups which can then be assessed to identify functional and prognostic similarities. The automated stratified groups in the current study were generated without any innate bias. The numbers and types of phenotypic clusters were not predetermined but were generated by the automated stratification process itself. Consequently, phenotypic features that might be overlooked with visual scores can be identified, and may be used to uncover populations with shared outcomes.

The comparable strength in risk stratification between the ILD-GAP model and the automated stratification model suggests that computer analysis and automated modelling could have a role as outcome measures in clinical trials. For example discrete therapeutic responses or adverse reactions may become apparent, with a sensitivity potentially surpassing functional indices alone.

Against the complexity of stratified mathematical modelling, the simplicity of a glyph representation translates dense numerical datapoints into a format with clinical pertinence. Patients can use a glyph to understand the nature and extent of their particular disease. For clinicians meanwhile, the combination of a glyph mapped to a stratified group allows the characterisation of a patients disease phenotype at a glance in a busy clinic setting.

The difference in quantitation of ground glass opacity and reticular pattern between visual and CALIPER scores was largely secondary to differences in classification of a pattern where ground glass opacity is overlaid by reticular pattern. Visual scorers often disagree on such a pattern, and in the current study the visual scorers considered the pattern to represent reticular pattern, whilst as previously described [2], CALIPER classifies such a pattern as ground glass opacity.

The prognostic implications of an increasing CT reticular pattern extent identified in the current analysis confirms findings from a previous HP study~~,~~ [27]. Fibrosis extent [27-30] and the decreased attenuation component of a mosaic attenuation pattern~~,~~ [27], also demonstrated in the current study, have similarly been previously implicated as prognostic variables in HP. Conversely however, CALIPER grade 1 DA which partly corresponds to air trapping on CT, demonstrated a mild protective effect on univariate analysis. Traction bronchiectasis, in contrast to previous studies in HP [14] was not found to be a powerful multivariate predictor of mortality.

There were limitations to the current study. CALIPER was unable to classify honeycombing on interspaced imaging. Nevertheless, despite the loss of a strong prognostic variable, a CALIPER variable, reticular pattern, remained an independent predictor of mortality across the entire HP cohort. PVV was over-represented on interspaced imaging secondary to its requirement, like honeycombing, for three-dimensional patterns for optimal characterization. PVV is recognized by CALIPER as contiguous tubes followed to the lung

edge using tubular filters and three-dimensional region growing software. The PVV signal in the current study may have related to linear tubular structures that were predominantly "in plane" on a single interspaced image. However it is also possible that some of the characterized vessels may have represented reticular densities that were in fact misclassified as vessels. Nevertheless, whilst there appears to be a degree of overlap in CALIPER scoring of reticular pattern and vessels, both factors remained independent predictors of mortality when CALIPER variables were analysed. A final limitation lies in the lack of an external validation cohort with which to confirm the study findings. The scarcity of large HP populations even within tertiary referral centres made validation of our results challenging, but remains an important aim for future studies.

In conclusion, we have shown for the first time that computer quantitation and automated stratification of CTs (by CALIPER) generate variables that are powerfully predictive of mortality in HP. Automated stratification is able to distinguish patients with differing disease phenotypes that correspond to discreet functional groups, and is equivalent to functional indices in the ILD-GAP model in their ability to risk stratify patients with hypersensitivity pneumonitis. When combined with quantitative CT analysis, prognostication using the ILD-GAP model was found to improve.

Formatted: Font: Not Bold

Formatted: Font: Not Bold

REFERENCES

1 Maldonado F, Moua T, Rajagopalan S, Karwoski RA, Raghunath S, Decker PA, et al. (2014) Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. European Respiratory Journal 43:204-212

2 Jacob J, Bartholmai B, Rajagopalan S, Kokosi M, Nair A, Karwoski R, et al. (2016) Automated quantitative CT versus visual CT scoring in idiopathic pulmonary fibrosis: validation against pulmoanry function. Journal of Thoracic Imaging 31:304-311

3 Gaxiola M, Buendía-Roldán I, Mejía M, Carrillo G, Estrada A, Navarro MC, et al. (2011) Morphologic diversity of chronic pigeon breeders disease: Clinical features and survival. Respiratory Medicine 105:608-614

4 Morell F, Villar A, Montero M-I, Muñoz X, Colby TV, Pipvath S, et al. (2013) Chronic hypersensitivity pneumonitis in patients diagnosed with idiopathic pulmonary fibrosis: a prospective case-cohort study. Lancet Respir Med 1:685-694

5 Elicker BM, Jones KD, Henry TS and Collard HR (2016) Multidisciplinary Approach to Hypersensitivity Pneumonitis. Journal of Thoracic Imaging 31:92-103

6 Flaherty KR, Mumford JA, Murray S, Kazerooni EA, Gross BH, Colby TV, et al. (2003) Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. AmJRespirCrit Care Med 168:543-548

7 Nagao T, Nagai S, Hiramoto Y, Hamada K, Shigematsu M, Hayashi M, et al. (2002) Serial evaluation of high-resolution computed tomography findings in patients with idiopathic pulmonary fibrosis in usual interstitial pneumonia. Respiration 69:413-419

8 Sumikawa H, Johkoh T, Colby TV, Ichikado K, Suga M, Taniguchi H, et al. (2008) Computed tomography findings in pathological usual interstitial pneumonia: relationship to survival. AmJRespirCrit Care Med 177:433-439

9 Edey AJ, Devaraj AA, Barker RP, Nicholson AG, Wells AU and Hansell DM (2011) Fibrotic idiopathic interstitial pneumonias: HRCT findings that predict mortality. European Radiology 21:1586-1593

10 Raghunath S, Rajagopalan S, Karwoski RA, Bartholmai BJ and Robb RA (2011) Referenceless stratification of parenchymal lung abnormalities. MedImage ComputComputAssistInterv 14:223-230

11 Bartholmai BJ, Raghunath S, Karwoski RA, Moua T, Rajagopalan S, Maldonado F, et al. (2013) Quantitative CT imaging of interstitial lung diseases. Journal of Thoracic Imaging 28:298-307

12 Raghunath S, Rajagopalan S, Karwoski A, Maldonado F, Peikert T, Moua T, et al. (2014) Quantitative Stratification of Diffuse Parenchymal Lung Diseases. PloS One 9:e93229

13 Ryerson CJ, Vittinghoff E, Ley B, Lee JS, Mooney JJ, Jones KD, et al. (2014) Predicting survival across chronic interstitial lung disease: The ild-gap model. CHEST Journal 145:723-728

14 Walsh SL, Sverzellati N, Devaraj A, Wells AU and Hansell DM (2012) Chronic hypersensitivity pneumonitis: high resolution computed tomography patterns and pulmonary function indices as prognostic determinants. European Radiology 22:1672-1679

15 Lynch DA, Newell JD, Logan PM, King TE, Jr. and Muller NL (1995) Can CT distinguish hypersensitivity pneumonitis from idiopathic pulmonary fibrosis? AJR 4:807-811

16 Adler BD, Padley SP, Muller NL, Remy-Jardin M and Remy J (1992) Chronic hypersensitivity pneumonitis: high-resolution CT and radiographic features in 16 patients. Radiology 185:91-95

17 Coleman AC and Colby TV (1988) Histologic diagnosis of extrinsic allergic alveolitis. American Journal of Surgical Pathology 2:514-518

18 Reyes CN, Wenzel FJ, Lawton BR and Emanuel DA (1982) The pulmonary pathology of farmer's lung. Chest 81:142-146

19 Wells AU, Desai SR, Rubens MB, Goh NS, Cramer D, Nicholson AG, et al. (2003) Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. American Journal of Respiratory and Critical Care Medicine 167:962-969

20 Frey BJ and Dueck D (2007) Clustering by Passing Messages Between Data Points. Science 315:972-976

21 Chinn S (1991) Repeatability and method comparison. Thorax 46:454-456

22 Uno H, Cai T, Pencina MJ, D'Agostino RB and Wei LJ (2011) On the C-statistics for Evaluating Overall Adequacy of Risk Prediction Procedures with Censored Survival Data. Statistics in Medicine 30:1105-1117

23 Han MK, Agusti A, Calverley PM, Celli BR, Criner G, Curtis JL, et al. (2010) Chronic Obstructive Pulmonary Disease Phenotypes. American Journal of Respiratory and Critical Care Medicine 182:598-604

24 Castaldi PJ, San José Estépar R, Mendoza CS, Hersh CP, Laird N, Crapo JD, et al. (2013) Distinct Quantitative Computed Tomography Emphysema Patterns Are Associated with Physiology and Function in Smokers. American Journal of Respiratory and Critical Care Medicine 188:1083-1090

25 Dirksen A and MacNee W (2013) The Search for Distinct and Clinically Useful Phenotypes in Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care Medicine 188:1045-1046

26 Choi SH, Lee HY, Lee KS, Chung MP, Kwon OJ, Han J, et al. (2014) The Value of CT for Disease Detection and Prognosis Determination in Combined Pulmonary Fibrosis and Emphysema (CPFE). PloS One 9:e107476

27 Lima MS, Coletta ENAM, Ferreira RG, Jasinowodolinski D, Arakaki JSO, Rodrigues SCS, et al. (2009) Subacute and chronic hypersensitivity pneumonitis: Histopathological patterns and survival. Respiratory Medicine 103:508-515

28 Mooney JJ, Elicker BM, Urbania TH, Agarwal MR, Ryerson CJ, Nguyen MLT, et al. (2013) Radiographic fibrosis score predicts survival in hypersensitivity pneumonitis. Chest 144:586-592

29 Vourlekis JS, Schwarz MI, Cherniack RM, Curran-Everett D, Cool CD, Tuder RM, et al. (2004) The effect of pulmonary fibrosis on survival in patients with hypersensitivity pneumonitis. Am J Med 116:662-668

30 Perez-Padilla R, Salas J, Chapel R, Sanchez M, Carrillo G, Perez R, et al. (1993) Mortality in Mexican patients with chronic pigeon breeder's lung compared to those with usual interstitial pneumonia. American Review of Respiratory Disease 148

Variable (Units are percentage unless stated)	All HP cases (n=98 unless stated)	Stratified Groups Group 1 (n=33)	Group 2 (n=40)	Group 3 (n=25)
Median Age (years)	59	57	61	56
Male/female	38/59	11/21	19/21	8/17
Survival (alive/dead)	47/51	26/7	15/25	6/19
Follow up time (months)	69.1 ± 43.3	96.9 ± 30.5	63.0 ± 43.4	42.0 ± 37.0
FEV1 % predicted (n=98)	68.7 ± 22.6	81.1 ± 20.5	70.8 ± 19.6	48.7 ± 15.5
FVC % predicted (n=98)	69.9 ± 24.6	85.5 ± 21.4	71.7 ± 20.5	46.7 ± 15.5
DLco % predicted (n=95)	41.8 ± 18.2	55.5 ± 15.9	37.6 ± 15.9	28.6 ± 11.2
Kco % predicted (n=95)	67.6 ± 19.0	74.9 ± 14.2	63.4 ± 20.3	64.6 ± 20.6
TLC% predicted (n=96)	72.1 ± 18.1	83.9 ± 15.3	71.5 ± 15.6	56.3 ± 13.1
RV% predicted (n=96)	83.7 ± 26.0	93.0 ± 27.5	81.2 ± 23.9	74.9 ± 24.3
CPI (n=95)	49.9 ± 16.6	37.2 ± 14.8	52.6 ± 13.4	64.8 ± 10.1
CALIPER ILD extent	24.3 ± 23.5	_3.6 ± 2.5	20.0 ± 11.0	58.5 ± 13.4
CALIPER GGO	16.6 ± 19.7	_1.0 ± 1.0	12.4 ± 9.9	44.0 ± 15.8
CALIPER Reticular pattern	_7.7 ± 5.4	_2.6 ± 1.7	_7.6 ± 2.8	14.4 ± 4.5
CALIPER Emphysema	_0.5 ± 1.2	_0.2 ± 0.5	_0.6 ± 1.6	_0.2 ± 0.6
Grade 1 DA	_0.5 ± 1.2	24.6 ± 25.6	_8.0 ± 10.7	_0.4 ± 0.7
CALIPER PVV	_7.8 ± 5.7	_2.8 ± 1.5	_7.7 ± 2.9	14.5 ± 5.9
CALIPER Normal lung	55.9 ± 24.1	68.7 ± 24.5	63.7 ± 9.4	26.5 ± 13.7
Visual ILD extent	33.5 ± 20.2	16.5 ± 12.4	37.8 ± 18.0	49.2 ± 15.1
Visual fibrosis extent	18.8 ± 14.7	8.3 ± 6.7	23.3 ± 14.9	25.4 ± 15.0
Visual GGO	9.2 ± 7.9	5.3 ± 6.7	9.5 ± 6.9	13.8 ± 8.4
Visual Reticular pattern	15.0 ± 11.2	7.8 ± 5.7	17.6 ± 11.3	20.2 ± 12.0
Visual Honeycombing	3.8 ± 5.5	0.5 ± 1.3	5.7 ± 6.6	5.2 ± 5.2
Visual Consolidation	7.6 ± 8.3	3.9 ± 6.5	7.1 ± 7.3	13.3 ± 9.1
Visual Mosaicism	17.2 ± 10.6	11.2 ± 9.1	19.6 ± 11.1	21.3 ± 8.3
Visual Emphysema	2.3 ± 6.4	1.2 ± 3.4	3.4 ± 9.1	2.0 ± 3.7
Visual TxBx (max score 18)	5.4 ± 1.9	4.3 ± 1.7	6.0 ± 1.7	6.0 ± 1.8

Formatted Table

Table 1. Patient age, gender and measures of pulmonary function indices and CALIPER scored CT parameters. Data represent mean values with standard deviations. 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, RV=residual volume, CPI=composite physiologic index, ILD=interstitial lung

disease, GGO=ground glass opacity, DA=decreased attenuation, PVV=pulmonary vessel volume, TxBx=traction bronchiectasis.

	Number of patients	Hazard ratio	P Value	95.0% Confidence Interval	
				Lower	Upper
CALIPER score					
Total ILD extent	98	1.02	0.001	1.01	1.03
Ground glass opacity	98	1.02	0.008	1.00	1.03
Reticular pattern	98	1.16	<0.0001	1.10	1.23
Emphysema	98		NS		
Grade 1 DA	98	0.98	0.03	0.96	1.00
Normal lung	98	0.98	<0.0001	0.97	0.99
PVV	98	1.14	<0.0001	1.09	1.19
Pulmonary Function					
FEV1 % predicted	987	0.98	0.001	0.96	0.99
FVC % predicted	987	0.97	<0.0001	0.96	0.99
TLC % predicted	965	0.96	<0.0001	0.94	0.98
RV % predicted	96	0.98	0.004	0.97	0.99
DLco % predicted	954	0.94	<0.0001	0.92	0.96
Kco % predicted	954	0.98	0.001	0.96	0.99
CPI % predicted	954	1.06	<0.0001	1.04	1.09
VISUAL score					
ILD extent	98	1.04	<0.0001	1.02	1.06
Fibrosis extent	98	1.06	<0.0001	1.04	1.08
Ground glass opacity	98		NS		
Reticular pattern	98	1.07	<0.0001	1.04	1.09
Honeycombing	98	1.07	<0.0001	1.04	1.09
Consolidation	98		NS		
Total emphysema	98		NS		
Mosaicism	98	1.04	0.001	1.02	1.07
TxBx severity	98	1.49	<0.0001	1.24	1.77
MULTIVARIATE MODEL					
CALIPER Reticular pattern		1.12	0.001	1.04	1.19
DLco % predicted		0.95	<0.0001	0.93	0.97

Table 2. Univariate Cox regression analysis demonstrating mortality according to CALIPER indices (top white), pulmonary function indices (light grey) and visual CT scores (dark grey). A multivariate model evaluated CALIPER and pulmonary function indices (lower white). 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, RV=residual volume, CPI=composite physiologic index, ILD=interstitial lung disease, DA=decreased attenuation, PVV = pulmonary vessel volume, TxBx=traction bronchiectasis.

	Group 1 vs 2	Group 1 vs 3	Groups 2 vs 3
ILD extent	< 0.0001	< 0.0001	< 0.0001
GGO	< 0.0001	< 0.0001	< 0.0001
Reticular pattern	< 0.0001	< 0.0001	< 0.0001
Normal lung	0.72*	< 0.0001	< 0.0001
Emphysema	0.56*	1.00*	0.53*
PVV	< 0.0001	< 0.0001	< 0.0001
FEV1 %	0.09*	< 0.0001	< 0.0001
FVC %	0.02	< 0.0001	< 0.0001
TLC %	0.003	< 0.0001	0.0006
RV%	0.16*	0.04	0.95*
Kco %	0.02	0.09*	1.00*
DLco %	< 0.0001	< 0.0001	0.06*
CPI	< 0.0001	< 0.0001	0.001
ILD extent	< 0.0001	< 0.0001	0.03
Fibrosis extent	< 0.0001	< 0.0001	1.00*
GGO	0.03	< 0.0001	0.08*
Reticular pattern	< 0.0001	< 0.0001	1.00*
Honeycombing	< 0.0001	< 0.0001	1.00*
Consolidation	0.17*	< 0.0001	0.01
Emphysema	0.54*	1.00*	1.00*
Mosaicism	0.003	< 0.0001	1.00*
Traction Bronchiectasis	< 0.0001	0.001	1.00*

Table 3. Functional and morphological differences between stratified groups, evaluated using pairwise T tests, following one-way ANOVA analysis and modified using the Bonferroni correction. CALIPER scores (white); Pulmonary function indices (light grey); Visual CT scores (dark grey). 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, RV=residual volume, CPI=composite physiologic index, ILD=interstitial lung disease, GGO=ground glass opacity, PVV = pulmonary vessel volume, *=not significant.

FIGURE LEGENDS:

Figure 1. CONSORT diagram illustrating the selection of patients for the final study population. HP = hypersensitivity pneumonitis, HRCT = high resolution computed tomography, CT = computed tomography.

Figure 2. Coronal three-dimensional rendering, ~~and~~ accompanying glyph of parenchymal patterns scored by CALIPER ~~using interspaced imaging~~ and corresponding colour overlay axial images at the lung bases using interspaced imaging (top images) and volumetric imaging (lower images). ~~The~~ Within the glyph the dark line separates right and left lungs and concentric circles overlying the glyph represented quintiles of lung volume. Examination of the glyph generated by the volumetric dataset demonstrates a shrunken left lung and right lower lobe both of which contain honeycombing. A greater volume of decreased attenuation lung is also evident in the right middle lobe when compared to the glyph generated from interspaced data. Dark green=normal lung, light green=grade 1 decreased attenuation, light and dark blue=emphysema, yellow=ground glass opacity, orange=reticular pattern, brown=honeycombing, white=pulmonary vessel volume.

Figure 3. Glyphs demonstrating the compositions of the three hypersensitivity pneumonitis groups (Group 1= left, group 2=middle, ~~group~~ 3=right) derived following CALIPER CT analysis. Dark green=normal lung, light green=grade 1 decreased attenuation, light and dark blue=emphysema, yellow=ground glass opacity, orange=reticular pattern, brown=honeycombing, white=pulmonary vessel volume.

~~Figure 4. Box and whisker plots demonstrating the functional relationships of the three hypersensitivity pneumonitis groups (G1, G2 and G3) stratified on the basis of CALIPER CT parenchymal pattern extents. 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, RV=residual volume, CPI=composite physiologic index.~~

Figure 45a. Kaplan Meier survival curves for the three hypersensitivity pneumonitis groups derived using automated stratification on the basis of CALIPER CT parenchymal pattern extents. Group 1 (blue) mean survival 120.4±4.9 months; n=33, group 2 (green) mean survival 74.6±8.0 months; n=40 and group 3 (yellow) mean survival 45.1±7.7 months, n=25. Log rank test p<0.0001.

Figure 45b. Kaplan Meier survival curves for patients with hypersensitivity pneumonitis stratified on the basis of the ILD-GAP model. Group 1 (blue) mean survival 122.6±5.1 months; n=21, group 2 (green) mean survival 98.3±7.3 months; n=35; Group 3 (yellow) mean survival 53.4±8.0 months; n=29; Group 4 (magenta) mean survival 46.1±15.6 months, n=10. Log rank test p<0.0001.

Figure 45c. Kaplan Meier survival curves for patients with hypersensitivity pneumonitis stratified on the basis of the Stratified-CT model combining automated stratified groups, patient age and gender. Group 1 (blue) mean survival 120.4±5.7 months; n=23, group 2 (green) mean survival 95.1±8.5 months; n=33; Group 3 (yellow) mean survival 66.0±8.5 months; n=32; Group 4 (magenta) mean survival 15.0±6.5 months, n=7. Log rank test p<0.0001.

Figure 5~~6~~. Receiver operating characteristic curves demonstrating sensitivity and specificity for mortality prediction using three models: ILD-GAP (blue, area under curve (AUC)=0.72, Confidence Interval (CI) 0.61-0.82, p=0.0002); Stratified-CT (green, area under curve (AUC)=0.76, CI 0.66-0.85, p<0.0001); Stratified-GAP (yellow, area under curve (AUC)=0.77, CI 0.6

Figure 1

Click here to download Figure Slide1.tiff

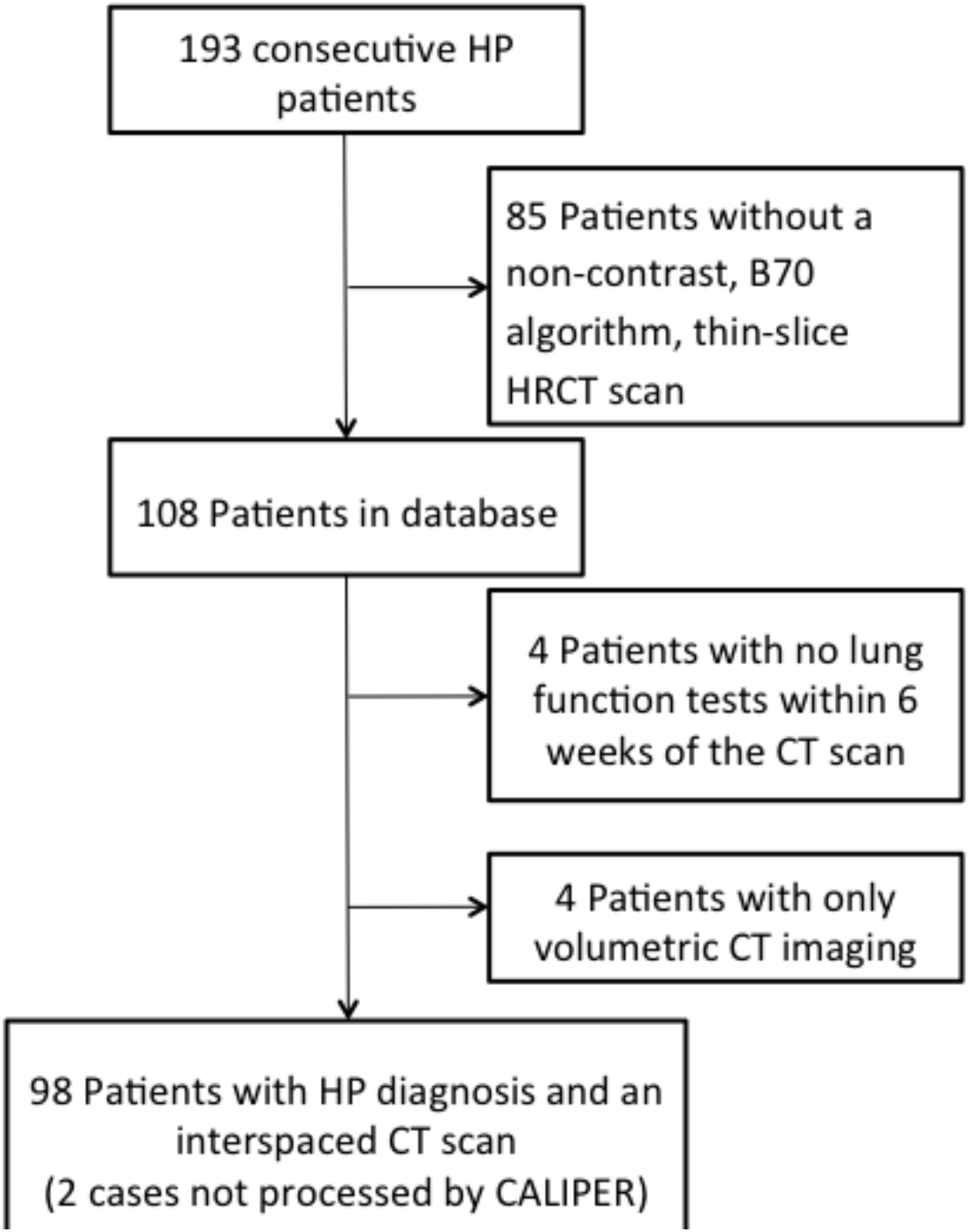

Figure 2a

Click here to download Figure Figure 2a.jpg

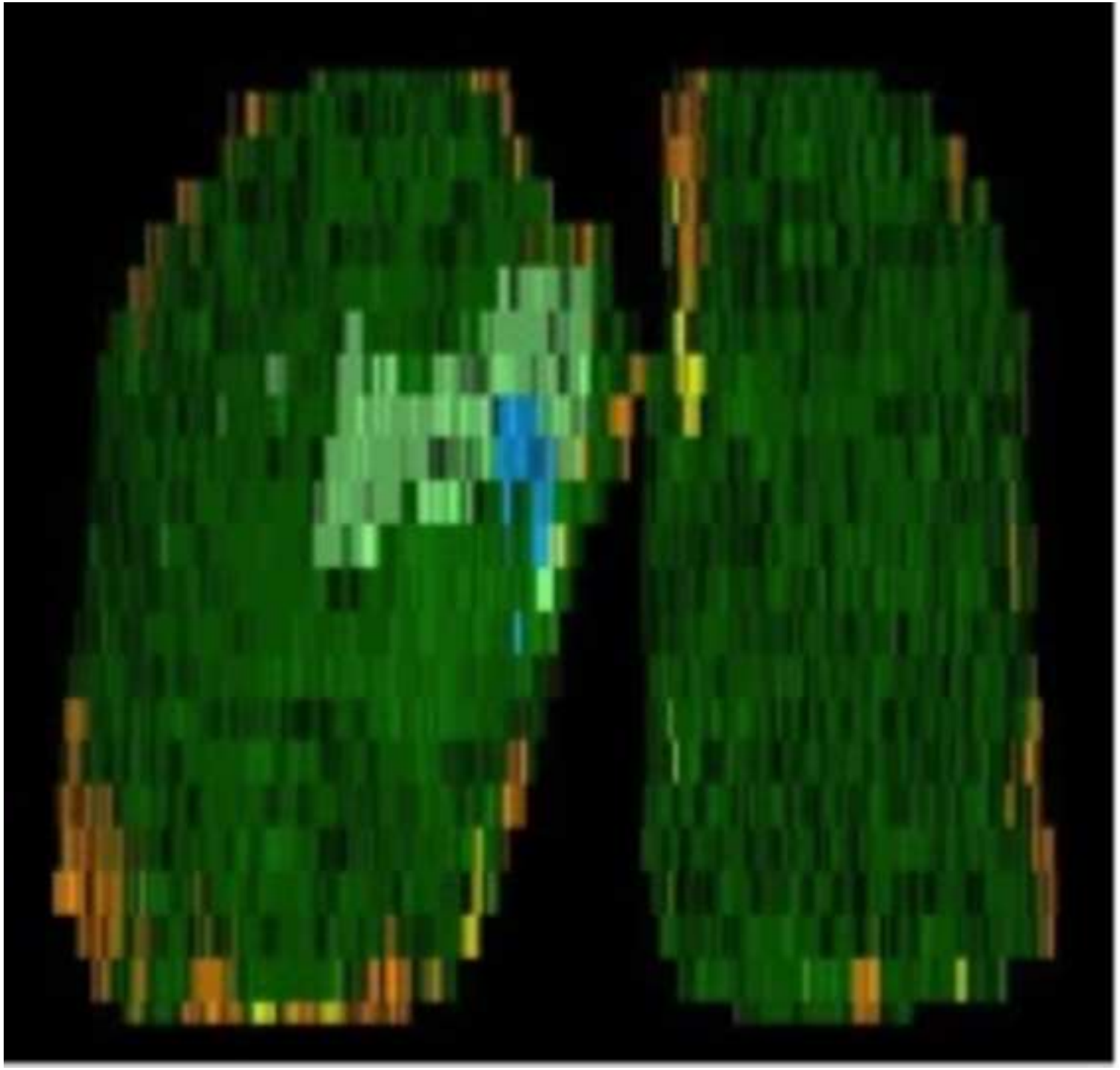

Figure 2b

Click here to download Figure Figure 2b.jpg

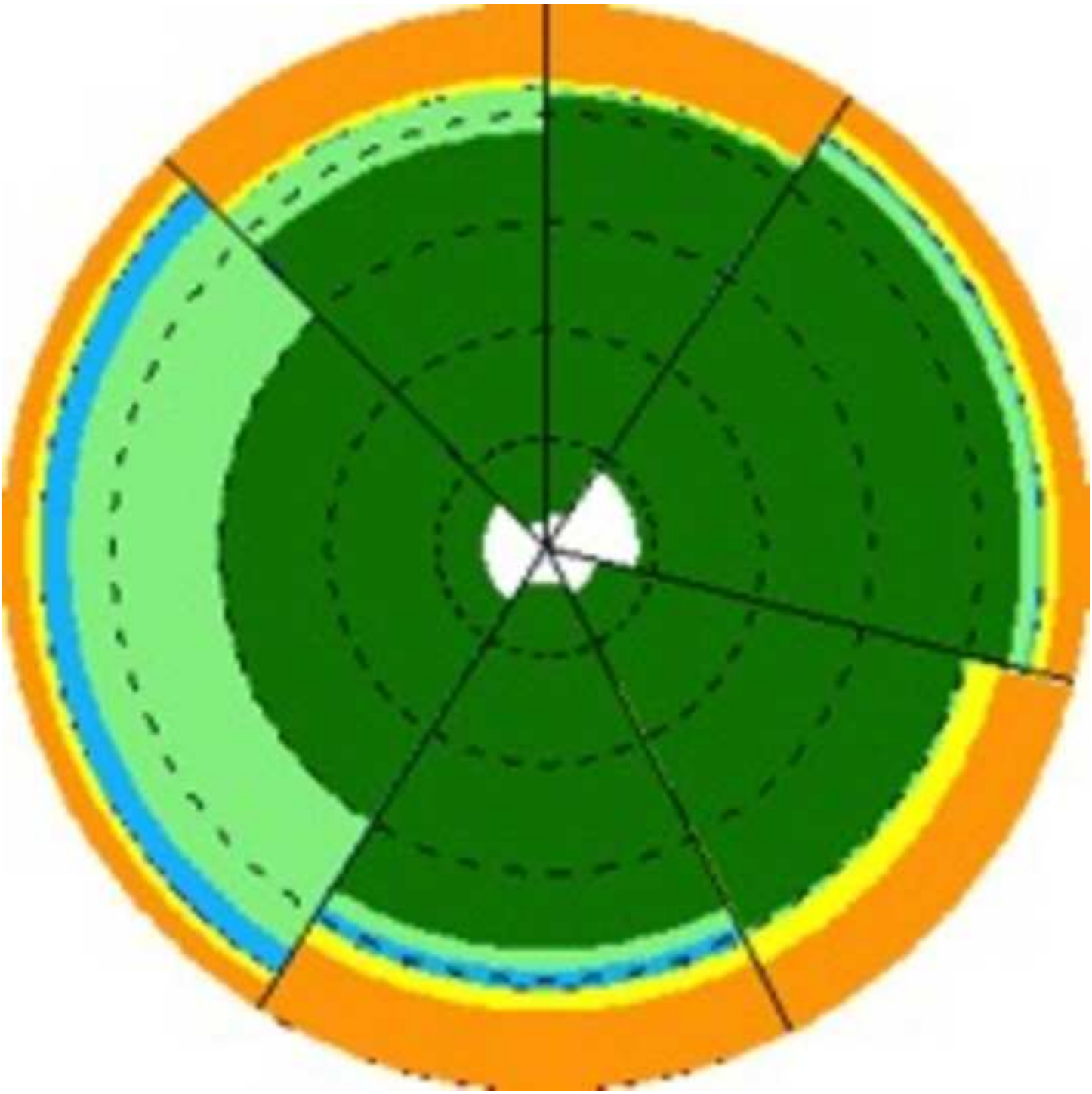

Figure 2c

Click here to download Figure Figure 2c.jpg

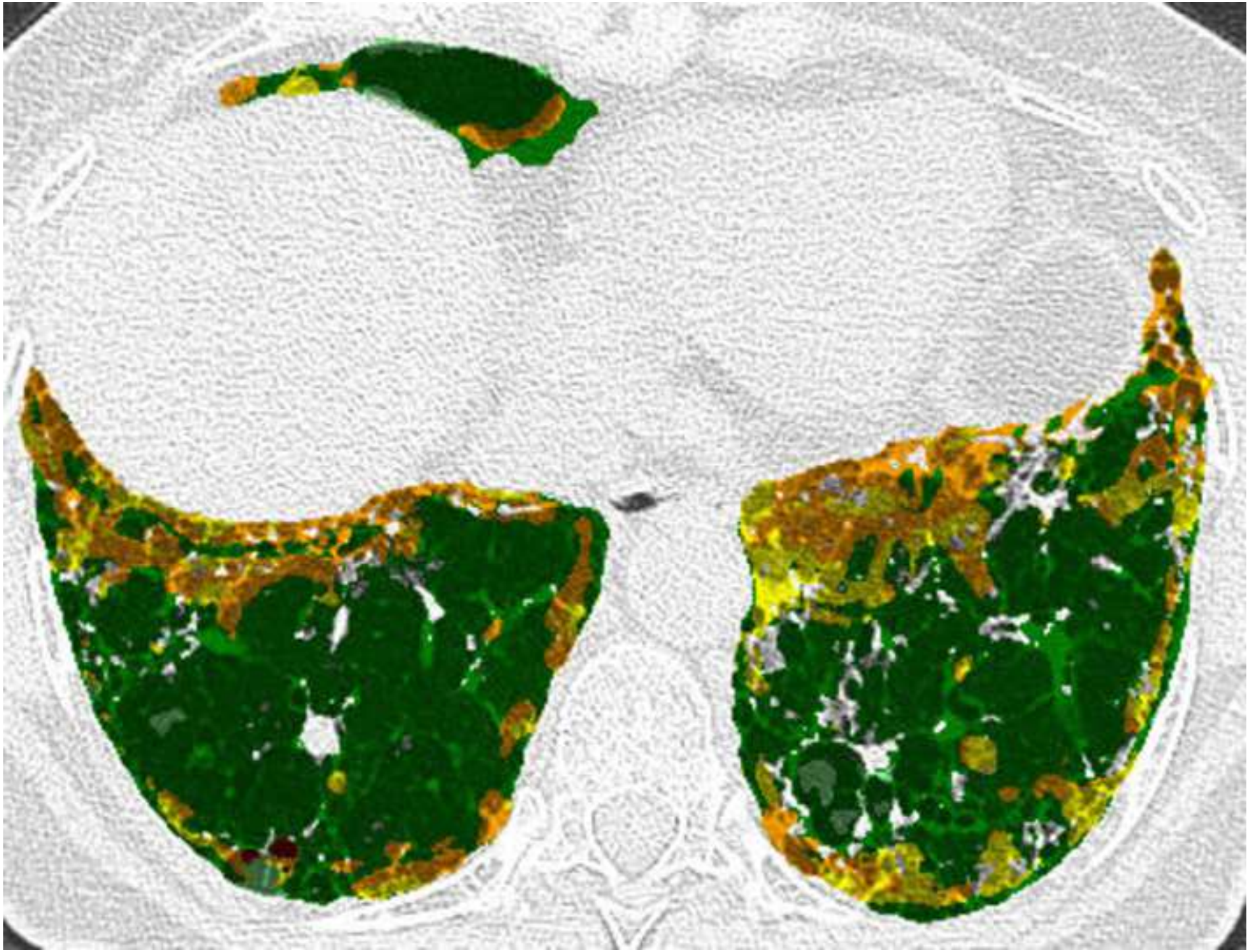

Figure 2d

Click here to download Figure Figure 2d.jpg

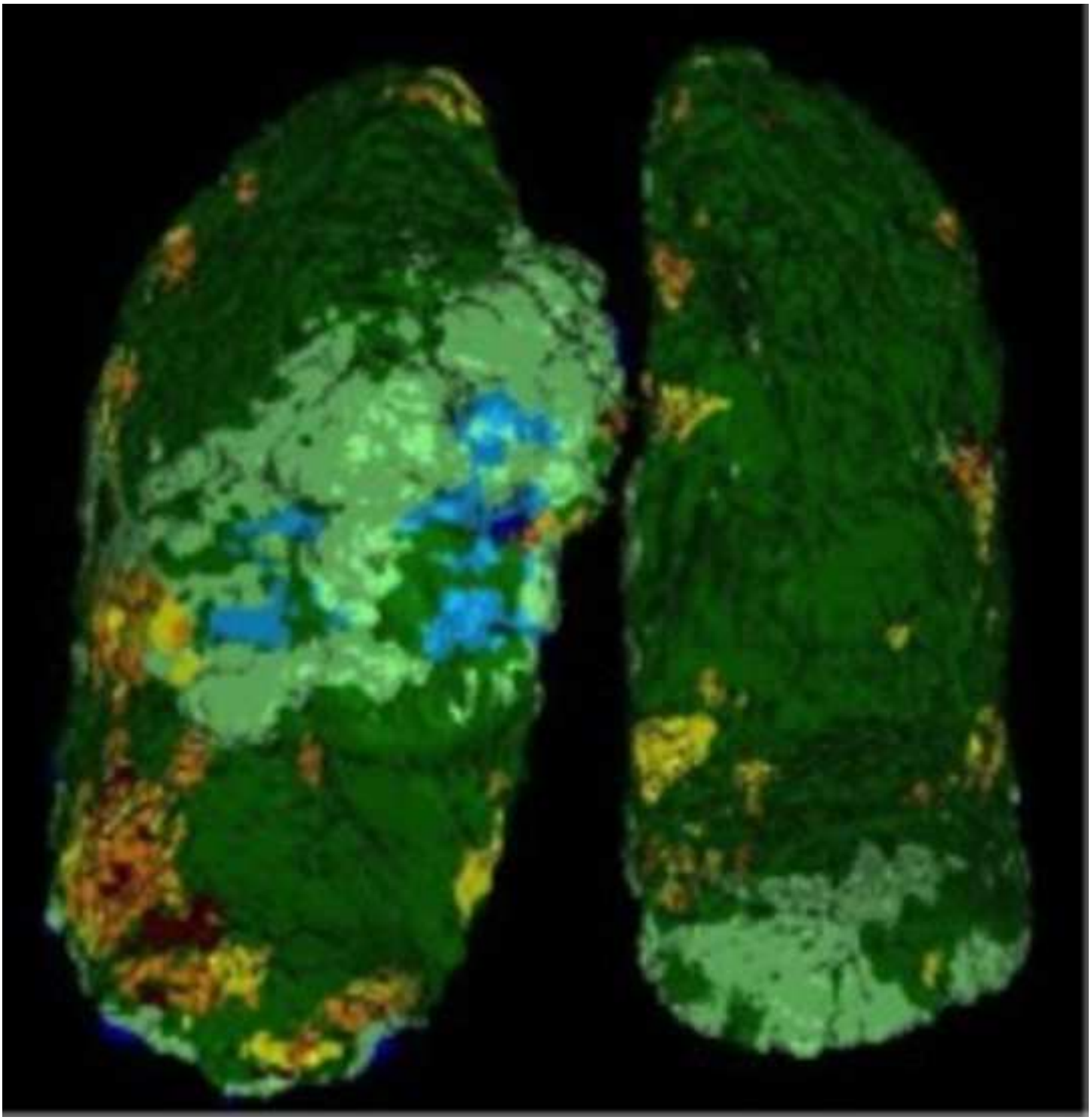

Figure 2e

Click here to download Figure Figure 2e.jpg

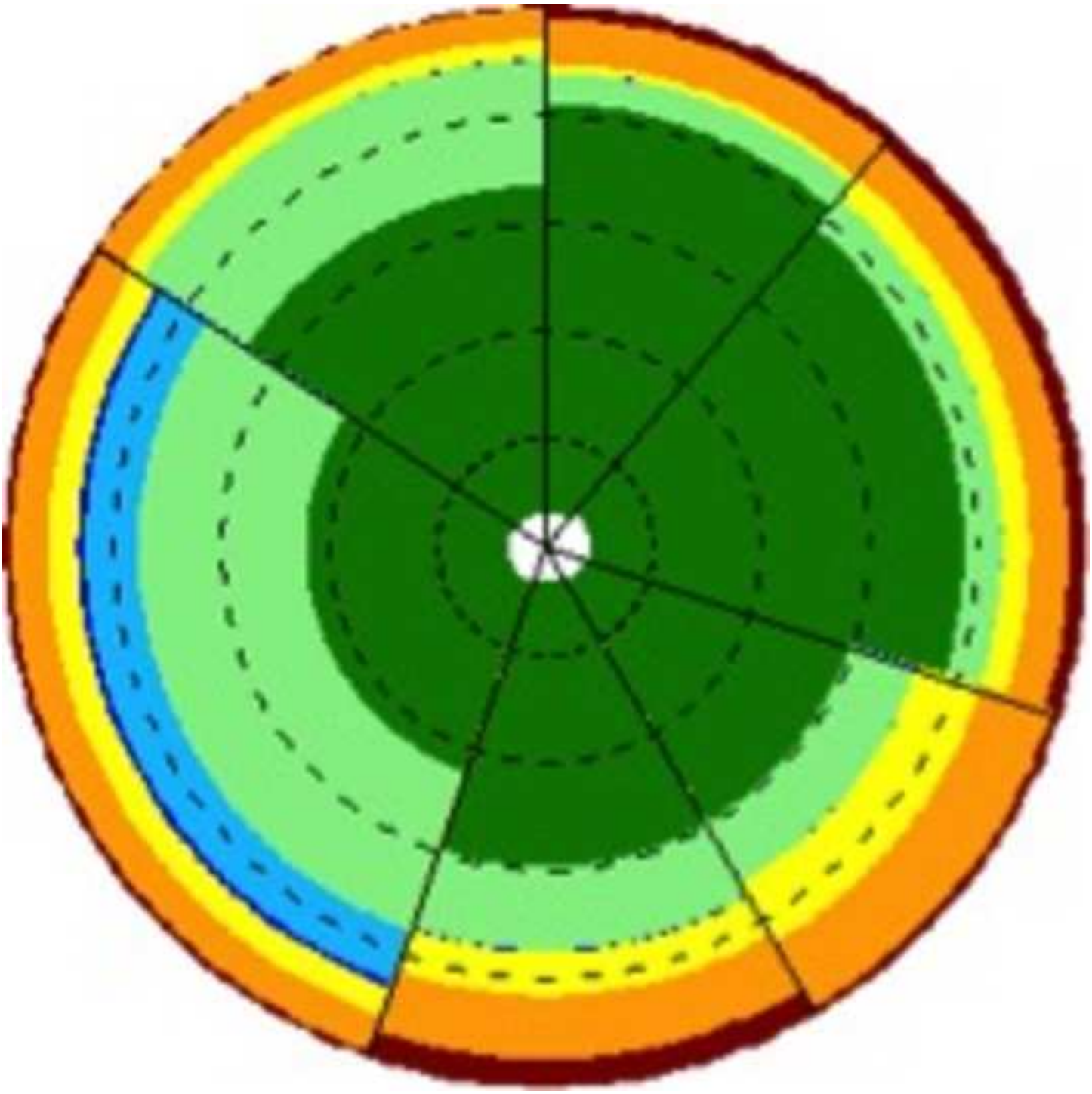

Figure 2f

Click here to download Figure Figure 2f.jpg

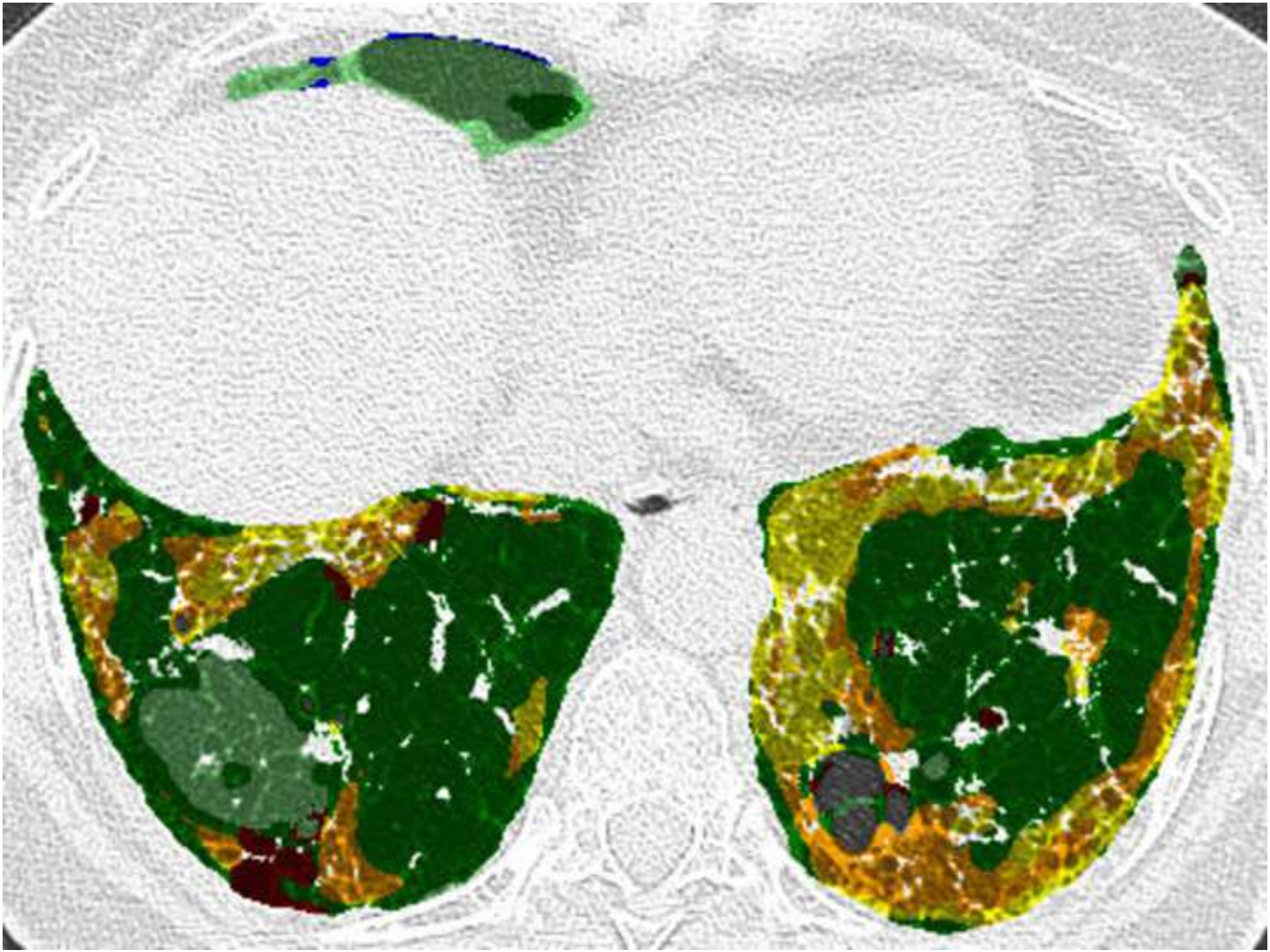

Figure 3a

Click here to download Figure Figure 3a.tiff

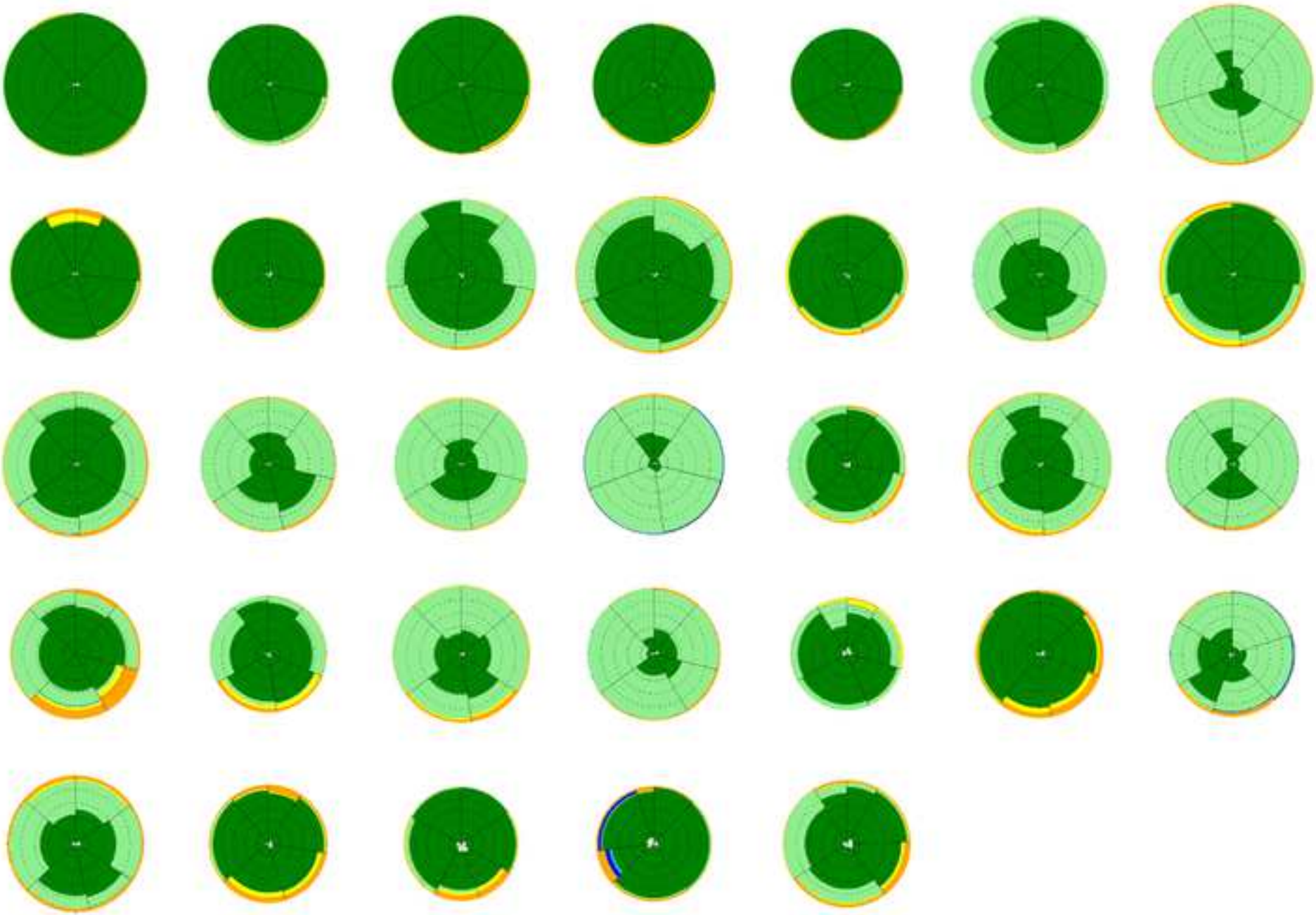

Figure 3b

Click here to download Figure Figure 3b.tiff

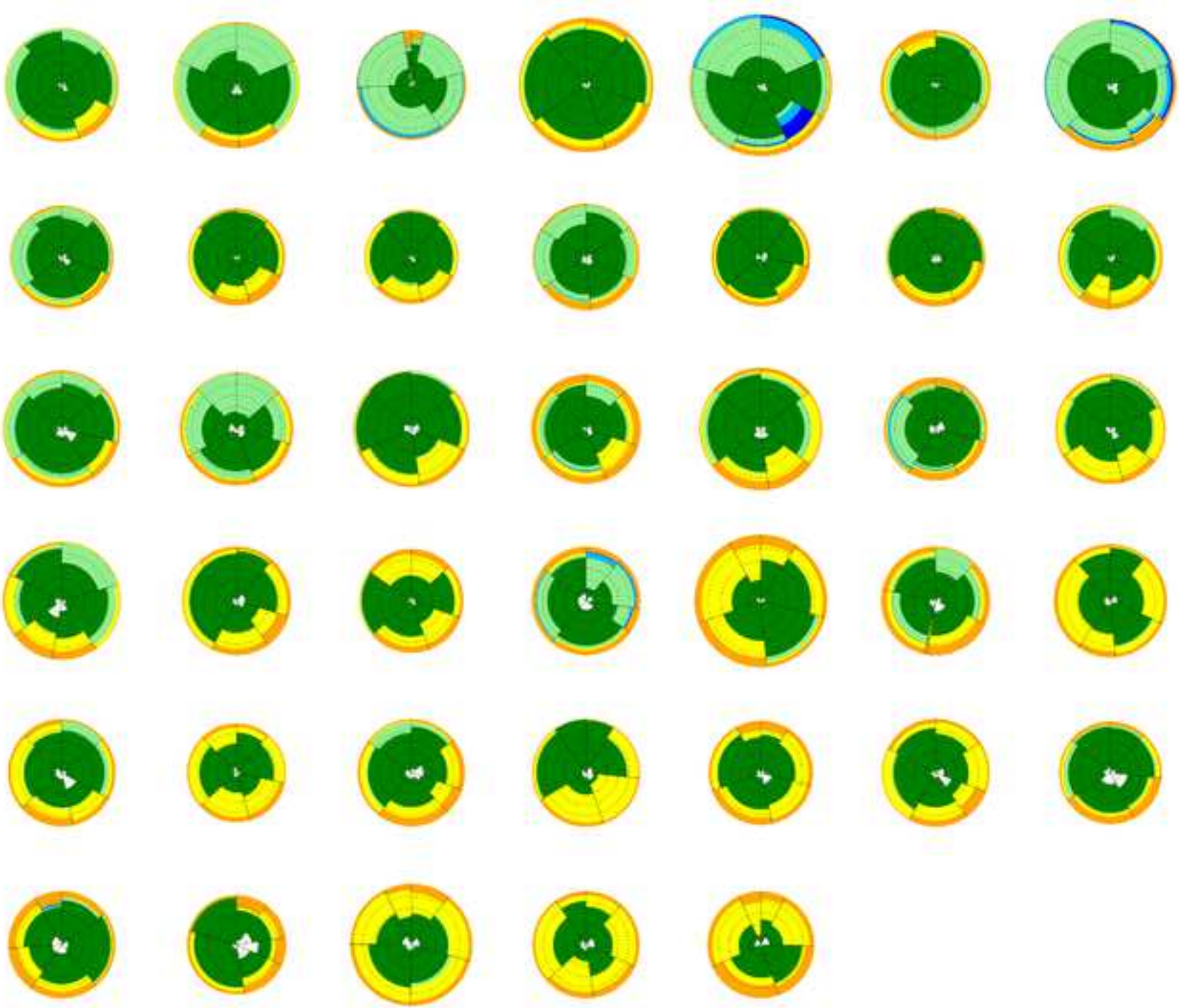

Figure 3c

Click here to download Figure Figure 3c.tiff

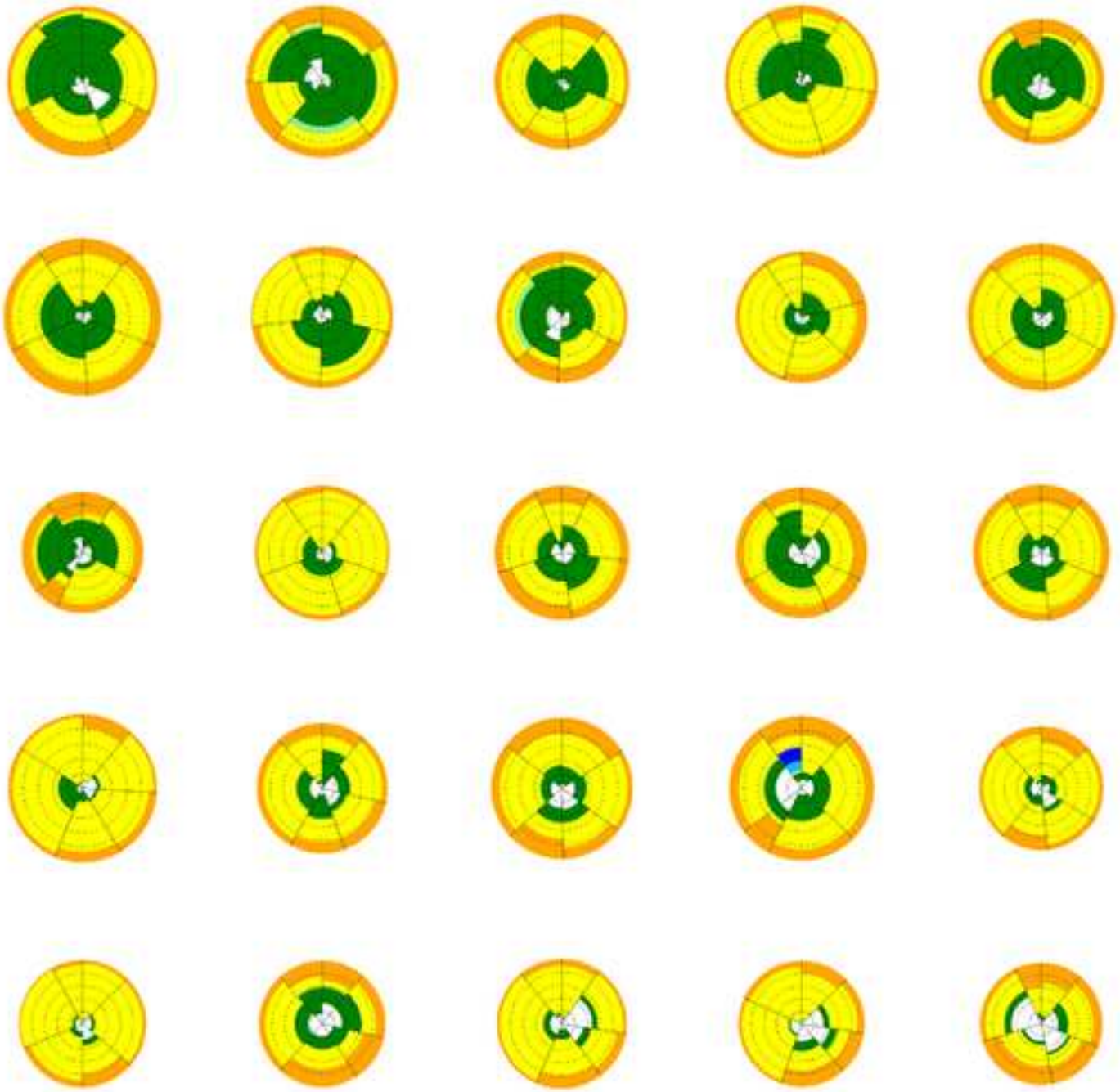

Figure 3d

Click here to download Figure Figure 3d.tiff

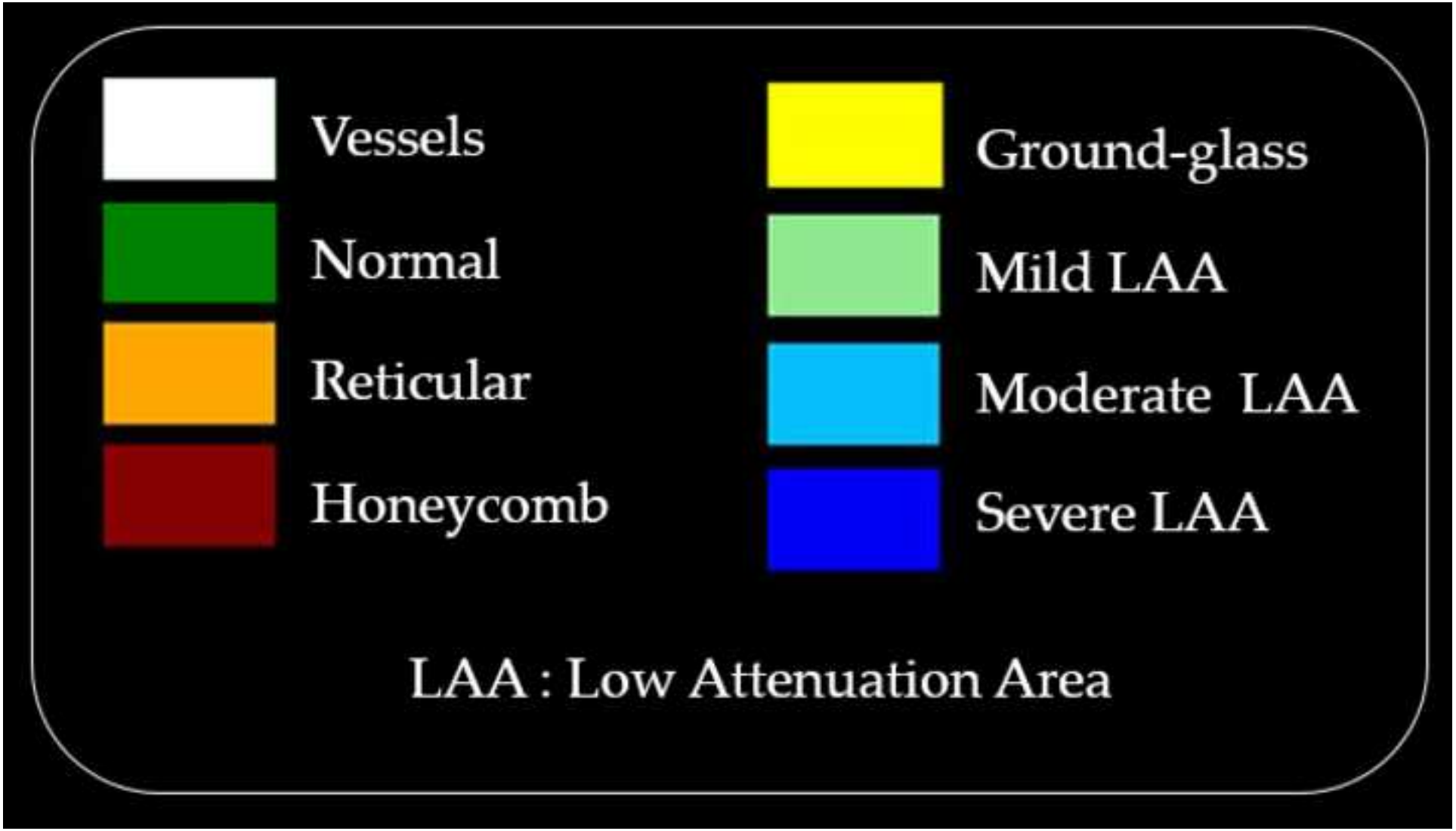

Figure 4a

Click here to download Figure Figure 4a.jpg

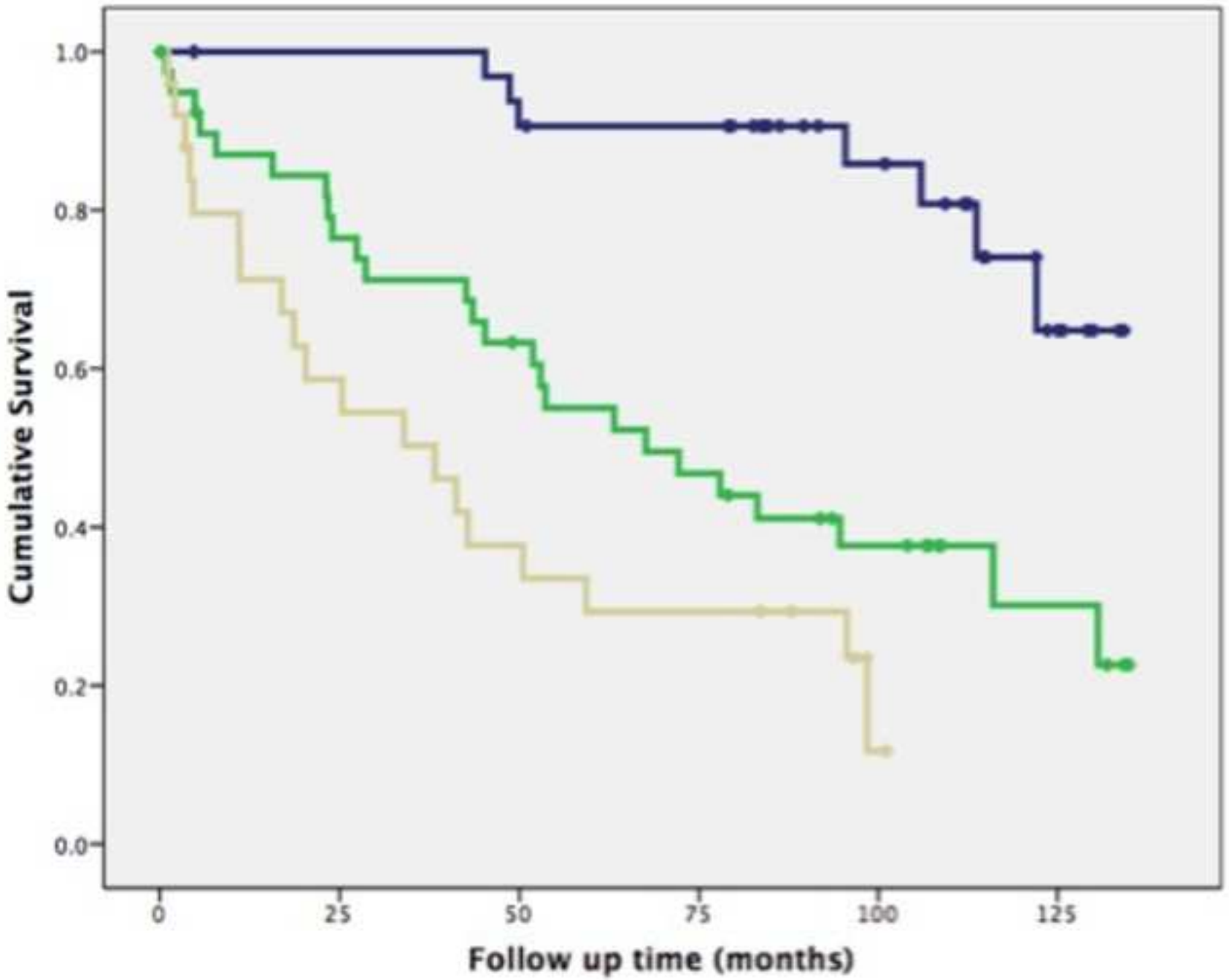

Figure 4b

Click here to download Figure Figure 4b.jpg

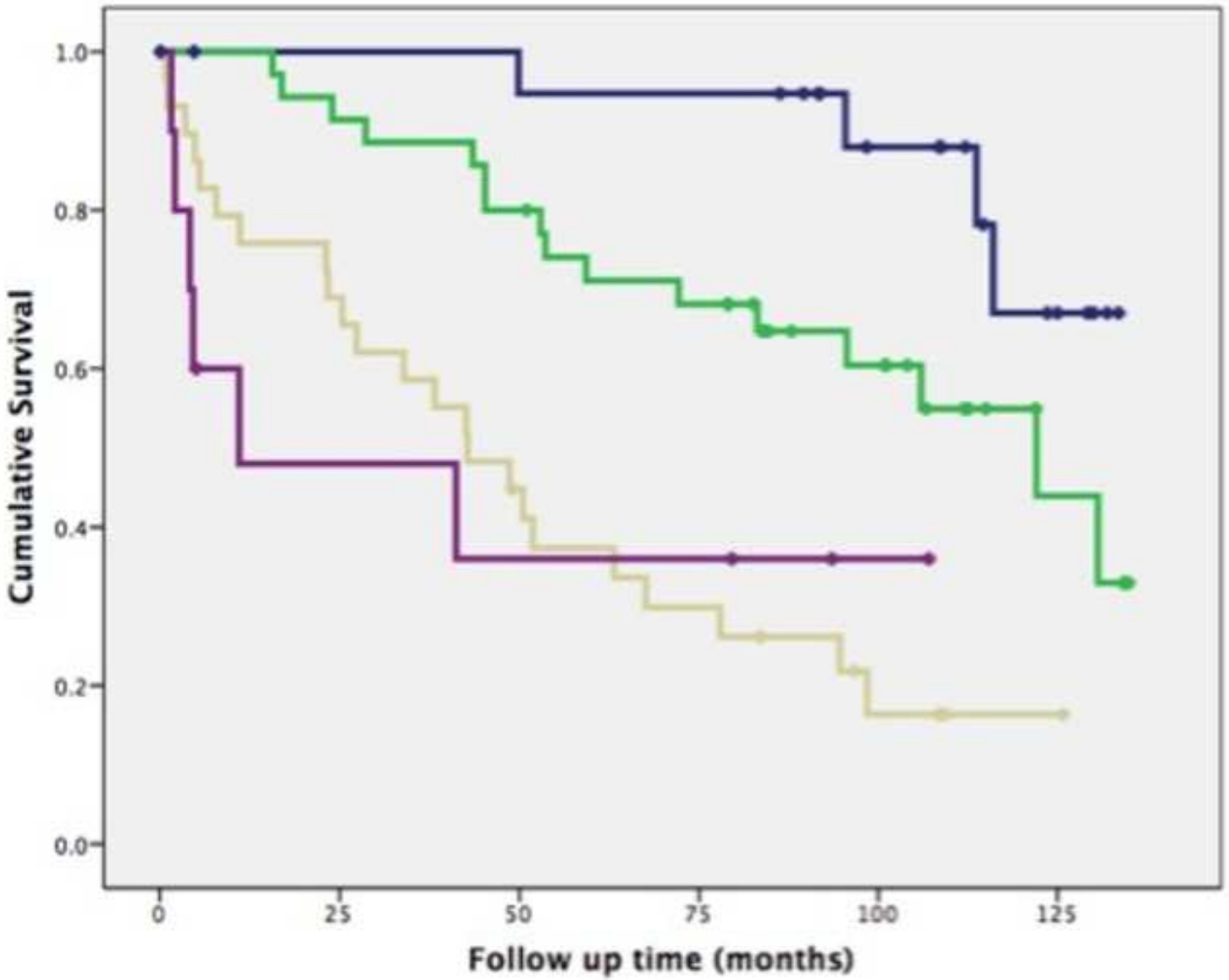

Figure 4c

Click here to download Figure Figure 4c.jpg

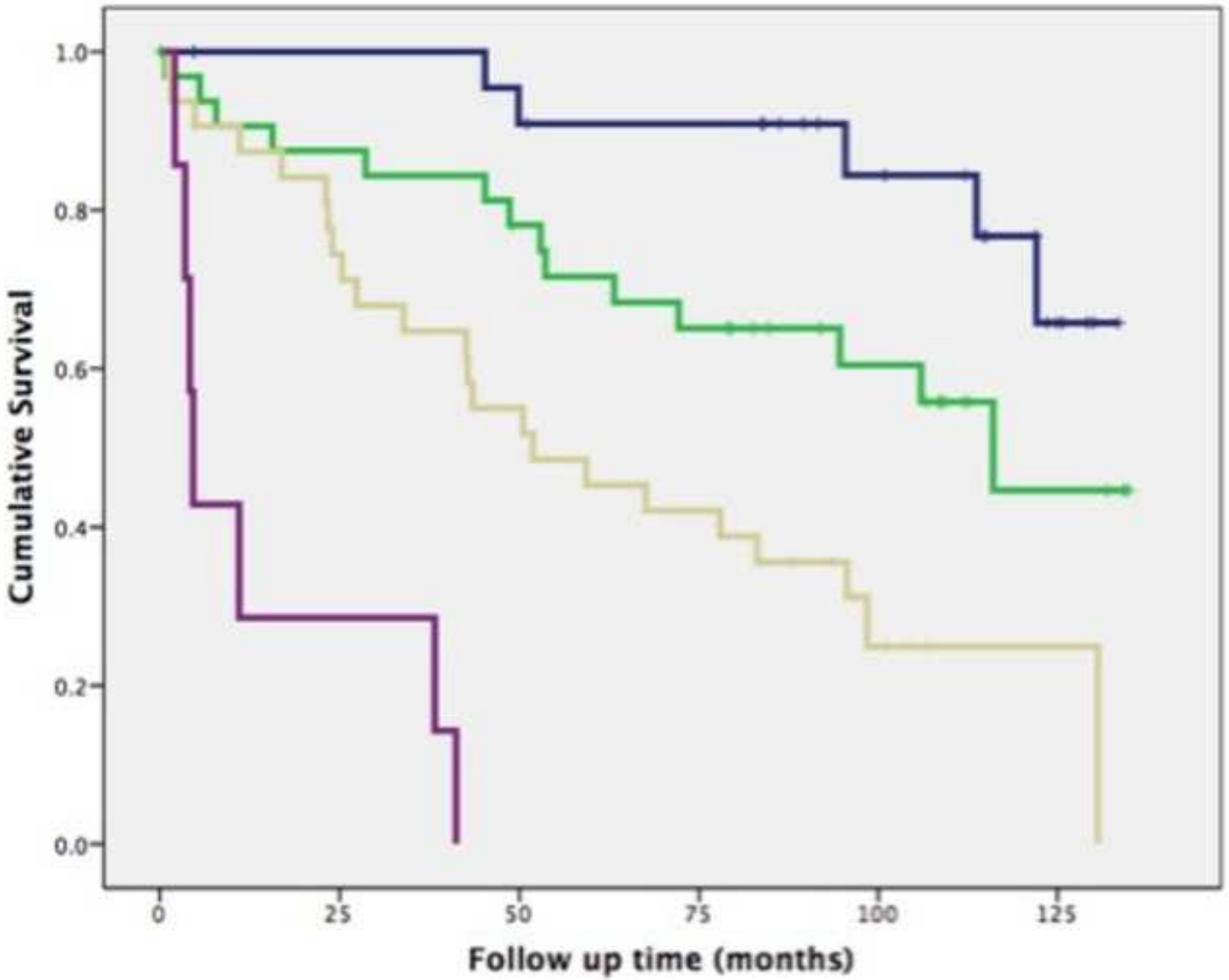

Figure 5

Click here to download Figure Figure 5.jpg

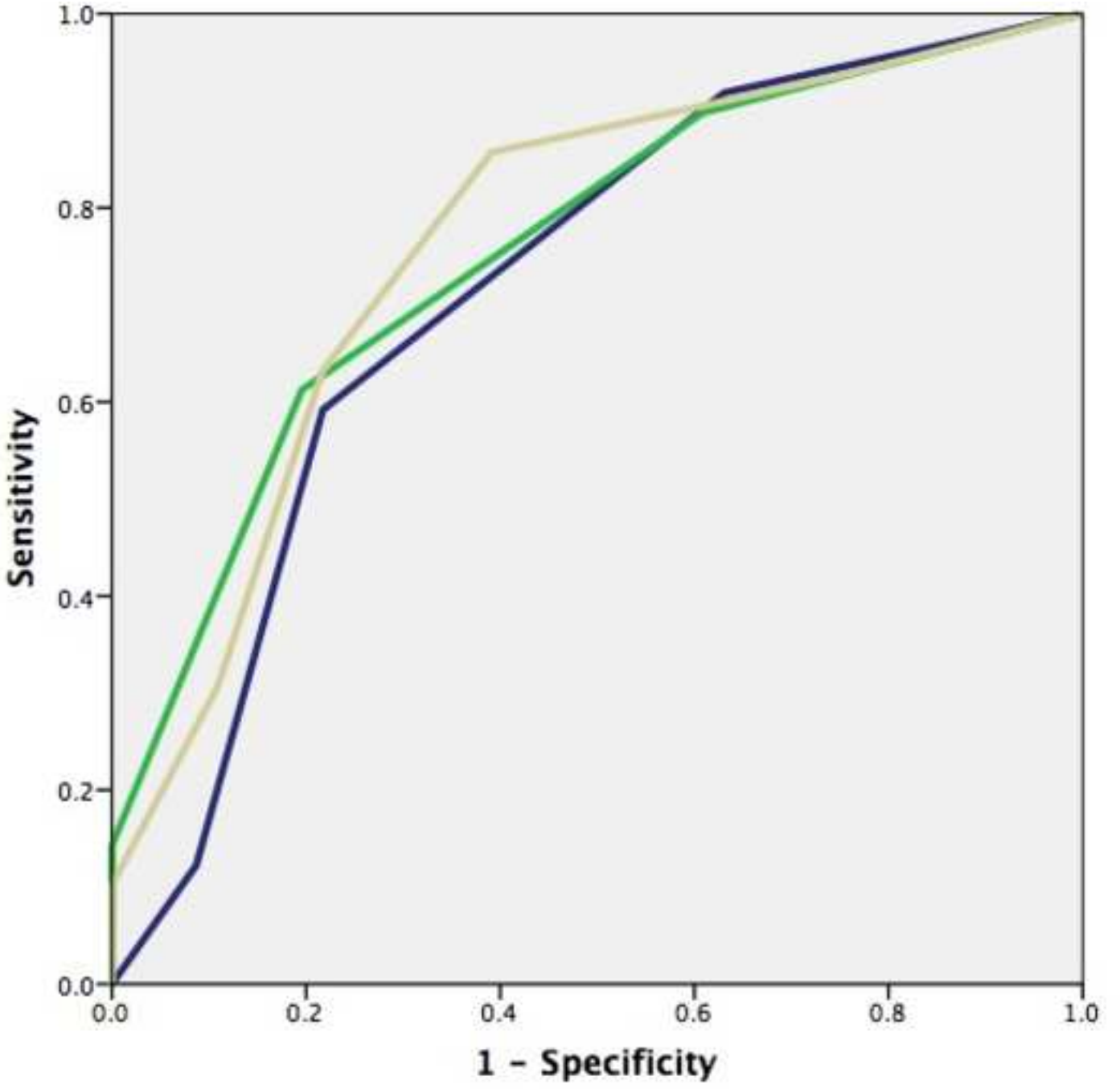

Figure Supp 1

Click here to download Figure Supplementary Figure 1.jpg