Unsupervised Discovery of Novel Emphysema Subtypes

Andrew Francis Laine, D.Sc.

Percy K. and Vida L. W. Hudson Professor of Biomedical Engineering and
Professor of Radiology (Physics)
Department of Biomedical Engineering
Columbia University
New York, NY 10027

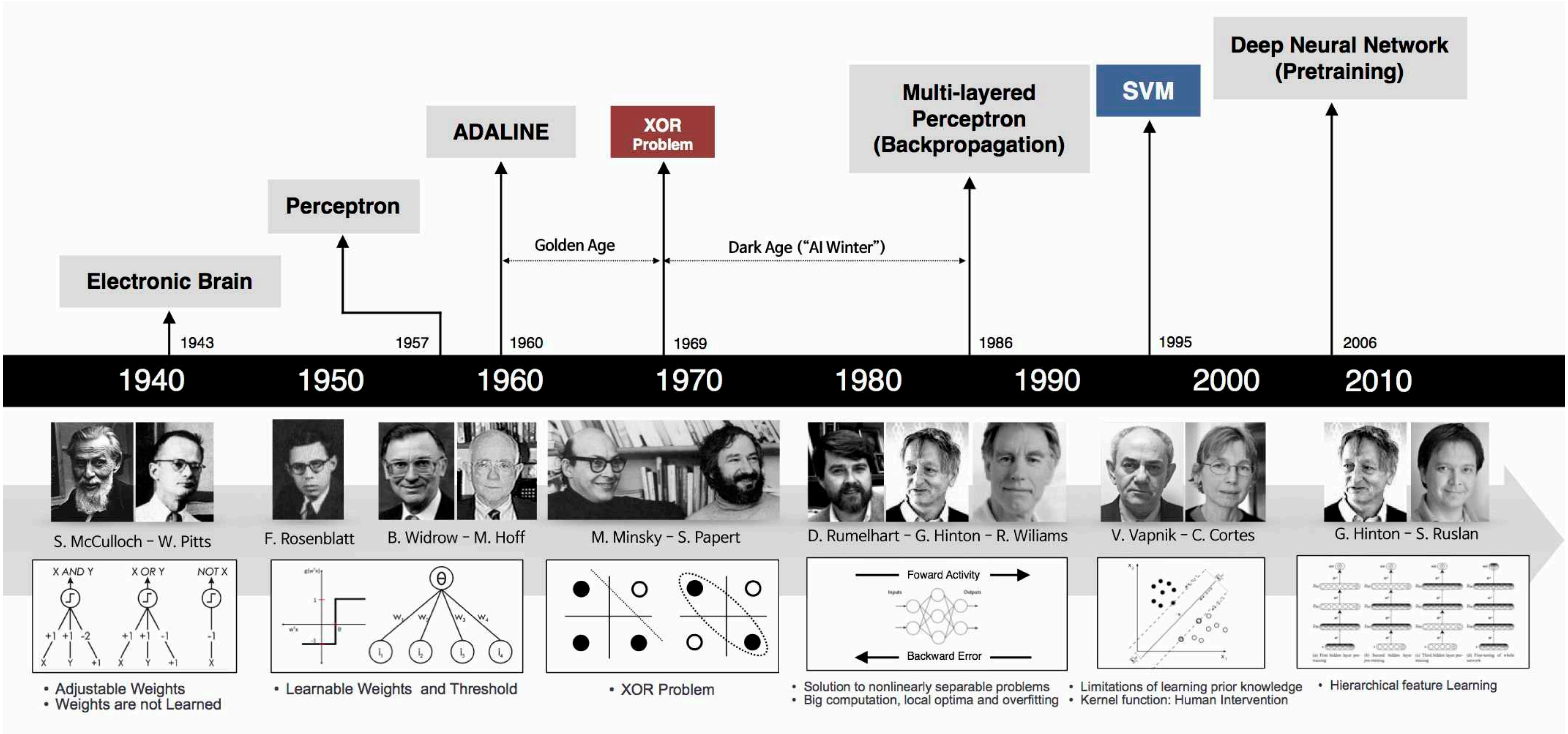
Collaborators:

R. Graham Barr, M.D., Ph.D.,
Department of Medicine, CUIMC

Elsa D. Angelini, Ph.D.
Department of Radiology
CUIMC

Jie Yang (Ph.D. Student, 2019)

Historical Overview of AI: Timeline



Machine Learning and AI

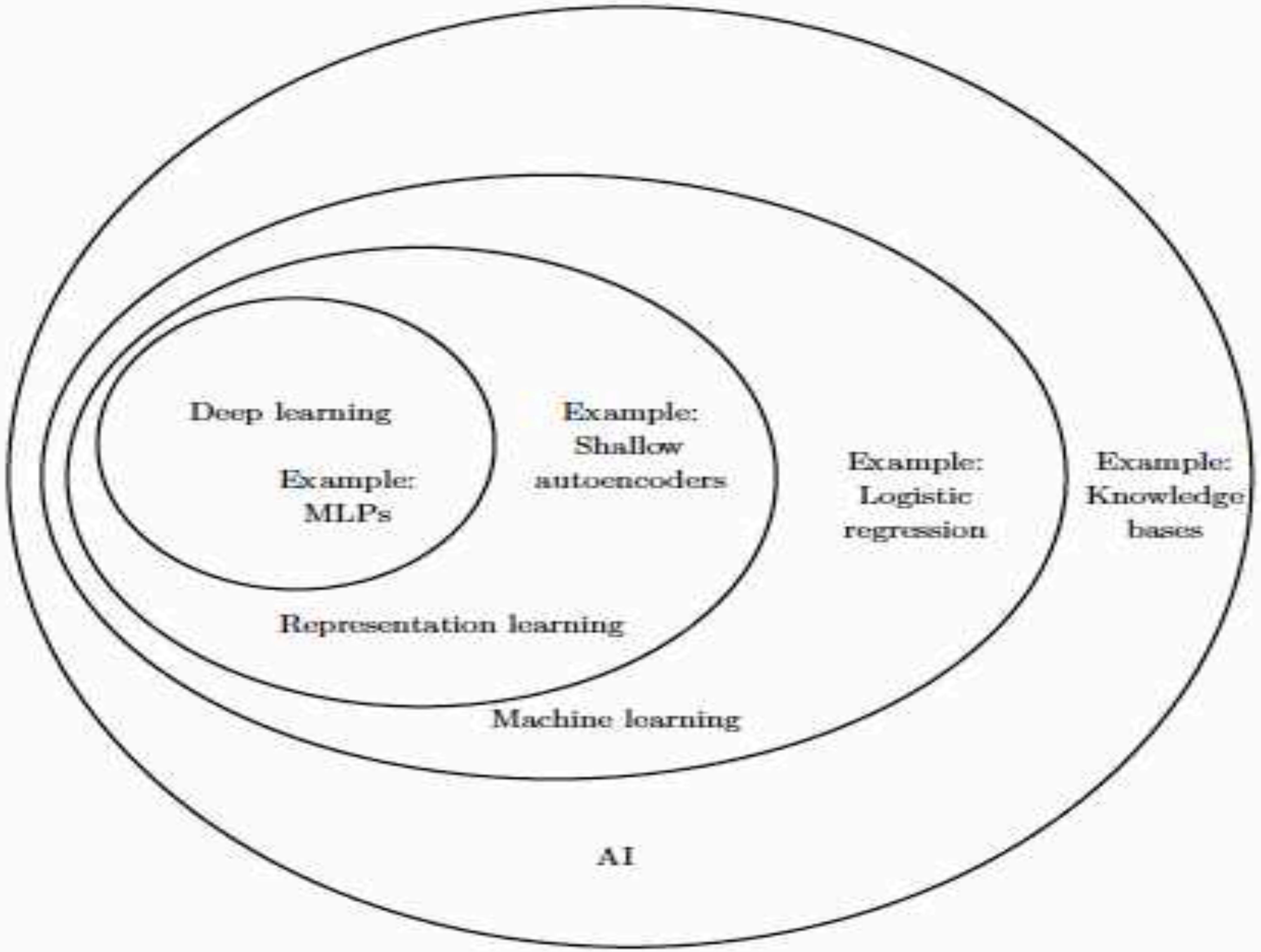


Figure 1.4

(Goodfellow 2016)

Representations Matter

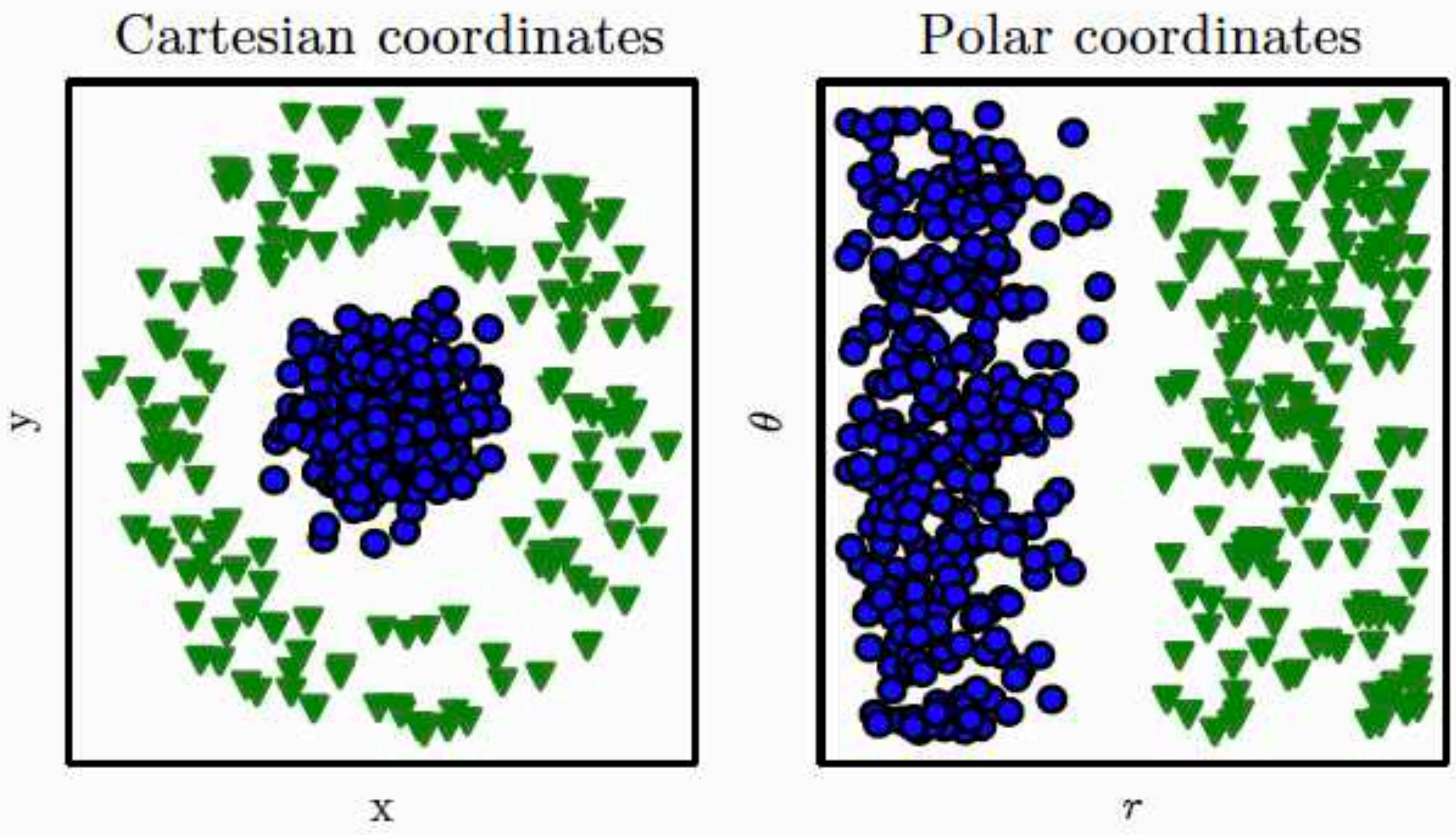
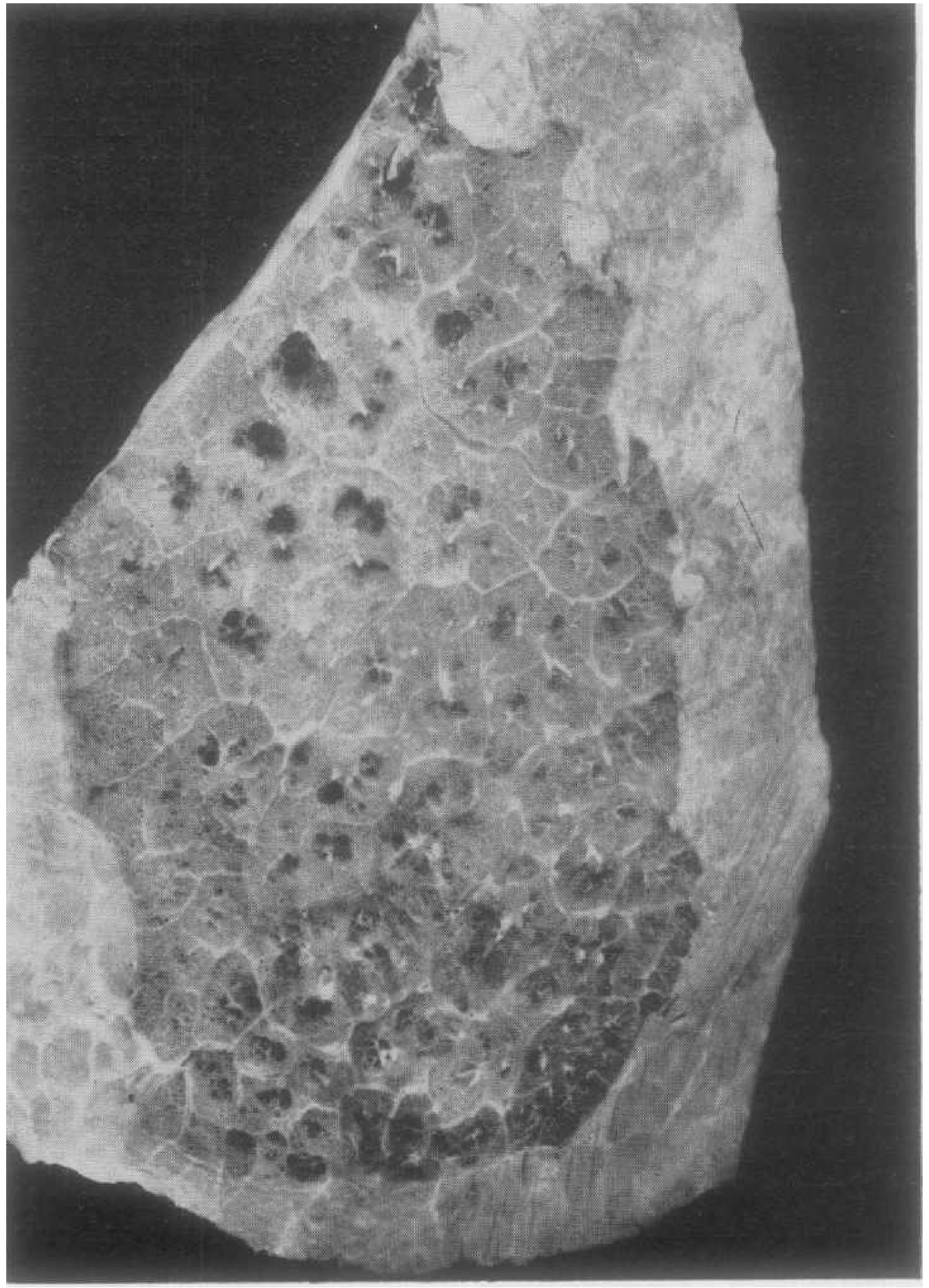


Figure 1.1

(Goodfellow 2016)

Emphysema



- Emphysema + COPD is 3rd leading cause of death in USA.
- Defined by loss of interalveolar septa
- Predicts mortality in patients with and without COPD

Leopold/Gough, Thorax, 1957
Johannessen, AJRCCM, 2013
Oelsner, Ann Intern Med, 2014

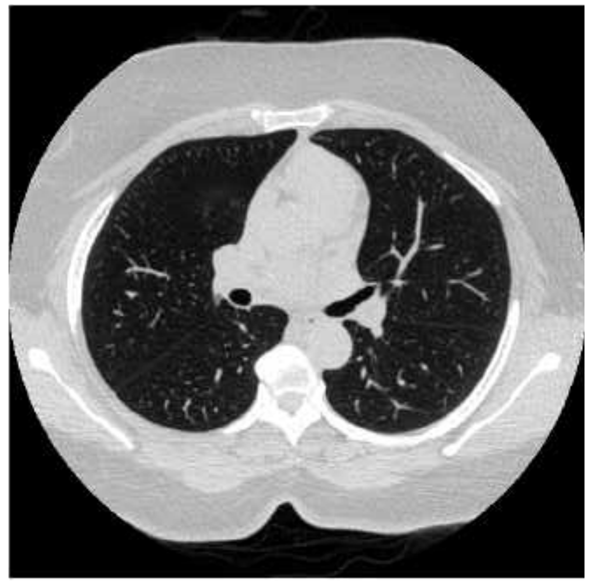
Background

Computed tomography (CT) used to analyze lung structure:

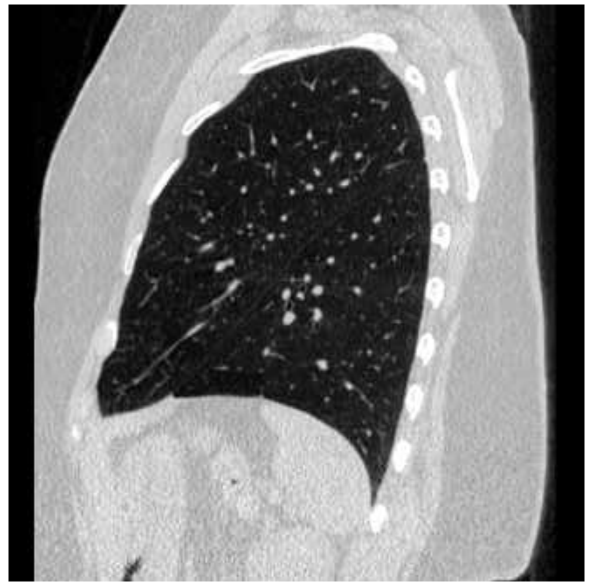
Resolution =
0.5×0.5×0.75 mm

Matrix size =
512×512×500 pixels

Intensity range =
[-1024 1024] HU

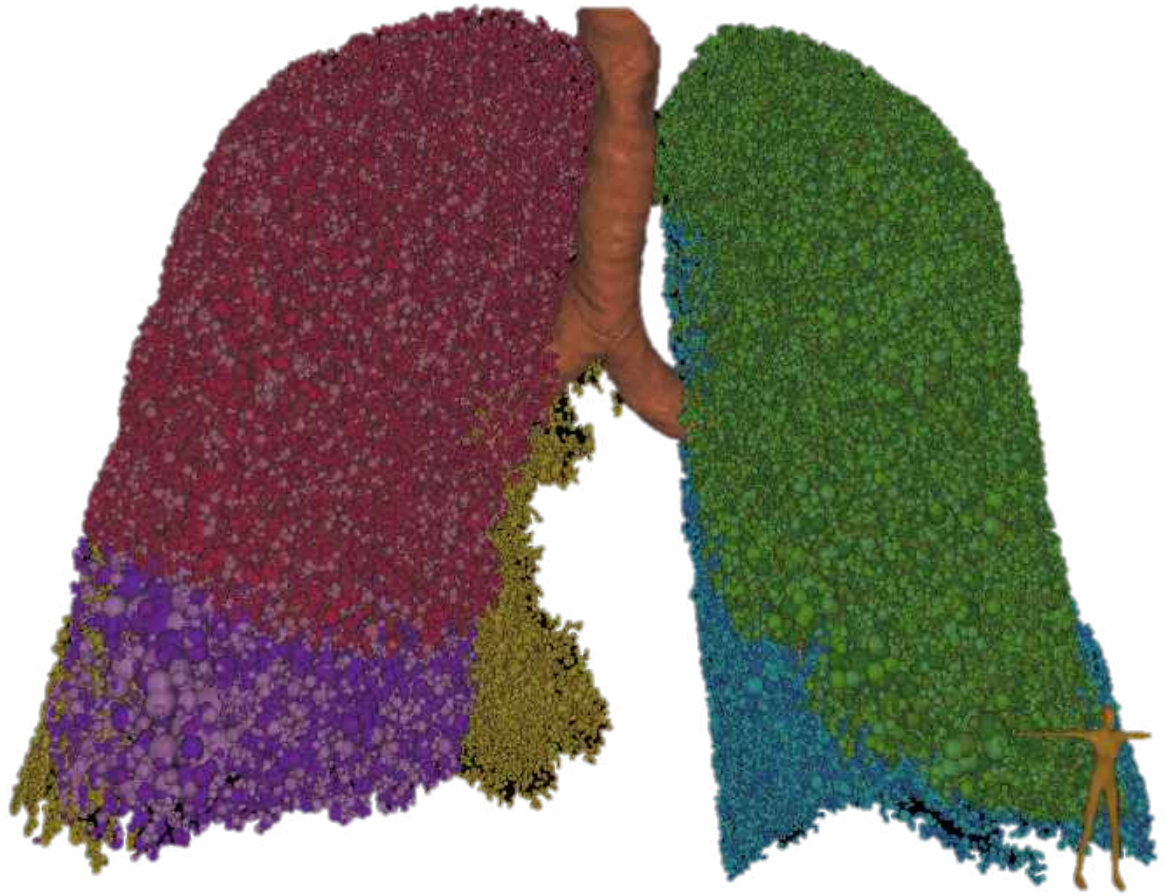


Axial



Sagittal

Coronal



40 megavoxels of the lung:

- Enable ***in vivo*** study of lung structure and disease patterns.

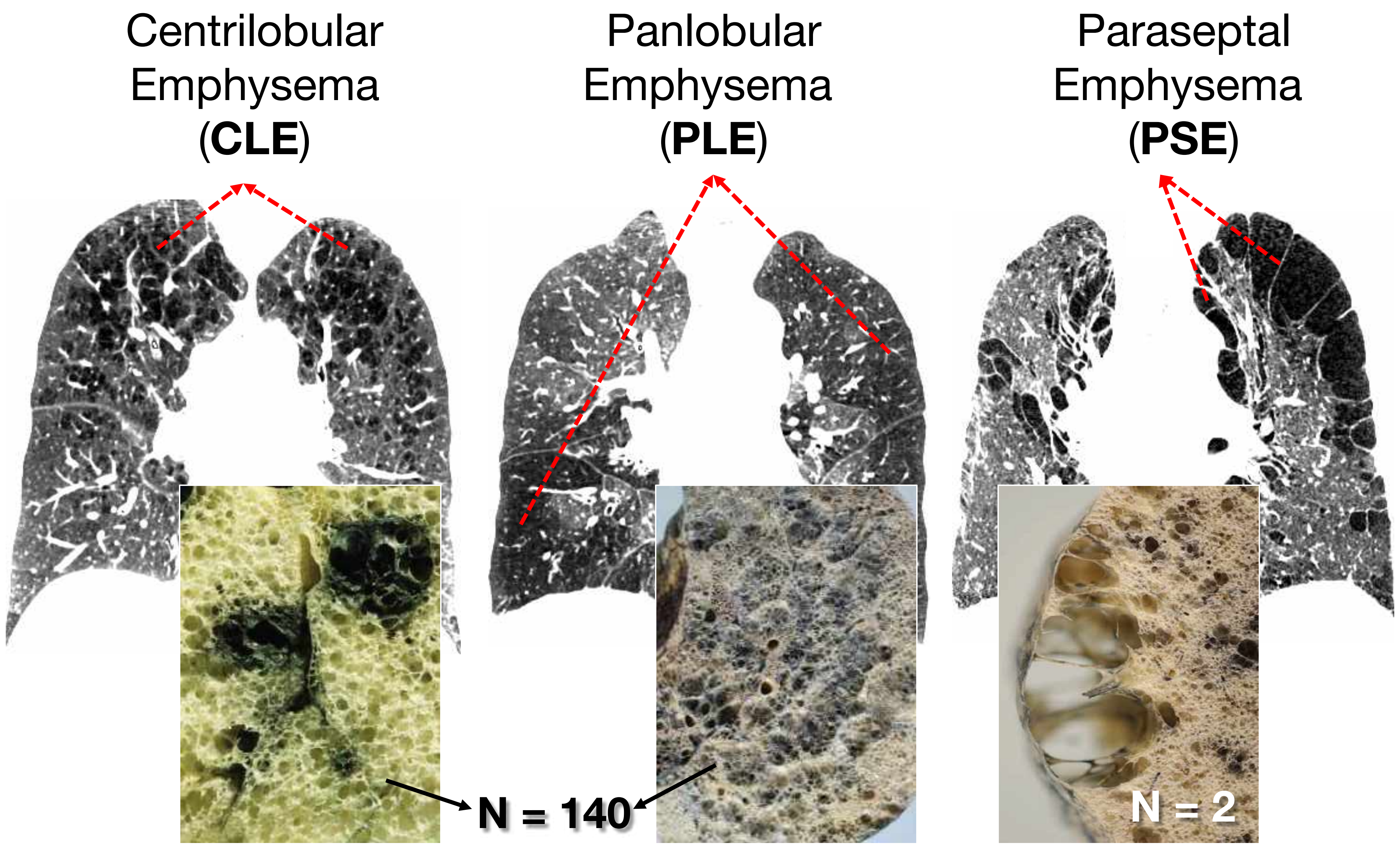
Background

Lung texture learning to characterize emphysema subtypes:

COPD and Emphysema

- ❑ Exact mechanism of developing COPD remains unknown;
- ❑ Three standard emphysema subtypes defined at autopsy [1]:
  - Limited inter-rater agreement.

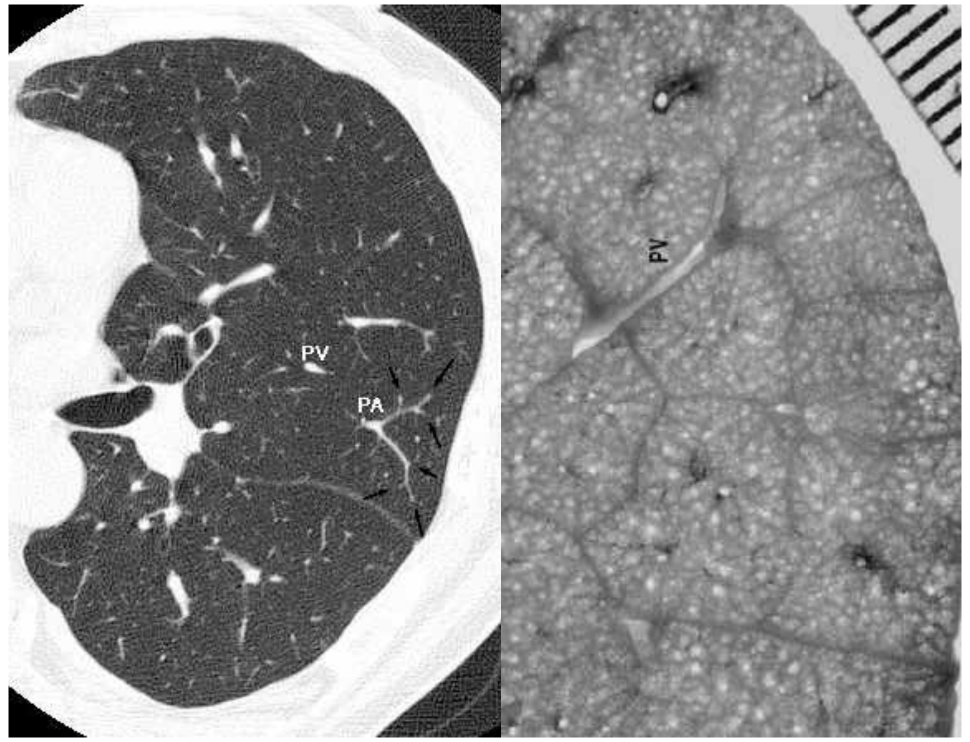
Lung texture learning for emphysema subtyping can advance disease understanding



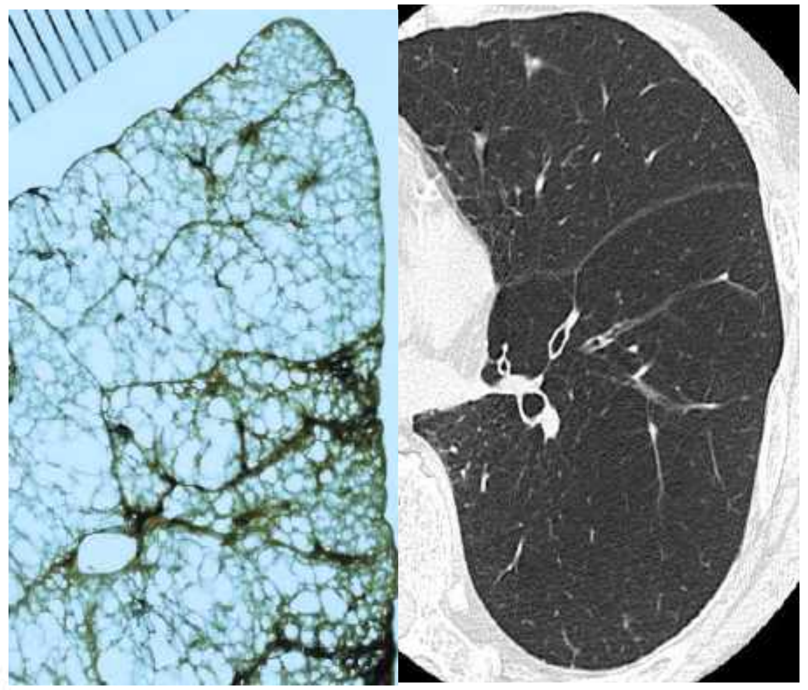
Emphysema Subcategories

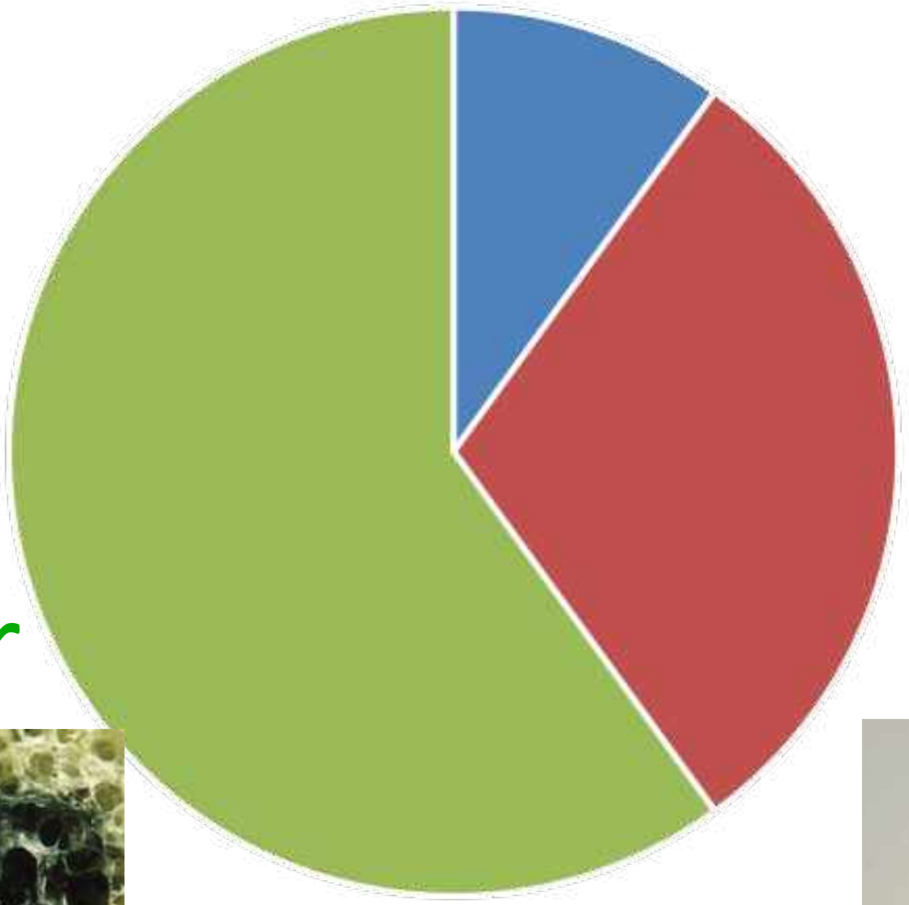
[1] Smith *et al.*, Pulmonary emphysema subtypes on computed tomography: the MESA COPD study, *The American Journal of Medicine*, 2014.

Classic Emphysema Subtypes



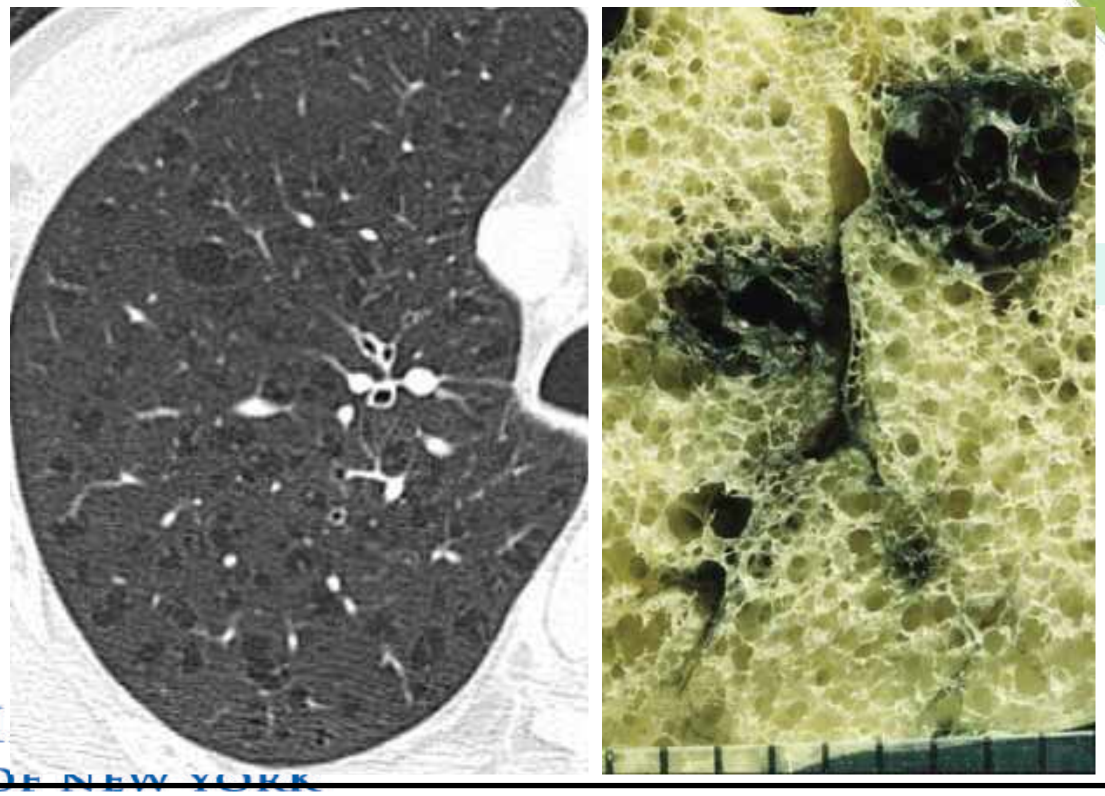
Pananlobular

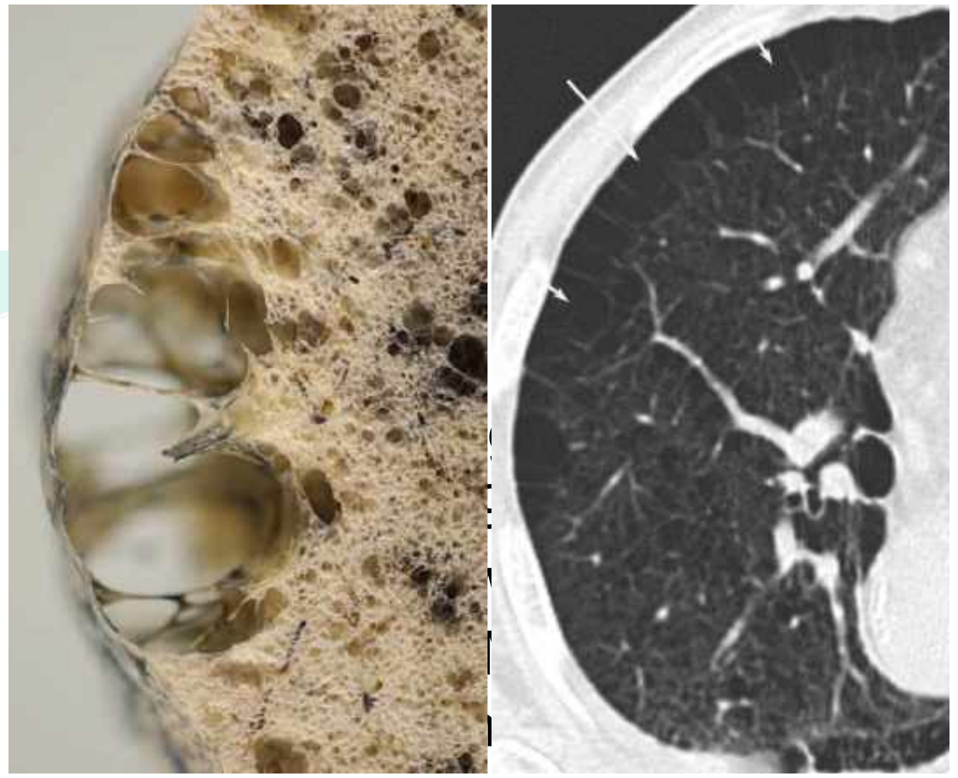




Centrilobular

Paraseptal





Background

Aimed to tackle the problem of CT-based lung texture learning exploiting ***spatial localization***, using ***unsupervised*** / ***weakly-supervised*** learning.

Aim 1: Develop an algorithm for **unsupervised** learning of **localized** texture patterns for **emphysema**.

Aim 2: Label the discovered localized texture patterns on **large datasets** of **cardiac** CT scans.

Aim 3: Examine possible correlations / hits with GWAS genomic information in MESA and SPIROMICS.

NIH R01-HL121270: **Novel Quantitative Emphysema Subtypes in MESA and SPIROMICS.**
(PIs Dr. R.G. Barr, Dr. A.F. Laine)

Hypothesis

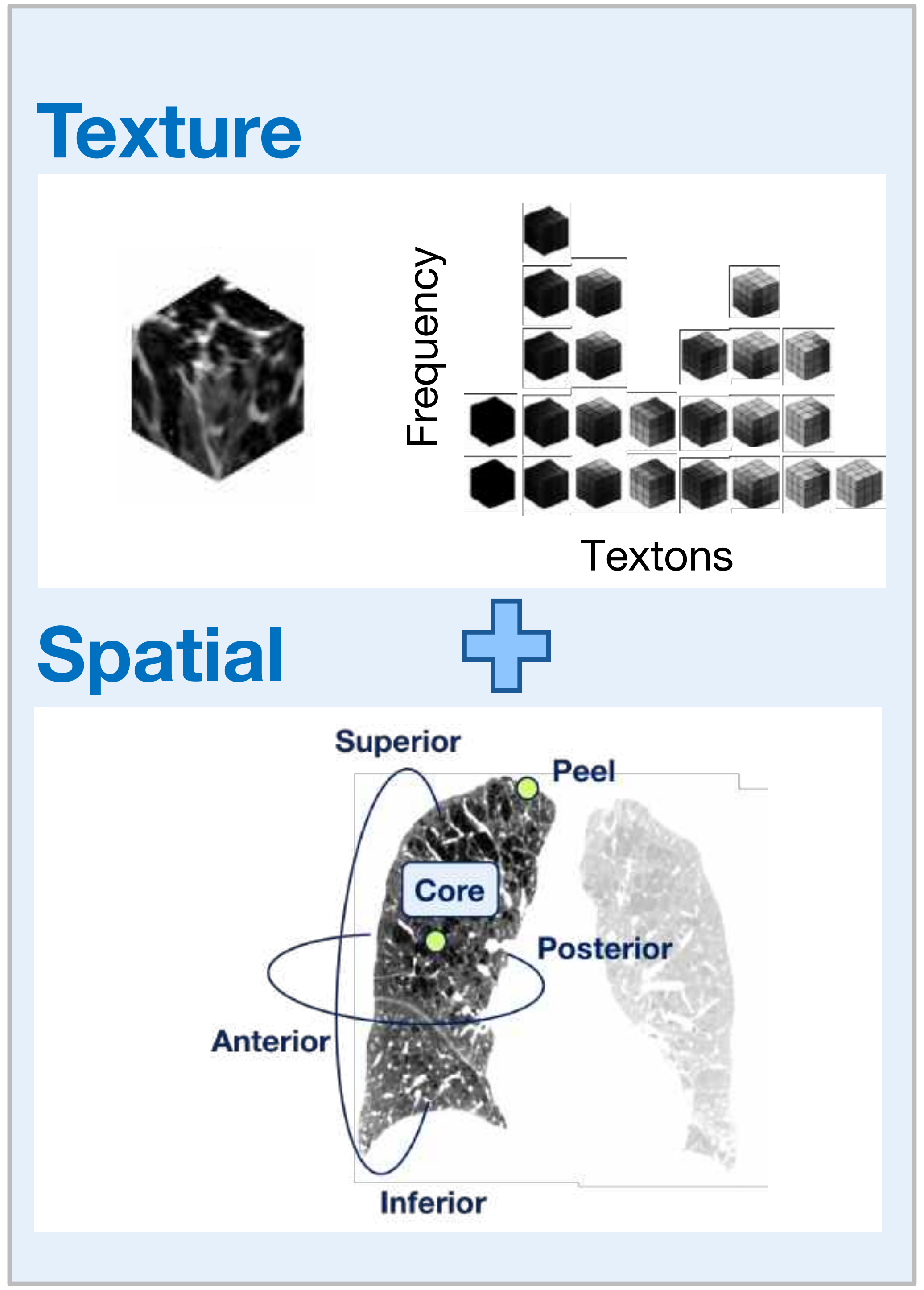
- Unsupervised learning of spatial lung texture patterns on research CT scans will yield novel emphysema subtypes.
  – Reproducible
  – Distinct symptoms
  – Specific histology and genetic basis

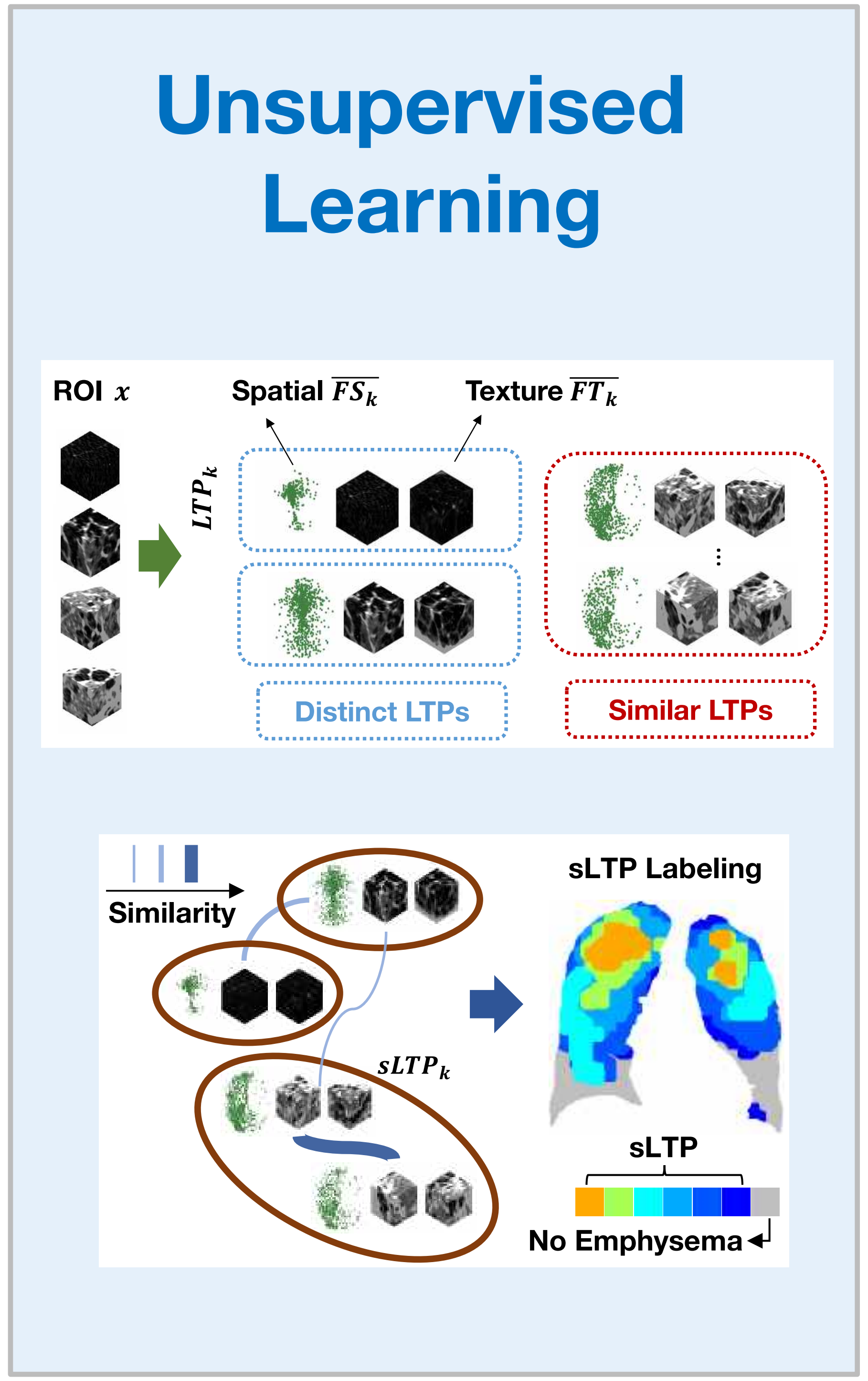
SPIROMICS and MESA

- ***S**ub**P**opulations and **I**ntermediate **O**utcome Measures In COPD Study*
  - COPD case-control study
  - 2,983 participants with CT scans (~6,000 scans)
  - Whole genome sequencing, multi-omics
- ***M**ulti-**E**thnic **S**tudy of **A**therosclerosis Lung Study*
  - Population-based, prospective cohort study
  - 3,205 participants with full-lung CT scans (~50,000 scans)
  - Whole genome sequencing, multi-omics

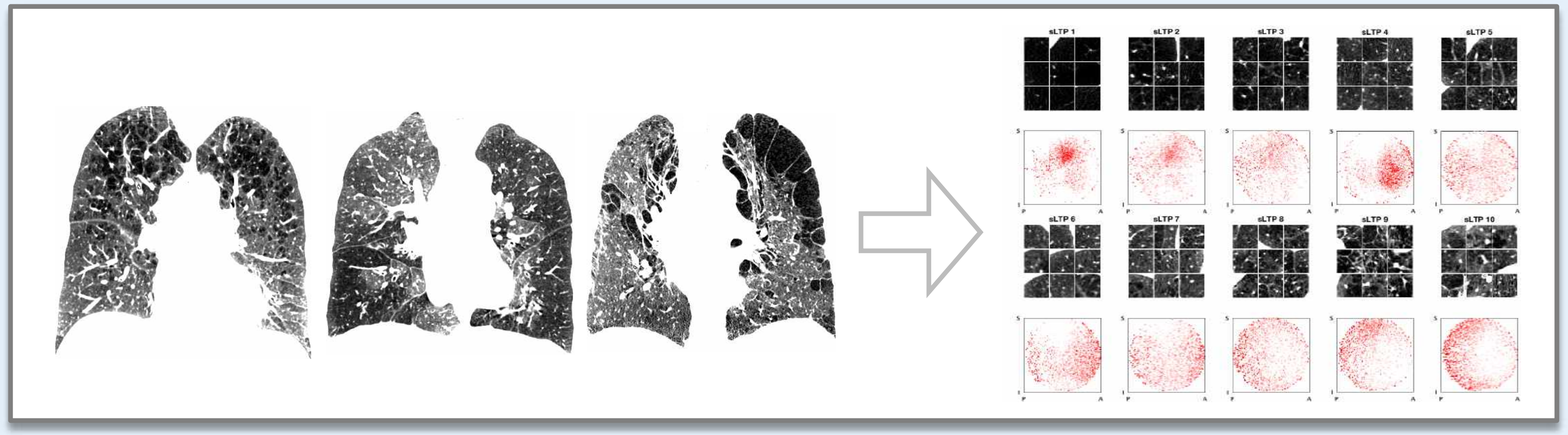
Background: Processing Overview

- Unsupervised machine learning defined novel emphysema patterns on CT.





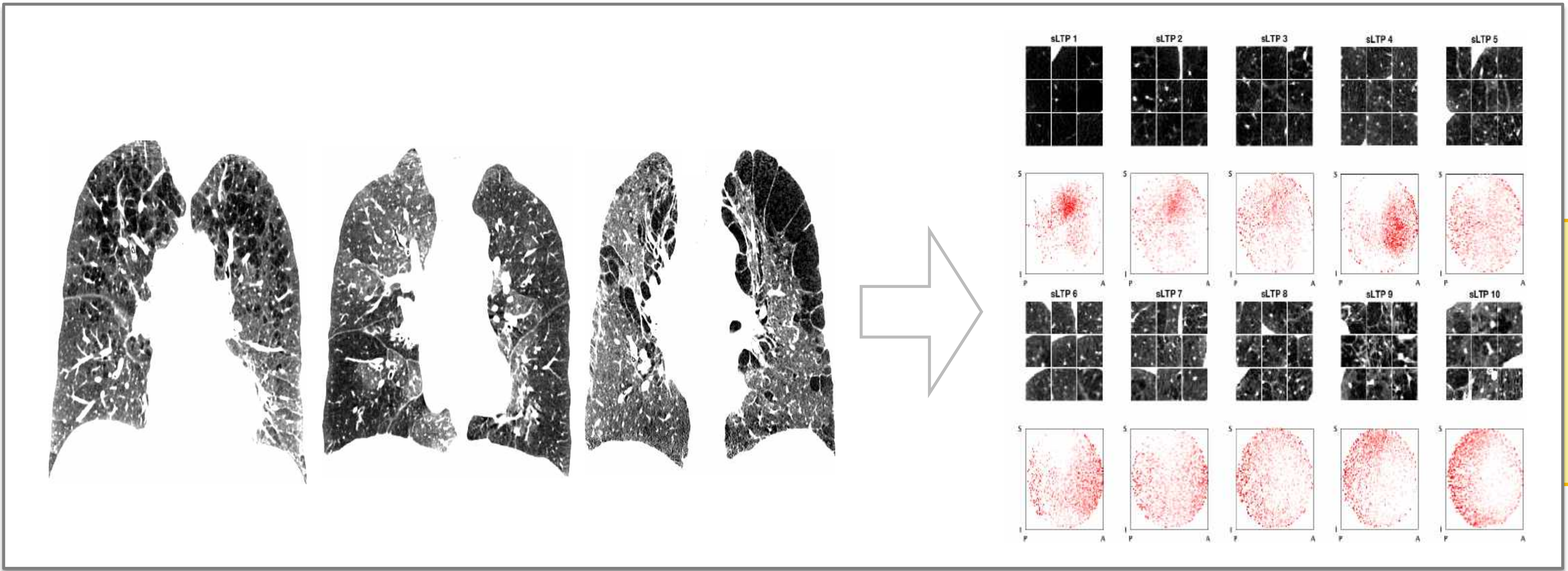
From 3 standard subtypes (1950's) to 10 novel spatial lung texture patterns (sLTP) for emphysema



COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Yang *et al*., Unsupervised Discovery of Spatially-Informed Lung Texture Patterns for Pulmonary Emphysema: The MESA COPD Study. *MICCAI*, 2017

Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema



Challenges

- Learning ***localized texture*** patterns in an ***unsupervised*** manner;
- ***Homogeneity*** vs. ***redundancy*** of the learned patterns.

Summary of Findings:

- Applied a method to standardize ***lung shape spatial mapping***;
- Developed a two-stage ***unsupervised*** framework combining ***spatial and texture*** information;
- Discovered emphysema patterns on ***large COPD cohorts*** with compelling clinical significance.

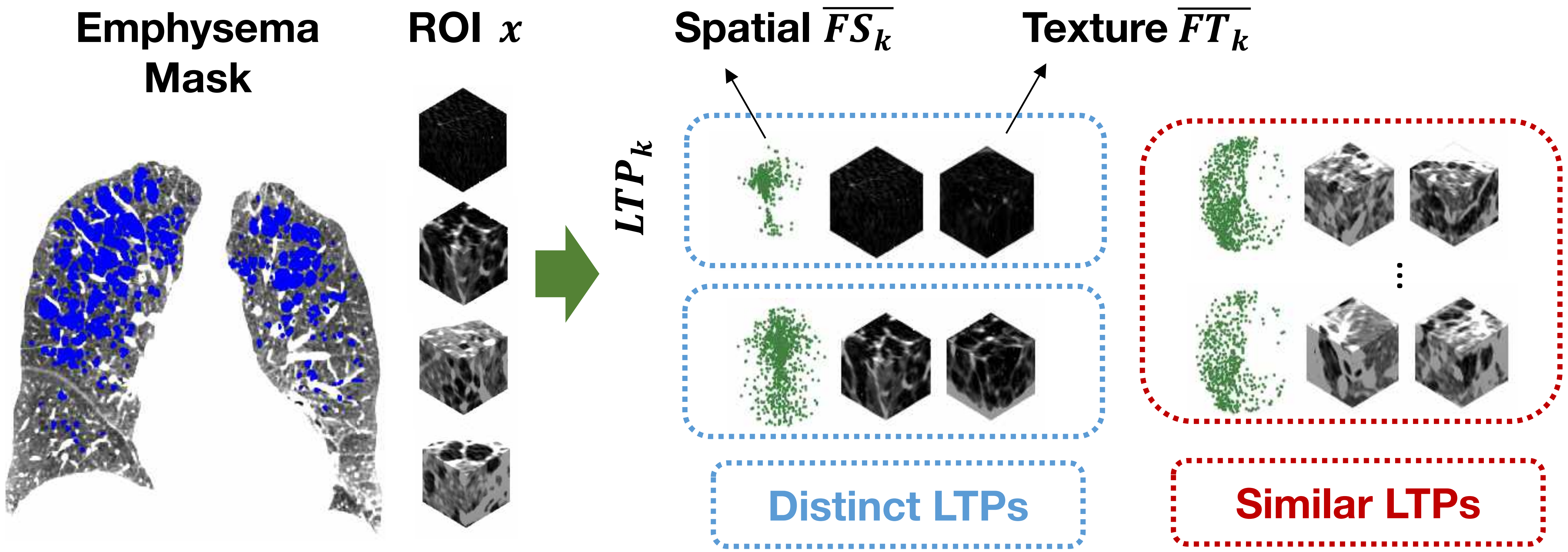
[1] **Jie Yang** *et al.*, Unsupervised Machine Learning to Define Quantitative Subtypes of Pulmonary Emphysema on CT, ***Science*** *(to submit in December 2018).*
[2] **Jie Yang** *et al.*, Unsupervised Discovery of Spatially-Informed Lung Texture Patterns for Pulmonary Emphysema: The MESA COPD Study. ***MICCAI***, 2017.

Method

Unsupervised learning of localized texture patterns for pulmonary emphysema

2 **Learning Stage 1:**

➢ **Augmented Lung Texture Patterns (LTPs)**



FT_x / FS_x = texture / spatial feature of ROI x

$\overline{FT_k}$ / $\overline{FS_k}$ = texture / spatial centroid of LTP_k

Iteratively update ROI assignment $\Lambda_k^{(t)}$ of $\mathrm{LTP_k}$, by ***minimizing a dedicated cost function*** [1] :

$$\chi^2\left(FT_x, \overline{FT_k^{(t-1)}}\right) \quad \text{Texture distance}$$

$$+$$

$$\omega \cdot W \cdot \left\| FS_x, \overline{FS_k^{(t-1)}} \right\|_2^2 \quad \text{Spatial regularization}$$

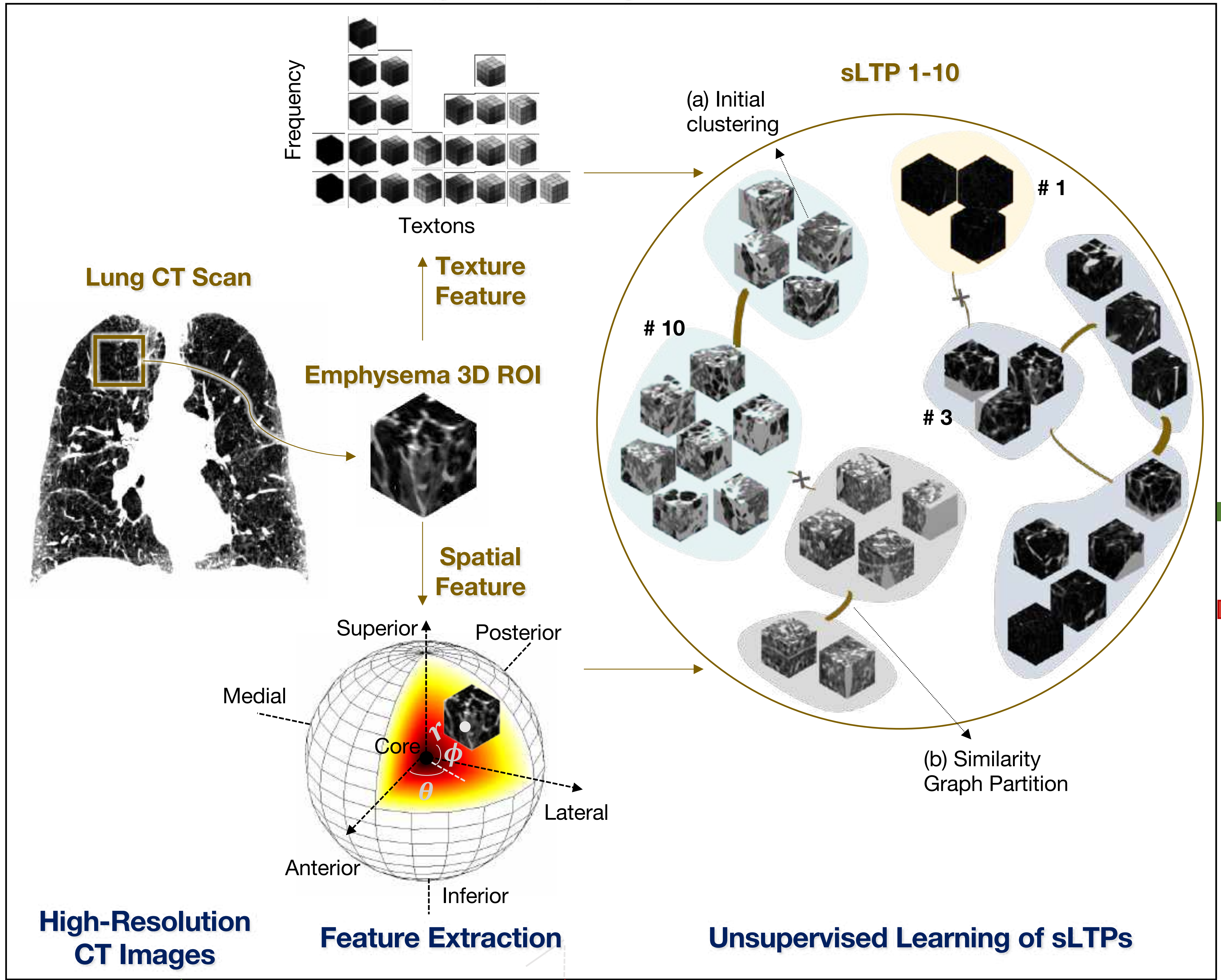
$$+$$

$$\gamma \cdot \mathbb{I}\left(\chi^2\left(FT_x, \overline{FT_k^{(t-1)}}\right) > thresh_{\chi^2}\right) \quad \text{Texture penalty}$$

Learning Pipeline and Results in SPIROMICS and MESA Lung Study

Unsupervised learning of localized texture patterns for pulmonary emphysema

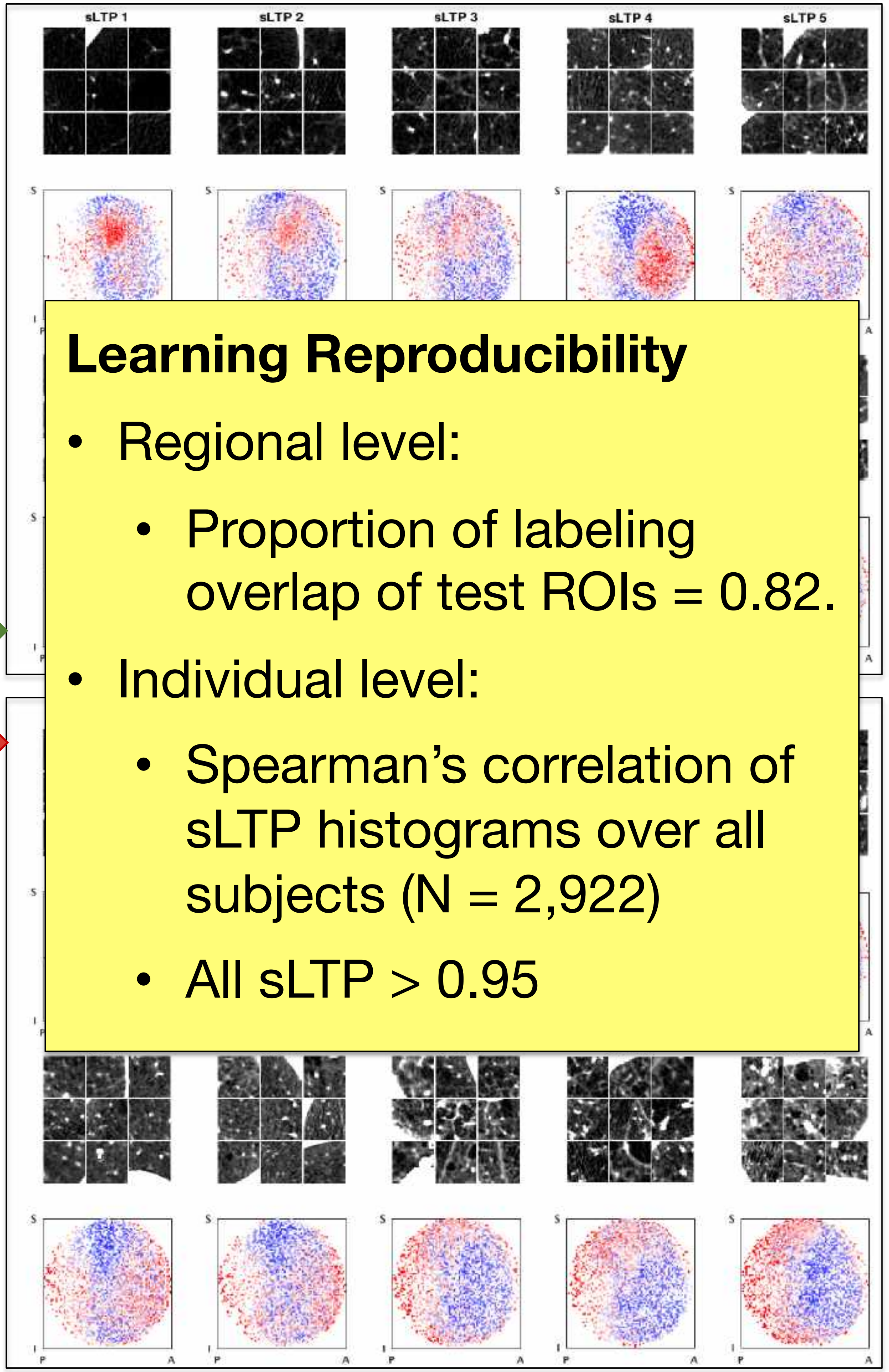
DATA: SPIROMICS (N=2,922)



Training Set 1 (N = 1,462) | Training Set 2 (N = 1,460)

Training Set 1 (N = 1,462)

Training Set 2 (N = 1,460)



Learning Reproducibility

- Regional level:
  - Proportion of labeling overlap of test ROIs = 0.82.
- Individual level:
  - Spearman's correlation of sLTP histograms over all subjects (N = 2,922)
  - All sLTP > 0.95

Spatial density

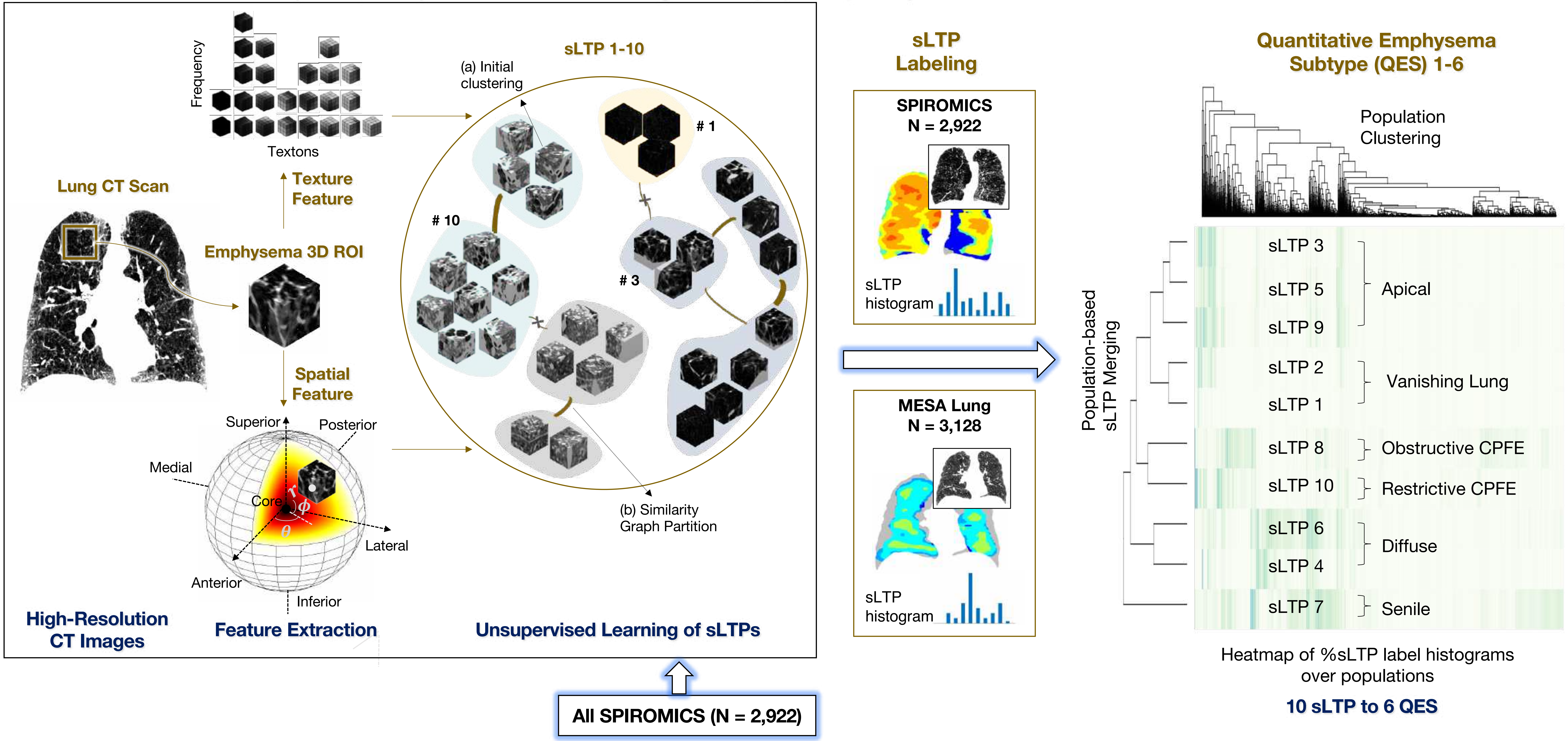
0 1 2 3 4

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Learning Pipeline and Results in SPIROMICS and MESA Lung Study

Unsupervised learning of localized texture patterns for pulmonary emphysema

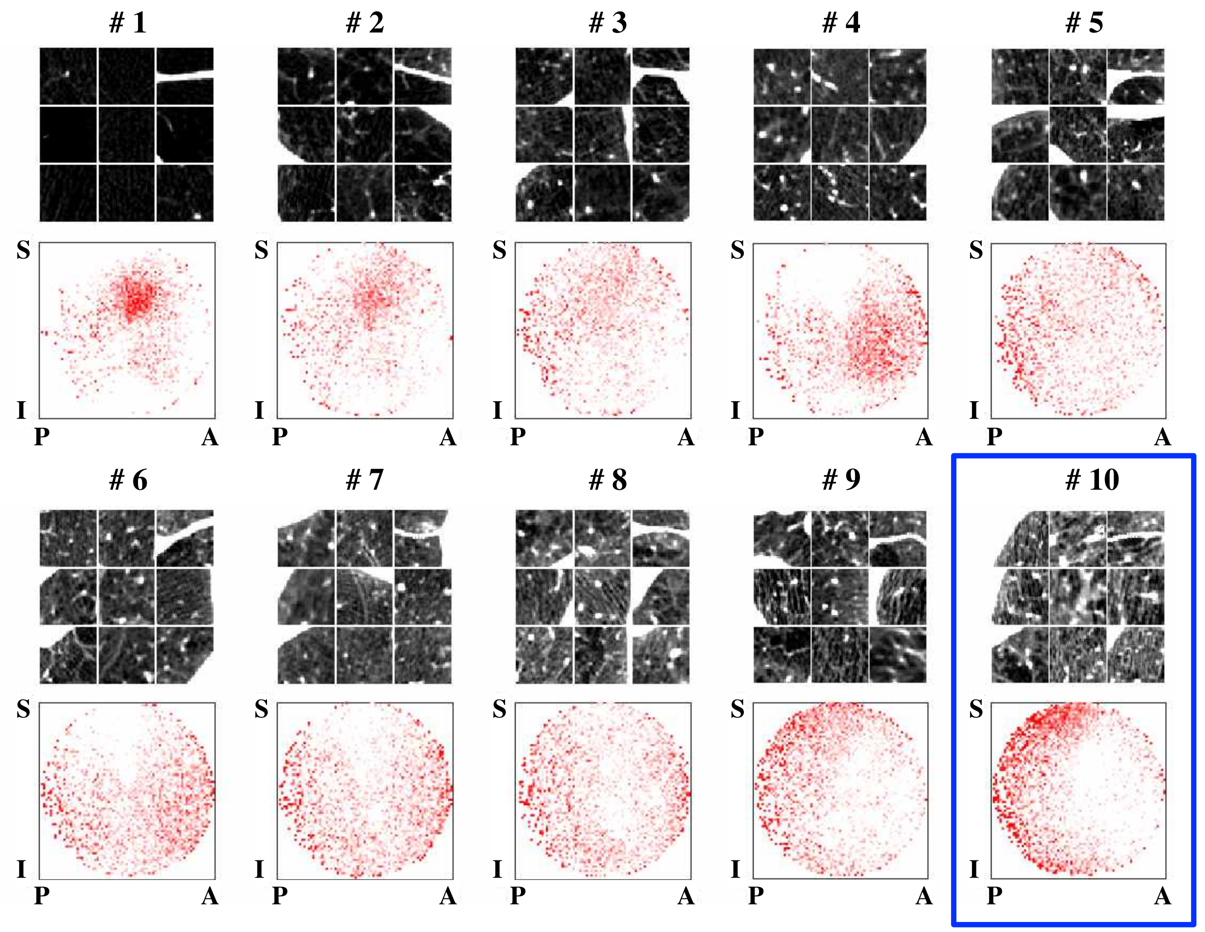
DATA: SPIROMICS (N=2,922) and MESA Lung Study (N=3,128)

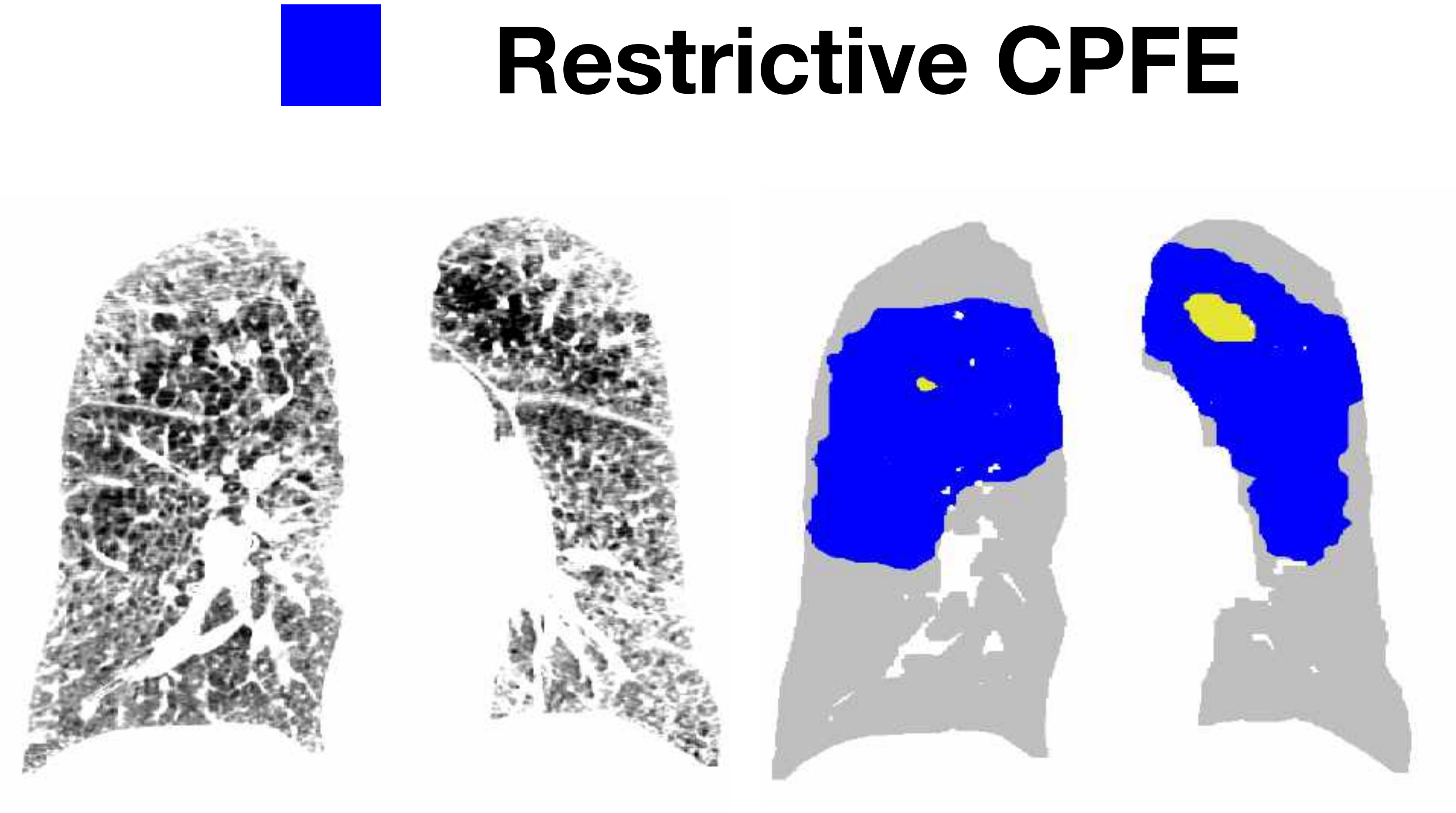


COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Experimental Results in SPIROMICS and MESA Lung Study

Unsupervised learning of localized texture patterns for pulmonary emphysema





Apical	Diffuse	Senile	Restrictive CPFE	Obstructive CPFE	Vanishing Lung

<u>Associated with</u>

Dyspnea
Hypoxemia at rest
Desaturation on exertion
↓↓6MWT
Exacerbations

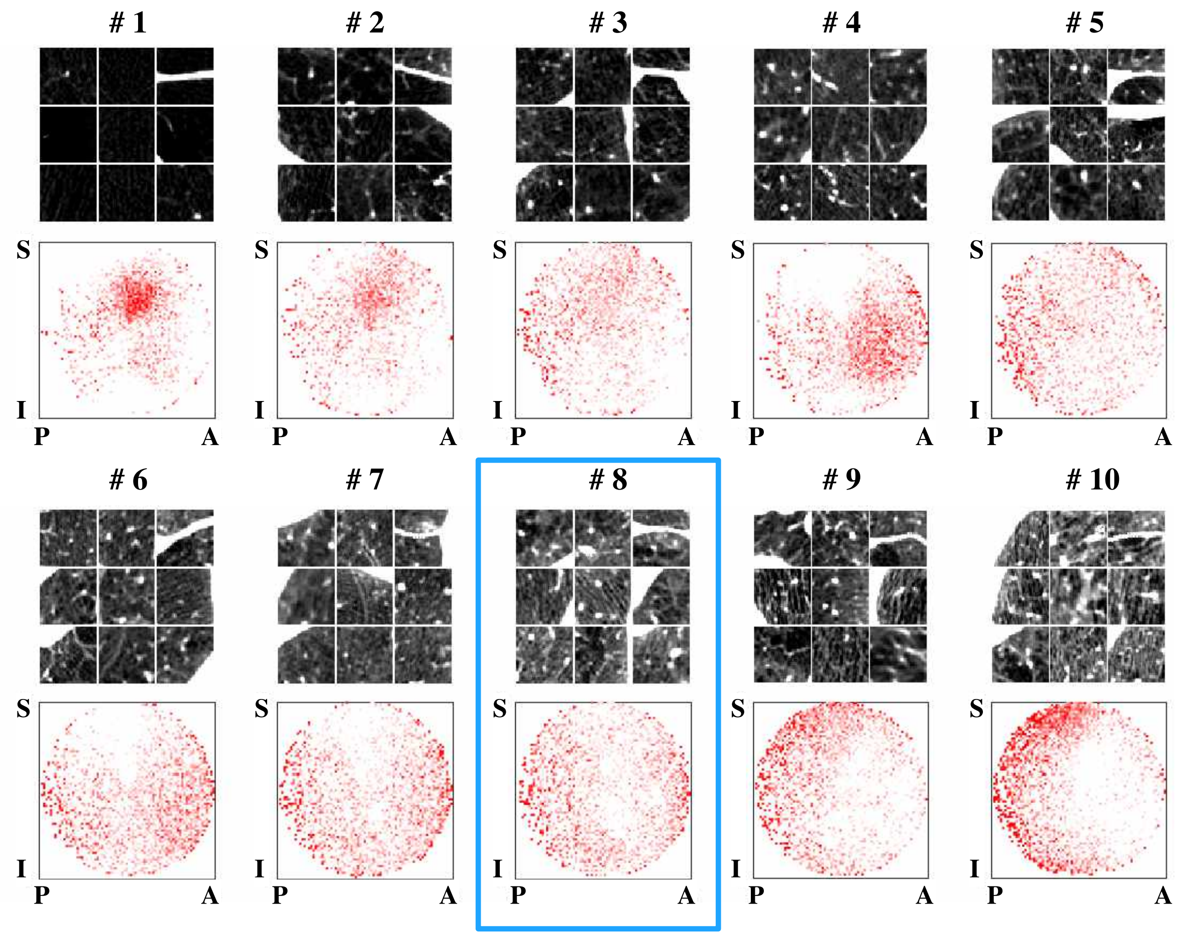
↓ FEV_1
FVC
FEV_1/FVC
↓↓ TLV on CT

Adjusted for age, sex, race/ethnicity, height, weight, smoking status, pack-years, COPD, scanner manufacturer, FEV1, other QES.

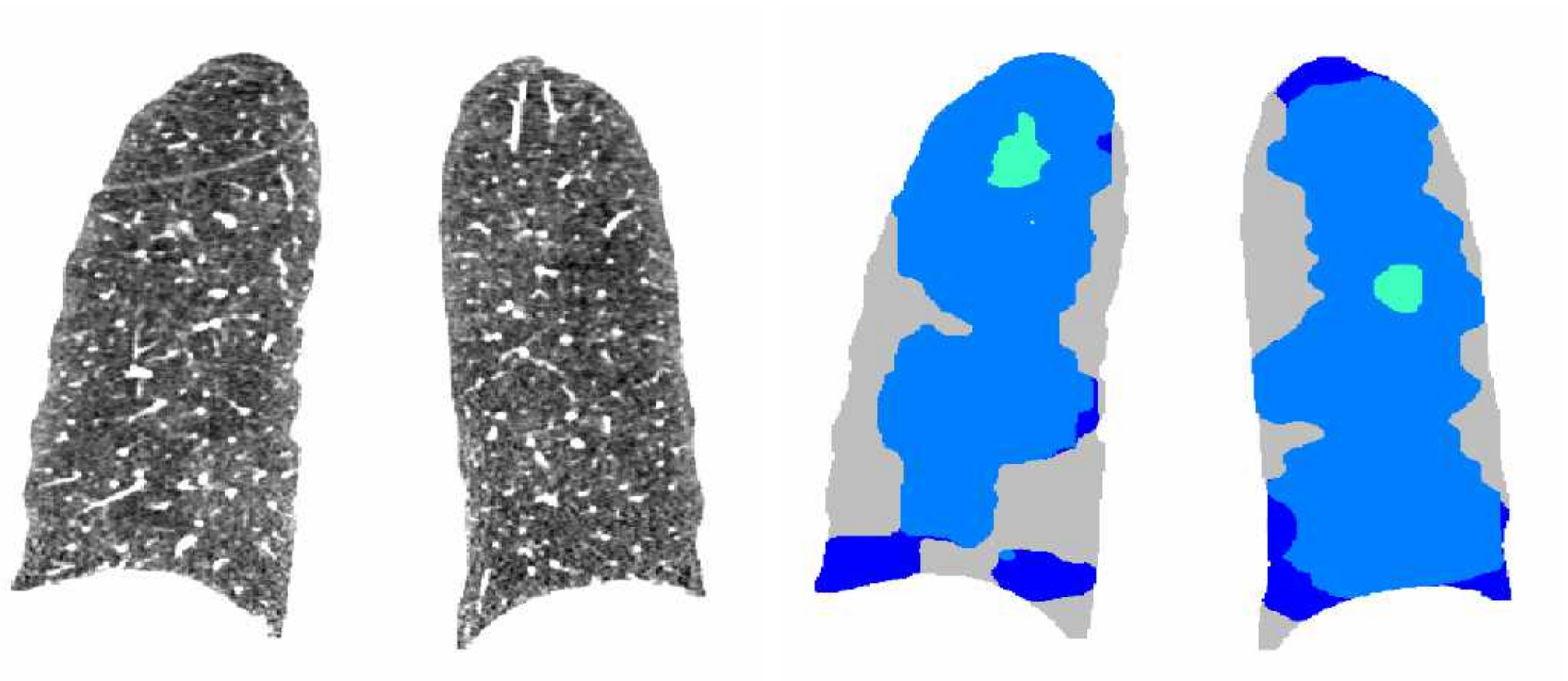
COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Experimental Results in SPIROMICS and MESA Lung Study

Unsupervised learning of localized texture patterns for pulmonary emphysema



Obstructive CPFE



Apical	Diffuse	Senile	Restrictive CPFE	Obstructive CPFE	Vanishing Lung

<u>Associated with</u>

Desaturation on exertion

↓ FEV_1
↓ FVC
↓ FEV_1/FVC
↑ TLV on CT

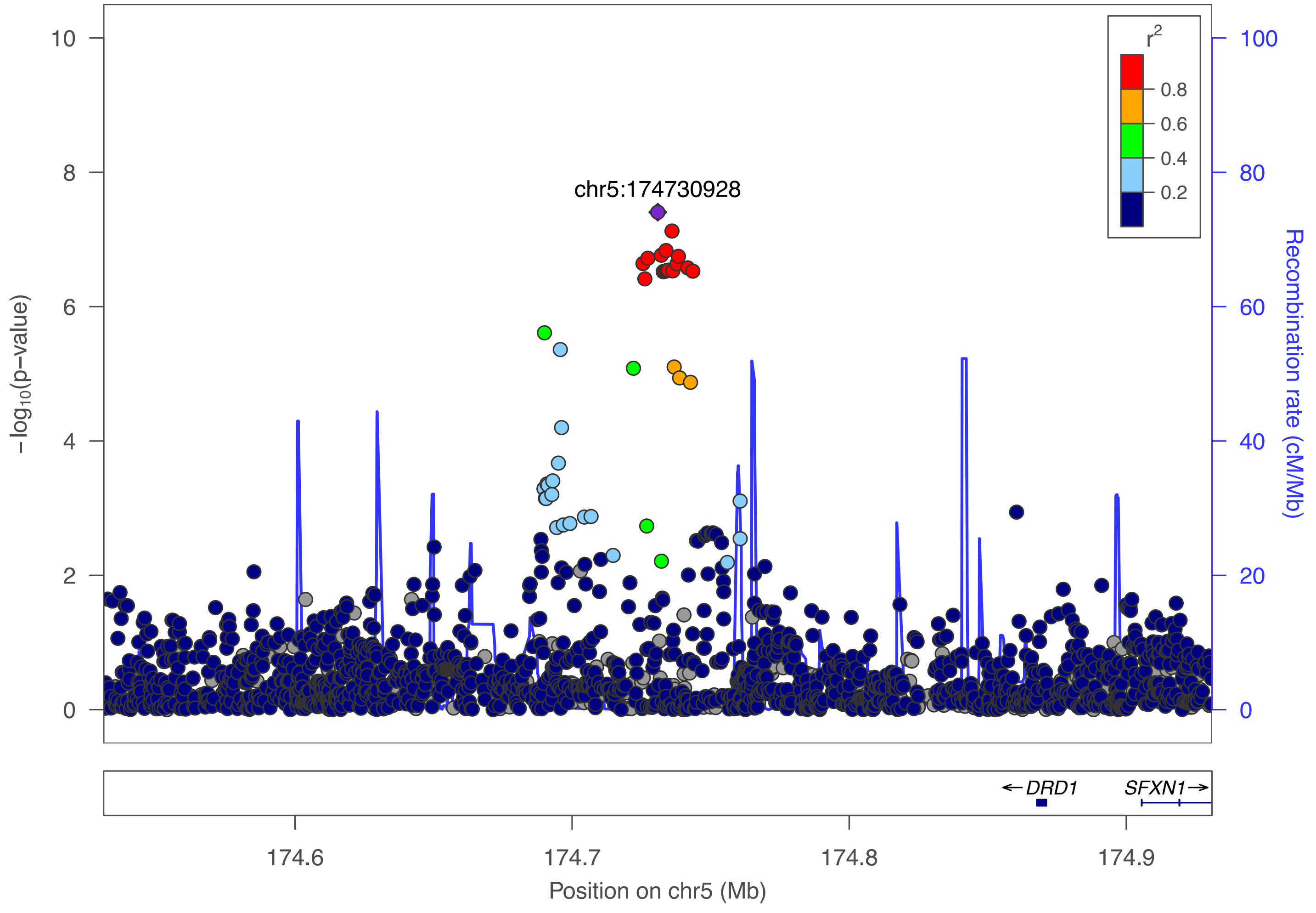
Adjusted for age, sex, race/ethnicity, height, weight, smoking status, pack-years, COPD, scanner manufacturer, FEV1, other QES.

COPD death

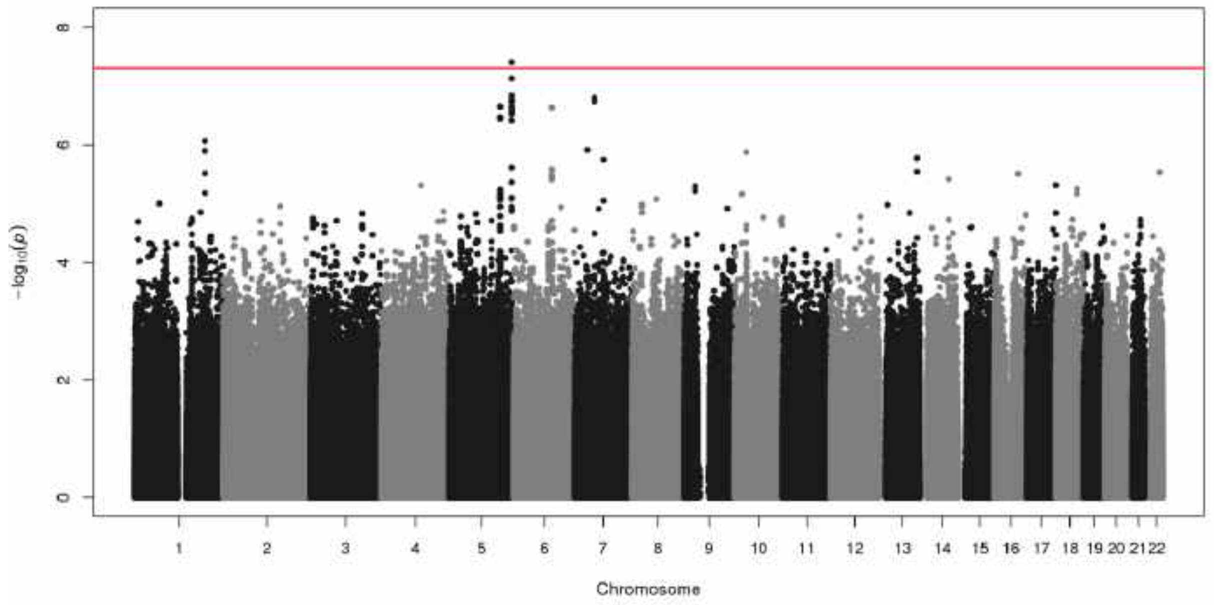
Columbia University
in the City of New York

Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema

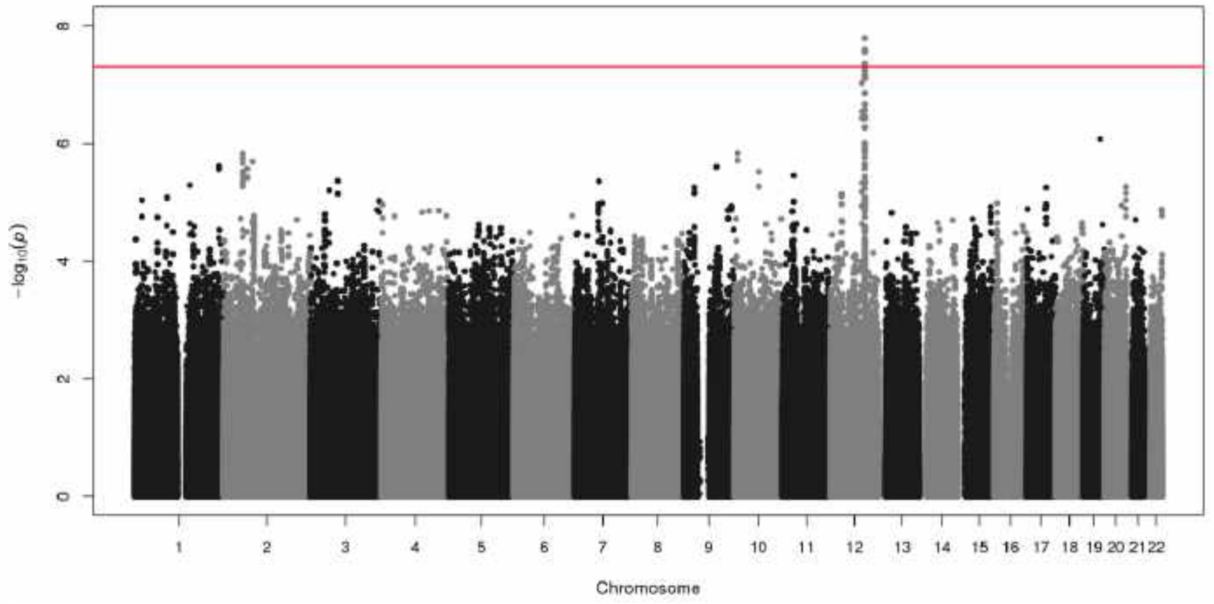
- **GWAS results : ***
  - **5 genetic variants for four QES**
  - **Apical QES: DRD1**



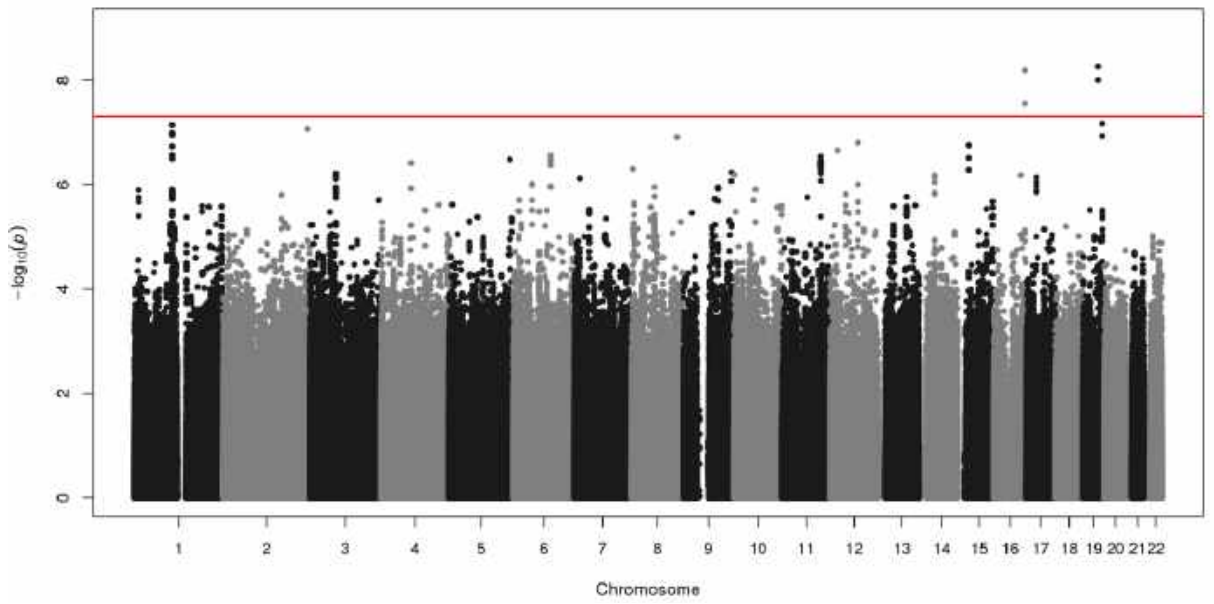
Severe Apical QES



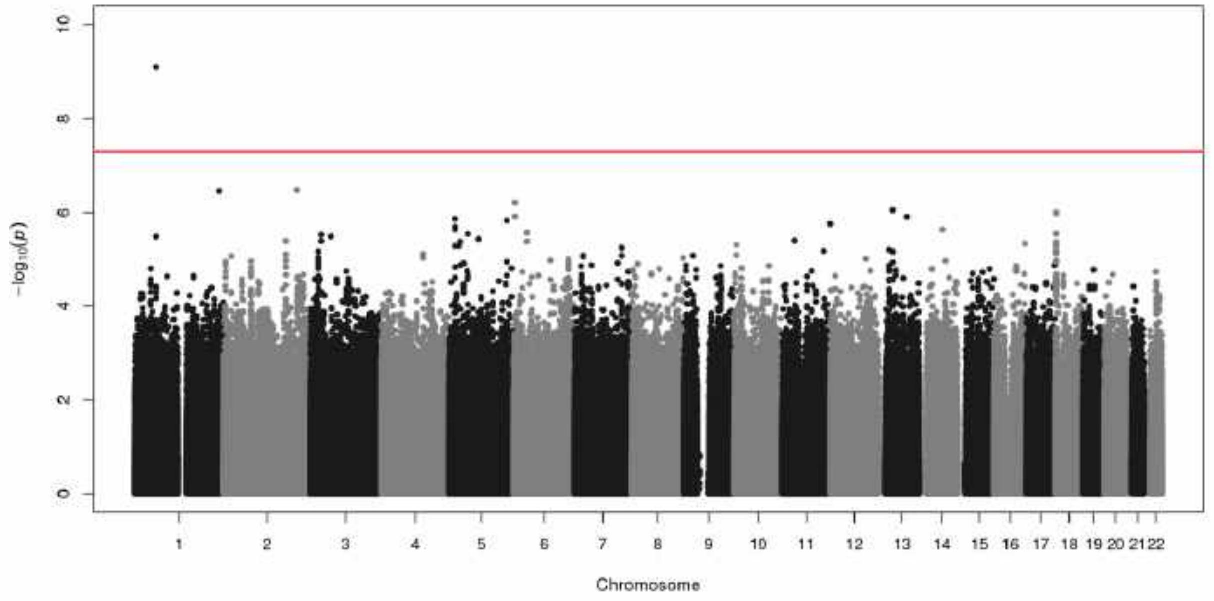
Restrictive CPFE QES



Obstructive CPFE QES



Severe Vanishing Lung QES



* University of Virginia

Summary

Unsupervised learning of localized texture patterns for pulmonary emphysema

- Novel **unsupervised learning** of emphysema patterns on CT:
  - A standardized lung shape **spatial mapping**;
  - A **two-stage** learning framework.
- Applied on large COPD and controls yielded:
  - 10 **highly-reproducible** sLTPs;
  - Six quantitative emphysema subtypes, associated independently with distinct **symptoms**, **lung function changes** and **mortality**.
- Enables:
  - **Novel definitions** of emphysema subtypes;
  - CT-based **emphysema-specific signatures (biomarkers)** of the lungs;
  - May facilitate future study for understanding COPD and emphysema, and the design of personalized / gene / drug therapies.

Other Evaluations:

- GWAS; 3 “hits” reproducible
- Extensive evaluation of sLTP reproducibility in MESA COPD (N = 317);
- Linking sLTP and standard emphysema subtypes in MESA COPD (N=317).

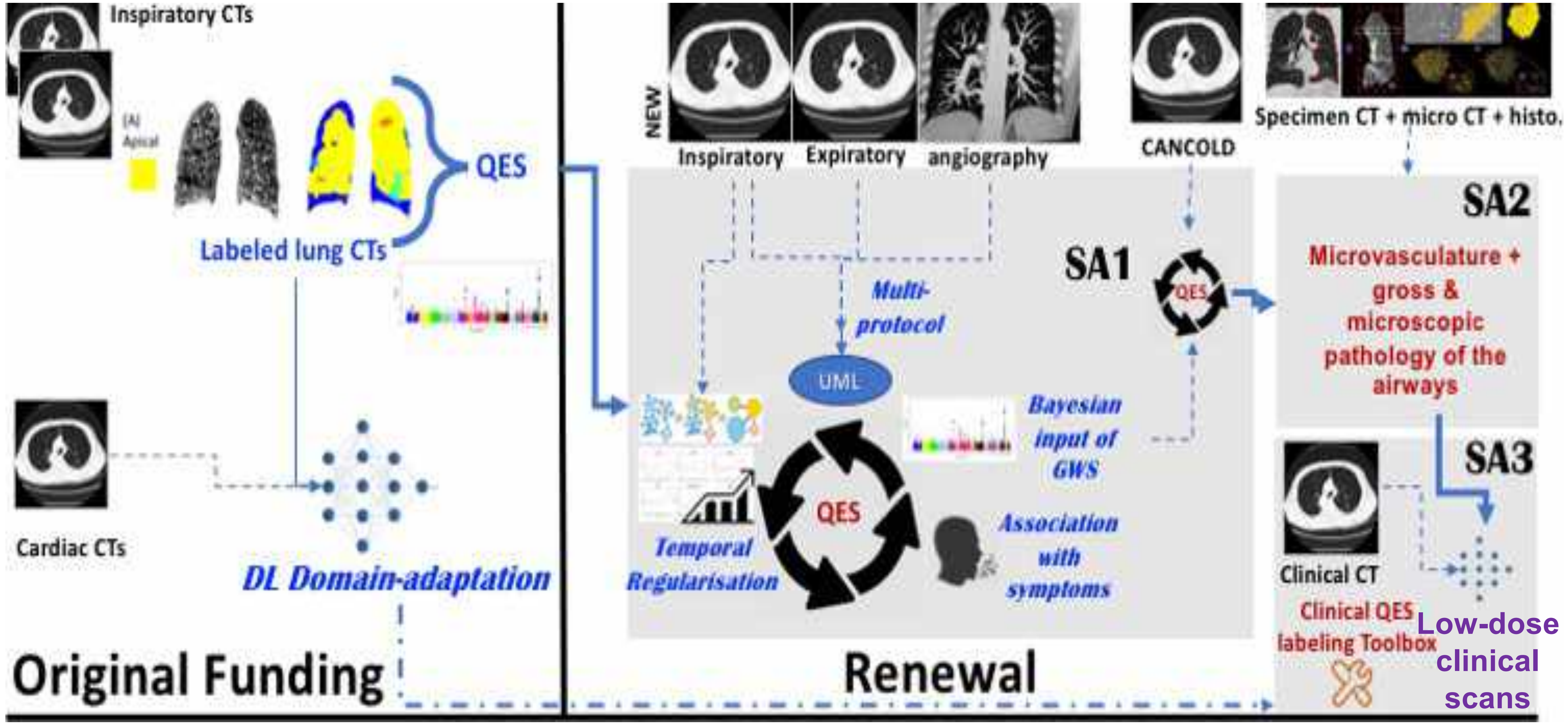
What’s next?

- **Deep & Molecular QES**

MESA Exam 6

& CANCOLD

Explanted lungs with microCT



Low-dose clinical scans

- **Quantitative airway tree subtyping**

Role of variants in COPD?

* COVID-19 vasculature

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Acknowledgements

Collaborators

Elsa D. Angelini, Ph.D.

R. Graham Barr, M.D., Ph.D.

Heffner Biomedical Imaging Lab

Jie Yang (Ph.D. Student, 2019)

Xinyang Feng (Ph.D. student, 2019)

Yrjö Häme (Ph.D. student)
Yu Gan (Post-doc)
Thomas Vetterli (M.Sc. Intern)

Additional Collaborators

Columbia University Medical Center:

Pallavi P. Balte, Ph.D.
John H.M. Austin, M.D.
Benjamin M. Smith, M.D.
Yifei Sun, Ph.D.
Wei Shen, M.D.

Iowa University:

Eric A. Hoffman, Ph.D.

University of Virginia:

Ani Manichaikul, Ph.D.

And the MESA, SPIROMICS Investigators!

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Heffner Biomedical Imaging Laboratory

https://hbil.bme.columbia.edu

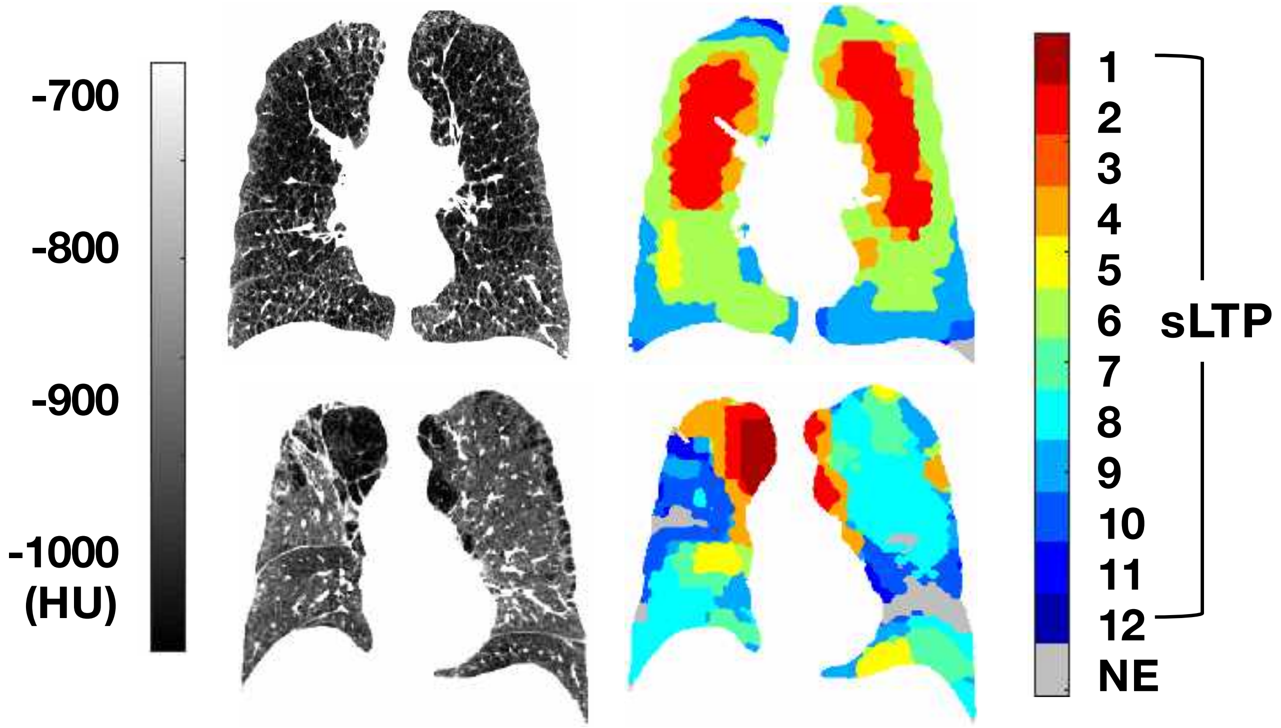
Thank you!

Extra Slides

Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema

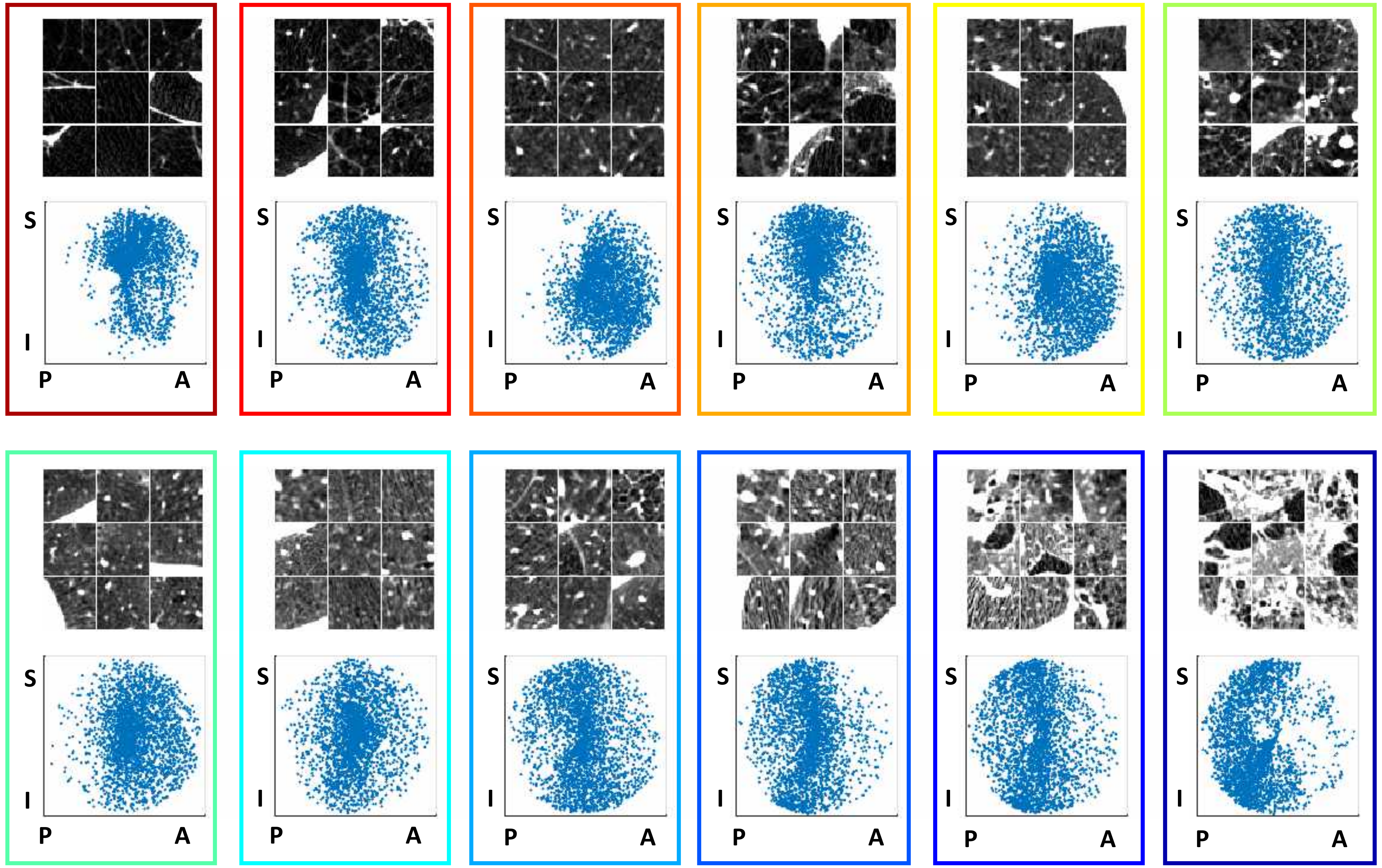
- **Learn sLTPs in MESA COPD study:**
  - N = 317 full-lung CT scans.

sLTP labeling: ***lung texture signature***

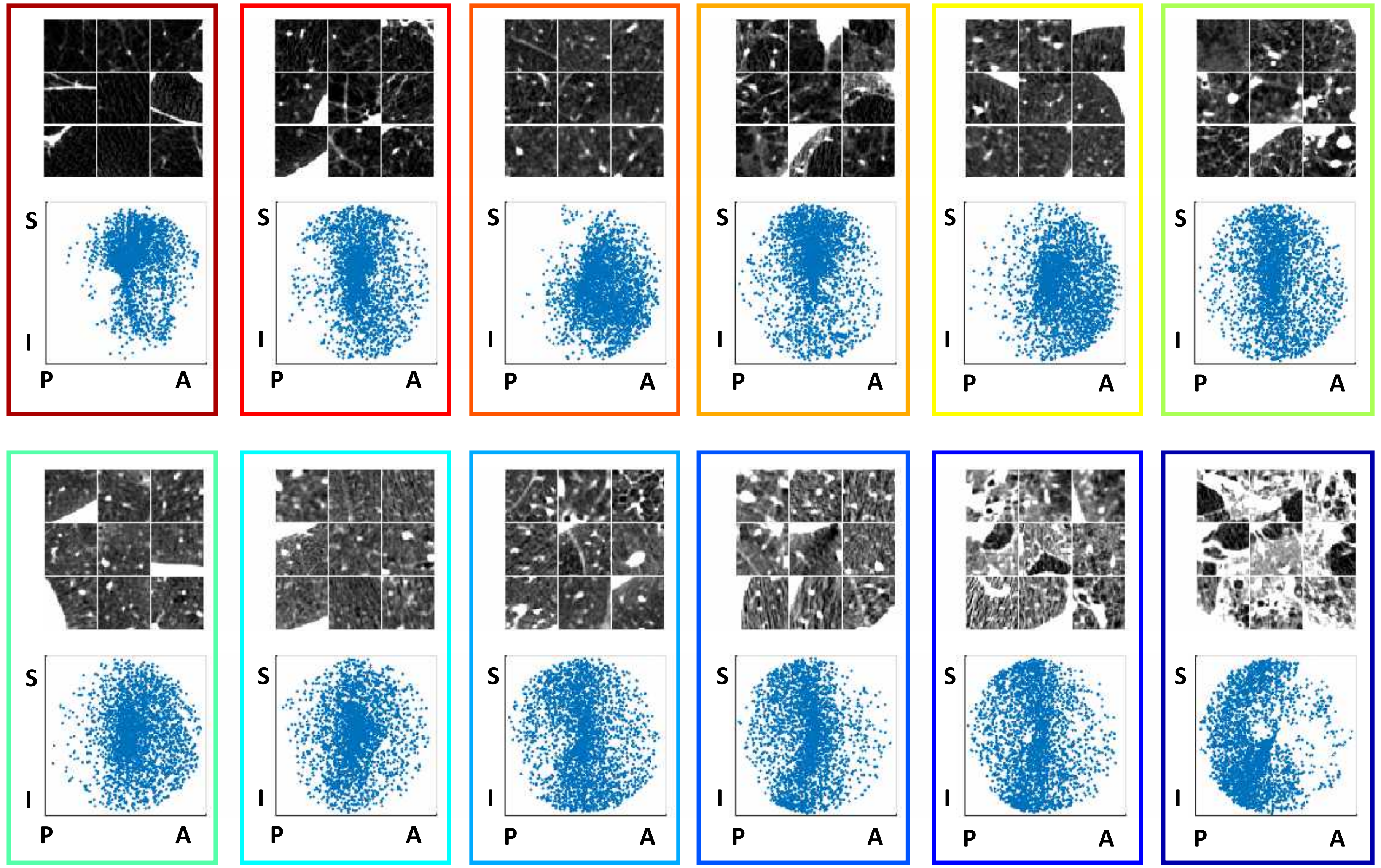


12 sLTPs discovered, ordered by average intensity

- **Training ROI:** $\%emph > 1\%$



Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema



Global label $H_g = [\%CLE, \%PLE, \%PSE, \%NE] = [P(L(x) = C_1), \dots, P(L(x) = C_4)]$

sLTP histogram $H_p = \left[\%sLTP_1, \dots, \%sLTP_{12}, \%\widehat{NE}\right] = [P(F(x) = p_1), \dots, P(F(x) = p_{13})]$

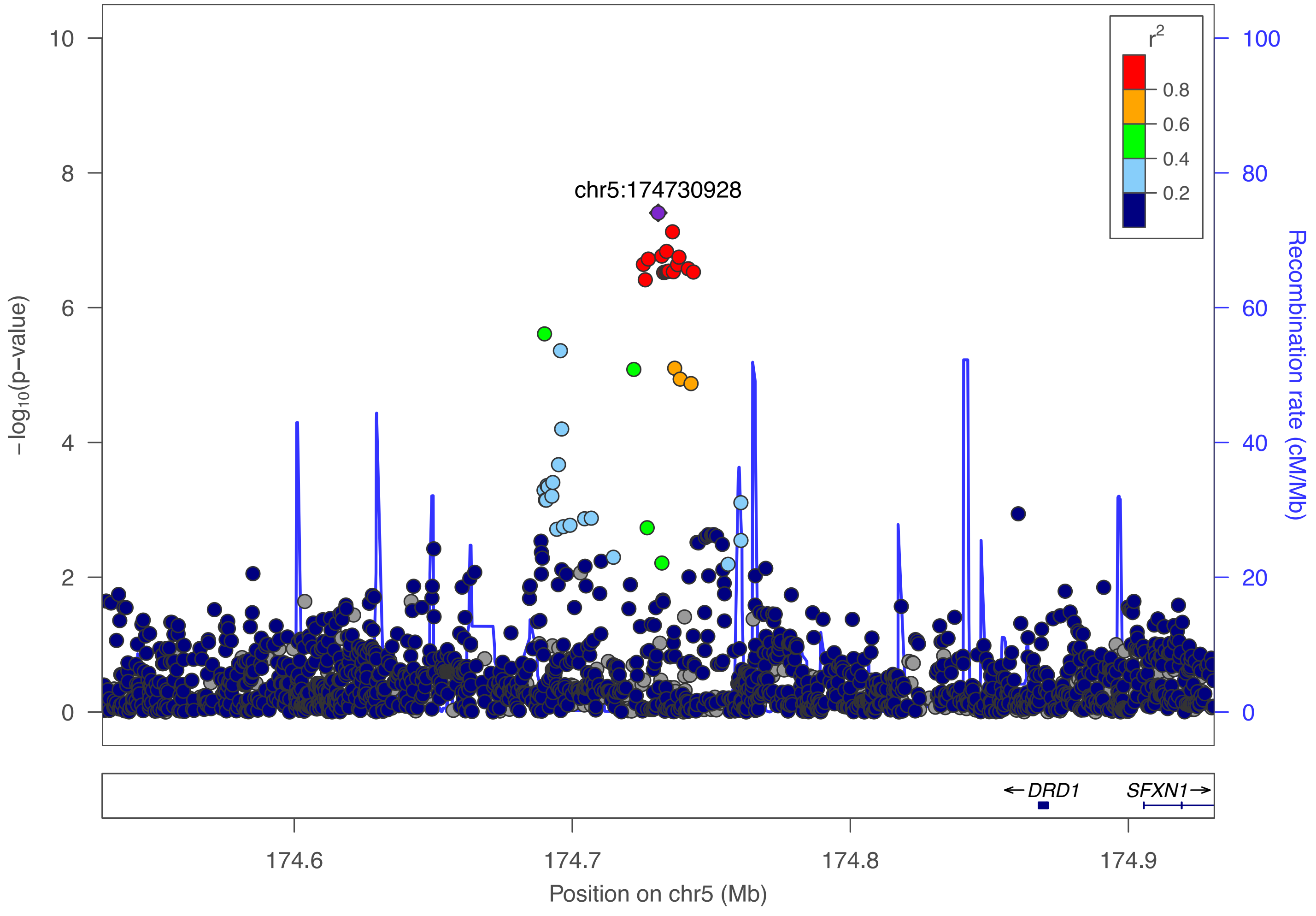
$$P(L(x) = C_i) = \sum_{k=1}^{13} P(L(x) = C_i | P(F(x) = p_k)) P(F(x) = p_k)$$

$$\boldsymbol{H_g = H_p A}$$

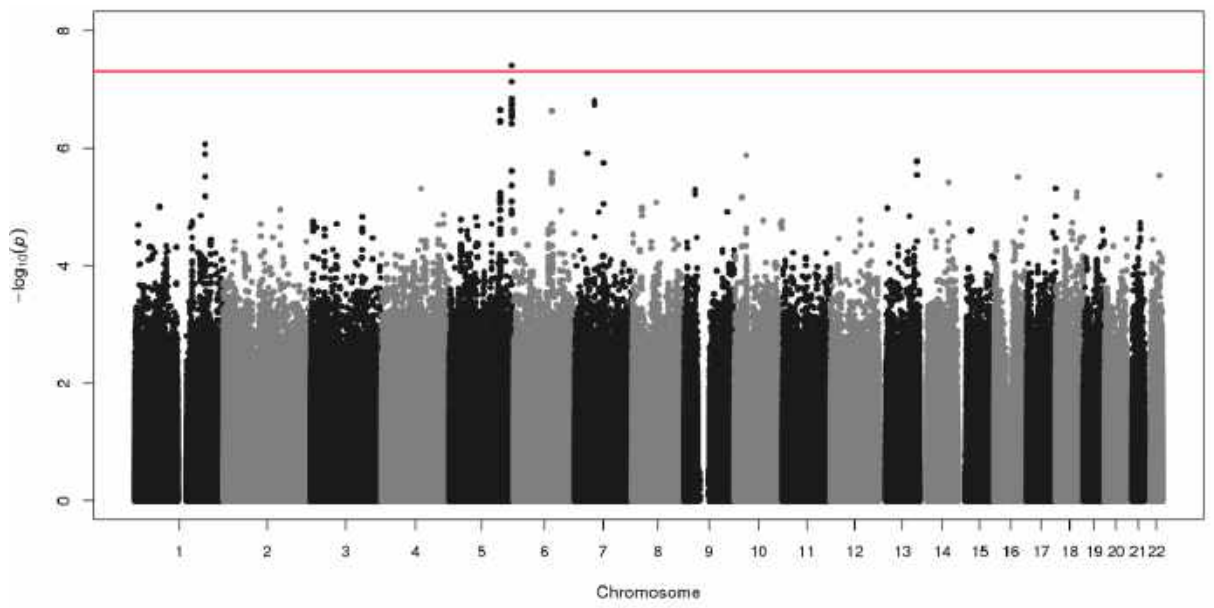
$$\boldsymbol{A_{k,i} = P(L(x) = C_i | F(x) = p_k)}$$

Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema

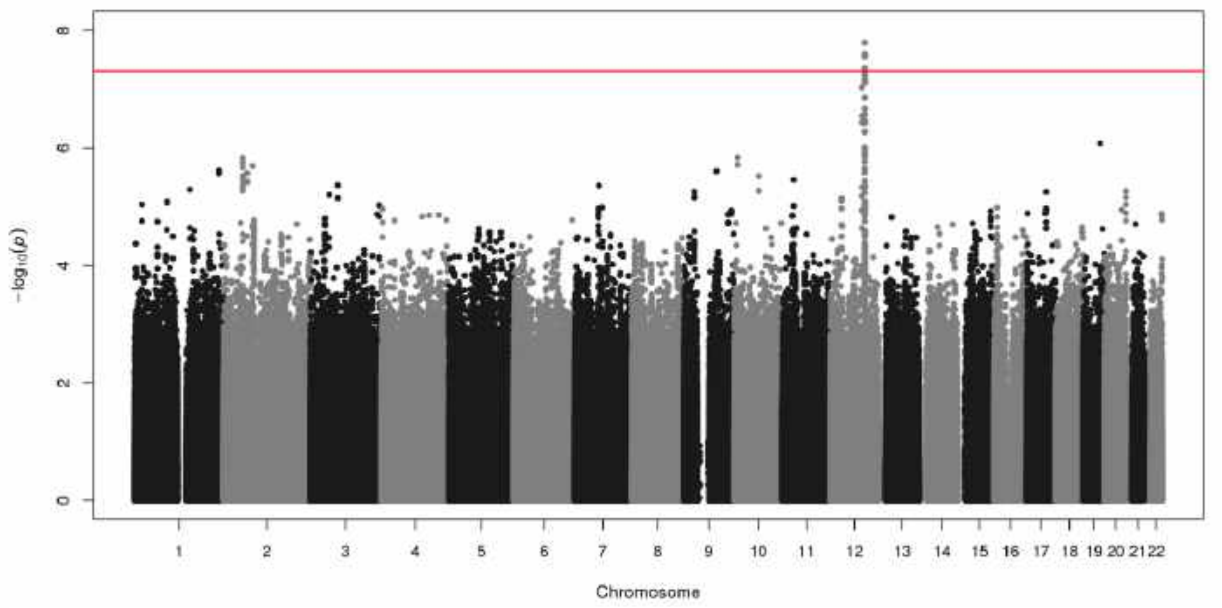
- **GWAS results:**
  - **5 genetic variants for four QES**
  - **Apical QES: DRD1**



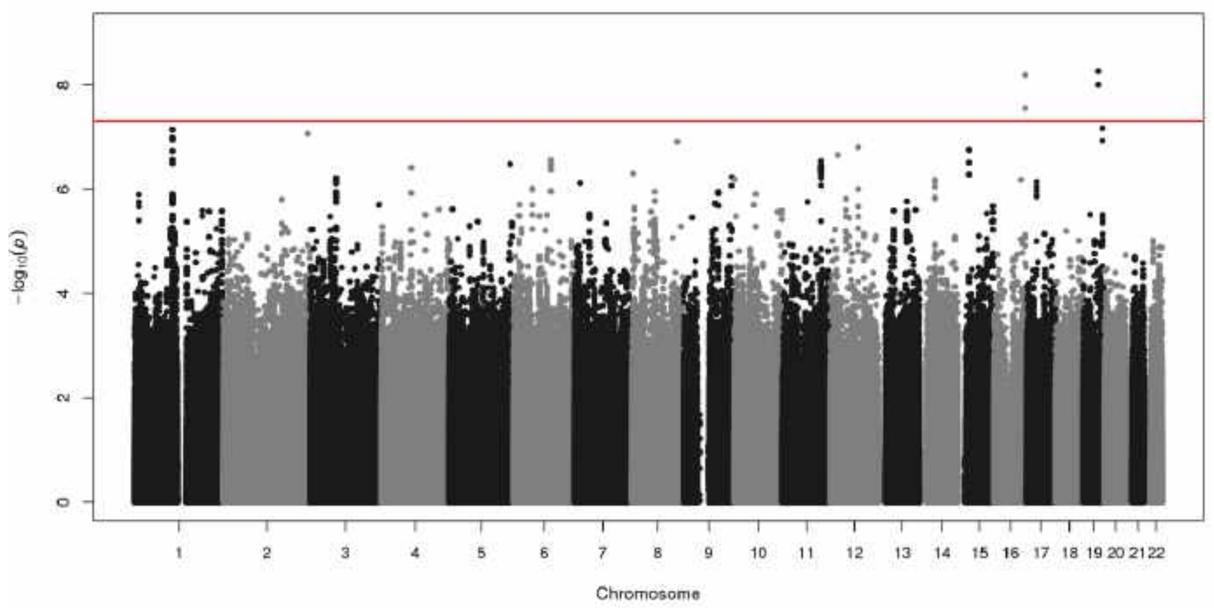
Severe Apical QES



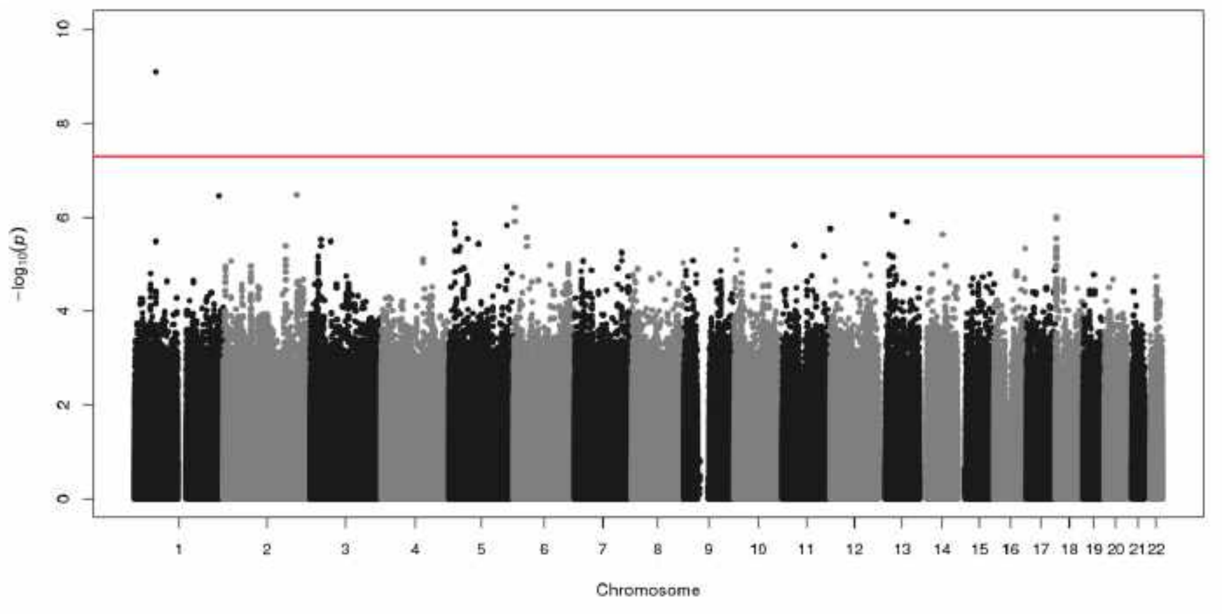
Restrictive CPFE QES



Obstructive CPFE QES

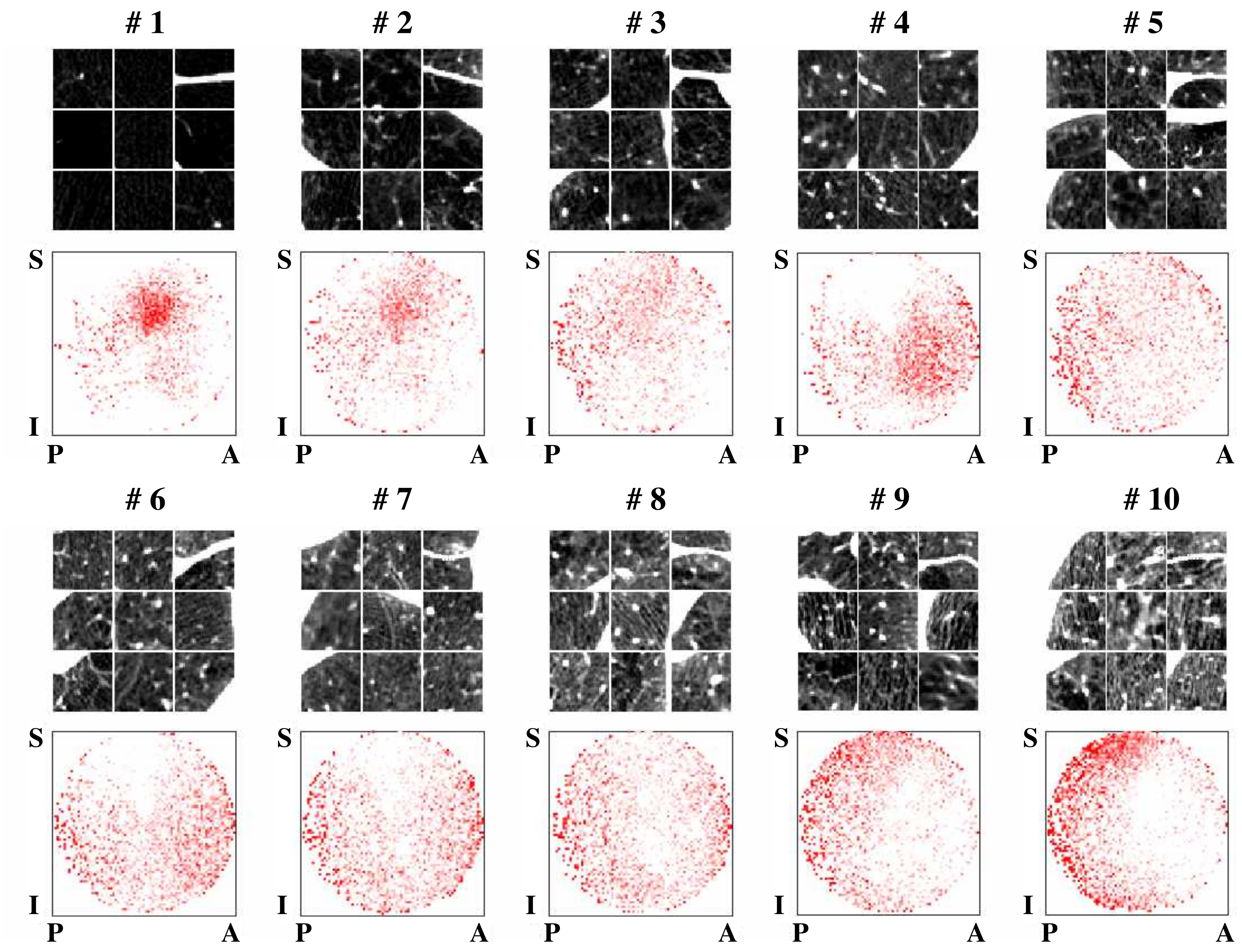


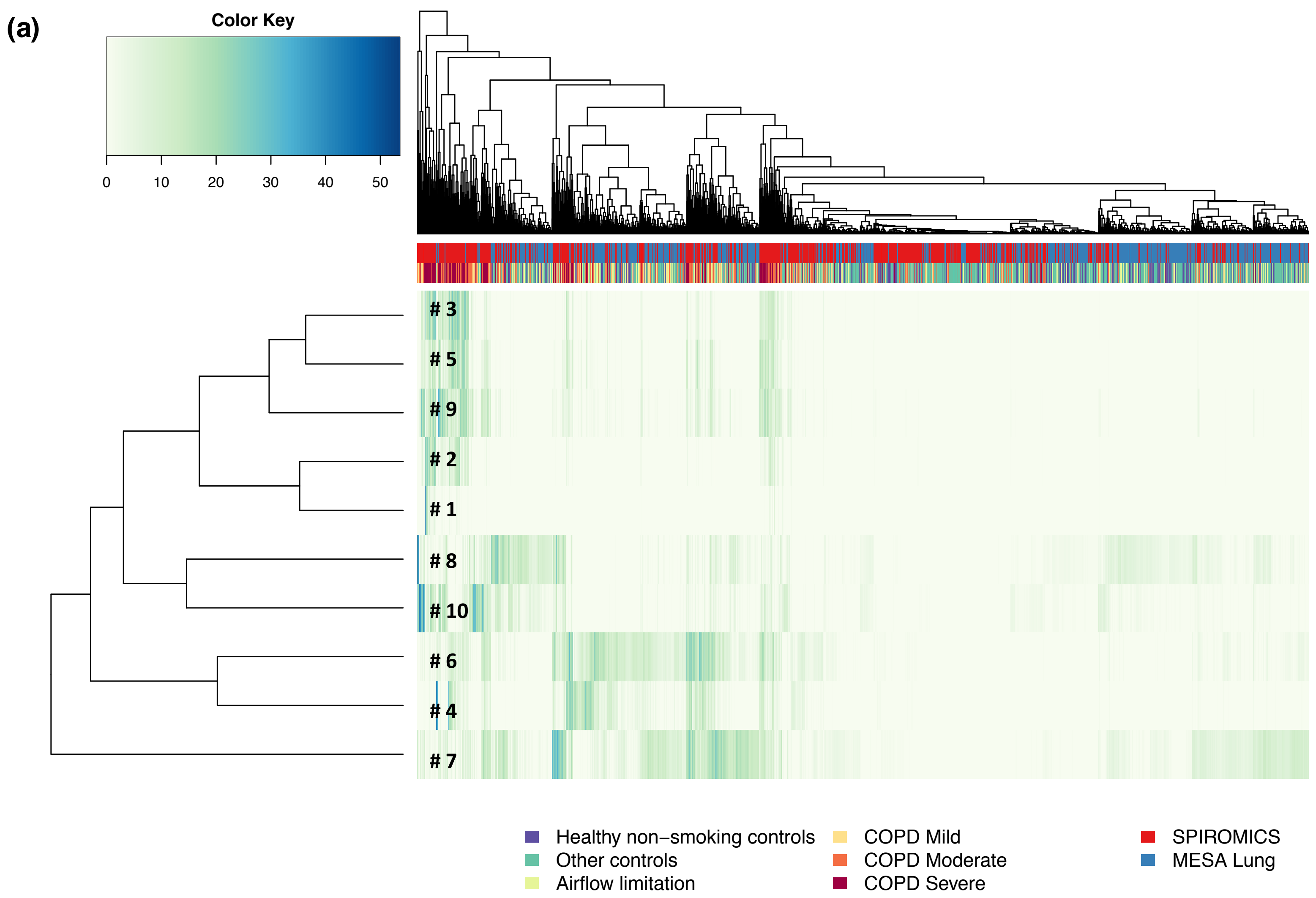
Severe Vanishing Lung QES

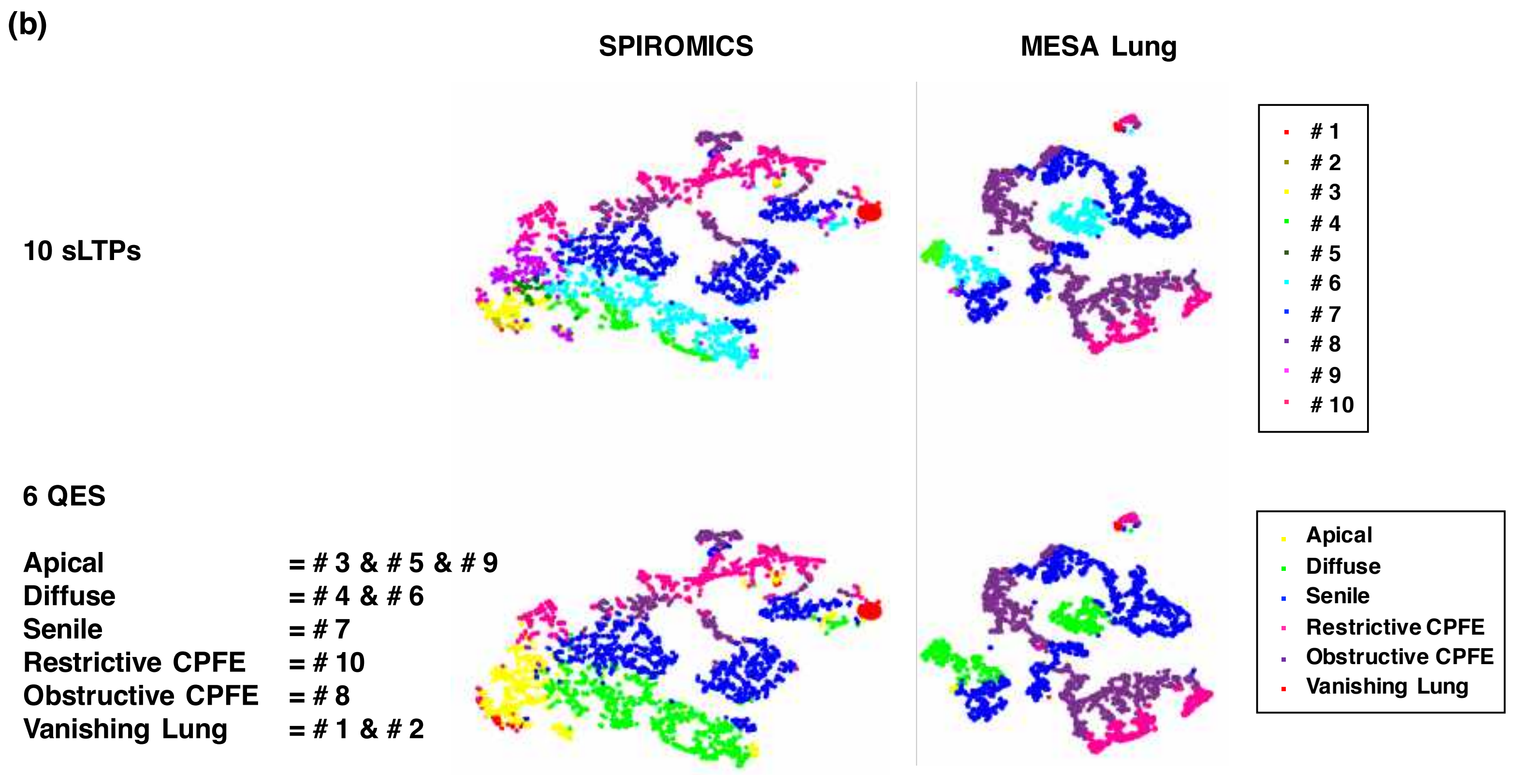


Columbia University
In the City of New York

Unsupervised Learning of Localized Texture Patterns for Pulmonary Emphysema

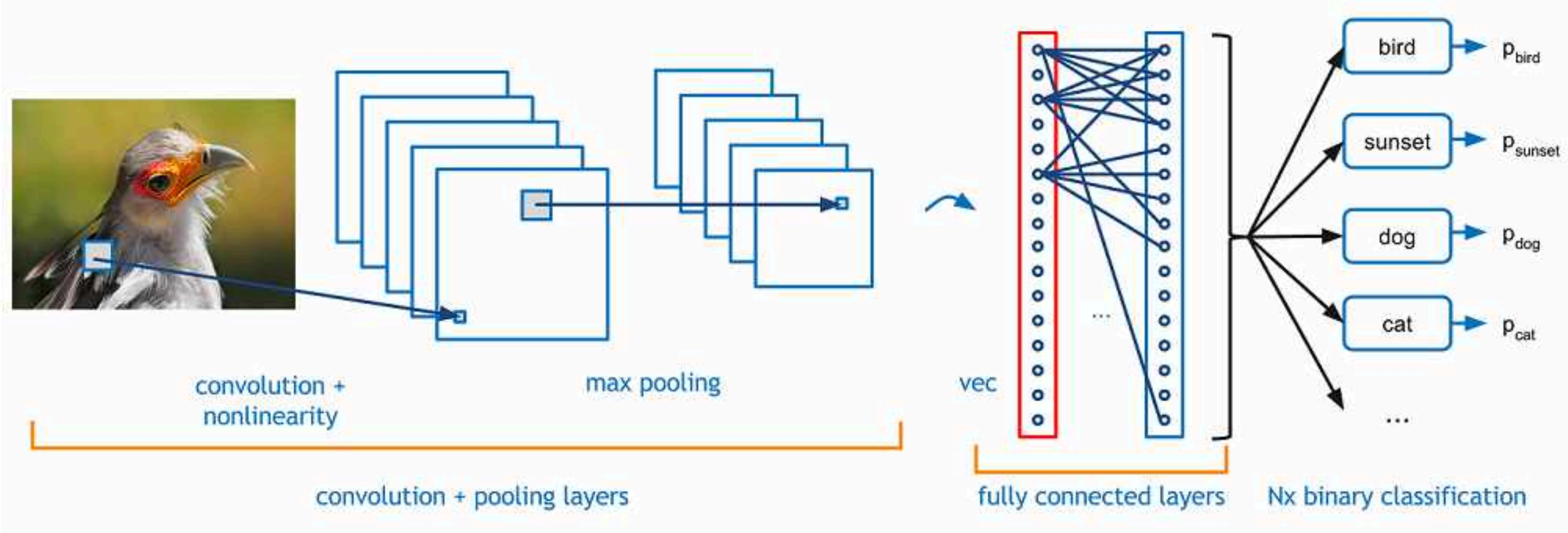






COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

CNN Background



h_i^l $=$ i-th feature map in layer l

h_k^{l-1} $=$ k-th feature map in layer $l-1$

$$h_i^l = \sigma \cdot (\sum_k h_k^{l-1} * W_{ki}^l + b_i^l)$$

$$h^l = \sigma \cdot (\sum_k h^{l-1} W^l + b_l)$$

h^l $=$ feature vector $\in \mathbb{R}^Q$ in layer l

h^{l-1} $=$ feature vector $\in \mathbb{R}^P$ in layer l-1