Workshop on Application of Biomechanics in Clinical Treatment and Rehabilitation
& 1st International Conference on "Medicine in Novel Technology and Devices
August 11, 2019, 13:55PM

Machine Learning of CT-based Imaging Clusters in Asthma and Chronic Obstructive Pulmonary Disease (COPD)

Ching-Long Lin 林慶龍
The University of Iowa, Iowa City, IA, USA

Motivation

- **(Part 1.)** Improve the efficacy of inhalational drug delivery in **asthmatics**
- **(Part 2.)** Improve management and reduce costs for development of new therapeutics in **COPD** patients
- **Strategies**
  - ✓Inter-subject variability
  - ✓Inter-cluster variability

  Cluster: homogeneous sub-groups with distinct structural and functional features

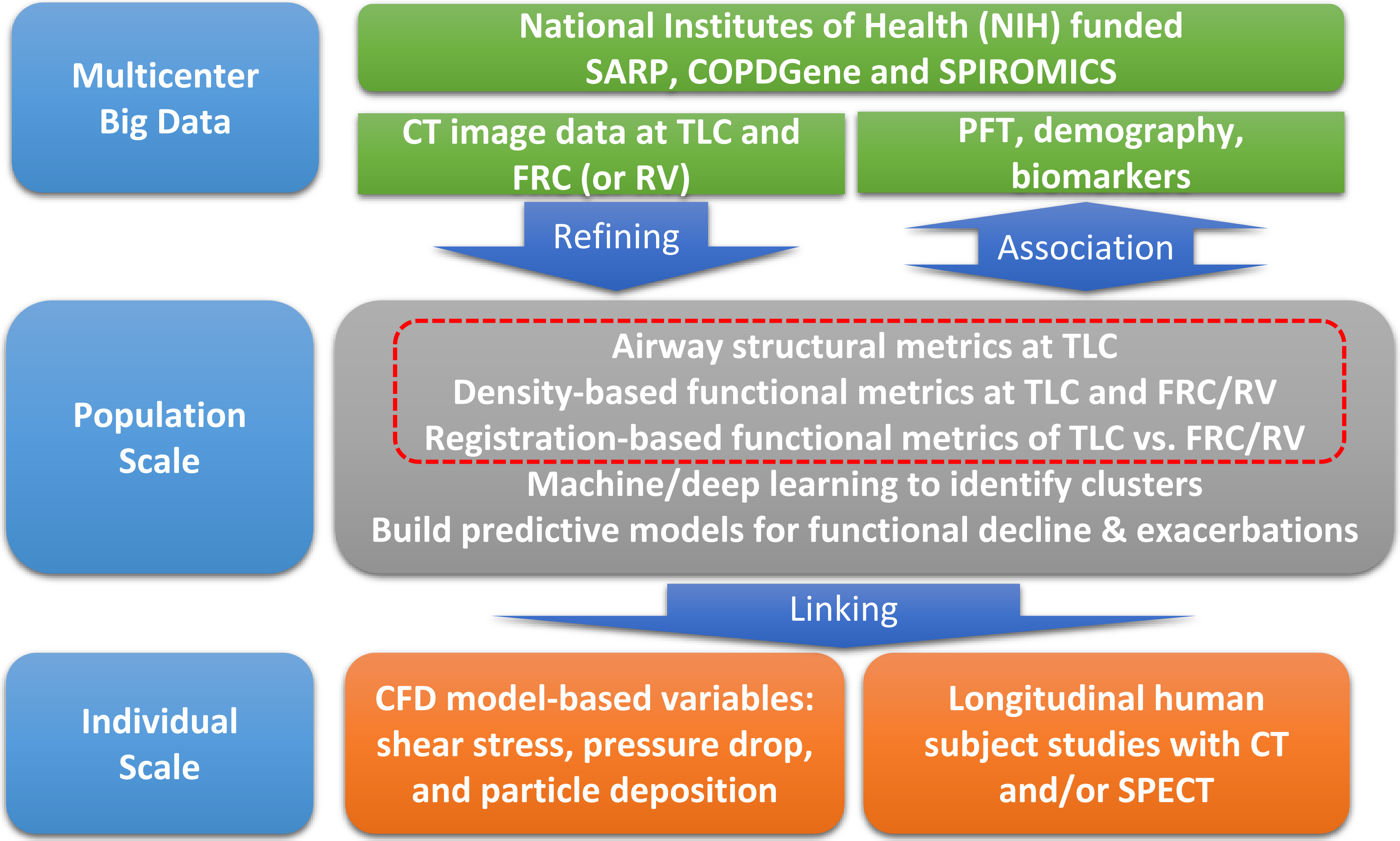
Why imaging-based variables instead of clinical variables?

- CT **imaging-based** variables could **sensitively** capture structural and functional alternations at both **local and global** scales during disease **progression**.
- Approach

  (1) Identify imaging-based clusters using imaging-based metrics

  (2) Establish their associations with clinical and biological characteristics

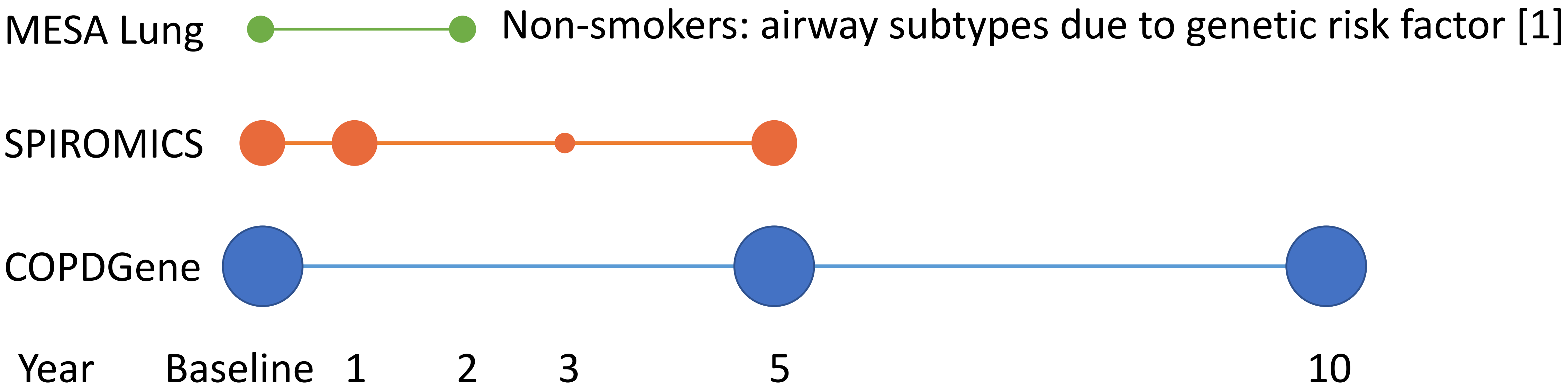
A Predictive Framework



TLC: Total Lung Capacity at full inspiration
FRC: Functional Residual Capacity at end of passive expiration
RV: Residual Volume at end of maximum exhalation
PFT: Pulmonary function tests
CFD: Computational Fluid Dynamics
CT: Computed Tomography
SPECT: Single-Photon Emission CT

SARP: Severe Asthma Research Program
COPDGene: Genetic Epidemiology of COPD
SPIROMICS: SubPopulations and InteRmediate Outcome Measures In COPD Study

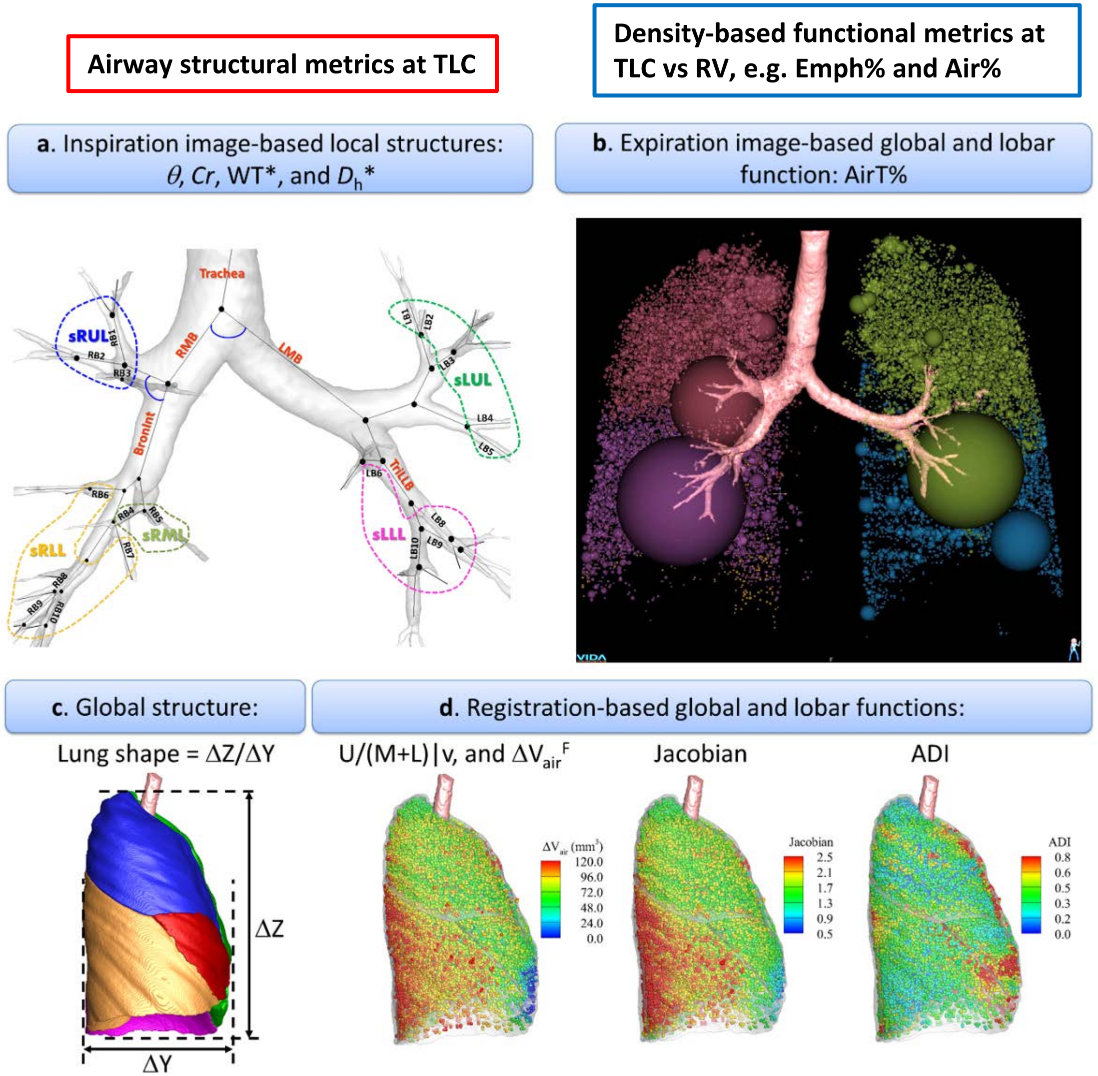
Longitudinal CT and Clinical Data



[1] Smith BM, et al. Human airway branch variation and chronic obstructive pulmonary disease. Proc Nat Acad Sci 2108

Machine Learning of Quantitative CT (QCT) Variables

75 QCT **hands-engineered** imaging-based variables, including **structural** variables via **segmentation** of TLC images and **functional** variables via **image-registration** of TLC and FRC (or RV) images.



Inter-site variability due to CT scanners & breath-hold coaches

$$I_{threshold} = \beta_{air,threshold} HU_{air,trachea} + (1 - \beta_{air,threshold}) HU_{tissue}$$

- Choi S, JAP 117:593-603, 2014

Inter-subject variability due to due to sex, age & height

Normalized airway wall thickness, WT*

Wall thickening is a phenotype for inflammation

Normalized luminal diameter, D_h*

Luminal narrowing is a phenotype for hyper-responsiveness

- Choi S, JAP 118(10):1286-98, 2015

- **Principal Component Analysis (PCA) for dimensional reduction**
- **K-means (machine learning) for clustering**

- Choi et al., "Quantitative computed tomography imaging-based clustering differentiates asthmatic subgroups with distinctive clinical phenotypes." Journal of Allergy and Clinical Immunology, 140(3), 2017.

(**Part 1.**) Major Features of 4 Asthma Clusters

	Imaging characteristics	Clinical characteristics
Cluster 1	• Normal airway structure • Increased lung deformation (Jacobian and ADI↑)	• Younger, early onset • Nonsevere asthma • Reversible lung function • Easy to control asthma symptoms
Cluster 2	• Airway luminal narrowing (D_h*↓) • No airway wall thickening (WT*) • Significant reduction of lung deformation (Jacobian and ADI↓)	• Nonsevere and severe asthma • Persistently altered lung function • Marginal to no inflammation • Difficult to control asthma symptoms
Cluster 3	• Airway wall thickening (WT*↑) • No airway luminal narrowing (D_h*) • Moderate reduction of lung deformation (Jacobian and ADI↓)	• Obese, female-dominant • Severe asthma • Reversible lung function • Blood lymphopenia • Difficult to control asthma symptoms
Cluster 4	• Airway luminal narrowing (D_h*↓) • Significant reduction of lung deformation (Jacobian and ADI↓) • Significant air-trapping (AirT%↑)	• Older, late onset, male-dominant • Severe asthma • Persistently altered lung function • Neutrophilic-dominant inflammation • Difficult to control asthma symptoms

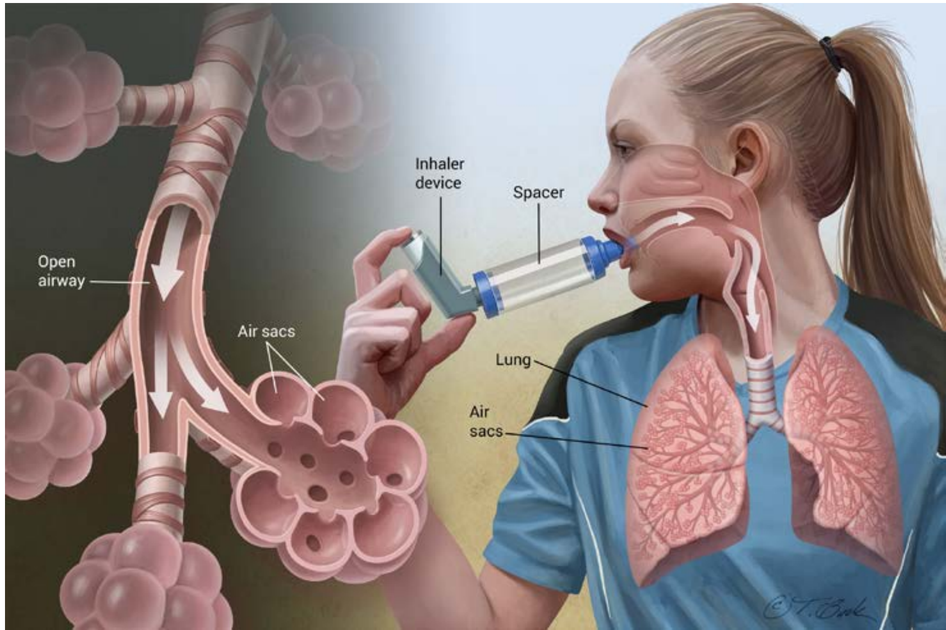
248 asthmatics

Choi S et al. *J Allergy Clin Immunol* (JACI) 2017;140(3):690-700.

Drug Aerosol Inhalation in Asthma

- Aerosol inhalation is a major way to deliver medication for treatment.
  - ➢ Aerosolized bronchodilators relax airway smooth muscle,
  - ➢ Corticosteroids reduce airway wall inflammation.

- However, the efficiency of delivery to the peripheral lung is limited due to:
  - ➢ structural and functional variability,
  - ➢ aerosol size,
  - ➢ inspiratory breathing patterns, and
  - ➢ device design and misuse.



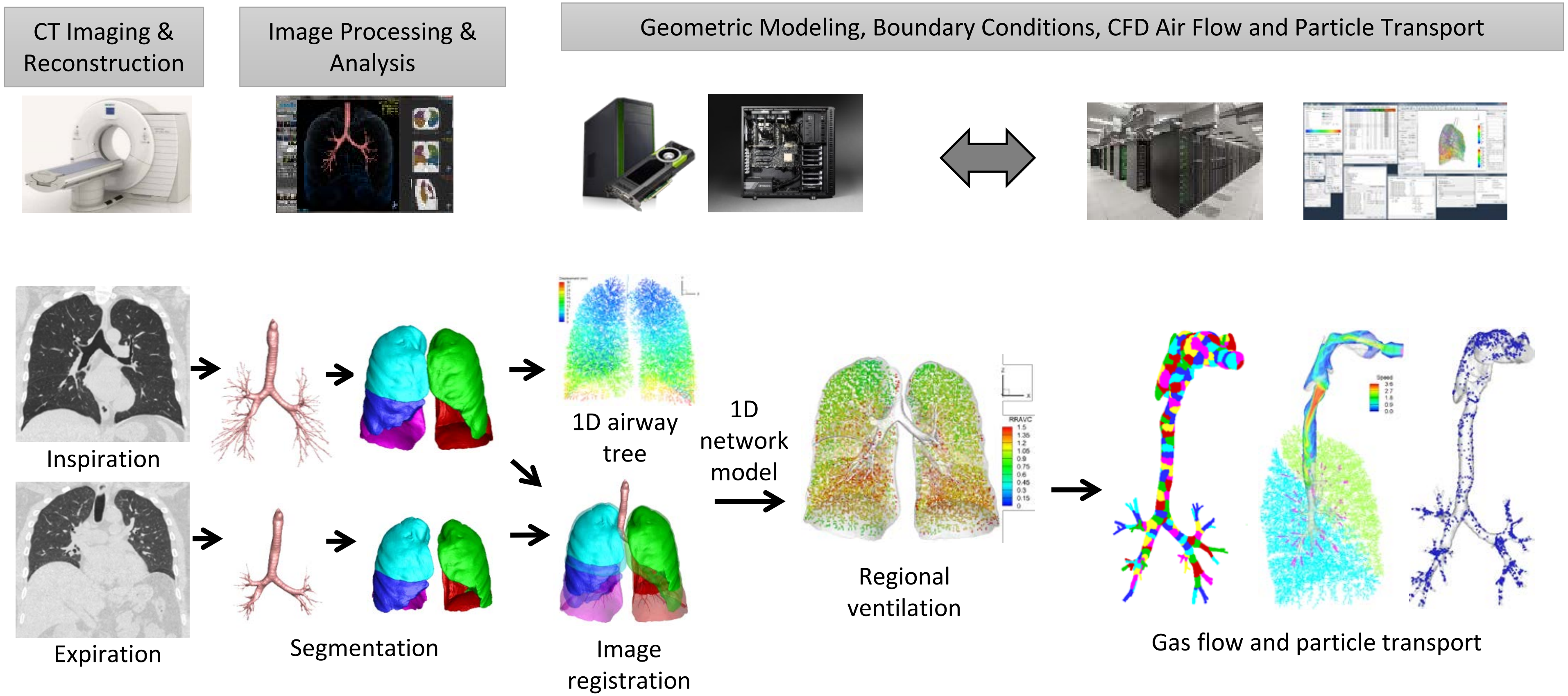
https://foundation.chestnet.org/patient-education-resources/asthma/

Cluster-guided Computational Fluid Dynamics (CFD) analysis for particle deposition in asthma

- We sought to identify particle deposition patterns in cluster representative subjects using CFD
- Cluster 2: non-severe/severe asthmatics had constricted airways in the left lower lobe (LLL).
- Cluster 3: female, obese & severe asthmatics had non-constricted airways
- Cluster 4: male, older & severe asthmatics had constricted airways in LLL.

[1] Choi J et al. *J Aerosol Med Pulm Drug Deliv* (JAMPDD) 2019

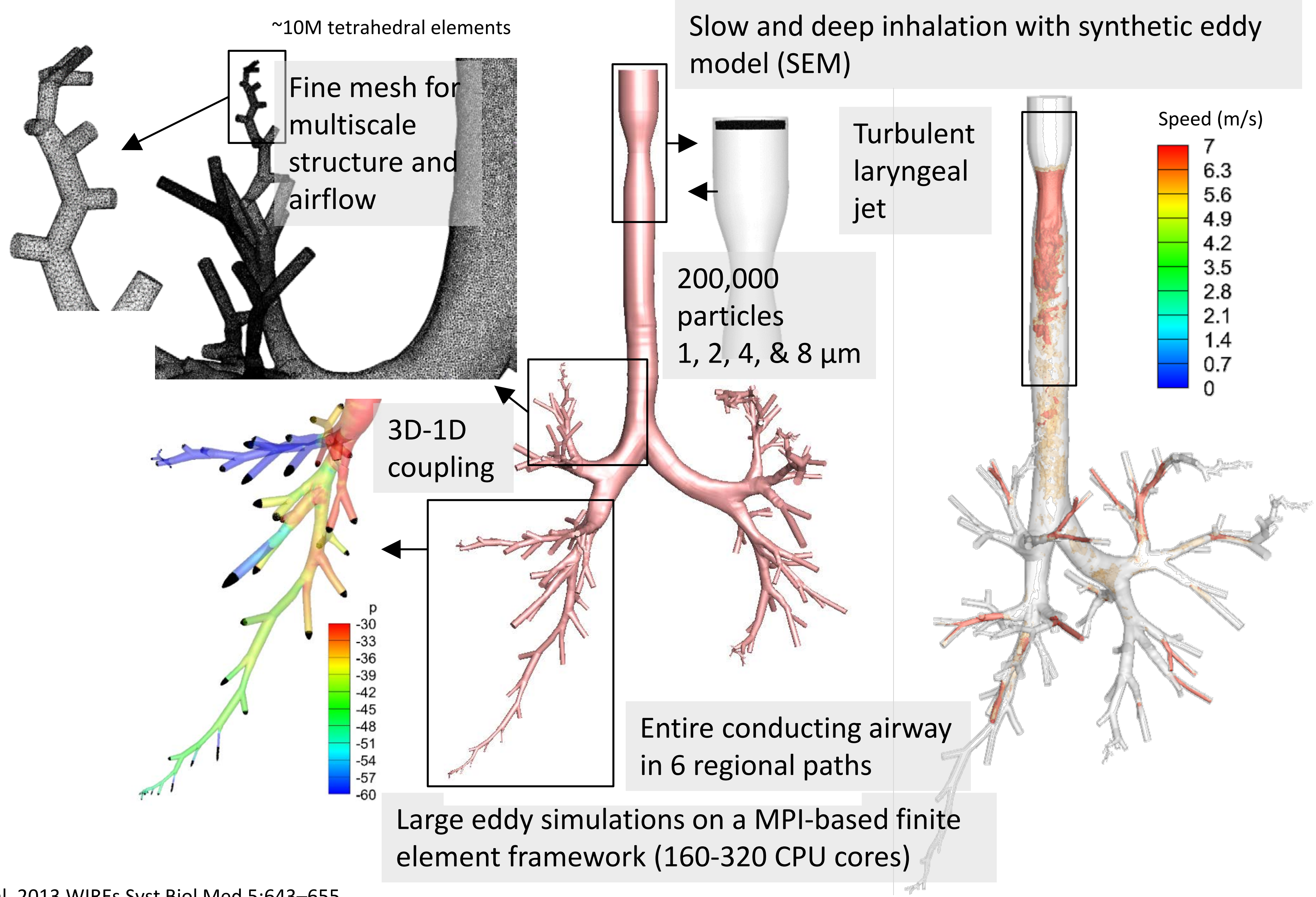
CT-based Subject-specific CFD Lung Model



- Anatomically accurate airway <u>structure</u> geometry
- Physiologically consistent regional lung <u>function</u>

➔ Multiscale subject-specific air flow and particle transport

Subject Specific Multiscale CFD Simulations



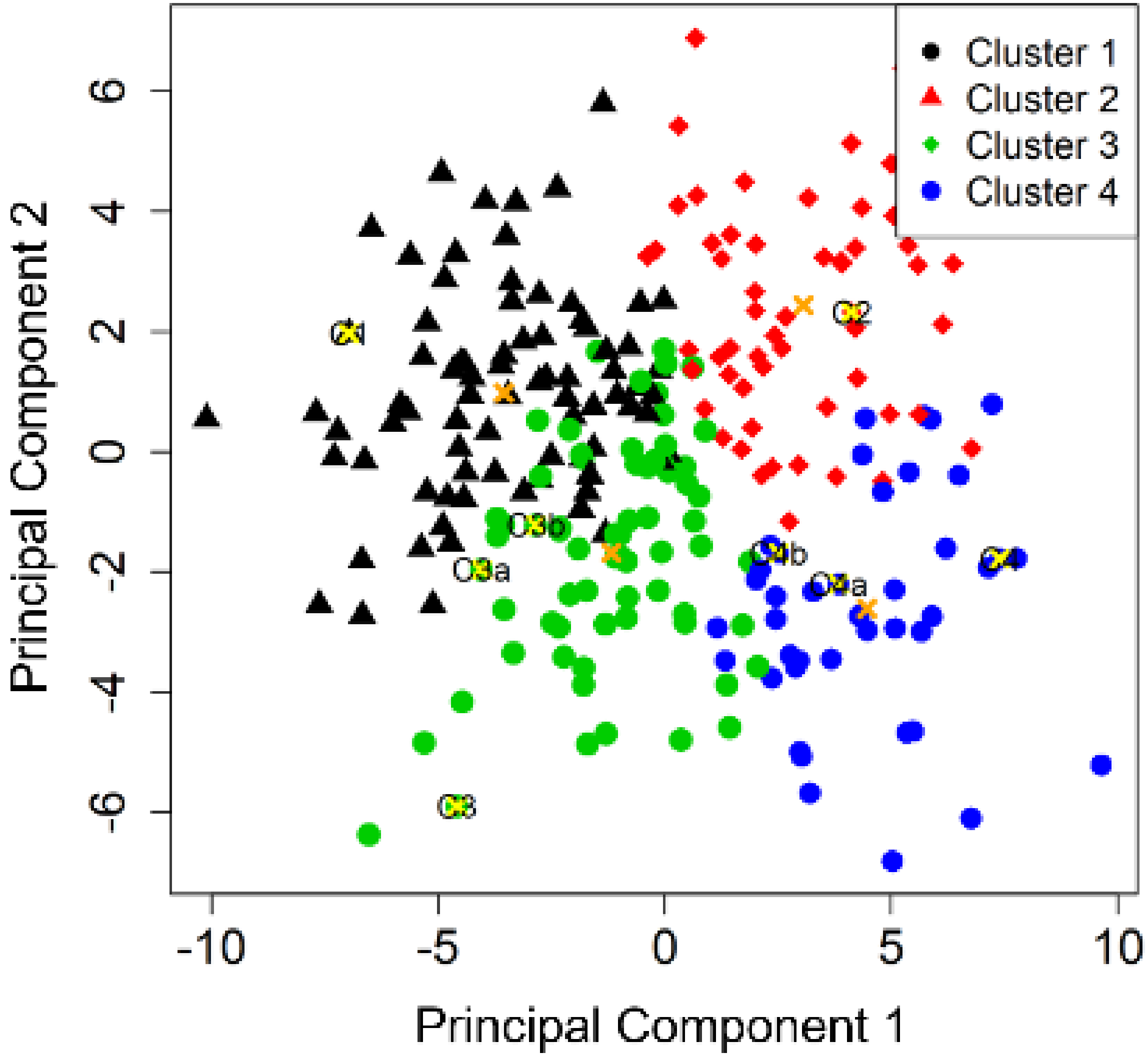
Lin et al. 2013 WIREs Syst Biol Med 5:643–655

Cluster-representative Subjects

10 subjects were selected for CFD simulations of air flow and particle transport.

- 1 healthy male subject
- 1 healthy female subject
- 1 cluster-1 subject
- 1 cluster-2 subject
- 3 cluster-3 subjects
- 3 cluster-4 subjects

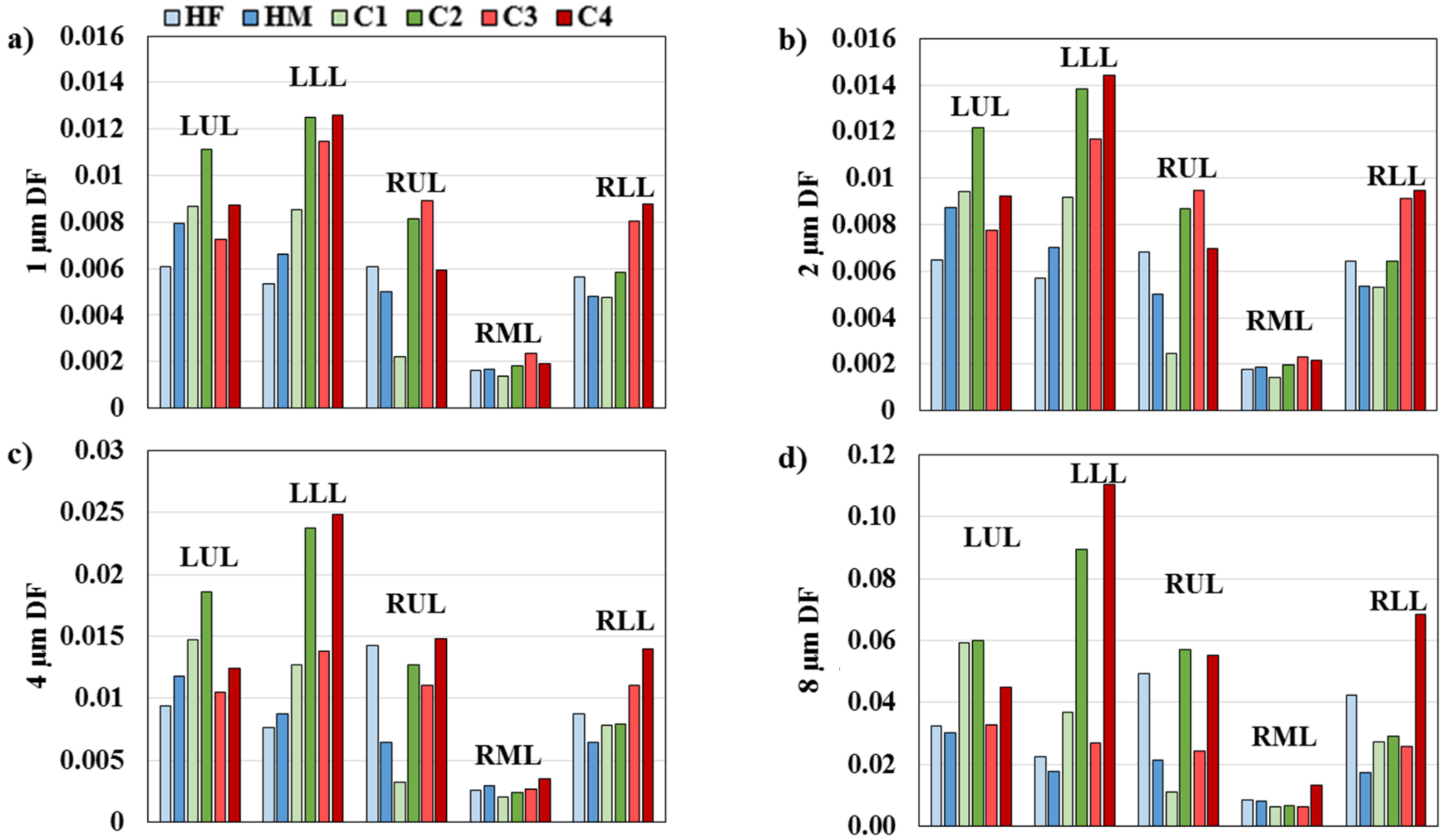
Choi et al. *JAMPDD* 2019



Projection of the four color-coded cluster subjects and their respective cluster means (“x”) on principal component (PC) 1 and PC 2 coordinates

Lobar Deposition Fractions (DF)

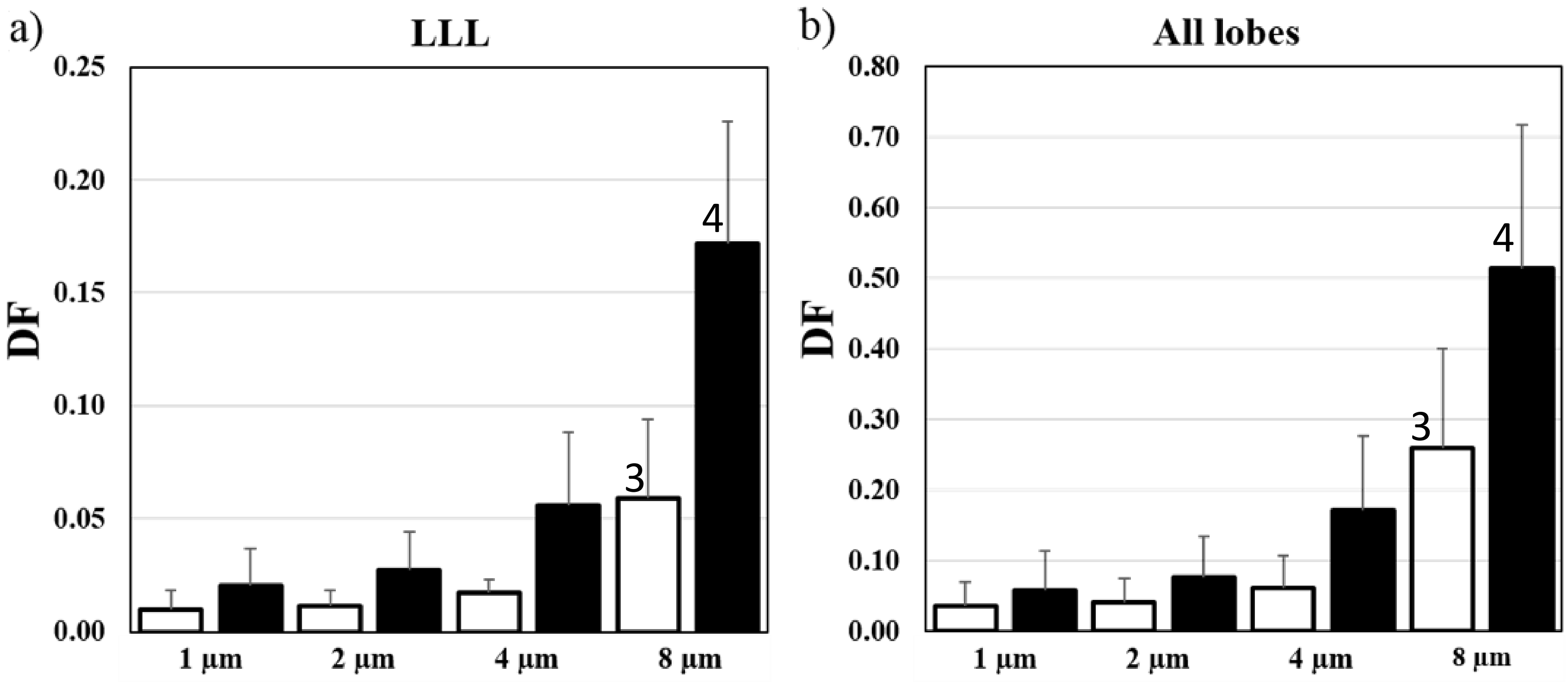
$$DF = \frac{N_{p,deposit}}{N_{p,enter}}$$



LUL, left upper lobe
LLL, left lower lobe
RUL, right upper lobe
RML, right middle lobe
RLL, right lower lobe

Cluster 2 and cluster 4 showed large DF in the lower left lobe (LLL).

Mean DFs of Severe Asthmatics in Clusters 3 & 4

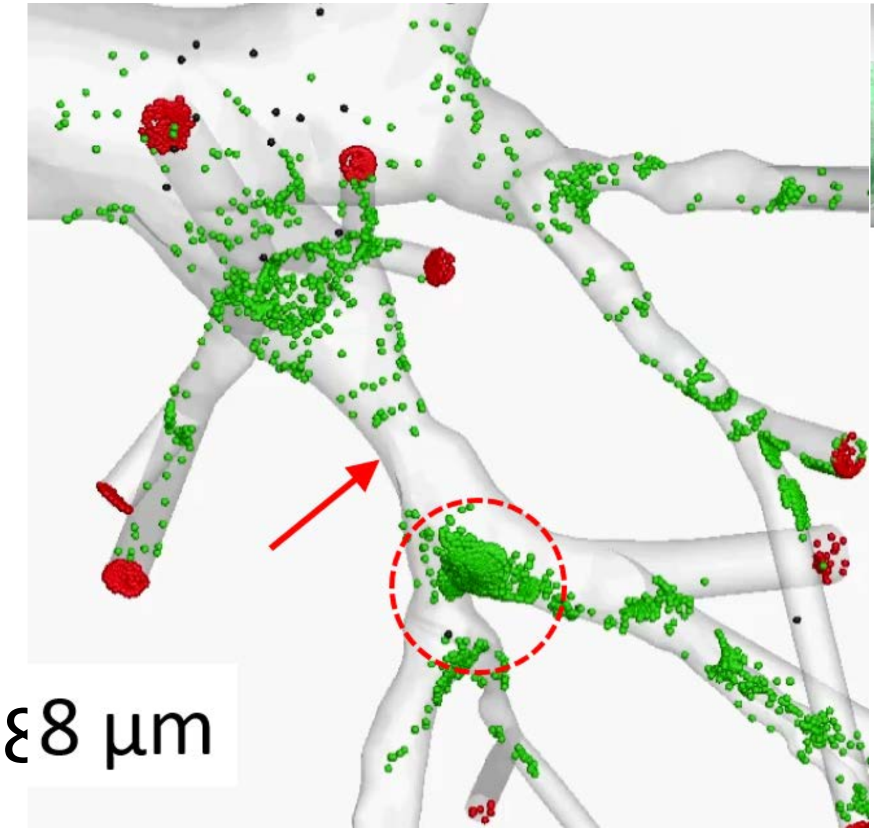


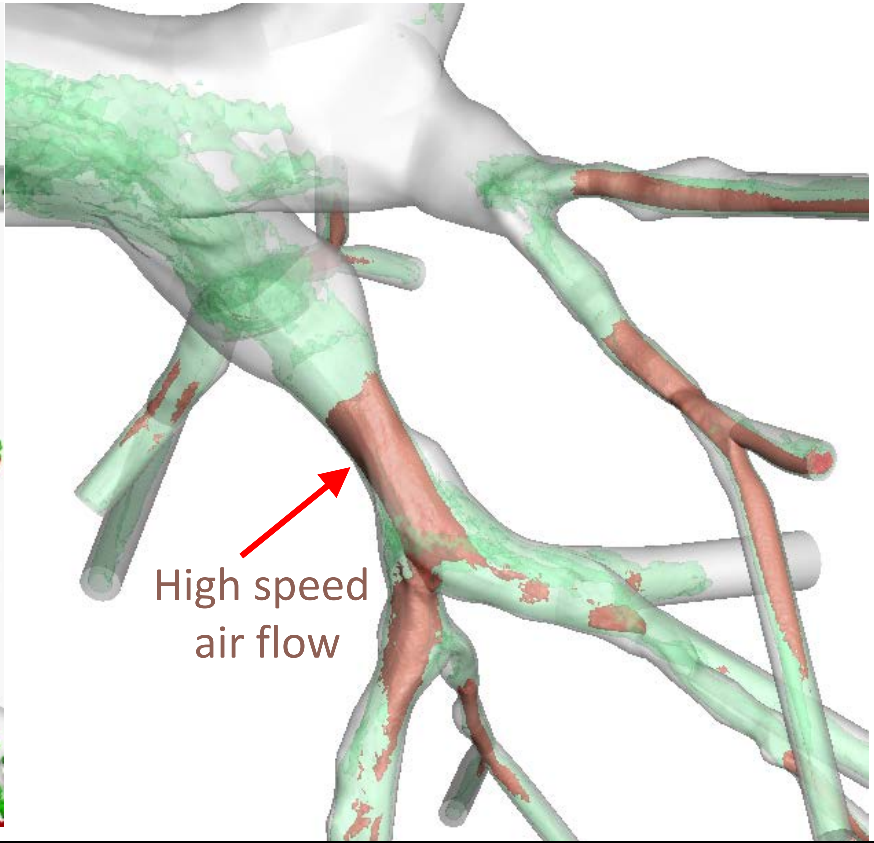
- DFs of 1, 2, 4, and 8 µm particles are compared in (a) LLL and (b) all the lobes for the three cluster 3 (blank) and cluster 4 (filled) subjects, respectively.
- DF is greater in cluster 4 than cluster 3. The difference increases with size.

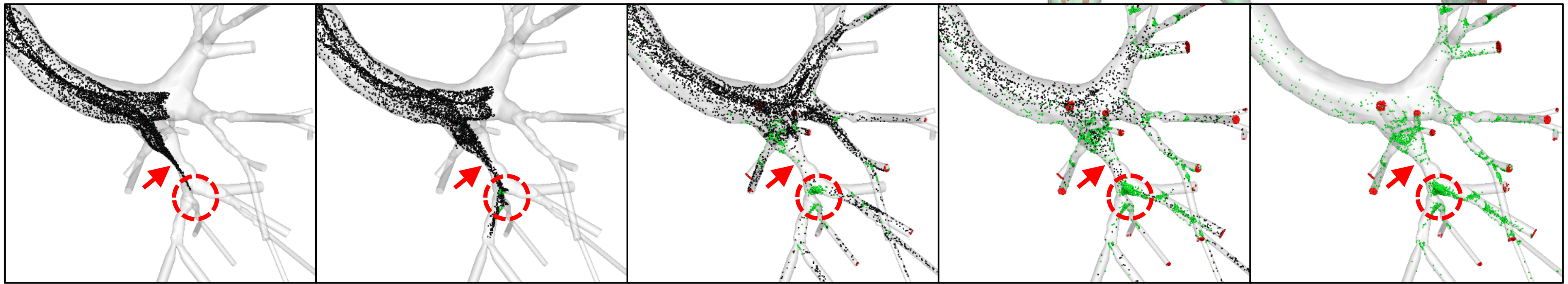
Choi et al. *JAMPDD* 2019

Airway Constriction in the Left Lower Lobe (LLL)

Cluster 4



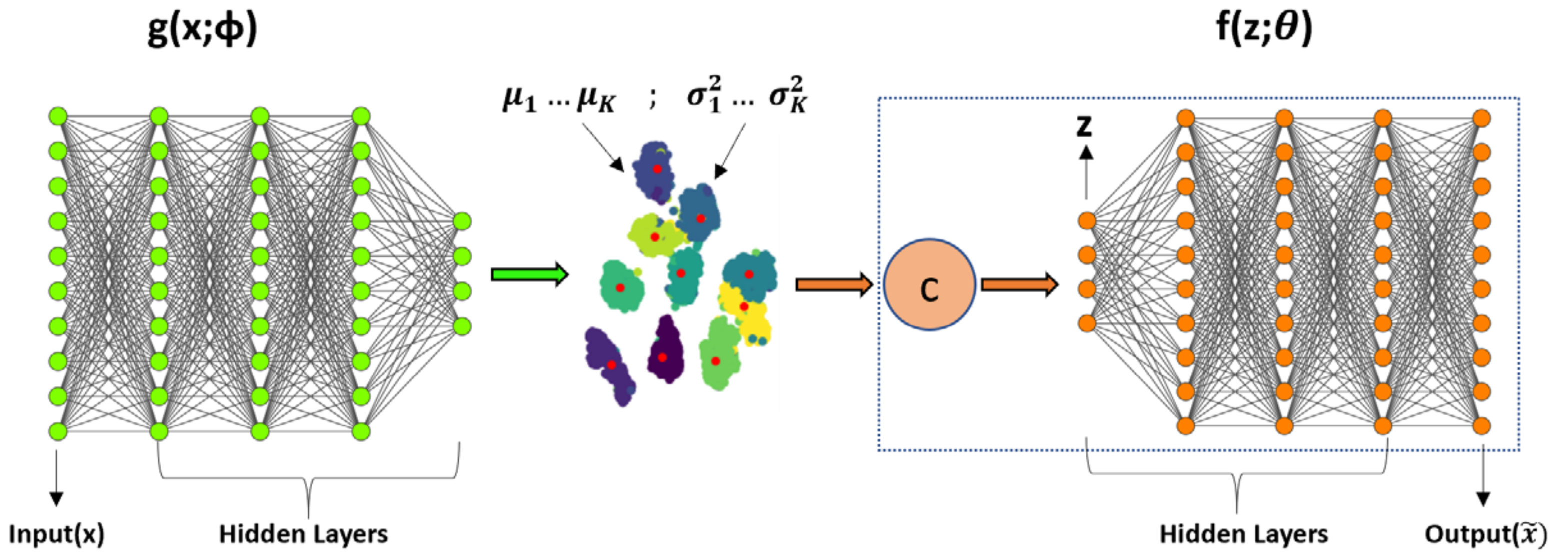




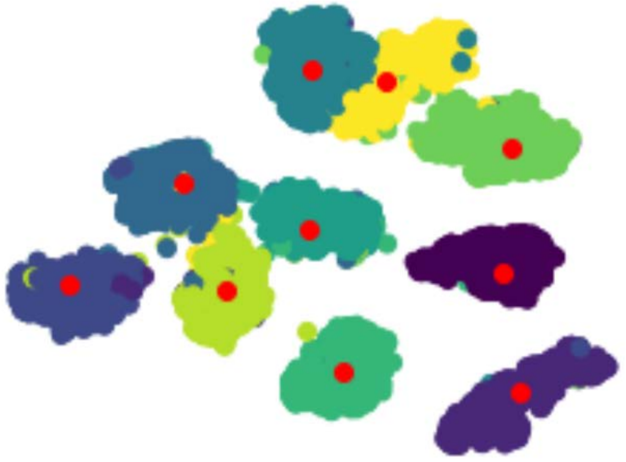
(**Part 2.**) COPD Former-Smoker Clustering

- COPD is a heterogeneous disease characterized by diversity of progressive respiratory symptoms and rates of functional decline.
- 528 former-smokers were analyzed.
- CT scans at two volumes TLC & RV and at baseline and one-year follow up
- Major improvements over our recent work on cross-sectional clustering in former smokers by B. Haghighi et al. Respiratory Research, 2019.
  - **PART 2.a.** PCA & K-means (traditional ML) vs. Variational Deep Embedding (VaDE): (a) linear vs non-linear, (b) VaDE trains embeddings and clustering simultaneously.
  - **PART 2.b.** Autoencoder (AE) region of interest (ROIs) deep learning algorithm to identity tissue-pattern clusters which can be used in conjunction with QCT **hands-engineered** imaging-based variables for analysis.

(Part 2.a.) Variational Deep Embedding VaDE



(a) VaDE architecture based on variation autoencoder VAE

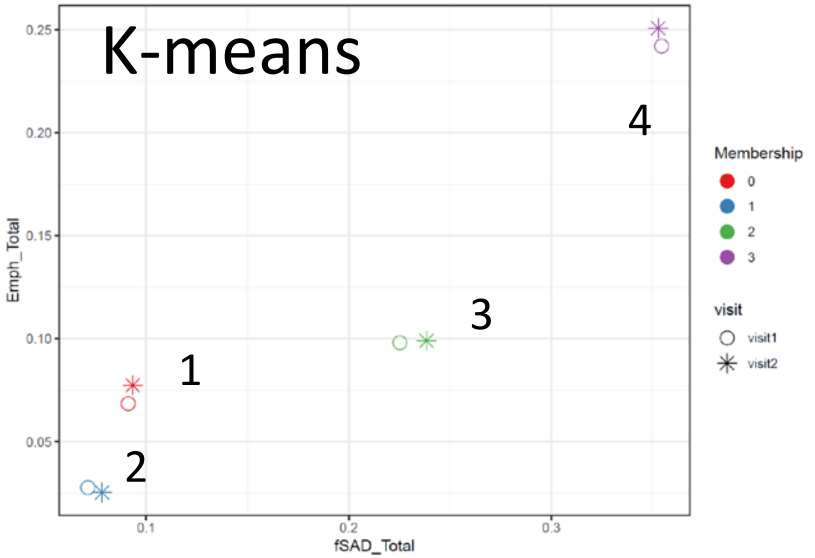


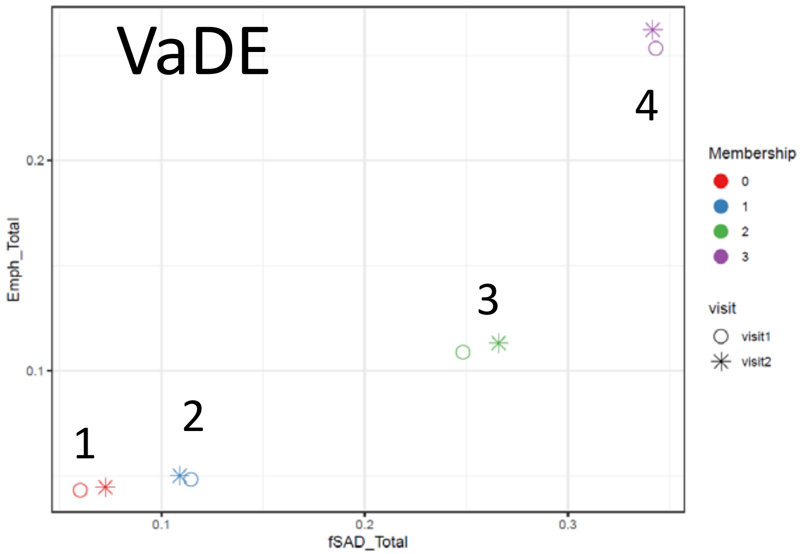
(b) VaDE learned embedding space using 5000 MNIST data.

- x and $\tilde{x}$ are input and generated data, respectively.
- z is the latent vector of means μ and variances σ_i^2.
- g and f are an encoder and a decoder. C is clustering.

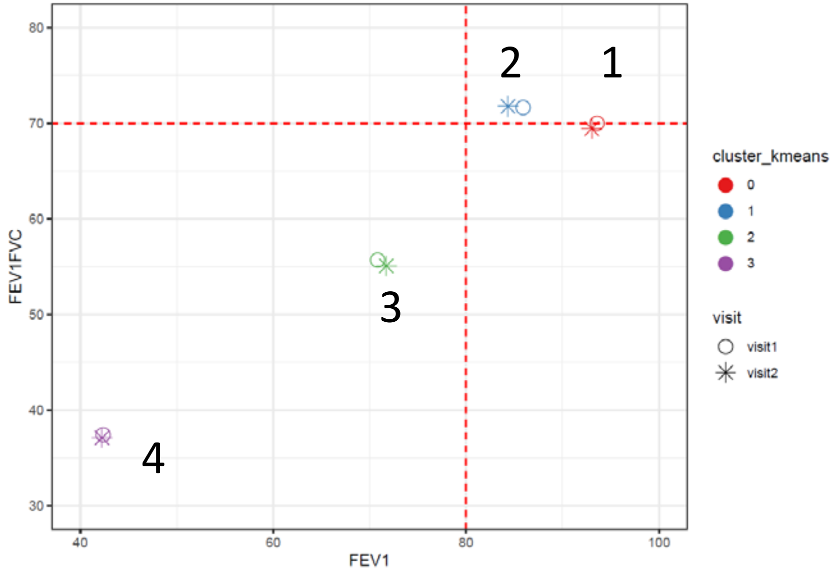
Longitudinal Clusters : K-means vs. VaDE with QCT Hands-engineered Variables

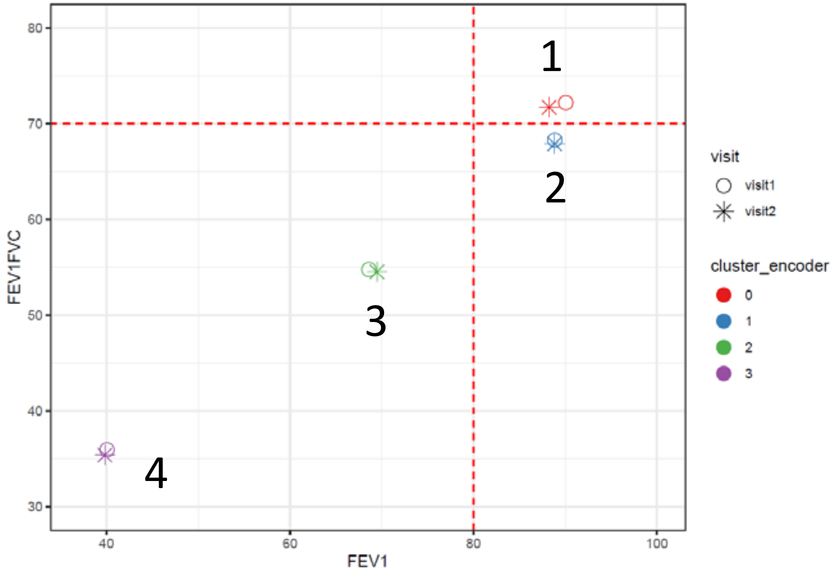
fSAD% vs. Emph%





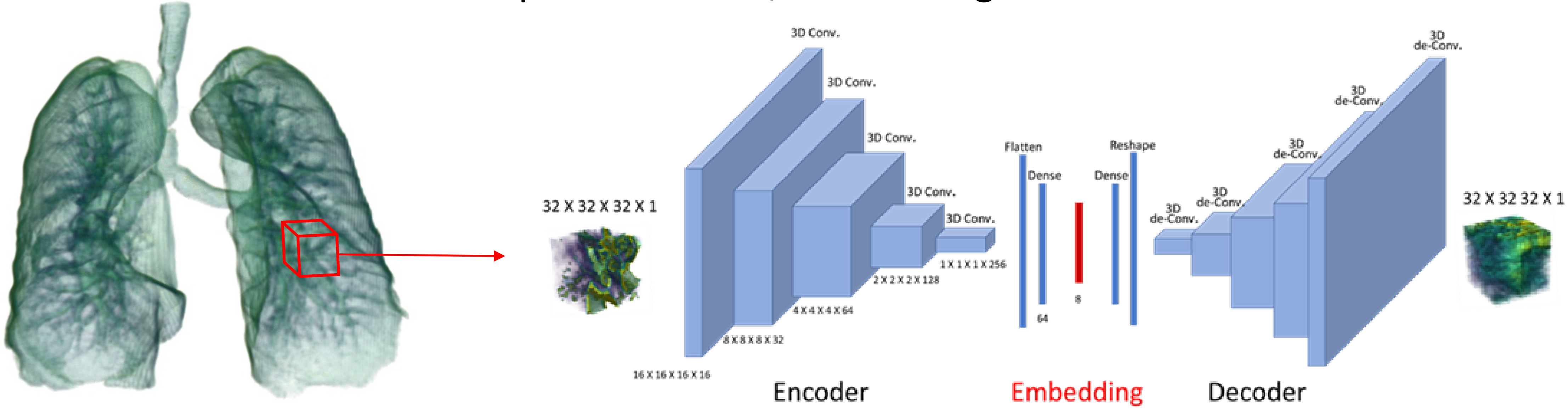
FEV1 vs. FEV1/FVC





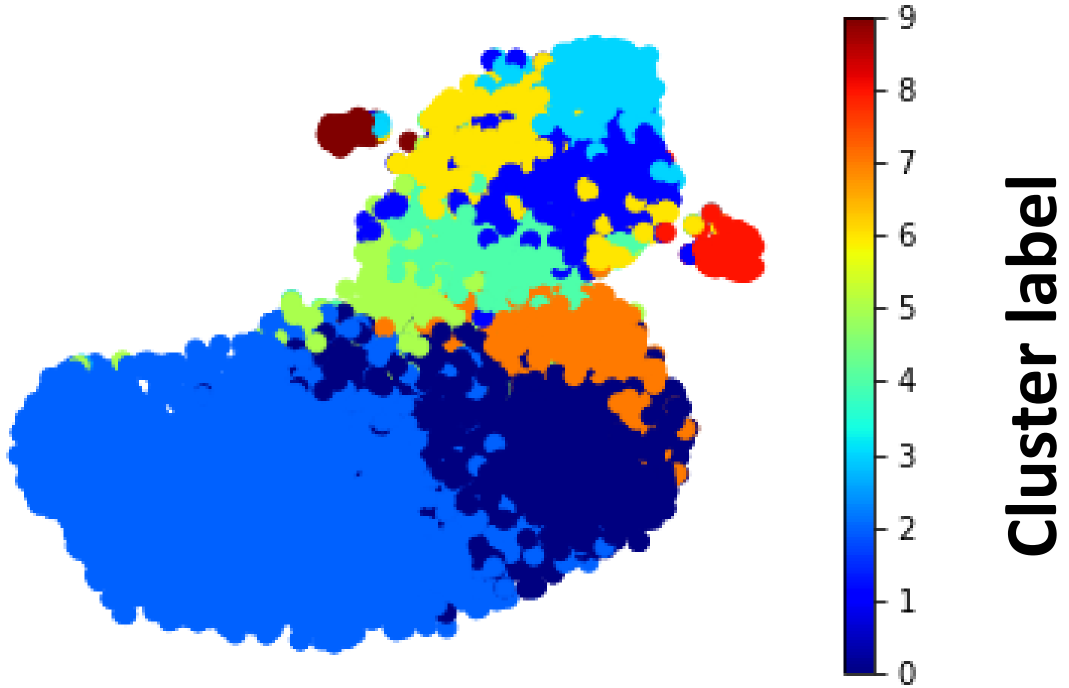
(Part 2.b.) Convolutional Autoencoder (CAE)

- The hands-engineered imaging-based variable may not be enough to describe all the features of the COPD lungs.
- The purpose is to identify undiscovered diseased patterns in the lungs using deep learning directly from CT images.
- A total of 10,000 three-dimensional ROIs were randomly extracted from 738 CT images at TLC of 369 former smokers.
- CAE is trained to learn 1D representations/embeddings of the ROIs.



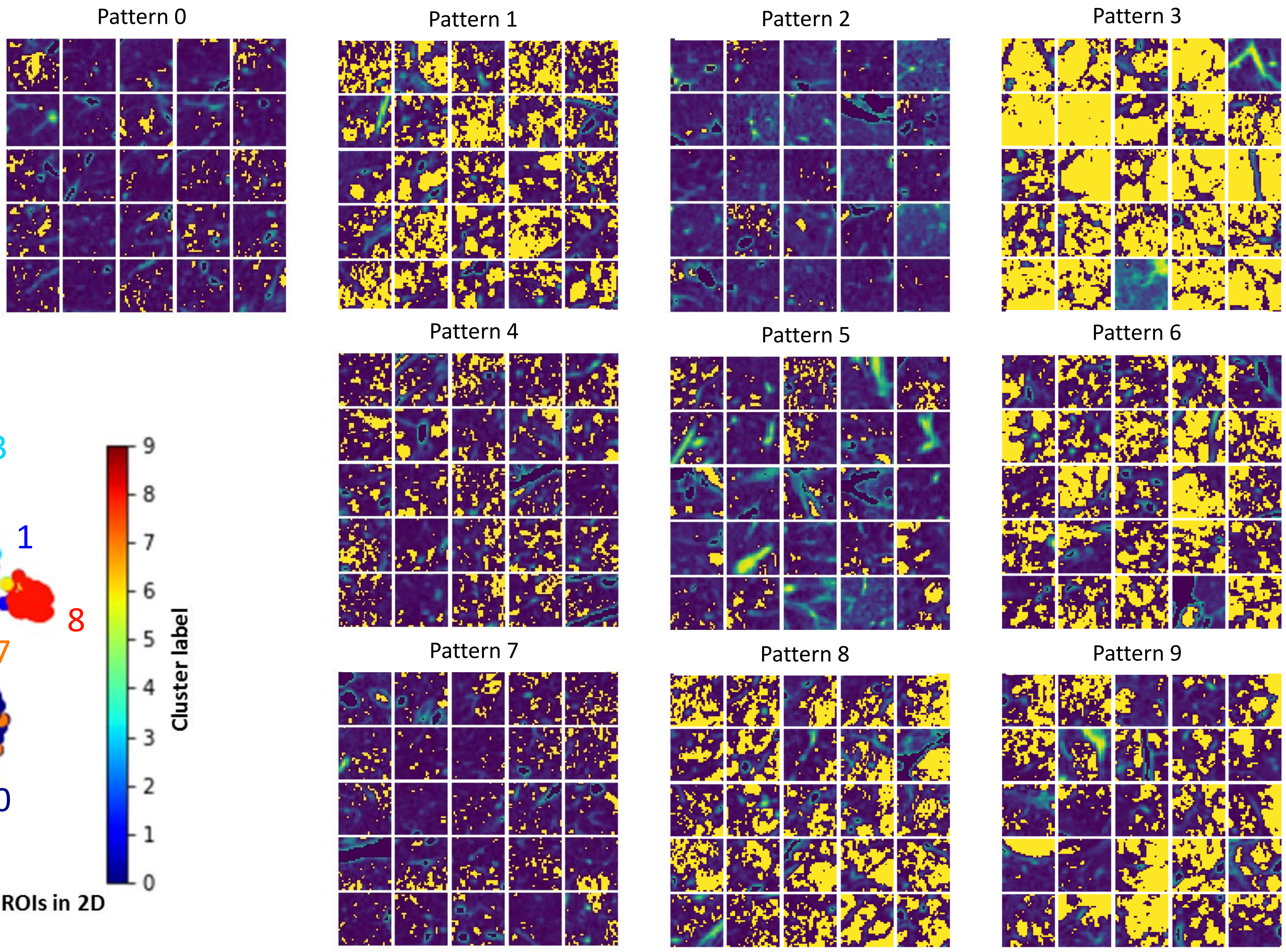
Identifying Tissue Patterns In The Lung

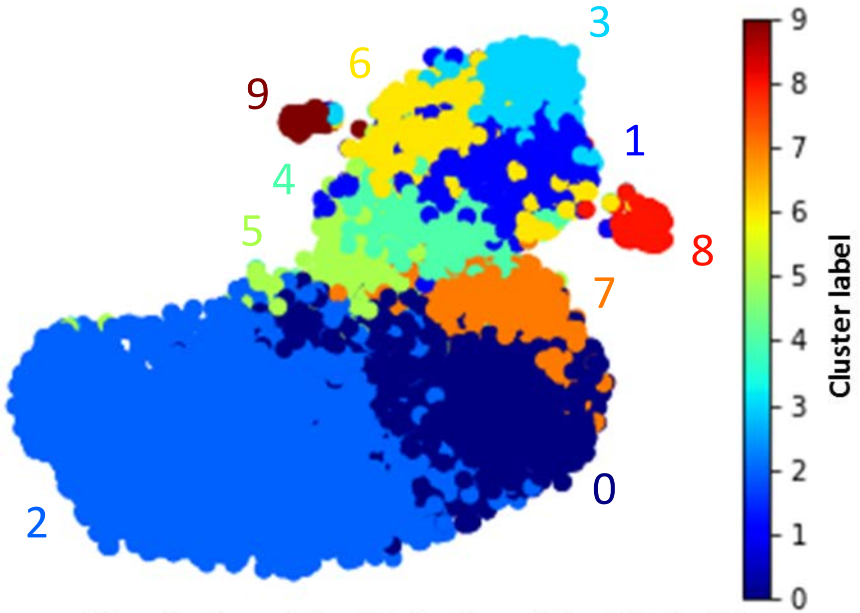
Group 1D representations/embeddings of the ROIs into 10 pattern-clusters using unsupervised learning.



t-SNE Visualization of the distribution of the ROIs in 2D

Visualization of the ROIs for each pattern-cluster





Visualization of the distribution of the ROIs in 2D

Potential Applications

- Quantify pattern-clusters in the whole lung to create a pattern histogram.
- Use pattern histograms to build a predictive model via supervised learning to predict lung function decline and exacerbation over time.

Visualization of the distribution of the ROIs in 2D

3D Pattern map

Pattern histogram

Summary

- We demonstrated the effects of cluster-specific imaging-based features on particle deposition in asthmatic subjects.
- The ability to differentiate severe asthmatics into clusters by imaging-based features may help devise strategies for improved inhalational drug delivery.
- Longitudinal (progression) clusters identified by non-linear deep-learning may be different from those of linear traditional machine-learning k-means.
- Convolutional autoencoder can identify tissue pattern clusters from CT images, enabling development of predictive models for precision medicine.

Acknowledgments

- FDA grant U01-FD005837, NIH grants U01-HL114494, R01-HL112986, S10-RR024738, S10-RR022421, and S10-18526, and NIEHS/NIH P30ES005605.
- Views expressed in this work do not necessarily reflect the official policies of the Department of Health and Human Services and may not be quoted as being made on behalf of a reflecting the position of the US Food and Drug Administration; nor does any mention of trade names, commercial practices, or organization imply endorsement by the United States Government.
- We also thank the Extreme Science and Engineering Discovery Environment (XSEDE) (allocation MCA07S015) sponsored by the National Science Foundation for computational time at San Diego Supercomputer Center (SDSC) and the Texas Advanced Computing Center (TACC).

Thank You!