Towards Understanding Spatial Lung Tissue Heterogeneity in Idiopathic Pulmonary Fibrosis

Akif B. Tosun[‡], Dimitris V. Manatakis[‡], Milica Vukmirovic[§], Robert Homer[§], Naftali Kaminski[§], Chakra S. Chennubhotla[‡], Panayiotis V. Benos[‡]

[§] Yale School of Medicine, Yale University [‡] Department of Computational and Systems Biology, University of Pittsburgh

Abstract

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive lung disease. IPF is consequence of fibrosis (irregular wound healing) and the microscopic appearance of fibrosis is heterogeneous. Identifying the causal associations between gene expression and histological structures will not only help understand molecular disease mechanisms involved, but it will also provide insights into potential therapeutic targets. The "lung DBP" (Driving Biomedical Problem), which is part of the Center for Causal Discovery (CCD), aims to study the causal genotype-phenotype mechanisms intrinsic to IPF by integrating and co-analyzing clinical variables reflecting disease progression, histopathological patterns in whole-slide H&E stained tissue images, and RNA-seq data collected from the same tissue.

Method

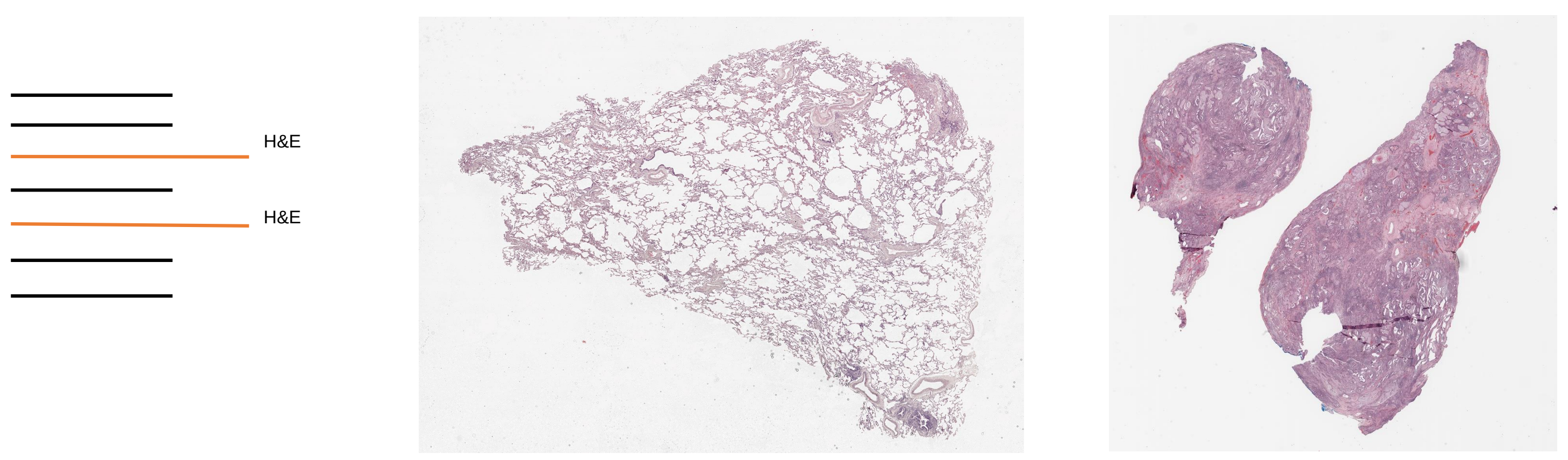
Lung tissue from IPF patients and controls has been extracted and three slides were cut sequentially. The top and the bottom slides were stained with H&E and scanned, while the middle slide was used to collect RNA-seq data.

Control (Normal Lung)
7 patients
16 tissue slide images

IPF (Idiopathic pulmonary fibrosis)

Pitt Cohort
23 patients
46 images

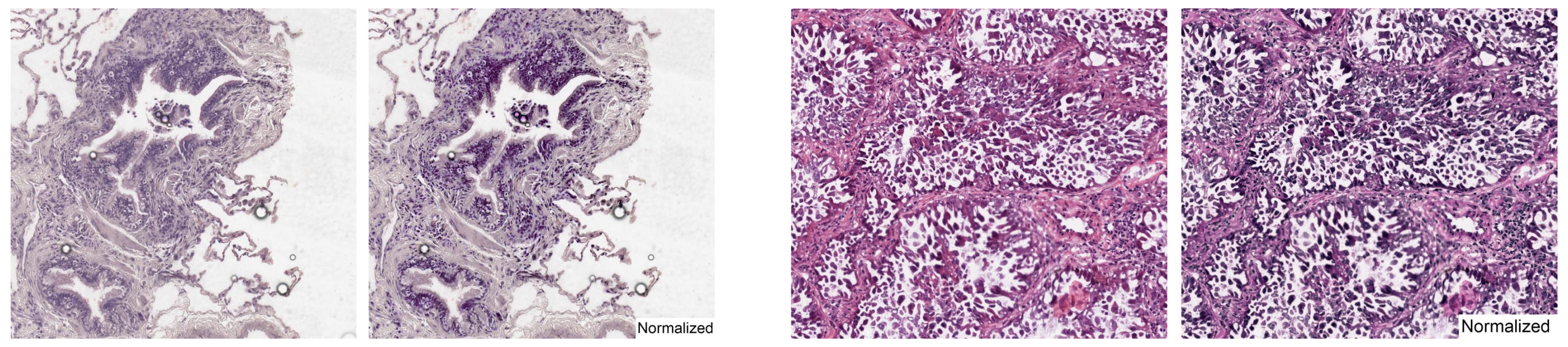
Yale Cohort
16 patients
32 images



We developed new computational pathology methods (i) to characterize fibrosis and other salient phenotypic features in IPF tissues, (ii) to quantify disease heterogeneity and (iii) to classify IPF samples from controls. We used a Mixed Graphical Model (MGM) causal algorithm to integrate these histopathological image features with the gene expression data from the same tissue and other clinical variables.

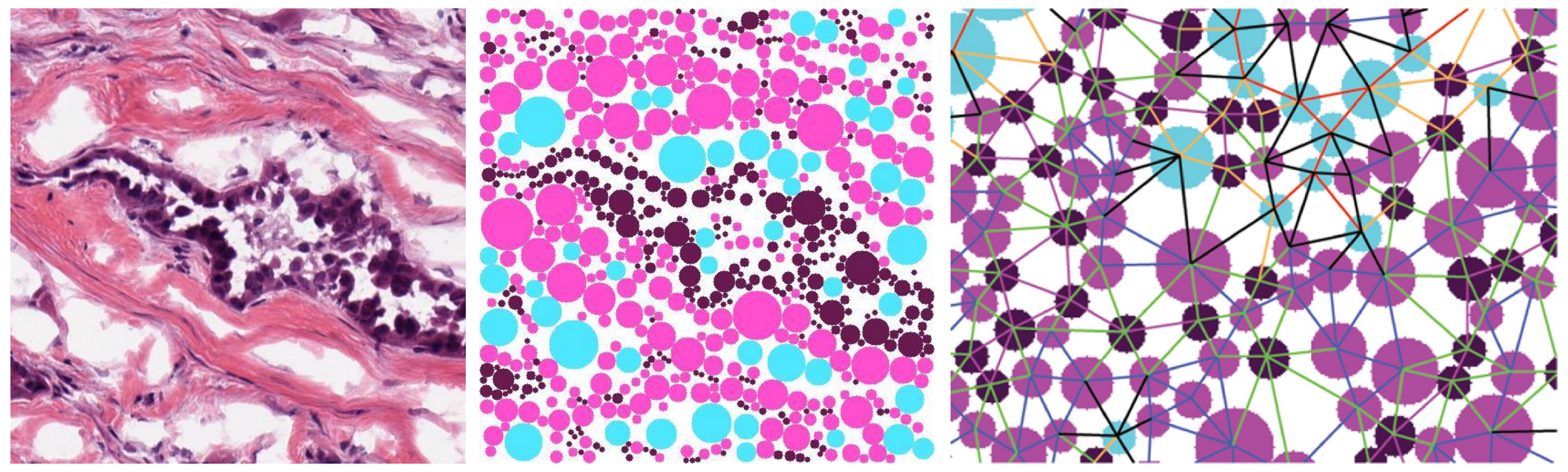
Image Color Normalization

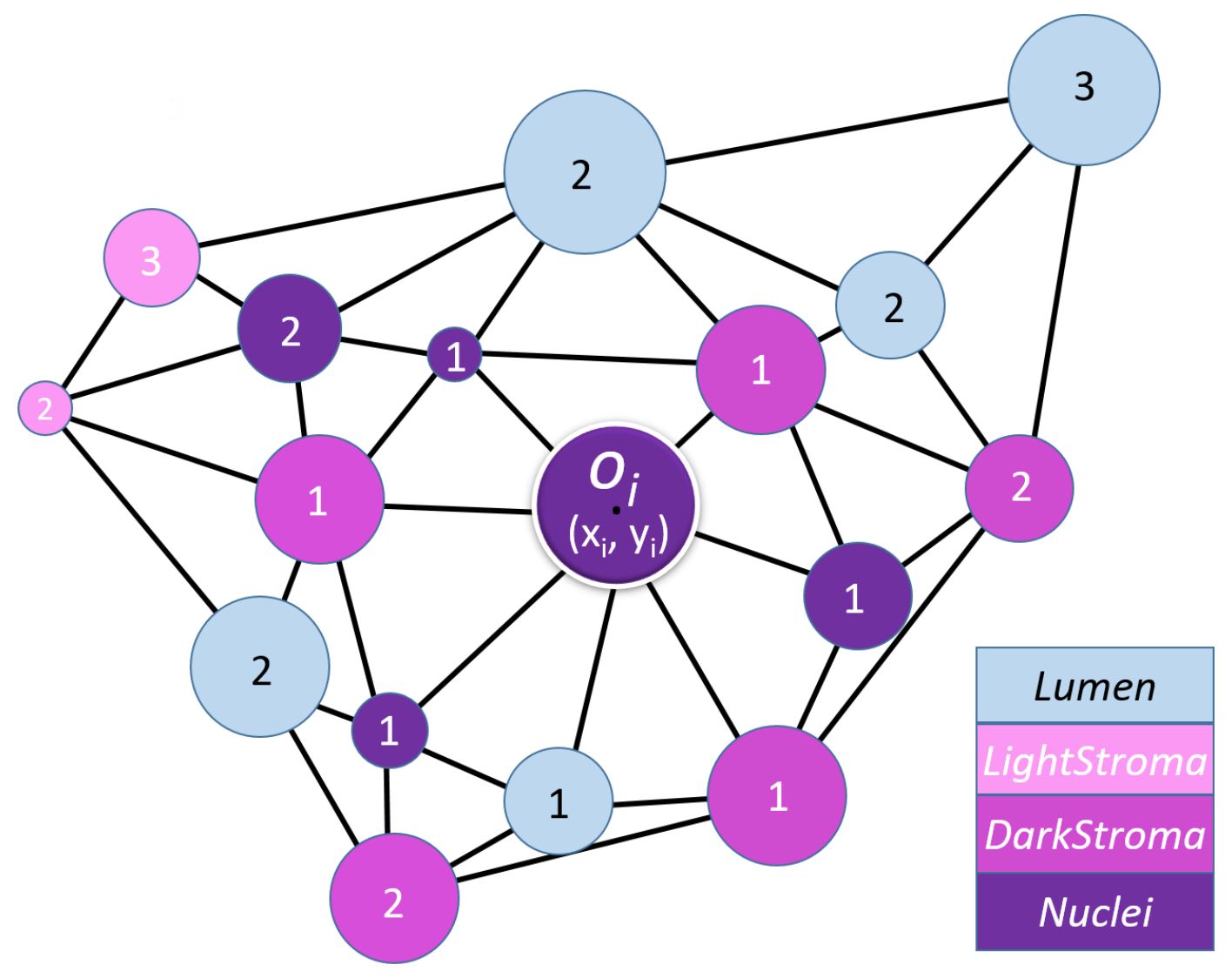
We first normalized H&E stained images (WSIs) to minimize the effects of uneven color distribution between images.



Method (cont.)

We segmented tissue components approximately; nuclei, light stroma, dark stroma, and lumen, into distinct objects using k-means. We then built an object graph using Delaunay triangulation originating from centers of each object. We encoded spatial context by collecting random walk statistics by defining a **breadth-first traversal (BFT)** for each object.

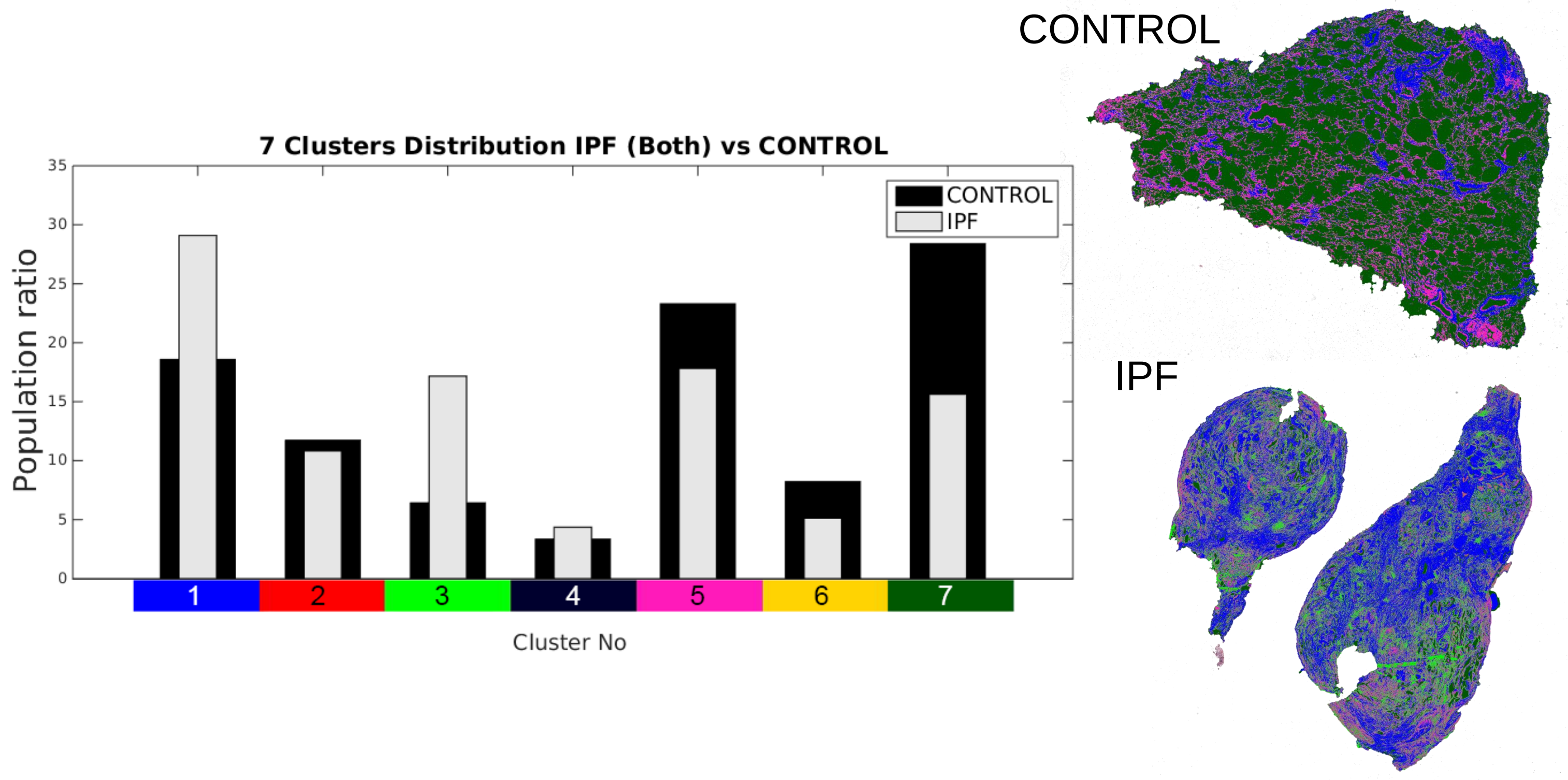




- For each depth level of BST we compute the probabilities of finding each type of object.
- Four types of objects makes ten different associations.
- For maximum depth of 10 hops, a set of 100 probability values describes the neighborhood statistics.

Finally, we cluster the neighborhood statistics into ***q*** clusters to find representative architectural patterns. For that, the principal components of the training data (Pitt Cohort) is calculated first and ***q*** is selected such that it will cover 90% of the input variance, for which ***q=7***. Following cluster center initialization, each of the nuclei neighborhoods is assigned to its closest cluster. For each patient, the proportion of 7 different types of clusters in two images are calculated as the image feature vector of the patient. This vector is also biologically interpretable.

Diagnostic Features to Discriminate Control from IPF

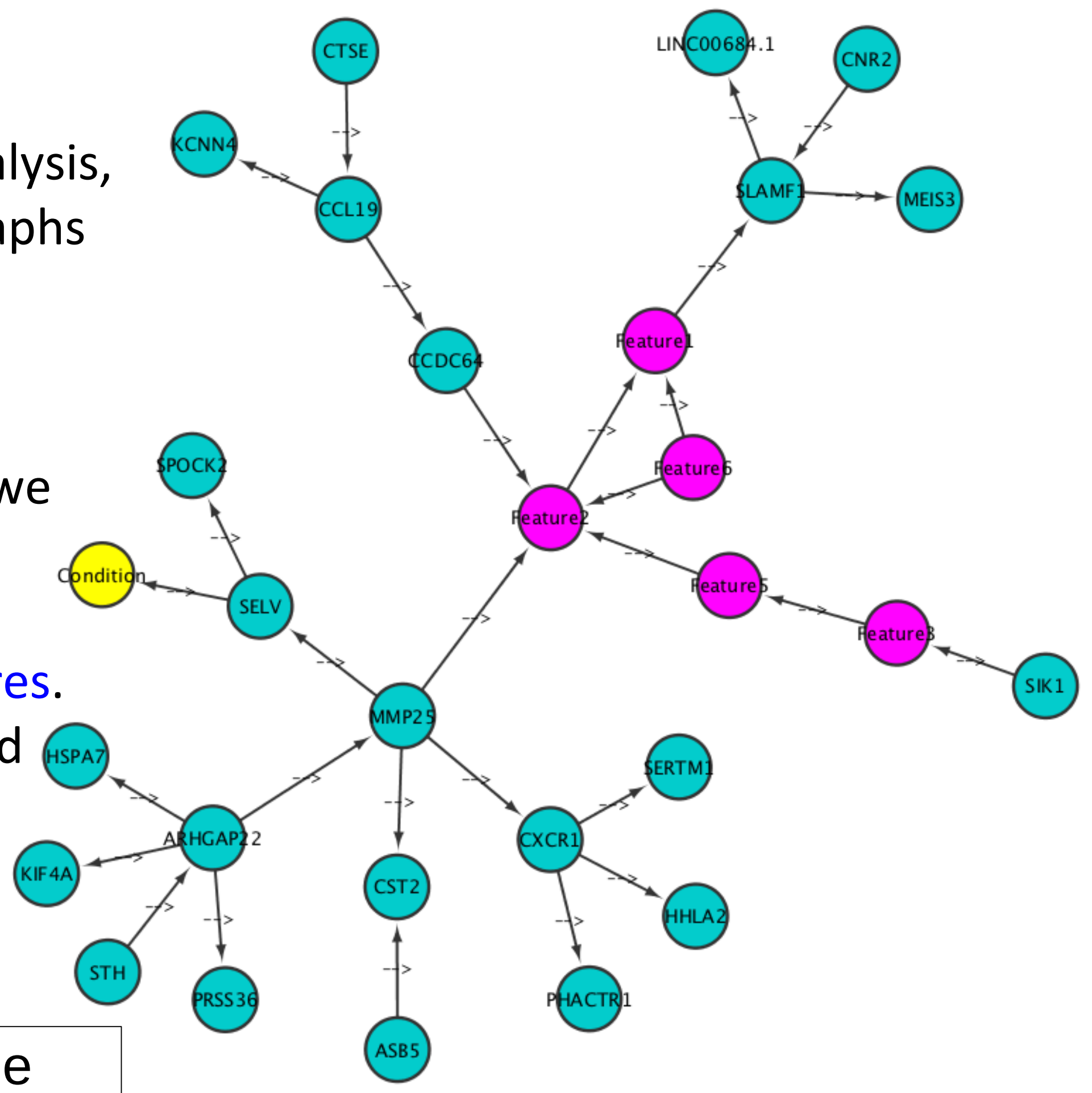


Discovering Causal Associations

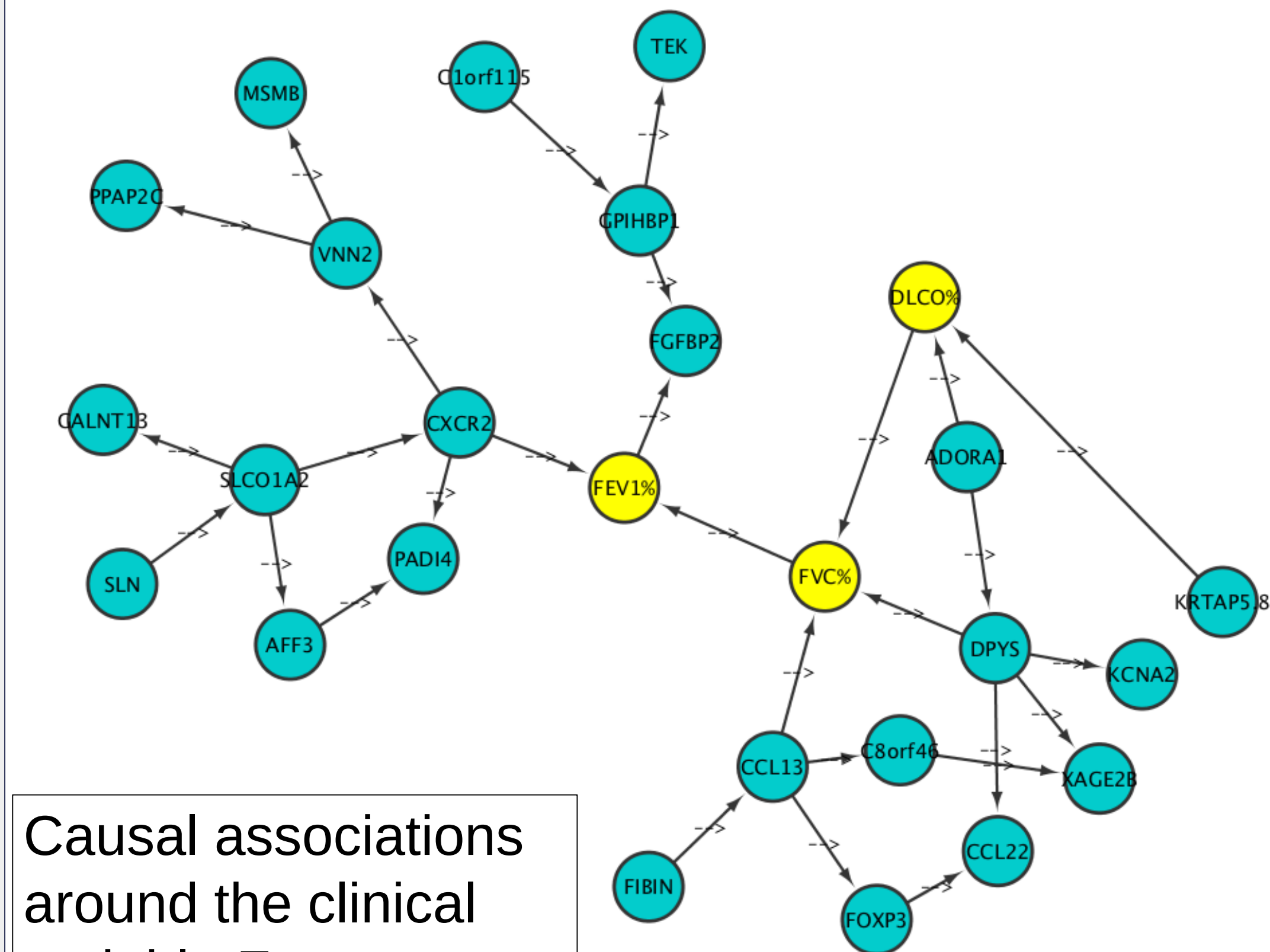
Our Mixed Graphical Model Learn (MGM-Learn) Algorithm[1] was developed to address a current bottleneck in biomedical data analysis, namely learn directed (causal) graphs over continuous and discrete variables.

Using the MGM-Learn algorithm we integrate the information of:

- Histopathological image features.
- Gene expression data extracted from the same tissue.
- Clinical variables.



Causal associations around the histopathological Image Feature 2



Causal associations around the clinical variable Force Percentual Expiratory Volume (FEV%).

Data analysis results show that the histopathological image features and the clinical variables are causally related to genes that are indicative to IPF.

We expect that these causal associations can elucidate the phenotype-genotype relation in the IPF lung disease.

Results

In this poster, we present new computational pathology algorithms to characterize the heterogeneity in the microscopic appearance of fibrosis in IPF whole-slide H&E stained tissue images. Furthermore, we present the genes and gene networks that are causally related to these features and the clinical variables that are indicative of IPF severity. These results constitute a first step in our understanding of the dynamic changes that potentially occur in a progressive fibrotic lung disease.

References

[1] Sedgewick AJ, Shi I, Donovan RM, Benos PV (2016). "Learning mixed graphical models with separate sparsity parameters and stability-based model selection." BMC Bioinformatics, 17(5), 307.

Acknowledgements

Funding from **NIH U54HG008540**.