Integrative clustering analysis to discover novel IPF disease subtypes in the IPF-PRO Registry

Yi Liu,[1] Huang Lin,[2] Julia F Soellner,[3] Janine Roy,[4] Richard Vinisko,[1] Ramona Schmid,[3] Benjamin Strobel,[3] John A Belperio,[5] Joao de Andrade,[6] Kevin R Flaherty,[7] Joseph A Lasky,[8] Tracy R Luckhardt,[9] Megan L Neely,[10,11] Christian Hesslinger,[3] Scott M Palmer,[10,11] Thomas B Leonard,[1] Jamie L Todd,[10,11] Margaret L Salisbury[12] on behalf of the IPF-PRO Registry investigators

[1]Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, Connecticut, USA; [2]Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; [3]Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany; [4]Staburo GmbH, Munich, Germany; [5]David Geffen School of Medicine at UCLA, Los Angeles, California, USA; [6]Vanderbilt University School of Medicine, Nashville, Tennessee, USA; [7]Division of Pulmonary and Critical Care Medicine, University of Michigan, Ann Arbor, Michigan, USA; [8]School of Medicine, Tulane University, New Orleans, Louisiana, USA; [9]Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA; [10]Duke Clinical Research Institute, Durham, North Carolina, USA; [11]Duke University Medical Center, Durham, North Carolina, USA; [12]Vanderbilt University, Nashville, Tennessee, USA.

INTRODUCTION

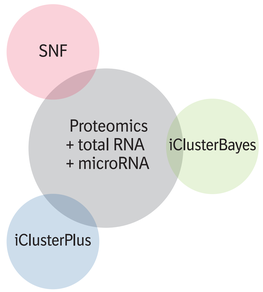
- Idiopathic pulmonary fibrosis (IPF) is a progressive fibrosing interstitial lung disease associated with high mortality.[1]
- The complexity and heterogeneity of the disease make phenotyping of patients with IPF a challenge.
- Leveraging different "omics" data may help us to establish an understanding of molecular-based IPF subtypes.
- The Idiopathic Pulmonary Fibrosis Prospective Outcomes (IPF-PRO) Registry (NCT01915511) is a multi-centre observational US registry of patients with IPF.[2]

AIMS

- Identify molecular IPF subtypes by simultaneous assessment of peripheral blood protein, total RNA and microRNA expression.
- Link the identified molecular subtypes with clinical factors at baseline and follow-up.

METHODS

- A subset of patients from the IPF-PRO Registry with proteomic, total RNA sequencing and microRNA sequencing data at baseline were used in integrative clustering analysis.
- Three approaches were applied: iClusterPlus,[3] iClusterBayes[4] and Similarity Network Fusion (SNF).[5]
- To determine the optimal number of clusters K, we leveraged Bayesian information criterion (BIC, the lower the better) and deviance ratio (% explained variation, the higher the better) for iClusterPlus and iClusterBayes, and Eigengap values for SNF (the higher the better). We then evaluated consistency in clustering membership across the models.
- To determine the clinical impact of the molecular-based subtypes identified, we compared measures of disease severity at baseline and time to the following composite outcomes between the two clusters identified using the SNF model:
  - death or lung transplant
  - death, lung transplant, or respiratory-related hospitalization
  - death, lung transplant, or decline in forced vital capacity (FVC) >10% predicted
  - death, lung transplant, or decline in diffusing capacity of the lung for carbon monoxide (DLco) >15% predicted.



CONCLUSIONS

- Integrative clustering approaches combining multi-omics data measured in the same set of patients with IPF provided a powerful tool for identifying molecular-based disease subtypes.
- Clustering results were generally consistent across methods. We observed significant differences in baseline measures of disease severity and outcomes related to disease progression between two clusters identified based on omics data.
- Future analyses will include investigation into proteins, genes and microRNAs that are differentially expressed between clusters.

RESULTS

- After data pre-processing, 243 patients had all three data types available for analysis.
- Overall, 1305 protein, 1051 total RNA and 472 microRNA analytes were included in the clustering analysis.
- The recommendation for the optimal number of clusters K varied across models:
  - iClusterPlus had monotonically decreasing BIC
  - iClusterBayes suggested three to four clusters
  - SNF suggested two clusters.
- There was large overlap in cluster memberships across the models (Table 1). Clusters identified using the SNF model were chosen for further evaluation as these represented the smallest number of clusters and captured the overlap in the other models.

Table 1. Cluster membership consensus across iClusterPlus, iClusterBayes and SNF

	iClusterPlus							
	Cluster	A1	A2	A3	A4	A5	A6	Total
SNF	C1	36	0	0	34	12	34	116
	C2	2	38	40	9	25	2	116
	Total	38	38	40	43	3	36	232

	iClusterBayes				
	Cluster	B1	B2	B3	Total
SNF	C1	63	45	8	116
	C2	7	31	78	116
	Total	70	76	86	232

Associations between SNF clusters and baseline measures of disease severity

- Results from the SNF model with two clusters showed that cluster 2 had significantly worse baseline disease severity based on mean DLco % predicted, FVC % predicted, and composite physiologic index (CPI)[6] (Table 2).

Table 2. Baseline disease severities between clusters identified using the SNF model

	Cluster 1 (N=116)	Cluster 2 (N=116)	P-value
DLco % predicted			
Median (IQR)	44.1 (32.8, 51.6)	39.3 (30.3, 47.0)	0.04*
Mean (SD)	43.0 (15.0)	38.9 (12.4)	
FVC % predicted			
Median (IQR)	73.1 (62.9, 84.1)	69.3 (60.1, 77.9)	0.03*
Mean (SD)	73.9 (18.2)	68.4 (15.0)	
FEV_1 % predicted			
Median (IQR)	78.7 (69.7, 91.5)	76.9 (67.2, 87.5)	0.21*
Mean (SD)	80.9 (20.3)	76.7 (16.3)	
CPI[6]			
Median (IQR)	51.5 (44.1, 59.4)	55.1 (49.2, 61.9)	0.01*
Mean (SD)	51.3 (11.8)	55.3 (9.8)	
GAP stage[7], n (%)			
I	36 (31.0)	26 (22.4)	0.33†
II	64 (55.2)	71 (61.2)	
III	16 (13.8)	19 (16.4)	
Diagnostic criteria[8], n (%)			
Definite IPF	87 (75.0)	85 (73.3)	0.23†
Probable IPF	27 (23.3)	24 (20.7)	
Possible IPF	2 (1.7)	7 (6.0)	

*Kruskal-Wallis test based on mean (SD). †Chi-square test. IQR, interquartile range. SD, standard deviation.

Association between SNF clusters and disease progression

- Based on the SNF model, cluster 2 had worse disease progression compared with cluster 1 (Figure). Time to death, lung transplant, or FVC decline >10% predicted was significantly different between the clusters (p=0.002).

Figure. Kaplan-Meier curves for composite disease progression outcomes by clusters identified using the SNF model

Log-rank test used to detect differences in outcomes.

REFERENCES

1. Raghu G et al. Am J Respir Crit Care Med 2018;198:e44–e68.
2. O'Brien EC et al. BMJ Open Respir Res 2016;3:e000108.
3. Mo Q et al. Proc Natl Acad Sci USA 2013;110:4245–50.
4. Mo Q et al. Biostatistics 2018;19:71–86.
5. Wang B et al. Nat Methods 2014;11:333–37.
6. Wells AU et al. Am J Respir Crit Care Med 2003;167:962–69.
7. Ley B et al. Ann Intern Med 2012;156:684–91.
8. Raghu G et al. Am J Respir Crit Care Med 2011;183:788–824.

ACKNOWLEDGEMENTS AND DISCLOSURES

The IPF-PRO™ Registry is funded by Boehringer Ingelheim Pharmaceuticals, Inc (BIPI) and co-ordinated by the Duke Clinical Research Institute. The authors meet criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE). The authors did not receive payment for the development of this poster. Editorial support and formatting assistance was provided by Julie Fleming and Wendy Morris of FleishmanHillard Fishburn, London, UK, which was contracted and funded by BIPI. BI was given the opportunity to review the poster for medical and scientific accuracy as well as intellectual property considerations.

Yi Liu is an employee of BIPI. Margaret Salisbury reports grants from NIH, personal fees from BIPI and personal fees from Orinove, Inc.

IPF-PRO™ Registry enrolling centers: Albany Medical Center, Albany, NY; Baylor College of Medicine, Houston, TX; Baylor University Medical Center at Dallas, Dallas, TX; Cleveland Clinic, Cleveland, OH; Columbia University Medical Center/New York Presbyterian Hospital, New York, NY; Duke University Medical Center, Durham, NC; Froedtert & The Medical College of Wisconsin Community Physicians, Milwaukee, WI; Houston Methodist Lung Center, Houston, TX; Lahey Clinic, Burlington, MA; Loyola University Health System, Maywood, IL; Lynchburg Pulmonary Associates, Lynchburg, VA; Medical University of South Carolina, Charleston, SC; National Jewish Health, Denver, CO; NYU Medical Center, New York, NY; Piedmont Healthcare, Austell, GA; Pulmonary Associates of Stamford, Stamford, CT; PulmonIx LLC, Greensboro, NC; Renovatio Clinical, The Woodlands, TX; Salem Chest and Southeastern Clinical Research Center, Winston Salem, NC; South Miami Hospital, South Miami, FL; St. Joseph's Hospital, Phoenix, AZ; Stanford University, Stanford, CA; Temple University, Philadelphia, PA; The Oregon Clinic, Portland, OR; Tulane University, New Orleans, LA; UNC Chapel Hill, Chapel Hill, NC; University of Alabama at Birmingham, Birmingham, AL; University of California, Davis, Sacramento, CA; University of California Los Angeles, Los Angeles, CA; University of Chicago, Chicago, IL; University of Cincinnati Medical Center, Cincinnati, OH; University of Louisville, Louisville, KY; University of Miami, Miami, FL; University of Michigan, Ann Arbor, MI; University of Minnesota, Minneapolis, MN; University of Pennsylvania, Philadelphia, PA; University of Pittsburgh, Pittsburgh, PA; University of Virginia, Charlottesville, VA; UT Southwestern Medical Center, Dallas, TX; Vanderbilt University Medical Center, Nashville, TN; Vermont Lung Center, Colchester, VT; Wake Forest University, Winston Salem, NC; Washington University, St. Louis, MO; Weill Cornell Medical College, New York, NY; Wilmington Health and PMG Research, Wilmington, NC; Yale School of Medicine, New Haven, CT.

Poster developed for the European Respiratory Society International Congress, 7–9 September 2020.