EDITORIALS

Integrated 'Omics in Idiopathic Pulmonary Fibrosis: Where Do We Go from Here?

Innovative studies with 'omics technologies have led to important insights into the pathophysiology of idiopathic pulmonary fibrosis (IPF). Studies using single-cell transcriptomics have demonstrated the role of novel epithelial and immune cell types in IPF lungs by identifying novel profibrotic alveolar macrophage populations (1) and dedifferentiated epithelial cell populations that secrete profibrotic mediators (2, 3). Genetic studies of familial pulmonary fibrosis have identified disease-causing genetic variants that disrupt genes involved in telomere maintenance, surfactant production, and mitochondrial functions, and studies of sporadic IPF have identified 14 genome-wide significant associations (4), including the particularly strong association with a promoter variant causing increased expression of *MUC5B* (5). Some of these discoveries map to known IPF-related pathways such as transforming growth factor-β (TGF-β) signaling and extracellular matrix production, but others point to novel processes or cell types. Through these discoveries and others, 'omics has successfully expanded the list of genes and biological processes involved in IPF, but can 'omics also generate hypotheses about how these disparate parts work together?

In this issue of the *Journal*, Konigsberg and colleagues (pp. 430–441) present a multiomic study of RNA sequencing, DNA methylation, and targeted proteomics generated from the same lung samples from 24 subjects with IPF and 14 comparison samples without chronic lung disease (6). The authors present differential expression results for each data type, highlighting both established and less well-known IPF-related mechanisms such as ciliary function, putative CUX1 regulation of type 1 collagen, and miR-21–associated TGF-β signaling. Pairwise relationships between data types were also examined to identify molecules exhibiting correlated behavior, and then the correlation patterns between all data types were examined using a modified version of canonical correlation analysis (DIABLO) to identify common axes of variability. Only one dimension of common variability or "latent variable" that differentiated IPF from control samples was identified, and the molecules most strongly associated with this latent variable included known IPF-associated genes such as MMP7 and LTBP1.

As in many other 'omics studies of IPF lung tissue, the sheer volume of associations can be overwhelming, and it is a challenge to make sense of the thousands of differentially associated RNAs, miRNAs, proteins, and methylated DNA regions. At times, 'omics analysis can be reminiscent of the hilarious scene from the movie *This Is Spinal Tap*, in which the fictional rock star Nigel Tufnel explains to a skeptical interviewer that the great thing about his amplifiers is that all the knobs can be turned to 11 rather than 10. When the interviewer asks whether simply changing the labeling on the knobs

actually corresponds to the amplifier being objectively louder, Tufnel seems mildly confused and simply repeats "These go to eleven." So it is with 'omics and biology, in which sheer data volume does not necessarily substitute for true biological insight or detailed elucidation of molecular mechanisms.

The analysis of Konigsberg and colleagues avoids this pitfall of "going to eleven" with 'omics by purposefully leveraging layered 'omics data to produce a number of functional hypotheses worthy of further investigation. The identification of the ciliary function–associated TMEM231 as a significantly upregulated protein in IPF extends previous work by this research group on a transcriptomic subtype of IPF characterized by genes involved in ciliary function (7). The identification of downregulation of the inflammatory mediator AGER at both the RNA and protein level adds to a growing number of interesting but unexplained molecular links between IPF and chronic obstructive pulmonary disease because s-RAGE, the soluble protein isoform of AGER, is one of the more promising biomarkers of emphysema (8, 9). Finally, focusing on MMP7, the authors identified multiple strong positive and negative correlations between *MMP7* RNA level and proteins, noncoding RNAs, and differentially methylated genomic regions, suggesting that some or all of these molecular entities may regulate or be regulated by *MMP7*.

This study has a number of analytic strengths. First, the use of multiple types of 'omics data provides complementary views of the same biological processes, allowing the authors to narrow the focus from thousands of significant results to a smaller set of molecular processes with correlated, reinforcing signals across data types. Second, pathway enrichment analysis leverages the multiplicity of 'omics measurements to identify biological processes in which the presence of multiple differentially expressed molecules can provide greater confidence in pathway-level associations. Third, the authors focus on established and novel pathways, using prior knowledge to anchor their interpretation of significant results and then using correlation analysis to identify connections between established IPF-associated molecules and putative functional partners.

There are also important limitations to this analysis. All data types are generated in bulk lung tissue; thus, differences in cell type proportions may explain some of the observed associations. The authors addressed this using algorithms to identify and adjust for estimated cell type differences, but such approaches are approximate and only partially mitigate the inherent limitations of bulk tissue analysis. Although extensive 'omics data are generated for each study sample, the sample size itself is small and insufficient to explore the molecular and clinical

heterogeneity of IPF. As there are few if any similarly characterized IPF lung collections, replication of these results in an independent cohort is yet to be performed. In this study, the IPF cases were purposefully selected to represent individuals with and without the *MUC5B* risk variant, but no molecules, including MUC5B, were significantly differentially expressed by genotype group. However, based on the effect estimates for *MUC5B*, it is likely that statistical significance would be achieved with a larger sample size. Finally, these multiomic data did not include genome-wide genotyping data, which can serve as a useful "causal backbone" for multiomics analysis.

At present, large multiomics data sets in thousands of samples from disease-based cohorts, including the Lung Tissue Research Consortium, are being generated through the National Heart, Lung, and Blood Institute Trans-Omics for Precision Medicine project. These data will enable robust replication of multiomics analyses like this one and will enable increasingly powerful new multiomics analyses. As the volume gets turned even higher for 'omics, we will do well to follow the example of Konigsberg and colleagues and strive to maximize innovation, interpretability, discovery, and validation of 'omics-based biological hypotheses. ■

Peter J. Castaldi, M.D., M.Sc.
Channing Division of Network Medicine
and

Division of General Internal Medicine and Primary Care
Brigham and Women's Hospital
Boston, Massachusetts

References

1. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, Chiu S, et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am J Respir Crit Care Med* 2019;199:1517–1536.
2. Xu Y, Mizuno T, Sridharan A, Du Y, Guo M, Tang J, et al. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 2016;1:e90558.
3. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* 2020;6: eaba1983.
4. Allen RJ, Guillen-Guio B, Oldham JM, Ma S-F, Dressen A, Paynton ML, et al. Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2020;201:564–574.
5. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011;364:1503–1512.
6. Konigsberg IR, Borie R, Walts AD, Cardwell J, Rojas M, Metzger F, et al. Molecular signatures of idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2021;65:430–441.
7. Yang IV, Coldren CD, Leach SM, Seibold MA, Murphy E, Lin J, et al. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. *Thorax* 2013;68:1114–1121.
8. Cheng DT, Kim DK, Cockayne DA, Belousov A, Bitter H, Cho MH, et al.; TESRA and ECLIPSE Investigators. Systemic soluble receptor for advanced glycation endproducts is a biomarker of emphysema and associated with AGER genetic variants in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013;188: 948–957.
9. Regan EA, Hersh CP, Castaldi PJ, Demeo DL, Silverman E, Crapo JD, et al. Omics and the search for blood biomarkers in COPD: Insights from COPDGene. *Am J Respir Cell Mol Biol* 2019;61:143–149.