Multiomics Approach to Asthma: Navigating the Network

With the advent of massive efforts such as the National Heart, Lung, and Blood Institute's Trans-Omics for Precision Medicine (TOPMed) (1) and the UK Biobank (2) aimed at generating omics data for thousands to millions of individuals, it is easy to imagine a day when biomedical scientists will no longer be limited by available sample sizes to gain insight into the etiology of human disease. However, omics data refer to more than genetic variation at the nucleotide level whose seemingly elusive sequence fueled the race for the first draft of the human genome over 15 years ago (3, 4). Omics data encompass multiple omes, including the genome, transcriptome, proteome, metabolome, epigenome, and microbiome. The challenge will soon be primarily focused on whether methodological approaches are available to make sense of omics data in order to advance our understanding of health and disease. In this issue of the *Journal*, Forno and colleagues (pp. 439–447) describe their approach to the daunting task of integrating findings from multiple omics approaches using a "vertical" analysis strategy to decipher the etiology of childhood asthma (5).

The authors analyzed multiomics data (including genotype, methylation, gene expression, and cytokine data) collected as part of a case-control study of 1,127 Puerto Rican children to gain insight into the etiology of childhood asthma. They began their vertical analysis by identifying 1,645 genes whose expression was associated ($P < 0.01$) with several cytokines (IL-10, IL-17A, IL-17F, IL-22, and IL-23). This list of genes was reduced to 269 genes involved in cytokine signaling based on gene-set enrichment analysis. The authors further screened the list of 269 genes down to one gene, *IL5RA*, using seven additional vertical-analysis steps incorporating omics data and clinical phenotypes. As the final step, they measured plasma levels of IL5-Rα in a subset of the children with asthma ($n = 130$) and found associations with age of onset and acute exacerbations of asthma.

The authors employed a vertical approach (which has also been called sequential analysis) to analyze omics data (6, 7). A challenge with using this form of sequential analysis is that the results become dependent on the order of the analytical steps. Replication of the findings of Forno and colleagues with IL5-Rα in an independent cohort would provide greater confidence in their validity. Further, the authors acknowledge a lack of correction for multiple testing in their analytical approach. However, permutation can be used to address this limitation of sequential analysis. Permutation in genomics analyses typically involves swapping case-control labels randomly many times, and calculating test statistics at each permutation to observe how often under the null hypothesis an observation as significant as the top findings occurs (8). The approach is computationally intensive but is parallelizable and thus can be facilitated by analyses on local and/or cloud-based computational clusters. Permutation can also be used to compare the findings from multiomics approaches with those obtained using a single omics approach. For example, a permutation approach was used to assess whether findings associated with chemotherapeutic sensitivity from a sequential analysis of multiomics data had a lower false-discovery rate than those from a genome-wide association study (9). Thus, one should consider using permutation when addressing analytical concerns in a sequential analysis of multiomics data.

In general, parallel or simultaneous analyses flag significant biomarkers in each set of omics data and short-list the intersection of top findings. Other methods have been developed to integrate omics data (reviewed in References 6, 7). One review (7) grouped integrative omics methods based on whether they were network based and/or used a Bayesian approach. Network-based methods typically apply approaches, such as integration of data from genotypes, gene expression levels and protein-protein interactions, to understand how perturbations in the system may lead to complex diseases (10, 11). Network-based methods offer a departure from reductionism by incorporating models that mirror biological systems. Investigators have used networks to gain insight into the disease properties of genes by examining the association of hub and peripheral node genes with disease (12). However, metrics other than node degree (the number of connections a node has to other nodes in the network) have also been used to gain insight into the important role that nodes in a biological network (which could be genes, metabolites, or proteins) can play in disease (reviewed in Reference 13). In a recent publication, a network-based Bayesian approach was used to generate a Conditional Gaussian Bayesian Network (CGBN) integrating metabolomics data with genome-wide genotype, gene expression and methylation data ($N = \sim 20$ children with asthma) (14). They first performed a parallel analysis to identify the top expression, SNP, and methylation probes associated with a panel of metabolites as input for the conditional Gaussian Bayesian network, and then performed a pathway overrepresentation analysis on the metabolite data alone. Both approaches implicated sphingolipid metabolism in asthma control; however, the analysis did not include replication of the findings in an independent population.

Overall, the field of integrative omics is still in a nascent stage. Approaches such as the one Forno and colleagues (5) employed in their current study provide a tangible starting point for researchers interested in analyzing multiomics data. Sequential or "vertical" analysis methods can be relatively easy to implement. When employing such methods, one should consider the importance of the order of the analytical steps, replication in an independent population, and adjustment by permutation. ■

This work was supported by grants from the National Institutes of Health (4R00HL121087) and the Parker B. Francis Foundation to M.-L.M.

Am J Respir Cell Mol Biol Vol 57, Iss 4, pp 381–382, Oct 2017

DOI: 10.1165/rcmb.2017-0220ED
Internet address: www.atsjournals.org

Author disclosures are available with the text of this article at www.atsjournals.org.

Merry-Lynn McDonald, M.Sc., Ph.D.
Division of Pulmonary, Allergy and Critical Care Medicine
University of Alabama at Birmingham
Birmingham, AL

ORCID ID: 0000-0002-2647-9422 (M.-L.M.).

References

1. National Heart, Lung, and Blood Institute. Trans-Omics for Precision Medicine (TOPMed) program [accessed 2017 Jun 12]. Available from: https://www.nhlbi.nih.gov/research/resources/nhlbi-precision-medicine-initiative/topmedweb
2. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, *et al*. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779.
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, *et al*.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
4. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, *et al*. The sequence of the human genome. *Science* 2001;291:1304–1351.
5. Forno E, Wang T, Yan Q, Brehm J, Acosta-Perez E, Colon-Semidey A, Alvarez M, Boutaoui N, Cloutier MM, Alcorn JF, *et al*. A multiomics approach to identify genes associated with childhood asthma risk and morbidity. *Am J Respir Cell Mol Biol* 2017;57:439–447.
6. Rotroff DM, Motsinger-Reif AA. Embracing integrative multiomics approaches. *Int J Genomics* 2016;2016:1715985.
7. Bersanelli M, Mosca E, Remondini D, Giampieri E, Sala C, Castellani G, Milanesi L. Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinformatics* 2016;17:15.
8. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet* 2014;15:335–346.
9. LaCroix B, Gamazon ER, Lenkala D, Im HK, Geeleher P, Ziliak D, Cox NJ, Huang RS. Integrative analyses of genetic variation, epigenetic regulation, and the transcriptome to elucidate the biology of platinum sensitivity. *BMC Genomics* 2014;15:292.
10. Silverman EK, Loscalzo J. Network medicine approaches to the genetics of complex diseases. *Discov Med* 2012;14:143–152.
11. Loscalzo J, Barabasi AL. Systems biology and the future of medicine. *Wiley Interdiscip Rev Syst Biol Med* 2011;3:619–627.
12. Goh K-II, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. *Proc Natl Acad Sci USA* 2007;104:8685–8690.
13. Furlong LI. Human diseases through the lens of network biology. *Trends Genet* 2013;29:150–159.
14. McGeachie MJ, Dahlin A, Qiu W, Croteau-Chonka DC, Savage J, Wu AC, Wan ES, Sordillo JE, Al-Garawi A, Martinez FD, *et al*. The metabolomics of asthma control: a promising link between genetics and disease. *Immun Inflamm Dis* 2015;3:224–238.