Cell Reports Medicine

CellPress
OPEN ACCESS

Perspective

Determining asthma endotypes and outcomes: Complementing existing clinical practice with modern machine learning

Anuradha Ray,[1,2,5,*] Jishnu Das,[2,3,5] and Sally E. Wenzel[1,2,4,5]
[1]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pittsburgh School of Medicine, 3459 Fifth Avenue, MUH 628 NW, Pittsburgh, PA 15213, USA
[2]Department of Immunology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[3]Center for Systems Immunology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA
[4]Department of Environmental Medicine and Occupational Health, School of Public Health, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[5]These authors contributed equally
*Correspondence: raya@pitt.edu
https://doi.org/10.1016/j.xcrm.2022.100857

SUMMARY

There is unprecedented opportunity to use machine learning to integrate high-dimensional molecular data with clinical characteristics to accurately diagnose and manage disease. Asthma is a complex and heterogeneous disease and cannot be solely explained by an aberrant type 2 (T2) immune response. Available and emerging multi-omics datasets of asthma show dysregulation of different biological pathways including those linked to T2 mechanisms. While T2-directed biologics have been life changing for many patients, they have not proven effective for many others despite similar biomarker profiles. Thus, there is a great need to close this gap to understand asthma heterogeneity, which can be achieved by harnessing and integrating the rich multi-omics asthma datasets and the corresponding clinical data. This article presents a compendium of machine learning approaches that can be utilized to bridge the gap between predictive biomarkers and actual causal signatures that are validated in clinical trials to ultimately establish true asthma endotypes.

INTRODUCTION

Over the past decade, technologies for deep profiling of the human system, across a range of disease contexts, have become readily available. These include a wide range of molecular profiles built on genomic, epigenomic, transcriptomic, proteomic, metabolomic, and antibody-omic datasets.[1] A key aspect of these developments has been the advent of single-cell methodologies.[2] These technologies are providing deep molecular snapshots of diverse cell types in which molecular interactions at different levels maintain homeostasis in health but are dysregulated in disease. To identify features that differentiate health from disease, existing studies have often focused on individual omic datasets, which provide a deep molecular profile of only one aspect of biological regulation. A combination of omic datasets can provide a significantly enhanced and holistic picture of interactions across the components underlying the pathophysiology of disease processes.[1,3–5] With rapid advances in modern machine learning, a suite of approaches has become available to integrate these high-dimensional multi-modal datasets. An important distinction here is between artificial intelligence (AI) and machine learning. Broadly, AI connotes ''imparting'' human intelligence to machines, and its application is evident in a range of contexts that includes clinical decision-making, biomarker discovery, drug discovery, 3D printing, and even self-driving cars.[6–12] Machine learning is a subset of AI focused on learning patterns or relationships (both discriminative and generative) from data.[6] In this article, we have focused on how modern machine learning techniques can be used to unravel the complexity and heterogeneity of asthma, both to define disease endotypes and outcomes, as well as to complement clinical practice with patterns learned from multi-omics datasets. These techniques encompass both data-centric approaches that attempt to improve model performance by augmenting the underlying data (i.e., by collecting additional layers of multi-modal data), as well as model-centric approaches that attempt to improve performance by improving the actual modeling approach. The two have complementary strengths and are appropriate in different contexts. We also draw key distinctions between the use of these machine learning approaches in prediction (e.g., defining correlates/predictive biomarkers) and inference of actual molecular mechanisms.

DEFINING ASTHMA ENDOTYPES

Despite the appreciation that ''asthma,'' as a chronic disease, is considered to be at the forefront of the world of precision medicine, the approaches and, importantly, their application in the

clinic remain rudimentary at best. While it is broadly appreciated that the term "asthma" is an umbrella term that encompasses multiple different clinical and molecular phenotypes,[13] there remains considerable controversy as to what are (1) the "true" endotypes and the factors driving them, (2) the long-term implications for both outcomes and treatment, and (3) how best to practically and inexpensively identify them. This section will explore the role that AI/machine learning has played in these areas to date, but perhaps more importantly, the role it can play moving forward.

Clinical heterogeneity of asthma has been appreciated for decades, due in part, to the broad, definition of the disease: nonspecific respiratory symptoms (wheeze, chest tightness, and/or shortness of breath) in the presence of physiologically confirmed reversible lower airway obstruction or hyperresponsiveness. Using *human-based* neurocognitive machine learning, astute clinicians observed distinct differences in clinical characteristics according to age at onset of symptoms and the relation to atopy/allergy, sinus disease, and asthmatic reactions to aspirin and other cyclooxygenase inhibitors.[14,15] However, the emergence of non-specific short-acting beta agonists and corticosteroids (CS) as the cornerstone of treatment *for most patients with asthma* for decades nearly eliminated the concept of heterogeneity. This instead enhanced the concept that all asthma was the same and should be treated the same, likely setting back studies of its molecular underpinnings and the development of successful targeted therapies.

THE EMERGENCE OF MOLECULAR PHENOTYPING

Since the early 2000s, it became increasingly recognized that, indeed, not all asthma patients responded to CS. Pathologic studies of these severe, poorly CS-responsive patients showed heterogeneity, and the first reports of efficacy of biologic agents in biomarker-targeted subgroups helped propel asthma into the age of precision medicine.[16–18] These efforts have now been translated into the clinic, and in almost all specialty practices, patients with a clinical diagnosis of asthma are characterized on the basis of at least one and sometime two biomarkers: blood eosinophils and fraction exhaled nitric oxide (FeNO). Variably identified as eosinophilic asthma or type-2 high (T2-Hi) asthma, these two simple biomarkers have identified asthma patients who generally respond to blockade of the type 2 (T2) cytokines, IL-4, -5, and/or -13. While this approach has improved clinical outcomes, this simple T2 biomarker approach remains limited in its ability to identify the highest risk patients or those who respond the best.

MOVING TOWARD ENDOTYPES

For many, this association of T2 biomarkers with efficacy of targeted therapy has promoted the concept of endotype. While the definition of an endotype is still controversial, there is increasing consensus that an endotype is a subtype of a health condition *defined* by a distinct functional or pathobiological mechanism. This necessity for a confirmed mechanism differentiates endotype from phenotype, which in its highest form is a collection of linked clinical, physiological, and/or cellular/molecular characteristics, without a known *essential* molecular driver. With effectiveness of monoclonal antibodies toward IL-4 receptor alpha (IL4Rα), IL-5, and its receptor (IL5RA) in patients with elevations in T2 biomarkers, the concept of a T2 endotype mechanistically driven by these cytokines has been promoted.[19–21] Yet, further inspection of the data regarding the essential nature of T2 cytokines to all patients fulfilling the current simple biomarker criteria reveals a vast range of responses to these therapies, from truly life changing to minimal improvement to actual worsening of disease (or concomitant side effects). It seems likely that in addition to the T2 cytokine profile, additional factors, including baseline clinical, metabolic, and immune characteristics, may collectively determine treatment efficacy.[21,22] Thus, it is likely that defining an endotype of T2-Hi asthma, using biomarker criteria alone, is premature. The unprecedented and even curative responses to T2 biologics in some patients do, however, support the identification of a true T2-Hi endotype in a minority of patients. In these patients, T2 pathways could be a true critical node for the pathobiology of the endotype. In others, with similar T2 biomarker elevations, T2 pathways could be only one of several intersecting non-critical nodes, reactionary, parallel, and even non-contributory. While clinical characteristics such as age at onset and concurrent nasal polyps may predict responses,[22] the complex *molecular* characteristics that determine responses to T2-targeted therapies remain unknown. Collectively, given that immune-mediated inflammatory mechanisms are clearly at play in asthma, the challenge is to identify the key cytokine hubs that orchestrate disease in each patient for maximum therapeutic efficacy, as also true in other immune-mediated inflammatory diseases.[23]

Molecular studies of asthma and severe asthma, where the current highest need exists, are challenging and expensive studies to complete. Patients manifest differing cellular and molecular processes by physical compartment, including differences in multiple lung compartments and in peripheral blood, so they require collection of samples from both the lungs and periphery. Lung sampling of asthma patients is not trivial and requires a dedicated participant/patient. Molecular patterns are also impacted by background therapy and duration of disease. Incorporating clinical/physiologic (including environmental and social determinants[24,25]) and radiologic characteristics[26] (used collectively with molecular, genetic, and other omic features) into the input variables (rather than as outcomes) is also important. Additionally, comorbidities can also influence asthma and the overall molecular profile.[27,28] These studies require generation of vast amounts of highly diverse data types that remain difficult to harmonize, and even more difficult to harmonize from study to study. Finally, and critically, endotyping requires confirmation only obtainable through directed and targeted interventions in which up- or downregulation of that pathway profoundly impacts asthma outcomes. A paradigm for approaching identification of endotypes is presented in Figure 1.

UNDERSTANDING ASTHMA HETEROGENEITY USING OMICS

Large-scale molecular phenotyping studies of asthma to date (typically bulk RNA-seq or microarrays) have focused on a single

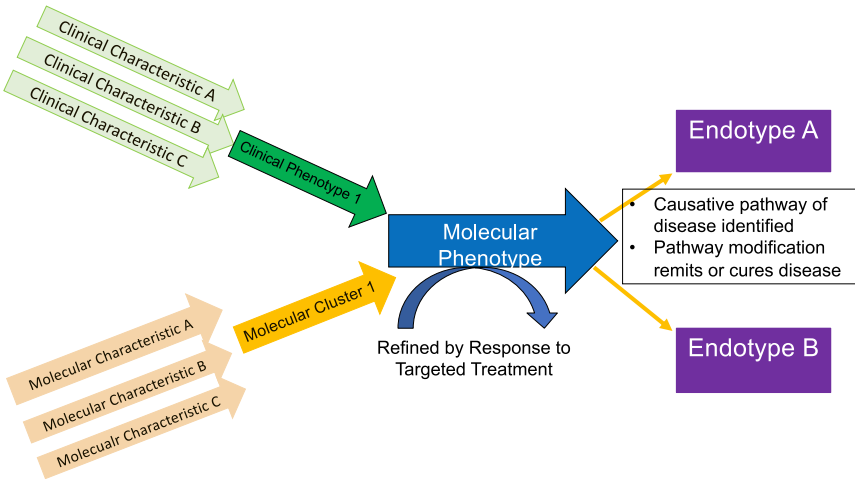

Figure 1. From clinical characteristics to endotypes

The process first requires integration of clinial information and molecular profiles derived from omics or other studies to identify molecular phenotypes. A molecular phenotype reaches endotype status when causality is demonstrated by targeting that specific pathway with relief of disease symptoms.

compartment, at a single time, and often only in relation to predefined asthma severity or biomarker endpoints. However, there is an increasing need for an integrated multi-omics approach to dissect the underlying pathophysiology of asthma for advancing precision medicine in this disease. Recent reviews have covered many of the major findings in these multi-omics studies,[29,30] although the immune diversity in the asthmatic airway has not yet been adequately reviewed. This is primarily because of a paucity of such data in the literature. In this article, we have discussed how machine learning would help to establish the broader significance of the findings in the omics studies and help in establishing asthma endotypes and predicting outcomes.

Use of an omics approach to understand disease pathogenesis is now standard practice in which access to the relevant biological specimens is feasible. In most studies, the goal has been to identify disease-associated gene signatures and/or biological pathways. Omics approaches must be able to handle many variables and many measurements. Each omic approach applied to study any disease captures only one aspect of molecular processes. This may include the genome, transcriptome, proteome, metabolome, epigenome, or the microbiome. There is no one set of rules that is used to analyze any omics data, and newer algorithms are constantly evolving to refine analysis.[31] Thus, any dataset can be analyzed in a multitude of ways, and therefore it is essential to validate findings in appropriate experimental systems.

Omics studies have their roots in genomics data, initially generated in the form of family-based linkage analysis transitioning to more advanced genome-wide association studies (GWAS). These studies have identified numerous candidate asthma risk genes that include the 17q12-21 locus,[32,33] which has been replicated in multiple independent studies.[34] Using asthma GWAS findings alone for therapeutic guidance has been challenging because of confounding variables present in the different cohorts. Given that GWAS studies have shown that a specific genetic variant can be associated with more than one disease,[35] in today's era of precision medicine, there is increasing interest in establishing the human genome-phenome relationship. How do asthma-associated genes relate to the human phenome? It seems plausible that the association of genetic variants identified in studies of asthma with the range of clinical phenotypic characteristics may be better realized when integrated directly with other omics data, potentially identifying biomarkers and clinically actionable pathways. An additional level of dimensionality on molecular phenotypes will be provided by incorporating single nucleotide polymorphisms in genes that functionally impact gene expression (expression quantitative trait loci -eQTLs) or protein sequence (protein quantitative trait loci -pQTLs) in a compartment-specific fashion, as shown for the *IL1RL1* (*ST2*) gene locus.[36]

While the bulk of omics studies in asthma have utilized transcriptomic data using RNA from bronchial and nasal brushings, endobronchial biopsies, and sputum and peripheral blood,[29,30] one study performed single-cell RNA sequencing (scRNA-seq) analysis of mild asthma patients.[37] The transcriptome of bronchoalveolar lavage (BAL) cells was interrogated in only two recent studies.[38,39] The older platform to study the transcriptome was DNA microarray, which has now been replaced by RNA sequencing, and more recently by scRNA-seq.[40] Each RNA-seq study generates a large number of data points, and integrating data from multiple compartments requires computational power and poses various statistical challenges. This is an important area of consideration as there is no gold standard for harmonizing data across different compartments.

How have transcriptomic or other omics data using specimens from different compartments helped in furthering our understanding of asthma endotypes? Are there overlapping signatures across different compartments? The latter is an important question in studies of asthma since specimens from blood, sputum, or nasal brushings can be obtained in a much less invasive fashion than those obtained by endobronchial biopsies or BAL. This helps to increase sample size and affords the opportunity for longitudinal analysis.

ENDOTYPING ASTHMA USING OMICS APPROACHES

To achieve consensus on asthma endotypes, an integrative approach is needed. Collectively, transcriptomic data are the richest information that can be tapped to refine asthma endotypes using machine learning. Samples from multiple compartments have been used to derive RNA-seq data with the objective of identifying biological pathways underlying the asthma disease spectrum from mild to most severe disease.

Two general approaches have been used to associate omics data with clinical phenotypes. The most common approach has associated gene expression data with a clinical characteristic.[41–45] In a second approach, unbiased methods have been used to cluster genes that in turn have been linked to specific characteristics.[46,47] Multiple transcriptomic studies have examined differentially expressed genes (DEGs) in relationship to clinical features such as persistent airflow limitation or inflammatory phenotype. One of the early studies associated DEGs with categorical variables such as asthma and absence of disease.[48] In other studies, using gene set variation analysis (GSVA), DEGs were evaluated for the presence of previously identified gene clusters derived from human *in vitro* or mouse-model studies.[41,42,49] These analyses were limited by the accuracy in classification of clinical phenotypes often based on categorical cut-points that lacked biologic validation. GSVA gene sets can also provide incorrect information on pathway analyses, especially when mouse comparator gene sets are included. Overall, it is likely that feature deficiencies in both the independent and dependent variables explained the modest associations found between GSVA-derived gene signatures and clinical characteristics. It is also plausible that data derived from gene set enrichment analysis may suffer from similar limitations. Despite the limitations, these studies independently associated eosinophilic or T2 cytokine (IL-4/5/13) gene sets with traits such as persistent airflow limitation or eosinophilic asthma phenotypes.[41,43] Events in the bronchial epithelium of severe asthmatics were evident by applying weighted gene co-expression analysis to airway epithelial cell microarray data.[46] Over 60 different gene modules were identified, grouped by similarities of genes with respect to biologic functions including previously unrecognized relationships between epithelial growth and repair, mitochondria genes and neural processes, both of which could underlie pathogenesis of severe asthma.

A BROAD T2 MOLECULAR PHENOTYPE: REACHING ENDOTYPE STATUS

A Th2 immune response was associated with the inflammatory response in the asthmatic airway in the early 1990s.[50] Since then, T2 immune phenotype has taken center stage as a broad molecular phenotype in asthma and has replaced the terminology Th2 because of production of the Th2-associated cytokines (IL-4, IL-5, and IL-13) in the asthmatic airway by other cell types such as ILC2s, mast cells, basophils, and even eosinophils. The broader cellular source of T2 cytokines also suggests the role of different mechanisms in driving disease.[51] A puzzling aspect related to the T2 gene signature in biological specimens collected from both mild and severe asthma patients is why the T2-Hi immune response in mild asthma is generally responsive to CS therapy[52] but is refractory to CS in severe asthma, even when used at high doses.[53] Moreover, not all patients selected for T2-directed therapy, based on the levels of T2-associated biomarkers, blood eosinophils, and FeNO, respond to therapy.[22,54] Why do some patients respond to these biologics with complete alleviation of symptoms and reduced requirement for CS, while others do not? How can one more accurately identify the endotype that is characterized by overactivation of IL-4Rα? Although epithelial transcriptomic studies show consistent increases in T2 signatures, additional pathological changes such as epithelial oxidative stress, lack of repair, embedded mast cell-derived mediators, and aberrant activation of innate pathways may also determine more severe disease.[39,46,55,56] To unmask the heterogeneity in disease mechanism and more precisely designate an asthma endotype that is not evident by assaying T2 biomarkers alone, it will be helpful to employ machine learning tools to integrate clinical data with multi-omics data derived from different biological compartments.

T2-HI AND T2-LO ASTHMA

Sputum transcriptomic data from adults have recently identified a T2-Lo category of adult asthma characterized by late onset of disease.[47] While the prevailing concept is that these patients are females with higher BMI, additional studies of independent cohorts are needed to substantiate this notion. Their disease is poorly responsive to CS therapy, and symptoms often worsen from long time use of CS. Is there a common mechanism of CS insensitivity in these T2-Hi and T2-Lo asthma populations that can be learned by machine learning? Identification of novel mechanisms underlying CS insensitivity may help in the development of next-generation CS whose application may transcend asthma.

IMMUNE PROFILING IDENTIFIES T1-HI SEVERE ASTHMA WITH VALIDATION ACROSS COHORTS USING MACHINE LEARNING

By integrating immune phenotypes with transcriptomic data and machine learning, we recently identified distinct immune phenotypes in severe asthma that were indistinguishable by existing biomarkers.[38] Although characterization of immune cells in the airways may provide deeper insight into immune dysregulation in the distal airway, there are few studies that have interrogated BAL cells by deep immune profiling. Our study identified two divergent immune phenotypes in patients with severe asthma by analyzing BAL immune cells using mass cytometry/CyTOF. We found that one group of patients had an abundance of IL-4+ and IL-5+ FceRIα+ innate immune cells distinct from mast cells or basophils, while the other group had a heightened adaptive immune response that was characterized by a dominance of $CD4^+$ and $CD8^+$ IFN-γ+ T cells comprising tissue resident memory cells and T effector/memory. This second group also had Th2 cells, albeit fewer than IFN-γ+ T cells, and still fewer Th17 cells.[38] Transcriptomic data from the T cell-enriched group also revealed signatures of mast cells. Clearly, the immune response in both groups was refractory to high doses of CS. Because of the existence of a common T2 response in both groups, the patients were indistinguishable by standard biomarker measurements (blood eosinophils and FeNO). Of note, although NOS2 is induced in the asthmatic epithelium by T2 cytokines, IFN-γ is also a potent inducer of NOS2 in epithelial cells.[57,58] Moreover, IFN-γ has been shown to promote eosinophil activation in humans.[59,60] Using a novel deconvolution algorithm, ICLite, gene modules derived from transcriptomic data were linked to the immune cells.[38,61] ICLite does not use pre-defined gene

expression sets for deconvolution. It breaks bulk transcriptional data into smaller sets of modules of correlated genes. The gene modules are then linked to specific cell populations within a mixture of cell types. Machine learning enabled validation of the data in an independent asthma cohort for which only BAL cell transcriptomic data were available.[38] Novel therapeutic targets beyond T2 were also unraveled in the study. Other available transcriptome deconvolution algorithms include CIBERSORT and AutoGeneS.[62,63] CIBERSORT relies on preselected gene expression signatures for deconvolution of bulk RNA-seq data to estimate cell types of interest. AutoGeneS does not require this prior knowledge and selects discriminative genes based on multiple criteria to infer cell-type proportions.

USE OF OMICS TO STUDY ASTHMA EXACERBATIONS

A particularly debilitating aspect of asthma is disease exacerbation.[64] Asthma exacerbations are episodic in nature and can cause rapid deterioration of lung function requiring immediate medical attention. What are the signatures of asthma exacerbations? Can they be detected in samples obtained by less invasive methods? It is important to note that using bioinformatics to predict exacerbations (asthma attacks) is likely as difficult as predicting one disease of "asthma," as reporting of exacerbations is highly patient dependent and subjective, with potentially different drivers by different types of exacerbations. Notwithstanding this consideration, a few studies are discussed that looked at asthma exacerbation using a bioinformatic approach. A single longitudinal analysis of nasal brushings and peripheral blood of 208 children reported signatures and predictors of asthma exacerbations *during* exacerbations.[65] The study identified distinct gene signatures across the time course of disease establishment. For example, epithelial-associated SMAD3 signaling was found early in exacerbations, which is interesting given that SMAD3 is downstream of TGF-β, which is associated with airway remodeling. A subsequent upregulation of epidermal growth factor receptor signaling, extracellular matrix production, mucus hypersecretion, and eosinophil activation was observed, highlighting the dynamic nature of molecular events during an exacerbation. However, the finding of a high T2 inflammation and low type I IFN response gene signature in nasal samples at baseline that predicted short-term exacerbation risk is particularly useful clinically. Also, it is in line with an earlier finding that showed that nasal IL-13 expression portends exacerbation risk.[66] Identification of distinct gene modules in virus- versus non-virus-mediated exacerbations was another important finding in this study. However, studies of lower airways are not feasible during exacerbations. A study of an adult population found that expression of a single gene, *CEACAM5*, in bronchial biopsies distinguished persistent versus persistent frequent exacerbators.[67] It is unclear whether higher *CEACAM5* expression during exacerbations is specific to the lower airways given that the study in children did not examine lower airway samples, which are difficult to obtain from children. A recent study analyzed BAL cells of asthma patients with historical exacerbations by scRNA-seq.[68] CD8 T cells and several monocyte clusters were found to be more abundant in the BAL fluid of patients with asthma exacerbations compared with that in healthy controls. Previous studies have also associated CD8 T cells with asthma exacerbations. A study performed RNA-seq of nasal brushings from 190 subjects and using a machine learning pipeline developed an asthma classifier gene module comprising 90 genes that helped to distinguish asthma and healthy control.[69] The classifier was validated across eight test sets that included RNA-seq data from an independent cohort.[69] While this study did not focus on asthma exacerbations, the ability to utilize machine learning to identify a gene signature in asthma using samples that can be accessed easily is encouraging. Undoubtedly, establishing gene signatures for key asthma traits is more challenging because of the need for adequate sample size and the importance of validation in independent cohorts. However, it is clear that machine learning tools can be applied to existing and emerging datasets to determine the broader significance of current data on asthma exacerbations and ultimately to establish causality.

LESSONS LEARNED FROM OTHER OMICS STUDIES

It is important to integrate data from other omics approaches with the transcriptomic datasets to establish asthma endotypes. However, there are only few omics studies that have explored epigenomics, proteomics, or metabolomics of human asthma.

Epigenomics is important to study given that asthma develops from gene-environment interactions. Studies of peripheral blood mononuclear cells have identified methylation in gene loci associated with T2 immunity that track with serum IgE levels.[70] An epigenomic study across tissues, including airway epithelial cells, identified CpG methylation sites in the airway cells of asthmatic patients.[71] Gene network analysis identified four modules, one of which associated with eosinophil levels in BAL fluid, another with FeNO, and the remaining two with CS use. The authors used a systems biology approach to integrate GWAS, epigenetic, and transcriptomic data to associate epigenetic signatures with distinct molecular pathways that would not have been possible using transcriptomic data alone.

Similarly, unraveling asthma metabolomics is also important since there is emerging evidence of an important role of cellular metabolism, influenced by not only the external environment but also by microbiota, in driving immune responses and cellular function as a whole. In one study, volatile organic metabolites in exhaled breath were found to be better indicators of response to CS and disease outcome than FeNO levels.[72] Using nuclear magnetic resonance (NMR) spectroscopy and machine learning, NMR spectra of exhaled breath condensate clustered asthmatic patients into three phenotypic groups. One of the clusters associated with neutrophilic asthma with low peripheral blood eosinophil and high neutrophil counts.[73] In another study, metabolomic data from exhaled breath allowed discrimination between peripheral blood eosinophil and neutrophil levels, as well as differences in oral CS use.[74] While these studies bode well for using metabolomics data for asthma endotyping, there was heterogeneity in the findings because of limited sample size. However, exhaled breath is an attractive source of biological material that lends well to asthma endotyping in both children and adults. As in the epigenomic study, it will be interesting to integrate the metabolomic data with other omics data available

from public databases to determine the significance of the findings beyond a single cohort.

There has been limited progress in the area of asthma proteomics with early studies devoted to assay of cytokines and chemokines in biological fluids by immunoassays.[75,76] The early immunoassay-based investigations have progressed to use of unbiased mass spectrometry to characterize the asthma proteome. Liquid chromatography-mass spectrometry analysis of the sputum proteome identified proteomic clusters and potential protein biomarkers that were associated with eosinophilic, neutrophilic, or highly atopic with low granulocytic inflammation.[77] Employing shotgun mass spectrometry, 18 proteins were identified that differed in abundance between allergic and nonallergic asthma, allergic rhinitis, and healthy controls.[78] As in all omics studies, there is a need to standardize proteomics protocols that would allow replication of data across independent cohorts in order to have confidence when associating with asthma traits. Also, no current technology is yet able to capture all components of the human proteome.

INTEGRATIVE OMICS FOR BETTER UNDERSTANDING OF ASTHMA ENDOTYPES AND TREATMENT STRATEGIES

To date, based on identification of a T2 phenotype in both CS-responsive mild allergic asthma and CS-refractory severe asthma, multiple clinical trials have targeted molecules associated with the T2 phenotype.[79] However, despite best efforts to select patients based on their biomarker status, primarily blood eosinophil levels and FeNO, the response to treatment has not been uniform.[80] In a recent study that targeted the alarmin IL-33, surprisingly, simultaneous targeting of two T2-associated molecules, IL-4Rα and IL-33, did not lead to a better outcome than when either was targeted alone.[81] This suggests that complete blockade of one immune pathway may shift the balance to another since immune responses cross-regulate each other.[82] It is clear that one subgroup of severe asthma patients harbor a high T1 (IFN-γ) immune response admixed with a T2 response. It is conceivable that complete blockade of T2 in some individuals harboring a mixed immune response would trigger an unchecked T1/IFN-γ response, as observed during the treatment of atopic dermatitis,[83] a high IFN-γ response having been associated with severe asthma in multiple studies.[38,53,84–90] These findings overall show that there is a need to integrate data derived from different platforms to make better informed decisions about targeting specific pathways to treat asthma. Integrative multi-omics has been applied to unravel disease mechanisms and host-microbe interactions in chronic obstructive pulmonary disease.[3,5]

No study has yet integrated molecular data from multiple compartments to identify novel *recognizable* phenotypes or endotypes, with treatment implications. In fact, few studies have attempted to harmonize multiple different types of data,[91,92] with none integrating clinical and biological/omics data. Novel unbiased or less biased approaches that harmonize distinct data types, ideally over time, to derive novel phenotypes are indeed rare, yet urgently needed. In addition to omics data, it would be important to integrate social, clinical, and environmental data associated with patients, as well as additional characterizations of the airways, including computed tomography imaging to better stratify asthma subtypes. For example, in addition to traditional methods of spirometry, the tool of impulse oscillometry shows promise in identifying airway dysfunction in both children and adults.[93] Disease characterization would also immensely benefit from inclusion of longitudinal omics and clinical data.

MACHINE LEARNING TOOLS TO IMPROVE CLINICAL CARE OF ASTHMA

As discussed above, multi-omics datasets (genomic/epigenomic, transcriptomic, proteomic, metabolomic, and lipidomic profiles) are now publicly available with associated clinical data that can be used to derive more precise information about molecular phenotypes and their relationship to asthma traits. In some cases, these molecular phenotypes can emerge as endotypes when they correspond to differential disease outcomes by pathway-targeted treatment modalities. Moving from molecular phenotypes to endotypes needs rigorous validation through relevant trials (Figures 1 and 2), which have not been performed yet. As has been shown, combining multi-omics profiles generated from the same individual can reveal important features not afforded by analysis of a single data type.[94] Consideration should be given to incorporating (and harmonizing) clinical data as well, so that it can also be analyzed taking into account relevant covariates. Multi-omic integration with machine learning approaches can be conceptually distinguished in different ways. The first is to categorize these approaches into unsupervised (i.e., do not use labels in the model training process) vs. supervised ones (i.e., incorporate labels in the model training process). They can also be categorized based on the underlying goals, e.g., predictive machine learning models focused on the discovery of predictive/correlative biomarkers vs. interpretable machine learning models that can provide inference of causal factors beyond biomarkers.[95] Within both sets of approaches (predictive vs. interpretable machine learning), a subset of techniques use prior knowledge, while others are solely based on the underlying data.

Unsupervised approaches (i.e., approaches that only take into account data without any corresponding group/outcome labels) for the integration of multi-omics datasets include techniques such as multiple canonical correlation analyses (MCCA),[96] multiple co-inertia analyses (MCIA),[97] multiple factor analyses (MFA),[98] and similarity network fusion (SNF).[99] MCCA is an extension of sparse canonical correlation analysis to more than two omics datasets by the addition of a LASSO-like L1 regularization penalty term. MCIA[97] relies on concatenations and projects different omics datasets into the same lower dimensional space. In this lower dimensional space, similar subjects are located close to each other. MFA[98] is similar to MCCA[96]; however, in this method the information content in each omic dataset is given equal weight. SNF[99] is a transformation-based approach that computes similarity graphs where vertices correspond to the samples. There are additional methods based on modern factor analysis such as multi-omics factor analysis (MOFA)[100] or MOFA+.[101] Some of these approaches, e.g., MOFA^{+}, are

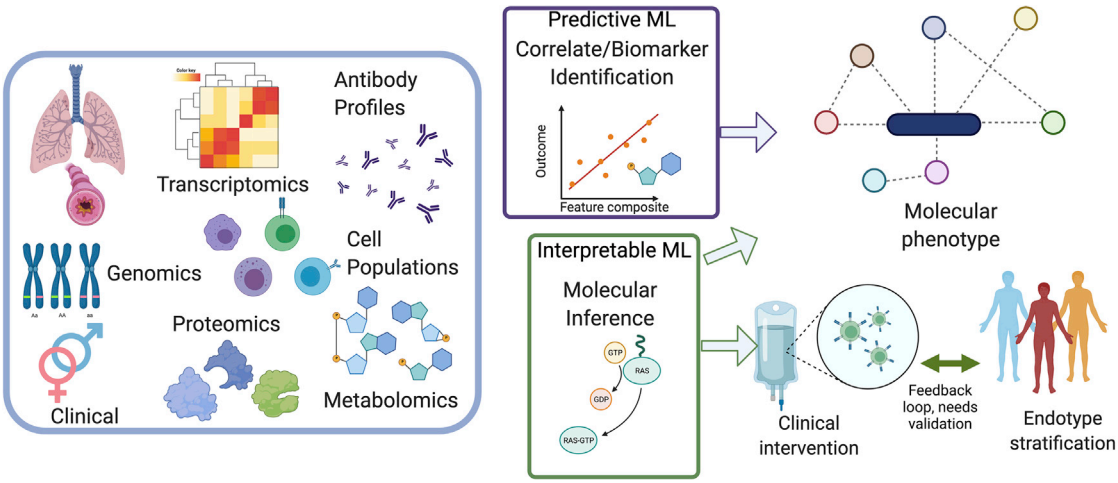

Figure 2. Predictive and interpretable machine learning approaches to integrate multi-omics datasets in asthma

designed primarily for multi-modal single-cell data;[101] i.e., they are based on structural and distributional assumptions germane to single-cell multi-omic datasets.[101] Unsupervised approaches detect broad structure in the data: some approaches may identify clusters that correspond to multiple molecular phenotypes, while others may be highly granular and split the same/similar molecular phenotypes into multiple clusters. It is important to keep in mind that these approaches are primarily meant for visualizing/understanding global data structure, and specific insights drawn from visualizations corresponding to overtuned hyperparameters may cause erroneous overinterpretations.

Beyond these unsupervised machine learning and dimensionality reduction/visualization approaches, one could also use supervised approaches (Figure 2). These align multi-omic datasets with specific outcomes labels/groups/endotypes of interest. They require pre-defined labels, but they are also able to home in on very specific signatures that track with the labels of interest. Importantly, the success of these approaches is inherently dependent on the quality of the training data, i.e., the availability of a high-quality truth set (reference labels/groups). Even the most appropriate model will not work with low-quality or poorly annotated data as machine learning approaches are inherently garbage in, garbage out. Within supervised approaches, there are two broad categories. The first category focuses on the identification of predictive biomarkers. Typically, multi-omic datasets are highly dimensional (i.e., the number of features exceed the number of samples). For these datasets, a key consideration of the analytical approaches is to avoid overfitting. An important consideration is the evaluation of these approaches with data held out using techniques like k-fold or leave-one-out cross-validation as well as the establishment of rigorous negative controls to benchmark model performance against (e.g., permutation testing[102]). We and others have successfully used these rigorous evaluation frameworks in a wide range of immune correlate analyses in infectious disease.[103–107] A range of predictive machine learning approaches including those that use regularized regression (e.g., LASSO,[108] L1 regularized regression, or Elastic Net, L1 + L2 regularized regression) or bootstrap aggregated decision trees (e.g., random forest[109]) are well adapted to avoid overfitting and identify predictive biomarkers. These approaches also include both linear (LASSO) and non-linear models (e.g., random forest). Other related approaches use classifying/regressing to outcomes of interest using dimensionality-reduced representations of data, e.g., principal components regression[110] or partial least squares (PLS) regression.[111] One example of such an approach is Immune Cell Linkage through Exploratory Matrices (ICLite) that allows deconvolution of bulk RNA-seq data derived from mixed cell populations.[38,61] ICLite constructs modules of functionally related genes and links them to specific lineages in mixed cell populations using a sparse PLS approach. These approaches are often tailored to specific disease contexts.

More recent machine learning approaches use embeddings to reduce high-dimensional feature spaces into efficient representations. Modern deep learning approaches such as convolution neural networks can be used on these. However, while such approaches are typically highly accurate in terms of prediction,[112,113] it is important to keep in mind that the signatures identified by these approaches may simply be correlative, i.e., not have anything to do with the underlying molecular mechanisms. Thus, insights from these approaches may not be directly useful for hypothesis generation and/or corresponding perturbation experiment design. However, they are still valuable as biomarkers in a clinical setting. Finally, while many of these approaches are based solely on available data, others leverage prior knowledge. The quality is often dependent on the nature and volume of available priors. For example, Data Integration Analysis for Biomarker Discovery using Latent Components (DIABLO)[114] is another recently developed multi-omics method that can simultaneously use various omics variables (e.g., transcriptome, proteome, metabolome) during the integration process to identify phenotypic groups.[114] It allows both modular-based analyses and cross-over study designs. Interestingly, in a study of peripheral blood after allergen challenges in patients with asthma, DIABLO[114] identified both known and novel multi-omics biomarkers consisting of mRNAs, miRNAs, epigenomic status, proteins, and metabolites. An important next step will

be to determine how to experimentally validate integrated findings such as these. These methods have been widely used for multi-omic integration across other clinical contexts and can be applied to multi-omics asthma datasets.

The second set of approaches moves beyond predictive biomarkers to inference of actual mechanisms underlying the group/outcome of interest. We and others have developed interpretable machine learning approaches for multi-omic integration that move beyond predictive biomarkers to actual causal signatures.[95] There has been extensive research in causal reasoning over the last 2 decades.[115] However, these approaches are suitable primarily for low-dimensional datasets due to the underlying computational complexity. Fortunately, this barrier has recently been overcome.[95] For example, we recently reported a novel latent factor regression framework, Essential Regression (ER) that integrates high-dimensional multi-omic datasets without any assumptions regarding underlying data-generating mechanisms.[95] ER clusters the observable features into latent factors with guarantees regarding identifiability, and then it identifies significant latent factors with putative cause-effect relationships to outcome. ER generates mechanistic hypotheses solely based on latent factors identified from multi-omic data without the incorporation of any prior knowledge. It is thus applicable to contexts where prior knowledge is weak or unavailable and is not limited by the nature and quality of available prior knowledge. Importantly, as ER does not make any assumptions about underlying data-generating mechanisms, it can also incorporate other data types (clinical, social determinants of health, etc.) beyond the standard multi-omic assays. Including these additional datasets in appropriate contexts can often significantly enhance the quality of the underlying inference.

Within approaches that provide inference beyond prediction, there are modern techniques that also incorporate prior knowledge. For example, Causal Oriented Search of Multi-Omics Space (COSMOS) integrates phosphoproteomics, transcriptomics, and metabolomics datasets, combining network-level causal reasoning to signaling networks.[116] COSMOS generates mechanistic hypotheses for experimental observations. We also developed another complementary approach that combines transcriptomic data with the modularity of the underlying protein interactome network to identify expression modules that underlie a clinical outcome of interest (rejection in the context of pediatric liver transplantation).[117] The same framework can be extrapolated to asthma multi-omic data. Overall, these inference frameworks without or with prior knowledge are complementary; while the former does not require any priors, the latter leverages higher-order structures in pathways or prior-knowledge biological networks. The former can be used for any multi-modal datasets, while the latter can be used for multi-omic datasets that directly measure or map to molecules that are represented in pathways (e.g., genes/proteins or metabolites). Further, while deep learning approaches have traditionally been hard to interpret, recent techniques have creatively used techniques like backpropagation to make these interpretable and amenable to inference beyond prediction.[118] These methods can integrate multi-omic datasets to uncover signatures of asthma molecular phenotypes that can be validated in hypothesis-driven experiments/corresponding perturbation systems to establish endotypes. Methods like ER that do not make assumptions regarding underlying data-generating mechanisms are especially well suited to integrate data across scales of organization and different tissues/cell types. And careful choice of an appropriate method keeping these considerations in mind leads to more reproducible and actionable science.

CONCLUDING REMARKS

To date, many of the systems approaches to study asthma produce interesting observations regarding potential pathway engagement, with less robust prediction of clinical features or treatment responses. Even with single compartment approaches, moving the list of pathways (genes/proteins) from hypothesis generation to development of clinically meaningful biomarkers and therapeutic targets remains challenging. As proposed in this article, machine learning approaches can be leveraged to identify the most actionable/highest priority pathways. The choice of approach (unsupervised vs. supervised, prediction vs. inference, etc.) depends both on the question and the nature of underlying data. For adequately powered cohorts with well-defined labels, supervised approaches that home in on putative causal factors are likely the most biologically relevant and clinically actionable. However, for an underpowered study, the same approach is unsuitable and can lead to overinterpetation and spurious inference. There, unsupervised approaches may be more appropriate. And incorporating complementary information and feedback (priors) is useful when there are reasonably strong context-specific priors. However, weak priors should be treated with caution as they can contradict strong inferences from a well-powered and well-designed study. Contradictions or irreproducibility can stem from a wide range of sources: from poor experimental design (e.g., data full of technical artifacts) to inappropriate analytical frameworks (e.g., failure to correct for batch effects) to low power. When comparing studies, it is important to keep these in mind and appropriately weigh for better designed and better powered studies.

There are also practical challenges to implementing AI/machine learning in independent studies in order to derive common actionable signatures that can be further tested in clinical trials. It is well known that a key challenge to implementing AI in clinical decision-making is the fairness of the underlying large clinical datasets.[119] These biases can lead to inherently flawed models. Similar challenges pertain to multi-omics integration of cellular and molecular datasets using machine learning methods. These multi-modal datasets, especially genomic and epigenomic data, can also suffer from similar sampling biases. Sampling biases can stem from a wide range of issues from the exclusion of under-represented minorities and under-resourced populations in cohort studies to non-uniform sampling across the socioeconomic spectrum. These sampling biases could translate to inherent discovery biases despite rigor of the analytical approaches. Further, running clinical trials requires enormous resources including long-term doctor-patient relationships, and the coupled machine learning approaches are often not matched in terms of power considerations. Collaborations across groups and cohorts to expand the datasets available for training and replication are critical to improve the quality of insights available

from these systems approaches. Further, if funding agencies enforce FAIR (findability, accessibility, interoperability, and reusability) principles for reusability of data, this would dramatically improve availability of data across cohorts for independent replication.[120] Overall, cross-talk between asthma researchers, clinicians, and systems immunologists would enable critical review of data from previous mouse models, GWAS studies, and clinical trials with the goal to identify the most critical factors in those pathways, and those most amenable to biomarker development or molecular targeting. Appropriate context-specific use of predictive vs. interpretable machine learning approaches will help distinguish molecular phenotypes and actual endotypes that are validated in clinical trials.

ACKNOWLEDGMENTS

This work was supported by NIH grants P01 AI106684 (A.R. and S.E.W.), R01 HL148604 (A.R.), R01 HL153058 (S.E.W.), DP2 AI164325, and U01 HG012041 (J.D.). We thank Mondira Ray for helpful discussions.

AUTHOR CONTRIBUTIONS

All authors contributed to conceptualization and manuscript preparation.

DECLARATION OF INTERESTS

A.R. has a research agreement with Pieris Pharmaceuticals. J.D. is a consultant for Seromyx Systems. S.E.W. is a consultant for AstraZeneca, Glaxo Smith-Kline, and Sanofi. She is also involved in clinical trials being run by Knopp, Sanofi, and AstraZeneca. She has a research agreement with Pieris Pharmaceuticals.

REFERENCES

1. Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. Genome Biol. *18*, 83. https://doi.org/10.1186/s13059-017-1215-1.
2. Stuart, T., and Satija, R. (2019). Integrative single-cell analysis. Nat. Rev. Genet. *20*, 257–272. https://doi.org/10.1038/s41576-019-0093-7.
3. Li, C.X., Wheelock, C.E., Sköld, C.M., and Wheelock, Å.M. (2018). Integration of multi-omics datasets enables molecular classification of COPD. Eur. Respir. J. *51*, 1701930. https://doi.org/10.1183/13993003.01930-2017.
4. Mao, K., Zhao, Y., Ding, B., Feng, P., Li, Z., Zhou, Y.L., and Xue, Q. (2022). Integrative analysis of multi-omics data-identified key genes with KLRC3 as the core in a gene regulatory network related to immune phenotypes in lung adenocarcinoma. Front. Genet. *13*, 810193. https://doi.org/10.3389/fgene.2022.810193.
5. Yan, Z., Chen, B., Yang, Y., Yi, X., Wei, M., Ecklu-Mensah, G., Buschmann, M.M., Liu, H., Gao, J., Liang, W., et al. (2022). Multi-omics analyses of airway host-microbe interactions in chronic obstructive pulmonary disease identify potential therapeutic interventions. Nat. Microbiol. *7*, 1361–1375. https://doi.org/10.1038/s41564-022-01196-8.
6. Kersting, K. (2018). Machine learning and artificial intelligence: two fellow travelers on the quest for intelligent behavior in machines. Front. Big Data *1*, 6. https://doi.org/10.3389/fdata.2018.00006.
7. Collins, G.S., and Moons, K.G.M. (2019). Reporting of artificial intelligence prediction models. Lancet *393*, 1577–1579. https://doi.org/10.1016/S0140-6736(19)30037-6.
8. Loftus, T.J., Tighe, P.J., Filiberto, A.C., Efron, P.A., Brakenridge, S.C., Mohr, A.M., Rashidi, P., Upchurch, G.R., Jr., and Bihorac, A. (2020). Artificial intelligence and surgical decision-making. JAMA Surg. *155*, 148–158. https://doi.org/10.1001/jamasurg.2019.4917.
9. Bender, A., and Cortes-Ciriano, I. (2021). Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data. Drug Discov. Today *26*, 1040–1052. https://doi.org/10.1016/j.drudis.2020.11.037.
10. Mann, M., Kumar, C., Zeng, W.F., and Strauss, M.T. (2021). Artificial intelligence for proteomics and biomarker discovery. Cell Syst. *12*, 759–770. https://doi.org/10.1016/j.cels.2021.06.006.
11. Zhu, Z., Ng, D.W.H., Park, H.S., and McAlpine, M.C. (2021). 3D-printed multifunctional materials enabled by artificial-intelligenceassisted fabrication technologies. Nat. Rev. Mater. *6*, 27–47.
12. Matschinske, J., Alcaraz, N., Benis, A., Golebiewski, M., Grimm, D.G., Heumos, L., Kacprowski, T., Lazareva, O., List, M., Louadi, Z., et al. (2021). The AIMe registry for artificial intelligence in biomedical research. Nat. Methods *18*, 1128–1131. https://doi.org/10.1038/s41592-021-01241-0.
13. Wenzel, S.E. (2012). Asthma phenotypes: the evolution from clinical to molecular approaches. Nat. Med. *18*, 716–725. https://doi.org/10.1038/nm.2678.
14. Rackemann, F.M. (1947). A working classification of asthma. Am. J. Med. *3*, 601–606. https://doi.org/10.1016/0002-9343(47)90204-0.
15. Samter, M., and Beers, R.F., Jr. (1967). Concerning the nature of intolerance to aspirin. J. Allergy *40*, 281–293. https://doi.org/10.1016/0021-8707(67)90076-7.
16. Haldar, P., Brightling, C.E., Hargadon, B., Gupta, S., Monteiro, W., Sousa, A., Marshall, R.P., Bradding, P., Green, R.H., Wardlaw, A.J., and Pavord, I.D. (2009). Mepolizumab and exacerbations of refractory eosinophilic asthma. N. Engl. J. Med. *360*, 973–984. https://doi.org/10.1056/NEJMoa0808991.
17. Pavord, I.D., Brightling, C.E., Woltmann, G., and Wardlaw, A.J. (1999). Non-eosinophilic corticosteroid unresponsive asthma. Lancet *353*, 2213–2214. https://doi.org/10.1016/S0140-6736(99)01813-9.
18. Wenzel, S.E., Schwartz, L.B., Langmack, E.L., Halliday, J.L., Trudeau, J.B., Gibbs, R.L., and Chu, H.W. (1999). Evidence that severe asthma can be divided pathologically into two inflammatory subtypes with distinct physiologic and clinical characteristics. Am. J. Respir. Crit. Care Med. *160*, 1001–1008. https://doi.org/10.1164/ajrccm.160.3.9812110.
19. Pavord, I.D., Korn, S., Howarth, P., Bleecker, E.R., Buhl, R., Keene, O.N., Ortega, H., and Chanez, P. (2012). Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. Lancet *380*, 651–659. https://doi.org/10.1016/S0140-6736(12)60988-X.
20. Wenzel, S.E. (2016). Emergence of biomolecular pathways to define novel asthma phenotypes. Type-2 immunity and beyond. Am. J. Respir. Cell Mol. Biol. *55*, 1–4. https://doi.org/10.1165/rcmb.2016-0141PS.
21. Wenzel, S.E. (2021). Severe adult asthmas: integrating clinical features, biology, and therapeutics to improve outcomes. Am. J. Respir. Crit. Care Med. *203*, 809–821. https://doi.org/10.1164/rccm.202009-3631CI.
22. Bleecker, E.R., Wechsler, M.E., FitzGerald, J.M., Menzies-Gow, A., Wu, Y., Hirsch, I., Goldman, M., Newbold, P., and Zangrilli, J.G. (2018). Baseline patient factors impact on the clinical efficacy of benralizumab for severe asthma. Eur. Respir. J. *52*, 1800936. https://doi.org/10.1183/13993003.00936-2018.
23. Schett, G., McInnes, I.B., and Neurath, M.F. (2021). Reframing immune-mediated inflammatory diseases through signature cytokine hubs. N. Engl. J. Med. *385*, 628–639. https://doi.org/10.1056/NEJMra1909094.
24. Federico, M.J., McFarlane, A.E., 2nd, Szefler, S.J., and Abrams, E.M. (2020). The impact of social determinants of health on children with asthma. J. Allergy Clin. Immunol. Pract. *8*, 1808–1814. https://doi.org/10.1016/j.jaip.2020.03.028.
25. Grant, T., Croce, E., and Matsui, E.C. (2022). Asthma and the social determinants of health. Ann. Allergy Asthma Immunol. *128*, 5–11. https://doi.org/10.1016/j.anai.2021.10.002.

26. Trivedi, A.P., Hall, C., Goss, C.W., Lew, D., Krings, J.G., McGregor, M.C., Samant, M., Sieren, J.P., Li, H., Schechtman, K.B., et al. (2022). Quantitative CT characteristics of cluster phenotypes in the severe asthma research program cohorts. Radiology *304*, 450–459. https://doi.org/10.1148/radiol.210363.

27. Gaffin, J.M., Castro, M., Bacharier, L.B., and Fuhlbrigge, A.L. (2022). The role of comorbidities in difficult-to-control asthma in adults and children. J. Allergy Clin. Immunol. Pract. *10*, 397–408. https://doi.org/10.1016/j.jaip.2021.11.016.

28. Juhn, Y.J. (2014). Risks for infection in patients with asthma (or other atopic conditions): is asthma more than a chronic airway disease? J. Allergy Clin. Immunol. *134*, 247–257, quiz 258-249. https://doi.org/10.1016/j.jaci.2014.04.024.

29. Gautam, Y., Johansson, E., and Mersha, T.B. (2022). Multi-omics profiling approach to asthma: an evolving paradigm. J. Pers. Med. *12*, 66. https://doi.org/10.3390/jpm12010066.

30. Tyler, S.R., and Bunyavanich, S. (2019). Leveraging -omics for asthma endotyping. J. Allergy Clin. Immunol. *144*, 13–23. https://doi.org/10.1016/j.jaci.2019.05.015.

31. Picard, M., Scott-Boyer, M.P., Bodein, A., Périn, O., and Droit, A. (2021). Integration strategies of multi-omics data for machine learning analysis. Comput. Struct. Biotechnol. J. *19*, 3735–3746. https://doi.org/10.1016/j.csbj.2021.06.030.

32. Moffatt, M.F., Gut, I.G., Demenais, F., Strachan, D.P., Bouzigon, E., Heath, S., von Mutius, E., Farrall, M., Lathrop, M., and Cookson, W.O.C.M.; GABRIEL Consortium (2010). A large-scale, consortium-based genomewide association study of asthma. N. Engl. J. Med. *363*, 1211–1221. https://doi.org/10.1056/NEJMoa0906312.

33. Halapi, E., Gudbjartsson, D.F., Jonsdottir, G.M., Bjornsdottir, U.S., Thorleifsson, G., Helgadottir, H., Williams, C., Koppelman, G.H., Heinzmann, A., Boezen, H.M., et al. (2010). A sequence variant on 17q21 is associated with age at onset and severity of asthma. Eur. J. Hum. Genet. *18*, 902–908. https://doi.org/10.1038/ejhg.2010.38.

34. Ober, C. (2016). Asthma genetics in the post-GWAS era. Ann. Am. Thorac. Soc. *13* (*Suppl 1*), S85–S90. https://doi.org/10.1513/AnnalsATS.201507-459MG.

35. Bush, W.S., Oetjens, M.T., and Crawford, D.C. (2016). Unravelling the human genome-phenome relationship using phenome-wide association studies. Nat. Rev. Genet. *17*, 129–145. https://doi.org/10.1038/nrg.2015.36.

36. Portelli, M.A., Dijk, F.N., Ketelaar, M.E., Shrine, N., Hankinson, J., Bhaker, S., Grotenboer, N.S., Obeidat, M., Henry, A.P., Billington, C.K., et al. (2020). Phenotypic and functional translation of IL1RL1 locus polymorphisms in lung tissue and asthmatic airway epithelium. JCI Insight *5*, 132446. https://doi.org/10.1172/jci.insight.132446.

37. Vieira Braga, F.A., Kar, G., Berg, M., Carpaij, O.A., Polanski, K., Simon, L.M., Brouwer, S., Gomes, T., Hesse, L., Jiang, J., et al. (2019). A cellular census of human lungs identifies novel cell states in health and in asthma. Nat. Med. *25*, 1153–1163. https://doi.org/10.1038/s41591-019-0468-5.

38. Camiolo, M.J., Zhou, X., Oriss, T.B., Yan, Q., Gorry, M., Horne, W., Trudeau, J.B., Scholl, K., Chen, W., Kolls, J.K., et al. (2021). High-dimensional profiling clusters asthma severity by lymphoid and non-lymphoid status. Cell Rep. *35*, 108974. https://doi.org/10.1016/j.celrep.2021.108974.

39. Weathington, N., O'Brien, M.E., Radder, J., Whisenant, T.C., Bleecker, E.R., Busse, W.W., Erzurum, S.C., Gaston, B., Hastie, A.T., Jarjour, N.N., et al. (2019). Bronchoalveolar lavage cell gene expression in severe asthma reveals mechanisms of severe disease and influences of medications. Am. J. Respir. Crit. Care Med. *200*, 837–856. https://doi.org/10.1164/rccm.201811-2221OC.

40. Ray, A., Camiolo, M., Fitzpatrick, A., Gauthier, M., and Wenzel, S.E. (2020). Are we meeting the promise of endotypes and precision medicine in asthma? Physiol. Rev. *100*, 983–1017. https://doi.org/10.1152/physrev.00023.2019.

41. Hekking, P.P., Loza, M.J., Pavlidis, S., De Meulder, B., Lefaudeux, D., Baribaud, F., Auffray, C., Wagener, A.H., Brinkman, P., Lutter, R., et al. (2017). Transcriptomic gene signatures associated with persistent airflow limitation in patients with severe asthma. Eur. Respir. J. *50*, 1602298. https://doi.org/10.1183/13993003.02298-2016.

42. Hekking, P.P., Loza, M.J., Pavlidis, S., de Meulder, B., Lefaudeux, D., Baribaud, F., Auffray, C., Wagener, A.H., Brinkman, P., Ir, Lutter, R., Ir., et al. (2018). Pathway discovery using transcriptomic profiles in adult-onset severe asthma. J. Allergy Clin. Immunol. *141*, 1280–1290. https://doi.org/10.1016/j.jaci.2017.06.037.

43. Kuo, C.H.S., Pavlidis, S., Loza, M., Baribaud, F., Rowe, A., Pandis, I., Sousa, A., Corfield, J., Djukanovic, R., Lutter, R., et al. (2017). T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. Eur. Respir. J. *49*, 1602135. https://doi.org/10.1183/13993003.02135-2016.

44. Baines, K.J., Simpson, J.L., Wood, L.G., Scott, R.J., and Gibson, P.G. (2011). Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. J. Allergy Clin. Immunol. *127*, 153–160. https://doi.org/10.1016/j.jaci.2010.10.024.

45. Baines, K.J., Simpson, J.L., Wood, L.G., Scott, R.J., Fibbens, N.L., Powell, H., Cowan, D.C., Taylor, D.R., Cowan, J.O., and Gibson, P.G. (2014). Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. J. Allergy Clin. Immunol. *133*, 997–1007. https://doi.org/10.1016/j.jaci.2013.12.1091.

46. Modena, B.D., and Wenzel, S.E. (2017). Consistency of T2 gene signatures in severe asthma. Key to effective treatments or merely the tip of the iceberg? Am. J. Respir. Crit. Care Med. *195*, 411–412. https://doi.org/10.1164/rccm.201609-1854ED.

47. Peters, M.C., Ringel, L., Dyjack, N., Herrin, R., Woodruff, P.G., Rios, C., O'Connor, B., Fahy, J.V., and Seibold, M.A. (2019). A transcriptomic method to determine airway immune dysfunction in T2-high and T2-low asthma. Am. J. Respir. Crit. Care Med. *199*, 465–477. https://doi.org/10.1164/rccm.201807-1291OC.

48. Woodruff, P.G., Boushey, H.A., Dolganov, G.M., Barker, C.S., Yang, Y.H., Donnelly, S., Ellwanger, A., Sidhu, S.S., Dao-Pick, T.P., Pantoja, C., et al. (2007). Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. Proc. Natl. Acad. Sci. USA *104*, 15858–15863. https://doi.org/10.1073/pnas.0707413104.

49. Rossios, C., Pavlidis, S., Hoda, U., Kuo, C.H., Wiegman, C., Russell, K., Sun, K., Loza, M.J., Baribaud, F., Durham, A.L., et al. (2018). Sputum transcriptomics reveal upregulation of IL-1 receptor family members in patients with severe asthma. J. Allergy Clin. Immunol. *141*, 560–570. https://doi.org/10.1016/j.jaci.2017.02.045.

50. Robinson, D.S., Hamid, Q., Ying, S., Tsicopoulos, A., Barkans, J., Bentley, A.M., Corrigan, C., Durham, S.R., and Kay, A.B. (1992). Predominant TH2-like bronchoalveolar T-lymphocyte population in atopic asthma. N. Engl. J. Med. *326*, 298–304. https://doi.org/10.1056/NEJM199201303260504.

51. Lambrecht, B.N., Hammad, H., and Fahy, J.V. (2019). The cytokines of asthma. Immunity *50*, 975–991. https://doi.org/10.1016/j.immuni.2019.03.018.

52. Woodruff, P.G., Modrek, B., Choy, D.F., Jia, G., Abbas, A.R., Ellwanger, A., Koth, L.L., Arron, J.R., and Fahy, J.V. (2009). T-helper type 2-driven inflammation defines major subphenotypes of asthma. Am. J. Respir. Crit. Care Med. *180*, 388–395. https://doi.org/10.1164/rccm.200903-0392OC.

53. Camiolo, M.J., Kale, S.L., Oriss, T.B., Gauthier, M., and Ray, A. (2021). Immune responses and exacerbations in severe asthma. Curr. Opin. Immunol. *72*, 34–42. https://doi.org/10.1016/j.coi.2021.03.004.

54. Brusselle, G., Germinaro, M., Weiss, S., and Zangrilli, J. (2017). Reslizumab in patients with inadequately controlled late-onset asthma and elevated blood eosinophils. Pulm. Pharmacol. Ther. *43*, 39–45. https://doi.org/10.1016/j.pupt.2017.01.011.

55. Modena, B.D., Tedrow, J.R., Milosevic, J., Bleecker, E.R., Meyers, D.A., Wu, W., Bar-Joseph, Z., Erzurum, S.C., Gaston, B.M., Busse, W.W., et al. (2014). Gene expression in relation to exhaled nitric oxide identifies novel asthma phenotypes with unique biomolecular pathways. Am. J. Respir. Crit. Care Med. *190*, 1363–1372. https://doi.org/10.1164/rccm.201406-1099OC.

56. Jevnikar, Z., Östling, J., Ax, E., Calvén, J., Thörn, K., Israelsson, E., Öberg, L., Singhania, A., Lau, L.C.K., Wilson, S.J., et al. (2019). Epithelial IL-6 trans-signaling defines a new asthma phenotype with increased airway inflammation. J. Allergy Clin. Immunol. *143*, 577–590. https://doi.org/10.1016/j.jaci.2018.05.026.

57. Voraphani, N., Gladwin, M.T., Contreras, A.U., Kaminski, N., Tedrow, J.R., Milosevic, J., Bleecker, E.R., Meyers, D.A., Ray, A., Ray, P., et al. (2014). An airway epithelial iNOS-DUOX2-thyroid peroxidase metabolome drives Th1/Th2 nitrative stress in human severe asthma. Mucosal Immunol. *7*, 1175–1185. https://doi.org/10.1038/mi.2014.6.

58. Spahn, J.D., Malka, J., and Szefler, S.J. (2016). Current application of exhaled nitric oxide in clinical practice. J. Allergy Clin. Immunol. *138*, 1296–1298. https://doi.org/10.1016/j.jaci.2016.09.002.

59. Lacy, P., Mahmudi-Azer, S., Bablitz, B., Hagen, S.C., Velazquez, J.R., Man, S.F., and Moqbel, R. (1999). Rapid mobilization of intracellularly stored RANTES in response to interferon-gamma in human eosinophils. Blood *94*, 23–32.

60. Ochiai, K., Tanabe, E., Ishihara, C., Kagami, M., Sugiyama, T., Sueishi, M., Koya, N., and Tomioka, H. (1999). Role of JAK2 signal transductional pathway in activation and survival of human peripheral eosinophils by interferon-gamma (IFN-gamma). Clin. Exp. Immunol. *118*, 340–343. https://doi.org/10.1046/j.1365-2249.1999.01068.x.

61. Camiolo, M.J., Wenzel, S.E., and Ray, A. (2021). Using ICLite for deconvolution of bulk transcriptional data from mixed cell populations. STAR Protoc. *2*, 100847. https://doi.org/10.1016/j.xpro.2021.100847.

62. Aliee, H., and Theis, F.J. (2021). AutoGeneS: automatic gene selection using multi-objective optimization for RNA-seq deconvolution. Cell Syst. *12*, 706–715.e4, e704. https://doi.org/10.1016/j.cels.2021.05.006.

63. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457. https://doi.org/10.1038/nmeth.3337.

64. FitzGerald, J.M., and Gibson, P.G. (2006). Asthma exacerbations . 4: prevention. Thorax *61*, 992–999. https://doi.org/10.1136/thx.2005.045195.

65. Altman, M.C., Gill, M.A., Whalen, E., Babineau, D.C., Shao, B., Liu, A.H., Jepson, B., Gruchalla, R.S., O'Connor, G.T., Pongracic, J.A., et al. (2019). Transcriptome networks identify mechanisms of viral and nonviral asthma exacerbations in children. Nat. Immunol. *20*, 637–651. https://doi.org/10.1038/s41590-019-0347-8.

66. Poole, A., Urbanek, C., Eng, C., Schageman, J., Jacobson, S., O'Connor, B.P., Galanter, J.M., Gignoux, C.R., Roth, L.A., Kumar, R., et al. (2014). Dissecting childhood asthma with nasal transcriptomics distinguishes subphenotypes of disease. J. Allergy Clin. Immunol. *133*, 670–678.e12. https://doi.org/10.1016/j.jaci.2013.11.025.

67. Hoda, U., Pavlidis, S., Bansal, A.T., Takahashi, K., Hu, S., Ng Kee Kwong, F., Rossios, C., Sun, K., Bhavsar, P., Loza, M., et al. (2022). Clinical and transcriptomic features of persistent exacerbation-prone severe asthma in U-BIOPRED cohort. Clin. Transl. Med. *12*, e816. https://doi.org/10.1002/ctm2.816.

68. Li, H., Wang, H., Sokulsky, L., Liu, S., Yang, R., Liu, X., Zhou, L., Li, J., Huang, C., Li, F., et al. (2021). Single-cell transcriptomic analysis reveals key immune cell phenotypes in the lungs of patients with asthma exacerbation. J. Allergy Clin. Immunol. *147*, 941–954. https://doi.org/10.1016/j.jaci.2020.09.032.

69. Pandey, G., Pandey, O.P., Rogers, A.J., Ahsen, M.E., Hoffman, G.E., Raby, B.A., Weiss, S.T., Schadt, E.E., and Bunyavanich, S. (2018). A nasal brush-based classifier of asthma identified by machine learning analysis of nasal RNA sequence data. Sci. Rep. *8*, 8826. https://doi.org/10.1038/s41598-018-27189-4.

70. Yang, I.V., Pedersen, B.S., Liu, A.H., O'Connor, G.T., Pillai, D., Kattan, M., Misiak, R.T., Gruchalla, R., Szefler, S.J., Khurana Hershey, G.K., et al. (2017). The nasal methylome and childhood atopic asthma. J. Allergy Clin. Immunol. *139*, 1478–1488. https://doi.org/10.1016/j.jaci.2016.07.036.

71. Nicodemus-Johnson, J., Myers, R.A., Sakabe, N.J., Sobreira, D.R., Hogarth, D.K., Naureckas, E.T., Sperling, A.I., Solway, J., White, S.R., Nobrega, M.A., et al. (2016). DNA methylation in lung cells is associated with asthma endotypes and genetic risk. JCI Insight *1*, e90151. https://doi.org/10.1172/jci.insight.90151.

72. Ibrahim, B., Basanta, M., Cadden, P., Singh, D., Douce, D., Woodcock, A., and Fowler, S.J. (2011). Non-invasive phenotyping using exhaled volatile organic compounds in asthma. Thorax *66*, 804–809. https://doi.org/10.1136/thx.2010.156695.

73. Sinha, A., Desiraju, K., Aggarwal, K., Kutum, R., Roy, S., Lodha, R., Kabra, S.K., Ghosh, B., Sethi, T., and Agrawal, A. (2017). Exhaled breath condensate metabolome clusters for endotype discovery in asthma. J. Transl. Med. *15*, 262. https://doi.org/10.1186/s12967-017-1365-7.

74. Brinkman, P., Wagener, A.H., Hekking, P.P., Bansal, A.T., Maitland-van der Zee, A.H., Wang, Y., Weda, H., Knobel, H.H., Vink, T.J., Rattray, N.J., et al. (2019). Identification and prospective stability of electronic nose (eNose)-derived inflammatory phenotypes in patients with severe asthma. J. Allergy Clin. Immunol. *143*, 1811–1820.e7. https://doi.org/10.1016/j.jaci.2018.10.058.

75. Brasier, A.R., Victor, S., Boetticher, G., Ju, H., Lee, C., Bleecker, E.R., Castro, M., Busse, W.W., and Calhoun, W.J. (2008). Molecular phenotyping of severe asthma using pattern recognition of bronchoalveolar lavage-derived cytokines. J. Allergy Clin. Immunol. *121*, 30–37.e6. https://doi.org/10.1016/j.jaci.2007.10.015.

76. Hastie, A.T., Steele, C., Dunaway, C.W., Moore, W.C., Rector, B.M., Ampleford, E., Li, H., Denlinger, L.C., Jarjour, N., Meyers, D.A., et al. (2018). Complex association patterns for inflammatory mediators in induced sputum from subjects with asthma. Clin. Exp. Allergy *48*, 787–797. https://doi.org/10.1111/cea.13129.

77. Schofield, J.P.R., Burg, D., Nicholas, B., Strazzeri, F., Brandsma, J., Staykova, D., Folisi, C., Bansal, A.T., Xian, Y., Guo, Y., et al. (2019). Stratification of asthma phenotypes by airway proteomic signatures. J. Allergy Clin. Immunol. *144*, 70–82. https://doi.org/10.1016/j.jaci.2019.03.013.

78. Nieto-Fontarigo, J.J., González-Barcala, F.J., Andrade-Bulos, L.J., San-José, M.E., Cruz, M.J., Valdés-Cuadrado, L., Crujeiras, R.M., Arias, P., and Salgado, F.J. (2020). iTRAQ-based proteomic analysis reveals potential serum biomarkers of allergic and nonallergic asthma. Allergy *75*, 3171–3183. https://doi.org/10.1111/all.14406.

79. Brusselle, G.G., and Koppelman, G.H. (2022). Biologic therapies for severe asthma. N. Engl. J. Med. *386*, 157–171. https://doi.org/10.1056/NEJMra2032506.

80. Yamane, M., Ohnishi, H., Tsuji, K., Anabuki, K., and Yokoyama, A. (2022). Dupilumab-induced peripheral neuropathy in a patient with severe asthma. Ann. Allergy Asthma Immunol. *128*, 611. https://doi.org/10.1016/j.anai.2022.01.028.

81. Wechsler, M.E., Ruddy, M.K., Pavord, I.D., Israel, E., Rabe, K.F., Ford, L.B., Maspero, J.F., Abdulai, R.M., Hu, C.C., Martincova, R., et al. (2021). Efficacy and Safety of itepekimab in patients with moderate-to-severe asthma. N. Engl. J. Med. *385*, 1656–1668. https://doi.org/10.1056/NEJMoa2024257.

82. Hu, X., and Ivashkiv, L.B. (2009). Cross-regulation of signaling pathways by interferon-gamma: implications for immune responses and autoimmune diseases. Immunity *31*, 539–550. https://doi.org/10.1016/j.immuni.2009.09.002.

83. Brøgger, P., Blom, L.H., Simonsen, S., Thyssen, J.P., and Skov, L. (2020). Antagonism of the interleukin 4 receptor alpha promotes TH 1-signalling

among T cells from patients with atopic dermatitis after stimulation. Scand. J. Immunol. *91*, e12835. https://doi.org/10.1111/sji.12835.

84. Duvall, M.G., Barnig, C., Cernadas, M., Ricklefs, I., Krishnamoorthy, N., Grossman, N.L., Bhakta, N.R., Fahy, J.V., Bleecker, E.R., Castro, M., et al. (2017). Natural killer cell-mediated inflammation resolution is disabled in severe asthma. Sci. Immunol. *2*, eaam5446. https://doi.org/10.1126/sciimmunol.aam5446.

85. Gauthier, M., Chakraborty, K., Oriss, T.B., Raundhal, M., Das, S., Chen, J., Huff, R., Sinha, A., Fajt, M., Ray, P., et al. (2017). Severe asthma in humans and mouse model suggests a CXCL10 signature underlies corticosteroid-resistant Th1 bias. JCI Insight *2*, 94580. https://doi.org/10.1172/jci.insight.94580.

86. Gauthier, M., Kale, S.L., Oriss, T.B., Scholl, K., Das, S., Yuan, H., Hu, S., Chen, J., Camiolo, M., Ray, P., et al. (2022). Dual role for CXCR3 and CCR5 in asthmatic type 1 inflammation. J. Allergy Clin. Immunol. *149*, 113–124.e7. https://doi.org/10.1016/j.jaci.2021.05.044.

87. Muehling, L.M., Heymann, P.W., Wright, P.W., Eccles, J.D., Agrawal, R., Carper, H.T., Murphy, D.D., Workman, L.J., Word, C.R., Ratcliffe, S.J., et al. (2020). Human TH1 and TH2 cells targeting rhinovirus and allergen coordinately promote allergic asthma. J. Allergy Clin. Immunol. *146*, 555–570. https://doi.org/10.1016/j.jaci.2020.03.037.

88. Oriss, T.B., Raundhal, M., Morse, C., Huff, R.E., Das, S., Hannum, R., Gauthier, M.C., Scholl, K.L., Chakraborty, K., Nouraie, S.M., et al. (2017). IRF5 distinguishes severe asthma in humans and drives Th1 phenotype and airway hyperreactivity in mice. JCI Insight *2*, 91019. https://doi.org/10.1172/jci.insight.91019.

89. Raundhal, M., Morse, C., Khare, A., Oriss, T.B., Milosevic, J., Trudeau, J., Huff, R., Pilewski, J., Holguin, F., Kolls, J., et al. (2015). High IFN-gamma and low SLPI mark severe asthma in mice and humans. J. Clin. Invest. *125*, 3037–3050. https://doi.org/10.1172/jci80911.

90. Wisniewski, J.A., Muehling, L.M., Eccles, J.D., Capaldo, B.J., Agrawal, R., Shirley, D.A., Patrie, J.T., Workman, L.J., Schuyler, A.J., Lawrence, M.G., et al. (2018). TH1 signatures are present in the lower airways of children with severe asthma, regardless of allergic status. J. Allergy Clin. Immunol. *141*, 2048–2060.e13. https://doi.org/10.1016/j.jaci.2017.08.020.

91. Wu, W., Bang, S., Bleecker, E.R., Castro, M., Denlinger, L., Erzurum, S.C., Fahy, J.V., Fitzpatrick, A.M., Gaston, B.M., Hastie, A.T., et al. (2019). Multiview cluster Analysis identifies variable corticosteroid response phenotypes in severe asthma. Am. J. Respir. Crit. Care Med. *199*, 1358–1367. https://doi.org/10.1164/rccm.201808-1543OC.

92. Wu, W., Bleecker, E., Moore, W., Busse, W.W., Castro, M., Chung, K.F., Calhoun, W.J., Erzurum, S., Gaston, B., Israel, E., et al. (2014). Unsupervised phenotyping of Severe Asthma Research Program participants using expanded lung data. J. Allergy Clin. Immunol. *133*, 1280–1288. https://doi.org/10.1016/j.jaci.2013.11.042.

93. Galant, S.P., Komarow, H.D., Shin, H.W., Siddiqui, S., and Lipworth, B.J. (2017). The case for impulse oscillometry in the management of asthma in children and adults. Ann. Allergy Asthma Immunol. *118*, 664–671. https://doi.org/10.1016/j.anai.2017.04.009.

94. Rappoport, N., and Shamir, R. (2018). Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Res. *46*, 10546–10562. https://doi.org/10.1093/nar/gky889.

95. Bing, X., Lovelace, T., Bunea, F., Wegkamp, M., Kasturi, S.P., Singh, H., Benos, P.V., and Das, J. (2022). Essential Regression: a generalizable framework for inferring causal latent factors from multi-omic datasets. Patterns (N Y) *3*, 100473. https://doi.org/10.1016/j.patter.2022.100473.

96. Witten, D.M., and Tibshirani, R.J. (2009). Extensions of sparse canonical correlation analysis with applications to genomic data. Stat. Appl. Genet. Mol. Biol. *8*, Article28. https://doi.org/10.2202/1544-6115.1470.

97. Meng, C., Kuster, B., Culhane, A.C., and Gholami, A.M. (2014). A multivariate approach to the integration of multi-omics datasets. BMC Bioinf. *15*, 162. https://doi.org/10.1186/1471-2105-15-162.

98. de Tayrac, M., Le, S., Aubry, M., Mosser, J., and Husson, F. (2009). Simultaneous analysis of distinct Omics data sets with integration of biological knowledge: multiple Factor Analysis approach. BMC Genom. *10*, 32. https://doi.org/10.1186/1471-2164-10-32.

99. Wang, B., Mezlini, A.M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., and Goldenberg, A. (2014). Similarity network fusion for aggregating data types on a genomic scale. Nat. Methods *11*, 333–337. https://doi.org/10.1038/nmeth.2810.

100. Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J.C., Buettner, F., Huber, W., and Stegle, O. (2018). Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol. Syst. Biol. *14*, e8124. https://doi.org/10.15252/msb.20178124.

101. Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J.C., and Stegle, O. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol. *21*, 111. https://doi.org/10.1186/s13059-020-02015-1.

102. Ojala, M., and Garriga, G.C. (2010). Permuattion tests for studying classifier performance. J. Mach. Lear.Res. *11*, 1833–1863.

103. Suscovich, T.J., Fallon, J.K., Das, J., Demas, A.R., Crain, J., Linde, C.H., Michell, A., Natarajan, H., Arevalo, C., Broge, T., et al. (2020). Mapping functional humoral correlates of protection against malaria challenge following RTS, S/AS01 vaccination. Sci. Transl. Med. *12*, eabb4757. https://doi.org/10.1126/scitranslmed.abb4757.

104. Das, J., Devadhasan, A., Linde, C., Broge, T., Sassic, J., Mangano, M., O'Keefe, S., Suscovich, T., Streeck, H., Irrinki, A., et al. (2020). Mining for humoral correlates of HIV control and latent reservoir size. PLoS Pathog. *16*, e1008868. https://doi.org/10.1371/journal.ppat.1008868.

105. Lu, L.L., Das, J., Grace, P.S., Fortune, S.M., Restrepo, B.I., and Alter, G. (2020). Antibody fc glycosylation discriminates between latent and active tuberculosis. J. Infect. Dis. *222*, 2093–2102. https://doi.org/10.1093/infdis/jiz643.

106. Ackerman, M.E., Das, J., Pittala, S., Broge, T., Linde, C., Suscovich, T.J., Brown, E.P., Bradley, T., Natarajan, H., Lin, S., et al. (2018). Route of immunization defines multiple mechanisms of vaccine-mediated protection against SIV. Nat. Med. *24*, 1590–1598. https://doi.org/10.1038/s41591-018-0161-0.

107. Das, J., Fallon, J.K., Yu, T.C., Michell, A., Suscovich, T.J., Linde, C., Natarajan, H., Weiner, J., Coccia, M., Gregory, S., et al. (2021). Delayed fractional dosing with RTS, S/AS01 improves humoral immunity to malaria via a balance of polyfunctional NANP6- and Pf16-specific antibodies. Med *2*, 1269–1286.e9. https://doi.org/10.1016/j.medj.2021.10.003.

108. Tibshirani, R. (1996). Regression Shrinkage and selection via the lasso. J. Roy. Stat. Soc. B *58*, 267–288.

109. Breiman, L. (2001). Random forests. Mach. Learn. *45*, 5–32.

110. Bair, E., Hastie, T., Paul, D., and Tibshirani, R. (2006). Prediction by supervised principal components. J. Am. Stat. Assoc. *101*, 119–137.

111. Boulesteix, A.L., and Strimmer, K. (2007). Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. Brief. Bioinform. *8*, 32–44. https://doi.org/10.1093/bib/bbl016.

112. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

113. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. Nature *596*, 590–596. https://doi.org/10.1038/s41586-021-03828-1.

114. Singh, A., Shannon, C.P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S.J., and Lê Cao, K.A. (2019). DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics *35*, 3055–3062. https://doi.org/10.1093/bioinformatics/bty1054.

115. Pearl, J. (2010). An introduction to causal inference. Int. J. Biostat. *6*, Article 7. https://doi.org/10.2202/1557-4679.1203.

116. Dugourd, A., Kuppe, C., Sciacovelli, M., Gjerga, E., Gabor, A., Emdal, K.B., Vieira, V., Bekker-Jensen, D.B., Kranz, J., Bindels, E.M.J., et al. (2021). Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. Mol. Syst. Biol. *17*, e9730. https://doi.org/10.15252/msb.20209730.

117. Ningappa, M., Rahman, S.A., Higgs, B.W., Ashokkumar, C.S., Sahni, N., Sindhi, R., and Das, J. (2022). A network-based approach to identify expression modules underlying rejection in pediatric liver transplantation. Cell Rep. Med. *3*, 100605. https://doi.org/10.1016/j.xcrm.2022.100605.

118. Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., and Zeitlinger, J. (2021). Base-resolution models of transcription-factor binding reveal soft motif syntax. Nat. Genet. *53*, 354–366. https://doi.org/10.1038/s41588-021-00782-6.

119. London, A.J. (2022). Artificial intelligence in medicine: overcoming or recapitulating structural challenges to improving patient care? Cell Rep. Med. *3*, 100622. https://doi.org/10.1016/j.xcrm.2022.100622.

120. Sinaci, A.A., Núñez-Benjumea, F.J., Gencturk, M., Jauer, M.L., Deserno, T., Chronaki, C., Cangioli, G., Cavero-Barca, C., Rodríguez-Pérez, J.M., Pérez-Pérez, M.M., et al. (2020). From raw data to FAIR data: the FAIRification workflow for health research. Methods Inf. Med. *59*, e21–e32. https://doi.org/10.1055/s-0040-1713684.