



Observational human studies in allergic diseases: design concepts and highlights of recent National Institute of Allergy and Infectious Diseases-funded research

Amanda K. Rudman Spergel and Alkis Togias

Purpose of review

To present and discuss key design concepts for optimizing the impact of observational studies in the field of allergy and to highlight recent findings from NIAID-funded research networks.

Recent findings

We discuss three concepts. First, the benefit of prospective, longitudinal observational studies exemplified by recent findings on the seasonal nature of all rhinitis phenotypes in children with asthma and the protective effects of high house dust allergen content during the first year of life on the development of asthma at age 7 years. Second, the benefit of detailed (deep) phenotyping exemplified by the identification of a *MALT1* gene variant as a strong genetic link to peanut allergy and the determination that atopic dermatitis with food allergy constitutes a distinct cutaneous endotype, compared with atopic dermatitis alone. Third, the benefit of hypothesis-generating research combined with prospective design and deep phenotyping as exemplified by the unveiling of potential pathophysiologic pathways leading to asthma exacerbations in children, after a 'cold'.

Summary

Observational studies can be highly impactful if designed well. Longitudinal study design, deep phenotyping, and hypothesis-generating research are three major design concepts that should be considered in the development of these studies.

Keywords

hypothesis-generating research, longitudinal studies, observational studies, phenotyping

INTRODUCTION

The mission of the National Institute of Allergy and Infectious Diseases (NIAID) at the National Institutes of Health (NIH) is 'to conduct and support basic and applied research to better understand, treat, and ultimately prevent infectious, immunologic, and allergic diseases.' As such, the NIAID constantly surveys these medical fields and identifies where programmatic emphasis should be placed. For the field of allergy, the current priorities are addressing the unmet scientific and medical needs for major conditions, such as asthma and allergic airways inflammation, food allergy and atopic dermatitis. In addition, NIAID has prioritized the need for understanding the mechanisms of immune tolerance in preventing or reversing allergic sensitization (Table 1).

In order to address these priorities, the NIAID funds both clinical trials and observational studies. Most people recognize the importance of clinical trials; such studies enable the medical and scientific community not only to determine the efficacy and/or safety of interventions that can prevent or ameliorate disease but also elucidate the understanding

Division of Allergy, Immunology and Transplantation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA

Correspondence to Amanda K. Rudman Spergel, MD, 5601 Fishers Lane Rm 6B54 MSC9827, Bethesda, MD 20892, USA.

Tel: +1 240 627 3840; fax: +1 301 480 4258;

e-mail: rudmana@niaid.nih.gov

Curr Opin Allergy Clin Immunol 2020, 20:000–000

DOI:10.1097/ACI.0000000000000620

KEY POINTS

- The field of allergy can be advanced with findings from observational studies.
- Longitudinal study design, deep phenotyping, and hypothesis-generating research are three major study design concepts that should be considered in the development of a meaningful observational study.

of pathophysiology. Observational studies are equally important, though. They can allow scientists to truly define phenotypes and endotypes of particular disease states and establish mechanistic pathways that are the basis for both health and disease. However, like clinical trials, observational studies require scientific rigor and careful design in order to produce reliable results of major impact. In this article, we discuss certain concepts that we believe are important in the design of an

observational study: the benefits of a longitudinal study design, deep phenotyping, and hypothesis-generating research. In discussing these concepts, we will also highlight recent clinical research findings deriving from the work of NIAID-funded clinical networks.

CONCEPT 1: LONGITUDINAL STUDY DESIGN

Observational studies on a specific chronic disease, such as asthma, typically assess disease activity. When developing the study, investigators can choose to make this assessment at a single timepoint via a cross-sectional approach. Alternatively, they can evaluate the activity over multiple visits for a length of time via prospective longitudinal studies. Cross-sectional studies that evaluate participants at a single timepoint are easier to conduct from a monetary and time standpoint. However, there are pitfalls with this approach as a single timepoint

Table 1. Priorities in the field of allergy

Asthma
Identify phenotypes and endotypes (the underlying mechanisms of disease) as they pertain to early-onset asthma, asthma exacerbations, and severe asthma
Understand innate and adaptive immunity interactions, including the role of airway epithelium and the glandular apparatus
Examine the role of environmental and airway microbiome in the induction and perpetuation of upper and lower airway disease
Improve sensitivity and accuracy of early disease prediction
Apply knowledge of disease mechanisms and predictors in clinical trials for disease prevention
Determine the specific roles of new and existing biologic therapies and allergen immunotherapy and microbiome manipulations in the management of severe asthma and in the modification of the natural history of chronic disease
Food allergy
Identify the mechanisms of food allergen sensitization and development of food allergy
Improve sensitivity and accuracy of early disease prediction
Develop reliable diagnostics for food allergy with the goal to minimize or eliminate the need for oral food challenges
Understand the interplay between the development of food allergy and other allergic disorders
Apply knowledge of disease mechanisms and predictors in clinical trials for disease prevention; that is, expanding and improving the concept of early dietary allergen introduction, considering early microbiome manipulations
Determine the role of various forms of allergen immunotherapy, new and existing biologic therapies, or other emerging approaches in the management and modification of the natural history of established food allergy
Atopic dermatitis
Identify phenotypes and endotypes as they pertain to early-onset atopic dermatitis, as well as to chronic and severe disease
Elucidate the role of the skin barrier and the skin microbiome in conjunction with environmental and genetic factors in the initiation and perpetuation of disease, and in the subsequent development of food allergy and allergic airway disease
Apply knowledge of disease mechanisms and predictors in clinical trials for disease prevention
Determine the specific role of new and existing biologic therapies and microbiome manipulations in the management of severe atopic dermatitis and in the modification of the natural history of chronic disease
Allergic sensitization and immune tolerance
Understand the process of allergen sensitization versus immune tolerance in early life and develop new strategies for clinical interventions that promote tolerance
Understand the pathophysiologic mechanisms that differentiate between allergen sensitization and development of allergic disease and develop strategies for clinical intervention that can interrupt this progression

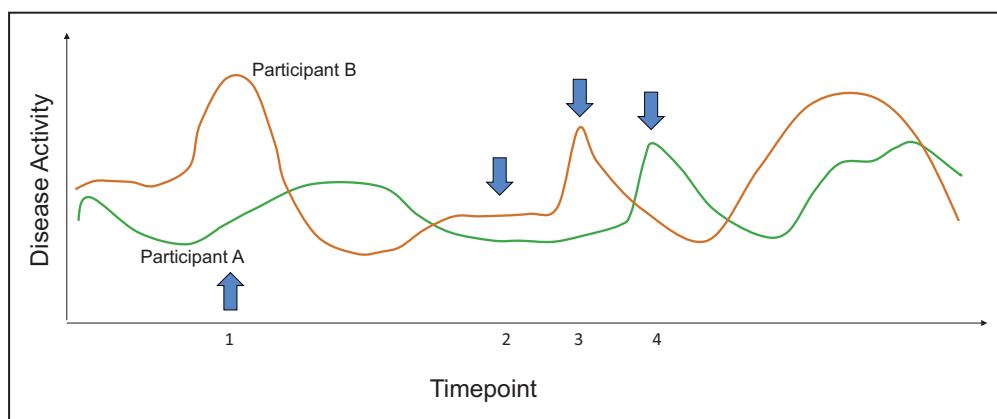


FIGURE 1. Disease activity in two individuals (participant A and participant B) at various timepoints. Downward arrows indicate examples of timepoints where a single-time, cross-sectional analysis could yield erroneous results. See text for detailed discussion.

only gives a narrow, and frequently misleading, view of the disease process. The benefit of investing in a prospective longitudinal study is the ability to look at the disease process at multiple timepoints; findings can change over time and, in reality, not be what they appear at first glance.

Take, for example, Fig. 1; if one compares the disease activity of participant A and participant B at timepoint 1, participant B looks much worse than participant A. If timepoint 1 was the single time assessed in the study, the researchers would conclude that participant B has more active disease than participant A. However, if timepoint 4 was the single time at which disease activity was assessed, the inverse conclusion would have been reached. Alternatively, if the investigators excluded disease exacerbations and evaluated each participant only at timepoint 2, there would have been no evidence of differential disease activity. If the investigators utilized a prospective longitudinal study and evaluated participants A and B at regular intervals, including periods of asthma exacerbations, they would have appreciated the variation of disease activity over time. For example, in comparing timepoints 3 and 4, the investigators would have known that both participants had an exacerbation. Furthermore, observing that participant B's exacerbation occurred before participant A's exacerbation, the investigators may have been able to attribute a cause for that variation, such as a participant-specific environmental exposure.

A real-life example of the utility of a prospective longitudinal study is the Asthma Phenotypes in the Inner City (APIC) study, which was conducted by the NIAID-funded Inner-City Asthma Consortium. In this study, 6–17-year-old children with asthma, regardless of the level of disease severity or control, underwent guideline-based management

for asthma and rhinitis for a 1-year period (six follow-up visits after enrollment) in order to better define the characteristics associated with Easy-To-Control versus Difficult-To-Control Asthma [1]. As part of this study, rhinitis was diagnosed and seasonal patterns were characterized via questionnaire at baseline and, in conjunction with allergy testing, APIC participants were categorized into one of five a priori rhinitis phenotypes and a small group of children (6.5%) who reported no or minimal symptoms of rhinitis. The longitudinal observational nature of the APIC study that included repetitive assessments of symptoms over an entire year in conjunction with optimal, standardized management allowed for validation of the a priori rhinitis phenotypes [2[¶]]. In addition, when the investigators studied how each group behaved over the course of the year, they noted that all predefined phenotypes exhibited seasonal oscillations, only differing in magnitude (Fig. 2). Interestingly, this observation also applied to the 'no rhinitis' group. These findings raise the question whether the seasonal oscillation is only because of allergen exposure or whether viral causes and climate conditions also play a major role with summer being the least symptomatic season for all phenotypes. In turn, such questions set up new hypotheses that could lead to new insight into the pathophysiology of rhinitis. The finding of seasonal oscillation in all types of rhinitis would not have been appreciated if a prospective longitudinal study approach had not been used.

In another example from the NIAID-funded Inner-City Asthma Consortium, the Urban Environment and Childhood Asthma (URECA) birth cohort study, which began in 2005, followed children from low-income, urban communities in four US cities, who were considered high risk for the development of atopic diseases and asthma based on parental

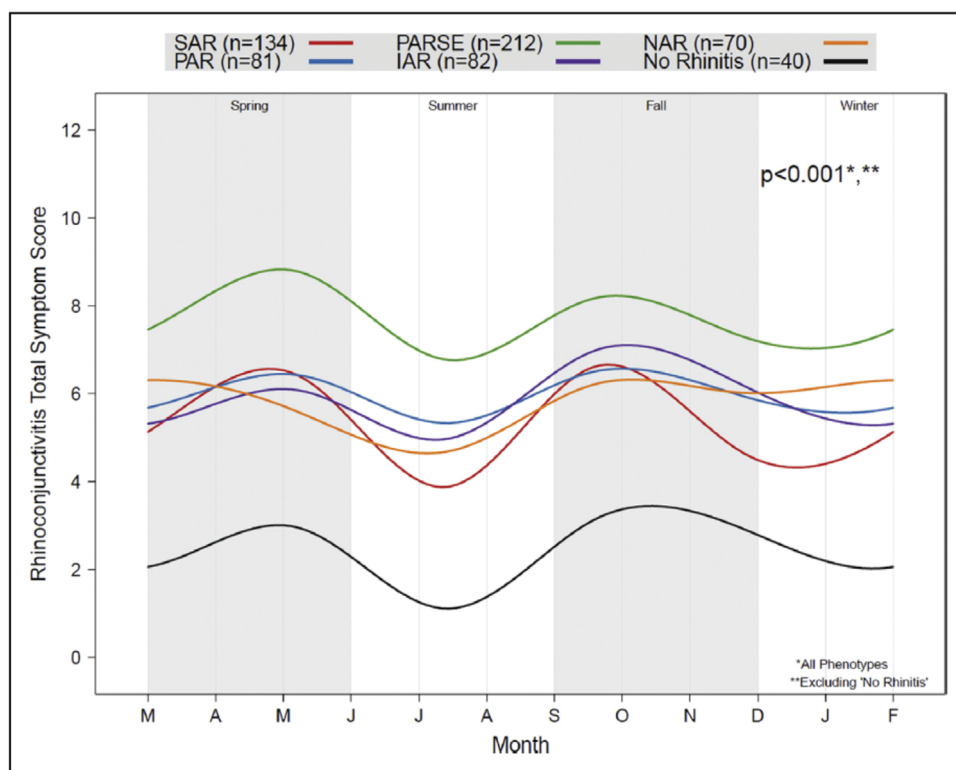


FIGURE 2. Seasonal variation in rhinoconjunctivitis symptoms scores by predefined rhinitis phenotypes and a 'no rhinitis' group among 619 children with asthma. *P* values correspond to a comparison of seasonal trend across all phenotypes, with and without the no rhinitis group. IAR, indeterminant atopic rhinitis; NAR, nonallergic rhinitis; PAR, perennial allergic rhinitis; PARSE, perennial allergic rhinitis with seasonal exacerbations; SAR, seasonal allergic rhinitis. Reproduced with permission from [2[¶]].

history [3]. At the age of 7 years, a determination was made, based on prespecified definitions, whether a participant had asthma [4^{¶¶}]. A major, unexpected observation was that participants who had higher cumulative allergen levels in house dust for cockroach, mouse, dog, and cat during the first year of life were at decreased risk of having asthma at age 7 years. Dust mite allergen exposure played no role. Furthermore, protection from asthma was not dependent on allergic sensitization as both sensitized and nonsensitized children who had high allergen exposure were at decreased risk of asthma. This observation led to the hypothesis that other factors associated with the presence of these allergens may mediate asthma protection; one of them being the house dust microbiome. Indeed, various bacterial taxa in house dust were found to be negatively or positively associated with asthma and the abundance of several of those taxa correlated with the levels of one or more of the protective allergens. If URECA had not been designed as a prospective longitudinal study, the connection between allergen exposure during the first year of life and asthma at age 7 would not have been realized.

CONCEPT 2: DEEP PHENOTYPING

The importance of establishing clear clinical phenotypes cannot be overstated when one is developing an observational study. Today's medicine is focused on personalized approaches to treating disease. In some fields, such as oncology, personalized medicine has already been introduced. However, lack of detailed (deep) phenotyping in the field of allergy has hindered advancement. For example, many genome-wide association studies have been conducted in asthma, but it has been difficult to identify strong associations [5]. This could partially be because of inadequate phenotyping. We now know, for example, that the term 'asthma' encompasses various types of airway inflammation, various forms of airway obstruction (reversible, partially reversible) and various clinical phenotypes related to age of onset, the presence of allergen sensitization, the propensity for frequent exacerbations, and so forth [6–8]. If ascertainment of asthma population in a large, cross-sectional genetic study is solely based on the criterion of 'physician diagnosis of asthma' or a similar, relatively superficial, phenotype, one can understand why genetic associations

may be diluted, and even if identified, they only explain a small portion of the disease. The same difficulty would likely apply to any allergic disease that is polygenic in nature.

Although not an observational study, the Learning Early About Peanut allergy (LEAP) trial, conducted by the NIAID-funded Immune Tolerance Network, which showed that early introduction of dietary peanut prevented peanut allergy by more than 80% in high-risk infants, is a good example of the importance of deep phenotyping in genetics [9]. By design, LEAP allowed for the creation of two highly characterized groups: children of high risk for peanut allergy who avoided peanut for 5 years (since infancy) and children of high risk for peanut allergy who were introduced to peanut-containing food in infancy and continued consuming such food regularly, for 5 years. In whole-genome sequencing performed in all study participants, a single nucleotide variant within the Mucosa-Associated Lymphoid Tissue Lymphoma Translocation (*MALT1*) gene, which is related to NF κ B activation in T and B cells after antigen binding to the T-cell or B-cell receptor was found to have a very strong association with peanut allergy and with peanut-specific IgE levels in the peanut avoidance arm, where the vast majority of peanut allergy cases were observed [10]. Strikingly, the genetic association of *MALT1* with both outcomes was eliminated in the peanut consumption arm offering even stronger evidence of a phenotype–genotype relationship. Such a strong association would have been unlikely to have been observed in a study that had simply ascertained children on the basis of established peanut allergy and had compared them to healthy controls without information on whether they had avoided or consumed peanut-containing foods earlier in life. The genetic link would have been most likely marred with peanut consumers versus avoiders lumped together.

Another recent example of the importance of deep phenotyping in observational research aiming at elucidating pathophysiologic mechanisms are the recent findings of a study conducted by the NIAID-funded Atopic Dermatitis Research Network [11[■]]. The study was designed to assess whether children with atopic dermatitis and food allergy (AD FA+) had different skin barrier function, compared with children with atopic dermatitis without food allergy (AD FA–) and nonatopic children. To characterize skin barrier function, a combination of transepithelial water loss (TEWL) and skin tape stripping was used in this study. AD FA+ was found to encompass a distinct endotype of atopic dermatitis characterized by increased TEWL as a function of the number of tape strips (Fig. 3). In addition, AD FA+ skin had

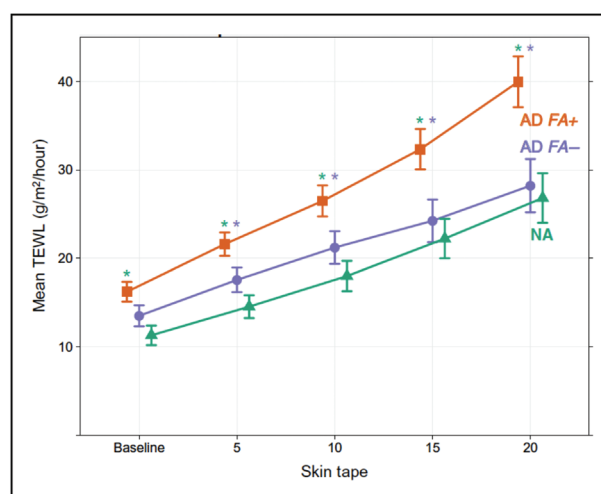


FIGURE 3. Transepithelial water loss measurements at baseline and after the consecutive application of 5, 10, 15, and 20 tapes on nonlesional skin of children with atopic dermatitis and food allergy (AD FA+), atopic dermatitis and no food allergy (AD FA–) and nonatopic children matched for age. Data are presented as means $\pm$ SEs. Colored asterisks indicate statistically significant differences: green represents a statistically significant difference between AD FA+ and NA, whereas purple compares AD FA+ to AD FA–. Reproduced with permission from [11[■]]. NA, nonatopic.

decreased filaggrin break down products and ω -esterified fatty acid sphingosine ceramide/nonhydroxy fatty acid sphingosine ceramide ratios. Those distinct differences characterizing children with AD FA+ would have not been unveiled if the investigators had not used such detailed phenotyping of skin function and structure.

CONCEPT 3: HYPOTHESIS-GENERATING RESEARCH

The final concept to consider when designing an observational study is that both hypothesis-driven and hypothesis-generating research need to be included in order to advance the field. Although hypothesis testing can address a specific predefined question, increased focus on the agnostic -omics and systems biology approaches in medical research will uncover connections in pathways that were previously unrecognized and will identify new target areas for future research. The fact that we can now confidently promote -omics research and systems biology analysis reflects the enormous advances in technology and bioinformatics that medicine has had access to starting only about 10 years ago.

The study by the NIAID-funded Inner-City Asthma Consortium entitled ‘Mechanisms Underlying Asthma Exacerbations Prevented and Persistent with Immune-Based Therapy: A Systems Approach

Phase 1 (MUPPITS-1)' is a good example where an agnostic hypothesis-generating approach was used in order to uncover new pathways associated with asthma exacerbations in children [12²²]. MUPPITS-1 was a longitudinal observational study of 208 children, 6–17 years of age, with difficult-to-control asthma designed to reveal changes in nasal cell gene expression associated with cold-induced asthma exacerbations. Nasal wash samples were collected

at baseline and serially following the onset of a cold that either led to an asthma exacerbation or spontaneously resolved. Furthermore, exacerbations were characterized as viral or nonviral based on nasal secretion testing. The primary analysis, using nasal cell RNA-seq transcriptomics, compared cold-associated asthma exacerbations with nonexacerbation colds. The investigators were able to determine four core gene expression modules, such as epithelial-

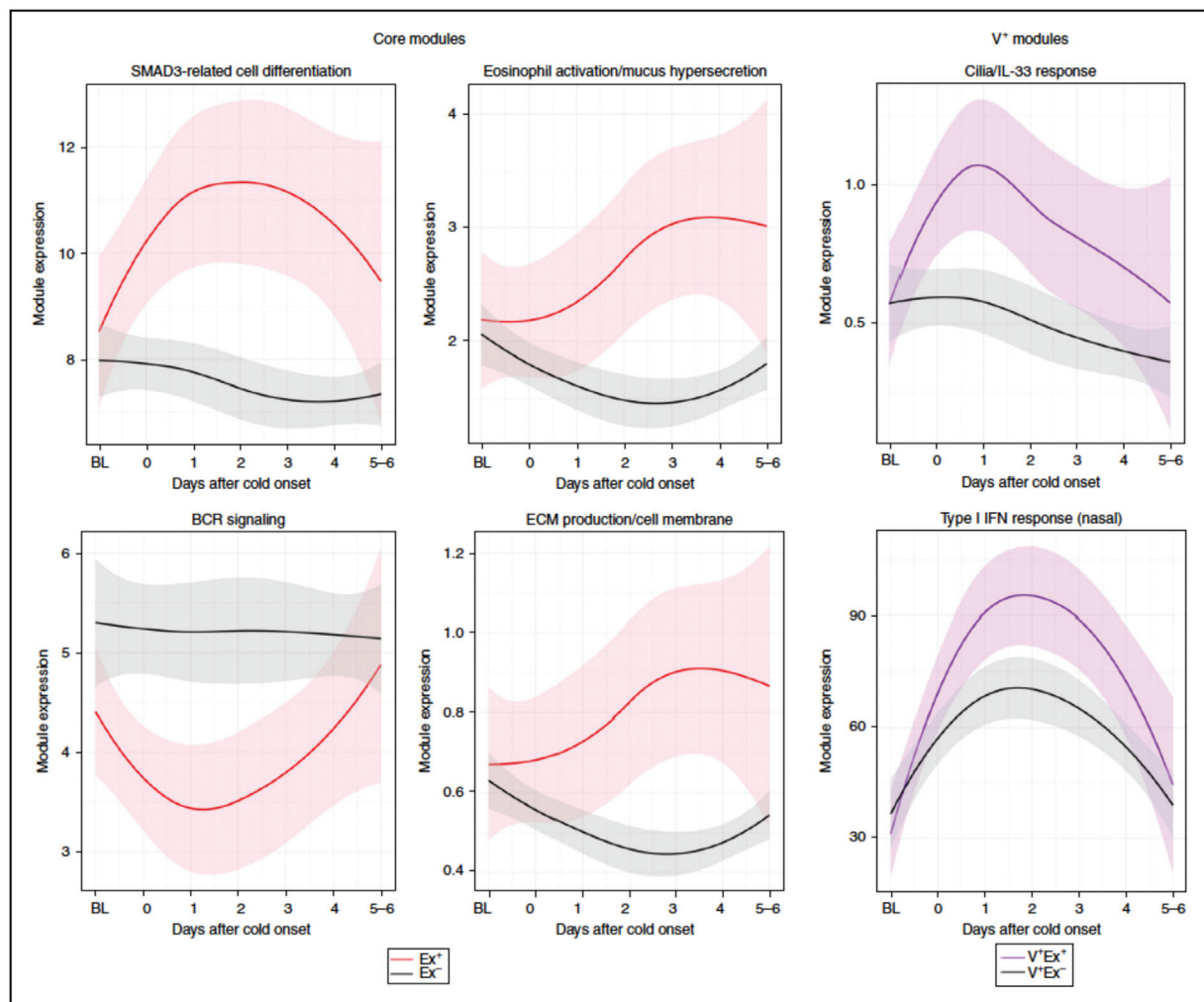


FIGURE 4. Examples of nasal cell gene expression modules and their time dynamics during the days following the onset of upper airway cold symptoms in children with severe asthma. The left four panels (pink) show major (core) modules that differentiated children who developed an exacerbation after the cold (Ex^+) from those whose cold resolved spontaneously (Ex^-). Among these core modules, expression in the epithelial-associated 'Mothers Against Decapentaplegic Homolog 3 (SMAD3)-related cell differentiation' module peaked early in the Ex^+ group, whereas expression in the 'B-cell receptor (BCR) signaling' module nadired early. Expression in the 'eosinophil activation/mucus hypersecretion' and the 'Extracellular matrix (ECM) production/cell membrane' modules increased over time in the Ex^+ but, if anything, decreased in the Ex^- group. Sample sizes are as follows: Ex^+ : $n = 92$; Ex^- : $n = 244$. The right two panels (purple) show modules that differentiated between children with colds where viruses were detected by PCR in nasal secretions and who either developed an asthma exacerbation (V^+Ex^+) or did not (V^+Ex^-). In V^+Ex^+ events increased expression of genes in the 'cilia/IL-33 response' module occurred early and was not present in the Ex^+V^+ group. In contrast, the expression of the nasal cell 'type I IFN response' module increased in both V^+Ex^+ and V^+E^- events, albeit less so in the V^+E^- . Sample sizes were as follows: V^+Ex^+ : $n = 67$; V^+Ex^- : $n = 157$. All data are plotted with 95% confidences (shaded region) estimated by local second-degree polynomial regression. Reproduced with permission from [12²²].

associated SMAD3 signaling and Epidermal Growth Factor Receptor signaling/cell–cell adhesion, that were upregulated in asthma exacerbation-associated colds while also identifying downregulated modules for B-cell receptor signaling and lymphocyte proliferation. Additionally, they were able to establish gene expression profiles that characterized viral versus nonviral asthma exacerbations (Fig. 4). This study and its findings highlight the point that an unbiased analytic approach applied in a prospective, longitudinal, study design with detailed clinical phenotyping and precise timing can allow for hypothesis-generating research to come to center stage. If the MUPPITS-1 investigators had not applied all three key concepts, we have highlighted for the design of this observational study, they likely would have not been able to identify new pathways that contribute to asthma exacerbations.

Hypothesis-generating research using -omics approaches has its own limitations including the lack of clarity as to the cellular source of products identified at the gene expression or protein level and the assumptions that are made during the analytic process of ‘simplifying’ the interpretation of the very large datasets that are generated. As a result, replication of findings is even more important than in nonbig data research. Furthermore, the cost is high. On the other hand, technology and bioinformatics are rapidly improving and the possibility that information at a single cell level becoming routine and being handled with multiomic analytics is approaching fast [13].

CONCLUSION

As the examples from the NIAID-funded clinical networks illustrate, observational studies can be equally illuminating as clinical trials. However, study design is important and investigators developing an observational study should consider longitudinal designs, conduct careful and detailed phenotyping, and incorporate hypothesis-generating research in order to ensure maximum impact.

Acknowledgements

None.

Financial support and sponsorship

None.

Conflicts of interest

There are no conflicts of interest.

Drs. Rudman Spergel's and Togias' co-authorship of this publication does not necessarily constitute endorsement by the National Institute of Allergy and Infectious Diseases, the National Institutes of Health or any other agency of the United States government.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Pongracic JA, Krouse RZ, Babineau DC, *et al.* Distinguishing characteristics of difficult-to-control asthma in inner-city children and adolescents. *J Allergy Clin Immunol* 2016; 138:1030–1041.
2. Togias A, Gergen PJ, Hu JW, *et al.* Rhinitis in children and adolescents with asthma: ubiquitous, difficult to control, and associated with asthma outcomes. *J Allergy Clin Immunol* 2019; 143:1003–1011.
3. Gern JE, Visness CM, Gergen PJ, *et al.* The Urban Environment and Childhood Asthma (URECA) birth cohort study: design, methods, and study population. *BMC Pulm Med* 2009; 9:17.
4. O'Connor GT, Lynch SV, Bloomberg GR, *et al.* Early-life home environment and risk of asthma among inner-city children. *J Allergy Clin Immunol* 2018; 141:1468–1475.
5. Kim KW, Ober C. Lessons learned from GWAS of asthma. *Allergy Asthma Immunol Res* 2019; 11:170–187.
6. Agache I, Akdis C, Jutel M, Virchow JC. Untangling asthma phenotypes and endotypes. *Allergy* 2012; 67:835–846.
7. Moore WC, Meyers DA, Wenzel SE, *et al.*, National Heart, Lung, and Blood Institute's Severe Asthma Research Program. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2010; 181:315–323.
8. Zoratti EM, Krouse RZ, Babineau DC, *et al.* Asthma phenotypes in inner-city children. *J Allergy Clin Immunol* 2016; 138:1016–1029.
9. Du Toit G, Roberts G, Sayre PH, *et al.*, LEAP Study Team. Randomized trial of peanut consumption in infants at risk for peanut allergy. *New Engl J Med* 2015; 372:803–813.
10. Winters A, Bahnson HT, Ruczinski I, *et al.*, Immune Tolerance Network LEAP Study Team. The MALT1 locus and peanut avoidance in the risk for peanut allergy. *J Allergy Clin Immunol* 2019; 143:2326–2329.
11. Leung DYM, Calatroni A, Zaramela LS, *et al.* The nonlesional skin surface distinguishes atopic dermatitis with food allergy as a unique endotype. *Sci Transl Med* 2019; 11:eaav2685.

This study compared the skin barrier function and superficial structure between children with atopic dermatitis and food allergy and those with atopic dermatitis without food allergy. Children with atopic dermatitis and food allergy present a distinct phenotype and endotype. This finding highlights the importance of deep phenotyping in the design of observational studies.

12. Altman MC, Gill MA, Whalen E, *et al.* Transcriptome networks identify mechanisms of viral and nonviral asthma exacerbations in children. *Nat Immunol* 2019; 20:637–651.

By utilizing a longitudinal, prospective design, detailed and time-sensitive phenotyping and nasal cell transcriptomics, this study identified molecular pathways involved in asthma exacerbations preceded by viral and nonviral upper airway cold-like symptoms.

13. Hu Y, An Q, Sheu K, *et al.* Single cell multi-omics technology: methodology and application. *Front Cell Dev Biol* 2018; 6:28.