Omics-oriented research illustrated with the LEAP study and the OASIS bioinformatics tool

Check for updates

Carolyn H. Baloh, MD,[a,b] Kanika Kanchan, PhD,[c] Gautam Shankar, MS,[c] Gerald T. Nepom, MD, PhD,[a,d] Rasika A. Mathias, ScD,[c] and James A. Perry, PhD[e] *Seattle, Wash; Boston, Mass; and Baltimore, Md*

Key words: *Peanut allergy, food allergy, allergy prevention, genetics, genomics, genome-wide association study, GWAS, mucosa-associated lymphoid tissue lymphoma translocation gene,* MALT1 *gene, bioinformatic tool(s)*

Automated, omics-oriented research processes can simplify and accelerate biologic discovery from massive quantities of analysis results generated by clinical and biologic studies. In this Paradigms and Perspectives article and as summarized in Fig 1, we describe our application of the preexisting Omics Analysis Search and Information System (OASIS) website[1] to integrate data from the Immune Tolerance Network (ITN) Learning Early About Peanut Allergy (LEAP) study to identify genetic variants, genes, and proteins implicated in immune tolerance phenotypes. OASIS is a preexisting website tool that brings together publicly available software.[1] We refer to our application of OASIS to ITN LEAP data and our planned application of it to additional ITN data collectively as ITN OASIS.

In Fig 1 we describe an integrative process using omics resources and biologic data repositories to identify genetic variants, genes, and proteins that are implicated in immune tolerance phenotypes. Genome-wide association study (GWAS) analysis allows initial identification of variants associated with phenotypes and diseases, which then need follow-up with functional mapping and validation. Colocalization analysis between GWAS results and quantitative trait loci (QTLs) from transcriptomics and/or proteomics data will facilitate this validation and help hone in on which variant(s) and gene(s) might be credible causal factors identified from the GWASs. This is further supported by functional mapping using a variety of public genomic annotation tools and direct integration with online tools such as the University of California Santa Cruz (UCSC) Genome Browser. Phenomewide associations with sentinel GWAS variants leveraging expansive and publicly available biobank data sets may help offer replication opportunities and reveal pleiotropic effects across correlated phenotypes. These steps have been automated within OASIS and are now available to researchers wishing to leverage genetic studies from the ITN. At present, ITN OASIS has extensive GWAS summary statistics data available from LEAP, described later in this article. Additional studies (eg, Oral Immunotherapy for Induction of Tolerance and Desensitization in Peanut-Allergic Children [IMPACT]) are being onboarded to facilitate cross-study hypothesis testing for the scientific community.[2] The remainder of this article provides an example of using this omics-oriented research process for the study of peanut allergy, with details illustrated in Fig 2.

CLINICAL STUDY OF A PHENOTYPE

The LEAP study was a randomized controlled trial designed to determine whether early introduction of peanut would prevent development of peanut allergy (Fig 2, *A*).[3] Participants were considered at high risk for peanut allergy if they had severe eczema and/or egg allergy. They were aged 4 to 11 months at randomization to either peanut consumption or peanut avoidance, with each randomization group containing infants with a skin prick test to peanut resulting in a wheal either less than 1 mm or between 1 and 4 mm. Regardless of the initial skin prick test results, early introduction of peanut decreased the frequency of peanut allergy at 60 months of age. Biomarkers of peanut allergy were also improved by early peanut introduction.[3]

GENOMICS: GWAS OF PEANUT ALLERGY

Whole genome sequencing was performed and analyzed by using a GWAS to identify genetic risk factors for peanut allergy.[4] The peak association with peanut allergy risk was located on chromosome 18 (Fig 2, *B*), a region containing the mucosa-associated lymphoid tissue lymphoma translocation (*MALT1*) gene.[4] This association was independent of baseline skin prick test results (sensitization status), egg allergy, or eczema. *MALT1* was also associated with peanut-specific IgE level, independent of total IgE level, and was found to be associated with component spreading with associations for *Arapis hypogaea* 1, 2 and 3. *MALT1* has also been associated with biomarkers of peanut allergy.[4]

FUNCTIONAL MAPPING OF TOP GWAS SIGNALS

Results from the GWAS of peanut allergy were loaded into OASIS, a web-based tool for mining GWAS results and connecting to other bioinformatics tools and online data repositories.

From [a]The Immune Tolerance Network, Benaroya Research Institute at Virginia Mason, Seattle; [b]the Department of Medicine, Harvard Medical School, Division of Allergy and Clinical Immunology, Brigham and Women's Hospital, Boston; [c]the Department of Medicine, School of Medicine, Johns Hopkins University, Division of Allergy and Clinical Immunology, Baltimore; [d]Benaroya Research Institute at Virginia Mason, Seattle; and [e]the Department of Medicine, University of Maryland School of Medicine, Baltimore.

This research was performed as a project of the Immune Tolerance Network, an international clinical research consortium headquartered at the Benaroya Research Institute, and supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (award UM1AI109565). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

Received for publication July 11, 2022; revised November 18, 2022; accepted for publication December 5, 2022.

Available online December 12, 2022.

Corresponding author: Rasika Mathias, ScD, 5501 Hopkins Bayview Circle, Baltimore, MD 21224. E-mail: rmathias@jhmi.edu.

J Allergy Clin Immunol 2023;151:416-9.

0091-6749/$36.00

© 2022 American Academy of Allergy, Asthma & Immunology

https://doi.org/10.1016/j.jaci.2022.12.003

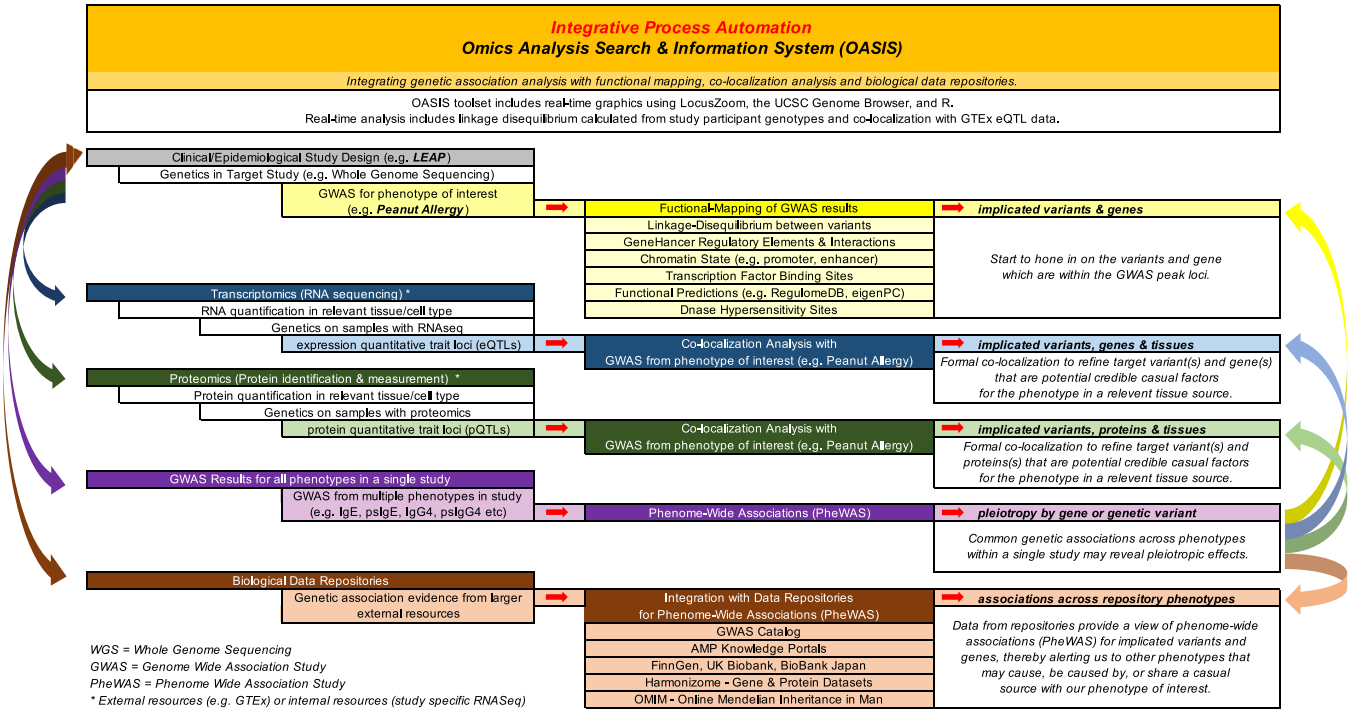

FIG 1. Overview of the ITN OASIS pipeline to demonstrate the types of data available for integration and the purpose of the integration of each data type, with primary posted GWAS summary statistics. *Arrows* are used to illustrate the nonlinear nature of these potential queries, and although not exhaustive in nature, they showcase the value of cross-data interrogation.

Linkage disequilibrium (LD) is computed by OASIS for all variants mapping to a selected GWAS peak (eg, chromosome 18 rs7665350 ± 200 kb) and then plotted with an integrated version of LocusZoom. The plot reveals that the GWAS signal comes from 12 variants that are in high LD and thus statistically indistinguishable (Fig 2, *C*), and the variants map to *MALT1*. OASIS's integration of the GWAS data with the University of California Santa Cruz Genome Browser[5] (Fig 2, *D*) allows for easy identification of each variant's potential functional implications for gene expression, such as proximity to enhancer locations, chromatin state, transcription factor binding sites, DNase hypersensitivity sites, and any other functional information available in the Genome Browser. From the display, it is immediately obvious that 2 variants are functionally implicated owing to intersections with GeneHancer regions and yellow (enhancer) chromatin state regions, as well as transcription factor binding sites and DNase hypersensitivity sites. The 2 variants are rs76653504 (in Fig 2, the top GWAS signal is highlighted with a lighter green line) and rs4940418 (the leftmost variant).

COLOCALIZATION ANALYSIS WITH TRANSCRIPTOMICS FROM GTEx

The Genotype-Tissue Expression (GTEx) project[6] provides tissue-specific expression QTL (eQTL) data for 18,000 genes and 54 tissues. OASIS provides real-time colocalization analysis[7] between the study phenotype GWAS signal and eQTL signals from GTEx. A positive colocalization result suggests that the variants driving the GWAS signal are also modulating gene expression, thus providing supportive evidence for the connection between the phenotype and a specific gene and tissue. Fig 2, *E* shows that rs76653504, the top GWAS signal, also has a significant eQTL signal (*above the red line*) from the GTEx analysis for skin tissue and *MALT1*. Most of the variants in LD with the top GWAS signal (*shown with red circles*) are also above the eQTL significance line. OASIS also performs formal colocalization using bayesian approaches (eg, *coloc*) against all available GTEx data for rapid and easy identification of credible gene(s) and variant(s).

PheWASs WITH BIOLOGIC DATA REPOSITORIES

Phenome-wide association studies (PheWASs) can alert us to other phenotypes that may cause, be caused by, or share a casual source. OASIS provides PheWASs (Fig 2, *F*) for any set of user-selected variants from a broad set of integrated resources, including FinnGen, UK Biobank, BioBank Japan, GWAS Catalog, Harmonizome, Online Mendelian Inheritance in Man (OMIM), and the Accelerating Medicines Partnership (AMP) knowledge portals.[7]

DISCUSSION

Identification and validation of variant and gene associations for a phenotype are greatly improved with an omics-oriented process and can provide a significant head start for functional studies. Here we have provided an illustration of querying on a single-peak locus, but OASIS also facilitates queries on candidate genes, specific variants, and cross-GWAS data mining. For the

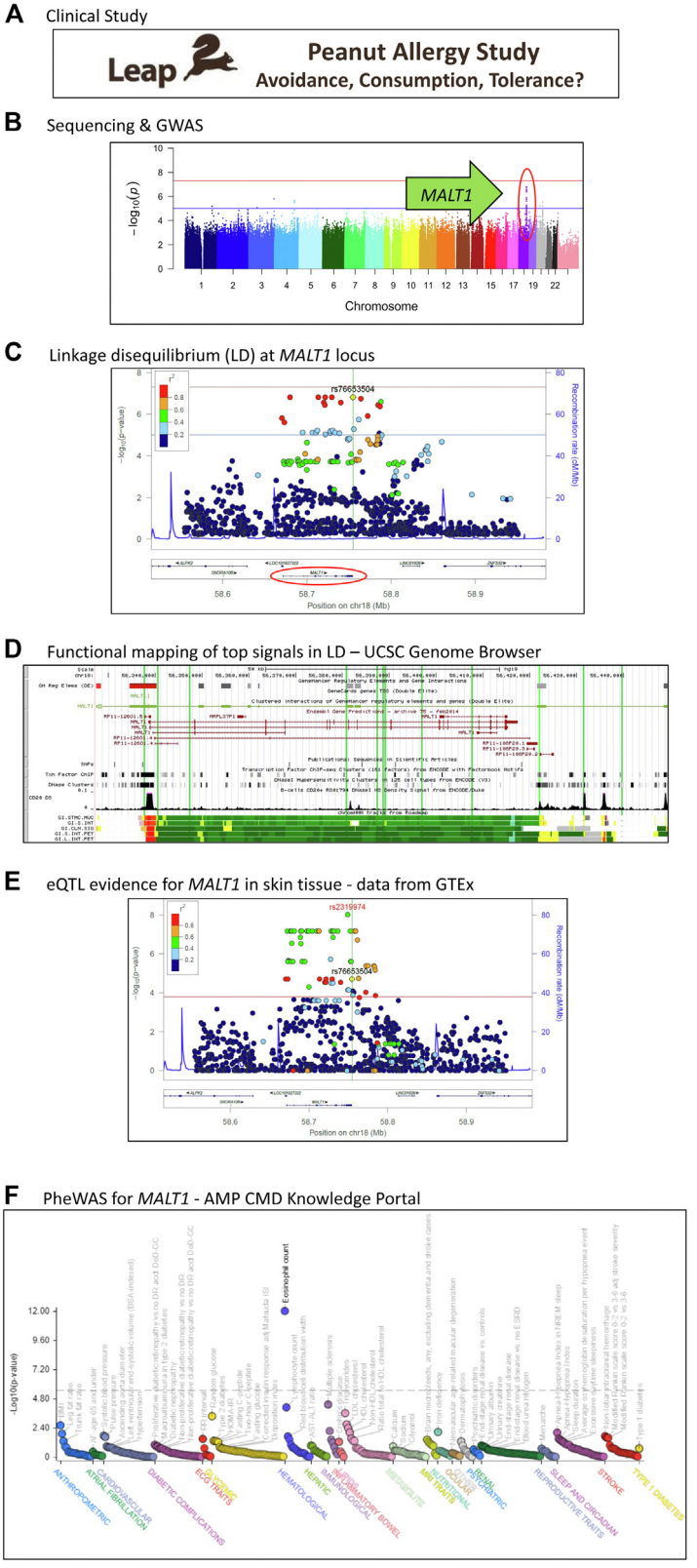

FIG 2. Illustration of the value of the OASIS automated pipeline in the GWAS for peanut allergy in LEAP from selecting the data set (**A**), viewing the full GWAS summary in Manhattan plots (**B**), zooming into specific loci and identifying mapped genes (**C**), functional mapping of the sentinel GWAS SNPs (**D**) (*green lines* represent the set of single-nucleotide polymorphisms in red from panel 3), and examining QTLs from public catalogs such as GTEx (formal colocalization is an available option) (**E**) to PheWAS examination in online resources (**F**). *AMP CMD*, Accelerating Medicines Partnership Program for Common Metabolic Diseases; *Chr*, chromosome; *ECG*, electrocardiogram.

LEAP study, the omics-oriented research process has been applied to a total of 55 GWAS outcomes and traits to date (including peanut allergy, biomarkers of allergy, and tolerance, including IgE, IgG4, and so forth), all of which are currently available in the web-based ITN OASIS bioinformatics tool. GWAS summary statistics results from additional immune tolerance studies will be available in the near future. Researchers wishing to leverage these expansive GWAS results and the tools included within ITN OASIS to mine these results may request access at https://edn.som.umaryland.edu/OASIS/ITN/. This link also provides unrestricted access to the ITN OASIS video library of tutorials and demonstrations.

ITN OASIS runs in a Federal Information Security Management Act of 2002–certified, secure web environment in which access to line-level and/or raw genotype data is restricted and only summary GWAS statistics and group-level genotype summaries are accessible to users. Additionally, access controls are available at the user and data set levels to control access to specific results by users. ITN OASIS anticipates continued expansion to new public resources (eg, onboarding more molecular QTLs for colocalization, more analysis options for fine-mapping, and additional biobank results). It is open to proposals for uploading additional study data sets (phenotype and genotype data) to allow the omics-oriented research process to be applied broadly to the ITN's mission. Potentially, the resource could be expanded to include GWAS summary statistics for other phenotypes across the atopic march in external GWASs (eg, asthma, allergic rhinitis) and also include the GWASs for food allergy that are publicly available. As these additional data are added to this resource, it may be possible to differentiate between types of allergens (eg, differentiate primary food allergens from cross-reacting allergens that illicit pollen food allergy) and facilitate research for common and distinct variants across allergy phenotypes.

Software availability

The OASIS code and all software tools and databases integrated into OASIS are available free of charge for noncommercial use. Visit the OASIS Resources site[1] for details and references.

We thank the LEAP study team members for their valuable research on peanut allergy, which provided the foundation for this work. We appreciate all of the LEAP study participants whose data are included in this study.

REFERENCES

1. Perry JA. OASIS resources, video library and contact information 2022. Available at: https://edn.som.umaryland.edu/OASIS/. Accessed October 24, 2022.
2. Jones SM, Kim EH, Nadeau KC, Nowak-Wegrzyn A, Wood RA, Sampson HA, et al. Efficacy and safety of oral immunotherapy in children aged 1-3 years with peanut allergy (the Immune Tolerance Network IMPACT trial): a randomised placebo-controlled study. Lancet 2022;399:359-71.
3. Du Toit G, Roberts G, Sayre PH, Bahnson HT, Radulovic S, Santos AF, et al. Randomized trial of peanut consumption in infants at risk for peanut allergy. N Engl J Med 2015;372:803-13.
4. Winters A, Bahnson HT, Ruczinski I, Boorgula MP, Malley C, Keramati AR, et al. The MALT1 locus and peanut avoidance in the risk for peanut allergy. J Allergy Clin Immunol 2019;143:2326-9.
5. Casper J, Zweig AS, Villarreal C, Tyner C, Speir ML, Rosenbloom KR, et al. The UCSC Genome Browser database: 2018 update. Nucleic Acids Res 2018;46:D762-9.
6. Lonsdale TJ, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580-5.
7. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014;10:e1004383.