RESEARCH LETTER

Whole blood transcriptomics identifies gene expression associated with peanut allergy in infants at high risk

To the Editor:

The prevalence of peanut allergy is increasing amongst children.[1] Early introduction of peanut by 4–6 months of age in infants at high risk for peanut allergy (defined by guidelines as those with severe atopic dermatitis, egg allergy or both diagnoses) may prevent peanut allergy.[2,3] Despite the availability of guidelines for early peanut introduction, there will continue to be infants who have peanut allergy. Once established, peanut allergy is often persistent with only 10%–20% of patients developing natural tolerance. With emerging therapeutic options, we must identify patients early and intervene with tertiary prevention.

Current testing modalities for evaluation of food allergy, including skin prick testing (SPT) and serum specific IgE (sIgE) measurement, are limited by false-positive results. Oral food challenges (OFC), the current standard for food allergy diagnosis, involve feeding incrementally increasing doses of allergenic protein to an individual to determine their food allergy status. These challenges present clinical risk, are expensive and time-consuming and require intensive nursing support with close clinician oversight. In infant populations, these challenges are more complex. Thus, there is a need for improved diagnostic tools for food allergy.

This study harnessed whole blood RNA sequencing technology to examine gene expression differences among infants at high risk for peanut allergy who are peanut allergic (PA), peanut sensitized but orally tolerant (PST) and peanut non-allergic (PNA). Our goal was to identify genes whose expression was sufficiently different as to discriminate between the three groups of subjects with effect sizes of at least one half log (~3 fold) with minimal overlap between groups. In our analysis of the data from this study, there were no differences in the whole blood transcriptome between these three groups that was large enough to distinguish the different clinical phenotypes. Gene expression studies in patients with food allergy have been used to identify molecular pathways underlying the development of sensitization and clinically reactive food allergy, elucidate genes implicated in food-induced reactions and determine if differences in gene expression can distinguish individuals with clinically reactive food allergy from those who are non-allergic.[4–8] These studies have primarily focussed on the use of antigen-stimulated or unstimulated peripheral blood mononuclear cells (PBMC) or T cells for evaluation of differential gene expression. Use of whole blood, requiring minimal processing, is practical for any clinical diagnostic test. Whole blood transcriptomics has been successfully used to identify gene expression signatures of other disease states; however, in this study, clinically meaningful differences were not identified.

We recruited infants aged 4–11 months with severe atopic dermatitis and/or egg allergy from the Allergy & Immunology clinic at the Ann and Robert H. Lurie Children's Hospital of Chicago (Chicago, IL). All data were obtained during their initial visit to allergy clinic, which was also the time of enrollment with written informed consent. The study was approved by the Institutional Review Board at the Ann & Robert H. Lurie Children's Hospital of Chicago. Clinically, seventeen infants underwent SPT and/or sIgE testing to peanut; the remaining 3 infants were orally tolerant of peanut at the time of enrollment so testing was not indicated. Subjects were categorized as PA, PST and PNA based on clinical history, sensitization data and/or OFC (see Supplemental Methods; Figure S2).

Of the 20 infants included in the analysis, 10 had severe eczema (50%), 7 had egg allergy (35%) and 3 had both diagnoses (15%) (Table 1). RNA sequencing of whole blood was used to evaluate differentially expressed genes between PA, PST and PNA subjects. We report genes with differential expression using criteria of a fold change ≥±1.3 and false discovery rate (FDR) p-value ≤0.05 (Table 2 and Figure S1).

There were eight significantly differentially expressed genes in the PST subjects compared to the PNA group (FDR p-value ≤0.05; Figure S1). The upregulated genes include *MGAM2*, *MCEMP1*, *C4BPA*, *S1000A8* (Figure S3), *CECR6* and *LINC01270*. *MEG3* and *DKK3* were downregulated. In the blood, *MGAM2* and *C4BPA* mRNA are primarily enriched in neutrophils. *MCEMP1*, *S100A8* and *CECR6* mRNA are primarily enriched in monocytes and neutrophils. *DKK3* mRNA expression is enriched in MAIT T-cells and naïve and memory CD8$^+$ T cells (Table 2).[9] Four genes were differentially expressed in PA versus PNA subjects (FDR p-value ≤0.05; Table 2), including *LOC283575*, *PLEKHD1*, *LOC100506159* and *ANKUB1*. In the blood, *PLEKHD1* mRNA expression is enriched in plasmacytoid dendritic cells. *ANKUB1* mRNA expression is detected primarily in the neutrophil population (Table 2).[9] Four genes were significantly differentially expressed in PA compared to PST subjects (FDR p-value

≤0.05; Figure S1), including *LOC101928126*, *FBXO2*, *COL18A1* and *TNFRSF10D* (Figure S3). *FBXO2* mRNA is detected in many cell types within the blood; however, it is enriched non-specifically in T- and B-cells. *COL18A1* mRNA expression can similarly be detected in many cell types; however, its expression is enriched in neutrophils. In the blood, *TNFRSF10D* is broadly expressed with some enhancement in naïve CD4[+] T cells (Table 2).[9] Deconvolution of the gene expression data showed that none of 22 targeted immune cell populations were significantly different by subject group (Figure S4).

In this study, differential gene expression in whole peripheral blood was present between PA versus PST, PA versus PNA and PNA versus PST infants; however, the fold changes ranged from 0.3 to 2.2. While statistical significance was identified for differentially expressed genes, the differences were too small to have clinical significance given the magnitude of changes and the variance between subjects within the same clinical group. There was no overlap of significant differentially expressed genes between the three groups. Notably, the genes that were differentially expressed were not consistent with what has been shown in studies of food allergic subjects using more targeted cell populations.[4,6,7] However, other targeted approaches may have missed some of the genes detected in our analysis (e.g. *S1000A8*), and this now provides further rationale for

> **Key Messages**
>
> - Gene expression in whole peripheral blood differed between peanut sensitized, allergic and non-allergic, non-sensitized subjects.
> - The use of whole blood to examine gene expression differences has limitations.
> - Future studies are required for the discovery of novel biomarkers for food allergy.

future studies investigating other important cell populations such as neutrophils that may have a role in the development of food allergy.

Examination of stimulated cord blood CD4[+] T cells from neonates subsequently diagnosed with IgE-mediated food allergy showed reduced expression of genes in the *NFKB* gene family compared to non-allergic neonates, suggesting that T-cell activation pathways may be deficient early in life amongst infants who go on to develop IgE-mediated food allergy.[7] This finding was further corroborated when the same group of investigators found that naïve CD4[+] T cells from egg allergic one year olds had decreased responsiveness upon

TABLE 1 Study participant characteristics

	Peanut Allergic (N = 8)	Peanut Sensitized, Tolerant (N = 5)	Peanut Non-Allergic (N = 7)	Overall (N = 20)
Age, months, median (IQR)	9.5 (3.5)	7 (0)	8 (4)	8 (3.5)
Gender, male, N (%)	6 (75)	4 (80)	3 (42.9)	13 (65)
Diagnosis of atopic dermatitis, N (%)	7 (87.5)	5 (100)	5 (71.4)	17 (85)
Diagnosis of severe atopic dermatitis, N (%)	4 (50)	3 (60)	3 (42.9)	13 (65)
Diagnosis of egg allergy, N (%)	3 (37.5)	1 (20)	3 (42.9)	10 (50)
Both severe atopic dermatitis and egg allergy, N (%)	1 (12.5)	1 (20)	1 (14.3)	3 (15)
Breastfed, N (%)[1]	8 (100)	5 (100)	7 (100)	18 (100)
Age of complementary feeding, months, median (IQR)[2]	6 (0)	5 (2)	5 (1)	6 (1)
Family history of food allergy, N (%)	3 (37.5)	5 (60)	5 (71.4)	11 (55)
Family history of atopic dermatitis, N (%)	4 (50)	3 (60)	3 (42.9)	10 (50)
Family history of allergic rhinitis, N (%)	4 (50)	4 (80)	6 (85.7)	14 (70)
Family history of asthma, N (%)	3 (37.5)	2 (40)	3 (42.9)	8 (40)
Peanut SPT wheal size, mm, median (IQR)[3]	7 (3)	4 (1)	0 (2)	4 (4.5)
Peanut sIgE, kU$_A$/L, median (IQR)[4]	14.9 (80.4)	1.1 (6.8)	0.1 (0)	9.5 (16.8)
Ara h 2 sIgE, kU$_A$/L, median (IQR)[5]	5.1 (42.9)	0 (0.12)	N/A	2.0 (37.6)

Abbreviations: IQR, interquartile range; sIgE, specific immunoglobulin E; SPT, skin prick testing.

[1]Data only available for N = 18.
[2]Data only available for N = 19.
[3]Data only available for N = 16.
[4]Data only available for N = 13.
[5]Data only available for N = 11.

TABLE 2 Differentially expressed genes by comparison of participant groups

	Gene Name	Up-regulated	Down-regulated	Fold Change	Known Function[10,11]	Predominant Blood Cell Expression[9]
PST vs. PNA	MGAM2 (maltase-glucoamylase 2)	X		2.2	-Carbohydrate binding -Glucan 1,4-alpha-glucosidase activity	Neutrophil
	MCEMP1 (mast cell expressed membrane protein 1)	X		2.2	-Protein binding	Monocyte, neutrophil
	C4BPA (complement component 4 binding protein alpha)	X		2.0	-Encodes the alpha chain of complement component 4 binding protein, which regulates complement activation	Neutrophil
	S100A8(S100 calcium binding protein A8)	X		2.0	-Calcium and zinc binding protein -Regulates inflammation -Acts as alarmin -Antimicrobial activity	Neutrophil, monocyte
	CECR6 (TMEM121B or transmembrane protein 121B)	X		2.0	-Candidate gene for Cat Eye Syndrome	Neutrophil, monocyte
	LINC01270 (long-intergenic non-protein coding RNA 1270)	X		2.1	-Non-coding RNA	N/A
	MEG3 (maternally expressed 3)		X	0.3	-Non-coding RNA	N/A
	DKK3 (dickkopf Wnt signaling pathway inhibitor 3)		X	0.5	-Wnt signalling antagonist	T-cell
PA vs. PNA	LOC283575 (long intergenic non-protein coding RNA 2288)		X	0.5	-Non-coding RNA	N/A
	PLEKHD1 (plecksstrin homology and coiled-coil domain containing D1)		X	0.6	N/A	Plasmacytoid dendritic cell
	LOC100506159 (long intergenic non-protein coding RNA 2470)	X		1.6	-Non-coding RNA	N/A
	ANKUB1 (Ankyrin repeat and ubiquitin domain containing 1)		X	0.6	N/A	Neutrophil
PA vs. PST	LOC101928126 (uncharacterized)	X		1.8	-Non-coding RNA	N/A
	FBXO2(F-box protein 2)		X	0.7	-Involved in protein ubiquitination and degradation	Multiple
	COL18A1 (collagen type XVIII alpha 1 chain)		X	0.5	-Extracellular matrix -Post-translational modification results in endostatin	Neutrophil
	TNFRSF10D(TNF receptor superfamily member 10D)	X		1.4	-Receptor for TRAIL and protective against apoptosis	Naïve CD4+ T-cell

Abbreviations: PA, peanut allergic; PNA, peanut non-allergic; PST, peanut sensitized, but tolerant.

polyclonal stimulation compared to non-atopic age-matched controls.[6] Examination of differential gene expression in peanut stimulated memory T cells from peanut allergic and peanut non-allergic children showed that *IL9* gene expression could best distinguish between these two groups.[4] We did not identify differential gene expression of *IL9* or *NFKB* family members between the three groups in our study, suggesting that whole blood is not sensitive to robust biomarkers identified in studies of specific cell populations.

A limitation of our study is the small sample size and that the subjects were characterized clinically with history, SPT, sIgE and/or OFC. There was lack of a non-atopic control group for this study. However, the purpose of the study was to assess gene expression differences among infants at high risk for peanut allergy as opposed to identifying differences from non-allergic controls. While deconvolution of whole blood transcriptomic data is a practical approach for identification of cell populations, particularly when the volume of a whole blood sample is limited, mass cytometry or spectral flow cytometry would allow for more refined characterization of these cell populations and should be considered in the future studies of this infant population.

Whole blood based techniques may not be refined enough to detect minor yet relevant gene expression changes. This was likely a major factor contributing to the magnitude of our findings. However, the use of whole blood may allow for detection of other relevant pathways that may have been missed by more focussed sequencing. Whilst this study was negative, it does suggest that studying other cell populations may help to elucidate novel and potentially important mechanisms underlying the loss of tolerance to food. In developing a clinical test, it is important to take into account the technical expertise required to perform the test. Use of whole blood for RNA extraction is simpler, less costly and requires limited technical expertise when compared to RNA extraction from an isolated cell population. However, for the detection of subtle gene expression changes or changes that may be occurring in a smaller cell population, the use of a highly processed sample or specific cell type is likely necessary. Similarly, there are differences between individuals in the cellular composition of whole blood that may contribute to gene expression findings. Limitations on blood volume collection in infants precluded us from obtaining more blood for complete blood counts or cell sorting by flow cytometry.

While a strong diagnostic test for food allergy with minimal sample processing is needed, the lack of clinically meaningful gene expression differences and limitations of using whole blood as described in this study suggest that a more targeted approach focusing on specific cell populations may be required.

ACKNOWLEDGEMENT

CONFLICTS OF INTEREST

RK is a consultant for Regeneron—Sanofi.

AUTHOR CONTRIBUTIONS

ALD—Conceptualization and design, data acquisition, analysis and interpretation of data, writing—original draft preparation. HF—Data acquisition, writing—reviewing and editing. MP—Analysis, visualization, and interpretation of data, writing—reviewing and editing. AP—Analysis, visualization, and interpretation of data, writing—reviewing and editing. DG—Conceptualization, data acquisition, writing—reviewing and editing. MS—Analysis, visualization, and interpretation of data, writing—reviewing and editing. JMD—Analysis, visualization, and interpretation of data, writing—reviewing and editing. SA—Analysis, visualization, and interpretation of data, writing—reviewing and editing. JAP—Conceptualization, writing—review and editing. LCK—Conceptualization, interpretation of data, writing—review and editing. RK—Conceptualization and design, data acquisition, analysis and interpretation of data, writing—review and editing.

ETHICAL STATEMENT

This study was approved by the Institutional Review Board at the Ann and Robert H. Lurie Children's Hospital of Chicago in Chicago, IL. The IRB ID number was 2019–1048. Written, informed consent was obtained prior to any study procedures.

FUNDING INFORMATION

DATA AVAILABILITY STATEMENT

The gene expression data generated and analysed during the current study will be made openly available in the short read archive NCBI GEO. The clinical data generated and analysed during the current study are available from the corresponding author on request.

Ashley L. Devonshire[1,2,3,4] (iD)

Hanli Fan[5]

Mario Pujato[6]

Aditi Paranjpe[6]

Demirkan Gursel[5]

Matthew Schipma[7]

Julia M. Dunn[3]

Sandra Andorf[3,4,6] (iD)

Jacqueline A. Pongracic[1,2]

Leah C. Kottyan[3,4,8] (iD)

Rajesh Kumar[1,2] (iD)

[1]Division of Allergy and Immunology, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA

[2]Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

[3]Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

[4]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA

[5]Pathology Core Facility, Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

[6]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

[7]Department of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

[8]Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Correspondence

Ashley L. Devonshire, 3333 Burnett Avenue, MLC 2000, Cincinnati, OH 45229, USA.
Email: ashley.devonshire@cchmc.org

ORCID

Ashley L. Devonshire https://orcid.org/0000-0002-6341-5794

Sandra Andorf https://orcid.org/0000-0002-3093-2568

Leah C. Kottyan https://orcid.org/0000-0003-3979-2220

Rajesh Kumar https://orcid.org/0000-0002-1962-7108

REFERENCES

1. Sicherer SH, Munoz-Furlong A, Godbold JH, Sampson HA. US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. *J Allergy Clin Immunol.* 2010;125(6):1322-1326.

2. Togias A, Cooper SF, Acebal ML, et al. Addendum guidelines for the prevention of peanut allergy in the United States: report of the National Institute of Allergy and Infectious Diseases-sponsored expert panel. *J Allergy Clin Immunol.* 2017;139(1):29-44.

3. Du Toit G, Roberts G, Sayre PH, et al. Randomized trial of peanut consumption in infants at risk for peanut allergy. *N Engl J Med.* 2015;372(9):803-813.

4. Brough HA, Cousins DJ, Munteanu A, et al. IL-9 is a key component of memory TH cell peanut-specific responses from children with peanut allergy. *J Allergy Clin Immunol.* 2014;134(6):1329-38 e10.

5. Do AN, Watson CT, Cohain AT, et al. Dual transcriptomic and epigenomic study of reaction severity in peanut-allergic children. *J Allergy Clin Immunol.* 2020;145(4):1219-1230.

6. Martino D, Neeland M, Dang T, et al. Epigenetic dysregulation of naive CD4+ T-cell activation genes in childhood food allergy. *Nat Commun.* 2018;9(1):3308.

7. Martino DJ, Bosco A, McKenna KL, et al. T-cell activation genes differentially expressed at birth in CD4+ T-cells from children who develop IgE food allergy. *Allergy.* 2012;67(2):191-200.

8. Watson CT, Cohain AT, Griffin RS, et al. Integrative transcriptomic analysis reveals key drivers of acute peanut allergic reactions. *Nat Commun.* 2017;8(1):1943.

9. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.

10. UniProt C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019;47(D1):D506-D515.

11. Gene [Internet]. 2004 [cited 12/14/2020]. https://www.ncbi.nlm.nih.gov/gene/

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.