Lee et al. World Allergy Organization Journal (2022) 15:100631
http://doi.org/10.1016/j.waojou.2022.100631

WORLD ALLERGY ORGANIZATION JOURNAL

Open Access

Identifying gene network patterns and associated cellular immune responses in children with or without nut allergy

Khui Hung Lee PhD[a], Anthony Bosco PhD[b], Michael O'Sullivan MBBS[c], Yong Song PhD[d], Jessica Metcalfe PhD[c], Kan Yu MSc(CompSc)Res[e], Benjamin J. Mullins PhD[a], Richard Loh MBBS[c] and Guicheng Zhang PhD[a,f,g]*

ABSTRACT

Background: Although evidence suggests that the immune system plays a key role in the pathophysiology of nut allergy, the precise immunological mechanisms of nut allergy have not been systematically investigated. The aim of the present study was to identify gene network patterns and associated cellular immune responses in children with or without nut allergy.

Methods: Transcriptome profiling of whole blood cells was compared between children with and without nut allergy. Three genes were selected to be validated on a larger cohort of samples ($n = 86$) by reverse transcription-polymerase chain reactions (RT-qPCR). The composition of immune cells was inferred from the transcriptomic data using the CIBERSORTx algorithm. A co-expression network was constructed employing weighted gene co-expression network analysis (WGCNA) on the top 5000 most variable transcripts. The modules were interrogated with pathway analysis tools (InnateDB) and correlated with clinical phenotypes and cellular immune responses.

Results: Proportions of neutrophils were positively correlated and $CD4^+$ T-cells and regulatory T-cells (Tregs) were negatively correlated with modules of nut allergy. We also identified 2 upregulated genes, namely *Interferon Induced With Helicase C Domain 1 (IFIH1)*, *DNA damage-regulated autophagy modulator 1 (DRAM1)* and a downregulated gene *Zinc Finger Protein 512B (ZNF512B)* as hub genes for nut allergy. Further pathway analysis showed enrichment of type 1 interferon signalling in nut allergy.

Conclusions: Our findings suggest that upregulation of type 1 interferon signalling and neutrophil responses and downregulation of $CD4^+$ T-cells and Tregs are features of the pathogenesis of nut allergy.

Keywords: RNA sequencing, Food allergy, Nut allergy, RT-qPCR, Biomarker, WGCNA

[a]School of Public Health, Curtin University of Technology, Bentley, 6102, Western Australia, Australia
*Corresponding author. School of Public Health, Curtin University of Technology, Kent St, Bentley, 6102, Western Australia, Australia. E-mail: brad.zhang@curtin.edu.au
Full list of author information is available at the end of the article

http://doi.org/10.1016/j.waojou.2022.100631

Received 22 February 2021; Received in revised from 23 November 2021; Accepted 18 January 2022
Online publication date xxx

1939-4551/© 2022 The Authors. Published by Elsevier Inc. on behalf of World Allergy Organization. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

INTRODUCTION

Today, food allergy affects up to 11% of infants and 3.8% of children, with progressively increasing prevalence over the last few decades.[1] Children with food allergy are at risk for potentially life-threatening allergic reactions including breathing difficulties, swollen tongue, throat tightness and wheezing, particularly when exposed to nut allergens.[2] Moreover, nut allergy tends to be persistent over the full duration of a person's lifetime and can cause anaphylactic reactions.[2,3] With the exception of treatment of peanut allergy,[4] immunotherapy for desensitizing in food allergy is still not recommended for clinical practice.[5] Hence, the best management for food allergies is to strictly avoid specific food allergens. As such, it is crucially essential to elucidate the fundamental mechanisms of food allergy to help develop effective prevention and treatment for this condition in children.

Food allergy is mechanistically characterized by the development of an overactive immune response to an otherwise harmless allergen, resulting in a T helper 2 (Th2) polarized cytokine response to the allergen and the production of an Immunoglobulin E (IgE) antibody response. Upon recognition of food allergens by antigen presenting cells, mainly dendritic cells (DCs), naïve T-cells are instructed to differentiate into allergen-specific Th2 cells,[6] which are characterized by the expression of type-2 cytokines including interleukin (IL)-4, IL-5, and IL-13.[7] These cytokines then mediate immune responses to food allergens by supporting B-cell proliferation,[8-10] promoting IgE isotype switching, and inducing the activation of mast cells and basophils.[11-14] Given that immune responses to food allergens are complex and involve multiple cell populations, we reasoned that an unbiased systems biology approach could reveal a unique level of insight into the underlying immunological mechanisms. RNA sequencing has been widely used for investigating the pathogenesis of complex human diseases,[15] because it enables the systematic study of the molecular states that underpin pathogenic states.[16] RNA sequencing data can be analysed with systems biology methods such as weighted gene co-expression network analysis (WGCNA), to elucidate the global architecture of the gene expression program and unmask systems-level properties of the biological systems under study.[17] In the current study, we applied an RNA sequencing approach coupled with cell deconvolution and weighted gene co-expression network analysis to identify gene network patterns and associated cellular immune responses in children with or without nut allergy, in order to better understand immunological mechanisms of nut allergy.

METHODS

This study was approved by the Curtin Human Research Ethics Committee (Curtin HREC HRE2016-0178) and Child and Adolescent Health Service (CAHS) Human Research Ethics Committee (CAHS HREC 2016046 EP) and conducted in accordance with the National Health and Medical Research Council National Statement on Ethical Conduct in Human Research. Written informed consent was obtained from all parents on behalf of participants. Our study included a total of 92 individuals (84 children with a nut allergy and 8 children without a nut allergy). Children (aged 1–16 years old) with suspected or diagnosed allergies to tree nuts and peanut were recruited from the outpatient clinic of the Immunology Department. Children that were older than 16 years were excluded. Diagnoses of nut allergy were determined by an immunologist on the day of recruitment based on the clinical outcomes of the patients. Nut allergy was defined by immediate symptoms (1–2 h) after nut ingestion combined with either failed oral food challenge or with positive skin prick test wheal diameter$\geq$3 mm to nuts). In contrast, a negative skin prick test and/or a passed oral food challenge for any food allergens defined the subject as children without a nut allergy. Venous blood was collected into PAXgene RNA blood tubes (PreAnalytiX, Qiagen, Hilden, Germany) by a trained phlebotomist.

RNA extraction

Total RNA samples were extracted using PAXgene Blood RNA Kit according to the manufacturer's instructions (Qiagen). Briefly, samples were incubated at room temperature for at least 2 h before RNA extraction. After washing with RNase-free water, the pellet was resuspended in 350 μl

resuspension buffer and incubated with 300 µl binding buffer and 40 µl proteinase K for 10 min at 55 °C. The lysate was transferred into a PAXgene shredder spin column and centrifuged for 3 min at 14 000 g. The flow-through fraction was mixed with 350 µl ethanol and transferred to a PAXgene RNA spin column. After washing the column with washing buffer 1, samples were incubated with 80 µl DNase 1 of (RNFD) incubation mix at room temperature for 15 min. The sample was washed several times with washing buffer before RNA was eluted with 40 µl elution buffer and stored at −80 °C. Total RNA concentration and purity were assessed by determining the A260/280 and A260/230 ratios, respectively (NanoDrop).

Library preparation, RNA sequencing and quality control (QC)

The library preparation, sequencing read, QC and read alignment of 30 samples were performed at the Australian Genome Research Facility (AGRF). Briefly, Ribosomal RNA was removed using the Ribo-Zero Gold rRNA removal kit prior to preparation of the cDNA libraries. Library was prepared using TruSeq stranded mRNA Library Prep Kit (Illumina) as per the manufacturer's instructions. The process included purification of mRNA purification via oligo (dT) beads, fragmentation of mRNA with divalent cations and heat, as well as the synthesis of first strand cDNA and 2nd strand cDNA. cDNA libraries were prepared by DNA fragment end repair, 3′ adenylation of DNA fragments, sequence adaptor ligation and amplification of library via PCR. In total, 30 cDNA libraries in 2 separate batches (13 samples and 17 samples) were constructed for sequencing. Samples were then sequenced on Illumina HiSeq 2500 platform with a 100 bp single end read.

Post run processing, including demultiplexing and generation of Fastq files was performed the Illumina bcl2fastq 2.20.0.422 pipeline. The quality of Fastq data were assessed for quality check using FastQC. The data were also screened for the presence of any Illumina adapter/overrepresented sequences and cross-species contamination.

The cleaned sequence reads were then aligned against the *Homo sapiens* genome (Build version HG38) by the Tophat aligner (v2.1.1). The transcripts were assembled with the Stringtie toolv1.3.3 (http://ccb.jhu.edu/software/stringtie/) utilising the reads alignment with hg38 and reference annotation based assembly option (RABT) generating assembly for known and potentially novel transcripts. The reads corresponding to each gene were summarized using the featureCounts v1.5.3 utility of the subread package (http://subread.sourceforge.net/).

Gene expression data of RNA sequencing

The raw gene read counts were converted into counts per million by using edgeR package of R. Genes with less than 20 counts in the sample with the lowest sequencing depth were excluded to reduce potential noise. Counts were then converted to log2-counts-per-million (logCPM) with precision weights, based on voom method using Limma package of R prior to statistical analysis.[18] Empirical Bayes statistics were applied to compute model statistics and calculate log-fold change (log2FC) of differential expressed genes. The cut-off criteria for significant analysis was: $|\log2FC|>0.3$ and $p < 0.05$. Age, sex and batch effect were considered as potential confounders.

Reverse transcription-polymerase chain reactions (RT-qPCR)

RNA was reverse transcribed to cDNA using the QuantiTect Reverse Transcription Kit (Qiagen, Germany). One µg of the total RNA from each sample was mixed with gDNA wipeout buffer and RNase-free water to prepare annealing mix, followed by incubation at 42 °C for 2 min. The annealing mix was stored on ice until reverse transcription master mix was added. The final mixture was incubated for 30 min at 42 °C and 3 min at 95 °C to inactivate reverse transcription reaction before storing at −80 °C.

RNA-specific primers for reverse transcription-polymerase chain reactions (RT-qPCR) were designed on the specific region of the genes (Supplemental Table 1) (Supplemental Figure 1). Quantitative PCR was performed on a ViiA7 ™ Real-Time PCR System (Thermo Fisher Scientific). qPCR thermal cycling was set as follows: initial denaturation at 95 °C for 5 min, followed by 40 cycles at 95 °C for 10 s and 55 °C for 30 s and a final extension at 95 °C for 15 s, 60 °C for 1min and 95 °C for 15 s 18s rRNA was measured as the

housekeeping gene for qPCR. The fold change in the expression of targeted genes was computed using the ΔΔCt method.[19]

Quantification of immune cells

The RNA-Seq data was then used to estimate the cellular composition of the whole blood samples using CIBERSORTx.[20] CIBERSORTx is a deconvolution algorithm, which can estimate the abundance of different cell types from RNA-Seq data based on cell type specific reference gene expression profiles. CIBERSORTx provides 22 types of functionally defined human immune cells (LM22 signature matrix) as a reference, and we focused on proportions of 12 human immune cells (B cells, $CD8^+$ T-cells, $CD4^+$ T-cells, follicular helper T-cells, Tregs, gamma delta T-cells, NK cells, monocytes, dendritic cells, mast cells, eosinophils and neutrophils). The cell type proportions in children with and without nut allergy were compared employing an independent *t*-test, and were correlated with coexpression modules derived from WGCNA.

Gene co-expression network construction

Genes with low variation from RNA sequencing data were then filtered out using a coefficient of variation cutoff (CV $>$ 0.05). This resulted in the identification of 9462 stably expressed genes in all the samples. Top 5000 genes with high coefficient of variation values were selected for additional analysis.

The selected genes were then used to construct a co-expression network by using the WGCNA.[21] First, the goodSamplesGenes function was used to filter genes with many mixing values. Then, hierarchical clustering of samples was analysed using the hclust function. The soft-thresholding power was then calculated by using a scale-free topology and the soft-thresholding power was determined based on a scale-free topology index (R_2) of 0.85 (Supplemental Figure 2). According to Zhang and Horvath,[21] the gene distribution conformed to a scale-free network if R2 value (model fitting index) is close to 1. Next, the selected soft-thresholding power of 8 was used to calculate the adjacency matrix (correlation strength) between the corresponding pair of genes. The adjacency matrix was then transformed into a Topological Overlap Matrix (TOM) to measure the connectivity of network of the genes and to remove any spurious association. TOM was later performed average linkage hierarchical clustering in order to identify modules of highly co-expressed genes. Network modules were subsequently identified using a dynamic tree cut algorithm with a minimum cluster size of 30 and merge cut height of 0.25 (default). The minimum size of the modules was set to 10 to ensure that small as well as large modules are detected. Subsequently, highly co-occurred genes were merged into modules and these modules later assigned to different colours for visualization.

Module trait relationship construction

Module eigengenes were used to perform principal component analysis of the expression matrix from each gene module. The correlation between module eigengenes (ME), proportions of immune cells, phenotype as well as demographics traits such as age, gender, and batch effect were calculated using Pearson correlation coefficient. Gene modules, which have p-value $<$0.05, were identified to have significant correlations with nut allergy and these modules were selected for further analysis.

Hub genes selection and visualization

Next, an intramodular analysis was performed to determine the hub genes in the selected modules by summing the connection strengths with other module genes. Hub genes were defined based on the standard cut off of module membership (MM) $>$ 0.8 and gene significance (GS) $>$ 0.5. Module membership (MM) is defined as the correlation between expression profile of a gene and each module eigengene. MM takes values between 0 and 1 and tells "how well a gene belongs to a module"; hub genes have an MM value closer to 1. Gene significance (GS) is defined as the correlation between gene expression and the outcome, nut allergy. The correlation between a module and nut allergy status is in fact a correlation between the module eigengene and nut allergy status. Hub genes of the significant modules were then visualized using Cytoscape v3.8.0.[22]

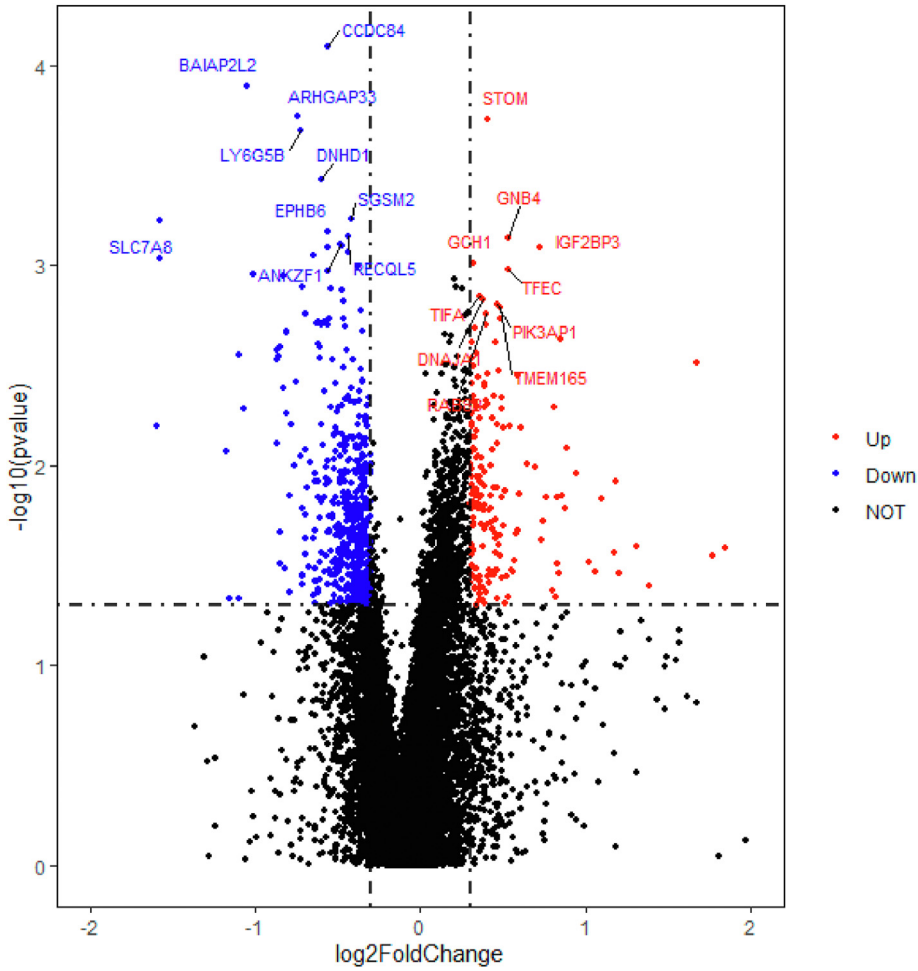


Fig. 1 The volcano plot for differentially expressed genes. The horizontal axis represents the log2 fold change, and the vertical axis represents the -log10 (pvalue). Red denotes upregulated genes with log2FC > 0.3 and p-value<0.05 while blue denotes downregulated genes with log2FC < −0.3 and p-value<0.05. The top 15 differentially expressed genes are labelled.

Functional enrichment analysis

The biological function and associated pathways enriched in each module was characterized by employing gene ontology (GO) and using InnateDB.[23]

RESULT

Study population

A total of 92 whole blood samples (30 whole blood samples for RNA sequencing and 86 whole blood samples for RT-qPCR) were obtained from children with nut allergy and children without nut allergy (Supplemental Table 2). The 2 study cohorts (RNA sequencing and RT-qPCR cohorts) were well-balanced ($p > 0.05$) for all the major demographic characteristics. Around sixty three percent of the subjects were boys, with the median age for RNA sequencing cohort and RT-qPCR cohort of 10.5 years and 9.3 years, respectively.

Gene expression profiling of whole blood in children with or without nut allergy

We utilized RNA sequencing to compare gene expression patterns of whole blood samples of children with nut allergy (n = 23) to whole blood samples of children without nut allergy (n = 7). RNA sequencing produced a total of 520 million reads with an average of 17 million mapped reads

Module colours	No. of genes	Associated biological processes	Adjusted p-value
Black	190	malonyl-CoA biosynthetic process	0.020
Blue	643	innate immune response	<0.001
Brown	384	negative regulation of dendritic cell differentiation	0.032
Green	266	positive regulation of humoral immune response	0.122
Greenyellow	90	chronic inflammatory response to antigenic stimulus	0.003
Magenta	132	establishment of T-cell polarity	0.037
Pink	187	gene expression	0.001
Purple	116	regulation of cytokine secretion	0.015
Red	229	IMP biosynthetic process	0.174
Tan	71	type 1 interferon signalling pathway	<0.001
Turquoise	1807	transcription, DNA-templated	<0.001
Yellow	349	protein import into peroxisome matrix, translocation	0.211

Table 1. Modules and associated biological processes.

per sample. A total of 12 523 genes were detected in both samples with and without nut allergy. Differential expression analysis adjusted for age, sex, and batch effect identified 184 upregulated genes (log2FC > 0.3) and 490 downregulated genes (log2FC < -0.3) (Fig. 1). However, it is noteworthy that these results were not significant after multiple corrections.

Validation of RNA sequencing results with RT-qPCR

We selected 3 genes, which were *G Protein Subunit Beta 4* (*GNB4*), *Golgi Associated Kinase 1B (GASK1B)* and *LysM Domain Containing 2* (*LYSMD2)*, for validation by RT-qPCR on a larger cohort of samples according to the following criteria: 1) dysregulated genes based on RNA-seq; 2) high abundance based on logCPM of RNA-seq data; 3) these genes are particularly interesting, associating with infection and inflammation based on literature. We observed the results of RT-qPCR were significantly correlated with RNA sequencing ($p < 0.01$) (Supplemental Figure 3). These results suggested that our RNA-Seq data was reliable.

Gene co-expression network construction

We constructed a coexpression network on the top 5000 most variable genes as described in the methods. Through WGCNA, we identified 12 modules of co-expressed genes, and each module comprised between 71 and 1807 genes (Table 1). Among the highly variable genes, only 536 genes (10%) were not assigned to any module, and these genes were clustered into the grey module as per default. Pathways analysis of the modules with InnateDb demonstrated that the modules were significantly enriched for coherent biological functions (Table 1).

Module trait relationship construction

The module eigengenes were further compared between children with and without nut allergy using module trait association analysis to identify the nut allergy-associated modules. Four modules out of 12 modules were identified to be significantly associated with nut allergy (Fig. 2), which included tan module ($r = 0.43$, $p = 0.03$), purple module ($r = 0.4$, $p = 0.04$), green module ($r = -0.48$, $p = 0.01$) and grey module ($r = -0.56$, $p = 0.003$). However, we did not try to find a hub

gene for grey module since genes in this module were not co-expressed. It was also observed that greenyellow module was marginally associated with nut allergy ($r = 0.4$, $p = 0.05$) and there was no hub gene identified in this module. The upregulated modules (tan and purple modules), were negatively correlated to $CD4^+$ T-cells and positively correlated to neutrophils. In addition, the purple module was also negatively correlated to Tregs. In contrast, the downregulated module (green module) was positively correlated to $CD4^+$ T-cells and Tregs as well as negatively correlated to neutrophils. However, two modules that were not significantly associated with nut allergy (red and greenyellow modules) were also observed to be strongly correlated with neutrophils.

The proportions of the 12 human immune cells in children with or without nut allergy are shown in Fig. 3. The cellular composition was dominated by neutrophils, Monocytes, $CD8^+$ T-cells, $CD4^+$ T-cells, B cells and Tregs. Among all the cell types, $CD4^+$ T-cells and macrophages were observed to have a significant difference between children with and without nut allergy.

Significant modules and associated biological process

The tan module was revealed to have the highest positive correlation with nut allergy. In the tan module, 71 genes were identified to be correlated with nut allergy and these genes were found to be involved in type 1 interferon signalling pathway. Particularly, *Interferon Induced With Helicase C Domain 1 (IFIH1)* was identified as a hub gene of the tan module (GeneSignificance>0.5 and Module Membership>0.8, Fig. 4).

The purple module was also positively correlated with nut allergy. In the purple module, 116 genes were identified to be correlated with nut allergy and these genes were found to be involved in the regulation of cytokine secretion. Particularly, *DNA damage-regulated autophagy modulator 1 (DRAM1)* was identified as a hub gene within the purple module (GeneSignificance>0.5 and Module Membership>0.8, Fig. 4).

In contrast, the green module was found to have a negative correlation with nut allergy. In the green module, 266 genes were identified to be correlated with nut allergy and these genes were found to be involved in the positive regulation of humoral immune response. Particularly, *Zinc Finger Protein 512B (ZNF512B)* was identified as a hub gene of the green module (GeneSignificance>0.5 and Module Membership>0.8, Fig. 4).

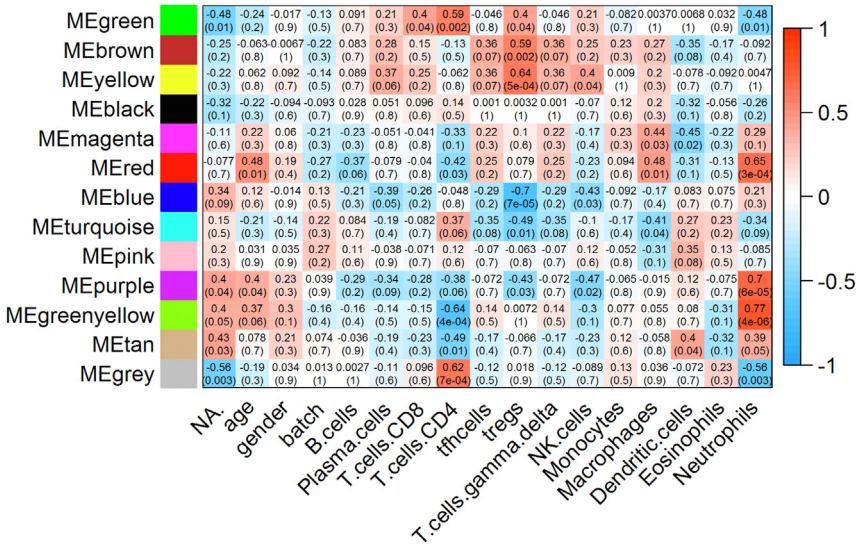


Fig. 2 Module-trait associations. Each row corresponds to a module eigengene (ME) while each column corresponds to either phenotype (NA: nut allergy) or demographic traits such as age and gender. Each cell contains the corresponding correlation coefficient (display at the top of the cell) and corresponding p-values for each module (display at the bottom of the cells within parentheses). Blue and red colours of the spectrum on the right denote low and high correlation, respectively.

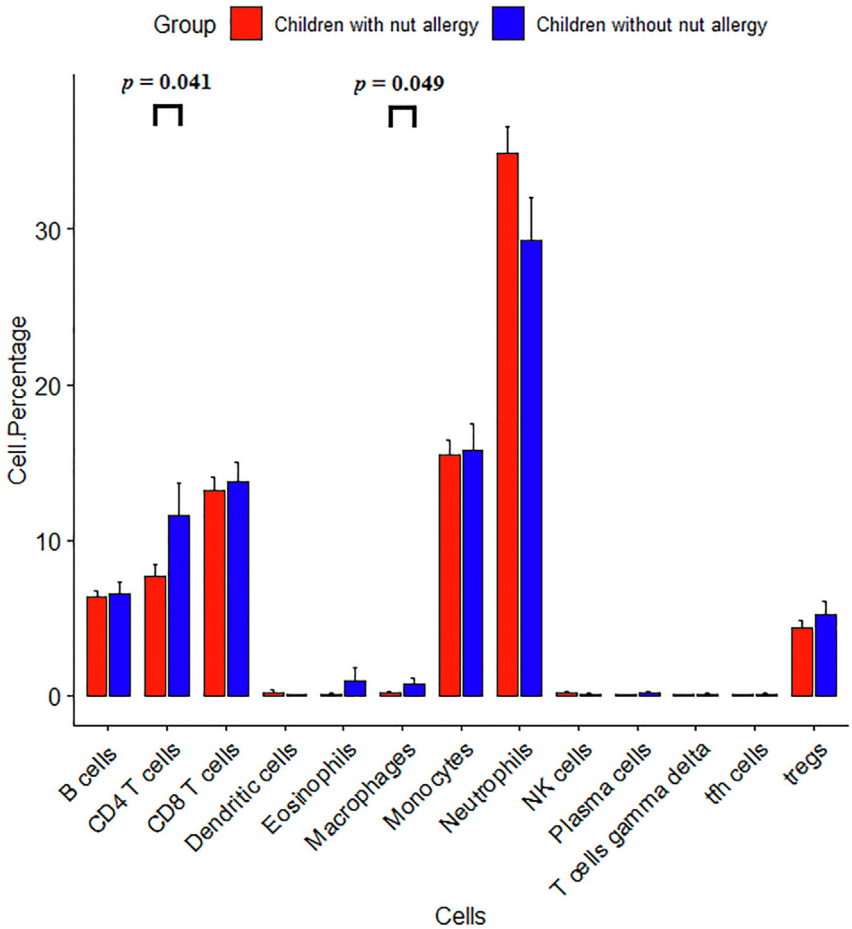


Fig. 3 Differential immune cell type expression was observed between children with and without nut allergy. Blue colour represents children without nut allergy, while red colour represents children with nut allergy.

DISCUSSION

Food allergy can cause severe, life-threatening anaphylactic reactions. Here, we profiled the transcriptome of immune cells to provide a window into the regulation of immune function in the context of food allergy and anaphylaxis. The application of WGCNA identifies the global connectivity structure of the gene expression program, and unmasks systems-level features of the underlying biology.[24,25] Herein, we analysed gene expression profiles obtained from whole blood transcriptome of children with and without nut allergy using WGCNA to identify molecular and cellular immune features associated with nut allergy**.** In the present study, we identified upregulation of the type I interferon production (tan) and cytokine production (purple) modules and downregulation of the humoral immune responses (green) module in children with nut allergy. These changes were positively correlated with proportions of neutrophils and negatively correlated with propotions of CD4 T cell/Treg. The upregulated modules were characterized by hub genes, namely *IFIH1*, *DRAM1,* which were involved in type 1 interferon signalling pathway and the regulation of cytokine secretion while the downregulated module was characterized by a hub gene, namely *ZNF512B*, which was involved in the positive regulation of humoral immune response. Together, our data suggest that the pathogenesis of nut allergy is associated with the upregulation of type 1 interferon and neutrophil responses, and downregulation of CD4 T cell/Treg responses. Consistent with our studies, previous studies have also identified the enrichment of type 1 interferons in subjects with allergic diseases compared to control.[26,27] The exposure to allergens activates toll like receptors, which in turn mediates the production of type 1 interferons, inducing phosphorylation of STAT1

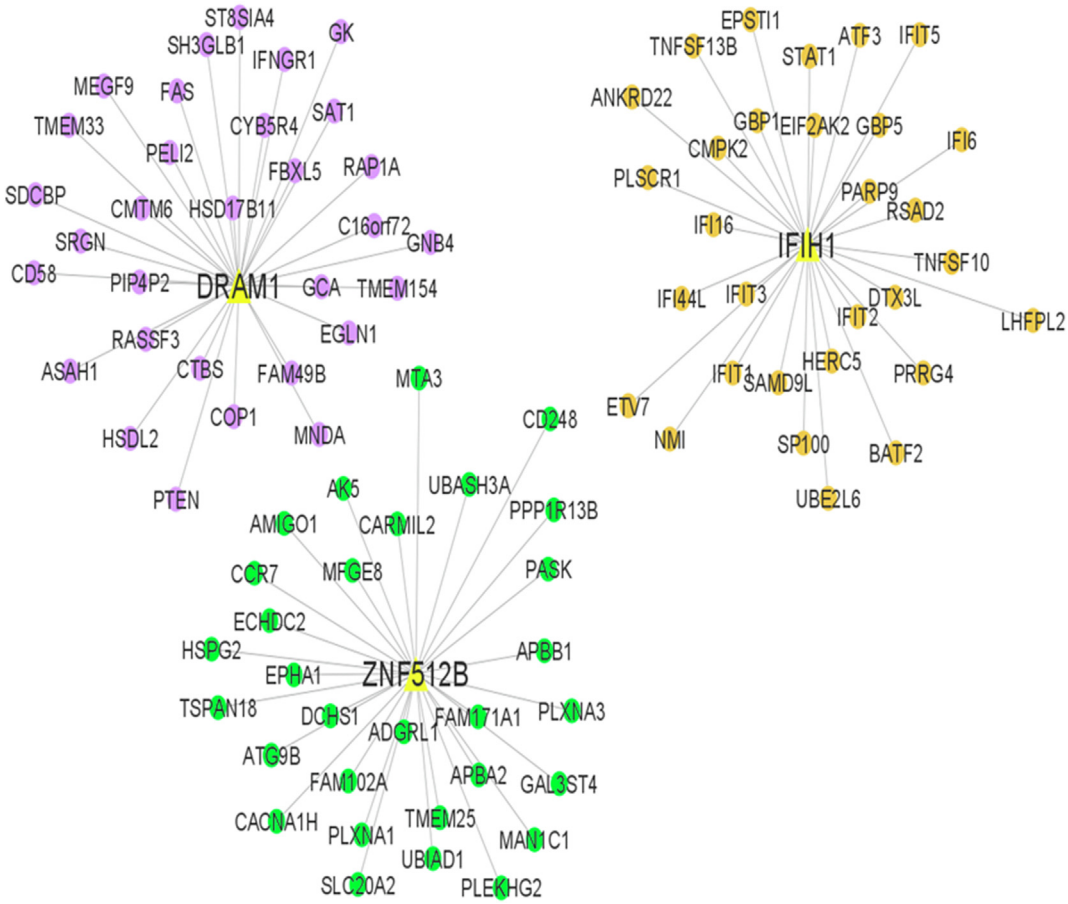


Fig. 4 Co-expression network of top 30 genes in 3 distinct modules (purple, tan and green), hub gene is indicated with yellow triangle shapes and other highly correlated genes are in round shapes and colour coded according to the module colour.

and inducing expression of pro-inflammatory cytokines.[27]

A limitation of gene expression profiles obtained from whole blood transcriptome is that the data are potentially confounded by variations in cellular composition. The integration of the co-expression network and cell deconvolution approaches allowed us to link co-expression patterns within each module with specific immune cells. $CD4^+$ T-cells and Tregs are already the focus of investigation in studies of food allergy.[28-30] Mouse models of food allergy demonstrated a critical role for CD4+Foxp3+ Treg cells in suppressing food allergy. These studies observed a significant increase in the percentage of CD4+Foxp3+ Treg cells in ovalbumin sensitized mice with mucosal tolerance induction compared to the intolerant group.[31-33] In addition, CD4+Foxp3+ Treg cells were found to inhibit the activation of dendritic cells, mast cells, basophils, and eosinophils, suppress the production of allergen-specific IgE, inhibit Th1, Th2, and Th17 migration patterns and effector functions as well as promote the secretion of IgG4.[34] Consistent with these studies, we also observed a downregulation of $CD4^+$ T-cell and Treg responses in the pathogenesis of nut allergy.

Our analysis also pointed to a possible contribution of neutrophils to the pathogenesis of nut allergy. However, we also observed a significant association of neutrophils with other modules that were not associated with nut allergy. These contrasting results could be due to the heterogeneity of neutrophils as previous studies have revealed variations in phenotype and functions of neutrophils in the development of allergic diseases. On the one hand, neutrophil Fc gamma receptors, FcγRIIIA and FcγRIV as well as several markers of neutrophil activation, *S100A8, S100A9, TLR4,*

TREM1, *S100A9*[35] and *interleukin 17A*,[36] are shown to induce acute anaphylaxis.[37,38] On the other hand, neutrophils are revealed to be a source of anti-inflammatory and immunomodulatory cytokines, such as interleukin-10.[39] Moreover, neutrophils can suppress NF-kB activation in the macrophage, which in turn reduces the expression of pro-inflammatory cytokines (tumor necrosis factor, chemokine ligand 8 and interleukin-6), leading to the resolution of inflammation.[40] Therefore our findings require further research on neutrophils-related gene networks with nut allergy.

Our study has several limitations. First, the sample sizes used for RNA sequencing are small (n = 30), but gene expression levels of three selected genes are validated using RT-qPCR in a larger sample (n = 86). Second, it is not known if the observed patterns of gene expression are related to the mechanisms that drive nut allergy or alternatively are reacting to nut allergy due to the cross-sectional nature of our study. Third, our study only focused on nut allergy, and therefore it is not known if the mechanisms we identified are relevant to the pathogenesis of other food allergies. Fourth, flow cytometry-based assays to target multiple cell populations were not feasible due to the volume restriction on blood collection from children. However, we have applied an unbiased deconvolution approach to infer the proportions of 12 human cells in whole blood transcriptome. Fifth, we were unable to deconvolve the co-linearity between the module, the phenotype and the cell counts. Therefore, the module train correlations were all explained by cell compositions. Lastly, our study does not define whether changes in cellular proportions precede or follow transcriptomic changes. Further experiments are needed to identify the cellular origin of the gene expression signals associated with nut allergy.

Our study suggests immunological mechanisms of nut allergy can possibly be explained by changes in cell composition. This is related to type 1 interferon signalling, regulation of cytokine responses and the humoral response. Our findings represent plausible pathways for further mechanistic investigation.

ABBREVIATIONS

DCs: Dendritic cells; *DRAM1: DNA damage-regulated autophagy modulator 1; GASK1B: Golgi Associated Kinase 1B; GNB4: G Protein Subunit Beta 4;* GS: Gene significance; *IFIH1: Interferon Induced With Helicase C Domain 1;* IgE: Immunoglobulin E; IL: Interleukin; *LYSMD2: LysM Domain Containing 2;* log2FC: Log-fold change; logCPM: Log2-counts-per-million; MM: Module membership; RT-qPCR: Reverse transcription-polymerase chain reactions; Th: T helper; Tregs: Regulatory T-cells; WGCNA: Weighted gene co-expression network analysis; *ZNF512B: Zinc Finger Protein 512B*.

Author contributions

K.H.L. prepared the manuscript and figures. K.H.L. carried out the experiments and data analysis with assistance from Y.S., K.Y., G.Z. and A.B. Y.S. designed primers and laboratory protocols used in the study. M.O, J.M., B.J.M. and R.L. helped to supervise the project. All authors conceived the manuscript and revised the manuscript. G.Z. coordinated all aspects of the study.

Ethics approval and consent to participate

This study was approved by the Curtin Human Research Ethics Committee (Curtin HREC HRE2016-0178) and Child and Adolescent Health Service (CAHS) Human Research Ethics Committee (CAHS HREC 2016046 EP) and conducted in accordance with the National Health and Medical Research Council National Statement on Ethical Conduct in Human Research. Written informed consent was obtained from all parents on behalf of participants.

Consent for publication

All authors agreed to publish the work.

Funding

This research was funded by Telethon-Perth Children's Hospital Research Fund (TPCHRF) and the Australian Government Research Training Program (RTP) Stipend Scholarship.

Availability of data

The data that support the findings of this study are available from the corresponding author, G.Z., upon reasonable request.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments
We sincerely thank all the participants and their parents for their participation in this study, as well as the staff of the Immunology Department, Perth Children's Hospital.

Appendix A. Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.waojou.2022.100631.

Author details
[a]School of Public Health, Curtin University of Technology, Bentley, 6102, Western Australia, Australia. [b]Telethon Kids Institute, University of Western Australia, Crawley, 6000, Western Australia, Australia. [c]Department of Immunology, Perth Children's Hospital, Nedlands, 6009, Western Australia, Australia. [d]The Menzies Institute for Medical Research, University of Tasmania, Hobart, 7000, Tasmania, Australia. [e]School of Science, Edith Cowan University, Joondalup, 6027, Western Australia, Australia. [f]Curtin Health Innovation Research Institute, Curtin University, Bentley, 6102, Western Australia, Australia. [g]Infection and Immunity, School of Biomedical Sciences, University of Western Australia, Crawley, 6000, Western Australia, Australia.

REFERENCES

1. Peters RL, Koplin JJ, Gurrin LC, et al. The prevalence of food allergy and other allergic diseases in early childhood in a population-based study: HealthNuts age 4-year follow-up. *J Allergy Clin Immunol*. Jul 2017;140(1):145–153. e8. https://doi.org/10.1016/j.jaci.2017.02.019.
2. Iweala OI, Choudhary SK, Commins SP. Food allergy. *Curr Gastroenterol Rep*. Apr 5 2018;20(5):17. https://doi.org/10.1007/s11894-018-0624-y.
3. Renz H, Allen KJ, Sicherer SH, et al. Food allergy. *Nat Rev Dis Prim*. Jan 4 2018;4:17098. https://doi.org/10.1038/nrdp.2017.98.
4. Hsiao KC, Ponsonby AL, Axelrad C, Pitkin S, Tang MLK, Team PS. Long-term clinical and immunological effects of probiotic and peanut oral immunotherapy after treatment cessation: 4-year follow-up of a randomised, double-blind, placebo-controlled trial. *The Lancet Child & adolescent health*. Oct 2017;1(2):97–105. https://doi.org/10.1016/S2352-4642(17)30041-X.
5. Burks AW, Sampson HA, Plaut M, Lack G, Akdis CA. Treatment for food allergy. *J Allergy Clin Immunol*. Jan 2018;141(1):1–9. https://doi.org/10.1016/j.jaci.2017.11.004.
6. Kapsenberg ML. Dendritic-cell control of pathogen-driven T-cell polarization. *Nat Rev Immunol*. Dec 2003;3(12):984–993. https://doi.org/10.1038/nri1246.
7. Turcanu V, Maleki SJ, Lack G. Characterization of lymphocyte responses to peanuts in normal children, peanut-allergic children, and allergic children who acquired tolerance to peanuts. *J Clin Invest*. Apr 2003;111(7):1065–1072. https://doi.org/10.1172/JCI16142.
8. Croote D, Darmanis S, Nadeau KC, Quake SR. High-affinity allergen-specific human antibodies cloned from single IgE B cell transcriptomes. *Science*. Dec 14 2018;362(6420):1306–1309. https://doi.org/10.1126/science.aau2599.
9. Kiewiet MBG, van Esch B, Garssen J, Faas MM, de Vos P. Partially hydrolyzed whey proteins prevent clinical symptoms in a cow's milk allergy mouse model and enhance regulatory T and B cell frequencies. *Mol Nutr Food Res*. Nov 2017;61:11. https://doi.org/10.1002/mnfr.201700340.
10. Hoh RA, Joshi SA, Lee JY, et al. Origins and clonal convergence of gastrointestinal IgE(+) B cells in human peanut allergy. *Science immunology*. Mar 6 2020;5:45. https://doi.org/10.1126/sciimmunol.aay4209.
11. Benede S, Berin MC. Mast cell heterogeneity underlies different manifestations of food allergy in mice. *PLoS One*. 2018;13(1), e0190453. https://doi.org/10.1371/journal.pone.0190453.
12. Chen CY, Lee JB, Liu B, et al. Induction of interleukin-9-producing mucosal mast cells promotes susceptibility to IgE-mediated experimental food allergy. *Immunity*. Oct 20 2015;43(4):788–802. https://doi.org/10.1016/j.immuni.2015.08.020.
13. Polukort SH, Rovatti J, Carlson L, et al. IL-10 enhances IgE-mediated mast cell responses and is essential for the development of experimental food allergy in IL-10-deficient mice. *J Immunol*. Jun 15 2016;196(12):4865–4876. https://doi.org/10.4049/jimmunol.1600066.
14. Hussain M, Borcard L, Walsh KP, et al. Basophil-derived IL-4 promotes epicutaneous antigen sensitization concomitant with the development of food allergy. *J Allergy Clin Immunol*. Jan 2018;141(1):223–234. e5. https://doi.org/10.1016/j.jaci.2017.02.035.
15. Costa V, Aprile M, Esposito R, Ciccodicola A. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. *Eur J Hum Genet : EJHG*. Feb 2013;21(2):134–142. https://doi.org/10.1038/ejhg.2012.129.
16. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. Jan 2009;10(1):57–63. https://doi.org/10.1038/nrg2484.
17. Read JF, Bosco A. Decoding susceptibility to respiratory viral infections and asthma inception in children. *Int J Mol Sci*. Sep 2 2020;21:17. https://doi.org/10.3390/ijms21176372.
18. Law CW, Alhamdoosh M, Su S, et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res*. 2018;5:1408.
19. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. Oct 1 2010;26(19):2460–2461. https://doi.org/10.1093/bioinformatics/btq461.
20. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. Jul 2019;37(7):773–782. https://doi.org/10.1038/s41587-019-0114-2.
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf*. Dec 29 2008;9:559. https://doi.org/10.1186/1471-2105-9-559.
22. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. Nov 2003;13(11):2498–2504. https://doi.org/10.1101/gr.1239303.

23. Breuer K, Foroushani AK, Laird MR, et al. InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. *Nucleic Acids Res*. Jan 2013;41(Database issue):D1228–D1233. https://doi.org/10.1093/nar/gks1147.

24. Hurgobin B, de Jong E, Bosco A. Insights into respiratory disease through bioinformatics. *Respirology*. Dec 2018;23(12): 1117–1126. https://doi.org/10.1111/resp.13401.

25. Jones C, Bosco A. Using network analysis to understand severe asthma phenotypes. *Am J Respir Crit Care Med*. 2017;195(11):1409–1411. https://doi.org/10.1164/rccm.201612-2572ED.

26. Kosoy R, Agashe C, Grishin A, et al. Transcriptional profiling of egg allergy and relationship to disease phenotype. *PLoS One*. 2016;11(10), e0163831. https://doi.org/10.1371/journal.pone.0163831.

27. Diez D, Goto S, Fahy JV, et al. Network analysis identifies a putative role for the PPAR and type 1 interferon pathways in glucocorticoid actions in asthmatics. *BMC Med Genom*. Jun 19 2012;5:27. https://doi.org/10.1186/1755-8794-5-27.

28. Martino D, Neeland M, Dang T, et al. Epigenetic dysregulation of naive CD4+ T-cell activation genes in childhood food allergy. *Nat Commun*. Aug 17 2018;9(1):3308. https://doi.org/10.1038/s41467-018-05608-4.

29. 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*. Feb 2012;67(2):191–200. https://doi.org/10.1111/j.1398-9995.2011.02737.x.

30. Papatriantafyllou M. Tolerance: the origins of colonic TReg cells. *Nat Rev Immunol*. Jun 2013;13(6):394. https://doi.org/10.1038/nri3468.

31. Bae MJ, Shin HS, See HJ, Jung SY, Kwon DA, Shon DH. Baicalein induces CD4(+)Foxp3(+) T cells and enhances intestinal barrier function in a mouse model of food allergy. *Sci Rep*. Aug 26 2016;6:32225. https://doi.org/10.1038/srep32225.

32. Rupa P, Mine Y. Oral immunotherapy with immunodominant T-cell epitope peptides alleviates allergic reactions in a Balb/c mouse model of egg allergy. *Allergy*. Jan 2012;67(1):74–82. https://doi.org/10.1111/j.1398-9995.2011.02724.x.

33. Smaldini PL, Orsini Delgado ML, Fossati CA, Docena GH. Orally-induced intestinal CD4+ CD25+ FoxP3+ Treg controlled undesired responses towards oral antigens and effectively dampened food allergic reactions. *PLoS One*. 2015;10(10), e0141116. https://doi.org/10.1371/journal.pone.0141116.

34. Zhang H, Kong H, Zeng X, Guo L, Sun X, He S. Subsets of regulatory T cells and their roles in allergy. *J Transl Med*. May 12 2014;12:125. https://doi.org/10.1186/1479-5876-12-125.

35. Francis A, Bosio E, Stone SF, et al. Markers involved in innate immunity and neutrophil activation are elevated during acute human anaphylaxis: validation of a microarray study. *Journal of innate immunity*. 2019;11(1):63–73. https://doi.org/10.1159/000492301.

36. Zbikowska-Gotz M, Palgan K, Gawronska-Ukleja E, et al. Expression of IL-17A concentration and effector functions of peripheral blood neutrophils in food allergy hypersensitivity patients. *Int J Immunopathol Pharmacol*. Mar 2016;29(1):90–98. https://doi.org/10.1177/0394632015617069.

37. Gillis CM, Jonsson F, Mancardi DA, et al. Mechanisms of anaphylaxis in human low-affinity IgG receptor locus knock-in mice. *J Allergy Clin Immunol*. Apr 2017;139(4):1253–1265. https://doi.org/10.1016/j.jaci.2016.06.058. e14.

38. Jonsson F, Mancardi DA, Kita Y, et al. Mouse and human neutrophils induce anaphylaxis. *J Clin Invest*. Apr 2011;121(4): 1484–1496. https://doi.org/10.1172/JCI45232.

39. De Santo C, Arscott R, Booth S, et al. Invariant NKT cells modulate the suppressive activity of IL-10-secreting neutrophils differentiated with serum amyloid A. *Nat Immunol*. Nov 2010;11(11):1039–1046. https://doi.org/10.1038/ni.1942.

40. Marwick JA, Mills R, Kay O, et al. Neutrophils induce macrophage anti-inflammatory reprogramming by suppressing NF-kappaB activation. *Cell Death Dis*. Jun 4 2018;9(6):665. https://doi.org/10.1038/s41419-018-0710-y.

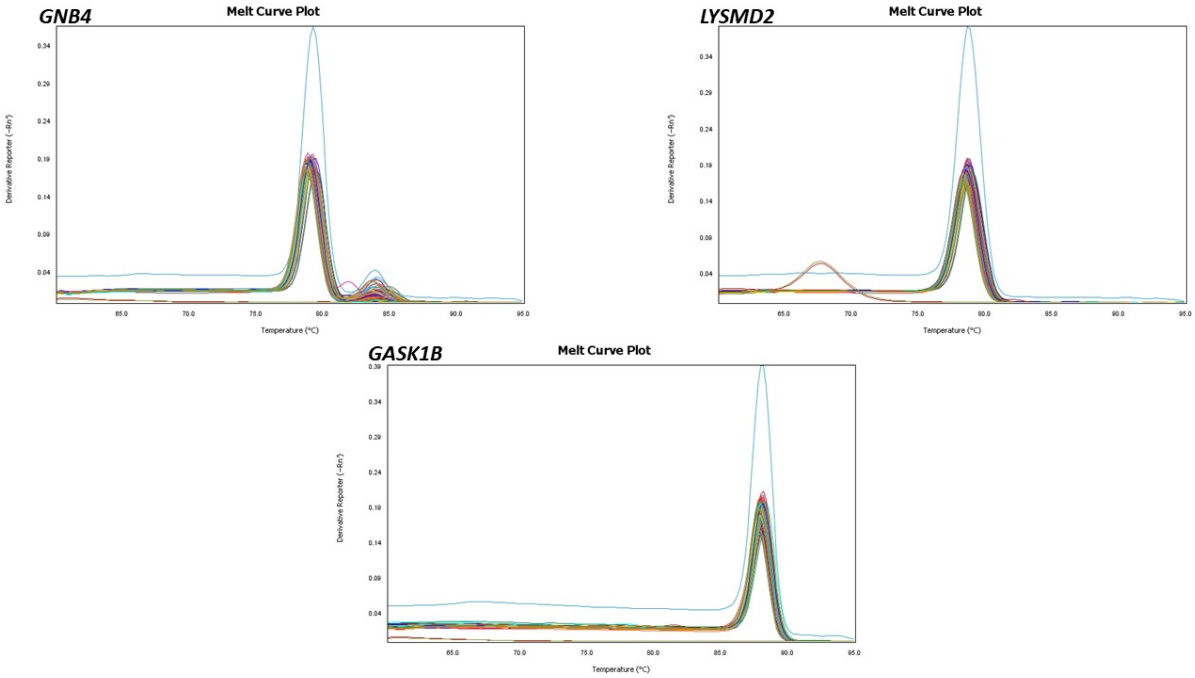
Supplemental Materials

Supplemental Table

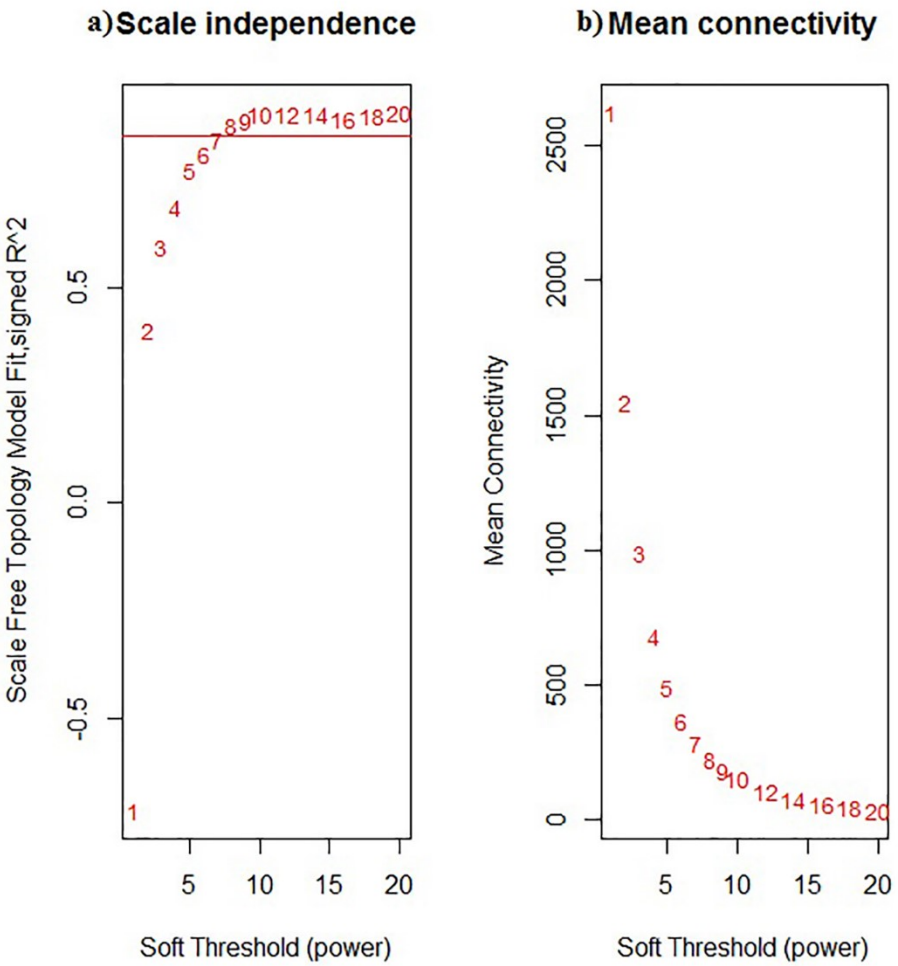
Supplemental Table 1. List of primers used for RT-qPCR analysis

Gene Name	Accession No.	Primer Sequence (5'-3')	Tm(°C)	Amplicon Length (bp)
GNB4	NM_021629	F: TAC GAT TCC AGG CTG CTA GTC A R: AGG GAG CAT AAG CAC AGG TCA	55	130
LYSMD2	NM_153374	F: ACT AGG TAG TTC TAC CTG CTG R: TAG GCT ACA ACC TTG CCA TC	55	104
GASK1B	NM_001031700	F: CGC AGC AAG CCG GCC AAT A R: TGA GCG TAC CCA GGT GCT ACA	55	161

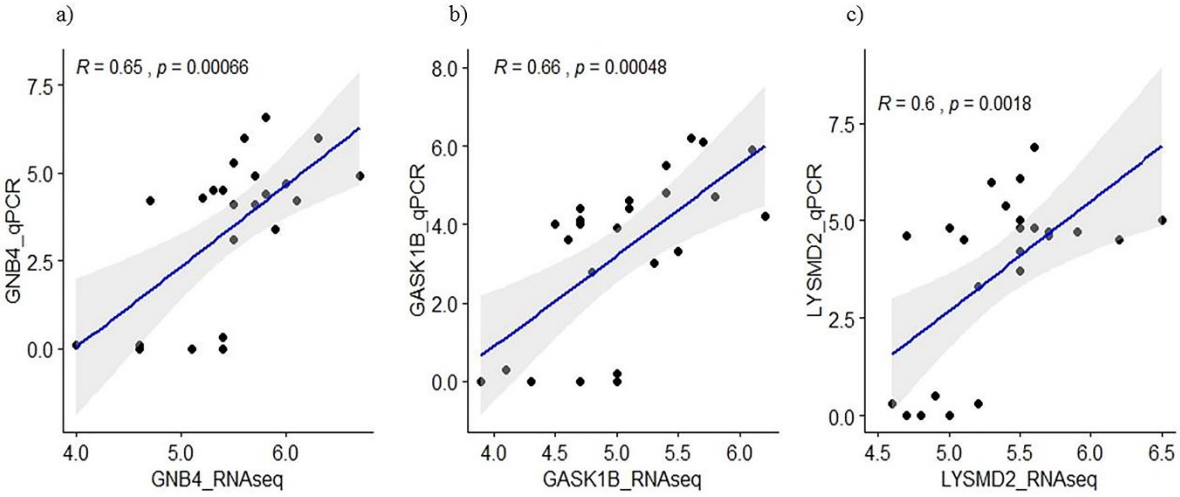
Supplemental Figures



Supplemental Figure 1. Melting curve to determine the specificity of three candidate reference gene primers by RT-qPCR.



Supplemental Figure 2. Determination of soft-thresholding power in the WGCNA. (a) The plot shows the scale-free topology fit index (y-axis) for different soft-thresholding powers (β) (x-axis). The higher the RΛ2 value, the closer to scale-free topology. Scale-free topology at β = 8 (b) Analysis of the mean connectivity (degree, y-axis) for various soft-thresholding powers (x-axis).



Supplemental Figure 3. Pearson correlation between expressions of RNA Sequencing and RT-qPCR in a) GNB4, b) GASK1B, c) LYSMD2 (p<0.01).