

Analysis of Differentially expressed genes and Molecular pathways in Peanut Allergy Induced Dynamic Changes involved in Peanut-Specific Immune Responses: A Systematic and Bioinformatics Approach

Authors: Glory Simon Parmar^{1*}, Manisha Gulati^{1*}, Md Khoshnade Rabby^{1*}, Ramchander Merugu^{1*}, Satabdi Mohanty^{1*}, Umme Kulsum^{1*}, Dhanshree Gajanan Pujari, Swapnja Rohit More, Om Prakash, Priyanka Shrikant Bhore, Deepshikha Satish[#]

Note: 1* Authors have contributed equally. # is corresponding Author

ABSTRACT

Peanut allergy poses a significant global health concern, triggering hypersensitivity reactions upon exposure to peanuts. Understanding the molecular mechanisms governing peanut-specific immune responses is imperative for developing effective therapeutic approaches. This study aimed to investigate differentially expressed genes (DEGs) and associated molecular pathways implicated in peanut allergy-induced immune responses. Employing a systematic and bioinformatics-driven approach, we analyzed gene expression profiles from peanut-allergic individuals and healthy controls using RNAseq Next-Generation Sequencing (NGS) data. Enrichment analysis of DEGs revealed their involvement in various inflammatory conditions, including autoinflammatory, allergic, and respiratory disorders. Additionally, pathway enrichment analysis highlighted perturbed molecular pathways such as Asthma, IL-17 signaling pathway, and Inflammatory bowel diseases, underscoring their role in modulating peanut-specific immune responses. Protein-protein interaction network analysis identified central regulatory hubs, elucidating the intricate molecular interplay underlying the immune response to peanut allergens. Overall, our findings offer comprehensive insights into the molecular landscape of peanut allergy, delineating key DEGs and pathways associated with peanut-specific immune responses. This systematic and bioinformatics-driven approach enhances our understanding of allergic reactions to peanuts, providing potential targets for the development of novel therapeutic interventions and diagnostic biomarkers for peanut allergy management.

KEYWORDS: *Peanuts; Allergy; NGS; RNASeq; DEG; Pathways.*

1. INTRODUCTION

Peanut allergy presents a significant global health challenge, particularly prominent in Western societies, where its incidence has surged over recent decades[1]. Marked by heightened sensitivity to peanut proteins, this allergy carries a substantial risk of severe, potentially life-threatening reactions, underscoring the importance of unraveling the molecular mechanisms governing immune reactions to peanuts. The pathogenesis of peanut allergy involves a multifaceted interplay of genetic predisposition, environmental influences, and immune system dysregulation. While the precise mechanisms remain partly elusive, recent strides in molecular biology and bioinformatics have furnished indispensable tools for delving into the intricate molecular pathways and identifying pivotal genes implicated in the immune responses specific to peanuts. This analysis lays the foundation for comprehending the development of allergies stemming from peanuts.[2]

The analysis of RNA-seq experiments requires precise and efficient software capable of managing the substantial sequencing volumes typically generated[3]. This process involves four primary tasks: aligning reads to the genome, assembling alignments into full-

length transcripts, quantifying gene and transcript expression levels, and calculating expression differences across experimental conditions. STAR efficiently aligns RNA-seq reads to the genome, discovering transcript splice sites while running much faster[4]. Cuffdiff, built upon Cufflinks, rigorously identifies differentially expressed transcripts between samples using statistical methods[5]. This paper presents a protocol for utilizing these tools in RNA-seq data analysis, all of which are thoroughly documented and actively maintained by developers.

This study focuses on analyzing differentially expressed genes (DEGs) and molecular pathways associated with peanut allergy-induced changes in immune responses. Through transcriptomic analysis of peanut-allergic individuals compared to non-allergic controls, we aim to identify genes significantly altered in response to peanut allergen exposure. These DEGs hold potential as biomarkers and therapeutic targets for peanut allergy. We will annotate the biological functions of identified DEGs and assess their enrichment in specific molecular pathways, unraveling the biological processes driving peanut-specific immune responses and allergic reactions. Integrating transcriptomic, proteomic, and

metabolomic data will provide a comprehensive understanding of molecular networks governing peanut allergy, revealing gene-gene interactions and regulatory mechanisms contributing to its pathogenesis[6].

Candidate DEGs and molecular pathways identified through bioinformatics analysis will undergo experimental validation using in vitro and in vivo models of peanut allergy, providing functional insights into their role in modulating immune responses. Computational analysis enables the use of gene expression differentials for functional annotation, pathway analysis, and molecular integration. Our research aims to elucidate the molecular mechanisms underlying peanut allergy, informing the development of targeted therapeutic interventions and diagnostic strategies. By combining sequencing technologies with rigorous analysis, we aim to uncover genetic intricacies within peanut samples, advancing personalized medicine and biological research[7].

2. MATERIALS AND METHODS

2.1. Data Acquisition and Processing:

The methodology initiated with the acquisition of raw RNA-seq reads sourced from the Sequence Read Archive (SRA) database

(<https://www.ncbi.nlm.nih.gov/sra>),

specifically utilizing the identifier SRX14128943 for investigating peanut allergy. Datasets pertinent to various experiments were gathered by accessing gene expression data stored within the database. The retrieval process involved querying the SRA database using keywords such as "Food allergen," "RNAseq," and "Homo sapiens" to identify datasets featuring relevant gene expression profiles. The dataset under investigation, SRX14128943, comprised twenty-five samples, encompassing seven and nineteen peanut allergens, as delineated in Table 1. Subsequently, gene expression data were extracted from the European Nucleotide Archive (ENA) database

(<https://www.ebi.ac.uk/ena/browser/>).

Table 1: List of selected individuals from the SRA Database, allergic to peanuts, with a control group of healthy individuals

Sr No	Sample ID	Sample Type
1	SRR17972672	Control
2	SRR17972673	Control
3	SRR17972674	Control
4	SRR17972675	Control
5	SRR17972676	Control
6	SRR17972677	Peanut Allergic
7	SRR17972678	Peanut Allergic
8	SRR17972679	Peanut Allergic
9	SRR17972680	Peanut Allergic

10	SRR17972681	Peanut Allergic
11	SRR17972682	Peanut Allergic
12	SRR17972683	Peanut Allergic
13	SRR17972684	Peanut Allergic
14	SRR17972685	Peanut Allergic
15	SRR17972686	Peanut Allergic
16	SRR17972687	Peanut Allergic
17	SRR17972688	Peanut Allergic
18	SRR17972689	Peanut Allergic
19	SRR17972690	Peanut Allergic
20	SRR17972691	Peanut Allergic
21	SRR17972692	Peanut Allergic
22	SRR17972693	Peanut Allergic
23	SRR17972694	Peanut Allergic
24	SRR17972695	Peanut Allergic
25	SRR17972696	Peanut Allergic

2.2. Data Preprocessing and Identification of DEGs

Utilizing advanced Next-Generation Sequencing (NGS) technology, the initial dataset underwent rigorous quality assessment using FastQC (v0.11.9)[8], followed by adapter trimming via Trimmomatic0.39. Subsequently, reads from each sample were aligned to the reference genome using STAR, yielding sorted BAM files[4]. These alignments were then subjected to normalization and abundance estimation using Cufflinks (v2.2.1), enabling the quantification of gene and isoform expression levels[5]. Cuffmerge was employed to assemble

genes for each dataset individually. Post-assembly, the complete set of assemblies was processed through Cuffdiff to compute differential gene expression, with resultant p-values and false discovery rates (FDR/q-values) determining significant differential expression between peanut allergens and controls[9].

We followed the stringent criteria, requiring $|\log(\text{fold change})| > 1$ and $p < 0.05$ to identify significant DEGs from the dataset. Upregulated DEGs were defined by $\log_{2}FC \geq 1$, while downregulated DEGs were identified by $\log_{2}FC \leq -1$. Subsequently, RStudio (v4.3.2) and the ggplot2 package were utilized to generate a volcano plot visualizing the significant DEGs. Identified DEGs were extracted from the dataset for further analysis.

To visualize gene expression patterns, a heatmap was constructed using the ComplexHeatmap library. This protocol is accessible without specialized programming skills but assumes familiarity with the Unix command-line interface and basic R scripting[10]. Users should feel comfortable executing commands in the Unix environment and editing text files. The study workflow is illustrated in Fig 1.

2.3. Functional Enrichment Analysis

Functional annotation and analysis of KEGG pathway enrichment were conducted using the web-based DAVID v6.8 tool, a reliable resource for comprehensive functional evaluation of high-throughput gene expression data[11]. The outcomes from DAVID were subsequently imported into R Studio for visualization using ggplot2. Additionally, ShinyGO (v 0.80) was employed to visualize functional and pathway enrichment of the top 10 Differentially Expressed Genes (DEGs). ShinyGO facilitates the integration of expression data with functional assessment results, enabling a deeper understanding of molecular, biological, and cellular functions, as well as pathway enrichment analysis[12].

2.4. Protein-Protein Interaction (PPI) Network Construction and Module Analysis

The protein-protein interaction (PPI) network was constructed using the STRING web-based tool (v12.0) to elucidate relationships among the DEGs identified in the datasets[13]. A medium confidence interaction score cutoff of ≥ 0.4 was applied to filter out inconsistent PPI interactions from the dataset, resulting in the establishment of a robust PPI network. Subsequently, the outcomes from the

STRING tool were integrated into Cytoscape software (v3.8.2) to visualize the PPI interactions among statistically significant DEGs. Hub genes within the PPI network were identified by sorting based on the degree of interaction. For enhanced visualization, the yFiles organic layout was applied to represent the network structure effectively.

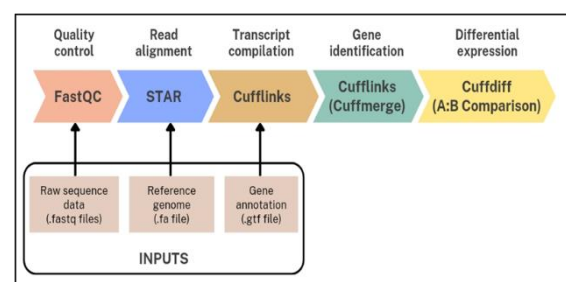


Fig 1. Shows RNA-Seq pipeline begins with quality control, which used FastQC to check the sequencing data's quality, assuring reliable downstream analysis. The readings were then aligned to a reference genome using STAR, which allows for more accurate mapping of RNA transcripts. Cufflinks are then used to assemble transcripts, allowing potential transcripts to be reconstructed from matched reads. Cuffmerge is used to identify genes by aggregating transcript annotations into a single set for further analysis. Finally, Cuffdiff is used for differential gene expression analysis, which allows the identification of genes that are differentially expressed across

experimental conditions, providing insights into biological mechanisms.

3. RESULTS

3.1. Identification of Differentially Expressed Genes (DEGs) and Data Visualization

The gene expression profiles of nineteen peanut allergen samples and seven control groups (CD154+ CD4 T-cells) were acquired from the SRA database (SRX14128943). Utilizing the Illumina HiSeq 2500 platform, DEGs between control samples and peanut allergens were identified employing the cuffdiff tool. DEGs were determined based on the criteria of $|\log_2FC| \geq 1.0$ and $p\text{-values} \leq 0.05$. A total of 444 genes were identified as biologically significant. Subsequently, a volcano plot was constructed using R studio and the ggplot2 package, visually representing the DEGs. In the plot (Fig 2), upregulated significant genes are depicted as red dots ($p\text{-value} \leq 0.05$, $\log_2FC \geq 1$), while downregulated significant genes are shown as green dots ($p\text{-value} \leq 0.05$, $\log_2FC \leq -1$). Furthermore, the top 20 DEGs were subjected to a heat map analysis using the ComplexHeatmap package, revealing the expression levels of these genes (Fig 3).

3.2. Protein-Protein Interaction (PPI) Network Construction and Module Analysis

To explore the physical and functional associations among DEG proteins, the STRING database was utilized. A minimum required interaction score of ≥ 0.4 was set, and the resultant interaction network was visualized using Cytoscape v3.8.2. The network comprised 264 nodes and 613 edges, with upregulated genes highlighted in red ($\text{cutoff } \log_2FC \geq 1$) and downregulated genes highlighted in green ($\text{cutoff } \log_2FC \leq -1$). Network analysis identified ATM, IL4, GATA3, and CTLA4 genes as highly interacted hub genes, with degrees of interaction ranging from 21 to 26.

3.3. Functional Enrichment Analysis using DAVID

Functional annotation of DEGs was performed using the DAVID v6.8 online server. KEGG pathway enrichment analysis and Gene Ontology (GO) classification were conducted to elucidate the biological roles of DEGs[14]. DEGs were found to be enriched in various biological processes, molecular functions, and cellular components. Specifically, upregulated genes were associated with processes such as steroid catabolism,

immune response, and cytokine activity, while downregulated genes were implicated in pathways related to adaptive immune response and antigen binding[15]. Additionally, DEGs were found to be involved in diverse cellular components, including lysosomes, endosomes, and the endoplasmic reticulum. KEGG pathway analysis revealed enrichment in pathways such as steroid hormone biosynthesis, cytokine-cytokine receptor interaction, and JAK-STAT signaling pathway, among others, highlighting their potential roles in peanut allergen response[16, 17].

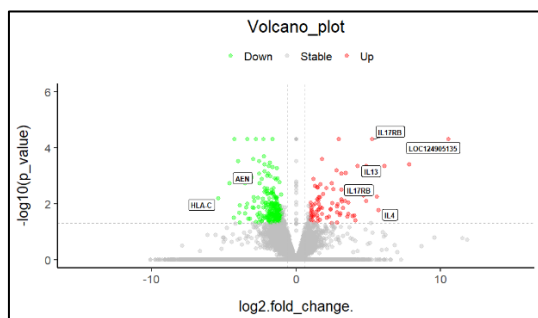


Fig 2: A volcano plot is a popular graphical depiction for differential gene expression studies.

It shows fold change on the x-axis and statistical significance (typically written as $-\log_{10}(p\text{-value})$) on the y-axis. Genes with significant changes in expression are shown as points above a specific threshold, which are often highlighted in the plot, generating a volcano-like shape with the most significant genes at the top.

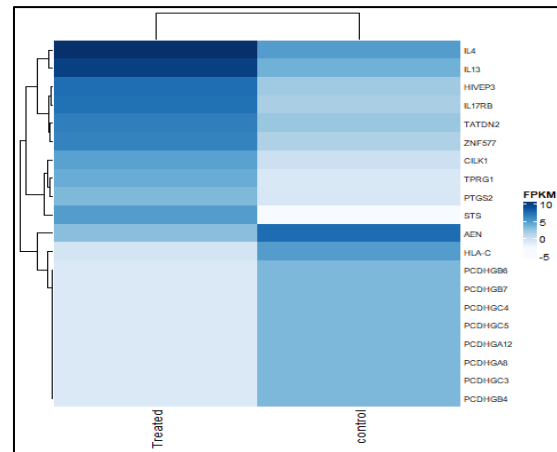
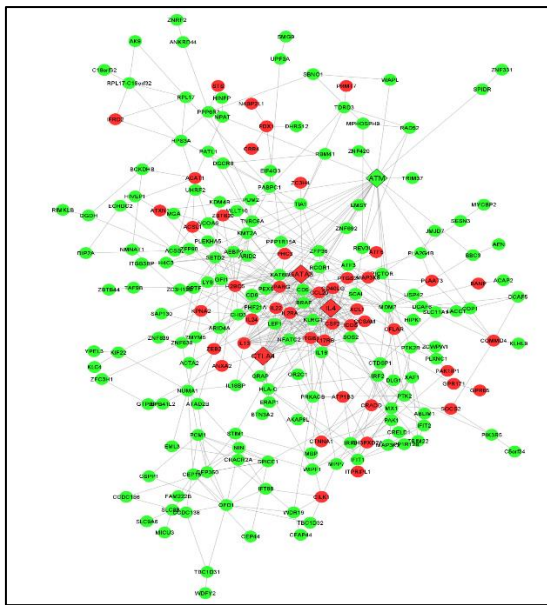


Fig 3: A heat map of gene expression is a visual representation of gene expression levels across various samples or conditions. Normally, rows indicate genes, columns represent samples, and colors represent expression levels, with brighter colors suggesting higher expression and darker shades indicating lower expression. Heat maps provide an overview of expression patterns, making it simple to identify clusters or trends in gene expression across different experimental conditions or biological samples.



imported into Cytoscape for visualisation and analysis, allowing researchers to investigate gene interconnectivity, functional modules, and putative biomolecular mechanisms driving complex biological processes. Cytoscape provides strong tools for visualizing and analyzing biological networks, making it easier to understand gene connections and their implications in a variety of biological contexts.

Fig 4: Gene network analysis is performed using String involves obtaining protein-protein interaction data to create a network graph in which genes are represented as nodes and interactions as edges. This network data is then able to be

Table 2: Enriched Gene Ontology (GO) Analysis Highlighting Upregulated Genes Identified via ShinyGO Database.

GO	nGenes	Genes
BIOLOGICAL PROCESS		
Negative reg. of cellular response to transforming growth factor beta	3	IL4 CFLAR PPARG
Positive reg. of T-helper 2 cell cytokine production	3	IL4 GATA3 XCL1
T-helper 2 cell cytokine production	3	IL4 GATA3 XCL1
Positive reg. of type 2 immune response	3	IL4 GATA3 XCL1
CD4-positive alpha-beta T cell cytokine production	3	IL4 GATA3 XCL1
Positive reg. of interleukin-10 production	4	IL4 CD40LG XCL1 IL13
Lymphocyte	4	IL4 ICOS CD40LG MAP3K8

stimulation		
Leukocyte cell-cell adhesion	9	IL4 CTLA4 ICOS CD40LG XCL1 ITGB1 GATA3 IL2RA MAP3K8
Leukocyte differentiation	10	IL4 CTLA4 ITGB1 GATA3 PPARG CSF2 ANXA2 CD40LG IL2RA SH3PXD2A
CELLULAR COMPONENTS		
Melanosome	4	ATP1B3 ITGB1 STX3 ANXA2
Pigment granule	4	ATP1B3 ITGB1 STX3 ANXA2
MOLECULAR FUNCTION		
Icosatetraenoic acid binding	2	PPARG STX3
Arachidonic acid binding	2	PPARG STX3
Icosanoid binding	2	PPARG STX3
Long-chain fatty acid binding	2	PPARG STX3
Protease binding	5	CRADD ANXA2 CFLAR SH3PXD2A ITGB1
Cytokine receptor binding	10	CD40LG CCL20 XCL1 CFLAR GATA3 IL4 SOCS2 CSF2 IL13 IL22
Cytokine activity	8	IL4 CCL20 XCL1 CSF2 CD40LG IL22 IL24 IL13
Receptor ligand activity	8	IL4 CCL20 XCL1 CSF2 CD40LG IL22 IL24 IL13
Signaling receptor activator activity	8	IL4 CCL20 XCL1 CSF2 CD40LG IL22 IL24 IL13

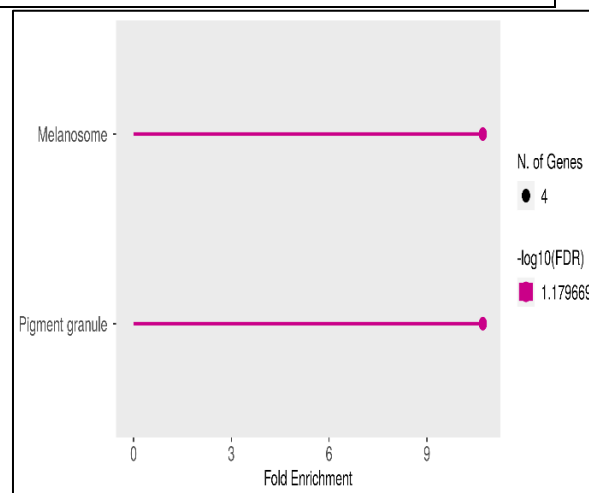
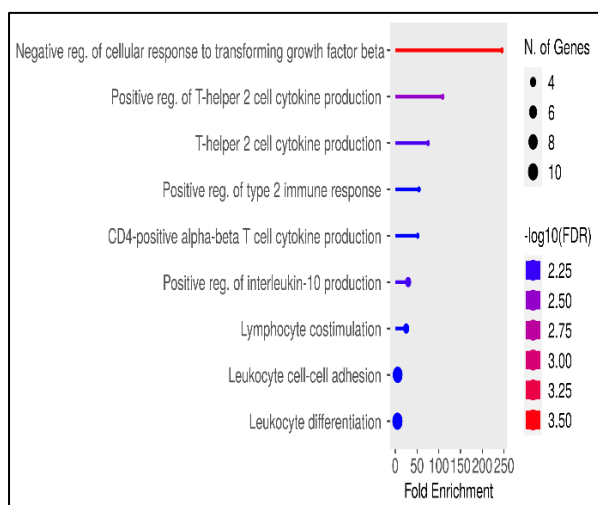
Table 3: Highly enriched Gene Ontology(GO) Analysis for Downregulated genes from ShinyGO database.

GO	nGenes	Genes
BIOLOGICAL PROCESS		
Homophilic cell adhesion via plasma membrane adhesion molecules	22	PCDHGA2 PCDHGA1 PCDHGC3 PCDHGC5 PCDHGC4 PCDHGA12 PCDHGB6 PCDHGA5 PCDHGA7 PCDHGA6 PCDHGA8 PCDHGA10 PCDHGA11 PCDHGB2 PCDHGB4 PCDHGB7 PCDHGB1 PCDHGA3 PCDHGA9 PCDHGB3 PCDHGA4 PCDHGB5
Cell-cell adhesion via plasma-membrane adhesion molecules	24	MBP PCDHGA2 PCDHGA1 PCDHGC3 PCDHGC5 PCDHGC4 PCDHGA12 PCDHGB6 PCDHGA5 PCDHGA7 PCDHGA6 PCDHGA8 PCDHGA10 PCDHGA11 PCDHGB2 PCDHGB4 PCDHGB7 PCDHGB1 PCDHGA3 PCDHGA9 PCDHGB3 PCDHGA4 PCDHGB5 CD6

Positive reg. of protein-containing complex assembly	10	BBC3 MPP7 SPIDR DLG1 RICTOR SLF2 PTK2B NUMA1 PAK1 ATM
Nuclear chromosome segregation	13	ZNF207 WAPL SPICE1 RNF212 ARID2 ZCWPW1 ATM AKAP8L KIF22 SLF2 NUMA1 EML3 RMDN1
Cell-cell adhesion	30	DLG1 CD5 ARID2 LEF1 PTK2 MBP PCDHGA2 PCDHGA1 PCDHGC3 PCDHGC5 PCDHGC4 PCDHGA12 PCDHGB6 PCDHGA5 PCDHGA7 PCDHGA6 PCDHGA8 PCDHGA10 PCDHGA11 PCDHGB2 PCDHGB4 PCDHGB7 PCDHGB1 PCDHGA3 PCDHGA9 PCDHGB3 PCDHGA4 PCDHGB5 CD6 FCHO1
Chromosome organization	31	ZNF207 WAPL KDM4B ATM CHD3 RNF212 H4C3 TDRD3 SLF2 ARID2 ACIN1 SMCHD1 PAK1 BPTF ARID4A PUM2 ZCWPW1 RSNB1 RCOR1 KMT2A PHF21A AEBP2 ITGB3BP EMSY SETD2 TSPYL2 AKAP8L KIF22 NUMA1 EML3 RMDN1
Negative reg. of transcription by RNA polymerase II	26	ZNF133 PHF21A SAP130 MIER3 ZNF75A CHD3 ZNF875 ZFP90 ARID4A HIVEP1 TRIM37 ZMYM5 MDM2 GFI1 ATF3 ZNF692 BPTF NFATC2 ZFP36 LEF1 AEBP2 HINFP DLG1 TAF9B IRF2 ZNF345
Cell cycle proc.	32	ZNF207 WAPL PCM1 CEP152 CSPP1 TRIM37 NUMA1 NPAT ATM SPICE1 ZFYVE19 RNF212 KLHL9 EXOC6 ARID2 LEF1 CEP44 CHD3 ZC3H12D SETD2 ZCWPW1 MDM2 CTDSP1 AKAP8L DLG1 KIF22 NIN SLF2 EML3 HINFP RMDN1 GFI1
Negative reg. of nucleobase-containing compound metabolic proc.	40	RCOR1 ZNF133 PHF21A SAP130 MIER3 ZNF75A CHD3 ZNF875 ZFP90 ARID4A PABPC1 HIVEP1 ACIN1 TRIM37 ZFP36 ZMYM5 MDM2 LEF1 GFI1 ATF3 ZNF692 BPTF HINFP TSPYL2 SLC11A1 LRCH4 NFATC2 TRDMT1 TRIM22 AEBP2 NPAT SCAI WAPL DLG1 SMCHD1 KMT2A ATM TAF9B IRF2 ZNF345
Negative reg. of RNA metabolic proc.	35	RCOR1 ZNF133 PHF21A SAP130 MIER3 ZNF75A CHD3 ZNF875 ZFP90 ARID4A PABPC1 HIVEP1 ACIN1 TRIM37 ZFP36 ZMYM5 MDM2 LEF1 GFI1 ATF3 ZNF692 BPTF HINFP SLC11A1 LRCH4 NFATC2 TRDMT1 TRIM22 AEBP2 NPAT SCAI DLG1 TAF9B IRF2 ZNF345
CELLULAR COMPONENTS		

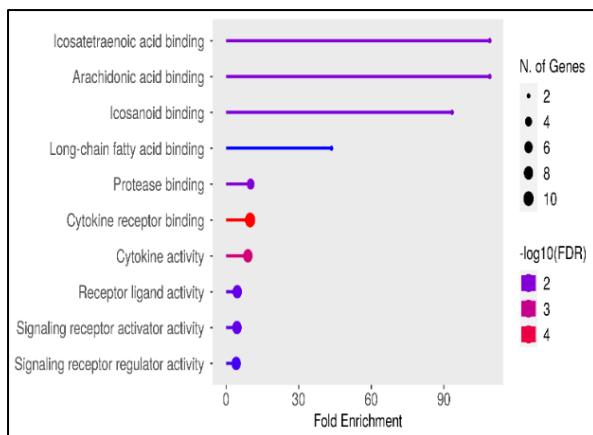
MPP7-DLG1-LIN7 complex	2	DLG1 MPP7
Exon-exon junction complex	3	TDRD3 ACIN1 UPF3A
Pericentriolar material	3	NIN CEP152 PCM1
Histone methyltransferase complex	5	KMT2A AEBP2 MGA NCOA6 TRIM37
Histone deacetylase complex	5	RCOR1 PHF21A SAP130 CHD3 ARID4A
Centriole	10	IFT88 MPHOSPH9 PCM1 NIN SPICE1 OFD1 CEP152 CEP350 CEP44 ODF2L
Spindle	14	NIN CSPP1 NUMA1 EML3 SPICE1 CEP44 RMDN1 WAPL KIF22 TBCK ZNF207 IRAG2 CEP350 ATM
Centrosome	20	OFD1 PCM1 NIN CEP152 CSPP1 ODF2L NUMA1 TBC1D31 SPICE1 CEP44 ZFYVE19 IFT88 MPHOSPH9 CEP350 PIK3R5 PRKACB HIPK1 CHD3 ATM RABGAP1
Microtubule organizing center	24	IFT88 OFD1 MPHOSPH9 PCM1 NIN CEP152 CSPP1 ODF2L CRACR2A NUMA1 TBC1D31 SPICE1 CEP44 ZFYVE19 CEP350 PIK3R5 PRKACB HIPK1 PTK2 CHD3 RABGAP1 IRAG2 PAK1 ATM
Microtubule cytoskeleton	36	IFT88 OFD1 MPHOSPH9 PCM1 KIF22 NIN CEP152 CSPP1 ACTA2 ODF2L CRACR2A KLC4 NUMA1 EML3 TBC1D31 MX1 SPICE1 CEP44 ZFYVE19 RMDN1 WAPL DLG1 CEP350 PIK3R5 PRKACB TBCK HIPK1 PTK2 CHD3 MYCBP2 ZNF207 RABGAP1 IRAG2 PAK1 ATM STIM1
MOLECULAR FUNCTION		
Acetate-CoA ligase activity	2	ACSS2 DIP2A
SUMO transferase activity	4	MDM2 UHRF2 KIAA1586 RNF212
Cation:cation antiporter activity	4	SLC11A1 SLC8B1 SLC9A6 TMCO3
Solute:cation antiporter activity	4	SLC11A1 SLC8B1 SLC9A6 TMCO3
Histone binding	13	MLLT10 ZCWPW1 ATAD2B SBNO1 CHD3 BPTF TSPYL2 TDRD3 KMT2A JMJD7 UHRF2 LEF1 HINFP

Calcium ion binding	32	CRACR2A MICU3 STIM1 PLA2G4B GLCE BRAF PCDHGA2 NIN CRELD1 MAN1B1 NELL2 PCDHGA1 PCDHGC3 PCDHGC5 PCDHGC4 PCDHGA12 PCDHGB6 PCDHGA5 PCDHGA7 PCDHGA6 PCDHGA8 PCDHGA10 PCDHGA11 PCDHGB2 PCDHGB4 PCDHGB7 PCDHGB1 PCDHGA3 PCDHGA9 PCDHGB3 PCDHGA4 PCDHGB5
Microtubule binding	11	ZNF207 KIF22 NUMA1 EML3 MX1 RMDN1 CEP44 STIM1 NIN CEP350 IRAG2
Tubulin binding	15	ZNF207 KIF22 NUMA1 EML3 MX1 RMDN1 PAK1 CEP44 STIM1 SETD2 NIN CEP350 OFD1 IRAG2 RABGAP1
DNA-binding transcription repressor activity RNA polymerase II-specif	11	ZNF133 ZNF75A ZNF875 ZFP90 GF11 ATF3 ZNF692 HIVEP1 NFATC2 LEF1 HINFP
DNA-binding transcription repressor activity	11	ZNF133 ZNF75A ZNF875 ZFP90 GF11 ATF3 ZNF692 HIVEP1 NFATC2 LEF1 HINFP



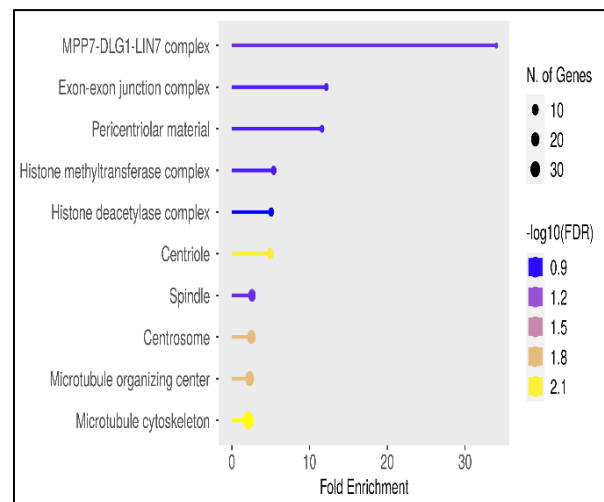
(A) Biological Process

(B) Cellular Component

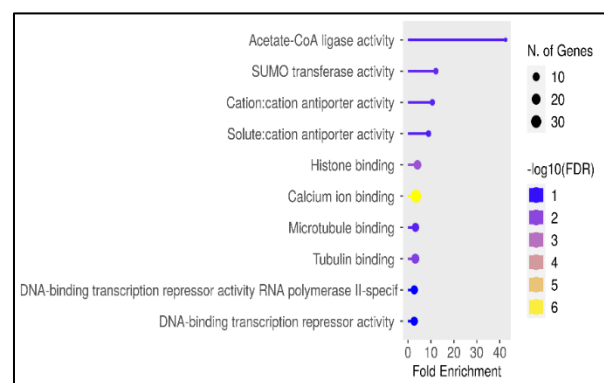


(C) Molecular Function

Fig 5: Shows the results of a functional enrichment study that focused on upregulated genes, highlighting the enriched (A) biological processes, (B) cellular components and (C) molecular functions associated with upregulated genes. The visualization provides insights into the biological pathways or mechanisms that may be activated or strengthened in response to experimental conditions, which aids in the understanding of gene expression changes.

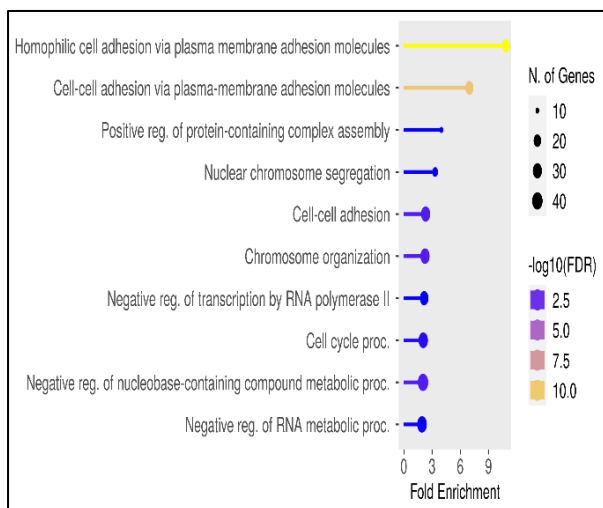


(B) Cellular Process



(C) Molecular Function

Fig 6: Shows the results of a functional enrichment study that focused on upregulated genes, highlighting the enriched (A) biological processes, (B) cellular components and (C) molecular functions associated with downregulated genes. The visualization provides insights into the biological pathways or mechanisms that may be activated or strengthened in response to experimental conditions, which aids in the understanding of gene expression changes.



(A) Biological Process

Table 4: Pathway Enrichment Analysis Performed Using the ShinyGO Database

Pathways	nGenes	Genes
hsa04657 IL-17 signalling pathway	6	CSF2 IL4 IL13 IL17RB PTGS2 CCL20

hsa05321 Inflammatory bowel disease	4	GATA3 IL4 IL13 IL22
hsa04660 T cell receptor signalling pathway	6	MAP3K8 CSF2 CTLA4 ICOS IL4 CD40LG
hsa01212 Fatty acid metabolism	3	ACSL1 ACAT1 CBR4
hsa04668 TNF signalling pathway	5	MAP3K8 CSF2 PTGS2 CCL20 CFLAR
hsa04658 Th1 and Th2 cell differentiation	4	GATA3 IL2RA IL4 IL13
hsa04630 JAK-STAT signalling pathway	7	IL24 CSF2 IL2RA IL4 IL13 IL22 SOCS2
hsa04061 Viral protein interaction with cytokine and cytokine receptor	4	IL24 IL2RA CCL20 XCL1
hsa04659 Th17 cell differentiation	4	GATA3 IL2RA IL4 IL22
hsa04060 Cytokine-cytokine receptor interaction	10	IL24 CSF2 IL2RA IL4 IL13 IL22 IL17RB CCL20 XCL1 CD40LG

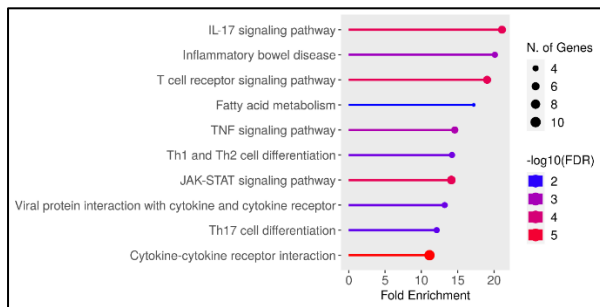


Fig 4: This figure depicts the results of a pathway enrichment analysis, emphasizing the enriched biological pathways linked to the experimental data. This graphic provides insights into the functional pathways that may be influenced or altered by the conditions under investigation.

4. DISCUSSION

In this study, we delved into the differential gene expression between nineteen peanut allergens and five controls using the SRA ID SRX14128943 as a reference. Our analysis scrutinized a total of 46,694 differentially expressed genes (DEGs), from which we

identified the top 444 significant DEGs for further investigation (Table 1). Notably, key genes such as ATM, IL4, GATA3, and CTLA4 emerged as highly interactive hub genes, suggesting their potential importance in the context of peanut allergy (Fig 4).

Moreover, our examination revealed three seed nodes—IL4, IL13, CCL20, and CSF2—potentially involved in the regulation of pathways implicated in peanut allergy[18]. To elucidate the functional roles of these DEGs, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses, which highlighted their involvement in crucial pathways such as cytokine-cytokine receptor interaction, JAK-STAT signaling, and T cell receptor signaling pathways, among others[14, 17].

Focusing on specific genes such as STS, IL13, IL4, IL17RB, HLA-C, and AEN, we explored their roles in modulating immune responses and allergic reactions to food allergens.

Steroid sulfatase (STS), for instance, implicated in steroid hormone metabolism, may contribute to allergic sensitization through immune modulation and epithelial barrier function. Similarly, cytokines IL-13 and IL-4 play pivotal roles in type 2 immune responses and allergic inflammation, with genetic variations in their genes associated with increased susceptibility to food allergies. IL17RB and HLA-C also contribute to allergic diseases through their roles in immune regulation and antigen presentation[19], while the AEN gene, involved in DNA repair and apoptosis regulation, may indirectly modulate allergic responses[18].

Further analysis of molecular functions (MF) and biological processes (BP) using GO revealed significant enrichment patterns among upregulated and downregulated DEGs. Upregulated genes were associated with cytokine receptor binding and leukocyte differentiation, while downregulated genes showed enrichment in calcium ion binding and negative regulation of nucleobase-containing compound metabolic processes, among others[20, 21]. Additionally, cellular component (CC) analysis identified enrichment in melanosome and microtubule cytoskeleton among upregulated and downregulated DEGs, respectively (Table 2 and Table 3).

This comprehensive exploration sheds light on the molecular mechanisms underlying peanut

allergy and provides valuable insights for potential therapeutic targets and intervention strategies[1]. Further studies are warranted to validate these findings and translate them into clinical applications for improved management of food allergies[18].

5. CONCLUSION

In conclusion, our study offers a thorough examination of the molecular intricacies of peanut allergy, elucidating crucial differentially expressed genes (DEGs) and pathways linked to peanut-specific immune reactions. Through our systematic and bioinformatics-driven methodology, we have deepened our comprehension of the fundamental mechanisms driving allergic responses to peanuts. These findings present promising avenues for the creation of innovative therapeutic strategies and diagnostic biomarkers, which could significantly advance the management of peanut allergies.

REFERENCES

1. Hourihane, J.O., *Peanut allergy*. *Pediatr Clin North Am*. 2011 Apr;58(2):445-58, xi. doi: 10.1016/j.pcl.2011.02.004., 2011.
2. Anagnostou, K. and A. Clark, *Peanut immunotherapy*. *Clin Transl Allergy*. 2014 Sep 25;4:30. doi: 10.1186/2045-7022-4-30. eCollection 2014., 2014.

3. Kukurba, K.R. and S.B. Montgomery, *RNA Sequencing and Analysis*. Cold Spring Harb Protoc. 2015 Apr 13;2015(11):951-69. doi: 10.1101/pdb.top084970., 2015.
4. Dobin, A.e.a.S., *ultrafast universal RNA-seq aligner*. *Bioinformatics* 29. 2013: p. 15-21.
5. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks*. *Nat Protoc*. 2012 Mar 1;7(3):562-78. doi: 10.1038/nprot.2012.016., 2012.
6. Imran, S., et al., *Epigenetic programming underpins B-cell dysfunction in peanut and multi-food allergy*. *Clin Transl Immunology*. 2021 Aug 24;10(8):e1324. doi: 10.1002/cti2.1324. eCollection 2021., 2021.
7. Du Toit, G., et al., *Randomized trial of peanut consumption in infants at risk for peanut allergy*. *N Engl J Med*. 2015 Feb 26;372(9):803-13. doi: 10.1056/NEJMoa1414850. Epub 2015 Feb 23., 2015.
8. Trivedi, U.H., et al., *Quality control of next-generation sequencing data without a reference*. *Front Genet*. 2014 May 6;5:111. doi: 10.3389/fgene.2014.00111. eCollection 2014., 2014.
9. McDermaid, A., et al., *Interpretation of differential gene expression results of RNA-seq data: review and integration*. *Brief Bioinform*. 2019 Nov 27;20(6):2044-2054. doi: 10.1093/bib/bby067., 2019.
10. Williams, J.R., et al., *Functional Heatmap: an automated and interactive pattern recognition tool to integrate time with multi-omics assays*. *BMC Bioinformatics*. 2019 Feb 15;20(1):81. doi: 10.1186/s12859-019-2657-0., 2019
11. Sherman, B.T., et al., *DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021Á update)*. *Nucleic Acids Res*. 2022 Jul 5;50(W1):W216-W221. doi: 10.1093/nar/gkac194., 2022.
12. Bang, H., et al., *Integrated bioinformatic analysis of gene expression profiling data to identify combinatorial biomarkers in inflammatory skin disease*. *Scientific Reports*, 2022. **12**(1): p. 5889.
13. Rao, V.S., et al., *Protein-protein interaction detection: methods and analysis*. *Int J Proteomics*. 2014;2014:147648. doi: 10.1155/2014/147648. Epub 2014 Feb 17., 2014.

14. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet. 2000 May;25(1):25-9. doi: 10.1038/75556., 2000. doi: 4;11:605688. doi: 10.3389/fimmu.2020.605688. eCollection 2020., 2020.
15. Jennifer Sasson, G. Brett Moreau, and W.A. Petri, *The role of interleukin 13 and the type 2 immune pathway in COVID-19: A review.* Ann Allergy Asthma Immunol., 2023
16. Dreskin, S.C., *Do HLA genes play a role in the genetics of peanut allergy?* 2006: Ann Allergy Asthma Immunol. 2006 Jun;96(6):766-8. doi: 10.1016/S1081-1206(10)61337-3.
17. Kanehisa, M., et al., *KEGG: new perspectives on genomes, pathways, diseases and drugs.* Nucleic Acids Research, 2016. 45(D1): p. D353-D361.
18. Sicherer, S.H. and H.A. Sampson, *Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment.* J Allergy Clin Immunol. 2014 Feb;133(2):291-307; quiz 308. doi: 10.1016/j.jaci.2013.11.020. Epub 2013 Dec 31., 2014.
19. Littera, R., et al., *Human Leukocyte Antigen Complex and Other Immunogenetic and Clinical Factors Influence Susceptibility or Protection to SARS-CoV-2 Infection and Severity of the Disease Course. The Sardinian Experience.* Front Immunol. 2020 Dec
20. Andrade, M.A., et al., *Automated genome sequence analysis and annotation.* Bioinformatics. 1999 May;15(5):391-412. doi: 10.1093/bioinformatics/15.5.391.
21. Hemler, J.A., et al., *The evolving story of human leukocyte antigen and the immunogenetics of peanut allergy.* Ann Allergy Asthma Immunol. 2015 Dec;115(6):471-6. doi: 10.1016/j.anai.2015.10.008. Epub 2015 Oct 27., 2015.