CORRELATION NETWORK ANALYSIS OF GENE EXPRESSION DATA FROM FAECAL TRANSCRIPTOME OF CHILDREN WITH ENVIRONMENTAL ENTEROPATHY: A SECONDARY DATA ANALYSIS

Word Count: 5327

Done By:
Madheshvaran S

REPORT
July 2018

Contents

1. Abstract

Environmental Enteric Dysfunction (EED) is a clinical condition characterized by abnormal intestinal permeability leading to intestinal and systemic inflammation, presumably caused by high pathogen load in the small intestine. A study conducted by Jinsheng Yu *et al.*, involving the characterization of faecal transcriptomes from 259 individuals from Malawi showed differential expression of genes associated with chronic inflammatory responses and epithelial tissue integrity in individuals with EED. We performed a secondary analysis of the published data to understand how genes are co-expressed in EED and whether inter-relationships between crucial genes is altered between healthy and EED individuals by constructing gene correlation coexpression networks. We find that significantly differentially expressed genes were also connected to each other differently between healthy controls and EED, with disappearance of nodes with negatively correlated connections, and strengthening of positive correlations. We also used this data to interrogate the relationship of TLR4 signaling (that would reflect effects of pathogen load) with specific pathways related to intestinal epithelial cell homeostasis (including development, differentiation, cell death, response to pathogens and inflammation) to understand which of the pathways are aberrant in EED. TLR-4 was significantly correlated with a genes involved in apoptosis, inflammatory responses, cell junction regulatory, TLR and TNFα signaling, but not intestinal development or differentiation. Together these analyses suggest alterations in inter-relationships between genes and intestinal homeostatic pathways in EED, that will form starting points for specific hypothesis-based studies.

2. Abbreviations

EED – Environmental Enteric Dysfunction

LM ratio / %L value – lactulose-mannitol ratio test

TLR-4 – Toll-Like Receptor 4

FDR – False Discovery Rate test

PCC / Pearson's r – Pearson's correlation coefficient

TNFα – Tumour Necrosis Factor

NF-κB – nuclear factor kappa-light-chain-enhancer of activated B cells

CD14 – Cluster of Differentiation 14

LAPTM5 – Lysosomal Multispanning Membrane Protein 5

IFITM1 – Interferon-Induced Transmembrane Protein

BIN2 – Bridging Integrator 2

CR1 – Complement C3b/C4b Receptor 1

S100A12 – S100 Calcium-Binding Protein A12

MD-2 – Lymphocyte antigen 96

PECAM1 – Platelet and Endothelial Cell Adhesion Molecule 1

3. Introduction

Environmental Enteric Dysfunction (EED) or Environmental enteropathy (EE) is a condition of chronic intestinal inflammation, which is caused by increased interaction between microbes and intestinal epithelial cells due to prolonged exposure to contaminated water and food[1,2,3]. It results in the disruption of epithelial cell junctions, chronic inflammatory responses such as increased T-cell proliferation and activation of neutrophils and macrophages, decreased absorption, and disturbed permeability of the gut[1,4]. Although there are minimal significant acute symptoms[5], there are severe long-term symptoms such as stunting[1,4,6], malnutrition[2,5], and impaired brain development[7,8].

One of the ways to detect EED is a simple dual sugar permeability test (lactulose - mannitol ratio test or more commonly, LM Ratio Test)[9]. However, the gold-standard test for EED is an intestinal biopsy, which is an expensive and invasive procedure[10]. The dual sugar permeability test works on the principle that lactulose (a poorly absorbable sugar) gets absorbed through the disturbed cell junctions in the intestine by passive transport from higher gradient (in the intestine) to lower gradient (in the blood). Since lactulose cannot be metabolised by the body, it gets excreted through the urine. Mannitol, which can pass through the cell membranes and cell junctions, is utilised to standardise lactulose take-up in the test. The fraction of lactulose to that of mannitol present in the urine is denoted by LM ratio or %L value[1,9]. This value is used to determine the nature of the intestine and hence, to diagnose EED[9] in the patient.

Although there is evidence that access to clean water, food and proper sanitation and promoting community hygiene are the best ways to prevent EED[11], there is insufficient understanding of the complex pathophysiological mechanisms that contribute to

development and progression of EED. There is insufficient evidence on relative importance of pathogen loads, nutrition, and inflammation on eliciting and sustaining cellular barrier dysfunction. It is also not clear whether the interplay of factors that elicit EED and/or the consequences of EED could alter intestinal stem cells to be epigenetically programmed differently, and alter its developmental potential, leading to permanently altered crypt-villus architecture and prolonged periods of intestinal dysfunction. This possibility is supported by studies in transgenic mouse models that show the effect of TLR4 signalling in reducing cell proliferation and increasing apoptosis in intestinal stem cells[12].

To understand the pathobiology of EED, Jinsheng Yu et al characterised the transcriptome in the faeces of 259 rural Malawian children who were evaluated as being at risk for environmental enteric dysfunction[1]. The faecal transcriptome was characterised by using high-density microarray to characterize gene expression patterns by two statistical methods:

(1) differential gene expression analysis between healthy controls and EED and

(2) correlation analysis of gene expression with the LM ratio values.

They observed that 51 genes that are significantly differentially expressed and have a high correlation with the %L values. This gene list was further characterized by pathway enrichment analysis using KEGG and GeneGO.

We did a secondary analysis of the raw gene expression data (available publicly on GEO database) to address the following additional questions:

(1) Is the degree of inter-relationships between the significantly different genes altered in EED vs healthy controls?

(2) Can we identify gene clusters that show significantly different correlated-coexpression between EED and healthy controls?

(3) Can we use gene correlation networks from this dataset to identify interactions between TLR4 signalling (that reflects bacterial pathogen exposure) and various pathways in intestinal epithelial cell homeostasis that can form predictions for mechanistic experiments?

4. Aim of the Project

- To test the degree of inter-relationships between the 51 significantly differentially expressed genes between healthy controls and EED (data from Jinsheng et al). This was achieved by constructing gene correlation co-expression networks separately for healthy controls and EED. This was a hypothesis-agnostic objective to identify and describe significant gene-nodes that show differences between healthy controls and EED.
- To test how TLR-4 correlates with the following gene sets that represent plausible pathophysiological processes in EED: Apoptosis, inflammation, cell junction regulation, intestinal epithelial cell development, intestinal epithelial cell differentiation, TLR4 signalling, WNT/β-Catenin signalling, TNFα signalling and NF-κB pathways. This was a hypothesis-driven objective to identify which of the pathways showed different patterns of co-expression with TLR4 in EED compared to healthy controls.

5. Materials and Methods

5.1 Source of Data

Given below are the sources of raw data that were obtained for this bioinformatic analysis.

- Raw data for microarray of the 259 children in the cohort was obtained from GEO database (GSE74681): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE74681
- Affymetrix probe annotations for Probe IDs and Gene Names: http://www.affymetrix.com/support/technical/byproduct.affx?product=human_transcriptome
- Gene sets containing Gene Names and their respective Probe IDs were obtained from the Broad Institute website: http://software.broadinstitute.org/gsea/msigdb/genesets.jsp

Spreadsheets of raw data and R codes used to analyze them are available at https://github.com/madheshvaran/EED

5.2 Software Used

The following software was used for the analysis of the EED data:

- **GEO2R**: This was used to compare two or more groups of samples to identify genes that are differentially expressed across experimental conditions.
- **Python 3.7**: This programming language was used to compare the demographics of the population involved in the study with their % L ratios.

- **R 3.6.0**: This programming language was used to create correlation matrixes and networks between high and low LM value groups. Moreover, R script was also used correlate TLR-4 with the different gene sets.

5.3 Packages Used

a. In Python 3.7

The packages that were used in python for this analysis were: Numpy and Matplotlib.pyplot

b. In R 3.6.0

The packages that were used in R for this analysis were: readxl, dplyr, reshape2, ggplot2, qgraph, hmisc, and readr.

5.4 Method of Analysis

a. Demographic details of the Individuals in the study

The list of individual's data along with their age, sex and %L value was downloaded from the GEO Database. The relationship of %L values with age and sex was analyzed. This analysis was done to verify whether the sample chosen was random and unbiased and to check how %L values of the individuals get distributed among different age groups in the sample.

b. Correlation Network between healthy and severe EED individuals

The study classifies the individuals into three groups based on their LM ratio values: the heathy group (%L value : 0 – 0.2), the intermediate EED group (%L value: 0.2 – 0.7) and severe EED group (%L value: 0.7 and above)[1]. The analysis was done

between two groups of thirty individuals belonging to healthy and severe EED groups. The sixty individuals' transcriptome was checked for its value distribution (a method which used to determine whether the value data are median-centred across the samples and thus, to check whether the groups are suitable for cross-comparison)[13]. The transcriptome expression values of 51 genes (that are found to be significantly differentially expressed according to the study) of the 60 individuals were extracted from the database and correlated within each group to see how the correlation network varies between those two groups.

c. Correlation of TLR-4 with various Gene Sets

Different gene sets that are hypothesized to be pathophysiologically relevant to EED were correlated with TLR-4. Correlations were done by calculating Pearson correlation coefficient. Also, their P values were acquired and adjusted by Benjamini & Hochberg adjustment method (also called as "FDR" method). The correlation of TLR4 with the gene sets of the two groups was plotted with correlation values of severe EED and healthy individuals on x and y axes, respectively. The genes that have significant P values (< 0.1) were marked with a different colour to visualise those genes in the plot.

6. Results

6.1 Demographic Analysis

a. Histogram of distribution of %L values

The data was analyzed to see how %L values vary with the frequency of individuals considered in the study. Fig 1 shows the density histogram showing distribution of %L values in the cohort. The distribution is right skewed and not normally distributed.

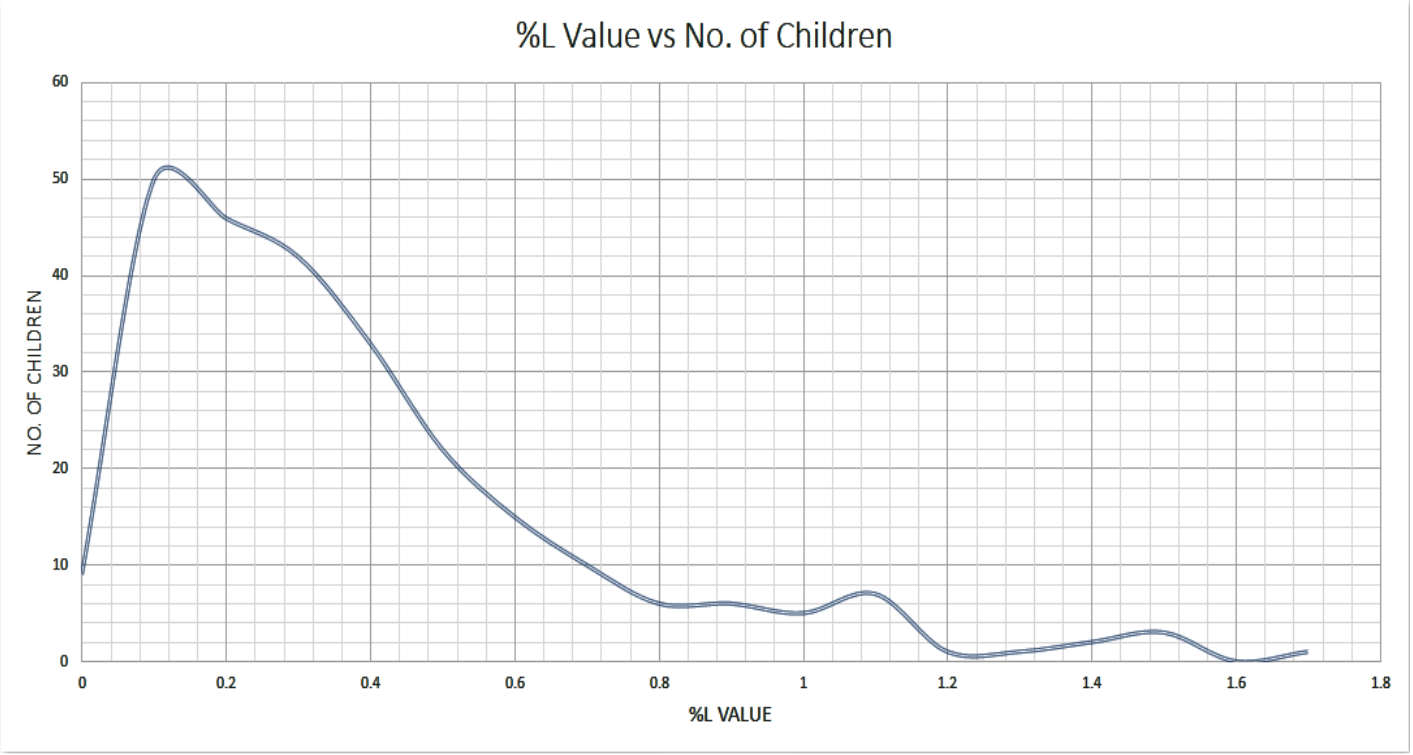


Fig 1: Histogram showing distribution of %L values of individuals

b. Distribution of %L values in different Age groups

The variation of %L values in different age groups was analysed using the individuals' data. Fig 2 shows the variation of %L values with age. There is an under-representation of children with high %L with increasing age. Most children with high %L were in the age group of 20 – 40 months.

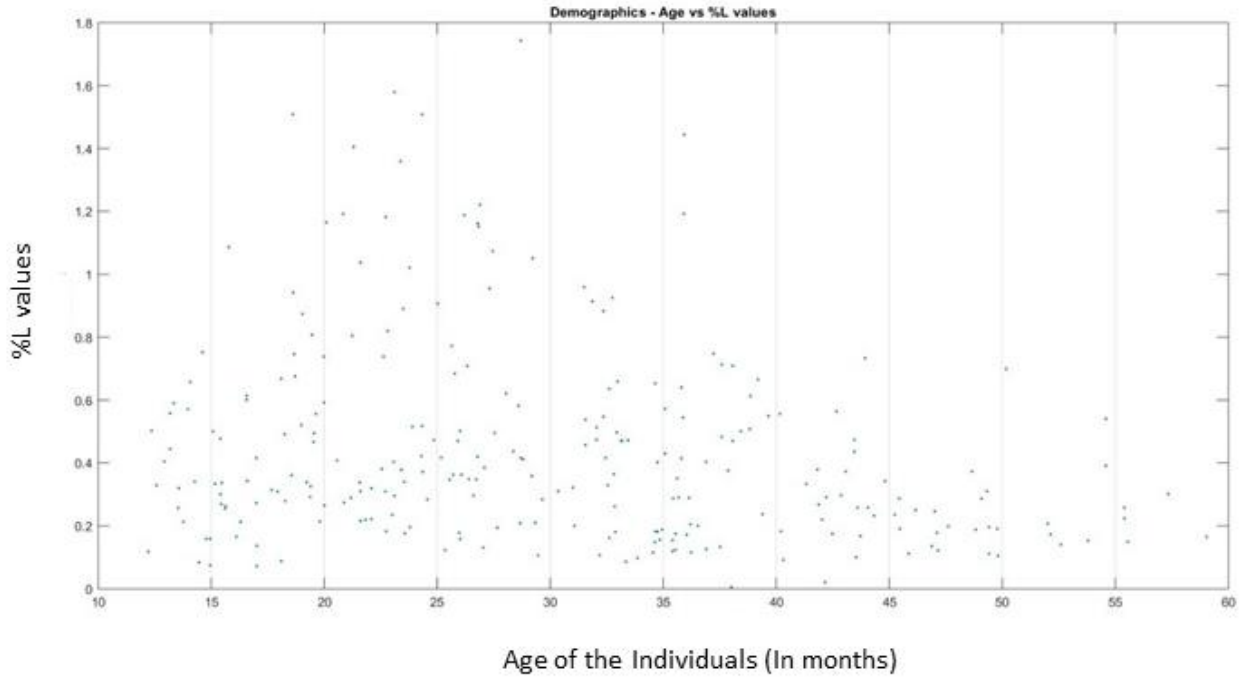


Fig 2: The plot of the age of the individual vs %L value of the individual.

c. Distribution of Gender among the different EED Groups

The individuals' information, belonging to one of the three EED groups as classified by the study (healthy, intermediate and severe EED groups – refer 'Materials and Methods' for their %L cutoff values), were analyzed based on gender. Figure 3 shows bar graphs of distribution of fraction of males and females belonging to the three groups – healthy, intermediate %L values (intermediate EED) and high %L values

(severe EED). Males and females were equally represented in the healthy group, whereas there is a higher proportion of males in the severe EED group.

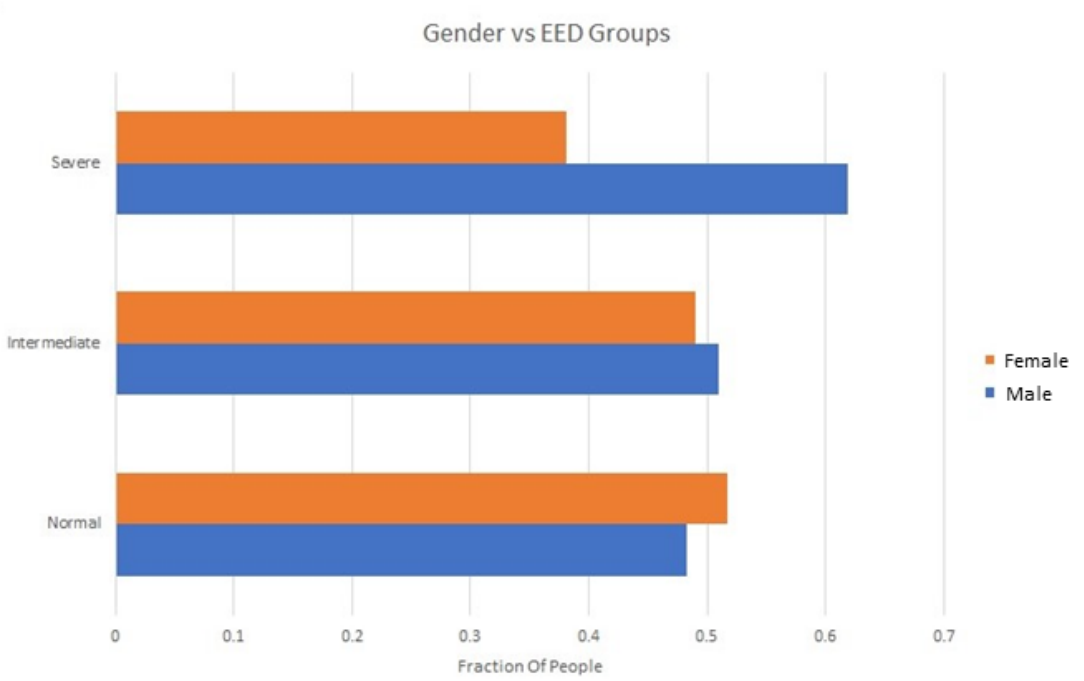


Fig 3: The plot between the fraction of males and females in all three groups: Healthy (Normal), Intermediate EED and Severe EED.

6.2 Transcriptome Analysis between Healthy and Severe EED Groups

Two groups, consisting of 30 individuals each and belonging to the Healthy and Severe EED groups, respectively, were analyzed for the value distribution of the transcriptome data. Based on the %L values, individuals were rank ordered and two groups were defined. Thirty individuals who had lowest values for %L formed the healthy controls and 30 individuals with highest %L values were defined as EED. All the 60 individuals had comparable median values and distribution of total gene expression profiles (Fig 4), ruling out batch effects and ensuring that data are normalized and cross-comparable.

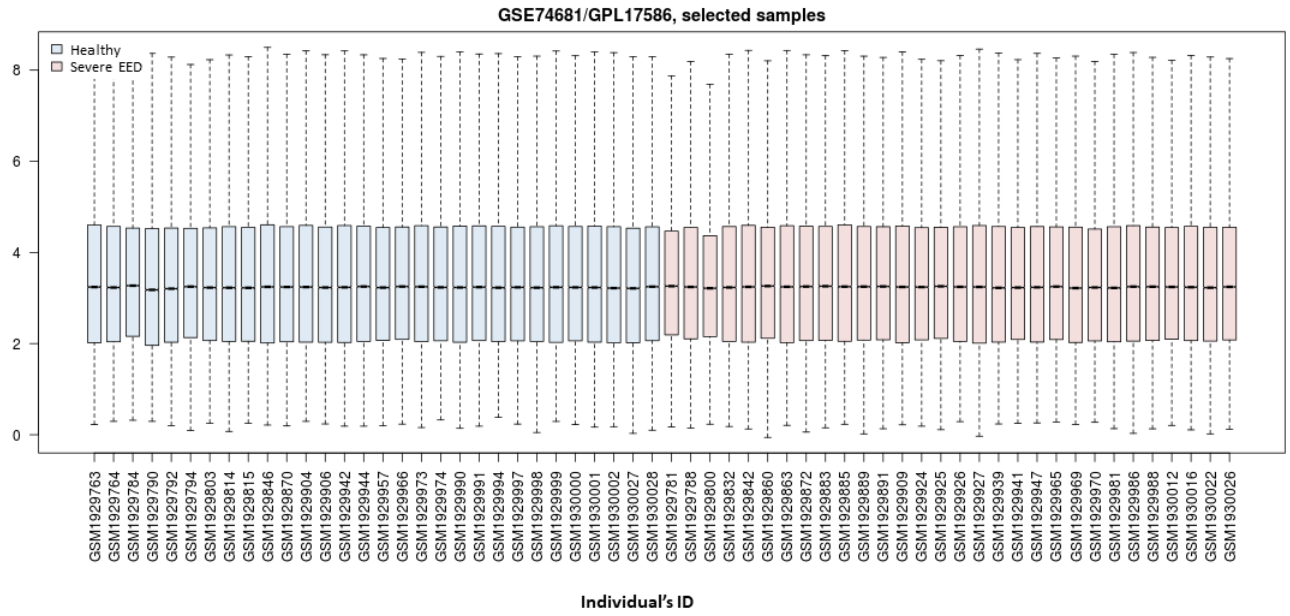


Fig 4: Boxplots showing distribution of total transcript abundance of the 60 individuals. Red and blue signifies that the individual belongs to 'EED' or 'Healthy' groups, respectively.

The expression values of those 51 genes in the individuals were extracted from the GEO database, and a correlation analysis was performed on the two groups individually. The data extracted were analysed in R Script by using 'cor' function, which correlates the data by using Pearson's correlation method and generates a correlation matrix consisting of Pearson's correlation coefficient (also called as Pearson's r value). This matrix can be visualised in the form of a correlation network by using the 'qgraph' package. The correlation network shows the connection between two genes in the form of lines with varying thickness and colour. Here, green and red lines signify positive and negative Pearson's r. The thickness of the line is indicative of the absolute value of Pearson's r; a thicker line indicates the more correlation between two genes and vice versa. The correlation networks were generated from the transcriptome data of the two groups and are shown below (Fig 5).

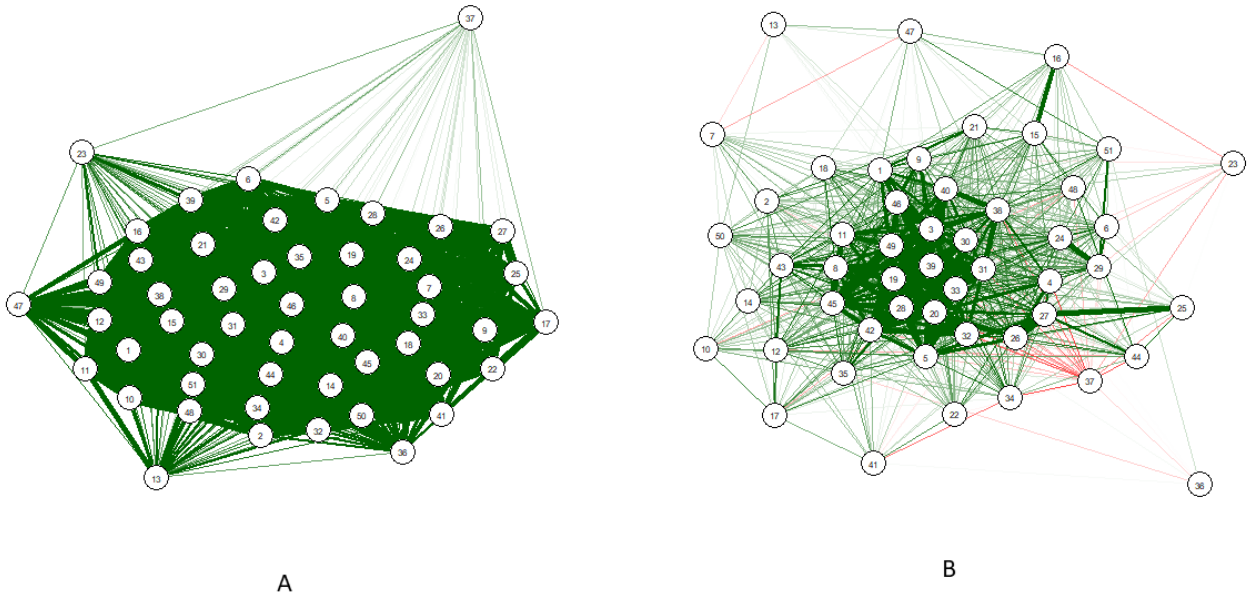


Fig 5: The correlation network of 51 genes in ‘Severe EED’ group (A) and ‘Healthy’ group (B).

In the case of the correlation network of ‘Severe EED’ group, most genes are strongly correlated (indicated by thick lines that fills up the whole space). Interestingly, the correlation network of ‘Healthy’ group has a weaker positive correlation between most of the genes and some negative correlation between some genes.

The genes which are positively correlated in the ‘Severe EED’ group but negatively correlated in ‘Healthy’ group are OR52D1, NOP10, LAPTM5, IFITM1, BIN2, CR1, S100A12. However, OR52D1 (an olfactory receptor protein-coding gene) and NOP10 (a ribonuclear protein-coding gene) were not considered for further analysis as their function cannot be correlated with the symptoms of EED. The remaining genes were analysed, and their role in EED was evaluated by looking at their functions.

6.3 Analysis of correlation of TLR-4 with Various Gene sets

To conclude TLR-4 as a suspectable biomarker for EED, it was correlated with different gene sets which have a common feature (such as apoptosis, inflammatory responses, cell junction regulators, etc.) that can be associated with the symptoms of EED. These gene sets were chosen based on the characteristic traits that can be seen in the patients of EED. They include Apoptotic gene sets, Inflammatory Gene Sets, Cell Junction gene sets, Gene sets responsible for intestinal epithelial cell development and differentiation, TLR4 signaling gene sets, WNT/β-Catenin Signaling gene sets and TNFα Signaling via NF-κB gene sets.

a. Correlation of TLR-4 with Apoptotic Gene sets

The apoptotic gene sets were correlated with TLR-4 in both the groups and the Pearson's r of those genes with TLR-4 were plotted by using 'ggplot' function (Fig 6). Moreover, the P values of correlation of those genes with TLR-4 less than 0.1 were marked differently from that of non-significant genes.

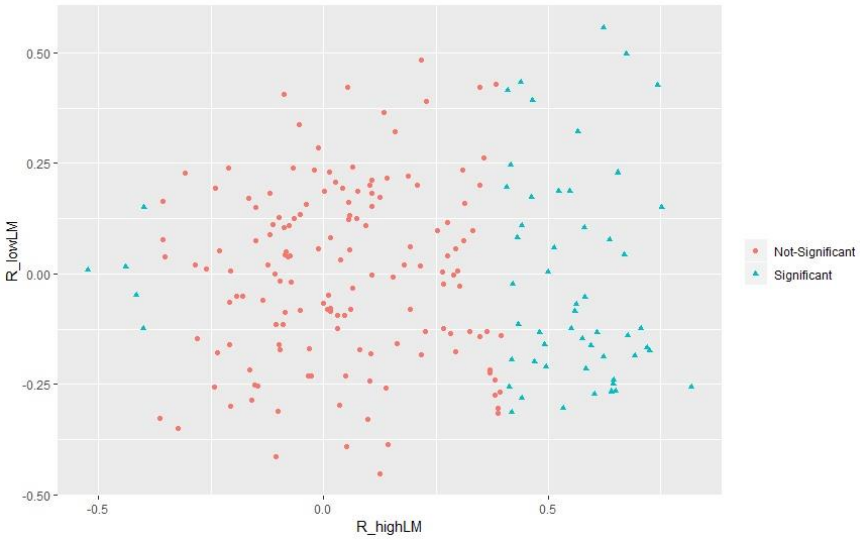


Fig 6: The scatter-plot of correlation of TLR-4 with apoptotic gene sets with X and Y axis representing the PCC of genes with TLR-4 in 'Severe EED' and 'Healthy' groups. The blue-triangle points represent genes with P value less than 0.1.

There were genes that are significantly and positively correlated with TLR-4 in Severe EED group (R_highLM) but negatively correlated with Healthy group (R_lowLM) and vice versa. These significantly correlated genes were listed and correlated among themselves (Fig 7). From the correlation network, CD14 (marked as 8) is negatively correlated with most of the genes in the Healthy group but positively correlates in the Severe EED group.

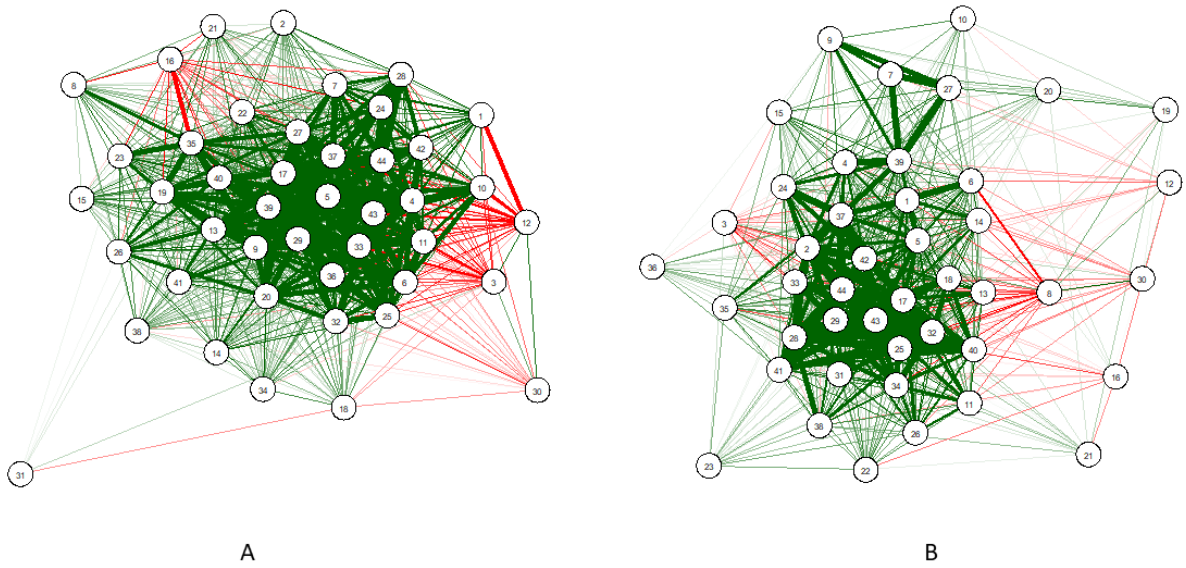


Fig 7: The correlation network of apoptotic genes (that significantly correlated with TLR-4) in 'Severe EED' group (A) and 'Healthy' group (B).

b. Correlation of TLR-4 with Inflammatory Gene sets

The inflammatory gene sets were correlated with TLR-4 in both the groups and the Pearson's r of those genes with TLR-4 were plotted (Fig 8). The significant correlation of genes with TLR-4 was marked differently from non-significant ones.

These significantly correlated genes are correlated among themselves, and correlation networks of these two groups were generated. By comparing the two correlation networks, the following genes were mostly negatively correlated with other

genes in ‘Healthy’ group but positively correlated with those genes in ‘Severe EED’ group: RNF144B, SERPINE1, NLRP3, OSM, IL18RAP, CD14, CSF3R, CYBB. Also, most of the genes in the severe EED group are highly correlated among each other.

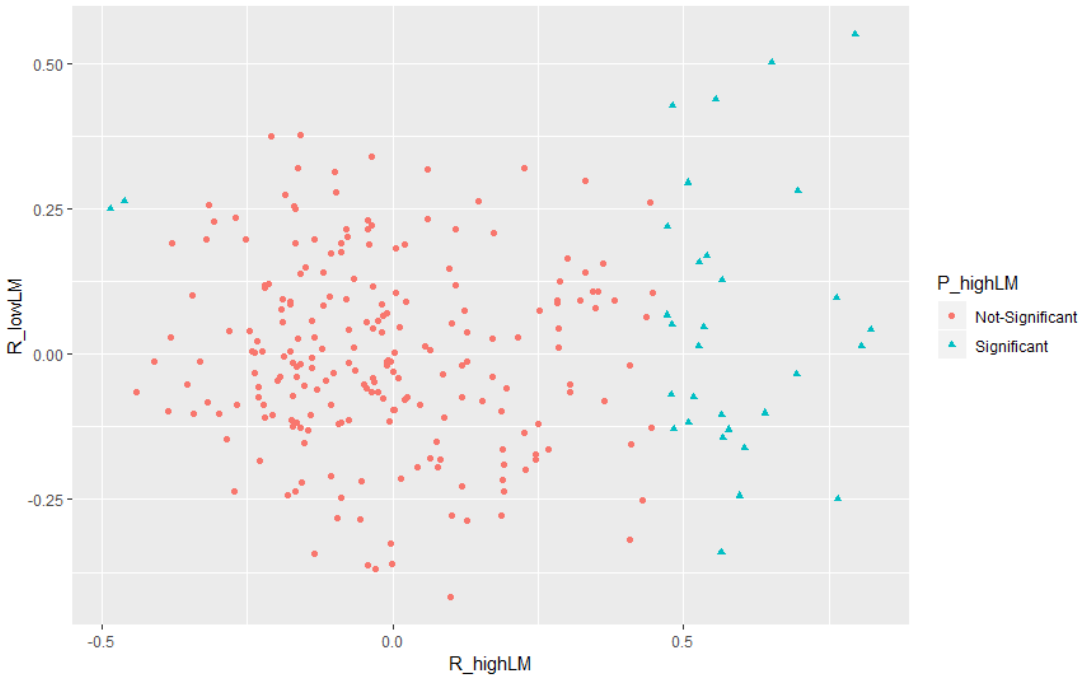


Fig 8: The scatter-plot of correlation of TLR-4 with inflammatory gene sets with X and Y axis representing the PCC of genes with TLR-4 in ‘Severe EED’ and ‘Healthy’ groups.

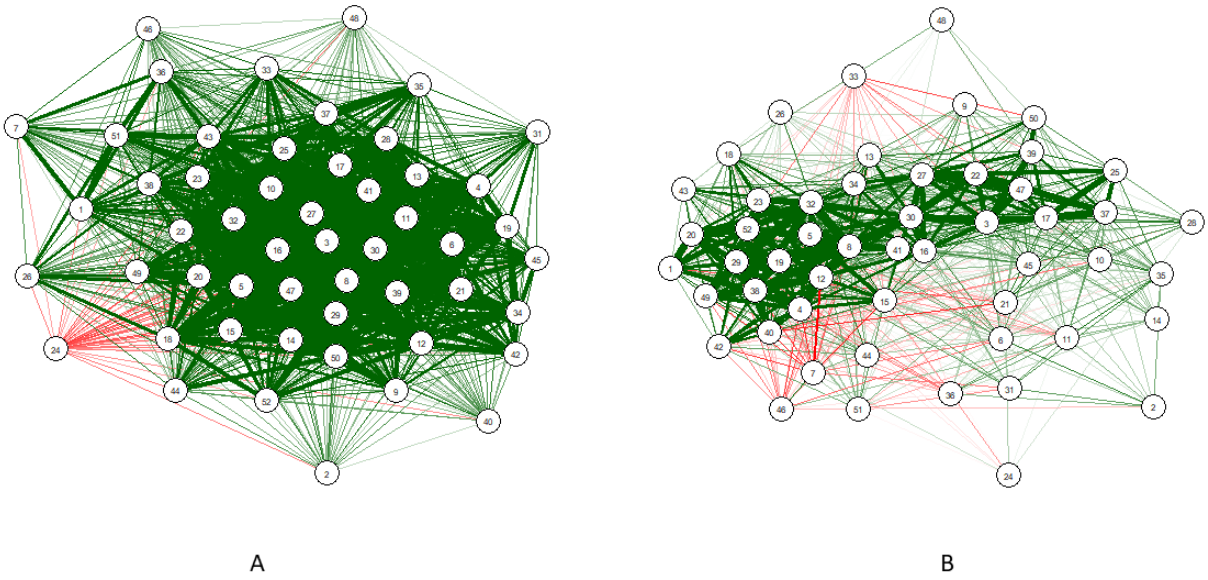


Fig 9: The correlation network of inflammatory genes (that significantly correlated with TLR4 in ‘Severe EED’ group (A) and ‘Healthy’ group (B).

c. Correlation of TLR-4 with Cell Junction Gene sets

A similar analysis was done, and the following were the results (Fig 10):

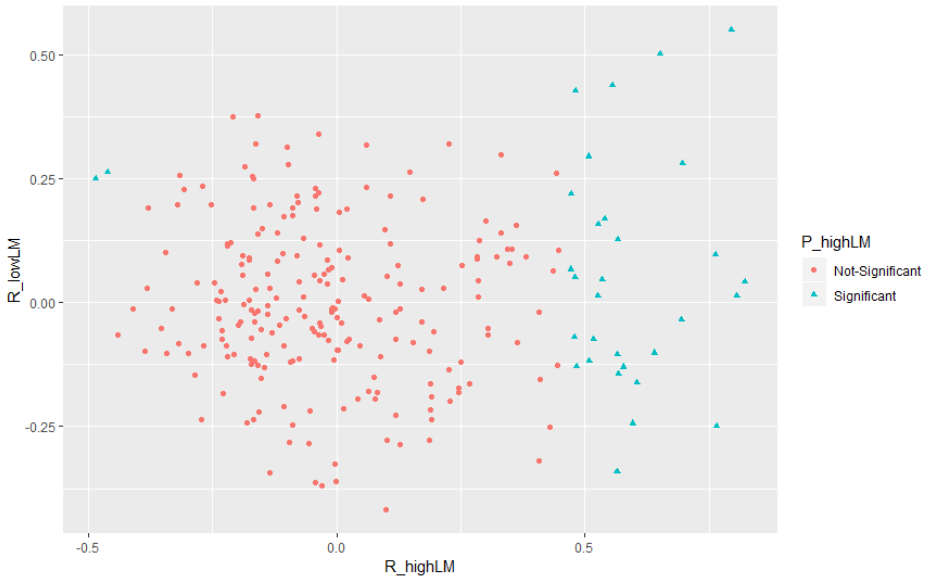


Fig 10: The scatter-plot of correlation of TLR-4 with cell junction gene sets with X and Y axis representing the PCC of genes with TLR-4 in 'Severe EED' and 'Healthy' groups. The blue-triangle points represent genes with P value less than 0.1.

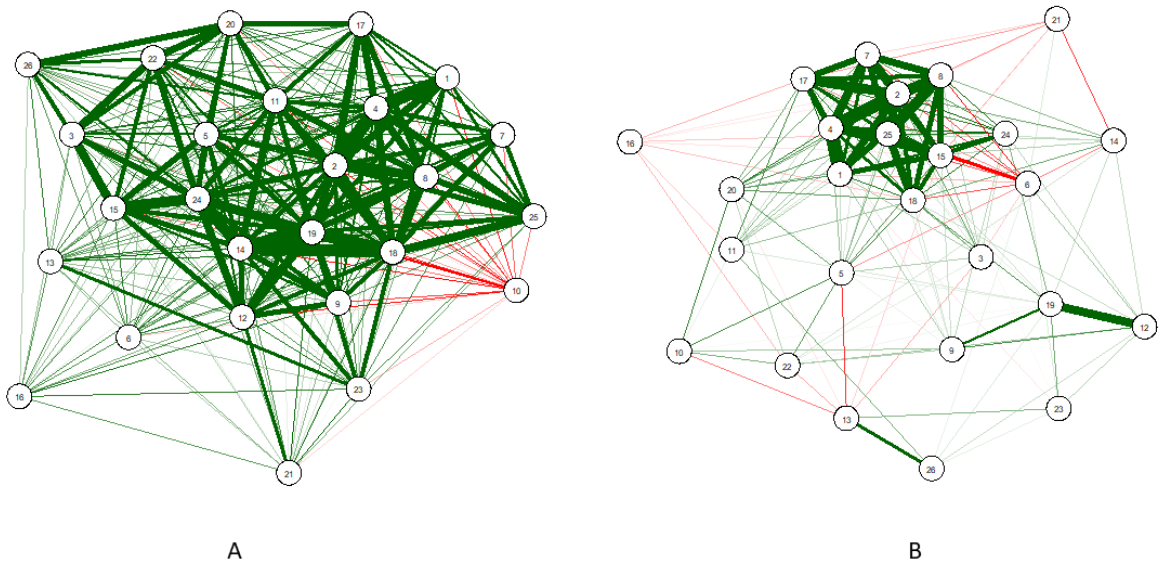


Fig 11: The correlation network of cell junction genes (that significantly correlated with TLR-4) in 'Severe EED' group (A) and 'Healthy' group (B).

From the correlation networks (Fig 11), interestingly enough, PECAM1 (CD31) gene was mostly negatively correlated with other genes in severe EED groups but positively correlated in Healthy groups. SKAP2, ZYX, SGCE, MAPK14 and ARHGEF6 were mostly negatively correlated with other genes in the 'Healthy' group but positively correlated with those genes in 'Severe EED' group.

d. Correlation of TLR-4 with Intestinal Developmental and Differentiation Gene sets

A scatter-plot showing the correlation of TLR-4 with the above gene sets in different groups was plotted (Fig 12).

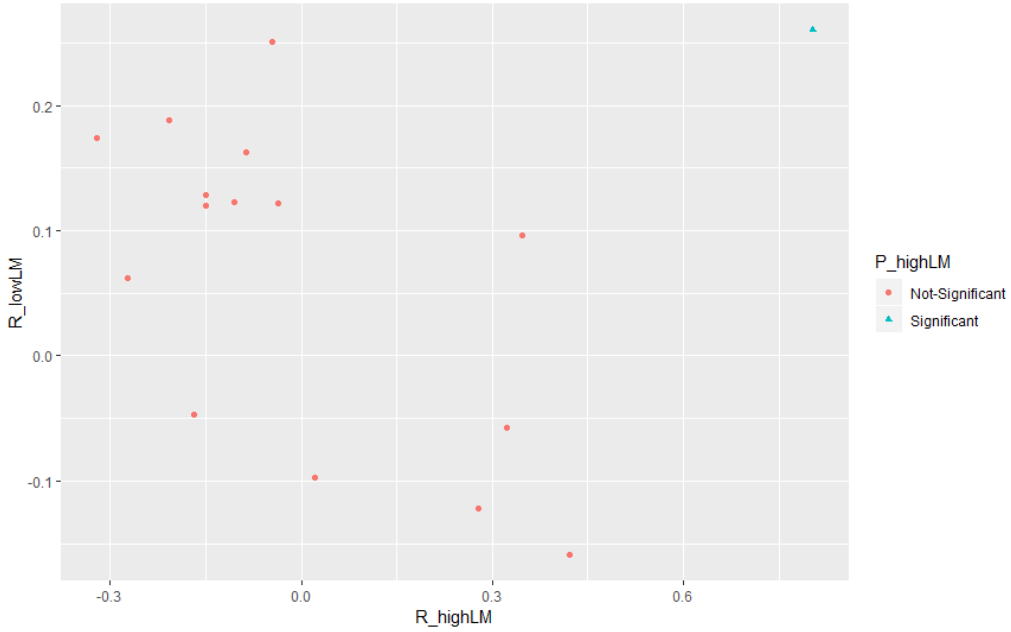


Fig 12: The scatter-plot of correlation of TLR-4 with intestinal development and differentiation gene sets with X and Y axis representing the PCC of genes with TLR-4 in 'Severe EED' and 'Healthy' groups. The blue-triangle points represent genes with P value less than 0.1.

Since the number of significant genes was low, a correlation network could not be plotted.

e. Correlation of TLR-4 with WNT/β-Catenin Signaling Gene sets

A scatter-plot showing the correlation of TLR-4 with the above gene sets in different groups was plotted by using 'ggplot' package in R Script (Fig 13). Again, since the number of significant genes was low, a correlation network could not be plotted.

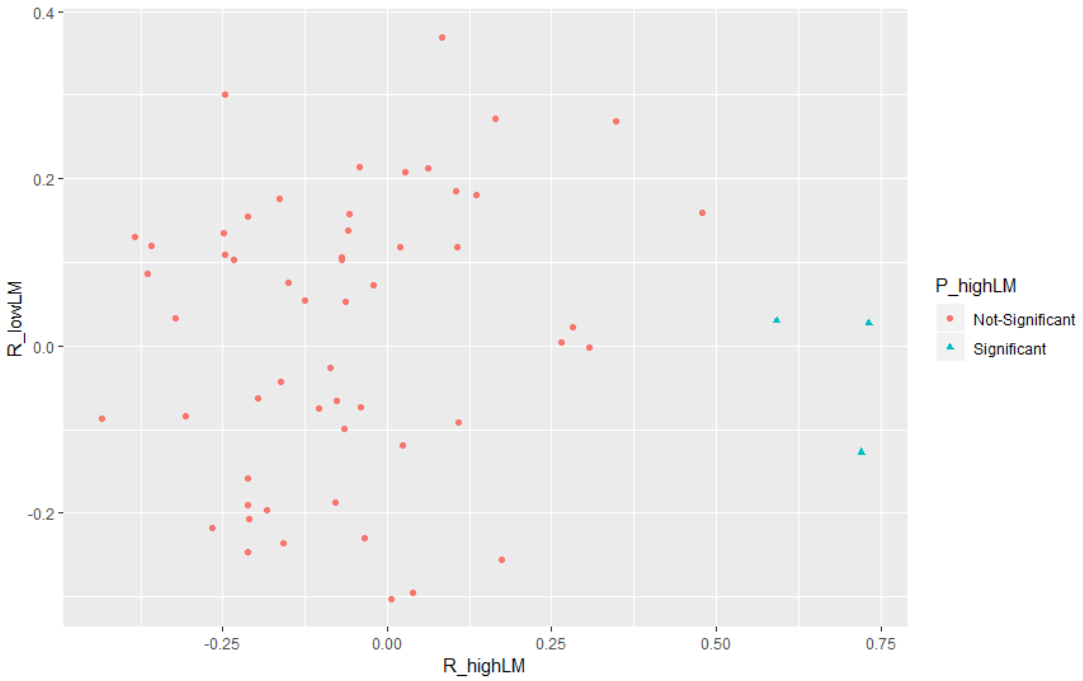


Fig 13: The scatter-plot of correlation of TLR-4 with WNT signaling gene sets with X and Y axis representing the PCC of genes with TLR-4 in 'Severe EED' and 'Healthy' groups. The blue-triangle points represent genes with P value less than 0.1.

f. Correlation of TLR-4 with TLR Signaling Gene sets

A similar analysis was done, and the following plot was produced (Fig 14). By comparing the two correlation networks (Fig 15), the following genes were mostly negatively correlated with other genes in 'Healthy' group but positively correlated with those genes in 'Severe EED' group: S100A12, TLR6, UBA52, NFKBIB, CD14, MAP2K6 and MAP3K1.

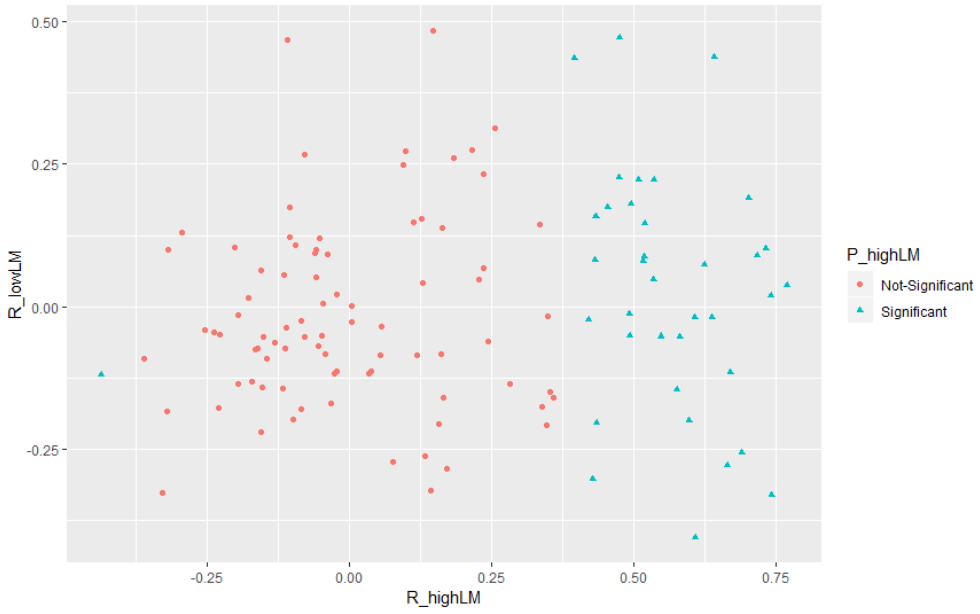
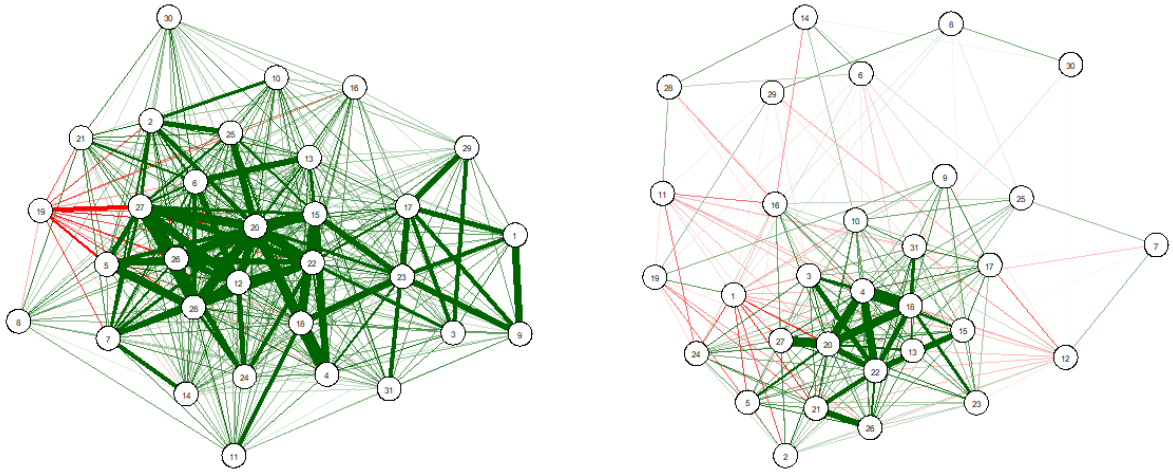


Fig 14: The scatter-plot of correlation of TLR-4 with TLR signaling gene sets with X and Y axis representing the PCC of genes with TLR-4 in 'Severe EED' and 'Healthy' groups. The blue-triangle points represent genes with P value less than 0.1.



A B

Fig 15: The correlation network of TLR signaling genes (that significantly correlated with TLR-4) in 'Severe EED' group (A) and 'Healthy' group (B).

g. Correlation of TLR-4 with TNFα Signaling via NF-κB Gene sets

The TNFα Signaling via NF-κB gene sets were correlated with TLR-4 in both the groups and the Pearson's r of those genes with TLR-4 were plotted (Fig 16)

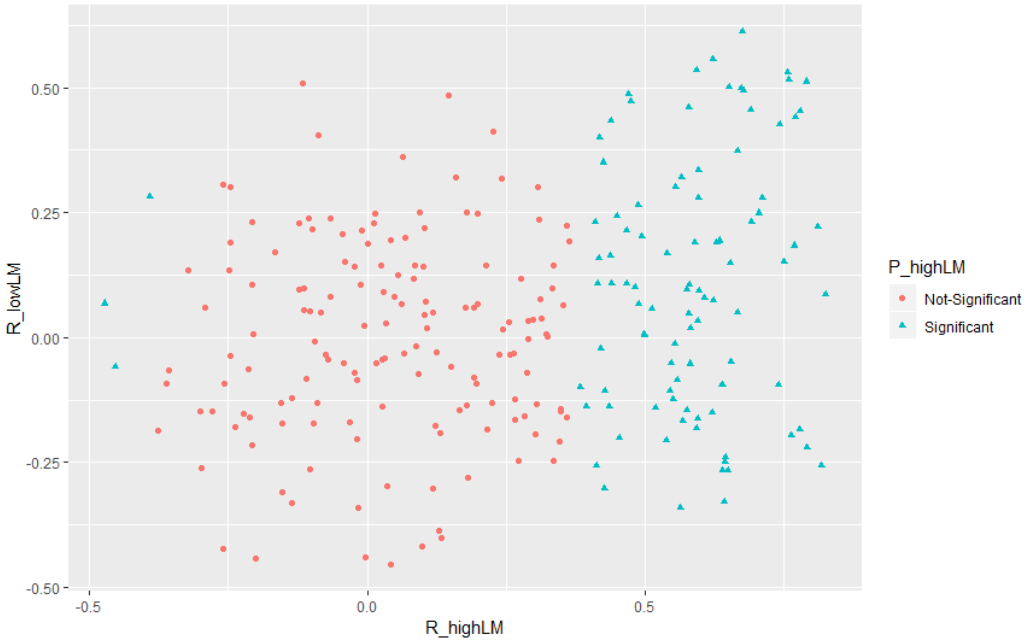
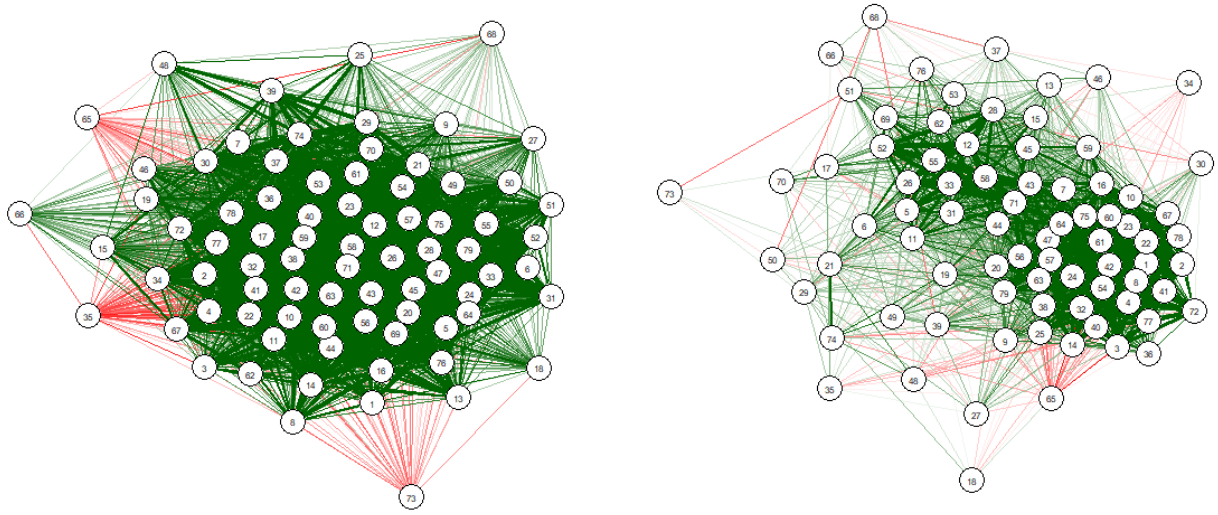


Fig 16: The scatter-plot of correlation of TLR-4 with TNFα signaling gene sets with X and Y axis representing the PCC of genes with TLR-4 in 'Severe EED' and 'Healthy' groups. The blue-triangle points represent genes with P value less than 0.1.

There were genes that are significantly and positively correlated with TLR-4 in Severe EED group (R_highLM) but negatively correlated with Healthy group (R_lowLM) and vice versa. These significantly correlated genes were listed and correlated among themselves. The correlation networks of the significant gene sets for the two groups were plotted using 'qgraph' package (Fig 17). SLC16A6, IL23A, DENND5A and DRAM1 were the genes that are mostly negatively correlated with other genes in the 'Healthy' group but positively correlated with those genes in 'Severe EED' group.



A B

Fig 17: The correlation network of TNFα signaling through NF-κB genes (that significantly correlated with TLR-4) in 'Severe EED' group (A) and 'Healthy' group (B).

7. Discussion

7.1 Demographic Analysis

The plot depicting the variation of %L values with the frequency of the individuals (Fig 1) shows that the individuals recruited for the study were at a risk group of EED as there was a large number of people who have a mild intestinal epithelial junction disruption (indicative of their %L values above 0.2). Among the 259 individuals, 60 of them belong to the 'Healthy' group, 161 belong to the 'intermediate EED' group, and remaining 42 belong to the 'Severe EED' group.

In fig 2, there were more individuals with high %L values in the age group of 20 – 40 months compared to that of individuals in the age groups below 20 months and above 40 months. This can conclude two things: EED can have an age-dependent effect, or the inclusion criteria of the study could have resulted in the omission of high %L value individuals in the age groups below 20 months and above 40 months. Most likely, the inclusion criteria may have caused this non-uniform sample. However, to check whether EED has an age-dependent effect, one can design a study which regularly estimates the variation of %L value of the individuals who were at risk for EED over a period of time.

In the case of distribution of gender within each group of the study (Fig 3), the trend observed was that the female fraction in each group decreases from 'Healthy' to 'Severe EED' groups and vice versa for the male fraction. This can be due to several reasons: the problem in inclusion criteria which leads to sampling error for this particular analysis, uneven recruitment (in the context of gender) of individuals for the study, the social stigma of the community, or even due to differential treatment between males and females in the community. In any case, this dominance of male

and female individuals in ‘Severe EED’ and ‘Healthy’ groups has led to genes that were sex-linked to be differentially expressed in the two groups when they were analysed by GEO2R’s ‘TOP250 Differentially Expressed Genes’ (This was a part of the analysis which was dropped due to the above reason).

7.2 Transcriptome Analysis between Healthy and Severe EED Groups

The correlation networks of 51 genes in ‘Severe EED’ and ‘Healthy’ groups were visually striking (Fig 5). In the correlation network of ‘Severe EED’ group, the 51 genes analysed were highly correlated among themselves whereas, in case of ‘Healthy’ group, the strength of the correlation decreases and even some of them were negatively correlated with most of the other genes. Those negatively correlated genes in the ‘Healthy’ group include LAPTM5, IFITM1, BIN2, CR1, S100A12.

LAPTM5 codes the protein that promotes caspase-independent cell death in mammalian cells[14] and a regulator of proinflammatory responses by macrophages through the activation of NF-κB and MAPK signaling[15]. IFITM1 influences interferon-gamma signaling[16], CR1 was responsible for binding and clearing of immune complexes that were complement activated[17], and S100A12 plays a vital role in TLR-4 signaling and regulates cytokine production[18]. In short, all these genes have pro-inflammatory responses which were activated due to microbial invasion. The change in microbiota in the gut due to poor sanitation and improper hygiene practices results in upregulation of inflammatory responses which distrusts the mucous layer and cellular integrity. Since this was a constant exposure, the immune system produces chronic inflammatory responses.

7.3 Correlation of TLR-4 with Different gene sets

A general trend in the correlation network comparison between the two groups was that most of the correlations between genes in the ‘Severe EED’ group have higher PCC compared to that of in the ‘Healthy’ group. This strongly suggests that the gene sets chosen for the analysis were mostly transcriptionally active (the transcriptome data was acquired from faecal mRNA) in individuals in ‘Severe EED’ group when compared to individuals in ‘Healthy’ group.

Another observation was that in the correlation of TLR-4 with apoptotic, inflammatory and TLR signaling gene sets, CD14 positively correlates with most of the genes in the respective gene set in ‘Severe EED’ group but negatively correlates with those same genes in the ‘Healthy’ group. This largely signifies that CD14, being a co-receptor of activation of LPS along with TLR-4 and MD-2 receptors[19], plays a role in the pathobiology of EED.

One of the interesting observations was that PECAM1 negatively correlates with other genes which regulate cell junction in the ‘Severe EED’ group but positively correlates with the same genes in the ‘Healthy’ group (Fig 11). PECAM1, or generally called as CD31, was mostly found in endothelial cells, monocytes, macrophages, DC and T cell subsets, and acts as an adhesion molecule in those cells[20]. It helps in the transmigration of immune cells[21] and plays a vital role in the inhibition of mitochondrial-mediated apoptosis[22]. Its negative correlation in ‘Severe EED’ group suggests its hyperresponsiveness in cell adhesion regulation, thus leading to disrupted small bowel in the individuals belonging to the ‘Severe EED’ group.

In the correlation network of inflammatory gene sets (Fig 9), the genes were strongly correlated among themselves in the ‘Severe EED’ group, but the correlation weakens

in the ‘Healthy’ group. This means that the genes that were responsible for inflammatory responses were upregulated in case of EED. This was expected as the study concludes stating that various pro-inflammatory responses associated with the %L values (indirectly, with the severity of EED)[1].

It can be seen that TLR-4 correlated significantly with a large number of genes in the apoptotic, inflammatory, cell junction regulatory, TLR and TNFα signaling gene sets. These gene sets were significantly expressed in EED affected individuals and their correlation with TLR-4 signifies that TLR-4 can be a biomarker for EED.

With this secondary analysis, there is a possibility of TLR-4 playing an important role in EED as TLR-4 significantly correlates (P-value < 0.1) with a high number of genes which have a characteristic feature (inflammatory, apoptotic, cell junction regulatory, TLR and TNFα signaling) that can found in the EED individuals. However, this must be validated experimentally before concluding TLR-4 as a potential biomarker in the pathobiology of EED.

8. Acknowledgement

I would like to express my deepest appreciation to all those who provided me with the chance to work and learn from CMC Vellore. I would like to thank Dr Gagandeep Kang, who gave me this opportunity to meet and work with many wonderful people and professionals in CMC Vellore. I express my special gratitude to Dr Savit B Prabhu, who in spite of being extraordinarily busy with his duties, took time out to hear, guide and keep me on the correct path throughout the project.

I also express my deepest thanks to Ms Akshaya Balasubramanian and Dr Punithavathy P M for their careful and precious guidance and advice, which were extremely valuable for my project both theoretically and practically.

I would also like to thank Mr Johnson J B and Ms Soundari B, who helped me to learn the technical aspects in the lab and supported me whenever I needed it. Furthermore, I would like to thank CMC Vellore for allowing me into their institution and help me build my passion for research.

9. Bibliography

1. Yu, J. *et al.* Environmental Enteric Dysfunction Includes a Broad Spectrum of Inflammatory Responses and Epithelial Repair Processes. *Cell. Mol. Gastroenterol. Hepatol.* **2**, 158-174.e1 (2016).
2. Prendergast, A. & Kelly, P. Enteropathies in the developing world: neglected effects on global health. *Am. J. Trop. Med. Hyg.* **86**, 756–63 (2012).
3. Mbuya, M. N. N. & Humphrey, J. H. Preventing environmental enteric dysfunction through improved water, sanitation and hygiene: an opportunity for stunting reduction in developing countries. *Matern. Child Nutr.* **12 Suppl 1**, 106–20 (2016).
4. Korpe, P. S. & Petri, W. A. Environmental enteropathy: critical implications of a poorly understood condition. *Trends Mol. Med.* **18**, 328–36 (2012).
5. Sadiq;, A. A. I., Iqbal, N. T. & Sadiq, K. Environmental enteropathy. *Curr. Opin. Gastroenterol.* **32**, 12–17 (2016).
6. Keusch, G. T. *et al.* Implications of Acquired Environmental Enteric Dysfunction for Growth and Stunting in Infants and Children Living in Low- and Middle-Income Countries. *Food Nutr. Bull.* **34**, 357–364 (2013).
7. Bhutta, Z. A., Guerrant, R. L. & Nelson, C. A. Neurodevelopment, Nutrition, and Inflammation: The Evolving Global Child Health Landscape. *Pediatrics* **139**, S12–S22 (2017).
8. John, C. C., Black, M. M., Nelson, C. A. & III. Neurodevelopment: The Impact of Nutrition and Inflammation During Early to Middle Childhood in Low-Resource Settings. *Pediatrics* **139**, S59–S71 (2017).
9. Kosek, M. *et al.* Assessment of Environmental Enteropathy in the MAL-ED Cohort Study: Theoretical and Analytic Framework. *Clin. Infect. Dis.* **59**, S239–S247 (2014).
10. Denno, D. M. *et al.* Use of the Lactulose to Mannitol Ratio to Evaluate Childhood Environmental Enteric Dysfunction: A Systematic Review. *Clin. Infect. Dis.* **59**, S213–S219 (2014).
11. Ngure, F. M. *et al.* Water, sanitation, and hygiene (WASH), environmental enteropathy, nutrition, and early child development: making the links. *Ann. N. Y. Acad. Sci.* **1308**, 118–128 (2014).
12. Neal, M. D. *et al.* Toll-like Receptor 4 Is Expressed on Intestinal Stem Cells and Regulates Their Proliferation and Apoptosis via the p53 Up-regulated Modulator of Apoptosis. *J. Biol. Chem.* **287**, 37296–37308 (2012).
13. About GEO2R - GEO - NCBI. Available at: https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html#value_distribution. (Accessed: 8th July 2019)
14. Mrschtik, M. & Ryan, K. M. Lysosomal proteins in cell death and autophagy. *FEBS Journal* **282**, 1858–1870 (2015).
15. Glowacka, W. K., Alberts, P., Ouchida, R., Wang, J. Y. & Rotin, D. LAPTM5 protein is a positive regulator of proinflammatory signaling pathways in macrophages. *J. Biol. Chem.* **287**, 27691–27702 (2012).
16. Tissue expression of IFITM1 - Summary - The Human Protein Atlas. Available at: https://www.proteinatlas.org/ENSG00000185885-IFITM1/tissue.

17. Tissue expression of CR1 - Summary - The Human Protein Atlas. Available at: https://www.proteinatlas.org/ENSG00000203710-CR1/tissue.

18. Tissue expression of S100A12 - Summary - The Human Protein Atlas. Available at: https://www.proteinatlas.org/ENSG00000163221-S100A12/tissue.

19. Abbas, A. K. & Lichtman, A. H. *Basic immunology : Functions and Disorders of the Immune System*. *Nature* **272**, (Elsevier Inc., 2005).

20. Watt, S. M., Gschmeissner, S. E. & Bates, P. A. PECAM-1: Its Expression and Function as a Cell Adhesion Molecule on Hemopoietic and Endothelial Cells. *Leuk. Lymphoma* **17**, 229–244 (1995).

21. Newman, D. K., Hamilton, C. & Newman, P. J. Inhibition of antigen-receptor signaling by platelet endothelial cell adhesion molecule-1 (CD31) requires functional ITIMs, SHP-2, and p56lck. *Blood* **97**, 2351–2357 (2001).

22. Jones, K. L. *et al.* PECAM-1 functions as a specific and potent inhibitor ofmitochondrial-dependent apoptosis. *Blood* **98**, 1456–1463 (2001).