ejpmr, 2022, 9(9), 130-140

SJIF Impact Factor 6.222

EUROPEAN JOURNAL OF PHARMACEUTICAL AND MEDICAL RESEARCH

www.ejpmr.com

Research Article

ISSN 2394-3211

EJPMR

TRANSCRIPTOMIC ANALYSIS OF DIFFERENTIAL GENE EXPRESSION IN *STAPHYLOCOCCUS AUREUS*-INDUCED PNEUMONIA IN PEDIATRICS BASED ON MICROARRAY ANALYSIS

Leonard O. Ehigie[1*], Fiyinfoluwa D. Ojeniyi[1], Adetayo Aborisade[2], Sulaimon Nassar[3], Christopher Igbeneghu[3], James Busayo Agboola[4] and Adeola F. Ehigie[1*]

[1]Department of Biochemistry, Ladoke Akintola University of Technology, Ogbomoso, Nigeria.
[2]Department of Oral Diagnostics Sciences, Bayero University, Kano.
[3]Department of Medical Laboratory Science, College of Health Sciences, Ladoke Akintola University of Technology, Ogbomoso, Nigeria.
[4]Department of Biochemistry and Molecular Biology, Faculty of Sciences, Obafemi Awolowo University, Ile-Ife.

***Corresponding Author: Leonard O. Ehigie**

Department of Biochemistry, Ladoke Akintola University of Technology, Ogbomoso, Nigeria.

***Article* Received on 11/07/2022 Article Revised on 01/08/2022 Article Accepted on 22/08/2022**

ABSTRACT

Background: *Staphylococcus aureus* is a leading cause of about 80% of infections in humans and up to 60–70% of hospital-acquired infections. Identification of an etiologic agent for pneumonia is critical in order to provide appropriate therapy and maintain epidemiological records. Study on the transcriptional profiling of patients infected with *S. aureus* is a pivot to the analysis of differentially expressed gene in the blood of patients infected with *S. aureus.* This study performed the analysis of gene expression dataset GSE30119 available on the Gene expression Omnibus (GEO) which is based on the hypothesis tested that patient clinical heterogeneity will be reflected in transcriptional profile heterogeneity. **Methods:** Gene expression profile dataset GSE30119 was obtained from Gene Expression Omnibus (GEO). We performed bioinformatic analysis to identify Differentially Expressed Genes in *S. aureus* infection induced Pneumonia from the transcriptional level. The study comprised 143 pediatric patients, with 44 healthy individuals, 81 pneumonia-free, and 18 pneumonia infection patients. Gene Ontology (GO) analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) functional, and pathway enrichment analyses were performed using Metascape. STRING and Cytoscape were used for the construction and visualization of biological networks. Hub-genes were identified through degree of association of interaction networks. **Results:** We discovered a total of 54 genes related to *S. aureus* infection and 612 genes associated with Pneumonia. According to Gene Ontology (GO) functional and pathway enrichment studies, the *S. aureus* infection associated genes are predominantly engaged in the innate immune response, calcium-mediated signaling, Neutrophil extracellular trap formation, Formation of Fibrin Clot (Clotting Cascade). Whereas the genes associated with Pneumonia are enriched in adaptive immune response, inflammatory, Interferon alpha/beta signaling, TCR signaling, Gα(i) signalling events. **Conclusions:** This study shows differentially expressed genes and their biological activities in relation to *S. aureus* infection and Pneumonia, and it may provide more light on the underlying molecular mechanisms and possibly important gene signatures in Pneumonia development.

KEYWORDS: STRING and Cytoscape were used for the construction and visualization of biological networks.

INTRODUCTION

Pneumonia is responsible for significantly more disease than cancer, diabetes, HIV/AIDS, malaria, and many other diseases that are regarded as important global health concerns (Mizgerd, 2012). Pneumonia has the highest global burden of disease, owing to the fact that it kills more children than any other disease (Rudan, 2008). In elderly people, pneumonia hospitalization is linked to a higher risk of death than any other common reason of hospitalization (Maurer, 2007). Pneumonia is a leading cause of death than any other infectious disease on the around the globe (Troeger *et al.,* 2018). In the United States, the financial repercussions are considerable, ranging from about 20 billion dollars to more than 80 billion dollars every year (Andrejko *et al.,* 2021). There are also indirect and long-term consequences, such as cognitive decline comparable to traumatic brain injury, increased incidence and severity of depression, worsened cardiovascular and cerebrovascular health, physical limitations, and a shorter life span, after all of this physical discomfort and expenditure (Leung *et al.,* 2020). Pneumonia demands special attention from the biomedical profession since it is a clear risk factor for mortality, as well as a factor in improper ageing and degeneration. Since microbial infection can lead to pneumonia, the host's response has an impact on the

disease's pathogenesis (Maggini *et al.,* 2018). Pneumonia is a respiratory disease in the host, however it is a complex disease involving numerous physiological systems working together. Despite being a brief sickness, it is associated with chronic conditions that have protracted consequences (Quinton *et al.,* 2018).

Pneumonia has a wide range of risk factors ranging from poor nutrition to microbial infections (Bacteria, Viruses and fungi). Microoganisms such as Streptococcus pneumoniae and Haemophilus influenza, corona virus, rhinovirus, human metapneumovirus, human bocavirus, parainfluenza, Respiratory syncytial virus, Mycoplasma pneumonia, S. pneumoniae, Staphylococcus aureus, Moraxella catarrhalis, H. influenzae, Mycoplasma pneumonia, Chlamydophila pneumoniae etc. The advent of molecular diagnostics techniques such as polymerase chain reachtion (PCR), Microarray techniques, and next-generation sequencing (NGS) has been helpful in the detection of the pathogens associated with Pneumonia (Bhuiyan *et al.,* 2018). However, NGS has proven to be superior over the traditional or conventional diagnostic techniques. In severe pneumonia, NGS may lead to a quick and effective diagnosis with a better clinical outcome than traditional detection approaches. It first demonstrates that NGS may swiftly provide etiology proof for severe pneumonia patients, guide clinic care, and ultimately reduce mortality (Li *et al.,* 2021).

Staphylococcus aureus being a common cause of hospital-acquired infection with Pneumonia recognized as the second most common hospital-associated infection, we assayed to study the biomarkers which may be linked to Pneumonia infection in *staphylococcus* infected patients. Due to the emergence antibiotic resistant strains of *Staphylococcus aureus*, e.g. Methicillin resistant *Staphylococcus aureus* (MRSA), there has been also increase in the number of infections arising from it. A longitudinal study of roughly 10 million pneumonia cases requiring hospitalization found that *Staphylococcus aureus* pneumonia was identified as the primary diagnosis in just 1.08 percent of the cases (Jacobs and Shaver, 2017). *Staphylococcus aureus* has long been recognised to play a big role in the development of pneumonia, and its importance as a pneumonia pathogen was recently demonstrated in an observational study in different Intensive care unit across Europe (Paling *et al.,* 2020). Furthermore, SARS-CoV-2 patient morbidity and death have recently been linked to *Staphylococcus aureus* pneumonia. Furthermore, SARS-CoV-2 patient morbidity and death have recently been linked to *Staphylococcus aureus* pneumonia (Lai *et al.,* 2020). The fact that S. aureus is multidrug-resistant adds to the problem's complexity. The nares and extranasal locations, such as the epidermis, perineum, and throat, have been proven to be colonised by S aureus, particularly MRSA (Gagnaire *et al.,* 2017). The absence of nasal colonization has been linked to a reduced risk of future MRSA infection (Kapali *et al.,* 2021). When the nares are colonized, S aureus has oppotunity to hide from the host's defenses, which can lead to infection if the host's defenses are breached (Ajayi, 2018). This study was carried out based on the hypothesis that "transcriptional profile heterogeneity will reflect patient clinical heterogeneity" and also identify gene signatures that may serve as biomarkers of *staphylococcus infection* in human. It is aimed at investigating the biomarker panel of pneumonia infection caused by *staphylococcus aureus.*

MATERIALS AND METHODS

Transcriptomic dataset

The original submitter-supplied dataset GSE30119 was obtained from GEO (http://www.ncbi.nlm.nih.gov/geo/), which was based on the platform of GPL6947 Illumina Human HT-12 V3.0 expression beadchip. The data was submitted by Banchereau et al., 2012 collected from Genome-wide analysis of whole blood transcriptional response to community-acquired *Staphylococcus aureus* infection in vivo Total RNA extracted from whole blood (lysed in Tempus tubes) drawn from pediatric patients with acute community-acquired *Staphylococcus aureus* infection. A total of 143 samples are included in this dataset, comprising 44 healthy individuals, 81 pneumonia-free, and 18 pneumonia infection patients. Total RNA extracted from whole blood was utilized for gene expression microarrays. This dataset was generated using the platform GPL6947 Illumina HumanHT-12 V3.0 expression beadchip.

Differential gene expression analysis

Data pre processing was performed using GEO2R (https: //www.ncbi.nlm.nih.gov/geo/geo2r) and was applied to screen Differentially Expressed Genes (DEGs) between the following groups: staphylococcus infection (SI) vs. Healthy (H), Pneumonia-free (PF) vs Healthy and Pneumonia infection (PI) vs. Healthy. GEO2R is a web-based tool that allows users to compare two or more groups of Samples in a GEO Series to find genes that are differentially expressed under different experimental settings. The results are supplied as a table of genes ordered by significance, as well as a collection of graphic plots to aid in the visualization of differentially expressed genes and the evaluation of data set quality. Using the Bioconductor project's GEOquery and limma R packages, GEO2R compares original submitter-supplied processed data tables. Bioconductor is an open source software project that provides tools for analyzing high-throughput genetic data. It is based on the R programming language. The R package GEOquery parses GEO data into R data structures that other R tools can use. Log transformation was applied to the data. The adjusted $P<0.01$ and $|\log_2$ fold change (FC)$| >1$ (i.e., FC >2) were selected as the threshold for each group.

Venn Diagram Analysis of DEGs

Venn diagram for DEGs of the comparison groups was constructed using Venny (http://bioinfogp.cnb.csic. es/tools/venny/index.html). The similarities and differences in three comparison groups were observed.

The DEGs that overlap the three comparison groups were recognized as genes associated with *S. aureus* infection. The other DEGs, observed between Pneumonia vs. Healthy but not Pneumonia free vs. Healthy were identified as Pneumonia-associated genes associated.

Functional, pathway enrichment Analysis

To undertake enrichment analysis for the DEGS, the Metascape database for annotation, visualization, and integrated discovery (http://metascape.org) was introduced. As an enrichment background, all genes in the genome were employed. Terms with a P-value less than .01, a minimum count of 3, and an enrichment factor more than 1.5 (the enrichment factor is the ratio between the observed counts and the counts expected by chance) are gathered and classified into clusters based on membership commonalities. To adjust for repeated testings, P-values are determined using the accumulative hypergeometric distribution (Karp *et al.,* 2021), and q-values are calculated using the Banjamini–Hochberg technique (Menyhart *et al.,* 2021). When doing hierarchical clustering on the enriched terms, Kappa scores (Gu and Huebschmann, 2021) are used as the similarity metric, and sub-trees with a similarity of >0.3 are deemed a cluster. Protein-protein interaction enrichment analysis was performed on each gene list using the STRING database (Szklarczyk *et al.,* 2016). In STRING (physical score > 0.132), only physical interactions are exploited. The resulting network comprises the proteins that have at least one physical contact with another member of the list. The Molecular Complex Detection (MCODE) algorithm has been used to discover highly coupled network components in networks with between 3 and 500 proteins.

Protein-protein Interaction analysis

Protein products of the differentially expressed genes were obtained from the String database (http://string-db.org/) and used to construct a network of protein-protein interaction profile. This database is one of the Cytoscape software 3.4.0 apps (Yan et al., 2017) that gives interaction information from three separate panels, including disease, protein, and PubMed queries. The strength of protein interactions can be fitted for the network construction (Szklarczyk et al., 2016). It's set to 0.4, which is the default value. The proteins were analyzed via the undirected edges methods by Cytoscape software. The gene expression data obtained from gene expression analysis was used to layout the network. The nodes (proteins) were mapped continuously using the log2foldchange value. Red colour indicates upregulation while blue colour indicates downregulation.

RESULTS

Identification of DEGs

The gene expression dataset GSE30119 was downloaded from the GEO database. DEGs between the disease and healthy samples were determined using the GEO2R tool. As presented in Fig. 1, a total of 821 DEGs were identified in the all the comparison groups using the threshold of P<0.05 and $|log_2FC|$ >1, including 488 upregulated genes and 333 downregulated genes. The top 10 up- and downregulated genes for each comparison group are listed in Table 1.

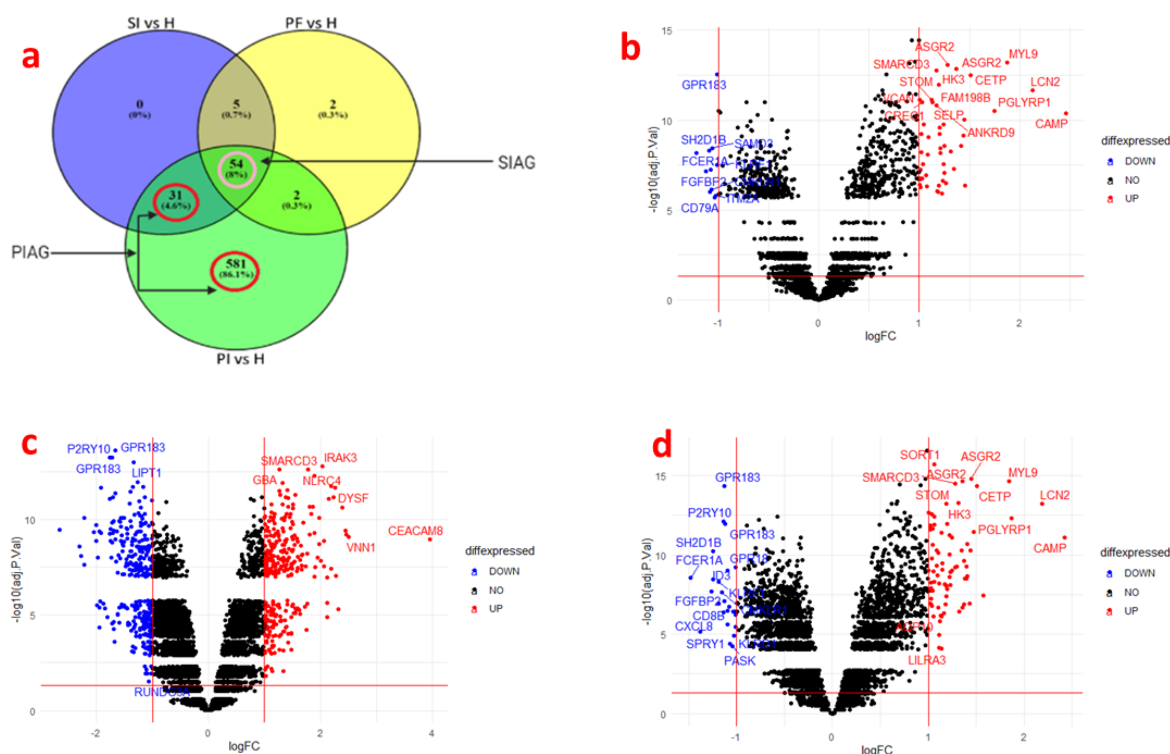

Figure 1: Differential gene expression analysis (a) Intersection of differentially expressed genes (DEGs). Staphylococcus infection associated genes (SIAGs) are marked by pink circles and Pneumonia infection

associated genes (PIAGs) are marked by red circle. Volcano Plot representation of the transcriptomic analysis of differentially expressed genes between (b) *Staphylococcus aureus* infected and healthy individuals (c) Pneumonia infected and healthy individuals (d) Pneumonia free and healthy individuals. The black points (NO) stands for genes that have a fold change less than 1.0. The blue points represent the genes which fold change is lower than -1.0 but their p-value is lower than 0.05 (down regulated genes). The genes depicted by red points have a p-value lower than 0.05 and a fold change higher than 1.0 (upregulated genes).

Table 1a: The top upregulated and downregulated genes between staphylococcus infection and healthy patients (ranked by Log2-fold change).

Gene.symbol	adj.P.Val	logFC	Gene.title	Regulation
CAMP	7.88E-12	2.420397	cathelicidin antimicrobial peptide	Upregulated
LCN2	6.11E-14	2.187087	lipocalin 2	Upregulated
PGLYRP1	5.09E-13	1.867231	peptidoglycan recognition protein 1	Upregulated
MYL9	2.16E-15	1.847294	myosin light chain 9	Upregulated
ANKRD22	3.55E-08	1.577498	ankyrin repeat domain 22	Upregulated
CETP	4.55E-15	1.509164	cholesteryl ester transfer protein	Upregulated
RNASE2	3.49E-12	1.475392	ribonuclease A family member 2	Upregulated
ASGR2	1.49E-15	1.452085	asialoglycoprotein receptor 2	Upregulated
PLOD2	5.07E-11	1.440901	procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	Upregulated
KLRF1	4.28E-07	-1.134533	killer cell lectin like receptor F1	Downregulated
P2RY10	7.89E-13	-1.1355418	purinergic receptor P2Y10	Downregulated
CMKLR1	2.38E-08	-1.1554959	chemerin chemokine-like receptor 1	Downregulated
CD79A	1.39E-06	-1.1817243	CD79a molecule	Downregulated
GZMB	1.13E-07	-1.1865535	granzyme B	Downregulated
SH2D1B	5.90E-11	-1.243959	SH2 domain containing 1B	Downregulated
FGFBP2	1.95E-08	-1.259197	fibroblast growth factor binding protein 2	Downregulated
CXCL8	7.15E-06	-1.3766733	C-X-C motif chemokine ligand 8	Downregulated
FCER1A	2.91E-09	-1.4779835	Fc fragment of IgE receptor Ia	Downregulated

Table 1b: The top 10 upregulated and downregulated genes between pneumonia free but staphylococcus infected and healthy patients (ranked by Log2-fold change).

Gene.symbol	adj.P.Val	logFC	Gene.title	Regulation
CAMP	4.34E-11	2.459041	cathelicidin antimicrobial peptide	Upregulated
LCN2	2.10E-12	2.126062	lipocalin 2	Upregulated
MYL9	6.18E-14	1.876577	myosin light chain 9	Upregulated
PGLYRP1	3.16E-11	1.751401	peptidoglycan recognition protein 1	Upregulated
CETP	3.12E-13	1.509559	cholesteryl ester transfer protein	Upregulated
ANKRD22	4.34E-07	1.457372	ankyrin repeat domain 22	Upregulated
RNASE2	9.74E-11	1.450099	ribonuclease A family member 2	Upregulated
NDUFAF3	7.08E-10	1.440268	NADH:ubiquinone oxidoreductase complex assembly factor 3	Upregulated
PLOD2	2.69E-09	1.413943	procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	Upregulated
ASGR2	1.39E-13	1.372498	asialoglycoprotein receptor 2	Upregulated
ITM2A	1.64E-06	-1.0254427	integral membrane protein 2A	Downregulated
CD79A	1.98E-06	-1.0341676	CD79a molecule	Downregulated
SAMD3	3.59E-09	-1.0604968	sterile alpha motif domain containing 3	Downregulated
GZMB	2.70E-06	-1.064022	granzyme B	Downregulated
CMKLR1	7.71E-07	-1.072244	chemerin chemokine-like receptor 1	Downregulated
FGFBP2	1.03E-06	-1.0850047	fibroblast growth factor binding protein 2	Downregulated
SH2D1B	4.98E-09	-1.0950522	SH2 domain containing 1B	Downregulated
CXCL8	1.11E-04	-1.1211511	C-X-C motif chemokine ligand 8	Downregulated
FCER1A	6.95E-09	-1.2169814	Fc fragment of IgE receptor Ia	Downregulated
CXCL8	6.93E-05	-1.2487535	C-X-C motif chemokine ligand 8	Downregulated

Table 1c: The top 10 upregulated and downregulated genes between staphylococcus infected with pneumonia infection and healthy patients (ranked by Log2-fold change).

Gene.symbol	adj.P.Val	logFC	Gene.title	Regulation
CEACAM8	1.11E-09	3.95031	carcinoembryonic antigen related cell adhesion molecule 8	Upregulated
VNN1	8.25E-10	2.507865	vanin 1	Upregulated
LCN2	6.17E-10	2.4617	lipocalin 2	Upregulated
OPLAH	3.75E-10	2.450587	5-oxoprolinase (ATP-hydrolysing)	Upregulated
PGLYRP1	2.34E-11	2.388465	peptidoglycan recognition protein 1	Upregulated
CYSTM1	5.03E-06	2.312637	cysteine rich transmembrane module containing 1	Upregulated
CA4	9.30E-08	2.276497	carbonic anhydrase 4	Upregulated
DYSF	2.18E-12	2.267905	dysferlin	Upregulated
CAMP	7.83E-07	2.246501	cathelicidin antimicrobial peptide	Upregulated
TLR5	6.67E-12	2.228208	toll like receptor 5	Upregulated
FAM102A	4.43E-05	-1.95014138	family with sequence similarity 102 member A	Downregulated
CXCL8	3.93E-06	-1.9523124	C-X-C motif chemokine ligand 8	Downregulated
LRRC26	3.19E-06	-1.99051709	leucine rich repeat containing 26	Downregulated
KLRF1	1.28E-09	-2.00306038	killer cell lectin like receptor F1	Downregulated
FGFBP2	4.77E-10	-2.04306243	fibroblast growth factor binding protein 2	Downregulated
KLRF1	2.30E-08	-2.21270306	killer cell lectin like receptor F1	Downregulated
S1PR5	8.14E-11	-2.2156796	sphingosine-1-phosphate receptor 5	Downregulated
TRAV20	8.49E-09	-2.27602061	T cell receptor alpha variable 20	Downregulated
KLRD1	2.59E-09	-2.30542009	killer cell lectin like receptor D1	Downregulated
FCER1A	3.35E-10	-2.65249304	Fc fragment of IgE receptor Ia	Downregulated

Enrichment analysis

Pathway and process enrichment analysis

For each given gene list, pathway and process enrichment analysis has been carried out with the following ontology sources: KEGG Pathway, GO Biological Processes, Reactome Gene Sets. All genes in the genome have been used as the enrichment background. Terms with a p-value < 0.01, a minimum count of 3, and an enrichment factor > 1.5 (the enrichment factor is the ratio between the observed counts and the counts expected by chance) are collected and grouped into clusters based on their membership similarities. The genes associated with staphylococcus infection, being 54 in number from differential gene expression analysis are enriched several pathways and processes as presented in Table 2. The table only shows the top 16 clusters with their representative enriched terms (one per cluster). "Count" is the number of genes in the user-provided lists with membership in the given ontology term. "%" is the percentage of all of the user-provided genes that are found in the given ontology term (only input genes with at least one ontology term annotation are included in the calculation). "Log10(P)" is the p-value in log base 10. "Log10(q)" is the multi-test adjusted p-value in log base 10. The enrichment category includes 9 GO Biological Processes (innate immune response, cell activation, muscle cell differentiation, calcium-mediated signaling, positive regulation of response to external stimulus, antimicrobial humoral immune response mediated by antimicrobial peptide, circulatory system process, cellular modified amino acid metabolic process, negative regulation of apoptotic signaling pathway), 4 Reactome Gene Sets (Neutrophil degranulation, Formation of Fibrin Clot/Clotting Cascade, Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell, Cell surface interactions at the vascular wall, Diseases of metabolism). Only one KEGG Pathway category (Neutrophil extracellular trap formation) and WikiPathways (miRNAs involvement in the immune response in sepsis) are included. While for the Pneumonia related genes, the top processes involved are response to cytokine, leukocyte activation, Neutrophil degranulation, Hematopoietic cell lineage, positive regulation of cell-cell adhesion, Cytokine Signaling in Immune system, Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell, positive regulation of cytokine production, Th1 and Th2 cell differentiation, positive regulation of cytokine production, Th1 and Th2 cell differentiation, positive regulation of immune response, regulation of defense response, inflammatory response, Network map of SARS-CoV-2 signaling pathway, negative regulation of immune system process, negative regulation of immune system process, T cell differentiation involved in immune response, Inflammatory bowel disease, positive T cell selection and Cell surface interactions at the vascular wall.

Table 2: Top 16 clusters with their representative enriched terms (one per cluster) for staphylococcus infection related genes.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
R-HSA-6798695	Reactome Gene Sets	Neutrophil degranulation	12	22.22	-10.37	-6.03
R-HSA-140877	Reactome Gene Sets	Formation of Fibrin Clot (Clotting Cascade)	4	7.41	-6.15	-2.18
GO:0045087	GO Biological Processes	innate immune response	10	18.52	-6.05	-2.18
GO:0001775	GO Biological Processes	cell activation	9	16.67	-5.76	-2.09
R-HSA-198933	Reactome Gene Sets	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	5	9.26	-5.41	-1.9
GO:0042692	GO Biological Processes	muscle cell differentiation	5	9.26	-4	-0.94
GO:0019722	GO Biological Processes	calcium-mediated signaling	4	7.41	-3.87	-0.88
GO:0032103	GO Biological Processes	positive regulation of response to external stimulus	6	11.11	-3.87	-0.88
WP4329	WikiPathways	miRNAs involvement in the immune response in sepsis	3	5.56	-3.7	-0.75
GO:0061844	GO Biological Processes	antimicrobial humoral immune response mediated by antimicrobial peptide	3	5.56	-3.55	-0.68
hsa04613	KEGG Pathway	Neutrophil extracellular trap formation	4	7.41	-3.43	-0.62
GO:0003013	GO Biological Processes	circulatory system process	5	9.26	-2.73	-0.19
R-HSA-202733	Reactome Gene Sets	Cell surface interactions at the vascular wall	3	5.56	-2.72	-0.18
GO:0006575	GO Biological Processes	cellular modified amino acid metabolic process	3	5.56	-2.38	0
GO:2001234	GO Biological Processes	negative regulation of apoptotic signaling pathway	3	5.56	-2.11	0
R-HSA-5668914	Reactome Gene Sets	Diseases of metabolism	3	5.56	-2.02	0

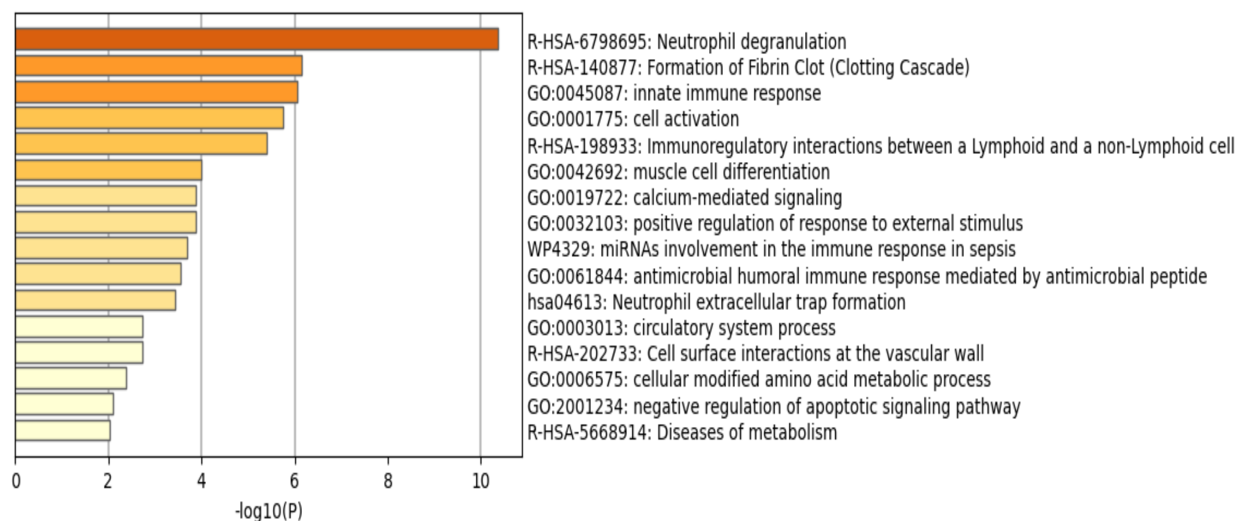

Figure 2: Bar-graph for top 16 clusters with their representative enriched terms (one per cluster) for staphylococcus infection related genes (colored by p-values). "Log10(P)" is the p-value in log base 10. "Log10(q)" is the multi-test adjusted p-value in log base 10.

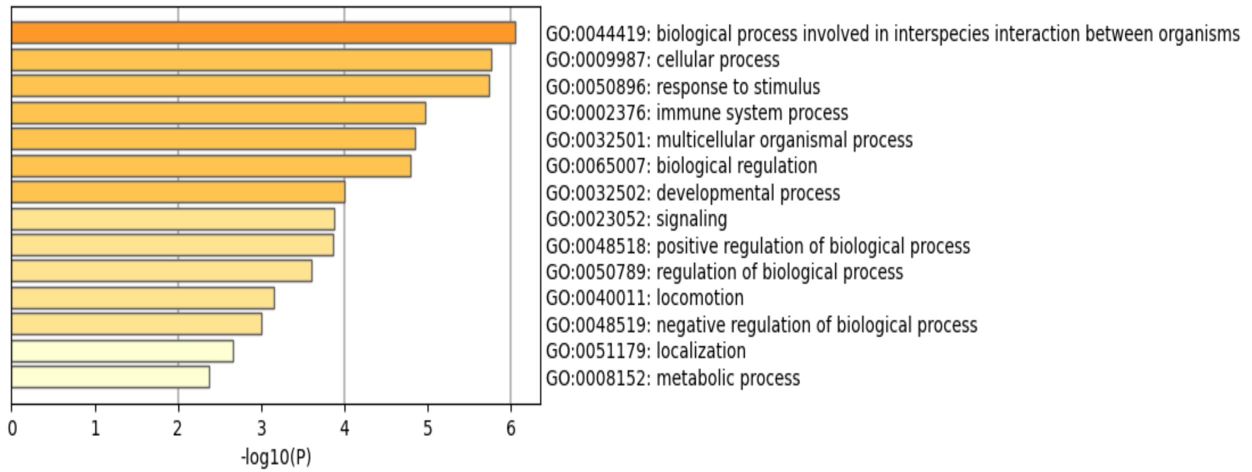

Figure 3: Bar-graph for top-level Gene Ontology biological processes enriched by *staphylococcus aureus* infection associated genes (colored by p-values). "Log10(P)" is the p-value in log base 10. "Log10(q)" is the multi-test adjusted p-value in log base 10.

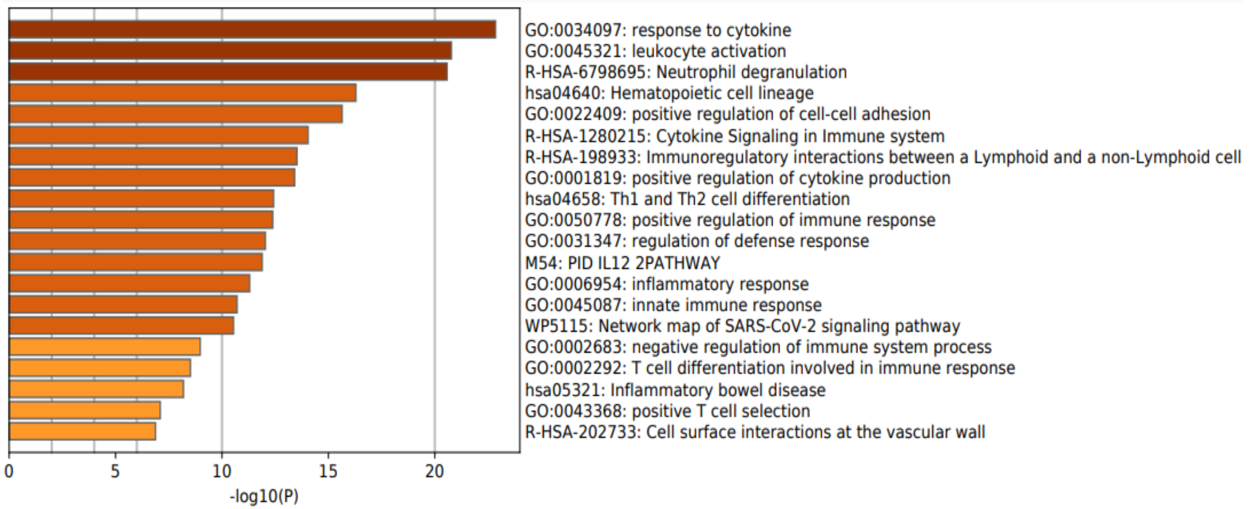

Figure 4: Bar-graph for top 20 clusters with their representative enriched terms (one per cluster) for pneumonia infection related genes (colored by p-values). "Log10(P)" is the p-value in log base 10. "Log10(q)" is the multi-test adjusted p-value in log base 10.

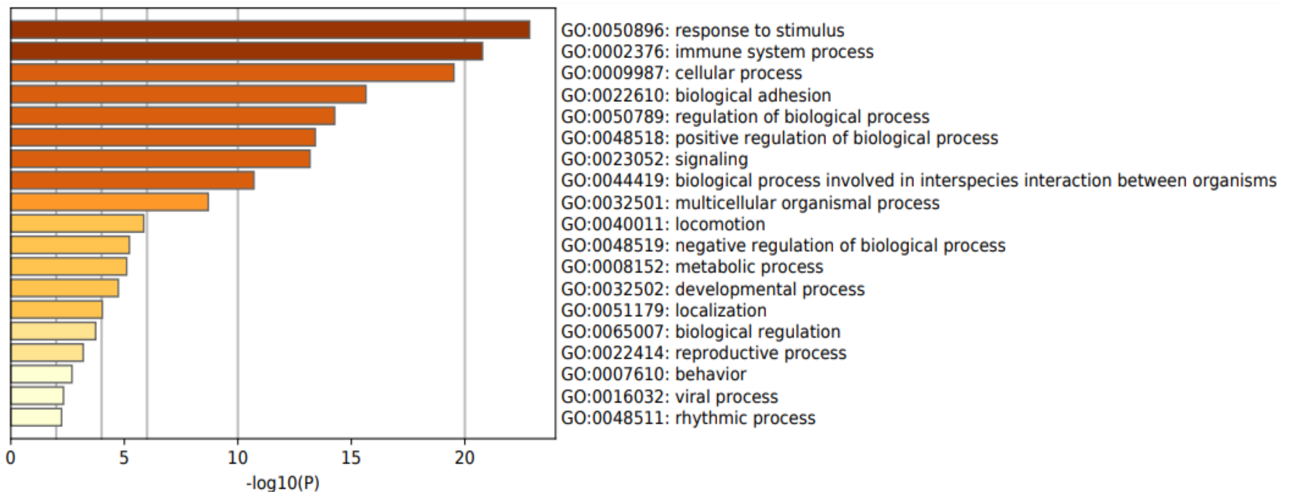

Figure 5: Bar-graph for top-level Gene Ontology biological processes enriched in pneumonia infection associated genes (colored by p-values). "Log10(P)" is the p-value in log base 10. "Log10(q)" is the multi-test adjusted p-value in log base 10.

Table 3. Top 20 clusters with their representative enriched terms (one per cluster). "Count" is the number of genes in the user-provided lists with membership in the given ontology term. "%" is the percentage of all of the user-provided genes that are found in the given ontology term (only input genes with at least one ontology term annotation are included in the calculation). "Log10(P)" is the p-value in log base 10. "Log10(q)" is the multi-test adjusted p-value in log base 10.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0034097	GO Biological Processes	response to cytokine	68	11.11	-22.85	-18.51
GO:0045321	GO Biological Processes	leukocyte activation	52	8.5	-20.78	-16.91
R-HSA-6798695	Reactome Gene Sets	Neutrophil degranulation	50	8.17	-20.58	-16.83
hsa04640	KEGG Pathway	Hematopoietic cell lineage	22	3.59	-16.29	-13.09
GO:0022409	GO Biological Processes	positive regulation of cell-cell adhesion	34	5.56	-15.64	-12.47
R-HSA-1280215	Reactome Gene Sets	Cytokine Signaling in Immune system	51	8.33	-14.04	-10.98
R-HSA-198933	Reactome Gene Sets	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	22	3.59	-13.52	-10.5
GO:0001819	GO Biological Processes	positive regulation of cytokine production	40	6.54	-13.41	-10.41
hsa04658	KEGG Pathway	Th1 and Th2 cell differentiation	18	2.94	-12.42	-9.54
GO:0050778	GO Biological Processes	positive regulation of immune response	42	6.86	-12.38	-9.51
GO:0031347	GO Biological Processes	regulation of defense response	45	7.35	-12.04	-9.21
M54	Canonical Pathways	PID IL12 2PATHWAY	15	2.45	-11.89	-9.07
GO:0006954	GO Biological Processes	inflammatory response	38	6.21	-11.3	-8.53
GO:0045087	GO Biological Processes	innate immune response	47	7.68	-10.71	-7.97
WP5115	WikiPathways	Network map of SARS-CoV-2 signaling pathway	24	3.92	-10.54	-7.83
GO:0002683	GO Biological Processes	negative regulation of immune system process	31	5.07	-8.97	-6.37
GO:0002292	GO Biological Processes	T cell differentiation involved in immune response	10	1.63	-8.51	-5.93
hsa05321	KEGG Pathway	Inflammatory bowel disease	12	1.96	-8.19	-5.65
GO:0043368	GO Biological Processes	positive T cell selection	8	1.31	-7.1	-4.65
R-HSA-202733	Reactome Gene Sets	Cell surface interactions at the vascular wall	15	2.45	-6.87	-4.44

Protein-protein interaction analysis

From the 54 identified genes related to staphylococcus infection, C-X-C motif chemokine ligand 8 (CXCL8) has the most significant interaction with the other proteins (Table S1). Its expression was however downregulated. Cathelicidin antimicrobial peptide (CAMP) and lipocalin 2 (LCN2) shows the highest expression values. It's however intriguing that both shows strong interaction as shown in Figure 4. About 613 genes are related to pneumonia infection in this study (Table S3), with carcinoembryonic antigen related cell adhesion molecule 8 (CEACAM8) and Fc fragment of IgE receptor Ia (FCER1A) emerging as the most upregulated and downregulated proteins respectively (Table 1c). However, FCER1A is not shown on the network in Figure 5 because there's no known functional and physical protein associations. In the network graph, the nodes represent proteins and the edges indicate both functional and physical protein associations existing among the nodes. The sources from which the interactions were obtained includes Text-mining from literature, empirical studies, Databases, co-expression, Neighborhood, gene fusion and co-ocurrence studies. Minimum required interaction score was set at 0.4 being the default value on string database. Nodes without any connection were excluded from the network.

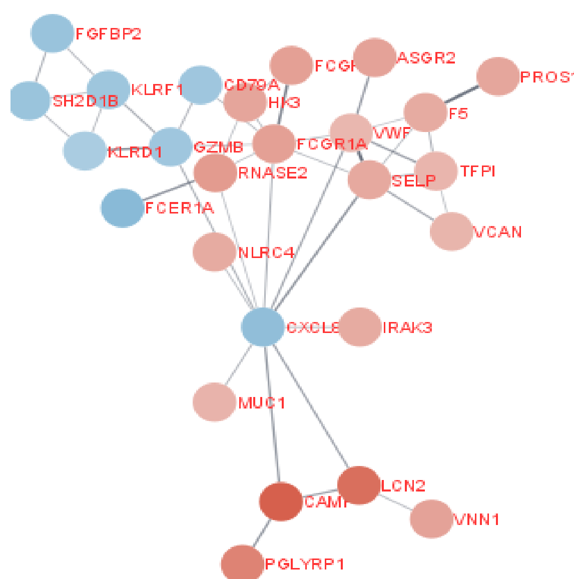

Figure 4: Protein-protein interaction network of staphylococcus infection related genes. Blue nodes are proteins whose expression were downregulated while the red nodes are those which were upregulated.

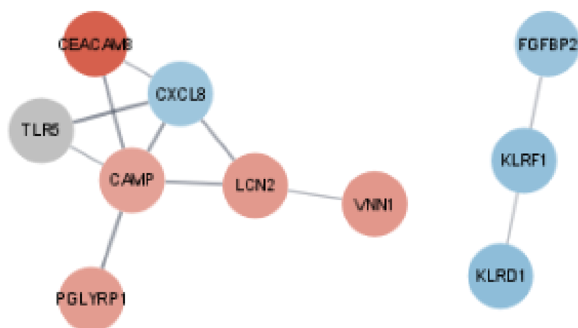

Figure 5: Protein-protein interaction network of pneumonia infection related genes. Blue nodes are proteins whose expression were downregulated while the red nodes are those which were upregulated. CEACAM8 has the highest fold change in expression value. Two clusters appear from the network. One basically comprises only downregulated proteins (right) while the other contains upregulated and downregulated proteins and a protein whose expression did not change significantly.

DISCUSSION

This study was carried out based on the hypothesis that "transcriptional profile heterogeneity will reflect patient clinical heterogeneity" and also identify gene signatures that may serve as biomarkers of *staphylococcus infection* in human". It is our goal to identify the genes which exhibit differential expression in pneumonia infection induced by *staphylococcus aureus.* One of the most common uses of sequencing data is differential gene expression (DGE) analysis. This method is commonly utilized in many sequencing data analysis applications since it enables for the identification of differentially expressed genes across two or more conditions. Due to the variety of formats based on the tool of choice and the multiple bits of information contained in these results files, interpreting DGE findings can be difficult and time consuming (Wang *et al.,* 2019). In the ICU, *Staphylococcus aureus* is the second most prevalent cause of pneumonia. Toxins and enzymes produced by the bacteria highlight its virulence, causing significant lung tissue damage. Clinical signs are insufficient to identify *Staphylococcus aureus* pneumonias from those caused by other pathogens, and clinical diagnosis suffers from the same limitations as other bacterial pneumonia causes (Hooper and Smith, 2012).

The comparison groups set for differential analysis in this study include *staphylococcus aureus* infected patients, staphylococcus infected patients with pneumonia infection and staphylococcus infected patients without pneumonia infection. The infection present aside pneumonia included bacteremia, osteomyelitis, suppurative arthritis, pyomyositis, empyema, abscess. Downregulation of gene expression is an indication of the inhibitory activity of the pathogen while the genes whose expression were upregulated are involved in the host defense against the pathogen (Shirahama *et al.,* 2020). The downregulated genes during staphylococcus infection include KLRF1, P2RY10, CMKLR1, CD79A, GZMB, KLRF1, SH2D1B, FGFBP2, CXCL8, FCER1A. KLRK1 is a type II transmembrane-anchored glycoprotein that is expressed on the surface of Natural Killer (NK) cells, gamma/delta TcR+ T cells, CD8+ T cells, and a modest subset of CD4+ T cells as a disulfide-linked homodimer. It binds to the DAP10 signaling protein non-covalently and sends activating or costimulatory signals to NK cells and T cells. NKG2D interacts to a family of glycoproteins called MICA, MICB, and ULBP1-6 membrane proteins in humans, which are commonly produced on cells that have been infected with pathogens or transformed. In comparison to adults, infants are more susceptible to many infections, which can be related to their undeveloped innate and adaptive immunity. Without pre-sensitization, natural killer cells provide first-line innate immune reactions against infected cells (Land, 2018). Lanier (2015), proved that the expression of KLRG1 on T cells improves during *M. tuberculosis* infection and declines after treatment, suggesting a correlation between KLRG1 expression and disease progression. The decrease in KLRG1 observed in this study may not be unconnected to the undeveloped immunity in pediatrics. The genes which contributed tremendously to the host defense against *staphylococcus aureus* infection include CAMP, LCN2, PGLYRP1, MYL9, ANKRD22, CETP, RNASE2, ASGR2, PLOD2. Cathelicidin antimicrobial peptide has the highest change in expression value. CAMP expression is also a member of the of the pneumonia infection associated/related genes (PIAGs) (figure 5).

Cathelicidin is an antibacterial peptide of the cathelicidin family. It is a small molecule (composed of 12-100 amino acids) with wide antibacterial activity that is thought to play a role in the innate immunity as the first line of defense against microbes. (Iacob and Iacob, 2014). When cathelicidin is produced enzymatically, it has an N-terminal prosequence followed by a C-terminal variable sequence with strong microbial activity. This antimicrobial peptide group is called cathelicidin because the structure of the prosequence is extremely similar to that of a protein called cathelin. Although the exact method of CAMP (Cathelicidin Antimicrobial Peptide) gene regulation is unknown, cathelicidin is reported to be upregulated when bacteria are present (Wang *et al.,* 2021). Bacterial compounds have been found to boost Cathelicidin production in cultured human cells, showing that Cathelicidin plays a role in infection resistance. Several compounds, including 1,25-dihydroxyvitamin D3 (1,25$(OH)_2$ D3), an active form of vitamin D, have been described as potent inducers of CAMP gene expression. Butyrate, Trichostatin A, Lithocholic acid, Interleukin-6, 1,25$(OH)_2$ D3 (Pineda *et al.,* 2019, Febriza *et al.,* 2019). The physical structure of cathelicidin, as well as its cationic and hydrophobic characteristics, are responsible for the majority of its antimicrobial

activities. The N-terminal helix is involved in chemotaxis and proteolisis defense, while the C-terminal helix is involved in antimicrobial activity. Cathelicidin binds to the surface of the microbial membrane, covers it, and perforates it, generating pores on the membrane that finally kill the bacteria. (Lv *et al.,* 2014). Unlike zwitterionic eukariotic membranes, cathelicidin attaches to cell membranes that contain lipopolysaccharide (Gram-negative) or teichoic acid (Gram-positive) with a negative charge. With the contact between the capsule membrane and the protein capsid, cathelicidin also exhibits antiviral properties (Steinbuch and Fridman, 2016). Cathelicidin binds to the bacterial membrane in oligomeric forms, altering the subsequent contact and permeabilization manner. Because the monomeric peptide is less susceptible to sequestration by serum or medium components, as well as components of the bacterial outer cell wall, this has a significant impact on antibacterial action. In staphylococcus infection and tuberculosis, cathelicidin can bind to lipoteikoic acid and lipoarabinomannan, preventing macrophage activation. In some circumstances, resistant bacteria's protolithic enzymes can destroy cathelicidin and other antimicrobial peptides (Rowe-Magnus *et al.,* 2019).

Among the pneumonia infection associated/related genes (PIAGs), Carcinoembryonic antigen related cell adhesion molecule 8 (CEACAM8) was highly upregulated and has interaction with Cathelicidin Antimicrobial Peptide (CAMP) which is also the highest upregulated among staphylococcus infection associated genes (SIAGs). The significant increase in the expression of CEACAM8 suggests the it may have a role in the interaction of *staphylococcus aureus* with neutrophils (Sarantis and Gray-Owen, (2012). CEACAM molecules are membrane glycoproteins that mediate intercellular interactions that influence cellular proliferation, immune cell activation, apoptosis, and tumor suppression. To establish a close interaction with host cells and tissues, a vast number of bacterial pathogens target cell adhesion molecules. Specific bacterial surface proteins typically identify members of the integrin, cadherin, and immunoglobulin-related cell adhesion molecule (IgCAM) families. Following cytoskeletal rearrangements triggered by receptor clustering, binding might cause bacterial internalization. Furthermore, signals from occupied receptors can cause cellular responses such as gene expression events, which affect the infected cell's phenotypic (Mix *et al.,* 2021).

CONCLUSION

The molecular mechanism of infection and the involvement of the host defense against pneumonia induced by *staphylococcus aureus* was critically examined. However, due to the fact that the study was carried out on pediatric patients, the results found may not be generalised on other age groups. There is a need for a comparative study to compare and contrast the mechanisms involved in other members of the population.

Compliance with Ethical Standards

Not applicable.

Funding

Not applicable.

Conflict of Interest

The authors declare that they have no competing interests.

Ethical approval

Not applicable.

Informed consent

Not applicable.

REFERENCES

1. Andrejko, K., Whittles, L. K., & Lewnard, J. A. (2021). Health-economic value of vaccination against group A Streptococcus in the United States. *Clinical Infectious Diseases*.
2. Bhuiyan, M. U., Snelling, T. L., West, R., Lang, J., Rahman, T., Granland, C., de Gier, C., Borland, M. L., Thornton, R. B., Kirkham, L. A. S., Sikazwe, C., Martin, A. C., Richmond, P. C., Smith, D. W., Jaffe, A., & Blyth, C. C. (2018). The contribution of viruses and bacteria to community-acquired pneumonia in vaccinated children: a case–control study. *Thorax*, *74*(3): 261–269. https://doi.org/10.1136/thoraxjnl-2018-212096
3. Febriza, A., Hatta, M., Natzir, R., Kasim, V. N., & Idrus, H. H. (2019). Activity of antimicrobial peptide; cathelicidin, on bacterial infection. *The Open Biochemistry Journal*, *13*(1).
4. Hooper, C., & Smith, W. (2012). Telavancin for the treatment of nosocomial pneumonia caused by methicillin-resistant Staphylococcus aureus (MRSA). *Therapeutics and Clinical Risk Management*, 131. https://doi.org/10.2147/tcrm.s23247
5. Iacob, S. A., & Iacob, D. G. (2014). Antibacterial Function of the Human Cathelicidin-18 Peptide (LL-37) between Theory and Practice. *Protein & Peptide Letters*, *21*(12): 1247–1256. https://doi.org/10.2174/0929866521666140805124855
6. Jacobs, D. M., & Shaver, A. (2017). Prevalence of and outcomes from Staphylococcus aureus pneumonia among hospitalized patients in the United States, 2009-2012. *American journal of infection control*, *45*(4): 404-409.
7. Land, W. G. (2018). Innate immune recognition molecules. In *Damage-Associated Molecular Patterns in Human Diseases* (pp. 43-108). Springer, Cham.
8. Lanier, L. (2015). KLRK1 (killer cell lectin-like receptor subfamily K, member 1). *Atlas of Genetics and Cytogenetics in Oncology and Haematology*, *3*. https://doi.org/10.4267/2042/56407

9. Leung, T. Y. M., Chan, A. Y. L., Chan, E. W., Chan, V. K. Y., Chui, C. S. L., Cowling, B. J., ... & Wong, I. C. K. (2020). Short-and potential long-term adverse health outcomes of COVID-19: a rapid review. *Emerging microbes & infections*, *9*(1): 2190-2199.
10. Lv, Y., Wang, J., Gao, H., Wang, Z., Dong, N., Ma, Q., and Shan, A. (2014). Antimicrobial Properties and Membrane-Active Mechanism of a Potential α-Helical Antimicrobial Derived from Cathelicidin PMAP-36. *PLoS ONE*, *9*(1): e86364. https://doi.org/10.1371/journal.pone.0086364
11. Maggini, S., Pierre, A., & Calder, P. C. (2018). Immune function and micronutrient requirements change over the life course. *Nutrients*, *10*(10): 1531.
12. Maurer, J. (2007). Trends in Hospitalizations for Pneumonia Among Persons Aged 65 Years or Older in the United States, 1988–2002. *Yearbook of Medicine*, *2007*; 252–253. https://doi.org/10.1016/s0084-3873(08)70171-8
13. Mix, A. K., Goob, G., Sontowski, E., and Hauck, C. R. (2021). Microscale communication between bacterial pathogens and the host epithelium. *Genes & Immunity*, *22*(5): 247-254.
14. Mizgerd, J. P. (2012). Respiratory Infection and the Impact of Pulmonary Immunity on Lung Health and Disease. *American Journal of Respiratory and Critical Care Medicine*, *186*(9): 824–829. https://doi.org/10.1164/rccm.201206-1063pp
15. Pineda Molina, C., Hussey, G. S., Eriksson, J., Shulock, M. A., Cárdenas Bonilla, L. L., Giglio, R. M., ... and Badylak, S. F. (2019). 4-hydroxybutyrate promotes endogenous antimicrobial peptide expression in macrophages. *Tissue Engineering Part A*, *25*(9-10): 693-706.
16. Quinton, L. J., Walkey, A. J., & Mizgerd, J. P. (2018). Integrative physiology of pneumonia. *Physiological reviews*, *98*(3): 1417-1464.
17. Rowe-Magnus, D. A., Kao, A. Y., Prieto, A. C., Pu, M., and Kao, C. (2019). Cathelicidin Peptides Restrict Bacterial Growth via Membrane Perturbation and Induction of Reactive Oxygen Species. *MBio*, *10*(5). https://doi.org/10.1128/mbio.02021-19
18. Rudan, I. (2008). Epidemiology and etiology of childhood pneumonia. *Bulletin of the World Health Organization*, *86*(5), 408–416. https://doi.org/10.2471/blt.07.048769
19. Sarantis, H., and Gray-Owen, S. D. (2012). Defining the roles of human carcinoembryonic antigen-related cellular adhesion molecules during neutrophil responses to Neisseria gonorrhoeae. Infection and immunity, 80(1): 345–358. https://doi.org/10.1128/IAI.05702-11
20. Shirahama, S., Miki, A., Kaburaki, T., and Akimitsu, N. (2020). Long non-coding RNAs involved in pathogenic infection. *Frontiers in Genetics*, *11:* 454.
21. Steinbuch, K. B., and Fridman, M. (2016). Mechanisms of resistance to membrane-disrupting antibiotics in Gram-positive and Gram-negative bacteria. *MedChemComm*, *7*(1): 86–102. https://doi.org/10.1039/c5md00389j
22. Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., ... and Von Mering, C. (2016). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*, gkw937.
23. Troeger, C., Blacker, B., Khalil, I. A., Rao, P. C., Cao, J., Zimsen, S. R., ... & Reiner Jr, R. C. (2018). Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet infectious diseases*, *18*(11): 1191-1210.
24. Wang, G., Li, Y., Yang, G., Yang, T., He, L., and Wang, Y. (2021). Cathelicidin antimicrobial peptide (CAMP) gene promoter methylation induces chondrocyte apoptosis. *Human Genomics*, *15*(1). https://doi.org/10.1186/s40246-021-00321-8
25. Wang, Z., Jin, S., & Zhang, C. (2019). A Method Based on Differential Entropy-Like Function for Detecting Differentially Expressed Genes Across Multiple Conditions in RNA-Seq Studies. *Entropy*, *21*(3): 242. https://doi.org/10.3390/e21030242
26. Yan, H., Li, Z., Shen, Q., Wang, Q., Tian, J., Jiang, Q., & Gao, L. (2017). Aberrant expression of cell cycle and material metabolism related genes contributes to hepatocellular carcinoma occurrence. *Pathology-Research and Practice*, *213*(4): 316-321.