e-ISSN 2795-3276
p-ISSN 2795-3270

GPH INTERNATIONAL JOURNAL
Research in
Biological Science
www.gphjournal.org || editor@gphjournal.org

10.5281/zenodo.10522531 Vol. 06 Issue 12 Dec - 2023 Manuscript ID: #1191

Gene Expression Profile in *Salmonella Enteritidis* in Egg Albumen Based on Pathogenicity via GEO-Analysis

By

Citra Sari[1], Fajar Shodiq Permata[2], Aryo Tedjo[3]

[1)]Laboratory of Veterinary Public Health, Faculty of Veterinary Medicine Universitas Brawijaya, Puncak Dieng Eksklusif Kalisongo Dau Malang, East Java, Indonesia, 65151,
ORCID ID: https://orcid.org/0009-0005-1850-4818

[2)]Laboratory of Veterinary Anatomy, Histology and Embriology, Faculty of Veterinary Medicine Universitas Brawijaya, Puncak Dieng Eksklusif Kalisongo Dau Malang, East Java, Indonesia, 65151,
ORCID ID: https://orcid.org/0000-0003-0971-6278

3) Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia, Jakarta, DKI Jakarta, Indonesia, 10430, ORCID ID: https://orcid.org/0000-0001-8998-3418

Corresponding Author: citrasari83@ub.ac.id

Abstract:

Salmonella enteriditis, a bacterium known for contaminating egg albumen, serves as a significant causative agent of foodborne illnesses in humans. These illnesses manifest with a spectrum of symptoms ranging from mild to severe, contingent upon the varying pathogenicity levels of *Salmonella enteritidis*. The central objective of this research endeavor was to meticulously analyze the gene expression profile of *Salmonella enteritidis* in egg albumin, correlating it with the pathogen's varying pathogenicity levels. This analysis was conducted utilizing the Gene Expression Omnibus (GEO) Analysis framework. A comprehensive examination of 18 genomic databases specific to *Salmonella enteritidis*, extracted from the GEO Dataset (GSE33102), was undertaken. These databases were methodically clustered in accordance with the pathogenicity gradations of the bacteria. Subsequently, an in-depth analysis and visualization of the data were performed using GEO2R. The analytical findings revealed a notable variance in gene expression, with 35-46 genes demonstrating significant differences (Padj<0.05) when comparing groups with High Pathogenicity and High-Medium Pathogenicity against those with Low Pathogenicity The study culminated in the identification of six distinct gene expressions that effectively discriminate between *Salmonella enteritidis* groups classified as High, High-Medium, and Low Pathogenicity. This discovery propels the hypothesis that these genes could potentially serve as specific markers for the presence of *Salmonella enteritidis* in contaminated eggs. Such markers would be instrumental in the early detection of foodborne diseases. However, it is imperative to conduct further research to ascertain the viability of these candidate genes as reliable indicators for the early detection of this pathogen in contaminated food sources.

Keywords:

contamination, foodborne, gene expression, GEO Analysis, pathogenicity

Introduction

Salmonella enteritidis contamination in eggs is a significant public health concern due to the potential risk of foodborne illness. Research has shown that vertical transmission is a common route of contamination, where the bacteria infect the layer hen oviduct, subsequently contaminating the eggs' internal contents during development(Jia et al., 2020). Additionally, delays in the growth of *S. enteritidis* in albumen have been observed, indicating the potential for contamination at different stages of egg development(Humphrey et al., 1991).

Efforts to address Salmonella contamination in the egg industry have been prompted by outbreaks of Salmonella-related illnesses, leading to joint initiatives by public health authorities and the egg industry to mitigate this concern(Chousalkar et al., 2017). Furthermore, longitudinal studies have been conducted to identify points in production where birds are most likely to be exposed to Salmonella, thus determining the highest risk of egg contamination(McWhorter & Chousalkar, 2019). The presence of Salmonella contamination in egg-packing plants has been identified as a significant hazard (Davies & Breslin, 2003). These findings highlight the complex nature of *Salmonella enteritidis* contamination in eggs and the multifaceted approaches required to address this public health challenge.

The pathogenicity of bacteria is a complex process involving various molecular mechanisms and interactions with the host. Studies have highlighted the importance of proteogenomics in understanding host-pathogen interactions from a bacterial perspective (Fels et al., 2017). High-throughput metagenomic approaches have been utilized to reveal the profile and fate of bacterial pathogens, providing a broad picture of the occurrence of bacterial pathogens and their advantages over traditional detection methods (Li et al., 2015)

The Gene Expression Omnibus (GEO) database is a valuable resource for researchers conducting gene expression analysis. It contains a vast amount of publicly available gene expression data from various organisms and biological phenomena(Clough & Barrett, 2016). Researchers have utilized the GEO database to identify differentially expressed genes (DEGs) in diseases(Liu et al., 2020; Song et al., 2020; Zhang et al., 2020). Furthermore, the GEO database has been employed in identifying potential biomarkers and constructing regulatory networks for various diseases(Qi & Chen, 2021; Shi et al., 2021; Wei et al., 2020).The GEO database has evolved over the years to accept high-throughput data for various applications, including genome methylation, chromatin structure, and genome–protein interactions(Clough & Barrett, 2016). It has also been used to rapidly discover potential drugs and identify molecular markers associated with disease progression and prognosis (Luo et al., 2019; Wei et al., 2020)

The current research landscape lacks reported data on the gene expression profiles delineating the pathogenicity of *Salmonella enteritidis* in egg albumen contamination; this study endeavored to bridge that knowledge gap.

Materials and Methods

1. Data mining

In this study, data acquisition was conducted via the National Center for Biotechnology Information (NCBI), utilizing the Gene Expression Omnibus (GEO) dataset. The specific dataset employed, bearing the accession number GSE33102, encompasses an array of *Salmonella enteritidis* strains derived from poultry. These strains exhibit a range of responses to stress factors

and varying survival rates within egg albumen. After the data retrieval, a comprehensive analysis was undertaken utilizing the GEO2R analytical tool, as depicted in **Figure 1.**Data mining is a necessary step for bioinformatics study (Sari et al., 2023)

The scope of the data analyzed encompassed a total of 18 samples, a detailed enumeration of which is provided in **Table 1**. Detailed information pertaining to each sample has been meticulously compiled and presented in Table 2 to facilitate a deeper understanding and enable a thorough interpretation. This arrangement ensures a structured and clear presentation of the data, effectively conveying findings and observations from the study.

2. Data analysis

The data procured in this study were meticulously categorized into three distinct groups based on their pathogenicity levels: High Pathogenicity (comprising six samples), High-Medium Pathogenicity (encompassing three samples), and Low Pathogenicity (including nine samples), as illustrated in Figure 1. This classification facilitated a focused and nuanced analysis of the gene expressions associated with the varying levels of pathogenicity exhibited by *Salmonella enteritidis* in egg albumen.

A detailed comparison of these gene expressions was conducted utilizing the GEO2R analysis tool, adhering to a threshold of P.adj<0.05 for statistical significance. The outputs from this analysis were visually represented through sophisticated graph clustering techniques. These included the generation of Volcano Plots, which provide a comprehensive view of statistically significant gene expression changes, and Mean Difference (MD) Plots. The MD Plots offer a clear visualization of the average differences in expression between the designated groups, thus enabling a more profound understanding of the pathogenetic variations among the Salmonella enteriditis strains. Such visual representations are crucial in elucidating the intricate relationships and differential gene expressions linked to the pathogenicity levels within the study's scope.

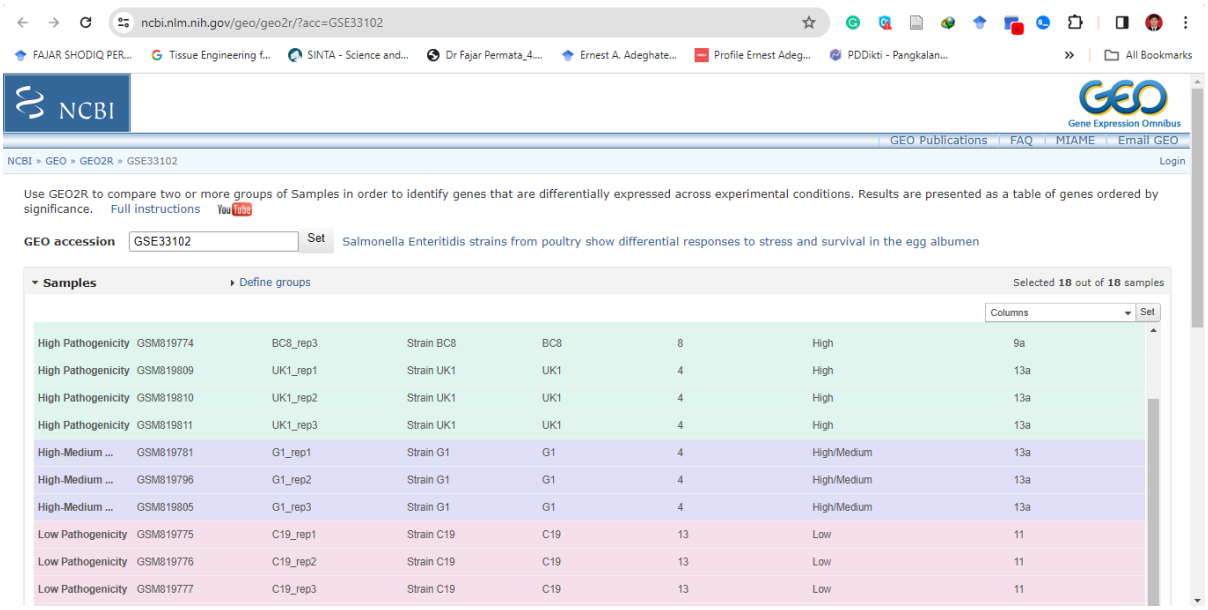


Figure 1. Clustering sample data on Genomic Expression of *Salmonella enteriditis* in Egg Albumen based on Pathogenicity in GEO application.

Table 1. Accession number of genomic database samples based on pathogenicity

Group	Sample GEO accession number
High pathogenicity (n=6)	GSM819772, GSM819773, GSM819774, GSM819809, GSM819810, GSM819811
High-Medium Pathogenicity (n=3)	GSM819781, GSM819796, GSM819805
Low pathogenicity (n=9)	GSM819775, GSM819776, GSM819777, GSM819778,GSM819779, GSM819780, GSM819806, GSM819807,GSM819808

Table 2. Information genomic data for each sample (*Salmonella enteritidis*, GSE33102)

No	GEO Accession number	Strain	Pathogenicity
1	GSM819796	G1	High-Medium
2	GSM819805	G1	High-Medium
3	GSM819773	BC8	High
4	GSM819774	BC8	High
5	GSM819776	C19	Low
6	GSM819777	C19	Low
7	GSM819778	C45	Low
8	GSM819780	C45	Low
9	GSM819806	C45	Low
10	GSM819807	C45	Low
11	GSM819809	UK1	High
12	GSM819810	UK1	High
13	GSM819811	UK1	High
14	GSM819772	BC8	High
15	GSM819775	C19	Low
16	GSM819779	C45	Low
17	GSM819781	G1	High-Medium
18	GSM819808	G45	Low

Results and Discussion

1. Results

The outcomes of the comparative analysis, focusing on the gene expression of Salmonella enteriditis in relation to varying levels of pathogenesis, revealed a discernible fluctuation in the expression of several genes. This fluctuation was effectively illustrated in the Volcano Plot **(Figure 2)** and the Mean Difference (MD) Plot **(Figure 3).** These graphical representations provided a detailed comparison of gene expression variations between the groups under study.

Additionally, a comprehensive recapitulation encompassing the enumeration of genes, categorized based on the rise and fall in their expression levels across different groups, has been meticulously compiled. This tabulated data was succinctly presented in **Table 3**, offering a clear and structured overview of the comparative gene expression levels in the context of the study's pathogenicity-based groupings.

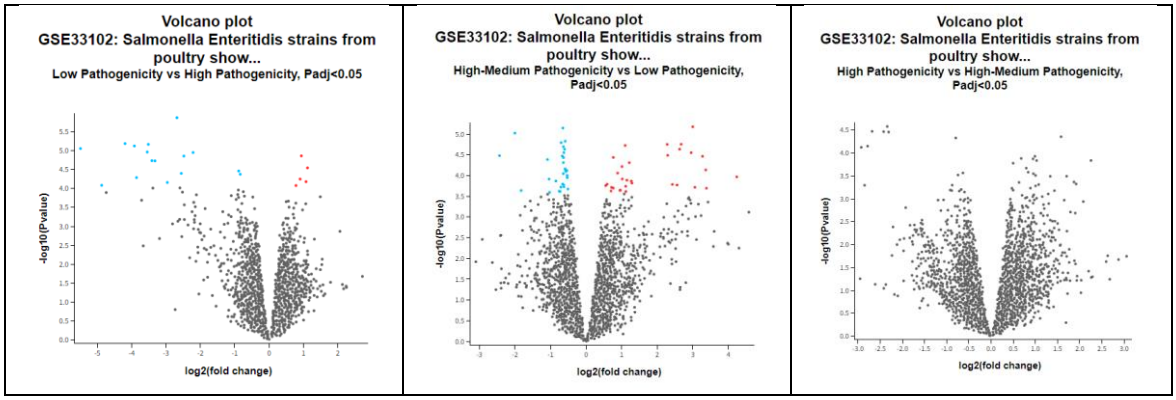


Fig 2. Volcano plots on comparisons between groups, namely Low Pathogenicity vs High Pathogenicity (left side), High-Medium Pathogenicity vs Low Pathogenicity (middle side), and High Pathogenicity vs High-Medium Pathogenicity (right side). Blue dot showed genes whose expression decreases significantly, while Red dot showed genes whose expression increases significantly. High Pathogenicity vs. High-Medium Pathogenicity did not show significantly different gene expression. (Picture from GEO2R Analysis)

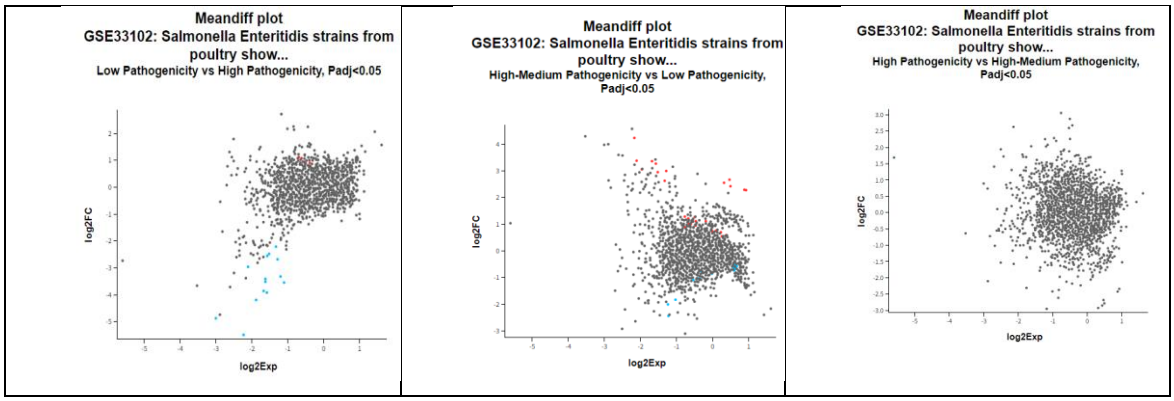


Fig 3. Mean Difference (MD) of Plot on comparison between groups, namely Low Pathogenicity vs High Pathogenicity (left side), High-Medium Pathogenicity vs Low Pathogenicity (middle side), and High Pathogenicity vs High-Medium Pathogenicity (right side). Blue dot showed genes whose expression decreases significantly, while Red dot shows genes whose expression increases significantly. High Pathogenicity vs. High-Medium Pathogenicity did not show significantly different gene expression. (Picture from GEO2R Analysis)

Table 3. Recapitulation Comparison of the number of Genes whose expression levels rise and fall between groups Pathogenicity of *Salmonella enteriditis* in egg albumen

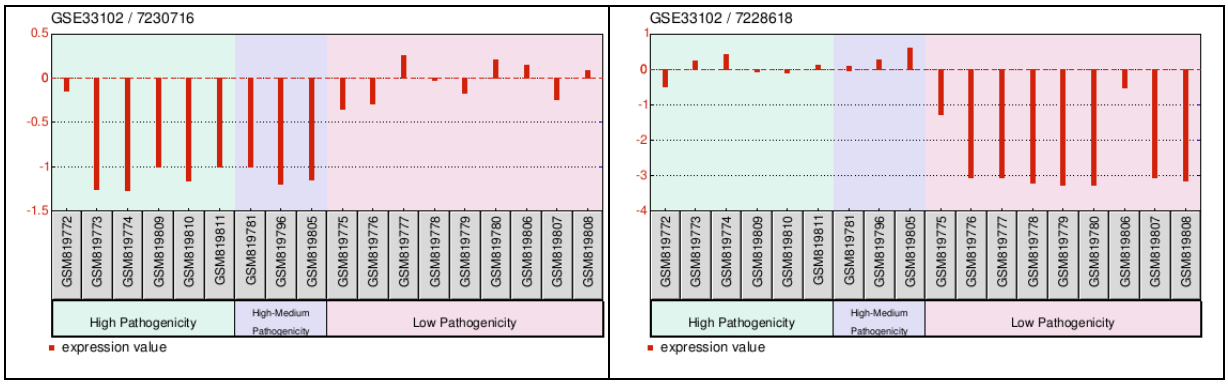
Comparison	Up genes	Down genes
High Pathogenicity vs Low Pathogenicity	5	16
High-Medium Pathogenicity vs Low Pathogenicity	31	30
High Pathogenicity vs High-Medium Pathogenicity	No significant genes	

The analysis revealed that six genes exhibited significantly divergent expressions, potentially serving as distinctive markers delineating the pathogenicity levels - High, High-Medium, and Low - of Salmonella enteriditis in egg albumen. These genes, which may function as genetic or protein markers, are detailed in **Table 4** and **Figure 4**. However, it is noteworthy that these identified

genes and proteins were not exclusive to Salmonella enteriditis, thus presenting a future challenge. The crux of this challenge lies in ensuring that the variations in gene expression profiles, which are correlated with the pathogenicity of Salmonella enteriditis in egg albumen, can be precisely utilized for the detection of harmful bacterial contamination in feed. This endeavor necessitates a heightened level of specificity to accurately identify and differentiate between benign and pathogenic bacterial presences.

Table 4. Candidate genetic marker to detect High Pathogenicity of *Salmonella enteriditis* contaminating egg albumen

No	ID	Gene name	Protein	Sequence
1	7230716	NA	phage protein	CGGGTGGAGAGCAGAAATACCGTCCGAAGTACTGGCTTGATAATAA
2	7228618	NA	heat shock protein HtpX	TTGAACAGCCTCGCAATGAAAGAGAACGCTGGTTGATGAACACCG
3	7232516	smpA	hypothetical protein	ACATGAGAACGTGACGCAGCAGACTCTGACGCTCACCTTTAACAG
4	7233752	Cca	multifunctional tRNA nucleotidyl	CCAGCAGGTAGGCCGTGATTTTCCTGTGTTTCTCCACCCGCAAAC
5	7236581	dcuB	anaerobic C4-dicarboxylate transporter	CACGTGGTCTACACCATTTTGCCGATTATCTATGACGTGGCGATC
6	7237221	yjjW	hypothetical protein	GCGCTTTAGTCAGTAAGATCATTGCGTTTTCCTGCGTCGATGGGC



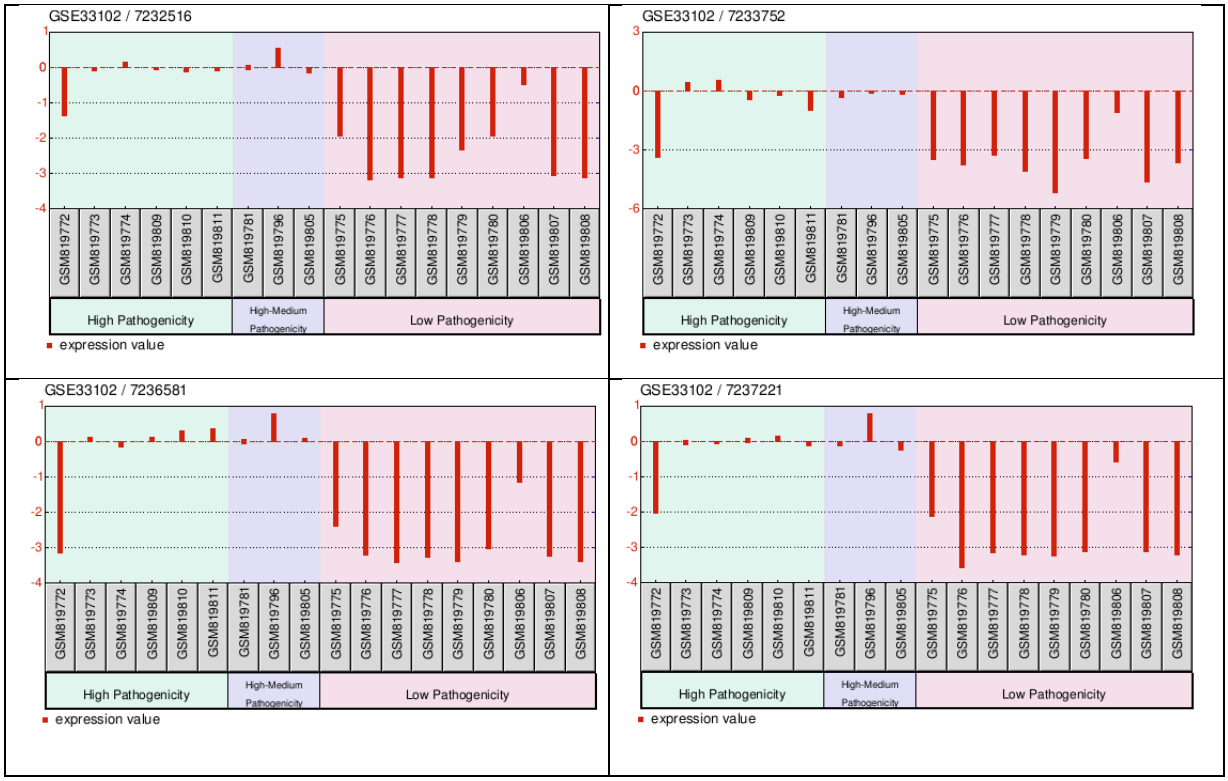


Fig 4. A total of 6 gene expression profiles of *Salmonella enteritidis* contaminating egg albumen that have similarities between High Pathogenicity and High-Medium Pathogenicity groups but are significantly different from Low Pathogenicity groups (Picture from GEO2R Analysis)

2. Discussion

Salmonella contamination in egg albumen is a significant concern due to its potential impact on public health. Research has shown that Salmonella enterica serovar Enteritidis can survive and contaminate egg albumen, leading to potential foodborne illness (Clavijo et al., 2006). It has been reported that Salmonella bacteria are often deposited within the albumen of eggs, particularly via the oviduct tissues, highlighting the risk of contamination in this part of the egg (Gantois et al., 2008). Furthermore, *Salmonella enteritidis* is frequently found in the albumen of contaminated eggs, emphasizing the importance of addressing this specific area of the egg to prevent contamination (Lu et al., 2003).

The survival characteristics of *Salmonella enteritidis* in chicken egg albumen have also been investigated, with findings indicating that the majority of Salmonella bacteria are deposited from the reproductive tract to the albumen, close to the yolk membrane or on the vitelline membrane, further underscoring the risk of contamination in this region (Kang et al., 2006)

The study found from GEO Data set accesion number GSE33102 there are three types of pathogenicity of *Salmonella enteritidis*contaminated egg albumen **(Table 2)** such as Low Pathogenicity (strain C19, G45 and C45), High-Medium Pathogenicity (strain G1) and High Pathogenicity (strain UK1 and BC8).The survival rate of UK, G1, and BC8 strainswas significantly higher as compared with the C19, C45, and G45 strains(Shah et al., 2012)

Based on this study that various up regulated genes and down regulated genes was observed significant differences between High Pathogenicity, High-Medium Pathogenicity versus Low

Pathogenicity **(Fig 2,3 and Table 3)**. There were six genes to be genetic marker candidates to confirm the pathogenicity of *Salmonella enteritidis*-contaminated egg albumin. The candidates were unspecified genes for particular expressed proteins**(Table 4)**.

The first gene was a phage protein-encoding gene (ID 7230716) and showed decreased expression levels in High, High-Medium Pathogenicity and increased in Low Pathogenicity **(Fig 4).**The phage protein in Salmonella plays a crucial role in the interaction between bacteriophages and the host bacterium. These proteins are involved in host recognition, attachment, and infection. For instance, the tailspike protein (TSP) from P22 bacteriophage has been shown to enable selective real-time detection of Salmonella, demonstrating its importance in identifying the pathogen(Singh et al., 2013). Additionally, the tail fiber protein YSD1_29 has been identified as a key component in the infection of Salmonella by the flagellotropic bacteriophage YSD1 (Dunstan et al., 2019).Furthermore, the gene product Gp17 protein of Salmonella phage P22 has been found to be associated with super infection exclusion, highlighting its role in phage therapy(Gendre et al., 2022).

Moreover, the phage tailspike protein alone has been demonstrated to effectively control Salmonella colonization and spread in chickens, indicating its potential for therapeutic applications(Miletic et al., 2016). The tailspike protein of bacteriophage Sf6 has been found to be functionally equivalent to that of Salmonella phage P22, mediating the attachment of the viral particle to the host cell-surface polysaccharide(Freiberg et al., 2003). Additionally, the tail spike protein (TSP) of Salmonella phage SS3e has been identified as a critical unit for host recognition and phage DNA ejection(Kim et al., 2012).

Furthermore, the E34 phage tailspike protein has shown promising results in protecting cells from Salmonella infection, suggesting its potential for therapeutic or preventive medicine(Ayariga et al., 2021). Additionally, the phage protein E4-54 has been found to be highly similar to the putative holin of Salmonella phage jersey, indicating its potential role in combating Salmonella infections (Torkashvand et al., 2022)

The second gene (ID 7228618)was the heat shock protein HtpX encoding gene. Expression of this gene significantly decreased in Low Pathogenicity of *Salmonella enteritidis***(Fig 4).**Heat shock protein HtpX in Salmonella is a crucial component for the bacterium's response to stress conditions. HtpX is a heat-inducible protein with sequence features of a membrane protein and a metalloprotease(Sakoh et al., 2005).It is part of the heat shock regulon and is involved in countering oxidative stress and surviving macrophage killing (Hews et al., 2019). Additionally, HtpX is a cytoplasmic membrane-bound Zn2+-dependent metalloprotease that is conserved in numerous bacteria(Lin et al., 2012). The gene encoding HtpX contributes to Salmonella Typhimurium persistence in intestine-associated tissues of pigs (Verbrugghe et al., 2015). Furthermore, the heat shock protein family gene coding for HSP15, which includes HtpX, is used for the detection of Salmonella in chicken meat (Tsen et al., 2013)

The function of HtpX is closely related to the heat shock response, a strategy widely used by bacteria to achieve protein quality control against misfolded and denatured proteins (Lin et al., 2012). HtpX is a heat shock gene induced by a temperature increase, and its over-expression leads to an increase in the degradation of abnormal proteins(Vickerman et al., 2002). Moreover, the induction of heat shock proteins, including HtpX, is crucial for Salmonella's response to stress conditions such as oxidative stress and macrophage killing(Hews et al., 2019).

The third gene (ID 7232516) and the sixth gene (ID 7237221) were both hypothetical protein-coding genes. Gene expression was also significantly decreased in the Low Pathogenicity group.Hypothetical proteins are proteins whose existence has been anticipated, but for which there are certain scarcities of experimental evidence about their structure, function, or linkage to any known genes(Galperin, 2004). These proteins are often predicted from nucleic acid sequences only and have unknown functions(Desler et al., 2009). When an open reading frame is annotated as a 'conserved hypothetical' protein, it does not necessarily mean that the function of its product is completely unknown, let alone that its very existence is questionable (Chandrasekaran et al., 2019). Conserved hypothetical proteins refer to proteins with phylogenetic lineages with no known definitive function(Galperin, 2004). The term "hypothetical protein" is used for a protein that is predicted to be expressed from an open reading frame but for which there is no experimental evidence of translation(Desler et al., 2009).

In the context of specific organisms, such as Salmonella, hypothetical proteins may play crucial roles in pathogenesis and virulence. For instance, the hypothetical protein EMK97_00595 in Salmonella has been studied to elucidate its structure and function using bioinformatics tools (Kader et al., 2022). Additionally, the comprehensive subcellular proteomic survey of Salmonella has provided insights into the subcellular localization of hypothetical proteins, aiding in understanding the bacterium's adaptation to intracellular environments.

The fourth gene (ID 7233752) is a nucleotidyl multifunctional tRNA protein-coding gene. The expression of this gene also showed a significant decrease in expression. Salmonella's nucleotidyl multifunctional tRNA protein is a crucial component involved in various cellular activities. This multifunctional protein is part of the multi-aminoacyl tRNA synthetase complex (MARS), which consists of aminoacyl tRNA synthetases (ARSs) and non-synthetase scaffold proteins(Eswarappa & Fox, 2013). The MARS complex plays a significant role in tRNA processing, RNA splicing, trafficking, apoptosis, and transcriptional and translational regulation(Ko et al., 2002). Additionally, it has been demonstrated that in higher eukaryotes, the multi-synthetase complex (MSC) is composed of eight aa-tRNA synthetases and three scaffold proteins, AIMP1, 2, and 3(Schwarz et al., 2018). This complex compartmentalizes amino acid-tRNA coupling, highlighting the multifunctional nature of the tRNA protein in cellular processes.

Furthermore, the tRNA protein is involved in RNA synthesis, mRNA capping, methylation, and polyadenylation(Matsumoto et al., 2015). Its involvement in these processes underscores its multifunctional nature and its significance in fundamental cellular activities. Additionally, the tRNA protein's role in Salmonella virulence is controlled by ubiquitin-dependent delivery, emphasizing its importance in bacterial pathogenesis(Thomas & Holden, 2009). Moreover, in E. coli and Salmonella, it has been shown that GidA and MnmE bind together to modify tRNA using a post-transcriptional mechanism, indicating the involvement of tRNA modification enzymes in bacterial pathogens(Shippy & Fadl, 2014).

The fifth gene (ID 7236581) is the anaerobic C4-dicarboxylate transporter encoding gene. This gene showed a significant decrease in expression in the Low Pathogenicity of *Salmonella enteritidis group*.The anaerobic C4-dicarboxylate transporter in Salmonella is a crucial component for the uptake and exchange of C4-dicarboxylates during anaerobic growth. Studies have identified multiple genes and proteins involved in this process in related bacteria such as *Escherichia coli* and *Actinobacillus succinogenes*. The dcuA and dcuB genes in E. coli are essential for anaerobic C4-dicarboxylate transport(Six et al., 1994), and the triple mutant (dcuA dcuB dcuC) in E. coli lacks C4-dicarboxylate transport during anaerobic growth(Zientz et al., 1996). Additionally, the DcuS-DcuR system in *E. coli*

controls gene expression in response to C4-dicarboxylates, particularly during anaerobic conditions(Golby et al., 1999). Furthermore, the E. coli homologue YchM (DauA) has been identified as a C4-dicarboxylic acid transporter, capable of C4-dicarboxylate exchange and uptake, with a preference for fumarate/succinate antiporters(Karinou et al., 2013).

In Salmonella, the anaerobic C4-dicarboxylate transporter likely plays a similar role to that in E. coli. The transport of C4-dicarboxylates is crucial for anaerobic growth and respiration, and the absence of specific transporters can lead to a complete loss of C4-dicarboxylate transport capabilities(Six et al., 1994; Zientz et al., 1996).

It can be seen that those six genes were critical in influencing the pathogenicity of *Salmonella enteritidis,* contaminating egg albumen.This research, while constrained by its reliance on bioinformatics investigations employing GEO Analysis of extant genomic databases, yields exceptionally promising findings in the identification of biomarker candidates for Salmonella enteritidis contamination in egg albumen. Additionally, the study delineates the sequences of each biomarker **(Table 4)**, potentially serving as probes for real-time PCR in future gene expression analyses. These analyses aimed to detect highly pathogenic strains of *Salmonella enteritidis* in egg albumen, as evidenced by the gene expression profiles presented in **Table 5.**

Table 5. The distinct of genetic biomarker expression for High pathogenicity of Salmonella enteritidis in egg albumen

No	Gene name	Protein	High, High-Medium Pathogenicity	Low Pathogenicity	Implications as High Pathogenicity
1	NA	phage protein	Strong negative regulated expression	Weak positive regulated expression	Negative expression induces to prevent Salmonella to be recognized by Immune cell
2	NA	heat shock protein HtpX	Weak positive regulated expression	Strong negative regulated expression	Positve expression induces to prevent protein of bacteria misfolded and denaturated
3	smpA	hypothetical protein	Weak positive regulated expression	Strong negative regulated expression	Positive expression induces bacteria adaptation capability toward environment
4	Cca	multifunctional tRNA nucleotidyl	Weak positive regulated expression	Strong negative regulated expression	Positive expression induces to tRNA regulation for bacterial activities
5	dcuB	anaerobic C4-dicarboxylate transporter	Weak positive regulated expression	Strong negative regulated expression	Positive expression induces the growth in anaerobic conditions.
6	yjjW	hypothetical protein	Weak positive regulated expression	Strong negative regulated expression	Positive expression induce bacteria adaptation capability toward environment

Conclusion

The culmination of this study is the identification of six distinct genes, characterized by their unique levels of expression, which potentially serve as biomarkers for differentiating the pathogenicity levels of *Salmonella enteritidis* in contaminated egg whites. However, this finding necessitates further investigative research, particularly focusing on the protein expression of Salmonella enteriditis in egg albumen. Such additional research is essential to definitively ascertain the specific levels of pathogenicity associated with these gene expressions, thereby enhancing our understanding of the bacterial behavior in contaminated environments.

Acknowledgments

The researcher expressed his gratitude to NCBI for data mining and open-source analysis with GEO. The researchers also thank Dr Davendra H Shah and the Department of Veterinary Microbiology and Pathology, Washington State University team for sharing the database in the GEO Dataset NCBI platform.

References

Ayariga, J. A., Gildea, L., Wu, H., & Villafane, R. (2021). *The E34 Phage Tailspike Protein: An in vitro characterization, Structure Prediction, Potential Interaction with* S. newington *LPS and Cytotoxicity Assessment to Animal Cell Line* [Preprint]. Microbiology. https://doi.org/10.1101/2021.09.20.461090

Chandrasekaran, M., Raman, C., Karthikeyan, K., & Paramasivan, M. (2019). Functional Annotation of Hypothetical Proteins Derived from Suppressive Subtraction Hybridization (SSH) Analysis Shows NPR1 (Non-Pathogenesis Related)-Like Activity. *Agronomy*, *9*(2), 57. https://doi.org/10.3390/agronomy9020057

Chousalkar, K. K., Sexton, M., McWhorter, A., Hewson, K., Martin, G., Shadbolt, C., & Goldsmith, P. (2017). *Salmonellatyphimurium* in the Australian egg industry: Multidisciplinary approach to addressing the public health challenge and future directions. *Critical Reviews in Food Science and Nutrition*, *57*(12), 2706–2711. https://doi.org/10.1080/10408398.2015.1113928

Clavijo, R. I., Loui, C., Andersen, G. L., Riley, L. W., & Lu, S. (2006). Identification of Genes Associated with Survival of *Salmonella enterica* Serovar Enteritidis in Chicken Egg Albumen. *Applied and Environmental Microbiology*, *72*(2), 1055–1064. https://doi.org/10.1128/AEM.72.2.1055-1064.2006

Clough, E., & Barrett, T. (2016). The Gene Expression Omnibus Database. In E. Mathé & S. Davis (Eds.), *Statistical Genomics* (Vol. 1418, pp. 93–110). Springer New York. https://doi.org/10.1007/978-1-4939-3578-9_5

Davies, R. H., & Breslin, M. (2003). Investigation of Salmonella contamination and disinfection in farm egg-packing plants. *Journal of Applied Microbiology*, *94*(2), 191–196. https://doi.org/10.1046/j.1365-2672.2003.01817.x

Desler, C., Suravajhala, P., Sanderhoff, M., Rasmussen, M., & Rasmussen, L. J. (2009). In Silico screening for functional candidates amongst hypothetical proteins. *BMC Bioinformatics*, *10*(1), 289. https://doi.org/10.1186/1471-2105-10-289

Dunstan, R. A., Pickard, D., Dougan, S., Goulding, D., Cormie, C., Hardy, J., Li, F., Grinter, R., Harcourt, K., Yu, L., Song, J., Schreiber, F., Choudhary, J., Clare, S., Coulibaly, F., Strugnell, R. A., Dougan, G., & Lithgow, T. (2019). The flagellotropic bacteriophage YSD1 targets *Salmonella* Typhi with a Chi-like protein tail fibre. *Molecular Microbiology*, *112*(6), 1831–1846. https://doi.org/10.1111/mmi.14396

Eswarappa, S. M., & Fox, P. L. (2013). Citric acid cycle and the origin of MARS. *Trends in Biochemical Sciences*, *38*(5), 222–228. https://doi.org/10.1016/j.tibs.2013.01.005

Fels, U., Gevaert, K., & Van Damme, P. (2017). Proteogenomics in Aid of Host–Pathogen Interaction Studies: A Bacterial Perspective. *Proteomes*, *5*(4), 26. https://doi.org/10.3390/proteomes5040026

Freiberg, A., Morona, R., Van Den Bosch, L., Jung, C., Behlke, J., Carlin, N., Seckler, R., & Baxa, U. (2003). The Tailspike Protein of Shigella Phage Sf6. *Journal of Biological Chemistry*, *278*(3), 1542–1548. https://doi.org/10.1074/jbc.M205294200

Galperin, M. Y. (2004). "Conserved hypothetical" proteins: Prioritization of targets for experimental study. *Nucleic Acids Research*, *32*(18), 5452–5463. https://doi.org/10.1093/nar/gkh885

Gantois, I., Eeckhaut, V., Pasmans, F., Haesebrouck, F., Ducatelle, R., & Van Immerseel, F. (2008). A comparative study on the pathogenesis of egg contamination by different serotypes of *Salmonella*. *Avian Pathology*, *37*(4), 399–406. https://doi.org/10.1080/03079450802216611

Gendre, J., Ansaldi, M., Olivenza, D. R., Denis, Y., Casadesús, J., & Ginet, N. (2022). Genetic Mining of Newly Isolated Salmophages for Phage Therapy. *International Journal of Molecular Sciences*, *23*(16), 8917. https://doi.org/10.3390/ijms23168917

Golby, P., Davies, S., Kelly, D. J., Guest, J. R., & Andrews, S. C. (1999). Identification and Characterization of a Two-Component Sensor-Kinase and Response-Regulator System (DcuS-DcuR) Controlling Gene Expression in Response to C_4 -Dicarboxylates in *Escherichia coli*. *Journal of Bacteriology*, *181*(4), 1238–1248. https://doi.org/10.1128/JB.181.4.1238-1248.1999

Hews, C. L., Pritchard, E. J., & Rowley, G. (2019). The Salmonella Specific, σE-Regulated, STM1250 and AgsA, Function With the sHsps IbpA and IbpB, to Counter Oxidative Stress and Survive Macrophage Killing. *Frontiers in Cellular and Infection Microbiology*, *9*, 263. https://doi.org/10.3389/fcimb.2019.00263

Humphrey, T. J., Whitehead, A., Gawler, A. H. L., Henley, A., & Rowe, B. (1991). Numbers of *Salmonella enteritidis* in the contents of naturally contaminated hens' eggs. *Epidemiology and Infection*, *106*(3), 489–496. https://doi.org/10.1017/S0950268800067546

Jia, S., McWhorter, A. R., Andrews, D. M., Underwood, G. J., & Chousalkar, K. K. (2020). Challenges in Vaccinating Layer Hens against Salmonella Typhimurium. *Vaccines*, *8*(4), 696. https://doi.org/10.3390/vaccines8040696

Kader, Md. A., Ahammed, A., Khan, Md. S., Al Ashik, S. A., Islam, Md. S., & Hossain, M. U. (2022). Hypothetical protein predicted to be tumor suppressor: A protein functional analysis. *Genomics & Informatics*, *20*(1), e6. https://doi.org/10.5808/gi.21073

Kang, H., Loui, C., Clavijo, R. I., Riley, L. W., & Lu, S. (2006). Survival characteristics of *Salmonella enterica* serovar Enteritidis in chicken egg albumen. *Epidemiology and Infection*, *134*(5), 967–976. https://doi.org/10.1017/S0950268806006054

Karinou, E., Compton, E. L. R., Morel, M., & Javelle, A. (2013). The *E scherichia coli* SLC26 homologue YCHM (DAUA) is a C_4 -dicarboxylic acid transporter. *Molecular Microbiology*, *87*(3), 623–640. https://doi.org/10.1111/mmi.12120

Kim, S.-H., Park, J.-H., Lee, B.-K., Kwon, H.-J., Shin, J.-H., Kim, J., & Kim, S. (2012). Complete Genome Sequence of Salmonella Bacteriophage SS3e. *Journal of Virology*, *86*(18), 10253–10254. https://doi.org/10.1128/JVI.01550-12

Ko, Y.-G., Park, H., & Kim, S. (2002). Novel regulatory interactions and activities of mammalian tRNA synthetases. *PROTEOMICS*, *2*(9), 1304–1310. https://doi.org/10.1002/1615-9861(200209)2:9<1304::AID-PROT1304>3.0.CO;2-E

Li, B., Ju, F., Cai, L., & Zhang, T. (2015). Profile and Fate of Bacterial Pathogens in Sewage Treatment Plants Revealed by High-Throughput Metagenomic Approach. *Environmental Science & Technology*, *49*(17), 10492–10502. https://doi.org/10.1021/acs.est.5b02345

Lin, T.-H., Huang, S.-C., & Shaw, G.-C. (2012). Reexamining Transcriptional Regulation of the *Bacillus subtilis htpX* Gene and the *ykrK* Gene, Encoding a Novel Type of Transcriptional Regulator, and Redefining the YkrK Operator. *Journal of Bacteriology*, *194*(24), 6758–6765. https://doi.org/10.1128/JB.01258-12

Liu, J., Feng, M., Li, S., Nie, S., Wang, H., Wu, S., Qiu, J., Zhang, J., & Cheng, W. (2020). Identification of molecular markers associated with the progression and prognosis of endometrial cancer: A bioinformatic study. *Cancer Cell International*, *20*(1), 59. https://doi.org/10.1186/s12935-020-1140-3

Lu, S., Killoran, P. B., & Riley, L. W. (2003). Association of *Salmonella enterica* SerovarEnteritidis YafD with Resistance to Chicken EggAlbumen. *Infection and Immunity*, *71*(12), 6734–6741. https://doi.org/10.1128/IAI.71.12.6734-6741.2003

Luo, D., Liang, X., Xu, B., Liu, J., Wei, C., & Li, G. (2019). Rapid Discovery of Potential Drugs for Osteonecrosis of Femoral Head Based on Gene Expression Omnibus Database and Connectivity Map. *Orthopaedic Surgery*, *11*(6), 1209–1219. https://doi.org/10.1111/os.12533

Matsumoto, Y., Ohta, K., Yumine, N., Goto, H., & Nishio, M. (2015). Identification of two essential aspartates for polymerase activity in parainfluenza virus L protein by a minireplicon system expressing secretory luciferase. *Microbiology and Immunology*, *59*(11), 676–683. https://doi.org/10.1111/1348-0421.12329

McWhorter, A. R., & Chousalkar, K. K. (2019). From hatch to egg grading: Monitoring of Salmonella shedding in free-range egg production systems. *Veterinary Research*, *50*(1), 58. https://doi.org/10.1186/s13567-019-0677-4

Miletic, S., Simpson, D. J., Szymanski, C. M., Deyholos, M. K., & Menassa, R. (2016). A Plant-Produced Bacteriophage Tailspike Protein for the Control of Salmonella. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.01221

Qi, D., & Chen, K. (2021). Bioinformatics Analysis of Potential Biomarkers and Pathway Identification for Major Depressive Disorder. *Computational and Mathematical Methods in Medicine*, *2021*, 1–11. https://doi.org/10.1155/2021/3036741

Sakoh, M., Ito, K., & Akiyama, Y. (2005). Proteolytic Activity of HtpX, a Membrane-bound and Stress-controlled Protease from Escherichia coli. *Journal of Biological Chemistry*, *280*(39), 33305–33310. https://doi.org/10.1074/jbc.M506180200

Sari, C., Ajeng Erika Prihastuti Haskito, Dinda Rahma Kurniasari, & Fajar Shodiq Permata. (2023). Big data analysis descriptively of Brucella abortus cases in Indonesia during 2006-2020. *International Journal of Science and Research Archive*, *10*(2), 1048–1061. https://doi.org/10.30574/ijsra.2023.10.2.1077

Schwarz, M. A., Lee, D. D., & Bartlett, S. (2018). Aminoacyl tRNA synthetase complex interacting multifunctional protein 1 simultaneously binds Glutamyl-Prolyl-tRNA synthetase and scaffold protein aminoacyl tRNA synthetase complex interacting multifunctional protein 3 of

the multi-tRNA synthetase complex. *The International Journal of Biochemistry & Cell Biology*, *99*, 197–202. https://doi.org/10.1016/j.biocel.2018.04.015

Shah, D. H., Casavant, C., Hawley, Q., Addwebi, T., Call, D. R., & Guard, J. (2012). *Salmonella* Enteritidis Strains from Poultry Exhibit Differential Responses to Acid Stress, Oxidative Stress, and Survival in the Egg Albumen. *Foodborne Pathogens and Disease*, *9*(3), 258–264. https://doi.org/10.1089/fpd.2011.1009

Shi, Y., Chen, D., Ma, S., Xu, H., & Deng, L. (2021). Identification of Potential Biomarkers of Depression and Network Pharmacology Approach to Investigate the Mechanism of Key Genes and Therapeutic Traditional Chinese Medicine in the Treatment of Depression. *Evidence-Based Complementary and Alternative Medicine*, *2021*, 1–14. https://doi.org/10.1155/2021/2165632

Shippy, D., & Fadl, A. (2014). tRNA Modification Enzymes GidA and MnmE: Potential Role in Virulence of Bacterial Pathogens. *International Journal of Molecular Sciences*, *15*(10), 18267–18280. https://doi.org/10.3390/ijms151018267

Singh, A., Poshtiban, S., & Evoy, S. (2013). Recent Advances in Bacteriophage Based Biosensors for Food-Borne Pathogen Detection. *Sensors*, *13*(2), 1763–1786. https://doi.org/10.3390/s130201763

Six, S., Andrews, S. C., Unden, G., & Guest, J. R. (1994). Escherichia coli possesses two homologous anaerobic C4-dicarboxylate membrane transporters (DcuA and DcuB) distinct from the aerobic dicarboxylate transport system (Dct). *Journal of Bacteriology*, *176*(21), 6470–6478. https://doi.org/10.1128/jb.176.21.6470-6478.1994

Song, S., Li, B., Jia, Z., & Guo, L. (2020). Sirtuin 3 mRNA Expression is Downregulated in the Brain Tissues of Alzheimer's Disease Patients: A Bioinformatic and Data Mining Approach. *Medical Science Monitor*, *26*. https://doi.org/10.12659/MSM.923547

Thomas, M., & Holden, D. W. (2009). Ubiquitination—A Bacterial Effector's Ticket to Ride. *Cell Host & Microbe*, *5*(4), 309–311. https://doi.org/10.1016/j.chom.2009.03.010

Torkashvand, N., Kamyab, H., Shahverdi, A. R., Khoshayand, M. R., & Sepehrizadeh, Z. (2022). *Isolation, characterization, and genome investigation of vB_SenS_TUMS_E4, a polyvalent bacteriophage against Salmonella enteritidis* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-1679990/v1

Tsen, H.-Y., Shih, C.-M., Teng, P.-H., Chen, H.-Y., Lin, C.-W., Chiou, C.-S., Wang, H.-T. T., Chang, H.-F. G., Chung, T.-Y., Lee, P.-Y., & Chiang, Y.-C. (2013). Detection of Salmonella in Chicken Meat by Insulated Isothermal PCR. *Journal of Food Protection*, *76*(8), 1322–1329. https://doi.org/10.4315/0362-028X.JFP-12-553

Verbrugghe, E., Van Parys, A., Leyman, B., Boyen, F., Haesebrouck, F., & Pasmans, F. (2015). HtpG contributes to Salmonella Typhimurium intestinal persistence in pigs. *Veterinary Research*, *46*(1), 118. https://doi.org/10.1186/s13567-015-0261-5

Vickerman, M. M., Mather, N. M., Minick, P. E., & Edwards, C. A. (2002). Initial characterization of the *Streptococcus gordonii htpX* gene. *Oral Microbiology and Immunology*, *17*(1), 22–31. https://doi.org/10.1046/j.0902-0055.2001.00000.x

Wei, R., Wang, Z., Zhang, Y., Wang, B., Shen, N., E, L., Li, X., Shang, L., Shang, Y., Yan, W., Zhang, X., Ma, W., & Wang, C. (2020). *Bioinformatic Analysis Revealing Mitotic Spindle Assembly regulated NDC80 and MAD2L1 as Prognostic Biomarkers in Non-Small Cell Lung Cancer Development* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-43770/v3

Zhang, X., Wang, Z., Zeng, Z., Shen, N., Wang, B., Zhang, Y., Shen, H., Lu, W., Wei, R., Ma, W., & Wang, C. (2020). *Bioinformatic Analysis Identifing FGF1 Gene as a New Prognostic Indicator in Clear Cell Renal Cell Carcinoma * [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-127479/v1

Zientz, E., Six, S., & Unden, G. (1996). Identification of a third secondary carrier (DcuC) for anaerobic C4-dicarboxylate transport in Escherichia coli: Roles of the three Dcu carriers in uptake and exchange. *Journal of Bacteriology*, *178*(24), 7241–7247. https://doi.org/10.1128/jb.178.24.7241-7247.1996