LOKMAN HEKIM HEALTH SCIENCES

DOI: 10.14744/lhhs.2023.70003
Lokman Hekim Health Sci 2023;3(3):153–160

lokmanhekimhs.com

ORIGINAL ARTICLE

Unraveling the Influence of *Escherichia coli* Strains on Gene Expression in Colorectal Cancer Cells through *in Silico* Analysis

In Silico Analizi Yaklaşımı ile Escherichia coli Suşlarının Kolorektal Kanser Hücrelerinde Gen İfadesi Üzerindeki Etkisinin Araştırılması

Ayşegül Yılmaz[1], Seyhan Türk[2], Ayriana Baesmat Safari[1], Can Türk[1]

[1]**Department of Medical Microbiology, Lokman Hekim University Faculty of Medicine, Ankara, Türkiye**
[2]**Department of Biochemistry, Hacettepe University Faculty of Pharmacy, Ankara, Türkiye**

Abstract

Introduction: Various studies have demonstrated a clear link between mucosa-adherent *Escherichia coli* and colorectal cancer (CRC). This *in silico* study aims to identify commonly differentially expressed genes in CRC cells treated with distinct *E. coli* strains, regardless of their pathogenicity.

Methods: The raw data from the GEO database were normalized with the Affy package in the R software. We used computational methods (limma analysis, linear regression analysis, hierarchical clustering, and gene set enrichment analysis) to identify the transcriptome alterations induced by *E. coli* O:157.H7 and *E. coli* K-12 strains.

Results: Out of the 80 genes that were significantly differentially expressed in both *E. coli*-treated groups, only three genes (HIST1H4E, JUN, and TMEM267) demonstrated a high correlation (r>0.9) with the incubation time. TMEM267 is downregulated as the incubation time increases whereas HIST1H4E and JUN become upregulated. In addition, we determined nine common genesets that were significantly enriched in the Caco2 cell line after treatment either with E.coli O:157.H7 or K-12 strains.

Discussion and Conclusion: In this study, the changes in gene expressions resulting from the treatment of colon cancer cells with different *E. coli* strains were investigated. These findings highlight the potential impact of *E. coli* on cancer development and suggest that certain genes may play a role in mediating the effects of *E. coli* in the context of CRC. Further research is warranted to elucidate the underlying mechanisms and specific pathways involved in the interactions between *E. coli* and cancer cells.

Keywords: Cancer molecular biology; Clinical microbiology; Colorectal cancer; *Escherichia coli*; Molecular biology; Transcriptome analysis

Colorectal cancer (CRC) is a highly lethal form of cancer worldwide.[1,2] Several bacteria, such as Streptococcus spp., Bacteroides fragilis, and pathogenic *Escherichia coli*, have been implicated in colorectal carcinogenesis.[3] Although *E. coli* is a bacterium that is part of the human microbiota and the most prevalent Gram-negative aerobes and

Cite this article as: *Yılmaz A, Türk S, Baesmat Safari A, Türk C. Unraveling the Influence of Escherichia coli Strains on Gene Expression in Colorectal Cancer Cells through in Silico Analysis. Lokman Hekim Health Sci 2023;3(3):153–160.*

Correspondence: Can Türk, M.D. Lokman Hekim Üniversitesi Tıp Fakültesi, Tıbbi Mikrobiyoloji Anabilim Dalı, Ankara, Türkiye
E-mail: can.turk@lokmanhekim.edu.tr **Submitted:** 26.07.2023 **Accepted:** 13.09.2023

anaerobes for culture, numerous studies have established an obvious association between mucosa-adherent *E. coli* and CRC.[4] Certain strains of intestinal *E. coli* have the potential to influence the initiation and progression of CRC by exploiting virulence factors and inflammatory pathways.[5] Most of the bacteria in the flora can also be pathogenic factors.[6] *E. coli* strains are classified into four phylogenic groups (A, B1, B2, and D) based on virulence factor acquisition. Phylogroups A and B1 typically lack pathogenicity, whereas phylogroups B2 and D are associated with intestinal and extra-intestinal diseases. Intriguingly, specific strains of *E. coli* from phylogroup B2 have been linked to Crohn's disease, a chronic bowel disease that has been implicated in an increased risk of CRC.[7] Moreover, studies have revealed a higher prevalence of mucosa-associated and mucosa-internalized *E. coli* in CRC patients compared to controls, underscoring the importance of these bacteria in CRC progress.[4] Pathogenic *E. coli* strains produce many virulence factors, such as cytotoxic necrotizing factor, cycle inhibiting factor, and colibactin.[8,9]

This *in silico* study aims to identify commonly differentially expressed genes in CRC cells treated with distinct *E. coli* strains, regardless of their pathogenicity.

Materials and Methods

Ethics Committee approval is not required for this study.

Microarray Gene Expression Data

Transcription profiles of human Caco2 colon epithelial cells isolated from colon tissue with colorectal adenocarcinoma and treated with either *E. coli* K-12 or O157:H7 were obtained from the Gene Expression Omnibus (GEO) database (GSE50040).[10]

Data Processing and Normalization

The raw data obtained from the GEO database were normalized using the Affy package in R software (version 3.6.3). The normalized transcription profiles included a total of 21,256 different genes/45,119 probe sets. The dataset consisted of three control samples without bacterial treatment (control 60 min, control 90 min, and control 120 min), as well as three samples treated with *E. coli* K-12 or O157:H7 at different time points (Caco-2 cocultured with *E. coli* K12: 60 min, 90 min, and 120 min, and Caco-2 cocultured with *E. coli* O157:H7: 60 min, 90 min, and 120 min).

Statistical Analysis

In the following sections, we describe the statistical analyses used to analyze the data collected in our study. In all analyses, a significance level of $p<0.05$ was considered statistically meaningful. The software programs used are specified in the relevant segments below.

Identification of Differentially Expressed Genes

The limma package (version 3.26.8) was used for differential expression analysis. The normalized gene expression data from the control groups and groups treated with *E. coli* K-12 or O157:H7 were compared separately to identify significant differentially expressed genes. Limma, a software tool in R/Bioconductor, offers various options for analyzing gene expression data, handling complex experimental designs, and addressing issues related to small sample sizes. The Venny tool (version 2.1.0) was employed to identify the differentially expressed common genes after treatment with both *E. coli* strains.

Linear Regression Analysis

Pearson correlation coefficient analysis was performed to identify genes highly correlated with the incubation time among the significantly differentially expressed genes. The correlation coefficient-based Pearson's correlation absolute p was calculated using GraphPad Prism 5.0. Genes with p above 0.05 were selected. In addition, genes with Pearson correlation coefficient values (R-values) above 0.9 were also selected.[11]

Hierarchical Clustering

Gene set enrichment analysis (GSEA) was conducted according to the guidelines and procedures outlined in the GSEA User Guide (http://software.broadinstitute.org/gsea/docGSEAUserGuideFrame.html). The GSE50040 data were utilized for this analysis, specifically comparing the control group with 120 min of incubation to the groups treated with *E. coli* K-12 or O157:H7 for 120 min, respectively. The goal of this analysis was to identify significantly enriched genes belonging to specific gene sets and determine which gene sets were enriched in each group. The Venny tool was employed to determine the common top significant gene sets.

GSEA

The process of GSEA was conducted following the GSEA guideline procedure (http://software.broadinstitute.org/gsea/docGSEAUserGuideFrame.html). The analysis utilized the GSE50040 dataset. Specifically, the analysis focused on comparing the control group incubated for 120 min with the groups treated with *E. coli* K-12 or O157:H7 for the same duration. The aim was to observe the patterns of *E. coli* treatment across these groups. The primary

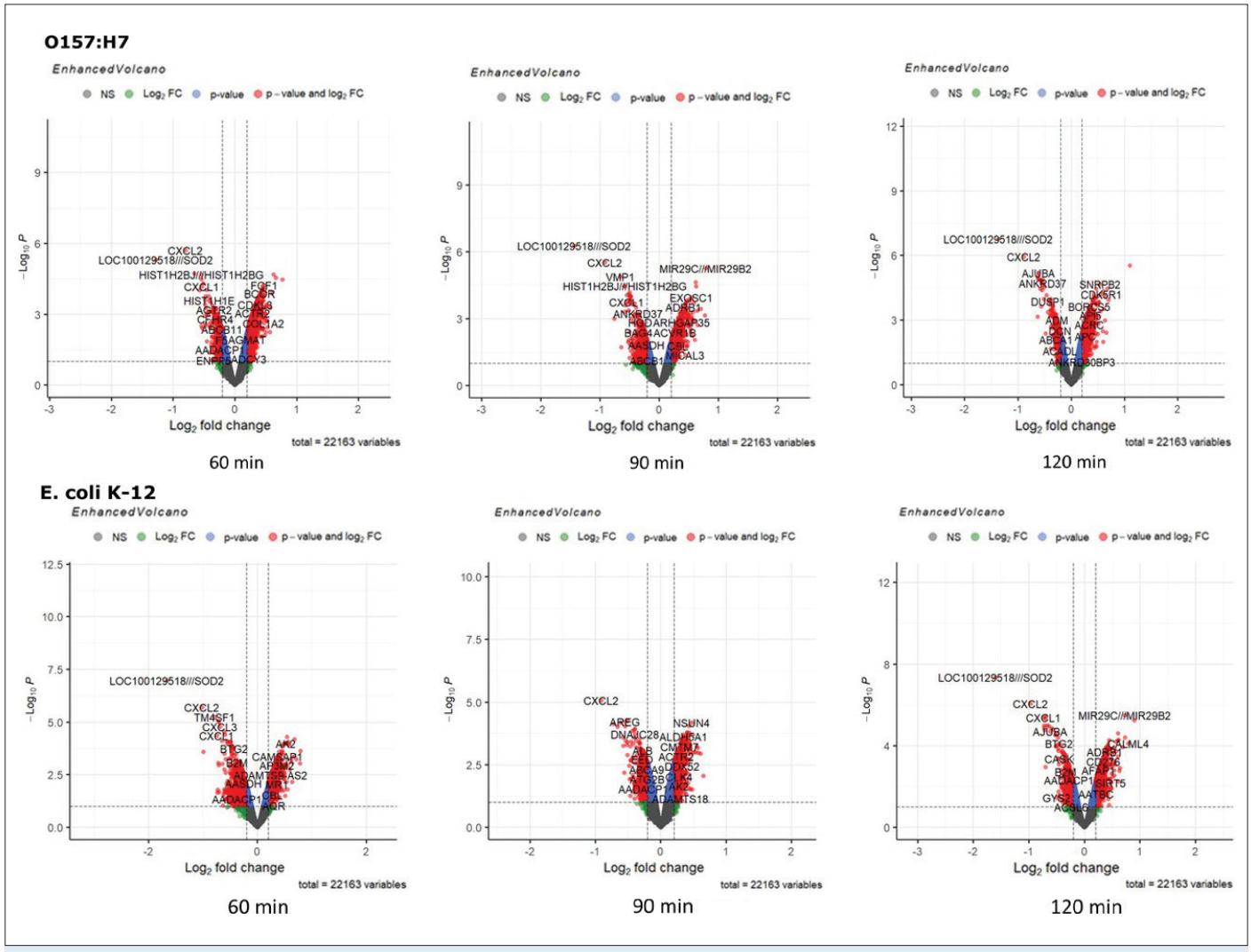

Figure 1. Comparison of differentially expressed genes in Caco2 cells treated with *E. coli* K-12 and E. coli O157:H7.

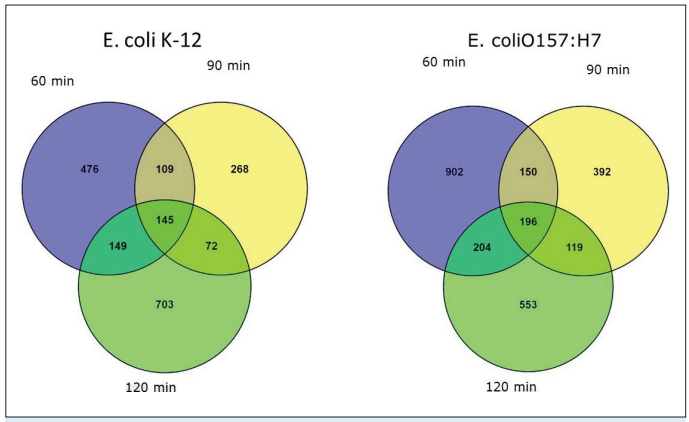

Figure 2. Overlap of differentially expressed genes in Caco2 cells treated with *E. coli* K-12 and E. coli O157:H7.

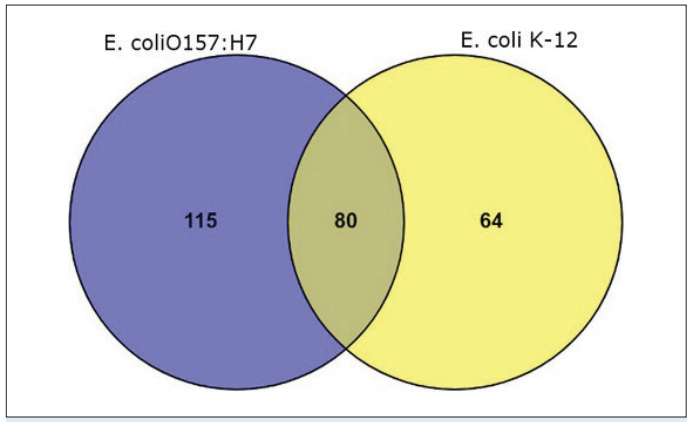

Figure 3. Identification of genes differentially expressed in both E. coli K-12 and *E. coli* O157:H7 treated Caco2 cells.

objective of this analysis was to identify gene sets within GSEA that exhibited significant enrichment for specific genes. In addition, the analysis aimed to determine which gene sets were enriched in the different groups. Ultimately, the Venny tool was employed to identify the common top significant gene sets.

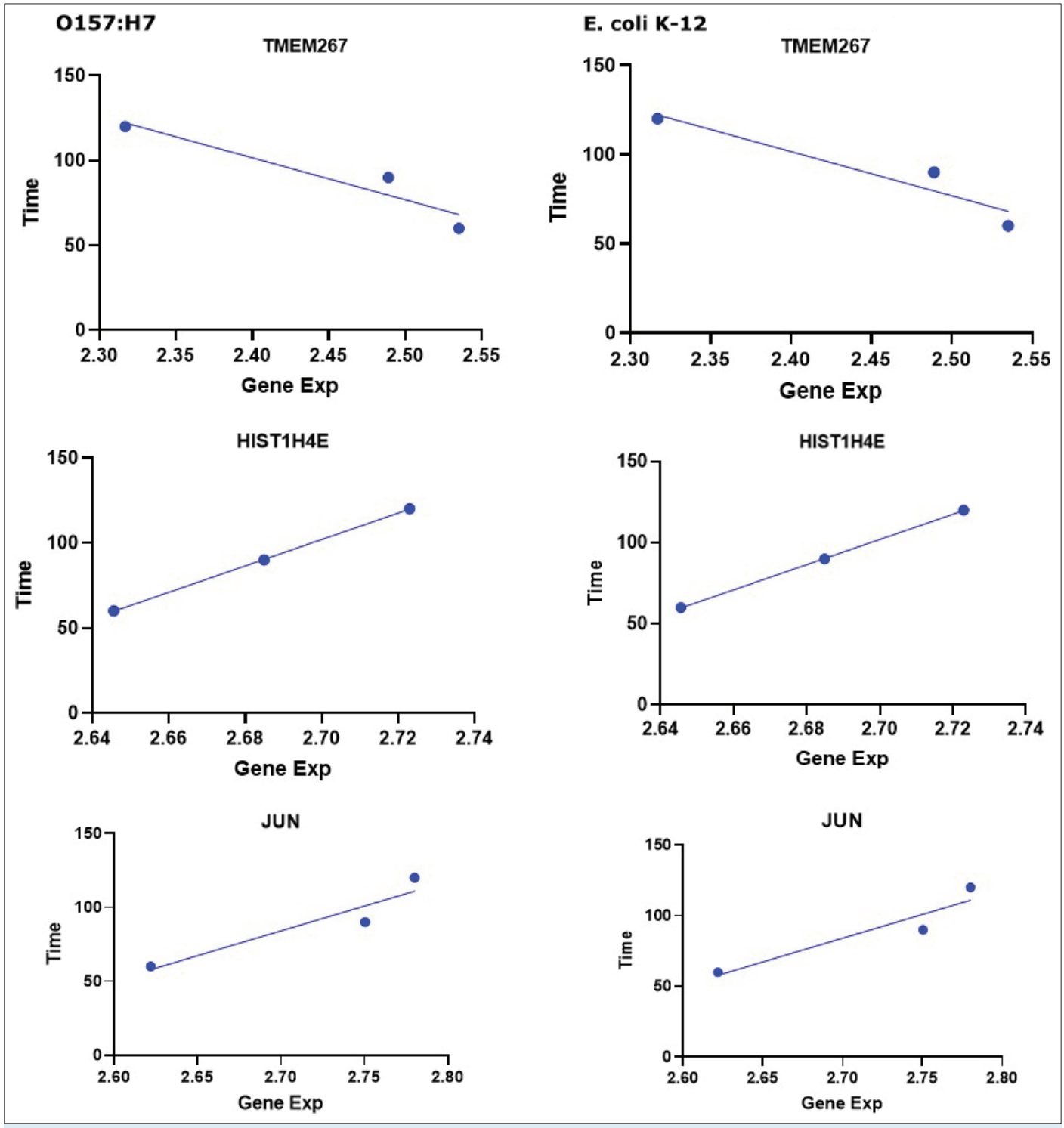

Figure 4. Correlation analysis of time-dependent gene expression changes in Caco2 cells treated with *E. coli* K-12 and E. coli O157:H7.

Results

Differential Gene Expression Analysis

Significant differences in gene expression were observed when comparing Caco2 cells treated with *E. coli* K-12 to the control groups. Specifically, 879 genes were differentially expressed between control 60 and *E. coli* K-12 60, 594 genes between control 90 and *E. coli* K-12 90, and 1069 genes between control 120 and *E. coli* K-12 120 ($p<0.05$). Similarly, comparing the control group to samples treated with *E. coli* O157:H7 revealed significant differential expression in 1452 genes between control 60 and *E. coli* O157:H7 60, 857 genes between

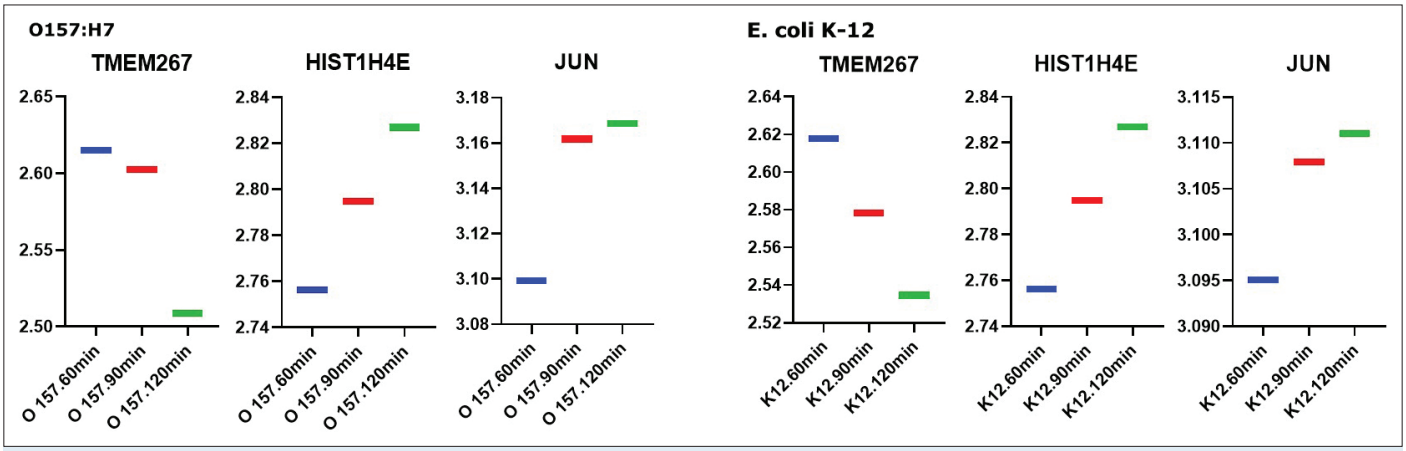

Figure 5. Heatmap clustering of Caco2 cell groups based on highly correlated genes (HIST1H4E, JUN, and TMEM267) in response to *E. coli* K-12 and *E. coli* O157:H7 treatment.

control 90 and *E. coli* O157:H7 90, and 1072 genes between control 120 and *E. coli* O157:H7 120 ($p<0.059$) (Fig. 1).

Identification of Common Differentially Expressed Genes

Distinct sets of differentially expressed genes were identified for each strain of *E. coli*. In Caco2 cells treated with *E. coli* K-12 for different periods, 145 genes showed differential expression. In contrast, 196 genes were found to be commonly differentially expressed in the Caco2 cells treated with *E. coli* O157:H7 (Fig. 2). Among these groups, 80 genes were significantly differentially expressed in both the *E. coli* K-12 and *E. coli* O157:H7 treated samples (Fig. 3).

Pearson Correlation Coefficient Analysis

Pearson correlation coefficient analysis was performed to explore the correlation between gene expression and incubation time. Out of the 80 genes that were significantly differentially expressed in both *E. coli*-treated groups, only three genes (HIST1H4E, JUN, and TMEM267) demonstrated a high correlation ($r>0.9$) with the incubation time. Notably, TMEM267 exhibited downregulation as the incubation time increased, whereas HIST1H4E and JUN showed upregulation (Fig. 4).

Hierarchical Clustering

Hierarchical clustering based on the expression patterns of the three correlated genes (HIST1H4E, JUN, and TMEM267) successfully distinguished between the different treatment groups, as demonstrated in the heatmap (Fig. 5).

GSEA

In Caco2 cells treated with either *E. coli* O157:H7 or K-12 for 120 min, nine common genesets were significantly enriched. These genesets included hormone activity, structural constituent of ribosome, signaling receptor regulator activity, ribosomal subunit, G protein-coupled receptor activity, large ribosomal subunit, neuropeptide hormone activity, and glutathione transferase activity (Fig. 6). Enrichment analysis results and corresponding p for each geneset in both bacteria-treated groups can be found in Table 1.

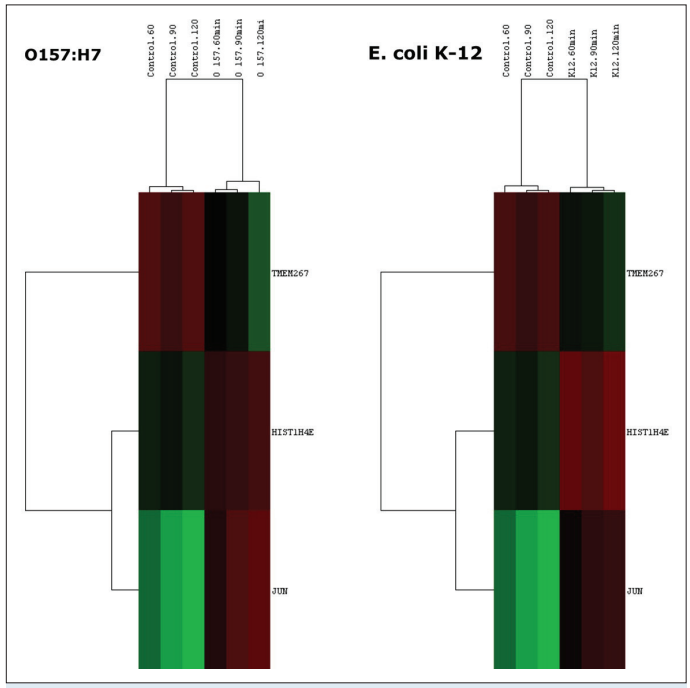

Figure 6. Enrichment analysis of common genesets in Caco2 cells treated with *E. coli* K-12 and *E. coli* O157:H7 for 120 min.

Discussion

E. coli is a Gram-negative bacterium commonly found in the human and animal gastrointestinal tract. While some strains, such as *E. coli* K12, are non-pathogenic, others such

Table 1. Enrichment and p of common genesets in Caco2 cells treated with *E. coli* K-12 and *E. coli* O157:H7

GS	ES		FDR q-val		FWER p-val	
	***E. coli* K-12**	**O157:H7**	***E. coli* K-12**	**O157:H7**	***E. coli* K-12**	**O157:H7**
GOMF_HORMONE_ACTIVITY	-0.34	-0.19	0.12	0.038	0.21	0.483
GOMF_STRUCTURAL_CONSTITUENT_OF_RIBOSOME	-0.26	-0.25	0.13	0.23	0.14	0.78
GOMF_SIGNALING_RECEPTOR_REGULATOR_ACTIVITY	-0.14	-0.1	0.14	0.005	0.18	0.063
GOCC_RIBOSOMAL_SUBUNIT	-0.23	-0.21	0.15	0.22	0.19	0.11
GOMF_G_PROTEIN_COUPLED_RECEPTOR_ACTIVITY	-0.13	-0.1	0.16	0.003	0.29	0.03
GOCC_LARGE_RIBOSOMAL_SUBUNIT	-3.16	-0.23	0.001	0.003	0.005	0.034
GOMF_NEUROPEPTIDE_HORMONE_ACTIVITY	-2.73	-0.38	0.005	0.041	0.086	0.594
GOMF_GLUTATHIONE_TRANSFERASE_ACTIVITY	-2.43	-0.37	0.028	0.124	0.575	0.983

as *E. coli* O157:H7, belonging to Shiga toxin-producing *E. coli* (STEC) serotypes, have been associated with outbreaks.[12] *E. coli* O157:H7 produces verotoxins that cause severe damage to the intestinal surface, leading to bloody diarrhea. In more vulnerable populations, such as young children and the elderly, this can progress to hemolytic uremic syndrome (HUS), a potentially life-threatening condition.[13] According to the WHO data, around 10% of individuals infected with STEC may develop HUS.[14]

CRC, composed of tumoral cells and microorganisms, is the third most common cancer worldwide and the second leading cause of cancer-related deaths.[15] Infections have been implicated in approximately 20% of all cancers, including sporadic CRC, where intestinal bacteria can contribute through mechanisms such as inducing inflammation, generating reactive oxygen species, and producing genotoxins.[9] Various studies have highlighted the association between colorectal carcinogenesis and bacterial species such as *B. fragilis*, *Helicobacter pylori*, and *Clostridium septicum*. Notably, there is a strong correlation between CRC and mucosa-associated *E. coli*, considering *E. coli*'s prevalence in the human intestine.[16,17] Recent investigations have also linked bacterial biofilms to human colon cancer.[18] Overall, *E. coli* appears to play a significant role in CRC carcinogenesis, although details regarding the characterization of colonic mucosa-associated *E. coli* in CRC patients are limited.[19]

Our study identified three genes (TMEM267, HIST1H4E, and JUN) that showed a high correlation when treating colon cells with both *E. coli* K12 and O157:H7 strains. Notably, the gene expressions of pathogenic and non-pathogenic strains exhibited similar changes over time. In our analysis, TMEM267 expression levels were observed to decrease with time in both strains. TMEM proteins are transmembrane proteins predicted to be present in various cell membranes, including mitochondria, endoplasmic reticulum, lysosomes, and Golgi membranes. Differential regulation of TMEM expression has been observed in various cancers.[2] A study has associated TMEM267 with tongue cancer.[20] In addition, Sun et al.[21] reported that loss of NUDT21 and the presence of oncogenes, including TMEM267, lead to unregulated tumor cell proliferation in hepatocellular carcinoma.

HIST1H4E, also known as H4 clustered histone 5, belongs to the histone H4 family. Several histone variants have been implicated in cancer progression due to their epigenetic roles.[22] Pozdeyev et al.[23] demonstrated high and specific expression of HIST1H4E, NOMO3, and NPIPA2 in medullary thyroid cancer. In our study, we observed an increase in HIST1H4E gene expression over time. Furthermore, a study showed that mRNA expression levels of both EML1 and HIST1H4E were elevated in metastatic cutaneous melanoma tissues.[24]

JUN genes (c-jun, jun-B, and jun-D) play crucial roles in various cellular functions such as proliferation, differentiation, and apoptosis.[25] c-Jun, encoded by the JUN gene, is an important transcription factor associated with cancer. It also plays a significant role in glucose metabolism and cancer metastasis. Zhu et al.[26] reported that Jun expression is downregulated in breast cancer development.

Peptide molecules, including hormones, neuropeptides, and cytokines, are involved in intercellular signaling across organs and tissues. A study identified a protein mimicking a peptide hormone in *E. coli*, suggesting the potential use of specific bacterial proteins as peptide-like drugs.[27] Ueta et al.[28] found that ribosomal protein L31 in *E. coli* contributes to ribosome subunit association and translation. In our study, we identified nine common genesets, including neuropeptide hormone activity and ribosomal subunit, significantly enriched in Caco2 cell lines after treatment with either *E. coli* O157:H7 or K-12 strains.

Conclusion

In this study, we investigated the gene expression changes induced by treating colon cancer cells with different strains of *E. coli*. The observed expression changes in three highly correlated genes suggest that *E. coli* may influence the progression and trajectory of cancer. Interestingly, treatment with both pathogenic and non-pathogenic strains resulted in expression changes in the same direction, indicating that the effects of *E. coli* on gene expression did not differ between the two types of strains.

These findings highlight the potential impact of *E. coli* on cancer development and suggest that certain genes may play a role in mediating the effects of *E. coli* in the context of CRC. Further research is warranted to elucidate the underlying mechanisms and specific pathways involved in the interactions between *E. coli* and cancer cells. Understanding the influence of *E. coli* on gene expression and its implications for cancer progression could provide valuable insights for the development of novel therapeutic strategies targeting CRC.

Peer-review: Externally peer-reviewed.

Authorship Contributions: Concept: CT, AY, ST, ABS; Design: CT, AY, ST, ABS; Supervision: CT, AY, ST, ABS; Data Collection or Processing: CT, AY, ST, ABS; Analysis or Interpretation: CT, ABS; Literature Search: AY, ST; Writing: AY, CT; Critical Review: CT, AY, ST, ABS.

Conflict of Interest: None declared.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Nouri R, Hasani A, Shirazi KM, Alivand MR, Sepehri B, Sotoodeh S, et al. *Escherichia coli* and colorectal cancer: unfolding the enigmatic relationship. Curr Pharm Biotechnol 2022;23(10):1257–68. [CrossRef]
2. Schmit K, Michiels C. TMEM proteins in cancer: a review. Front Pharmacol 2018;9:1345. [CrossRef]
3. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science 2012;338(6103):120–3.
4. Martin HM, Campbell BJ, Hart CA, Mpofu C, Nayar M, Singh R, et al. Enhanced *Escherichia coli* adherence and invasion in Crohn's disease and colon cancer. Gastroenterology 2004;127(1):80–93. [CrossRef]
5. Gagnière J, Raisch J, Veziant J, Barnich N, Bonnet R, Buc E, et al. Gut microbiota imbalance and colorectal cancer. World J Gastroenterol 2016;22(2):501–18. [CrossRef]
6. Türk C, Göçer S, Yılmaz A, Çelik G, Özgüven ŞV. Antibiotic resistance patterns of methicillin-resistant coagulase negative staphylococci isolated from blood cultures at a university hospital in Turkey. Malays J Microbiol 2022;18(6):602–11. [CrossRef]
7. Darfeuille-Michaud A, Boudeau J, Bulois P, Neut C, Glasser AL, Barnich N, et al. High prevalence of adherent-invasive *Escherichia coli* associated with ileal mucosa in Crohn's disease. Gastroenterology 2004;127(2):412–21. [CrossRef]
8. Lax AJ. Opinion: Bacterial toxins and cancer--a case to answer? Nat Rev Microbiol 2005;3(4):343–9. [CrossRef]
9. Bonnet M, Buc E, Sauvanet P, Darcha C, Dubois D, Pereira B, et al. Colonization of the human gut by *E. coli* and colorectal cancer risk. Clin Cancer Res 2014;20(4):859–67. [CrossRef]
10. He X, Mishchuk DO, Shah J, Weimer BC, Slupsky CM. Cross-talk between *E. coli* strains and a human colorectal adenocarcinoma-derived cell line. Sci Rep 2013;3:3416. [CrossRef]
11. Turk S, Turk C, Temirci ES, Malkan UY, Ucar G, Ozguven SV. Assessing the genetic impact of Enterococcus faecalis infection on gastric cell line MKN74. Ann Microbiol 2021;71(1):8. [CrossRef]
12. Melton-Celsa AR. Shiga toxin (Stx) classification, structure, and function. Microbiol Spectr 2014;2(4):EHEC-0024-2013. [CrossRef]
13. Rangel JM, Sparling PH, Crowe C, Griffin PM, Swerdlow DL. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982-2002. Emerg Infect Dis 2005;11(4):603–9. [CrossRef]
14. Siddiqui S, Yuan J. Binding characteristics study of DNA based Aptamers for *E. coli* O157:H7. Molecules 2021;26(1):204. [CrossRef]
15. Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi RE, Corcione F. Worldwide burden of colorectal cancer: a review. Updates Surg 2016;68(1):7–11. [CrossRef]
16. Raisch J, Buc E, Bonnet M, Sauvanet P, Vazeille E, de Vallée A, et al. Colon cancer-associated B2 *Escherichia coli* colonize gut mucosa and promote cell proliferation. World J Gastroenterol 2014;20(21):6560–72. [CrossRef]
17. Buc E, Dubois D, Sauvanet P, Raisch J, Delmas J, Darfeuille-Michaud A, et al. High prevalence of mucosa-associated *E. coli* producing cyclomodulin and genotoxin in colon cancer. PLoS One 2013;8(2):e56964. [CrossRef]
18. Coleman OI, Haller D. Microbe-mucus interface in the pathogenesis of colorectal cancer. Cancers (Basel) 2021;13(4):616.
19. Nouri R, Hasani A, Masnadi Shirazi K, Alivand MR, Sepehri B, Sotoudeh S, et al. Mucosa-Associated *Escherichia coli* in colorectal cancer patients and control subjects: variations in the prevalence and attributing features. Can J Infect Dis Med Microbiol 2021;2021:2131787. [CrossRef]
20. Reddy RB, Khora SS, Suresh A. Molecular prognosticators in clinically and pathologically distinct cohorts of head and neck squamous cell carcinoma-A meta-analysis approach. PLoS One 2019;14(7):e0218989. [CrossRef]
21. Sun M, Ding J, Li D, Yang G, Cheng Z, Zhu Q. NUDT21 regulates 3'-UTR length and microRNA-mediated gene silencing in hepatocellular carcinoma. Cancer Lett 2017;410:158–68. [CrossRef]
22. Boulard M, Bouvet P, Kundu TK, Dimitrov S. Histone variant nucleosomes: structure, function and implication in disease. Subcell Biochem 2007;41:71–89. [CrossRef]
23. Pozdeyev N, Erickson TA, Zhang L, Ellison K, Rivard CJ, Sams S, et al. Comprehensive immune profiling of medullary thyroid cancer. Thyroid 2020;30(9):1263–79. [CrossRef]

24. He Y, Liu H, Luo S, Amos CI, Lee JE, Yang K, et al. Genetic variants of EML1 and HIST1H4E in myeloid cell-related pathway genes independently predict cutaneous melanoma-specific survival. Am J Cancer Res 2021;11(6):3252–62.
25. Neyns B, Katesuwanasing, Vermeij J, Bourgain C, Vandamme B, Amfo K, et al. Expression of the jun family of genes in human ovarian cancer and normal ovarian surface epithelium. Oncogene 1996;12(6):1247–57.
26. Zhu P, Liu G, Wang X, Lu J, Zhou Y, Chen S, et al. Transcription factor c-Jun modulates GLUT1 in glycolysis and breast cancer metastasis. BMC Cancer 2022;22(1):1283. [CrossRef]
27. Fetissov SO, Legrand R, Lucas N. Bacterial protein mimetic of peptide hormone as a new class of protein- based drugs. Curr Med Chem 2019;26(3):546–53. [CrossRef]
28. Ueta M, Wada C, Bessho Y, Maeda M, Wada A. Ribosomal protein L31 in *Escherichia coli* contributes to ribosome subunit association and translation, whereas short L31 cleaved by protease 7 reduces both activities. Genes Cells 2017;22(5):452–71. [CrossRef]