RESEARCH ARTICLE

The Effect of Salmonella AvrA Infection on Colon Cancer: An *in-silico* Study

Bogi Pratomo Wibowo[1]*, Handono Kalim[2], Husnul Khotimah[3], Hidayat Sujuti[4], Reza Ardi Wibowo[5]

Abstract

Objective: Colorectal cancer (CRC) is one of the main causes of morbidity and mortality due to cancer. The purpose of this in-silico study was to examine the relationship of chronic infection mechanisms caused by Salmonella Anti virulence agent A (AvrA) to gene mutations in the carcinogenic process of CRC. **Methods:** Gene expression data on the mouse colon was obtained from the GSE22215 dataset | Gene Expression Omnibus (GEO). Adjusted p-value was calculated using Benjamini & Hochberg False Discovery Rate (FDR<0.01). Gene expression in colon adenocarcicoma tumors was obtained from The Cancer Genome Atlas's (TCGA) Genomic Data Commons (GDC) dataset containing 458 colon tumor samples. **Result:** Expressions of MLH1, MSH2, EPCAM, APC, and PMS2 in cases of colon adenocarcinoma tumor showed a correlation with genes that underwent changes due to Salmonella AvrA infection. Among the gens of interest, EPCAM was the gene that had the highest correlation compared to other genes (MLH1, MSH2, APC, and PMS2) (n= 514, Gene r-p value < 0.01 =22355). There were 514 genes that had a correlation with cases of AvrA infection. Tumor Necrosis Factor (TNF), which is a gene that is upregulated in AvrA infection and correlates negatively with EPCAM, had the highest BC value compared to other gens (p= 0.0000768). Survival probability showed that EPCAM was highly expressed and it can increase survival time. In addition to TNF, our study indicated that IL1B (p= 0.000419), S100A8 (p= 2.02E-05), S100A9 (p=0.000419) correlated with the gene of interest. **Conclusion:** Late Salmonella AvrA infection affects the expression of genes involved in inflammation in colorectal cancer samples.

Keywords: Colon Cancer- Salmonella- AvrA- Infection

Introduction

Colorectal cancer (CRC) is a malignant tumor in the colon and rectum. It is one of the leading causes of morbidity and mortality due to cancer, this malignancy ranks the fourth leading cause of death due to cancer worldwide. The WHO estimates there were 694,000 deaths caused by CRC in 2012 (Bernard et al., 2014). Lynch Syndrome is the dominant autosomal disorder that causes the most CRC of inherited genetic disorders. This syndrome is caused by mutase from one of the DNA mismatch-repair genes: MLH1, MSH2, MSH6, PMS2 or EPCAM. The second most common cause of hereditary CRC is Familial Adenomatous Polyposis caused by mutases from the adenomatous polyposis coli (APC) gene, which controls the activity of the Wnt signal pathway.

There is no optimal way for the prevention and treatment of CRC. Several studies have been conducted for this purpose, including efforts to discover the various pathomechanisms of CRC. A risk-sharing approach was used to learn more about CRC pathomechanism. The recent findings suggest that infectious agents and changes in gut microecology (Compare and Nardone, 2014) and inflammatory bowel disease (Lutgens et al., 2013) are also risk factors for CRC.

Salmonella is one of the most common infectious agents in the world. Every year approximately 93.8 million people are infected with Salmonella (Majowicz et al., 2010). Salmonella infection can become chronic, increasing the risk of cancer (Mughini-Gras et al., 2018). AvrA is a bacterial protein from Salmonella enterica that plays a significant role in causing chronic infections. AvrA is a pathogenic product of Salmonella, secreted as a T3SS effector protein that affects eukaryotic cell pathways (Liu et al., 2010) by regulating ubiquitination and a selection thereby inhibiting apoptosis and increasing

[1]Doctoral Program, Faculty of Medicine, Universitas Brawijaya, Indonesia. [2]Department of Internal Medicine, Faculty of Medicine, Universitas Brawijaya, Indonesia. [3]Department of Pharmacology, Faculty of Medicine, Universitas Brawijaya, Indonesia. [4]Department of Ophalmology Faculty of Medicine, Universitas Brawijaya, Indonesia. [5]Department of Internal Medicine, Faculty of Medicine, Universitas Indonesia, Indonesia. *For Correspondence: dr.rezaardiw@gmail.com

intestinal cell proliferation (Jones et al., 2008) which eventually increases tumorigenesis (Lu et al., 2014). AvrA activates the Wnt/β catenin signaling pathways (Lu et al., 2014) and STAT3 (Rong Lu et al., 2016) which increase carcinogenesis in the mouse colon. Salmonella AvrA is also detected in CRC of human (Rong Lu et al., 2017). The mechanism through which AvrA induces other eukaryotic signaling pathways needs further studies.

Infectious agents such as bacteria play a role in the development of CRC through inflammatory processes, induction of DNA damage with toxins, metabolites, and or manipulating host cell signaling pathways during the infection cycle (Chumduri et al., 2016; Gagnaire et al., 2017). The Aim of this study was to examine the link between the mechanisms of chronic infection caused by Salmonella AvrA to gene mutations in the carcinogenic process of CRC.

Materials and Methods

Sample Collection 1

Gene expression data on the mouse colon was obtained from the GSE22215 dataset | Gene Expression Omnibus (GEO). The dataset contained gene expressions undergone changes in the late phase (4 days of treatment) with Salmonella strain SL1344 (AvrA expression) infection and HBSS sterile treatment (control). The instrument used in the GSE22215 dataset was microarrays [MoGene-1_0-st] Affymetrix Mouse Gene 1.0 ST Array [transcript (gene) version]. Data was processed using GEO2R to obtain upregulated and downregulated gene lists. The adjustment p-value was calculated using Benjamini & Hochberg (False discovery rate) (FDR <0.01).

Gene Correlation of AvrA Infection with Gene Expression in Colon Adenocarcinoma Tumor Samples 2

Gene expression in colon adenocarcinoma tumors was obtained from the cancer genome atlas's (TCGA) genomic data commons (GDC) dataset containing 458 colon tumor samples. The data was normalized using Fragments Per Kilobase of transcript per Million mapped reads upper quartile (FPKM-UQ) gencode22. The gene interests used were MLH1, MSH2, MSH6, PMS2, EPCAM, and APC which play a role in repairing systems. Only MLH1, MSH2, PMS2, EPCAM, and APC can be analyzed in correlation with gene expression obtained at stage 1. The correlation score r(+) can be interpreted as a positive correlation in the gene, while r(-) means that the gene is negatively correlated. Negative correlation means that the gene analysis has the opposite expression. Data was processed using R^2: Genomics Analysis and Visualization Platform. False Rate Discovery <0.01 was used for data filtering (Figure 2).

Gene Annotation Analysis 3

The gene list that had a correlation of the influence of AvrA infection and gene expression in colon tumor cells was then performed functional analysis using Database for Annotation, Visualization and Integrated Discovery (DAVID). The purpose of this analysis was to find out the role of biological processes, molecular functions, cellular components, and pathways of the inputted gene list. The displayed data is terminology that has FDR < 0.05 (Table 2).

Network Analysis 4

Analysis of protein interaction was conducted using STRING DB V.11.5 (March 16, 2022, 4:30 PM). Gene list correlation AvrA-EPCAM analyzed its interaction using STRING DB with inputs of Homo sapiens organisms, active interaction sources (Textmining, Experiment, Databases) with confidence 0.7. This data was then processed using CytoScape V.9.0 to better visualize the data. CytoScape can also be used to analyze which genes have the most significant role in the pathway. Analysis using cytoscape is called Topological (Graph) Network Analysis. The parameters used included Betweennes Centrality (BC) and Degree (D). BC was used to examine the protein nodes that have the most essential information in a network. While Degree was used to examine the most targeted nodes (Figure 3).

Results

Gene Expression due to Salmonella INFECTION SL1344 (AvrA expression) 1

The results of the analysis using GEO2R showed that there were 1,219 genes with significant differences in expression between treatment and control. The P-value FDR used was < 0.01. The results of the analysis are displayed in the form of a volcano plot (Figure 1). Red plots are the genes that are upregulated, while blue-colored plots are the downregulated genes.

Gene Correlation of AvrA Infection with Gene Expression in Colon Adenocarcinoma Tumor Samples 2

The correlation of gene expression that significantly changes due to AvrA infection with EPCAM is stronger (Table 1). At least 514 significant genes appeared in the correlation of EPCAM and AvrA (Table 1). Although based on The Human Protein Atlas database, EPCAM has a low cancer specificity, its expression is found in many cases of colorectal cancer (TCGA Dataset) (Figure 4). When the EPCAM expression is high, it increases the patient's survival probability (Figure 5, Figure 8 and Figure 9).

Gene Ontology Analysis 3

Based on DAVID analysis, out of the 514 genes inputted (correlated with EPCAM) there were 473 identified in the Biological Process, 490 in cellular

Table 1. Gene Correlation affected by AvrA with Gene Interest Cancer Colon

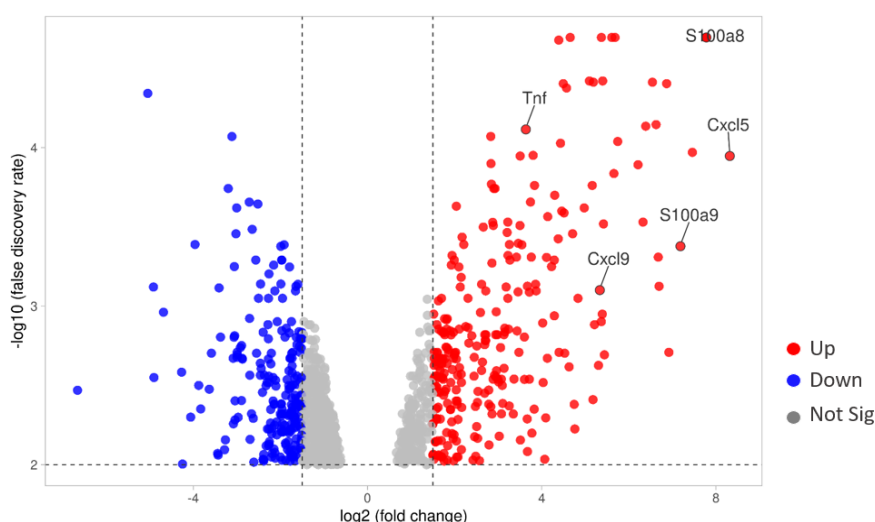| Gene Correlation	Gene r-p value < 0.01	Common Element AvrA Infection
EPCAM	22355	514
MLH1	8580	307
APC	14867	276
MSH2	4661	177
PMS2	6231	130

Figure 1. Volcano Plot Expression Avra vs Control

components, 480 in molecular function, and 243 in KEGG pathway (Table 3).

Network Analysis 4

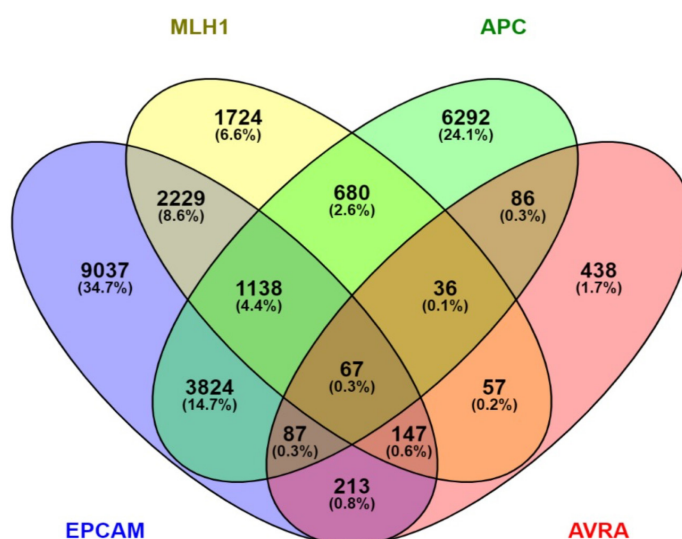Analysis of protein interaction was conducted using STRING DB V.11.5 (March 16, 2022, 4:30 PM). Gene list

Figure 2. Gene Correlation Affected by AvrA with EPCAM, APC and MLH1

Table 2. EPCAM Correlation with Genes Affected by AvrA Infection

ID	Gene	r-value	r-p value	Correlation	AvrA Infection	Fold Change	Adj. P value
ENSG00000143546.8	*S100A8*	-0.26761837	7.81E-08	negative	upregulated	7.7818787	0.0000202
ENSG00000232810.3	*TNF*	-0.21351842	2.96E-05	negative	upregulated	3.6358676	0.0000768
ENSG00000163735.6	*CXCL5*	-0.15829339	0.00272531	negative	upregulated	8.3203182	0.000113
ENSG00000125538.10	*IL1B*	-0.19950517	0.00010704	negative	upregulated	4.3797648	0.0003761
ENSG00000163220.10	*S100A9*	-0.28141872	1.33E-08	negative	upregulated	7.1868046	0.0004189
ENSG00000169245.5	*CXCL10*	-0.23612645	2.99E-06	negative	upregulated	3.8588917	0.0007279
ENSG00000138755.5	*CXCL9*	-0.31013654	2.36E-10	negative	upregulated	5.3351624	0.0007924
ENSG00000188257.9	*PLA2G2A*	-0.23473621	3.47E-06	negative	upregulated	4.2899871	0.0011514
ENSG00000100906.9	*NFKBIA*	-0.29675622	1.64E-09	negative	upregulated	1.4397641	0.0014125
ENSG00000067182.6	*TNFRSF1A*	-0.13854428	0.00980114	negative	upregulated	1.4393666	0.0014324
ENSG00000118503.13	*TNFAIP3*	-0.24896691	7.22E-07	negative	upregulated	1.4136701	0.0017402
ENSG00000132170.18	*PPARG*	0.463178142	2.53E-23	positive	downregulated	-1.1629524	0.0057457
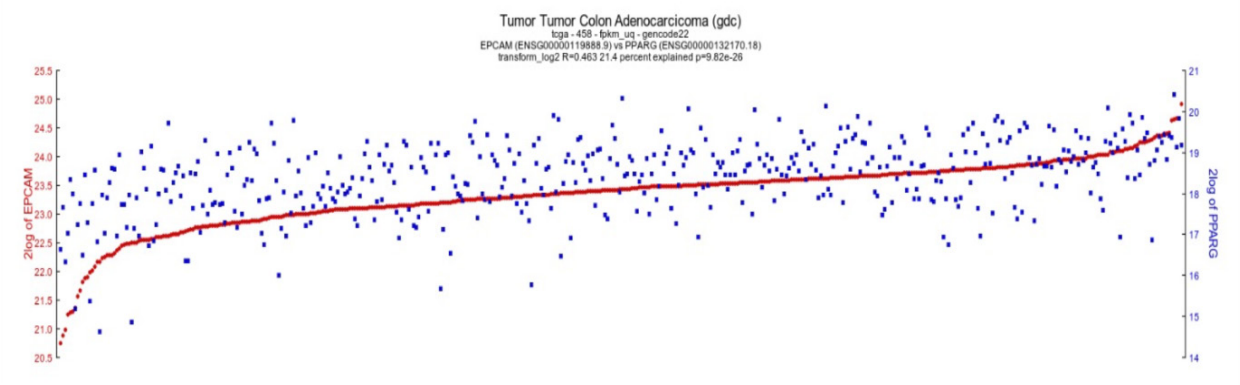
Figure 3. EPCAM and PPARG Expression Correlation

Table 3. Gene Ontology Analysis

Gene Ontology	Gene	Percent (%)
Biological Process	473 /514	94.0
Cellular Component	490/514	97.4
Molecular Function	480/514	95.4
KEGG Pathway	243/514	48.3

Table 4. Network Analysis

	Betweenness Centrality	Degree
TNF	0.400992	22
PPARA	0.390468	7
STAT1	0.340045	25
NFKBIA	0.120111	10
EPCAM	0.085485	5
PPARG	0.080588	8

correlation AvrA-EPCAM analyzed its interaction using STRING Database with inputs of Homo sapiens organisms, active interaction sources (Textmining, Experiment, Databases) with confidence 0.7. STRING database aims to integrate all known and predicted associations between proteins, including both physical interactions as well as functional associations. The data was then processed using CytoScape V.9.0 to better visualize the data. CytoScape can be also used to determine which genes have the most significant role in the pathway being analyzed (Table 4). Analysis using cytoscape is called Topological (Graph) Network Analysis. Parameters used included Betweennes Centrality (BC) and Degree (D). BC was used to examine the protein nodes that have the most valuable information

in a network. The degree was used to examine the most targeted nodes. TNF, which is a gene that is upregulated in AvrA infection and correlates negatively with EPCAM, had the highest BC value compared to other nodes. The higher the BC value, the greater the role of the node in the network. The network is indicated by the larger size of the node, the higher the BC score (Figure 6 and Figure 7).

Discussion

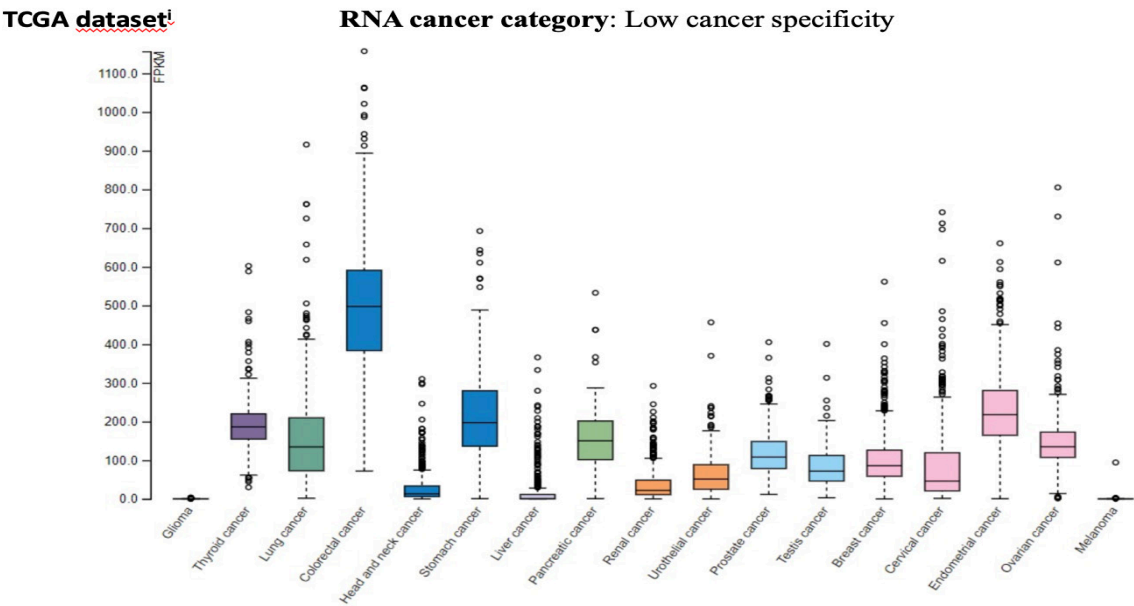Inflammatory bowel disease and ulcerative colitis

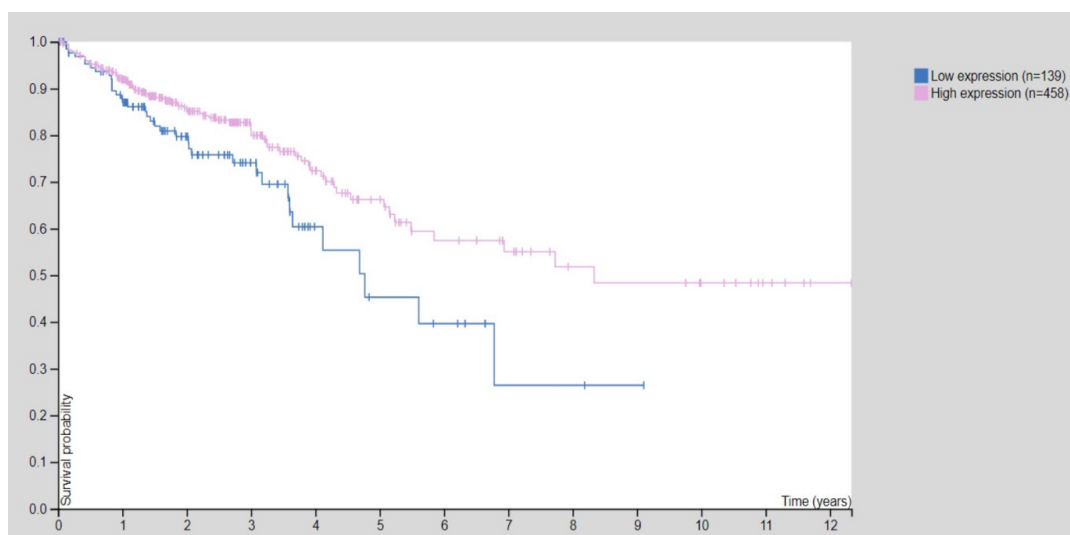Figure 4. EPCAM Expression on Cancer-Sharing (The Human Protein Atlas)

Figure 5. Survival Probability EPCAM in Cases of Colorectal Cancer (The Human Protein Atlas – TCGA)

are inflammation-related diseases of colorectal cancer. Certain cytokines, such as TNF-α and IFN-γ, are the key factors in determining the contribution of inflammatory processes to CRC. TNF-α regulates the induction of Associated Metastasis in Colon Cancer 1 (MACC1) through the NF-κB p65 subunit and c-Jun transcription factor in CRC cells (Kobelt et al., 2020). MACC1 can transcribe the genes involved in epithelial-mesenchymal transition (EMT), including those capable of directly inducing metastasis such as c-MET. Thus, it impacts the migration and invasion of tumor cells, and induces metastasis in solid cancers (Radhakrishnan et al., 2018). Immunohistochemistry shows elevated levels of TNF-R1 and p-JNK expression in adenoma epithelial

Table 5. Gene Interest Correlation with Gene Experiencing Expression Changes due to AvrA Infection

| Gene		r-value	r-p value	Correlation	AvrA Infection	Fold Change	Adj. P value
APC	S100A9	-0.18305909	0.00026	negative correlation	upregulated	7.186805	0.000419
EPCAM	S100A9	-0.28141872	1.33E-08	negative correlation	upregulated	7.186805	0.000419
MLH1	IL1B	-0.17298089	0.00197	negative correlation	upregulated	4.379765	0.000376
EPCAM	IL1B	-0.19950517	0.00011	negative correlation	upregulated	4.379765	0.000376
MLH1	S100A8	-0.15934096	0.00521	negative correlation	upregulated	7.781879	2.02E-05
EPCAM	S100A8	-0.26761837	7.81E-08	negative correlation	upregulated	7.781879	2.02E-05
MLH1	TLR4	0.183302989	0.00089	positive correlation	upregulated	1.06889	0.005959
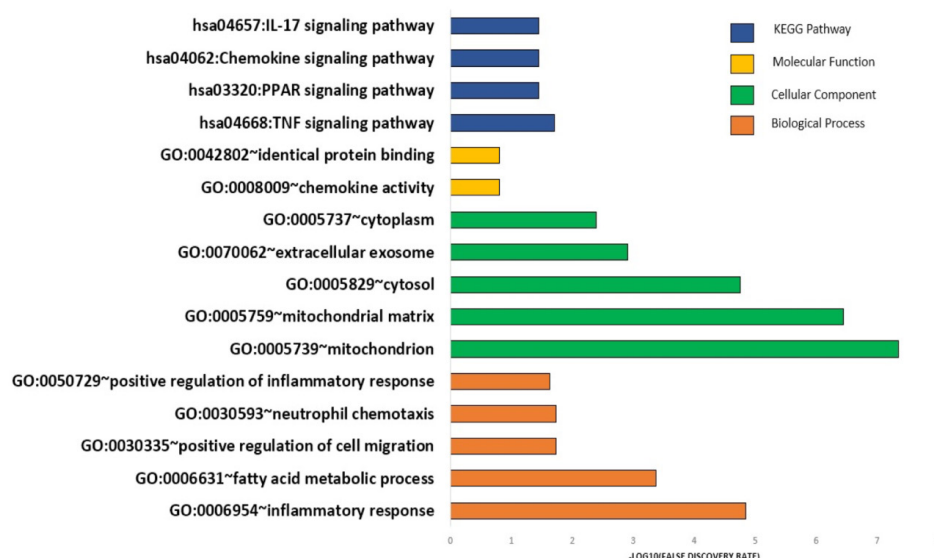
Figure 6. Gene Ontology Analysis
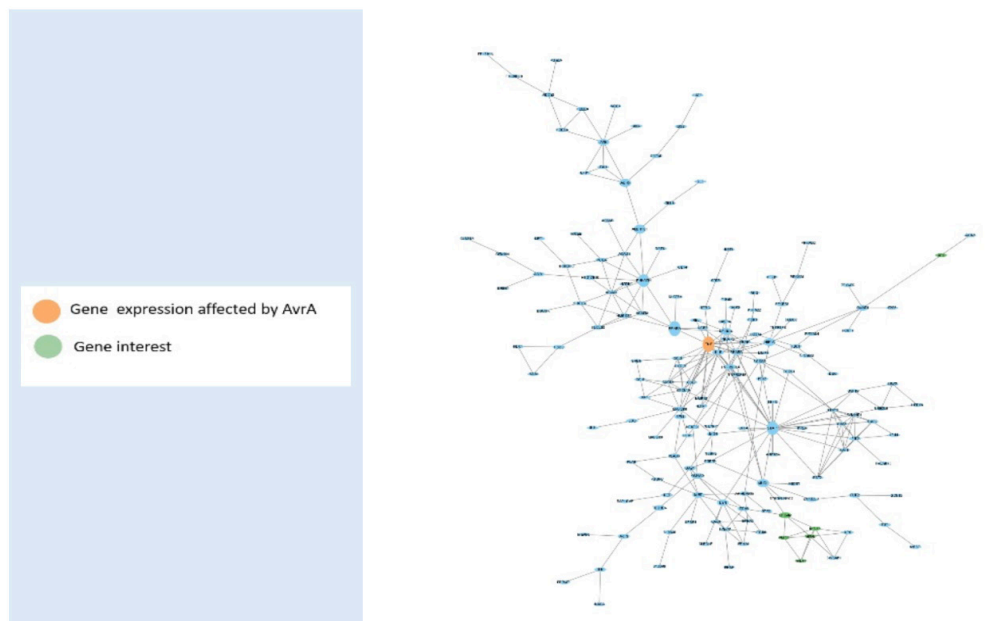
Figure 7. Protein-Protein Interaction

cells. Furthermore, high incidences of co-localization of TNF-R1 and p-JNK were identified in adenoma tissue. TNF-R1 may be a promising biomarker of colorectal adenomas and may also play a significant role in the early stages of colorectal carcinogenesis (Hosono et al., 2012).

EPCAM activity is negatively regulated by TNF-α, which induces inhibition to EPCAM expression mediated by TNF receptor 1 through TNF receptor-associated death domain protein (TRADD) and through the activation of nuclear factor kappaB (NF-kappaB). NF-kappaB can repress EPCAM expression by competing to bind transcriptional coactivator p300/CREB binding protein (p300/CBP) (Schmitt et al., 2001).

EPCAM expression is a poor prognosis of CRC. Studies have shown that EPCAM is overexpressed in the primary stage of cancer; however, its expression decreases during the developmental stage of the disease (Mokhtari and Zakerzade, 2017).

APCs play a significant role in several cellular processes, such as tumorigenesis suppression, cell development by lowering the Wnt pathway, actin tissue suppression and microtubules, chromosomal segregation, and cell adhesion, and migration. Therefore, mutations in APCs can play a role in tumorigenesis in the colon. (Yang et al., 2021) In the case of colon cancer, high APC expression can increase survival probability. APC is negatively correlated with S100A9; so that, when the expression of APC increases, the expression of S100A9 decreases. In the cases of late infection, AvrA shows a decrease in expression S100A9 and S100A8.

S100A8 (S100 calcium binding proteins A8 (calgranulin A) and S100A9 (S100 calcium binding protein A9 (calgranulin B) increase in more than 50% of CRC tissue and their expression in tumor cells is associated with differentiation, stage Dukes, and metastasis of lymph nodes. Studies have shown that S100A8 and S100A9
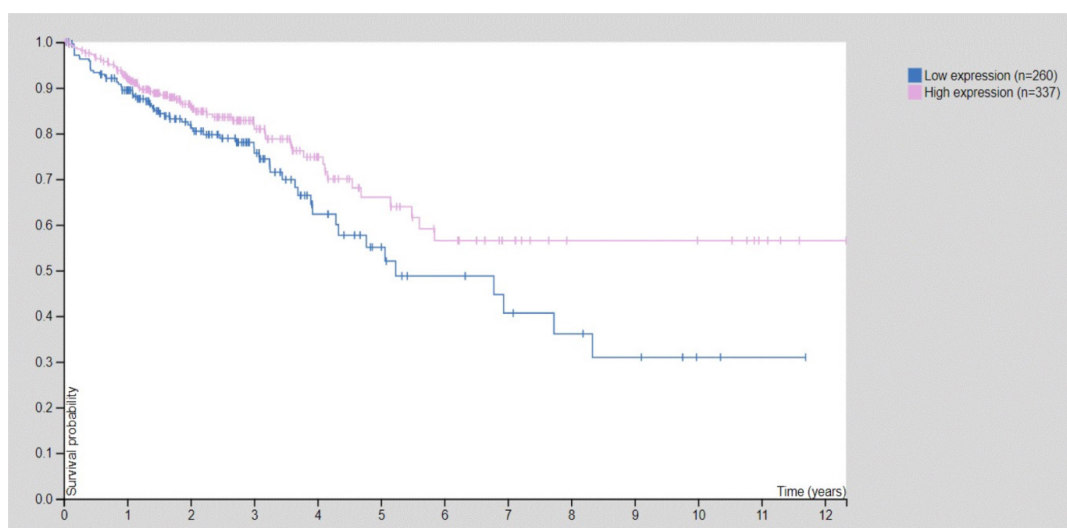

Figure 8. Survival Probability of APC Expression in Colon Cancer Cases (The Human Protein Atlas -TCGA)
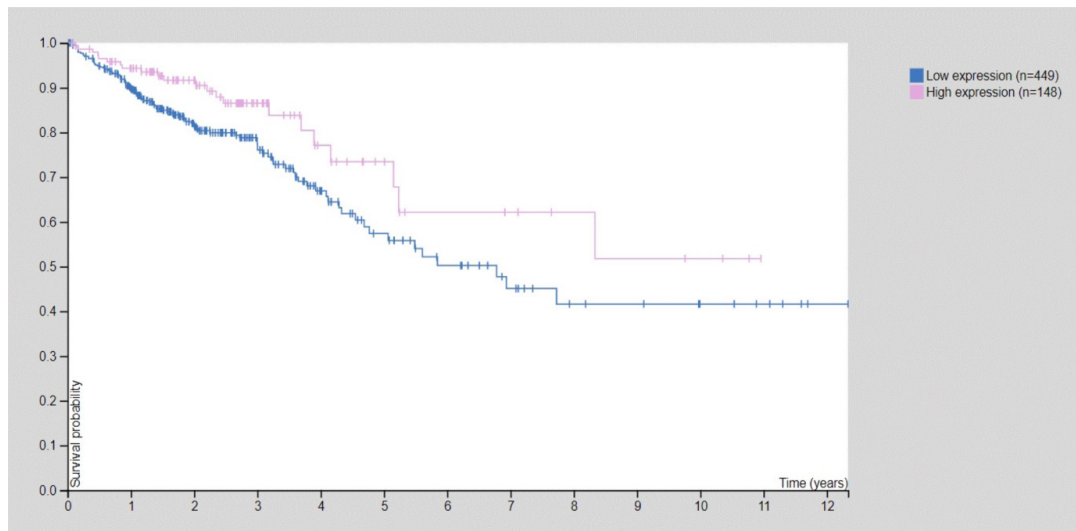
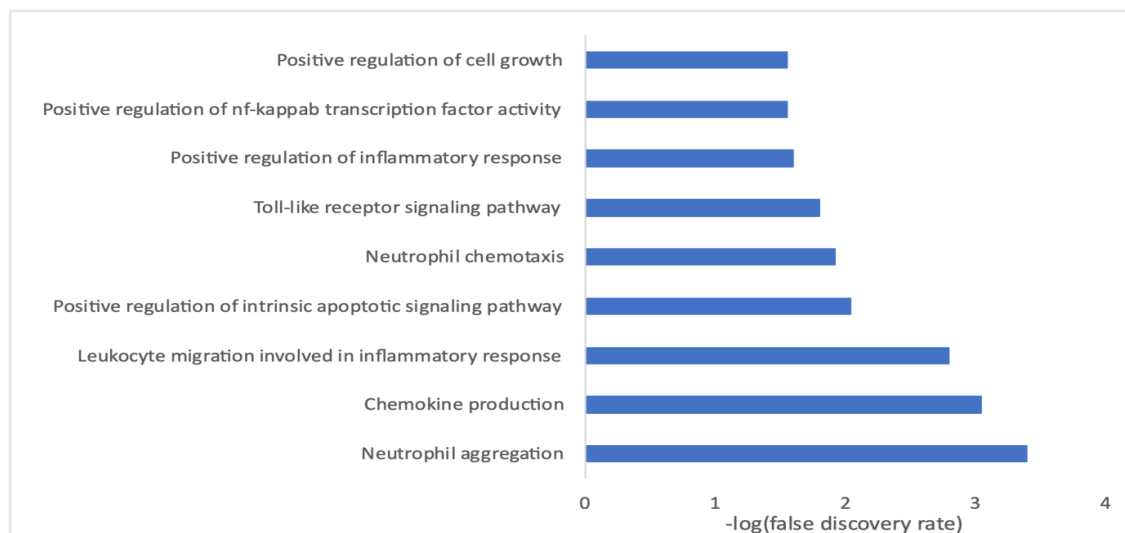Figure 9. Survival Probability of MLH1 Expression in Colon Cancer Cases (The Human Protein Atlas -TCGA)

Figure 10. Gene Ontology Biological Process S100A8 and S100A9

are related to the development of CRC, and one of the underlying molecular mechanisms is that extracellular S100A8 and S100A9 proteins contribute to the survival and migration of CRC cells through the Wnt/β-catenin pathway (Duan et al., 2013). In addition, S100A8 and S100A9 play a role in inflammatory responses and apoptosis (Figure 10).

In conclusion, late Salmonella AvrA infection (four days) affects the expression of genes involved in inflammation. Expression of MLH1, MSH2, EPCAM, APC, PMS2 in the case of colon adenocarcinoma tumor showed a correlation with genes that undergo changes due to Salmonella AvrA infection. In-Silico's analysis with gene ontology and pathway approaches suggested that genes affected by AvrA expression play a role in inflammation. Among gene interest, EPCAM had the highest correlation compared to other gene interests (Table 1). There are 514 genes that have a correlation with cases of AvrA infection. TNF is upregulated in AvrA infection and correlates negatively with EPCAM and it has the highest BC value compared to other nodes.

Survival probability (Figure 5) showed that EPCAM has a High-Expression and it can increase survival probability. In addition to TNF, IL1B, S100A8, S100A9 are genes related to inflammation and correlated with gene interest (Table 5).

Author Contribution Statement

All authors contributed equally in this study.

Acknowledgements

References

Bernard WS, Christopher PW, Editors (2014). World Cancer Report 2014. In International Agency for Research on Cancer.

Chumduri C, Gurumurthy RK, Zietlow R, Meyer TF (2016). Subversion of host genome integrity by bacterial pathogens. *Nat Rev Mol Cell Biol*, **17**, 659–73.

Compare D, Nardone G (2014). The bacteria-hypothesis of colorectal cancer : pathogenetic and therapeutic implications. *Transl Gastrointest Cancer*, **3**, 44–53.

Duan L, Wu R, Ye L, et al (2013). S100A8 and S100A9 Are Associated with Colorectal Carcinoma Progression and Contribute to Colorectal Carcinoma Cell Survival and Migration via Wnt/β-Catenin Pathway. *PLoS One*, **8**, 1–13.

Gagnaire A, Nadel B, Raoult D, Neefjes J, Gorvel JP (2017). Collateral damage: Insights into bacterial mechanisms that predispose host cells to cancer. *Nat Rev Microbiol*, **15**, 109–28.

Hosono K, Yamada E, Endo H, et al (2012). Increased tumor necrosis factor receptor 1 expression in human colorectal adenomas. *World J Gastroenterol*, **18**, 5360–8.

Jones TF, Ingram LA, Cieslak PR, et al (2008). Salmonellosis outcomes differ substantially by serotype. *J Infect Dis*, **198**, 109–14.

Kobelt D, Zhang C, Clayton-Lucey IA, et al (2020). Pro-inflammatory TNF-α and IFN-γ Promote Tumor Growth and Metastasis via Induction of MACC1. *Front Immunol*, **11**, 1–15.

Liu X, Lu R, Xia Y, Sun J (2010). Global analysis of the eukaryotic pathways and networks regulated by Salmonella typhimurium in mouse intestinal infection in vivo. *BMC Genomics*, **11**.

Lu R, Wu S, Zhang YG, et al (2014). Enteric bacterial protein AvrA promotes colonic tumorigenesis and activates colonic beta-catenin signaling pathway. *Oncogenesis*, **3**, 1–10.

Lu R, Bosland M, Xia Y, et al (2017). Presence of Salmonella AvrA in colorectal tumor and its precursor lesions in mouse intestine and human specimens. *Oncotarget*, **8**, 55104–15.

Lu R, Wu S, Zhang YG, et al (2016). Salmonella Protein AvrA Activates the STAT3 Signaling Pathway in Colon Cancer. *Neoplasia*, **18**, 307–16.

Lutgens MWMD, Van Oijen MGH, Van der Heijden GJMG, et al (2013). Declining risk of colorectal cancer in inflammatory bowel disease: an updated meta-analysis of population-based cohort studies. *Inflamm Bowel Dis*, **19**, 789–99.

Majowicz SE, Musto J, Scallan E, et al (2010). The global burden of nontyphoidal salmonella gastroenteritis. *Clin Infect Dis*, **50**, 882–9.

Mokhtari M, Zakerzade Z (2017). EPCAM Expression in Colon Adenocarcinoma and its Relationship with TNM Staging. *Adv Biomed Res*, **6**, 56–6.

Mughini-Gras L, Schaapveld M, Kramers J, et al (2018). Increased colon cancer risk after severe Salmonella infection. *PLoS One*, **13**, 1–19.

Radhakrishnan H, Walther W, Zincke F, et al (2018). MACC1-the first decade of a key metastasis molecule from gene discovery to clinical translation. *Cancer Metastasis Rev*, **37**, 805–20.

Schmitt B, Münz M, Gires O, et al (2001). Tumor Necrosis Factor alpha Negatively Regulates the Expression of the Carcinoma-Associated Antigen Epithelial Cell Adhesion Molecule. *Cancer*, **92**, 620–8.

Yang J, Wen Z, Li W, et al (2021). Immune Microenvironment: New Insight for Familial Adenomatous Polyposis. *Front Oncol*, **11**, 1–11.