Determination of Eight Hub Genes and Functional Pathways Affecting both the Survival of Early- and Late-Stage Colon Cancer Patients

Gizem AYNA DURAN,1 Fatma SERT2

1Department of Biomedical Engineering, İzmir University of Economics Faculty of Engineering, İzmir-Türkiye
2Department of Radiation Oncology, Ege University Faculty of Medicine, İzmir-Türkiye

OBJECTIVE
The stage of colon cancer (CC) and therefore the level at which the treatment is initiated affects the survival of CC patients. In our study, we aimed to identify the common survival-related genes in both early- and late-stage CC patients.

METHODS
Information on the demographic characteristics of 581 patients and microarray expression profiles (GSE39582) were obtained from the gene expression omnibus database. Survival analysis was performed using univariate and multivariate Cox regression methods with the help of R3.53 programming language and Kaplan–Meier graphics through the R software “Survival” package. ShinyGO v0.741 gene ontology enrichment analysis was performed to clarify the common and functional pathways related to both early- and late-stage CC cancer patients’ data.

RESULTS
Cox regression analysis indicated that overall survival and relapse-free survival of CC patients were strongly influenced by stage. Genes that significantly affect prognosis and survival in early- and late-stage CC patients were identified. As a result of gene enrichment analysis, arginine binding, oxidoreductase activity, and methylcytosine dioxygenase activity and related eight hub genes (TM4SF5, NOS3, Ten eleven translocation [TET1], TET3, JMJD7, AKR1C1, prenylcysteine oxidase 1 like, Methionine sulfoxide reductase A) were identified.

CONCLUSION
According to our results, it might be considered that developing new treatment strategies based on eight hub-genes related to arginine binding, oxidoreductase activity, and methylcytosine dioxygenase activity detected at different stages of CC might increase the success of targeted therapies.

Keywords: Arginine binding; colon cancer stage; methylcytosine dioxygenase activity; oxidoreductase activity; survival.

INTRODUCTION
Colon cancer (CC) is the second most lethal cancer and the fourth-most diagnosed cancer type. In line with GLOBOCAN data, the incidence of both colon and rectal cancer is increasing, and the increase is seen, especially in developing countries trying to adapt to the western way of life. Environmental factors such as eat-
ing habits, obesity, lack of physical activity, cigarette/alcohol consumption, malnutrition, and excessive consumption of some vitamins and minerals, some diseases (diabetes and insulin resistance) can cause CC. In addition, 7–10% of CC can be passed from generation to generation through inherited mutations. Approximately, 30% of patients tend to develop a familial neoplasm. For example, the probability of being diagnosed with CC maybe 2–4 times higher in individuals whose first-degree relatives have been diagnosed with CC.[1–3] Some hereditary syndromes such as inflammatory bowel disease, familial adenomatous polyposis, and hereditary non-polyposis colorectal cancer syndrome can higher the risk of CC development.[1] CC is one of the most interesting cancers which includes a lot of questions that need to be solved in terms of both the effect of the genetic background of the disease and treatment options according to new targets. For instance, one of the most important reasons for the decrease in treatment success and shortening of survival in CC is whether cancer is diagnosed in the early stages or not. The stage of cancer diagnosis and therefore the level at which treatment is initiated affects survival. There are studies investigating how the survival of early and advanced CC patients varies in different populations (For example, USA and Kuwait patient populations).[4] In addition, there are studies comparing the survival of early-stage and late-stage CC patients based on other demographic characteristics.[3,5,6] The number of studies in which gene screening is performed according to the stage of cancer based on the gene densities of early- and late-stage CC patients is also quite limited.[7]

In light of this information, we divided patients diagnosed with CC into two groups as early- and late-stage cancer patients, and the genes that significantly affected patient survival were separately identified in both groups. By doing this, we aimed to further evaluate genetic factors underlying the treatment responses and outcomes for the patients with CC. In our study, not only the differences in genes that affect survival in early- and late-stage CC s were determined but also whether the detected genes were hazardous, or survival triggering was also shown. Thus, genes that differ according to the stage in CC patients were identified and brought to literature for use in treatment approaches.

**MATERIALS AND METHODS**

**Data Acquisition and Data Processing**

About 581 CC patients (January 2012 and December 2019) were included in this study. Information on the demographic characteristics and microarray expression profiles were obtained from the Gene Expression Omnibus (GEO) database (GSE39582). Gene intensity values of CC patients’ genes (MAS5 log2 normalized intensity values of the genes) were used from [HG-U133_Plus_2] Affymetrix Human Genome U133 plus 2.0 Array (GPL570) platform with 54,697 probes (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39582). Furthermore, gene enrichment analysis was carried out to clarify the common and functional pathways related to both early- and late-stage CC cancer patients’ data. For this purpose, ShinyGO v0.741: Gene Ontology Enrichment Analysis was performed on September 23, 2022 (http://bioinformatics.sdstate.edu/go/). Totally 2579 cancer-related genes were used as background genes (http://www.bushmanlab.org/links/genelists).

**Statistical Analysis**

Descriptive statistics of variables such as mean and standard deviation were provided. In this study, survival analysis was performed using univariate and multivariate Cox regression method with the help of R3.53 programming language and Kaplan–Meier graphics through R software “Survival” package (version 3.5.1; https://CRAN.R-project.org/package=survival). The patients were divided into two groups AJCC stage 1,2 (early stage) and AJCC stage 3,4 (late stage). R survival package was used to perform multivariate Cox regression analysis in early- and late-stage CC patient groups. Kaplan–Meier and Cox regression analysis were performed. P<0.05 was accepted as a statistical significance.

**Ethics Committee Approval**

A publicly available dataset of CC patients was used. Therefore, ethics committee approval is not required.

**RESULTS**

**Demographic, Clinical, and Pathological Findings**

The characteristics of patients, diseases, and tumors were detailed in Table 1. The median age of the patients was 68.55±13.18 (22–97 age range). There were 290 patients below the median age and 290 patients above the median age. Women gender was 54.91% of the included patients. The number and rate of the patients AJCC stage 1, 2, 3, and 4 were 38 (6.56%), 271 (46.80%), 210 (36.26%), and 60 (10.36%), respectively. The distribution of CC patients according to TNM classification, tumor location, and mutation status could be seen in Table 1. Whereas the number of adjuvant therapy receiving patients was 240 (42.7%), the number of pa-
patients not receiving adjuvant therapy was 322 (57.3%). About 18.83% of the patients in stages 1 and 2 received adjuvant therapy while 71.65% of the patients in stages 3 and 4 received adjuvant therapy (Table 2).

**Results of Kaplan–Meier and Multivariate Cox Regression Analysis in CC Patients**

According to our results, after a median 53- and 45-month follow-up time, oxide synthase (OS) and relapse-free survival (RFS) times for the whole cohort were 58.23 and 49.89 months, respectively. For the patients with early-stage CC, OS was 64.49 months and RFS was 58.24 months. On the other hand, for the patients with late-stage CC, OS was 51.05 months and RFS was 40.32 months. The survival rate for the overall patients was 67.02% (OS), and the RFS was 68.60%. In patients diagnosed with early-stage CC (stages 1 and 2), the survival rate was 72.70% (OS) and RFS was 79.61%. In patients diagnosed with late-stage CC (stages 3 and 4), the survival rate was 60.53% (OS) and RFS was 56.02%.

In the Cox regression (both univariate and multivariate) analysis, it was seen that the OS was significantly affected by stage, age, t, and chemotherapy, while only the stage and t variables significantly affected RFS. It was found that the stage variable significantly and strongly affects both OS and RFS. Besides, the hazard ratios of the stage variable were found to be significantly above 1. More importantly, the hazard ratio of the stage variable is strongly significant in all situations, regardless of the dependent variable (RFS or OS). According to these results, it is shown that having early- or late-stage CC is the most effective variable on survival in the diagnosis of patients (Table 2, Figs. 1, 2).

**Results of Kaplan–Meier multivariate Cox Regression Analysis by Stage in CC Patients**

In our study, genes that significantly affect prognosis and survival in early- and late-stage CC patients (n=581) were identified. First, multivariate Cox regression analysis was applied, in which density values of 54,697 genes and other independent variables were included. Whereas 4056 genes were found to be statistically significant for early-stage patients, and 3499 genes were found to be statistically significant for late-stage patients (p<0.05). For both stage patient groups, the total number of significant genes with a p=1% was found to be 74. Among these common genes, the hazard ratio of the highest 20 and the lowest 20 genes can be seen in Table 3.
Results of Functional Gene Ontology (GO) Enrichment Analysis

According to ShinyGO v0.741: Gene Ontology Enrichment Analysis results, arginine binding, oxidoreductase activity, and methylcytosine dioxygenase activity and related eight hub genes (TM4SF5, NOS3, Ten eleven translocation \([TET1]\), \([TET3]\), \(JMJD7\), \(AKR1C1\), prenylcysteine oxidase 1 like \([PCYOX1L]\), Methionine sulfoxide reductase A \([MSRA]\)) were identified both early-stage and late-stage CC patients (Fig. 3 and Table 4).

DISCUSSION

In the present study, 581 CC patient’s gene intensity data obtained from the NCBI GEO database was used. Since overall and RFS in all CC patients was strongly influenced by stage, we aimed to identify the common survival-related genes in both early- and late-stage CC patients. According to analysis, common 74 genes were identified as hazardous and they were related to common functional biological path-
Ayna Duran and Sert.
Genes Affecting Survival Regardless of Stage in Colon Cancer

ways such as arginine binding, oxidoreductase activity, and methylcytosine dioxygenase activity.

In our study, one of the targeted biological processes is the arginine binding pathway. Arginine is one of the natural amino acids, and tumor cells are auxotrophic for arginine. For this reason, arginine-depriving therapy can lead to tumor cell proliferation and inhibit cell death pathways using arginine deiminase enzyme.

Table 3 Top 20 most hazardous and survival-triggering genes affecting the survival of both early and late-stage CC patients

<table>
<thead>
<tr>
<th>Gene</th>
<th>Hazard ratio</th>
<th>p</th>
<th>Gene</th>
<th>Hazard ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR1</td>
<td>24.017287</td>
<td>0.003405</td>
<td>MAPRE2</td>
<td>0.049295</td>
<td>0.00477</td>
</tr>
<tr>
<td>PPP2R3A</td>
<td>17.0081124</td>
<td>0.0087</td>
<td>KLK4</td>
<td>0.15981</td>
<td>0.000977</td>
</tr>
<tr>
<td>1560743_a_at</td>
<td>15.2527398</td>
<td>0.003268</td>
<td>LINC00313</td>
<td>0.165643</td>
<td>0.00391</td>
</tr>
<tr>
<td>CDHB</td>
<td>13.576319</td>
<td>0.001595</td>
<td>DOCK1</td>
<td>0.16797</td>
<td>0.009341</td>
</tr>
<tr>
<td>CCDC38</td>
<td>8.788534</td>
<td>0.00944</td>
<td>NPTXR</td>
<td>0.19395</td>
<td>0.00204</td>
</tr>
<tr>
<td>TMEM74</td>
<td>7.334637</td>
<td>0.00981</td>
<td>BCAT2</td>
<td>0.215049</td>
<td>0.000995</td>
</tr>
<tr>
<td>PCDH15</td>
<td>6.525729</td>
<td>0.000537</td>
<td>POLH</td>
<td>0.21771</td>
<td>0.00000222</td>
</tr>
<tr>
<td>LPA</td>
<td>6.084885</td>
<td>0.00436</td>
<td>FXN</td>
<td>0.228755</td>
<td>0.000856</td>
</tr>
<tr>
<td>HAPLN4</td>
<td>5.347975</td>
<td>0.00954</td>
<td>CCDC174</td>
<td>0.300549</td>
<td>0.00961</td>
</tr>
<tr>
<td>SDC1</td>
<td>4.946807</td>
<td>0.00403</td>
<td>GFI1</td>
<td>0.3427379</td>
<td>0.00872</td>
</tr>
<tr>
<td>CYLD</td>
<td>4.516423</td>
<td>0.0095</td>
<td>TRIM11</td>
<td>0.36504</td>
<td>0.00723</td>
</tr>
<tr>
<td>ZSCAN30</td>
<td>4.475585</td>
<td>0.0062</td>
<td>ZNF562</td>
<td>0.378879</td>
<td>0.00114</td>
</tr>
<tr>
<td>LINC00310</td>
<td>4.293516</td>
<td>0.006504</td>
<td>1566003_x_at</td>
<td>0.394778</td>
<td>0.00761</td>
</tr>
<tr>
<td>SLC14A2-AS1</td>
<td>4.176029</td>
<td>0.00981</td>
<td>DOK2</td>
<td>0.4120158</td>
<td>0.00682</td>
</tr>
<tr>
<td>1563414_at</td>
<td>3.773802</td>
<td>0.00538</td>
<td>WTAP</td>
<td>0.43471</td>
<td>0.001212</td>
</tr>
<tr>
<td>DOCK9</td>
<td>3.304124</td>
<td>0.00928</td>
<td>ZNF443</td>
<td>0.44579</td>
<td>0.0014</td>
</tr>
<tr>
<td>214862_x_at</td>
<td>3.284441</td>
<td>0.00758</td>
<td>TCF3</td>
<td>0.474085</td>
<td>0.00314</td>
</tr>
<tr>
<td>PAX8</td>
<td>3.24541</td>
<td>0.00276</td>
<td>THAP12</td>
<td>0.487329</td>
<td>0.00485</td>
</tr>
<tr>
<td>ATG28</td>
<td>3.192219</td>
<td>0.002788</td>
<td>RAB15</td>
<td>0.49615</td>
<td>0.00105</td>
</tr>
<tr>
<td>GDI1</td>
<td>3.039679</td>
<td>0.00171</td>
<td>RAB15</td>
<td>0.496569</td>
<td>0.00105</td>
</tr>
</tbody>
</table>

Top 20 most influential common genes were sorted with respect to hazard ratio and RFS was accepted as the dependent variable. Genes left in the form of probe names were not found and were left as such.

Fig. 3. Gene ontology enrichment analysis results of common hazardous genes of both early- and late-stage colon cancer patients (function). FDR: False discovery rate.
Both, cell line-based studies as well as clinical Phase I, II, and II trials have shown the effect of arginine deprivation therapy in various cancers mostly such as hepatocellular carcinoma (HCC), breast cancer, melanoma, lymphoma, lung cancer, and less commonly for CC.[8–10] These studies are based on enzyme therapy and gene therapy-based studies specific to the arginine binding pathway have not yet been broadly performed. According to our studies, transmembrane 4 superfamily member 5 (TM4SF5), endothelial NOS, and eNOS (NOS3) genes were related to arginine binding pathway which can be possible anti-cancer therapy for CC patients commonly for both early- and late-stage (Table 4). In the literature, it has been shown that TM4SF5 gene is overexpressed in pancreatic cancers when compared to normal tissue which makes this gene as a candidate anti-cancer target gene.[11] Furthermore, TM4SF5 gene is also shown as aberrantly expressed in hepatocellular and CCs and chimeric monoclonal antibodies targeting TM4SF5 gene may be used as a personalized anti-cancer agent.[12,13] CC-specific antibody and cancer vaccine studies for this gene have been limited to mice, and our knowledge, there is no clinical trial-level study. Therefore, its importance in CC is worth investigating and promising.[14,15] Furthermore, it has been shown that NOS3 gene overexpression, which is known as the endothelial form of nitric oxide synthases (NOS), is also related to certain cancers such as pancreatic, stomach adenocarcinoma, shown in mouse and patient models.[16,17] Like the TM4SF5 gene, the number of CC-based studies for the NOS3 gene is very limited.[18]

<table>
<thead>
<tr>
<th>Enrichment FDR</th>
<th>nGenes</th>
<th>Pathway genes</th>
<th>Fold enrichment</th>
<th>Pathway</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.014534</td>
<td>7</td>
<td>835</td>
<td>5.454671</td>
<td>Oxidoreductase activity</td>
<td>TET1, PCYOX1L, NOS3, AKR1C1, TET3, JMJJD7</td>
</tr>
<tr>
<td>0.010448</td>
<td>5</td>
<td>184</td>
<td>11.419878</td>
<td>Oxidoreductase activity, acting on paired donors, with incorporation or reductio</td>
<td>TET1, NOS3, AKR1C1, TET3, JMJJD7</td>
</tr>
<tr>
<td>0.025913</td>
<td>3</td>
<td>54</td>
<td>16.558823</td>
<td>2-oxoglutarate-dependent dioxygenase activity</td>
<td>TET1, TET3, JMJJD7</td>
</tr>
<tr>
<td>0.043353</td>
<td>3</td>
<td>121</td>
<td>12.419117</td>
<td>Monoxygenase activity</td>
<td>NOS3, AKR1C1, JMJJD7</td>
</tr>
<tr>
<td>0.014534</td>
<td>2</td>
<td>10</td>
<td>66.235294</td>
<td>Arginine binding</td>
<td>TM4SF5, NOS3</td>
</tr>
<tr>
<td>0.0259136</td>
<td>2</td>
<td>4</td>
<td>44.156862</td>
<td>Methylcytosine dioxygenase activity</td>
<td>TET1, TET3</td>
</tr>
<tr>
<td>0.042780</td>
<td>2</td>
<td>61</td>
<td>33.117647</td>
<td>Amino acid binding</td>
<td>TM4SF5, NOS3</td>
</tr>
</tbody>
</table>

FDR: False discovery rate; TET1: Ten eleven translocation-1; TET3: Ten eleven translocation-3; PCYOX1L: Prenylcysteine oxidase 1 like; NOS3 and eNOS: Endothelial nitric oxide synthase; AKR1C1: Human 20-hydroxysteroid dehydrogenase; JmjC and JMJJD7: Jumonji-C domain-containing 7; MSRA: Methionine sulfoxide reductase A; TM4SF5: Transmembrane 4 superfamily member 5

Zhang et al.[19] have shown that certain differentially expressed genes are enriched in oxidoreductase activity and they also related these DEGs to HCC initiation and progression. On the other hand, glutathione-dependent oxidoreductase activity has been indicated to have a role in invasiveness in the MCF-7 breast cancer cell line.[20] Inhibiting oxidoreductase activity may lead to a decrease in the apoptosis activity of colorectal cancer cells collected from clinical samples.[21]

DNA methylation/demethylation patterns have been known to be critical in cancer development. TET enzymes are known as tumor suppressor enzymes. TET1 is determined as the first methyl-cytosine dioxygenase enzyme (5-methylcytosine, [5mC]) and TET2 and TET3 family members were later characterized. TET family genes (including TET1, TET2, and TET3) are genes that catalyze 5mC oxidation and encode proteins that can modulate DNA methylation due to this oxidation.[22] Furthermore, TET enzymes are shown to have roles in the carcinogenesis of various cancers such as hematological cancers, breast, melanoma, lung, and thyroid cancers through inducing DNA demethylation.[23,24] Among TET enzymes, TET2 is the one which is more frequently associated with cancers.[25] However, in our study, TET1/TET3 genes are determined as two of the key common genes in both early- and late-stage of CC patients. In literature, it has been indicated that TET1 is downregulated through the activation of NF-kB in basal-like breast cancer as well as lung, melanoma, and thyroid cancers.[26] Furthermore, the epigenetic
inactivation of TET3 has been associated with head-and-neck squamous cell carcinomas.[27]

The human 20-hydroxysteroid dehydrogenase (AKR1C1) gene and its coded protein is one of the members of the aldo-keto reductase superfamily and its overexpression has been related to the drug resistance in several cancers such as cervix, lung, uterine, skin, ovary, and CCs.[28,29]

On the other hand, another gene which is named as Jumonji-C (jmjC) domain-containing 7 (JMJD7) is known as a 2-oxoglutarate (2OG)-dependent oxygenase. This gene has been shown to have roles in survival, apoptosis, and cell proliferation in cancer cells such as head/neck squamous cell carcinoma, breast and prostate cancers, and its inhibitors are shown to be potent in preventing the carcinogenesis in certain cancer cell lines such as breast cancer, lymphocytic leukemia, cervical carcinoma, and neuroblastoma cell lines [30–33].

MSRA gene downregulation has been related to increased aggressiveness in breast cancers and it has been shown that it can lead to suppress of metastasis in hepatocellular cancers.[34,35] Furthermore, it can also be accepted as a candidate biomarker for endometrial cancers.[36] On the other hand, PCYOX1L gene is indicated to be upregulated in renal cell carcinoma tissues in accordance with normal ones and this gene is associated with the metastatic state and placental invasion in melanoma cancer patients.[37] To our knowledge, unlike some other cancers mentioned above, there is no published article on the expression level of the MSRA and PCYOX1L gene in CCs. In this context, our study includes novel findings.

The number of studies focused on these eight hub genes (TM4SF5, NOS3, TET1, TET3, JMJD7, AKR1C1, PCYOX1L, MSRA) is very limited in the case of CC. In summary, the treatment options in CC can be optimized regardless of the stage through targeting these pathways and hub genes.

CONCLUSION

In the study, it was concluded that gene expression varying according to the stage of cancer in CC significantly affects the survival of CC patients. Based on these results, it should be considered that developing new targeted therapies based on eight hub genes related to arginine binding, oxidoreductase activity, and methylcytosine dioxygenase activity detected at different stages of CC may increase the success of treatment.

Peer-review: Externally peer-reviewed.

Conflict of Interest: All authors declared no conflict of interest.

Financial Support: None declared.


REFERENCES


