Research Article

Bioinformatics Analysis Reveals Most Prominent Gene Candidates to Distinguish Colorectal Adenoma from Adenocarcinoma

Nina Hauptman ⓘ**, Emanuela Boštjančič** ⓘ**, Margareta Žlajpah,**
Branislava Ranković, and Nina Zidar

Institute of Pathology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

Correspondence should be addressed to Nina Hauptman; nina.hauptman@mf.uni-lj.si

Colorectal cancer (CRC) is one of the leading causes of death by cancer worldwide. Bowel cancer screening programs enable us to detect early lesions and improve the prognosis of patients with CRC. However, they also generate a significant number of problematic polyps, e.g., adenomas with epithelial misplacement (pseudoinvasion) which can mimic early adenocarcinoma. Therefore, biomarkers that would enable us to distinguish between adenoma with epithelial misplacement (pseudoinvasion) and adenoma with early adenocarcinomas (true invasion) are needed. We hypothesized that the former are genetically similar to adenoma and the latter to adenocarcinoma and we used bioinformatics approach to search for candidate genes that might be potentially used to distinguish between the two lesions. We used publicly available data from Gene Expression Omnibus database and we analyzed gene expression profiles of 252 samples of normal mucosa, colorectal adenoma, and carcinoma. In total, we analyzed 122 colorectal adenomas, 59 colorectal carcinomas, and 62 normal mucosa samples. We have identified 16 genes with differential expression in carcinoma compared to adenoma: *COL12A1, COL1A2, COL3A1, DCN, PLAU, SPARC, SPON2, SPP1, SULF1, FADS1, G0S2, EPHA4, KIAA1324, L1TD1, PCKS1,* and *C11orf96*. In conclusion, our *in silico* analysis revealed 16 candidate genes with different expression patterns in adenoma compared to carcinoma, which might be used to discriminate between these two lesions.

1. Introduction

Colorectal cancer (CRC) is developed by multistep process from normal epithelium to adenoma and adenocarcinoma, which can eventually metastasize to different organs [1]. The model of development of CRC was introduced in 1990, where *APC, KRAS, TP53,* and *DCC* were proposed as genes promoting the progression of CRC [2]. Since, many studies have investigated underlying molecular mechanisms of CRC. It is accepted that CRC arises from accumulation of genetic and epigenetic events that alter signaling in pathways, such as Wnt, PIK3CA, and TGF-β. Three major accepted pathways in the pathogenesis of CRC are chromosome instability pathway, microsatellite instability pathway, and CpG island methylator phenotype. There are many CRCs that lack the changes described in above pathways, suggesting that other mechanisms are involved in the development of CRC [1].

CRC is one of the leading causes of death by cancer worldwide. In Europe, CRC is the second and the third cause of death by cancer in men and women, respectively [3]. Five-year survival for patients with early CRC is 90%, while for patients with advanced CRC, survival drops to only 8-12% [4]. The prognosis can improve significantly with the introduction of population screening. Bowel cancer screening programs enable us to detect early lesions, including adenomas and adenomas with early adenocarcinoma (malignant polyps). However, they also generate a significant number of problematic polyps which contain dysplastic glands in the submucosa. This phenomenon has been referred to as epithelial misplacement (pseudoinvasion). It can be the result of a torsion or intraluminal trauma of large pedunculated polyps of the distal colon, or it may be a consequence of a previous biopsy. Adenomas with epithelial misplacement (pseudoinvasion) can be difficult to distinguish from adenomas

with early adenocarcinoma [5–7]. The correct diagnosis is crucial for the choice of optimal treatment. For adenoma and adenoma with epithelial misplacement, endoscopic removal is sufficient, whereas malignant adenomas (early carcinomas) may require surgical treatment, since they are capable of metastasizing [7].

Despite well-defined morphologic features of epithelial misplacement and early invasion, there are a significant number of lesions with ambiguous features leading to divergent diagnostic opinions among pathologists [7]. Biomarkers that would enable to distinguish between adenoma with epithelial misplacement (pseudoinvasion) and adenoma with early adenocarcinoma (true invasion) are needed. We hypothesized that the former is genetically similar to adenoma and the latter to adenocarcinoma and we used bioinformatics approach to search for candidate genes that might be potentially used to distinguish between the two lesions.

Gene expression in CRC was widely studied by microarray technique, usually comparing carcinomas to normal mucosa tissue, studying microsatellite instable CRC, or establishing CRC subtypes based on gene expression patterns [8–10]. Some of the studies have focused on the gene expression difference between colorectal adenomas and carcinomas [11–15]. The downside of these studies is limitation in number of samples. Our goal was to minimize any variabilities arising from different microarrays and procedures, to identify the genes and subsequently pathways associated with adenoma progression to carcinoma. Due to the aim of the study, we have chosen five different sets of data, containing normal, adenoma, and carcinoma samples, where two of them were not published yet.

2. Materials and Methods

2.1. Microarray Data. Several projects (GSE10714, GSE37364, GSE41657, GSE50114, and GSE50115) with gene expression profiles of colon normal, adenoma, and carcinoma samples were downloaded from the public functional genomics data repository-Gene Expression Omnibus database (GEO, http://www.ncbi.nlm.nih.gov/geo) of the National Center for Biotechnology Information (NCBI). In total, 7 CRC, 5 adenomas, and 3 normal mucosa specimens were included in GSE10714, while 27 CRC, 29 adenomas, and 38 normal mucosa specimens were included in GSE37364 (both on platform GPL570 Affymetrix Human Genome U133 Plus 2.0 array). GSE41657 was composed of 25 CRC, 51 adenomas, 12 normal mucosa samples, and GSE50114 combined with GSE50115 contained 9 CRC, 37 adenoma, and 9 normal mucosa samples (all three on platform GPL6480 Agilent Whole Human Genome Microarray 4x44K G4112F). In total, 252 samples of colonic biopsies, including 62 normal, 122 adenomas, and 59 CRC samples, were included in this study.

2.2. Data Processing. For all projects, the original data files were downloaded and further normalized in R language (https://www.r-project.org/). For projects on Affymetrix arrays (GSE10714, GSE37364) package affy was used to convert CEL files into expression data using robust multichip average function, which performs background correction and normalization in one step [16]. For projects on Agilent arrays (GSE41657, GSE50114, and GSE50115) package limma was used to perform background correction and normalization between arrays [17]. After data normalization gene filter was used to remove probes that had intensity less than 100 in more than 20% of samples in each project.

Differentially expressed genes (DEG) were identified on probe level using limma package in R for each individual project [18]. We constructed three contrast matrices (adenoma compared to normal, carcinoma compared to adenoma, and carcinoma compared to normal) for each GEO project. The cut-off conditions were set to adjusted p value < 0.05 and absolute value of log fold change (log FC) > 1.5. Every comparison (adenoma compared to normal, carcinoma compared to adenoma, and carcinoma compared to normal) was overlapped among the projects to obtain the DEGs common to all projects.

2.3. Functional Analysis and Protein-Protein Interactions Network. For functional analysis and construction of protein-protein interactions (PPI) network, the Search Tool for the Retrieval of Interacting Genes (STRING) database was employed (https://string-db.org/). PPI network analysis is one of the important tools for interpretation of molecular mechanisms in the process of carcinogenesis. STRING offers integrative tools for uncovering the biological meaning behind large sets of genes, providing besides constructing PPI networks and also functional and pathway enrichment analysis. Gene ontology (GO) analysis including biological process, molecular function, and cellular component and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were conducted for selected DEGs with STRING. The statistical significance threshold was set to p < 0.05.

In this study, we constructed PPI networks of DEGs for carcinoma compared to normal, adenoma compared to normal, and carcinoma compared to adenoma. The PPI network was constructed under the cut-off of interaction score of 0.4. Visualization of all three networks together was done in Cytoscape version 3.5.1 (http://www.cytoscape.org/).

3. Results

Data from each microarray was separately analyzed to obtain DEGs for each comparison, carcinoma compared to normal, adenoma compared to normal, and carcinoma compared to adenoma. We identified 172 genes overlapping in all projects for carcinoma compared to normal (568 in GSE10714, 845 in GSE37364, 1057 in GSE41657, and 806 in GSE50114 combind with GSE50115), 137 genes overlapping in all projects for adenoma compared to normal (530 in GSE10714, 412 in GSE37364, 927 in GSE41657, and 555 in GSE50114 combind with GSE50115), and 26 genes overlapping in all projects for carcinoma compared to adenoma (252 in GSE10714, 392 in GSE37364, 116 in GSE41657, and 348 in GSE50114 combind with GSE50115) (Figure 1). We also constructed heatmap with union of all genes differentially expressed in every individual project, to confirm that samples belong to three distinct

Carcinoma vs Normal

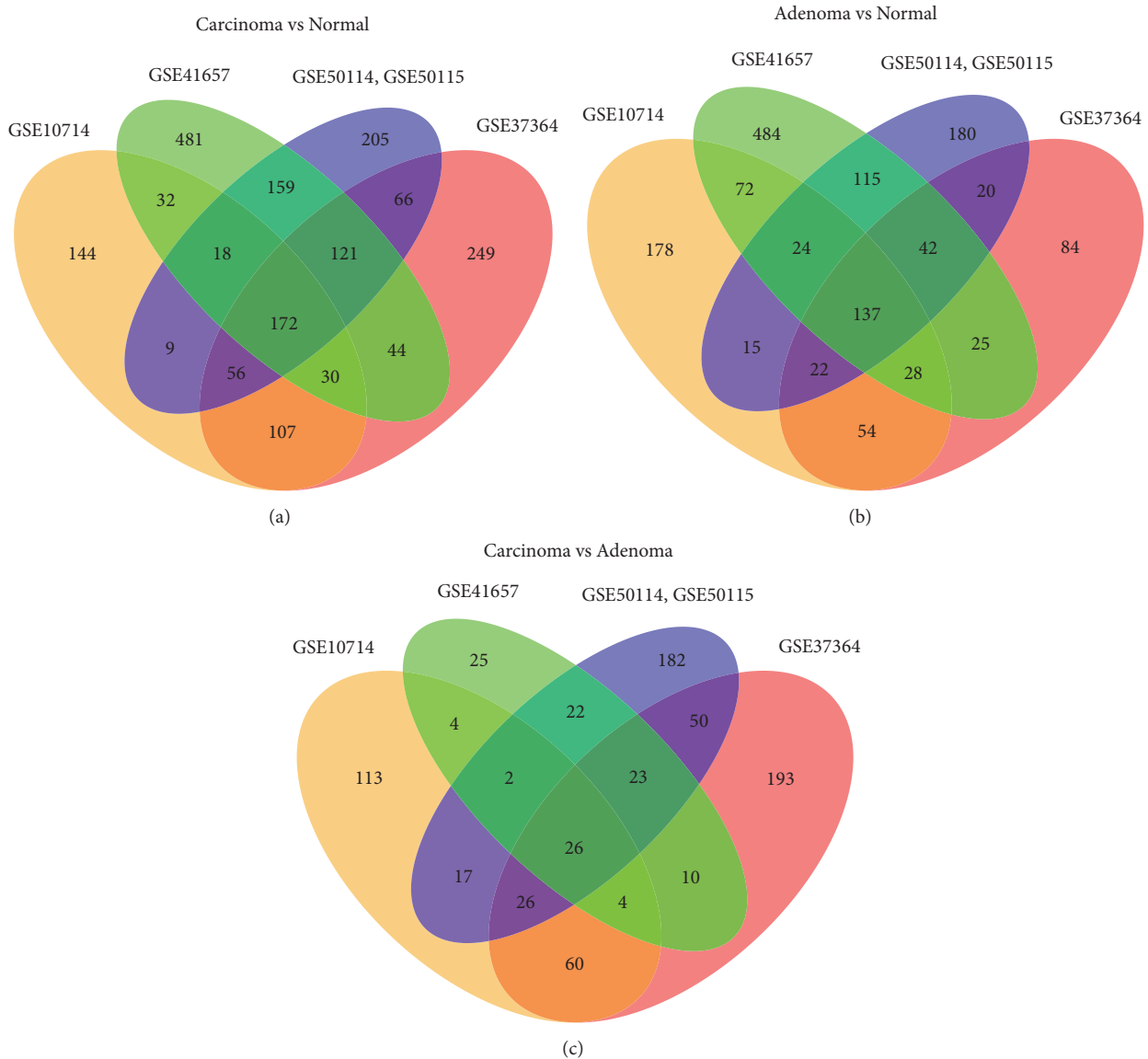(a)

Adenoma vs Normal

(b)

Carcinoma vs Adenoma

(c)

FIGURE 1: Identification of differentially expressed genes in gene expression datasets (GSE10714, GSE37364, GSE41657, GSE50114, and GSE50115) and their overlaps.

groups, namely, carcinoma, adenoma, and normal mucosa samples (Figure 2).

In order to investigate our selected DEGs, we overlapped the genes in each comparison, to obtain the unique set of genes characteristic for each comparison (Supplementary Figure 1). As expected, the most DEGs were found in carcinoma compared to normal mucosa group (172), somewhat less in adenoma compared to normal group (137), and just 26 DEGs in carcinoma compared to adenoma group (Supplementary Table 1). Interestingly, there were no DEG common to all three comparisons.

3.1. Protein-Protein Interaction Networks.

The PPI network was constructed on the basis of STRING database and visualized using Cytoscape software. Figure 3 represents network of genes differentially expressed in our analysis. In the whole network, the top hub genes are IGF1 (21), MYC (20), FN1 (14), CXCL12 (14), GCG (13), AGT (10), and BCL2 (10). The number in brackets represents the number of interaction each gene has with other genes in network.

We identified top hub genes in each group, where there are at least four connections for a gene. In adenoma compared to normal top hub genes are APOE (7), NR3C1 (4), and NMU (4), in carcinoma compared to normal top hub genes are AGT (10), BCL2 (10), AURKA (9), MMP3 (6), CDC6 (6), TPX2 (6), PRKACB (6), UB2C (5), SULT1A1 (4), KLF4 (4), ECT2 (4), and MMP1 (4), and in carcinoma compared to adenoma top hub genes are COL3A1 (6), COL1A2 (6), SPARC (5), DCN (5), and SPP1 (4).

3.2. Functional Enrichment Analysis.

The top five significant terms of GO and KEGG enrichment analysis are presented

TABLE 1: Gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of differentially expressed genes in each comparison group.

Pathway ID	Pathway description	Number of observed genes	FDR	Number of genes up/down regulated
Carcinoma vs normal				
Biological process				
GO.0001932	Regulation of protein phosphorylation	29	7.58E-05	16↑/13↓
GO.0006730	One-carbon metabolic process	7	7.58E-05	1↑/6↓
GO.0006820	Anion transport	17	0.000315	4↑/13↓
GO.0009719	Response to endogenous stimulus	30	0.000315	8↑/22↓
GO.0015701	Bicarbonate transport	6	0.000315	0↑/6↓
Cellular component				
GO.0005576	Extracellular region	63	9.38E-05	14↑/49↓
GO.0031982	Vesicle	53	0.000163	12↑/41↓
GO.0031988	Membrane-bounded vesicle	52	0.000163	12↑/40↓
GO.0044421	Extracellular region part	55	0.000163	13↑/42↓
GO.0070062	Extracellular exosome	42	0.00196	9↑/31↓
Molecular function				
GO.0004089	Carbonate dehydratase activity	5	0.0001	0↑/5↓
GO.0003824	Catalytic activity	70	0.000147	15↑/55↓
GO.0005179	Hormone activity	7	0.00231	2↑/5↓
GO.0005488	Binding	106	0.00306	35↑/71↓
GO.0008237	Metallopeptidase activity	9	0.00509	3↑/6↓
KEGG				
910	Nitrogen metabolism	5	5.28E-05	0↑/5↓
4964	Proximal tubule bicarbonate reclamation	4	0.00167	0↑/4↓
4976	Bile secretion	6	0.00167	0↑/6↓
5204	Chemical carcinogenesis	6	0.00167	0↑/6↓
4972	Pancreatic secretion	6	0.00513	0↑/6↓
Adenoma vs normal				
Biological process				
GO.0006820	Anion transport	15	5.23E-05	3↑/12↓
GO.0006730	One-carbon metabolic process	5	2.40E-03	0↑/5↓
GO.0015701	Bicarbonate transport	5	0.0024	0↑/5↓
GO.0015711	Organic anion transport	11	0.0024	3↑/8↓
GO.0006811	Ion transport	17	0.0359	3↑/14↓
Cellular component				
GO.0005576	Extracellular region	47	4.53E-06	10↑/37↓
GO.0005615	Extracellular space	23	1.15E-05	6↑/17↓
GO.0044421	Extracellular region part	41	1.15E-05	8↑/33↓
GO.0031988	Membrane-bounded vesicle	37	0.000128	6↑/31↓
GO.0098589	Membrane region	19	0.000128	5↑/14↓
Molecular function				
GO.0004089	Carbonate dehydratase activity	4	0.00115	0↑/4↓
KEGG				
910	Nitrogen metabolism	4	1.89E-04	0↑/4↓
4976	Bile secretion	6	0.000189	1↑/5↓
4964	Proximal tubule bicarbonate reclamation	4	0.000204	0↑/4↓
Carcinoma vs adenoma				
Biological process				
GO.0022617	Extracellular matrix disassembly	6	3.11E-05	6↑/0↓
GO.0030198	Extracellular matrix organization	8	3.11E-05	8↑/0↓
GO.0009888	Tissue development	11	0.0017	8↑/3↓
GO.0060279	Positive regulation of ovulation	2	0.0128	2↑/0↓
GO.0018149	Peptide cross-linking	3	0.0138	3↑/0↓

TABLE 1: Continued.

| Pathway ID	Pathway description	Number of observed genes	FDR	Number of genes up/down regulated
Cellular component				
GO.0005615	Extracellular space	14	4.03E-08	11↑/3↓
GO.0044420	Extracellular matrix component	6	5.01E-06	6↑/0↓
GO.0098644	Complex of collagen trimers	3	0.00151	3↑/0↓
GO.0005581	Collagen trimer	4	0.00166	4↑/0↓
GO.0044421	Extracellular region part	14	0.00637	10↑/4↓
Molecular function				
GO.0050840	Extracellular matrix binding	3	0.0478	3↑/0↓
KEGG				
4512	Extracellular-receptor interaction	4	0.00111	4↑/0↓
4510	Focal adhesion	4	0.0155	4↑/0↓
4974	Protein digestion and absorption	3	0.0155	3↑/0↓
5146	Amoebiasis	3	0.0219	3↑/0↓
4151	PI3K-Akt signaling pathway	4	0.0461	4↑/0↓
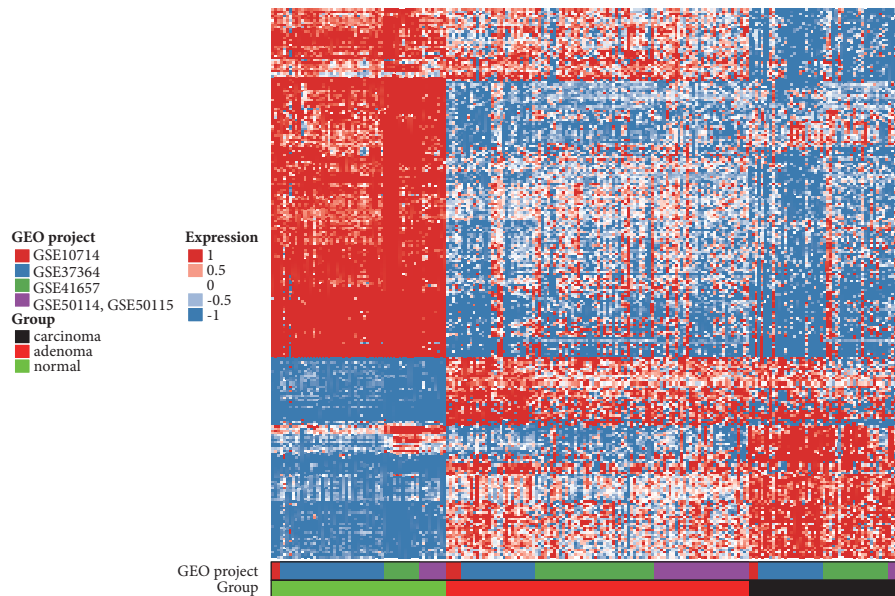
FIGURE 2: Heatmap of union of genes differentially expressed in each dataset (GSE10714, GSE37364, GSE41657, GSE50114, and GSE50115).

in Table 1, while all terms can be viewed in Supplementary Table 2. The group carcinoma compared to normal exhibits enrichment in biological process of regulation of protein phosphorylation, one-carbon metabolic process, anion transport, response to endogenous stimulus, and bicarbonate transport. As for molecular function, these genes are enriched in carbonate dehydratase activity, catalytic activity, hormone activity, binding, and metallopeptidase activity. Cellular function is enriched for genes which are included in extracellular region, vesicle, membrane-bounded vesicle, extracellular region part, and extracellular exosome. The biological processes enriched in adenoma compared to normal group were anion transport, one-carbon metabolic process, bicarbonate transport, organic anion transport, and ion transport. In this group, only one molecular function term was enriched, namely, carbonate dehydratase activity. Genes

were enriched in cellular component of extracellular region, extracellular space, extracellular region part, membrane-bounded vesicle, and membrane region. It is interesting that in both described groups of carcinoma compared to normal and adenoma compared to normal the same KEGG pathways were enriched, i.e., nitrogen metabolism, bile secretion, and proximal tubule bicarbonate reclamation. Additionally, in cancer compared to normal group two more KEGG pathways were found, namely, chemical carcinogenesis and pancreatic secretion.

3.3. Carcinoma Compared to Adenoma. The most interesting is the comparison between adenoma and carcinoma. Construction of contrast matrix enables us to compare the two groups, yet we have no information about the third group. To compare all three groups, we constructed a figure of
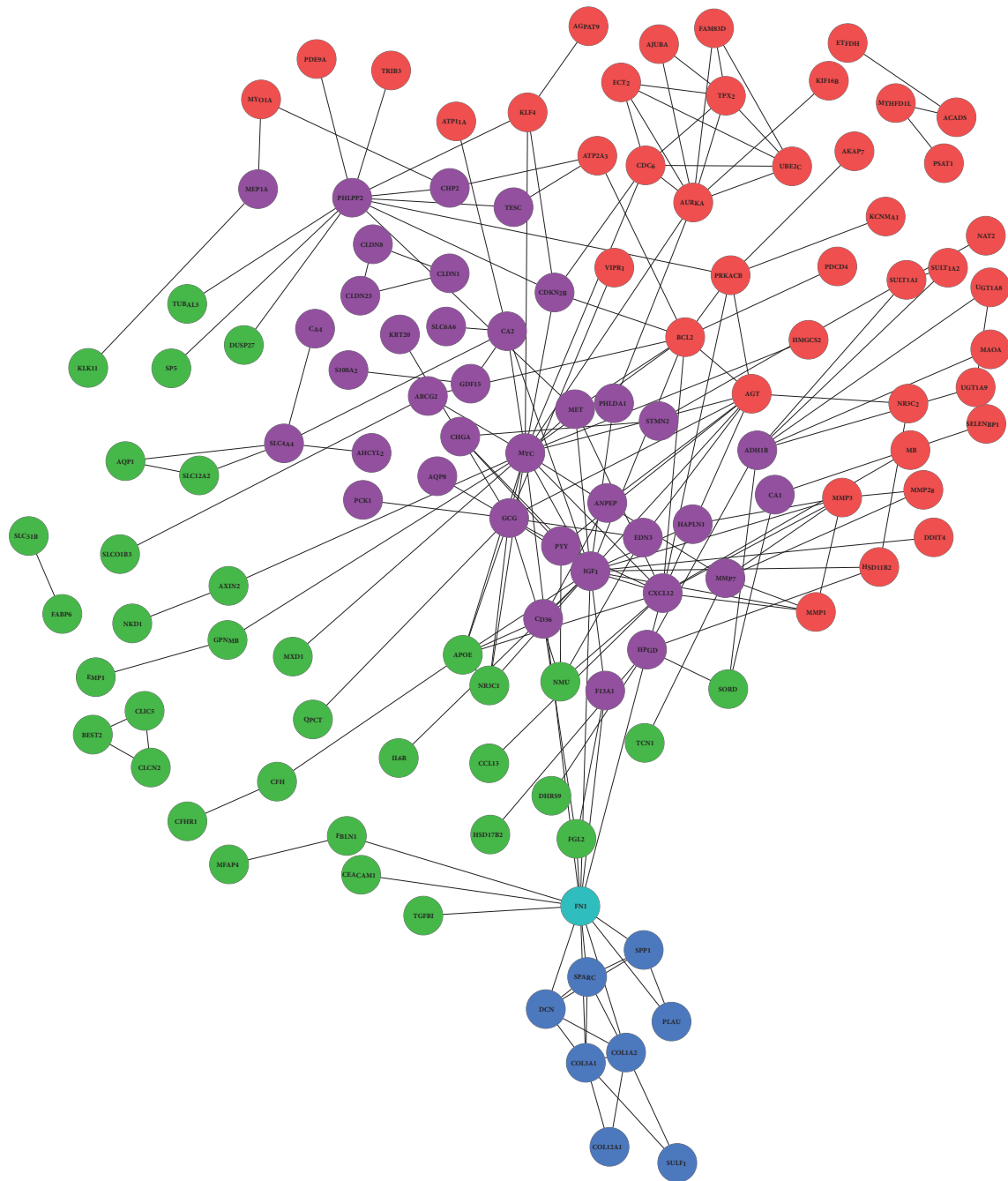
FIGURE 3: The protein-protein interactions (PPI) network of differentially expressed genes (DEGs) for each comparison. Red, green, and blue hubs represent carcinoma compared to normal, adenoma compared to normal, and carcinoma compared to adenoma, respectively. Purple hubs represent genes which are common to carcinoma compared to normal and adenoma compared to normal groups, while turquoise represents genes common to adenoma compared to normal and carcinoma compared to adenoma groups.

logarithmic average intensity values, comparing normal, adenoma, and carcinoma samples (Figure 4). The figure shows that the 16 genes unique to carcinoma compared to adenoma group are also distinguishable from average intensities of normal samples. There are four types of changes in expression. *COL12A1* follows the first pattern and has similar expression in normal and adenoma, while in carcinoma the expression is elevated. The other pattern is that expression is similar in

normal and adenoma, and reduced expression is observed in carcinoma. Genes that follow this pattern are *KIAA1324* and *PCKS1*. *EPHA4* and *L1TD1* follow the third pattern, which higher expression in adenoma and lower in normal and carcinoma. All the other genes *C11orf96*, *COL1A2*, *COL3A1*, *DCN*, *FADS1*, *G0S2*, *PLAU*, *SPARC*, *SPON2*, *SPP1*, and *SULF1* follow the fourth pattern, where expression is decreased in adenoma and increased in normal and carcinoma.
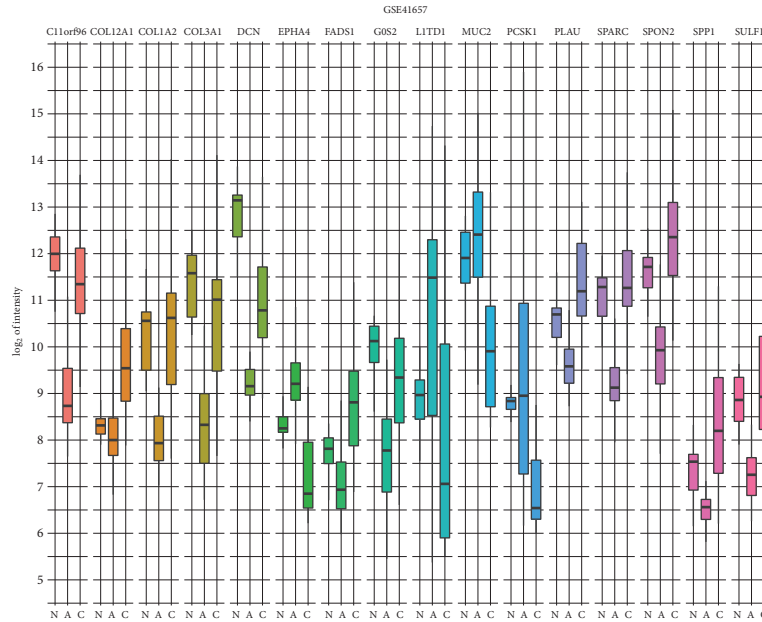
FIGURE 4: Logarithmic values of average intensities for normal (N), adenoma (A), and carcinoma (C) samples for GSE41657. Logarithmic values of average intensities for GSE37364, GSE10714, GSE50114, and GSE50115 can be found in Supplementary Figure 2.

4. Discussion

The CRC can arise through the progression of adenoma, which is the consequence of genetic and epigenetic events in epithelial cells. Some microarray studies have already identified gene expression profiles of adenoma and carcinoma [11, 19–24]. However, a study conducted by Nannini et al. revealed there is a rather weak overlap of gene expression profiles among different studies. They assigned this to several reasons: technical variability arising from collection of samples, protocols used for sample preparation, type of microarray used and subsequent data analysis pipeline used, and lack of large scale study [25]. We overcame some of these limitations by using more datasets on two different platforms, Affymetrix Human Genome U133 Plus 2.0 array and Agilent Whole Human Genome Microarray 4x44K G4112F. We used four raw datasets of microarray gene expression studies (GSE10714–Gambo et al. [19], GSE37364–Valcz et al. [26], GSE41657, GSE50114 and GSE50115–the latter three unpublished) and conducted our procedure of normalization, summation, and filtration, irrespective of procedures supplied by authors of the data.

The aim of this study was to investigate the differences in gene expression profiles of colorectal adenoma compared to adenocarcinoma, using normal mucosa samples as the reference. Our analysis showed many changes occur in adenoma compared to the normal group, suggesting that adenoma is an intermediate state between normal and carcinoma, although not all the changes found in carcinoma were found in adenoma. We identified 16 gene expression patterns unique to carcinoma compared to adenoma, suggesting that these 16 genes have a role in promoting progression of adenoma to carcinoma. Some of these genes have already been reported in

adenoma compared to carcinoma, such as *SPON2* [15], *SPP1*, and *SPARC* [11], which is validation for our own analysis.

Functional analysis of genes in carcinoma compared to adenoma group revealed that the most significant biological processes and KEGG pathways are connected to extracellular matrix (ECM). Top two significant biological processes in this comparison are ECM disassembly and the other ECM organization; furthermore the top KEGG pathway is ECM-receptor interaction. Genes involved in these two biological process pathways are similar; *COL12A1*, *COL1A2*, *COL3A1*, *DCN*, *FN1*, and *SPP1* are involved in ECM disassembly and the same genes with addition of *SPARC* and *SULF1* are involved in ECM organization (Supplementary Table 2). Genes involved in these pathways are all upregulated in carcinoma compared to adenoma, indicating that the process of ECM organization is involved in the progression of adenoma to carcinoma.

The ECM is a superstructure, which has a supportive role, but on the other hand, it also delivers signals to cells, which determines their behavior. Therefore, the EMC is directly involved in process of EMT during malignant transformation and plays a major role in the pathology of cancer [27]. Results of our analysis show that nine out of 16 genes, which showed differential expression in carcinoma compared to adenoma, are components of ECM. These genes are all three collagen genes, *DCN, PLAU, SPARC, SPON2, SPP1*, and *SULF1*. They all showed an increase in expression in carcinoma compared to adenoma in our study.

Two collagen I proteins (*COL1A1*, *COL1A2*) were found significantly upregulated in cancer group compared to normal tissue. The study revealed higher expression of collagen I in stage II tumors, suggesting that the activation of collagen I is an early event in CRC progression. The finding suggests that expression of collagen I is higher at early stages of CRC and

that collagen I is needed for tumor invasiveness [28]. Studies on cell lines suggest that adherence to collagen I promotes intracellular signaling pathways, including AKT pathways; furthermore collagen I was demonstrated to induce EMT-like changes, associated with tumor progression and metastasis [29, 30]. Expression of *COL3A1* gene was shown to be upregulated in CRC compared to normal controls. Wang et al. used Kaplan-Meier survival analysis to show that increased *COL3A1* protein in cancer epithelial cells predicted a worse prognosis [31]. The study was expanded to plasma samples, where soluble extracellular protein *COL3A1* was also significantly higher in patients with CRC compared to normal controls. Also, *COL3A1* was found to promote CRC cell proliferation by activating AKT signaling pathway [31]. One study used microarray data (GSE20219) and experimentally validated *COL12A1* gene. Its expression continuously increased from normal, through adenoma to carcinoma. Moreover, expression of *COL12A1* was reported to clearly distinguish between normal, adenoma, and carcinoma group and may have further diagnostic potential [32]. Besides collagens, EMC contains also other proteins, such as proteoglycans, sulfatases, and phosphoproteins. DCN is a fibril-associated proteoglycan, found in EMC. Although upregulated when compared carcinoma to adenoma, the overall expression of *DCN* is downregulated when carcinoma to normal and adenoma to normal is compared. The role of *DCN* both in vivo and in vitro suggested that its role is tumor suppressive in stromal and epithelial cells [33]. A negative correlation between the immunoreactivity of *DCN* and malignant potential was observed [34].

The *PLAU* is a urokinase-like plasminogen activator (uPA), which is secreted serine protease that converts plasminogen into active plasmin. Binding of uPA to its receptor, uPA-R, activates its proteolytic activity, which promotes ECM degradation and subsequently the invasion and migration of tumor cells. The *PLAU* is found to be upregulated in CRC. Furthermore, increased activity of the plasmin/plasminogen system leads to tumor budding, which is also significantly related to lymph node metastasis [35]. SPARC is a member of the family of matricellular proteins, a calcium-binding protein. Studies have shown that *SPARC* expression in mesenchymal and stromal cells (MSC) was significantly higher compared to expression in cancer cells and in normal mucosa tissues. Low expression of *SPARC* is an independent unfavorable prognostic factor of colorectal cancer [36]. Another secreted ECM protein is SPON2, which belongs to mindin/F-spondin family. Spondin proteins play important role in different signaling pathways, important in cancer. SPON2 is found to be upregulated in many cancers, including CRC. SPON2 was also tested as a biomarker in plasma of CRC patients, where it was upregulated and downregulated after surgery was performed, indicating SPON2 to be associated with tumor burden [37]. SPP1 is phosphoprotein found upregulated in many cancers, including CRC. It was found to promote cell proliferation and metastasis by activating EMT [38]. The last ECM component that was significantly differentially expressed between adenoma and carcinoma is sulfatase, *SULF1*. Sulfatases are overexpressed in CRC and contribute to cell proliferation, migration, and invasion [39].

Other genes are not connected to ECM but are included in the progression of carcinoma. Two more genes had increased expression in carcinoma compared to adenoma, and these are *FADS1* and *G0S2*. FADS1 is a member of the fatty acid desaturase gene family and has been suggested to regulate inflammation by modifying the metabolite profiles of fatty acids, which may influence the progression of cancer. Decreased expression of FADS1 benefits development and growth of cancer cells, whereas increased expression was observed to be a protective factor in esophageal squamous cell carcinoma (ESCC). Decreased expression was associated with poor prognosis in patients with ESCC [40]. Expression of *G0S2* is downregulated in a wide variety of cancer cell types and has the properties of tumor suppressor. The upregulation of *G0S2* has shown a significant reduction in tumor cell growth and motility. Since the *G0S2* is a negative regulator of triglyceride catabolism, an altered lipid metabolism is present in transformation of cells from normal to cancerous [41].

The last four genes in our study were downregulated in carcinoma compared to adenoma, and those are *EPHA4, KIAA1324*, *L1TD1*, and *PCKS1*. Upregulation of EPHA4 was observed in various cancers, including CRC. The study shows that activated EPHA4 is associated with highly aggressive EMT-like phenotype. Also, activation of EPHA4 reduced E-cadherin expression and controlled cell migration and invasion through PI3K signaling [42]. The KIAA1324 is a transmembrane protein, also known as EIG121 (estrogen-induced gene 121). It was shown that KIAA1324 acts as a tumor suppressor in gastric cancer cell lines, where the induction of *KIAA1324* gene expression significantly reduced tumor size [43]. L1TD1 is RNA-binding protein, which is highly expressed in pluripotent cells. Depletion of L1TD1 leads to reduction in levels of OCT4 and NANOG and increased differentiation in human embryonic stem cells (hESCs). L1TD1 is required for self-renewal of hESCs and is reported as one of the key regulators of stem cell fate [44]. One study reported an increased expression of *PSKC1* in nasal polyps compared to the normal nasal mucosa. Furthermore, they showed that increased expression of *PSKC1* induces EMT-like process in airway epithelial cells. The cell lines displayed a morphological transformation from typical epithelial-like shape to an elongated, spindle morphology. The overexpression in *PSKC1* resulted in an enhanced cell proliferation and exhibits a significant increase in cell migration after wounding. Also the cells displayed reduced expression of epithelial markers and increased expression of mesenchymal markers [45].

In conclusion, distinguishing adenomas with epithelial misplacement (pseudoinvasion) from adenomas with early carcinomas (true invasion) is of great importance, in order to choose the optimal treatment. For this purpose, we identified 16 candidate genes with different expression patterns in adenoma compared to carcinoma, with a potential to discriminate between these two lesions, which will be the basis of our future work, where we will experimentally validate genes on selected tissue sections of adenomas with epithelial misplacement and adenomas with early carcinomas.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Supplementary Materials

Supplementary 1. Supplementary Figures: Figure 1 shows the number of genes differentially expressed in each comparison and their intersection; Figure 2 shows the logarithmic values of average intensities for normal, adenoma, and carcinoma samples for (a) GSE10714, (b) GSE37364, (c) GSE50114 and GSE50115.

Supplementary 2. Supplementary Table 1 shows a list of genes common in each comparison and their expressions for each project.

Supplementary 3. Supplementary Table 2 shows gene ontology and KEGG of differentially expressed genes in each comparison group.

References

[1] C. Balch, J. B. Ramapuram, and A. K. Tiwari, "The Epigenomics of Embryonic Pathway Signaling in Colorectal Cancer," *Frontiers in Pharmacology*, vol. 8, 2017.

[2] E. R. Fearon and B. Vogelstein, "A genetic model for colorectal tumorigenesis," *Cell*, vol. 61, no. 5, pp. 759–767, 1990.

[3] M. Malvezzi, G. Carioli, P. Bertuccio et al., "European cancer mortality predictions for the year 2017, with focus on lung cancer," *Annals of Oncology*, vol. 28, no. 5, pp. 1117–1123, 2017.

[4] H. Cao, E. Xu, H. Liu, L. Wan, and M. Lai, "Epithelial-mesenchymal transition in colorectal cancer metastasis: A system review," *Pathology - Research and Practice*, vol. 211, no. 8, pp. 557–569, 2015.

[5] N. A. Shepherd and R. K. L. Griggs, "Bowel cancer screening-generated diagnostic conundrum of the century: Pseudoinvasion in sigmoid colonic polyps," *Modern Pathology*, vol. 28, pp. S88–S94, 2015.

[6] N. C. Panarelli, T. Somarathna, W. S. Samowitz et al., "Diagnostic Challenges Caused by Endoscopic Biopsy of Colonic Polyps: A Systematic Evaluation of Epithelial Misplacement with Review of Problematic Polyps from the Bowel Cancer Screening Program, United Kingdom," *The American Journal of Surgical Pathology*, vol. 40, no. 8, pp. 1075–1083, 2016.

[7] R. K. L. Griggs, M. R. Novelli, D. S. A. Sanders et al., "Challenging diagnostic issues in adenomatous polyps with epithelial misplacement in bowel cancer screening: 5 years' experience of the Bowel Cancer Screening Programme Expert Board," *Histopathology*, vol. 70, no. 3, pp. 466–472, 2017.

[8] M. Sheffer, M. D. Bacolod, O. Zuk et al., "Association of survival and disease progression with chromosomal instability: A genomic exploration of colorectal cancer," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 106, no. 17, pp. 7131–7136, 2009.

[9] D. Barras, E. Missiaglia, P. Wirapati et al., "BRAF V600E mutant colorectal cancer subtypes based on gene expression," *Clinical Cancer Research*, vol. 23, no. 1, pp. 104–115, 2017.

[10] M. Bianchini, E. Levy, C. Zucchini et al., "Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa," *International Journal of Oncology*, vol. 29, pp. 83–94, 2006.

[11] O. Galamb, F. Sipos, S. Spisak et al., "Potential biomarkers of colorectal adenoma-dysplasia-carcinoma progression: mRNA expression profiling and in situ protein detection on TMAs reveal 15 sequentially upregulated and 2 downregulated genes," *Cellular Oncology*, vol. 31, pp. 19–29, 2009.

[12] M. Skrzypczak, K. Goryca, T. Rubel et al., "Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability," *PLoS ONE*, vol. 5, no. 10, Article ID e13091, 2010.

[13] H. Tang, Q. Guo, C. Zhang et al., "Identification of an intermediate signature that marks the initial phases of the colorectal adenoma-carcinoma transition," *International Journal of Molecular Medicine*, vol. 26, no. 5, pp. 631–641, 2010.

[14] A. H. Sillars-Hardebol, B. Carvalho, M. De Wit et al., "Identification of key genes for carcinogenic pathways associated with colorectal adenoma-to-carcinoma progression," *Tumor Biology*, vol. 31, no. 2, pp. 89–96, 2010.

[15] B. Carvalho, A. H. Sillars-Hardebol, C. Postma et al., "Colorectal adenoma to carcinoma progression is accompanied by changes in gene expression associated with ageing, chromosomal instability, and fatty acid metabolism," *Cellular Oncology*, vol. 35, no. 1, pp. 53–63, 2012.

[16] L. Gautier, L. Cope, B. M. Bolstad, and R. A. Irizarry, "Affy—analysis of Affymetrix GeneChip data at the probe level," *Bioinformatics*, vol. 20, no. 3, pp. 307–315, 2004.

[17] M. E. Ritchie, B. Phipson, D. Wu et al., "*limma* powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, 2015.

[18] G. K. Smyth, "Linear models and empirical Bayes methods for assessing differential expression in microarray experiments," *Statistical Applications in Genetics and Molecular Biology*, vol. 3, no. 1, article 3, 2004.

[19] O. Galamb, F. Sipos, N. Solymosi et al., "Diagnostic mRNA expression patterns of inflamed, benign, and malignant colorectal biopsy specimen and their correlation with peripheral blood results," *Cancer Epidemiology, Biomarkers & Prevention*, vol. 17, no. 10, pp. 2835–2845, 2008.

[20] O. Kitahara, Y. Furukawa, T. Tanaka et al., "Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia," *Cancer Research*, vol. 61, no. 9, pp. 3544–3549, 2001.

[21] S. Lechner, U. Müller-Ladner, B. Renke, J. Schölmerich, J. Rüschoff, and F. Kullmann, "Gene expression pattern of laser microdissected colonic crypts of adenomas with low grade dysplasia," *Gut*, vol. 52, no. 8, pp. 1148–1153, 2003.

[22] D. A. Notterman, U. Alon, A. J. Sierk, and A. J. Levine, "Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays," *Cancer Research*, vol. 61, no. 7, pp. 3124–3130, 2001.

[23] E. Staub, J. Groene, M. Heinze et al., "Genome-wide expression patterns of invasion front, inner tumor mass and surrounding normal epithelium of colorectal tumors," *Molecular Cancer*, vol. 6, article no. 79, 2007.

[24] A. H. Wiese, J. Auer, S. Lassmann et al., "Identification of gene signatures for invasive colorectal tumor cells," *Cancer Epidemiology*, vol. 31, no. 4, pp. 282–295, 2007.

[25] M. Nannini, M. A. Pantaleo, A. Maleddu, A. Astolfi, S. Formica, and G. Biasco, "Gene expression profiling in colorectal cancer using microarray technologies: Results and perspectives," *Cancer Treatment Reviews*, vol. 35, no. 3, pp. 201–209, 2009.

[26] G. Valcz, Á. V. Patai, A. Kalmár et al., "Myofibroblast-Derived SFRP1 as Potential Inhibitor of Colorectal Carcinoma Field Effect," *PLoS ONE*, vol. 9, no. 11, p. e106143, 2014.

[27] G. Tzanakakis, R.-M. Kavasi, K. Voudouri et al., "Role of the extracellular matrix in cancer-associated epithelial to mesenchymal transition phenomenon," *Developmental Dynamics*, vol. 247, no. 3, pp. 368–381, 2017.

[28] X. Zou, B. Feng, T. Dong et al., "Up-regulation of type I collagen during tumorigenesis of colorectal cancer revealed by quantitative proteomic analysis," *Journal of Proteomics*, vol. 94, pp. 473–485, 2013.

[29] S. C. Kirkland, "Type i collagen inhibits differentiation and promotes a stem cell-like phenotype in human colorectal carcinoma cells," *British Journal of Cancer*, vol. 101, no. 2, pp. 320–326, 2009.

[30] L. Krasny, N. Shimony, K. Tzukert et al., "An in-vitro tumour microenvironment model using adhesion to type i collagen reveals Akt-dependent radiation resistance in renal cancer cells," *Nephrology Dialysis Transplantation*, vol. 25, no. 2, pp. 373–380, 2010.

[31] X. Wang, Z. Tang, D. Yu et al., "Epithelial but not stromal expression of collagen alpha-1(III) is a diagnostic and prognostic indicator of colorectal carcinoma," *Oncotarget*, vol. 7, no. 8, 2016.

[32] M. Mikula, T. Rubel, J. Karczmarski, K. Goryca, M. Dadlez, and J. Ostrowski, "Integrating proteomic and transcriptomic high-throughput surveys for search of new biomarkers of colon tumors," *Functional & Integrative Genomics*, vol. 11, no. 2, pp. 215–224, 2011.

[33] X. Bi, N. M. Pohl, Z. Qian et al., "Decorin-mediated inhibition of colorectal cancer growth and migration is associated with E-cadherin *in vitro* and in mice," *Carcinogenesis*, vol. 33, no. 2, pp. 326–330, 2012.

[34] K. Augoff, J. Rabczynski, R. Tabola, L. Czapla, K. Ratajczak, and K. Grabowski, "Immunohistochemical study of decorin expression in polyps and carcinomas of the colon," *Medical Science Monitor*, vol. 14, no. 10, pp. CR530–CR535, 2008.

[35] B. Märkl, I. Renk, D. V. Oruzio et al., "Tumour budding, uPA and PAI-1 are associated with aggressive behaviour in colon cancer," *Journal of Surgical Oncology*, vol. 102, no. 3, pp. 235–241, 2010.

[36] J. Liang, H. Wang, H. Xiao et al., "Relationship and prognostic significance of SPARC and VEGF protein expression in colon cancer," *Journal of Experimental & Clinical Cancer Research*, vol. 29, no. 1, p. 71, 2010.

[37] Q. Zhang, X. Wang, J. Wang et al., "Upregulation of spondin-2 predicts poor survival of colorectal carcinoma patients," *Oncotarget*, vol. 6, no. 17, 2015.

[38] C. Xu, L. Sun, C. Jiang et al., "SPP1, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating EMT pathway," *Biomedicine & Pharmacotherapy*, vol. 91, pp. 1167–1177, 2017.

[39] C. M. Vicente, M. A. Lima, E. A. Yates, H. B. Yates, and L. Toma, "Enhanced tumorigenic potential of colorectal cancer cells by extracellular sulfatases," *Molecular Cancer Research*, vol. 13, no. 3, pp. 510–523, 2015.

[40] Y. Du, S.-M. Yan, W.-Y. Gu et al., "Decreased expression of FADS1 predicts a poor prognosis in patients with esophageal squamous cell carcinoma," *Asian Pacific Journal of Cancer Prevention*, vol. 16, no. 12, pp. 5089–5094, 2015.

[41] R. Zagani, W. El-Assaad, I. Gamache, and J. G. Teodoro, "Inhibition of adipose triglyceride lipase (ATGL) by the putative tumor suppressor G0S2 or a small molecule inhibitor attenuates the growth of cancer cells," *Oncotarget*, vol. 6, no. 29, pp. 28282–28295, 2015.

[42] P. G. de Marcondes and J. A. Morgado-Díaz, "The Role of EphA4 Signaling in Radiation-Induced EMT-Like Phenotype in Colorectal Cancer Cells," *Journal of Cellular Biochemistry*, vol. 118, no. 3, pp. 442–445, 2017.

[43] J. M. Kang, S. Park, S. J. Kim et al., "KIAA1324 suppresses gastric cancer progression by inhibiting the oncoprotein GRP78," *Cancer Research*, vol. 75, no. 15, pp. 3087–3097, 2015.

[44] E. Närvä, N. Rahkonen, M. R. Emani et al., "RNA-binding protein L1TD1 interacts with LIN28 via RNA and is required for human embryonic stem cell self-renewal and cancer cell proliferation," *Stem Cells*, vol. 30, no. 3, pp. 452–460, 2012.

[45] S.-N. Lee, D.-H. Lee, M. H. Sohn, and J.-H. Yoon, "Over-expressed proprotein convertase 1/3 induces an epithelial-mesenchymal transition in airway epithelium," *European Respiratory Journal*, vol. 42, no. 5, pp. 1379–1390, 2013.