Identification of Prognostic Biomarker Signatures and Candidate Drugs in Colorectal Cancer: Insights from Systems Biology Analysis

Rezanur Rahman 1,2,#, Tania Islam 1,#, Esra Gov 3, Beste Turanli 4,5, Gizem Gulfidan 4, Shahjaman 6, Nilufer Akhter Banu 1, Nurul Haque Mollah 7, Kazim Yalcin Arga 4 and Mohammad Ali Moni 8,*

1 Department of Biotechnology and Genetic Engineering, Islamic University, Kushtia, Bangladesh; rezanur12@yahoo.com (M.R.R.), taniaislam1304@gmail.com (T.I.).
2 Department of Biochemistry and Biotechnology, School of Biomedical Science, Khwaja Yunus Ali University, Sirajgonj, Bangladesh; rezanur12@yahoo.com (M.R.R.)
3 Department of Bioengineering, Adana Science and Technology University, Adana, Turkey; egov@adanabtu.edu.tr (E.G.).
4 Department of Bioengineering, Marmara University, Istanbul, Turkey; bcalimlioglu@gmail.com (B.T.); gizemgulfidan@gmail.com (G.G.); kazim.arga@marmara.edu.tr (K.Y.A.).
5 Department of Bioengineering, Istanbul Medeniyet University, Istanbul, Turkey; bcalimlioglu@gmail.com (B.T.).
6 Department of Statistics, Begum Rokeya University, Rangpur, Bangladesh; shahjaman_brur@yahoo.com (M.S.).
7 Laboratory of Bioinformatics, Department of Statistics, University of Rajshahi, Rajshahi, Bangladesh; mollah.stat.bio@ru.ac.bd (M.N.H.M.).
8 The University of Sydney, Sydney Medical School, School of Medical Sciences, Discipline of Biomedical Science, Sydney, New South Wales, Australia; mohammad.moni@sydney.edu.au (M.A.M.).

*Correspondence: E-mail: mohammad.moni@sydney.edu.au (M.A.M.); rezanur12@yahoo.com (M.R.R.).
Tel.: +8861293519522.

Abstract: Background and objectives: Colorectal cancer (CRC) is the second most common cause of cancer-related death in the world, but early diagnosis ameliorates the survival of CRC. This report directed to identify molecular biomarker signatures in CRC. Materials and Methods: We analyzed two microarray datasets (GSE35279 and GSE21815) from Gene Expression Omnibus (GEO) to identify mutual differentially expressed genes (DEGs). We integrated DEGs with protein-protein interaction and transcriptional/post-transcriptional regulatory networks to identify reporter signaling and regulatory molecules; utilized functional overrepresentation and pathway enrichment analyses to elucidate their roles in biological processes and molecular pathways; performed survival analyses to evaluate their prognostic performance; and applied drug repositioning analyses through Connectivity map (CMap) and geneXpharma tools to hypothesize possible drug candidates targeting reporter molecules. Results: A total of 727 up-regulated and 99 down-regulated DEGs were detected. The PI3K-Akt signaling, Wnt signaling, ECM-interaction, and cell cycle were identified as significantly enriched pathways. Ten hub proteins (ADNP, CCND1, CD44, CDK4, CEBPB, CENPA, CENPH, CENPN, MYC, and RFC2), 10 transcription factors (ETS1, ESR1, GATA1, GATA2, GATA3, AR, YBX1, FOXP3, E2F4, and PRDM14) and 2 miRNAs (miR-193b-3p and miR-615-3p) were detected as reporter molecules. The survival analyses through Kaplan Meier curves indicated remarkable performance of reporter molecules in estimation of survival probability in CRC patients. In addition, several drug candidates including anti-neoplastic and immunomodulating agents were repositioned. Conclusions: This study presents biomarker signatures at protein and RNA levels with prognostic capability in CRC. We think that the molecular signatures and candidate drugs presented in this study might be useful in future...
studies indenting development of accurate diagnostic and/or prognostic biomarker screens and efficient therapeutic strategies in CRC.

**Keywords:** Colorectal cancer; differentially expressed genes; biomarkers; protein-protein interaction; reporter biomolecules; candidate drugs; systems biology; drug repositioning.

### 1. Introduction

Colorectal cancer (CRC) is the second most common cause of mortality of male and female in the world [1]. The number of CRC cases is still increasing, and the global burden of CRC is expected to increase by 60% to more than 2.2 million new cases and 1.1 million deaths by 2030 [2]. Like other cancers, a number of factors such as genetic factors, epigenetic alterations, diet, and environmental factors contribute to the progression and metastasis of CRC [3,4]. Despite the comprehensive studies (as reviewed by [5]), the molecular mechanisms of CRC pathogenesis is only partially understood. Several biomarkers (KRAS and BRAF) are used to detect the CRC, but these biomarkers are not sufficiently sensitive and specific; consequently there is an urgent need for identification of efficacious biomarkers, therapeutic targets and agents for early diagnosis, prevention, and personalized therapy in CRC [6].

The gene expression profiling technologies have been employed for years to identify genetic alterations at transcriptional level that pave the way to candidate biomarkers in human diseases including cancers [7–9]. These biomarkers may be used in early detection and/or serve as novel therapeutic targets. The hundreds of differentially expressed genes (DEGs) have been identified in CRC from microarray data [10,11]; however, their roles within human signaling network and their transcriptional regulatory mechanisms via transcription factors (TFs) and microRNAs (miRNAs) were not studied in detailed within a network biomedicine approach. The regulatory biomolecules might be attractive biomarkers since several reports proposed miRNAs that act as key players in CRC as prognostic biomarkers [12,13].

The power of multi-omics analyses within network biomedicine perspective [14] in elucidation of molecular signatures in human diseases was previously shown in many human diseases such as head and neck cancers [15], esophageal squamous cell carcinoma [16], triple negative breast cancer [17], cervical cancer [18], ovarian cancer [19] and ovarian diseases [20], psoriasis [21] and type 2 diabetes [22]. Therefore, in this study, systems-based approaches have been considered to explore the potential biomarker signatures at protein (i.e., hub proteins and TFs) and RNA levels (i.e., miRNAs and mRNAs) (Figure 1). For this purpose, we considered mutual DEGs identified from two independent gene expression profiling studies to maintain robustness, integrated this information with human biomolecular networks (namely, protein-protein interaction and transcriptional/post-transcriptional regulatory networks) to identify reporter signaling and regulatory molecules, utilized functional overrepresentation and pathway enrichment analyses to elucidate the roles of reporter molecules in biological processes and molecular pathways, and performed survival analyses to evaluate their prognostic performance as potential biomarkers in CRC. In addition, several candidate drugs were repositioned in CRC using *in silico* drug repositioning tools, Connectivity map (CMap) [23] and geneXpharma [24], considering these biomarker signatures as therapeutic targets.
Figure 1: The integrative analytical pipeline employed in the present study. (A) The colorectal cancer (CRC) datasets were analyzed under Bioconductor platform in R. We used limma to detect the DEGs in CRC compared to normal samples. (B) GO terms and molecular pathways were identified by DEGs enrichment via DAVID. (C) The hub proteins were identified by PPI analysis. (D) The reporter feature algorithm was used to identify reporter biomolecules as transcriptional regulatory elements. (E) The survival analysis of the hub biomolecules through TCGA CRC datasets via SurvExpress and oncomiR. (F) The candidate drug molecules identified by cMap and geneXpharma.

2. Materials and Methods

2.1. High-throughput Microarray Gene Expression Datasets

To analyze mRNA signatures in CRC samples compared to normal tissues, two gene expression datasets obtained using Agilent microarrays in independent experiments, GSE35279 [25] and GSE21815 [26], were downloaded from the Gene Expression Omnibus (GEO) database [27], which is a public functional genomics data repository supporting MIAME compliant data submissions. Consequently, a total of 220 specimens (206 CRC specimens and 14 normal samples) were comparatively analyzed.

2.2. Identification of Differentially Expressed Genes

To characterize differentially expressed genes (DEGs), each dataset was normalized by means of the Robust Multi-Array Average (RMA) expression measure [28] and DEGs were identified from the normalized log-expression values using the multiple testing option of LIMMA (linear models for microarray data) [29] using R/Bioconductor platform (version Rx64 3.4.1). Benjamini-Hochberg’s method was used to control the false discovery rate. An adjusted p-value threshold of 0.01 with a fold-change cutoff of 2 was used to determine the statistical significance of differential expression.

2.3. Gene Ontology and Pathway Analysis

Clustering of DEGs and reporter molecules into functional groups (i.e., biological processes, and molecular pathways) was performed via DAVID’s functional annotation tool [30]. In the analyses, Kyoto Encyclopedia of Genes and Genomes (KEGG) [31] was preferably used as the pathway database and Gene Ontology (GO) project [32] was used as the annotation source for biological processes and molecular functions. Fisher’s exact test was used to evaluate the statistical significance. p-values were corrected via Benjamini–Hochberg’s method, and an adjusted p value threshold of adj-p < 0.05 was used for all enrichment analyses.
2.4. Reconstruction and Analysis of Protein–Protein Interaction Network in CRC

We recruited the previously reconstructed high-confidence PPI network of Homo sapiens [33] consisting of 288,033 physical interactions between 21,052 proteins to construct a PPI subnetwork around the proteins encoded by the identified DEGs. The subnetwork was visualized and analyzed via Cytoscape (v3.4 and 2.8.3) [34]. The topological analysis was performed to characterize the network properties through Cyto-Hubba plugin [35]. The dual-metric approach [17,22] utilizing a local (i.e., degree) and a global (i.e., betweenness centrality) metric simultaneously was employed to define hub proteins. The modules in the PPI sub-networks were identified using MCODE plug-in [36] in Cytoscape. The modules were further analyzed through enrichment analyses in DAVID’s functional annotation tool [30].

2.5. Identification of Reporter Biomolecules

To identify reporter regulatory molecules (i.e., TFs, and miRNAs) around which significant changes occur at transcriptional level, we employed the comprehensive human transcriptional and post-transcriptional regulatory network [37], consisting of the experimentally verified TF-target gene and miRNA-target gene interactions from HTRIdb [38] and miRTarbase (Release 6.0) [39] databases. The reporter features algorithm [40] was used and implemented as described previously [15, 18, 20] to obtain z-scores and corresponding p values of the molecules. The p-values were corrected via Benjamini–Hochberg’s method, and statistically significant (adj-p < 0.01) results were considered as reporter biomolecules.

2.6. Evaluation of the prognostic performance of reporter molecules

The prognostic power of reporter biomolecules (i.e., hubs, TFs, and miRNAs) was analyzed via multivariate Cox regression analysis as implemented in SurvExpress [41] and OncomiR [42] by using independent gene expression (RNA-Seq or miRNA-Seq) datasets obtained from The Cancer Genome Atlas (TCGA). The RNA-Seq dataset consists of 467 samples with their clinical information, whereas the miRNA-Seq data includes 424 patients. The patients were partitioned into low- and high-risk groups according to their prognostic index determined by SurvExpress or OncomiR. The differences in gene expression levels between the risk groups were represented via box-plots, and the statistical significance of the differences was estimated by Student t-test. The survival signatures of reporter biomolecules were evaluated by Kaplan-Meier plots, and a log-rank p-value < 0.05 was considered as the cut-off to describe statistical significance in all survival analyses.

2.7. Identification of Candidate Drugs

We used simultaneously the Connectivity Map (CMap) database [23] and geneXpharma tool [24] to identify potential candidate drugs. CMap stores the expression profiles from cultured human cells exposed to various small molecular agents. A total of 50,304 gene–drug interactions comprising 4344 genes and 11,939 drugs was presented in geneXpharma. The hypergeometric probability test was used to statistically associate drugs to CRC.
3. Results

3.1. Identification of Differentially Expressed Genes

We studied two microarray CRC datasets (GSE35279 and GSE21815) from independent experiments to detect DEGs dysregulated in CRC samples compared to normal tissues. The analyses presented 727 up-regulated and 99 down-regulatory genes mutually differentiated in both CRC datasets (Figure 2). Then, we performed gene set overrepresentation analyses to obtain the GO annotations (in terms of molecular function, biological process, and cellular component) and KEGG pathways significantly associated with DEGs. Top 5 GO terms for up-regulated and down-regulated DEGs were summarized in Table 1, and the significant molecular pathways altered in CRC were shown in Figure 3.

![Figure 2: Identification of Differentially Expressed Genes (DEGs) in colorectal cancer (CRC) from microarray CRC datasets. (A) The up-regulated genes in the CRC expression profiling datasets. (B) The down-regulated genes in the CRC expression profiling datasets.](image)
Table 1. Functional overrepresentation of differentially expressed genes in colorectal cancer.

<table>
<thead>
<tr>
<th>Gene Ontology</th>
<th>GO term</th>
<th># of genes</th>
<th>Coverage (%)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Up-regulated genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Biological Process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>collagen fibril organization</td>
<td>11</td>
<td>1.62</td>
<td>4.53×10⁻⁷</td>
</tr>
<tr>
<td></td>
<td>extracellular matrix organization</td>
<td>22</td>
<td>3.24</td>
<td>2.94×10⁻⁶</td>
</tr>
<tr>
<td></td>
<td>male gonad development</td>
<td>14</td>
<td>2.06</td>
<td>1.53×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td>positive regulation of transcription from RNA polymerase II promoter</td>
<td>58</td>
<td>8.56</td>
<td>3.90×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td>collagen catabolic process</td>
<td>11</td>
<td>1.62</td>
<td>5.07×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td>Extracellular region</td>
<td>84</td>
<td>12.4</td>
<td>2.40×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td>cytoplasm</td>
<td>216</td>
<td>31.9</td>
<td>5.80×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cellular Component</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>extracellular space</td>
<td>70</td>
<td>10.3</td>
<td>1.50×10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>basement membrane</td>
<td>11</td>
<td>1.62</td>
<td>2.56×10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>extracellular matrix</td>
<td>23</td>
<td>3.39</td>
<td>3.34×10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>protein binding</td>
<td>354</td>
<td>52.3</td>
<td>8.10×10⁻⁶</td>
</tr>
<tr>
<td></td>
<td>protein homodimerization activity</td>
<td>42</td>
<td>6.20</td>
<td>7.54×10⁻⁴</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Molecular Function</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>growth factor activity</td>
<td>15</td>
<td>2.21</td>
<td>1.04×10⁻³</td>
</tr>
<tr>
<td></td>
<td>extracellular matrix binding</td>
<td>6</td>
<td>0.88</td>
<td>1.47×10⁻³</td>
</tr>
<tr>
<td></td>
<td>amino acid transmembrane transporter activity</td>
<td>7</td>
<td>1.03</td>
<td>4.43×10⁻³</td>
</tr>
<tr>
<td><strong>Down-regulated genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Biological Process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>bicarbonate transport</td>
<td>5</td>
<td>4.90</td>
<td>5.89×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td>chloride transmembrane transport</td>
<td>4</td>
<td>3.92</td>
<td>4.00×10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>nervous system development</td>
<td>7</td>
<td>6.86</td>
<td>2.63×10⁻⁰</td>
</tr>
<tr>
<td></td>
<td>regulation of chloride transport</td>
<td>2</td>
<td>1.96</td>
<td>9.62×10⁻³</td>
</tr>
<tr>
<td></td>
<td>plasma membrane</td>
<td>31</td>
<td>30.4</td>
<td>0.0108</td>
</tr>
<tr>
<td></td>
<td>extracellular space</td>
<td>14</td>
<td>13.7</td>
<td>0.0135</td>
</tr>
<tr>
<td><strong>Cellular Component</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>integral component of membrane</td>
<td>36</td>
<td>35.3</td>
<td>0.0163</td>
</tr>
<tr>
<td></td>
<td>anchored component of membrane</td>
<td>4</td>
<td>3.92</td>
<td>0.0179</td>
</tr>
<tr>
<td></td>
<td>integral component of plasma membrane</td>
<td>13</td>
<td>12.7</td>
<td>0.0421</td>
</tr>
<tr>
<td></td>
<td>carbonate dehydratase activity</td>
<td>4</td>
<td>3.92</td>
<td>4.16×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td>hormone activity</td>
<td>5</td>
<td>4.90</td>
<td>0.0012</td>
</tr>
<tr>
<td><strong>Molecular Function</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>zinc ion binding</td>
<td>15</td>
<td>14.7</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>UDP-galactose:beta-N-acetylgalcosamine beta-1,3-galactosyltransferase activity</td>
<td>3</td>
<td>2.94</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>chloride channel activity</td>
<td>4</td>
<td>3.92</td>
<td>0.0025</td>
</tr>
</tbody>
</table>

The overrepresentation analyses indicated the up-regulation of collagen associated processes, extracellular matrix (ECM) organization, and male gonad development. The up-regulated proteins were mainly having protein binding activities and localized in extracellular environments or cytoplasm. On the other hand, transport process, most specifically bicarbonate and chloride transport, were down-regulated in CRC. Down-regulated proteins were mostly showing zinc ion...
binding, hormone and chloride channel activities and were localized in the integral component of plasma membrane (Table 1). In parallel to GO enrichment results, the PI3K-Akt signaling pathway, Wnt signaling pathway, cell cycle, lung cancer, ECM-receptor interaction, protein digestion and absorption, pathways in cancer, and TGF-beta signaling pathway were up-regulated in CRC (Figure 3A). Contrarily, nitrogen metabolism, pancreatic secretion, axon guidance, retinol metabolism, renin secretion, and chemical carcinogenesis pathways were down-regulated in CRC (Figure 3B).

Figure 3: The significant pathways altered in colorectal cancer. (A) Up-regulated pathways in colorectal cancer. (B) Down-regulated pathways in colorectal cancer.

3.2. Analysis of Protein-Protein Interaction Network to Identify Hub Proteins

To identify hub proteins, a PPI sub-network around proteins encoded by the DEGs was constructed, and its topological analysis was performed. Following the scale-free degree distribution and small-world properties of biological networks, the presence of 10 hub proteins (ADNP, CCND1, CD44, CDK4, CEBPB, CENPA, CENPH, CENPN, MYC, and RFC2) was detected using degree and betweenness centrality metrics. These hub proteins may play significant key roles in signal transduction during the progression of CRC (Table 2). Two functional modules were revealed from the PPI network: Module 1, consisting of IPO5, RBP2, and RAN, was associated with intracellular protein transport, and module 2, consisting of CENPN, CENPA, and CENPH, was enriched with sister chromatid cohesion, kinetochore and nucleosome assembly (data not shown).
Identification of Regulatory Biomolecules

To identify reporter regulatory molecules (i.e., TFs, and miRNAs) around which significant changes occur at transcriptional level, we integrated DEGs with human transcriptional and post-transcriptional regulatory network and employed the adopted version of reporter features algorithm [20, 40] for each dataset. Considering a statistical significance level of adj-\(p < 0.01\), we identified 10 TFs (ETS1, ESR1, GATA1, GATA2, GATA3, AR, YBX1, FOXP3, E2F4, and PRDM14) and 10 miRNAs (miR-16-5p, miR-26b-5p, miR-124-3p, let-7b-5p, miR-92a-3p, miR-192-5p, miR-155-5p, miR-93-5p, miR-193b-3p, and miR-17-5p) as the mutual transcriptional regulatory components in both CRC datasets (Table 3).

<table>
<thead>
<tr>
<th><strong>Table 2.</strong> Summary of hub proteins in colorectal cancer.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symbol</strong></td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>ADNP</td>
</tr>
<tr>
<td>CEBPB</td>
</tr>
<tr>
<td>CCND1</td>
</tr>
<tr>
<td>CD44</td>
</tr>
<tr>
<td>CDK4</td>
</tr>
<tr>
<td>CENPA</td>
</tr>
<tr>
<td>CENPH</td>
</tr>
<tr>
<td>RFC2</td>
</tr>
<tr>
<td>MYC</td>
</tr>
<tr>
<td>CENPN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Table 3.</strong> Summary of reporter regulators in colorectal cancer.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symbol</strong></td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>AR</td>
</tr>
<tr>
<td>GATA1</td>
</tr>
<tr>
<td>GATA2</td>
</tr>
<tr>
<td>GATA3</td>
</tr>
<tr>
<td>E2F4</td>
</tr>
<tr>
<td>ETS1</td>
</tr>
</tbody>
</table>
YBX1  Y-Box Binding Protein 1  Aberrant expression is associated with cancer
PRADM14 PR/SET Domain 14  Involved in breast cancer
ESR1  Estrogen Receptor 1  Involved in breast cancer
FOXP3 Forkhead Box P3 (afflicted with colorectal cancer)  DNA binding

Reporter microRNAs

miR-193b-3p  MicroRNA 193  Afflicted with CRC and epidermal squamous cell carcinoma
miR-615-3p  MicroRNA 615  Afflicted with CRC
miR-16-5p  MicroRNA 16  Potential biomarkers in gastric cancer
miR-26b-5p  MicroRNA 26  Afflicted with CRC
let-7b-5p  MicroRNA 7  Afflicted with CRC
miR-92a-3p  MicroRNA 92  Afflicted with CRC
miR-124-3p  MicroRNA 124  Afflicted with CRC, gastric and breast cancer
miR-484  MicroRNA 484  Afflicted with CRC
miR-192-5p  MicroRNA 192  Afflicted with CRC
miR-93-5p  MicroRNA 93  Afflicted with head and neck cancer

3.4. Survival Analysis of Biomolecules

We performed the survival analysis of biomolecules (i.e., 10 hubs, 10 TFs, and 10 miRNAs) using CRC datasets from TCGA. Based on expression levels of reporter biomolecules and estimated survival probabilities, the patients were partitioned into two groups (i.e., high-risk and low-risk groups). The differential gene expression levels in high- and low-risk groups were represented by the box-plots and the estimated the survival probabilities were represented by Kaplan-Meier plots. In simulations, hub proteins, reporter TFs and reporter miRNAs were considered as separate biomarker sets.

Almost all of the hub proteins (except RFC2) contributed the discrimination of risk groups as seen in statistical powers represented in box-plot (Figure 4A), and the hub proteins as a group demonstrated statistically significant prognostic capability with a hazards ratio of 2.57 (log-rank p=9.56x10^-6) (Figure 4B). The reporter TFs (log-rank p=0.0185) were also indicative of CRC prognosis with hazards ratios 1.75 (Figure 5B). Among those TFs, GATA1, GATA2, E2F4, ESR1, and PRDM14 were the major discriminators (Figure 5A). In addition, the survival analysis of a subset of reporter miRNAs, consisting of miR-193b-3p and miR-615-3p, showed a prognostic signature (log-rank p=0.014) (Figure 6).
Figure 4: The survival analysis of the hub genes in the prognosis of colorectal cancer. (A) The box plot represents the differential expression of the 10 hub genes in two risks groups. (B) Kaplan-Meier plot represents the prognostic ability of the hub gene signatures in CRC.
Figure 5: The survival assessment of the reporter TFs signatures in the prognosis of colorectal cancer. (A) The box plot represents the differential expression of the 10 TFs between two risks groups. (B) Kaplan-Meier plot represents the prognostic power of the TFs signatures in colorectal cancer.
Figure 6: The survival analysis of the reporter miRNAs signatures in colorectal cancer. Kaplan-Meier plot represents the prognostic ability of miRNA signatures (miR-193b-3p and miR-615-3p) in colorectal cancer.

3.5. Identification of Candidate Drugs through in silico Drug Repositioning

Regarding the hub proteins and TFs as potential drug targets in CRC, we identified potential drugs based on the transcriptome signatures guided drug repositioning tool, geneXpharma and CMap database. We considered only the common drugs between two databases for CRC. Statistical evaluation revealed 45 candidate drugs targeting 6 proteins (Table 4). The drugs were classified according to the anatomical sites and development stages (Figure 7). Among the 10 hub proteins considered as a drug target, 3 hub proteins i.e., CCND1, CDK4, MYC were targeted by 9 drugs (Table 4). Contrarily, among the 10 reporters TFs, 3 reporter TFs were targeted by 23 drugs (Table 4). The repositioned drugs were classified based on the Anatomical Therapeutic Chemical classification system and found that 16.12% were antineoplastic, 22.58% were antineoplastic and immunomodulating agents. The hormones and contraceptives agents (9.67%) were following the antineoplastic and immunomodulating agents. The repositioned drugs were analyzed and found that 49% of drugs were approved, whereas 48% were still under investigation and 3% were in the experimental stage (Figure 7).
Table 4. Selected repositioned drugs in colorectal cancer.

<table>
<thead>
<tr>
<th>Target</th>
<th>Repositioned Drug</th>
<th>Drug Class/Status/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Hub protein</strong></td>
<td></td>
</tr>
<tr>
<td>Gefitinib</td>
<td>Antineoplastic Agents/ Approved, Investigational/ used in the treatment of cancer</td>
<td></td>
</tr>
<tr>
<td>Hydrocortisone</td>
<td>Anti-Inflammatory Agents/Approved/ used in the treatment of inflammation, allergy, collagen diseases, asthma, and some neoplastic conditions</td>
<td></td>
</tr>
<tr>
<td>Irinotecan</td>
<td>Antineoplastic Agents/Approved, Investigational/ used in the treatment of colorectal cancer</td>
<td></td>
</tr>
<tr>
<td>Letrozole</td>
<td>Antineoplastic Agents/Approved, Investigational/ introduced for treatment of breast cancer</td>
<td></td>
</tr>
<tr>
<td>CCND1</td>
<td>Lidocaine</td>
<td>Anesthetics/Approved/ A local anesthetic and used as an antiarrhythmia agent</td>
</tr>
<tr>
<td>Methotrexate</td>
<td>Antimetabolites, Antineoplastic/Approved/ antineoplastic antimetabolite with immunosuppressant properties</td>
<td></td>
</tr>
<tr>
<td>Sirolimus</td>
<td>Antineoplastic and Immunomodulating Agents/Approved, Investigational/ a potent immunosuppressant and possesses both antifungal and antineoplastic properties</td>
<td></td>
</tr>
<tr>
<td>Tamoxifen</td>
<td>Antineoplastic and Immunomodulating agents/Approved/ for the treatment and prevention of breast cancer</td>
<td></td>
</tr>
<tr>
<td>Gefitinib</td>
<td>Antineoplastic Agents/ Approved, Investigational/ used in the treatment of cancer</td>
<td></td>
</tr>
<tr>
<td>CDK4</td>
<td>Lidocaine</td>
<td>Anesthetics/Approved/ local anesthetic and used as an antiarrhythmia agent</td>
</tr>
<tr>
<td>Sirolimus</td>
<td>Antineoplastic and Immunomodulating Agents/Approved, Investigational/ a potent immunosuppressant and possesses both antifungal and antineoplastic properties</td>
<td></td>
</tr>
<tr>
<td>Gefitinib</td>
<td>Antineoplastic Agents/ Approved, Investigational/ used in the treatment of cancer</td>
<td></td>
</tr>
<tr>
<td>MYC</td>
<td>Tamoxifen</td>
<td>Antineoplastic and Immunomodulating Agents/Approved/ for the treatment and prevention of breast cancer</td>
</tr>
<tr>
<td>Simvastatin</td>
<td>Cardiovascular System/Approved/ a lipid-lowering agent</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Reporter TFs</strong></td>
<td></td>
</tr>
<tr>
<td>Azathioprine</td>
<td>Antineoplastic and Immunomodulating agents/ Approved/ immunosuppressive antimetabolite pro-drug</td>
<td></td>
</tr>
<tr>
<td>Daunorubicin</td>
<td>Antineoplastic and Immunomodulating Agents/ Approved/ used in treatment of leukemia and other neoplasms</td>
<td></td>
</tr>
<tr>
<td>Dexamethasone</td>
<td>Antineoplastic Agents/Approved, Investigational, Vet approved/ for the treatment of endocrine disorders, rheumatic disorders, collagen diseases, dermatologic diseases</td>
<td></td>
</tr>
<tr>
<td>GATA3</td>
<td>Doxorubicin</td>
<td>Antineoplastic and Immunomodulating agents/Approved, Investigational/used neoplastic conditions like acute lymphoblastic leukemia,</td>
</tr>
<tr>
<td>Mercaptopurine</td>
<td>Antimetabolite antineoplastic agent with immunosuppressant properties/ Approved/ in the treatment of leukemia</td>
<td></td>
</tr>
<tr>
<td>Methotrexate</td>
<td>Antimetabolites, Antineoplastic/Approved/antineoplastic antimetabolite with immunosuppressant properties</td>
<td></td>
</tr>
<tr>
<td>Drug Name</td>
<td>Classification</td>
<td>Approval Status</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Clomifene</td>
<td>Estrogen Agonist/Antagonist/Approved, Investigational/used mainly in female</td>
<td></td>
</tr>
<tr>
<td>Daunorubicin</td>
<td>Antineoplastic and Immunomodulating Agents/ Approved/ used in treatment of</td>
<td></td>
</tr>
<tr>
<td>Dexamethasone</td>
<td>Antineoplastic Agents/Approved, Investigational/ for the treatment of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estriol</td>
<td>Estradiol Congeners/Approved, Investigational/ used as a test to determine the</td>
<td></td>
</tr>
<tr>
<td>Estrone</td>
<td>Hormones/Approved/ used for management of perimenopausal and</td>
<td></td>
</tr>
<tr>
<td>Etoposide</td>
<td>Antineoplastic Agents/Approved/ used in the treatment of refractory testicular</td>
<td></td>
</tr>
<tr>
<td>Fulvestrant</td>
<td>Antineoplastic and Immunomodulating Agents/Approved, Investigational/ a drug</td>
<td></td>
</tr>
<tr>
<td>Glibenclamide</td>
<td>Oral Hypoglycemics/Approved/ used for the treatment of non-insulin-dependent</td>
<td></td>
</tr>
<tr>
<td>ESR1</td>
<td>Imipramine/ Approved/ antidepressant used for the relief of symptoms of</td>
<td></td>
</tr>
<tr>
<td>Letrozole</td>
<td>Antineoplastic agents/Approved, Investigational/ introduced for the treatment</td>
<td></td>
</tr>
<tr>
<td>Megestrol</td>
<td>Antineoplastic and Immunomodulating Agents/Approved, Investigational/used in</td>
<td></td>
</tr>
<tr>
<td>Mifepristone</td>
<td>Investigational/ For the medical termination of intrauterine pregnancy. Also</td>
<td></td>
</tr>
<tr>
<td>Progesterone</td>
<td>Contraceptive Agents/Approved, Vet approved/ Progesterone acts on the uterus,</td>
<td></td>
</tr>
<tr>
<td>Raloxifene</td>
<td>Estrogen Agonist/Antagonist/Approved, Investigational/ used to prevent</td>
<td></td>
</tr>
<tr>
<td>Tamoxifen</td>
<td>Antineoplastic and Immunomodulating Agents/Approved/ for the treatment and</td>
<td></td>
</tr>
<tr>
<td>Testosterone</td>
<td>Androgens and Estrogens/Approved, Investigational/ In men, testosterone is</td>
<td></td>
</tr>
<tr>
<td>Cyproterone</td>
<td>Antineoplastic Agents and Hormone Antagonists/Approved, Investigational/ It is</td>
<td></td>
</tr>
<tr>
<td>Flufenamic acid</td>
<td>Antiinflammatory and Antirheumatic /Experimental/ analgesic,</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>Antineoplastic Agents, Hormonal/Approved, Investigational/ For the management</td>
<td></td>
</tr>
<tr>
<td>Flutamide</td>
<td>Contraceptive Agents/Approved, Investigational/ For the treatment of</td>
<td></td>
</tr>
<tr>
<td>Levonorgestrel</td>
<td>Abortifacient Agents and Blood Glucose Lowering Agents/Approved, Investigational/</td>
<td></td>
</tr>
<tr>
<td></td>
<td>indicated to control hyperglycemia</td>
<td></td>
</tr>
</tbody>
</table>
Spironolactone Agents causing hyperkalemia /Approved/ Used primarily to treat low-renin hypertension, hypokalemia, and Conn's syndrome
Testosterone Androgens and Estrogens/Approved, Investigational/ In men, testosterone is produced primarily by the interstitial cells of the testes. Functions to maintain libido and general wellbeing in women.

Figure 7: Drug repositioning results in colorectal cancer. (A). Classification of repurposed drugs according to drug development stages. (B) Distribution of approved drugs into anatomical therapeutic chemical drug classes.
4. Discussion

Colorectal cancer (CRC) is complex disease, and the molecular mechanisms of CRC pathogenesis is only partially understood. The augmenting effect of genetic, endocrinological perturbations, and epigenetic aberrations contribute to the pathobiology of CRC [4-6]. The high-throughput gene expression profiling technology has been considered as one of the efficient sources for screening of biomarker candidates [7-9]. Understanding the disease pathways and exploration of biomarkers requires integration of omics data from different levels, and the power of this multi-omics approach in elucidation of molecular signatures in human diseases was previously shown in many human diseases [14-22]. Consequently, we employed a system biology approach to explore the in-depth mechanism of CRC in the present study.

Analysis of differential gene expression in CRC using two different high-throughput experimentation resulted with identification of 727 up-regulated and 99 down-regulated DEGs. The pathway enrichment analyses revealed significant molecular pathways including Wnt signaling pathway and inflammatory signaling pathways, which were already implicated in the pathogenesis of CRC [43]. TGF-β pathway behaves as tumor suppressor or tumor promoter depending on the context in different cancers, and the TGF-β was proposed as a target for cancer therapy [44]. Considering the significant alterations in these pathways during the progression of the CRC, we propose their components of as potential therapeutic targets in CRC.

Analysis of the PPI provides insights into central mechanisms on the pathobiology of cancers [45]. The PPI networks were reconstructed in order to clarify the interaction among the identified DEGs. Several hub proteins came into prominence as the reporter signaling mediators in CRC associated PPI. The prognostic survival analysis showed that these hub genes were significantly associated with the worse survival outcomes in the CRC patients (Figure 5). Among the hub protein, ADNP is dysregulated in CRC with high WNT activity [46]; CEBPB is afflicted with colorectal cancer and glioblastoma cells [47,48]; CCND1 dysregulation contributes to the pathogenesis of CRC [49,50]; CD44 plays diverse roles in cancer cells [51]; the CDK4 is the target for different cancer treatment including colorectal cancer [51,52]; CENPA is associated in pathobiology of CRC [53]; CENPH was also implicated in CRC [54]; RFC2 is implicated in hematologic cancers [55,56]; MYC is dysregulated in CRC [57-59]. CENPN is a protein that in humans is involved in cell cycle process showing direct binding of CENPN to CENPA [60]. The modules significantly contained the nodes (i.e., CENPA, CENPN, and CENPH) which are associated with different cancer and disease progression as discussed above.

Significant TFs regulating the DEGs were also characterized. Among the reporter TFs, AR is dysregulated in the prostate cancer [61]; ETS is involved in a different type of cancers [62]; GATA2 is deregulated in CRC with poor survival outcomes [63]; GATA3 and GATA4 was proposed to be implicated in different cancers [64]; YBX1 and FOXP3 are markers of cancers [65-67]; the E2F4 disruption is involved in cancers [68,69]; the dysregulation of PRDM14 and ESR1 are found in breast cancers [70-72].

Expression of 500 miRNAs is mentioned in CRC [6]. Thus, we evaluated the biomarker potentiality of the miRNAs in CRC since they regulate genes involved in cell cycle [12,73,74]. We identified relevant miRNAs signatures (miR-193b-3p and miR-615-3p), and survival analysis showed their significant potential as biomarkers in CRC. Recently, Wu et al., found that dysregulation of miR-193b-3p affects the growth of CRC via TGF-beta and regulation of SMAD signaling pathway [75]. Our pathway enrichment results also showed the dysregulation of TGF-beta signaling pathway. Moreover, miR-193b-3p is predictive biomarkers of renal cell carcinoma [76]. The high expression of miR-615-3p was associated with pathogenesis of CRC and gastric cancer [77,78]. Researches on these miRNAs might provide therapeutic target for CRC.

The survival analysis of the hub genes, TFs, and miRNAs clarified that those gene signatures (MYC, CENPN, RFC, CENPA, CEBPB, ADNP, CDK4, CCND1, CENPH and CD44) have high potentiality of being prognostic biomarkers in CRC. It was found that high expression of reporter TFs signatures (AR, GATA1, GATA2, GATA3, EST1, YBX1, PRADM14, ESR1, E2F4, and FOXP3) were associated with worse survival outcomes of the CRC patients. The survival analysis of the
miRNAs signatures (miR-193b-3p and miR-615-3p) also showed significant prognostic power in CRC. In addition, we here identified 45 candidate repositioned drugs, which were mostly antineoplastics, antiabetics, and endocrinologicals.

Despite the tremendous significance of the computational finding of this present works, further experiments at transcription and protein expression levels (such as western blot, qRT-PCR, CRISPR/Cas9 gene editing, etc.) and in vitro and in vivo cell culture assays for potential drugs may be performed for confirmation of the above results.

5. Conclusions

We employed a well-established systems biomedicine framework where transcriptome datasets were incorporated with genome-scale human molecular networks to reveal molecular biomarker signatures at RNA (i.e., miRNAs, miRNAs) and protein (i.e., hub proteins and TFs) levels in CRC. The prognostic survival analysis of the identified reporter biomolecules revealed proteomic signatures consisting of hub proteins (MYC, CENPN, RFC, CENPA, CEBPB, ADNP, CDK4, CCND1, CENPH and CD44), and regulatory signatures consisting of TFs (AR, GATA1, GATA2, GATA3, EST1, YBX1, PRADM14, ESR1, E2F4, and FOXP3) and miRNAs (miR-193b-3p and miR-615-3p) as prognostic biomarker candidates in CRC. In addition, candidate repositioned drugs targeting hub proteins and TFs were identified. The identified biomarker signatures and candidate repositioned drugs in this study deserve further experimentation since they show importance as candidate biomarkers and therapeutics for precision medicine approaches to treat CRC.


Funding: This article did not receive no external funding.

Acknowledgments: We would like to express our thanks to the Department of Biotechnology and Genetic Engineering, Islamic University, Kushtia, Bangladesh, and Laboratory of Bioinformatics, Department of Statistics, University of Rajshahi, Bangladesh for providing the bioinformatics laboratory facility. This work was supported by Islamic University research grant (2017-2018). The financial support to Dr. Kazim Yalcin Arga by The Scientific and Technological Research Council of Turkey (TUBITAK) through projects 116M014 and 117S489, and Marmara University Research Fund (BAPKO) through projects FEN-DRP-250816-0417 and FEN-C-YP-170118-0013 are acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References


Calimlioglu, B.; Karagoz, K.; Sevimoglu, T.; Kilic, E. Gov, E. Arga, K.Y. Tissue-Specific Molecular Biomarker Signatures of Type 2 Diabetes: An Integrative Analysis of Transcriptomics and...
Protein-Protein Interaction Data. OMICS 2015,19,563–73; doi:10.1089/omi.2015.0088.


[53] Tomonaga, T.; Matsushita, K.; Yamaguchi, S. Overexpression and Mistargeting of Centromere Protein-A in Human Primary Colorectal Cancer Overexpression and Mistargeting of Centromere


Jenster, G. The role of the androgen receptor in the development and progression of prostate cancer. Semin Oncol 1999;26:407–421.


