



OPEN

## Integrated whole transcriptome and small RNA analysis revealed multiple regulatory networks in colorectal cancer

Hibah Shaath<sup>1,2</sup>, Salman M. Toor<sup>1</sup>, Mohamed Abu Nada<sup>3</sup>, Eyad Elkord<sup>4</sup> & Nehad M. Alajez<sup>1,2</sup>✉

Colorectal cancer (CRC) remains a global disease burden and a leading cause of cancer related deaths worldwide. The identification of aberrantly expressed messenger RNA (mRNA), long non-coding RNA (lncRNA), and microRNA (miRNA), and the resulting molecular interactions and signaling networks is essential for better understanding of CRC, identification of novel diagnostic biomarkers and potential development of therapeutic interventions. Herein, we performed microRNA (miRNA) sequencing on fifteen CRC and their non-tumor adjacent tissues and whole transcriptome RNA-Seq on six paired samples from the same cohort and identified alterations in miRNA, mRNA, and lncRNA expression. Computational analyses using Ingenuity Pathway Analysis (IPA) identified multiple activated signaling networks in CRC, including ERBB2, RABL6, FOXM1, and NFkB networks, while functional annotation highlighted activation of cell proliferation and migration as the hallmark of CRC. IPA in combination with in silico prediction algorithms and experimentally validated databases gave insight into the complex associations and interactions between downregulated miRNAs and upregulated mRNAs in CRC and vice versa. Additionally, potential interaction between differentially expressed lncRNAs such as *H19*, *SNHG5*, and *GATA2-AS1* with multiple miRNAs has been revealed. Taken together, our data provides thorough analysis of dysregulated protein-coding and non-coding RNAs in CRC highlighting numerous associations and regulatory networks thus providing better understanding of CRC.

CRC is one of the most prevalent types of cancer worldwide<sup>1,2</sup>, with an estimated 9% and 8% of all diagnosed cancers in males and females respectively, to be attributed to CRC in the United States<sup>1</sup>. Better understanding of CRC biology and the development of new strategies and therapies targeting CRC is crucial, as the current traditional options remain limited to chemotherapy, radiotherapy and surgery, which can be highly invasive, involving the use of highly toxic drugs with many undesirable side effects<sup>3,4</sup>. In recent years, scientists are realizing the importance of post-transcriptional regulation and their effects on cellular processes, and in particular, the roles they play in cancer onset and progression. Studying the changes in expression levels of mRNAs, miRNAs, and lncRNAs has become of major importance, as well as understanding the effects of their aberrant expression patterns on downstream effectors and alterations in important regulatory pathways.

Multiple molecular alterations and genetic signatures have been associated with the onset and progression of CRC, which can serve as potential prognostic and predictive biomarkers for the identification of high-risk patients. Of the most common gene signatures associated with CRC is the mutation of *Adenomatous polyposis coli* (*APC*) gene. The encoded tumor suppressor, in addition to preventing uncontrolled growth of cancerous tumors, regulates beta-catenin, affecting its interactions with E-cadherin responsible for cell adhesion<sup>5</sup>. *Apc*-mutant cells were recently found to be enriched for transcripts encoding several secreted WNT antagonists such as *notum*. The inhibition of NOTUM in *Apc*-mutant cells prevented their expansion and ability to form intestinal adenomas, therefore identifying *notum* expression as a key mediator in early mutation fixation in CRC<sup>6</sup>.

MiRNAs are a large and diverse class of non-coding RNAs, around 18–24 nucleotides in length that are known to play a role in cell proliferation, migration and invasion, and have been suggested as potential strategy to overcome CRC chemo-resistance in vitro and in experimental animals<sup>7,8</sup>. Abnormal expression patterns of

<sup>1</sup>College of Health and Life Sciences, Hamad Bin Khalifa University (HBKU), Qatar Foundation (QF), Doha, Qatar. <sup>2</sup>Translational Cancer and Immunity Center (TCIC), Qatar Biomedical Research Institute (QBRI), Hamad Bin Khalifa University (HBKU), Qatar Foundation (QF), PO Box 34110, Doha, Qatar. <sup>3</sup>Department of Surgery, Hamad Medical Corporation, Doha, Qatar. <sup>4</sup>Biomedical Research Center, School of Science, Engineering and Environment, University of Salford, Manchester, UK. ✉email: nalajez@hbku.edu.qa

Patient no.	Age (years)	Gender	TNM stage	Tumor location	RNA-Seq	miRNA-Seq
1	62	Male	I	Rectosigmoid	No	Yes
2	47	Male	III	Ascending colon	No	Yes
3	65	Female	IV	Sigmoid colon	No	Yes
4	65	Male	II	Cecum	No	Yes
5	65	Male	III	Transverse colon	No	Yes
6	67	Male	III	Descending colon	No	Yes
7	96	Male	II	Sigmoid colon	Yes	Yes
8	59	Female	IV	Ascending colon	Yes	Yes
9	60	Male	II	Rectosigmoid	No	Yes
10	75	Male	III	Rectosigmoid	Yes	Yes
11	65	Male	I	Ascending colon	Yes	Yes
12	52	Male	II	Transverse colon	Yes	Yes
13	69	Male	I	Rectosigmoid	No	Yes
14	60	Female	I	Cecum	No	Yes
15	67	Female	IV	Ascending colon	Yes	Yes

**Table 1.** Characteristic features of study populations. All patients presented with moderately differentiated tumors. CRC colorectal cancer.

miRNAs, detected through microarrays, and more recently, next-generation sequencing, have been implicated in a number of human cancers including breast cancer<sup>9,10</sup>, cervical squamous cell carcinoma<sup>11</sup>, bladder cancer<sup>12</sup>, and melanoma<sup>13</sup>. We previously reported miRNA-320, downregulated in primary CRC, to suppress CRC by targeting SOX4, FOXM1, and FOXQ1. Upon lentiviral mediated re-expression of miR-320, growth and migration of CRC cells in vitro was inhibited, sensitizing them to 5-Fluorouracil (5-FU) therapy, and inhibited tumor formation in SCID mice<sup>14</sup>.

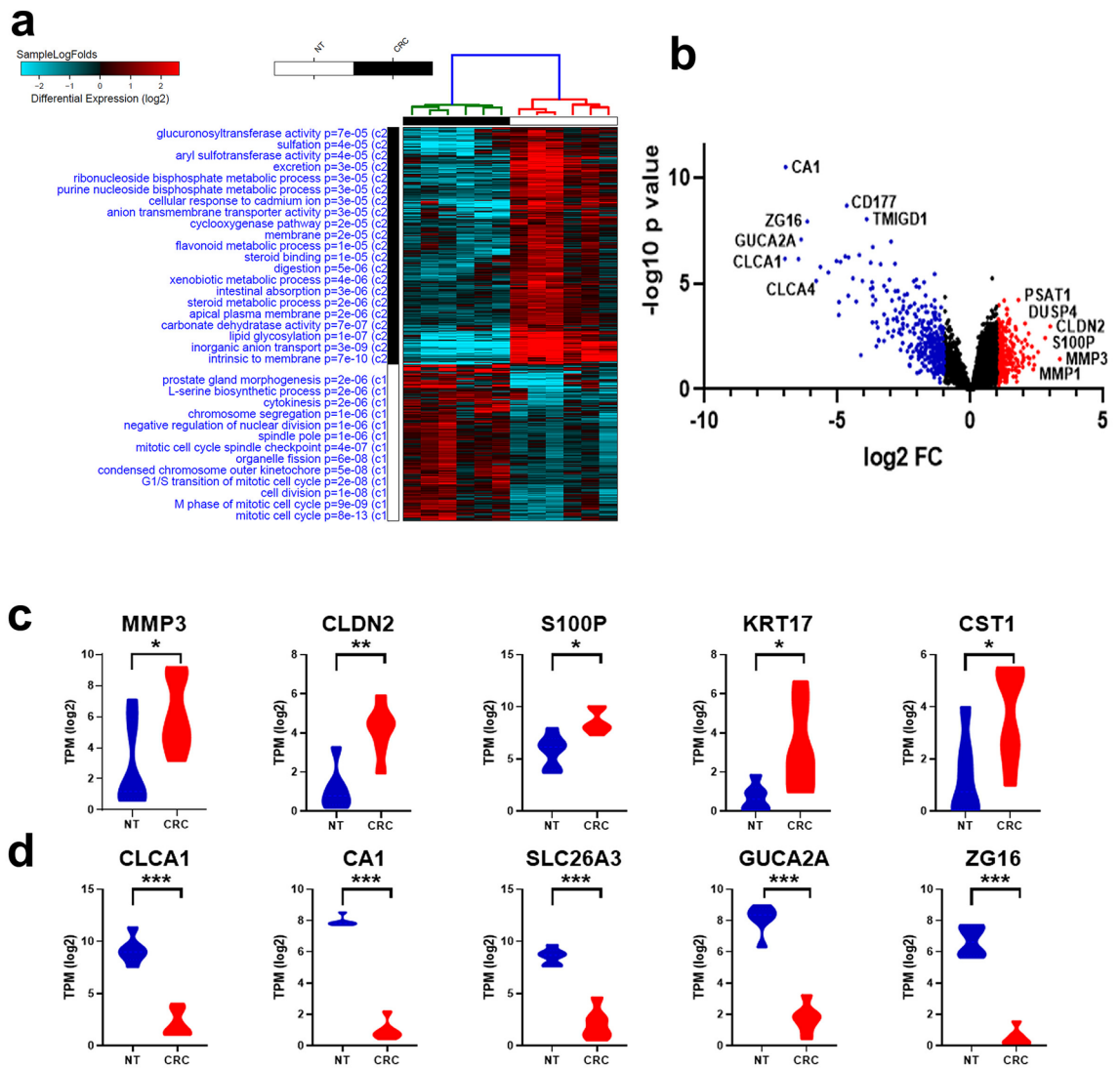
MiRNAs in cancer seldom work alone, and extensive research has provided us with known associations with lncRNAs, another group of non-coding RNAs of around 200 nucleotides<sup>15</sup>, working independently or together as part of an axis in governing specific cellular processes in cancer<sup>16</sup>. As an example, a recent study by Li et al., demonstrates the suppression of CRC cell growth upon knockdown of lncRNA XIST, via regulating microRNA-338-3p/PAX5 axis. This observation may serve as a potential target for the prevention and treatment of CRC<sup>16</sup>. XIST has also been described, in an earlier study, to inhibit 5-FU -induced cell cytotoxicity through promoting thymidylate synthase expression, suggesting the silencing of XIST to be beneficial in overcoming chemo-resistance during of CRC therapy<sup>17</sup>.

Our previous genome-wide mRNA and miRNA expression profiling revealed multiple regulatory networks in CRC including Wnt (wingless-type MMTV integration site family member), matrix metalloproteinase, and TGF- $\beta$  pathways. Pharmacological inhibition of these pathways led to dose- and time-dependent inhibition of CRC cell growth<sup>18</sup>. In another study, we revealed a number of systemic alterations in gene expression in circulation (peripheral blood mononuclear cells (PBMCs)), compared to the tumor microenvironment of CRC patients, which revealed activation of immune cell trafficking and cellular movement, and suppression of cellular processes related to cell death<sup>19</sup>.

In current study, we performed total RNA-seq and miRNA-Seq on paired CRC and adjacent normal tissue. Computational analyses identified a myriad of affected mRNA, miRNA and lncRNA interactions in CRC, resulting in multiple affected signaling networks. Functional annotation analysis also highlighted activation of cell proliferation and migration as the hallmarks of CRC. More in depth understanding of the roles played by the identified miRNAs and lncRNAs in CRC is crucial in the identification of potential biomarkers.

## Results

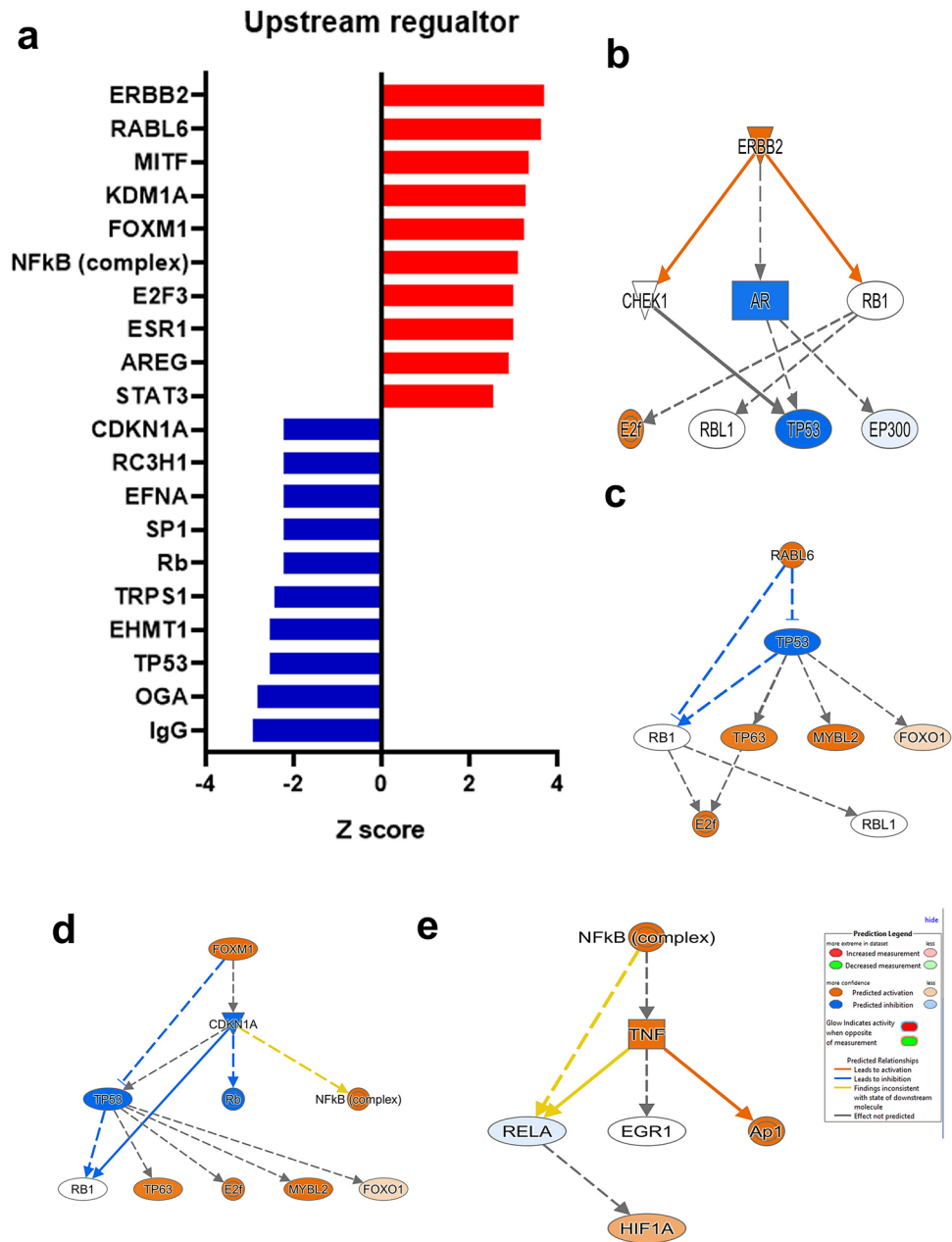
**Alterations in mRNA expression in CRC compared to NT.** CRC tissue and their adjacent NT from six patients were subjected to total RNA sequencing. Patients' characteristics and information on tumor site are shown in Table 1. Differential analysis identified several upregulated and downregulated mRNAs in CRC compared to NT (supplementary table 1). Hierarchical clustering of CRC and NT based on differentially expressed (log<sub>2</sub>) mRNAs between the two groups is illustrated in Fig. 1a. The heat map depicts remarkable differences in the expression levels of several mRNAs (each row representing an mRNA), grouped in relation to their indicated biological processes. Functions related to prostate gland morphogenesis, L-serine biosynthetic process, cytokinesis, chromosomes segregation, and several other cell cycle processes were significantly upregulated (red) in CRC tissue compared to NT, as seen in Fig. 1a. Downregulated functions in CRC tissue compared to NT include inorganic anion transport, lipid glycosylation, steroid metabolic processes and carbonate dehydratase activity (blue). Alterations in mRNA expression between CRC and NT are presented as volcano plot (Fig. 1b), highlighting upregulated mRNAs (red dots) based on log<sub>2</sub> FC including PSAT1, DUSP4, CLDN2, S100P, MMP3 and MMP1, among several others. Downregulated mRNAs are presented as blue dots on the volcano plot which include CA1, CD177, TMIGD1, ZG16, GUCA2A, CLCA1 and CLCA4. Figure 1c,d show the top 5 upregulated mRNAs (MMP3, CLDN2, S100P, KRT17 and CST1) and top five downregulated mRNAs (CLCA1, CA1,



**Figure 1.** Alteration in mRNA expression in CRC compared to NT. **(a)** Hierarchical clustering of CRC ( $n=6$ ) and adjacent normal tissue ( $n=6$ ) based on differentially expressed mRNAs between the two groups. Each column represents one sample and each row represents an mRNA. Expression level of each mRNA ( $\log_2$ ) in a single sample is depicted according to the color scale. **(b)** Volcano plot representation altered mRNAs in CRC vs. NT. Red and blue colors indicate the genes with significant increased or decreased expression ( $-2.0 \geq FC \geq 2.0$ ,  $p < 0.05$ ), respectively, while black color indicates genes with no significant change. Expression of top five upregulated **(c)** or downregulated **(d)** mRNAs in CRC vs NT based on RNA-Seq analysis.

*SLC26A3*, *GUCA2A* and *ZG16* in CRC vs NT based on RNA-Seq analysis ( $\log_2$  TPM). The top differentially expressed genes identified in our study were compared against the TCGA colon adenocarcinoma (COAD) dataset. Expression of the top 20 differentially expressed mRNAs in our cohort (10 upregulated and 10 downregulated) in the TCGA COAD dataset is shown in supplementary Fig. 1, revealing high concordance with our data.

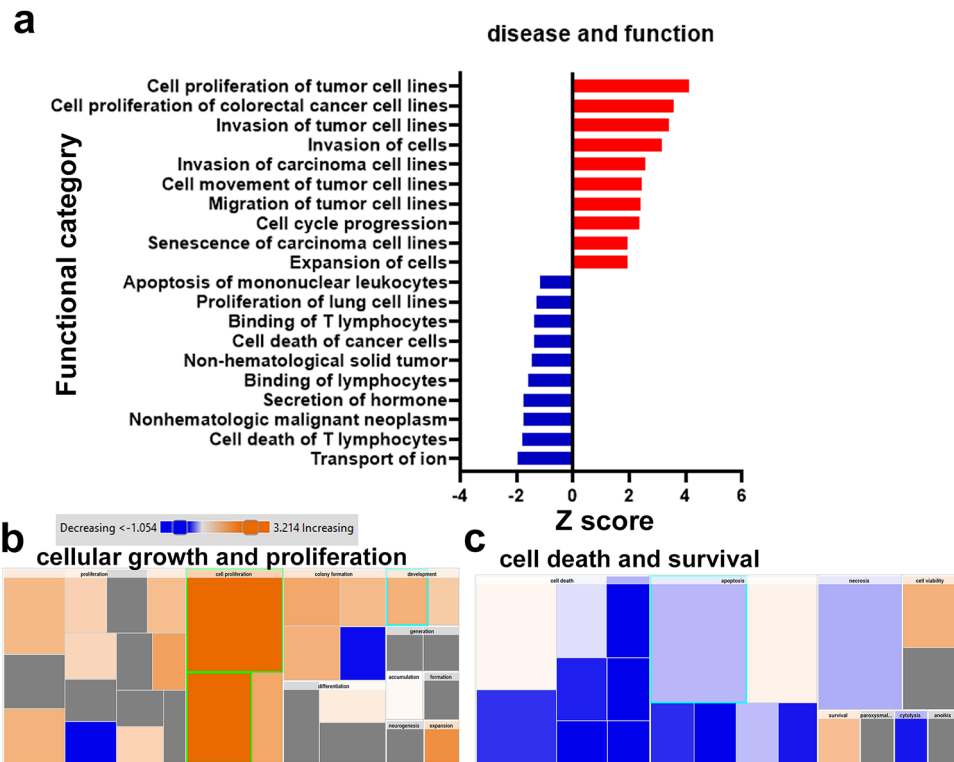
**Multiple affected signaling network in CRC compared to NT.** Differentially expressed mRNAs were subsequently subjected to IPA analysis revealing multiple affected signaling networks in CRC compared to NT (Fig. 2a). Those upstream regulators most activated with highest Z scores were ERBB2, RABL6, MITF, KDM1A, FOXM1, NFkB complex, ERF3, ESR1, AREG, and STAT3. Suppressed upstream regulators with negative Z scores were IgG, OGA, TP53, EHMT1, TRPS1, Rb, SP1, EFNA, RC3H1 and CDKN1A. Figure 2b–e show the ERBB2, RABL6, FOXM1, and NFkB complex networks in more detail. ERBB2 is predicted to activate RB1, which in turn activates E2f. ERBB2's activation of CHEK1, along with its inhibition of AR, simultaneously inhibit TP53 (Fig. 2b). The activation of RABL6 is also predicted to have an inhibitory effect on TP53, with both molecules also inhibiting RB1. The inhibition of TP53 however, leads to TP63, MYBL2 and FOXO1 activation, with RB1 and TP63 both activating E2f as a third level downstream effect of RABL6 activation (Fig. 2c). FOXM1 activation is predicted to lead to the inhibition of the CDKN1A network, which inhibits Rb but activates NFkB complex (inconsistent) (Fig. 2d). CDKN1A also inhibits TP53, which has its own downstream consequences as seen in



**Figure 2.** Multiple affected signaling networks in CRC. **(a)** Top altered upstream regulator networks in CRC compared to NT based on IPA analysis. Y-axis indicated the upstream regulator network and the x-axis represent the activation Z score. **(b–e)** are representation of ERBB2, RABL6, FOXM1, and NFkB networks, respectively.

Fig. 2c. Figure 2e highlights the downstream effects of NFkB activation, which activates TNF, which in turn activates Ap1 but suppresses RELA (inconsistent). This inhibition of RELA is predicted to activate HIF1A as an indirect effect of NFkb activation.

**Functional category annotation analysis highlights activation of cell proliferation and migration as the hallmark of CRC.** Functional category annotation of differentially expressed mRNAs brought forward disease and functions predicted to be altered in CRC. Activated disease and functional categories with higher Z scores include tumor cell proliferation, invasion, movement, and cell cycle progression. Those suppressed disease functions include transport of ions, cell death of T lymphocytes, non-hematologic malignant neoplasm, secretion of hormones, binding of T lymphocytes and cell death of cancer cells (Fig. 3a). Figure 3b and c show heat maps illustrating affected downstream biological processes including cellular growth and proliferation, and cell death and survival, respectively. Orange boxes depict activation function whereas blue highlights suppressed functions. As expected in cancer cells, functional categories associated with cellular



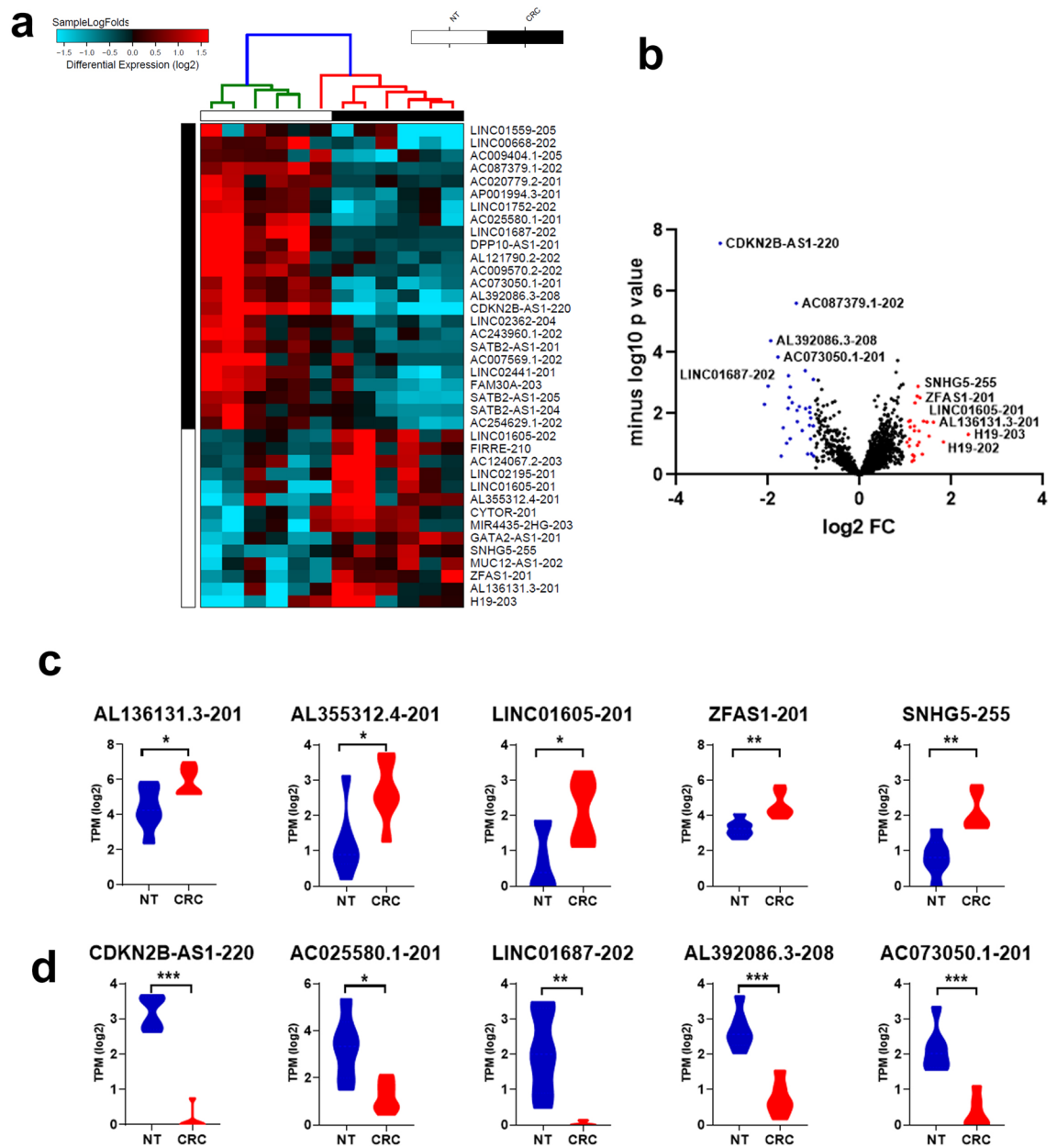
**Figure 3.** Functional category annotation analysis highlights activation of cell proliferation and migration as the hallmark of CRC. **(a)** Disease and function analysis of differentially expressed mRNAs in CRC. Y-axis indicated the functional category and the x-axis represent the activation Z score. Heatmap representation of cellular growth and proliferation **(b)** and cell death and survival **(c)** functional categories. Activation Z score is depicted according to the color scale.

growth and proliferation, colony formation and expansion are activated whereas those associated with cell death, apoptosis, and necrosis are suppressed, characteristic of cancer cell persistence.

**Alterations in lncRNA expression in CRC compared to NT.** lncRNA transcripts were also analyzed in the same cohort of six CRC and matched NT revealing several differentially expressed lncRNA transcripts (supplementary table 2). Hierarchical clustering of CRC (n=6) and adjacent NT (n=6) based on differentially expressed lncRNA transcripts (log<sub>2</sub>) is shown in Fig. 4a. Upregulated transcripts in CRC tissues (red) include *LINC01605-202*, *FIRRE-210*, *AC124067.2-203*, *LINC0219-201*, *LINC01605-201*, *AL355312.4-201*, *CYTOR-201*, *MIR4435-2HG-203*, *GATA2-AS1-201*, *SNHG5-255*, *MUC12-AS1-201*, *MUC12-AS1-202*, *ZFAS1-201*, *AL136131.3-201* and *H19-203*, while several other lncRNAs were downregulated (Fig. 4a). Figure 4b is a volcano plot displaying significantly altered lncRNA transcripts in CRC vs. NT. Blue dots represent the most significantly downregulated lncRNA transcripts whereas the red dots depict the most significantly upregulated lncRNA transcripts in CRC vs. NT. Figure 4c and d show the expression of top five upregulated (*AL136131.3-201*, *AL355312.4-201*, *LINC01605-201*, *ZFAS1-201* and *SNHG5-255*) and downregulated (*CDKN2B-AS1-220*, *AC025580.1-201*, *LINC01687-202*, *AL392086.3-208* and *AC073050.1-201*) lncRNAs in CRC compared to NT. In agreement with our data, several of the top upregulated and downregulated lncRNAs from our study showed similar expression pattern in the TCGA COAD dataset (supplementary Fig. 2).

**Alterations in miRNA expression in CRC compared to NT.** To provide a more comprehensive view of alterations in noncoding RNA transcriptome in CRC, miRNA expression profiling was conducted on 15 CRC tissues compared to the corresponding normal adjacent tissue using miRNA-Seq. Differential expression analysis revealed several differentially expressed miRNAs in CRC compared to NT (supplementary table 3). Hierarchical clustering based on miRNA expression revealed close clustering of CRC away from NT, except for few samples (Fig. 5a). Principle Component Analysis (PCA), showing the degree of relatedness of the samples to each other based on PC1 and PC2, also highlighted the clear segregation of the two groups (NT in red and CRC in blue), with a slight overlap of a few samples as evident in the hierarchical clustering analysis (Fig. 5b). *Hsa-miR-133a-3p*, *hsa-miR-363-3p*, *hsa-miR-145-5p*, *hsa-miR-363-3p*, and *hsa-miR-195-3p*, were among the downregulated while *hsa-miR-135b-5p*, *hsa-miR-552-5p*, *hsa-miR-224-5p*, *hsa-miR-183-5p* and *hsa-miR-552-3p* were among those upregulated in CRC. We subsequently validated top 10 upregulated and top 10 downregulated miRNAs from our study in the TCGA COAD dataset which showed high degree of concordance except for *hsa-miR-138-5p*, *hsa-miR-1-3p*, and *hsa-miR-9-5p* which were not concordant (supplementary Fig. 3). When

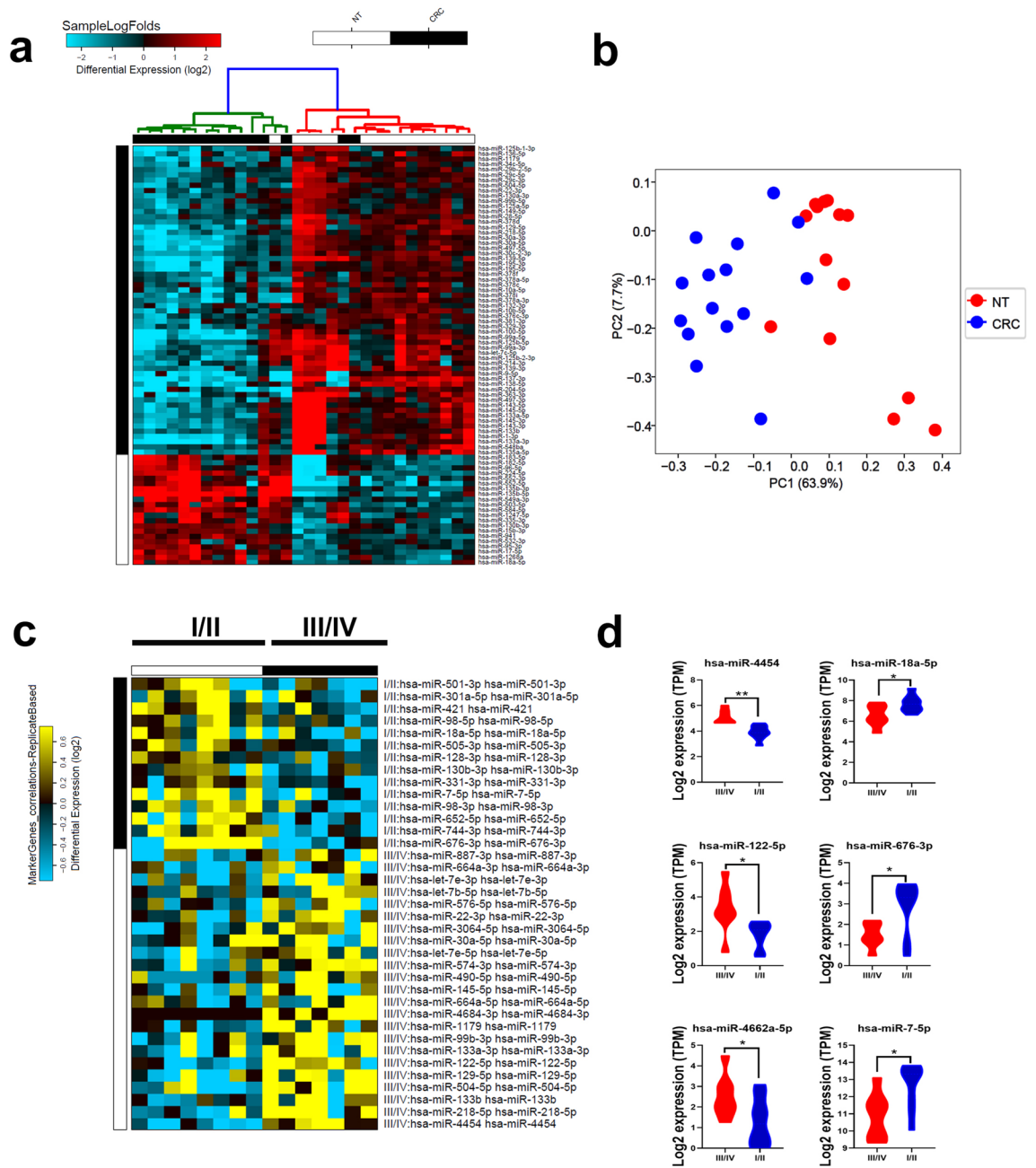




**Figure 4.** Alteration in lncRNA expression in CRC vs. NT. **(a)** Hierarchical clustering of CRC ( $n=6$ ) and adjacent normal tissue ( $n=6$ ) based on differentially expressed lncRNA transcripts between the two groups. Each column represents one sample and each row represents single lncRNA transcript. Expression levels of each transcript (log2) in a single sample is depicted according to the color scale. **(b)** Volcano plot representation of significantly altered lncRNA transcripts in CRC vs. NT. Red and blue colors indicate the transcripts with increased or decreased expression, respectively. Black color indicates no significant change. Expression of the top five upregulated **(c)** or downregulated **(d)** lncRNA transcripts in CRC vs. NT based on RNA-Seq analysis.

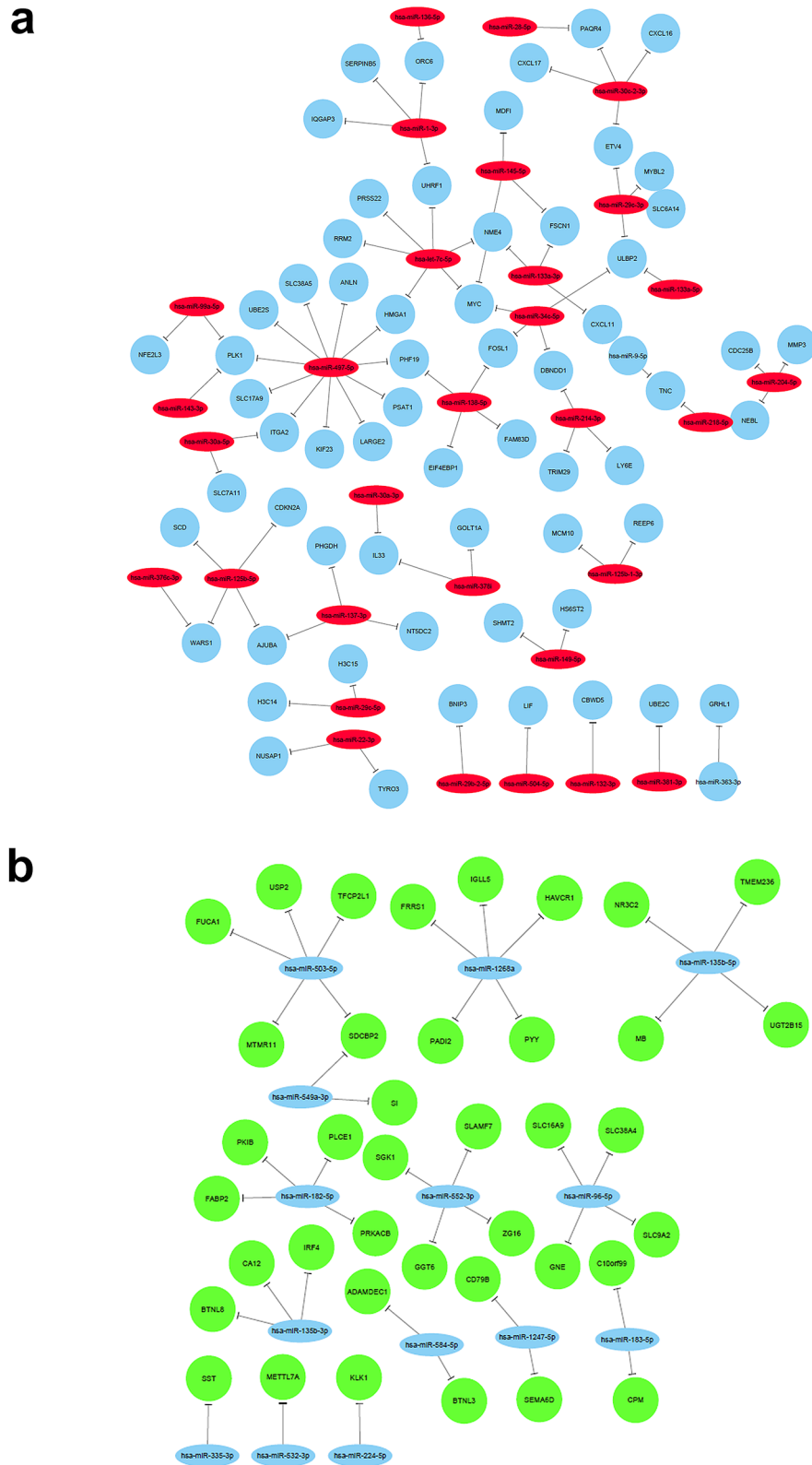
patients were grouped according to disease stage (stage III/IV vs. stage I/II), several miRNAs were found to correlate with disease stage as depicted in Fig. 5c. Among those, *hsa-miR-501-3p*, *hsa-miR-421*, *hsa-miR-18a-5p*, *hsa-miR-128-3p*, *hsa-miR-7-5p*, and *hsa-miR-744-3p*, were downregulated while *hsa-miR-let-7b-5p*, *hsa-miR-30a-5p*, *hsa-miR-145-5p*, *hsa-miR-1179*, *hsa-miR-133a-3p*, *hsa-miR-504-5p*, and *hsa-miR-218-5p* were upregulated in stage III/IV compared to stage I/II. Figure 5d shows the expression of three upregulated (*hsa-miR-4454*, *hsa-miR-122-5p*, and *hsa-miR-4662a-5p*) and three downregulated (*hsa-miR-18a-5p*, *hsa-miR-676-3p*, and *hsa-miR-7-5p*) miRNAs in CRC stage III/IV compared to stage I/II.

**miRNA-mRNA-lncRNA computational interaction prediction.** We subsequently integrated the miRNA and mRNA expression data using the microRNA filter with IPA thus connecting expression data from the current study with in silico prediction and the IPA manually curated database which identified many regulatory networks (supplementary table 4). We subsequently focused on the differentially expressed miRNAs and



**Figure 5.** Alteration in miRNA expression in CRC vs. NT. **(a)** Hierarchical clustering of CRC (n = 15) and adjacent normal tissue (n = 15) based on differentially expressed miRNAs between the two groups. Each column represents one sample and each row represents a miRNA. Expression level of each miRNA (log<sub>2</sub>) in a single sample is depicted according to the color scale. **(b)** PCA analysis depicting relatedness of the samples based on PC1 and PC2. **(c)** Heat map illustrating the most differentially expressed miRNAs in CRC in Stage I/II vs. Stage III/IV. Expression level in each sample is depicted according to the color scale. **(d)** Expression of three upregulated (*hsa-miR-4454*, *hsa-miR-122-5p*, and *hsa-miR-4662a-5p*) and three downregulated (*hsa-miR-18a-5p*, *hsa-miR-676-3p*, and *hsa-miR-7-5p*) in CRC stage III/IV vs. stage I/III.

mRNAs in CRC and those which were experimentally observed or predicted with high confidence based on IPA analysis which led to the identification of 33 downregulated and 14 upregulated miRNAs and their respected mRNA targets. Visualization of the miRNA-mRNA interactions are depicted using cystoscope (Fig. 6). Starting with *hsa-miR-1-3p*, the downregulation of this miRNA is predicted to upregulate *IQGAP3*, *SERPINB5*, *ORC6* (also associated with the downregulation of *hsa-miR-136-5p*), and *UHRF1*. *UHRF1*, along with *NME4*, *MYC*, *HMGA1*, *RRM2*, and *PRSS22*, interact with the downregulated *hsa-miR-7c-5p* and so on. Figure 6a illustrates in detail, the complex associations and interactions between downregulated miRNAs and upregulated mRNAs in CRC, including several experimentally validated interactions and associations with high confidence according to IPA predictions. Figure 6b shows interactions between upregulated miRNAs and downregulated mRNAs in



**Figure 6.** Networks illustrating the interactions between differentially expressed miRNAs and mRNAs in CRC. **(a)** Network illustrating the interaction between downregulated miRNAs and upregulated mRNAs in CRC based on Targetscan prediction and IPA analysis. **(b)** Network illustrating the interaction between upregulated miRNAs and downregulated mRNAs in CRC based on Targetscan prediction and IPA analysis. Only experimentally validated and interactions with high confidence are included in both networks.



CRC. For example, upregulated *hsa-miR-1268a* could downregulate *FRRS1*, *IGLL5*, *HAVCR1*, *PYY* and *PADI2*, whereas upregulated *hsa-miR-1356-5p* is predicted to downregulate *NR3C2*, *TMEM236*, *MB* and *UGT2B15*. Other mRNAs can be downregulated by more than one miRNA, such as in the case of *SDCBP2*, which is downregulated by the upregulation of both *hsa-miR-503-5p* and *hsa-miR-549a-3p*.

LncRNAs are believed to act as sponge to sequester mature miRNAs and to exert their biological function. Therefore, we explored the DIANA-LncBase database connecting differentially expressed lncRNAs and their corresponding downregulated miRNAs from the current study. For example, Fig. 7a highlights the complex interactions between miRNAs and lncRNAs in CRC in more detail, illustrating the complex associations predicted in CRC. LncRNA *H19* is shown to interact with more than 20 downregulated miRNAs including, *hsa-miR-378a-3p*, *hsa-miR-30c-2-3p*, *hsa-miR-22-3p*, and *hsa-miR-125a-5p*. Other miRNAs that interact with *H19*, also interact with other lncRNAs such as *hsa-miR-138-5p*, which interacts with *H19*, *SNHG5* and *GATA2-AS1*. Inversely, interactions between upregulated miRNAs and downregulated lncRNAs in CRC based on DIANA-LncBase database show *hsa-miR-503-5p* to interact with *FAM30A*, *hsa-miR-17-5p* to interact with *SATB2-AS1*, and *hsa-miR-18a-5p* to interact with both *LINC02362* and *CDKN2B-AS1*, which in turn, interacts with all of *hsa-miR-15b-3p*, *hsa-miR-130b-3p*, and *hsa-miR-183-5p*. LncRNA *AC009404.1* also shows association with upregulated miRNA *hsa-miR-183-5p*, in addition to both *hsa-miR-182-5p* and *hsa-miR-335-5p* (Fig. 7b).

## Discussion

Colorectal cancer remains one of the most commonly diagnosed cancers worldwide, and a leading cause of cancer related deaths, posing an international burden which is anticipated to worsen according to future predictions<sup>20</sup>. With this knowledge, studies into alternative methods of identifying susceptibility, early diagnosis and treatment is imperative to those affected by CRC. Exploring new avenues through widening our knowledge on the roles played by non-coding RNAs, including miRNA and lncRNA in CRC etiology and progression provides great prospects in improving survival rates of CRC patients. We performed RNA-Seq and computation analyses to characterize the altered mRNA, miRNA and lncRNA expression in CRC compared to NT, as well as identifying the affected signaling networks and predicting the interaction between differentially expressed miRNAs and mRNAs, and miRNAs and lncRNAs in CRC.

The most upregulated mRNAs in CRC according to our cohort includes *MMP3*, *CLDN2*, *S100P*, and *KRT17*. In a recent study on colon adenocarcinoma (COAD), the most common type of CRC, Zeng and colleagues looked into transcriptome data from the TCGA COAD dataset and identified *MMP3*, in addition to several other genes, as prognostic factors for patients with COAD<sup>21</sup> with functional studies showing improved viability of colon cancer cell lines and decreased apoptosis rate<sup>22</sup>. Other studies have also implicated *MMP3* in resistance mechanisms, where expression levels of *MMP3*, *MMP9* and *MMP14*, as well as some cytokines was increased in drug-resistant CRC<sup>23</sup>.

Much evidence has also shown the role played by claudin-2 (*CLDN2*) in breast and CRC<sup>24–27</sup>. Claudins are a family of proteins making up the most important components of tight junctions. Whereas one study shows the upregulation of *CLDN1* and *CLDN2* in CRC, other CLDNs such as *CLDN5*, 7, 8, and 23 were down-regulated, indicating a wide range of roles played by these proteins in CRC. Others such as *CLDN11*, *CLDN12*, and *CLDN23* were associated with longer overall survival<sup>27</sup>. The expression of *KRT17* transcript and *KRT17* protein as a marker were also significantly associated with poor relapse-free survival (RFS) in stage II CRC<sup>28,29</sup>. Nonetheless, our data highlighted a central role for TP53 in CRC<sup>30</sup>. Signaling networks affected by these alternations in mRNA in CRC include ERBB2, RABL6, FOXM1, and NFKB networks, some of which we have previously reported in breast<sup>31,32</sup> and CRC<sup>19,32</sup>.

Aside from protein coding RNA transcripts, our data unraveled multiple dysregulated lncRNAs in CRC. Upregulated lncRNA *LINC01605* in our data, otherwise known as *LincDUSP*, has been found to regulate cell cycle genes to resist apoptosis in a study by Forrest et al., where the knockdown of *lincDUSP* in colon tumor cell lines increased the accumulation of cells in early S-phase, indicating increased DNA damage response induction<sup>33</sup>. In a different study, over 80% of CRC samples showed upregulated expression of *LINC01605*. Depletion of *LINC01605* led to suppression of proliferation, migration and invasion ability of CRC cells in vitro. Furthermore, our data identified *LINC01605* to be involved in the miR-3960/SOX11 regulatory axis, which was confirmed experimentally by Hu et al.<sup>34</sup>. *H19* was the most upregulated lncRNA in CRC in our study. Upregulated *H19* correlated with poor prognosis and induces epithelial-to-mesenchymal transition (EMT) in CRC. The knock down of hnRNPA2B1, directly bound to *H19*, attenuated CRC cells ability to migrate and invade<sup>35</sup>. Furthermore, single-nucleotide polymorphisms (SNPs) identified in the promoter region of *H19* significantly increased the risk of CRC<sup>36</sup>. *H19* lncRNA also acts as a sponge for other miRNAs such as miR-22-3P involved in reduced HDAC2 expression, which strongly promoted CRC lung metastasis<sup>37</sup>. Multiple studies highlight a role for *ZFAS1* in CRC tumorigenesis, where *ZFAS1* silencing reduced CRC cell line migration and invasion ability through miR-34b direct interaction with *ZFAS1* 3' untranslated region (3'UTR)<sup>38</sup>. The *ZFAS1/DDX21/POLR1B* signaling regulation axis may also be a new biomarker for targeted CRC treatment, where *ZFAS1* knockdown dramatically reduced cancer cell properties and increase apoptosis. This inhibitory effect could be reversed by *DDX21* overexpression<sup>39</sup>. Expanding future studies to include the effects of snoRNAs in CRC is also relevant. LncRNA *ZFAS1* activities extend to rRNA 2'-O-methylation (Me) in CRC initiation and development through oncogenic upregulation of NOP58 and SNORD12C/78 expression in CRC<sup>37</sup>.

Other lncRNAs highlighted in this study are novel transcripts with limited annotations including *AL355312.4–201*, simply referred to as antisense to LRP11. Such lncRNAs provide us with further opportunities to data mine for future studies; exploring their biological roles in CRC as well as their associations and interactions with other mRNAs and miRNAs.



A recent study shows the lncRNA *MCF2L-AS1* to provoke proliferation, invasion and glycolysis of CRC via the crosstalk with miR-874-3p/FOXM1 signaling axis<sup>40</sup>, highlighting the importance of altered interactions between mRNAs, miRNAs and lncRNAs in CRC. In addition to this, lncRNA *SNHG16* was found to promote CRC cell proliferation, migration, and epithelial-mesenchymal transition through miR-124-3p/MCP-1<sup>41</sup>. Our data identified several novel lncRNA transcripts including one of the most significantly upregulated, *SNHG5*, while *SNHG16* was found to be part of a regulatory circuit involving *miR-124-3p*, we found *SNHG5* to interact with *miR-132-3p* (Fig. 7a), which has previously been reported to affect cell proliferation, metastasis and migration of CRC through regulating interaction with CREB5<sup>42</sup>. We identified a number of other different interactions of *SNHG5* with miRNAs including *hsa-let-7c-5p*, *hsa-miR-363-3p*, *hsa-miR-125b-2-3p*, *hsa-miR-214-3p*, *hsa-miR-9-5-p*, *hsa-miR-132-3p*, *hsa-miR-29c-3p*, and *hsa-miR-10a-5p*, amongst others, while other groups have also reported on the significance of *SNHG5* in CRC<sup>42</sup>. Elevated expression of the downregulated *miR-let-7c* in CRC was found to be associated with higher disease control rate (DCR) in metastatic CRC patients treated with anti-EGFR mAbs<sup>43,44</sup>. The upregulation of *miR-let-7c* via the silencing of *Lin28* in another study, promoting apoptosis in CRC cells, indicating the *Lin28/let-7c* axis as a potential route for novel therapeutic target in CRC<sup>45</sup>. *Let-7c* is reported to also function as a metastasis suppressor by targeting MMP11 and PBX3 in CRC, with its downregulation being significantly associated with metastases, advanced TNM stages and poor survival of CRC patients<sup>46</sup>. Our data also records the significant downregulation of lncRNA *CDKN2B-AS1*, which was also found to be aberrantly expressed and is suggested as an optimal diagnostic lncRNA biomarker for COAD by Huang et al.,<sup>47</sup>. *CDKN2B-AS1* has also been confirmed to be downregulated in another study on CRC staging and progression<sup>48</sup>.

Emerging data, including our study, have served to solidify the imperative role played by mRNAs, miRNAs, and lncRNAs, and their resulting interactions within networks and axes as a result of their aberrant expression in CRC. While we have highlighted several aberrantly expressed transcripts described in literature, we have also identified multiple novel transcripts and have therefore identified a number of interactions that warrant further investigation, which could serve as potential biomarkers, giving us a clearer picture on how mRNA and ncRNA play a role in the etiology CRC.

## Materials and methods

**Ethics statement.** All patients included in the study were treatment-naïve and provided a written informed consent prior to sample collection. Tumor tissues (TT) and adjacent normal tissues (NT) taken away from tumor margins, were cut from freshly resected tissues by pathologists. The study was performed under ethics committee approvals from Hamad Medical Corporation, Doha, Qatar (Protocol no. MRC-02-18-012) and Qatar Biomedical Research Institute, Doha, Qatar (Protocol no. 2017-006). The characteristics of patients included in this study are provided in Table 1. All experiments were performed in accordance with the ethical principles of the Declaration of Helsinki.

**Sample preparation, RNA extraction, and small RNA library preparation.** Colorectal cancer tissue and their non-tumor adjacent tissue specimens were cut into small pieces then snap-frozen and were stored in liquid nitrogen for use in subsequent experiments. Tissues were homogenized using Tissue Homogenization Set from Bioneer Corporation (Daejeon, South Korea) according to manufacturer instructions. Homogenized tissue lysates were used to extract RNA using the miRNeasy Mini Kit- QIAGEN (Hilden, Germany) according to manufacturer instructions. In brief, tissue were lysed and kept at room temperature (15–25 °C) for 5 min, chloroform is added, shaking vigorously at room temperature for additional 2–3 min. Samples were centrifuged for 15 min at 12,000×g at 4 °C. The aqueous phase was washed with 100% ethanol and added into an RNeasy Mini column in a 2 ml collection tube. RWT and RPE buffers are added, then finally RNA is eluted in water in a fresh tube. The concentration and purity of extracted RNA were measured using NanoDrop 2000c (Thermo Scientific, Waltham, MS, USA). The quantity and quality of extracted RNA was checked using the Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA, USA) and the Agilent 2100 Bioanalyzer (Agilent Technologies, USA) as per the manufacturer's instructions. Samples with RNA Integrity Number (RIN) >9 were used for subsequent analyses.

**TruSeq stranded total RNA library preparation.** The TruSeq Stranded Total RNA Library Prep Gold kit (Cat #: 20020598) from Illumina was used for total RNA library preparation following the manufacturer's protocol. Briefly, 100 ng of total RNA extracted from TT and NT samples was subjected to rRNA depletion and then to fragmentation. The first-strand cDNA synthesis was performed with random hexamers and SuperScript II Reverse Transcriptase (Cat#: 18064014) from ThermoFisher Scientific. The second cDNA strand synthesis was performed by substitution of dTTP with dUTP. The double-stranded cDNA is then end-repaired and adenylated. Barcoded DNA adapters were ligated to both ends of the double-stranded cDNA and then amplified. The libraries quality was checked on an Agilent 2100 Bioanalyzer system and quantified using Qubit 2.0 fluorometer (Invitrogen).

**QIaseq small RNA (miRNA) library preparation.** For small RNA library preparation, 100 ng of RNA was subjected to 3' ligation followed by 5' ligation and reverse transcription. The next steps included QIaseq miRNA NGS (QMN) bead preparation, cDNA cleanup, and finally library amplification using tube indices, QIaseq miRNA NGS 96 Index IL kit (QIAGEN, Hilden, Germany). The yield of cDNA libraries was quantified using Qubit dsDNA HS assay kit (Invitrogen) and size distribution of the cDNA libraries were determined using the Agilent 2100 Bioanalyzer DNA1000 chip (Agilent Technologies). Libraries were subsequently pooled and were subjected to illumine sequencing at QBRI genomic core.

**Illumina sequencing and bioinformatics analyses.** TruSeq Stranded Total RNA libraries were pooled, clustered on a cBot platform, and sequenced on an Illumina HiSeq 4000 at a minimum of 50 million paired end reads ( $2 \times 75$  bp) per sample. Similarly, miRNA libraries were pooled and were sequenced using the same platform (5 million reads per sample). Data were subsequently demultiplexed using illumina suggested protocols and the resulting FASTQ files were used for further analysis. For total RNA-Seq analysis, FASTQ files were subsequently pseudo aligned to the Gencode Release 33 index (mRNA and lncRNA) and reads were subsequently counted using KALLISTO 0.42.1 as previously described<sup>10,19,49</sup>. Normalized Transcripts Per Million (TPM) expression values were subsequently subjected to differential expression analysis using 2.0-fold change and  $< 0.05$  p-value cut-off. Transcripts with raw expression values  $< 1.0$  TPM were excluded from the analysis. Hierarchical clustering was performed using cosine for columns and cosine for rows as described before<sup>10,19,50</sup>. Volcano plots was used to illustrate most differentially expressed genes (log two fold change) vs  $-\log_{10}$  p value.

For microRNA analysis, FASTQ files were subjected to small-RNA quantification and differential expression analysis in CLC genomics workbench 20.2 using built-in workflow. Initially, raw expression data were mapped to the miRbase v22 reference genome. Read counts of mature miRNA were then normalized using the trimmed mean of M values (TMM) normalization method. Mature miRNA expression values were then subjected to differential analysis as described above. Low abundance miRNAs with expression values  $< 5.0$  count per million (CPM) were excluded from the analysis. The microRNA Target Filter in IPA (using the default database) was subsequently employed to identify miRNA-mRNA networks based on differentially expressed mRNAs and miRNAs from current study. This bioinformatics approach provides insights into the biological effects of microRNAs, using corresponding mRNA expression data as well as experimentally validated interactions from several databases, and the predicted microRNA-mRNA interactions from TargetScan. Only highly predicted and experimentally validated interactions were included in the final analysis. DIANA-LncBase v3 was utilized to identify miRNA-lncRNA interactions. Cytoscape 3.8.1 was used to construct miRNA-mRNA and miRNA-lncRNA networks. Expression values from whole transcriptome RNA-Seq and miRNA-Seq are provided as supplementary tables 5 and 6, respectively.

Received: 24 February 2021; Accepted: 21 June 2021

Published online: 14 July 2021

## References

- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **70**(1), 7–30 (2020).
- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J. Clin.* **68**(1), 7–30 (2018).
- Marijnen, C. A. *et al.* Acute side effects and complications after short-term preoperative radiotherapy combined with total mesorectal excision in primary rectal cancer: Report of a multicenter randomized trial. *J. Clin. Oncol.* **20**(3), 817–825 (2002).
- Glynn-Jones, R. & Debus, J. Improving chemoradiotherapy in rectal cancer. *Oncologist* **6**(Suppl 4), 29–34 (2001).
- Markowitz, S. D. & Bertagnolli, M. M. Molecular origins of cancer: Molecular basis of colorectal cancer. *N. Engl. J. Med.* **361**(25), 2449–2460 (2009).
- Flanagan, D. J. *et al.* NOTUM from Apc-mutant cells biases clonal competition to initiate cancer. *Nature* (2021).
- Yang, K. *et al.* Rosmarinic acid inhibits migration, invasion, and p38/AP-1 signaling via miR-1225-5p in colorectal cancer cells. *J. Recept. Signal Transduct. Res.* 1–10 (2020).
- Vaghari-Tabari, M. *et al.* MicroRNAs and colorectal cancer chemoresistance: New solution for old problem. *Life Sci.* 118255 (2020).
- Chen, H. *et al.* MicroRNA-200c inhibits the metastasis of triple-negative breast cancer by targeting ZEB2, an epithelial-mesenchymal transition regulator. *Ann. Clin. Lab Sci.* **50**(4), 519–527 (2020).
- Elango, R. *et al.* MicroRNA expression profiling on paired primary and lymph node metastatic breast cancer revealed distinct microRNA profile associated with LNM. *Front. Oncol.* **10**, 756 (2020).
- Bozgeyik, E., Tepe, N. B., & Bozdogan, Z. Identification of microRNA expression signature for the diagnosis and prognosis of cervical squamous cell carcinoma. *Pathol. Res. Pract.* **216**(11), 153159 (2020).
- Ratert, N. *et al.* miRNA profiling identifies candidate miRNAs for bladder cancer diagnosis and clinical outcome. *J. Mol. Diagn.* **15**(5), 695–705 (2013).
- Wozniak, M., Mielczarek, A. & Czyz, M. miRNAs in melanoma: Tumor suppressors and oncogenes with prognostic potential. *Curr. Med. Chem.* **23**(28), 3136–3153 (2016).
- Vishnubalaji, R. *et al.* MicroRNA-320 suppresses colorectal cancer by targeting SOX4, FOXM1, and FOXQ1. *Oncotarget* **7**(24), 35789–35802 (2016).
- Zhang, H. *et al.* Long non-coding RNA: A new player in cancer. *J. Hematol. Oncol.* **6**, 37 (2013).
- Li, W., He, Y. & Cheng, Z. Long noncoding RNA XIST knockdown suppresses the growth of colorectal cancer cells via regulating microRNA-338-3p/PAX5 axis. *Eur. J. Cancer Prevent.* (2020).
- Xiao, Y., Yurievich, U. A. & Yosypovych, S. V. Long noncoding RNA XIST is a prognostic factor in colorectal cancer and inhibits 5-fluorouracil-induced cell cytotoxicity through promoting thymidylate synthase expression. *Oncotarget* **8**(47), 83171–83182 (2017).
- Vishnubalaji, R. *et al.* Genome-wide mRNA and miRNA expression profiling reveal multiple regulatory networks in colorectal cancer. *Cell Death Dis.* **6**, e1614 (2015).
- Shaath, H. *et al.* Transcriptomic analyses revealed systemic alterations in gene expression in circulation and tumor microenvironment of colorectal cancer patients. *Cancers (Basel)* **11**(12) (2019).
- Araghi, M. *et al.* Global trends in colorectal cancer mortality: Projections to the year 2035. *Int. J. Cancer* **144**(12), 2992–3000 (2019).
- Zeng, C. & Chen, Y. HTR1D, TIMP1, SERPINE1, MMP3 and CNR2 affect the survival of patients with colon adenocarcinoma. *Oncol. Lett.* **18**(3), 2448–2454 (2019).
- Bufu, T. *et al.* Celestrol inhibits colorectal cancer cell proliferation and migration through suppression of MMP3 and MMP7 by the PI3K/AKT signaling pathway. *Anticancer Drugs* **29**(6), 530–538 (2018).
- Lai, M. *et al.* Hsa\_circ\_0079662 induces the resistance mechanism of the chemotherapy drug oxaliplatin through the TNF-alpha pathway in human colon cancer. *J. Cell Mol. Med.* **24**(9), 5021–5027 (2020).
- Tabaries, S. *et al.* Lyn modulates Claudin-2 expression and is a therapeutic target for breast cancer liver metastasis. *Oncotarget* **6**(11), 9476–9487 (2015).
- Mezheyeuski, A. *et al.* Treatment-related survival associations of claudin-2 expression in fibroblasts of colorectal cancer. *Virchows Arch.* **472**(3), 395–405 (2018).



26. Shi, Y. J. *et al.* Toll-like receptor 4 (TLR4) deficiency aggravates dextran sulfate sodium (DSS)-induced intestinal injury by down-regulating IL6, CCL2 and CSF3. *Ann. Transl. Med.* **7**(23), 713 (2019).
27. Cherradi, S. *et al.* Claudin gene expression profiles and clinical value in colorectal tumors classified according to their molecular subtype. *Cancer Manag. Res.* **11**, 1337–1348 (2019).
28. Ujiie, D. *et al.* KRT17 as a prognostic biomarker for stage II colorectal cancer. *Carcinogenesis* **41**(5), 591–599 (2020).
29. Kim, C. Y. *et al.* Proteomic analysis reveals overexpression of moesin and cytokeratin 17 proteins in colorectal carcinoma. *Oncol. Rep.* **27**(3), 608–620 (2012).
30. Michel, M. *et al.* The role of p53 dysfunction in colorectal cancer and its implication for therapy. *Cancers (Basel)* **13**(10) (2021).
31. Vishnubalaji, R. *et al.* Integrated transcriptome and pathway analyses revealed multiple activated pathways in breast cancer. *Front. Oncol.* **9**, 910 (2019).
32. Elango, R. *et al.* Concurrent targeting of BMI1 and CDK4/6 abrogates tumor growth in vitro and in vivo. *Sci. Rep.* **9**(1), 13696 (2019).
33. Forrest, M. E. *et al.* Colon cancer-upregulated long non-coding RNA lincDUSP regulates cell cycle genes and potentiates resistance to apoptosis. *Sci. Rep.* **8**(1), 7324 (2018).
34. Hu, S. S. *et al.* LINC01605 regulates proliferation, migration and invasion of colorectal cancer cells via targeting miR-3960/SOX11. *Eur. Rev. Med. Pharmacol. Sci.* **25**(3), 1322–1329 (2021).
35. Zhang, Y. *et al.* Long non-coding RNA H19 promotes colorectal cancer metastasis via binding to hnRNPA2B1. *J. Exp. Clin. Cancer Res.* **39**, 141. <https://doi.org/10.1186/s13046-020-01619-6> (2020).
36. Qin, W. *et al.* Functional polymorphisms of the lncRNA H19 promoter region contribute to the cancer risk and clinical outcomes in advanced colorectal cancer. *Cancer Cell Int.* **19**, 215 (2019).
37. Wu, H. *et al.* Long noncoding RNA ZFAS1 promoting small nucleolar RNA-mediated 2'-O-methylation via NOP58 recruitment in colorectal cancer. *Mol. Cancer* **19**, 95. <https://doi.org/10.1186/s12943-020-01201-w> (2020).
38. Deng, H. *et al.* ZFAS1 promotes colorectal cancer metastasis through modulating miR-34b/SOX4 targeting. *Cell Biochem. Biophys.* **79**(2), 387–396 (2021).
39. Wang, X. *et al.* Long non-coding RNA ZFAS1 promotes colorectal cancer tumorigenesis and development through DDX21-POLR1B regulatory axis. *Aging (Albany NY)* **12**(22), 22656–22687 (2020).
40. Zhang, Z. *et al.* LncRNA MCF2L-AS1 aggravates proliferation, invasion and glycolysis of colorectal cancer cells via the crosstalk with miR-874-3p/FOXO1 signaling axis. *Carcinogenesis* **42**, 263–271 (2021).
41. Chen, Z.Y. *et al.* LncRNA SNHG16 promotes colorectal cancer cell proliferation, migration, and epithelial-mesenchymal transition through miR-124-3p/MCP-1. *Gene Ther.* (2020).
42. Zhang, M. *et al.* LncRNA SNHG5 affects cell proliferation and migration of colorectal cancer through regulating miR-132-3p/CREB5. *Cancer Biol. Ther.* **20**(4), 524–536 (2019).
43. Fiala, O. *et al.* Association of miR-125b, miR-17 and let-7c dysregulations with response to anti-epidermal growth factor receptor monoclonal antibodies in patients with metastatic colorectal cancer. *Cancer Genomics Proteomics* **17**(5), 605–613 (2020).
44. Cappuzzo, F. *et al.* MicroRNA signature in metastatic colorectal cancer patients treated with anti-EGFR monoclonal antibodies. *Clin. Colorectal Cancer* **13**(1), 37–45 e4 (2014).
45. Zhang, H. *et al.* Silencing Lin28 promotes apoptosis in colorectal cancer cells by upregulating let7c targeting of antiapoptotic BCL2L1. *Mol. Med. Rep.* **17**(4), 5143–5149 (2018).
46. Han, H. B. *et al.* Let-7c functions as a metastasis suppressor by targeting MMP11 and PBX3 in colorectal cancer. *J. Pathol.* **226**(3), 544–555 (2012).
47. Huang, W. *et al.* Identification of long noncoding RNAs biomarkers for diagnosis and prognosis in patients with colon adenocarcinoma. *J. Cell Biochem.* **120**(3), 4121–4131 (2019).
48. Wu, M. *et al.* Comprehensive analysis of the expression profiles of long non-coding RNAs with associated ceRNA network involved in the colon cancer staging and progression. *Sci. Rep.* **9**(1), 16910 (2019).
49. Bray, N. L. *et al.* Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**(5), 525–527 (2016).
50. Emig, D. *et al.* AltAnalyze and DomainGraph: Analyzing and visualizing exon expression data. *Nucleic Acids Res.* **38**(web server issue), W755–W762 (2010).

## Acknowledgements

This work was supported by Qatar Biomedical Research Institute (QBRI) start-up fund (grant no. QB13) for Nehad M. Alajez.

## Author contributions

H.S. and S.M.T. performed experiments, M.A.N. provided clinical materials and data, E.E. and N.M.A. conceived study. H.S. wrote initial manuscript draft. N.M.A. obtained funding and finalized manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-93531-y>.

**Correspondence** and requests for materials should be addressed to N.M.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.





**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021