

Transcriptome and Network Dissection of Microsatellite Stable and Highly Instable Colorectal Cancer

Vahid Akbari^{1*}, Marzieh Kallhor², Behrouz Mollashahi³, Abolfazl Movafagh^{4*}

Abstract

Background: Colorectal cancer (CRC) is one of the most common cancers worldwide with high number of mortality every year. Microsatellite instability (MSI) is a considerable feature of CRC which affects prognosis and treatment. High level of MSI or MSI-high (MSI-H) colorectal cancer has better prognosis and immunotherapy response, while microsatellite stable (MSS) CRC has better response to 5-fluorouracil (5-FU)-based chemotherapy. More studies are needed, specifically on MSS CRC which has worse prognosis, to further reveal biological differences and similarities between MSS and MSI colorectal cancer, which may equip us with the knowledge to develop more promising therapeutic approaches to target both types or be more effective for each type. **Methods:** We aimed to find affected biological processes and their regulators in both type, MSS and MSI-H, of CRC; as well as reveal specific ones in each type. We applied meta- and network analysis on freely available transcriptome data in MSS and MSI-H colorectal cancer from gene expression omnibus (GEO) database to detect common differentially expressed genes (DEGs) and critical biological processes and predict their most significant regulators. **Results:** Our results demonstrate considerable up and downregulation in cell cycle and lipid catabolism processes, respectively; and introduced MYC and *FOXMI* as two central and up-stream regulators of DEGs in both type of CRC. Chemokine-mediated processes displayed up-regulation in MSI-H type, while metastasis-related processes showed more activation in MSS CRC. Additionally, *DACHI* and *TP53* were detected as two important transcription factors that differentially expressed just in MSS and MSI-H, respectively. **Conclusion:** Our results can explain why MSI and MSS CRC display different immunotherapy response, prognosis, and metastasis feature. Moreover, our predicted up-stream regulators in the regulatory networks may be promising therapeutic targets.

Keywords: Biological processes- transcription factor- *FOXMI*- *DACHI*- *TP53*

Asian Pac J Cancer Prev, 20 (8), 2445-2454

Introduction

Colorectal cancer is one of the major causes of cancer mortality in the world with an estimation of more than 1.8 million new colorectal cancer cases and 881,000 deaths in 2018 (Bray et al., 2018). CRC usually begins as an abnormal growth on the rectum or colon, known as polyps. Development of polyps into cancerous status can lead to the formation of a tumor on the wall of the colon or rectum (Kinzler and Vogelstein, 1996). Sporadic CRC comprises majority of cases, while ~30% of cases are familial type. Somatic alterations in mismatch repair (MMR) genes, especially MSH2 and MLH1, are frequent in sporadic CRC and their germline mutations associated with lynch syndrome, the most common hereditary CRC (Li and Martin, 2016).

Microsatellite instability (MSI), variations in the length of tandem nucleotide repeats (microsatellite

fragments), is a result of mutations in MMR genes. Although the most attention to MSI is directed to familial CRC, only about 3% of all CRC cases come from the families with Lynch syndrome (Hampel et al., 2005, 2008). Approximately 12%–17% of whole colorectal tumors have microsatellite instability and majority of them are sporadic (Ward et al., 2001; Popat et al., 2005). The main criteria for classification a tumor as MSS, low level of MSI or MSI-low (MSI-L), or MSI-H is based on five microsatellites recommended by the National Cancer Institute (NCI) including BAT26, D17S250, D2S123, BAT25 and D5S346. If two or more of the five microsatellite sequences have been mutated, tumor will be considered as MSI-H and if just one of the microsatellites has been mutated it considered as MSI-L. MSS refers to tumor without MSI in any of the markers (Kurzwski et al., 2004). It has been demonstrated that there is a correlation between MSI-H and colorectal cancer prognosis; and the

¹Department of Medical Genetics, Faculty of Medicine, University of British Columbia, Vancouver, BC Canada, ²Department of Medical Genetics, Faculty of Medical Sciences, Tarbiat Modares University, ³Department of Cellular-Molecular Biology, Faculty of Biological Sciences and Technologies, ⁴Department of Medical Genetics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran. *For Correspondence: Movafagh.a@sbm.ac.ir, vahid.akbari@alumni.ubc.ca. Vahid Akbari, Marzieh Kallhor and Behrouz Mollashahi have equal contribution in this study.

existence of high-frequency MSI predicts a relatively favorable outcome and smaller likelihood of metastasis in CRC (Gryfe et al., 2000; Popat et al., 2005; Sagaert et al., 2014). Microsatellite instability gives distinguished features to the colorectal tumors; for instance, this type of tumors have a tendency to appear in the proximal colon, infiltration across the lymphocytes, and also they have featured as weakly differentiated, signet ring or mucinous appearance (Thibodeau et al., 1993). Moreover, MSS and MSI-H types of CRC are different in the response to chemotherapeutics and treatments. Tumors with MSI-H phenotype have better response to immunotherapy, while MSS and MSI-low Tumors display better response to 5-fluorouracil (5-FU)-based chemotherapy (Peltomäki et al., 1993; Kim et al., 2016; Overman et al., 2018).

Regarding the differences between MSI-H and MSS colorectal cancer in some aspects such as prognosis and treatment, there need to be more studies to reveal discrepancies and similarities between these two class of CRC in biological pathways and their regulation, and interplay between MMR and other biological processes. When we armed with this knowledge, we may be able to develop more therapeutic approaches which can be effective on both type of CRC, or develop more powerful drugs which can effectively target each type. In the current study, we applied meta- and network analysis on freely available expression data of MSS and MSI-H colorectal cancer patients deposited in gene expression omnibus (GEO) database to find common biological processes and transcription factors for their regulation in MSS and MSI-H CRC. Moreover we mined these expression data to detect differences between these two types of CRC in transcriptomics and regulatory context.

Materials and Methods

Data Collection

We searched Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>) deeply using the following key terms “Colorectal cancer”, “MSI CRC”, and “MSS CRC”. We included studies with following criteria: on human tissue samples, at least three samples, sporadic CRC, with control samples from CRC individuals, and studies that used array platforms which cover majority of human genes. Finally, 7 appropriate datasets which contain expression data of 402 MSS patients, 101 MSI-H patients and 113 normal controls, were found (Watanabe et al., 2006; Vilar et al., 2009; Hong et al., 2010; Ågesen et al., 2011, 2012; Gröne et al., 2011; Sveen et al., 2011; Cordero et al., 2014).

Differential Expression (DE) analysis and Common DEGs Detection

Log₂ transformed gene expression matrix of each study was obtained using Biobase, GEOquery, and limma R packages (Davis and Meltzer, 2007; Huber et al., 2015; Ritchie et al., 2015). Gene expression matrixes were annotated using the approved gene symbols, entrez gene ID or RefSeq ID from HUGO Gene Nomenclature Committee (HGNC, <http://www.genenames.org/>) (Yates et al., 2016) and imported into the INMEX web tool (Xia et

al., 2013) for differential expression analysis. Expression tables were quantile normalized and DE analysis was performed based on linear model of limma algorithm (Ritchie et al., 2015) in INMEX. The results of DE analysis were filtered based on P value <0.05. In order to obtain common DEGs in majority of studies, the filtered results from DE analysis of all studies were merged. Finally, those genes that consistently up or down-regulated in at least 2 out of 3 (for MSI-H vs control) and 3 out of 4 (for MSS vs control) studies with |fold change (FC)|>1.5 in each study were considered as common DEGs. In MSS vs MSI-H comparison, genes that consistently up or down-regulated in at least 3 out of 6 studies with |FC|>2 or 4 out of 6 studies with |FC|>1.5 in each study were considered as common DEGs.

Subsequently, Cluster 3 and Java TreeView (Eisen et al., 1998; Saldanha, 2004) were used for cluster analysis, gene arrangement based on similarity in expression, and heatmap demonstration of common DEGs. We applied correlation (uncentered) and centroid linkage analysis for cluster analysis.

Gene Ontology (GO) and Pathway analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) (Sherman and Lempicki, 2009) with annotation from Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2017) was used for pathway analysis. The list of common DEGs was imported into the DAVID and the results of KEGG pathway analysis were extracted. In order to address the most important terms, resulted pathways were filtered based on P value<0.05.

Gene ontology enrichment analysis was performed using the BiNGO plugin of Cytoscape (Shannon et al., 2003; Maere et al., 2005). At first, the most updated ontology and gene association data were downloaded from gene ontology consortium (Consortium, 2016) and imported into the BiNGO for GO analysis. Common DEGs were imported into the Cytoscape and gene ontology analysis was performed based on biological process. Overrepresentation with the hypergeometric test was used and Bonferroni Family-Wise Error Rate (FWER) correction <0.05 was considered as a threshold in order to address the most important terms in GO analysis.

Gene regulatory network construction and analysis

To find and predict master regulators of common DEGs in each comparison, we performed regulatory network analysis. Gene regulatory networks was constructed using transcription factors and visualized by Cytoscape. To examine TF regulatory network in a comprehensive manner we used TF binding sites information from two extensive databases which include transcription binding sites profile resulted from various techniques including ChIP enrichment analysis (ChEA) database (mainly comprised of TF binding sites obtained from ChIP-chip, ChIP-seq, ChIA-PET, and DamID techniques) (Lachmann et al., 2010) and ENCODE TF ChIP-seq database (Transcription factor DNA-binding by ChIP-seq) (Consortium, 2004). Gene expression data were mapped to the TF-binding sites from mentioned databases. TFs

with P value<0.05 and FC>1.5 in at least 2 out of 3 (for MSI-H vs control) and 3 out of 4 (for MSS vs control) studies with the same expression pattern were considered as differentially expressed transcription factors (DE-TFs). For MSS vs MSI-H, TFs with P value<0.05 and FC>2 in at least 3 out of 6 or FC>1.5 in at least 4 out of 6 studies with the same expression pattern were considered as DE-TFs.

In order to obtain hub regulators in regulatory networks degree parameter for the directed network was calculated using CentiScaPe plugin of Cytoscape (Scardoni et al., 2014). To rank DE-TFs, we used out-degree which indicates the number of target genes for each TF. As well as, core regulatory networks were extracted to investigate regulatory interactions between regulators of DEGs and find upstream DE-TFs.

Protein-protein interaction (PPI) network construction and analysis

Protein-protein interactions drive biological processes. They are crucial for all intra and extracellular functions and the technologies to analyze these interactions are widely used throughout the diverse fields of biological sciences. To dissect the biological meaning of DEGs at the protein level and find DE-TFs and DEGs with the highest number of PPIs, we also constructed PPI networks for common DEGs. We used information derived from several databases to examine PPI networks at a more comprehensive level. For this aim, PPI networks were constructed using the most updated information from General Repository for Interaction Datasets (BioGRID) (Stark et al., 2006), The Molecular INTeraction Database (MINT) (Licata et al., 2011), Human Protein Reference Database (HPRD) (Keshava Prasad et al., 2008) and Search Tool for the Retrieval of Interacting Genes/Proteins (String) (Jensen et al., 2008). The interactions from string were filtered based on interaction score ≥ 0.7 which means high confidence. We applied topology analysis to find densely connected regions (clusters) in the PPI networks which may represent significant molecular complexes. Topology analysis was performed using MCODE algorithm (Bader and Hogue, 2003) to find crucial protein complexes from the PPI networks,

subsequently GO analysis of found protein complexes was performed using BinGO.

Results

To evaluate the gene expression and find common and also unique biological processes and their regulators in both MSS and MSI-H colorectal cancer we applied comprehensive transcriptome and network mining of CRC expression data deposited in GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). We searched GEO database to find appropriate studies on human colorectal cancer tissues. Finally, seven studies (Table 1) have been found from our search in GEO (Watanabe et al., 2006; Vilar et al., 2009; Hong et al., 2010; Ågesen et al., 2011, 2012; Gröne et al., 2011; Sveen et al., 2011; Cordero et al., 2014). Complete set of samples from original studies which used in this study is provided in Supplementary file 1. We used these studies to find significant DEGs and their regulators for colorectal cancer. We detected 1401 DEGs (764 downregulated and 637 upregulated genes) in MSS vs control cases (Figure 1a, Supplementary file 2), 2011 DEGs (1,083 downregulated and 928 upregulated genes) in MSI-H vs control cases (Figure 2a, Supplementary file 3), and 188 DEGs (65 downregulated and 123 upregulated genes) in MSS vs MSI-H cases (Figure 3a, Supplementary file 4).

Regarding GO analyses of DEGs (Figures 1b and 2b), up-regulated genes in both MSS and MSI-H versus control cases were involved in biological processes related to cell cycle and cell division, while down-regulated genes were mainly involved in processes related to lipid and fatty acid metabolism, specifically catabolism and oxidation of fatty acids. The results of pathway analyses also highlighted the results of GO analyses (Figures 1c and 2c). Based on pathway analysis, in both comparisons (MSS and MSI-H versus control cases) cell cycle related pathways were the most overexpressed pathways, while fatty acid degradation pathway was among the most down-regulated ones.

In order to reveal differences between MSS and MSI cases we performed comparison of expression data

Table 1. Microarray Datasets Used in Our Survey

Experiment	Organism	Comparison	Accession Number	Chip Type
Sveen A et al, 2011	Human	21 MSI-H* vs 13 normal 110 MSS† vs 13 normal 110 MSS vs 21 MSI-H	GSE24551	Affymetrix Human Exon 1.0 ST Array [transcript (gene) version]
Ågesen et al, 2011	Human	5 MSI-H vs 4 normal 38 MSS vs 4 normal 38 MSS vs 5 MSI-H	GSE25071	ABI Human Genome Survey Microarray Version 2
Ågesen et al, 2012	Human	10 MSI-H vs 2 normal 34 MSS vs 2 normal 34 MSS vs 10 MSI-H	GSE29638	Affymetrix Human Exon 1.0 ST Array [transcript (gene) version]
Cordero et al, 2014	Human	98 MSS vs 98 normal	GSE44076	Affymetrix Human Genome U219 Array
Watanabe et al, 2006	Human	50 MSS vs 33 MSI-H	GSE4554	Affymetrix Human Genome U133 Plus 2.0 Array
Vilar et al, 2009	Human	38 MSS vs 13 MSI-H	GSE11543	Affymetrix Human Full Length HuGeneFL Array
Gröne et al, 2011	Human	34 MSS vs 19 MSI-H	GSE18088	Affymetrix Human Genome U133 Plus 2.0 Array

*, Microsatellite instability-high; †, Microsatellite stable

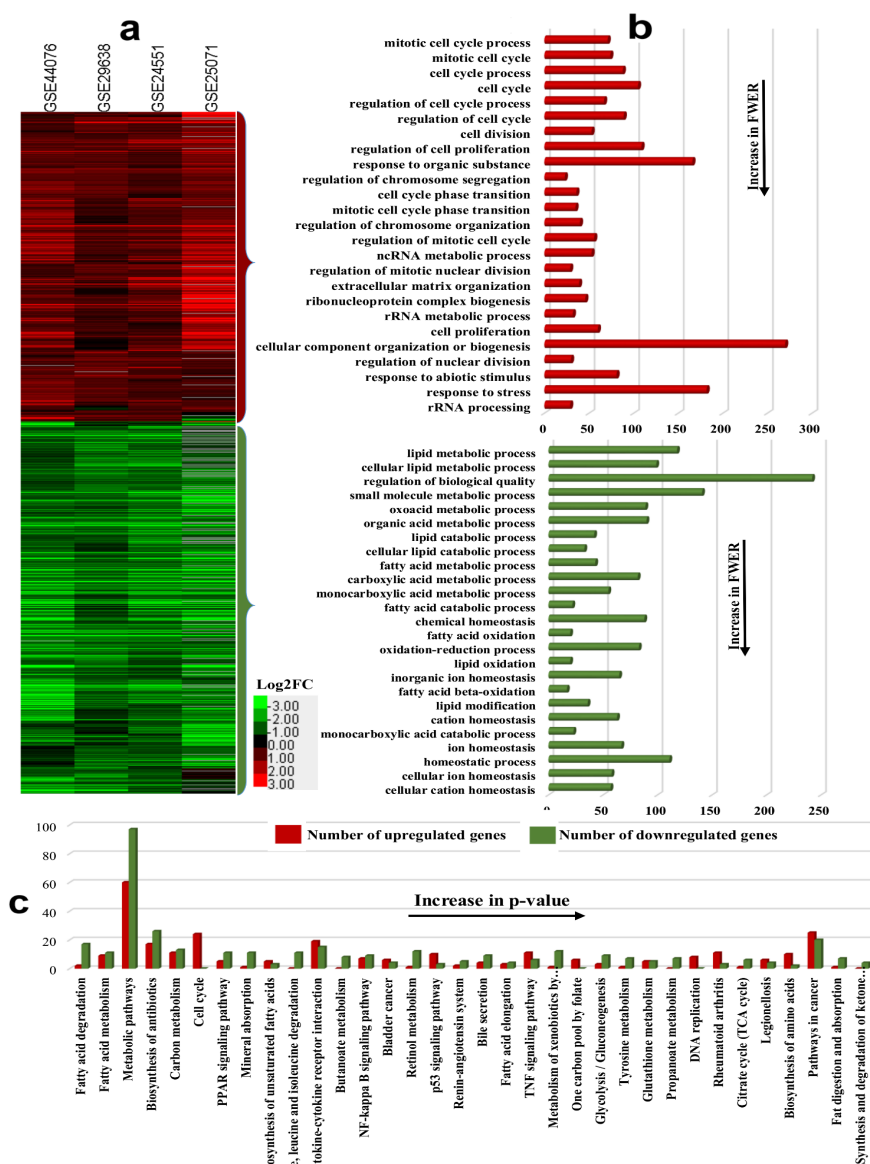


Figure 1. Resulted Heatmap for Common DEGs in MSS vs Control Comparisons (a), and the results of their GO; (b), and pathway analysis; (c), MSS, microsatellite stable; DEG, differentially expressed genes; GO, gene ontology

in MSS sample versus MSI-H cases. Moreover, we compared the resulted DEGs of this comparison with the results of MSS and MSI-H vs controls in order to find direction of changes; for example genes and biological processes that represented downregulation in MSS vs MSI-H comparison may actually up-regulated in MSI-H cases and when we consider MSI-H cases as controls and compare the expression of MSS cases to them (MSS vs MSI-H) we see downregulation. Finally, we found that immune-related processes are more active in MSI-H type of CRC while processes related to metastasis including migratory and vasculature development processes are more overexpressed in MSS cases. As well as, comparison of pathway analysis results of MSS and MSI-H vs controls suggested more up-regulation of MMR pathway in MSI-H CRC.

Mining regulatory network and finding critical regulators of common DEGs

Regulatory networks were constructed by information

from two databases to draw a comprehensive picture of DEGs regulation. We found 24 DE-TFs in MSI-H and 24 DE-TFs in MSS CRC cases (Figure 4a). Twelve DE-TFs were common in both comparisons which among them VDR (vitamin D receptor), KLF4 (Kruppel like factor 4), KLF6, TCF21 (transcription factor 21), JUND, and KAT2B (lysine acetyltransferase 2B) were down-regulated and MYC (MYC proto-oncogene, bHLH transcription), E2F7 (E2F transcription factor 7), BHLHE40 (basic helix-loop-helix family member e40), BATF (Basic Leucine Zipper ATF-Like Transcription Factor), STAT1 (Signal Transducer And Activator Of Transcription 1), and FOXM1 (forkhead box M1) were up-regulated.

On the other hand, we found specific DE-TFs for each condition that significantly up- or down-regulated just in MSS or MSI-H cases. For this aim we considered all found DE-TFs from all comparisons to find the most important ones. TP53 showed up-regulation in MSI-H cases and its expression did not change significantly in any

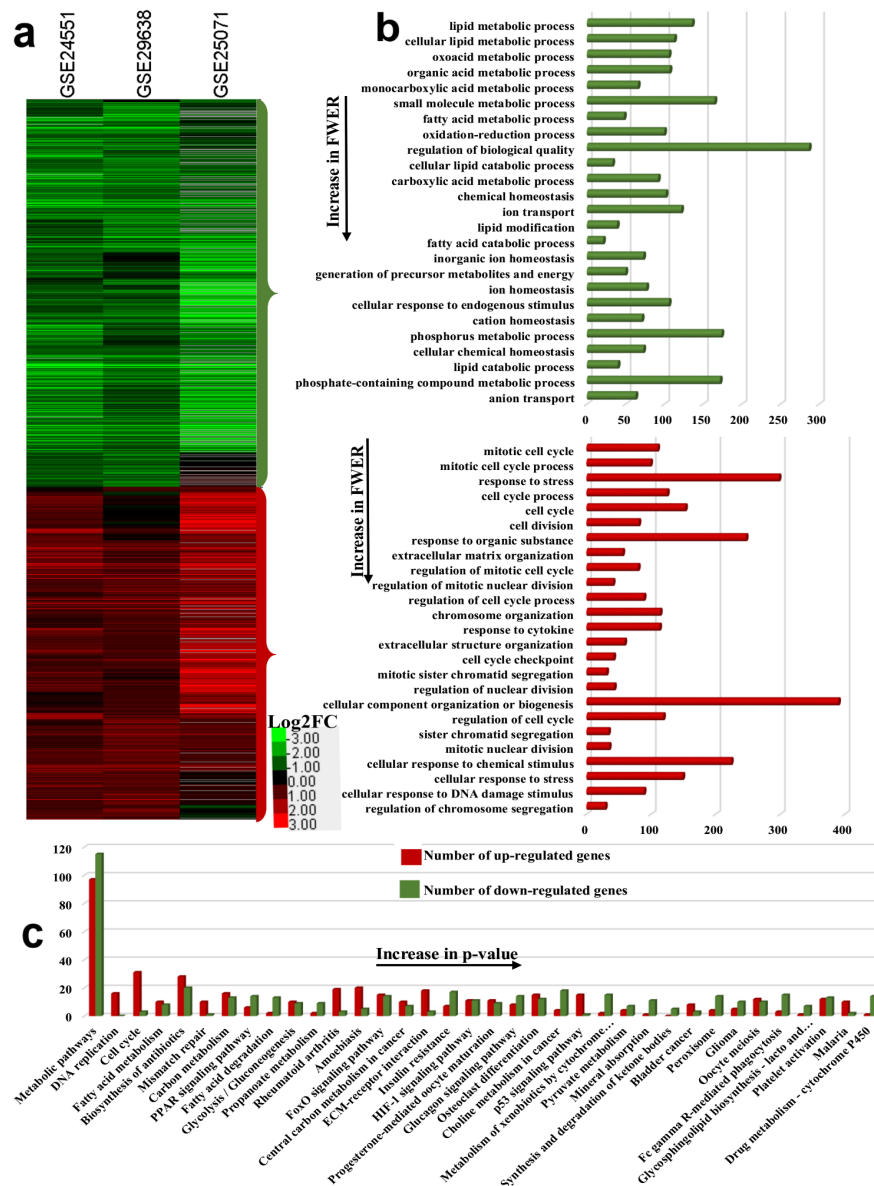


Figure 2. Resulted Heatmap for Common DEGs in MSI-H vs Control Comparisons (a), and the results of their GO; (b), and pathway analysis; (c), MSI-H, microsatellite instable-high; DEG, differentially expressed genes; GO, gene ontology

MSS cases, as well as *TP53* showed down-regulation in 3 out of 6 studies in MSS versus MSI-H comparison with fold change > 1.5 which highlights its overexpression in MSI-H patients. Contrariwise, *DACH1* and *FOXA2* were up-regulated and *FOXA1* was down-regulated in MSS cases and their expression did not change significantly in any MSI-H cases; among them, *FOXA1* showed downregulation pattern in all MSS vs MSI-H cases (in 2 cases with FC > 1.5) and *DACH1* displayed significant upregulation in 4 out of 6 MSS vs MSI-H comparisons.

By analysis the regulatory network of found DEGs from comparing the expression data of MSS versus MSI-H cases, we also found that *CDX2* (Caudal Type Homeobox 2), *ZMYND8* (Zinc Finger MYND-Type Containing 8), and *DACH1* (Dachshund Family Transcription Factor 1 DACH) were up-regulated in the majority of MSS vs MSI-H comparisons. *CDX2* showed downregulation in one of the MSI-H vs control comparisons and the same pattern of expression in others.

DACH1, as we mentioned previously, was significantly upregulated in MSS vs control studies, and *ZMYND8* showed downregulation pattern in all MSI-H vs control cases with FC > 1.5 in one and FC ~ 1.5 in another study. All in all, *TP53* and *DACH1* display exclusive up-regulation in MSI-H or MSS CRC, respectively. The exclusive up-regulation of these two TFs were consistent in all comparisons (MSS vs control, MSI-H vs control, and MSS vs MSI-H).

We applied centrality and core regulatory analysis to find central and upstream DE-TFs. Out-degree (number of targets) parameter was calculated for each node (gene) in the regulatory networks and DE-TFs were sorted based on their number of targets in each comparison (Figures 4, b and c). Moreover, core regulatory networks were extracted to examine regulatory interactions between DE-TFs and find upstream regulators (Figures 4, d and e). These two analyses revealed *MYC*, *JUND*, *FOXM1*, *BHLHE40*, and *VDR* as significant regulators in both types of CRC.

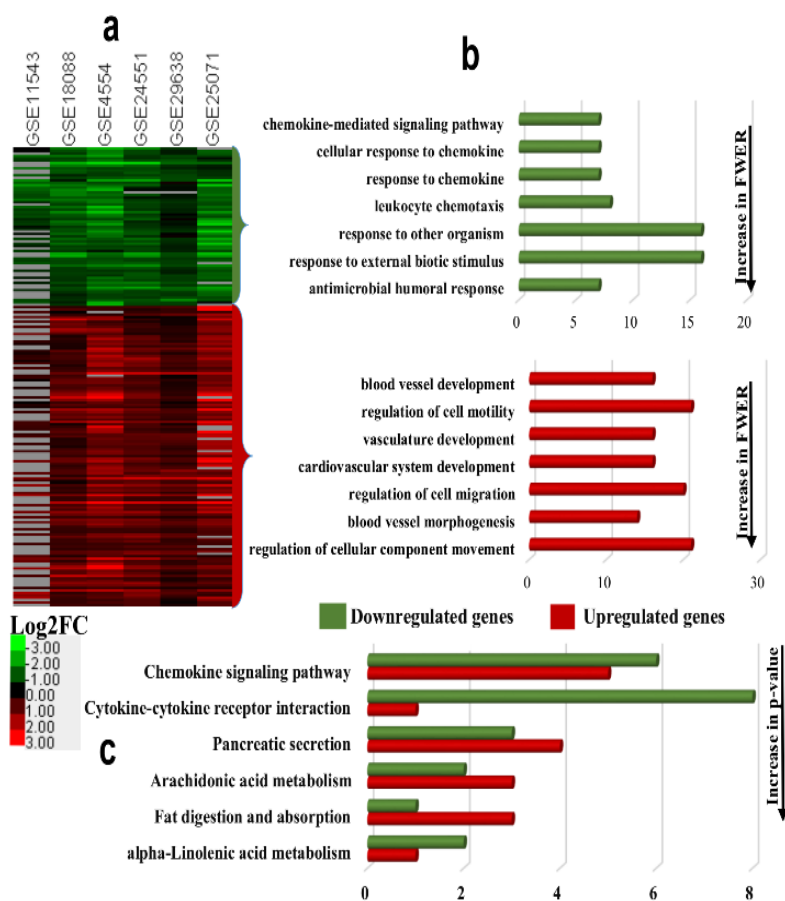


Figure 3. Resulted Heatmap for Common DEGs in MSS vs MSI-H Comparisons (a), and the results of their GO; (b), and pathway analysis; (c), MSS, microsatellite stable; MSI-H, microsatellite instable-high; DEG, differentially expressed genes; GO, gene ontology

PPI network analysis

To examine the biological meaning of DEGs at the protein level in a comprehensive manner, we used information from several databases for PPIs. Final PPI networks were constructed by information from BioGRID, MINT, HPRD and STRING databases. PPI network of DEGs found in MSS vs control cases comprised of 1139 nodes and 6569 edges and PPI network of DEGs found in MSI-H vs control cases consisted of 1,697 nodes and 14,779 edges. We used topology analysis by MCODE algorithm to find crucial modules with score >4 in each comparison and examined their GO by BiNGO plugin (Supplementary file 5, Figures 1 and 2). As well as, we calculated the degree parameter to find the number of interactions of each node in the networks. Degree parameter was used to sort DE-TFs in the PPI networks based on their PPIs (Supplementary file 5, Figure 3). GO analysis of significant protein complexes revealed up-regulation in modules involved in cell cycle processes in both MSS and MSI-H cases, as well as dysregulation of modules involved in fatty acid oxidation processes. FOXM1 was the only DE-TF which represented in the most significant modules from MCODE analysis of MSS and MSI-H PPI networks and these modules critically involved in cell cycle processes which accentuates the crucial role of FOXM1 as a central regulator for cell cycle in CRC. On the other hand, MYC, FOXM1, and STAT1 had the highest number of PPIs in comparison with the

other common DE-TFs in both types of CRC.

Discussion

Among all found DE-TFs in both MSS and MSI-H CRC, MYC and FOXM1 were common in both type based on our criteria of DEG selection, and also had high number of targets and were among the most up-stream DE-TFs based on centrality and core regulatory network analysis. They also showed a high number of interactions in PPI networks in comparison with the other DE-TFs. Moreover, FOXM1 was in the most important module resulted from MCODE topology analysis in both MSS and MSI-H protein-protein interaction networks.

MYC is a proto-oncogene which involves in many cancers and its significance in the cell cycle, cell proliferation, and various metabolic processes has been reported (Stine et al., 2015). It has been demonstrated that elimination of MYC function can result in tumor regression which indicates its importance as a therapeutic target, however MYC is an elusive target for cancer therapy (Castell and Larsson, 2015). Our results also displayed significant upregulation of MYC in both MSS and MSI-H CRC and its importance as a critical regulator of DEGs involved in cell cycle and various metabolic processes. However, MYC is somehow undruggable and hard for targeted therapy, but some strategies have been proposed that in one of them suggested for CRC is to

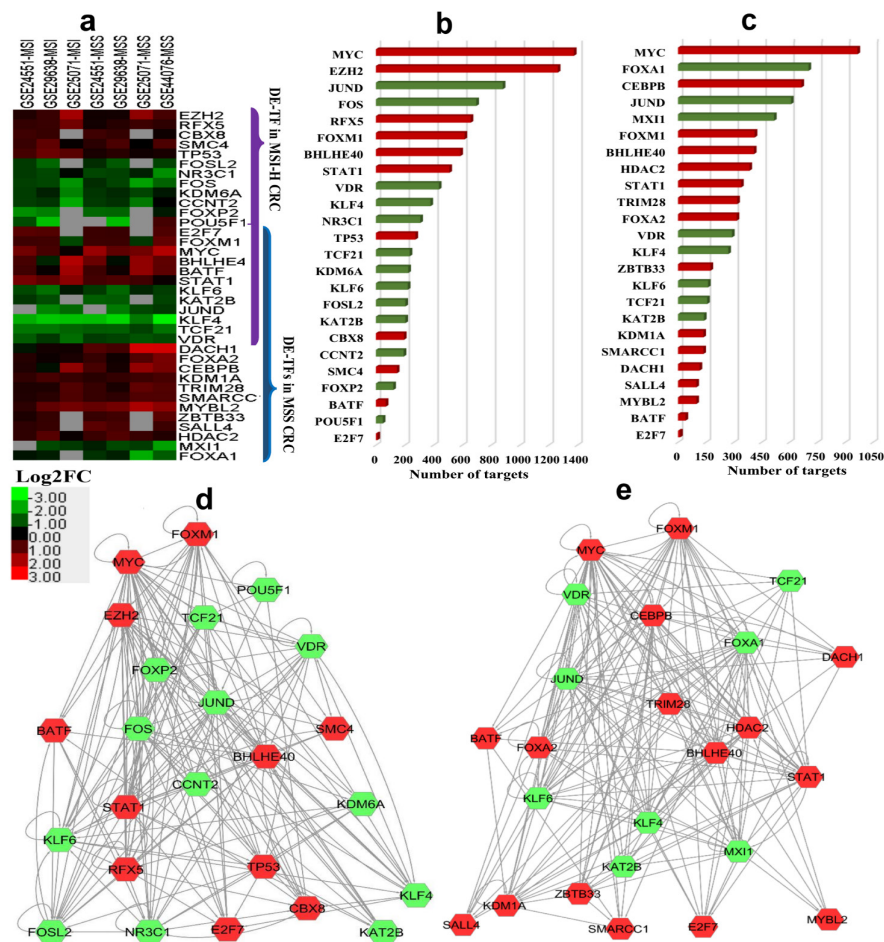


Figure 4. (a), Resulted Heatmap for DE-TFs in MSS and MSI-H vs Normal Comparisons; (b), results of centrality analysis of DE-TFs in MSI-H colorectal cancer and their arrangement based on number of targets; (c), results of centrality analysis of DE-TFs in MSS colorectal cancer and their arrangement based on number of targets; (d), extracted core regulatory network from MSI-H whole regulatory network; and (e), extracted core regulatory network from MSS whole regulatory network. DE-TFs, differentially expressed transcription factors; MSS, microsatellite stable; MSI-H, microsatellite instable-high

target MYC at translation level by directly targeting the translation initiation factor eIF4A (Castell and Larsson, 2015). Another way to target MYC can be targeting its upstream regulators which indirectly affect MYC activation. However, when we target a regulator of gene expression it affects many other down-stream target genes; but if we target a MYC regulator, which also deregulated in CRC and plays a role in its development, may be an effective approach. Our results suggest FOXM1 as a critical and central regulator in both type of CRC and our core regulatory network represented positive regulation of MYC by FOXM1 in CRC. Upregulation of FOXM1 in various type of cancer and its association with drug resistance have been proved (Myatt and Lam, 2008; Li et al., 2014; Gu et al., 2018). Moreover, the crucial role of FOXM1 in cell cycle and its use as a candidate for targeted therapy in a number of cancers have also been reported (Myatt and Lam, 2008; Wang et al., 2010; Buchner et al., 2015; Gu et al., 2016; Lu et al., 2018; Akbari et al., 2019). Furthermore, transactivation of MYC promoter by FOXM1 has been demonstrated (Wierstra and Alves, 2006, 2007). Taken together, these data suggest positive regulation of MYC by FOXM1 in CRC, therefore the use

of FOXM1 as an alternative therapeutic target for MYC may be of great value.

We also mined the expression data to unravel the differences between MSS and MSI-H CRC. Immune-related processes, mainly chemokine mediated processes, displayed more overexpression in MSI-H, while processes related to metastasis including migratory and vasculature development showed more upregulation in MSS. These results may explain why MSI-H colorectal cancer has better immunotherapy response and lower metastasis rather than MSS CRC. It has been proven that the existence of tumor-infiltrating lymphocytes (TILs) in many cancer types is a positive prognostic factor and increased TILs density have been demonstrated in MSI-H tumors, but the reason of that is not well understood (Pagès et al., 2005; Lee et al., 2016). Chemokines play important role in recruitment and infiltration of lymphocytes within the tumor microenvironment. As well as, the role of chemokine-chemokine receptor network in T-cell-mediated antitumor immune response has been demonstrated (Franciszkievicz et al., 2012). We detected up-regulation of chemokine mediated processes (Figure 3 b and c) in MSI-H cases. CXCL10, CXCL9, CCL5,

CXCL13, and CCL18 genes which all encode chemokines and CXCR4 gene which encodes a chemokine receptor were common in all of those processes in our study. Taken together, up-regulation of these processes and cytokines may be responsible for increased density of TILs and better immunotherapy response in MSI-H CRC.

Additionally, MMR pathway showed more upregulation in MSI-H cases. Defective MMR is characterized by MSI, and hereditary nonpolyposis colorectal cancer (HNPCC) or Lynch syndrome, the most common form of hereditary colorectal cancer, caused by inherited MMR genes mutations, most commonly mutations in MLH1 and MSH2 (Li and Martin, 2016). Even though MMR displayed upregulation; but one of the major genes in this process, MLH1, displayed downregulation just in MSI-H CRC without any significant change in MSS CRC studies. As well as, MLH1 showed upregulation in 5 out of 6 MSS vs MSI-H studies (which is due to downregulation of MLH1 in MSI-H cases). In fact, MLH1 was the only mismatch repair gene that showed significant downregulation which highlights the importance of this gene in sporadic MSI-H CRC than the other MMR genes.

Further, we found *TP53* (upregulated in MSI-H CRC), *FOXA1* (downregulated in MSS CRC), *DACHI* (upregulated in MSS CRC), *CDX2*, and *ZMYND8* (downregulated in MSI-H CRC) as the main DE-TFs which showed differential expression just in one type of CRC, MSS or MSI-H. Among them, *TP53* and *DACHI* not only represent upregulation in all of the MSI-H and MSS versus control studies, respectively; but also displayed differential expression in a considerable number of MSS vs MSI-H cases. Previous studies reported an inverse relation between p53 gene alterations and MSI. On the other hand, some other studies reported that overexpression of p53 is more common in MSI than MSS CRC and suggested the possible use of TP53 expression to predict MSI-H CRC prior to MSI typing (Nyiraneza et al., 2011). Our result not only demonstrate upregulation of *TP53* in MSI-H CRC but also suggest that this TF may be a significant regulator of DEGs in this type of cancer. Else ways, *DACHI* showed significant upregulation just in MSS CRC. Previous studies reported methylation of *DACHI* promoter and its downregulation in cancer and demonstrated that *DACHI* can inhibit breast tumor invasion and growth by suppressing epithelial-mesenchymal transition (EMT) and proliferation through repression of Snail and Cyclin D1, respectively (Wu et al., 2006; Zhao et al., 2015). Bu, Xiao-Na and colleagues illustrated that *DACHI* upregulation can promote pancreatic cancer growth and invasion; downregulation of *DACHI* activity with shRNA can repress cell proliferation and tumor invasion by mostly inducing apoptosis and inhibiting EMT in pancreatic cancer cells through modulating Bcl-2 (pro-survival regulator) and E-cadherin, respectively (Bu et al., 2016). Our data have shown significant upregulation of *DACHI* in MSS CRC; moreover Cyclin D1 and Snail represent upregulation in 3 out of 4 MSS studies; on the other hand, Bcl-2 was downregulated and E-cadherin did not represent significant change in expression in MSS studies. These results may refute the inhibitory role of *DACHI* in tumor growth and invasion in MSS CRC through Snail

and Cyclin D1 and also may reject the involvement of *DACHI* in tumor growth and invasion through Bcl-2 and E-cadherin in MSS CRC. Lee, Jae-Woong, and colleagues have shown that *DACHI* can elevate cell cycle progression through upregulation of cyclin D1, D3, F, and Cdk 1, 4, and 6 and decrease in p21Cip1 in myeloid cells (Lee et al., 2012). In our survey, Cyclin D1, F, and Cdk 4 were shown significant upregulation in most of the MSS studies; as well as Cdk 1 showed upregulation pattern in all MSS studies (FC>1.5 in 3 studies with significant P value in 2 of them) and p21Cip1 represented downregulation pattern in all MSS studies (with significant downregulation in 1 study). These results may suggest the involvement of *DACHI* in cell cycle progression in MSS CRC mainly through upregulation of cyclin D, F, Cdk 1 and 4.

In conclusion, the present study revealed significant upregulation of cell cycle and downregulation of metabolic, specifically fatty acid catabolism, processes in CRC and predicted MYC and FOXM1 as two critical and central regulators of DEGs in CRC. On the other hand, we demonstrated that immune-related processes are more active in MSI-H cases, while in MSS cases metastasis related processes including migratory and vasculature development are more active; also we found *TP53* and *DACHI* as two DE-TFs which differentially expressed just in one type of CRC.

Acknowledgments

We would like to appreciate the authors of studies whose data was used in our study for making their data freely available.

Ethics

The informed consent and research ethics board review processes for ethical approval were performed by the original studies; and as the data are freely available and that the data or samples are anonymous there is no ethical issues concerning their use.

Conflicts of interest

The authors declare that there is no conflict of interest and source of funding.

References

- Ågesen TH, Berg M, Clancy T, et al (2011). CLC and IFNAR1 are differentially expressed and a global immunity score is distinct between early- and late-onset colorectal cancer. *Genes Immun*, **12**, 653.
- Ågesen TH, Sveen A, Merok MA, et al (2012). ColoGuideEx: a robust gene classifier specific for stage II colorectal cancer prognosis. *Gut*, **61**, 1560–7.
- Akbari V, Kallhor M, Akbari MT (2019). Transcriptome mining of non-BRCA1/A2 and BRCA1/A2 familial breast cancer. *J Cell Biochem*, **120**, 575– 83.
- Bader GD, Hogue CW V (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, **4**, 2.
- Bray F, Ferlay J, Soerjomataram I, et al (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA*

- Cancer J Clin*, **68**, 394–424.
- Bu X-N, Qiu C, Wang C, Jiang Z (2016). Inhibition of *DACH1* activity by short hairpin RNA represses cell proliferation and tumor invasion in pancreatic cancer. *Oncol Rep*, **36**, 745–54.
- Buchner M, Park E, Geng H, et al (2015). Identification of FOXM1 as a therapeutic target in B-cell lineage acute lymphoblastic leukaemia. *Nat Commun*, **6**, 6471.
- Castell A, Larsson L-G (2015). Targeting myc translation in colorectal cancer. *Cancer Discov*, **5**, 701–3.
- Consortium GO (2016). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res*, **45**, 331–8.
- Consortium TEP (2004). The ENCODE (ENCyclopedia of DNA elements) project. *Science*, **306**, 636–40.
- Cordero D, Solé X, Crous-Bou M, et al (2014). Large differences in global transcriptional regulatory programs of normal and tumor colon cells. *BMC Cancer*, **14**, 708.
- Davis S, Meltzer PS (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, **23**, 1846–7.
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, **95**, 14863–8.
- Franciszkiwicz K, Boissonnas A, Boutet M, Combadière C, Mami-Chouaib F (2012). Role of chemokines and chemokine receptors in shaping the effector phase of the antitumor immune response. *Cancer Res*, **72**, 6325–32.
- Gröne J, Lenze D, Jurinovic V, et al (2011). Molecular profiles and clinical outcome of stage UICC II colon cancer patients. *Int J Colorectal Dis*, **26**, 847–58.
- Gryfe R, Kim H, Hsieh ETK, et al (2000). Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *N Engl J Med*, **342**, 69–77.
- Gu C, Yang Y, Sompallae R, et al (2016). FOXM1 is a therapeutic target for high-risk multiple myeloma. *Leukemia*, **30**, 873.
- Gu C, Jing X, Holman C, et al (2018). Upregulation of FOXM1 leads to diminished drug sensitivity in myeloma. *BMC Cancer*, **18**, 1152.
- Hampel H, Frankel WL, Martin E, et al (2005). Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med*, **352**, 1851–60.
- Hampel H, Frankel WL, Martin E, et al (2008). Feasibility of screening for Lynch syndrome among patients with colorectal cancer. *J Clin Oncol*, **26**, 5783.
- Hong Y, Downey T, Eu KW, Koh PK, Cheah PY (2010). A ‘metastasis-prone’ signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics. *Clin Exp Metastasis*, **27**, 83–90.
- Huber W, Carey VJ, Gentleman R, et al (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*, **12**, 115–21.
- Jensen LJ, Kuhn M, Stark M, et al (2008). STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*, **37**, 412–6.
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*, **45**, 353–61.
- Keshava Prasad TS, Goel R, Kandasamy K, et al (2008). Human protein reference database-2009 update. *Nucleic Acids Res*, **37**, 767–72.
- Kim CG, Ahn JB, Jung M, et al (2016). Effects of microsatellite instability on recurrence patterns and outcomes in colorectal cancers. *Br J Cancer*, **115**, 25.
- Kinzler KW, Vogelstein B (1996). Lessons from hereditary colorectal cancer. *Cell*, **87**, 159–70.
- Kurzawski G, Suchy J, Debniak T, Kladny J, Lubinski J (2004). Importance of microsatellite instability (MSI) in colorectal cancer: MSI as a diagnostic tool. *Ann Oncol*, **15**, 283–4.
- Lachmann A, Xu H, Krishnan J, et al (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics*, **26**, 2438–44.
- Lee J-W, Kim H-S, Kim S, et al (2012). *DACH1* regulates cell cycle progression of myeloid cells through the control of cyclin D, Cdk 4/6 and p21Cip1. *Biochem Biophys Res Commun*, **420**, 91–5.
- Lee V, Murphy A, Le DT, Diaz LA (2016). Mismatch repair deficiency and response to immune checkpoint blockade. *Oncologist*, **21**, 1200–11.
- Li SKH, Martin A (2016). Mismatch repair and colon cancer: mechanisms and therapies explored. *Trends Mol Med*, **22**, 274–89.
- Li X, Yao R, Yue L, et al (2014). FOXM1 mediates resistance to docetaxel in gastric cancer via up-regulating Stathmin. *J Cell Mol Med*, **18**, 811–23.
- Licata L, Briganti L, Peluso D, et al (2011). MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res*, **40**, 857–61.
- Lu X-F, Zeng D, Liang W-Q, et al (2018). FoxM1 is a promising candidate target in the treatment of breast cancer. *Oncotarget*, **9**, 842.
- Maere S, Heymans K, Kuiper M (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**, 3448–9.
- Myatt SS, Lam EW-F (2008). Targeting foxm1. *Nat Rev Cancer*, **8**, 242.
- Nyiraneza C, Jouret-Mourin A, Kartheuser A, et al (2011). Distinctive patterns of p53 protein expression and microsatellite instability in human colorectal cancer. *Hum Pathol*, **42**, 1897–1910.
- Overman MJ, Ernstoff MS, Morse MA (2018). Where we stand with immunotherapy in colorectal cancer: deficient mismatch repair, proficient mismatch repair, and toxicity management. *Am Soc Clin Oncol Educ B*, **38**, 239–47.
- Page F, Berger A, Camus M, et al (2005). Effector memory T cells, early metastasis, and survival in colorectal cancer. *N Engl J Med*, **353**, 2654–66.
- Peltomäki P, Lothe RA, Aaltonen LA, et al (1993). Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome. *Cancer Res*, **53**, 5853–5.
- Popat S, Hubner R, Houlston RS (2005). Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol*, **23**, 609–18.
- Ritchie ME, Phipson B, Wu D, et al (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, **43**, 47.
- Sagaert X, Van Cutsem E, Tejpar S, Prenen H, De Hertogh G (2014). MSI versus MSS sporadic colorectal cancers: Morphology, inflammation, and angiogenesis revisited. *J Clin Oncol*, **32**, 495.
- Saldanha AJ (2004). Java Treeview-extensible visualization of microarray data. *Bioinformatics*, **20**, 3246–8.
- Scardoni G, Tosadori G, Faizan M, et al (2014). Biological network analysis with CentiScaPe: centralities and experimental dataset integration. *F1000Research*, **3**, doi: 10.12688/f1000research.4477.2
- Shannon P, Markiel A, Ozier O, et al (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, **13**, 2498–2504.
- Sherman BT, Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, **4**, 44–57.

- Stark C, Breitkreutz B-J, Reguly T, et al (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res*, **34**, 535–9.
- Stine ZE, Walton ZE, Altman BJ, Hsieh AL, Dang CV (2015). MYC, metabolism, and cancer. *Cancer Discov*, **5**, 1024–39.
- Sveen A, Ågesen TH, Nesbakken A, et al (2011). Transcriptome instability in colorectal cancer identified by exon microarray analyses: Associations with splicing factor expression levels and patient survival. *Genome Med*, **3**, 32.
- Thibodeau SN, Bren G, Schaid D (1993). Microsatellite instability in cancer of the proximal colon. *Science*, **260**, 816–9.
- Vilar E, Mukherjee B, Kuick R, et al (2009). Gene expression patterns in mismatch repair-deficient colorectal cancers highlight the potential therapeutic role of inhibitors of the phosphatidylinositol 3-kinase-AKT-mammalian target of rapamycin pathway. *Clin Cancer Res*, **15**, 2829–39.
- Wang Z, Ahmad A, Li Y, et al (2010). Forkhead box M1 transcription factor: a novel target for cancer therapy. *Cancer Treat Rev*, **36**, 151–6.
- Ward R, Meagher A, Tomlinson I, et al (2001). Microsatellite instability and the clinicopathological features of sporadic colorectal cancer. *Gut*, **48**, 821–9.
- Watanabe T, Kobunai T, Toda E, et al (2006). Distal colorectal cancers with microsatellite instability (MSI) display distinct gene expression profiles that are different from proximal MSI cancers. *Cancer Res*, **66**, 9804–8.
- Wierstra I, Alves J (2006). FOXM1c transactivates the human c-myc promoter directly via the two TATA boxes P1 and P2. *FEBS J*, **273**, 4645–67.
- Wierstra I, Alves J (2007). FOXM1c and Sp1 transactivate the P1 and P2 promoters of human c-myc synergistically. *Biochem Biophys Res Commun*, **352**, 61–8.
- Wu K, Li A, Rao M, et al (2006). *DACH1* is a cell fate determination factor that inhibits cyclin D1 and breast tumor growth. *Mol Cell Biol*, **26**, 7116–29.
- Xia J, Fjell CD, Mayer ML, et al (2013). INMEX—a web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Res*, **41**, 63–70.
- Yates B, Braschi B, Gray KA, et al (2016). Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Res*, **45**, 619–25.
- Zhao F, Wang M, Li S, et al (2015) *DACH1* inhibits SNAI1-mediated epithelial–mesenchymal transition and represses breast carcinoma metastasis. *Oncogenesis*, **4**, e143.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.