

A comprehensive investigation of colorectal cancer progression, from the early to late-stage, a systems biology approach

Mohammad Ghorbani¹, Yazdan Asgari^{1,*}

¹ Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences, Tehran, Iran

*Corresponding Author

Yazdan asgari's orchid number: 0000-0001-6993-6956

Yazdan asgari's email address: yasgari@tums.ac.ir

Mohammad ghorbani's orchid number: 0000-0001-8278-6024

Mohammad ghorbani's email: m28.biotech@gmail.com

Abstract

Colorectal cancer is a widespread malignancy with a concerning mortality rate. It could be curable at the first stages, but the progress of the disease and reaching to the stage-4 could make shift the treatments from curative to palliative. In this stage, the survival rate is meager, and therapy options are limited. The question is, what are the hallmarks of this stage and what genes are involved? What mechanism and pathways could drive such a malign shift from stage-1 to stage-4? In this study, first we identified the core modules for both the stage-1 and stage-4 which four of them have a significant role in stage-1 and two of them have a role in stage-4. Then we investigated the gene ontology and hallmarks analysis for each stage. According to the results, the immune-related process, especially interferon-gamma, impacts stage-1 in colorectal cancer. Concerning stage-4, extracellular matrix ontologies, and metastatic hallmarks are in charge. At last, we performed a differentially expressed gene analysis of stage-4 vs. stage-1 and analyzed their pathways which reasonably undergone a hypo/hyperactivity or being abnormally regulated through the cancer progression. We found that lncRNA in canonical WNT signaling and colon cancer has the most significant pathways, followed by WNT signaling, which means that these pathways may be the driver for the development from early-stage to late-stage. Of these lncRNAs, we had two upregulated kind, H19, and HOTAIR, which both can be involved and mediate metastasis and invasion in colorectal cancer.

Keywords: Systems biology, colorectal cancer, WGCNA, lncRNA, WNT, metastasis

Introduction

From submucosa it arose, growing to beyond the inner layer; no invasion, manageable and curable; this is what you can call an early-stage(stage-1) of colorectal cancer [1]. Usually, it needs imaging and colonoscopy to diagnose [2]. The primary strategy at this point is resection of the tumor [3]. Even if therapy has become successful, constant surveillance must continue to secure that the tumor never comes back, yet there is still a risk of recurrence [4]. One of the reasons for this ominous turn of events and progress of the disease is Micrometastasis in stage-1/2, which enables tumor cells to reach the nearby lymph node and hide from the available detection method. These cells could be responsible for later relapse [5]. Still, the survival rate for early-stage is about 93.2% percent. As the disease progress, survival goes down with it; Stage-2 has a survival rate of 85 to 70 percent, Stage three is about 44 percent, and finally, only eight percent of late-stage(stage-4) patients could pass the five-year survival [6]. In Stage-2, cancer might be grown into the wall of the bowl, to the adjacent tissues [7]. Surgery and adjuvant chemotherapy and different kinds of combination therapy have been considered as an option at this point. Despite all of this, the risk of recurrence in this stage is highly increased [8]. By the Stage-3, tumor growth further to bowel wall and even into the nearby organs. It has already entered lymph nodes [7], and it is safe to say that it shows far aggressive behavior. At this level, recurrence is more common, the survival rate drops drastically, and managing tumor becomes harder until the late-stage occurs [9]. Stage-4 of colorectal cancer begins with the spread throughout the body [10]. Usually liver is the first place of the invasion which followed by the lung, bones, and brain; this is what we can call advanced colorectal cancer with a grave prognosis [11]. Late-stage usually response partially to treatment, and of course, at this stage, relapse and recurrence are a prevalent phenomenon [12, 13]. Indeed, one of the hallmarks of late-stage is the invasion [14].

So, there is an urgent need to understand what could contribute to this transformation from stage-1 to stage-4. What pathway involves and what modules are in charge? Systems biology could provide an answer for these questions by analyzing high-throughput data. It enables us to infer from the expression profile, so what modules are involving in different Stages of the disease and what pathways make the difference. It expands our insight and opens new possibilities to treat and handle the diseases [15]. Since we must consider the worst scenario, we should have a good idea about the mechanism of early and late stages of colorectal cancer and have a piece of proper knowledge about what could make this progress happens. If there is a possibility to stop this transformation from stage-1 to stage-4, it must be considered and taken into effect.

In this investigation, we analyze the most valuable modules by using the weighted co-expression network analysis (WGCNA). Based on the integrative results of WGCNA modules, we identified the essential genes in each stage. Then we conducted a thorough Gene Ontology (GO), Hallmark's analysis for both the early and the late-stage of colorectal cancer. To identify the accountable pathway for this process, we performed a differentially expressed gene(DEG) analysis between stage-4 and stage-1. The outcome of our DEG investigation went through a pathway analysis in order to identify the responsible genes and pathways.

Materials and methods

Research design

Our main question was the primary features of the first and late stages of colorectal cancer and how this progression occurs. We used the system biology approach to answer these questions. We had found primary modules in each stage, using hallmarks and gene ontology analysis, then we identified the main features. Nonetheless, what could bring such a contrast? We conducted a differentially expressed gene analysis to answer this question, and the outcome underwent a pathway analysis. Our conclusion was based on The main pathways and their relationship with their respected genes.(Figure 1)

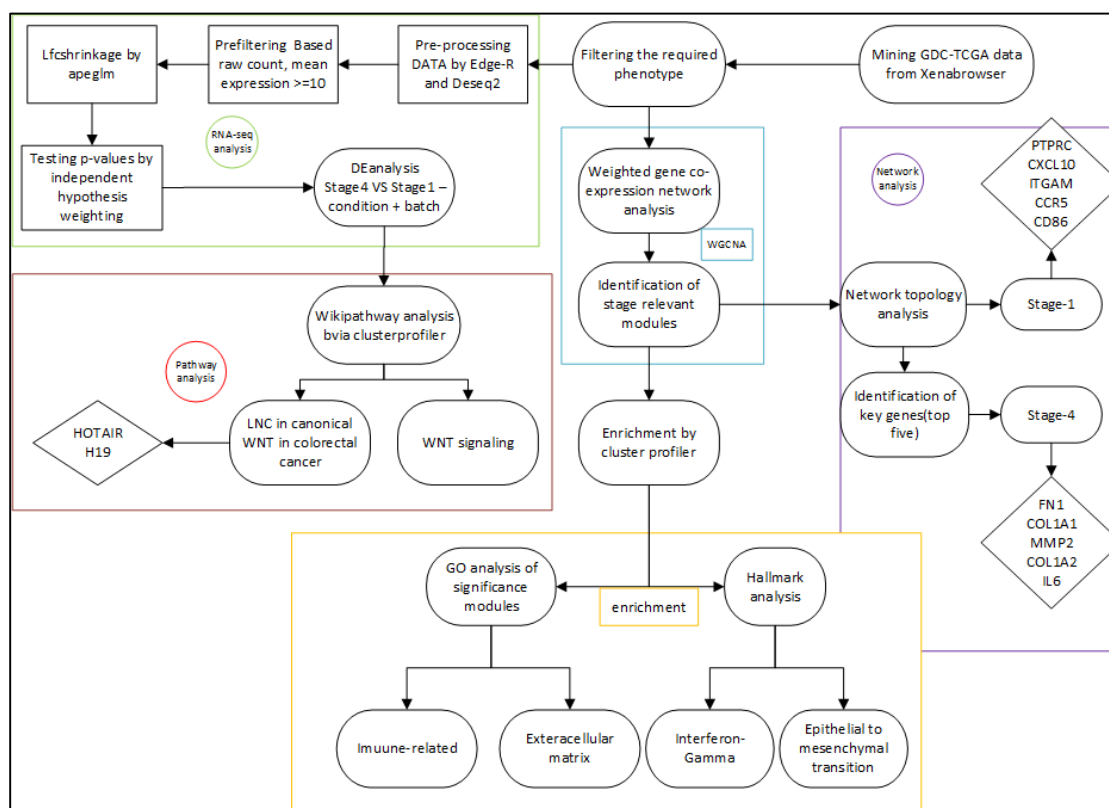


Figure 1. Workflow of the study

Adjusting phenotype to different stages

First, we obtained phenotype data from xenabrowser, a TCGA tool. Data comes from "GDC TCGA COAD" dataset 2019 version [16]. In order to filter typical phenotypes based on our purpose, which was identifying functional module and key-genes in stage-4 and stage-1 of colon cancer, we set the parameter accordingly: sample type on "primary tumor" because we need to explore stage-4 and stage-1. Hence the healthy solid tissues were unrelated. Primary diagnosis on "Adenocarcinoma", and disease type was on "Adenomas and adenocarcinomas" because it comprises the most cases of colon cancer. Regarding tumor stages, Stage-I and stage-Ia considered as stage-1 with 68 samples, and stages-IV/IVa/IVb considered as stage-4 with 56 samples. We

also included Tumor Node Metastasis (TNM) data since it could be directly applied to colorectal cancer and can define different stages in this disease. Then we matched the identified sample id from phenotype sheet to count and FPKM data for further analysis (Supp1).

Weighted Gene Co-expression Network Analysis

Weighted Gene Co-expression Network Analysis (WGCNA) is an R package that can find clusters of highly correlated genes that are valuable in a biological context and relate them to a set of distinct traits, and this could directly guide you to hub-genes, hub-modules, and even significance pathways concerning to that specific trait [17]. First, we pre-processed FPKM data, related to early and late stages of colorectal cancer, and removing outlier rows with zero median expression, since they do not have any biological value. Then we chose the top quartile of our network by variance. This process conducted by R version 3.6.3 [18]. Mentioned pre-processed data again, going through sample clustering via WGCNA to trim outlier samples. For this analysis, we chose the Step-by-step network construction method for the identification of correlated modules. After drawing scale-free topology and choosing seven as soft power, the process of building the adjacency matrix started, which then transformed into the topological overlap matrix and then to its respective similarity (TOMsimilarity) and dissimilarityTOM matrix. Finally, a cluster dendrogram established based on TOM-dissimilarity and cutting of the eigengenes modules tree. To discover the essential genes, we merged all modules which met our given criteria (significance>0.1 and P-value <0.1). For stage-4, midnight blue and black modules were statistically meaningful, and in stage-1, there were red, green yellow, grey60, and magenta.

Topological Analysis

Obtained integrated modules from late-stage and early-stage were imported to Cytoscape to assemble a Protein-Protein Interaction (PPI) network by STRING database (Search Tool for the Retrieval of Interacting Genes/Proteins) [19, 20]. Confidence cutoff was set on 0.4, and the maximum additional interaction was zero. Obtained PPI network undergoes a centrality-based analysis via CytoNCA, a Cytoscape plugin; CytoNCA is a centrality ranking method and investigates an assigned network by eight centralities method [21].

Enrichment

To find a significant pathway in the most correlated modules, we ran a GO over-representation analysis (ORA) through an R package called clusterprofiler [22-24]. Then we use Msigdb analysis (another clusterprofiler features), which allows us to perform hallmarks analysis [13, 25]. Additionally, to find a significantly enriched pathway in differentially expressed genes, we ran a gene set enrichment analysis (GSEA) by clusterprofiler through wikipathway database in a sort of pre-ranked list. We used fold change as a weight for enrichment. The criteria for these analyses was $\log_2\text{foldchange} \geq 1$ and adjusted-value ≤ 0.05 . Besides, we used enrichplot (an R package) to draw plots and figures to be more visually illustrative [26].

Identifying differentially expressed genes

First, we downloaded raw count data from xenabrowser, GDC TCGA COAD dataset. However, data was in the form of $\log_2(\text{count}+1)$ and needed to be transformed to the original count table (count without log) before using it as raw material for DEanalysis packages in R. According to

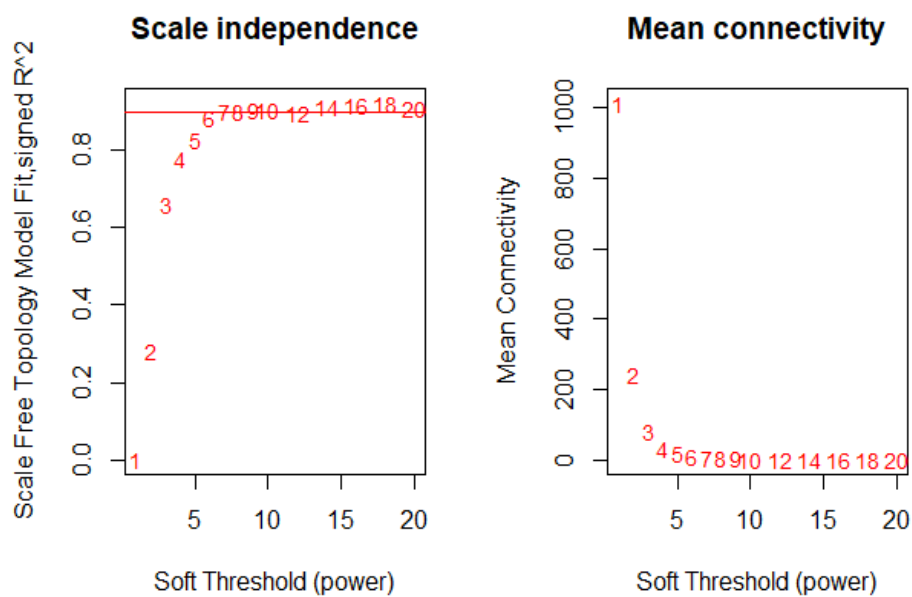
phenotype choice, stage-1 consider as control by 68 samples. and stage-4 consider as the case by 56 Samples. Then we used EdgeR (an R package) to get annotations through “org.Hs.eg.db” only to keep rows that have Entrez Gene ID [27, 28]. In the next step, we used Deseq2 (another R package) to keep rows that meet our criteria for counts expression rowsums \geq 10 [29]. Moreover, LFC shrinkage applied by apeglm method and Independent hypothesis weighting estimates weighs for each P-value through the Benjamini-Hochberg method [30, 31]. We only considered genes statistically meaningful if their Adjusted P-value was less than or equal to 0.05. all computational analyses were performed by R (v 3.6.3) and Rstudio [32] in a windows10 OS.

Results

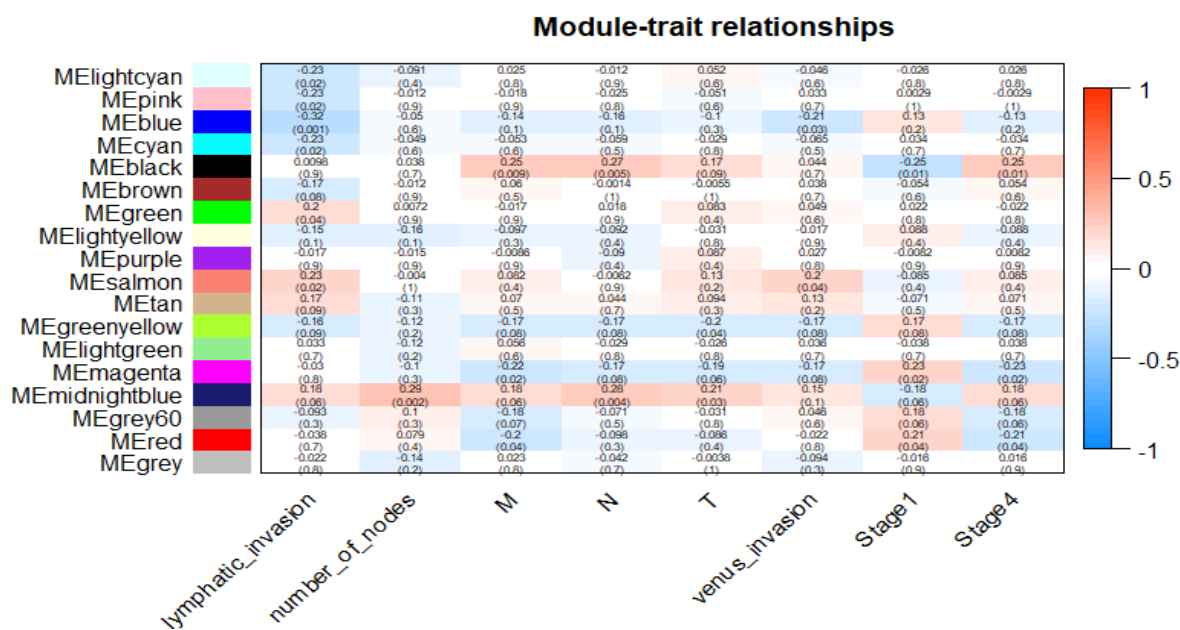
Identification of primary-modules

Our sample clustering detected and removed 19 outliers from our first collection of samples. Scale-free topology identified seven as an optimal power (Figure 2a). We cut the eigengenes module tree by 0.30 to correlate to 0.70 similarity. Based on our inquiry 17 modules found in the network that two of them can be correlated to stage-4 (black and midnight blue) and four of them met the criteria to be considered as significant in stage-1 category (green yellow, red, grey60, and magenta) (Figures 2b, 2c). Also, Midnight blue Module appeared to correlate to N and T stages in T.N.M staging system (Figures 2b, 2c). But the Black module seemed to be more associated with M stage also, the difference was very marginal (Figures 2b, 2c). Then we integrated all statistically meaningful modules in each stage in order to make a gene ontology analysis. In this analysis, our networks considered as a signed network since both negatively and positively correlated genes considered important in the related modules.

a



b



c

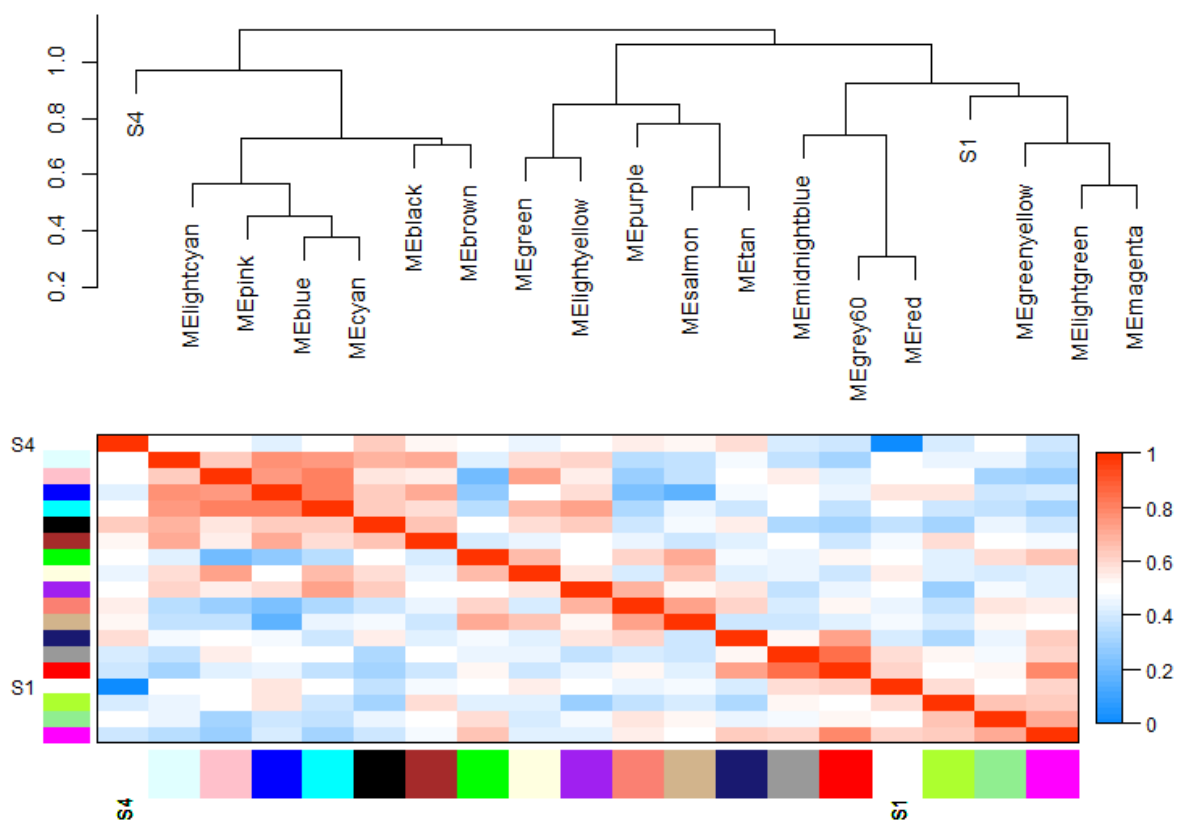


Figure 2: a) Scale-free topology to obtain the optimal soft power, b) Module trait relationship in which black and midnight blue modules came as significant for stage-4, and green yellow, red, grey60, and magenta came as significance for stage-1, c) Correlation between modules and stages using plotEigengeneNetworks.

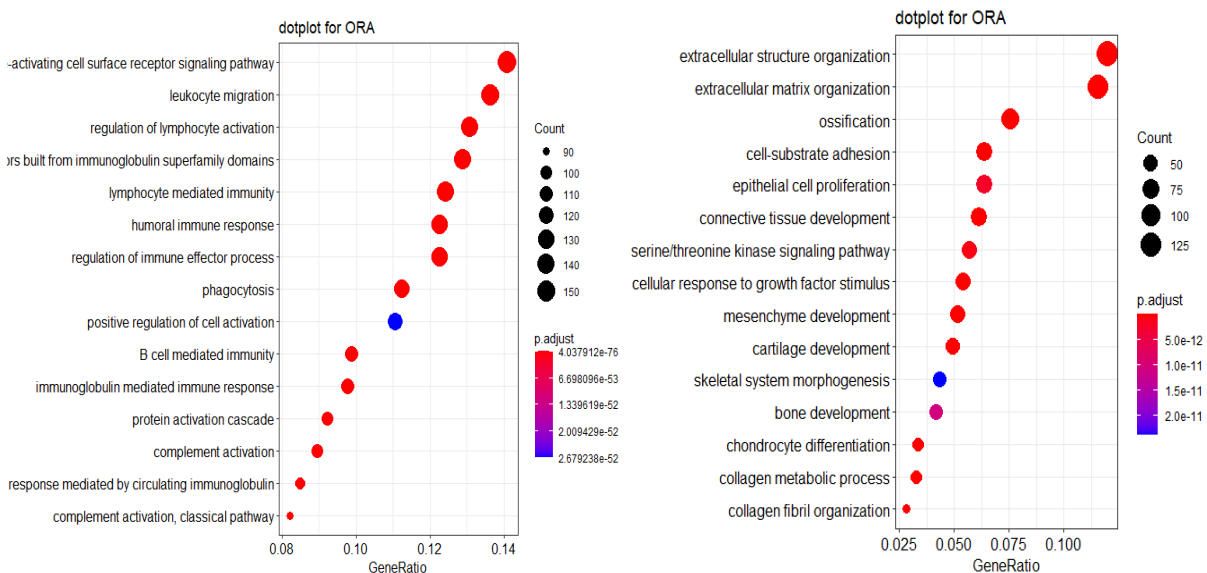
Module Enrichment

To get a deep insight into the functionality of the modules, we performed a GO over-representation analysis via clusterprofiler. GO analysis implemented in three categories: biological pathway, cellular component, and molecular function [22].

In stage-1, four modules met our criteria for GO enrichment (green yellow, magenta, red, and grey60). We integrated all genes from those modules and conducted a GO analysis in three categories. Biological pathway mostly pointed out immune-related process (Figure 3). Furthermore, cellular component comprised of immune system features, like immunoglobulin complex and MHC classes. Molecular function was considered chemokine and cytokine activity, antigen binding, and other immune like trends as significance. Since stage-4 only had two statistically notable modules, we merged these two as one unified block (black and midnight blue). The GO results from cellular component, biological pathway, and molecular function for this specific block was very consistent. In biological pathway, extracellular matrix organization and extracellular structure organization by far had the most meaningful adjusted-p-value, q-value, and gene ratio. However, the majority of results involved in matrix, collagen, and adhesion processes (Figure 3). In cellular component category, collagen-containing extracellular matrix had the best score by far, and most of the top ten outcomes led us to the extracellular matrix and cell adhesion. There was also evident in the top ten molecular function.

We also run an MSigDb analysis in order to find hallmarks for each stage [25]. According to stage-1 hallmarks, "interferon-gamma (IFN- γ) response" had the most significance adjusted p-value and q-value (Figure 4). Results of the stage-4 modules (black and midnight blue) hallmarks were a closure to our GO analysis (Figure 4). hallmarks analysis had a solid answer to our GO results, and that was Epithelial to mesenchymal transition, which is very consistent with late-stage behavior and proved by many experimental and clinical studies [33, 34].

a



b

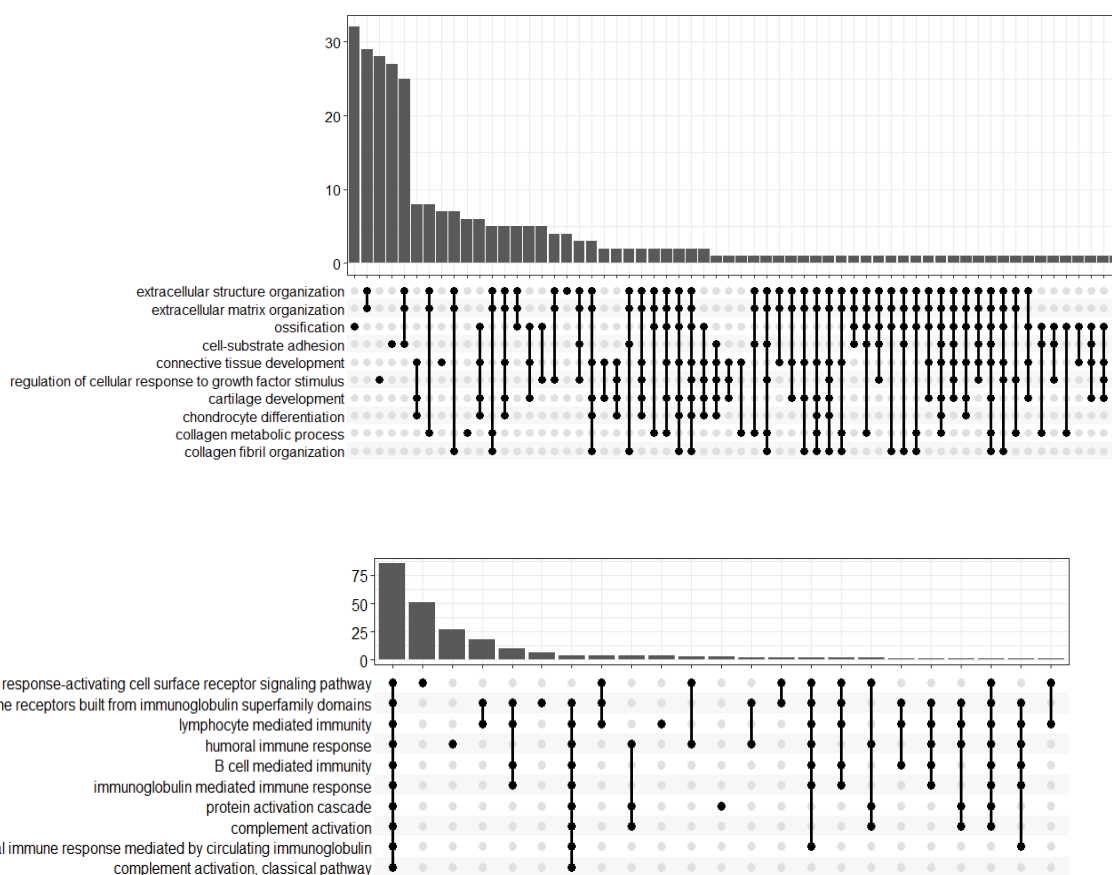
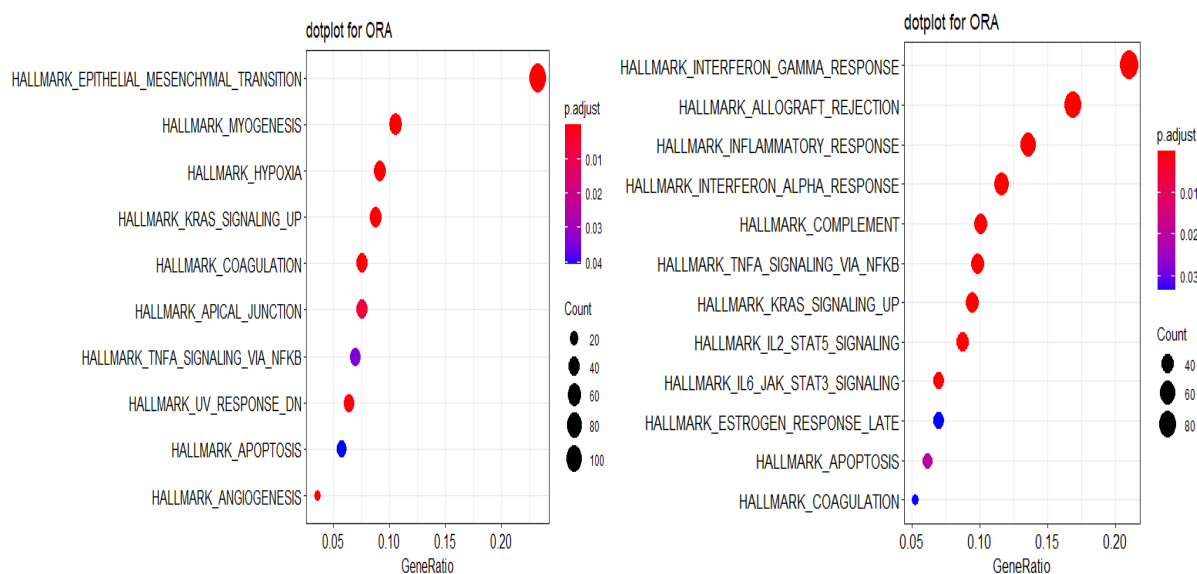


Figure 3: a) A dot plot of GO biological pathways, terms are categorized based on their gene ratios and adjusted p-values, **b)** upset plot for biological pathways, upset plot shows association, and overlapping between different components of a complex system.

a



b

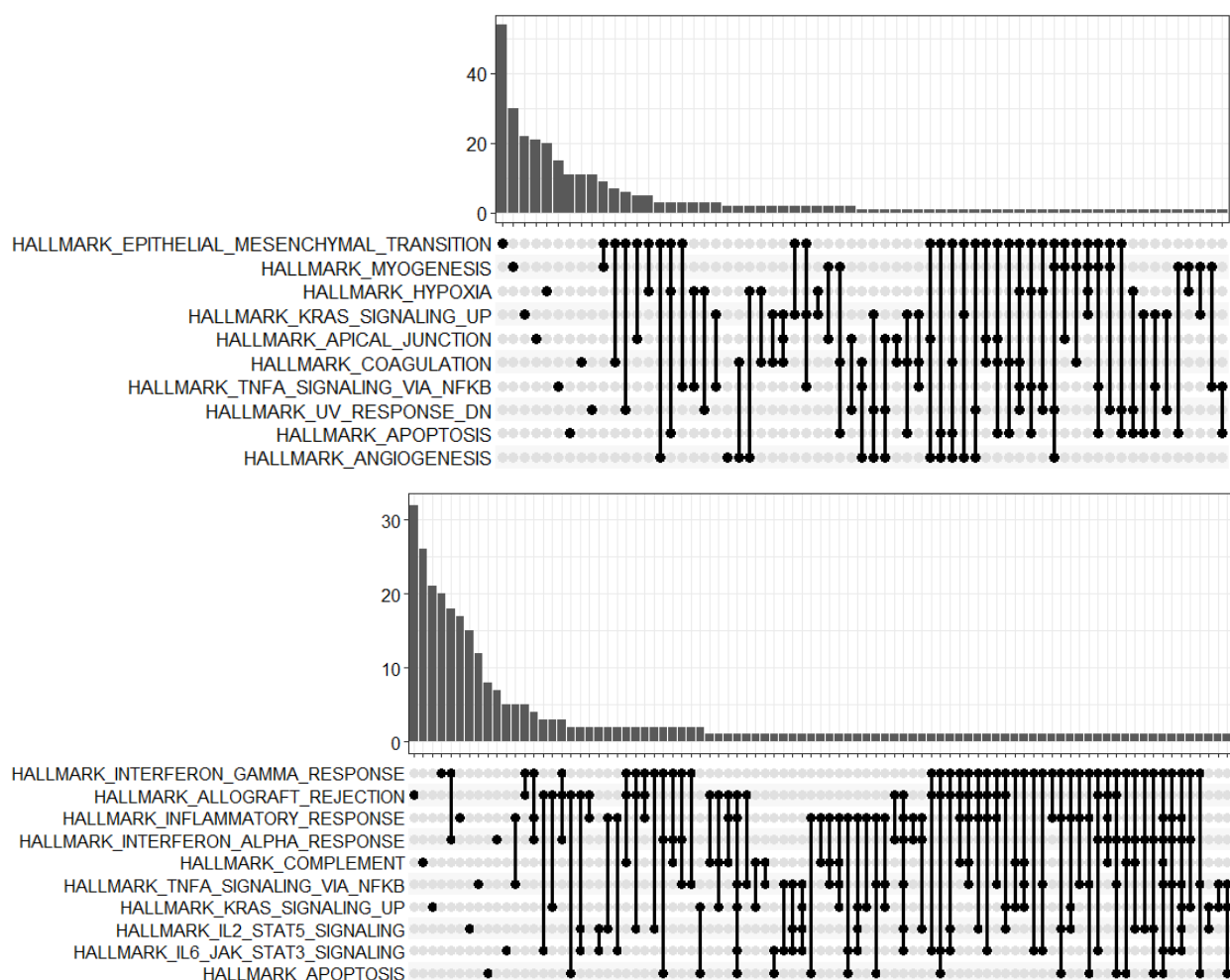


Figure 4: a) Dot plot for hallmark analysis, sorted by adjustedP-value, b) upset plot for visualizing hallmark analyzing for stage-1 and stage-4 modules

Identifying key-genes

In order to find key genes in primary modules for either stage, first of all, we constructed a PPI network via STRING. From 1345 genes of four significance modules of stage-1, 941 genes established a network, and from 1276 genes from two statistically meaningful modules of stage-4, 1095 genes were able to assemble a network. We chose eigenvector as a preferred centrality to identify the key genes. Opposed to degree index, which only measures the number of neighbors a node has, eigenvector judges a node based on the significance of its neighbors. According to this concept, if a node connects to more central nodes, it will have a higher value. We chose the top ten genes in the eigenvector category for each stage. Here are some examples of the experimental evidence for some of the key genes. With respect to stage-1 key genes (Figure 5), CD86 gene polymorphism has an association with colorectal cancer risk [35]. ITGAM and ITGAX both could use as potential markers for colorectal cancer [36]. CXCL8 (also known as IL-8) involves in

tumorigenesis and progression of CRC. Secretion of CXCL10 and CCL5 by colorectal tumor cells correlate with immune cells infiltration and absence of metastasis [37]. Among the stage-4 key genes (Figure 5), Fibronectine-1 contributes to metastasis in colorectal cancer and has anti-apoptotic effects [38]. COL3A1 also has a role in metastasis and can stimulate proliferation via pi3k-akt axis [39]. MMP9 is a key factor in the reconstruction of ECM, invasion, and regulation of tumor micro-environment, and angiogenesis in colorectal cancer [40]. Furthermore, MMP2 has more than ten times expression in CRCs than healthy tissues. It is one of the prognostic markers of CRC, additionally, it is more involved in local spread than distant metastasis [41].

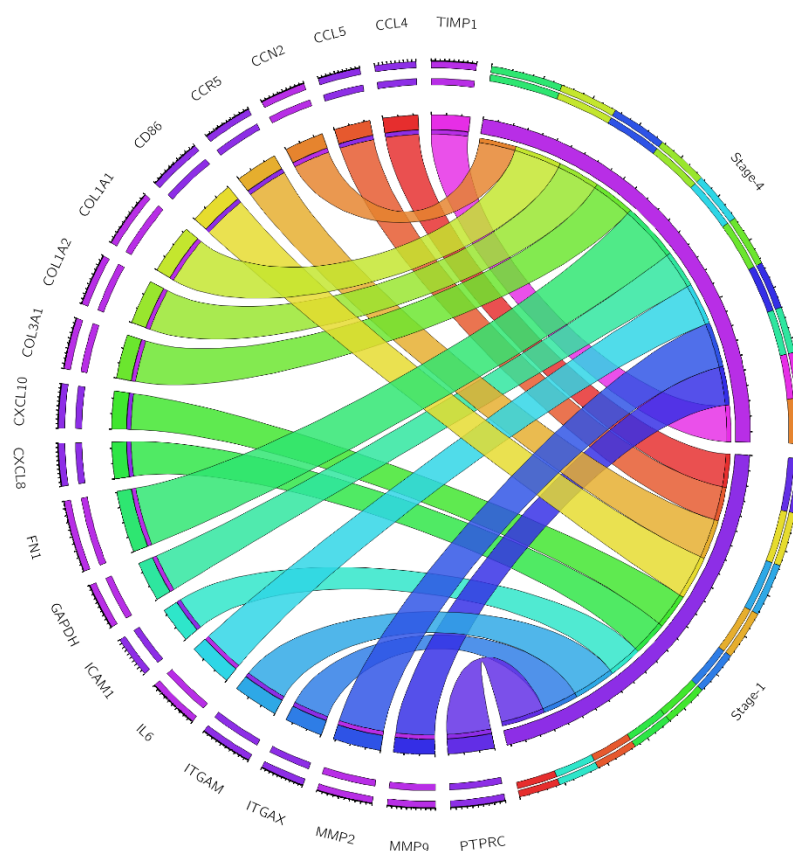


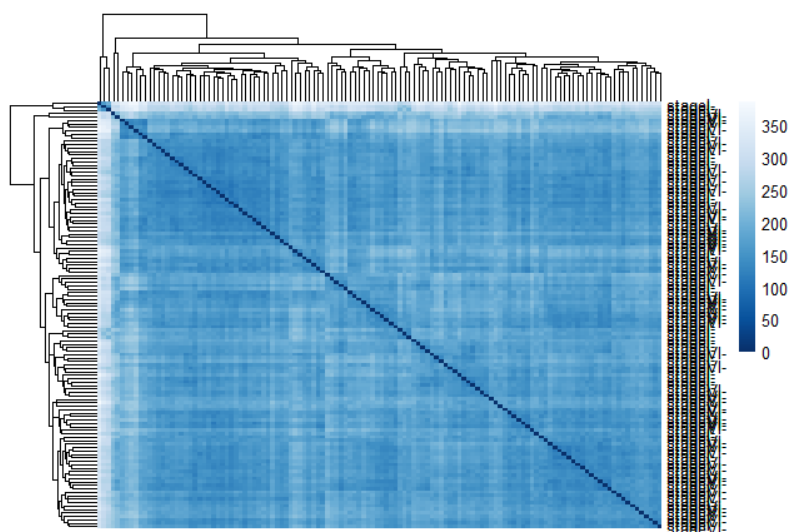
Figure 5: Circus table for genes with the highest eigenvectors' scores for stage-4 and stage-1. We considered eigenvector as a centrality index in order to identify key genes.

DEGs results

For DEGs analyze the whole of 25,550 annotated ENTREZ IDs were used. There were 68 samples as stage-1 and 56 samples as Stage-4. The whole study design was stage-4 vs. stage-1 (batch plus condition). Using deseq2 tool, there is no need to remove a batch effect. Instead, it is possible to model a batch effect as a covariate into the analyses during normalization by including batch and condition as a design. Log2fold change was set on 0.75, and adjusted-p value < 0.05. According to the setting, 925 DEGs acquired in which 525 genes were down-regulated (top four of them were L1TD1, CLDN18, DEFA6, and CCL25) (supp2). Also, 400 genes were up-regulated in which HAND1, IGFL1, TMPRSS11E, and FOXG1 were in the top four. This output was used to map

the most significant pathway in this process, and finally, we could find out what was the driver path from stage-1 condition to metastatic behavior of late-stage in colorectal cancer (Figure 6).

a



b

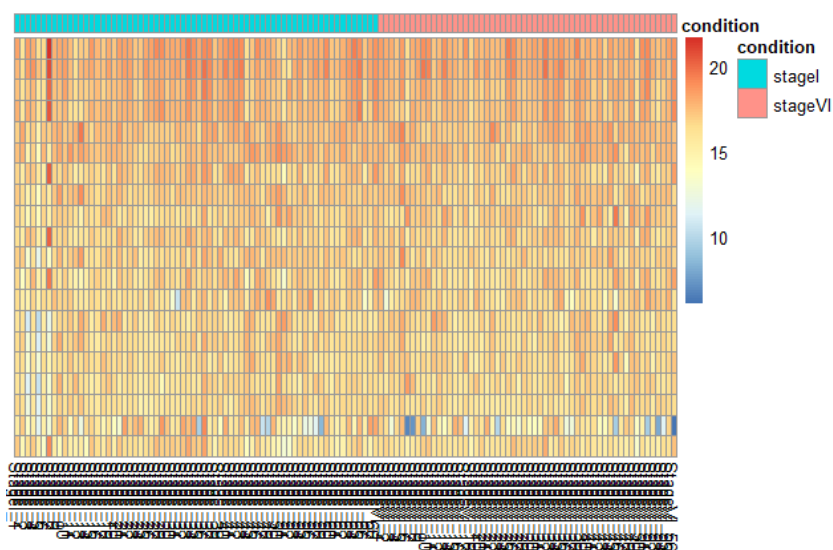


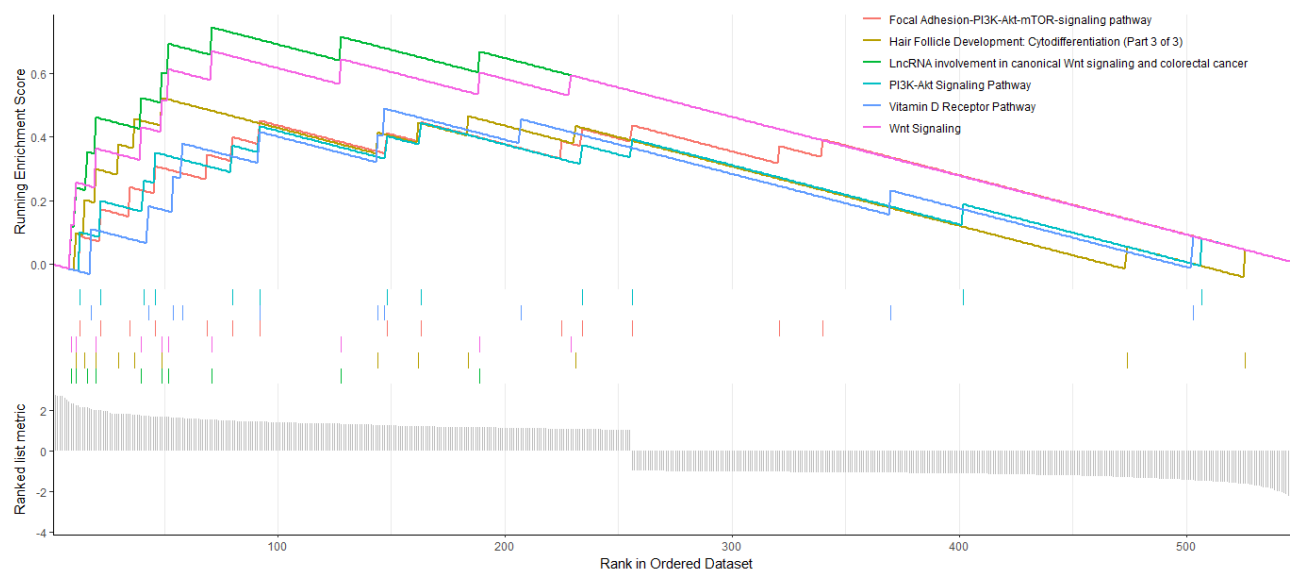
Figure 6: a) Heatmaps of sample to sample distance, b) Count matrix pheatmap for the DEGs analysis through deseq2 tool.

Pathway analysis

Based on the results of the differentially expressed gene, we conducted a pre-ranked enrichment analysis via clusterprofiler through wikipathway database. We filtered the DEGs by setting adjusted-p-value ≤ 0.05 and fold change ≥ 1.0 . Based on the analysis, “LNC in canonical WNT

signaling pathway in colorectal cancer” had the best NES (normalized enrichment score) followed by WNT signaling (Figure 7). As it appeared in the analysis, WNT-related pathways have a vital role in the progression from stage-1 to stage-4 of colorectal cancer. Nonetheless, we had two of the most important WNT-LNC related genes: HOTAIR and H19. Since it has already proven by experimental works, WNT signaling might be contributed to tumorigenesis and metastasis in colorectal cancer [42]. Also, it has recently proven that long non-coding RNAs play a role in metastasis of colorectal cancer by modulating WNT signaling pathway [43, 44]. Further, Pi3k-AKT-mTOR pathway also came out as meaningful, especially because of the role that they play in the focal adhesion.

a



b

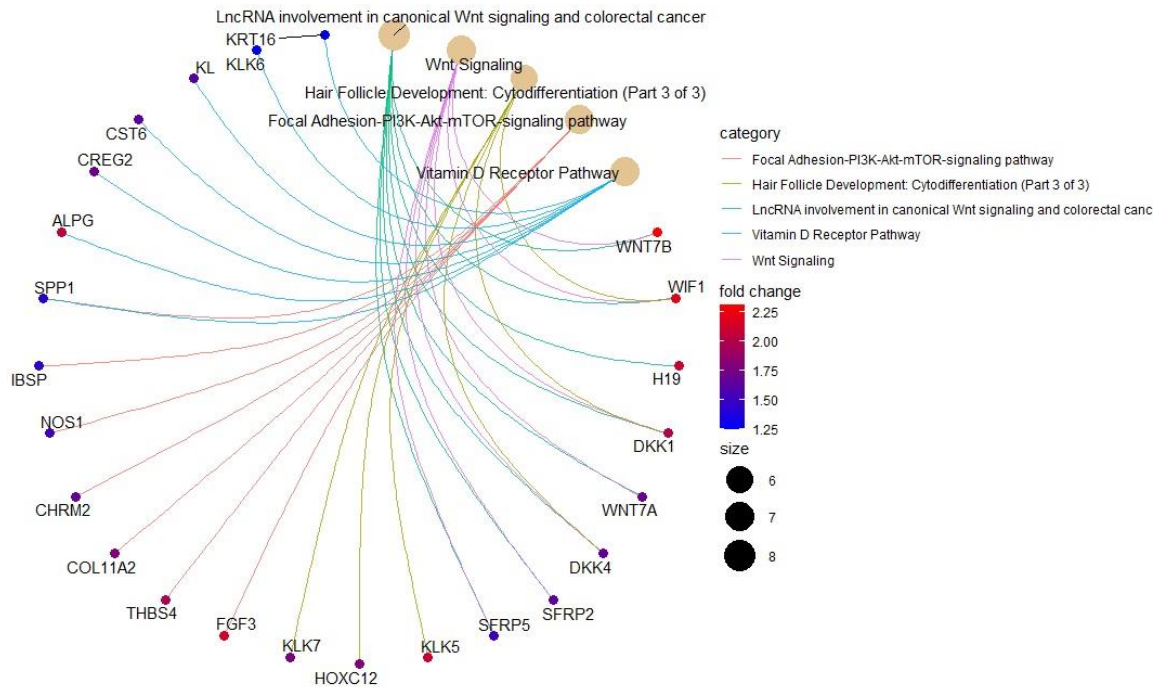


Figure 7: a) GSEA plot for our pathway analysis, based on the results, WNT-related pathways were at the outmost of important, b) Cnetplot to visualize pathway analysis based on gene-gene interactions and overlapping pathways components

Discussion

Despite many milestones in treating cancer, still, stage-4 cancers have been problematic. Many cases diagnosed while they have progressed to stage-4, but the disappointment is when a patient shows a complete response to treatment and enters the disease-free survival. Still, after years, cancer will eventually recur, this time more aggressive, more resilient, and even more advanced. It seems it is necessary to have a deep understanding of the underlying mechanism in which cancer progress from stage-1 to stage-4. Knowing the pathways which have direct or indirect involvement in this process could help us to find better and more efficient treatments. In this study, we tried to identify significance modules in the initial and late stages of colorectal cancer, and then analyze them accordingly. Based on the results, an immune-related process, specifically interferon-gamma response, has a distinct role in stage-1 colon adenocarcinoma, but when it comes to stage-4, the pathways and processes are linked to extracellular matrix and invasion. We also performed a topological analysis on the networks obtained from these modules and identified PTPRC, CD86, ITGAM, CCR5, and CXCL10 as the top five critical genes in stage-1, whereas FN1, COL1A1, MMP2, COL1A2, and IL6 were identified as the top five in stage-4 of adenocarcinoma. Interestingly, interferon gamma (IFN- γ) uses as a treatment for the colorectal cancer and also many other forms of malignancies, since it is one the essential component of the TH-1 response [45]. But recent researches found out that there could be more to tell about the role of IFN- γ in tumor and tumor microenvironment. In one study, researchers concluded that low-dose of IFN- γ could contribute to cancer stemness and metastasis and its effects are dose-dependent. Also there is some reports about contribution of IFN- γ in aggressiveness and proliferation of tumor cells [46, 47].

According to the differentially expressed gene analysis, the WNT signaling and LNC in canonical WNT signaling in colorectal cancer might be one of the main reasons for the malign transition from stage-1 to stage-4. It is clear that WNT signaling is one the most frequently hyperactivated path in all types of colon cancers, But the most types of CRCs have WNT ligands-independent hyperactivity, which means direct neutralization of WNT-ligands cannot be effective [48]. Nonetheless, in our analysis, not only the signaling itself but also the WNT-ligands were up-regulated. Still worth of mention that according to the previous studies, WNT signaling might be responsible for checkpoint inhibitor resistance and immunological difficulties in treating colorectal cancer [48-50]. As it appear in our study, lncRNA in canonical WNT signaling in colorectal cancer also has a significant role in the transition from stage-1 to stage-4; this notion additionally has been proven by experimental works [43, 51]. HOTAIR is one of these LNCs that could contribute to the overexpression of WNT signaling [52]. HOTAIR has a role in metastasis, drug resistance, and proliferation [53]. According to our DEGs analysis, HOTAIR is one of the most expressed differentially from the stage-1 to stage-4. H19 is another related LNC in WNT signaling in metastasis of colorectal cancer. H19 has a role in metastasis, proliferation, and drug resistance in many types of cancers [54]. Increased expression of H19, also correlates with poor prognosis [55]. H19 has many important pathways under its control like MYC and let-7, methylation of the genome, regulation of cell cycle, and of course WNT-Beta catenin [56-58]. In our DEGs analysis, the expression of H19, was meaningfully increased.

Conclusion

In the past studies, it has been shown that inhibition of HOTAIR or H19 could significantly reduce invasion, proliferation, and drug resistance in colorectal cancer [59, 60]. Now, our systems biology study could systematically support the idea that the WNT- signaling and LNCs mediated in WNT signaling, specifically HOTAIR and H19, might have a notable effect in the metastatic process in colorectal cancer. Therefore their inhibition could be considered as a therapeutic option to prevent the metastasis and progression of colorectal cancer.

Declarations

Funding

This study was funded and supported by Tehran University of Medical Sciences (TUMS) [Grant no. 98-01-87-42014, 98-01-87-42015, 98-01-87-42016].

Conflicts of interest

The authors declare that there is no conflict of interest.

Ethics approval

This study does not need any ethical approvals or informed consent.

Consent for publication

Not Applicable.

Supplementary data and materials

Supplementary file 1: Phenotypes and samples information

Supplementary file 2: Deg results

Code availability

All software have been mentioned in the manuscript and all scripts would be available upon request.

Authors' Contributions

MG designed the study, generated and analyzed the data, and drafted the manuscript. YA supervised the study, validated the data, and reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors acknowledge the reviewers for their helpful comments.

References

1. Freeman, H.J., *Early stage colon cancer*. World journal of gastroenterology, 2013. **19**(46): p. 8468-8473.
2. Hirano, D., et al., *Clinicopathologic and endoscopic features of early-stage colorectal serrated adenocarcinoma*. BMC gastroenterology, 2017. **17**(1): p. 158-158.
3. Lafitte, M., A. Sirvent, and S. Roche, *Collagen Kinase Receptors as Potential Therapeutic Targets in Metastatic Colon Cancer*. Frontiers in oncology, 2020. **10**: p. 125-125.
4. van der Stok, E.P., et al., *surveillance after curative treatment for colorectal cancer*. Nature Reviews Clinical Oncology, 2017. **14**(5): p. 297-315.
5. Lips, D.J., et al., *The influence of micrometastases on prognosis and survival in stage I-II colon cancer patients: the Enroute ⊕ Study*. BMC surgery, 2011. **11**: p. 11-11.
6. Van Cutsem, E. and J. Oliveira, *Primary colon cancer: ESMO Clinical Recommendations for diagnosis, adjuvant treatment and follow-up*. Annals of Oncology, 2009. **20**: p. iv49-iv50.
7. Akkoca, A.N., et al., *TNM and Modified Dukes staging along with the demographic characteristics of patients with colorectal carcinoma*. International journal of clinical and experimental medicine, 2014. **7**(9): p. 2828-2835.
8. Fotheringham, S., et al., *Challenges and solutions in patient treatment strategies for stage II colon cancer*. Gastroenterology report, 2019. **7**(3): p. 151-161.
9. Chan, G.H.J. and C.E. Chee, *Making sense of adjuvant chemotherapy in colorectal cancer*. Journal of gastrointestinal oncology, 2019. **10**(6): p. 1183-1192.
10. Feo, L., M. Polcino, and G.M. Nash, *Resection of the Primary Tumor in Stage IV Colorectal Cancer: When Is It Necessary?* The Surgical clinics of North America, 2017. **97**(3): p. 657-669.
11. Riihimäki, M., et al., *Patterns of metastasis in colon and rectal cancer*. Scientific Reports, 2016. **6**(1): p. 29765.
12. Zhou, M., L. Fu, and J. Zhang, *Who will benefit more from maintenance therapy of metastatic colorectal cancer?* Oncotarget, 2017. **9**(15): p. 12479-12486.
13. Mathonnet, M., et al., *Hallmarks in colorectal cancer: angiogenesis and cancer stem-like cells*. World journal of gastroenterology, 2014. **20**(15): p. 4189-4196.
14. Kow, A.W.C., *Hepatic metastasis from colorectal cancer*. Journal of gastrointestinal oncology, 2019. **10**(6): p. 1274-1298.
15. Kuenzi, B.M. and T. Ideker, *A census of pathway maps in cancer systems biology*. Nature reviews. Cancer, 2020. **20**(4): p. 233-246.
16. Goldman, M., et al., *The UCSC Xena platform for public and private cancer genomics data visualization and interpretation*. bioRxiv, 2019: p. 326470.
17. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis*. BMC Bioinformatics, 2008. **9**(1): p. 559.
18. Team, R.C., *A language and environment for statistical computing*. R Foundation for Statistical Computing. 2020.
19. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets*. Nucleic acids research, 2019. **47**(D1): p. D607-D613.
20. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. Genome research, 2003. **13**(11): p. 2498-2504.
21. Tang, Y., et al., *CytoNCA: A cytoscape plugin for centrality analysis and evaluation of protein interaction networks*. Biosystems, 2015. **127**: p. 67-72.
22. Yu, G., et al., *clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters*. OMICS: A Journal of Integrative Biology, 2012. **16**(5): p. 284-287.
23. The Gene Ontology Consortium, *The Gene Ontology Resource: 20 years and still GOing strong*. Nucleic acids research, 2019. **47**(D1): p. D330-D338.
24. Ashburner, M., et al., *Gene ontology: tool for the unification of biology*. The Gene Ontology Consortium. Nature genetics, 2000. **25**(1): p. 25-29.
25. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection*. Cell systems, 2015. **1**(6): p. 417-425.
26. Yu, G., *enrichplot: Visualization of Functional Enrichment Result*. R package version 1.6.1. 2019.

27. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics (Oxford, England), 2010. **26**(1): p. 139-140.
28. Carlson, M., *org.Hs.eg.db: Genome wide annotation for Human. R package version*. 2019.
29. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. Genome biology, 2014. **15**(12): p. 550-550.
30. Zhu, A., J.G. Ibrahim, and M.I. Love, *Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences*. Bioinformatics, 2018. **35**(12): p. 2084-2092.
31. Ignatiadis, N., et al., *Data-driven hypothesis weighting increases detection power in genome-scale multiple testing*. Nature methods, 2016. **13**(7): p. 577-580.
32. Team, R., *RStudio: Integrated Development for R. RStudio, Inc., Boston, MA*. 2015.
33. Yokota, J., *Tumor progression and metastasis*. Carcinogenesis, 2000. **21**(3): p. 497-503.
34. Giannelli, G., et al., *Role of epithelial to mesenchymal transition in hepatocellular carcinoma*. Journal of Hepatology, 2016. **65**(4): p. 798-808.
35. Azimzadeh, P., et al., *Association of co-stimulatory human B-lymphocyte antigen B7-2 (CD86) gene polymorphism with colorectal cancer risk*. Gastroenterology and hepatology from bed to bench, 2013. **6**(2): p. 86-91.
36. Gong, Y.-Z., et al., *Diagnostic and prognostic values of integrin α subfamily mRNA expression in colon adenocarcinoma*. Oncology reports, 2019. **42**(3): p. 923-936.
37. Zumwalt, T.J., et al., *Active secretion of CXCL10 and CCL5 from colorectal cancer microenvironments associates with GranzymeB+ CD8+ T-cell infiltration*. Oncotarget, 2015. **6**(5): p. 2981-2991.
38. Cai, X., et al., *Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration, and invasion*. Journal of Cellular Biochemistry, 2018. **119**(6): p. 4717-4728.
39. Wang, X.-Q., et al., *Epithelial but not stromal expression of collagen alpha-1(III) is a diagnostic and prognostic indicator of colorectal carcinoma*. Oncotarget, 2016. **7**(8): p. 8823-8838.
40. Jonsson, A., et al., *Stability of matrix metalloproteinase-9 as biological marker in colorectal cancer*. Medical Oncology, 2018. **35**(4): p. 50.
41. Dong, W., et al., *Matrix metalloproteinase 2 promotes cell growth and invasion in colorectal cancer*. Acta Biochimica et Biophysica Sinica, 2011. **43**(11): p. 840-848.
42. Basu, S., G. Haase, and A. Ben-Ze'ev, *Wnt signaling in cancer stem cells and colon cancer metastasis*. F1000Research, 2016. **5**: p. F1000 Faculty Rev-699.
43. Zarkou, V., et al., *Crosstalk mechanisms between the WNT signaling pathway and long non-coding RNAs*. Non-coding RNA Research, 2018. **3**(2): p. 42-53.
44. Yu, J., et al., *LncRNA SLCO4A1-AS1 facilitates growth and metastasis of colorectal cancer through β -catenin-dependent Wnt pathway*. Journal of experimental & clinical cancer research : CR, 2018. **37**(1): p. 222-222.
45. Zaidi, M.R. and G. Merlino, *The two faces of interferon- γ in cancer*. Clinical cancer research : an official journal of the American Association for Cancer Research, 2011. **17**(19): p. 6118-6124.
46. Mojic, M., K. Takeda, and Y. Hayakawa, *The Dark Side of IFN- γ : Its Role in Promoting Cancer Immuno-evasion*. International journal of molecular sciences, 2017. **19**(1): p. 89.
47. Song, M., et al., *Low-dose IFN- γ induces tumor cell stemness in the tumor microenvironment of non-small cell lung cancer*. Cancer Research, 2019: p. canres.0596.2019.
48. Schatoff, E.M., B.I. Leach, and L.E. Dow, *Wnt Signaling and Colorectal Cancer*. Current colorectal cancer reports, 2017. **13**(2): p. 101-110.
49. Oderup, C., M. LaJevic, and E.C. Butcher, *Canonical and noncanonical Wnt proteins program dendritic cell responses for tolerance*. Journal of immunology (Baltimore, Md. : 1950), 2013. **190**(12): p. 6126-6134.
50. Spranger, S., R. Bao, and T.F. Gajewski, *Melanoma-intrinsic β -catenin signalling prevents anti-tumour immunity*. Nature, 2015. **523**(7559): p. 231-235.
51. Hu, X.-Y., et al., *The roles of Wnt/ β -catenin signaling pathway related lncRNAs in cancer*. International Journal of Biological Sciences, 2018. **14**(14): p. 2003-2011.
52. Ge, X.-S., et al., *HOTAIR, a prognostic factor in esophageal squamous cell carcinoma, inhibits WIF-1 expression and activates Wnt pathway*. Cancer Science, 2013. **104**(12): p. 1675-1682.
53. Xiao, Z., et al., *LncRNA HOTAIR is a Prognostic Biomarker for the Proliferation and Chemoresistance of Colorectal Cancer via MiR-203a-3p-Mediated Wnt/ β -Catenin Signaling Pathway*. Cellular Physiology and Biochemistry, 2018. **46**(3): p. 1275-1285.

54. Javed, Z., et al., *LncRNA & Wnt signaling in colorectal cancer*. Cancer cell international, 2020. **20**: p. 326-326.
55. Han, D., et al., *Long noncoding RNA H19 indicates a poor prognosis of colorectal cancer and promotes tumor growth by recruiting and binding to eIF4A3*. Oncotarget, 2016. **7**.
56. Ma, C., et al., *H19 promotes pancreatic cancer metastasis by derepressing let-7's suppression on its target HMGA2-mediated EMT*. Tumor Biology, 2014. **35**(9): p. 9163-9169.
57. Kallen, A.N., et al., *The imprinted H19 lncRNA antagonizes let-7 microRNAs*. Molecular cell, 2013. **52**(1): p. 101-112.
58. Liang, W.-C., et al., *The lncRNA H19 promotes epithelial to mesenchymal transition by functioning as miRNA sponges in colorectal cancer*. Oncotarget, 2015. **6**(26): p. 22513-22525.
59. Yang, Q., et al., *H19 promotes the migration and invasion of colon cancer by sponging miR-138 to upregulate the expression of HMGA1*. Int J Oncol, 2017. **50**(5): p. 1801-1809.
60. Yang, X.-D., et al., *Knockdown of long non-coding RNA HOTAIR inhibits proliferation and invasiveness and improves radiosensitivity in colorectal cancer*. Oncol Rep, 2016. **35**(1): p. 479-487.