ORIGINAL ARTICLE

RBINF®

Exploring the biological processes driving metastasis in colorectal cancer patients

Syed Jafri[1]

[1]**Cumming School of Medicine, University of Calgary, Calgary, Canada**

ABSTRACT

The progression of metastasis in primary colorectal cancer (CRC) is an invasive cancerogenic process which increases decreases the 5-year relative survival rate for Canadians by 77%. Hallmarks of cancer such as inflammation and angiogenesis are known to play a crucial role in the process of cancer growth, yet the specific biological processes that contribute to these processes are lessor known. This study compares liver metastatic CRC to primary CRC and normal tissue using differential expression, gene ontology and KEGG pathway analyses. We identified anion transport and extracellular matrix degradation as biological processes that significantly contribute to ROS generation, inflammation and angiogenesis. Furthermore, complement and coagulation cascade pathways were found to be significantly affected in the progression of CRC. We propose that therapies targeting these biological processes need to be further explored in addition to the effectiveness of high-antioxidant diets in CRC patients.

KEY WORDS: Colorectal cancer, angiogenesis, metastasis, inflammation, ROS

Introduction

Colorectal cancer (CRC) is the second most common cancer in Canada, with an estimated 26,800 new diagnoses and 9,400 deaths each year.[1] While the 5-year relative survival rate for localized CRC (limited to colon or rectum) is 90%, the 5-year relative survival rate for metastasized CRC is only 13%.[2] Therefore, research into early signature biomarkers and a deeper understanding of the pathways involved in CRC metastasis is necessary.

Inflammation is one of the hallmarks of cancer due to its complex role in promoting angiogenesis, a precursor for cancer metastasis. Angiogenesis induced by inflammation provides nutrients and transport of tumor cells and factors to distal sites. Previous attempts at classifying biomarkers and mechanisms of CRC through gene analysis have primarily been conducted using Affymetrix arrays.[3,4,5] The overlap of genes between microarray and RNA-seq data has been previously estimated to be only 66%~68%, and RNA-seq approaches are known to identify pathways previously missed by microarrays.[6]

This research extends prior studies that have focused on the primary CRC site by comparing primary and metastatic liver CRC to analyze biomarkers and pathways specific to metastatic sites present in CRC. The analysis of differentially expressed genes (DEGs) in these groups and their enriched biological functions and pathways will lead to a deeper understanding of the biomarkers that are involved in CRC and potential therapies that can be developed for metastatic CRC patients.

Methods

Dataset collection and preprocessing

A dataset containing 54 samples from 18 colorectal cancer patients was obtained from the NCBI-Gene Expression Omnibus database (NCBI-GEO) with accession code GSE50760.[7] In this dataset, each patient contains a normal colon, a primary CRC tumor, and a liver metastasis sample sequenced using an Illumina Hiseq-200 in a paired end layout.[7]

The SRA files from NCBI were converted to fastq files using the fastq-dump software in the SRA toolkit.[8] Before sequence alignment, a Kallisto transcriptome index was built using the full Homo Sapiens transcriptome reference GRCh38 from Ensembl.[9] Transcript abundances were then generated using Kallisto for all 54 samples.[10] All data preparation was conducted on the Synergy compute cluster at the Cummings School of Medicine Centre for Health Genomics and Informatics in Calgary, Alberta.

Differentially Expressed Genes (DEGs)

Transcript abundance data obtained from Kallisto was imported into R using the tximport package.[11] Differential expression (DE) analysis was conducted through the DESeq2 R package, which estimates size factors and dispersion and uses these estimates to fit a Negative Binomial (NB) distribution.[12] LFC shrinkage in the DESeq2 package was applied to obtain shrunken log2 fold changes for further downstream analysis.

Methods (continued)

Clustering Analysis

For clustering analysis using PCA plot and heatmap, a parametric variance stabilizing transformation was initially applied to the DE results. A PCA plot was generated from the top 500 genes in each of the three sample conditions using the DESeq2 package in R.[12] A heat-map was generated from the top 25 differentially expressed genes in each coefficient using the gplots package in R.

Gene Ontology Enrichment

The cellular changes and functions of identified DEGs were explored by performing GO enrichment analysis using the clusterProfiler in R.[14] The impact of DEGs on biological process (BP), molecular function (MF) and cellular components (CC) was analyzed. The cut-off criterion of $p < 0.05$ and abs(log2FC) > 1 was used in the analysis in conjunction with a p-value adjustment using the Holm-Bonferroni method. The identified enriched GO terms were plotted using a standard dotplot in R.

KEGG Pathway Analysis

The changes in the function of cells was further explored by investigating the changes in pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The Holm-Bonferroni method was used for p value correction, and cut-off criterion of $p < 0.05$ and abs(log2FC) 1 was used to identify significant signaling pathways.

Results

Clustering Analysis

Distinct clustering patterns were identified for the three groups in the PCA plot shown in Figure 1. Minor overlap was discovered between primary CRC compared to normal group, and primary CRC compared to metastatic liver CRC. The heatmap analysis showed large levels of overexpression for the metastatic liver CRC group as shown in Figure 2.

Differential Expression Analysis

A total of 35,193 genes were identified with nonzero total read count. The counts of up-regulated and down-regulated genes with a cut-off criterion of $p < 0.10$ and abs(log2FC) > 1 in the three coefficients are shown in Table 1.

Table 1. Summary of log-folded shrunken DEGs out of 35193 genes with nonzero read count

Coefficient	Up-regulated (+)	Down-regulated (-)
Primary vs Normal	556 (1.58%)	252 (0.72%)
Metastatic vs Normal	1567 (4.45%)	811 (2.30%)
Metastatic vs Primary	386 (1.10%)	563 (1.60%)

Gene Ontology Enrichment

We found 10 enriched terms for the primary CRC compared to normal group, 52 enriched terms for the metastatic CRC compared to normal group and 39 enriched terms for the metastatic CRC compared to primary CRC groups. The top 10 enriched terms for each coefficient are shown in Figure 3.

KEGG Analysis

We found 2 enriched pathways comparing the primary CRC and normal groups, 3 enriched pathways comparing the metastatic CRC and normal group and 1 enriched pathway comparing the primary and metastatic liver CRC groups, as shown in Table 2. The 19 genes associated with the complement and coagulation cascade pathway in metastatic versus primary group were VWF, F2RL2, PLG, SERPINE1, FGB, C4BPB, VTN, C8B, C3, F9, C5, SERPINC1, C4BPA, C9, SERPIND1, MBL2, F13B, FGA and C8A.

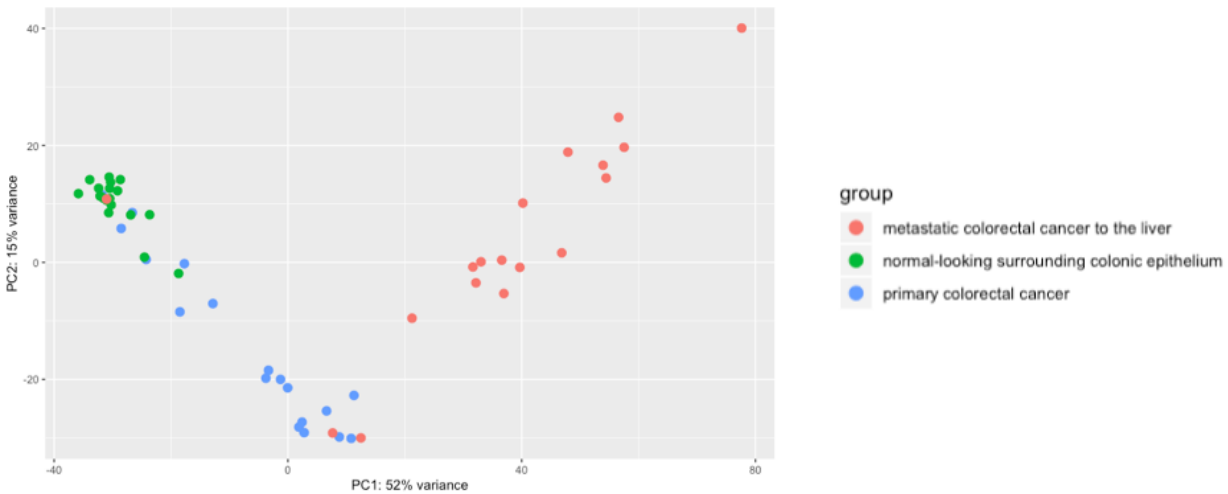

Fig. 1. Principle Component Analysis of 54 samples consisting of primary CRC, metastatic liver cancer and normal groups.

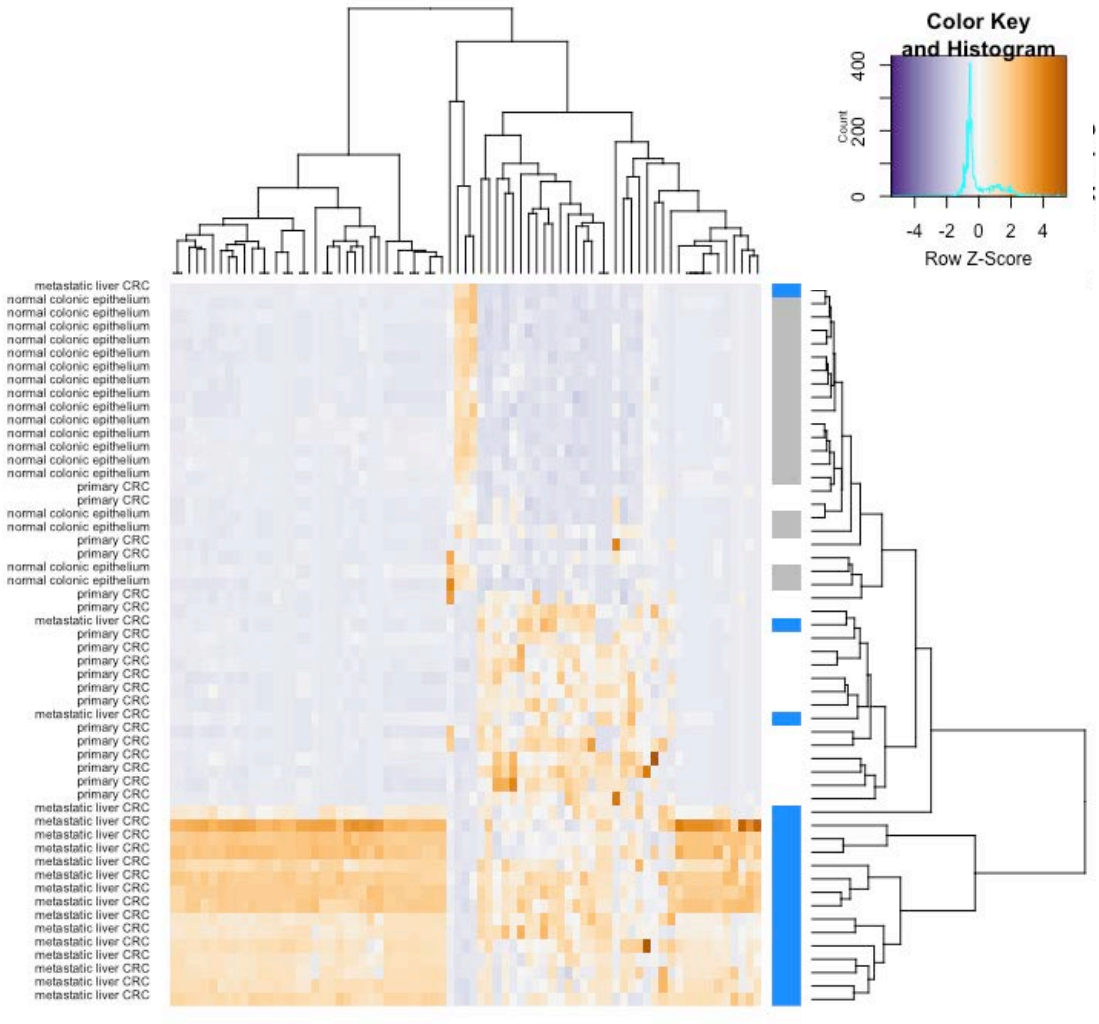

Fig. 2. Heatmap analysis of gene expression consisting of the 25 DEGs from each of the three groups ordered by absolute shrunken log2 fold change.

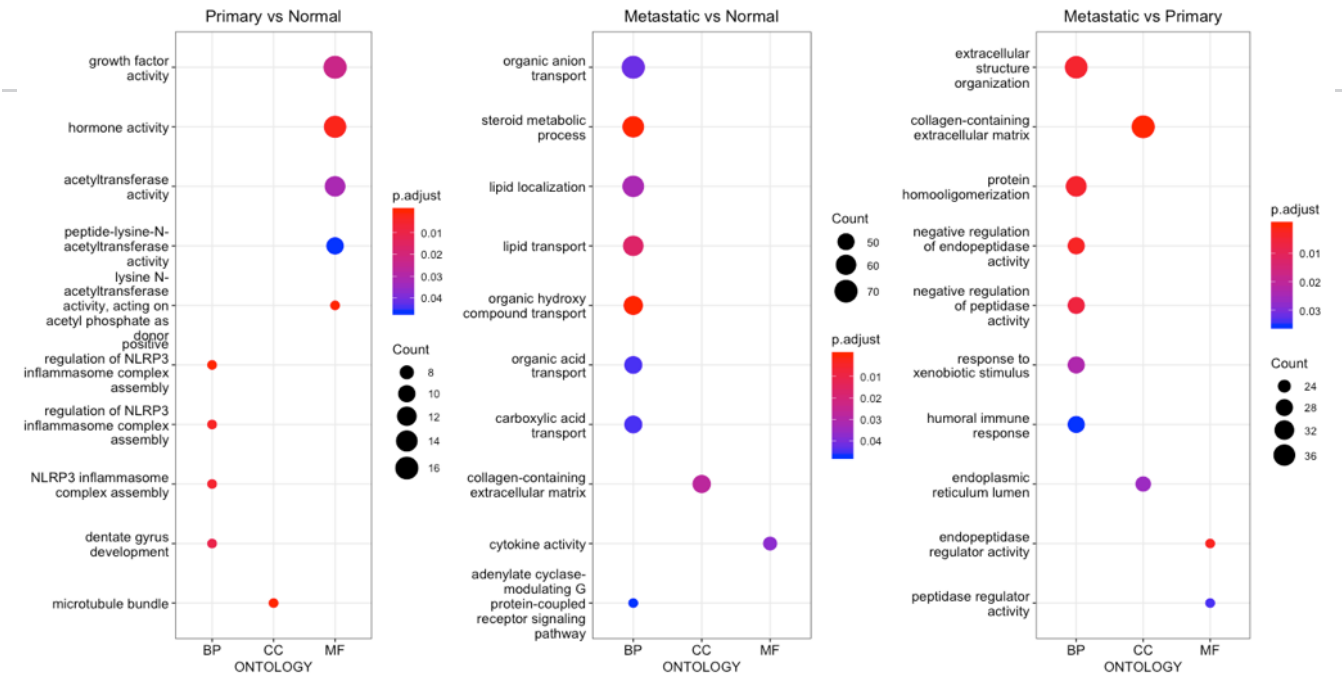

Fig. 3. For the group, the top 10 enriched GO processes are plotted ranked by gene ratio. The size of each dot represents the count of genes found to be DE in relation to the GO term. The color of each dot represents the p-value as adjusted using the Holm-Bonferroni method.

Table 2. Significant KEGG pathways found from DEGs

Coefficient	KEGG ID	Description	Gene Ratio	Adj. P-Value
Primary vs Normal	hsa04080	Neuroactive ligand-receptor interaction	29/275	0.00185
	hsa04060	Cytokine-cytokine receptor interaction	24/275	0.02644
Metastatic vs Normal	hsa04610	Complement and coagulation cascades	30/810	1.93488e-08
	hsa04976	Bile secretion	19/810	2.72108e-02
	hsa04979	Cholesterol metabolism	15/810	3.09442e-02
Metastatic vs Primary	hsa04610	Complement and coagulation cascades	19/322	7.09251e-08

Discussion

Inflammation and sustained angiogenesis display an intertwined relationship in the tumor microenvironment which often contributes to the proliferation of malignant cells in the human body.[15] The aim of the current study was to compare differences in metastatic CRC to primary CRC and normal tissue to gain a deeper understanding of the expression differences between the three groups that may lead to metastatic development of CRC.

In the current study, the PCA analysis confirmed that primary CRC, metastatic liver CRC and normal samples group distinctly with high levels of similarity. We hypothesize that the single metastatic oddity seen in the normal group was due to the high level of variability between consensus molecular subtypes (CMC) in CRC.[16] Our dataset did not have CMS subtypes available as a data point to analyze. The heatmap cluster analysis further solidified our hypothesis that a distinct pattern of expression separates metastatic CRC tissue from primary CRC and normal groups. The distinct pattern is further supported by the comparison of primary CRC to metastatic CRC which shows a one-fold overexpression in 386 (1.10%) and a one-fold underexpression in 563 (1.60%) of all DEGs.

The current study identified 39 significantly enriched GO terms between the metastatic liver CRC and primary CRC groups (Figure 3). Extracellular structure and matrix GO terms suggest a key role of ECM macromolecules in CRC. Collagens and other types of ECM proteins are known to migrate in vivo through the degradation of the ECM and the development of migration tracks.[17] Additionally, changes in ECM proteins modulates the flow of both immune cells and drugs into the cell, altering the effectiveness of the immune system in the host.[18] Further clinical data on the effectiveness of therapies targeting these ECM macromolecules in stopping the metastatic cascade is needed for CRC patients. Comparing the metastatic samples to the normal samples, we find GO themes relating to transport of organic anions, organic hydroxides and carboxylic acids. Reactive oxygen species (ROS) play a crucial role in migration and invasion of cancer cells.[19]

Recent evidence demonstrates that ROS contribute to the destabilization of the cell-cell junction through VE-cadherin.[19] The ROS cascade of ROS-dependent Src kinase activation and subsequent VE-cadherin tyrosine phosphorylation that leads to angiogenesis can be prevented by antioxidant compounds.[19] This presents an exploratory opportunity into the diets of individuals affected by CRC, to study the potential effects of a high-antioxidant diet.

In the current study, complement and coagulation cascade pathways were found to be the most significant pathway affected between the metastatic liver CRC and primary CRC groups (Table 2). In humans, the complement system plays a crucial role in modulating the immune response and maintaining homeostasis.[20] Complement activation by the expression of C3a and C5a genes leads to the decreased expression of E-cadherin, a gene that prevents cancer invasion.[20] The underexpression of E-cadherin is known to induce the Epithelial to mesenchymal transition, leading to metastasis.[20]

The current study has shown that large expression differences exist between metastatic liver CRC, primary CRC and normal tissue, which may serve as the basis of targeted therapies towards metastatic CRC. With the poor prognosis of individuals diagnosed with metastatic CRC, there is a need for further research into prevention strategies such as high-antioxidant diets.

References

1. Cancer.ca. (2018). Canadian Cancer Statistics:. [online] Available at: http://www.cancer.ca/~/media/cancer.ca/CW/cancer%20information/cancer%20101/Canadian%20cancer%20statistics/Canadian-Cancer-Statistics-2018-EN.pdf [Accessed 25 Feb. 2019].

2. National Cancer Institute. (2014). Surveillance Epidemiology and End Results (SEER). Bethesda, MD: National Cancer Institute. Section 6: Colon and Rectum. [online] Available at: http://www.cancer.ca/en/cancer-information/cancer-type/colorectal/prognosis-and-survival/survival-statistics/?region=ab#ixzz5gs7OlKFR [Accessed 25 Feb. 2019].

References (continued)

3. Guo Y, Bao Y, Ma M, Yang W. Identification of Key Candidate Genes and Pathways in Colorectal Cancer by Integrated Bioinformatical Analysis. International Journal of Molecular Sciences [Internet]. 2017 Mar 28;18(4):722. Available from: http://dx.doi.org/10.3390/ijms18040722

4. Yang Q, Feng M, Ma X, Li H, Xie W. Gene expression profile comparison between colorectal cancer and adjacent normal tissues. Oncology Letters [Internet]. 2017 Sep 8; Available from: http://dx.doi.org/10.3892/ol.2017.6915

5. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value. Kemp C, editor. PLoS Medicine [Internet]. 2013 May 21;10(5):e1001453. Available from: http://dx.doi.org/10.1371/journal.pmed.1001453

6. Xu X, Zhang Y, Williams J, Antoniou E, McCombie W, Wu S, et al. Parallel comparison of Illumina RNA-Seq and Affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. BMC Bioinformatics [Internet]. 2013;14(Suppl 9):S1. Available from: http://dx.doi.org/10.1186/1471-2105-14-S9-S1

7. Kim S-K, Kim S-Y, Kim J-H, Roh SA, Cho D-H, Kim YS, et al. A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. Molecular Oncology [Internet]. 2014 Jul 4;8(8):1653–66. Available from: http://dx.doi.org/10.1016/j.molonc.2014.06.016

8. Leinonen R, Sugawara H, Shumway M. The Sequence Read Archive. Nucleic Acids Research [Internet]. 2010 Nov 9;39(Database):D19–21. Available from: http://dx.doi.org/10.1093/nar/gkq1019

9. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Research [Internet]. 2017 Nov 16;46(D1):D754–61. Available from: http://dx.doi.org/10.1093/nar/gkx1098

10. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology [Internet]. 2016 Apr 4;34(5):525–7. Available from: http://dx.doi.org/10.1038/nbt.3519

11. Love M. Tximport: Differential Analyses For Rna-Seq: Transcript-Level Estimates Improve Gene-Level Inferences [Internet]. Zenodo; 2015. Available from: https://zenodo.org/record/35123

12. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology [Internet]. 2014 Dec;15(12). Available from: http://dx.doi.org/10.1186/s13059-014-0550-8

13. Warnes et al. (2012) Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B. 2012. gplots: various R programming tools for plotting data. http://CRAN.R-project.org/package=gplots.

14. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology [Internet]. 2012 May;16(5):284–7. Available from: http://dx.doi.org/10.1089/omi.2011.0118

15. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. Carcinogenesis [Internet]. 2009 May 25;30(7):1073–81. Available from: http://dx.doi.org/10.1093/carcin/bgp127

16. Müller MF, Ibrahim AEK, Arends MJ. Molecular pathological classification of colorectal cancer. Virchows Archiv [Internet]. 2016 Jun 20;469(2):125–34. Available from: http://dx.doi.org/10.1007/s00428-016-1956-3

17. Paul CD, Mistriotis P, Konstantopoulos K. Cancer cell motility: lessons from migration in confined spaces. Nature Reviews Cancer [Internet]. 2016 Dec 2;17(2):131–40. Available from: http://dx.doi.org/10.1038/nrc.2016.123

18. Venning FA, Wullkopf L, Erler JT. Targeting ECM Disrupts Cancer Progression. Frontiers in Oncology [Internet]. 2015 Oct 20;5. Available from: http://dx.doi.org/10.3389/fonc.2015.00224

19. Lee DJ, Kang SW. RETRACTED ARTICLE: Reactive oxygen species and tumor metastasis. Molecules and Cells [Internet]. 2013 Feb;35(2):93–8. Available from: http://dx.doi.org/10.1007/s10059-013-0034-9

20. Pio R, Corrales L, Lambris JD. The Role of Complement in Tumor Growth. In: Advances in Experimental Medicine and Biology [Internet]. Springer New York; 2013. p. 229–62. Available from: http://dx.doi.org/10.1007/978-1-4614-5915-6_11