Sex-specific Transcriptome Meta-analysis for Colorectal Cancer

Eunyoung Choi[1], Jaeseung Song[1], Wonhee Jang[1,*]

[1] Department of Life Sciences, Dongguk University-Seoul, 04620, Seoul, Republic of Korea

Abstract

Colorectal cancer (CRC) is the third most common cancer and the second cause of cancer-related death, which is known to occur more commonly in males. This suggests that gender may be an important factor in CRC. To investigate the sex-specific transcriptional differences of CRC, we performed transcriptome meta-analysis by integrating three gene expression data consists of 131 tumor and 124 normal adjacent samples. Batch effects between data sets were removed with ComBat function and differential expression analysis was conducted in sex-specific manner with limma. We identified 64 and 2347 differentially expressed genes (DEGs) in females and males, respectively. Because the number of female DEGs was not enough to provide sufficient statistical power, gene set enrichment analysis (GSEA) and network analysis were only performed for males to find out the enriched biological terms and functional connection between DEGs. GSEA showed significant enrichment in cell division and gene repair. In the network analysis, we identified 8 modules on the global network and 3 modules on the colon-specific network and the functions of these modules were well conformed to the GSEA results. By conducting systematic transcriptome analysis, herein, we suggested potential CRC biomarkers for male that may be useful for effective CRC diagnosis.

Workflow

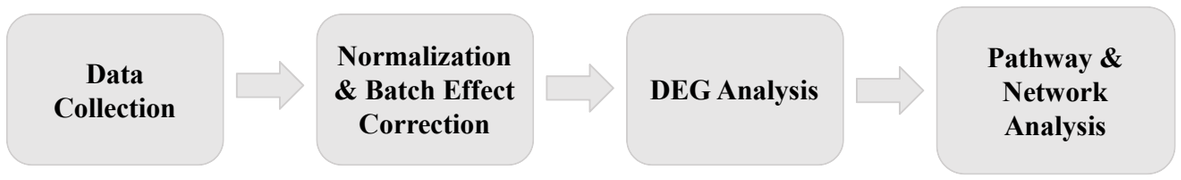

Figure 1. Workflow of overall study. Microarray data for CRC and adjacent normal tissues were collected from gene expression omnibus (GEO). For meta-analysis, each data was normalized separately and batch effect was corrected. DEG analysis was performed using limma package, and pathway analysis for male DEG was performed through GSEA. The network was analyzed using HumanBase functional module detection.

Dataset

GEO ID	Source	Platform	Case/Control	Male Case/Control	Female Case/Control
GSE44076	Colorectal adenocarcinoma	Affymetrix Human Genome U219 Array	98/98	71/71	27/27
GSE74460		Affymetrix Human Transcriptome Array 2.0	27/20	21/13	6/7
GSE84984		Affymetrix Human Transcriptome Array 2.0	6/6	1/1	5/5
Total	**-**	**-**	**131/124**	**93/85**	**38/39**

Table 1. Table of datasets used in this study. Three microarray datasets were integrated and used in this study. Each dataset was extracted from CRC tissue and adjacent normal tissue.

Meta-analysis

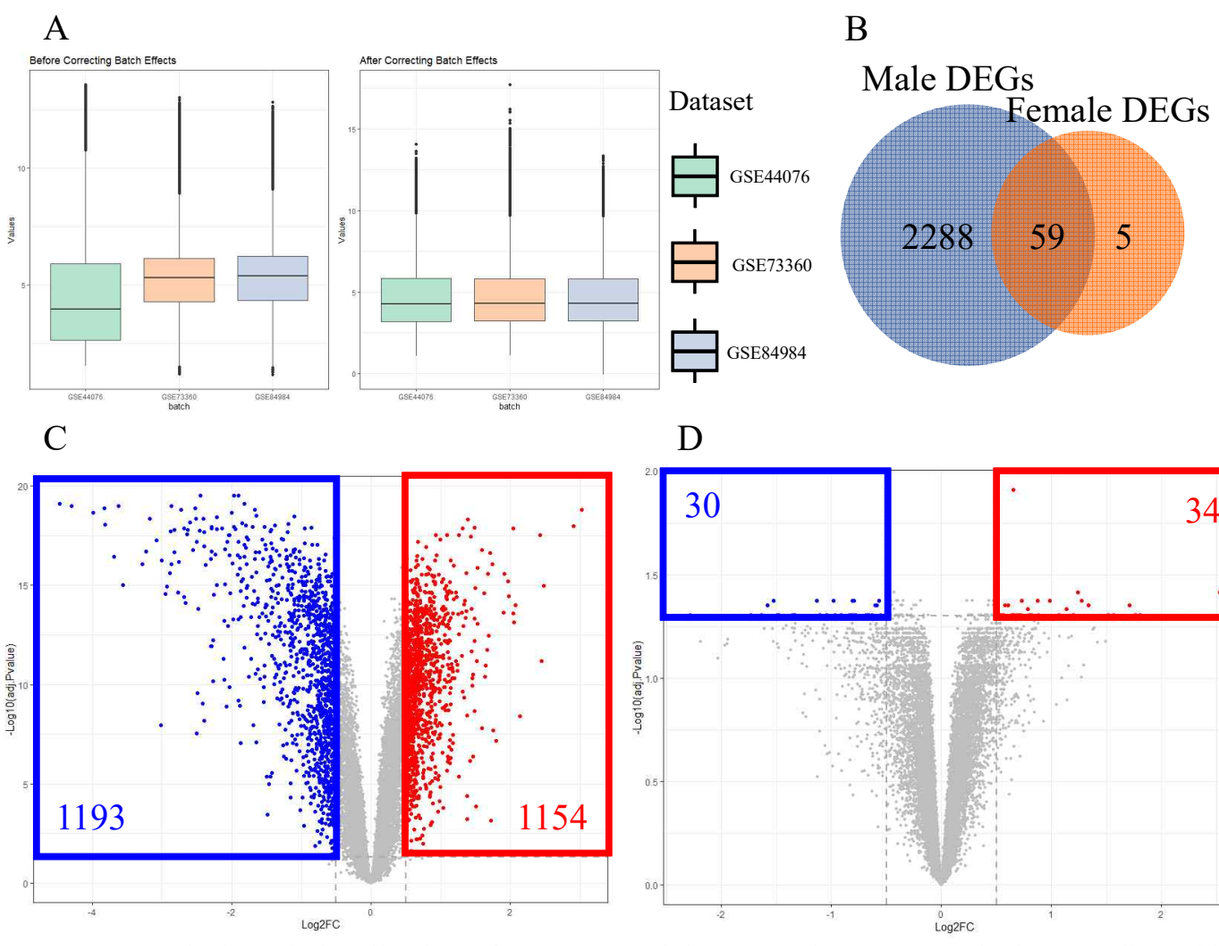

Figure 2. Analysis and visualization of preprocessed datasets using Meta-analysis. We adjusted batch effects to use microarray datasets from three different studies. (A) The boxplots showing the expression levels of each statistic before and after adjusting batch effect. (B) The Venn diagram shows the number of overlapping DEGs divided into males and females (adj P-val < 0.05, |log2-FC| > 0.5). Volcano plots for (C) male and (D) female, showing the number of over-expressed and under-expressed genes.

GSEA

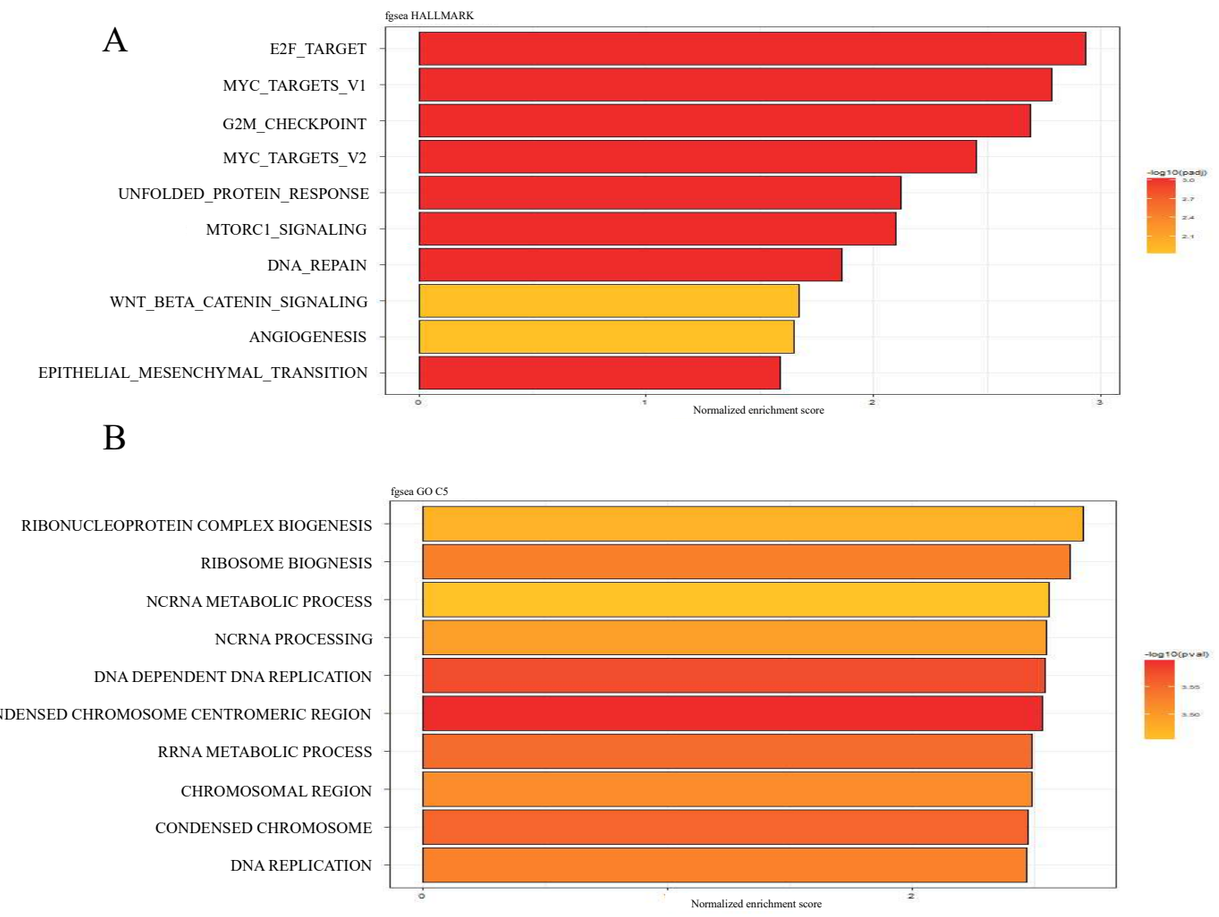

Figure 3. Pathway analysis and GSEA result plots. (A) A bar plot of the result for Hallmark pathways. (B) A bar plot of the result for GO. Both plots were sorted in descending order based on the normalized enrichment score (NES). (C) A GSEA plot for E2F_target, which showed the strongest result in plot (A). (D) A GSEA plot for ribonucleoprotein complex biogenesis, which showed the strongest result in plot (B).

Network Analysis

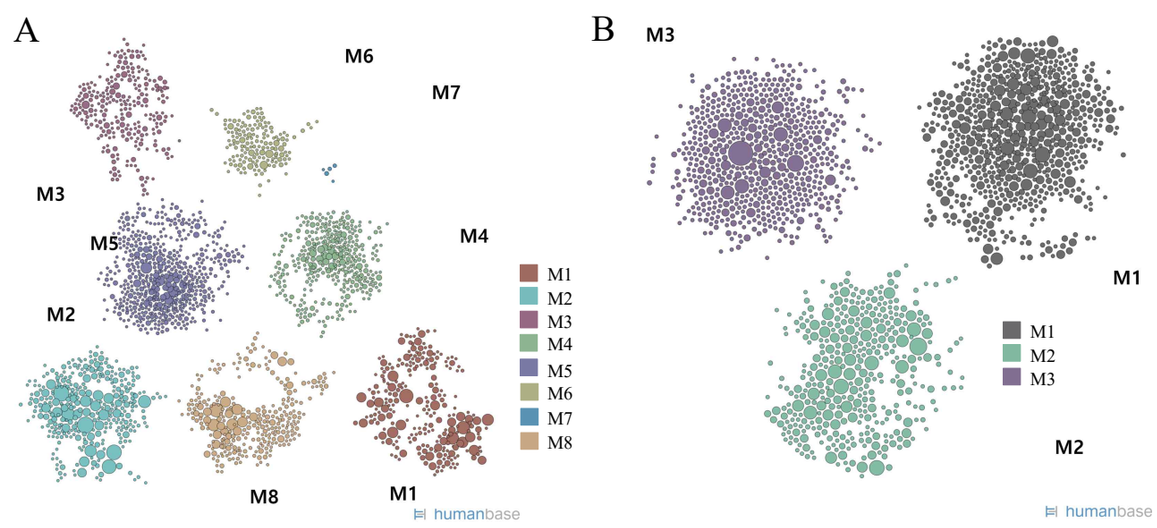

Figure 4. Network analysis and visualization using Male DEGs. (A) A result of functional module detection using HumanBase, 8 global networks were identified. Top terms in the global networks are chromosome segregation(M1) and ribosome biogenesis(M2). (B) A result of network analysis of colon, three networks were identified. The top term was chromosome segregation (M1), which is consistent with the results of global network analysis.

Conclusion

In this study, we identified sex-biased DEGs according to sex for CRC through meta-analysis of three datasets. Through pathway and network analysis, it was confirmed in which functional module the genes work. These genes can be presented as potential biomarkers for CRC.

Acknowledgement

- "This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2021-2020-0-01789) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)"