Meta-analysis of Female-specific Epigenetic Markers for Colorectal Cancer

Yubin Lee[1],Jaeseung Song[1], Wonhee Jang[1*]
[1]*Department of Life Sciences, Dongguk University, Seoul, 04620, Republic of Korea*

Abstract

Colorectal cancer (CRC) is the third most common cancer in the world, and the incidence of CRC increases steadily. In particular, the diagnosis of CRC in females is delayed because of confusion with menstrual issues and the relative survival rate of females is lower compared to males. To identify female-specific epigenetic markers for diagnosis of CRC, we conducted a meta-analysis using four DNA methylation array datasets of CRC patients. In this study, we identified 7639 differentially methylated CpG sites associated with 2261 differentially methylated genes (DMGs) between CRC tissues and non-tumor adjacent tissues. We then performed functional enrichment analysis with DAVID to investigate biological mechanisms of DMGs and observed that DMGs were enriched in the several oncogenic pathways. In order to analyze the functional connectivity between DMGs, we conducted a tissue-specific network analysis to reveal relationships between DMGs. We identified two colon-specific subnetwork modules consisting of 339 and 371 DMGs, respectively. Our study may suggest the underlying biological mechanisms to explain female-specific epigenetic effects of CRC and provide potential genetic markers for female-specific precise diagnosis of CRC.

Dataset

ArrayExpress/ GEO ID	Source	Platform	Control/ Disease	Female/ Male
GSE77954	Colorectal adenocarcinoma	Illumina Infinium HumanMethylation450 BeadChip	4/13	6/11
GSE101764			149/112	90/171
E-MTAB-3027			24/24	22/26
E-MTAB-7036			32/216	142/106
Total	**-**	**-**	**209/365**	**260/314**

Table 1. A table of datasets used for this study. Four DNA methylation array datasets were pre-processed using Minfi R package.

Workflow

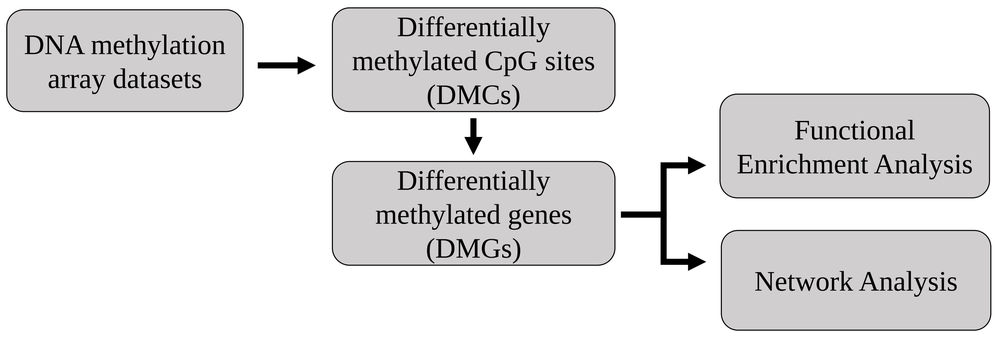

Figure 1. A workflow of overall analysis. Four DNA methylation array datasets were combined with SVA R package. Differentially methylated CpG sites (DMCs) and differentially methylated genes (DMGs) were obtained by using limma R package. Functional enrichment analysis with DAVID was performed for DMGs to identify biological mechanisms. Also, tissue-specific network analysis was conducted to investigate the functional connectivity between DMGs by HumanBase, a web-based tool.

DMCs and DMGs Analysis

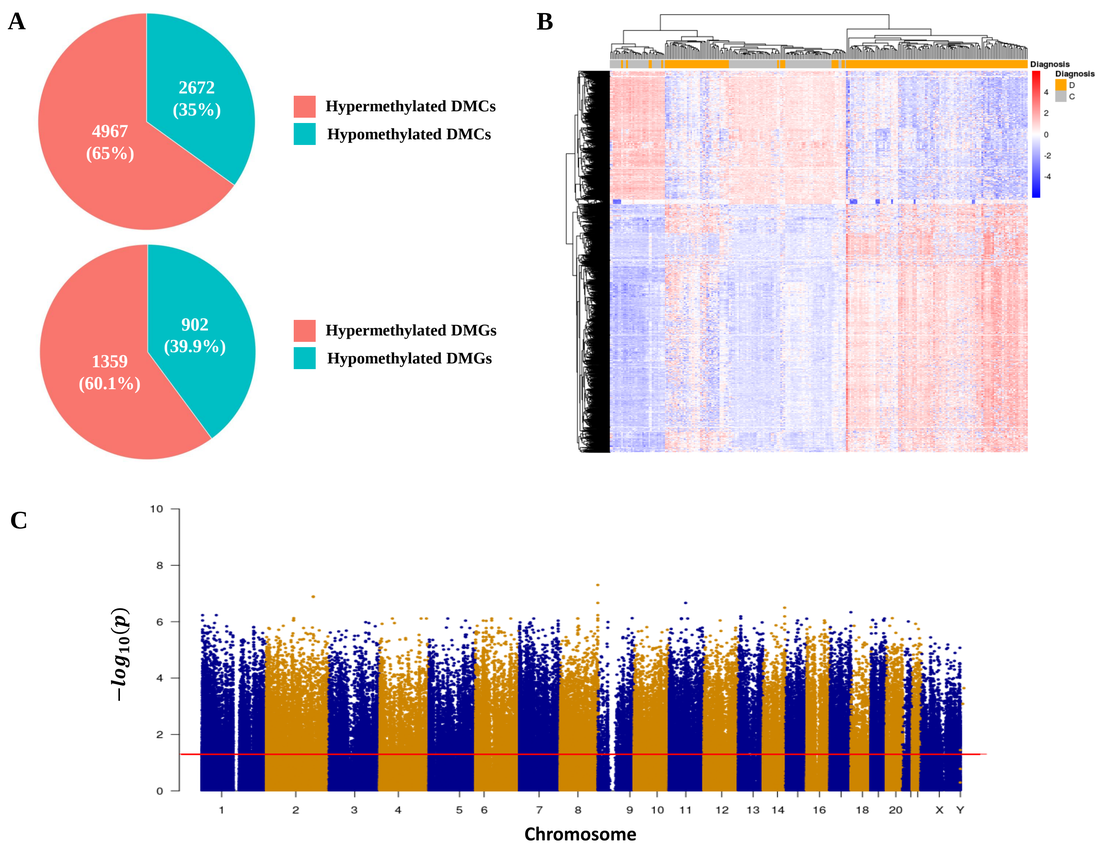

Figure 2. DMCs and DMGs between colorectal tissues and non-tumor adjacent tissues of female patients. (A) The pie charts showing the number of CpG sites and genes that have significant changes in the combined microarray dataset of female. (adj.P.val < 0.05 and $|log_2FC| > 0.5$) (B) A heatmap of 7639 DMCs in the merged dataset. Orange and gray represent disease group and control group, respectively. (C) A Manhattan plot of all DMCs in the comparison of colorectal cancer tissues and non-tumor adjacent tissues. The red line indicates the significance threshold. (adj.P.val < 0.05)

Functional Enrichment Analysis

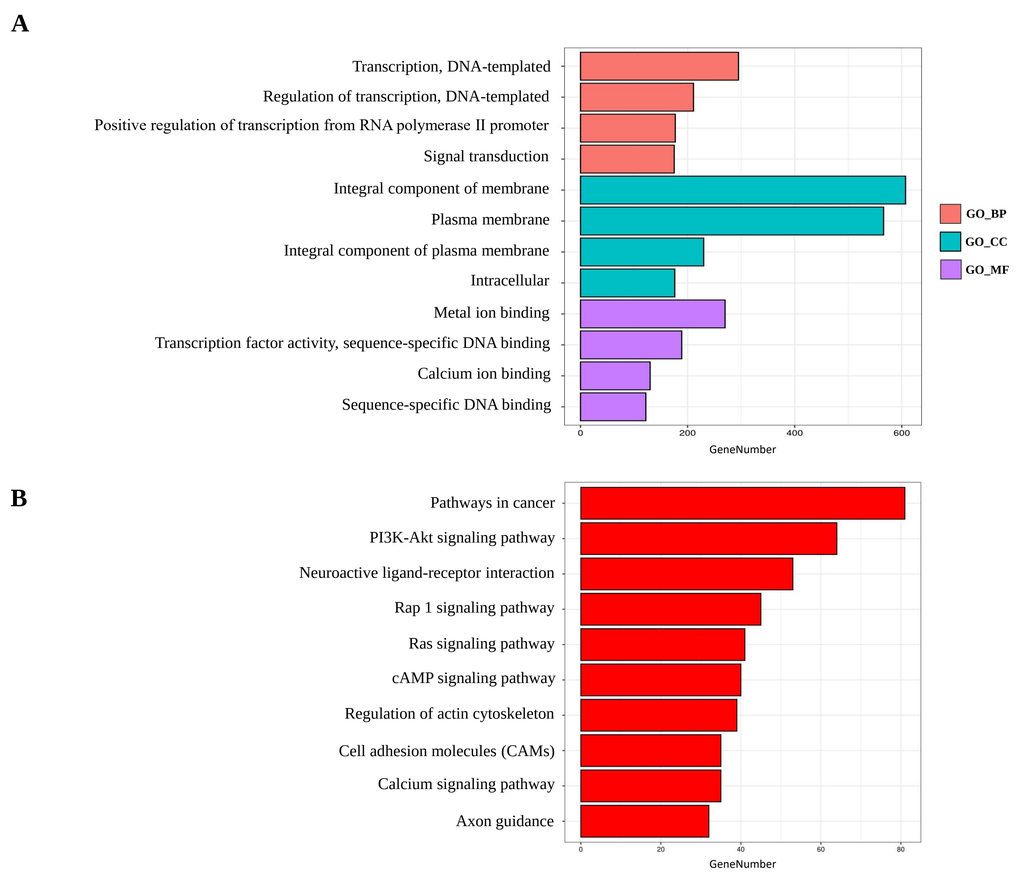

Figure 3. Results of functional enrichment analysis for 2261 DMGs. (A) The bar plots showing the result of Gene Ontology (GO) analysis from DMGs. Most terms were related to transcriptional regulation and signal transduction in Biological Process (BP) category. (B) The bar plots showing the result of KEGG pathway analysis with DMGs. The top enriched KEGG term was pathways in cancer, and most of the terms were associated with the signaling pathway known as the oncogenic pathway in colorectal cancer.

Network Analysis of DMGs

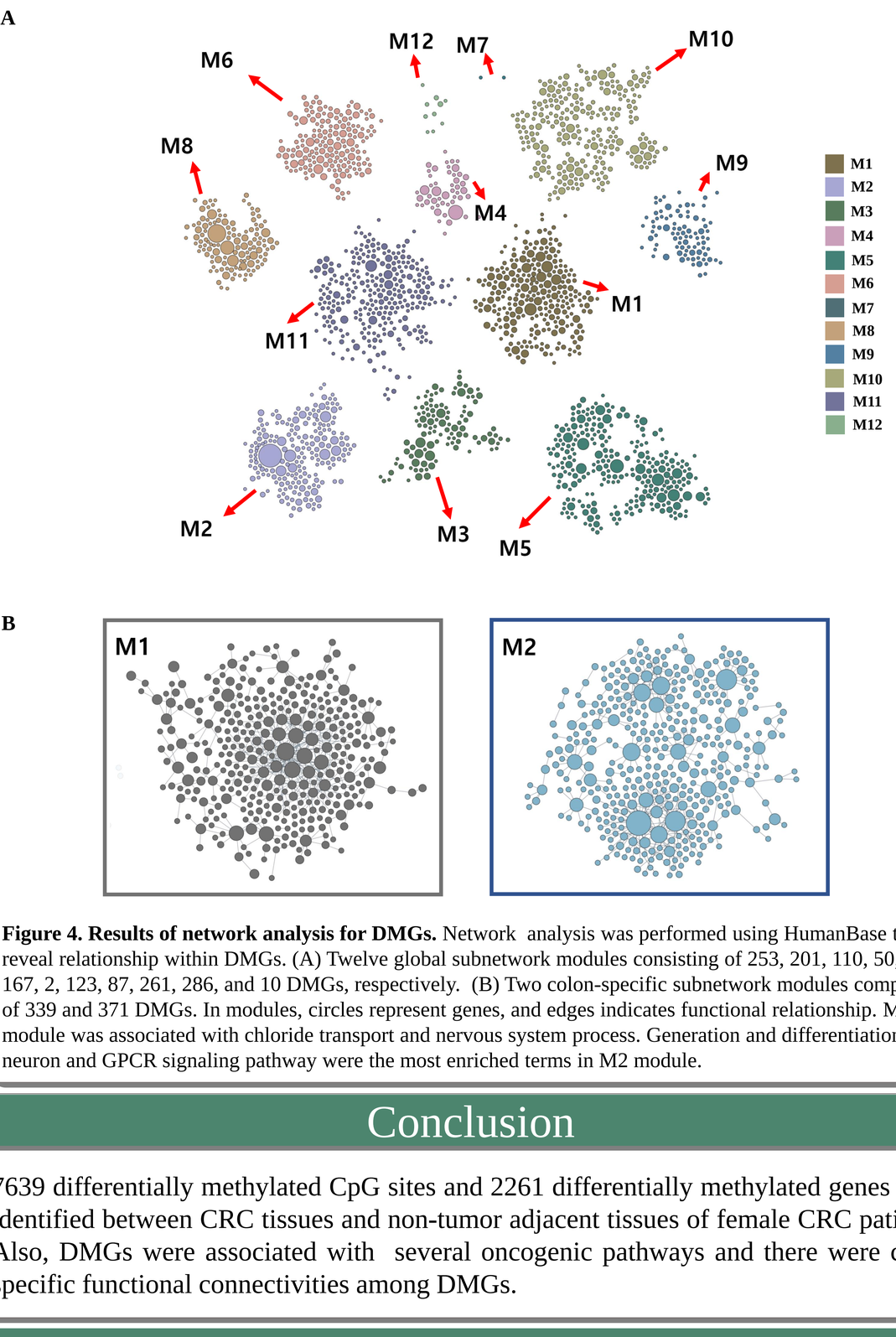

Figure 4. Results of network analysis for DMGs. Network analysis was performed using HumanBase to reveal relationship within DMGs. (A) Twelve global subnetwork modules consisting of 253, 201, 110, 50, 263, 167, 2, 123, 87, 261, 286, and 10 DMGs, respectively. (B) Two colon-specific subnetwork modules composed of 339 and 371 DMGs. In modules, circles represent genes, and edges indicates functional relationship. M1 module was associated with chloride transport and nervous system process. Generation and differentiation of neuron and GPCR signaling pathway were the most enriched terms in M2 module.

Conclusion

7639 differentially methylated CpG sites and 2261 differentially methylated genes were identified between CRC tissues and non-tumor adjacent tissues of female CRC patients. Also, DMGs were associated with several oncogenic pathways and there were colon specific functional connectivities among DMGs.

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)"

dongguk UNIVERSITY
Developmental biology Laboratory
Dongguk University, Seoul, Korea