Abstract. Microsatellite instability (MSI) has emerged as one of the key biological features of colorectal cancer (CRC). However, controversies remain regarding the association between the MSI status and clinicopathological characteristics of CRC. Therefore, it is crucial to identify potential key genes and pathways associated with MSI in CRC. In the present study, the GSE25071 gene expression profile was retrieved, with thirty-eight cases of microsatellite stable (MSS), five of MSI-High (MSI-H) and three of MSI-Low (MSI-L) CRC patients. The differentially expressed genes (DEGs) were analyzed by Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes pathway enrichment, gene set enrichment analysis (GSEA) and gene co-expression network analysis. Weighted gene correlation network analysis (WGCNA) was used for the gene modules and correlation of clinical traits. A total of forty-nine DEGs were identified between MSI-H and MSS, including six upregulated and forty-three down-regulated DEGs. Only the DEGs of MSI-H and MSS were subjected to subsequent analysis (limited number of DEGs of MSI-L and MSS, MSI-H and MSI-L). RNA metabolic process, endoplasmic reticulum and chemokine receptor binding were the top ranked terms in GO enrichment. The hub genes of co-expression network of DEGs included zinc finger protein (ZNF) 813, ZNF426, ZNF611, ZNF320 and ZNF573. The GSEA of MSI-H and MSS indicated that the mammalian target of rapamycin complex 1 signaling was significantly enriched with a nominal P-value of 0.038 and normalized enrichment score of 0.446. The WGCNA results showed that the pink module was the top in correlation with MSI status ($R^2=0.5$, P=0.0004). The genes in the pink module were significantly enriched in proteins targeting to endoplasmic reticulum, cytosolic part, structural constituent of ribosome and ribosome pathway. The hub genes identified in the pink module were ribosomal protein L12 (RPL12), RPS3A, RPS9, RPL27A, RPL7, RPL28, RPL14, RPS17, mitochondrial ribosomal protein L16, and G elongation factor, mitochondrial 2.

Introduction

Colorectal cancer (CRC) is among the most common malignancies worldwide, with a high incidence in the United States and Western Europe (1). In 2017, >130,000 newly cases were diagnosed and >50,000 individuals succumbed to mortality in the United States (1). The incidence rates and mortality rates have declined in patients with CRC aged ≥50 years, however, the incidence rates have increased by 22% in patients aged <50 years and the mortality rates have increased by 13% in the last decade (1). In China, the incidence and mortality rates of CRC have markedly increased (2). Despite substantial efforts in the establishment of early detection systems and chemotherapy reagents, the prognosis of CRC remains far from satisfactory for the majority of patients (3).

The therapeutic responses and survival outcomes of CRC are constrained by the clinical heterogeneity (4,5). Therefore, molecular markers emerge as efficient classifiers for CRC (6,7).
Materials and methods

Gene expression profile from the Gene Expression Omnibus (GEO) database. The gene expression profile, GSE25071, which contained 38 MSS colorectal cancer cases, 17 MSI-H and 3 MSI-L cases, was downloaded from GEO database (GEO: www.ncbi.nlm.nih.gov/geo) (18). The highest ranked three modules were identified using the Molecular Complex Detection (MCODE) program (29). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2. Given the limited DEGs in MSS vs. MSI-L and MSS vs. MSI-H comparison, only the expression data of DEGs identified in MSS vs. MSI-H were included for the subsequent analytic processes. All DEGs were imported into GeneMANIA, a web-based gene-gene interactions identification database (http://www.genemania.org) (24-26). The Benjamini-Hochberg procedure was applied to correct for multiple testing. DEGs with adjusted P values <0.05 and |log fold change (logFC)|>2 were selected for further analysis.

Identification of the DEGs. The DEGs were selected for further functional enrichment analysis and pathway enrichment analysis based on the GO and KEGG databases. The DEGs were further enriched in biological processes (BP), cellular components (CC) and molecular functions (MF) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web-based platform (27). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2. Given the limited DEGs in MSS vs. MSI-L and MSS vs. MSI-H comparison, the DEGs were then enriched for biological processes (BP), cellular components (CC) and molecular functions (MF) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web-based platform (27). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2.

Gene ontology (GO) and KEGG Encyclopedia of Genes and Genomes (KEGG) GO and KEGG enrichment analyses were performed using the GO and KEGG databases. The DEGs were selected for further functional enrichment analysis and pathway enrichment analysis based on the GO and KEGG databases. The DEGs were further enriched in biological processes (BP), cellular components (CC) and molecular functions (MF) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web-based platform (27). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2. Given the limited DEGs in MSS vs. MSI-L and MSS vs. MSI-H comparison, the DEGs were then enriched for biological processes (BP), cellular components (CC) and molecular functions (MF) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web-based platform (27). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2.

Results

Microsatellite instability (MSI) is one of the most promising markers investigated to date for CRC, bioinformatics analysis of the GSE25071 gene expression profile, including 38 MSS, five MSI-H and three MSI-L cases, was conducted to identify potential key genes and pathways associated with MSI. MSI is one of the most promising markers investigated to date for CRC, bioinformatics analysis of the GSE25071 gene expression profile, including 38 MSS, five MSI-H and three MSI-L cases, was conducted to identify potential key genes and pathways associated with MSI.

The GSE25071 profile was further annotated by the profile in the GEO database (GEO: www.ncbi.nlm.nih.gov/geo) (18). The gene expression profile, GSE25071, which contained 38 MSS colorectal cancer cases, 17 MSI-H and 3 MSI-L cases, was downloaded from GEO database (GEO: www.ncbi.nlm.nih.gov/geo) (18). The highest ranked three modules were identified using the Molecular Complex Detection (MCODE) program (29). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2. Given the limited DEGs in MSS vs. MSI-L and MSS vs. MSI-H comparison, only the expression data of DEGs identified in MSS vs. MSI-H were included for the subsequent analytic processes. All DEGs were imported into GeneMANIA, a web-based gene-gene interactions identification database (http://www.genemania.org) (24-26). The Benjamini-Hochberg procedure was applied to correct for multiple testing. DEGs with adjusted P values <0.05 and |log fold change (logFC)|>2 were selected for further analysis.

Identification of the DEGs. The DEGs were selected for further functional enrichment analysis and pathway enrichment analysis based on the GO and KEGG databases. The DEGs were further enriched in biological processes (BP), cellular components (CC) and molecular functions (MF) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web-based platform (27). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2. Given the limited DEGs in MSS vs. MSI-L and MSS vs. MSI-H comparison, only the expression data of DEGs identified in MSS vs. MSI-H were included for the subsequent analytic processes. All DEGs were imported into GeneMANIA, a web-based gene-gene interactions identification database (http://www.genemania.org) (24-26). The Benjamini-Hochberg procedure was applied to correct for multiple testing. DEGs with adjusted P values <0.05 and |log fold change (logFC)|>2 were selected for further analysis.

Identification of the DEGs. The DEGs were selected for further functional enrichment analysis and pathway enrichment analysis based on the GO and KEGG databases. The DEGs were further enriched in biological processes (BP), cellular components (CC) and molecular functions (MF) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web-based platform (27). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2. Given the limited DEGs in MSS vs. MSI-L and MSS vs. MSI-H comparison, only the expression data of DEGs identified in MSS vs. MSI-H were included for the subsequent analytic processes. All DEGs were imported into GeneMANIA, a web-based gene-gene interactions identification database (http://www.genemania.org) (24-26). The Benjamini-Hochberg procedure was applied to correct for multiple testing. DEGs with adjusted P values <0.05 and |log fold change (logFC)|>2 were selected for further analysis.

Identification of the DEGs. The DEGs were selected for further functional enrichment analysis and pathway enrichment analysis based on the GO and KEGG databases. The DEGs were further enriched in biological processes (BP), cellular components (CC) and molecular functions (MF) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web-based platform (27). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2. Given the limited DEGs in MSS vs. MSI-L and MSS vs. MSI-H comparison, only the expression data of DEGs identified in MSS vs. MSI-H were included for the subsequent analytic processes. All DEGs were imported into GeneMANIA, a web-based gene-gene interactions identification database (http://www.genemania.org) (24-26). The Benjamini-Hochberg procedure was applied to correct for multiple testing. DEGs with adjusted P values <0.05 and |log fold change (logFC)|>2 were selected for further analysis.

Identification of the DEGs. The DEGs were selected for further functional enrichment analysis and pathway enrichment analysis based on the GO and KEGG databases. The DEGs were further enriched in biological processes (BP), cellular components (CC) and molecular functions (MF) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web-based platform (27). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2. Given the limited DEGs in MSS vs. MSI-L and MSS vs. MSI-H comparison, only the expression data of DEGs identified in MSS vs. MSI-H were included for the subsequent analytic processes. All DEGs were imported into GeneMANIA, a web-based gene-gene interactions identification database (http://www.genemania.org) (24-26). The Benjamini-Hochberg procedure was applied to correct for multiple testing. DEGs with adjusted P values <0.05 and |log fold change (logFC)|>2 were selected for further analysis.

Identification of the DEGs. The DEGs were selected for further functional enrichment analysis and pathway enrichment analysis based on the GO and KEGG databases. The DEGs were further enriched in biological processes (BP), cellular components (CC) and molecular functions (MF) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web-based platform (27). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2. Given the limited DEGs in MSS vs. MSI-L and MSS vs. MSI-H comparison, only the expression data of DEGs identified in MSS vs. MSI-H were included for the subsequent analytic processes. All DEGs were imported into GeneMANIA, a web-based gene-gene interactions identification database (http://www.genemania.org) (24-26). The Benjamini-Hochberg procedure was applied to correct for multiple testing. DEGs with adjusted P values <0.05 and |log fold change (logFC)|>2 were selected for further analysis.

Identification of the DEGs. The DEGs were selected for further functional enrichment analysis and pathway enrichment analysis based on the GO and KEGG databases. The DEGs were further enriched in biological processes (BP), cellular components (CC) and molecular functions (MF) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web-based platform (27). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2. Given the limited DEGs in MSS vs. MSI-L and MSS vs. MSI-H comparison, only the expression data of DEGs identified in MSS vs. MSI-H were included for the subsequent analytic processes. All DEGs were imported into GeneMANIA, a web-based gene-gene interactions identification database (http://www.genemania.org) (24-26). The Benjamini-Hochberg procedure was applied to correct for multiple testing. DEGs with adjusted P values <0.05 and |log fold change (logFC)|>2 were selected for further analysis.

Identification of the DEGs. The DEGs were selected for further functional enrichment analysis and pathway enrichment analysis based on the GO and KEGG databases. The DEGs were further enriched in biological processes (BP), cellular components (CC) and molecular functions (MF) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) web-based platform (27). The interaction list of the DEGs included adj. P<0.05 and |log fold change (logFC)|>2. Given the limited DEGs in MSS vs. MSI-L and MSS vs. MSI-H comparison, only the expression data of DEGs identified in MSS vs. MSI-H were included for the subsequent analytic processes. All DEGs were imported into GeneMANIA, a web-based gene-gene interactions identification database (http://www.genemania.org) (24-26). The Benjamini-Hochberg procedure was applied to correct for multiple testing. DEGs with adjusted P values <0.05 and |log fold change (logFC)|>2 were selected for further analysis.
Protein-protein interaction (PPI) networks. All DEGs were further input to the Search Tool for the Retrieval of Interacting Genes/Proteins database (STRING version 10.0; www.string-db.org) for interaction network at the protein level (30). The results were further visualized using Cytoscape with predefined cut-off values: Node degree, ≥2; k-score(value=2); node score, 0.2; max. depth, 100 (28).

Gene set enrichment analysis (GSEA). GSEA, released by the Broad Institute (software.broadinstitute.org/gsea/index.jsp), was used to cluster significant gene sets associated with given annotation terms (31). In the present study, the MSS and MSI-H samples were analyzed by GSEA with the annotation of ‘hallmark gene sets’. The significant cut-off value was defined as P<0.05.

Weighted gene correlation network analysis (WGCNA). WGCNA was used for co-expression network construction based on the correlations among genes and identification of top-ranked gene modules and hub genes. The ‘WGCNA’ R package was implemented for all the genes in 46 samples (38 MSS, five MSI-H and three MSI-L, normal samples were excluded). Initially, Pearson’s correlation of each pair genes was calculated and an adjacency matrix was constructed based on the Pearson’s results and a predefined soft-thresholding parameter (β). Subsequently, the topological overlap matrix of the included genes and adjacency matrix was produced. Genes with similar expression trends were classified as the same modules eigengenes for further clinical traits correlation (32,33). The genes of the most correlated module were extracted for GO and KEGG analyses and PPI network construction. Hub genes were defined with the highest degrees.

Prognostic values of hub genes in WGCNA. The overall survival of hub genes determined by WGCNA was further examined in the PrognoScan database, a comprehensive platform for prognostic annotation (dna00.bio.kyutech.ac.jp/PrognoScan/) (34).

Results

Identification of upregulated and downregulated DEGs. A total of 49 DEGs, including six upregulated and 43 downregulated DEGs, were identified between MSI-H and MSS groups, illustrated by the bidirectional clustering heat map (Fig. 1A). However, no significant DEGs were identified between MSI-L and MSS, and only one DEG was identified between MSI-H and MSI-L. Therefore, subsequent investigations focused on the DEGs between MSI-H and MSS.

GO enrichment and KEGG pathway analyses of DEGs. A total of 15 BP, three CC and five MF terms were significantly enriched. Specifically, RNA metabolic process, endoplasmic reticulum and chemokine receptor binding were the top-ranked in each term, respectively (Fig. 1B). No significant KEGG pathway was enriched in the DEGs.

Co-expressed genes network analysis. To delineate the biological functions of the DEGs, a co-expression network of the DEGs with correlated additional genes was established based on the GeneMANIA program. A total of 66 nodes and 678 edges were determined (Fig. 2A). The top five hub genes with the highest degree were identified, including zinc finger protein (ZNF) 813, ZNF426, ZNF611, ZNF320 and ZNF573. The top three modules with the highest scores were identified using the MCODE plugin (Fig. 2B-D). Among them, no particular KEGG pathway was significantly enriched. However, in the GO enrichment, the regulation of transcription, DNA-templated term was the highest ranked BP for module 1, and negative regulation of hydroxylase activity for module 2. Subsequently, the hub genes were externally validated in GSE18088, GSE13067 and GSE78220. The expression levels of ZNF426, ZNF320 and ZNF573 were significantly reduced in MSI-H compared with MSS in GSE18088. The expression levels of ZNF813, ZNF426 and ZNF573 were significantly reduced in the MSI-H group compared with the MSS group.
in GSE13067. The expression levels of ZNF813 and ZNF573 were significantly reduced in the pembrolizumab no-response group compared with the response group (Fig. 3A-C).

**PPI network analysis.** The minimum required interaction score of STRING was medium confidence (0.4) and the cut-off degree for the included nodes in Cytoscape was ≥1. The PPI network constructed by the differentially expressed genes and additional correlated genes from the GeneMANIA program is shown in Figure 2. (A) Co-expression network; (B) top modules with highest scores; (C) second highest module; (D) third highest module. Red indicates upregulation; blue indicates downregulation; lines between nodes indicate interactions between genes.
networks included 14 nodes and eight edges (Fig. 4), being distinct from the co-expression networks (Fig. 2A).

**GSEA results.** GSEA was used to determine the functions of gene sets between the MSI-H and MSS groups. Only one gene set, the mammalian target of rapamycin complex 1 (mTORC1) signaling (nominal P-value, 0.038; normalized enrichment score, 0.446) was significantly enriched in MSI-H, with none significantly enriched in MSS (Fig. 5). Furthermore, the 50 top ranked genes correlated with each
phenotype (MSI-H and MSS) are illustrated with a heat map (Fig. 6).

WGCNA of the gene expression profile in all tumor samples. To further investigate the potential gene modules associated with MSI status, the WGCNA was conducted with R package. A total of 46 cases, including 38 MSS, eight MSI (five MSI-H and three MSI-L) were clustered (Fig. 7A). The power of $\beta=9$ was defined as the soft-thresholding value (scale free $R^2=0.95$; slope=-1.67; Fig. 7B-D). A total of 25 modules were identified (Fig. 8A). Noteworthy, the pink module was the most significantly correlated with MSI status ($R^2=0.5$, $P=0.0004$). Of note,
the pink module also featured a high correlation ($R^2=0.53$, $P=0.0002$) with tumor stage (Fig. 8B).

Identification of hub genes in the pink module associated with MSI status. The genes in the pink module were significantly enriched by protein targeting to endoplasmic reticulum in BP terms, cytosolic part in CC terms, structural constituent of ribosome in MF terms and ribosome in KEGG pathways. Subsequently, a PPI network (degree $\geq 1$), with 55 nodes and 96 edges, was constructed based on the genes in the pink module. The top 10 hub genes with the highest degrees were determined, including ribosomal protein L12 (RPL12), ribosomal protein S3A (RPS3A), ribosomal protein S9 (RPS9), ribosomal protein L27a (RPL27A), ribosomal protein L7 (RPL7), ribosomal protein L28 (RPL28), ribosomal protein L14 (RPL14), ribosomal protein S17 (RPS17), mitochondrial ribosomal protein L16 (MRPL16) and G elongation factor, mitochondrial 2 (GFM2), as shown in Fig. 9.

Prognostic values of the hub genes in the pink module. The prognostic values of the hub genes were assessed in various independent datasets with different probe IDs and array types. In all, none of the genes exhibited significant prognostic values (Table I).

Discussion

The present study is the first, to the best of our knowledge, to use multiple bioinformatics analysis approaches to demonstrate the potential key genes and pathways associated with MSI status in patients with CRC. Among the significant GO terms, RNA metabolic process, endoplasmic reticulum and chemokine receptor binding were the top-ranked terms. No significant KEGG pathway was identified, however, using GSEA, the MTORC1 signaling pathway was significantly enriched. MTOR signaling pathways, consisting of at least two complexes, mTORC1 and mTORC2, receive a plethora of input factors and modulate a broad spectrum of downstream molecules (35,36). Noteworthy, the inhibition of mTOR1 only leads to mild protein synthesis reduction and potential influence upon the cell cycle process (35,37).

Previously, Choi et al investigated somatic mutational, intratumoral heterogeneity and expressional alterations of mTOR pathway-related genes in cancer with MSI, including PIK3CB, insulin receptor substrate 1/2 (IRS1), RPS6, eukaryotic translation initiation factor 4B (EIF4B), RPS6KA5 and PRKAA2 (38). Of the patients with MSI-H CRC, 8.9% harbored IRS1 frameshift mutations, whereas 10.1% harbored mutations in EIF4B and 3.8% in RPS6KA5. Noteworthy, no mutations was identified in MSS or MSI-L (38). The study by Choi et al and the present study demonstrated the potential roles of the ribosomal protein family associated with MSI status. In addition, differing from previous search strategies in published results (38), the present study illustrated how WGCNA can be implemented to predict new genes in the regulation of MSI in CRC.

Lin et al analyzed the mutations of 113 MSS and 29 MSI-H cases of CRC (39). Mutations of PIK3CA, phosphatase and
tensin homolog and/or AKT1 in the mTOR pathway were found in 59% of the MSI-H patients compared with 19% of the patients with MSS (39). Methodologically, a 50-gene AmpliSeq Cancer Hotspot Panel was introduced by Lin et al., whereas the expression profile generated using the ABI Human Genome Survey Microarray was analyzed in the present study for GSEA and WGCNA. Collectively, the present study provided an insightful target and further complemented the results reported by Choi et al. and Lin et al. with regards to the multiple bioinformatics strategies.

In the co-expressed gene network, ZNF813, ZNF426, ZNF611, ZNF320 and ZNF573 were the top-ranked hub genes closely associated with MSI-H. Reduced expression levels of ZNF813 and ZNF573 were found in the MSI-H group and pembrolizumab no-response group. Of note, the check-point inhibitor pembrolizumab significantly prolonged the prognostic outcomes of patients with MSI-H CRC compared with those in the MSS group (7). This indicated that potential mechanisms exist between MSI status and the outcomes of check-point inhibitor treatment, further highlighting the predictive role of ZNFs.

ZNFs are one of the most common proteins in the eukaryotic system, with a broad range of biological functions, including DNA recognition, RNA transcription, apoptosis and protein structure (40). The five hub genes are mainly located in the nucleus and are involved in DNA-binding and transcription regulation. Novel topologies of numerous ZNF domains have provided evidence for structure/function relationships (40). To the best of our knowledge, the present study is the first highlighting the potential association between ZNFs and MSI in CRC.
<table>
<thead>
<tr>
<th>ID_NAME</th>
<th>Dataset</th>
<th>Array type</th>
<th>Probe ID</th>
<th>N</th>
<th>COX P-value</th>
<th>HR (95% CI lower- CI upper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFM2</td>
<td>GSE17536</td>
<td>HG-U133_Plus_2</td>
<td>231917_at</td>
<td>177</td>
<td>0.52</td>
<td>1.20 (0.69-2.10)</td>
</tr>
<tr>
<td>GFM2</td>
<td>GSE17536</td>
<td>HG-U133_Plus_2</td>
<td>231918_s_at</td>
<td>177</td>
<td>0.43</td>
<td>1.19 (0.78-1.81)</td>
</tr>
<tr>
<td>GFM2</td>
<td>GSE17537</td>
<td>HG-U133_Plus_2</td>
<td>225392_at</td>
<td>55</td>
<td>0.35</td>
<td>1.33 (0.73-2.42)</td>
</tr>
<tr>
<td>GFM2</td>
<td>GSE17537</td>
<td>HG-U133_Plus_2</td>
<td>231917_at</td>
<td>55</td>
<td>0.64</td>
<td>0.75 (0.22-2.54)</td>
</tr>
<tr>
<td>GFM2</td>
<td>GSE17537</td>
<td>HG-U133_Plus_2</td>
<td>231918_s_at</td>
<td>55</td>
<td>0.54</td>
<td>1.24 (0.63-2.44)</td>
</tr>
<tr>
<td>MRPL16</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>217980_s_at</td>
<td>62</td>
<td>0.82</td>
<td>1.12 (0.41-3.08)</td>
</tr>
<tr>
<td>MRPL16</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>200809_x_at</td>
<td>62</td>
<td>0.39</td>
<td>107.59 (0.00-460056831)</td>
</tr>
<tr>
<td>RPL12</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>214271_x_at</td>
<td>62</td>
<td>0.76</td>
<td>2.21 (0.01-326.11)</td>
</tr>
<tr>
<td>RPL12</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>200809_y_at</td>
<td>62</td>
<td>0.39</td>
<td>107.59 (0.00-460056831)</td>
</tr>
<tr>
<td>RPL12</td>
<td>GSE17536</td>
<td>HG-U133_Plus_2</td>
<td>214271_x_at</td>
<td>177</td>
<td>0.29</td>
<td>2.03 (0.55-7.52)</td>
</tr>
<tr>
<td>RPL12</td>
<td>GSE17536</td>
<td>HG-U133_Plus_2</td>
<td>200809_y_at</td>
<td>177</td>
<td>0.34</td>
<td>1.91 (0.50-7.28)</td>
</tr>
<tr>
<td>RPL12</td>
<td>GSE17537</td>
<td>HG-U133_Plus_2</td>
<td>214271_x_at</td>
<td>55</td>
<td>0.47</td>
<td>1.76 (0.38-8.13)</td>
</tr>
<tr>
<td>RPL14</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>213588_x_at</td>
<td>62</td>
<td>0.45</td>
<td>7654467.33 (0.00-65...</td>
</tr>
<tr>
<td>RPL14</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>219138_at</td>
<td>62</td>
<td>0.76</td>
<td>0.78 (0.17-3.69)</td>
</tr>
<tr>
<td>RPL14</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>200003_s_at</td>
<td>62</td>
<td>0.90</td>
<td>1.07 (0.40-2.83)</td>
</tr>
<tr>
<td>RPL28</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>213223_at</td>
<td>62</td>
<td>0.24</td>
<td>2.03 (0.62-6.63)</td>
</tr>
<tr>
<td>RPL28</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>200003_s_at</td>
<td>62</td>
<td>0.58</td>
<td>1.34 (0.48-3.73)</td>
</tr>
<tr>
<td>RPL28</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>213223_at</td>
<td>177</td>
<td>0.95</td>
<td>0.98 (0.53-1.81)</td>
</tr>
<tr>
<td>RPL28</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>200003_s_at</td>
<td>177</td>
<td>0.81</td>
<td>0.83 (0.18-3.74)</td>
</tr>
<tr>
<td>RPS17</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>212578_x_at</td>
<td>62</td>
<td>0.66</td>
<td>1.91 (0.11-33.65)</td>
</tr>
<tr>
<td>RPS17</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>201665_x_at</td>
<td>62</td>
<td>0.86</td>
<td>0.82 (0.08-8.15)</td>
</tr>
<tr>
<td>RPS17</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>212578_x_at</td>
<td>177</td>
<td>0.082</td>
<td>4.17 (0.84-20.77)</td>
</tr>
</tbody>
</table>
Table I. Continued.

<table>
<thead>
<tr>
<th>ID_NAME</th>
<th>Dataset</th>
<th>Array type</th>
<th>Probe ID</th>
<th>N</th>
<th>COX P-value</th>
<th>HR (95% CI lower-CI upper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPS17</td>
<td>GSE17536</td>
<td>HG-U133_Plus_2</td>
<td>201665_x_at</td>
<td>177</td>
<td>0.17</td>
<td>2.96 (0.64-13.76)</td>
</tr>
<tr>
<td>RPS17</td>
<td>GSE17537</td>
<td>HG-U133_Plus_2</td>
<td>212578_x_at</td>
<td>55</td>
<td>0.53</td>
<td>2.16 (0.20-23.51)</td>
</tr>
<tr>
<td>RPS17</td>
<td>GSE17537</td>
<td>HG-U133_Plus_2</td>
<td>201665_x_at</td>
<td>55</td>
<td>0.37</td>
<td>2.99 (0.27-32.95)</td>
</tr>
<tr>
<td>RPS3A</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>201257_x_at</td>
<td>62</td>
<td>0.80</td>
<td>25.91</td>
</tr>
<tr>
<td>RPS3A</td>
<td>GSE17536</td>
<td>HG-U133_Plus_2</td>
<td>201257_x_at</td>
<td>177</td>
<td>0.97</td>
<td>1.04 (0.11-9.64)</td>
</tr>
<tr>
<td>RPS3A</td>
<td>GSE17537</td>
<td>HG-U133_Plus_2</td>
<td>201257_x_at</td>
<td>55</td>
<td>0.65</td>
<td>0.48 (0.02-11.00)</td>
</tr>
<tr>
<td>RPS9</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>217747_s_at</td>
<td>62</td>
<td>0.85</td>
<td>0.90 (0.31-2.63)</td>
</tr>
<tr>
<td>RPS9</td>
<td>GSE12945</td>
<td>HG-U133A</td>
<td>214317_x_at</td>
<td>62</td>
<td>0.77</td>
<td>5.03 (0.00-240092.23)</td>
</tr>
<tr>
<td>RPS9</td>
<td>GSE17536</td>
<td>HG-U133_Plus_2</td>
<td>1557981_at</td>
<td>177</td>
<td>0.0018</td>
<td>0.06 (0.01-0.35)</td>
</tr>
<tr>
<td>RPS9</td>
<td>GSE17536</td>
<td>HG-U133_Plus_2</td>
<td>214317_x_at</td>
<td>177</td>
<td>0.81</td>
<td>1.22 (0.24-6.07)</td>
</tr>
<tr>
<td>RPS9</td>
<td>GSE17536</td>
<td>HG-U133_Plus_2</td>
<td>217747_s_at</td>
<td>177</td>
<td>0.66</td>
<td>1.30 (0.40-4.17)</td>
</tr>
<tr>
<td>RPS9</td>
<td>GSE17537</td>
<td>HG-U133_Plus_2</td>
<td>217747_s_at</td>
<td>55</td>
<td>0.99</td>
<td>1.01 (0.36-2.84)</td>
</tr>
<tr>
<td>RPS9</td>
<td>GSE17537</td>
<td>HG-U133_Plus_2</td>
<td>1557981_at</td>
<td>55</td>
<td>0.25</td>
<td>0.08 (0.00-5.87)</td>
</tr>
<tr>
<td>RPS9</td>
<td>GSE17537</td>
<td>HG-U133_Plus_2</td>
<td>214317_x_at</td>
<td>55</td>
<td>0.70</td>
<td>1.50 (0.19-11.58)</td>
</tr>
</tbody>
</table>

HR, hazard ratio; CI, confidence intervals; GFM, G elongation factor, mitochondrial 2; MRPL16, mitochondrial ribosomal protein L16; RP, ribosomal protein.

Figure 9. Protein-protein interaction networks of the extracted genes from the pink module.
The genes in the pink module of the WGCNA demonstrated the highest correlation with MSI. By performing further PPI network analysis, the top 10 hub genes were identified, including RPL12, RPS3A, RPS9, RPL27A, RPL7, RPL28, RPL14, RPS17, MRPL16 and GFM2. Given the high proportion of RPs in the hub gene list, RPs are of interest for further discussion and may be of significance to the mechanism underlying the MSI in CRC. The synthesis of RPs is the basis for the biological processes in each cell (41). The newly decoded crystal structures of ribosomes provide multiple traits associated with PPIs, RNA-protein and protein-drugs interactions (41). Noteworthy, the mutation of RPS20, part of the small ribosome subunit, can render individuals with MSS predisposition (42), highlighting the association between ribosome and MSI/MSS status. In addition, RPs are associated with biological RNA synthesis, one of the predominant features of cancer cells exposed to 5-FU treatment (43,44). Therefore, this clarified the role of MSI in FU-non-responders, at least in part. Of note, the prognostic evaluation of the hub genes indicated that their potential roles may not be directly associated with survival status.

Previously, Timmermann et al fully investigated the whole exome next generation sequencing of 454 patients with CRC and identify the significant 359 mutations in MSI and 45 mutations in MSS (45). In addition to the MSI and CIN subtypes, a third subtype associated with sessile-serrated adenomas was proposed (46). This newly added third subtype may partially clarify the limited DEGs identified in MSI-L vs. MSS and MSI vs. MSS in the present study.

The limitations of the present study include the comparably small sample size in MSI-H and lack of experimental validation for hub genes. Larger CRC samples with MSI/MSS and molecular biological experiments are required to specifically confirm the functions and mechanisms of the hub genes underlying MSI in CRC. However, to reduce the potential confounding factors produced by a single bioinformatics approach, the present study employed multiple bioinformatics patterns, including DEG analysis, GSEA and WGCNA. In addition, due to the limited number of patients and lack of survival data in original files, the prognostic values of hub genes identified by WGCNA were examined using the PrognoScan database.

In conclusion, the bioinformatics analysis performed in the present study identified key genes and pathways associated with MSI, and further elucidated insightful traits for potential mechanisms.

Acknowledgements

The authors would like to thank the Shanghai Institute of Digestive Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine (Shanghai, China) for their academic support. The authors would also like to thank Dr Ernest Johann Helwig (Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China) for his helpful contributions with discussions and language editing.

Funding

This study was financially supported by the National Natural Science Foundation of China (grant nos. 81402423 and 81572818) and the Shanghai Municipal Commission of Health and Family Planning (grant no. 2017YQ062).

Availability of data and materials

The datasets supporting the conclusion of this article are included within the article.

Authors’ contributions

CY, HH, SZ and JS performed experiments and data analysis; CY, HH, YZ, JM and MZ drafted the manuscript; CY, AL, JS, MZ, YZ and JM participated in the study design, data collection and revision process. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.