Neoadjuvant chemoradiation alters biomarkers of anticancer immunotherapy responses in locally advanced rectal cancer

Incheon Seo, Hye Won Lee, Sang Jun Byun, Jee Young Park, Hyeonji Min, Sung Hwan Lee, Ju-Seog Lee, Shin Kim, Sung Uk Bae

ABSTRACT
Background Neoadjuvant chemoradiation therapy (CRT) is a widely used preoperative treatment strategy for locally advanced rectal cancer (LARC). However, a few studies have evaluated the molecular changes caused by neoadjuvant CRT in these cancer tissues. Here, we aimed to investigate changes in immunotherapy-related immunogenic effects in response to preoperative CRT in LARC.

Methods We analyzed 60 pairs of human LARC tissues before and after irradiation from three independent LARC cohorts, including a LARC patient RNA sequencing (RNA-seq) dataset from our cohort and GSE15781 and GSE94104 datasets.

Results Gene ontology analysis showed that preoperative CRT significantly enriched the immune response in LARC tissues. Moreover, gene set enrichment analysis revealed six significantly enriched Kyoto Encyclopedia of Genes and Genomes pathways associated with downregulated genes, including mismatch repair (MMR) genes, in LARC tissues after CRT in all three cohorts. Radiation also induced apoptosis and downregulated various MMR system-related genes in three colorectal cancer cells. One patient with LARC showed a change in microsatellite instability (MSI) status after CRT, as demonstrated by the loss of MMR protein and PCR for MSI. Moreover, CRT significantly increased tumor mutational burden in LARC tissues.

CIBERSORT analysis revealed that the proportions of M2 macrophages and CD8 T cells were significantly increased after CRT in both the RNA-seq dataset and GSE94104. Notably, preoperative CRT increased various immune biomarker scores, such as the interferon-γ signature, the cytolytic activity and the immune signature.

Conclusions Taken together, our findings demonstrated that neoadjuvant CRT modulated the immune-related characteristics of LARC, suggesting that neoadjuvant CRT may enhance the responsiveness of LARC to immunotherapy.

BACKGROUND
Neoadjuvant chemoradiation therapy (CRT) and total mesorectal excision are commonly incorporated into the multimodal treatment of locally advanced rectal cancer (LARC). This therapeutic strategy is associated with downstaging of the tumor, more frequent use of sphincter-preserving procedures by preoperatively shrinking the tumor, decreased local recurrence and reduced toxicity compared with postoperative adjuvant chemoradiation. CRT delivers ionizing radiation directly to target cells, with the goal of causing genetic damage, such as radiation-induced DNA double-strand breaks, which are repaired through double-strand break repair mechanisms. In addition to the direct cytotoxic effects of radiotherapy, this treatment may increase neoantigens, activate the major histocompatibility complex class I system, activate tumor-infiltrating lymphocytes and induce the abscopal effect, wherein localized radiation provokes distant antitumor effects. The abscopal effect can be enhanced by combining an immune checkpoint blocker (ICB) with radiotherapy.

Cancer immunotherapy with ICBs, such as antiprogrammed death-1/programmed death-ligand 1 and anticytotoxic T lymphocyte antigen-4 inhibitors, is a novel treatment strategy in the field of immuno-oncology. To improve the effectiveness of this treatment strategy, several studies have recently evaluate various biomarkers, such as microsatellite instability (MSI), mismatch repair (MMR) deficiency, tumor mutational burden (TMB) and immune biomarker scores, for prediction of responses to immunotherapies, such as ICBs, in cancer.

Identifying molecular signatures using next-generation sequencing (NGS) has become an emerging focus in the field of cancer research, facilitating the diagnosis and prediction of prognosis for patients. However, a few studies have performed transcriptomic profiling of biomarkers for
predicting responses to immunotherapy in rectal cancer before and after preoperative CRT. Accordingly, in this study, we investigated changes in the molecular profiles of LARC after CRT with regard to improved responses to immunotherapy through bioinformatics analysis of NGS transcriptome and Gene Expression Omnibus (GEO) datasets. Furthermore, we validated the results from these analyses using colorectal cancer (CRC) cell lines.

METHODS

Patients and sample collection

Formalin-fixed paraffin-embedded (FFPE) block specimens from preoperative biopsy via sigmoidoscopy and surgical resection of the primary tumor were obtained from 54 patients with rectal adenocarcinoma who underwent neoadjuvant concurrent CRT between August 2016 and December 2017. Preoperative clinical stage was determined by abdominal and chest CT scans and pelvic MRI and the inclusion criterion was clinical stages 2–3 (T3 or T4 and/or node positive) rectal adenocarcinoma. The exclusion criteria were stage IV tumors, cancer related to familial adenomatous polyposis or hereditary nonpolyposis CRC, synchronous or previous malignancies, distant metastasis during neoadjuvant CRT and patients who refused the definite surgery or were lost to follow-up. Based on these criteria, we included 11 pairs of human CRC tissues obtained before and after CRT (online supplemental table 1) and performed RNA sequencing (RNA-seq) analysis. The detailed study design and enrolled samples at each stage are illustrated in the workflow shown in figure 1. All patients received conventional long-course neoadjuvant CRT (five cycles of 5-fluorouracil (5-FU)-based chemotherapy and 50.4 Gy radiation) online supplemental method 1. Additionally, all patients underwent staging abdominopelvic and chest CT scan, rectal MRI, colonoscopy, biopsy and positron emission tomography scans. Total mesorectal excision was performed within 6–8 weeks of the final CRT.

Cell cultures and materials

DLD-1 and SW480 human colorectal adenocarcinoma cells and HCT-116 human CRC cells were obtained from the American Type Culture Collection (Rockville, Maryland, USA) and grown in GIBCO RPMI1640 medium (Thermo Fisher Scientific, Wilmington, Delaware, USA) supplemented with 10% heat-inactivated fetal bovine serum (Thermo Fisher Scientific), 2 mM l-glutamine, 100 µg/mL streptomycin and 100 µg/mL penicillin. 5-FU was purchased from Sigma-Aldrich (St. Louis, Missouri, USA).

Gene expression databases

The rectal cancer gene expression profiles used in this study were downloaded from the publicly available GEO database (National Institutes of Health, Bethesda, Maryland, USA; http://www.ncbi.nlm.nih.gov/geo/), with accession numbers GSE1578119 and GSE94104. The gene expression profiles (ID, GSE15781) were previously produced using the ABI Human Genome Survey Microarray V.2 (Thermo Fisher Scientific). Briefly, specimens from GSE15781 were obtained from

---

**Figure 1** Diagram of patient selection and study workflow. Samples before neoadjuvant CRT were obtained at the time of diagnosis using endoscopy. Samples after neoadjuvant CRT were obtained after surgical resection. Samples were obtained from 11 patients after enrollment and were prepared and used for RNA sequencing analysis. The 11 patients underwent curative surgery. CRC, colorectal cancer; CRT, chemoradiation therapy; IHC, immunohistochemistry; KEGG, Kyoto Encyclopedia of Genes and Genomes; LARC, locally advanced rectal cancer; MMR, mismatch repair; MSI, microsatellite instability.
13 patients with LARC who received 50 Gy (delivered as 25 fractions of 2 Gy) over a 5-week period. In addition, patients received Capecitabine (Xeloda; Roche) 2500 mg/mL daily throughout the treatment period. Resection of the rectum was performed 4–6 weeks after preoperative radiation therapy. Among the specimens in the GSE15781 dataset, nine paired tumor tissues before and after CRT were analyzed in this study. An additional gene expression profile (ID, GSE94104) was previously produced using an Illumina Human HT-12 WG-DASL V4.0 R2 expression beadchip (Illumina, San Diego, California, USA). Briefly, the specimens in the GSE94104 dataset were obtained from 40 patients with LARC who received 45 Gy in 25 fractions over a 5-week period along with 5-FU or capecitabine. Patients underwent total mesorectal excision at an interval of 1–14 weeks (mean=9 weeks) following treatment. Prior to bioinformatics analysis, we converted the probe ID to HUGO gene symbols (accessed on December 15 2019) according to the annotation information for the probes from platforms GPL2986 and GPL14951. When there were multiple probe IDs for one gene, only the probe ID with the largest absolute deviation between samples was selected. The probe intensity was normalized using the quantile normalization method and then log2-transformed for gene expression analysis. For gene set enrichment analysis (GSEA) and CIBERSORT analyses, natural scale probe intensities were used.

RNA extraction, library preparation and sequencing
Total RNA was extracted using TRIzol RNA Isolation Reagent (Life Technologies) from the 11 pairs of human LARC. The quantity and quality of the total RNA were evaluated using an Agilent 2100 bioanalyzer RNA kit (Agilent). The isolated total RNA was processed for preparation of an RNA-seq library using an Illumina TruSeq Stranded mRNA Sample Preparation kit (Illumina) according to the manufacturer’s protocol. The quality and size of libraries were assessed using an Agilent 2100 bioanalyzer DNA kit (Agilent). All libraries were quantified by quantitative real-time PCR (qPCR) using a CFX96 Real Time System (Bio-Rad Laboratories, Hercules, California, USA) and sequenced on NextSeq500 sequencers (Illumina) with a paired-end 76 bp plus single 6 bp index read run.

RNA-seq, quality control and mapping
For quality control, low-quality bases and adapter sequences were trimmed from the raw sequencing reads using Trimmomatic software21 with default parameters, except that the minimum length of reads to be dropped was 38 bp (online supplemental table 2). Trimmed sequencing reads were mapped to the hg38 human reference genome22 using STAR aligner23 with default parameters. Reads per gene was counted simultaneously by STAR using the ‘--quantMode GeneCounts’ parameter.

Analysis of differentially expressed genes
Count normalization and differentially expressed gene (DEG) identification in pairwise comparisons between before and after CRT were performed using R software V.3.5.0 (The R Foundation for Statistical Computing, 2018) using DESeq2 package.24 For GSE15781 and GSE94104, we used limma package25 with default parameters. DEGs were defined as satisfying both fold change greater than or equal to two and Benjamini and Hochberg adjusted p value less than 0.05.

Gene ontology and Kyoto Encyclopedia of Genes and Genomes
pathway analyses
The biological significance of DEGs was analyzed using The Database for Annotation, Visualization, and Integrated Discovery (DAVID) 6.8 (http://david.ncifcrf.gov, accessed on May 15 2020).26 The bar plot was generated to show significantly enriched terms of designated genes involved in biological processes. GSEA was performed27 for 186 curated Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets (MsigDB collection name: CP:KEGG of C2) with default parameters, except that the permutation type was set to ‘gene_set’. The significance cut-off for the false-discovery rate q value was set to 0.05.

Analysis of tumor-infiltrating immune cells
The relative proportions of the 22 tumor-infiltrating immune cells (TIICs) in CRC tissues were estimated using CIBERSORT (http://cibersort.stanford.edu/)28 with the leucocyte gene signature matrix (LM22). CIBERSORT analysis was carried out with DESeq2-normalized counts for RNA-seq and natural scale probe intensity for arrays using 500 permutations with quantile normalization. The abundance weighted Bray-Curtis dissimilarity was calculated from the relative proportions of TIICs in each sample. Then, a permutation multivariate analysis of variance with 1000 permutations was conducted to test statistical differences in TIIC compositions between before and after CRT. Principal component analysis (PCA) was performed on the ratio of TIICs in order to explore the similarities among samples. A volcano plot showing the fold change ratios compared with Wilcoxon signed-rank test’s p values was created to identify statistically different TIICs.

Analysis of immune biomarker scores
Three immune biomarker scores were calculated to compare the immune activities of the specimens before and after neoadjuvant CRT. The average of log2-transformed normalized expression counts or intensities (geometric mean) of the genes listed below was used for each score calculation. The gene lists included in each score calculation were as follows: CXCL9, CXCL10, IDO1, IFNG, HLA-DRA, and STAT1; GZMA and PRF1; and CD247, CD2, CD3E, GZMH, NKG7, PRF1 and GZMK for the interferon-γ (IFN-γ) signature;16 the cytolytic activity17 and the immune signature,18 respectively.
Table 1  Primer sequences of mismatch repair-related genes evaluated by quantitative PCR

| Primer name | Sequences |
|-------------|-----------|
| EXO1 sense  | 5'-TGAGAGAGCTATAGAACAGCCAGGT -3' |
| EXO1 antisense | 5'-AGTTTTTCTACACAGCAATAGCC -3' |
| MLH1 sense  | 5'-CTCTCCATCACCATCTGATGGA -3' |
| MLH1 antisense | 5'-GCAGATGCCTGACATACTGTTG -3' |
| LIG1 sense  | 5'-GAAGGAGGCATATAAACAGCAG -3' |
| LIG1 antisense | 5'-ACTGCTGGACACACTCCATT -3' |
| MLH3 sense  | 5'-ACAGCATCATATATTCCATCTGG -3' |
| MLH3 antisense | 5'-TTCAGCTCATCATACTGTTGAG -3' |
| MSH2 sense  | 5'-AGGGATCCAGAGGAGATATGTTG -3' |
| MSH2 antisense | 5'-GGATCCACATACATCCAACTG -3' |
| MSH3 sense  | 5'-GAGCGACCGGATATAGGCTGA -3' |
| MSH3 antisense | 5'-AAAGGGCAGCTCAATTCCGG -3' |
| MSH6 sense  | 5'-CCAGGGGAGAAAAACTCCGA -3' |
| MSH6 antisense | 5'-ACCAGGGGTAAACCTCCATC -3' |
| PCNA sense  | 5'-CTGTTGAGGATATTGCTCCA -3' |
| PCNA antisense | 5'-CAGCGTGAAGGCGTGAAGC -3' |
| PMS2 sense  | 5'-CTATTGATCGGAAGTCCAAGTCA -3' |
| PMS2 antisense | 5'-CTACATCTCCTTATACGAGTCG -3' |
| POLD1 sense | 5'-ATCCAGATATATCTCAGCTGC -3' |
| POLD1 antisense | 5'-ACGGCATGACGTGTTAGGA -3' |
| POLD2 sense | 5'-CCATCAGGCAAACATGCGAC -3' |
| POLD2 antisense | 5'-CTAGCGGCGGATGTTGGA -3' |
| POLD3 sense | 5'-GAGTTCGTCACGGACCAAAAAC -3' |
| POLD3 antisense | 5'-GCCAGACCAAGTATGTAAC -3' |
| POLD4 sense | 5'-ACCAAAGAATCTCGAGGACAG -3' |
| POLD4 antisense | 5'-AGTTGAGCTCTTGACACTCCT -3' |
| RFC1 sense  | 5'-TGAGAGGCAGCAGTGATGAAG -3' |
| RFC1 antisense | 5'-CCTTCTGAGCCTTTTTGCT -3' |
| RFC2 sense  | 5'-GTAGAGGAGTCTAGGCTTGCT -3' |
| RFC2 antisense | 5'-TGAGTTCACATGCGACTCTG -3' |
| RFC3 sense  | 5'-GTGGACCAATGTGCGCTGG -3' |
| RFC3 antisense | 5'-TGATGTTGGCTGACTACAGCT -3' |
| RFC4 sense  | 5'-CGCGTGAACAGGATCAGG -3' |
| RFC4 antisense | 5'-AGGGGACGCGTGGCTTTT -3' |
| RFC5 sense  | 5'-GAAGCGAACGCCATGACACTG -3' |
| RFC5 antisense | 5'-GACCAGGACGGAACTCGT -3' |
| RPA1 sense  | 5'-GGGAGTACCAAAACATAAGCCCAGAC -3' |
| RPA1 antisense | 5'-CGATAACGGGAGGACGATATT -3' |
| RPA2 sense  | 5'-GCACCTTCTCAGGCGGAGAA -3' |
| RPA2 antisense | 5'-CCTTACATCTCTGAGGCAA -3' |
| RPA3 sense  | 5'-AGCTTCAATTCATGCGACAGCC -3' |

Continued

Table 1 Continued

| Primer name | Sequences |
|-------------|-----------|
| RPA3 antisense | 5'-TCTTCTCAAGGGGTCCATCA -3' |
| RPA4 sense  | 5'-GTGCAACACTGTTGAGAGAG -3' |
| RPA4 antisense | 5'-TACACGGTGACACGCTGTTGGA -3' |
| SSBP1 sense | 5'-TGAGTTCGAAACAACTACCTGCT -3' |
| SSBP1 antisense | 5'-CCTGATGCAGGCACATCTCCATTG -3' |
| β-actin sense | 5'-CAGCATCATGCTGTTGTCACCATC -3' |
| β-actin antisense | 5'-AGGCTCCAGGCGGATGCTGATGGG -3' |
| β-glucuronidase sense | 5'-CCCAGCTACTGAGAAATGGGTA -3' |
| β-glucuronidase antisense | 5'-CACAAAACCCAGGCGCAGAAA -3' |

TMB estimation

To evaluate changes in TMB before and after neoadjuvant CRT using RNA-seq data, the following three steps were performed. First, variants, including single-nucleotide polymorphisms (SNPs) and small insertions and deletions, were called from the RNA-seq data according to GATK’s best practice (https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNA-seq-short-variant-discovery-SNPs-Indels). Briefly, sequencing reads were mapped to hg38 using STAR aligner two-pass mode with the ‘sjdbOverhang=75’ parameter. The Picard tool (http://broadinstitute.github.io/picard/) was used for adding read group information, sorting, marking duplicates and indexing to the mapped reads. Mapped reads were split into exon segments, and any overhanging reads were hard-clipped into the intronic regions using SplitNCigarReads tool. Base quality scores were recalibrated based on build 138 of the dbSNP using BaseRecalibrator tool. Variant calling and filtering were performed using HaplotypeCaller and VariantFiltration tools, respectively. The number of variants called from each sample is described in online supplemental table 3. Second, to exclude germline mutations from the called variants, shared mutations between samples collected before and after CRT for a single patient were listed and filtered using the CombineVariants tool in the GATK toolkit. Finally, the number of somatic variants was normalized to the total number of reliable bases having sufficient mapped transcripts of respective sequencing samples because the number of somatic variants estimated from RNA-seq data could not be normalized by the size of the whole exome as in the conventional method. The total number of reliable bases for each sample was counted using the CallableLoci tool in the GATK toolkit. Finally, normalized values were converted into mutations per million base pairs, a conventional unit for the TMB.

Chemoradiation

To estimate the response to chemoradiation, DLD-1 (0.1×10^6 cells/well), HCT-116 (0.07×10^6 cells/well) and SW480 (0.15×10^6 cells/well) cells were plated in 6-wells...
plates and incubated at 37°C under humidified conditions in an atmosphere containing 5% CO₂. The cells were treated with 2.5 μM 5-FU and/or without irradiated with 10 or 20 Gy X-rays radiation in one or two fractions (6 mega-voltage; dose rate: 200 cGy/s) using a linear accelerator (VitalBeam; Varian, USA) for two consecutive days. After radiation, culture medium was exchanged to exclude the effects of radiation on the contents of the culture medium.

RNA isolation and reverse transcription-qPCR
Total cellular RNA was extracted from tissues using TRIzol reagent (Molecular Research Center, Inc., Cincinnati, Ohio, USA). RNA was quantified using a NanoDrop 1000 (Thermo Fisher Scientific). cDNA was synthesized from 2 μg total RNA using MMLV reverse transcriptase (Promega, Madison, Wisconsin, USA) according to the manufacturer’s protocol. qPCR was performed on a LightCycler 480 real-time PCR system (Roche Diagnostics, Mannheim, Germany) using the specific primer pairs described in table 1 and SYBR Green Premix (Toyobo, Japan). β-Actin and β-glucuronidase were used as housekeeping genes for normalization, and a no-template sample was used as a negative control. qPCR data were analyzed using ΔCt values. All experiments were performed with three replicates, and similar results were obtained.

Immunohistochemistry
Immunohistochemistry (IHC) was performed for four MMR proteins (MutL homolog (MLH) 1, MSH2, MSH6 and PMS1 homolog 2 (PMS2)). All human primary LARC samples were FFPE tissues. The representative blocks were selected for each case after review of hematoxylin and eosin slides. We built up sets of tissue microarrays from biopsy and paired surgical specimens. The staining was performed using a BenchMark ULTRA automated staining system (Ventana Medical Systems, Tucson, Arizona, USA), according to the manufacturer’s protocol. The monoclonal antibodies used in this study were as follows: anti-MLH1 (clone M1, ready to use), anti-MSH2 (clone G219-1129, ready to use), anti-MSH6 (clone SP93, ready to use) and anti-PMS2 (clone A16-4, ready to use); all of these antibodies were from Ventana (Roche/Ventana Medical Systems). Bound antibodies were visualized using an OptiView DAB Detection Kit (Ventana Medical Systems). All sections were evaluated by an experienced pathologist (HWL) who was blinded to the clinicopathological features or clinical outcome. Unequivocal nuclear staining of tumor cells compared with adjacent stromal cells or intratumoral lymphocytes (internal control) was regarded as positive. The absolute absence of nuclear staining was considered to indicate a defect in an MMR protein. There were no equivocal findings of IHC in the biopsy specimens. Five surgical specimens showed ambiguous findings, such as reduced expression in both tumor cells and the internal control or aberrant expression. These results could be explained by poor fixation or treatment effects. In these cases, MSI analysis using PCR was also performed for validation.

MSI analysis
MSI assays were performed on DNA extracted from FFPE and matched normal tissues. Tumor and adjacent normal areas were separately marked and collected. Genomic DNA was extracted using a QiAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s protocol. The extracted DNA was amplified by PCR with fluorescent dye-labeled primers targeting five microsatellite loci: BAT25, BAT26, D5S346, D2S123 and D17S250, as recommended by the National Cancer Institute (NCI) guidelines. For MSI analysis, differences in amplified PCR fragments between tumor and normal tissues were detected using a Qsep100 gene analyzer (Biopict, Taiwan, China). In accordance with the NCI criteria, MSI-high tumors were defined as having instability in two or more microsatellite loci; MSI-low tumors were defined as having instability in only one locus; and microsatellite stable (MSS) tumors were defined as showing no apparent instability.

Statistical analysis
Most statistical analyses were performed in R. Graphs, related to R statistical analyses, were drawn using the ggplot2 package in R. Differences in CRC tissues from patients with LARC before and after CRT were statistically analyzed using paired t-tests or Wilcoxon signed-rank tests. Results with p values of less than 0.05 were considered statistically significant.

RESULTS
Neoadjuvant CRT induced MMR deficiency in LARC
To categorize biological processes of DEGs between LARC tissues before and after CRT from the same patient group, we used DAVID and three LARC datasets. As shown in figure 2A and online supplemental figure 2A, significantly enriched biological processes in the same patients with LARC after CRT compared with that before CRT were cell adhesion, extracellular matrix organization, inflammatory response, immune response and response to lipopolysaccharide. Additionally, the significantly enriched biological processes in the same patients with LARC before CRT compared with that after CRT were cell division, DNA replication, mitotic nuclear division, G1/S transition and sister chromatid cohesion (figure 2A and online supplemental figure 2B). To predict the functions and expression trends for genes involved in modulating the effects of CRT in LARC tissues, we applied GSEA using the three LARC cohorts. GSEA yielded a variety of gene pathways and categories that were significantly modified on average across LARC tissues after CRT. Dysregulated mRNAs were associated with six biological pathways, including DNA replication, cell cycle, ribosome, base excision repair, MMR and peroxisome, in all three cohorts (figure 2B and online supplemental figure 2C).
As shown in online supplemental table 4, GSEA of RNA-seq, GSE15781 and GSE94104 datasets showed statistically significant associations with 46, 60 and 23 biological pathways, respectively. GSEA results for all three datasets showed downregulation of the MMR system, with normalized enrichment scores of −2.37 to −1.76 and −2.02, respectively (q<0.001, 0.019 and 0.001, respectively). Then, to investigate whether the 23 MMR system-related genes were altered in our cohort of LARC tissues after CRT compared with that before CRT, statistical comparisons were performed using Wilcoxon signed-rank tests for paired LARC tissues. As shown in online supplemental figure 1, the average mRNA expression levels of the 23 MMR-related genes tended to decrease in LARC tissues after CRT.

Radiation downregulated the mRNA levels of MMR genes in various CRC cells

To validate the observed results regarding the effects of CRT on LARC tissues, we exposed CRC cell lines to 2.5 µM 5-FU and/or radiation (figure 3A) and then examined changes in the expression levels of MMR system-related genes in CRC cells using qPCR. Prior to investigating the effects of radiation on MMR system-related genes, we explored whether radiation induced apoptosis in various CRC cell lines, including DLD-1, HCT-116 and...
SW480 cells. As shown in figure 3B,C, radiation induced morphological changes consistent with apoptosis, such as shrinkage, cellular detachment from the plate and rounding of cells, and increased accumulation of sub-G1-phase cells. Importantly, as demonstrated in figure 3D, mRNA expression levels of almost the 23 MMR system-related genes were markedly decreased, whereas the mRNA levels of EXO1, MLH1 and MLH3 at 10 Gy and RFC3 were not decreased in DLD-1 cells. Moreover, the mRNA expression levels of EXO1, LIG1, MSH2, MSH6, POLD1, POLD3, RFC1, RFC2, RFC3, RFC4, RFC5 and RPA3 mRNAs were markedly decreased in HCT-116 cell lines treated with 5-FU, 10 Gy and 5-FU, and 10 Gy, 10 Gy, and 5-FU. Additionally, in SW480 cells treated with 5-FU and/or radiation, a remarkable decrease was observed in all genes except EXO-1.

**Neoadjuvant CRT altered MSI status and increased TMB in LARC tissues**

To investigate the effects of CRT on MSI status and TMB in LARC tissues, biopsy specimens of LARC tissues before CRT were evaluated. As shown in figure 4A, nuclear expression was maintained in tumor cells for all four MMR system-related proteins, including MLH1, MSH2, MSH6 and PMS2. Moreover, neoadjuvant CRT resulted in loss of MSH6 in one of 11 LARC tissues (figure 4A). In this same case, DNA electropherograms showed that MSI-low was detected in LARC tissue after CRT, whereas the
LARC specimen before CRT for the same patient showed MSS (figure 4B). Furthermore, the TMB was significantly higher in LARC tissues after CRT (p=0.0049) than in that before CRT (figure 4C).

Comparative analysis of TIIC compositions in LARC tissues before and after CRT

To determine the diversity and landscape of TIICs, a gene expression-based deconvolution algorithm, CIBERSORT, was applied to the three cohorts. First, we performed PCA on the relative proportions of TIICs to visualize the effects of irradiation. As shown in figure 5A, D and G, PCA plots exhibited similar compositional shifts in TIICs in all three cohorts. Next, to investigate which TIIC fractions were significantly changed by CRT, we analyzed the proportions of TIICs in the three cohorts. As shown in figure 5C and I, we found that the proportions of M2 macrophages (p=0.01367 and p=0.00009) and CD8 T cells (p=0.03461 and p=0.00009) were significantly increased after CRT in both the RNA-seq and GSE94104 datasets, respectively. Moreover, the proportions of activated memory CD4 T cells, plasma cells, M0 and M1 macrophages, and monocytes in the GSE94104 dataset (p=0.00451, p=0.00196, p=0.01070, p=0.00454 and p=0.02100, respectively, figure 5I) were also significantly altered. Analysis of the GSE94104 dataset showed that the proportions of activated mast cells, neutrophils, resting dendritic cells and naïve CD4 T cells were significantly decreased after CRT in the GSE94104 dataset (p<0.00001, p=0.00183, p=0.00238 and p=0.03098, respectively; figure 5I).

Analysis of immune biomarker scores in LARC tissues before and after CRT

Accumulating evidence has demonstrated that immune biomarker scores could be used to predict responses to immunotherapy.16–18 To investigate the possibility of improving responsiveness to immunotherapy after CRT, three immune biomarker scores (the IFN-γ signature, the cytolytic activity and the immune signature) before and after CRT were analyzed. Scores for the cytolytic activity and the immune signature were significantly increased in all three datasets (figure 6B,C,E,F,H, and I), and the IFN-γ signature tended to increase after CRT in the RNA-seq and GSE15781 dataset (figure 6A–D).

DISCUSSION

Neoadjuvant CRT is a promising therapeutic strategy for patients with LARC to increase rates of tumor
downstaging, clinical response and pathological response and to improve surgical resectability. The main mechanism mediating the anticancer effects of radiotherapy is direct tumor cell killing owing to double-strand breakage of DNA in cancer cells. Nevertheless, recent studies have reported that radiotherapy also has anticancer effects by activation of the immune system through regulation of tumor immunogenicity, including improving intratumoral immune cell infiltration, enriching neoantigens, generating tumor-associated antigen-specific immune cells, facilitating the activity of major histocompatibility complex class I by increasing the intracellular peptide pool and releasing high-mobility group protein box 1. ICBs were approved by the US Food and Drug Administration as innovative anticancer therapies, and synergistic therapeutic strategies using ICBs and radiotherapy have been explored for the treatment of various types of cancer, including hepatocellular carcinoma, CRC, breast cancer, oral cancer and esophageal cancer. As the importance of precision medicine has become clear, studies have also evaluated the effects of CRT on various bioinformatic biomarkers that can predict the reactivity of human cancers to ICBs. Accordingly, we designed this study using three cohorts to explore whether CRT could alter the expression levels of potential biomarkers of the response to immunotherapies, such as ICBs, in LARC.

In this study, we investigated the immune-related efficacy of CRT in LARC and evaluated whether CRT increased immunogenicity or induced immune system activation. After processing our transcriptomic data, we first performed gene ontology (GO) analysis and functional enrichment analysis in our LARC cohort. The results revealed that preoperative CRT significantly enriched the immune response (GO:0006955) in LARC tissues. Additionally, from our CIBERSORT analyses, we found that the proportions of CD8+ T cells and M2 macrophages were significantly increased in the LARC microenvironment after CRT. Ionizing radiation has been shown to recruit M2 macrophages to radiated tissues through directly increasing epithelial-to-mesenchymal transition-related markers and secretion of various chemokines. Therefore, it is possible that CRT may induce the recruitment of M2 macrophages into the LARC microenvironment. However, further studies are needed to investigate the signaling mechanisms and roles of M2 macrophages in the LARC microenvironment after CRT.

Among GO classifications, biological adhesion involves the adhesion of the symbiont to the host, cell adhesion, intermicrovillar adhesion and multicellular organism adhesion. Additionally, radiotherapy leads to upregulation of intercellular adhesion molecule 1 and vascular cell adhesion molecule 1 in the tumor microenvironment. Our results demonstrated that CRT enriched the biological adhesion process and enhanced focal adhesion and cell adhesion molecules in LARC tissues. Among the significantly enriched KEGG gene sets in the all three
cohorts, the DNA replication gene set was the most significantly involved, suggesting impairment of DNA replication by CRT.

MMR is an important and highly conserved biological process that functions to maintain genomic stability.\(^\text{47}\) Notably, the MMR gene set was significantly involved in all three cohorts in this study. Moreover, CRT downregulated MMR system-related genes in the three CRC cell lines examined in this study. Interestingly, 5-FU deepened 10 Gy radiation-induced downregulation of MMR system-related genes in DLD-1 and HCT-116 cells. On the other hand, 20 Gy in two fractions did not affect the 5-FU in the three CRC cell lines (figure 3D). These results should be interpreted with caution because each CRC cell line have different genetic background and CRC cell lines do not always fully reflect the genetic characteristics of LARC. Dose and fractionation regimens of radiotherapy play a crucial role in radiation-induced immunogenic cell death and efficacy of combined treatment with ICD. For example, fractionated radiotherapy (8 Gy x 3) stimulates IFN-1 signal more than single-dose radiotherapy (20 Gy x 1).\(^\text{48}\) Although our experimental conditions of CRT dose not match the reported regimens of CRT, MMR system-related genes were decreased by CRT. These results might be helpful in establishing guideline for combining CRT and immunotherapy.

Importantly, dysregulation of MMR system-related genes causes MSI,\(^\text{49}\) and the four MMR proteins (MLH1, MSH2, MSH6 and PMS2) showing loss in this study by IHC or PCR are important established biomarkers that can predict the response to ICB.\(^\text{50, 51}\) Therefore, we also investigated changes in the MMR system-related proteins and MSI status in LARC tissues. Our results revealed that CRT induced the loss of MSH6 protein and changed the MSI status from MSS to MSI-low in LARC tissue. In samples from the same patient showing loss of MSH6 protein by IHC, the mRNA expression levels of MSH2, MSH6 and PMS2 were downregulated, and that of MLH1
was upregulated (online supplemental figure 1A), marked with red dots). This discrepancy between the results of RNA-seq and IHC may be related to differences in the post-transcriptional regulation of genes and sample status used in RNA-seq and IHC.

A higher TMB is an effective and independent predictive biomarker of responsiveness to ICB in various cancers.13 Our results showed that CRT significantly increased TMB in our LARC cohort, suggesting that CRT may have the potential to increase responsiveness to ICB. A recent study showed that concurrent CRT in esophageal squamous cell carcinoma (ESCC) induced immunogenomic changes, including increased immune scores, enriched immune signaling pathways and increased neutrophil proportions; however, CRT significantly reduced TMB in ESCC tissues after CRT. These outcomes were somewhat different from our results because of differences in the genetic characteristics, CRT regimens and sample collection time points between the two studies. Alternatively, immune biomarker scores may also have applications as biomarkers to predict responsiveness to ICB in cancer.14–16 In the current study, we found that CRT not only significantly increased the cytolytic activity and the immune signature scores but also tended to increase the IFN-γ signature score in both cohorts.

Our study had some limitations. First, the RNA-seq data included only a small number of samples owing to CRT-induced necrosis of tumors. Additionally, few samples were available for IHC analysis because of weak expression and inconsistent results for MSI and MMR system-related genes. TMB analysis using the RNA-seq dataset also did not include sufficient exome sequencing. Finally, survival analysis was difficult because the time after patients enrolled was insufficient, and the experimental model of CRC cell lines was used for evaluation on the single dose of chemotherapy 5-FU (2.5 μM) and the two doses of radiation (10 or 20 Gy X-rays radiation in one or two fractions). Therefore, future studies are needed to confirm and validate our findings. Nevertheless, our findings provided strong in silico and vitro evidence of changes in bioinformatics biomarkers of the response to immunotherapy induced by CRT in patients with LARC. These results provide a molecular basis for combined treatment with radiation and immunotherapy in patients with LARC.

Author affiliations
1Department of Microbiology, Dongguk University College of Medicine, Gyeongju, Gyeongsangbuk-do, Korea (the Republic of)
2Department of Pathology, Keimyung University Dongsan Medical Center, Daegu, Korea (the Republic of)
3Institute for Cancer Research, Keimyung University, Daegu, Korea (the Republic of)
4Department of Radiation Oncology, Keimyung University School of Medicine, Daegu, Korea (the Republic of)
5Department of Immunology, Keimyung University School of Medicine, Daegu, Korea (the Republic of)
6Department of Surgery, CHA University - Bundang Campus, Seongnam, Gyeonggi-do, Korea (the Republic of)
7Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Acknowledgements We would like to thank all the members of our research group for their enthusiastic participation in this study.

Contributors IS, SK and SUB contributed to the conception and design of the study. IS, SHL, J-SL and SK performed the in silico studies. HWL, JYP, HM and SK performed the in vitro studies. HWL, SJB and SUB provided guidance regarding the clinical implications of the current study. IS, JYP and HM contributed to acquisition of data. IS, HWL, SJB, SUB and SK wrote the manuscript. IS, SUB and SK reviewed and edited the manuscript. All authors read and approved the manuscript and agree to be accountable for all aspects of the research.

Funding The current study was supported by the National Research Foundation of Korea (grant funded by the Korean Government Ministry of Science, ICT and Future Planning; grant nos. 201-1R1A5A2010008, 2017R1C1B5076880 and NRF-2019R1I1A301063114).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The experimental study was approved by the Institutional Review Board of Keimyung University Dongsan Medical Center (IRB approval no. 2016-08-020).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. The datasets used in the current study are available from the corresponding author on reasonable request. The primary and processed data used to generate the analyses presented here can be downloaded by registered users from GEO databases (http://www.ncbi.nlm.nih.gov/geo/).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See https://creativecommons.org/licenses/by/4.0/.

ORCID iDs
Sang Jun Byun http://orcid.org/0000-0002-6115-2804
Sung Hwan Lee http://orcid.org/0000-0003-3365-0096
Shin Kim http://orcid.org/0000-0002-1099-5027

REFERENCES
1 Minsky BD, Cohen AM, Kemeny N, et al. The efficacy of preoperative 5-fluorouracil, high-dose leucovorin, and sequential radiation therapy for unresectable rectal cancer. Cancer 1993;71:3486–92.
2 van Gijn W, Marijnissen CAM, Nagtegaal ID, et al. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer: 12-year follow-up of the multicentre, randomised controlled trial. Lancet Oncol 2011;12:575–82.
3 Lim DR, Bae SU, Hur H, et al. Long-Term oncological outcomes of robotic versus laparoscopic total mesorectal excision of mid-low rectal cancer following neoadjuvant chemoradiation therapy. Surg Endosc 2017;31:1728–37.
4 Kim NK, Kim MS, Al-Asari SF. Update and debate issues in surgical treatment of middle and low rectal cancer. J Korean Soc Coloproctol 2012;28:290–40.
5 Khanna KK, Jackson SP. Dna double-strand breaks: signaling, repair and the cancer connection. Nat Genet 2001;27:247–54.
6 Demaria S, Ng B, Devitt ML, et al. Ionizing radiation inhibition of distant untreated tumors (abscopal effect) is immune mediated. Int J Radiat Oncol Biol Phys 2004;58:862–70.
7 Vanpouille-Box C, et al. Combinations of immunotherapy and radiation in cancer therapy. Front Oncol 2014;4:325.
8 Vanpouille-Box C, et al. Immune recognition of irradiated cancer cells. Immunol Rev 2017;280:220–30.
9 Kalbasi A, June CH, Haas N, et al. Radiation and immunotherapy: a synergistic combination. J Clin Invest 2013;123:2756–63.
10 Liu Y, Dong Y, Kong L, et al. Abscopal effect of radiotherapy combined with immune checkpoint inhibitors. J Hematol Oncol 2018;11:104.
11 Darvin P, Toor SM, Sasidharan Nair V, et al. Immune checkpoint inhibitors: recent progress and potential biomarkers. Exp Mol Med 2018;50:1–11.
12 Chang L, Chang M, Chang HM, et al. Microsatellite instability: a predictive biomarker for cancer immunotherapy. Appl Immunohistochem Mol Morphol 2018;26:615–21.
13 Zhang L, Li L, Jiang J, et al.Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy. J Hematol Oncol 2019;12:54.
14 Hellmann MD, Ciuleanu T-E, Pluzanski A, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. N Engl J Med 2018;378:2093–104.
15 Goodman AM, Kato S, Bazhenova L, et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. Mol Cancer Ther 2017;16:2598–608.
16 Gao SY, Burtenshaw-Kerr B, Mera R, et al. Safety and clinical activity of pembrolizumab for treatment of recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-012): an open-label, multicentre, phase 1b trial. Lancet Oncol 2016;17:956–65.
17 Rooney MS, Shukla SA, Wu CJ, et al. Molecular and genetic properties of combination with local immune cytolytic activity. Cell 2015;160:48–61.
18 Davoli T, Uno H, Wooten EC, et al. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science 2017;355:aaw8399.
19 Snapstad T, Fenton CG, Kjaever J, et al. New specific molecular targets for radio-chemotherapy of rectal cancer. Mol Oncol 2010;4:52–64.
20 Li J, Ma S, Lin T, et al. Comprehensive analysis of therapy-related messenger RNAs and long noncoding RNAs as novel biomarkers for advanced colorectal cancer. Front Genet 2019;10:803.
21 Bolger AM, Lohse M, Usadel B. Trimomatic: a flexible trimmer for illumina sequence data. Bioinformatics 2014;30:2114–20.
22 Schneider VA, Graves-Lindsay T, Howe K, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res 2017;27:849–64.
23 Dobin A, Davis CA, Schlesinger F, et al. Star: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.
24 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
25 Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-seq and microarray studies. Nucleic Acids Res 2015;43:4412–21.
26 Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene Lists using David bioinformatics resources. Nat Protoc 2009;4:44–57.
27 Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
28 Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453–7.
Supplementary Table 1. Characteristics and postoperative pathologic outcomes of patients in our LARC cohort

| Case_ID | Age | Sex | TRG | Mandard Testing | MSI Testing | BMI | cT stage | ypT stage | cN stage | ypN stage | Differentiation | Tumor budding | Perineural invasion (PNI) | Extraneal tumor deposit | Exon 2 K-ras mutation | BRAF mutation |
|---------|-----|-----|-----|-----------------|-------------|-----|----------|-----------|----------|-----------|-----------------|--------------|------------------------|----------------------|----------------------|--------------|
| 1       | 69  | M   | NR  | L               |             | 23.9| 2        | 1         | 1        | 0         | 1               | N            | N                      | N                    | Y                    | N            |
| 2       | 47  | M   | NR  | L               |             | 26.3| 4        | 1         | 2        | 0         | 1               | N            | N                      | N                    | N                    | N            |
| 3       | 66  | M   | NR  | L               |             | 20.9| 3        | 3         | 1        | 0         | 1               | N            | N                      | N                    | N                    | Y            |
| 4       | 72  | M   | R   | H               |             | 23.8| 3        | 0         | 0        | 0         | 1               | N            | N                      | N                    | N                    | Y            |
| 5       | 67  | M   | NR  | L               |             | 26.2| 3        | 3         | 2        | 0         | 1               | N            | Y                      | N                    | N                    | N            |
| 6       | 72  | M   | R   | L               |             | 19.5| 3        | 0         | 1        | 0         | 1               | N            | N                      | N                    | Y                    | N            |
| 7       | 72  | M   | R   | L               |             | 22.9| 4        | 3         | 2        | 2         | 1               | N            | N                      | Y                    | N                    | N            |
| 8       | 55  | M   | R   | L               |             | 18.9| 3        | 2         | 2        | 0         | 2               | N            | N                      | N                    | N                    | N            |
| 10      | 68  | M   | NR  | L               |             | 20.3| 3        | 3         | 0        | 0         | 1               | Y            | N                      | N                    | N                    | N            |
| 12      | 54  | M   | R   | L               |             | 29.0| 3        | 0         | 1        | 0         | 1               | N            | N                      | N                    | Y                    | Y            |
| 14      | 76  | M   | NR  | L               |             | 21.3| 3        | 3         | 2        | 1         | 1               | N            | N                      | N                    | N                    | N            |

M, male; R, response to neoadjuvant chemoradiation; NR, non-response to neoadjuvant chemoradiation; H, microsatellite instability-high; L, microsatellite instability-low; MSI, microsatellite instability; BMI, body mass index; cT, clinical primary tumor; ypT, posttreatment primary tumor; cN, clinical regional lymph node; ypN, posttreatment regional lymph node; LARC, locally advanced rectal cancer; N, no; Y, yes
Supplementary Table 2. Read lengths.

| ID    | Raw reads  | Reads after trimming | Mapped reads |
|-------|------------|----------------------|--------------|
| 1-PRERT | 18,058,563 | 16,733,633 (92.66)  | 16,364,242 (97.79) |
| 1-POSTRT | 20,301,918 | 18,765,530 (92.43)  | 18,239,808 (97.20) |
| 2-PRERT | 16,825,688 | 15,542,904 (92.38)  | 15,178,983 (97.66) |
| 2-POSTRT | 18,838,467 | 17,544,965 (93.13)  | 17,072,947 (97.31) |
| 3-PRERT | 16,907,244 | 15,676,022 (92.72)  | 15,333,695 (97.82) |
| 3-POSTRT | 20,538,117 | 19,021,628 (92.62)  | 18,565,691 (97.60) |
| 4-PRERT | 18,861,843 | 17,519,576 (92.88)  | 17,175,773 (98.04) |
| 4-POSTRT | 19,061,963 | 17,662,984 (92.66)  | 17,254,851 (97.69) |
| 5-PRERT | 17,305,382 | 16,216,182 (93.71)  | 15,913,698 (98.13) |
| 5-POSTRT | 17,803,704 | 16,501,411 (92.69)  | 16,000,822 (96.97) |
| 6-PRERT | 18,461,008 | 16,971,725 (91.93)  | 16,609,666 (97.87) |
| 6-POSTRT | 19,629,324 | 18,158,133 (92.51)  | 17,711,303 (97.54) |
| 7-PRERT | 17,973,926 | 16,020,523 (89.13)  | 15,631,032 (97.57) |
| 7-POSTRT | 18,535,525 | 17,099,368 (92.25)  | 16,661,351 (97.44) |
| 8-PRERT | 17,792,102 | 16,057,480 (90.25)  | 15,706,844 (97.82) |
| 8-POSTRT | 17,698,298 | 16,334,788 (92.30)  | 15,803,813 (96.75) |
| 10-PRERT | 18,363,116 | 17,885,076 (92.37)  | 17,454,124 (97.59) |
| 10-POSTRT | 19,596,452 | 16,730,855 (92.97)  | 16,390,228 (97.96) |
| 12-PRERT | 18,641,037 | 17,177,027 (92.15)  | 16,753,695 (97.54) |
| 12-POSTRT | 16,671,573 | 13,756,025 (82.51)  | 13,098,322 (95.22) |
| 14-PRERT | 18,552,789 | 16,459,995 (88.72)  | 15,759,179 (95.74) |
| 14-POSTRT | 21,771,211 | 20,579,574 (94.53)  | 20,220,822 (98.26) |

Note: the minimum length of reads to be dropped was 38 bp.
### Supplementary Table 3. Number of variants called from each sample in our LARC cohort.

| ID    | No. of total variants | Common variants | Callable loci | TMB (somatic variants per Mbp) |
|-------|-----------------------|-----------------|---------------|--------------------------------|
| 1-PRERT | 64,484               | 39,143          | 48,897,606    | 518.25                         |
| 1-POSTRT | 80,526              | 39,143          | 55,051,650    | 751.71                         |
| 2-PRERT  | 63,193               | 31,316          | 47,655,625    | 668.90                         |
| 2-POSTRT | 78,075               | 31,316          | 53,864,994    | 868.08                         |
| 3-PRERT  | 61,853               | 33,744          | 47,055,069    | 597.36                         |
| 3-POSTRT | 78,569               | 33,744          | 53,810,905    | 833.01                         |
| 4-PRERT  | 61,386               | 33,781          | 46,184,868    | 597.71                         |
| 4-POSTRT | 69,435               | 33,781          | 50,501,224    | 706.00                         |
| 5-PRERT  | 63,835               | 32,970          | 47,978,747    | 643.31                         |
| 5-POSTRT | 69,500               | 32,970          | 49,766,227    | 734.03                         |
| 6-PRERT  | 66,198               | 35,764          | 50,631,976    | 601.08                         |
| 6-POSTRT | 70,490               | 35,764          | 50,129,685    | 692.72                         |
| 7-PRERT  | 62,945               | 36,842          | 47,746,549    | 546.70                         |
| 7-POSTRT | 72,143               | 36,842          | 52,064,400    | 678.03                         |
| 8-PRERT  | 45,503               | 27,090          | 38,551,811    | 477.62                         |
| 8-POSTRT | 64,777               | 27,090          | 47,335,812    | 796.16                         |
| 10-PRERT | 69,064               | 38,430          | 51,882,343    | 590.45                         |
| 10-POSTRT | 73,558              | 38,430          | 54,195,701    | 648.17                         |
| 12-PRERT | 56,591               | 25,743          | 42,621,344    | 723.77                         |
| 12-POSTRT | 58,303              | 25,743          | 44,070,288    | 738.82                         |
| 14-PRERT | 62,445               | 27,241          | 41,921,885    | 839.75                         |
| 14-POSTRT | 60,936              | 27,241          | 43,213,310    | 779.74                         |

LARC, locally advanced rectal cancer; Mbp, million base pairs; TMB, tumor mutational burden
Supplementary figure 2B

Down-regulated biological process after CRT

recombination

replication

DNA replication initiation

DNA synthesis involved in DNA replication

Telomere maintenance via recombination

Regulation of transcription involved in G1/S transition of mitotic cell cycle

DNA replication

DNA replication initiation

G1/S transition of mitotic cell cycle

Replication of DNA-Directed DNA polymerase activity

Mitotic sister chromatid segregation

Mitotic nuclear division

Cell division

Cohesion

DNA repair

Double-strand break repair via homologous recombination

S phase enrichment

Mitotic sister chromatid segregation

Mitotic nuclear division

Cell division

Cohesion

DNA replication

DNA synthesis involved in DNA replication

Telomere maintenance via recombination

Regulation of transcription involved in G1/S transition of mitotic cell cycle

DNA replication

DNA replication initiation

G1/S transition of mitotic cell cycle

Replication of DNA-Directed DNA polymerase activity

Mitotic sister chromatid segregation

Mitotic nuclear division

Cell division

Cohesion

DNA repair

Double-strand break repair via homologous recombination

S phase enrichment

Mitotic sister chromatid segregation

Mitotic nuclear division

Cell division

Cohesion
Supplementary Table 4. Significantly enriched biological processes in GSEA of the three cohorts.

Supplementary Table 4-1. Significantly enriched biological processes in GSEA of the RNA-seq cohort

| Biological process                                                      | Size | ES       | NES       | Nom p value | FDR q-value | FWER p value |
|------------------------------------------------------------------------|------|----------|-----------|-------------|-------------|--------------|
| DILATED_CARDIOMYOPATHY                                                 | 90   | 0.5125893| 1.9825155 | 0           | 0.006476901 | 0.006        |
| CALCIUM_SIGNALING_PATHWAY                                              | 175  | 0.4641802| 1.9757535 | 0           | 0.003748867 | 0.007        |
| HYPERTROPHIC_CARDIOMYOPATHY_HCM                                       | 83   | 0.5067126| 1.908026  | 0           | 0.007400138 | 0.021        |
| HEMATOPOIETIC_CELL_LINEAGE                                            | 86   | 0.4956405| 1.872708  | 0           | 0.008608614 | 0.032        |
| VASCULAR_SMOOTH_MUSCLE_CONTRACTION                                    | 114  | 0.4644208| 1.865802  | 0           | 0.006886891 | 0.032        |
| ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC                  | 74   | 0.4987064| 1.8555245 | 0           | 0.007351546 | 0.041        |
| ECM_RECEPTOR_INTERACTION                                               | 84   | 0.4827511| 1.8464414 | 0           | 0.007072384 | 0.046        |
| ASTHMA                                                                | 30   | 0.6003459| 1.8422378 | 0           | 0.006594927 | 0.049        |
| GRAFT_VERSUS_HOST_DISEASE                                              | 41   | 0.5428953| 1.7973992 | 0           | 0.010876824 | 0.09         |
| AUTOImmune_THYROID_DISEASE                                             | 51   | 0.5156497| 1.794059  | 0           | 0.01050675  | 0.097        |
| CELL_ADHESION_MOLECULES_CAMS                                           | 130  | 0.4365804| 1.793934  | 0           | 0.009550614 | 0.097        |
| NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION                                | 270  | 0.3942687| 1.7656447 | 0           | 0.012545356 | 0.14         |
| REGULATION_OF_AUTOPHAGY                                                | 34   | 0.5398197| 1.7385498 | 0.003824092 | 0.01590422  | 0.186        |
| FOCAL_ADHESION                                                        | 197  | 0.4028898| 1.7303211 | 0           | 0.014768204 | 0.186        |
| ALLOGRAFT_REJECTION                                                   | 37   | 0.5302213| 1.7213805 | 0.001934236 | 0.016476285 | 0.217        |
| LONG_TERM_DEPRESSION                                                  | 70   | 0.4613604| 1.707969  | 0           | 0.017353414 | 0.238        |
| LONG_TERM_POTENTATION                                                 | 70   | 0.4372975| 1.6331328 | 0.003816794 | 0.03305391  | 0.419        |
| OLFACtORY_TRANSDuCTION                                                | 379  | 0.35552466| 1.6294472 | 0           | 0.032004103 | 0.428        |
| SPLICEOSOME                                                           | 125  | -0.6826143| -2.765868 | 0           | 0           | 0            |
| DNA_REPLICATION                                                        | 36   | -0.8515296| -2.7587698| 0           | 0           | 0            |
| Pathway                                   | ES  | FDR   | FWER  | GSEA   |
|------------------------------------------|-----|-------|-------|--------|
| CELL_CYCLE                                | 124 | -0.6641761 | -2.6726406 | 0 | 0 | 0 |
| RIBOSOME                                  | 84  | -0.66559803 | -2.5256746 | 0 | 0 | 0 |
| BASE_EXCISION_REPAIR                      | 33  | -0.7684398 | -2.4396727 | 0 | 0 | 0 |
| PROTEASOME                                | 43  | -0.7071348 | -2.4239998 | 0 | 0 | 0 |
| MISMATCH_REPAIR                           | 23  | -0.80656695 | -2.3662093 | 0 | 0 | 0 |
| AMINOACYL_TRNA_BIOSYNTHESIS               | 39  | -0.7060635 | -2.3642282 | 0 | 0 | 0 |
| HOMOLOGOUS_RECOMBINATION                  | 26  | -0.77329946 | -2.327796  | 0 | 0 | 0 |
| PYRIMIDINE_METABOLISM                     | 97  | -0.58644736 | -2.3116546 | 0 | 0 | 0 |
| RNA_POLYMERASE                            | 29  | -0.73242897 | -2.261553  | 0 | 0 | 0 |
| RNA_DEGRADATION                           | 53  | -0.620992   | -2.1442883 | 0 | 0 | 0 |
| NUCLEOTIDE_EXCISION_REPAIR                | 44  | -0.617002   | -2.063633  | 0 | 0 | 0 |
| N_GLYCAN_BIOSYNTHESIS                     | 46  | -0.5494952  | -1.8698767 | 0 | 0.002188098 | 0.032 |
| ALANINE ASPARTATE_AND_GLUTAMATE_METABOLISM | 31  | -0.5881689  | -1.8699325 | 0 | 0.002169371 | 0.04  |
| CYSTEINE_AND_METHIONINE_METABOLISM        | 34  | -0.57071894 | -1.8656715 | 0 | 0.002088731 | 0.035 |
| PURINE_METABOLISM                         | 155 | -0.43708715 | -1.8541654 | 0 | 0.00225027 | 0.04  |
| OOCYTE_MEOISIS                            | 111 | -0.45886746 | -1.832368  | 0 | 0.002783426 | 0.052 |
| BASAL_TRANSCRIPTION_FACTORS               | 35  | -0.5689937  | -1.8084297 | 0.002066116 | 0.003596376 | 0.07  |
| SELENOAMINO_ACID_METABOLISM               | 25  | -0.6029439  | -1.8009231 | 0 | 0.003688005 | 0.076 |
| PENTOSE_PHOSPHATE_PATHWAY                 | 26  | -0.57771033 | -1.7341917 | 0.004166667 | 0.007883973 | 0.162 |
| PROGESTERONE_MEDIATED_OOCYTE_MATURATION   | 85  | -0.44431958 | -1.6868293 | 0.002118644 | 0.012527867 | 0.26  |
| GLYOXYLATE_AND_DICARBOXYLATE_METABOLISM   | 16  | -0.6329198  | -1.6816388 | 0.012552301 | 0.012731631 | 0.274 |
| PEROXISOME                                | 78  | -0.43489367 | -1.6427   | 0.00212766 | 0.018011903 | 0.377 |
| ONE_CARBON_POOL_BY_FOLATE                 | 17  | -0.59410673 | -1.6189371 | 0.018907564 | 0.0206849 | 0.445 |
| STEROID_BIOSYNTHESIS                      | 16  | -0.5842019  | -1.557012  | 0.03050109 | 0.035570364 | 0.638 |
| HUNTINGTONS_DISEASE                       | 160 | -0.364217   | -1.5333275 | 0 | 0.0424422 | 0.719 |
| CITRATE_CYCLE_TCA_CYCLE                   | 30  | -0.48704192 | -1.5312638 | 0.01629328 | 0.04156437 | 0.724 |

ES, enrichment score; FDR, false-discovery rate; FWER, familywise-error rate; GSEA, gene set enrichment analysis; NES, normalized
enrichment score

Supplementary Table 4-2. Significantly enriched biological processes in GSEA in the GSE15781 cohort

| Biological process                                      | Size | ES               | NES               | Nom p value | FDR q-value | FWER p value |
|---------------------------------------------------------|------|------------------|-------------------|-------------|-------------|--------------|
| ASTHMA                                                  | 28   | 0.73588747       | 2.5179844         | 0           | 0           | 0            |
| ALLOGRAFT_REJECTION                                    | 32   | 0.72345424       | 2.498432          | 0           | 0           | 0            |
| GRAFT_VERSUS_HOST_DISEASE                              | 34   | 0.68916994       | 2.3831823         | 0           | 0           | 0            |
| LEISHMANIA_INFECTION                                   | 64   | 0.58616734       | 2.3514082         | 0           | 0           | 0            |
| INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION           | 42   | 0.6263371        | 2.304866          | 0           | 0           | 0            |
| COMPLEMENT_AND_COAGULATION_CASCADES                    | 63   | 0.577753         | 2.2939477         | 0           | 0           | 0            |
| GLYCOsAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE     | 19   | 0.7334301        | 2.2069645         | 0           | 0           | 0            |
| AUTOIMMUNE_THYROID_DISEASE                             | 46   | 0.5839816        | 2.1899981         | 0           | 0           | 0            |
| FOCAL_ADHESION                                         | 186  | 0.445508         | 2.1271925         | 0           | 2.17E-04    | 0.002        |
| ECM_RECEPTOR_INTERACTION                               | 77   | 0.5060727        | 2.0891776         | 0           | 1.96E-04    | 0.002        |
| VIRAL_MYOCARDITIS                                      | 63   | 0.5139724        | 2.016566          | 0           | 8.57E-04    | 0.01         |
| CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION                 | 243  | 0.41357958       | 2.0133467         | 0           | 8.63E-04    | 0.011        |
| HEMATOPOIETIC_CELL_LINEAGE                            | 80   | 0.48226136       | 1.9927986         | 0           | 0.001312858 | 0.018        |
| CELL_ADHESION_MOLECULES_CAMS                           | 122  | 0.4314273        | 1.9497426         | 0           | 0.002034219 | 0.03         |
| TYPE_1_DIABETES_MELLITUS                              | 38   | 0.53817123       | 1.9339674         | 0           | 0.002226526 | 0.034        |
| SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT             | 35   | 0.5134649        | 1.8601531         | 0           | 0.004540115 | 0.07         |
| VASCULAR_SMOOTH_MUSCLE_CONTRACTION                     | 103  | 0.43035984       | 1.8598087         | 0           | 0.00427305  | 0.07         |
| CHEMOKINE_SIGNALING_PATHWAY                            | 172  | 0.39330772       | 1.8331753         | 0           | 0.005704433 | 0.098        |
| Pathway                                             | Score | Z-Score | FDR   | P-Value  | Adj.P-Value |
|-----------------------------------------------------|-------|---------|-------|----------|-------------|
| Leukocyte Transendothelial Migration                 | 106   | 0.4017 | 1.794 | 0.0087   | 0.155       |
| Systemic Lupus Erythematosus                        | 106   | 0.3905 | 1.732 | 0.0137   | 0.249       |
| Nod Like Receptor Signaling Pathway                 | 62    | 0.4337 | 1.729 | 0.0134   | 0.253       |
| Prion Diseases                                     | 35    | 0.4806 | 1.691 | 0.0186   | 0.351       |
| Glycosphingolipid Biosynthesis_Ganglio Series       | 15    | 0.6067 | 1.690 | 0.0180   | 0.352       |
| MAPK Signaling Pathway                              | 253   | 0.3397 | 1.663 | 0.0222   | 0.435       |
| Regulation of Actin Cytoskeleton                    | 195   | 0.3381 | 1.622 | 0.0307   | 0.564       |
| Hypertrophic Cardiomyopathy_HCM                    | 79    | 0.3885 | 1.616 | 0.0310   | 0.578       |
| Epithelial Cell Signaling in Helicobacter_Pylori Infection | 66    | 0.3997 | 1.580 | 0.0404   | 0.697       |
| Natural Killer Cell Mediated Cytotoxicity           | 123   | 0.3487 | 1.570 | 0.0422   | 0.726       |
| Melanoma                                            | 69    | 0.3817 | 1.570 | 0.0410   | 0.728       |
| Neurotrophin Signaling Pathway                      | 120   | 0.3550 | 1.570 | 0.0397   | 0.729       |
| Rig I Like Receptor Signaling Pathway               | 64    | 0.3918 | 1.560 | 0.0413   | 0.756       |
| Dilated Cardiomyopathy                             | 85    | 0.3692 | 1.556 | 0.0417   | 0.774       |
| Pathways in Cancer                                 | 316   | 0.3039 | 1.549 | 0.0428   | 0.788       |
| Antigen Processing and Presentation                | 76    | 0.3769 | 1.536 | 0.0467   | 0.825       |
| Lysosome                                            | 115   | 0.3475 | 1.525 | 0.0496   | 0.847       |
| Toll Like Receptor Signaling Pathway                | 100   | 0.3528 | 1.522 | 0.0489   | 0.848       |
| Citrate Cycle_TCA Cycle                            | 30    | -0.6289 | -2.191 | 0       | 0           |
| Aminoacyl TRNA Biosynthesis                        | 32    | -0.6083 | -2.132 | 0       | 0           |
| DNA Replication                                    | 33    | -0.5951 | -2.104 | 0       | 0           |
| Base Excision Repair                               | 33    | -0.5867 | -2.035 | 0       | 0.0010       |
| Pyrimidine Metabolism                              | 90    | -0.4641 | -2.006 | 0       | 0.0014       |
| Spliceosome                                        | 113   | -0.4429 | -1.996 | 0       | 0.0017       |
| Drug Metabolism Other Enzymes                      | 39    | -0.5204 | -1.897 | 0       | 0.0061       |
| Steroid Biosynthesis                               | 16    | -0.6416 | -1.848 | 0       | 0.0108       |
| Biological process                                      | Size | ES          | NES          | Nom \(p\) value | FDR q-value | FWER \(p\) value |
|--------------------------------------------------------|------|-------------|--------------|-----------------|-------------|-----------------|
| PEROXISOME                                             | 75   | -0.43664226 | -1.8302426   | 0               | 0.012082151 | 0.098           |
| RNA_POLYMERASE                                         | 26   | -0.550258   | -1.7956698   | 0.011857707     | 0.015288933 | 0.13            |
| VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION              | 42   | -0.4746235  | -1.7846826   | 0.004140787     | 0.015903868 | 0.149           |
| MISMATCH_REPAIR                                        | 22   | -0.5634375  | -1.7611866   | 0.002079002     | 0.018849427 | 0.191           |
| TERPENOID_BACKBONE_BIOSYNTHESIS                        | 15   | -0.6244792  | -1.7600522   | 0.006147541     | 0.017642075 | 0.192           |
| ALANINE_ASPARTATE_AND_GLUTAMATE_METABOLISM             | 31   | -0.4917142  | -1.7519854   | 0.004338395     | 0.017693076 | 0.204           |
| CELL_CYCLE                                             | 122  | -0.3796363  | -1.74143     | 0               | 0.018191693 | 0.224           |
| SELENOAMINO_ACID_METABOLISM                            | 21   | -0.53938395 | -1.7279283   | 0.010416667     | 0.018887721 | 0.246           |
| RIBOSOME                                               | 75   | -0.40516287 | -1.7208233   | 0               | 0.018710503 | 0.259           |
| PARKINSONS_DISEASE                                     | 97   | -0.39299116 | -1.7123795   | 0               | 0.019501364 | 0.278           |
| GLYCOSPHINGOLIPID_BIOSYNTHESIS_LACTO_AND_NEOLACTO_SERIES | 26   | -0.5284643  | -1.7038212   | 0.008639309     | 0.019893743 | 0.295           |
| GLUTATHIONE_METABOLISM                                 | 44   | -0.44670752 | -1.6992595   | 0               | 0.019898029 | 0.312           |
| HUNTINGTONS_DISEASE                                    | 150  | -0.35769758 | -1.6822829   | 0               | 0.021907598 | 0.348           |
| PROXIMAL_TUBULE_BICARBONATE_RECLAMATION                | 23   | -0.5226012  | -1.6757207   | 0.015151516     | 0.02179761  | 0.359           |
| PROPAANOATE_METABOLISM                                 | 30   | -0.46956128 | -1.6182915   | 0.00631579      | 0.034719527 | 0.525           |
| THYROID_CANCER                                         | 28   | -0.4587294  | -1.5823514   | 0.022177419     | 0.045327876 | 0.643           |

ES, enrichment score; FDR, false-discovery rate; FWER, familywise-error rate; GSEA, gene set enrichment analysis; NES, normalized enrichment score.

Supplementary Table 4-3. Significantly enriched biological processes in GSEA in the GSE94104 cohort.
| Term                                        | ES       | NES       | FDR          | FWER          | NES         |
|---------------------------------------------|----------|-----------|--------------|---------------|-------------|
| DILATED_CARDIOMYOPATHY                      | 0.5738871| 2.09093   | 9.49E-04     | 0.001         |             |
| MELANOMA                                    | 0.5887409| 2.0711582 | 9.36E-04     | 0.002         |             |
| HYPERTROPHIC_CARDIOMYOPATHY_HCM             | 0.5681370| 2.0289567 | 6.24E-04     | 0.002         |             |
| CALCULI_SIGNALING_PATHWAY                   | 0.4930849| 1.9452127 | 0.003931471  | 0.016         |             |
| ECM_RECEPTOR_INTERACTION                    | 0.5363607| 1.9236966 | 0.004036177  | 0.021         |             |
| FOCAL_ADHESION                              | 0.4612941| 1.8581474 | 0.008099408  | 0.051         |             |
| ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC | 0.5191158| 1.8241311 | 0.010827938  | 0.077         |             |
| CELL_ADHESION_MOLECULES_CAMS                | 0.4693440| 1.8116379 | 0.011433065  | 0.092         |             |
| VIRAL_MYOCARDITIS                           | 0.5030668| 1.7191666 | 0.008128142  | 0.266         |             |
| DNA_REPLICATION                             | -0.7007715| -2.301029 | 0             | 0             |             |
| CELL_CYCLE                                  | -0.5712032| -2.295487 | 0             | 0             |             |
| MISMATCH_REPAIR                             | -0.6837358| -2.020403 | 7.03E-04     | 0.002         |             |
| PEROXISOME                                  | -0.5047265| -1.893322 | 0.005765674  | 0.021         |             |
| ASCORBATE_AND_ALDARATE_METABOLISM           | -0.6515238| -1.8501037| 0.008128142  | 0.037         |             |
| BASE_EXCISION_REPAIR                        | -0.5739228| -1.8270599| 0.008980554  | 0.047         |             |
| MATURITY_ONSET_DIABETES_OF_THE_YOUNG        | -0.6287635| -1.8215218| 0.008178927  | 0.05          |             |
| RIBOSOME                                    | -0.4652022| -1.7553846| 0.016389811  | 0.112         |             |
| HOMOLOGOUS_RECOMBINATION                    | -0.5782895| -1.7548293| 0.014813413  | 0.114         |             |
| PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS    | -0.5949544| -1.7142693| 0.0222204    | 0.181         |             |
| PORPHYRIN_AND_CHLOROPHYLL_METABOLISM        | -0.5163599| -1.672384| 0.03282529   | 0.276         |             |
| RIBOFLOVIN_METABOLISM                       | -0.6053013| -1.637684| 0.04302932   | 0.371         |             |
| FRUCTOSE_AND_MANNOSE_METABOLISM             | -0.5114403| -1.6306827| 0.04293906   | 0.394         |             |
| DRUG_METABOLISM_OTHER_ENZYMES               | -0.4819766| -1.6222355| 0.0436709    | 0.425         |             |

ES, enrichment score; FDR, false-discovery rate; FWER, familywise-error rate; GSEA, gene set enrichment analysis; NES, normalized enrichment score
Supplemental Figure Legends

Supplementary Figure 1. Changes in mRNA expression profiles of mismatch repair (MMR) system-related genes in locally advanced rectal cancer (LARC) tissues before and after chemoradiation therapy (CRT). A, Expression levels (log2[normalized count+1]) of 23 MMR system-related genes in the RNA-seq cohort of LARC tissues before and after CRT. B, Expression levels (log2[normalized count+1]) of 23 MMR system-related genes in the GSE15781 cohort of LARC tissues before and after CRT. C, Expression levels (log2[normalized count+1]) of 23 MMR system-related genes in the GSE94104 cohort of LARC tissues before and after CRT. *P values were calculated by Wilcoxon signed-rank tests. Pre, LARC tissues before CRT; Post, LARC tissues after CRT.

Supplementary Figure 2. Functional enrichment analysis of differentially expressed genes (DEGs) in locally advanced rectal cancer (LARC) tissues before and after CRT. A, Significantly enriched biological processes in LARC tissues after CRT compared with that before CRT. B, Significantly enriched biological processes in LARC tissues before CRT compared with that after CRT. C, Gene set enrichment analysis (GSEA) plots (score curves) of the three cohorts. Representative GSEA of DEGs in tumor samples before and after CRT in the RNA-seq, GSE15781, and GSE94104 datasets.
Supplementary Method 1. Preoperative chemoradiation regimen of patients in our LARC cohort.

Patients underwent fluoropyrimidine-based CONCURRENT chemoradiation therapy prior to surgery. A dose of 50.4 Gy was given, which included 45 Gy in 25 fractions for the pelvis and a 5.4 Gy boost in 3 fractions for RESIDUAL primary tumor(s) over 5.5 weeks. The chemotherapy regimens are: 2 cycles of intravenous bolus of fluorouracil (400 mg/m² per day) and leucovorin (20 mg/m² per day) for 5 days in the first and fifth weeks of radiation therapy.