의학박사 학위논문

국소 진행성 직장 선암의 유전체 지형: 선행 항암방사선치료 전후의 비교 및 유전적 바이오마커의 임상적 결과와 종양 반응에 대한 영향

Genomic landscape of locally advanced rectal adenocarcinoma: comparison of before and after neoadjuvant chemoradiation, and impact of genetic biomarkers on clinical outcomes and tumor response

2023년 8월

서울대학교 대학원

임상의과학과

이 태 훈

Ph.D. Dissertation of Clinical Medical Science

Genomic landscape of locally advanced rectal adenocarcinoma: comparison of before and after neoadjuvant chemoradiation, and impact of genetic biomarkers on clinical outcomes and tumor response

국소 진행성 직장 선암의 유전체 지형:
선행 항암방사선치료 전후의 비교 및 유전적 바이오마커의 임상적 결과와 종양 반응에 대한 영향

August 2023

Department of Clinical Medical Science
Seoul National University Graduate School

Tae Hoon Lee

국소 진행성 직장 선암의 유전체 지형: 선행 항암방사선치료 전후의 비교 및 유전적 바이오마커의 임상적 결과와 종양 반응에 대한 영향

지도 교수 지 의 규

이 논문을 의학박사 학위논문으로 제출함
2023년 4월

서울대학교 대학원
임상의과학과
이 태 훈

이태훈의 의학박사 학위논문을 인준함
2023년 7월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

Genomic landscape of locally advanced rectal adenocarcinoma: comparison of before and after neoadjuvant chemoradiation, and impact of genetic biomarkers on clinical outcomes and tumor response

by

Tae Hoon Lee

A thesis submitted to the Department of Clinical Medical Science in partial fulfilment of the requirements for the Degree of Doctor of Philosophy at Seoul National University Graduate School

July 2023

Approved by Thesis Committee:

Chair	_____(Seal)
Vice Chair	_____(Seal)
Examiner	_____(Seal)
Examiner	_____(Seal)
Examiner	_____(Seal)

Abstract

Purpose: To explore genomic biomarkers in rectal cancer by performing whole-exome sequencing (WES).

Materials and Methods: Pre-chemoradiation (CRT) biopsy and post-CRT surgical specimens were obtained from 27 patients undergoing neoadjuvant CRT followed by definitive resection. Exomes were sequenced to a mean coverage of 30x. Somatic single nucleotide variants (SNVs) and insertions/deletions (indels) were identified. Tumor mutational burden was defined as the number of nonsynonymous mutations per megabase pair. Mutational signatures were extracted and fitted to COSMIC reference signatures. Tumor heterogeneity was quantified with a Mutant-Allele Tumor Heterogeneity (MATH) score. Genetic biomarkers and frequently occurred copy number alterations (CNAs) were compared between pre- and post-CRT specimens. Their associations with tumor regression grade (TRG) and clinical outcomes were explored.

Results: Top five mutated genes were *APC*, *TP53*, *NF1*, *KRAS*, and *NOTCH1* for pre-CRT samples and *APC*, *TP53*, *NF1*, *CREBBP*, and *ATM* for post-CRT samples. Several gene mutations including *RUNX1*, *EGFR*, and *TP53* in pre-CRT samples showed significant association with clinical outcomes, but not with TRG. However, no such association was found in post-CRT samples. Discordance of driver mutation status was found between pre- and post-CRT samples. In tumor mutational burden analysis, higher number of mutations per megabase pair was associated with worse treatment outcomes. Six single-base substitution (SBS) signatures identified were

SBS1, SBS30, SBS29, SBS49, SBS3 and SBS44. The MATH score decreased after CRT on paired analysis. Less than half of CNAs frequent in post-CRT samples were present in pre-CRT samples.

Conclusion: Pre- and post-CRT samples showed different genomic landscape. Potential genetic biomarkers of pre-CRT samples found in the current analysis call for external validation.

Keyword: Rectal cancer, neoadjuvant chemoradiation, whole exome sequencing

Student number: 2021-35331

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CEA – Carcinoembryonic antigen

CNA – Copy number alteration

COSMIC – Catalog of Somatic Mutations in Cancer

CRT – Chemoradiation

DBS – Doublet-base substitution

DMFR – Distant metastasis-free rate

FFPE – Formalin fixed paraffin embedded

H&E – Hematoxylin and eosin

IGV – Integrated Genomics Viewer

Indel – Insertion/deletion

IRB – Institutional Review Board

LRCR – Locoregional control rate

MATH – Mutant-Allele Tumor Heterogeneity

MSI – Microsatellite instability

OSR – Overall survival rate

PFSR – Progression-free survival rate

SBS – Single-base substitution

SNV – Single nucleotide variant

TCGA – The Cancer Genome Atlas

TRG – Tumor regression grade

VAF – Variant allele frequency

WES – Whole exome sequencing

Introduction

Neoadjuvant chemoradiation (CRT) with radical surgery is a standard treatment for locally advanced rectal cancer. It is well-known that response after neoadjuvant CRT can significantly impact clinical outcomes [1]. Several studies have reported that patients with complete response after neoadjuvant therapy may show good rectal preservation and pelvic tumor control with a "Watch-and-Wait" strategy [2, 3]. Furthermore, applying more intense CRT for those expected to have a poor response or prognosis prior to treatment initiation might improve overall outcome. Based on these approaches, prediction of response and prognosis prior to neoadjuvant CRT could help tailor individualized treatment. Statically significant models to predict tumor response to CRT based on integrated clinical factors have been reported [4]. However, no model is widely accepted for clinical practice.

There have been several attempts to utilize genomic features to predict tumor response and prognosis in rectal cancer. It is widely accepted that mutations of some specific genes such as *KRAS* and *TP53* are correlated with worse prognosis or poor tumor response after neoadjuvant CRT [5, 6]. In addition, several studies have reported promising results of predicting outcomes based on gene expression profiling, although these studies lack concordance to integrate such results into routine practices [7]. Recent advance in molecular biology has enabled the sequencing of large amounts of DNA in a short period of time. Such high-throughput sequencing has been applied to various cancers. Several studies have analyzed whole exome sequencing (WES) materials from rectal tumors to predict tumor response

and prognosis after neoadjuvant CRT [8, 9]. While each study has shown statistical significance on its own to predict outcomes, inconsistency among studies remains as an obstacle to introduce these models to clinical practices. Therefore, more robust and validated studies are needed to explore clinically useful genetic markers.

Recent trends in neoadjuvant CRT involve treatment intensification. Administering neoadjuvant chemotherapy before or after neoadjuvant radiotherapy or CRT has shown a better response rate and improved treatment outcomes [10, 11]. This strategy, known as total neoadjuvant therapy, has recently been incorporated into clinical guidelines. Patients who do not achieve a pathologic complete response after standard neoadjuvant CRT would benefit the most from such intensified treatment. Therefore, identifying specific changes in the genomic landscape of non-responders, understanding their clinical significance, and incorporating this into prediction models would be stepping stone for personalized treatment in patients with locally advanced rectal cancer. The purpose of this study was to provide genetic data using tissues acquired before and after CRT and find clinically significant biomarkers for predicting prognosis or tumor response by performing WES for materials obtained from patients with locally advanced rectal cancer who underwent neoadjuvant CRT and surgery.

Materials and Methods

1. Tumor samples

This study was approved by the Institutional Review Board (IRB) of Seoul National University Hospital (approval no. H-2011-047-1171). Tissues analyzed in this study were previously donated to the institutional repository with consents of the patients for further research after pathologic diagnosis. Additional informed consent for this study was waived by the IRB. Formalin fixed paraffin embedded (FFPE) tissues were obtained from patients with locally advanced rectal adenocarcinoma (cT3-4 or N+ without systemic metastasis) who underwent neoadjuvant CRT and definitive surgery from 2008 to 2016. These patients underwent neoadjuvant radiation therapy for the rectum and pelvic lymph nodal area with concurrent 5-fluorouracil or capecitabine. Total mesorectal excision was performed at 6 to 8 weeks after the end of neoadjuvant treatment. Post-operative chemotherapy was administered at the discretion of practicing medical oncologists. Tumors with complete response in post-CRT surgical samples were excluded as there was no analyzable post-CRT tumor tissue in such samples. Normal samples were obtained from unaffected and disease-free sites of biopsy and surgical samples. Pathologic tumor response was evaluated using three-point tumor regression grade (TRG) system proposed by Ryan et al. [12].

2. Sample processing

Two tumor samples and one normal sample were retrieved from pre-CRT biopsy and

post-CRT surgical specimens of each patient for sequencing. Tumor and normal areas were identified in hematoxylin and eosin (H&E) stained slides and superimposed to unstained slides. Manual macrodissection was performed for corresponding area using a scalpel. Macrodissected FFPE tissue was digested in a cell lysis solution with Proteinase K to extract DNA.

3. Whole exome sequencing

To build standard exome capture libraries, an Agilent SureSelect Target Enrichment protocol for Illumina paired-end sequencing library and 1 μg input genomic DNA were used. In all cases, the SureSelect Human All Exon V6 probe (Agilent Technologies Inc., Santa Clara, CA, USA) set was used. FFPE genomic DNA was sheared using a Covaris LE220 Focused-ultrasonicator (Covaris LLC, Woburn, MA, USA). A SureSelect All Exon Capture Library was used for exome capture according to the manufacturer's protocol. Indexed libraries were sequenced to a mean coverage of 30x using a NovaSeq 6000 platform (Illumina Inc., San Diego, CA, USA) by Macrogen Incorporated (Seoul, Republic of Korea).

4. Single nucleotide variant and indel analysis

Generated FASTQ files were aligned to the GRCh38 reference genome using BWA-MEM algorithm [13]. Duplicate reads were marked using Picard Tools (Broad Institute, Boston, MA, USA). Base quality score was recalibrated and applied to mark-duplicated BAM files using GATK 4.1.0.0 (Broad Institute, Boston, MA,

USA). These procedures were processed by Clara Parabricks 3.6.1 (NVIDIA Corporation, Santa Clara, CA, USA). Somatic single nucleotide variants (SNVs) and insertions/deletions (indels) were identified with Mutect2 of GATK 4.2.4.0. As two tumor samples were obtained from each pre-CRT biopsy and post-CRT surgical specimens, joint analysis of two BAM files from the same specimen was performed. Identified SNVs and indels were filtered using FilterMutectCalls with additional condition of minimum variant allele fraction (VAF) of 5% and annotated using Funcotator of GATK 4.2.4.0.

Annotated mutations by Catalog of Somatic Mutations in Cancer (COSMIC) (https://cancer.sanger.ac.uk/cosmic) database v84 were selected for clinical enrichment. Correlation of specific gene mutations with clinical outcomes and TRG measured in post-CRT specimen was analyzed using Maftools [14]. SNVs and indels from pre- and post-CRT samples were also compared to find changes caused by neoadjuvant CRT. Shared mutations were defined as SNVs and indels located in the gene mutated in both samples. Percentages of shared mutations between pre- and post-CRT samples from the same patient were measured and correlated with clinical outcomes and TRG.

Mutational signatures were extracted and fitted to COSMIC reference signatures (single-base substitution [SBS] and doublet-base substitution [DBS], and indel) [15] using Sigflow 1.5 [16]. Automatic extraction of signatures using Bayesian variant of non-negative matrix factorization algorithm was performed. Cosine similarity analysis was applied to extracted signatures and COSMIC reference signatures for fitting. Difference of mutational signatures between pre- and post-CRT samples and correlation of percentages of each mutational signatures with clinical

outcomes and TRG were analyzed.

Tumor mutational burden was defined as the number of identified nonsynonymous mutations per megabase pair. The total length of protein-coding regions was assumed to be 35 megabase pairs. Difference of tumor mutational burden between pre- and post-CRT samples and correlation of mutational burden with clinical outcomes and TRG were also analyzed.

Mutant-Allele Tumor Heterogeneity (MATH) score calculated as 100 x median absolute deviation divided by median of VAF was used for tumor heterogeneity quantification [17]. MATH scores were calculated for each sample without joint analysis of two samples from the same tumor specimen. Changes of MATH score caused by neoadjuvant CRT were identified by comparing MATH scores of pre- and post-CRT samples. Effects of MATH score on clinical outcomes and TRG were analyzed. SNVs and indels incorporated in tumor mutational burden, mutational signature, and MATH score analyses were not filtered by COSMIC database.

5. Copy number analysis

BAM files generated in the previous analysis were processed using Sequenza 2.1.2 [18] for copy number alteration (CNA) detection and cellular/ploidy estimation. CNAs were analyzed per sample without joint analysis. Significantly amplified or deleted regions were identified with GISTIC2 [19] when q-value adjusted by false discovery rate was below 0.1. Frequently occurred CNAs were compared between pre- and post-CRT samples. CNA burden was defined as the ratio of the summation

of the length of copy number-altered segments to the total length of analyzed segments. Ploidy was considered when determining the copy number of a certain segment, by setting the closest integer number to the estimated ploidy number as the standard copy number for the specific sample. The correlation of the existence of certain CNAs and CNA burden with treatment outcomes and TRG was analyzed.

6. Clinical outcomes and statistical analysis

Clinical outcomes analyzed in the current study were locoregional control rate (LRCR), distant metastasis-free rate (DMFR), progression-free survival rate (PFSR), and overall survival rate (OSR). An LRCR event was defined as a recurrence within the pelvis including anastomotic site and regional pelvic nodal area. A DMFR event was defined as occurrence of distant metastasis. For PFSR, the event was defined as any recurrence or death, while it was death irrespective of cause for OSR. These events were measured from the date of the completion of radiation therapy. Clinical outcomes at defined time points were calculated using Kaplan-Meier estimator.

A multivariate analysis was performed on four clinical outcomes in conjunction with clinical variables and genetic biomarkers identified in previous steps. Clinical variables included old age (≥60 years), high-risk T3–4 disease (T3 with mesorectal fat >5 mm extension and/or mesorectal fascia involvement, or T4), N2 disease, proximal rectal cancer (anal verge ≥10cm), differentiation (moderate vs. well-differentiated), and elevated carcinoembryonic antigen (CEA) levels at pre-CRT. Univariate analysis was first performed to identify clinical variables with statistical significance or marginal significance ($p<0.1$), and selected clinical

variables and genetic biomarkers were incorporated in the final multivariate analysis. MATH score and CNA analysis results were not included in this multivariate analysis because the MATH score was calculated per tumor sample and not per patient, and the CNA analysis was only performed on patients with analyzable normal sample pairs. Cox proportional hazards model was applied in the univariate and multivariate analyses. A backward stepwise selection based on Akaike information criterion was used to select variables independently associated with the four clinical outcomes.

Associations of genomic biomarkers with clinical outcomes were calculated using univariate Cox proportional hazards model or log-rank test. Associations between genomic biomarkers and TRG were calculated using Kruskal-Wallis test or Wilcoxon test. P-value lower than 0.05 was considered as statistically significant. All statistical analyses were performed using R 4.2.0 (The R Foundation for Statistical Computing, Vienna, Austria).

Results

1. Patient characteristics and analyzable samples

Paired set of samples were available from 27 patients, including 18 (66.7%) males and 9 (33.3%) females. The median age at diagnosis was 60 years (range, 31–81 years). Total radiation dose ranged from 50.4–54.0 Gy in 1.8 Gy per fractions. Concurrent capecitabine was administered to 14 (51.9%) patients and 5-FU was given to 13 (48.1%) patients. Sphincter sparing surgery, either low anterior resection or ultralow anterior resection, was performed in 13 (48.1%) and 12 (44.4%) patients, respectively. The remaining two (7.4%) patients underwent abdominoperineal resection. The median time from CRT to surgery was 50 days (range, 39-76 days). Post-operative chemotherapy was applied to 24 (88.9%) patients. The median follow-up of patients was 69.7 months (range, 4.9–127.6 months). Ten (37.0%) patients experienced recurrence, with distant metastasis being the most frequent one at nine (33.3%) events. There were two isolated local failures and two locoregional failures. Nine (33.3%) patients succumbed to the disease.

Several samples failed to pass quality check of exome capture. A total of 25 patients were able to be analyzed using pre-CRT samples. Both pre-CRT tumor and normal samples were analyzable for 17 patients, whereas pre-CRT normal samples were not analyzable for 8 patients. For post-CRT samples, tumor and normal samples from 12 patients were able to be sequenced. Eleven patients had analyzable pairs of pre- and post-CRT tumor samples. One patient had only post-CRT normal sample that passed that quality check. This patient was excluded from further analysis.

Among 25 patients with analyzable pre-CRT samples, 8 (32.0%), 13 (52.0%), and 4 (12.0%) patients had TRG 1, TRG 2, TRG 3, respectively. Among 12 patients with analyzable post-CRT samples, 2 (16.7%), 6 (50.0%), and 4 (33.3%) patients had TRG 1, TRG 2, TRG 3, respectively.

2. Single nucleotide variants and indels

Pre-CRT mutational landscape for 25 patients is illustrated in Figure 1A. Genes that mutated in at least four patients were included. Top five frequently mutated genes were *APC*, *TP53*, *NF1*, *KRAS*, and *NOTCH1*. Corresponding genes for colorectal adenocarcinoma data from The Cancer Genome Atlas (TCGA) PanCancer Atlas (https://datacatalog.mskcc.org/dataset/10411, N=594) were *APC*, *TP53*, *TTN*, *KRAS*, and *PIK3CA*. Integrated Genomics Viewer (IGV) screenshots for the top five frequently mutated genes in pre-CRT samples were illustrated in Figure 2A. *APC*, *TP53*, and *KRAS* were compared with TCGA data as they were major driver mutations of colorectal cancer [20]. Patients in this study had slightly lower rates of mutations of these genes compared to TCGA data (*APC*, 56.0% vs. 72.5%; *TP53*, 48.0% vs. 58.8%; *KRAS*, 32.0% vs. 40.8%). Other notable genes such as *EGFR*, *BRAF*, *NRAS*, *PIK3CA*, and *SMAD4* were also compared. Patients in this study showed higher mutation rates of *EGFR* (16.0% vs. 2.8%), *BRAF* (16.0% vs 11.6%), and *SMAD4* (16.0% vs. 12.7%) but a lower mutation rate of *PIK3CA* (12.0% vs. 27.5%) than TCGA data. No *NRAS* and *HRAS* mutation was observed in pre-CRT samples. Four genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*) associated with microsatellite instability (MSI) were also explored [21]. There were two (8.0%) patients with

MLH1 mutations, 2 (8.0%) patients with *MSH2* mutations, and 4 (16.0%) patients with *MSH6* mutations. No *PSM2* mutation was observed. Six (24.0%) patients had at least one of these MSI-related gene mutations. Gene mutations in other members of *EGFR* family were observed (*ERBB2*, 2 [8.0%]; *ERBB3*, 2 [8.0%]; *ERBB4*, 3 [12.0%]). Co-mutation plot for pre-CRT mutations was illustrated in Figure 3A.

Associations of clinical outcomes with specific gene mutations were explored by Cox proportional hazards model. Thirty genes that mutated in at least four patients were analyzed. *SMAD4* and *EGFR* mutations were significantly associated with worse LRCR. *RUNX1* mutation was significantly associated with worse DMFR. *RUNX1*, *GATA1*, *BAP1*, *KMT2A*, and *TP53* mutations showed statistically significant associations with worse PFSR. *RUNX1*, *EGFR*, *STAG2*, and *TP53* mutations showed statistically significant associations with worse OSR, while *APC* mutations were associated with improved OSR. Hazard ratio, confidence interval, and *P*-value are summarized in Table 1. Kaplan-Meier curves of clinical outcomes affected by these gene mutations are illustrated in Figure 4. Mutations of MSI genes did not influence clinical outcomes (LRCR, HR 1.635, 95% CI 0.169–15.82, p=0.671; DMFR, HR 1.379, 95% CI 0.286–6.656, p=0.689; PFSR, HR 1.743, 95% CI 0.461–6.599, p=0.413; OSR, HR 2.484, 95% CI 0.616–10.01, p=0.201). Relationships of mutations in pre-CRT samples with TRG were also explored. No gene mutation significantly associated with TRG was discovered, although no patient with TRG 1 harbored *BAP1* mutation in pre-CRT samples (TRG 1, 0 in 8 [0.0%]; TRG 2, 5 in 13 [38.5%]; TRG 3, 1 in 4 [25.0%]; p=0.134).

Post-CRT mutational landscape for 12 patients is illustrated in Figure 1B. Genes showing mutations in at least two patients are included in the figure. One

patient had none of the 39 frequently mutated genes. Thus, this patient was not included in the figure. Top five frequently mutated genes were *APC*, *TP53*, *NF1*, *CREBBP*, and *ATM*. IGV screenshots for the top five frequently mutated genes in post-CRT samples were illustrated in Figure 2B. Top three genes remained, whereas mutation rate of *KRAS* and *NOTCH1* decreased from 32.0% to 16.7%. No *NRAS* mutation was observed, but there was 2 (1.67%) patients with *HRAS* mutation. Among four MSI genes were explored, 2 (16.7%) patients harbored *MLH1* mutations. *MSH2*, *MSH6*, and *PSM2* showed no mutations. There was 1 (8.3%) patient with *EGFR* mutation and 1 (8.3%) patient with *ERBB4* mutation. No *ERBB2* and *ERBB3* mutations were observed. Co-mutation plot for pre-CRT mutations was illustrated in Figure 3B. Association of clinical outcomes and specific gene mutations were explored for twelve genes showing mutations in at least three patients. Clinical outcomes showed no statistically significant associations with specific gene mutations. When relationships between mutations in post-CRT samples and TRG were explored, no significant association was found. Although statistically non-significant, no patients with TRG 1 had *BRAF* mutation in post-CRT samples. A high percentage of *BRAF* mutation was observed in patients with TRG 3 (TRG 1, 0 in 2 [0.0%]; TRG 2, 1 in 6 [16.7%]; TRG 3, 3 in 4 [75.0%]; p=0.087). In addition, all patients with TRG 3 had *TP53* mutation in post-CRT samples (TRG 1, 1 in 2 [50.0%]; TRG 2, 2 in 6 [33.3%]; TRG 3, 4 in 4 [100.0%]; p=0.108).

Figure 1. Mutational landscape of (A) pre-chemoradiation samples, and (B) post-chemoradiation samples. Each column represents a patient and each row represents a gene. Total numbers of identified single nucleotide variants (SNVs) and indels of each patient are shown in the top. Mutation rates of specific genes are illustrated in the right. Genes are sorted by mutation rates. (C) Comparison of gene mutation rates between pre- and post-chemoradiation samples in patients with analyzable sample pair. Maftools [14] package of R 4.2.0 was used for generating the figure.

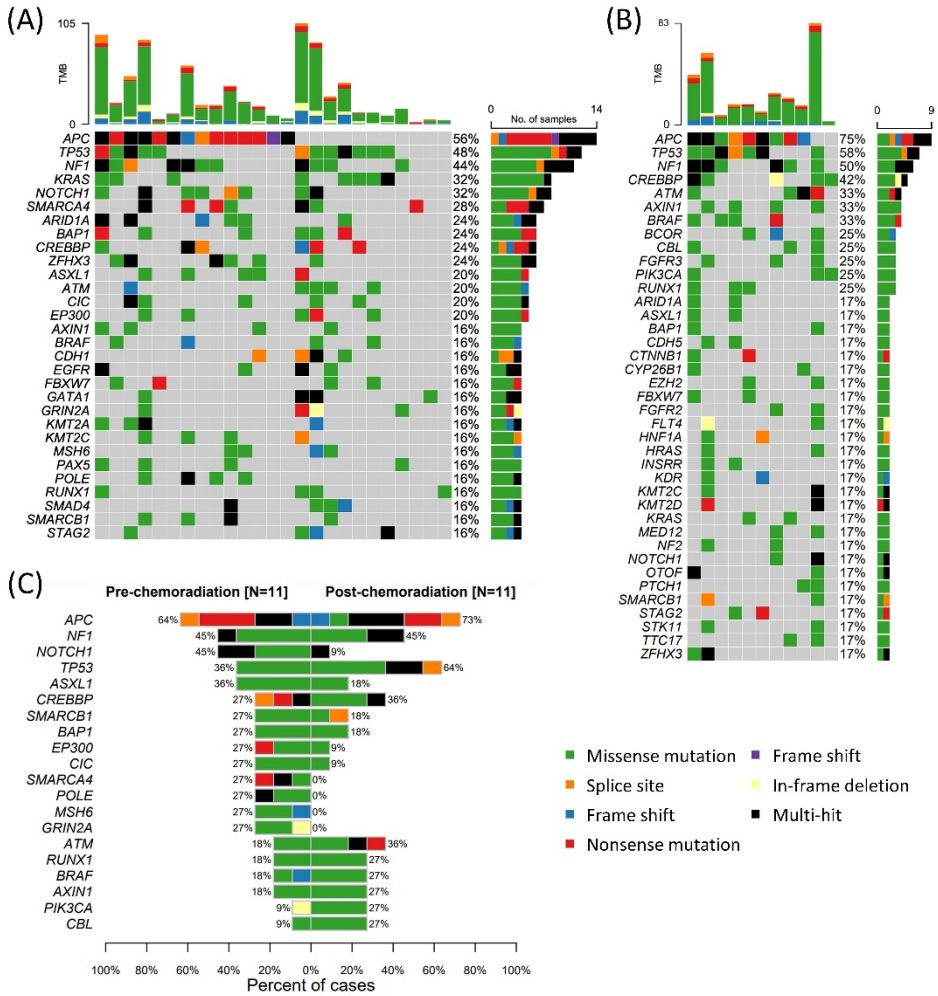
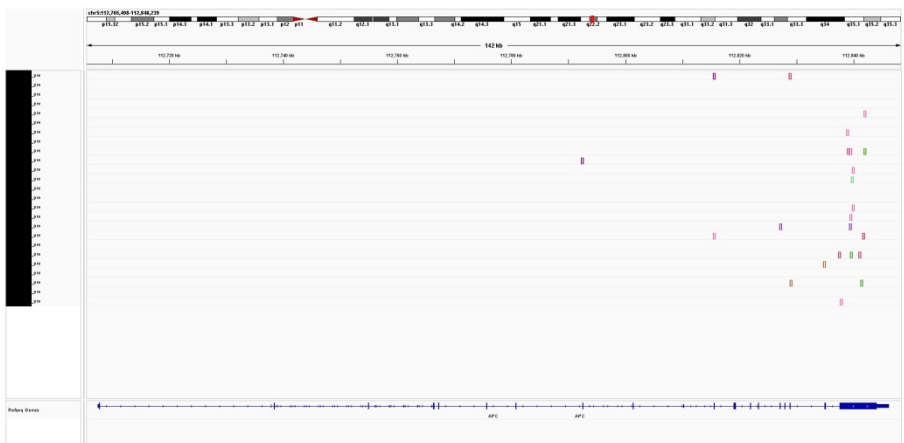


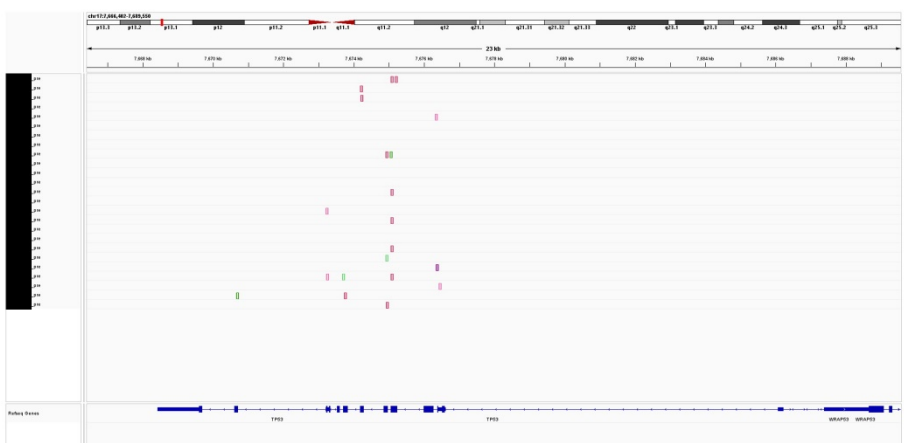
Figure 2. Integrative Genomics Viewer screenshots for the top five frequently mutated genes in (A) pre- and (B) post-chemoradiation samples.

(A) Pre-chemoradiation samples

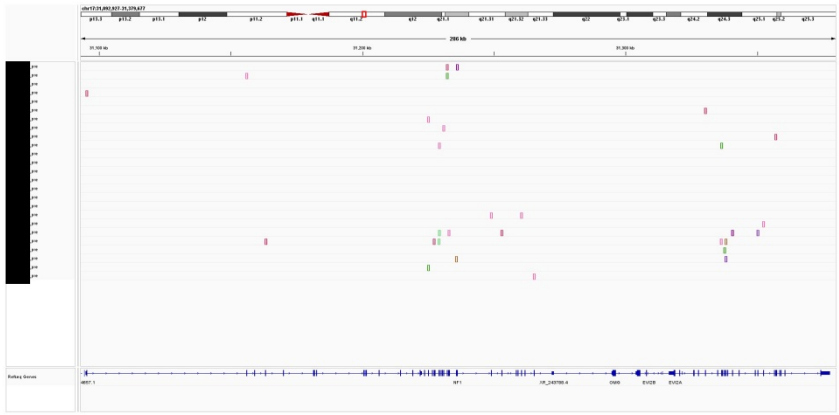
APC



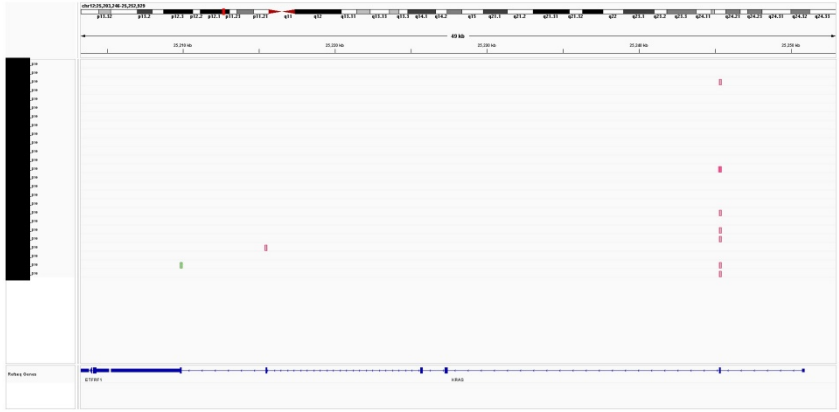
TP53



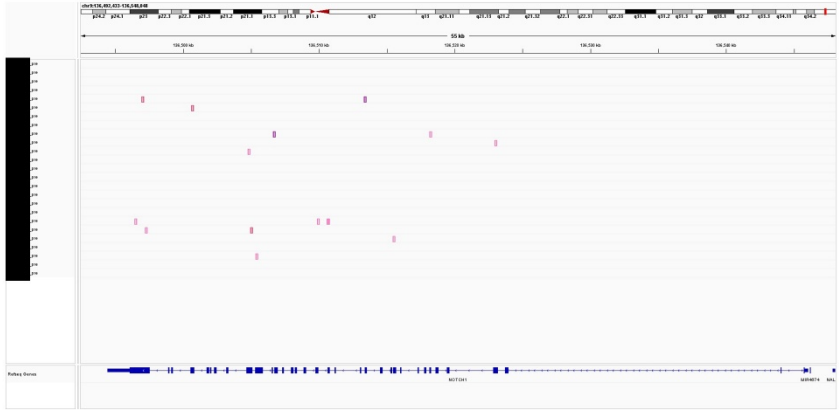
NF1



KRAS

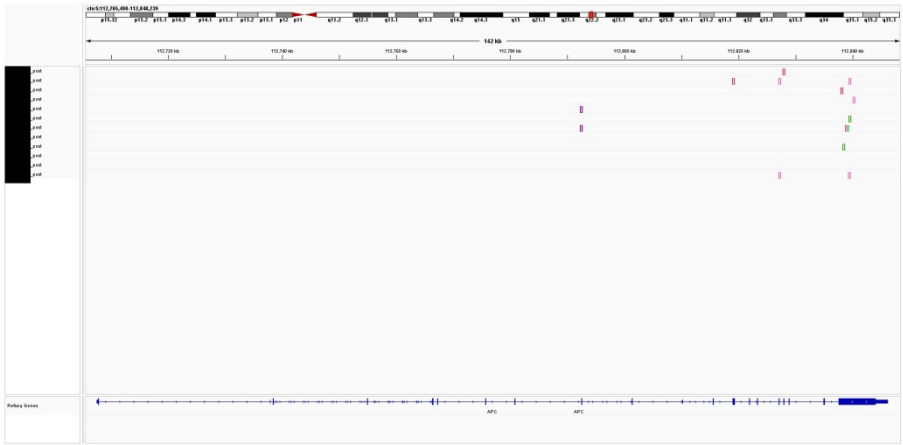


NOTCH1

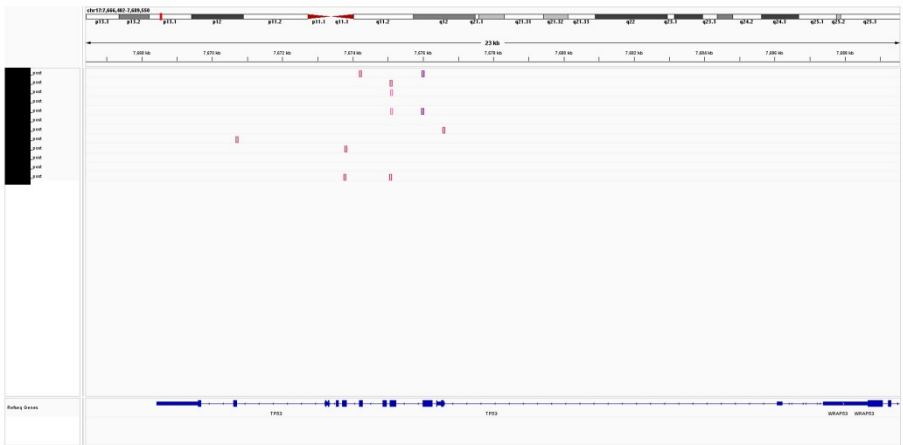


(B) Post-chemoradiation samples

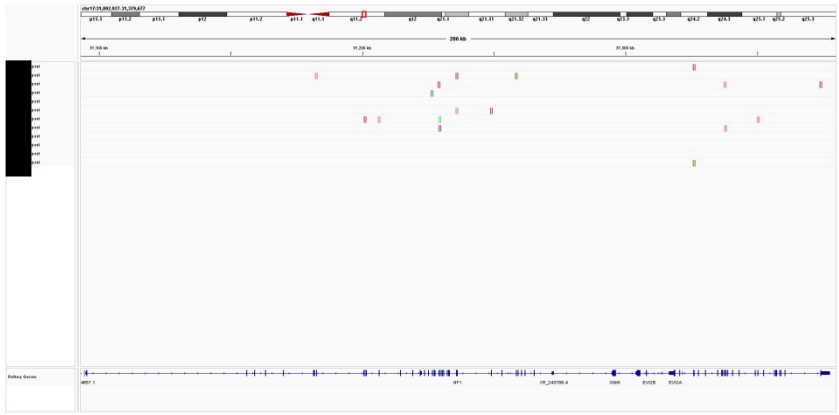
APC



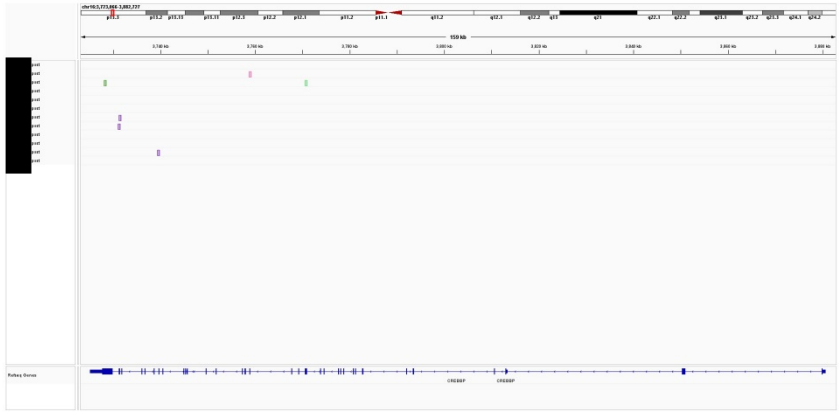
TP53



NF1



CREBBP



ATM

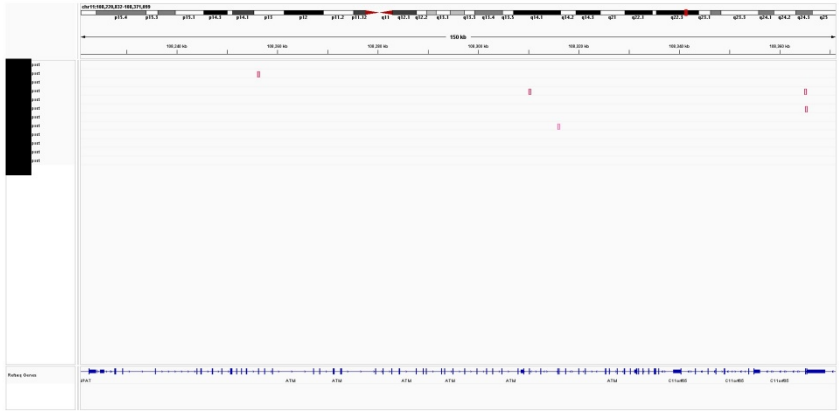


Figure 3. Co-mutation plots for gene mutations detected in (A) pre- and (B) post-chemoradiation samples. Maftools [14] package of R 4.2.0 was used for generating the figure.

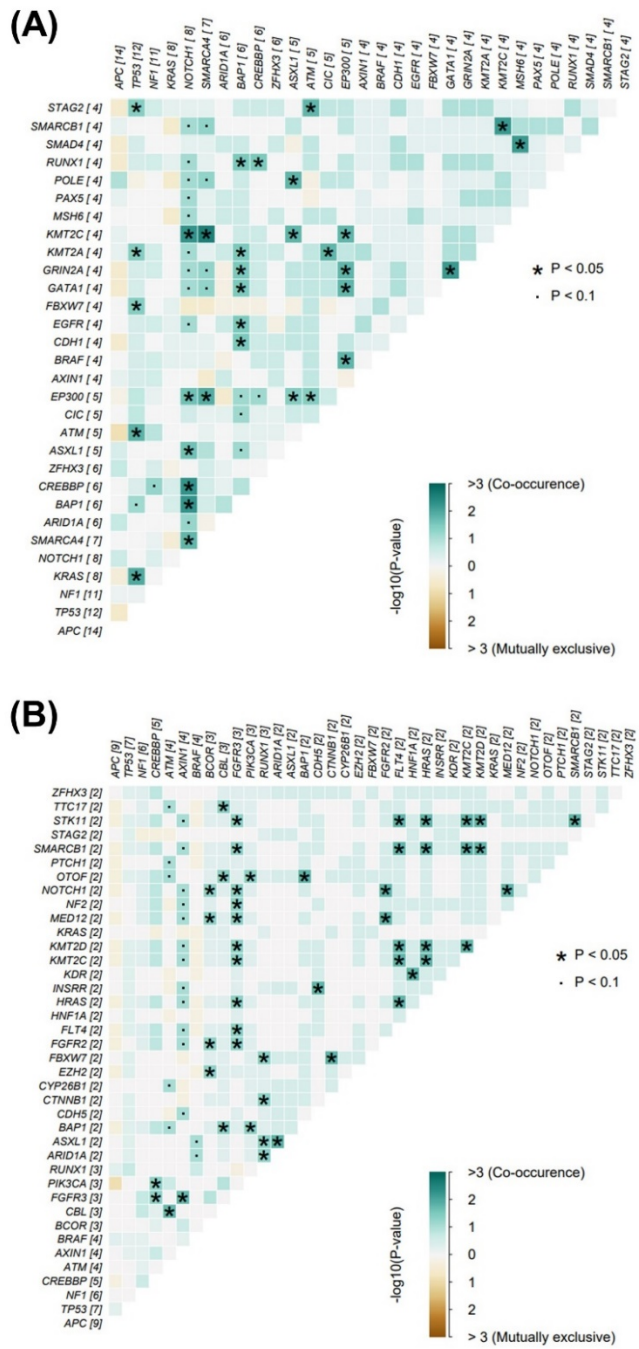
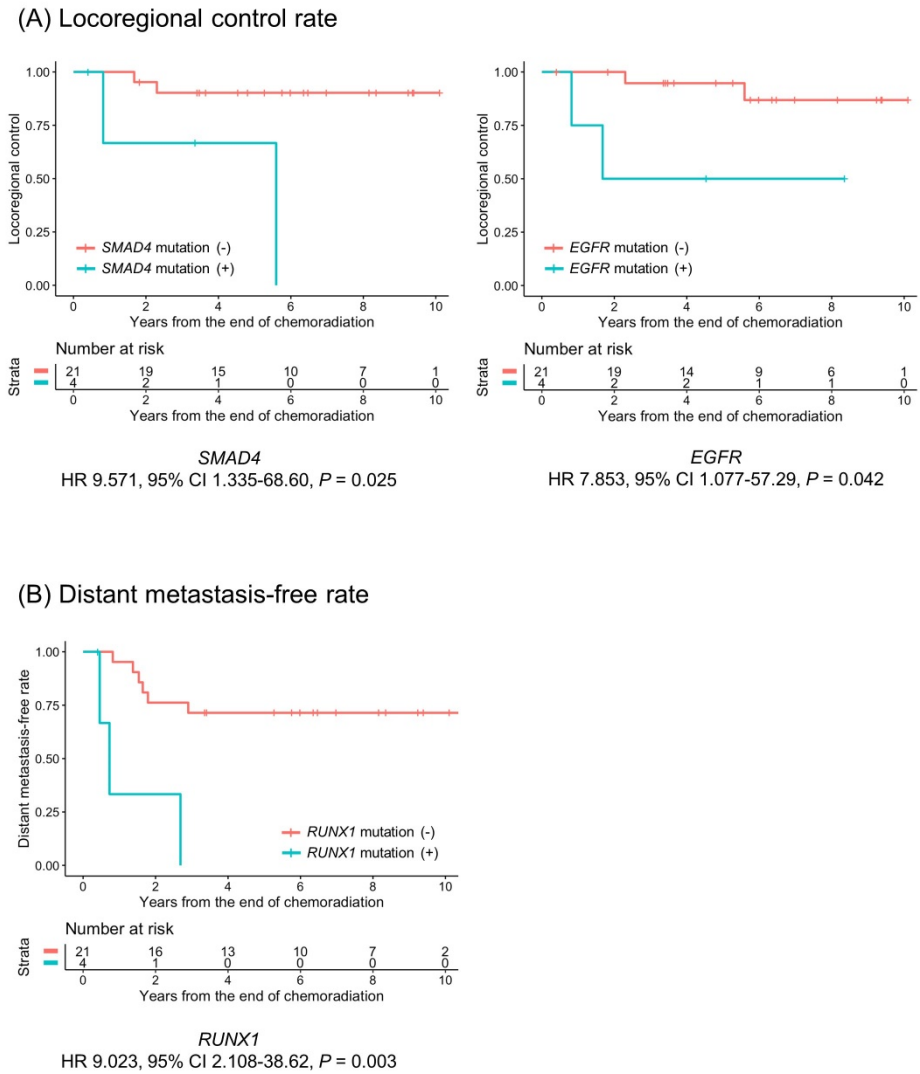


Table 1. Mutations in pre-chemoradiation samples associated with clinical outcomes

	Hazard ratio	95% confidence interval	P-value
Locoregional control rate			
SMAD4	9.571	1.335–68.60	0.025
EGFR	7.853	1.077–57.29	0.042
Distant metastasis-free rate			
RUNX1	9.023	2.108–38.62	0.003
Progression-free survival rate			
RUNX1	8.993	2.457–32.91	0.001
GATA1	5.019	1.402–17.97	0.013
BAP1	3.681	1.119–12.11	0.032
KMT2A	4.045	1.047–15.63	0.043
TP53	3.971	1.046–15.07	0.043
Overall survival rate			
RUNX1	14.64	3.118–68.72	0.001
APC	0.187	0.039–0.907	0.038
EGFR	4.591	1.075–19.61	0.038
STAG2	4.528	1.064–19.27	0.041
TP53	4.878	1.008–23.59	0.049

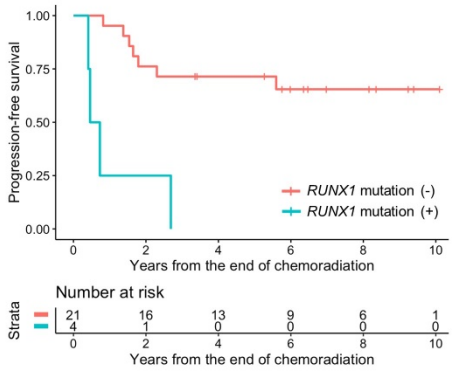
Cox proportional hazards model was used for calculation.

Figure 4. Kaplan-Meier curves of clinical outcomes significantly associated with gene mutations in pre-chemoradiation samples.

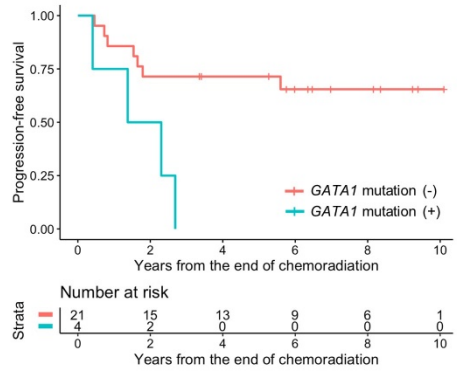


Abbreviations: CI, confidence interval; HR, hazard ratio.

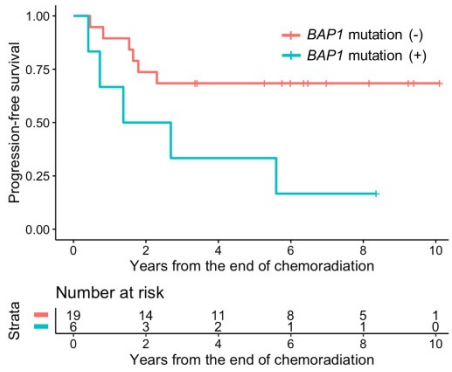
(C) Progression-free survival rate



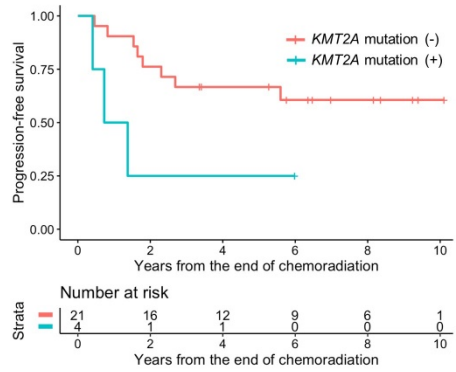
RUNX1
HR 8.993, 95% CI 2.457-32.91, *P* = 0.001



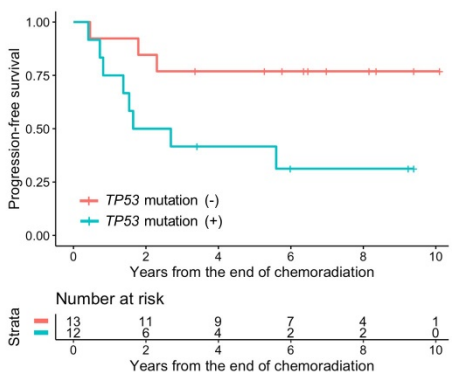
GATA1
HR 5.019, 95% CI 1.402-17.97, *P* = 0.013



BAP1
HR 3.681, 95% CI 1.119-12.11, *P* = 0.032

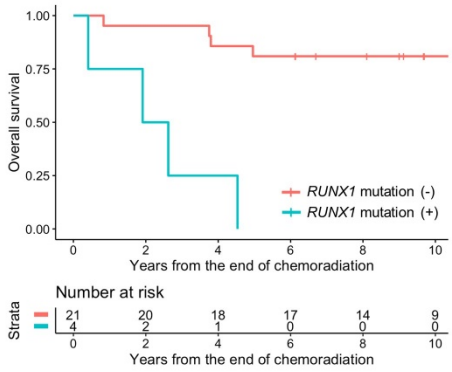


KMT2A
HR 4.045, 95% CI 1.047-15.63, *P* = 0.043

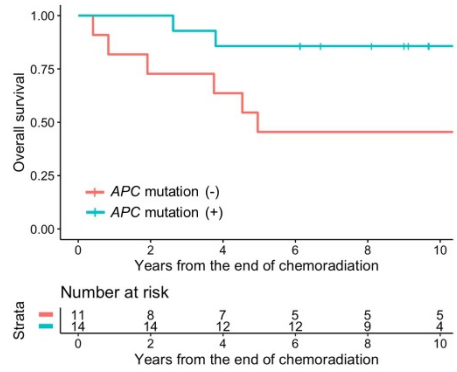


TP53
HR 3.971, 95% CI 1.046-15.07, *P* = 0.043

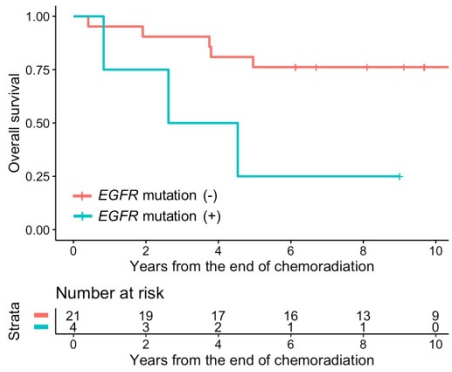
(D) Overall survival rate



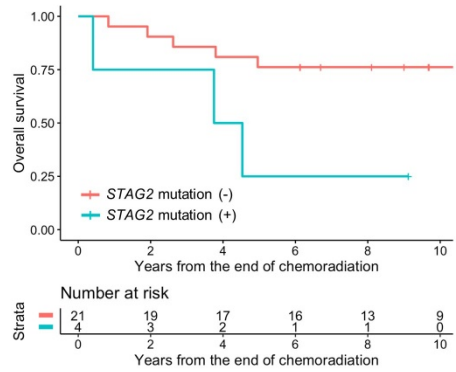
RUNX1
HR 14.64, 95% CI 3.118-68.72, *P* = 0.001



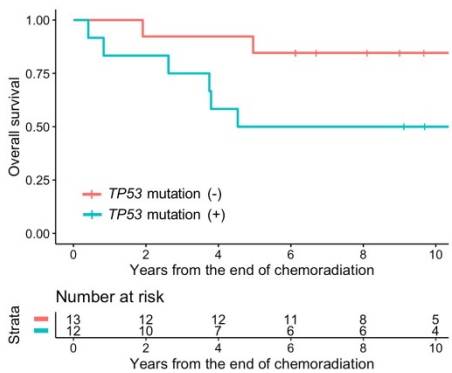
APC
HR 0.187, 95% CI 0.039-0.907, *P* = 0.038



EGFR
HR 4.591, 95% CI 1.075-19.61, *P* = 0.038



STAG2
HR 4.528, 95% CI 1.064-19.27, *P* = 0.041



TP53
HR 4.878, 95% CI 1.008-23.59, *P* = 0.049

3. Comparison of mutations between pre- and post-chemoradiation samples

Comparison of gene mutations between pre- and post-CRT samples was performed for 11 analyzable patients. There were 20 genes that mutated in at least three patients in pre- or post-CRT samples. There were no statistically significant differences in mutation rates of these genes. Comparison of mutation rates of these genes is illustrated in Figure 1C. Concordance of driver mutations such as *APC*, *TP53*, and *KRAS* between pre- and post-CRT samples was examined. Four (36.4%) patients retained *TP53* status. Two (18.2%) patients had *TP53* mutation present in pre-CRT samples but not in post-CRT samples. *TP53* mutation appeared in post-CRT samples of 5 (45.5%) patients, which was not present in pre-CRT samples. *KRAS* status was maintained in 9 (81.8%) patients. Two (18.2%) patients gained *KRAS* mutation in post-CRT samples. Four (36.4%) patients had retained *APC* status. Three (27.3%) patients lost *APC* mutation, whereas four (36.4%) patients gained *APC* mutation.

Number of shared mutations that presented in the same gene in both pre- and post-CRT samples was searched for 11 analyzable patients. It is illustrated in Figure 5. The median rate of shared mutation among pre-CRT samples was 13.8% (range, 0.0%–75.0%). The median rate of shared mutation among post-CRT samples was 18.6% (range, 0.0%–50.0%). There was no significant association between clinical outcomes or TRG and rate of shared mutation. These results are summarized in Table 2 and Figure 6, respectively.

Figure 5. Numbers of shared and non-shared mutations in pre- and post-chemoradiation samples. Each row represents a patient.

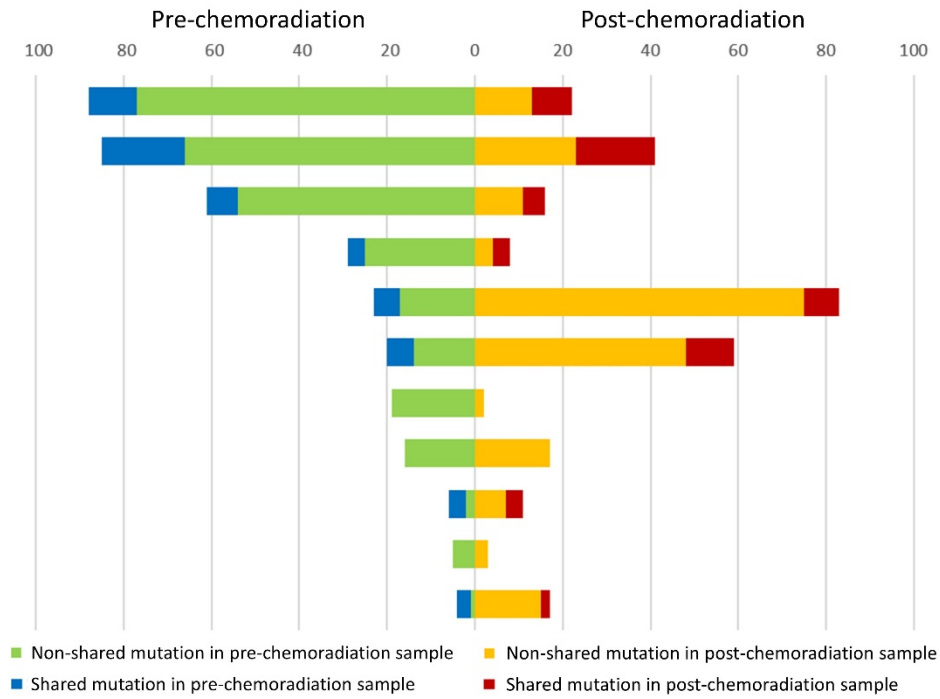
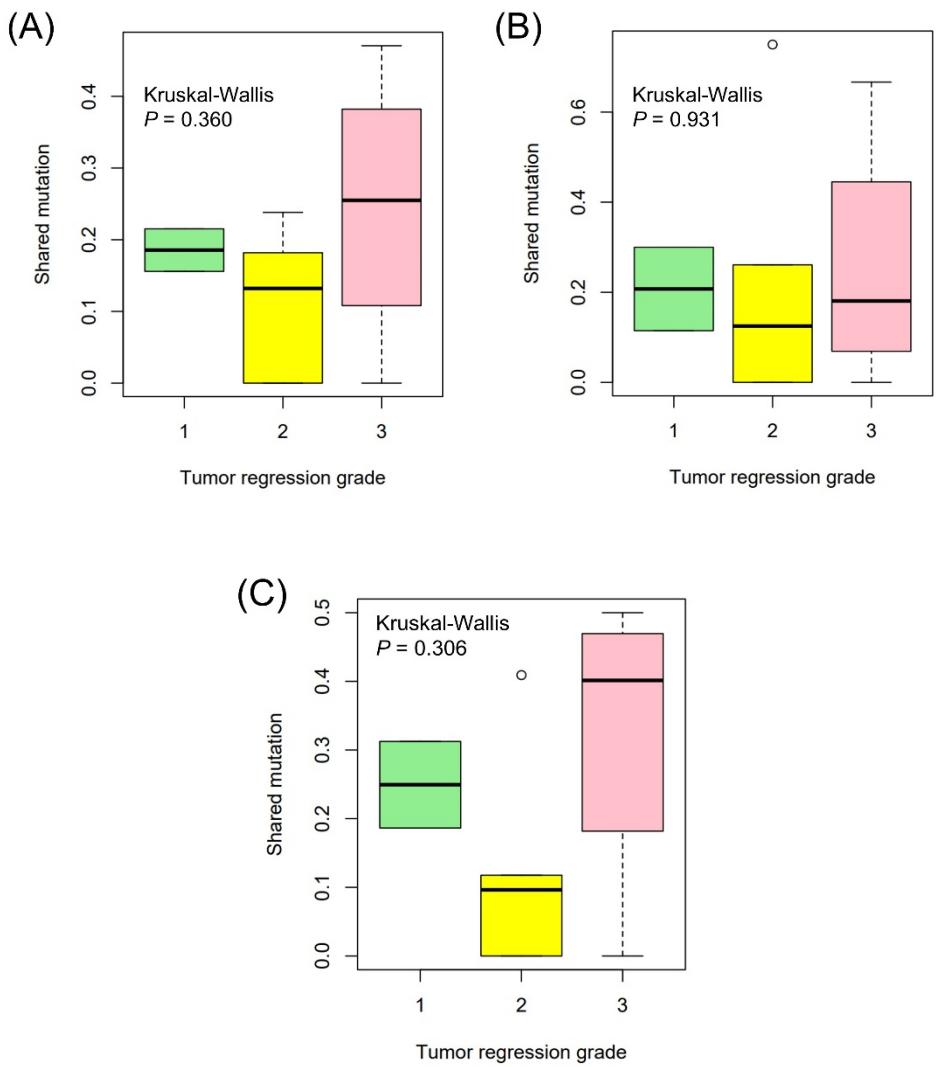


Table 2. Association of percentage of shared mutation with clinical outcomes

Clinical outcomes	Hazard ratio	95% confidence interval	P-value
Percentage of shared mutation in both pre- and post-chemoradiation samples			
Locoregional control rate	1.027	0.903–1.167	0.689
Distant metastasis-free rate	1.004	0.941–1.072	0.893
Progression-free survival rate	1.014	0.959–1.073	0.622
Overall survival rate	1.014	0.950–1.082	0.685
Percentage of shared mutation in pre-chemoradiation samples			
Locoregional control rate	0.980	0.882–1.090	0.713
Distant metastasis-free rate	1.010	0.971–1.050	0.622
Progression-free survival rate	1.007	0.973–1.043	0.679
Overall survival rate	1.008	0.972–1.045	0.669
Percentage of shared mutation in post-chemoradiation samples			
Locoregional control rate	-	-	-
Distant metastasis-free rate	1.028	0.968–1.092	0.362
Progression-free survival rate	1.039	0.982–1.099	0.184
Overall survival rate	1.027	0.968–1.090	0.383

These values were calculated based on Cox proportional hazards model.

Figure 6. Association of tumor regression grade with rate of shared mutation in (A) both pre- and post-chemoradiation samples, (B) pre-chemoradiation samples, and (C) post-chemoradiation samples



4. Mutational signature

Relative compositions of SBS mutational signatures for each patient are illustrated in Figure 7. Six COSMIC SBS signatures were identified: SBS1 (spontaneous deamination of 5-methylcytosine), SBS30 (defective DNA base excision repair due to NTHL1 mutations), SBS29 (tobacco chewing), SBS49 (possible sequencing artifact), SBS3 (defective homologous recombination DNA damage repair), and SBS44 (defective DNA mismatch repair). When mutational signatures of pre- and post-CRT samples for 11 analyzable patients were compared, the relative rate of SBS29 was significantly decreased after CRT from an average of 9.91% to 3.92% (paired Wilcoxon singed rank test p=0.024). When associations of clinical outcomes with pre-CRT mutational signatures were explored, a high SBS30 rate was associated with a worse DMFR (Cox proportional hazards model HR: 1.070 per 1%, 95% CI: 1.014–1.129, p=0.014) and PFSR (HR: 1.065 per 1%, 95% CI: 1.015–1.117, p=0.011). There were no significant associations between clinical outcomes and post-CRT mutational signatures. TRG showed no significant association with pre- or post-CRT mutational signatures. Mutation of MSI gene was not associated with relative rate of SBS44 in pre- or post-CRT samples.

Two COSMIC DBS signatures were identified, namely DBS1 (ultraviolet light exposure) and DBS10 (defective DNA mismatch repair). There were no significant differences in relative rates of DBS signatures between pre- and post-CRT samples. Clinical outcomes and TRG showed no significant associations with pre- or post-CRT DBS signatures. MSI gene mutations and relative rate of DBS10 showed no significant relationship. Two COSMIC indel signatures were identified: ID6 (defective homologous recombination DNA damage repair) and ID2 (slippage

during DNA replication of the template DNA strand). There were no significant differences in relative rates of indel signatures between pre- and post-CRT samples. Indel signatures did not impact clinical outcomes. Pre-CRT indel signatures did not show any significant relationship with TRG. However, in post-CRT samples, lower ID6 composition was observed in TRG 1 (average relative rate: TRG 1, 87.3%; TRG 2, 100.0%; TRG 3, 97.8%; Kruskal-Wallis test p=0.016). Relative compositions of DBS and indel signatures for each patient are illustrated in Figure 8. Association of compositions of mutational signatures and clinical outcomes were summarized in Table 3.

Figure 7. Relative compositions of single-base substitution signatures in pre- and post-chemoradiation samples sorted by rates of SBS1. Each column represents a patient.

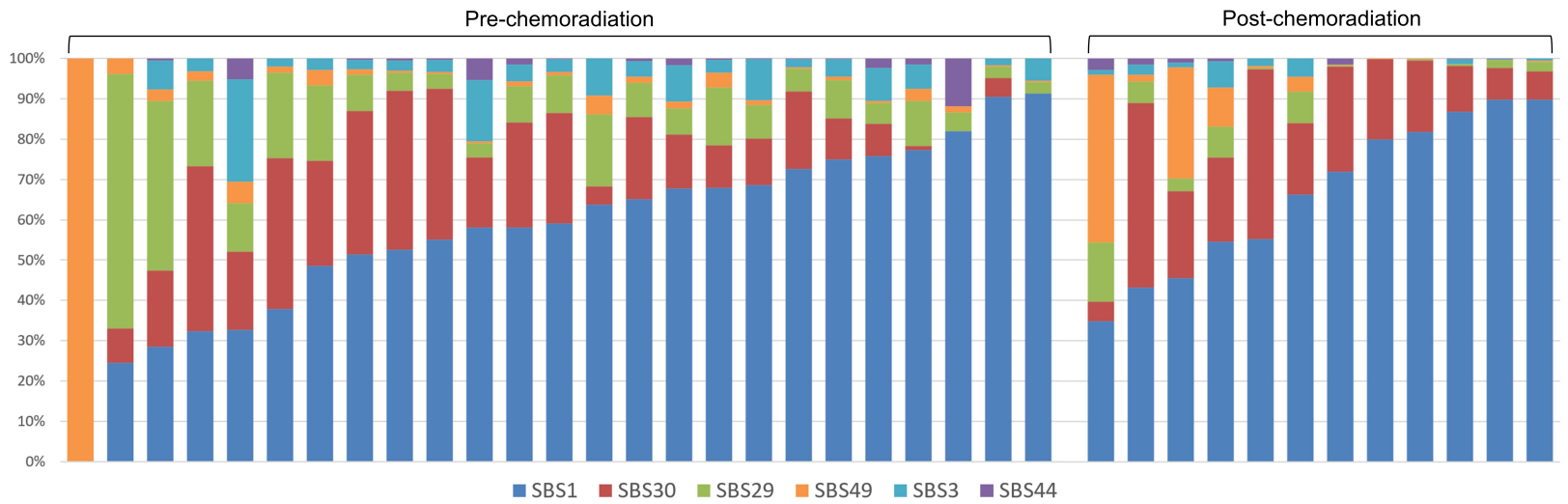


Figure 8. (A) Relative compositions of doublet-base substitution sorted by DBS1, and (B) Relative compositions of indel signatures sorted by ID6 in pre- and post-chemoradiation samples.

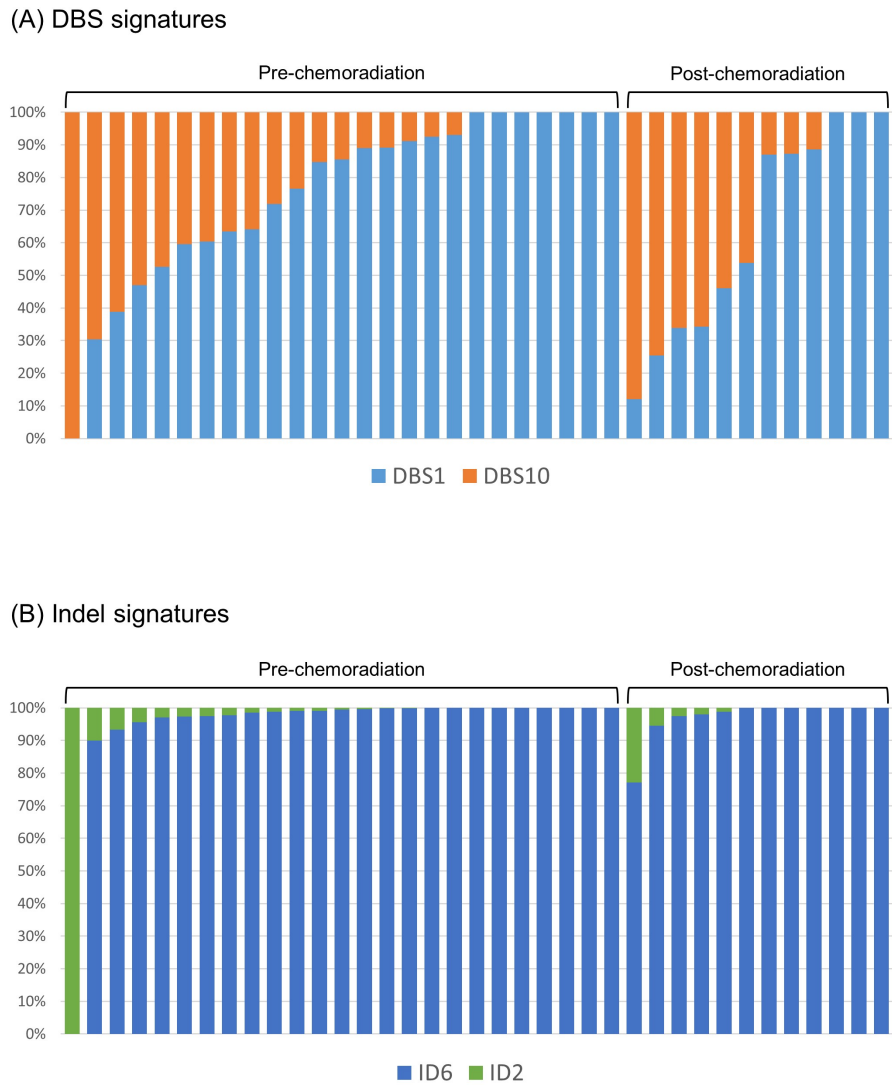


Table 3. Association of clinical outcomes with composition of single-base substitution (SBS), doublet-base substitution (DBS), and indel signitures in pre- and post-chemoradiation samples

Clinical outcomes	Hazard ratio	95% confidence interval	P-value
Pre-chemoradiation			
SBS signiture			
SBS1			
Locoregional control rate	0.987	0.949–1.027	0.525
Distant metastasis-free rate	1.000	0.973–1.030	0.951
Progression-free survival rate	0.996	0.972–1.021	0.754
Overall survival rate	1.003	0.973–1.034	0.831
SBS30			
Locoregional control rate	1.062	0.980–1.151	0.143
Distant metastasis-free rate	1.070	1.014–1.129	0.014
Progression-free survival rate	1.065	1.015–1.117	0.011
Overall survival rate	1.050	0.996–1.108	0.070
SBS29			
Locoregional control rate	1.022	0.969–1.079	0.426
Distant metastasis-free rate	0.983	0.927–1.043	0.574
Progression-free survival rate	1.002	0.962–1.043	0.940
Overall survival rate	0.986	0.934–1.041	0.605
SBS49			
Locoregional control rate	0.833	0.427–1.626	0.592
Distant metastasis-free rate	0.771	0.492–1.210	0.258
Progression-free survival rate	0.800	0.534–1.197	0.277
Overall survival rate	0.619	0.362–1.058	0.080
SBS3			
Locoregional control rate	0.964	0.772–1.203	0.745
Distant metastasis-free rate	0.951	0.822–1.101	0.503
Progression-free survival rate	0.959	0.842–1.092	0.526
Overall survival rate	0.983	0.865–1.117	0.793
DBS signiture			

DBS1			
Locoregional control rate	0.993	0.958–1.028	0.682
Distant metastasis-free rate	1.001	0.977–1.025	0.930
Progression-free survival rate	1.000	0.978–1.023	0.988
Overall survival rate	0.996	0.975–1.019	0.751
Indel signature			
ID6			
Locoregional control rate	2.376	0.483–11.69	0.287
Distant metastasis-free rate	1.081	0.785–1.488	0.633
Progression-free survival rate	1.139	0.810–1.601	0.455
Overall survival rate	1.052	0.824–1.343	0.686

Post-chemoradiation			
SBS signiture			
SBS1			
Locoregional control rate	0.969	0.864–1.086	0.585
Distant metastasis-free rate	0.990	0.941–1.042	0.704
Progression-free survival rate	1.000	0.955–1.048	0.989
Overall survival rate	0.987	0.939–1.039	0.624
SBS30			
Locoregional control rate	1.003	0.862–1.166	0.974
Distant metastasis-free rate	0.985	0.912–1.064	0.695
Progression-free survival rate	0.984	0.917–1.056	0.660
Overall survival rate	0.995	0.922–1.074	0.897
SBS29			
Locoregional control rate	1.154	0.812–1.639	0.425
Distant metastasis-free rate	1.034	0.853–1.254	0.733
Progression-free survival rate	0.996	0.822–1.207	0.969
Overall survival rate	1.045	0.862–1.268	0.652
SBS49			
Locoregional control rate	1.010	0.886–1.153	0.879
Distant metastasis-free rate	1.019	0.960–1.081	0.539
Progression-free survival rate	1.010	0.952–1.071	0.749
Overall survival rate	1.016	0.955–1.080	0.617
SBS3			

Locoregional control rate	-	-	-
Distant metastasis-free rate	1.535	0.974–2.419	0.065
Progression-free survival rate	1.310	0.891–1.925	0.170
Overall survival rate	1.484	0.947–2.324	0.085
DBS signiture			
DBS1			
Locoregional control rate	1.029	0.935–1.132	0.558
Distant metastasis-free rate	1.028	0.984–1.075	0.213
Progression-free survival rate	1.013	0.982–1.044	0.421
Overall survival rate	1.019	0.982–1.058	0.327
Indel signature			
ID6			
Locoregional control rate	0.959	0.770–1.195	0.709
Distant metastasis-free rate	1.059	0.836–1.341	0.636
Progression-free survival rate	1.052	0.860–1.286	0.625
Overall survival rate	1.028	0.852–1.241	0.772

These values were calculated based on Cox proportional hazards model.

5. Tumor mutational burden

The median number of mutations per megabase pair in pre-CRT samples without COSMIC database filtering was 36.5 (range, 5.4–239.4). The median number of mutations per megabase pair in post-CRT samples was 31.9 (range, 7.3–208.1). When all pre- and post-CRT samples were compared, no significant difference in the number of mutations per megabase pair (Wilcoxon rank sum test p=0.713) was observed. When paired pre- and post-CRT analyzable samples were compared within 11 patients, there was no significant difference in the number of mutations per megabase pair (paired Wilcoxon signed rank test p=0.465). Box plots of mutational burden by pre- and post-CRT are illustrated in Figure 9A. When correlations for numbers of mutations per megabase pair between pre- and post-CRT samples were explored in 11 analyzable patients, no statistically significant correlation was found with assumption of a linear relationship between pre- and post-CRT numbers (Pearson's correlation coefficient 0.025, p=0.941). Scatter plots by numbers of mutations per megabase pair between pre- and post-CRT samples are illustrated in Figure 10A.

Tumor mutational burden of pre-CRT samples was associated with clinical outcomes. Higher number of mutations per megabase pair was associated with worse PFSR (HR: 1.008 95% CI: 1.001–1.015, p=0.020). There were no significant relationships between tumor mutational burden of post-CRT samples and clinical outcomes. Associations between tumor mutational burden and clinical outcomes are summarized in Table 4. No significant relationship between TRG and mutational burden was observed. Box plots of mutational burden by TRG are illustrated in Figures 10B and 10C.

Figure 9. (A) Box plots for number of mutations per megabase pair, (B) Box plots for MATH scores in pre- and post-chemoradiation samples.

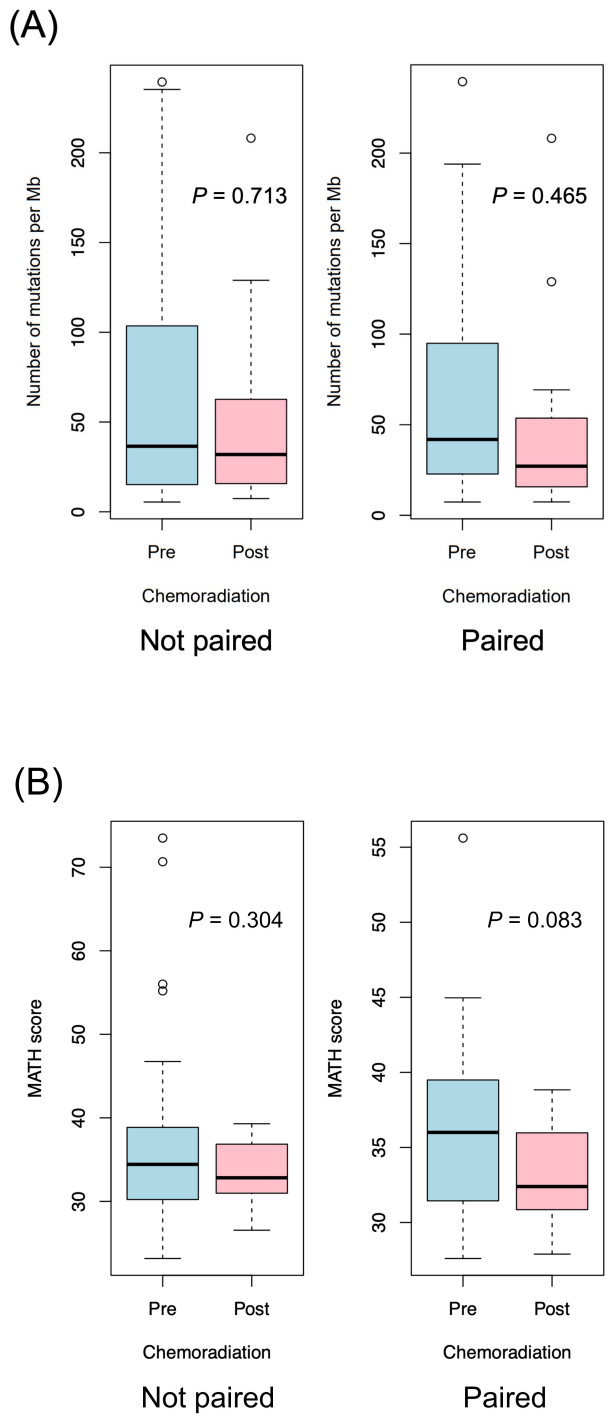


Figure 10. (A) Scatter plots for number of mutations per megabase pair by pre- and post-chemoradiation (CRT) samples, and box plots for number of mutations per megabase pair by tumor regression grade in (B) pre- and (C) post-CRT samples.

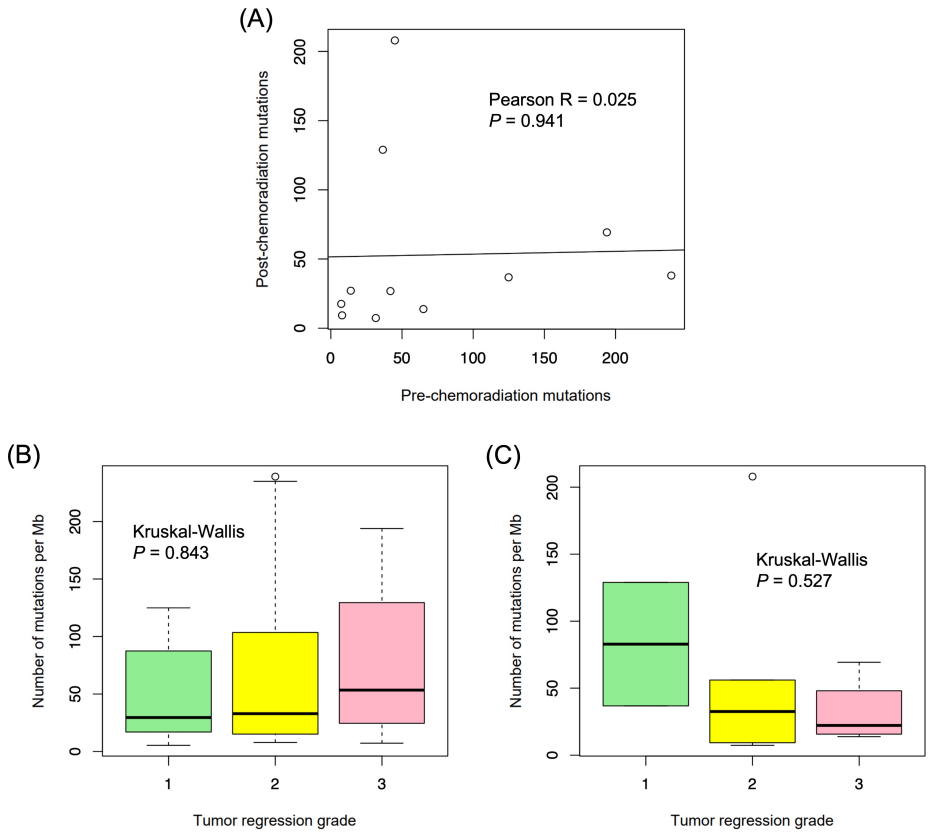


Table 4. Associations of tumor mutational burden with clinical outcomes

	Hazard ratio (per 1 mutation per megabase pair)	95% confidence interval	P-value
Pre-chemoradiation			
Locoregional control rate	1.025	0.984–1.067	0.244
Distant metastasis-free rate	1.025	0.997–1.054	0.081
Progression-free survival rate	1.029	1.004–1.054	0.024
Overall survival rate	1.019	0.992–1.047	0.180
Post-chemoradiation			
Locoregional control rate	0.759	0.359–1.606	0.471
Distant metastasis-free rate	0.931	0.799–1.084	0.357
Progression-free survival rate	0.969	0.900–1.044	0.413
Overall survival rate	0.964	0.876–1.060	0.446

Cox proportional hazards model was used for calculation.

Abbreviation: SNV, single nucleotide variation.

6. MATH score

Tumor heterogeneity is quantified by MATH score for each sample without joint analysis of two samples from the same biopsy or surgical specimens. For pre-CRT samples (N=50), the median MATH score was 34.43 (range, 23.17–73.50). For post-CRT samples (N=24), it was 32.83 (range, 26.55–39.60). When all pre- and post-CRT samples were compared, no difference in MATH score was observed (Wilcoxon rank sum test p=0.304). Paired match was carried out for 11 patients with both sequenced pre- and post-CRT samples. Average MATH score of two tumor samples for each pre- and post- CRT per patient was used for this analysis. MATH score decreased with marginal significance after CRT (average: 37.06 vs. 32.90, paired Wilcoxon signed rank test p=0.083). Box plots of MATH score by pre- and post-CRT are illustrated in Figure 9B. No relationship between pre- and post-CRT MATH scores was observed (p=0.686). No significant associations between clinical outcomes or TRG and MATH score of pre- or post-CRT samples were found. Box plots by TRG and scatter plots by pre- and post-CRT of MATH score are illustrated in Figure 11. No significant association between clinical outcomes or TRG and MATH score of pre- or post-CRT samples was found, and this result was summarized in Table 5.

Figure 11. (A) Box plots for Mutant-Allele Tumor Heterogeneity (MATH) score by tumor regression grade, (B) Scatter plots for MATH score by pre- and post-chemoradiation samples.

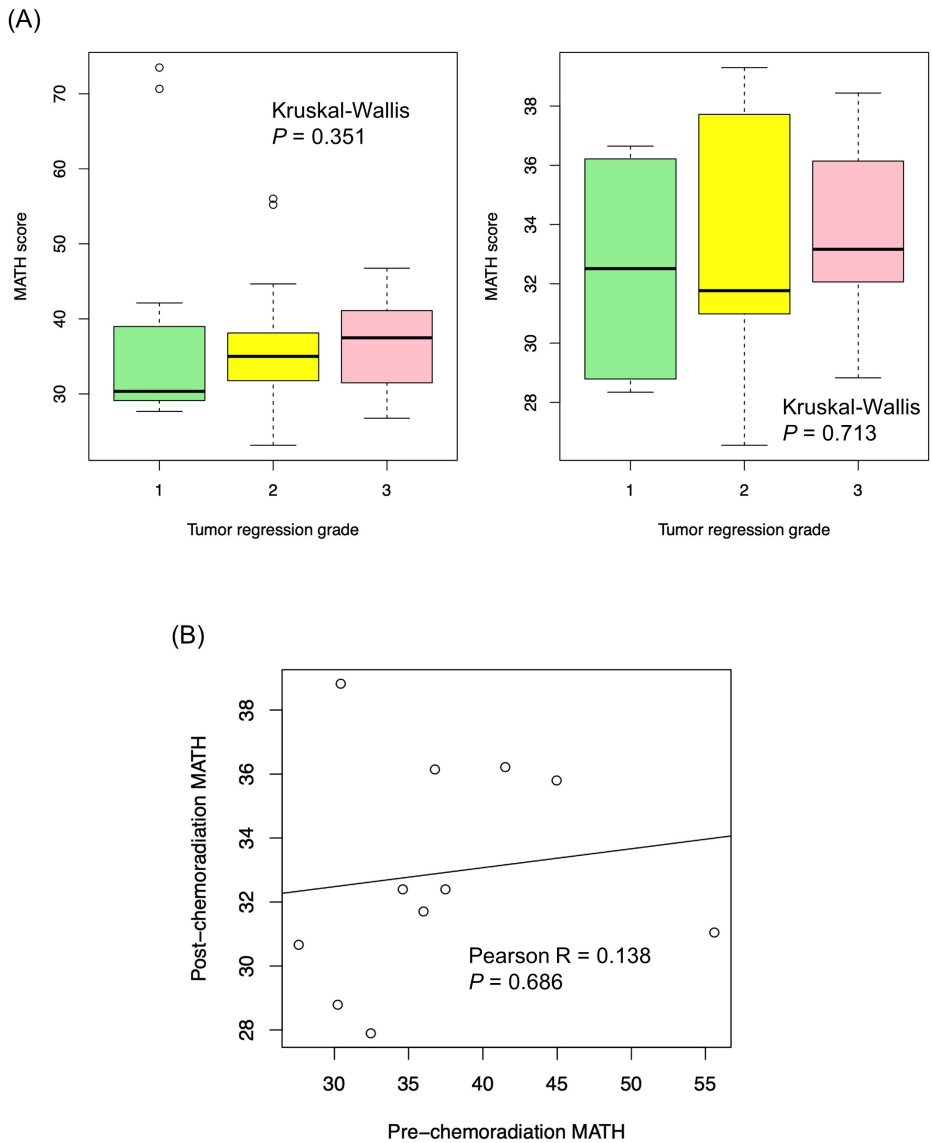


Table 5. Association of Mutant-Allele Tumor Heterogeneity (MATH) score with clinical outcomes

Clinical outcomes	Hazard ratio	95% confidence interval	P-value
Pre-chemoradiation			
Locoregional control rate	1.012	0.946-1.082	0.733
Distant metastasis-free rate	1.014	0.977-1.053	0.459
Progression-free survival rate	1.014	0.979-1.051	0.431
Overall survival rate	1.025	0.989-1.063	0.174
Post-chemoradiation			
Locoregional control rate	1.213	0.792-1.858	0.375
Distant metastasis-free rate	0.956	0.802-1.140	0.620
Progression-free survival rate	0.952	0.811-1.118	0.548
Overall survival rate	0.961	0.799-1.156	0.676

These values were calculated based on Cox proportional hazards model.

7. Copy number alteration

CNA analysis was performed for 34 pre-CRT samples of 17 patients and 24 post-CRT samples of 12 patients as normal pair was required for Sequenza analysis. Estimated sample purity and ploidy with Sequenza is summarized in Table 6. CNAs frequently found in pre- and post-CRT tumor samples are illustrated in Figure 12. The full list of these CNAs is summarized in Table 7. Seven amplification peaks and 21 deletion peaks were found to be significant in pre-CRT samples. In post-CRT samples, 5 amplification peaks and 4 deletion peaks were significantly altered. Among 9 CNAs in post-CRT samples, 4 CNAs overlapped with CNA peaks in pre-CRT samples. The association of CNAs appearing in at least 4 patients with clinical outcomes or TRG was explored. CNAs appearing in post-CRT samples with a minimum of 3 patients were analyzed in the same way. The results were summarized in Table 8 and Table 9. Some types of CNAs showed a significant association with clinical outcomes. For pre-CRT samples, 6p21.1 amplification, 4p11 deletion, 6p21.32 deletion, 14q11.2 deletion, and 14q32.33 deletion were associated with significantly worse prognosis, and 12q12 amplification and 20q11.1 deletion were associated with better prognosis. For post-CRT samples, 4p11 deletion was associated with better prognosis. Some CNAs showed a significant relationship with TRG. 6p21.1 amplification, 5q11.2 deletion, and 6p21.32 deletion were related to higher TRG in pre-CRT samples, and 6p21.1 and 11q13.3 amplification were related to higher TRG in post-CRT samples.

The median CNA burden of pre-CRT samples was 25.0% (range: 3.3%–78.1%), and for post-CRT samples, it was 51.5% (range: 0.2%–77.5%). No difference in CNA burden was observed between pre- and post-CRT samples

(Wilcoxon rank sum test, p=0.867). A paired match was carried out for 9 patients with both analyzable pre- and post-CRT CNAs. The average CNA burden of two samples for each pre- and post-CRT per patient was used for this analysis, and no difference was observed in the CNA burden of paired pre- and post-CRT samples (p=0.250). There was a significant association between pre-CRT CNA burdens and post-CRT CNA burdens (Pearson's correlation coefficient 0.776, p=0.014). These results were illustrated in Figure 13A and B. No statistically significant association was found between CNA burdens and clinical outcomes in both pre- and post-CRT samples, as summarized in Table 10. A significant association between CNA burdens of pre- and post-CRT samples with TRG was observed, with higher CNA burden being associated with worse TRG in both pre- and post-CRT samples (pre-CRT, p=0.040; post-CRT, p=0.004). Box plots for this analysis were illustrated in Figure 13C.

Table 6. Estimated sample purity and ploidy

Patient Number	Sample Number	Estimated Sample Purity	Estimated Ploidy
Pre-chemoradiation			
1	1	0.40	1.8
1	2	0.51	1.9
2	1	0.28	6.4
2	2	0.25	2.4
5	1	0.41	1.8
5	2	0.43	1.8
8	1	0.66	2.0
8	2	0.65	2.0
12	1	0.10	1.3
12	2	0.09	1.2
14	1	0.26	6.1
14	2	0.25	6.4
15	1	0.17	1.8
15	2	0.12	2.0
16	1	0.09	2.3
16	2	0.05	2.4
17	1	0.22	2.0
17	2	0.14	2.1
18	1	0.27	5.0
18	2	0.30	4.8
19	1	0.18	2.3
19	2	0.12	2.8
21	1	0.15	3.1
21	2	0.15	3.4
22	1	0.23	4.8
22	2	0.25	4.9
23	1	0.22	5.9

23	2	0.23	5.7
24	1	0.30	2.1
24	2	0.27	2.1
25	1	0.42	2.1
25	2	0.42	2.2
26	1	0.38	2.8
26	2	0.39	2.8
Post-chemoradiation			
7	1	0.37	2.1
7	2	0.38	2.4
10	1	0.96	1.9
10	2	0.39	1.7
14	1	0.51	2.9
14	2	0.52	2.9
15	1	0.08	2.1
15	2	0.75	2.0
16	1	0.06	1.0
16	2	0.07	1.4
18	1	0.08	2.3
18	2	0.07	2.4
20	1	0.13	2.1
20	2	0.16	2.2
22	1	0.15	4.7
22	2	0.14	4.5
23	1	0.13	1.1
23	2	0.13	1.3
24	1	0.19	1.9
24	2	0.07	1.6
25	1	0.44	2.0
25	2	0.45	2.0
26	1	0.19	3.6
26	2	0.18	3.9

Figure 12. Copy number alterations frequently found in (A) pre- and (B) post-chemoradiation tumor samples. Maftools [14] package of R 4.2.0 was used for generating the figure.

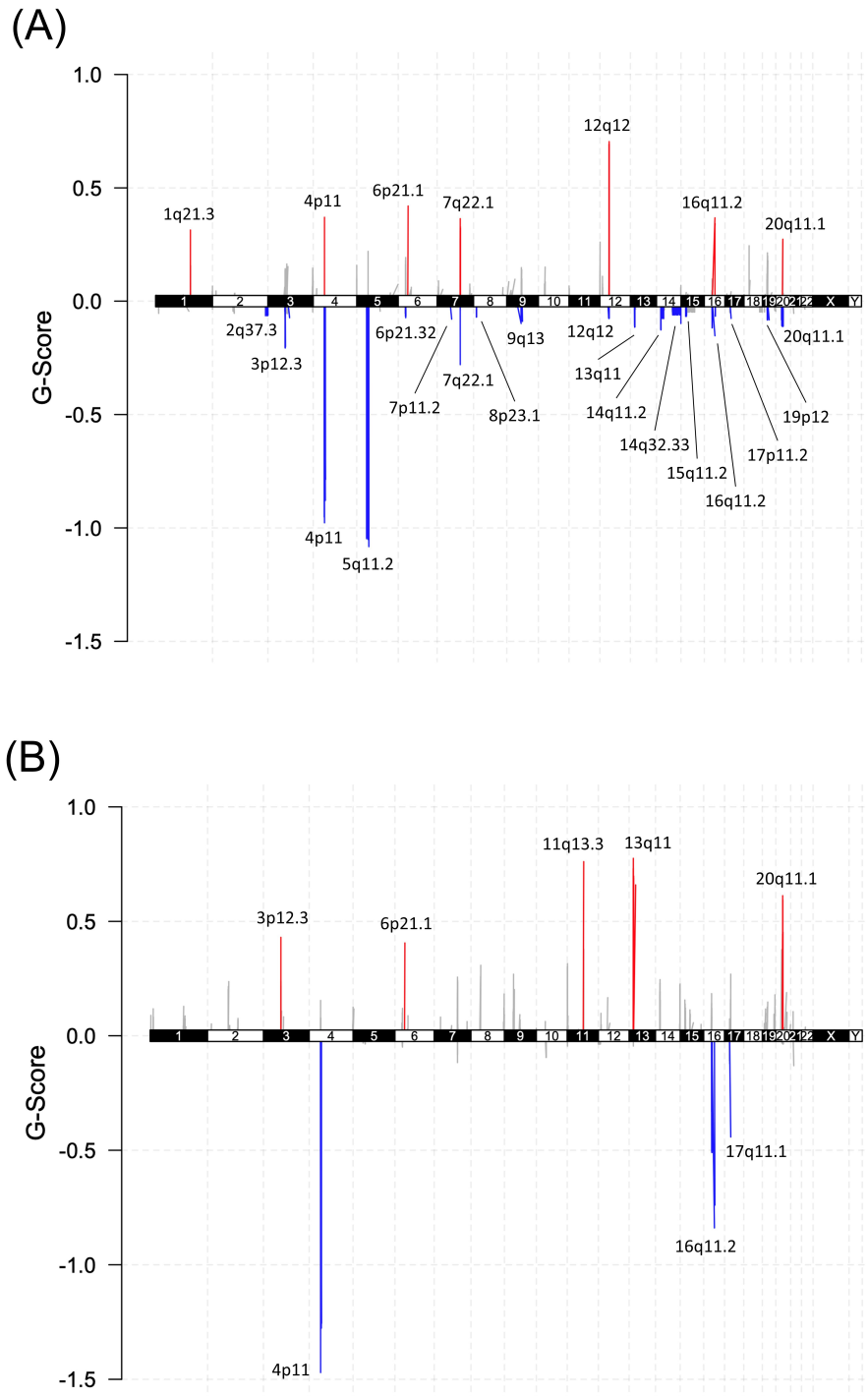


Table 7. List of GISTIC2 amplification and deletion peaks in pre- and post-chemoradiation samples

Type	Descriptor	Wide Peak Limits	Q-value
Pre-chemoradiation			
Amplification	1q21.3	chr1:152214708-152358763	0.014
Amplification	4p11	chr4:48561089-52917612	0.002
Amplification	6p21.1	chr6:41584796-43443364	0.001
Amplification	7q22.1	chr7:100954062-100956479	0.003
Amplification	12q12	chr12:39854856-40489822	< 0.001
Amplification	16q11.2	chr16:34158851-46401173	0.002
Amplification	20q11.1	chr20:26071629-30815511	0.055
Deletion	2q37.3	chr2:231493530-242193529	0.097
Deletion	3p12.3	chr3:75672123-75737023	< 0.001
Deletion	4p11	chr4:1-190214555	< 0.001
Deletion	4p11	chr4:48551167-52907613	< 0.001
Deletion	5q11.2	chr5:49651846-52871266	< 0.001
Deletion	6p21.32	chr6:32589734-32741592	0.093
Deletion	7p11.2	chr7:56077388-63629923	0.085
Deletion	7q22.1	chr7:100810077-100969663	< 0.001
Deletion	8p23.1	chr8:11769691-12455364	0.097
Deletion	9q13	chr9:42784720-62801650	0.043
Deletion	12q12	chr12:1-133275309	0.085
Deletion	12q12	chr12:34209511-40488161	0.085
Deletion	13q11	chr13:18230172-19178748	0.027
Deletion	14q11.2	chr14:1-18994699	0.010
Deletion	14q32.33	chr14:106284978-107043718	0.046
Deletion	15q11.2	chr15:1-23575742	0.097

Deletion	16q11.2	chr16:34158851-48200415	0.001
Deletion	17p11.2	chr17:21545188-27560503	0.085
Deletion	19p12	chr19:1-58617616	0.085
Deletion	19p12	chr19:21163934-29613828	0.085
Deletion	20q11.1	chr20:26061756-32216519	0.027

Post-chemoradiation			
Amplification	3p12.3	chr3:75665273-75738424	0.023
Amplification	6p21.1	chr6:41584802-43350622	0.045
Amplification	11q13.3	chr11:69077921-71557391	0.001
Amplification	13q11	chr13:1-18273528	0.001
Amplification	20q11.1	chr20:29678346-29800133	0.005
Deletion	4p11	chr4:48906358-52597557	< 0.001
Deletion	16q11.2	chr16:35473236-48200415	< 0.001
Deletion	17q11.1	chr17:1-83257441	< 0.001
Deletion	17q11.1	chr17:21703502-27301793	< 0.001

Table 8. Association of copy number alteration with clinical outcomes

CNA and clinical outcomes	Events in samples without CNA	Events in samples with CNA	Hazard ratio	95% confidence interval	*P*-value
Pre-chemoradiation samples					
Amplification					
6p21.1					
LRCR	2/26 (7.7%)	2/8 (25.0%)	4.707	0.648-34.19	0.126
DMFR	8/26 (30.8%)	6/8 (75.0%)	5.259	1.783-15.51	0.003
PFSR	8/26 (30.8%)	6/8 (75.0%)	5.259	1.783-15.51	0.003
OSR	4/26 (15.4%)	6/8 (75.0%)	11.01	2.991-40.54	< 0.001
7q22.1					
LRCR	2/26 (7.7%)	2/8 (25.0%)	3.061	0.431-21.75	0.263
DMFR	8/26 (30.8%)	6/8 (75.0%)	2.417	0.834-7.006	0.104
PFSR	8/26 (30.8%)	6/8 (75.0%)	2.417	0.834-7.006	0.104
OSR	6/26 (23.1%)	4/8 (50.0%)	2.161	0.609-7.671	0.233
12q12					
LRCR	4/22 (18.2%)	0/12 (0.0%)	-	-	0.114*
DMFR	14/22 (63.6%)	0/12 (0.0%)	-	-	< 0.001*
PFSR	14/22 (63.6%)	0/12 (0.0%)	-	-	< 0.001*
OSR	10/22 (45.5%)	0/12 (0.0%)	-	-	0.008*
20q11.1					
LRCR	3/16 (18.8%)	1/18 (5.6%)	0.308	0.032-2.967	0.308
DMFR	5/16 (31.3%)	9/18 (50.0%)	2.067	0.691-6.185	0.194
PFSR	5/16 (31.3%)	9/18 (50.0%)	2.067	0.691-6.185	0.194
OSR	5/16 (31.3%)	5/18 (27.8%)	0.989	0.283-3.382	0.972
Deletion					
3p12.3					
LRCR	3/21 (14.3%)	1/13 (7.7%)	0.479	0.050-4.610	0.524
DMFR	11/21 (52.4%)	3/13 (23.1%)	0.327	0.091-1.175	0.087
PFSR	11/21 (52.4%)	3/13 (23.1%)	0.327	0.091-1.175	0.087
OSR	7/21 (33.3%)	3/13 (23.1%)	0.583	0.151-2.256	0.434
4p11					
LRCR	0/6 (0.0%)	4/28 (14.3%)	-	-	0.327*
DMFR	0/6 (0.0%)	14/28 (50.0%)	-	-	0.043*
PFSR	0/6 (0.0%)	14/28 (50.0%)	-	-	0.043*
OSR	0/6 (0.0%)	10/28 (35.7%)	-	-	0.105*
5q11.2					
LRCR	4/27 (14.8%)	0/7 (0.0%)	-	-	0.280*
DMFR	13/27 (48.1%)	1/7 (14.3%)	0.227	0.030-1.741	0.154

PFSR	13/27 (48.1%)	1/7 (14.3%)	0.227	0.030-1.741	0.154
OSR	10/27 (37.0%)	0/7 (0.0%)	-	-	0.074*
6p21.32					
LRCR	1/24 (4.2%)	3/10 (30.0%)	8.004	0.832-77.02	0.072
DMFR	6/24 (25.0%)	8/10 (80.0%)	5.090	1.722-15.05	0.003
PFSR	6/24 (25.0%)	8/10 (80.0%)	5.090	1.722-15.05	0.003
OSR	3/24 (12.5%)	7/10 (70.0%)	8.262	2.107-32.40	0.002
7q22.1					
LRCR	3/23 (13.0%)	1/11 (9.1%)	0.732	0.076-7.035	0.787
DMFR	12/23 (52.2%)	2/11 (18.2%)	0.314	0.070-1.406	0.130
PFSR	12/23 (52.2%)	2/11 (18.2%)	0.314	0.070-1.406	0.130
OSR	8/23 (24.8%)	2/11 (18.2%)	0.523	0.111-2.465	0.413
8p23.1					
LRCR	2/13 (15.4%)	2/21 (9.5%)	0.595	0.084-4.224	0.603
DMFR	6/13 (46.2%)	8/21 (38.1%)	0.744	0.258-2.147	0.584
PFSR	6/13 (46.2%)	8/21 (38.1%)	0.744	0.258-2.147	0.584
OSR	2/13 (15.4%)	8/21 (38.1%)	2.515	0.533-11.87	0.244
13q11					
LRCR	2/25 (8.0%)	2/9 (22.2%)	3.077	0.433-21.86	0.261
DMFR	10/25 (40.0%)	4/9 (44.4%)	1.219	0.382-3.891	0.739
PFSR	10/25 (40.0%)	4/9 (44.4%)	1.219	0.382-3.891	0.739
OSR	6/25 (24.0%)	4/9 (44.4%)	2.249	0.633-7.986	0.210
14q11.2					
LRCR	0/18 (0.0%)	4/16 (25.0%)	-	-	0.030*
DMFR	8/18 (44.4%)	6/16 (37.5%)	0.803	0.278-2.318	0.684
PFSR	8/18 (44.4%)	6/16 (37.5%)	0.803	0.278-2.318	0.684
OSR	6/18 (33.3%)	4/16 (25.0%)	0.710	0.200-2.521	0.596
14q32.33					
LRCR	0/22 (0.0%)	4/12 (33.3%)	-	-	0.005*
DMFR	6/22 (27.3%)	8/12 (66.7%)	3.060	1.056-8.864	0.039
PFSR	6/22 (27.3%)	8/12 (66.7%)	3.060	1.056-8.864	0.039
OSR	4/22 (18.2%)	6/12 (50.0%)	3.085	0.869-10.95	0.081
15q11.2					
LRCR	2/22 (9.1%)	2/12 (16.7%)	1.675	0.236-11.91	0.606
DMFR	10/22 (45.5%)	4/12 (33.3%)	0.591	0.185-1.886	0.374
PFSR	10/22 (45.5%)	4/12 (33.3%)	0.591	0.185-1.886	0.374
OSR	6/22 (27.3%)	4/12 (33.3%)	1.121	0.316-3.974	0.860
17p11.2					
LRCR	0/12 (0.0%)	4/22 (18.2%)	-	-	0.142*
DMFR	4/12 (33.3%)	10/22 (45.5%)	1.419	0.444-4.532	0.555
PFSR	4/12 (33.3%)	10/22 (45.5%)	1.419	0.444-4.532	0.555
OSR	2/12 (16.7%)	8/22 (36.4%)	2.289	0.486-10.79	0.295

20q11.1					
LRCR	4/26 (15.4%)	0/8 (0.0%)	-	-	0.239*
DMFR	14/26 (53.8%)	0/8 (0.0%)	-	-	0.014*
PFSR	14/26 (53.8%)	0/8 (0.0%)	-	-	0.014*
OSR	10/26 (38.5%)	0/8 (0.0%)	-	-	0.050*

Post-chemoradiation samples					
Amplification					
3p12.3					
LRCR	0/14 (0.0%)	2/10 (20.0%)	-	-	0.112*
DMFR	4/14 (28.6%)	4/10 (40.0%)	1.320	0.330-5.285	0.695
PFSR	6/14 (42.9%)	4/10 (40.0%)	0.883	0.249-3.133	0.848
OSR	4/14 (28.6%)	4/10 (40.0%)	1.425	0.356-5.707	0.617
6p21.1					
LRCR	0/18 (0.0%)	2/6 (33.3%)	-	-	0.018*
DMFR	6/18 (33.3%)	2/6 (33.3%)	0.939	0.189-4.662	0.939
PFSR	8/18 (44.4%)	2/6 (33.3%)	0.709	0.150-3.342	0.663
OSR	6/18 (33.3%)	2/6 (33.3%)	1.078	0.217-5.355	0.927
11q13.3					
LRCR	2/19 (10.5%)	0/5 (0.0%)	-	-	0.432*
DMFR	8/19 (42.1%)	0/5 (0.0%)	-	-	0.078*
PFSR	10/19 (52.6%)	0/5 (0.0%)	-	-	0.056*
OSR	8/19 (42.1%)	0/5 (0.0%)	-	-	0.101*
13q11					
LRCR	2/13 (15.4%)	0/11 (0.0%)	-	-	0.147*
DMFR	5/13 (38.5%)	3/11 (27.3%)	0.439	0.104-1.847	0.262
PFSR	7/13 (53.8%)	3/11 (27.3%)	0.340	0.087-1.324	0.120
OSR	6/13 (46.2%)	2/11 (18.2%)	0.285	0.057-1.421	0.126
20q11.1					
LRCR	0/4 (0.0%)	2/20 (10.0%)	-	-	0.495*
DMFR	2/4 (50.0%)	6/20 (30.0%)	0.439	0.104-1.847	0.262
PFSR	2/4 (50.0%)	8/20 (40.0%)	0.340	0.087-1.324	0.120
OSR	0/4 (0.0%)	8/20 (40.0%)	-	-	0.155*
Deletion					
4p11					
LRCR	2/18 (11.1%)	0/6 (0.0%)	-	-	0.375*
DMFR	8/18 (44.4%)	0/6 (0.0%)	-	-	0.045*
PFSR	10/18 (55.6%)	0/6 (0.0%)	-	-	0.030*
OSR	8/18 (44.4%)	0/6 (0.0%)	-	-	0.064*
16q11.2					
LRCR	1/13 (7.7%)	1/11 (9.1%)	1.458	0.091-23.30	0.790

DMFR	5/13 (38.5%)	3/11 (27.3%)	0.826	0.197-3.460	0.793
PFSR	5/13 (38.5%)	5/11 (45.5%)	1.349	0.390-4.667	0.637
OSR	5/13 (38.5%)	3/11 (27.3%)	0.794	0.189-3.327	0.752
17.q11.1					
LRCR	2/20 (10.0%)	0/4 (0.0%)	-	-	0.495*
DMFR	6/20 (30.0%)	2/4 (50.0%)	1.524	0.305-7.607	0.608
PFSR	8/20 (40.0%)	2/4 (50.0%)	1.155	0.244-5.467	0.856
OSR	8/20 (40.0%)	0/4 (0.0%)	-	-	0.155*

Cox proportional hazards model was used for calculation.

Abbreviation: CNA, copy number alteration; DMFR, distant metastasis-free rate; LRCR, locoregional control rate; OSR, overall survival rate; PFSR, progression-free survival rate.

* These p-values were calculated by log-rank test.

Table 9. Association of copy number alteration with tumor regression grade

Copy number alteration	TRG 1	TRG 2	TRG 3	*P*-value
Pre-chemoradiation samples				
Amplification				
6p21.1				0.007
Yes	0 (0.0%)	4 (50.0%)	4 (50.0%)	
No	12 (42.9%)	12 (42.9%)	4 (14.3%)	
7q22.1				0.143
Yes	2 (25.0%)	6 (75.0%)	0 (0.0%)	
No	10 (38.5%)	10 (38.5%)	6 (23.1%)	
12q12				0.389
Yes	6 (50.0%)	4 (33.3%)	2 (16.7%)	
No	6 (27.3%)	12 (54.5%)	4 (18.2%)	
20q11.1				0.246
Yes	6 (33.3%)	7 (38.9%)	5 (27.8%)	
No	6 (37.5%)	9 (56.3%)	1 (6.3%)	
Deletion				
3p12.3				0.578
Yes	6 (46.2%)	5 (38.5%)	2 (15.4%)	
No	6 (28.6%)	11 (52.4%)	4 (19.0%)	
4p11				0.690
Yes	9 (32.1%)	14 (50.0%)	5 (17.9%)	
No	3 (50.0%)	2 (33.3%)	1 (16.7%)	
5q11.2				0.009
Yes	1 (14.3%)	2 (28.6%)	4 (57.1%)	
No	11 (40.7%)	14 (51.9%)	2 (7.4%)	
6p21.32				0.016
Yes	0 (0.0%)	8 (80.0%)	2 (20.0%)	
No	12 (50.0%)	8 (33.3%)	4 (16.7%)	
7q22.1				0.057

Yes	7 (63.6%)	3 (27.3%)	1 (9.1%)	
No	5 (21.7%)	13 (56.5%)	5 (21.7%)	
8p23.1				0.940
Yes	7 (33.3%)	10 (47.6%)	4 (19.0%)	
No	5 (38.5%)	6 (46.2%)	2 (15.4%)	
13q11				0.630
Yes	2 (22.2%)	5 (55.6%)	2 (22.2%)	
No	10 (40.0%)	11 (44.0%)	4 (16.0%)	
14q11.2				0.460
Yes	6 (37.5%)	6 (37.5%)	4 (25.0%)	
No	6 (33.3%)	10 (55.6%)	2 (11.1%)	
14q32.33				0.108
Yes	2 (16.7%)	6 (50.0%)	4 (33.3%)	
No	10 (45.5%)	10 (45.5%)	2 (9.1%)	
15q11.2				0.968
Yes	4 (33.3%)	6 (50.0%)	2 (16.7%)	
No	8 (36.4%)	10 (45.5%)	4 (18.2%)	
17p11.2				0.389
Yes	6 (27.3%)	12 (54.5%)	4 (18.2%)	
No	6 (50.0%)	4 (33.3%)	2 (16.7%)	
20q11.1				0.286
Yes	4 (50.0%)	4 (50.0%)	0 (0.0%)	
No	8 (30.8%)	12 (46.2%)	6 (23.1%)	
Post-chemoradiation samples				
Amplification				
3p12.3				0.651
Yes	1 (10.0%)	6 (60.0%)	3 (30.0%)	
No	3 (21.4%)	6 (42.9%)	5 (35.7%)	
6p21.1				< 0.001
Yes	0 (0.0%)	0 (0.0%)	6 (100.0%)	
No	4 (22.2%)	12 (66.7%)	2 (11.1%)	

11q13.3				0.043
Yes	0 (0.0%)	1 (20.0%)	4 (80.0%)	
No	4 (21.1%)	11 (57.9%)	4 (21.1%)	
13q11				0.920
Yes	2 (18.2%)	5 (45.5%)	4 (36.4%)	
No	2 (15.4%)	7 (53.8%)	4 (30.8%)	
20q11.1				0.091
Yes	4 (20.0%)	8 (40.0%)	8 (40.0%)	
No	0 (0.0%)	4 (100.0%)	0 (0.0%)	
Deletion				
4p11				0.108
Yes	0 (0.0%)	2 (33.3%)	4 (66.7%)	
No	4 (22.2%)	10 (55.6%)	4 (22.2%)	
16q11.2				0.432
Yes	2 (18.2%)	4 (36.4%)	5 (45.5%)	
No	2 (15.4%)	8 (61.5%)	3 (23.1%)	
17q11.1				0.472
Yes	0 (0.0%)	3 (75.0%)	1 (25.0%)	
No	4 (20.0%)	9 (45.0%)	7 (35.0%)	

P-value was calculated by Fisher's exact test.

Abbreviation: TRG, tumor regression grade.

Figure 13. Analysis of copy number alteration burden. (A) Box plots for copy number alteration (CNA) burden by pre- and post-chemoradiation samples. (B) Scatter plot for CNA burden by paired pre- and post-chemoradiation samples. (C) Box plots for CNA burden by tumor regression grade.

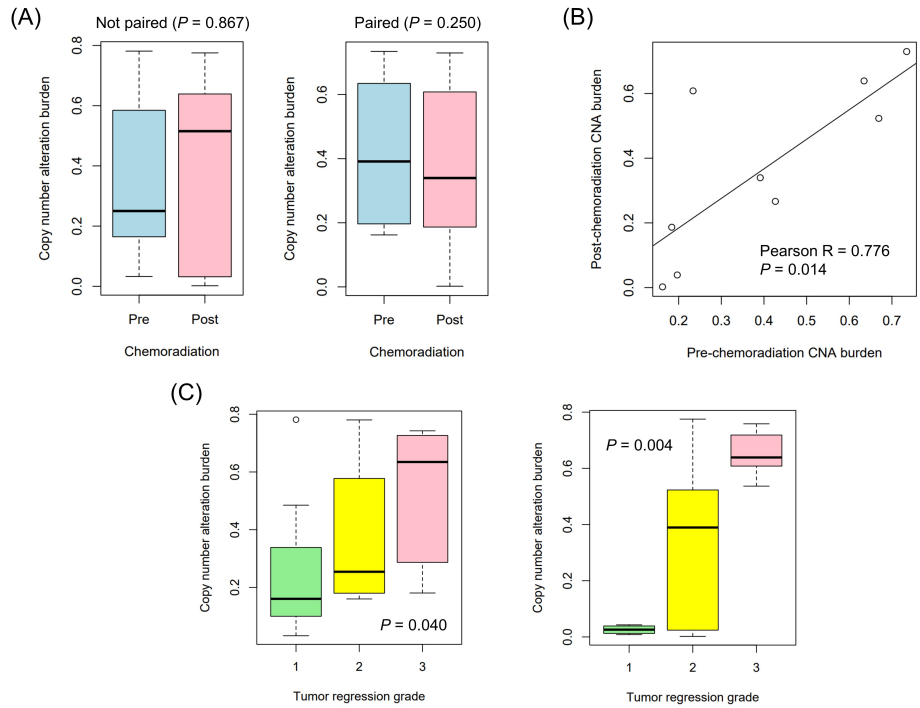


Table 10. Association of copy number alteration burden with clinical outcomes

Clinical outcomes	Hazard ratio	95% confidence interval	P-value
Pre-chemoradiation samples			
Locoregional control rate	2.542	0.223-724.5	0.218
Distant metastasis-free rate	3.467	0.409-29.42	0.255
Progression-free survival rate	3.467	0.409-29.42	0.255
Overall survival rate	8.527	0.662-109.9	0.100
Post-chemoradiation samples			
Locoregional control rate	>1000	0.006- >1000	0.133
Distant metastasis-free rate	0.790	0.076-8.271	0.844
Progression-free survival rate	1.485	0.175-12.64	0.717
Overall survival rate	2.450	0.197-30.45	0.486

These values were calculated based on Cox proportional hazards model.

8. Multivariate analysis in conjunction with clinical variables and genetic biomarkers

Results of the univariate analysis of the four clinical outcomes with clinical variables were summarized in Table 11. Elevated CEA was associated with all four clinical outcomes with statistical or marginal significance and was selected for further multivariate analysis. The results of the backward stepwise variable selection for the multivariate analysis, in conjunction with clinical variables and genetic biomarkers, were summarized in Table 12. Along with genetic biomarkers, elevated CEA was independently associated with DMFR, PFSR, and OSR.

Table 11. Univariate analysis of clinical outcomes with clinical variables

Characteristics (comparison vs. reference)	Locoregional control rate			Distant metastasis-free rate			Progression-free survival rate			Overall survival rate		
	HR	95% CI	P value	HR	95% CI	P value	HR	95% CI	P value	HR	95% CI	P value
Age (≥ 60 vs. < 60 years)	No event in "≥ 60 years"			0.463	0.116–1.853	0.276	0.516	0.151–1.767	0.292	0.761	0.204–2.839	0.684
High-risk T3–4 disease (yes vs. no)	2.868	0.297–27.66	0.362	1.155	0.310–4.308	0.830	1.570	0.459–5.375	0.472	1.879	0.468–7.540	0.374
N2 disease (yes vs. no)	0.510	0.053–4.913	0.560	0.810	0.202–3.246	0.766	0.921	0.269–3.153	0.896	0.820	0.204–3.291	0.779
Proximal rectal cancer (yes vs. no)	No event in "yes"			0.904	0.113–7.235	0.924	0.804	0.102–6.356	0.836	No event in "yes"		
Differentiation (moderate vs. well-differentiated)	No event in "well-differentiated"			1.365	0.170–10.93	0.770	1.780	0.227–13.94	0.583	1.251	0.156–20.02	0.833
Elevated CEA (yes vs. no)	6.404	0.873–46.98	0.068	3.951	0.955–16.34	0.058	4.492	1.277–15.80	0.019	5.670	1.466–21.93	0.012

Abbreviations: CEA, carcinoembryonic antigen; CI, confidence interval; HR, hazard ratio.

Table 12. Multivariate analysis in conjunction with clinical variables and genetic biomarkers

Characteristics (comparison vs. reference)	Locoregional control rate			Distant metastasis-free rate			Progression-free survival rate			Overall survival rate		
	HR	95% CI	P value	HR	95% CI	P value	HR	95% CI	P value	HR	95% CI	P value
Elevated CEA (yes vs. no)	-	-	-	1.511	1.034–19.85	0.045	1345	5.002–>10000	0.012	6.439	1.472–28.173	0.013
SMAD4 mutation (yes vs. no)	2.826	1.470–193.6	0.023	-	-	-	-	-	-	-	-	-
EGFR mutation (yes vs. no)	2.662	1.196–171.5	0.036	-	-	-	-	-	-	-	-	-
RUNX1 mutation (yes vs. no)	-	-	-	-	-	-	999.6	5.189–>10000	0.010	13.48	2.630–69.07	0.002
GATA1 mutation (yes vs. no)	-	-	-	-	-	-	>10000	33.783–>10000	0.010	-	-	-
BAP1 mutation (yes vs. no)	-	-	-	-	-	-	>10000	2.854–>10000	0.032	-	-	-
KMA2A mutation (yes vs. no)	-	-	-	-	-	-	>10000	7.228–>10000	0.014	-	-	-
APC mutation (yes vs. no)	-	-	-	-	-	-	-	-	-	0.222	0.041–1.214	0.083
Relative composition of SBS30 mutational signature (continuous, per 1%)	-	-	-	1.080	1.016–1.147	0.013	1.981	1.169–3.358	0.011	-	-	-

Tumor mutational burden (continuous, per 1 / megabase pair)	-	-	-	-	-	-	0.890	0.812–0.974	0.012	-	-	-

Abbreviations: CEA, carcinoembryonic antigen; CI, confidence interval; HR, hazard ratio.

Discussion

This study explored WES data from pre-CRT biopsy and post-CRT surgical specimens of patients with locally advanced rectal adenocarcinoma. Some changes in somatic SNVs, indels, genetic biomarkers derived from mutations, and CNAs between pre- and post-CRT samples were observed. Several somatic mutations and genetic biomarkers potentially affecting clinical outcomes were found. The initial intended approach of this study was to identify specific changes in the genomic landscape of non-responders after neoadjuvant CRT and incorporate those changes into prediction models. Although this study presents several differences between pre- and post-CRT tumor samples, the establishment of clinically significant changes in the genomic landscape after CRT was not possible, likely due to frequent quality check failures of the samples. However, several potential genetic biomarkers were found in the pre-CRT samples, and a statistically significant prediction model for treatment outcomes could be established.

This study found several somatic gene mutations in pre-CRT associated with clinical outcomes. The strongest association was observed in *RUNX1* mutation. In this study, *RUNX1* mutation was associated with worse DMFR, PFSR, and OSR. *RUNX1* mutation has been mainly studied in hematologic malignancy [22], and previous reports have shown that *RUNX1* mutation is correlated with poor prognosis [23, 24]. Recent studies have reported that *RUNX1* expression is also prevalent in solid tumor [25]. Several mechanisms for function of *RUNX1* as an oncogene or a tumor suppressor have been proposed [26]. Previous reports have analyzed the role

of *RUNX1* in colorectal cancer and concluded that *RUNX1* may facilitate cancer cell migration [27, 28]. In this study, *RUNX1* mutation impacted DMFR but not LRCR. It might indicate that worse PFSR and OSR of patients with *RUNX1* mutation were due to more frequent distant metastasis affected by *RUNX1* activity.

In this study, *EGFR* mutation was associated with worse LRCR and OSR. Although EGFR mutations are uncommon in colorectal cancer, it has been reported that *EGFR* signaling pathway can regulate cellular events and lead to progression of several cancers [29]. In pre-CRT samples of this study, 16% of analyzable patients had EGFR mutation. This high percentage of mutation might be due to the small sample size of this study. It might also be due to racial difference as a previous Korean study reported that the EGFR mutation rate was 22.41% in 58 colorectal cancer patients [30]. Impact of *EGFR* mutation on prognosis of colorectal cancer is questionable. Reports of *EGFR* mutation are scarce due to its rareness, although several studies have explored the association between *EGFR* expression and prognosis [31]. Contrary to the current analysis, the above-mentioned Korean study concluded that there was no association between *EGFR* mutation status and OSR [30]. Validation with a larger cohort of specific ethnicity may clarify this difference. *TP53* mutation was also associated with PFSR and OSR. It is well-known that *TP53* has a role in carcinogenesis of colorectal adenocarcinoma [32]. Furthermore, mutational analysis of prospective series has revealed that *TP53* mutation is associated with worse PFSR [6], concordant with the current finding.

SMAD4 mutation was associated with worse LRCR in this study. It has been reported that loss of *SMAD4* is involved in colorectal carcinogenesis [33]. Worse prognosis in patients with *SMAD4* mutation has been previously reported [34]. Four

relatively rare gene mutations (*GATA1*, *BAP1*, *KMT2A*, and *STAG2*) were significantly associated with worse PFSR or OSR in this study. *GATA1* mutation was associated with worse PFSR. GATA1 was considered to play a role in the hematopoietic system [35]. It might also play a role in colorectal cancer progression, as *GATA1* expression may affect prognosis of colorectal cancer [36]. *BAP1* mutation was also associated with worse PFSR. Although *BAP1* mutation is prevalent in other cancer types such as uveal melanoma and malignant mesothelioma, its prognostic significance in colorectal cancer has also been reported [37]. *KMT2A* is responsible for histone methyltransferase. The mechanism of its impact on colorectal cancer development has been proposed, and mutation in this gene is associated with worse PFSR [38]. *STAG2* mutation was associated with OSR. *STAG2* encodes a cohesion subunit. It has been proposed that its mutational inactivation might lead to aneuploidy by dysfunctional chromosomal segregation [39]. However, impact of *STAG2* mutation in colorectal cancer remains questionable [40]. These are uncommon mutations in colorectal adenocarcinoma. Their proposed prognostic impact requires independent external validation. *APC* mutation is one of the most frequent mutations in colorectal adenocarcinoma. It is also one of well-known driver mutations. In this study, *APC* mutation was associated with better OSR, concordant with previous literature [41].

No gene mutations in post-CRT samples impacted clinical outcomes. This was presumably due to the low number of analyzable post-CRT samples compared to pre-CRT samples. Statistically significant relationship between specific gene mutations and TRG was not found either.

Several differences of genomic landscape between pre- and post-CRT samples were found in this study. Although differences in mutation rates of the same genes between paired pre- and post-CRT samples were not significant, not all mutations observed in pre-CRT samples were retained in post-CRT samples, including driver mutations such as *APC*, *TP53*, and *KRAS*. Yang et al. [42] have compared genomic landscape of 28 pairs of locally advanced rectal cancer samples between before and after neoadjuvant CRT. Analyzed tumors were non-responders to CRT. Similar to this study, several genomic differences between pre- and post-CRT samples were found. Low number of retained driver genes was observed. The average shared mutation rate was 8.21%. Changes in CNA were also found. Toomey et al. [43] have used a similar approach with different conclusion. WES was performed for pre-treatment biopsy specimen, on-treatment biopsy specimen, and surgical tissues. There was no newly found driver mutation when WES results of pre-treatment and on-treatment biopsy specimen were compared. Furthermore, the vast majority of driver mutations retained when WES results of pre-treatment biopsy and surgical specimen were compared, although variant allele fractions might differ. Kamran et al. [44] have also compared pre- and post-CRT mutations and reported that gene mutation status occurring frequently in colorectal cancer such as *TP53*, *KRAS* and *APC* is grossly similar between pre- and post-CRT samples. Mutational differences between pre- and post-CRT samples reported in this study and the study by Yang et al. might be affected by intra-tumoral heterogeneity. Different tissues obtained in the same rectal tumor might harbor different mutations. In addition, simple bulk sequencing might not be sufficient to demonstrate variance of clonal or subclonal mutations of rectal cancer [45]. Discordant results between this study and above-mentioned studies might also indicate that analyzing mutations in rectal

cancer and effect of CRT on mutations is susceptible to difference in detailed methodology of respective studies such as sequencing depth and ways to obtain tissues due to substantial intra-tumoral heterogeneity of rectal cancer. Additional sequencing data with deeper depth might be one way to address this issue.

Six COSMIC SBS signatures were observed in WES materials of this study. SBS1 and SBS44 are known to be relatively frequent in colorectal adenocarcinoma, but SBS30, SBS29, or SBS3 [15]. Current findings on mutational signature changes between pre- and post-CRT samples and association between mutational signature and clinical outcomes involve SBS29 and SBS30. Although studies on other primaries have shown significant relationships between certain mutational signatures and clinical outcomes [46, 47], it is quite difficult to interpret current findings for mutational signatures due to unknown mechanisms behind rare mutational signatures in colorectal adenocarcinoma. Further studies may help clarify this. This is also true for DBS and indel signatures reported in this study.

Effects of tumor mutational burden on clinical outcomes and TRG were explored in this study. The exact definition of tumor mutational burden is different among studies, including the number of mutations and the density of mutations from sequencing materials. Previous studies with larger cohort have reported that higher tumor mutational burden is associated with better prognosis in colorectal cancer [48, 49]. Furthermore, higher rate of response to immune checkpoint inhibitor has been reported in colorectal tumors with higher mutational burden [50]. However, the relationship of tumor mutational burden with clinical outcomes in this study showed contrasting results, with higher burden in pre-CRT samples showing significant

associations with worse PFSR. Limited size of the analyzed cohort might have led to such discordant results.

Tumor heterogeneity was quantified by MATH score in this study. Changes of MATH score before and after CRT and the relationship between MATH score and clinical outcomes or pathologic response were determined. In this study, MATH score decreased with marginal statistical significance after CRT in patients with paired pre- and post-CRT samples, whereas clinical outcomes and TRG did not show significant associations with MATH score. In contrast, increase of MATH score after CRT [51], associations of MATH score with higher disease stage [51] and pathologic response [52] have been reported. This discordance might be due to the limited size of the cohort. However, decreased MATH score after CRT might be possible if clonal selection by CRT is dominant. The effect of CRT on tumor heterogeneity might be different depending on clinical situations such as response.

There have been some studies that have correlated CNA burden with clinical outcomes, but there are several different definitions of CNA burden in existence. In this study and in previous literature, the ratio of the sum of the length of copy number-altered segments to the total length of analyzed segments was used [53]. Other definitions include the number of copy number-altered genes [54] and algorithm-based scoring [55]. Despite using different definitions, several studies have shown a significant association between CNA burden and clinical outcomes. However, no impact of CNA burden on clinical outcomes was observed in this study, although an association with TRG was reported. In addition, this study found that tumors with high CNA burden generally retained their high burden after neoadjuvant CRT. CNA burden may be a useful biomarker if further studies yield clinically useful

results, but the definition of CNA burden needs to be standardized to be applied in a broader aspect of genomics study.

This study performed a multivariate analysis of clinical outcomes in conjunction with clinical variables and potential genetic biomarkers. Although the sample size and number of events were small, elevated CEA and genetic biomarkers were independently associated with clinical outcomes. Elevated CEA is a well-known risk factor for poor prognosis in locally advanced rectal cancer [4]. Clinical risk factors remain effective in the era of genetic testing, and both clinical and genetic factors should be considered when constructing a model to predict tumor regression and prognosis.

The main limitation of this study was the limited number of available specimens followed by frequent failure of quality checks on the obtained specimens. The normal samples from pre-CRT biopsies and the tumor samples from post-CRT surgical specimens had high rates of quality check failure. These high failure rates were likely due to insufficient DNA material in the normal samples from pre-CRT biopsies and the tumor samples from post-CRT surgical specimens. The pre-CRT biopsy samples were mostly obtained from tumor-enriched tissues, resulting in only a small fraction of normal tissues available for analysis. On the other hand, the tumor samples from post-CRT surgical specimens were obtained from areas predominantly containing malignant tumors based on pathological examination, but the surviving tumor fractions in those areas were often low, reaching near complete response status. Due to the limited number of paired pre- and post-CRT samples, it was challenging to identify differences between the pairs. This ultimately became the main limitation against initially intended aim of this study.

This study has several minor limitations. First, the cohort size of this study might be too small for reliability. Thus, further clarification by independent external validation is needed. Second, the sequencing depth of an average of x30 was relatively low. Considering substantial heterogeneity of rectal cancer, this might have obscured clonal variations of mutations. Contrary to several studies, blood samples, which were used as normal references in other studies, were not obtainable. Despite these limitations, this study showed the genomic landscape of Korean patients with locally advanced rectal adenocarcinoma before and after CRT and generated hypotheses for further analysis.

In conclusion, differences in the genomic landscape between pre- and post-CRT samples were observed. Despite the aforementioned limitations preventing the achievement of the initial goal, several potential genetic biomarkers were identified in pre-CRT samples for prognosis prediction. Significant prediction models for treatment outcomes, incorporating both clinical and genetic factors, were successfully generated, suggesting the potential for a clinically useful prediction model utilizing genetic variables. However, independent external validation is necessary in the future to further validate these findings.

References

1. Garcia-Aguilar J, Hernandez de Anda E, Sirivongs P, Lee SH, Madoff RD, Rothenberger DA: A pathologic complete response to preoperative chemoradiation is associated with lower local recurrence and improved survival in rectal cancer patients treated by mesorectal excision. *Dis Colon Rectum* 2003, 46(3):298-304.

2. Smith JJ, Strombom P, Chow OS, Roxburgh CS, Lynn P, Eaton A, Widmar M, Ganesh K, Yaeger R, Cercek A *et al*: Assessment of a Watch-and-Wait Strategy for Rectal Cancer in Patients With a Complete Response After Neoadjuvant Therapy. *JAMA Oncol* 2019, 5(4):e185896.

3. van der Valk MJM, Hilling DE, Bastiaannet E, Meershoek-Klein Kranenbarg E, Beets GL, Figueiredo NL, Habr-Gama A, Perez RO, Renehan AG, van de Velde CJH *et al*: Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study. *Lancet* 2018, 391(10139):2537-2545.

4. Kang BH, Song C, Kang SB, Lee KW, Lee HS, Kim JS: Nomogram for Predicting the Pathological Tumor Response from Pre-treatment Clinical Characteristics in Rectal Cancer. *Anticancer Res* 2020, 40(4):2171-2177.

5. Zhou P, Goffredo P, Ginader T, Thompson D, Hrabe J, Gribovskaja-Rupp I, Kapadia M, Hassan I: Impact of KRAS status on tumor response and

survival after neoadjuvant treatment of locally advanced rectal cancer. *J Surg Oncol* 2021, 123(1):278-285.

6. Sclafani F, Wilson SH, Cunningham D, Gonzalez De Castro D, Kalaitzaki E, Begum R, Wotherspoon A, Capdevila J, Glimelius B, Rosello S *et al*: Analysis of KRAS, NRAS, BRAF, PIK3CA and TP53 mutations in a large prospective series of locally advanced rectal cancer patients. *Int J Cancer* 2020, 146(1):94-102.

7. Ryan JE, Warrier SK, Lynch AC, Ramsay RG, Phillips WA, Heriot AG: Predicting pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a systematic review. *Colorectal Dis* 2016, 18(3):234-246.

8. Lee IH, Kang K, Kang BW, Lee SJ, Bae WK, Hwang JE, Kim HJ, Park SY, Park JS, Choi GS *et al*: Genetic variations using whole-exome sequencing might predict response for neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Med Oncol* 2018, 35(11):145.

9. Xiao WW, Li M, Guo ZW, Zhang R, Xi SY, Zhang XG, Li Y, Wu DQ, Ren YF, Pang XL *et al*: A Genotype Signature for Predicting Pathologic Complete Response in Locally Advanced Rectal Cancer. *Int J Radiat Oncol Biol Phys* 2021, 110(2):482-491.

10. Kasi A, Abbasi S, Handa S, Al-Rajabi R, Saeed A, Baranda J, Sun W: Total Neoadjuvant Therapy vs Standard Therapy in Locally Advanced Rectal Cancer: A Systematic Review and Meta-analysis. *JAMA Netw Open* 2020,

3(12):e2030097.

11. Liu S, Jiang T, Xiao L, Yang S, Liu Q, Gao Y, Chen G, Xiao W: Total Neoadjuvant Therapy (TNT) versus Standard Neoadjuvant Chemoradiotherapy for Locally Advanced Rectal Cancer: A Systematic Review and Meta-Analysis. *Oncologist* 2021, 26(9):e1555-e1566.

12. Ryan R, Gibbons D, Hyland JM, Treanor D, White A, Mulcahy HE, O'Donoghue DP, Moriarty M, Fennelly D, Sheahan K: Pathological response following long-course neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *Histopathology* 2005, 47(2):141-146.

13. Li H: Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In.; 2013: arXiv:1303.3997.

14. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP: Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* 2018, 28(11):1747-1756.

15. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN *et al*: The repertoire of mutational signatures in human cancer. *Nature* 2020, 578(7793):94-101.

16. Wang S, Tao Z, Wu T, Liu XS: Sigflow: an automated and comprehensive pipeline for cancer genome mutational signature analysis. *Bioinformatics* 2021, 37(11):1590-1592.

17. Mroz EA, Rocco JW: MATH, a novel measure of intratumor genetic

heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol* 2013, 49(3):211-215.

18. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, Szallasi Z, Eklund AC: Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* 2015, 26(1):64-70.

19. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G: GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011, 12(4):R41.

20. Smit WL, Spaan CN, Johannes de Boer R, Ramesh P, Martins Garcia T, Meijer BJ, Vermeulen JLM, Lezzerini M, MacInnes AW, Koster J *et al*: Driver mutations of the adenoma-carcinoma sequence govern the intestinal epithelial global translational capacity. *Proc Natl Acad Sci U S A* 2020, 117(41):25560-25570.

21. Poynter JN, Siegmund KD, Weisenberger DJ, Long TI, Thibodeau SN, Lindor N, Young J, Jenkins MA, Hopper JL, Baron JA *et al*: Molecular characterization of MSI-H colorectal cancer by MLHI promoter methylation, immunohistochemistry, and mismatch repair germline mutation screening. *Cancer Epidemiol Biomarkers Prev* 2008, 17(11):3208-3215.

22. Sood R, Kamikubo Y, Liu P: Role of RUNX1 in hematological malignancies. *Blood* 2017, 129(15):2070-2082.

23. Grossmann V, Kern W, Harbich S, Alpermann T, Jeromin S, Schnittger S,

Haferlach C, Haferlach T, Kohlmann A: Prognostic relevance of RUNX1 mutations in T-cell acute lymphoblastic leukemia. *Haematologica* 2011, 96(12):1874-1877.

24. Mendler JH, Maharry K, Radmacher MD, Mrozek K, Becker H, Metzeler KH, Schwind S, Whitman SP, Khalife J, Kohlschmidt J *et al*: RUNX1 mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and MicroRNA expression signatures. *J Clin Oncol* 2012, 30(25):3109-3118.

25. Tuo Z, Zhang Y, Wang X, Dai S, Liu K, Xia D, Wang J, Bi L: RUNX1 is a promising prognostic biomarker and related to immune infiltrates of cancer-associated fibroblasts in human cancers. *BMC Cancer* 2022, 22(1):523.

26. Hong D, Fritz AJ, Gordon JA, Tye CE, Boyd JR, Tracy KM, Frietze SE, Carr FE, Nickerson JA, Van Wijnen AJ *et al*: RUNX1-dependent mechanisms in biological control and dysregulation in cancer. *J Cell Physiol* 2019, 234(6):8597-8609.

27. Li Q, Lai Q, He C, Fang Y, Yan Q, Zhang Y, Wang X, Gu C, Wang Y, Ye L *et al*: RUNX1 promotes tumour metastasis by activating the Wnt/beta-catenin signalling pathway and EMT in colorectal cancer. *J Exp Clin Cancer Res* 2019, 38(1):334.

28. Lu C, Yang Z, Yu D, Lin J, Cai W: RUNX1 regulates TGF-beta induced migration and EMT in colorectal cancer. *Pathol Res Pract* 2020, 216(11):153142.

29. Krasinskas AM: EGFR Signaling in Colorectal Carcinoma. *Patholog Res Int* 2011, 2011:932932.

30. Oh BY, Lee RA, Chung SS, Kim KH: Epidermal growth factor receptor mutations in colorectal cancer patients. *J Korean Soc Coloproctol* 2011, 27(3):127-132.

31. Del Carmen S, Corchete LA, Gervas R, Rodriguez A, Garcia M, Alcazar JA, Garcia J, Bengoechea O, Munoz-Bellvis L, Sayagues JM *et al*: Prognostic implications of EGFR protein expression in sporadic colorectal tumors: Correlation with copy number status, mRNA levels and miRNA regulation. *Sci Rep* 2020, 10(1):4662.

32. Nakayama M, Oshima M: Mutant p53 in colon cancer. *J Mol Cell Biol* 2019, 11(4):267-276.

33. Zhao M, Mishra L, Deng CX: The role of TGF-beta/SMAD4 signaling in cancer. *Int J Biol Sci* 2018, 14(2):111-123.

34. Fang T, Liang T, Wang Y, Wu H, Liu S, Xie L, Liang J, Wang C, Tan Y: Prognostic role and clinicopathological features of SMAD4 gene mutation in colorectal cancer: a systematic review and meta-analysis. *BMC Gastroenterol* 2021, 21(1):297.

35. Gutierrez L, Caballero N, Fernandez-Calleja L, Karkoulia E, Strouboulis J: Regulation of GATA1 levels in erythropoiesis. *IUBMB Life* 2020, 72(1):89-105.

36. Yu J, Liu M, Liu H, Zhou L: GATA1 promotes colorectal cancer cell proliferation, migration and invasion via activating AKT signaling pathway. *Mol Cell Biochem* 2019, 457(1-2):191-199.

37. Tang J, Xi S, Wang G, Wang B, Yan S, Wu Y, Sang Y, Wu W, Zhang R, Kang T: Prognostic significance of BRCA1-associated protein 1 in colorectal cancer. *Med Oncol* 2013, 30(2):541.

38. Fang Y, Zhang D, Hu T, Zhao H, Zhao X, Lou Z, He Y, Qin W, Xia J, Zhang X *et al*: KMT2A histone methyltransferase contributes to colorectal cancer development by promoting cathepsin Z transcriptional activation. *Cancer Med* 2019, 8(7):3544-3552.

39. Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, Harris BT, Toretsky JA, Rosenberg SA, Shukla N, Ladanyi M *et al*: Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science* 2011, 333(6045):1039-1043.

40. Kim MS, Kim SS, Je EM, Yoo NJ, Lee SH: Mutational and expressional analyses of STAG2 gene in solid cancers. *Neoplasma* 2012, 59(5):524-529.

41. Schell MJ, Yang M, Teer JK, Lo FY, Madan A, Coppola D, Monteiro AN, Nebozhyn MV, Yue B, Loboda A *et al*: A multigene mutation classification of 468 colorectal cancers reveals a prognostic role for APC. *Nat Commun* 2016, 7:11743.

42. Yang J, Lin Y, Huang Y, Jin J, Zou S, Zhang X, Li H, Feng T, Chen J, Zuo Z *et al*: Genome landscapes of rectal cancer before and after preoperative

chemoradiotherapy. *Theranostics* 2019, 9(23):6856-6866.

43. Toomey S, Gunther J, Carr A, Weksberg DC, Thomas V, Salvucci M, Bacon O, Sherif EM, Fay J, Kay EW *et al*: Genomic and Transcriptomic Characterisation of Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Cancers (Basel)* 2020, 12(7).

44. Kamran SC, Lennerz JK, Margolis CA, Liu D, Reardon B, Wankowicz SA, Van Seventer EE, Tracy A, Wo JY, Carter SL *et al*: Integrative Molecular Characterization of Resistance to Neoadjuvant Chemoradiation in Rectal Cancer. *Clin Cancer Res* 2019, 25(18):5561-5571.

45. Liu M, Liu Y, Di J, Su Z, Yang H, Jiang B, Wang Z, Zhuang M, Bai F, Su X: Multi-region and single-cell sequencing reveal variable genomic heterogeneity in rectal cancer. *BMC Cancer* 2017, 17(1):787.

46. Xing R, Zhou Y, Yu J, Yu Y, Nie Y, Luo W, Yang C, Xiong T, Wu WKK, Li Z *et al*: Whole-genome sequencing reveals novel tandem-duplication hotspots and a prognostic mutational signature in gastric cancer. *Nat Commun* 2019, 10(1):2037.

47. Meng F, Zhang L, Ren Y, Ma Q: The genomic alterations of lung adenocarcinoma and lung squamous cell carcinoma can explain the differences of their overall survival rates. *J Cell Physiol* 2019, 234(7):10918-10925.

48. Domingo E, Camps C, Kaisaki PJ, Parsons MJ, Mouradov D, Pentony MM, Makino S, Palmieri M, Ward RL, Hawkins NJ *et al*: Mutation burden and

other molecular markers of prognosis in colorectal cancer treated with curative intent: results from the QUASAR 2 clinical trial and an Australian community-based series. *Lancet Gastroenterol Hepatol* 2018, 3(9):635-643.

49. Innocenti F, Ou FS, Qu X, Zemla TJ, Niedzwiecki D, Tam R, Mahajan S, Goldberg RM, Bertagnolli MM, Blanke CD *et al*: Mutational Analysis of Patients With Colorectal Cancer in CALGB/SWOG 80405 Identifies New Roles of Microsatellite Instability and Tumor Mutational Burden for Patient Outcome. *J Clin Oncol* 2019, 37(14):1217-1227.

50. Schrock AB, Ouyang C, Sandhu J, Sokol E, Jin D, Ross JS, Miller VA, Lim D, Amanam I, Chao J *et al*: Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. *Ann Oncol* 2019, 30(7):1096-1103.

51. Bettoni F, Masotti C, Correa BR, Donnard E, Dos Santos FF, Sao Juliao GP, Vailati BB, Habr-Gama A, Galante PAF, Perez RO *et al*: The Effects of Neoadjuvant Chemoradiation in Locally Advanced Rectal Cancer-The Impact in Intratumoral Heterogeneity. *Front Oncol* 2019, 9:974.

52. Greenbaum A, Martin DR, Bocklage T, Lee JH, Ness SA, Rajput A: Tumor Heterogeneity as a Predictor of Response to Neoadjuvant Chemotherapy in Locally Advanced Rectal Cancer. *Clin Colorectal Cancer* 2019, 18(2):102-109.

53. Hieronymus H, Schultz N, Gopalan A, Carver BS, Chang MT, Xiao Y, Heguy A, Huberman K, Bernstein M, Assel M *et al*: Copy number alteration

burden predicts prostate cancer relapse. *Proc Natl Acad Sci U S A* 2014, 111(30):11139-11144.

54. Lu Z, Chen H, Li S, Gong J, Li J, Zou J, Wu L, Yu J, Han W, Sun H *et al*: Tumor copy-number alterations predict response to immune-checkpoint-blockade in gastrointestinal cancer. *J Immunother Cancer* 2020, 8(2):e000374.

55. Bassaganyas L, Pinyol R, Esteban-Fabro R, Torrens L, Torrecilla S, Willoughby CE, Franch-Exposito S, Vila-Casadesus M, Salaverria I, Montal R *et al*: Copy-Number Alteration Burden Differentially Impacts Immune Profiles and Molecular Features of Hepatocellular Carcinoma. *Clin Cancer Res* 2020, 26(23):6350-6361.

국문 초록

국소 진행성 직장 선암의 유전체 지형: 선행 항암방사선치료 전후의 비교 및 유전적 바이오마커의 임상적 결과와 종양 반응에 대한 영향

목적: 직장암의 예후에 영향을 미치는 유전적 바이오마커를 전체 엑솜 염기서열 분석을 이용하여 확인하고자 하였다.

연구대상 및 방법: 선행 항암방사선치료 이후 근치적 절제를 시행 받은 27명의 환자에 대해 항암방사선치료 전 생검 및 항암방사선치료 후 수술 검체에서 시료를 확보하였다. 엑솜 염기서열 분석의 평균 커버리지는 30x 였다. 시료에 존재하는 체세포 단일 염기 변이와 삽입/결손을 확인하였다. 종양 불균질성은 변이 대립 유전자 종양 분균질성 (MATH) 점수를 이용하여 정량하였다. 항암방사선치료 전후의 유전적 바이오마커와 흔히 발생하는 유전자 복제수 변이를 비교하였다. 종양 퇴행 등급과 임상적 결과에 대해서 유전적 바이오마커와의 연관성을 분석하였다.

결과: 가장 흔히 변이된 다섯 가지 유전자는 항암방사선치료 전

검체에서는 *APC*, *TP53*, *NF1*, *KRAS*, *NOTCH1* 였고, 항암방사선치료 후 검체에서는 *APC*, *TP53*, *NF1*, *CREBBP*, *ATM* 였다. *RUNX1*, *EGFR*, *TP53* 등 항암방사선치료 전 검체의 몇몇 유전자 변이들은 임상적 결과와의 연관성을 보였으나, 종양 퇴행 등급과는 연관이 없었다. 항암방사선치료 후 검체에서는 이러한 연관성을 확인할 수 없었다. 항암방사선치료 전후의 검체 사이에서 발암 변이 상태가 일치하지 않는 경우가 있었다. 종양 변이 부담 분석에서는 백만 염기 당 돌연변이 수가 불량한 예후와 연관성이 있었다. 여섯 가지의 단일 염기 치환 (SBS) 이 확인되었다 (SBS1, SBS30, SBS29, SBS49, SBS3, SBS44). MATH 점수는 대응표본 분석에서 항암방사선치료 이후 감소하였다. 항암방사선치료 후 검체에서 확인된 유전자 복제수 변이 중 절반 이하가 항암방사선치료 전 검체에서도 확인되었다.

결론: 항암방사선치료 전후의 검체는 서로 다른 유전체 지형을 보여주었다. 항암방사선치료 전 검체에서 잠재적인 유전적 바이오마커를 확인할 수 있었으나, 외부 검증을 통한 추가 검증이 필요하겠다.

주요어: 직장암, 선행 항암방사선치료, 전체 엑솜 염기서열

학번: 2021-35331