Supplementary Information

Contents

1. Supplementary methods

1.1. Ground truth of microsatellite status

In this study, the microsatellite status based on polymerase chain reaction (PCR) analysis detection of microsatellite instability (MSI-PCR) was used as the ground truth. DNA mismatch repair (MMR) protein immunohistochemistry (IHC) staining and targeted next-generation sequencing (NGS) were used as ground truth labels only in the additional explorative analysis; performance evaluation was performed by incorporating our deep-learning-based assay with MMR IHC (Figure 4) and the modified ground truth determined by targeted NGS (Supplementary Table 4).

1) MSI-PCR

Four 4- to 5-μm thick, formalin-fixed, paraffin-embedded (FFPE) tissue sections were used for assessment by PCR analysis or MMR IHC. Standard pentaplex PCR amplification of a panel of five microsatellites (BAT-25, BAT-26, D5S346, D2S123, and D17S250) was performed. Tumors with instability in two or more out of five microsatellite loci were classified into the MSI-high (MSIH) group, those with instability in only one locus were considered MSI-low (MSIL), and tumors showing no instability in all loci were considered microsatellite stable (MSS). In our experiment, cases with MSIL or MSS were assigned to the MSS group, and MSIH was designated as MSI [1-3].

For MMR IHC, tissue was processed using the streptavidin-biotin method with an LSAB kit (Dako, Carpinteria, CA). Cases with retained staining for all four MMR proteins (MLH1, MSH2, MSH6, or PMS2) were classified as MSS, and cases with loss of staining for one or more MMR proteins were defined as MSI [4].

2) NGS

We additionally performed targeted NGS analysis on 9 cases that showed wide discrepancies in probability cores of the model compared to the results of MSI-PCR.

- DNA extraction

Two to five slices that were 6 μm–thick from each specimen were used for genomic DNA extraction. After processing with xylene and ethanol for deparaffinization, genomic DNA was isolated using a NEXprep FFPE Tissue kit (NexK-9000; Genes Laboratories, Gyeonggi-do, Korea), according to the manufacturer's protocol.

- DNA library preparation and sequencing

Targeted NGS analysis was performed using the MiSeq platform (Illumina, San Diego, CA) with OncoPanel AMC version 3, which was designed by Asan Medical Center through SureDesign (Agilent Technologies, Santa Clara, CA) using the GRCh37 reference version. Entire exons for four MMR genes (MLH1, MSH2, MS6, and PMS2) were covered by the OncoPanel AMC version 3 panel [5]. Using the SureSelect XT reagent kit (Agilent Technologies, Santa Clara, CA, USA), DNA fragmentation, repair, A-tailing, and ligation was performed with the TruSeq adaptor. Each library was constructed with sample-specific barcodes 6 bp in size and quantified using the Qubit kit. Eight libraries were pooled (yielding a total of 720 ng) for hybrid capture using the Agilent SureSelect XT custom kit (OP-AMCv3 RNA bait, 1.2 Mb; Agilent Technologies). The concentration of the enriched target was measured using quantitative polymerase chain reaction (qPCR; Kapa Biosystems, Inc., Woburn, MA, USA). DNA libraries that passed quality checks were sequenced using MiSeq [6].

- Genomic data analysis and quality control

Sequenced reads were mapped to the human reference genome (hg 19) using the Burrows-Wheeler Aligner (version 0.5.9). IndelRealigner, and TableRecalibration,in Genome Analysis Tool Kit (GATK) were used for local alignment optimization and duplication marking. Variant calling in tumor tissue was performed with matched normal tissue using MuTect (1.1.7) and SomaticIndelocator in GATK. Germline variants were filtered out using the common dbsnp database (build 141; found in ≥1% of samples) and a panel of normal samples. Filtered somatic variants were annotated with the Variant Effect Predictor (v79), and then converted to MAF

file using vcf2maf (v1.6.12). Quality checks for FASTQ files were performed by FastQC. CalculateHSMetrics and MarkDuplicates modules of the Picard package were used for analysis-ready BAM files. Integrative Genomics Viewer (v2.4) was used to view the BAM file26 [6].

1.2. Model development for automated tumor segmentation

To train an automatic tumor detector for whole-slide images (WSIs) of colorectal cancer, we used an internal training set (n=351), which was same as that used for MSI prediction. Tumor areas in the WSIs were manually annotated by two expert pathologists using Automated Slide Analysis Platform (ASAP) Software (version 1.9, https://computationalpathologygroup.github.io/ASAP/) [7]. Image preprocessing was performed to remove non-tissue regions and normalize the stain per patch [8,9]. To classify tumor areas versus normal tissue at the patch level, we trained and used transfer learning of a residual convolutional neural network (CNN) with 18 layers (ResNet18), which uses the ImageNet dataset pretrained weights as is standard practice in deep learning literature [10]. For automated segmentation of an entire WSI, a sliding window approach was used to obtain the per tissue patch probability with subsequent thresholding chosen empirically using the internal training set. Finally, automated tumor segmentation was applied to all internal and external test tests.

1.3. Image preprocessing and model development for microsatellite instability prediction

The automatically detected tumor areas were subdivided into non-overlapping and color-normalized patch images at a resolution of 0.5 μm/pixel at 20x magnification. To account for severe variations in staining that may not have been resolved by color normalization, we applied color augmentation techniques (randomly adjusting hue, saturation, contrast and brightness with probability 0.25) to all the normalized patches during training and testing [8,9]. Subsequently, our models were robust to variations in staining between the internal and external datasets. By employing an automated tumor detector as the first step in our pipeline, we could easily apply our MSI predictor to any WSI when only global annotations were available (i.e., absence of expert annotations that are laborious to obtain).

For the development of the deep learning framework for MSI prediction in colorectal cancer, we trained and used transfer learning of another ResNet18; we used two separate ResNet18 models for tumor detection and MSI patch-level prediction in tumor detection areas [10,11]. Herein, the models were each trained for 20 epochs with a learning rate of 0.0001 using the Adam optimizer, including the use of balanced sampling of all images and color-based augmentation mentioned earlier to avoid overfitting.

Subsequently, the pretrained MSI patch network was used as a feature extractor to obtain spatial feature maps that served as input in the final WSI MSI model that predicts the microsatellite status. For this purpose, we designed a network inspired by the prior work, and the designed network employed spatial feature maps as input and was learned to aggregate all relevant features into a single vector per slide to predict MSI probability [12]. The spatial features were relevant to accurately predict the microsatellite status. Our final model is technically novel from existing approaches that ignore spatial context and only employ flattened latent vectors with complex scoring strategies (i.e., heuristic majority voting of patch level predictions, whereas our model is end-to-end with full automation). This model was trained for 40 epochs to minimize the binary cross-entropy loss, using an initial learning rate of 0.01 that is gradually decreased at epochs 10, 20 and 30. Moreover, during training we randomly sampled 32 patches per slide, whereas all available patches per slide were used in testing. Finally, the WSI-level prediction of either MSS or MSI was based on an optional probability threshold of 0.5.

1.4. Assessment of pathologic features

1) Tumor differentiation

Well-differentiated adenocarcinoma was defined if gland formation accounted for more than 95% of the tumor, moderately differentiated adenocarcinoma showed 50–95% gland formation, and poorly differentiated adenocarcinoma showed <50% gland formation [13]. If the worst grade of tumor was seen at 10% or more, it was applied for the overall grade, and tumor budding was not counted as poor differentiation [4,14].

Tumors with greater than 50% area showing extracellular mucin were classified as mucinous carcinoma,

and tumors were classified as signet ring cell carcinomas if greater than 50% of their area showed signet ring cell differentiation [14]. The amount of mucinous component within the tumor area was manually quantified as the percentage of mucinous component per a total tumor area. Medullary carcinoma was defined as a tumor that was poorly differentiated or undifferentiated with a well-circumscribed margin with a marked lymphocytic infiltrate in the peritumoral and intratumoral area [4].

2) Expansile growth pattern

An advancing edge of the tumor at low power with a pushing/expansile pattern was considered to be the presence of an expansile growth pattern, and otherwise was considered to be the absence of this pattern (infiltrative pattern). If the advancing edge of the tumor was not present, this field was coded as unknown [14].

3) Tumor-infiltrating lymphocytes (TILs)

The presence of intra-epithelial TILs (eTILs) was considered when there were at least 5 intra-epithelial lymphocytes/high-power field (HPF) and at least 10 HPFs had been searched on each WSI [4]. Stromal TILs (sTILs) were measured as a percentage of mononuclear inflammatory cells in the stromal area as counted in 5 HPFs in the invasive front and in the center of the tumor with the exception of tumor areas with crush artifacts, necrosis, or regressive hyalinization. For statistical analysis, three levels of infiltration of sTILs were determined: 1, weak (0–10% of sTILs); 2, moderate (20–40% of sTILs); and 3, strong (50–90% of sTILs). Those levels were grouped into levels 1–2 and level 3 [15].

4) Crohn's-like lymphocytic reaction

The presence of Crohn's-like lymphocytic reaction was scored when at least 4 nodular lymphoid aggregates were counted in a low-power field (4×) beyond the advancing edge of the tumor and generally within the subserosa or mesenteric fat [4]. If the advancing edge of the tumor was not present, this field was recorded as unknown [14].

5) Dirty necrosis

Dirty necrosis is often considered a characteristic of colorectal carcinomas, and a lack of dirty necrosis is considered an MSI-associated feature. If a rare focus of necrosis of less than 10% was present, then the tumor was classified as a lack of dirty necrosis. Infarcted areas were not included [13,14].

6) Neutrophilic and eosinophilic infiltration

Neutrophils or eosinophils were measured and graded as the number of neutrophils or eosinophils per HPF; a score of 1 represented the absence of to low count of eosinophils (< 10/HPF), a score of 2 represented an intermediate count (10-50/HPF), and a score of 3 represented a high count (>50/HPF) [16]. Those levels were grouped into levels 1–2 as the “low” group and level 3 as the “high” group.

2. Supplementary Results

2.1. Supplementary Tables

Supplementary Table 1. Dataset partition

Cohort		Number of patients	Material & staining	Number of slides	Number of patches (k= x10^3)	MSS:MSI (MSI-PCR)
Training		351	FFPE, H&E	351	~124k	231:120
Validation		90	FFPE, H&E	90	~30k	59:31
Test	CRC-TEST	198	FFPE, H&E	198	~71k	119:79
	CRC-NCM	154	FFPE, H&E	154	~50k	100:54
	CRC-C	52	FFPE, H&E	52	~26k	32:20
	CRC-M	47	FFPE, H&E	47	~23.5k	27:20
	External test set	196	FFPE, H&E	196	~79k	176:20
	TCGA	95	FFPE, H&E	95	~34k	85:10

MSS, microsatellite stable; MSI, microsatellite instability; FFPE, formalin-fixed paraffin-embedded; H&E, hematoxylin and eosin

Supplementary table 2. Univariate logistic regression analysis on the CRC-TEST set for microsatellite instability prediction

Covariates	HR (95% CI)	*P* value
Age ≥ 50 years	1.659 (0.866–3.177)	0.127
Right-sided location	4.365 (2.191–8.695)	<0.001
Tumor size of ≥ 6 cm	2.469 (1.286–4.739)	0.007
Poorly differentiation	5.56 (2.77–11.16)	<0.001
Mucinous or signet ring cell carcinoma	45.62 (5.89–353.44)	<0.001
Medullary carcinoma	6.40 (1.64–25.03)	0.008
Expansile growth	9.81 (4.18–22.98)	<0.001
eTILs	3.47 (1.83–6.58)	<0.001
Crohn's-like lymphocytic reaction	1.43 (0.76–2.67)	0.265
Necrosis	0.17 (0.09–0.33)	<0.001
Neoadjuvant chemoradiation therapy	0.35 (0.12–1.08)	0.067
Metastatic colorectal cancer	0.81 (0.28–2.38)	0.700
MSI-PCR	10.74 (5.31–21.73)	<0.001

TILs, tumor-infiltrating lymphocytes; MSI, microsatellite instability; PCR, polymerase chain reaction; HR, hazard ratio; CI, confidential interval.

Supplementary table 3. Multivariate logistic regression analysis with stepwise proportion of mucinous or signet ring cell carcinoma for microsatellite instability prediction

	CTC-TEST set		External test set*	
Covariates	**HR (95% CI)**	***P* value**	**HR (95% CI)**	***P* value**
Poorly differentiation	0.93 (0.23–3.73)	0.933	-	-
Medullary carcinoma	8.36 (0.65–107.0)	0.102	-	-
Expansile growth	3.45 (0.75–15.79)	0.111	2.64 (0.82–8.48)	0.103
eTILs	1.26 (0.38–4.22)	0.710	1.88 (0.78–4.57)	0.161
Crohn's-like lymphocytic reaction	0.38 (0.12–1.17)	0.091	-	-
Necrosis	1.07 (0.32–3.63)	0.913	0.38 (0.15–0.92)	0.033
Mucinous or signet ring cell carcinoma				
<5%	reference	-	reference	-
≥5% & <50%	1.14 (0.29–4.44)	0.849	2.76 (0.97–7.86)	0.057
≥50%	18.60 (1.34–257.64)	0.029	5.11 (1.02–26.68)	0.048
Neoadjuvant chemoradiation therapy	0.03 (0.00–0.26)	0.026	-	-
Metastatic colorectal cancer	0.01 (0.00–0.44)	0.013	-	-
MSI-PCR	21.30 (6.08–74.60)	<0.001	24.36 (4.72–125.79)	<0.001

TILs, tumor-infiltrating lymphocytes; MSI, microsatellite instability; PCR, polymerase chain reaction; HR, hazard ratio; CI, confidential interval.

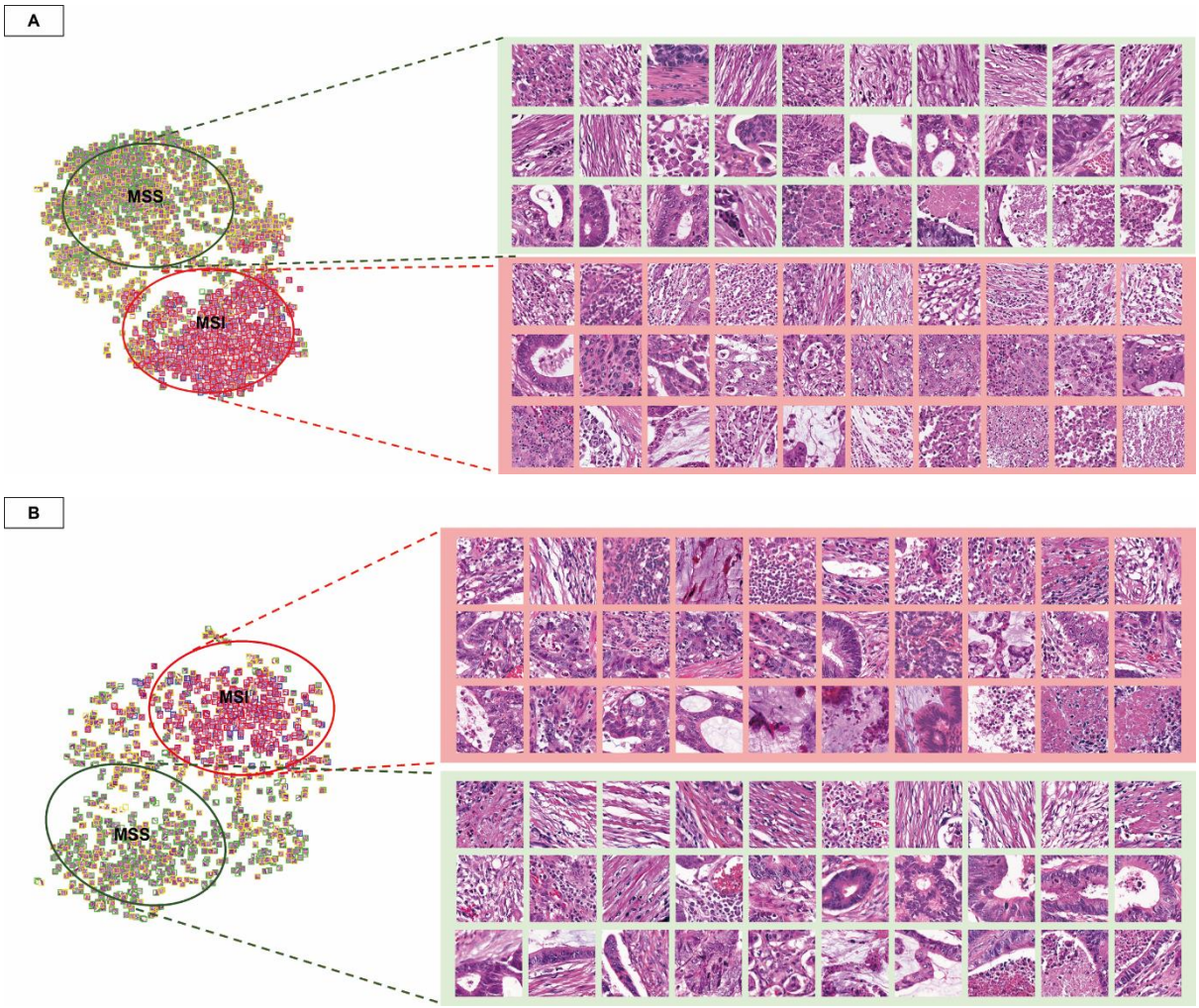
*Variables including poorly differentiated, medullary carcinoma, Crohn's-like lymphocytic reaction, neoadjuvant chemoradiation therapy, and metastatic colorectal cancer were excluded from the multivariate analysis due to an insufficient number of positive samples.

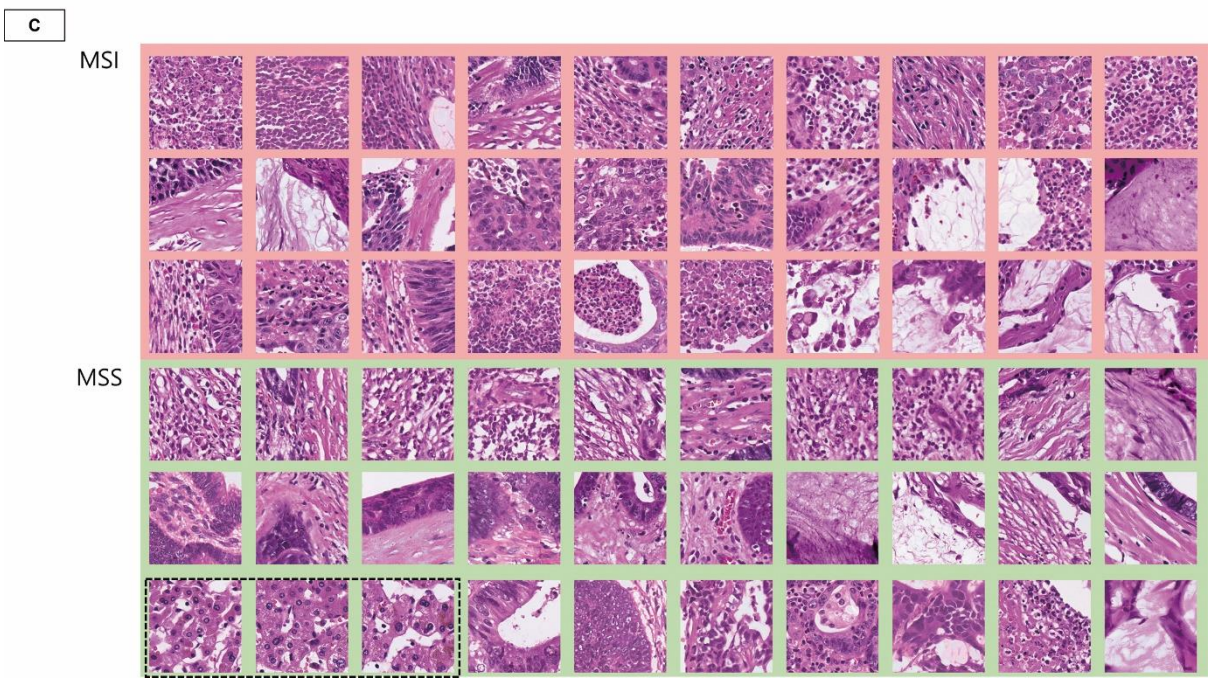
Supplementary table 4. Performance of the deep learning model for microsatellite instability prediction with the modified ground truth of targeted next-generation sequencing analysis

	Accuracy (weighted)	Sensitivity	Specificity	AUC
CRC-NCM (n=154)	0.851	0.80	0.87	0.910

AUC, area under the receiver operating characteristic curve

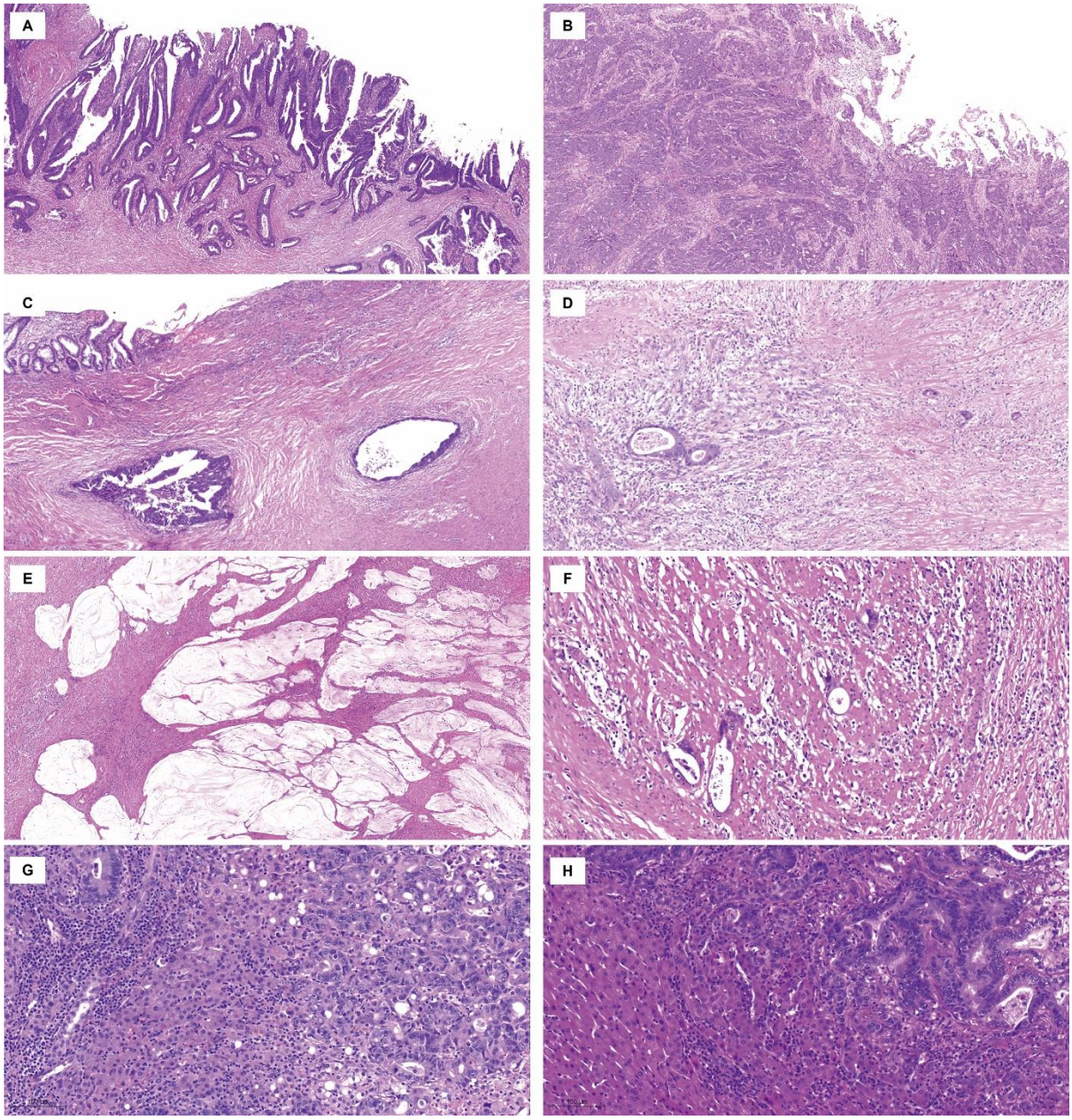
2.2. Supplementary figures





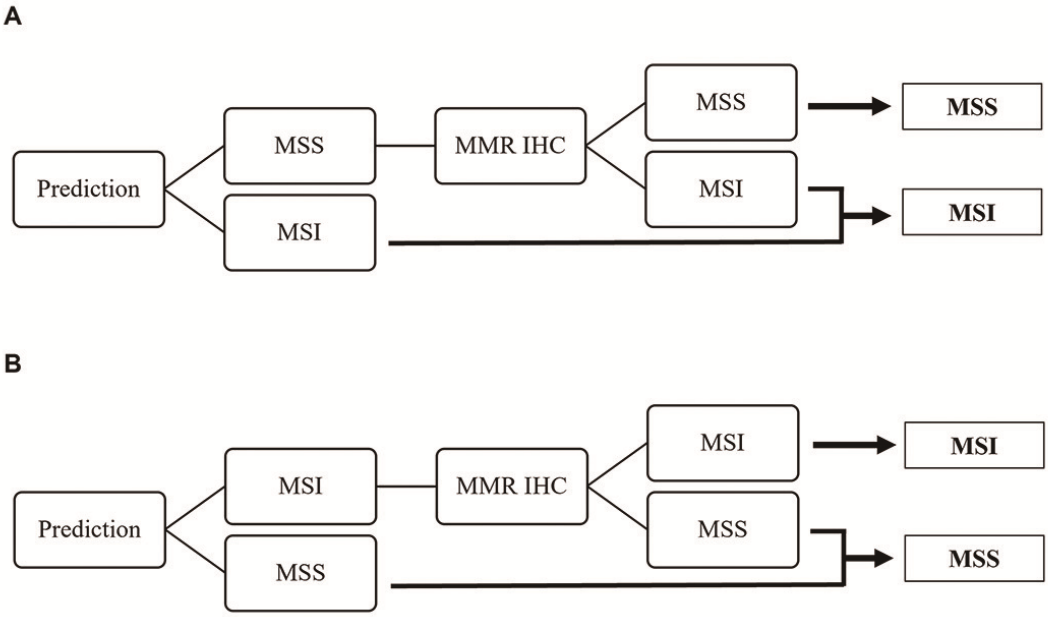
Supplementary Figure 1. Representative patch images extracted from the predicted MSS and MSI groups on (*A*) the CRC-test set, (*B*) external test set, and (*C*) CRC-C/CRC-M sets.

The presence of poorly differentiated areas, mucinous components, and infiltration of stromal inflammatory cells including mononuclear cells, neutrophils, and eosinophils, were pronounced in cases predicted to be MSI than in those predicted to be MSS. Several extracted patches of normal hepatic parenchyma were identified in cases predicted to be MSS in the CRC-C set (*C*, black box). MSS, microsatellite stable; MSI, microsatellite instability



Supplementary Figure 2. Representative histological H&E images of colorectal cancer in the (*A* and *B*) CRC-NCM, (*C*–*F*) CRC-C, and (*G* and *H*) CRC-M test sets.

Compared with the images of the CRC-NCM set (*A* and *B*, x50 magnification), residual tumor cells with neoadjuvant chemoradiation therapeutic changes exhibit stromal fibroinflammatory reaction with replacement of the tumor cells (*C*, x50 magnification; *D*, x100 magnification), predominant colloid change (*E*, x100 magnification), or nuclear atypia with eosinophilic cytoplasmic change (*F*, x400 magnification). Metastatic tumor cells (right side) and peripheral remaining hepatic parenchyma (left side; *G* and *H*, x200 magnification).



Workflow	Sensitivity	Specificity
Panel *A*	0.94	0.86
Panel *B*	0.69	1
Model's prediction, only	0.78	0.88

Supplementary Figure 3. Flow diagram showing the two panels predicting microsatellite status in colorectal cancer using a combination of model's prediction and MMR IHC.

On the upper panel "*A*", cases with predicted MSS and retained MMR expression are classified as the MSS group, and cases with predicted MSI or loss of MMR expression are classified as the MSI group. On the lower panel "*B*", cases with predicted MSS or retained MMR expression are classified as the MSS group, and cases with predicted MSI and loss of MMR expression are classified as the MSI group. Panel "*A*" shows a higher sensitivity of 0.94, and panel "*B*" shows a higher specificity of 1.0 than those of the model's prediction alone.

3. References

1 Boland, C. R. *et al.* A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* **58**, 5248-5257 (1998).

2 Loukola, A. *et al.* Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC). *Cancer Res* **61**, 4545-4549 (2001).

3 Gibson, J., Lacy, J., Matloff, E. & Robert, M. Microsatellite instability testing in colorectal carcinoma: a practical guide. *Clin Gastroenterol Hepatol* **12**, 171-176.e171, doi:10.1016/j.cgh.2013.11.001 (2014).

4 Jenkins, M. A. *et al.* Pathology features in Bethesda guidelines predict colorectal cancer microsatellite instability: a population-based study. *Gastroenterology* **133**, 48-56, doi:10.1053/j.gastro.2007.04.044 (2007).

5 Schrock, A. B. *et al.* Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. *Ann Oncol* **30**, 1096-1103, doi:10.1093/annonc/mdz134 (2019).

6 Park, H. *et al.* Detection of chromosome structural variation by targeted next-generation sequencing and a deep learning application. *Sci Rep* **9**, 3644, doi:10.1038/s41598-019-40364-5 (2019).

7 Marée, R. Open Practices and Resources for Collaborative Digital Pathology. *Front Med (Lausanne)* **6**, 255, doi:10.3389/fmed.2019.00255 (2019).

8 Reinhard, E., Adhikhmin, M., Gooch, B. & Shirley, P. Color transfer between images. *IEEE Computer graphics and applications* **21**, 34-41 (2001).

9 Macenko, M. *et al.* in *2009 IEEE international symposium on biomedical imaging: from nano to macro.* 1107-1110 (IEEE).

10 He, K., Zhang, X., Ren, S. & Sun, J. in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 770-778.

11 Deng, J. *et al.* in *2009 IEEE conference on computer vision and pattern recognition.* 248-255 (Ieee).

12 Zaheer, M. *et al.* Deep sets. *Advances in neural information processing systems* **30** (2017).

13 Fleming, M., Ravula, S., Tatishchev, S. F. & Wang, H. L. Colorectal carcinoma: Pathologic aspects. *Journal of gastrointestinal oncology* **3**, 153 (2012).

14 Greenson, J. K. *et al.* Pathologic predictors of microsatellite instability in colorectal cancer. *The American journal of surgical pathology* **33**, 126 (2009).

15 Jakubowska, K., Kisielewski, W., Kańczuga‐Koda, L., Koda, M. & Famulski, W. Stromal and intraepithelial tumor‐infiltrating lymphocytes in colorectal carcinoma. *Oncology letters* **14**, 6421-6432 (2017).

16 Harbaum, L. *et al.* Peritumoral eosinophils predict recurrence in colorectal cancer. *Modern*

Pathology **28**, 403-413 (2015).