Supplementary materials

- **Supplementary Methods**
- **Supplementary Tables**
- **Supplementary Figure**

Supplementary Methods

Tissue processing and genomic DNA extraction

Formalin-fixed paraffin-embedded (FFPE) tissue sections were evaluated for tumor cell content using hematoxylin and eosin (H&E) staining. Only samples with a tumor content of ≥20% were eligible for subsequent analyses. FFPE tissue sections were placed in a 1.5 microcentrifuge tube and deparaffinized with mineral oil. Samples were incubated with lysis buffer and proteinase K at 56 ° C overnight until the tissue was completely digested. The lysate was subsequently incubated at 80 °C for 4 hours to reverse formaldehyde crosslinks. Genomic DNA was isolated from tissue samples using the ReliaPrep™ FFPE gDNA Miniprep System (Promega) and quantified using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific) following the manufacturer's instructions.

Blood sample processing and cfDNA isolation

Blood samples were centrifuged in Streck tubes within 2 hours of collection at 1,600 g at 4°C for 10 minutes. About 5 ml plasma supernatant was transferred to a new 5 ml microfuge tube and centrifuged at 16,000 g at 4 ℃ for 10 minutes to remove residual cells and debris. Supernatant was transferred into a new tube, followed by cfDNA extraction using the QiAmp Circulating Nucleic Acid Kit (Qiagen) following the manufacturer's instructions. DNA concentration was quantified with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Genomic DNA (gDNA) from white blood cells was extracted using the QIAamp DNA Mini Kit (Qiagen).

Library preparation and targeted capture

DNA extracts (30-200 ng) were sheared to 250 bp fragments using an S220 focused-ultrasonicator (Covaris). Libraries were prepared using the KAPA Hyper Prep Kit (KAPA Biosystems) following the manufacturer's protocol. The concentration and size distribution of each library were determined using a Qubit 3.0 fluorometer (Thermo Fisher Scientific) and a LabChip GX Touch HT Analyzer (PerkinElmer) respectively.

For targeted capture, indexed libraries were subjected to probe-based hybridization with

a customized NGS panel targeting exons of 733 cancer-related genes, where the probe baits were individually synthesized 5′ biotinylated 120 bp DNA oligonucleotides (IDT). The xGen® Hybridization and Wash Kit (IDT) was employed for hybridization enrichment. Briefly, 500 ng indexed DNA libraries were pooled to obtain a total amount of 2 μg of DNA. The pooled DNA sample was then mixed with human cot DNA and xGen Universal Blockers-TS Mix and dried down in a SpeedVac system. The Hybridization Master Mix was added to the samples and incubated in a thermal cycler at 95℃ for 10 min, before being mixed and incubated with 4 μl of probes at 65℃ overnight. The target regions were captured following the manufacturer's instructions. The concentration and fragment size distribution of the final library were determined using a Qubit 3.0 fluorometer (Thermo Fisher Scientific) and a LabChip GX Touch HT Analyzer (PerkinElmer) respectively.

DNA sequencing, data processing, and variant calling (for tissue-based testing)

The captured libraries were loaded onto a NovaSeq 6000 platform (Illumina) for 100 bp paired-end sequencing.Raw data of paired samples (an FFPE sample and its normal tissue control) were mapped to the reference human genome hg19 using the Burrows-Wheeler Aligner (v0.7.12). PCR duplicate reads were removed and sequence metrics were collected using Picard (v1.130) and SAMtools (v1.1.19), respectively. Variant calling was performed only in the targeted regions. Somatic single nucleotide variants (SNVs) were detected using an in-house developed R package to execute a variant detection model based on binomial test. Local realignment was performed to detect indels. Variants were then filtered by their unique supporting read depth, strand bias, base quality as previously described. All variants were then filtered using an automated false positive filtering pipeline to ensure sensitivity and specificity at an allele frequency (AF) of ≥ 1%. Single-nucleotide polymorphism (SNPs) and indels were annotated by ANNOVAR against the following databases: dbSNP (v138), 1000Genome and ESP6500 (population frequency > 0.015). Only missense, stopgain, frameshift and non-frameshift indel mutations were kept. Copy number variations (CNVs) and gene rearrangements were detected as described previously. Germline variants were

identified using paired normal tissues adjacent to the cancer. Pathogenic or likely pathogenic mutation was determined as per a joint consensus of the latest literature or reports from clinical trials and the recommendation of American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP).

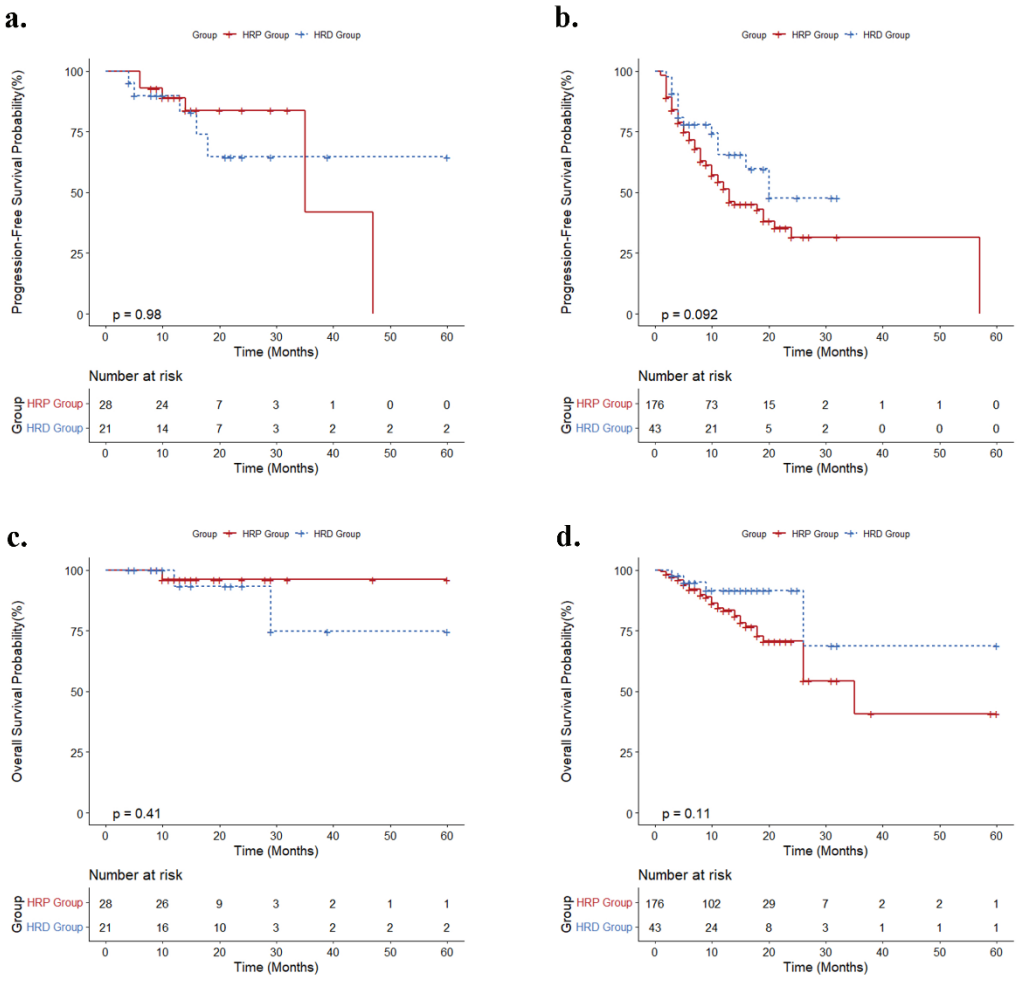
Supplementary Table 1. HRD-related genes.

No.	**Gene**
1	*ABL1*
2	*ATM*
3	*ATR*
4	*BAP1*
5	*BARD1*
6	*BLM*
7	*BRCA1*
8	*BRCA2*
9	*BRIP1*
10	*CDK12*
11	*CHEK1*
12	*CHEK2*
13	*DNMT3A*
14	*ERCC1*
15	*ERCC4*
16	*FANCA*
17	*FANCC*
18	*FANCD2*
19	*FANCE*
20	*FANCF*
21	*FANCG*
22	*FANCL*
23	*MRE11*
24	*NBN*
25	*NONO*
26	*PALB2*
27	*RAD50*
28	*RAD51*
29	*RAD51B*
30	*RECQL4*
31	*RMI2*
32	*SFPQ*
33	*WRN*

Supplementary Table 2. Univariate and multivariable logistic regression analysis of predictors for the risk of HRD group.

Variable	Univariable			Multivariable		
	OR	**95% CI**	***P***	**OR**	**95% CI**	***P***
Gender (female vs. male)	1.15	0.65-2.05	0.64	\	\	\
Age (>60 vs. ≤60 years)	0.53	0.28-1.00	0.05	\	\	\
BMI						
<18.5	reference	\	\	\	\	\
18.5-24	0.38	0.13-1.10	0.08	\	\	\
>24	0.36	0.12-1.09	0.07	\	\	\
Differentiation (Moderately& Well &Chronic inflammation after neoadjuvant therapy vs. Poorly)	0.54	0.27-1.06	0.08	\	\	\
Pathological Type						
Chronic inflammation after neoadjuvant therapy	reference	\	\	\	\	\
Adenocarcinoma	0.25	0.06-1.03	0.06	\	\	\
Special histologic type adenocarcinoma	1.33	0.23-7.63	0.75	\	\	\
Mixed	0.43	0.09-2.10	0.30	\	\	\
Tumor location						
Left Colon	reference	\	\	reference	\	\
Right Colon	2.60	1.42-4.75	< 0.01*	1.80	0.85-3.80	0.13
Mixed	2.15	0.19-24.33	0.54	0.42	0.01-32.61	0.70
Tumor size (<4 cm vs. >=4 cm)	0.72	0.40-1.29	0.27	\	\	\
AJCC stage (III& IV vs. I& II)	0.33	0.17-0.63	< 0.01*	0.66	0.27-1.59	0.35
MSI status (MSI-H vs. MSS)	138.90	18.29-1054.52	< 0.01*	118.56	14.52-968.10	< 0.01*

OR: Odd Ratio; 95%CI: 95% Confidence interval; MSI: Microsatellite Instability; *: significant difference.



Supplementary Figure 1 Kaplan-Meier curves of PFS and OS in HRD/HRP cohort. Survival curves were compared, and P values were determined using a 2-sided log-rank test. A& C: patients with stage I and II CRC. B& D: patients with stage III and IV CRC. PFS: Progression-Free Survival. OS: Overall Survival. HRD: Homologous Recombination Deficiency. HRP: Homologous Recombination Proficient. CRC: Colorectal Cancer.