

Research Article

Long-Read Nanopore Sequencing Identifies Mismatch Repair-Deficient Related Genes with Alternative Splicing in Colorectal Cancer

Hao Qu,¹ Zhenjun Wang ,¹ Yudong Zhang ,¹ Baocheng Zhao,¹ Shuai Jing,¹ Jianwei Zhang,¹ Chunxiang Ye ,¹ Yaohan Xue,^{1,2} and Lei Yang ^{1,3}

¹Department of General Surgery, Beijing Chao-Yang Hospital, Capital Medical University, Beijing 100020, China

²Department of Colorectal and Anal Surgery, Beijing Rectum Hospital, Beijing 100120, China

³Medical Research Center, Beijing Chao-Yang Hospital, Capital Medical University, Beijing 100020, China

Correspondence should be addressed to Zhenjun Wang; drzhenjun@163.com and Lei Yang; yl6649084@163.com

Received 2 February 2022; Revised 15 June 2022; Accepted 23 June 2022; Published 21 July 2022

Academic Editor: Zhijie Xu

Copyright © 2022 Hao Qu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Alternative splicing (AS) plays a crucial role in regulating the progression of colorectal cancer (CRC), but its distribution remains to be explored. Here, we aim to investigate the genes edited by AS which show differential expression in patients with mismatch repair deficiency (dMMR)/microsatellite instability (MSI). **Materials and Methods.** We applied long-read nanopore sequencing to determine the mRNA profiles and screen AS genes using Oxford Nanopore Technologies (ONT) method in ten paired CRC tissues. CRC tissue and plasma samples were used to validate the differential genes with AS using real-time fluorescent quantitative PCR, immunohistochemistry, and enzyme-linked immunosorbent assay. **Results.** ONT sequencing identified 404 genes were downregulated, and 348 genes were upregulated in MSI cancer tissues compared with microsatellite stability (MSS) cancer tissues. In total, 6,200 AS events were identified in 2,728 mRNA transcripts. WGCNA revealed dMMR/MSI-correlated gene modules, including INHBA and RPL22L1, which were upregulated; conversely, HMGCS2 was downregulated in MSI cancer. Overexpression of RPL22L1, INHBA, and CAPZA1 was further confirmed in CRC tissues. INHBA was found to be associated with tumor lymphatic metastasis. Importantly, the levels of INHBA in CRC plasma were significantly increased compared with those in noncancer plasma. INHBA showed a higher level in dMMR/MSI CRC than in MSS CRC, indicating that INHBA is a useful biomarker. **Conclusion.** Our results showed that ONT-identified genes provide a pool to explore AS-associated markers for dMMR/MSI CRC. We demonstrated INHBA as a promising signature for clinical application in predicting tumor lymphatic metastasis and screening dMMR/MSI candidates.

1. Introduction

Colorectal cancer (CRC) is one of the most important malignant gastrointestinal tumors and has demonstrated high rates of mortality and morbidity [1]. CRC is classified as mismatch repair deficient (dMMR) or microsatellite instability (MSI) by the absence of at least one of four mismatch repair proteins (MLH1, MSH2, MSH6, or PMS2) [2]. The characteristics of MSI CRC include a high tumor mutational burden, generating

hyperneoantigens that are responsible for favorable clinical responses to immune checkpoint blockade (ICB) [3]. However, approximately 10~15% of CRC cases display dMMR/MSI [4].

Currently, dMMR/MSI screening is helpful to select the most appropriate ICB for CRC and is also used as a biomarker to predict the therapeutic response to chemotherapy [5]. Patients with metastatic CRC (mCRC) that are high in dMMR/MSI are less responsive to conventional chemotherapy

with or without drugs targeting VEGF or EGFR [6]. The KEYNOTE-177 study showed that pembrolizumab (anti-PD1) monotherapy provides a clinically meaningful improvement in progression-free survival (PFS) compared with chemotherapy for dMMR/MSI mCRC (median PFS: 16.5 months vs. 8.2 months) [7]. Additionally, mCRC patients with dMMR/MSI treated with first-line pembrolizumab showed better health-related quality of life than those treated with chemotherapy [2].

Cancer-associated mutations in genes encoding RNA alternative splicing (AS), such as MLH1, commonly occur in CRC [8]. AS is a rich source of tumor-specific neoantigen targets for ICB [9]. To some extent, MSI patients can benefit from ICB depending on large numbers of immunogenic neoantigens that can be recognized by immune cells [10]. Mutations in MMR proteins confer an increased lifetime risk of cancer development in affected individuals, while neoantigens result from the hypermutable nature of dMMR [11]. AS is partly responsible for transcript variation and proteome diversity [2]. MLH1 deficiency triggers chromosomal abnormalities and the release of nuclear DNA into the cytoplasm, resulting in the activation of the cyclic GMP-AMP synthase- (cGAS-) STING pathway, whose key molecules undergo AS [12]. cGAS acts as a specific pattern recognition receptor and is activated by type I interferons (IFNs), which mediate the Janus kinase- (JAK-) signal transducer and activator of transcription (STAT) pathway [13]. Through AS, STAT3 is translated into two distinct isoforms, STAT3 α and STAT3 β [14]. Therefore, identification dMMR-related AS genes would be helpful on either illustrating novel mechanism for ICB or screening biomarker for predicting therapeutic effect.

Comparing with microsatellite stability (MSS) CRCs, dMMR CRCs exhibited distinct immunological feature. Activation of cGAS-STING and type I IFN signaling promotes the overexpression of CCL5 and CXCL10 and recruitment CD8⁺ T cells into dMMR CRC TME [15]. CXCL10 is upregulated in homologous recombination-deficient tumors and binds to CXCR3 [16]. CXCR3 is primarily expressed on activated CD8⁺ T cells and is suppressed by transforming growth factor beta (TGF β) [17]. TGF β signaling is the major pathway that excludes immune cells from tumors. MSI CRC is unresponsive to TGF β because of TGFBR2 mutations [18]. Thus, different molecular patterns exist between MSI and MSS patients, but their identities remain elusive.

High-throughput sequencing technology has become increasingly available to identify differential mRNA profiles or genomic variations. Nanopore sequencing technology with highly accurate quantification of transcripts is a third-generation sequencing method recently developed to sequence full-length RNA transcripts as well as distinct AS transcripts [19]. In the present study, we performed transcriptome sequencing to survey the full-length CRC transcriptome landscape offered by Oxford Nanopore Technologies (ONT). We further compared the differential profiles between MSI and MSS tumors. Our data provide the first full-length MSI and MSS CRC transcriptome databases, allowing us to identify INHBA as biomarkers to discern among MSI CRC patients using blood plasma.

2. Materials and Methods

2.1. Patient Selection. For ONT RNA sequencing, 20 fresh samples were collected from 10 CRC patients from 2020 to 2021 at Beijing Chao-Yang Hospital. These CRC patients enrolled five dMMR/MSI patients. dMMR/MSI was defined when any of these MMR proteins were completely absent in the cancer tissue but presented in adjacent benign tissue. Microsatellite stability (MSS) patients were defined when all of MMR (MLH1, MSH2, MSH6, and PMS2) were positive expression in both tumor and adjacent benign tissue. Immunohistochemistry (IHC) was assayed to detect MLH1, MSH2, MSH6, and PMS2 separately. Patients who showed dMMR were defined as MSI. For validation, 23 paired tumor samples and tumor-adjacent normal fresh tissues were collected and stored at -80°C until to use. Forty-six normal intestinal tissues and 49 cancer samples were formalin-fixed and formalin-embedded and reassessed by two pathologists. Sixty-six peripheral plasma samples from CRC patients and 20 plasma samples from non-tumor donors were separated and stored at -80°C until use. This study was approved by the Institutional Review Board at Beijing Chao-Yang Hospital. All the enrolled subjects signed written informed consent forms.

2.2. ONT Sequencing. One microgram of total RNA was prepared for cDNA libraries using a cDNA-PCR Sequencing Kit (SQK-PCS109; ONT). The products were then subjected to ONT adaptor ligation using T4 DNA ligase (NEB). The final cDNA libraries were added to FLO-MIN109 flow cells and run on the PromethION platform at Biomarker Technology Company (Beijing, China). Raw reads were first filtered (minimum average read quality score = 7 minimum read length = 500 bp). Ribosomal RNAs were discarded after mapping to the rRNA database. Clusters of full-length, nonchimeric (FLNC) transcripts and consensus isoforms were obtained after mapping to the reference genome (GRCh38_release95) with minimap2 (version 2.7-r654) in spliced alignment mode after polishing within each cluster using pinfish [20]. Mapped reads were further collapsed by the cDNA_Cup cake package with min-coverage = 85% and min-identity = 90%. The criteria for fusion candidates are that a single transcript must map to ≥ 2 loci (minimum coverage 5%, ≥ 1 bp; total coverage $\geq 95\%$, at least 10 kb). The transcripts were validated against known reference transcript annotations (hg38) with Gffcompare (version 0.1.26) [21]. Alternative splicing (AS) events, including IR (intro retention), ES (exon skipping), AA (Alternative acceptor), and MEE (mutually exclusive exon), were identified using the AStalavista tool (version 3.0) [22]. Simple Sequence Repeat (SSR) of the transcriptome were identified using MICOsaTellite identification tool (version 2.1) [23]. Alternative polyadenylation (APA) analysis was conducted with TAPIS (Transcriptome Analysis Pipeline for Isoform Sequencing, version 1.2.1) [24]. Coding Sequence (CDS) were predicted by TransDecoder (version 5.5.0) [25]. The data were submitted to the NCBI SRA database and are accessible through the citation accession number PRJNA758110. For quantification, the expression levels were estimated by reads per gene/transcript per 10,000 reads mapped. Differential expression analysis was performed using the DESeq2 R package (version 1.6.3). These bioinformatic

analyses were performed by Biomarker Technology Company. Genes with an adjusted P value < 0.01 and fold change (FC) ≤ -1.5 or ≥ 1.5 found by DESeq2 were assigned as differentially expressed. Gene Ontology (GO) enrichment analysis of the differentially expressed genes (DEGs) was implemented using the DAVID Bioinformatics Resources (version 6.8) tool [26] and plotted using GOplot R packages (version 1.0.2) [27]. GO terms included biological process (BP), cellular component (CC), and molecular function (MF). The top eight terms of biological process (BP) were statistically analyzed according to P value (< 0.05) and FC ≤ -1.5 or ≥ 1.5 .

2.3. TCGA Dataset Validation. COAD (enrolled 41 normal tissue samples and 457 cancer tissue samples) and READ (enrolled 10 normal tissue samples and 166 cancer tissue samples) datasets were downloaded from The Cancer Genome Atlas (TCGA) data portal (<http://www.cbiportal.org/>). The R package “limma” [28] was used to screen for the differentially expressed genes between cancer and normal samples according to the criteria of an adjusted P value < 0.01 and fold change (FC) ≥ -1.5 or ≥ 1.5 .

2.4. Estimating the Proportion of Immune and Cancer (EPIC) and CIBERSORT Analysis. EPIC and CIBERSORT were used to estimate the fraction of immune cells of sequenced samples. EPIC was performed using an online tool (<http://epic.gfellerlab.org/>). The R package “CIBERSORT” (R script version 1.03) was applied to assess the fractions of the 22 types of tumor-infiltrating immune cells [29]. Outputs were considered to be accurate after screening to meet P value < 0.05 .

2.5. Reverse Transcription Quantitative Real-Time PCR (RT-qPCR). Total RNA was purified from 46 samples using TRIzol Reagent (Invitrogen, CA, USA). A QuantiTect Reverse Transcription kit (Qiagen, Germany) was used for reverse transcription. An ABI7500 Real-Time PCR System (Applied Biosystems, USA) was used to detect the expression of candidate genes. The primers verified in this study are shown as follows:

- (1) INHBA: forward—5'-CAGTGCCAATACCATG AAGAGGA-3'; reverse—5'-ATGCAAACTAGGG AAGAGAACCC-3'
- (2) RPL22L1: forward—5'-AAAGACAGGAAGCCCA AGAGG-3'; reverse—5'-AGGGCATCCTGGCAGC AAAA-3'
- (3) CAPZA1: forward—5'-CGAATGAAGCCCAAAC TGCC-3'; reverse—5'-TTTGTTGCGGGTAACT GGAA-3'
- (4) HMGCS2: forward—5'-AGTGGTAATGCTCGTC CCAC-3'; reverse—5'-ATATGGGTTCCCCTCA GCCC-3'
- (5) 18S (reference gene): forward—5'-AAACGGCTA CCACATCCA-3'; reverse—5'-CACCAGACTTG CCCTCCA-3'

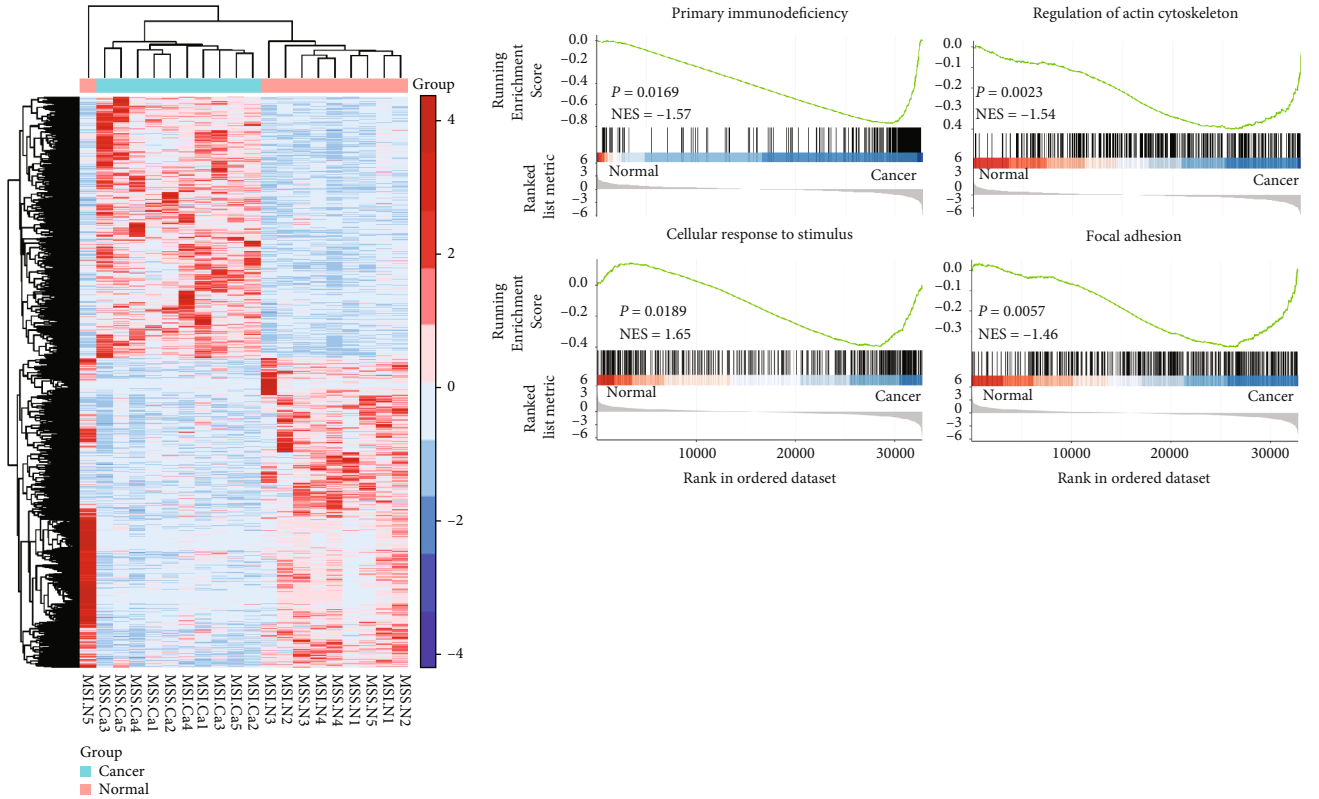
2.6. IHC Assays. The primary antibodies used in IHC for tissue slides were as follows: anti-INHBA (Proteintech; #10651-1-AP; 1:50 dilution), anti-RPL22L1 (Cusabio; #CSB-PA740724A01HU; 1:50 dilution), and anti-CAPZA1 (CSB-PA004510LA01HU; 1:50 dilution). The slice was deparaffinized and rehydrated, pretreated with citric acid antigen retrieval solution (pH = 6.8), and rinsed in PBS. The sections were then blocked in 2% goat serum and incubated with the primary antibody overnight at 4°C. The streptavidin-peroxidase method (ZSGG-BIO, China; #PV9001) was used to show the levels of stained proteins. The percentage of positive cells was scored as follows: 1 (0–25%), 2 (26–50%), 3 (51–75%), and 4 (>75%). The intensity of positive staining was classified into 4 scales as follows: 0 (negative), 1 (weak), 2 (moderate), and 3 (strong). The levels were semiquantitatively determined as percentages multiplied by intensity.

2.7. Enzyme-Linked Immunosorbent Assay (ELISA). The amount of INHBA in blood plasma was quantified using the Human INHBA ELISA Kit (#ml022686; mlbio, China) according to the manufacturer's instructions. Briefly, 50 μ l of plasma was added to 96-well anti-INHBA-coated plates and incubated at 4°C overnight. After that, the samples were removed and washed with washing buffer, and then horseradish peroxidase-labeled antibodies were added and incubated at 37°C for 1 hour. Finally, substrate solution was added and terminated by adding stop solution. Each well was measured at an Optical Density (O.D.) of 450 nm using a microtiter plate reader within 15 minutes. The INHBA concentrations were calculated using a standard curve.

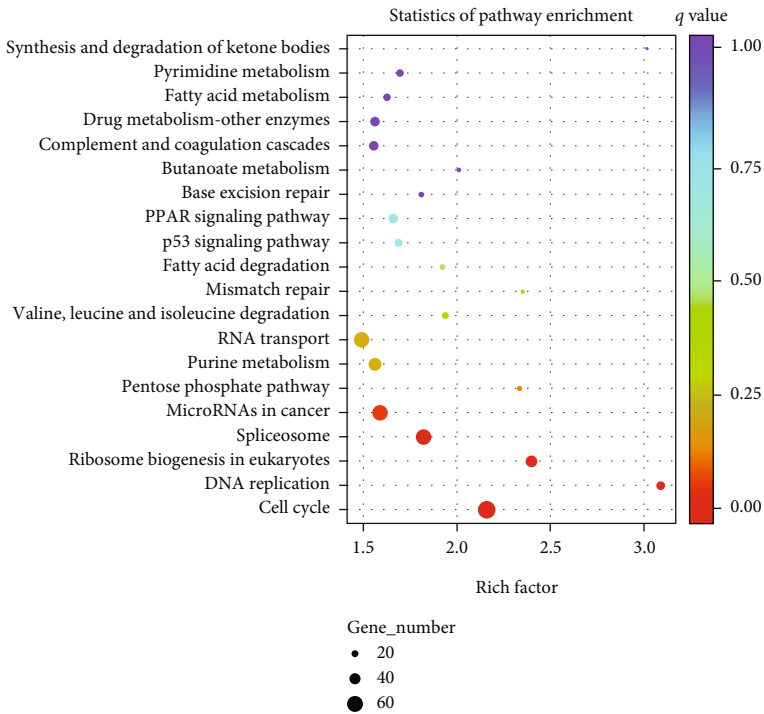
2.8. Statistical Analysis. We used unpaired Student's t -test and the Mann-Whitney test as indicated, and binary outcomes were compared. A P value less than 0.05 was considered statistically significant. Analysis was performed using GraphPad Prism software (version 8.0.1).

3. Results

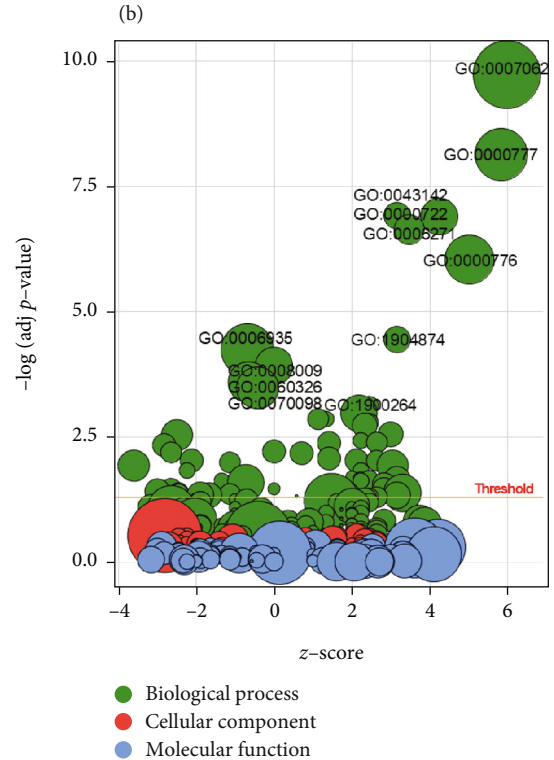
3.1. Long Read RNA Sequencing Establishes CRC-Related Differential Profiles. To investigate the differential mRNA profiles of CRC, we used full-length RNA sequencing by ONT, which is a third-generation RNA-seq technique, and constructed transcriptome landscapes in 20 CRC samples, which were paired normal and cancer tissues (Figure S1, Figure S2, Supplementary Table S1). ONT sequencing identified 8,566 transcripts with fold changes ≤ -1.5 or ≥ 1.5 and a P value < 0.01 , encoding 4,179 genes (Figure 1(a)). The downregulated genes numbered 2,632, and the upregulated genes numbered 1,547 in cancer tissues. Gene set enrichment analysis (GSEA) indicated that these genes are involved in the primary immunodeficiency, regulation of actin cytoskeleton, cellular response to stimulus, and focal adhesion (Figure 1(b)). Additionally, pathway enrichment analysis demonstrated that these genes were significantly enriched in DNA replication and cell cycle pathways (Figure 1(c)). To explore the biological function, Gene Ontology (GO) enrichment analysis was performed, and the eight top GO terms associated with



(a)



(c)



(d)

FIGURE 1: Continued.

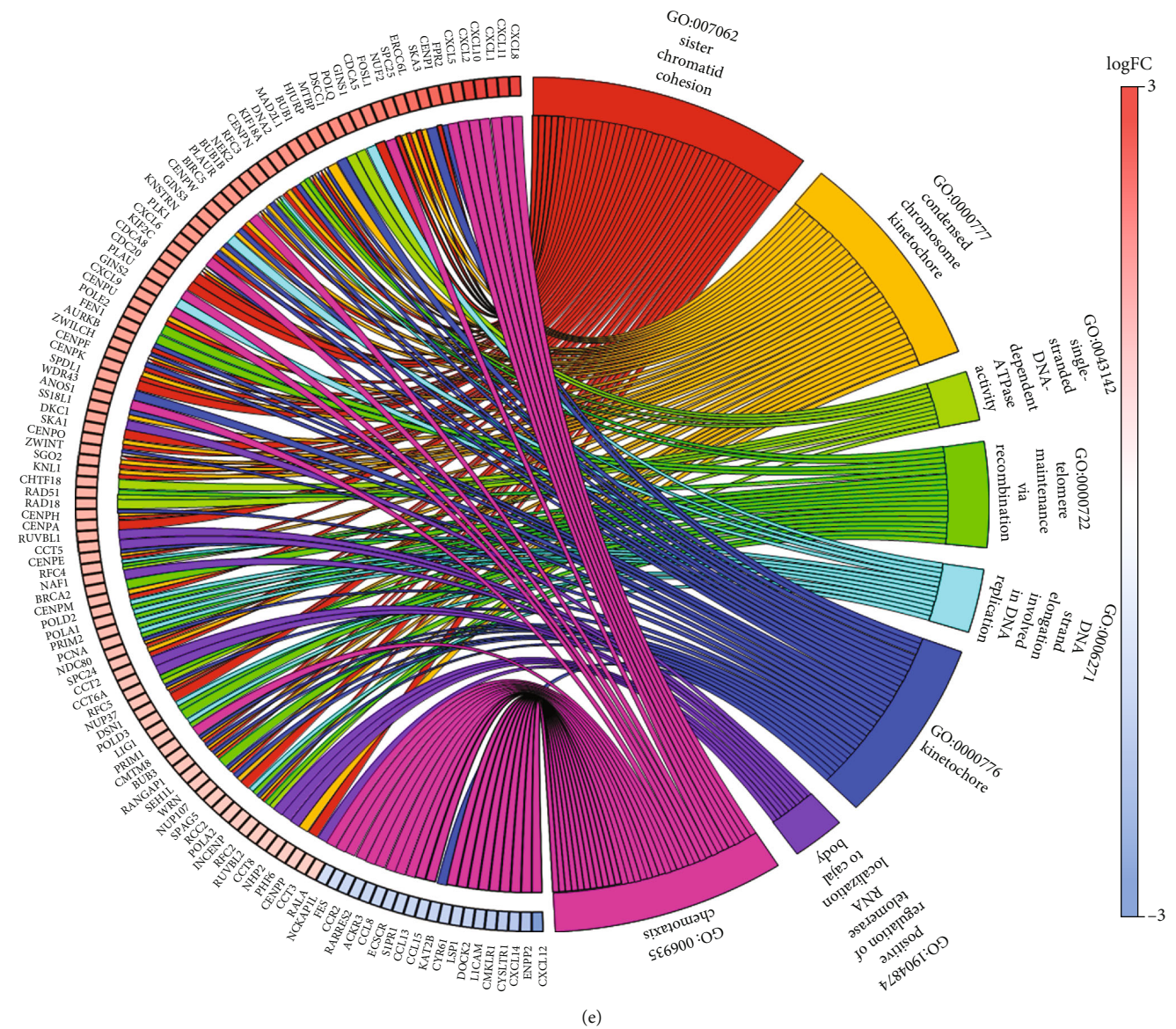


FIGURE 1: Differential mRNA profiles of CRC ONT sequencing reveal the most dysregulated immune-related pathways and genes. (a) Heatmap clusters of the mRNA levels of 4,179 genes in 20 samples. (b) GSEA enriches the pathways for differentially expressed genes between cancer and normal tissues. (c) KEGG analysis investigates the pathway for these differential genes involved. (d). Gene Ontology (GO) identifies the biological process (BP), cellular component (CC), and molecular function (MF) for differentially expressed genes. (e) The top eight BP terms are plotted, and the genes enriched in these terms are shown with fold changes.

immune response and DNA replication were enriched (Figures 1(d) and 1(e)). Several chemokines enriched in chemotaxis terms, including CXCL8/11/1/10/5, were significantly increased in cancer tissues (Figure 1(e)). Our GO analyses enriched many genes associated with sister chromatid cohesion and condensed chromosome kinetochores, which are involved in DNA replication, one of the most important preconditions for the cell cycle (Figure 1(e)).

To validate our findings, we identified differential mRNAs in TCGA colorectal cancer datasets (COAD and READ) (Figure S3A). In total, 5,121 and 4,924 differentially expressed genes were identified in both datasets. Additionally, 1,417 overlapping genes showed differential expression by either ONT sequencing or TCGA analysis (Figure S3B). GO analysis

was also used to enrich the biological processes for COAD differential genes, indicating that these genes were involved in GO terms concerning the chemokine-mediated signaling pathway, cellular response to tumor necrosis factor, and cellular response to interleukin-1 (Figure S3C). For READ differential genes, GO terms concerning the chemokine-mediated signaling pathway, cell chemotaxis, and chemotaxis were enriched as the most important biological processes for these genes involved (Figure S3C). The above results showed that chemokines play a key role in regulating CRC progression.

3.2. *dMMR/MSI Related Genes*. We further addressed the differentially expressed genes between MSI and MSS to screen the biomarkers associated with dMMR. ONT sequencing

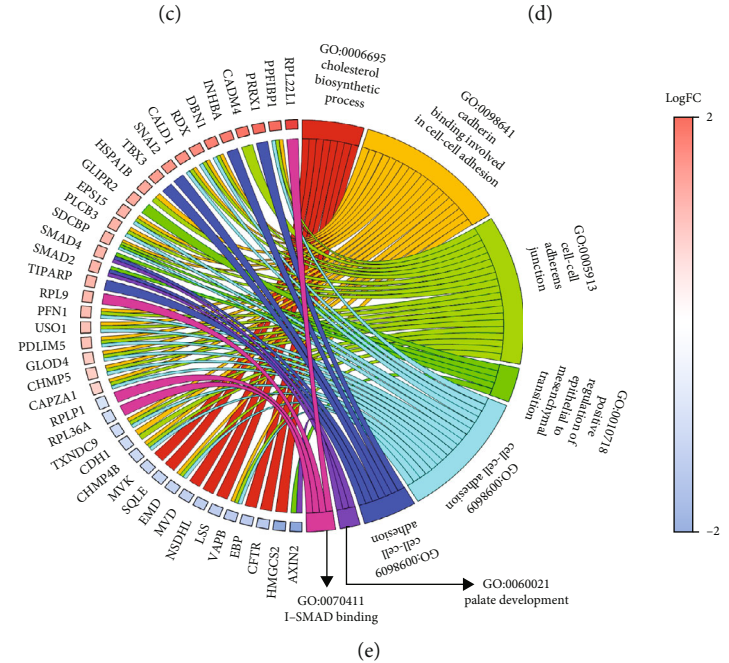
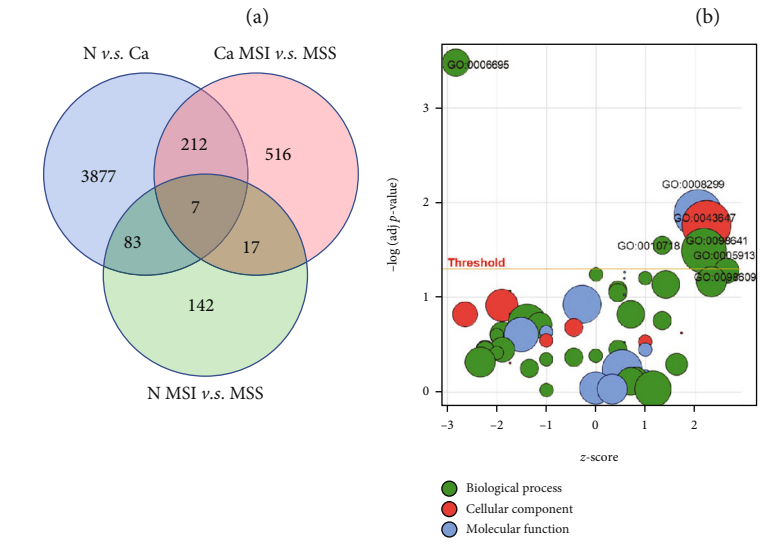
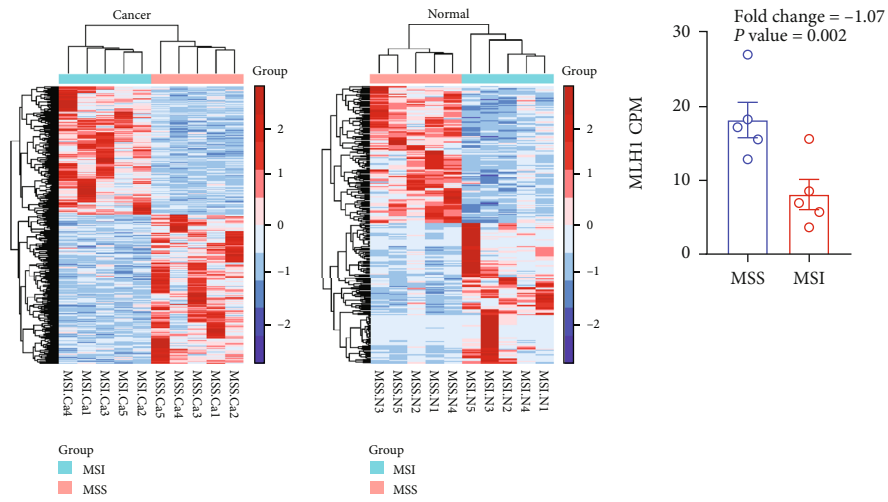


FIGURE 2: Continued.

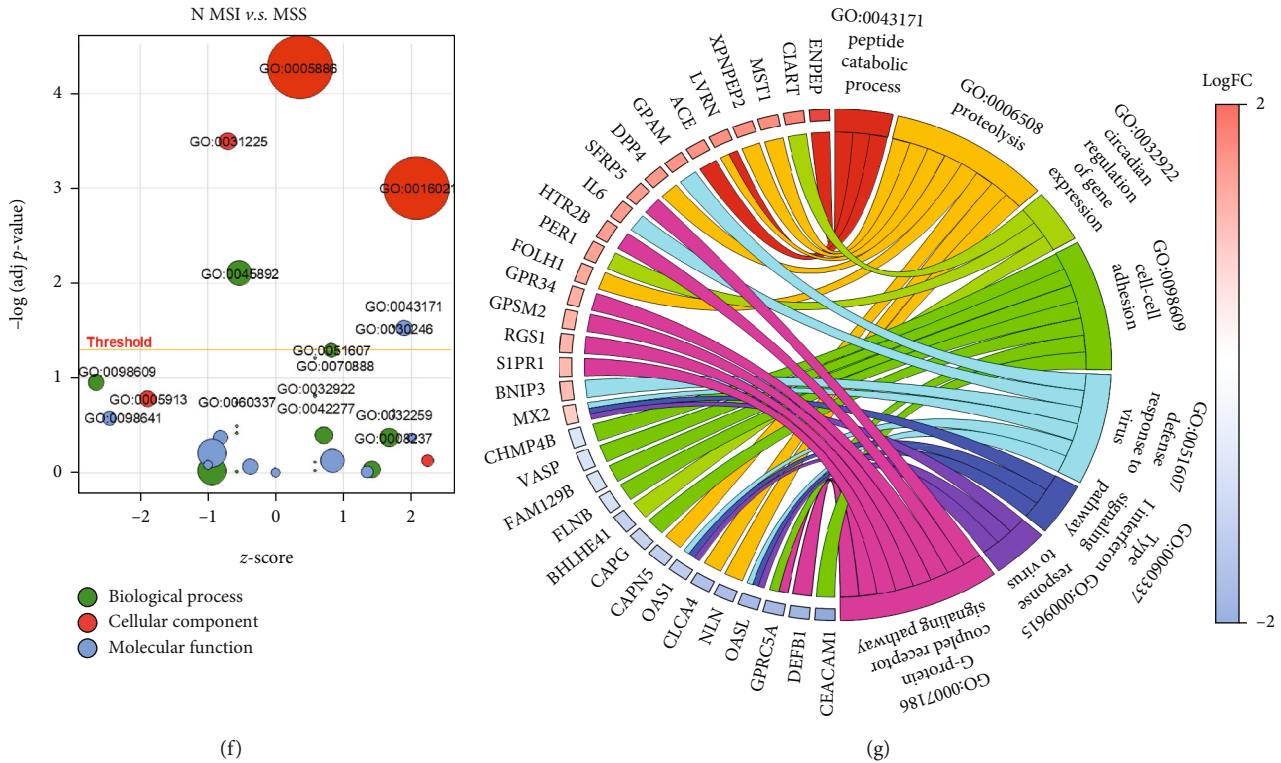


FIGURE 2: Identification of dMMR/MSI-related genes between cancer and normal CRC tissues. (a) Heatmap illustrating the differentially expressed genes in dMMR/MSI and MSS CRC. The dysregulated genes are clustered in cancer (left) and normal (right) tissues. (b) MLH1 is the marker for dMMR, and its levels were compared between MSS ($n = 5$) and MSI ($n = 5$) tissues using ONT sequencing data. (c) Venn diagram analysis of the significant overlap of upregulated or downregulated genes between dMMR/MSI and MSS and comparing these genes with those showing differential levels in cancer (Ca) or normal (N) CRC tissues. (d) GO analysis enriches dMMR/MSI-related genes in cancer tissues. (e) Circos plot displaying the GO BP terms and dMMR/MSS-related genes. (f) GO enrichment shows the genes identified in dMMR/MSI normal tissues. (g) Circos plot delineates the GO BP terms and their enriched genes.

analyzed five MSIs and five MSSs. In cancer tissues, 752 differentially expressed genes were identified; 404 genes were downregulated, and 348 genes were upregulated in MSI cancer tissues compared with MSS cancer tissues (Figure 2(a)). In normal tissues, 249 differential genes were identified in MSI, comprising 126 upregulated genes and 123 downregulated genes (Figure 2(a)). The MLH1 gene is a hallmark of dMMR and was significantly decreased in MSI cancer tissues in our ONT analysis (Figure 2(b)). Two hundred thirty-six overlapping genes with differential expression were identified when comparing cancer-MSI versus (*vs.*) cancer-MSS or cancer *vs.* cancer tissue (Figure 2(c)). Additionally, 107 overlapping genes were detected when comparing normal-MSI *vs.* normal-MSS or cancer *vs.* normal tissues (Figure 2(c)).

To assess the biological features of these differentially expressed genes, we employed GO analysis to feature the terms for MSI-related genes. In cancer tissues, the most significant GO terms for MSI-specific genes included GO:0006695 (cholesterol biosynthetic), GO:0008299 (isoprenoid biosynthetic), and GO:0098609 (cell-cell adhesion), which referred to biological process, molecular functions, and cellular component, respectively (Figure 2(d)). In normal tissues, the most significant GO terms were GO:0005886 (plasma membrane) and GO:0045892 (negative regulation of transcription, DNA-templated), which were classified as cellular component and biological process, respectively (Figure 2(f)). Additionally, both cancer

and normal tissues MSI-related differential genes showed similar GO terms, and most of these genes were involved in metabolic-related processes, such as cholesterol biosynthetic and peptide catabolic process (Figures 2(e) and 2(g)).

3.3. Exploration of MSI-Related Alternative Splice (AS) Events. To identify AS landscape, these transcripts determined using ONT sequencing were aligned to against known reference transcript annotations (hg38, UCSC) [30]. This analysis allows identification of the following 7 types of alternative splice events: (1) alternative 3' splice site, (2) alternative 5' splice site, (3) exon skipping, (4) exon skipping/mutually exclusive exon, (5) intron retention, (6) mutually exclusive exon, and (7) mutually exclusive exon/exon skipping. In total, ONT sequencing detected 6,200 AS events in 20 samples, which were found in 2,728 genes. The frequencies of AS were significantly higher in cancer tissues than in normal tissues, indicating that AS-related variations play critical roles in regulating tumor progression (Figure 3(a)). Additionally, AS frequencies were significantly increased in MSI compared with MSS and revealed that AS might be the signature for dMMR (Figure 3(a)). Exon skipping was the most important AS event, accounting for nearly 60% of the total AS, and was significantly increased in cancer tissues (Figure 3(b)). However, alternative 5' splice sites were significantly decreased in cancer tissues (Figure 3(b)). In total, 932 genes with AS events were significantly differentially

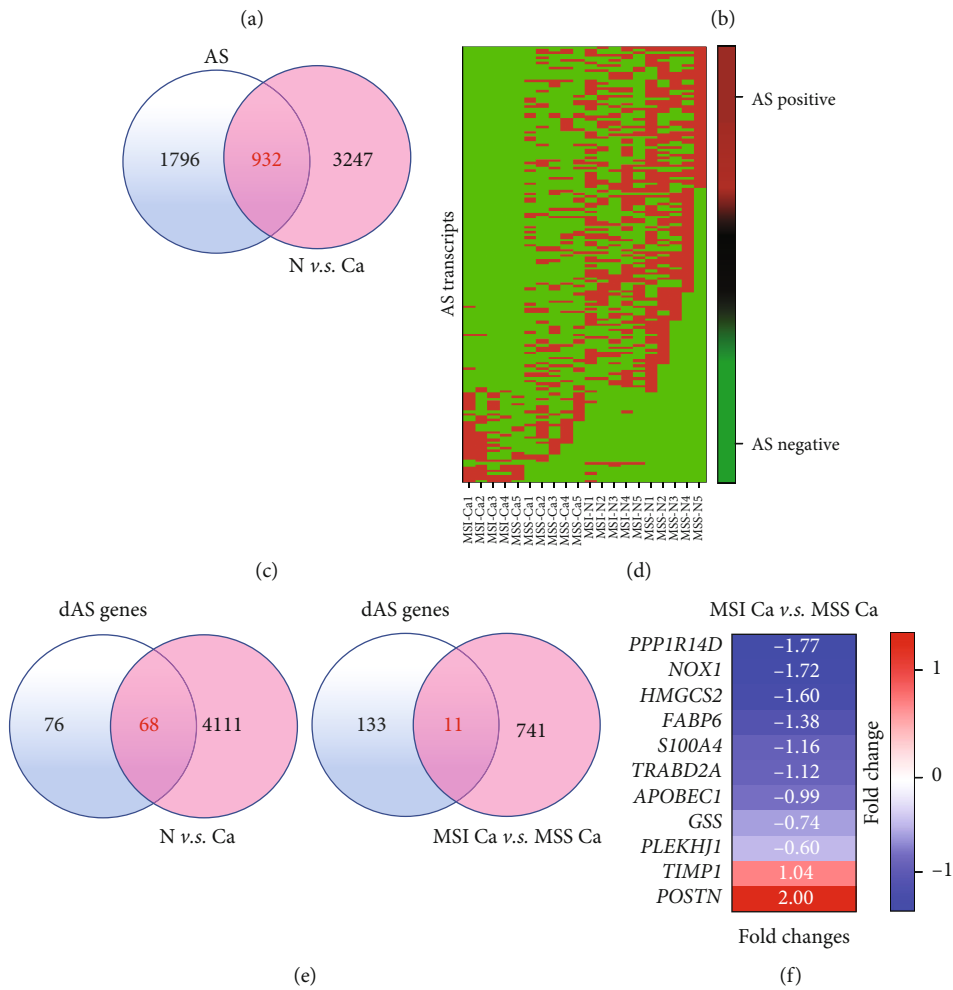
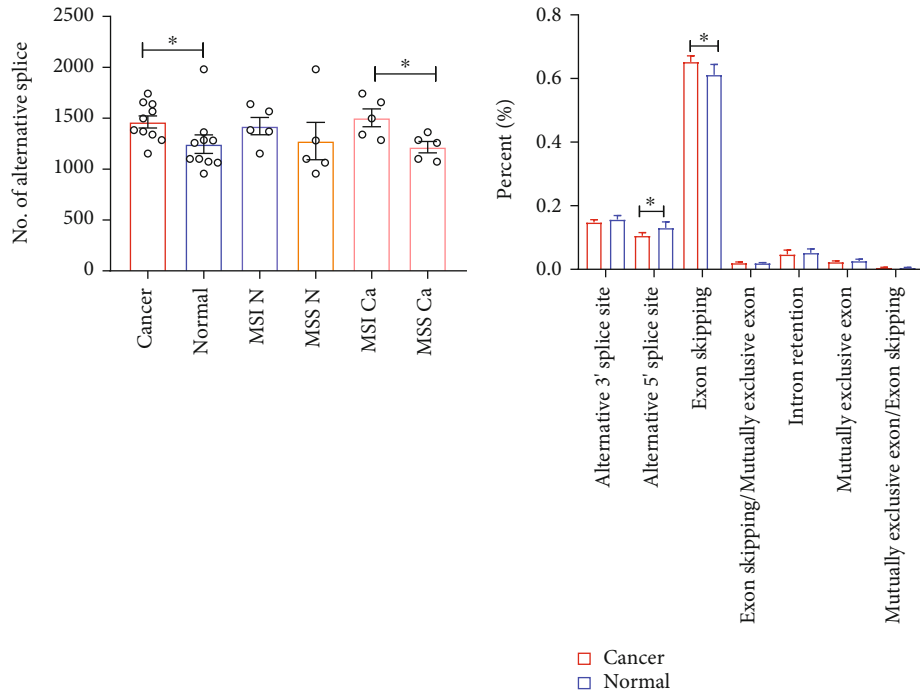


FIGURE 3: Continued.

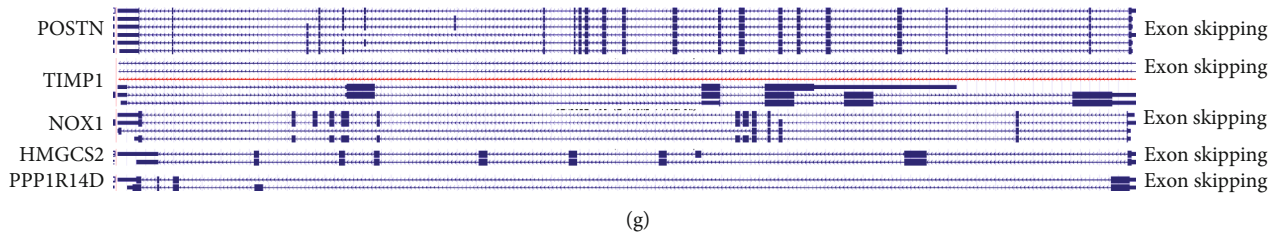


FIGURE 3: Alternative splicing (AS) is significantly associated with dMMR CRC. (a) The number (no.) of AS-related events was compared in cancer vs. normal controls. Normal tissue, MSI N vs. MSS N, and MSI Ca. vs. MSS Ca. (b) The percentages of seven AS type forms were compared between cancer and normal tissues. (c) Venn diagrams representing the overlapping genes that were differentially expressed and edited by AS. (d) Screening of AS-edited genes specifically associated with dMMR/MSI. Rows represent alternative splicing transcripts. E. Venn diagrams of dMMR/MSI-specifically related AS- (dAS-) edited genes as well as differentially expressed genes between cancer and normal tissues (left) and genes differentially upregulated or downregulated in MSI (right). (f) Eleven AS edited genes are listed with fold changes. (g) Graphical depiction of ENSEMBL transcripts from POSTN, TIMP1, NOX1, HMGCS2, and PPP1R14D. P values were obtained by nonparametric Mann-Whitney test (a, b) data are depicted by bar chart with error \pm standard deviation of mean (SEM). * $P < 0.05$.

expressed in cancer tissues (Figure 3(c)). To screen AS events specifically expressed in MSI, we selected AS events that occurred in three or more of these 20 samples, as a result, 170 MSI-related AS transcripts were produced which were transcribed by 144 genes (Figure 3(d)). Among 144 genes, 68 genes were differentially expressed when comparing their levels between cancer and normal tissues (Figure 3(e)). Eleven AS-edited genes had differential levels when compared their expression in MSI with their levels in MSS (Figures 3(e) and 3(f)). PPP1R14D, NOX1, and HMGCS2 were downregulated, while POSTN and TIMP1 were upregulated in MSI tissues (Figure 3(f)). According to UCSC analysis [31], both POSTN and TIMP1 had six transcripts and were transcribed through exon skipping, and NOX1 was encoded by four transcripts while transcribed by exon skipping. HMGCS2 and PPP1R14D were edited by exon skipping to transcribe two types of transcripts (Figure 3(g)).

3.4. Infiltrated Immune Cells in Tumor Microenvironment Were Identified by EPIC and CIBERSORT Analysis. Estimating the proportion of immune and cancer (EPIC) and CIBERSORT are designed to identify infiltrated immune cells and cancer cells from bulk RNA sequencing. Both tools are used to assess the proportion of immune cells and cancer cells in ONT sequencing data. EPIC identified 8 types of cells in 20 samples (Figure S4A), and immune cells, including CD8 T cells, NK cells, and B cells, were higher in normal tissues than in cancer tissues (Figure S4B). By contrast, endothelial cells and macrophages were decreased in cancer tissues (Figure S4B). Cancer-associated fibroblasts (CAFs) and NK cells were significantly increased in MSI (Figure S4C). CIBERSORT recognized 22 cell types, which included several subtypes of immune cells. Plasma cells accounted for the highest proportion in our ONT data (Figure S4D), and these cells were decreased in cancer tissues (Figure S4E). Additionally, activated mast cells, activated/resting dendritic cells, activated NK cells, and naïve B cells were increased, while M2 macrophages and resting CD4 memory cells were decreased in cancer tissues (Figure S4E). In CIBERSORT analysis, activated mast cells were abundant in MSI cancer (Figure S4F). EPIC and

CIBERSORT analysis indicated significantly differential immune cell patterns between cancer and normal tissues.

3.5. WGCNA Identifies MSI-Specific Clusters. Weighted gene coexpression network analysis (WGCNA) was performed across all the sequencing samples to address the specific clusters associated with MSI. In total, these genes can be divided into 17 clusters, which are indicated with different colors in Figure 4(a). Three clusters showed differential expression in MSI that might represent signatures. Cluster 1 included 246 genes; among them, 20 overlapped genes were simultaneously shown to be differentially expressed in MSI and cancer vs. normal tissues (Figure 4(b)). INHBA, MMP1, and SERPINE1 comprised a protein-protein interacting (PPI) network according to STRING network analysis, and these genes were significantly increased in cancer tissues or MSI cancer tissues (Figure 4(b)). In cluster 2, HMGCS2, FABP1, and SELENBP1 were significantly decreased in MSI according ONT sequencing result, consistent with an interacting network (Figure 4(c)). Cluster 3 included several upregulated genes in MSI and failed to enrich any PPI network (Figure 4(d)). We further addressed the overall survival (OS) relevant to the TCGA dataset using the GEPIA tool. The results demonstrated that the overexpression of INHBA, TNFAIP6, and TIMP1 was associated with a lower OS rate (Figure 4(e)). By contrast, increases in HMGCS2 and SELENBP1 were associated with a higher OS rate (Figure 4(e)).

3.6. Validation of MSI-Associated Genes. Our ONT sequencing combined tremendous information in CRC, particularly for MSI, allows identification of biomarkers to screen CRC or MSI candidates. As mentioned above, the third-generation sequencing technique allowed us to perform integrative analysis of AS and gene levels at the RNA level. Our GO enrichment revealed that RPL22L1, INHBA, and CAPZA1 had a significantly higher expression levels in MSI than in MSS, and these genes were also upregulated in CRC cancer tissues compared with their levels in matched normal tissues. RPL22L1 transcripts were edited by alternative 3' splicing and were identified in all 20 samples. HMGCS2 was downregulated in CRC cancer

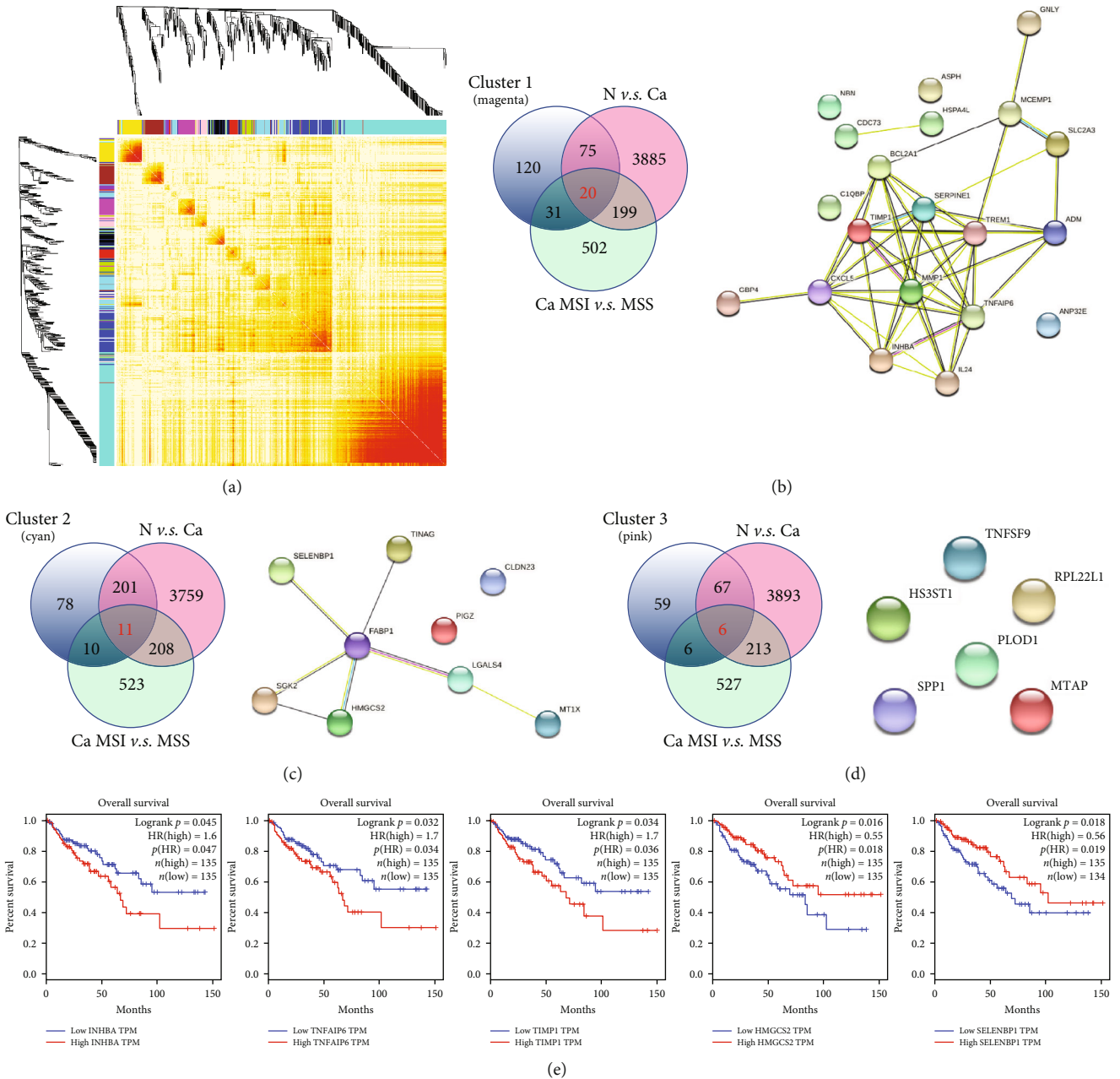


FIGURE 4: dMMR/MSS-related gene modules mined through weighted gene coexpression network analysis (WGCNA). (a) WGCNA analyzes ONT sequencing data. Gene clustering tree (dendrogram), module colors (main modules and submodules), and labels. (b) WGCNA identifies cluster 1, which contained 20 genes with differential levels either comparing cancer and normal samples or comparing dMMR/MSS and MSS samples (left). The protein-protein networks (PPIs) were investigated using the STRING tool (<https://www.string-db.org/>) (right). (c) Cluster 2 contained 11 genes (left) used to construct the PPI (right). (d) Cluster 3 included 6 genes (left) that showed no interactive relationship (right). (e) The clustered genes were associated with overall survival (OS) rates using TCGA COAD datasets according to GEPIA (<http://gepia.cancer-pku.cn/>).

tissue and showed lower levels in MSI than in MSS. Additionally, HMGCS2 is an AS transcript that is not expressed in MSI and serves as a signature for MSI. Importantly, RPL22L1, INHBA, and HMGCS2 were involved in the I-SMAD binding biological process (Figure 2(e)). Therefore, we selected these four genes to further explore their differential expression using RT-qPCR in 23 paired cancer and normal tissues. As expected, the mRNA levels of RPL22L1, INHBA, and CAPZA1 were

significantly higher in cancer than in normal tissues; conversely, HMGCS2 was significantly downregulated in cancer (Figure 5(a)). Next, we questioned whether these genes could be signatures for MSI. We compared their fold changes between MSI and MSS, and INHBA showed significantly increased levels in MSI (Figure 5(b)). In cancer tissues, RPL22L1 and INHBA were significantly upregulated in CRC with positive lymph nodes. In normal tissues, RPL22L1,

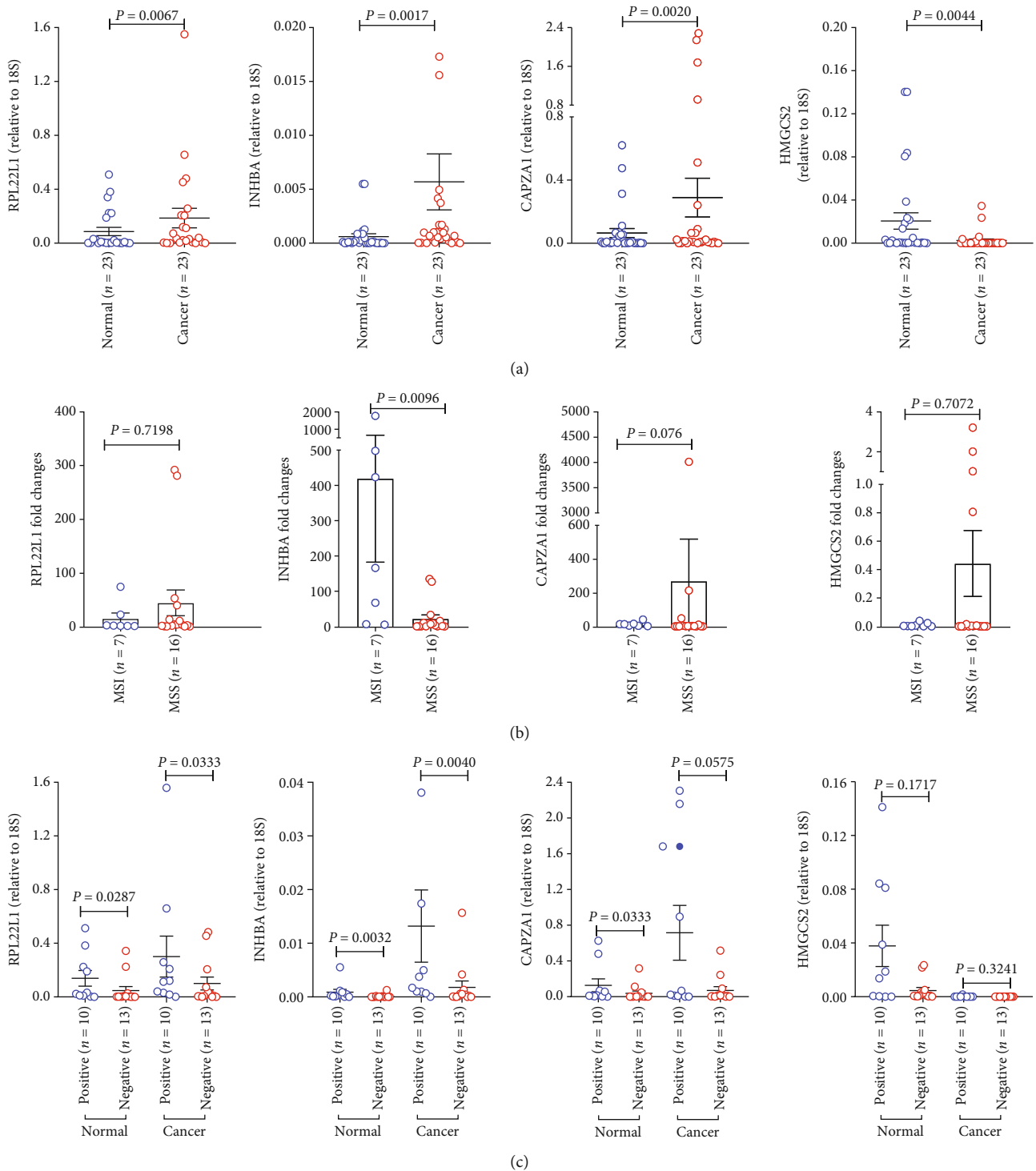


FIGURE 5: RT-qPCR assays validate the differential expression of selected genes. (a) RPL22L1, INHBA, and CAPZA1 are increased in cancer tissues; conversely, HMGCS2 is decreased. (b) The fold changes of INHBA are significantly upregulated in MSI compared with those in MSS. (c) RPL22L1 and INHBA expression are associated with tumor lymphatic metastasis. P values were obtained by paired Student's *t*-test (a) and nonparametric Mann-Whitney test (b, c); data are depicted by scatter plots (a, c) with mean \pm SEM or bar chart and scatter plots (b) with error \pm SEM.

INHBA, and CAPZA1 were found to be significantly increased in CRC patients for whom experienced tumor lymphatic metastasis (Figure 5(c)).

Next, we applied IHC to determine the protein levels of INHBA, RPL22L1, and CAPZA1. Forty-nine cancer tissues and 47 normal tissues were enrolled for IHC analysis. Their

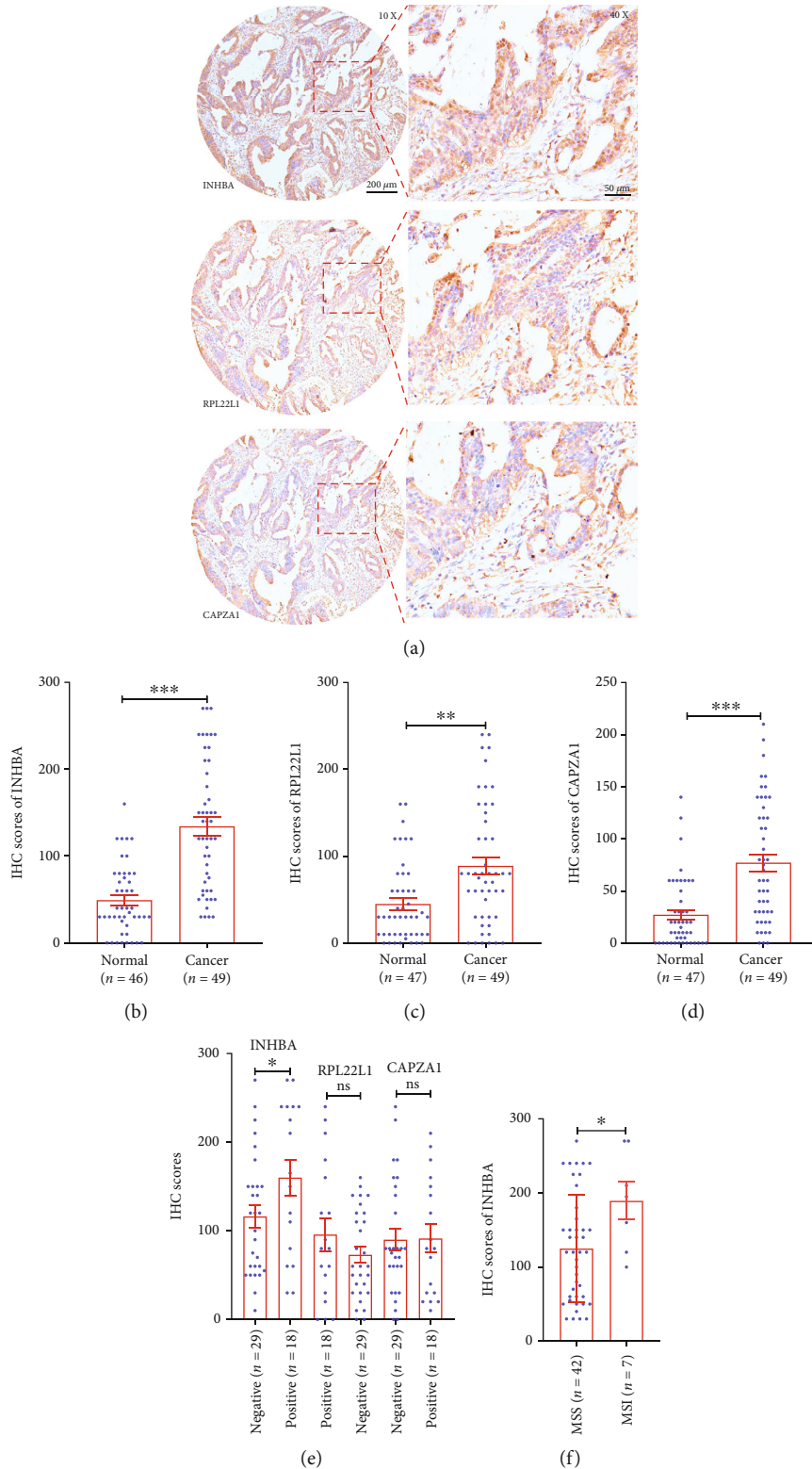


FIGURE 6: Immunohistochemistry (IHC) assays determine the overexpression of screened genes. (a) Representative IHC pictures show the expression of INHBA, RPL22L1, and CAPZA1 in tissue splices obtained from the same sample. (b) Comparison of the levels of INHBA validates its overexpression in cancer tissues. (c) RPL22L1 levels were compared in cancer and normal tissues. (d) CAPZA1 shows higher levels in cancer than in normal tissues. (e) The levels of INHBA, RPL22L1, and CAPZA1 were determined in lymph node-positive metastatic tumors and lymph node-negative metastatic tumors (two patients missing information). (f) INHBA expression was significantly increased in dMMR/MSI cancer. *P* values were obtained by nonparametric Mann–Whitney test (b–f), data are depicted by bar chart and scatter plots (b–f) with error \pm SEM. **P* < 0.05, ***P* < 0.01, and ****P* < 0.001; ns: not significant.

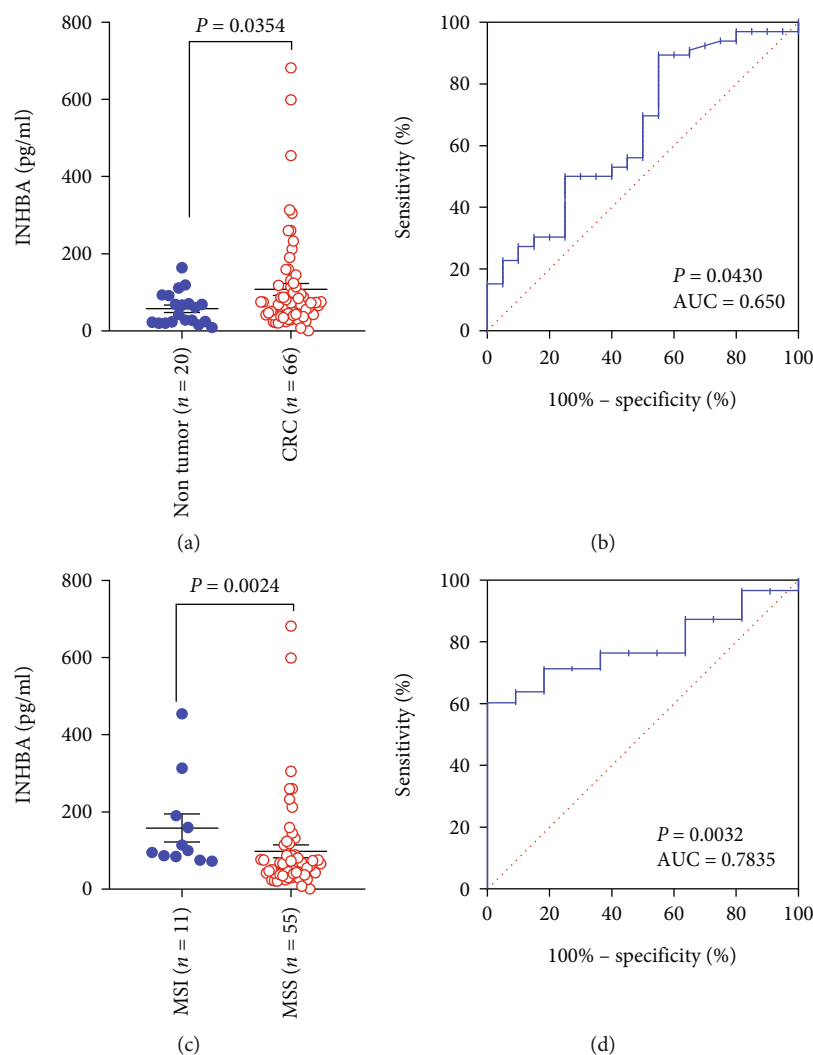


FIGURE 7: The detection of INHBA in blood plasma identifies a biomarker for CRC. (a) In plasma, the levels of INHBA were significantly increased in patients with CRC. (b) Receiver operating characteristic (ROC) curve analysis of the diagnostic efficacy. (c) INHBA levels in dMMR/MSI patients were higher than those in MSS patients. (d) INHBA is a biomarker for diagnosing dMMR/MSI candidates. P values were obtained by nonparametric Mann–Whitney test (a, c); ROC curves were plotted using GraphPad Prism software to calculate P and area under the curve (AUC) (b, d). Data are depicted by scatter plots (a, c) with mean value \pm SEM.

levels were scored using semiquantitative analysis, demonstrating that these three genes were significantly overexpressed in cancer tissues (Figures 6(a)–6(d)). IHC assays were consistent with our RT-qPCR analyses. We also addressed clinicopathological relevance and found that INHBA was significantly increased in patients diagnosed with lymph node metastasis tumors (Figure 6(e)). INHBA shown higher levels in male CRC patients than their levels in female (Table S2). Other clinicopathological factors included tumor size, differentiation, TNM stage, and vascular and nerve invasion; INHBA levels did not show statistical differences (Tables S2). We inferred based on both mRNA and protein that INHBA expression can be used as a biomarker to predict tumor lymphatic metastasis. Seven patients who were enrolled in the IHC analysis were diagnosed with MSI. We compared the expression of

INHBA in these patients and found significantly higher levels in patients with MSI than in MSS patients (Figure 6(f)).

3.7. INHBA Is a Signature for MSI Diagnosis in Blood Plasma.

Because INHBA is a protein secreted into the extracellular space by tumor cells, we applied ELISA to determine its plasma levels. Our analysis enrolled 20 plasma samples from patients diagnosed with nontumor benign disease and 66 plasma samples from CRC patients. As expected, the CRC plasma levels of INHBA were significantly increased compared with those of samples from nontumor benign disease patients (Figure 7(a)). Receiver operating characteristic (ROC) curve analysis was performed to examine the diagnostic efficiency when INHBA was recognized as a biomarker. Area under the curve (AUC) analysis indicated that secreted INHBA was a

promising biomarker that could be used in clinical practice (AUC = 0.650; $P = 0.0430$) (Figure 7(b)). We further examined whether secreted INHBA served as a marker to discern dMMR/MSI CRCs. As expected, the levels of INHBA in dMMR/MSI patients were significantly higher than those in patients with MSS (Figure 7(c)). ROC curve analysis proved that INHBA could diagnose dMMR/MSI CRCs (AUC = 0.7835; $P = 0.0032$) (Figure 7(d)).

4. Discussion

In the present study, we constructed CRC mRNA profiles using ONT long read sequencing, which has a unique advantage in identifying the AS of mRNA transcripts. dMMR/MSI tumors exhibit a high frequency of AS, which is associated with a large burden of destabilizing mutations [32]. Utilizing integrative bioinformatic analysis, we demonstrated that immune chemotactic genes are the most differentially expressed in cancer tissues and mediate the recruitment of immune cells [33]. Both B and T cells are the major class of immune cells in the adaptive immune system and have multiple functions in facilitating the coevolution of cancer and its microenvironment [34]. Our ONT data consistently indicated that these cells are decreased in cancer tissues. Innate immune cells, including NK cells, macrophages, and mast cells, were dysregulated and showed differential levels in MSI-related CRC. These findings are associated with the characteristics of innate cells, which infiltrate the microenvironment and display tumor phagocytosis or promote tumor growth [35].

Despite extensive investigation of genomic mutations and many genetic drivers, few studies have focused on AS in dMMR/MSI [36]. Here, despite using a small cohort, our ONT sequencing determined thousands of differentially expressed genes and AS events, which showed a higher efficacy than next-generation sequencing. Tumor mutational burdens predict the response efficiency to ICB treatment in dMMR/MSI tumors [37]. In our study, AS of mRNA transcripts was significantly correlated with dMMR/MSI and we screened several potential biomarkers to validate in CRC tissues or plasma. INHBA, RPL22L1, CAPZA1, and HMGCS2 showed significant differential levels in dMMR/MSI cancer tissues and were finally selected for further detection. RPL22L1, a ribosomal protein, is overexpressed in ovarian cancer and promotes cancer cell metastasis via the epithelial-to-mesenchymal transition [38]. Consistently, we validated its overexpression in CRC and found that it was associated with dMMR/MSI or tumor lymphatic metastasis. CAPZA1 regulates actin polymerization, which controls cell motility by binding to the barbed ends of actin filaments [39]. CAPZA1 is upregulated in hepatocellular carcinoma [40] and gastric cancer [39]. HMGCS2 is downregulated in CRC, serving as a critical downstream target of SLC38A4 to regulate tumor cell metabolism [41].

INHBA is a susceptibility gene for pancreatic cancer [42] that is involved in the TGF- β signaling pathway and was edited by AS in our ONT analysis. INHBA is associated with the CRC OS rate and plays a role in promoting tumor cell

proliferation and migration [43]. We demonstrated that INHBA is significantly increased in CRC cancer cells and, more importantly, is a potential biomarker for discerning dMMR/MSI candidates.

The most important concept conveyed in this study is a comprehensive investigation of mRNA profiles using third-generation sequencing, which allows us to explore AS events related to CRC dMMR/MSI for the first time. These newly constructed ONT data identified multiple genes that could be applied as markers associated with CRC progression. We surveyed their potential clinical applications in small sample sizes, such as INHBA, a promising dMMR/MSI biomarker. Two major limitations of our study are its retrospective nature and single-center design. Further study is required to address the clinical relevance of INHBA or other identified genes as markers for MSI-H/dMMR solid tumors in a larger cohort.

Data Availability

The data were submitted to the NCBI SRA database and are accessible through the citation accession number PRJNA758110.

Ethical Approval

The present study was approved by institutional review board at Beijing Chao-Yang Hospital.

Consent

All the enrolled subjects signed written informed consent for the use of their clinical specimens.

Conflicts of Interest

The authors declare no competing interests.

Authors' Contributions

Lei Yang and Zhenjun Wang are assigned to the conceptualization. Lei Yang, Hao Qu, and Chunxiang Ye are responsible for the methodology. Lei Yang, Hao Qu, Baocheng Zhao, Yudong Zhang, and Yaohan Xue did the investigation. Lei Yang is assigned to the writing—review and editing. Lei Yang acquired funding. Lei Yang is assigned to the supervision.

Acknowledgments

This project was supported by grants from the National Natural Science Foundation of China (81802349 and 81802353), Beijing Natural Science Foundation (7192070), Beijing Municipal Administration of Hospitals Incubating Program (PX2018013), Scientific Research Project of Beijing Educational Committee (KM201910025016), and Open Project of Key Laboratory of Cardiovascular Disease Medical Engineering, Ministry of Education (2019XXG-KFKT-03).

Supplementary Materials

Figure S1: a schematic diagram illustrates how samples were selected and ONT sequencing carried out. Figure S2: hematoxylin-eosin staining shows normal tissue, MSS cancer tissue, and MSI cancer tissue. Figure S3: TCGA COAD and READ analysis validate the identified differential genes using ONT sequencing. A. Volcano plots showing the upregulated or downregulated genes in the COAD (left) and READ (right) datasets. B. Venn diagram determines the overlapping dysregulated genes using ONT sequencing detection as well as validation in COAD and READ. C. GO enrichment enriched BP terms for COAD (left) and READ (right). Figure S4: cell types in the tumor microenvironment (TME) are analyzed using EPIC and CIBERSORT tools. A. EPIC analysis shows the cell fractions. B. Comparison of the distributions per cell type in cancer and normal tissues was performed by EPIC. C. dMMR/MSI tumor has higher CAF and NK cells. D. Cell compositions in the TME were calculated using the CIBERSORT method. E. Cell compositions analyzed by CIBERSORT were compared between cancer and normal TMEs. F. The compositions of activated mast cells are increased in MSI. *P* values were obtained by nonparametric Mann–Whitney test (B, C, E); data are depicted by box and scatter plots (B, C, E) with whiskers (Min–Max). **P* < 0.05, ***P* < 0.01, and ****P* < 0.001; ns: not significant. (*Supplementary Materials*)

References

- [1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, “Cancer statistics, 2021,” *CA: a Cancer Journal for Clinicians*, vol. 71, no. 1, pp. 7–33, 2021.
- [2] T. Andre, M. Amonkar, J. M. Norquist et al., “Health-related quality of life in patients with microsatellite instability-high or mismatch repair deficient metastatic colorectal cancer treated with first-line pembrolizumab versus chemotherapy (KEYNOTE-177): an open-label, randomised, phase 3 trial,” *The Lancet Oncology*, vol. 22, no. 5, pp. 665–677, 2021.
- [3] C. Lu, J. Guan, S. Lu et al., “DNA sensing in mismatch repair-deficient tumor cells is essential for anti-tumor immunity,” *Cancer Cell*, vol. 39, no. 1, pp. 96–108.e6, 2021.
- [4] G. Picco, C. M. Cattaneo, E. J. van Vliet et al., “Werner helicase is a synthetic-lethal vulnerability in mismatch repair-deficient colorectal cancer refractory to targeted therapies, chemotherapy, and immunotherapy,” *Cancer Discovery*, vol. 11, no. 8, pp. 1923–1937, 2021.
- [5] Z. Cai, Z. Wang, C. Liu et al., “Detection of microsatellite instability from circulating tumor DNA by targeted deep sequencing,” *The Journal of Molecular Diagnostics*, vol. 22, no. 7, pp. 860–870, 2020.
- [6] F. Innocenti, F. S. Ou, X. Qu 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,” *Journal of Clinical Oncology*, vol. 37, no. 14, pp. 1217–1227, 2019.
- [7] T. André, K.-K. Shiu, T. W. Kim et al., “Pembrolizumab in microsatellite-Instability-high advanced colorectal cancer,” *The New England Journal of Medicine*, vol. 383, no. 23, pp. 2207–2218, 2020.
- [8] D. Gao, E. Morini, M. Salani et al., “A deep learning approach to identify gene targets of a therapeutic for human splicing disorders,” *Nature Communications*, vol. 12, no. 1, p. 3332, 2021.
- [9] L. Wang, K. Shamardani, H. Babikir et al., “The evolution of alternative splicing in glioblastoma under therapy,” *Genome Biology*, vol. 22, no. 1, p. 48, 2021.
- [10] D. J. McGrail, J. Garnett, J. Yin et al., “Proteome instability is a therapeutic vulnerability in mismatch repair-deficient cancer,” *Cancer Cell*, vol. 37, no. 3, pp. 371–386.e12, 2020.
- [11] J. A. Nowak, M. B. Yurgelun, J. L. Bruce et al., “Detection of mismatch repair deficiency and microsatellite instability in colorectal adenocarcinoma by targeted next-generation sequencing,” *The Journal of Molecular Diagnostics*, vol. 19, no. 1, pp. 84–91, 2017.
- [12] J. Liang, Z. Hong, B. Sun, Z. Guo, C. Wang, and J. Zhu, “The alternatively spliced isoforms of key molecules in the cGAS-STING signaling pathway,” *Frontiers in Immunology*, vol. 12, article 771744, 2021.
- [13] J. B. Fan, S. Miyauchi, H. Z. Xu et al., “Type I interferon regulates a coordinated gene network to enhance cytotoxic T cell-mediated tumor killing,” *Cancer Discovery*, vol. 10, no. 3, pp. 382–393, 2020.
- [14] M. Tolomeo and A. Cascio, “The multifaced role of STAT3 in cancer and its implication for anticancer therapy,” *International Journal of Molecular Sciences*, vol. 22, no. 2, p. 603, 2021.
- [15] C. Mowat, S. R. Mosley, A. Namdar, D. Schiller, and K. Baker, “Anti-tumor immunity in mismatch repair-deficient colorectal cancers requires type I IFN-driven CCL5 and CXCL10,” *The Journal of Experimental Medicine*, vol. 218, no. 9, article e20210108, 2021.
- [16] Z. Shi, J. Shen, J. Qiu, Q. Zhao, K. Hua, and H. Wang, “CXCL10 potentiates immune checkpoint blockade therapy in homologous recombination-deficient tumors,” *Theranostics*, vol. 11, no. 15, pp. 7175–7187, 2021.
- [17] A. J. Gunderson, T. Yamazaki, K. McCarty et al., “TGFβ suppresses CD8+ T cell expression of CXCR3 and tumor trafficking,” *Nature Communications*, vol. 11, no. 1, p. 1749, 2020.
- [18] R. Derynck, S. J. Turley, and R. J. Akhurst, “TGFβ biology in cancer progression and immunotherapy,” *Nature Reviews Clinical Oncology*, vol. 18, no. 1, pp. 9–34, 2021.
- [19] A. D. Tang, C. M. Soulette, M. J. van Baren et al., “Full-length transcript characterization of *SF3B1* mutation in chronic lymphocytic leukemia reveals downregulation of retained introns,” *Nature Communications*, vol. 11, no. 1, p. 1438, 2020.
- [20] H. Li, “Minimap2: pairwise alignment for nucleotide sequences,” *Bioinformatics*, vol. 34, no. 18, pp. 3094–3100, 2018.
- [21] G. Pertea and M. Pertea, “GFF Utilities: GffRead and GffCompare,” *F1000Research*, vol. 9, 2020.
- [22] S. Foissac and M. Sammeth, “ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets,” *Nucleic Acids Research*, vol. 35, Web Server, pp. W297–W299, 2007.
- [23] S. Beier, T. Thiel, T. Münch, U. Scholz, and M. Mascher, “MISA-web: a web server for microsatellite prediction,” *Bioinformatics*, vol. 33, no. 16, pp. 2583–2585, 2017.
- [24] S. E. Abdel-Ghany, M. Hamilton, J. L. Jacobi et al., “A survey of the sorghum transcriptome using single-molecule long reads,” *Nature Communications*, vol. 7, no. 1, article 11706, 2016.
- [25] B. J. Haas, A. Papanicolaou, M. Yassour et al., “De novo transcript sequence reconstruction from RNA-seq using the

- Trinity platform for reference generation and analysis,” *Nature Protocols*, vol. 8, no. 8, pp. 1494–1512, 2013.
- [26] D. W. Huang, B. T. Sherman, and R. A. Lempicki, “Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists,” *Nucleic Acids Research*, vol. 37, no. 1, pp. 1–13, 2009.
- [27] D. J. Nolan, M. Ginsberg, E. Israely et al., “Molecular signatures of tissue-specific microvascular endothelial cell heterogeneity in organ maintenance and regeneration,” *Developmental Cell*, vol. 26, no. 2, pp. 204–219, 2013.
- [28] M. E. Ritchie, B. Phipson, D. Wu et al., “Limma powers differential expression analyses for RNA-sequencing and microarray studies,” *Nucleic Acids Research*, vol. 43, no. 7, article e47, 2015.
- [29] A. E. Teschendorff, C. E. Breeze, S. C. Zheng, and S. Beck, “A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies,” *BMC Bioinformatics*, vol. 18, no. 1, p. 105, 2017.
- [30] H. Li, “Minimap and minimap: fast mapping and de novo assembly for noisy long sequences,” *Bioinformatics*, vol. 32, no. 14, pp. 2103–2110, 2016.
- [31] B. J. Raney, T. R. Dreszer, G. P. Barber et al., “Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser,” *Bioinformatics*, vol. 30, no. 7, pp. 1003–1005, 2014.
- [32] C. Antoniotti, W. M. Korn, F. Marmorino et al., “Tumour mutational burden, microsatellite instability, and actionable alterations in metastatic colorectal cancer: next-generation sequencing results of TRIBE2 study,” *European Journal of Cancer*, vol. 155, pp. 73–84, 2021.
- [33] V. Chew, Y. H. Lee, L. Pan et al., “Immune activation underlies a sustained clinical response to Yttrium-90 radioembolisation in hepatocellular carcinoma,” *Gut*, vol. 68, no. 2, pp. 335–346, 2019.
- [34] C. Hu, C. Chu, L. Liu et al., “Dissecting the microenvironment around biosynthetic scaffolds in murine skin wound healing,” *Science Advances*, vol. 7, no. 22, 2021.
- [35] K. Ganesh, Z. K. Stadler, A. Cercek et al., “Immunotherapy in colorectal cancer: rationale, challenges and potential,” *Nature Reviews Gastroenterology & Hepatology*, vol. 16, no. 6, pp. 361–375, 2019.
- [36] M. Chalabi, L. F. Fanchi, K. K. Dijkstra et al., “Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers,” *Nature Medicine*, vol. 26, no. 4, pp. 566–576, 2020.
- [37] K. Chida, A. Kawazoe, M. Kawazu et al., “A low tumor mutational burden and PTEN mutations are predictors of a negative response to PD-1 blockade in MSI-H/dMMR gastrointestinal tumors,” *Clinical Cancer Research*, vol. 27, no. 13, pp. 3714–3724, 2021.
- [38] G. Kan, Z. Wang, C. Sheng et al., “Dual inhibition of DKC1 and MEK1/2 synergistically restrains the growth of colorectal cancer cells,” *Advanced Science*, vol. 8, no. 10, p. 2004344, 2021.
- [39] H. Tsugawa, C. Kato, H. Mori et al., “Cancer stem-cell marker CD44v9-positive cells arise from *Helicobacter pylori*-infected CAPZA1-overexpressing cells,” *Cellular and Molecular Gastroenterology and Hepatology*, vol. 8, no. 3, pp. 319–334, 2019.
- [40] D. Huang, L. Cao, J.-x. S. Le Xiao, Y.-j. Zhang, P. Zheng, and S.-g. Zheng, “Hypoxia induces actin cytoskeleton remodeling by regulating the binding of CAPZA1 to F-actin via PIP2 to drive EMT in hepatocellular carcinoma,” *Cancer Letters*, vol. 448, pp. 117–127, 2019.
- [41] J. Li, M.-h. Li, T.-t. Wang et al., “SLC38A4 functions as a tumour suppressor in hepatocellular carcinoma through modulating Wnt/ β -catenin/MYC/HMGCS2 axis,” *British Journal of Cancer*, vol. 125, no. 6, pp. 7465–7476, 2021.
- [42] J. Zhong, A. Jermusyk, L. Wu et al., “A transcriptome-wide association study identifies novel candidate susceptibility genes for pancreatic cancer,” *Journal of the National Cancer Institute*, vol. 112, no. 10, pp. 1003–1012, 2020.
- [43] X. Yang, J. Yan, Y. Jiang, and Y. Wang, “An immune-related model based on INHBA, JAG2 and CCL19 to predict the prognoses of colon cancer patients,” *Cancer Cell International*, vol. 21, no. 1, p. 299, 2021.